Compressive Sensing With Side Information: How to Optimally Capture This Extra Information for GMM Signals?

Meng-Yang Chen, *Student Member, IEEE*, Francesco Renna, *Member, IEEE*, and Miguel R. D. Rodrigues, *Senior Member, IEEE*

Abstract—This paper studies how to optimally capture side information to aid in the reconstruction of high-dimensional signals from low-dimensional random linear and noisy measurements, by assuming that both the signal of interest and the side information signal are drawn from a joint Gaussian mixture model. In particular, we derive sufficient and (occasionally) necessary conditions on the number of linear measurements for the signal reconstruction minimum mean squared error (MMSE) to approach zero in the low-noise regime; moreover, we also derive closed-form linear side information measurement designs for the reconstruction MMSE to approach zero in the low-noise regime. Our designs suggest that a linear projection kernel that optimally captures side information is such that it measures the attributes of side information that are maximally correlated with the signal of interest. A number of experiments both with synthetic and real data confirm that our theoretical results are well aligned with numerical ones. Finally, we offer a case study associated with a panchromatic sharpening (pan sharpening) application in the presence of compressive hyperspectral data that demonstrates that our proposed linear side information measurement designs can lead to reconstruction peak signal-to-noise ratio (PSNR) gains in excess of 2 dB over other approaches in this practical application.

Index Terms—Compressive sensing, compressive sensing with side information, random projection kernels, linear projection kernel design, Gaussian mixture models (GMM), pan-sharpening.

I. INTRODUCTION

COMPRESSIVE sensing (CS) is a promising technology that enables the successful recovery of high-dimensional signals from low-dimensional linear measurements. In particular, by assuming that the signals admit a parsimonious (sparse) representation in some domain [2]–[6], the reconstruction of the signal of interest from the signal measurements can be performed with tractable ℓ_1-norm minimization methods [2], [4], [6], or with iterative algorithms [7], [8].

Other signal models beyond sparsity that can also be leveraged to enable the reconstruction of a high-dimensional signal from low-dimensional linear measurements include wavelet trees [9], [10], manifolds [11], [12], union-of-subspaces models [10], [13]–[15], or statistical models [16], [17]. A widely used statistical model is the Gaussian mixture model (GMM) that can be seen as the Bayesian counterpart of (linear or affine) union-of-subspaces models, where each subspace corresponds to the image of the (possibly low-rank) covariance matrix associated to each Gaussian component in the mixture. In fact, low-rank GMM priors can also be shown to approximate well compact manifolds [11], and they were shown to provide state-of-the-art results in practical problems in dictionary learning [11], image processing [18], [19], image classification [20], and compressive video acquisition [21]. In particular, GMM priors with a moderate number of classes ($\approx$10 to 20) have been shown to reliably model real-world data, such as, for example, small portions (e.g., 8×8 pixels) of natural images or video frames [21], [22]. Given a set of training data, such GMM priors can be estimated by using different learning methods such as, for example, the well-known expectation maximization (EM) algorithm [23] or the non-parametric approach described in [11]. Another advantage of GMM models over other structured models relates to the fact that the reconstruction of signals drawn from a GMM distribution can be very effectively performed via an optimal conditional mean estimator available in closed-form [11].

The use of side information – in the form of other signals that may be related to the signal of interest – has also been recently used to improve the performance of CS systems. For example, recent advances in compressive sensing have proposed different techniques to effectively incorporate side information into signal reconstruction algorithms [24]–[31]. In particular, [24] studies the case when side information is available at the decoder in the form of partial information about the signal support; and [25] the case when side information is given as an additional noisy version of the signal of interest.

Furthermore, [28], [29] provide a theoretical framework to determine sufficient conditions for perfect reconstruction with ℓ_1/ℓ_1 and mixed ℓ_1/ℓ_2 reconstruction strategies in the presence of side information. Reference [30] shows how reference signals can be used to enhance reconstruction performance via a reweighted ℓ_1/ℓ_1 minimization scheme that uses the sample complexity bound to update its weights. Other related works, including [32]–[34], provide sufficient and (occasionally) necessary conditions on the number of random measurements taken from the signal of interest and the side information signal to guarantee that the reconstruction error tends to zero in the low-noise regime. In particular, [33] introduces a GMM model that generalizes the joint sparsity models JSM-1 and JSM-3 proposed for distributed compressive sensing (DCS) in [35] and [36].[1] [34] studies the impact of projection kernel design associated with the signal of interest on the reconstruction performance in the presence of side information. Another advance in compressive sensing relates to the use of side information at the encoder [37].

The use of side information in CS has also led to significant results in medical imaging. For example, [26] and [27] use information from former scans of the same patient, taken from magnetic resonance images and dynamic tomographic images, respectively, and prove enhanced reconstruction performance with a smaller number of measurements than traditional CS approaches. The use of side information has also led to gains in various other applications. For example, [31] considers RGB images as side information to recover high-resolution hyperspectral images from compressive measurements. In turn, [38] shows how side information can be used in image separation problems arising in art-investigation applications.

The availability of side information in compressive sensing systems also leads naturally to the question:

How can we optimally capture side information in compressive sensing systems in order to aid in the reconstruction of the signal of interest from the signal measurements?

This question is also motivated by the observation that, although some devices might impose a particular structure to the acquired measurements (e.g., magnetic resonance imaging (MRI), computed tomography (CT), radio interferometry (RI), seismic tonography, etc.), there exist applications where it is possible to design the measurement matrices, at least to a certain extent. Some of these applications include compressive imaging, fluorescence microscopy, compressive hyperspectral imaging, and all compressive video and imaging devices based on coded aperture snapshot architectures (e.g., coded aperture snapshot spectral imaging (CASSI) [39], coded aperture compressive temporal imaging (CACTI) [40], etc.). In fact, the design of linear measurements for CS systems has been shown to lead to reconstruction gains in various applications [19], [41]–[43]; however, the design of linear measurements for CS systems in the presence of side information is less understood (note that [34] is only applicable to the scenario where we attempt to

optimally capture the signal of interest rather than the side information).

This paper answers this question by considering a scenario where the signal reconstruction device has access to noisy (random) linear measurements of the signal of interest and noisy (optimally designed) linear measurements of the side information. By postulating that both the signal of interest and the side information admit a joint GMM model, our contributions are:

- Sufficient and – where possible – necessary conditions on the number of optimal linear side information measurements for the reconstruction minimum mean-square error (MMSE) to approach zero in the low-noise regime.
- Concrete optimal or nearly optimal linear measurement designs for the reconstruction MMSE to approach zero in the low-noise regime.
- A case study associated with a pan-sharpening application involving the reconstruction of a high-resolution RGB image from a high-resolution panchromatic (gray scale) one in the presence of low-resolution compressive hyperspectral measurements acquired with a CASSI camera. This case study showcases that our proposed measurement designs have the potential to considerably improve performance in relation to random measurement strategies; moreover, this case study also showcases that our proposed modeling approach based on GMM models has advantages over other state-of-the-art modeling approaches.

Finally, this work differs from our previous work in [1], [33], [34] as follows: (a) [33] analyses the performance of compressive sensing systems with side information with random measurement strategies whereas this paper analyses the performance of these systems with designed measurement strategies; (b) [34] considers how to optimally capture the signal of interest in compressive sensing systems with side information whereas this paper considers instead how to optimally capture the side information in these systems; and finally (c) this work augments the work in [1] in various ways by providing additional insights about the behavior of the reconstruction error, additional simulation results, a case-study, and detailed proofs.

A. Organization

The paper is organized as follows: Section II defines the signals and system models. Section III determines sufficient conditions and (occasionally) necessary conditions on the number of measurements required for the reconstruction MMSE to approach zero in the low-noise regime, when the measurement kernel associated with the signal of interest is random and the measurement kernel associated with the side information is designed, as well as closed-form measurement strategies. Section IV presents numerical results with synthetic data drawn from both Gaussian and GMM distributions and with real imaging data that illustrate the validity of the theoretical derivations. An application example associated with the reconstruction of high-resolution RGB images from gray scale images using low-resolution, compressive, hyperspectral measurements as side information is presented in Section V. Conclusions are drawn

---

[1]Another difference between this work and the work in [33] relates to the fact that [33] focuses on random measurement strategies whereas our work focuses on designed ones.
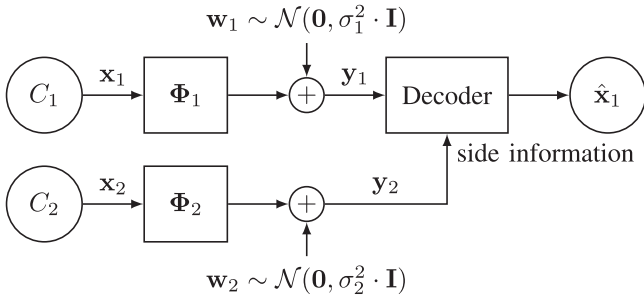
Fig. 1. Compressive sensing model in the presence of side information.

in Section VI. The proofs of all theorems are reported in the Appendices.

B. Notation

We adopt the following notation: boldface upper-case letters denote matrices ($\mathbf{X}$) and boldface lower-case letters denote column vectors ($\mathbf{x}$). 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). $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\dagger}$, $\mathrm{rank}(\cdot)$, and $\mathbb{E}[\cdot]$ represent the transpose, Moore-Penrose pseude-inverse, rank, and expectation operators, respectively. The Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. $\mathrm{Im}(\cdot)$, $\mathrm{Null}(\cdot)$, and $\dim(\cdot)$ denote the (column) image of a matrix, null space of a matrix, and the dimension of a linear space, respectively. $\| \cdot \|_F$ denotes the Frobenius norm of a matrix.

II. SYSTEM MODEL

We consider a setting where one aims to reconstruct a high-dimensional signal from noisy, linear, compressive measurements in the presence of side information, as depicted in Fig. 1. In particular, we wish to reconstruct the signal of interest $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ from measurements $\mathbf{y}_1 \in \mathbb{R}^{m_1}$, with $n_1 \geq m_1$, given by

$$\mathbf{y}_1 = \boldsymbol{\Phi}_1 \mathbf{x}_1 + \mathbf{w}_1, \qquad (1)$$

as well as noisy, linear, compressive measurements $\mathbf{y}_2 \in \mathbb{R}^{m_2}$, associated with another signal $\mathbf{x}_2 \in \mathbb{R}^{n_2}$, with $n_2 \geq m_2$, given by

$$\mathbf{y}_2 = \boldsymbol{\Phi}_2 \mathbf{x}_2 + \mathbf{w}_2. \qquad (2)$$

The matrices $\boldsymbol{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ and $\boldsymbol{\Phi}_2 \in \mathbb{R}^{m_2 \times n_2}$ represent linear projection kernels and $\mathbf{w}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \cdot \sigma_1^2)$ and $\mathbf{w}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \cdot \sigma_2^2)$ are additive white Gaussian noise (AWGN) vectors that model possible distortion introduced by the sensing process and noise.

We will be assuming that the signal of interest $\mathbf{x}_1$ and the side information signal $\mathbf{x}_2$ – which are correlated – are described by a joint Gaussian mixture model. In particular, $\mathbf{x}_1$ is associated with underlying class labels $C_1 \in \{1, \dots, K_1\}$, while $\mathbf{x}_2$ is associated with underlying class labels $C_2 \in \{1, \dots, K_2\}$. The class labels obey the joint probability mass function (pmf) $P_{C_1, C_2}(i, k)$, and the signals $\mathbf{x}_1$ and $\mathbf{x}_2$, conditioned on the un-

derlying class labels $C_1 = i$ and $C_2 = k$, obey the joint Gaussian probability density function (pdf):

$$p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}, \boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}), \qquad (3)$$

and so $\mathbf{x}_1$ and $\mathbf{x}_2$ follow the GMM given by

$$p(\mathbf{x}_1, \mathbf{x}_2) = \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} P_{C_1, C_2}(i, k) p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k)$$

$$= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} P_{C_1, C_2}(i, k) \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}, \boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}), \qquad (4)$$

where

$$\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)} \\ \boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)} & \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} \\ \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} & \boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}. \qquad (5)$$

Here, $\boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)}$ are the mean and covariance matrix of $\mathbf{x}_1$, conditioned on class labels $C_1 = i, C_2 = k$, and likewise $\boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ are the mean and covariance matrix of $\mathbf{x}_2$, conditioned on class labels $C_1 = i, C_2 = k$, respectively. The cross-covariance between $\mathbf{x}_1$ and $\mathbf{x}_2$ given the class labels $C_1 = i$ and $C_2 = k$ is given by $\boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}$.

We will also be assuming that the linear spaces associated with the images of the covariance matrices $\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}$, $\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)}$, and $\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ for different class labels $C_1 = i$ and $C_2 = k$ are independently drawn at random from a continuous pdf over the corresponding Grassmann manifold[2] with $r_{\mathbf{x}}^{(i,k)} = \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}) \leq n_1 + n_2$, $r_{\mathbf{x}_1}^{(i,k)} = \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)}) \leq n_1$, and $r_{\mathbf{x}_2}^{(i,k)} = \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}) \leq n_2$. Note that, conditioned on class labels $C_1 = i, C_2 = k$, such ranks can also act as proxy to measure the correlation between $\mathbf{x}_1$ and $\mathbf{x}_2$. In fact, for all (i, k), the overall rank $r_{\mathbf{x}}^{(i,k)}$ is upper bounded by $r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)}$ and lower bounded by $\max\{r_{\mathbf{x}_1}^{(i,k)}, r_{\mathbf{x}_2}^{(i,k)}\}$; the upper bound is attained by uncorrelated signals (conditioned on class labels $C_1 = i, C_2 = k$) and the lower bound is attained by perfectly correlated ones (conditioned on class labels $C_1 = i, C_2 = k$).

We also assume that the decoder has knowledge about the joint pdf $p(\mathbf{x}_1, \mathbf{x}_2)$, the linear projection kernels $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$, and the noise variances σ_1^2, σ_2^2.[3] We are interested in optimal mean-squared error performance. Therefore, reconstruction of the signal of interest $\mathbf{x}_1$ from the measurements $\mathbf{y}_1$ and $\mathbf{y}_2$ is obtained via the conditional mean estimator given by

$$\hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{y}_2) = \mathbb{E}[\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2] = \int_{-\infty}^{+\infty} \mathbf{x}_1 p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2) d\mathbf{x}_1, \quad (6)$$

where $p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2)$ is the *a posteriori* pdf of $\mathbf{x}_1$ given the measurements $\mathbf{y}_1$ and $\mathbf{y}_2$, so that the minimum mean-squared error

---

[2]Note that this assumption on the linear spaces occupied by signals in different classes reflects well the behavior of many real data ensembles for various applications such as face recognition, video motion segmentation, and digits classification [20], [44]. In addition, such assumption will enable us to simplify the statement of some of our results.

[3]Although our analysis assumes that the decoder can access the true signal distributions, in Sections IV and V we also conduct experiments with real datasets to assess scenarios where the decoder does not know the true distributions but rather approximate ones, that are learnt from training data.

is given by

$$\text{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = \mathbb{E}[\|\mathbf{x}_1 - \hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{y}_2)\|^2]. \quad (7)$$

Note that we express the MMSE explicitly as a function of the noise variances σ_1^2, σ_2^2 and the linear projection kernels $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$.

Our objective is to derive sufficient and (where possible) necessary conditions on the number of measurements for the MMSE to tend to zero in the low-noise regime, i.e.,

$$\lim_{\sigma_1^2, \sigma_2^2 \to 0} \text{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = 0. \quad (8)$$

In particular, we consider the scenario where the measurement matrix associated with the signal of interest, $\mathbf{\Phi}_1$, is randomly constituted,[4] whereas the measurement matrix associated with the side information, $\mathbf{\Phi}_2$, is optimally designed as follows:

$$\mathbf{\Phi}_2^\star = \arg\min_{\mathbf{\Phi}_2} \text{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2),$$

$$\text{subject to } \text{tr}(\mathbf{\Phi}_2 \mathbf{\Phi}_2^\mathsf{T}) \le m_2, \quad (9)$$

where the trace constraint in (9) limits the average energy corresponding to the projection kernel.

The challenge associated with the characterization of sufficient and necessary conditions on the number of measurements for reliable reconstruction (i.e., such that (8) holds) is due to the fact that (7) does not admit a closed form expression, even though (6) does for GMMs (the problem is also compounded in view of the fact that the solution to (9) cannot be analytically characterized). Therefore, our ensuing analysis will rely on bounds on the MMSE as a means to characterize conditions and linear projection kernel designs for reliable reconstruction.

III. BOUNDS ON THE NUMBER OF MEASUREMENTS FOR RELIABLE RECONSTRUCTION

We focus now on determining sufficient conditions (and necessary conditions in some cases) on the minimum number of measurements m_1 and m_2 for reliable reconstruction in the low-noise regime (i.e., such that (8) holds), assuming that the measurement matrix associated to the signal of interest is randomly constituted, and the measurement matrix associated to the side information is optimally designed. We consider both the simpler scenario where the signal of interest and the side information obey a joint multivariate Gaussian distribution as well as the more challenging scenario where the signals obey a joint GMM.

A. Gaussian Sources

We start by considering the case where $\mathbf{x}_1$ and $\mathbf{x}_2$ are described by a joint Gaussian distribution with mean $\boldsymbol{\mu}_\mathbf{x}$ and covariance matrix $\mathbf{\Sigma}_\mathbf{x}$, i.e., when $K_1 = K_2 = 1$ in (4).

---

[4]In this work, random measurement matrices are drawn from left-rotationally invariant distributions. A random matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ is said to be (left or right) rotation-invariant if the joint pdf of its entries $p(\mathbf{M})$ satisfies $p(\mathbf{\Theta M}) = p(\mathbf{M})$, or $p(\mathbf{M\Psi}) = p(\mathbf{M})$, respectively, for any orthogonal matrix $\mathbf{\Theta}$ or $\mathbf{\Psi}$. 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 commonly used in CS.

Theorem 1: Consider the measurements model in (1) and (2), where $\mathbf{x}_1$ and $\mathbf{x}_2$ are described by a joint Gaussian distribution with mean $\boldsymbol{\mu}_\mathbf{x}$ and covariance $\mathbf{\Sigma}_\mathbf{x}$, such that $r_\mathbf{x} = \text{rank}(\mathbf{\Sigma}_\mathbf{x})$, $r_{\mathbf{x}_1} = \text{rank}(\mathbf{\Sigma}_{\mathbf{x}_1})$, and $r_{\mathbf{x}_2} = \text{rank}(\mathbf{\Sigma}_{\mathbf{x}_2})$. Assume that $\mathbf{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ is drawn from a left-rotationally invariant distribution, and $\mathbf{\Phi}_2 = \mathbf{\Phi}_2^\star$ is the optimal projection matrix corresponding to the solution of the optimization problem in (9). Then, it holds

$$\lim_{\sigma_1^2, \sigma_2^2 \to 0} \text{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2^\star) = 0 \Leftrightarrow \begin{cases} m_1 + m_2 \ge r_{\mathbf{x}_1} \\ m_1 \ge r_\mathbf{x} - r_{\mathbf{x}_2} \end{cases}. \quad (10)$$

Proof: See Appendix A. ∎

Note that Theorem 1 establishes conditions on the number of measurements m_1 and m_2 for reliable reconstruction. These conditions suggest that:

1) The total number of measurements $m_1 + m_2$ must be equal to at least the dimension of the space spanned by the signal of interest $r_{\mathbf{x}_1}$. We recall that, without side information we would need $m_1 \ge r_{\mathbf{x}_1}$, which means that the m_2 measurements taken from the side information signal allow to decrease the number of measurements m_1 required for reliable reconstruction by m_2. Note that this does not happen in general for the case when side information measurements are random, since, in that case, the number of measurements $m_1 + m_2$ must be equal to at least $r_\mathbf{x} \ge r_{\mathbf{x}_1}$ [33, Theorem 3].

2) The number of measurements taken from $\mathbf{x}_1$ must also be at least equal to the codimension of the intersection of the space spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$ in the space spanned by $\mathbf{x}_1$. In other terms, this means that the number of measurements taken from $\mathbf{x}_1$ must be large enough to capture the specific attributes of $\mathbf{x}_1$ that cannot be inferred from the observation of $\mathbf{x}_2$.

The necessary and sufficient conditions for reliable reconstruction for Gaussian sources in Theorem 1 also suggest a concrete linear projection kernel to acquire the side information (see Appendix A). Such design is obtained from the generalized singular value decomposition (GSVD) [45] associated with two matrices related to the covariance matrices of the sources, $\mathbf{A} = (\mathbf{\Sigma}_{\mathbf{x}_2})^{1/2}$ and $\mathbf{B} = (\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^\dagger \mathbf{\Sigma}_{\mathbf{x}_{12}})^{1/2}$. In particular, these matrices can be written as follows:

$$\mathbf{A} = \mathbf{UCX}^\mathsf{T} \quad , \quad \mathbf{B} = \mathbf{VSX}^\mathsf{T}, \quad (11)$$

where $\mathbf{U} \in \mathbb{R}^{n_2 \times n_2}$, $\mathbf{V} \in \mathbb{R}^{n_2 \times n_2}$ are unitary matrices, $\mathbf{X} \in \mathbb{R}^{n_2 \times n_2}$ is non-singular, and $\mathbf{C} = [\mathbf{\Lambda_A} \ \mathbf{0}]$ and $\mathbf{S} = [\mathbf{\Lambda_B} \ \mathbf{0}]$ are diagonal matrices with

$$\mathbf{\Lambda_A} = \begin{bmatrix} \overset{r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_\mathbf{x}}{\mathbf{I}} & \overset{r_\mathbf{x} - r_{\mathbf{x}_1}}{} \\ & \mathbf{D_A} \end{bmatrix}, \quad (12)$$

$$\mathbf{\Lambda_B} = \begin{bmatrix} \overset{r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_\mathbf{x}}{\mathbf{0}} & \overset{r_\mathbf{x} - r_{\mathbf{x}_1}}{} \\ & \mathbf{D_B} \end{bmatrix}, \quad (13)$$

where we have reported explicitly the dimensions of the square blocks in $\mathbf{\Lambda_A}$ and $\mathbf{\Lambda_B}$.

Then, the proposed design is given by

$$
\boldsymbol{\Phi}_2 = \begin{bmatrix} \mathbf{I}_{m_2'} & \mathbf{0}_{m_2' \times (n_2 - m_2')} \\ \mathbf{0}_{(m_2 - m_2') \times m_2'} & \mathbf{0}_{(m_2 - m_2') \times (n_2 - m_2')} \end{bmatrix} \mathbf{X}^{-1}. \tag{14}
$$

where $m_2' = \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}$.

This shows – in view of the sharpness of the necessary and sufficient conditions in (10) – that an optimal matrix design exhibits the following attribute: it captures the portion of the linear space contained in the intersection between the image of $\boldsymbol{\Sigma}_{\mathbf{x}_2}$ and the null space of the matrix $\boldsymbol{\Sigma}_{\mathbf{x}_2} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}} \boldsymbol{\Sigma}_{\mathbf{x}_1}^{\dagger} \boldsymbol{\Sigma}_{\mathbf{x}_{12}}$. Then, noting that the conditional distribution $p(\mathbf{x}_2|\mathbf{x}_1)$ is Gaussian with covariance matrix given by $\boldsymbol{\Sigma}_{\mathbf{x}_2} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}} \boldsymbol{\Sigma}_{\mathbf{x}_1}^{\dagger} \boldsymbol{\Sigma}_{\mathbf{x}_{12}}$, we can observe that an optimal matrix design is aligned with the linear space spanned by signals drawn from the distribution $p(\mathbf{x}_2)$ which is not occupied by signals drawn from the conditional distribution $p(\mathbf{x}_2|\mathbf{x}_1)$ in order to improve the reconstruction performance. Intuitively, such design aims to capture the most of signal components from $\mathbf{x}_2$ that are maximally correlated with $\mathbf{x}_1$.

The result in Theorem 1 together with the result in [33, Theorem 3] also showcase the merit of an optimal measurement matrix $\boldsymbol{\Phi}_2^{\star}$ in comparison to a random one, in terms of the number of measurements necessary and sufficient for reliable reconstruction.

Corollary 2: Consider the measurement and source model in Theorem 1. Let $\mathcal{M}_G^R$ be the set of pairs (m_1, m_2) such that $\lim_{\sigma_1^2, \sigma_2^2 \to 0} \mathrm{MMSE}(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) = 0$ for random $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$, and let $\mathcal{M}_G^D$ be the set of pairs (m_1, m_2) such that $\lim_{\sigma_1^2, \sigma_2^2 \to 0} \mathrm{MMSE}(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2^{\star}) = 0$ for random $\boldsymbol{\Phi}_1$ and the optimal $\boldsymbol{\Phi}_2^{\star}$. Then, it holds

$$
\mathcal{M}_G^R \subseteq \mathcal{M}_G^D. \tag{15}
$$

Proof: See Appendix A. ∎

The advantage of using a designed matrix in relation to a random one to capture side information is also illustrated in Fig. 2. The shaded region of Fig. 2(a) represents the set $\mathcal{M}_G^D$; and the shaded region in Fig. 2(b) represents the set $\mathcal{M}_G^R$, as reported in [33, Theorem 3].

We can note that, depending on the specific values of $r_{\mathbf{x}}, r_{\mathbf{x}_1}, r_{\mathbf{x}_2}$, i.e., depending on the correlation between $\mathbf{x}_1$ and $\mathbf{x}_2$, proper design of the linear projection kernel $\boldsymbol{\Phi}_2$ can guarantee a significant advantage with respect to random kernels in terms of number of measurements required for reliable reconstruction. In particular, it is possible to show that $\mathcal{M}_G^R \subset \mathcal{M}_G^D$ if and only if $r_{\mathbf{x}} < r_{\mathbf{x}_1} + r_{\mathbf{x}_2}$ and $r_{\mathbf{x}} > r_{\mathbf{x}_1}$. The proof of this result relies on a simple comparison of the expressions for the minimum number of measurements required for reliable reconstruction with random and designed measurements, and it is not reported here for reasons of space.

This is somewhat surprising because it has been shown that optimal design of the linear projection kernel used to acquire the signal of interest does not reduce significantly the minimum number of measurements required for reliable reconstruction [34].

This fact can be justified by the following observation. In the case of Gaussian sources, the MMSE associated to the

(a) Designed projection kernel.
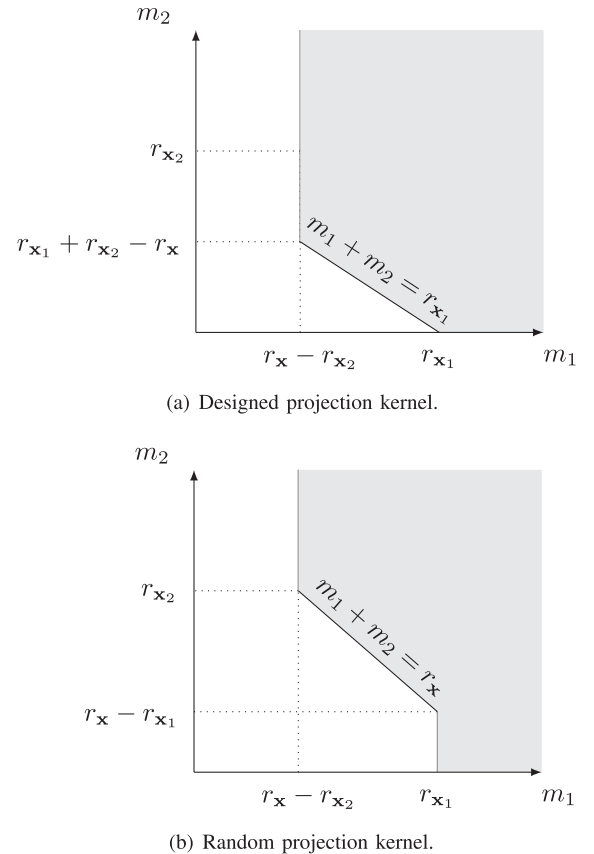

(b) Random projection kernel.

Fig. 2. Representation of the conditions on m_1 and m_2 for MMSE phase transition for Gaussian sources, for the designed kernel case (a) and random kernel case (b).

estimation of $\mathbf{x}_1$ from the observation of $\mathbf{y}_1$ and $\mathbf{y}_2$ is equivalent to the estimation of the vector $\mathbf{z} \sim p(\mathbf{x}_1|\mathbf{y}_2)$ from the observation of the vector $\boldsymbol{\Phi}_1 \mathbf{z} + \mathbf{w}_1$. Moreover, as it was shown in [22], the fundamental limits on the number of measurements needed to verify (8) are dictated by the dimension of the linear subspace spanned by the signal of interest, for both cases when $\boldsymbol{\Phi}_1$ is random and designed. Therefore, careful design of $\boldsymbol{\Phi}_1$ does not have an impact on such limit, whereas careful design of $\boldsymbol{\Phi}_2$ can reduce the dimension of the space spanned by signals $\mathbf{z} \sim p(\mathbf{x}_1|\mathbf{y}_2)$ in the low-noise regime, thus leading to a reduction of the minimum number of measurements needed for reliable reconstruction.

We will next leverage the intuition derived from the analysis of measurement designs for the Gaussian sources to analyze measurement designs for GMM sources. In particular, by leveraging the measurement design in (14), it will be possible to conceive new measurement strategies that outperform random ones for capturing side information in joint GMM models.

B. GMM Sources

We now consider the scenario where $\mathbf{x}_1$ and $\mathbf{x}_2$ obey a joint GMM. In this case, we only provide sufficient conditions rather than sufficient and necessary conditions for reliable reconstruction.

Theorem 3: Consider the measurements model in (1) and (2), where $\mathbf{x}_1$ and $\mathbf{x}_2$ conditioned on the underlying class labels $C_1 = i$, $C_2 = k$ obey a joint Gaussian distribution with mean $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}$ and covariance $\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}$ such that $r_{\mathbf{x}}^{(i,k)} = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)})$, $r_{\mathbf{x}_1}^{(i,k)} = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})$, and $r_{\mathbf{x}_2}^{(i,k)} = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}) \, \forall i,k$. Consider also that the linear spaces associated to the images of the co-variance matrices $\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}$, $\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)}$, and $\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ are independently drawn from a continuous pdf over the corresponding Grassmann manifold. Assume that $\boldsymbol{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ is random, drawn from a left-rotationally invariant distribution, and $\boldsymbol{\Phi}_2 = \boldsymbol{\Phi}_2^{\star}$ is the optimal projection matrix corresponding to the solution of the optimization problem in (9). Then, sufficient conditions on the number of projections m_1 and m_2 for

$$\lim_{\sigma_1^2, \sigma_2^2 \to 0} \text{MMSE}(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2^{\star}) = 0 \qquad (16)$$

are given by
$m_1 >$

$$\begin{cases} r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)} & , \text{if } m_2 \le r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} \\ \min\{r_{\mathbf{x}}^{(i,k)} - m_2, \\ \quad r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)}\} & , \text{if } r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} < m_2 \le r_{\mathbf{x}_2}^{(i,k)} \\ r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} & , \text{if } m_2 > r_{\mathbf{x}_2}^{(i,k)} \end{cases},$$
$$(17)$$

for $i = 1, \ldots, K_1$ and $k = 1, \ldots, K_2$, where $m_2^{(i,k)} \in \mathbb{N}$ are such that $m_2^{(i,k)} \le r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)} - r_{\mathbf{x}}^{(i,k)}$ and $\sum_{i,k} m_2^{(i,k)} = m_2$.

Proof: See Appendix B. ∎

Note that Theorem 3 also establishes conditions on the number of measurements m_1 and m_2 for reliable reconstruction. However, in this case the number of projections taken from the signal of interest is expressed in terms of the number of projections taken from the side information signal.

The sufficient conditions for reliable reconstruction for GMM sources embodied in Theorem 3 are obtained by considering a specific (suboptimal) design for $\boldsymbol{\Phi}_2$, which is inspired by the projection design that achieves the necessary and sufficient conditions for reliable reconstruction for Gaussian sources. This suboptimal design is obtained from the GSVD associated with the pair of matrices

$$\mathbf{A}^{(i,k)} = (\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})^{1/2}, \qquad (18)$$

$$\mathbf{B}^{(i,k)} = (\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger} \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)})^{1/2}, \qquad (19)$$

for all (i,k). In particular, we write these matrices as follows:

$$\mathbf{A}^{(i,k)} = \mathbf{U}^{(i,k)} \mathbf{C}^{(i,k)} (\mathbf{X}^{(i,k)})^{\mathrm{T}}, \qquad (20)$$

$$\mathbf{B}^{(i,k)} = \mathbf{V}^{(i,k)} \mathbf{S}^{(i,k)} (\mathbf{X}^{(i,k)})^{\mathrm{T}}, \qquad (21)$$

where $\mathbf{U}^{(i,k)} \in \mathbb{R}^{n_2 \times n_2}$, $\mathbf{V}^{(i,k)} \in \mathbb{R}^{n_2 \times n_2}$ are unitary matrices, $\mathbf{X}^{(i,k)} \in \mathbb{R}^{n_2 \times n_2}$ is non-singular, and $\mathbf{C}^{(i,k)} = [\boldsymbol{\Lambda}_{\mathbf{A}}^{(i,k)} \, \mathbf{0}]$ and $\mathbf{S}^{(i,k)} = [\boldsymbol{\Lambda}_{\mathbf{B}}^{(i,k)} \, \mathbf{0}]$ are diagonal matrices with

$$\boldsymbol{\Lambda}_{\mathbf{A}}^{i,k} = \begin{bmatrix} \overset{r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)} - r_{\mathbf{x}}^{(i,k)}}{\mathbf{I}} & \overset{r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)}}{} \\ & \mathbf{D}_{\mathbf{A}}^{(i,k)} \end{bmatrix}, \qquad (22)$$

$$\boldsymbol{\Lambda}_{\mathbf{B}}^{i,k} = \begin{bmatrix} \overset{r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)} - r_{\mathbf{x}}^{(i,k)}}{\mathbf{0}} & \overset{r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)}}{} \\ & \mathbf{D}_{\mathbf{B}}^{(i,k)} \end{bmatrix}, \qquad (23)$$

where we have reported explicitly the dimensions of the square blocks in $\boldsymbol{\Lambda}_{\mathbf{A}}^{(i,k)}$ and $\boldsymbol{\Lambda}_{\mathbf{B}}^{(i,k)}$.

We then define the matrix

$$\bar{\bar{\boldsymbol{\Phi}}}_2 = [(\bar{\bar{\boldsymbol{\Phi}}}_2^{(1,1)})^{\mathrm{T}}, \ldots, (\bar{\bar{\boldsymbol{\Phi}}}_2^{(K_1, K_2)})^{\mathrm{T}}]^{\mathrm{T}} \qquad (24)$$

and the matrices

$$\bar{\bar{\boldsymbol{\Phi}}}_2^{(i,k)} = \left[\mathbf{I}_{r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)} - r_{\mathbf{x}}^{(i,k)}} \, \mathbf{0} \right] (\mathbf{X}^{(i,k)})^{-1}. \qquad (25)$$

Finally, the suboptimal measurement matrix $\boldsymbol{\Phi}_2$ is given by picking any m_2 rows from $\bar{\bar{\boldsymbol{\Phi}}}_2$. More specifically, $\boldsymbol{\Phi}_2$ is obtained by picking $m_2^{(i,k)}$ rows from $\bar{\bar{\boldsymbol{\Phi}}}_2^{(i,k)}$ so that $\sum_{i,k} m_2^{(i,k)} = m_2$.[5] In fact, the $m_2^{(i,k)}$ measurements picked from $\bar{\bar{\boldsymbol{\Phi}}}_2^{(i,k)}$ capture the portion of the linear space spanned by signals drawn from the distribution $p(\mathbf{x}_2 | C_1 = i, C_2 = k)$ which is not occupied by signals drawn from the distribution $p(\mathbf{x}_2 | \mathbf{x}_1, C_1 = i, C_2 = k)$. The remaining $m_2 - m_2^{(i,k)}$ measurements act as random measurements for signals in class $C_1 = i, C_2 = k$. We note in passing that the derivation of the sufficient conditions for reliable reconstruction in Theorem 3 does not follow immediately from the proof of Theorem 1, since, in the GMM case, it is necessary to quantitatively determine the effect that measurements designed for a specific Gaussian class have on a different Gaussian class within the Gaussian mixture model. The details of such derivation are reported in Appendix B.

Via Theorem 3 and [33, Theorem 4] it is also possible to showcase the merit of an optimal measurement matrix $\boldsymbol{\Phi}_2^{\star}$ in comparison to a random one, in terms of the number of measurements that guarantee reliable reconstruction.

Corollary 4: Consider the measurement and source model in Theorem 3. Let $\mathcal{M}_{GMM}^R$ be the set of pairs (m_1, m_2) that verify the sufficient conditions for $\lim_{\sigma_1^2, \sigma_2^2 \to 0} \text{MMSE}(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) = 0$ in [33, Theorem 4] for random $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$, and let $\mathcal{M}_{GMM}^D$ be the set of pairs (m_1, m_2) that verify the sufficient conditions for $\lim_{\sigma_1^2, \sigma_2^2 \to 0} \text{MMSE}(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2^{\star}) = 0$ in Theorem 3, for random $\boldsymbol{\Phi}_1$ and optimal $\boldsymbol{\Phi}_2^{\star}$. Then, it holds

$$\mathcal{M}_{GMM}^R \subseteq \mathcal{M}_{GMM}^D. \qquad (26)$$

Proof: See Appendix B. ∎

---

[5] It is also possible to tighten further the sufficient conditions in Theorem 3 by choosing the values of $m_2^{(i,k)}$ that minimize the maximum value of the right hand side of (17) for all (i,k), subject to the constraints $m_2^{(i,k)} \le r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)} - r_{\mathbf{x}}^{(i,k)}$, $\forall (i,k)$ and $\sum_{i,k} m_2^{(i,k)} = m_2$.

We note in passing that the sets $\mathcal{M}_{GMM}^{R}$ and $\mathcal{M}_{GMM}^{D}$ are obtained as the intersection of the regions associated to Gaussian distributions corresponding to the different class labels (i, k). Moreover, it is easy to verify that, if $r_{\mathbf{x}}^{(i,k)} < r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)}, \forall (i,k)$, $r_{\mathbf{x}}^{(i,k)} > r_{\mathbf{x}_1}^{(i,k)}, \forall (i,k)$ and $m_2^{(i,k)} > 0, \forall (i,k)$, then $\mathcal{M}_{GMM}^{R} \subset \mathcal{M}_{GMM}^{D}$. This means that, if at least one designed measurement is associated to each class in the GMM, then side information designed kernels guarantee reliable reconstruction with a strictly lower number of measurements than random ones. The proof of this results is also based on a simple comparison of the expression of the minimum number of measurements required for reliable reconstruction for random and designed measurements, and it is not reported here due to reasons of space.

We also note that, as for Gaussian sources, designed projection kernels $\mathbf{\Phi}_2$ for GMM sources allow to capture the information contained in the side information $\mathbf{x}_2$, which is mostly correlated with $\mathbf{x}_1$. In fact, designed kernels aim at neglecting the information contained in the space spanned by signals drawn from the distribution $p(\mathbf{x}_2|\mathbf{x}_1)$, which can be interpreted as the innovation component of $\mathbf{x}_2$ with respect to $\mathbf{x}_1$.

Finally, we emphasize that our proposed measurement designs in (14), (24), and (25) are applicable to any GMM model independently of the subspaces dimension or the ambient dimension. In the sequel, we apply our approach to problems involving imagery data where it is typical to extract low-dimensional patches from the high-dimensional images to be modeled via a GMM [22], [33].

IV. NUMERICAL RESULTS

We now provide numerical results to showcase that our analysis is aligned with practice. We first report results for synthetic data, considering both cases of Gaussian sources and GMM sources, that highlight the value of using designed measurement kernels with respect to random ones. Then, similar trends are also reported for real world imaging data. In all experiments, we set $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

A. Synthetic Data: Gaussian Sources

We start by considering the case when $\mathbf{x}_1$ and $\mathbf{x}_2$ are drawn from a joint Gaussian distribution. We assume that signal dimensions are $n_1 = 14$ and $n_2 = 6$. Both signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are zero-mean, i.e., $\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{0}$, and the corresponding covariance matrices are independently drawn from a continuous pdf over the corresponding Grassmann manifold such that $r_{\mathbf{x}} = 5$, $r_{\mathbf{x}_1} = 4$, and $r_{\mathbf{x}_2} = 3$. We assume that the projection kernel $\mathbf{\Phi}_1$ is generated with i.i.d., zero-mean, Gaussian entries with fixed variance. We fix the number of measurements extracted from the side information signal, $m_2 = 2$, and we consider three different cases: i) the case when $\mathbf{\Phi}_2$ is random, with i.i.d., zero-mean, Gaussian entries with fixed variance; ii) the case when $\mathbf{\Phi}_2$ is obtained via numerical solution of the problem (9); iii) the case when $\mathbf{\Phi}_2$ is obtained via the projection kernel design in (14).

Fig. 3 illustrates the MMSE vs. $1/\sigma^2$. We can note that the MMSE tends to zero when $m_1 \geq 3$ and $m_2 = 2$ for the case of random projection kernels, thus matching the conditions reported in [33, Theorem 3]. In contrast, we observe that
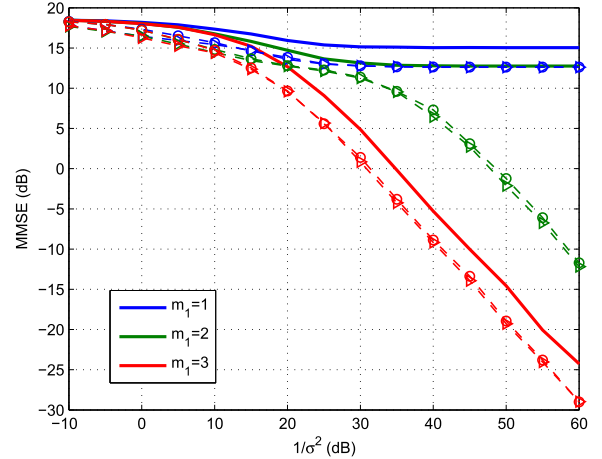

Fig. 3. MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3$ and $m_2 = 2$ for joint Gaussian sources. Side information with random projection kernel (solid lines). Side information with designed projection kernel $\mathbf{\Phi}_2^{\star}$ (numerical solution) (triangles) and suboptimal design in (14) (circles).

designed projection kernels guarantee that the reconstruction MMSE approaches zero in the low-noise regime when $m_1 \geq 2$ and $m_2 = 2$, as predicted by Theorem 1.

The results in Fig. 3 also show that both kernel designs ii) and iii) provide similar values of MMSE. Moreover, as predicted by the theory developed in Section III, the numerical solution of (9) provides a measurement kernel that requires the same number of measurements as the design in (14) in order to drive the MMSE to zero in the low-noise regime.

We also note in passing that it is possible to characterize further the MMSE in the moderate SNR regime by leveraging a first-order expansion of the MMSE. This is obtained by leveraging (49) in Appendix A and the result in [22, Theorem 1]:

$$\text{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2^{\star}) = \mathcal{A}_G + \mathcal{B}_G \cdot \sigma_1^2 + o(\sigma_1^2), \quad (27)$$

where $\mathcal{A}_G$ and $\mathcal{B}_G$ are obtained by defining

$$\mathbf{\Sigma}_{\mathbf{z}} = \mathbf{\Sigma}_{\mathbf{x}_1} - \mathbf{\Sigma}_{\mathbf{x}_{12}}(\mathbf{\Phi}_2^{\star})^{\text{T}}(\mathbf{\Phi}_2^{\star}\mathbf{\Sigma}_{\mathbf{x}_2}(\mathbf{\Phi}_2^{\star})^{\text{T}} + \mathbf{I}\sigma_2^2)^{\dagger}\mathbf{\Phi}_2^{\star}\mathbf{\Sigma}_{\mathbf{x}_{21}}, \quad (28)$$

and by considering the following eigenvalue decomposition:

$$\mathbf{\Theta} = \mathbf{\Sigma}_{\mathbf{z}}^{\frac{1}{2}} \mathbf{\Phi}_1^{\text{T}} \mathbf{\Phi}_1 \mathbf{\Sigma}_{\mathbf{z}}^{\frac{1}{2}} = \mathbf{U}_{\mathbf{\Theta}} \mathbf{\Lambda}_{\mathbf{\Theta}} \mathbf{U}_{\mathbf{\Theta}}^{\text{T}}. \quad (29)$$

Namely, on defining $r_{\mathbf{\Theta}} = \text{rank}(\mathbf{\Theta})$, on recalling $r_{\mathbf{z}} = \text{rank}(\mathbf{\Sigma}_{\mathbf{z}})$, on writing $\mathbf{\Lambda}_{\mathbf{\Theta}} = \text{diag}(\lambda_{\mathbf{\Theta},1}, \ldots, \lambda_{\mathbf{\Theta},r_{\mathbf{\Theta}}}, 0, \ldots, 0)$, and on denoting by $\mathbf{u}_{\mathbf{\Theta},t}$ the t-th column of $\mathbf{U}_{\mathbf{\Theta}}$, we have

$$\mathcal{A}_G = \sum_{t=r_{\mathbf{\Theta}}+1}^{r_{\mathbf{z}}} \mathbf{u}_{\mathbf{\Theta},t}^{\text{T}} \mathbf{\Sigma}_{\mathbf{z}} \mathbf{u}_{\mathbf{\Theta},t}, \quad (30)$$

and

$$\mathcal{B}_G = \sum_{t=1}^{r_{\mathbf{\Theta}}} \frac{1}{\lambda_{\mathbf{\Theta},t}} \mathbf{u}_{\mathbf{\Theta},t}^{\text{T}} \mathbf{\Sigma}_{\mathbf{z}} \mathbf{u}_{\mathbf{\Theta},t}. \quad (31)$$

Then, we note that, when the number of measurements m_1 is such that $\mathcal{A}_G = 0$, i.e., such that the conditions in Theorem 1 are verified, we have

$$\log \text{MMSE}_G(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2^{\star}) \approx \log \mathcal{B}_G + \log \sigma_1^2, \quad (32)$$

where the term $\log \mathcal{B}_G$ represents a power offset (in a dB-dB scale) and the term $\log \sigma_1^2$ explains the linear behavior of the MMSE as a function of the SNR. In fact, in this case the expression of $\mathcal{B}_G$ could in principle be leveraged in order to optimize further the measurement designs in terms of power offset. Namely, on denoting by $\mathcal{D}_G$ the set of matrices $\boldsymbol{\Phi}_2$ obtained from the expression (14) by replacing $\mathbf{I}_{m'_2}$ with a diagonal matrix with nonzero diagonal elements, we can further optimize the measurement designs by solving the optimization problem:

$$\arg \min_{\boldsymbol{\Phi}_2 \in \mathcal{D}_G} \mathcal{B}_G, \tag{33}$$

$$\text{subject to } \operatorname{tr}(\boldsymbol{\Phi}_2 \boldsymbol{\Phi}_2^{\mathrm{T}}) \leq m_2. \tag{34}$$

We also note that the MMSE behavior is consistent with that in [46] and our analysis generalizes that in [47] because we provide a closed-form expansion for the behavior of the MMSE whereas [47] provide only fixed-point equations to characterize the MMSE.

B. Synthetic Data: GMM Sources

We now consider a joint GMM distribution for signals $\mathbf{x}_1$ and $\mathbf{x}_2$, with dimensions $n_1 = 14$ and $n_2 = 6$, $K_1 = K_2 = 2$, and $P_{C_1, C_2}(1, 2) = P_{C_1, C_2}(2, 1) = 0$. All means associated with the various Gaussian distributions are zero, i.e., $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \mathbf{0}$, and the covariance matrices are randomly generated such that $r_{\mathbf{x}}^{(i,k)} = 6$, $r_{\mathbf{x}_1}^{(i,k)} = 3$ and $r_{\mathbf{x}_2}^{(i,k)} = 5$ for $i = 1, 2$ and $k = 1, 2$. The images of the covariance matrices associated with different classes are drawn uniformly at random from the corresponding Grassmann manifold. We also assume that the projection kernel $\boldsymbol{\Phi}_1$ has i.i.d., zero-mean, Gaussian entries with fixed variance. We fix $m_2 = 3$ and we also consider three different cases: i) the case when $\boldsymbol{\Phi}_2$ is random, with i.i.d., zero-mean, Gaussian entries with fixed variance; ii) the case when $\boldsymbol{\Phi}_2$ is obtained via numerical solution of the problem (9); iii) the case when $\boldsymbol{\Phi}_2$ is obtained via the projection kernel design in (25). In this last case, we report numerical results for two different measurement configurations: the first one is obtained by setting $m_2^{(1,1)} = 1$ and $m_2^{(2,2)} = 2$, whereas the second one corresponds to choosing $m_2^{(1,1)} = 2$ and $m_2^{(2,2)} = 1$.

Fig. 4 reports the MMSE vs. $1/\sigma^2$. We observe that random projection kernels guarantee reliable reconstruction when $m_1 + m_2 > r_{\mathbf{x}}^{(i,k)} = 6$, as predicted by the results in [33, Theorem 4].[6] On the other hand, designing the projection kernel $\boldsymbol{\Phi}_2$ leads to reliable reconstruction with a lower number of measurements, as the MMSE is observed to approach zero in the low-noise regime when $m_1 > 2$ and $m_2 = 3$. Such behavior is well aligned with the sufficient condition expressed in Theorem 3, since $r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)} \leq 2, \forall (i,k)$. Moreover, we observe that, in this case, it turns out that the design in (25) requires the same number of projections as the optimal design in (9) in order to guarantee reliable reconstruction, but this may not always be the case.

[6]The analysis carried out in [33] considers a slightly different signal model in which $\mathbf{x}_1$ and $\mathbf{x}_2$ obey only an approximately low-rank joint GMM distribution and linear measurements are noiseless. However, the results on the number of random measurements required for reliable reconstruction can be easily modified to fit the signal model considered in this work.
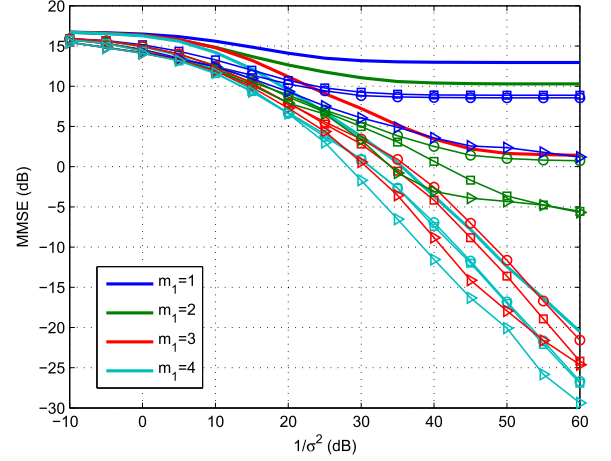

Fig. 4. MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3, 4$ and $m_2 = 3$ for joint GMM sources. Side information with random projection kernel (solid lines). Side information with designed projection kernel $\boldsymbol{\Phi}_2^{\star}$ obtained via numerical solution of the problem (9) (triangles). Side information with suboptimal design in (25) with $m_2^{(1,1)} = 1$ and $m_2^{(2,2)} = 2$ (circles). Side information with suboptimal design in (25) with $m_2^{(1,1)} = 2$ and $m_2^{(2,2)} = 1$ (squares).

We also note that, in this case, different choices of the number of measurements assigned to each Gaussian class, $m_2^{(1,1)}$ and $m_2^{(2,2)}$, do not have an impact on the asymptotic behavior of the MMSE in the low-noise regime. However, such different measurement configurations can lead to different values of the MMSE in intermediate noise regimes.

Akin to the Gaussian case, we can provide a characterization of the MMSE for the case of GMM sources in the finite noise regime. Such characterization builds upon an expansion of a lower bound on the MMSE, which we denote by $\mathrm{MSE}_{GMM}^{LB}(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2^{\star})$. Such expansion, which is akin to that reported in [33, Lemma 1] is given by:

$$\mathrm{MSE}_{GMM}^{LB}(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2^{\star}) = \mathcal{A}_{GMM} + \mathcal{B}_{GMM} \cdot \sigma_1^2 + o(\sigma_1^2), \tag{35}$$

where $\mathcal{A}_{GMM}$ and $\mathcal{B}_{GMM}$ are obtained in a similar way as done for $\mathcal{A}_G$ and $\mathcal{B}_G$. In particular, we have

$$\mathcal{A}_{GMM} = \sum_{(i,k)} p_{C_1, C_2}(i, k) \mathcal{A}^{(i,k)}, \tag{36}$$

and

$$\mathcal{B}_{GMM} = \sum_{(i,k)} p_{C_1, C_2}(i, k) \mathcal{B}^{(i,k)}. \tag{37}$$

Then, the terms $\mathcal{A}^{(i,k)}$ and $\mathcal{B}^{(i,k)}$ are obtained by defining

$$\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)} = \boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} (\boldsymbol{\Phi}_2^{\star})^{\mathrm{T}}$$
$$\cdot (\boldsymbol{\Phi}_2^{\star} \boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} (\boldsymbol{\Phi}_2^{\star})^{\mathrm{T}} + \mathbf{I} \sigma_2^2)^{\dagger} \boldsymbol{\Phi}_2^{\star} \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}, \tag{38}$$

and by considering the following eigenvalue decomposition:

$$\boldsymbol{\Theta}^{(i,k)} = (\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)})^{\frac{1}{2}} \boldsymbol{\Phi}_1^{\mathrm{T}} \boldsymbol{\Phi}_1 (\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)})^{\frac{1}{2}}$$
$$= \mathbf{U}_{\boldsymbol{\Theta}}^{(i,k)} \boldsymbol{\Lambda}_{\boldsymbol{\Theta}}^{(i,k)} (\mathbf{U}_{\boldsymbol{\Theta}}^{(i,k)})^{\mathrm{T}}. \tag{39}$$

Namely, on denoting $r_{\boldsymbol{\Theta}}^{(i,k)} = \operatorname{rank}(\boldsymbol{\Theta}^{(i,k)})$, on recalling $r_{\mathbf{z}}^{(i,k)} = \operatorname{rank}(\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)})$, on writing $\boldsymbol{\Lambda}_{\boldsymbol{\Theta}}^{(i,k)} = \operatorname{diag}(\lambda_{\boldsymbol{\Theta},1}^{(i,k)},$

$\ldots, \lambda_{\Theta, r_\Theta}^{(i,k)}, 0, \ldots, 0)$, and on denoting by $\mathbf{u}_{\Theta, t}^{(i,k)}$ the t-th column of $\mathbf{U}_{\Theta}^{(i,k)}$, we have

$$\mathcal{A}^{(i,k)} = \sum_{t=r_\Theta^{(i,k)}+1}^{r_{\mathbf{z}}^{(i,k)}} (\mathbf{u}_{\Theta, t}^{(i,k)})^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)} \mathbf{u}_{\Theta, t}^{(i,k)}, \qquad (40)$$

and

$$\mathcal{B}^{(i,k)} = \sum_{t=1}^{r_\Theta^{(i,k)}} \frac{1}{\lambda_{\Theta, t}^{(i,k)}} (\mathbf{u}_{\Theta, t}^{(i,k)})^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)} \mathbf{u}_{\Theta, t}^{(i,k)}. \qquad (41)$$

Moreover, such expansion of the lower bound of the MMSE for GMM sources is shown to be tight when the number of measurements m_1 is such that the conditions in [33, Lemma 4] are verified. Therefore, also in this case, the expression of $\mathcal{B}_{GMM}$ could in principle be leveraged in order to optimize further the measurement designs in terms of power offset. Namely, on denoting by $\mathcal{D}_{GMM}$ the set of matrices obtained from (24) and (25) by replacing the identity matrix $\mathbf{I}_{r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)} - r_{\mathbf{x}}^{(i,k)}}$ with a diagonal matrix with nonzero diagonal entries, we can express the power-offset designed kernel as:

$$\arg \min_{\boldsymbol{\Phi}_2 \in \mathcal{D}_{GMM}} \mathcal{B}_{GMM}, \qquad (42)$$

$$\text{subject to } \mathrm{tr}(\boldsymbol{\Phi}_2 \boldsymbol{\Phi}_2^{\mathrm{T}}) \leq m_2. \qquad (43)$$

C. Real Imaging Data

It is now natural to ask whether the measurement designs identified in (9) and (25) lead to better reconstruction results with real data. We use a high-resolution image "Lena" (512 × 512 pixels) as the input signal, and a low-resolution version of the same subject (128 × 128 pixels) as side information. Both images are partitioned into non-overlapping patches, so that vectors $\mathbf{x}_1$ represent 8×8 non-overlapping patches extracted from the high-resolution image, and vectors $\mathbf{x}_2$ represent 2×2 patches from the side information image, which correspond to the same spatial portion of the subject. The vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are assumed to be described by a joint GMM prior with $K = K_1 = K_2 = 20$ classes. The parameters of the joint GMM, i.e., prior class probabilities, class-conditioned means and class-conditioned covariance matrices, are estimated using the EM algorithm [23] over a set of patches extracted from images in the "Caltech 101" dataset [48]. Also in this case, we assume that the projection kernel $\boldsymbol{\Phi}_1$ has i.i.d., zero-mean, Gaussian entries with fixed variance.

In Fig. 5 we report reconstruction examples obtained with $m_1 = 10$ and $m_2 = 4$ linear measurement for each non-overlapping patch, and with noise level $\sigma^2 = -40$ dB. From left to right, the reconstruction images correspond to different scenarios: (a) $\boldsymbol{\Phi}_2$ is random, with i.i.d., zero-mean Gaussian entries with fixed variance; (b) $\boldsymbol{\Phi}_2$ is obtained numerically via the solution of the problem in (9); (c) $\boldsymbol{\Phi}_2$ is designed via the GSVD-based design in (25).[7]

(a) Random kernels.　(b) Designed kernel (numerical solutions).　(c) Designed kernel (GSVD).

Fig. 5. Reconstruction results of image "Lena" for $\sigma^2 = -40$ dB. From left to right, the reconstruction PSNR values are 25.4 dB, 28.1 dB, and 28.1 dB. Note that, in this case, the GSVD-based design described in (25) achieves approximately the same PSNR as the design corresponding to the numerical solution of the problem in (9).

The reconstruction results for both designed cases in Fig. 5(b) and (c) exhibit a peak signal-to-noise ratio (PSNR) gain of approximately 3 dB with respect to the case of random projection kernels (Fig. 5(a)). Moreover, we observe that the design described in (25) achieves similar PSNR values as when the projection kernel is designed via more computationally expensive numerical solution of the optimization problem in (9). This suggests that (25) can be used as an off-the-shelf measurement design procedure applicable to compressive sensing systems with side information where the signals are modeled via GMMs.

Finally, it is also instructive to describe the computational complexity of the proposed approach as a function of the total ambient dimension $n_1 + n_2$, the number of Gaussian components in the mixture $K_1 K_2$,[8] the total number of linear measurements $m_1 + m_2$, and the number of training samples T. In particular, there are three major computational aspects: (1) the first relates to the complexity associated with learning the Gaussian mixture model, (2) the second relates to the complexity associated with the measurement design method, and (3) the third is associated with the reconstruction procedure.

The expectation maximization algorithm is well established in the pattern recognition literature (see, for example, [23, Chapter 9]). The EM algorithm is an iterative method, where the computational complexity of each iteration depends on the ambient dimension of the signal vectors, the number of Gaussian components in the mixture, and the number of training samples, as follows: $\mathcal{O}((n_1 + n_2)T + K_1 K_2 T^2)$.

The reconstruction method, i.e., the conditional mean estimator, can be written in closed form. It involves the inversion of $K_1 K_2$ matrices with dimension $(m_1 + m_2) \times (m_1 + m_2)$. Therefore, the complexity of the reconstruction method is $\mathcal{O}(K_1 K_2 (m_1 + m_2)^3)$.

The design method involves the computation of the GSVD of $K_1 K_2$ pairs of matrices of dimension $n_2 \times n_2$. Therefore, the computational complexity of the design method is $\mathcal{O}(K_1 K_2 n_2^3)$.

---

[7] In this case, the matrix $\bar{\boldsymbol{\Phi}}_2 \in \mathbb{R}^{80 \times 4}$ was generated by picking the first 4 rows of the matrix $(\mathbf{X}^{(i,k)})^{-1}$ in (25), for all (i, k). Then, $\boldsymbol{\Phi}_2$ was obtained by picking the $m_2 = 4$ rows from $\bar{\boldsymbol{\Phi}}_2$ that corresponded to the lowest reconstruction MMSE. In general, $\boldsymbol{\Phi}_2$ could be formed by randomly picking $m_2 = 4$ rows from $\bar{\boldsymbol{\Phi}}_2$.

[8] In general, the number of Gaussian components in the GMM is given by $K_1 K_2$. However, in the numerical examples reported in Sections IV-C and V, the class labels associated to the signal of interest and the side information signal are assumed to be perfectly correlated. In this case, the total number of classes in the GMM is equal to $K = K_1 = K_2$.

V. CASE STUDY: PAN-SHARPENING OF RGB IMAGES WITH COMPRESSIVE HYPERSPECTRAL SIDE INFORMATION

Finally, we present an example of how the linear projection design scheme used to capture side information can also be leveraged in order to improve the performance of real world imaging applications. In particular, we consider a pan-sharpening application, which involves the recovery of a high-resolution, color image from a high-resolution panchromatic snapshot and low-resolution hyperspectral images of the same scene.

This problem arises often in the context of remote sensing applications [49], where panchromatic sensors are typically used to produce high-resolution images, and hyperspectral cameras are used to provide lower resolution images. The use of spatial information from the high-resolution panchromatic image and color information from the low-resolution hyperspectral snapshots can then be combined in order to generate a high-resolution RGB image of the scene. Our case study involves the use of a compressive hyperspectral camera (the CASSI in [39]) – in *lieu* of a standard hyperspectral camera – that provides for a mechanism to design the hyperspectral measurements in order to improve performance further.

A. System Model

We consider a high-resolution (512×512 pixels), RGB image of a scene as the signal of interest (Fig. 6(a)), and a lower resolution version (256×256 pixels) of a hyperspectral datacube ($N_\lambda = 33$ different channels) of the same scene as the side information signal (Fig. 6(d)).

In this case, vectors $\mathbf{x}_1$ represent 8×8 patches extracted from the signal of interest and vectors $\mathbf{x}_2$ represent 4×4 patches corresponding to the same spatial location extracted from the hyperspectral image datacube. In particular, vectors extracted from the signal of interest and the side information are subdivided into overlapping patches with overlap stride equal to 4 and 2 pixels,[9] respectively.

The vector $\mathbf{y}_1$ contains high-resolution, panchromatic measurements associated with patches of the image of interest (an example of panchromatic measurements is reported in Fig. 6(b)). The linear projection kernel $\mathbf{\Phi}_1$ models the conversion from high-resolution RGB images to a gray scale, panchromatic, high-resolution version of the same image. In particular, on neglecting the effect of gamma compression and expansion,[10] and by arranging the RGB channels of the image patches in the vectors $\mathbf{x}_1^R, \mathbf{x}_1^G$, and $\mathbf{x}_1^B$, so that,

$$\mathbf{x}_1 = \begin{bmatrix} \mathbf{x}_1^R \\ \mathbf{x}_1^G \\ \mathbf{x}_1^B \end{bmatrix}, \tag{44}$$

(a) Original RGB image. (b) Gray scale image. (c) CASSI measurement.
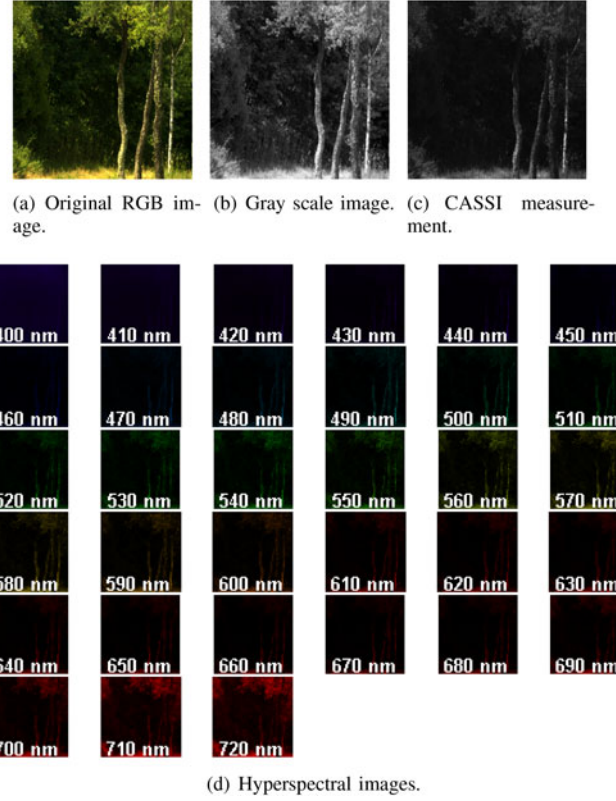

(d) Hyperspectral images.

Fig. 6. (a) Original RGB image, (b) gray scale image, (c) CASSI measurement, (d) hyperspectral images corresponding to different wavelengths.

we can express $\mathbf{\Phi}_1$ as

$$\mathbf{\Phi}_1 = \begin{bmatrix} c_r \mathbf{I} & c_g \mathbf{I} & c_b \mathbf{I} \end{bmatrix}. \tag{45}$$

Here, c_r, c_g, c_b are constant values that are determined by the measured intensity perception of human eyes associated to the three RGB channels. Typical values of such constants are $c_r = 0.2126, c_g = 0.7152, c_b = 0.0722$.

Finally, the vector $\mathbf{y}_2$ contains the side information measurements captured by the compressive hyperspectral imaging device, i.e., the CASSI camera (an example of CASSI measurements is reported in Fig. 6(c)). Such measurements are based on spectrally shifting the 3D spatio-spectral information (i.e., hyperspectral datacube), through a coded aperture and a dispersive element onto a 2D detector. More precisely, a scene is imaged through imaging optics onto a binary-value coded aperture over all different wavelengths and a dispersive element [51]–[53] then projects these multiplex signals of the coded scene with wavelength-dependent shifts (Fig. 8).[11] Then, on writing the vectors $\mathbf{x}_2$, which contain the data in the hyperspectral datacube corresponding to patches, as

$$\mathbf{x}_2 = \begin{bmatrix} \mathbf{x}_2^1 \\ \vdots \\ \mathbf{x}_2^{N_\lambda} \end{bmatrix}, \tag{46}$$

---

[9]The overlap stride denotes the distance between corresponding pixel locations in adjacent image patches.

[10]Gamma compression and expansion are one-to-one, nonlinear transformation of the RGB components in each pixel, hence they can be ignored in our measurement model [50].

[11]For more details on the actual implementation of the CASSI camera, please refer to [39].

(a) Training scene 1.   (b) Training scene 2.   (c) Training scene 3.

(d) Training scene 4.   (e) Training scene 5.   (f) Training scene 6.

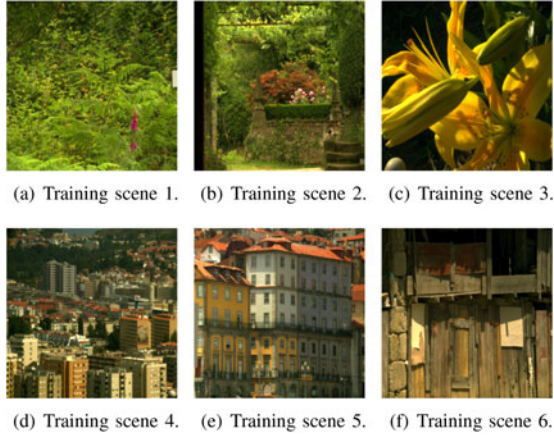Fig. 7.   Training dataset.

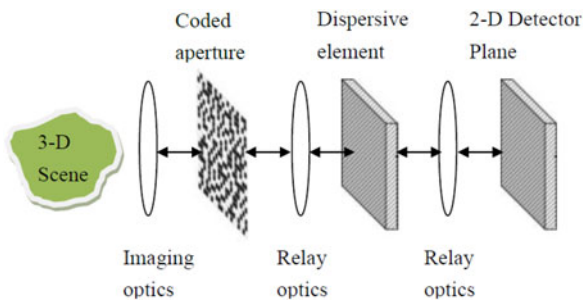

Fig. 8.   CASSI architecture.

where $\mathbf{x}_2^n$ contains the pixels associated to the n-th wavelength channel, we can express $\mathbf{\Phi}_2$ as follows
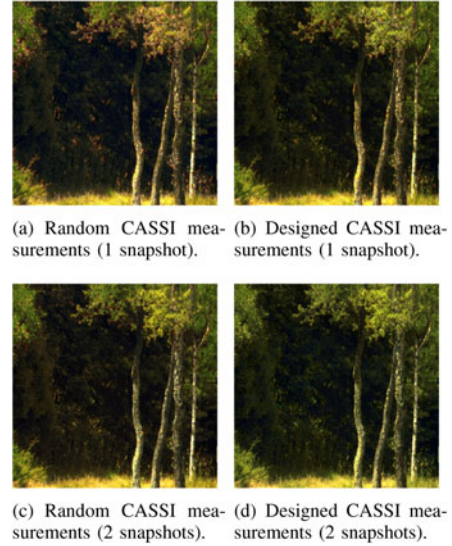
$$\mathbf{\Phi}_2 = \begin{bmatrix} \mathbf{M}^1, \dots, \mathbf{M}^{N_\lambda} \end{bmatrix}. \tag{47}$$

Here, the matrices $\mathbf{M}^1, \dots, \mathbf{M}^{N_\lambda}$ are diagonal with binary diagonal entries that take values 0 and 1 according to the corresponding pixel values in the coding mask.

It is important to note that, since the effects of the coding mask for different wavelengths are obtained by translations of the coded aperture in the device, the elements in the diagonal entries of the different matrices $\mathbf{M}^1, \dots, \mathbf{M}^{N_\lambda}$ are related one to the other, so it is not possible to choose them independently. In particular, in order to increase the diversity of compressive hyperspectral measurements, and in order to decrease the compression ratio, several snapshots of the same scene can be captured with the CASSI device, by changing the coding mask used for each snapshot. This means that that the design of the measurement matrix $\mathbf{\Phi}_2$ for the CASSI architecture involves the choice of the coded masks pixel values, i.e., the diagonal entries of the matrices $\mathbf{M}^1, \dots, \mathbf{M}^{N_\lambda}$.

B. Numerical Results

We report MMSE reconstruction results associated with this pan-sharpening applications, by modeling the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ via a joint GMM with $K = K_1 = K_2 = 20$ classes. The parameters of the corresponding distributions, i.e., prior class probabilities, class-conditioned means, and class-conditioned covariance matrices, are learnt by using the EM algorithm on the training set described in [39], which contains RGB images

(a) Random CASSI mea-   (b) Designed CASSI mea-
surements (1 snapshot).   surements (1 snapshot).

(c) Random CASSI mea-   (d) Designed CASSI mea-
surements (2 snapshots).   surements (2 snapshots).

Fig. 9.   Reconstruction examples for random and designed CASSI measurements, 1 snapshot ($m_2 = 16$) and 2 snapshots ($m_2 = 32$).

and hyperspectral snapshots of 6 different scenes, where 33 hyperspectral images corresponds to 33 different wavelengths ($N_\lambda = 33$) from 400 nm to 720 nm (the training scenes are shown in Fig. 7).

We also report MMSE reconstruction results corresponding to two different measurement settings. In the first case, binary coding mask values for CASSI measurements are random, in particular, they are drawn i.i.d. from a Bernoulli distribution with parameter 0.5. In the second case, the elements of the coding mask are designed according to the construction described in Section III.

However, in view of the fact that the CASSI architecture only permits one to constitute a certain class of coded apertures (see (47)) – our design is obtained as follows:

$$\mathbf{\Phi}_2^{\text{CASSI}} = \underset{\mathbf{\Phi}_2 \in \mathcal{F}_{\text{CASSI}}}{\arg\min} \left\| \mathbf{\Phi}_2 - \mathbf{\Phi}_2^{\text{GSVD}} \right\|_F, \tag{48}$$

where $\mathbf{\Phi}_2^{\text{CASSI}}$ embodies the CASSI coded aperture design, $\mathbf{\Phi}_2^{\text{GSVD}}$ embodies the design in (25), and $\mathcal{F}_{\text{CASSI}}$ denotes the set of coded aperture designs consistent with the CASSI architecture. That is, the proposed design $\mathbf{\Phi}_2^{\text{CASSI}}$ is the closest in Frobenius norm to the GSVD design $\mathbf{\Phi}_2^{\text{GSVD}}$ in (25), that is consistent with the CASSI constraints.[12]

Fig. 9 shows reconstruction examples of the RGB image in Fig. 6(a). Figs. 9(a) and (b) show reconstructions obtained with random and designed CASSI measurements, respectively, obtained with a single snapshot. We can observe that designed CASSI measurements can guarantee significant improvement with respect to random measurements. In particular, designed

[12]Note that the optimization problem (48) can be solved with computational complexity growing linearly with the number of binary pixel values corresponding to the coding mask. In fact, the diagonal elements of the matrices $\mathbf{M}^1, \dots, \mathbf{M}^{N_\lambda}$ in $\mathbf{\Phi}_2$ are generated by shifting the binary pattern in the coding mask. In particular, on considering single pixel shifts in a given direction, patches of size 4×4 and $N_\lambda = 33$ channels, the total number of pixels used to code a given patch is $4 \times (4 + 33 - 1) = 144$. Moreover, we have used the same mask for all patches contained in the image of interest.

TABLE I
PSNR VALUES OF THE RECONSTRUCTION EXAMPLES IN FIG. 9 AND FIG. 10

	Random CASSI	Designed CASSI	CDL Approach
1 snapshot	25.7831 dB	28.3577 dB	26.1185 dB
2 snapshots	28.6507 dB	30.7845 dB	28.0251 dB

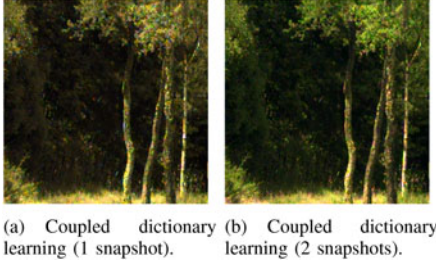(a) Coupled dictionary learning (1 snapshot). (b) Coupled dictionary learning (2 snapshots).

Fig. 10. Reconstruction examples for coupled dictionary learning (CDL) based approach, 1 snapshot ($m_2 = 16$) and 2 snapshots ($m_2 = 32$).

measurements are particularly effective in reducing the red color artifacts which are easily observable in Fig. 9(a). Figs. 9(c) and (d), which contain reconstructions examples for random and designed measurements corresponding to two CASSI snapshots, depict similar effects. In fact, as reported in Table I, for both cases of one and two CASSI snapshots, the proposed measurement design scheme leads to a PSNR gain of approximately 2 dB.

Finally, we also compare our GMM approach to linking the two different data modalities, i.e., the signal of interest and the side information signal, to another approach based on sparse representations induced by coupled dictionary learning (CDL) [38], [54] (note that this approach to linking different data modalities generalizes the JSM-1 and JSM-3 models proposed in [35] and [36]). In particular, we consider dictionaries with 512 atoms and common and innovation components with sparsities equal to 4 and 2, respectively. We use the training data shown in Fig. 7 to learn dictionaries that link the hyperspectral patches and the RGB patches using the method described in [38], [54]. The dimensions and strides of the patches used for dictionary learning are the same as those used in the previous experiment. We then use the testing data (as shown in Fig. 6(a)) to determine the reconstruction performance. In particular, we use the orthogonal matching pursuit (OMP) algorithm to reconstruct the signal of interest from noisy random linear projections of the signal of interest as well as noisy random linear projections from the side information. Here, note that we consider random measurements (as described earlier) rather than designed ones because it is not immediate how to conceive optimized linear measurement designs to capture the side information for this approach to linking the different data modalities. We note again that our proposed design leads to PSNR gains of about 2 dB in relation to the CDL approach. See Table I and also Fig. 10.

Overall, these gains are particularly significant by taking into account that the optimization of the measurement procedure has been effected under the strong constraints imposed by the CASSI architecture. Therefore, this also leads us to conclude that (25) or (48) for GMM sources can be used to guide how to capture side information in compressive sensing systems.

VI. CONCLUSION

This paper has investigated how to capture side information to aid in the reconstruction of a high-dimensional signal from low-dimensional linear measurements. In particular, by assuming that both the signal of interest and the side information signal are described by a joint GMM distribution, the paper has provided concrete linear side information measurement designs that guarantee the signal reconstruction error to approach zero in the low-noise regime.

It was shown that these proposed designs capture the information contained in the side information which is mostly correlated to the signal of interest. It was also shown that these designs can lead to substantial gains over random linear designs. This is somewhat surprising because it has been established that optimization of the acquisition process for the signal of interest does not have a significant impact on the number of measurements required for reliable reconstruction [34], but in contrast optimization of the side information acquisition process does.

We have also considered how to apply our measurement designs to a real-world pan-sharpening application in the presence of compressive hyper-spectral data. We have shown that our linear measurement designs deliver better reconstruction results in relation to random ones, both when one couples the different data modalities using a GMM model or else a CDL model. Of particular relevance, one of the advantages of using GMM models in relation to the CDL ones is that one can come up with concrete strategies to define optimal measurement designs.

APPENDIX A

A. Proof of Theorem 1

We now provide necessary and sufficient conditions on the number of measurements m_1 and m_2, for the MMSE to approach zero in the low-noise regime.

We will use the symbol $\mathrm{MMSE}(\mathbf{x}|\mathbf{y})$ to explicitly denote the MMSE in recovering $\mathbf{x}$ from the observation of $\mathbf{y}$. Then, by noting that the Gaussian MMSE does not depend on the value of the mean of the input signal, and by taking the expectation in the MMSE expression with respect to the random variables $\mathbf{x}_1|\mathbf{y}_2$ and $\mathbf{y}_2$, separately, it is possible to show that

$$\mathrm{MMSE}(\mathbf{x}_1|\mathbf{y}_1, \mathbf{y}_2) = \mathrm{MMSE}(\mathbf{z}|\mathbf{\Phi_1}\mathbf{z} + \mathbf{w}_1), \quad (49)$$

where $\mathbf{z} \sim p(\mathbf{x}_1|\mathbf{y}_2)$. By leveraging [22], we observe that $\mathrm{MMSE}(\mathbf{z}|\mathbf{\Phi_1}\mathbf{z} + \mathbf{w}_1)$ tends to be zero as $\sigma_1^2 \to 0$, if and only if

$$m_1 \geq \mathrm{rank}(\mathbf{\Sigma_z}) = r_{\mathbf{z}}. \quad (50)$$

Note that the covariance matrix $\mathbf{\Sigma_z}$ can be written as

$$\mathbf{\Sigma_z} = \mathbf{\Sigma_{x_1}} - \mathbf{\Sigma_{x_{12}}} \mathbf{\Phi_2^T}(\mathbf{\Phi_2}\mathbf{\Sigma_{x_2}}\mathbf{\Phi_2^T} + \mathbf{I}\sigma_2^2)^\dagger \mathbf{\Phi_2}\mathbf{\Sigma_{x_{21}}}, \quad (51)$$

and in particular, when $\sigma_2^2 \to 0$, $\mathbf{\Sigma_z}$ tends to

$$\mathbf{\Sigma_z} = \mathbf{\Sigma_{x_1}} - \mathbf{\Sigma_{x_{12}}} \mathbf{\Phi_2^T}(\mathbf{\Phi_2}\mathbf{\Sigma_{x_2}}\mathbf{\Phi_2^T})^\dagger \mathbf{\Phi_2}\mathbf{\Sigma_{x_{21}}}. \quad (52)$$

Note that $\mathbf{\Sigma_z}$ in (52) is the generalized Schur complement [55] of the block $\mathbf{\Phi_2}\mathbf{\Sigma_{x_2}}\mathbf{\Phi_2^T}$ of the positive semidefinite matrix

$$\mathbf{\Sigma_{x_1 \Phi_2 x_2}} = \begin{bmatrix} \mathbf{\Sigma_{x_1}} & \mathbf{\Sigma_{x_{12}}}\mathbf{\Phi_2^T} \\ \mathbf{\Phi_2}\mathbf{\Sigma_{x_{21}}} & \mathbf{\Phi_2}\mathbf{\Sigma_{x_2}}\mathbf{\Phi_2^T} \end{bmatrix}, \quad (53)$$

so that we have [55]

$$\text{rank}(\Sigma_{\mathbf{x}_1 \Phi_2 \mathbf{x}_2}) = r_{\mathbf{z}} + \text{rank}(\Phi_2 \Sigma_{\mathbf{x}_2} \Phi_2^T). \qquad (54)$$

In addition, on considering the matrix

$$\Sigma_{\Phi_2 \mathbf{x}_2 \mathbf{x}_1} = \begin{bmatrix} \Phi_2 \Sigma_{\mathbf{x}_2} \Phi_2^T & \Phi_2 \Sigma_{\mathbf{x}_{21}} \\ \Sigma_{\mathbf{x}_{12}} \Phi_2^T & \Sigma_{\mathbf{x}_1} \end{bmatrix}, \qquad (55)$$

and on applying the same rank computation, we also have

$$\text{rank}(\Sigma_{\Phi_2 \mathbf{x}_2 \mathbf{x}_1}) = r_{\mathbf{x}_1} + \text{rank}\left(\Phi_2 (\Sigma_{\mathbf{x}_2} - \Sigma_{\mathbf{x}_{21}} \Sigma_{\mathbf{x}_1}^\dagger \Sigma_{\mathbf{x}_{12}})\Phi_2^T\right). \qquad (56)$$

By substituting (54) and (56), we can express $r_{\mathbf{z}}$ as

$$r_{\mathbf{z}} = r_{\mathbf{x}_1} - \text{rank}(\Phi_2 \Sigma_{\mathbf{x}_2} \Phi_2^T)$$
$$+ \text{rank}(\Phi_2 (\Sigma_{\mathbf{x}_2} - \Sigma_{\mathbf{x}_{21}} \Sigma_{\mathbf{x}_1}^\dagger \Sigma_{\mathbf{x}_{12}})\Phi_2^T). \qquad (57)$$

The sufficient conditions for the MMSE phase transition can be obtained by considering the MMSE associated to a suboptimal projection matrix Φ_2 with respect to the optimization problem (9). Based on the GSVD of two matrices $\Sigma_{\mathbf{x}_2}$ and $\Sigma_{\mathbf{x}_2} - \Sigma_{\mathbf{x}_{21}} \Sigma_{\mathbf{x}_1}^\dagger \Sigma_{\mathbf{x}_{12}}$, the projection matrix Φ_2 is obtained in (14).

Then, it is straightforward to verify that $\text{rank}(\Phi_2 \Sigma_{\mathbf{x}_2} \Phi_2^T) = m_2'$ and $\text{rank}(\Phi_2(\Sigma_{\mathbf{x}_2} - \Sigma_{\mathbf{x}_{21}} \Sigma_{\mathbf{x}_1}^\dagger \Sigma_{\mathbf{x}_{12}})\Phi_2^T) = 0$, thus implying that, when adopting the side information projection matrix Φ_2 in (14), the corresponding MMSE approaches zero in the low-noise regime if

$$m_1 \geq r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}, \qquad (58)$$

which is equivalent to the condition (10).

Conversely, we can show that the condition in (58) is also necessary to achieve the MMSE phase transition when $\Phi_2 = \Phi_2^\star$. In order to do that, we show that, for any choice of $\Phi_2 \in \mathbb{R}^{m_2 \times n_2}$, it holds

$$r_{\mathbf{z}} \geq r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}. \qquad (59)$$

In particular, on leveraging the Sylvester's rank Theorem [56], which states that

$$\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{B}) - \dim(\text{Im}(\mathbf{B}^T) \cap \text{Null}(\mathbf{A})), \qquad (60)$$

we can write

$$\Delta(\Phi_2) = \text{rank}(\Phi_2 \Sigma_{\mathbf{x}_2} \Phi_2^T)$$
$$- \text{rank}(\Phi_2 (\Sigma_{\mathbf{x}_2} - \Sigma_{\mathbf{x}_{21}} \Sigma_{\mathbf{x}_1}^\dagger \Sigma_{\mathbf{x}_{12}})\Phi_2^T) \qquad (61)$$
$$= \dim(\text{Im}(\Phi_2^T) \cap \text{Null}(\Sigma_{\mathbf{x}_2} - \Sigma_{\mathbf{x}_{21}} \Sigma_{\mathbf{x}_1}^\dagger \Sigma_{\mathbf{x}_{12}}))$$
$$- \dim(\text{Im}(\Phi_2^T) \cap \text{Null}(\Sigma_{\mathbf{x}_2})). \qquad (62)$$

Then, on observing that $\Sigma_{\mathbf{x}_2} - \Sigma_{\mathbf{x}_{21}} \Sigma_{\mathbf{x}_1}^\dagger \Sigma_{\mathbf{x}_{12}}$ is the generalized Schur complement of $\Sigma_{\mathbf{x}_1}$ in $\Sigma_{\mathbf{x}}$ [57], and by leveraging [57, Lemma 4.1] in conjunction with [58, Theorem 4.3], we have that

$$\text{Null}(\Sigma_{\mathbf{x}_2}) \subseteq \text{Null}(\Sigma_{\mathbf{x}_2} - \Sigma_{\mathbf{x}_{21}} \Sigma_{\mathbf{x}_1}^\dagger \Sigma_{\mathbf{x}_{12}}), \qquad (63)$$

and

$$\dim(\text{Null}(\Sigma_{\mathbf{x}_2} - \Sigma_{\mathbf{x}_{21}} \Sigma_{\mathbf{x}_1}^\dagger \Sigma_{\mathbf{x}_{12}})) = n_2 - r_{\mathbf{x}} + r_{\mathbf{x}_1}, \qquad (64)$$
$$\dim(\text{Null}(\Sigma_{\mathbf{x}_2})) = n_2 - r_{\mathbf{x}_2}. \qquad (65)$$

Finally, on leveraging the following lemma, we can prove that $\Delta(\Phi_2) \leq \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}$.

Lemma 5: Consider two linear spaces $\mathcal{D}_1, \mathcal{D}_2 \in \mathbb{R}^n$, such that $\mathcal{D}_2 \subseteq \mathcal{D}_1$, with dimensions d_1 and $d_2, d_1 \geq d_2$, respectively. Consider a third linear space $\mathcal{V} \in \mathbb{R}^n$, with dimension v. Then,

$$\dim(\mathcal{V} \cap \mathcal{D}_1) - \dim(\mathcal{V} \cap \mathcal{D}_2) \leq \min\{v, d_1 - d_2\}. \qquad (66)$$

Proof: It is straightforward to note that the difference in (66) is always upper bounded by v, since $\dim(\mathcal{V} \cap \mathcal{D}_1) \leq \dim(\mathcal{V}) = v$.

Then, we divide the proof into two cases. In the first case, we assume $\dim(\mathcal{V} \cap \mathcal{D}_1) \leq d_1 - d_2$, from which (66) follows immediately. In the second case, we assume $\dim(\mathcal{V} \cap \mathcal{D}_1) > d_1 - d_2$, and we write $\dim(\mathcal{V} \cap \mathcal{D}_1) = d_1 - d_2 + k$, with $k \in \{0, 1, \ldots, d_2\}$. Then, we have that

$$\dim(\mathcal{V} \cap \mathcal{D}_2) \geq \dim((\mathcal{V} \cap \mathcal{D}_1) \cap \mathcal{D}_2) \qquad (67)$$
$$\geq [d_1 - d_2 + k + d_2 - d_1]^+ = k, \qquad (68)$$

where, in the last inequality, we have used fact that both $\mathcal{V} \cap \mathcal{D}_1$ and $\mathcal{D}_2$ are linear spaces contained in $\mathcal{D}_1$, which has dimension d_1. Therefore, from (68) we obtain $\dim(\mathcal{V} \cap \mathcal{D}_1) - \dim(\mathcal{V} \cap \mathcal{D}_2) \leq d_1 - d_2$, which concludes the proof of (66). ∎

B. Proof of Corollary 2

Note that, given any value of m_2, the minimum value of m_1 that verifies the necessary and sufficient conditions for the MMSE to approach zero in the low-noise regime when both Φ_1 and Φ_2 are random can be written as [33, Theorem 3]

$$m_1^R = r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_2}\} + \min\{m_2, r_{\mathbf{x}} - r_{\mathbf{x}_1}\}, \qquad (69)$$

whereas the minimum value of m_1 that verifies the necessary and sufficient condition in Theorem 1 can be written as

$$m_1^D = r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}. \qquad (70)$$

Then, the proof of Corollary 2 is obtained by comparing (69) and (70) for different values of m_2:
- If $m_2 \leq r_{\mathbf{x}} - r_{\mathbf{x}_1}$, we have $m_1^R = r_{\mathbf{x}_1} \geq r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\} = m_1^D$;
- If $r_{\mathbf{x}} - r_{\mathbf{x}_1} < m_2 < r_{\mathbf{x}_2}$, we need to consider two further cases: if $m_2 \geq r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}$, then $m_1^R = r_{\mathbf{x}} - m_2 > r_{\mathbf{x}} - r_{\mathbf{x}_2} = m_1^D$; on the other hand, if $m_2 < r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}$, then $m_1^R = r_{\mathbf{x}} - m_2 \geq r_{\mathbf{x}_1} - m_2 = m^D$.
- If $m_2 \geq r_{\mathbf{x}_2}$, then $m_1^R = r_{\mathbf{x}} - r_{\mathbf{x}_2} = m_1^D$.

Fig. 11 illustrates with different shading colors the different sets of values (m_1, m_2) considered in the proof of Corollary 2.

APPENDIX B

A. Proof of Theorem 3

When the signals $\mathbf{x}_1, \mathbf{x}_2$ are described via a joint GMM, the corresponding MMSE can be expressed as

$$\text{MMSE}(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2) = \mathbb{E}[\text{MMSE}(\mathbf{z} | \Phi_1 \mathbf{z} + \mathbf{w}_1)], \qquad (71)$$

where $\mathbf{z} \sim p(\mathbf{x}_1 | \mathbf{y}_2)$ and where the outer expectation is taken with respect to $\mathbf{y}_2$. Note that, for a given value of $\mathbf{y}_2$, $p(\mathbf{z})$ is a
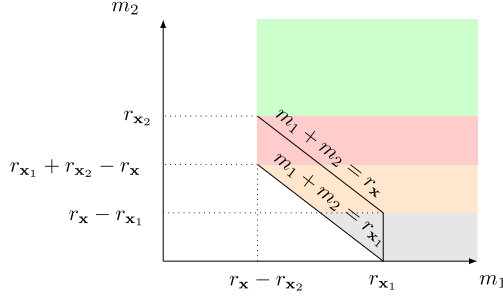
Fig. 11. Representation of the conditions on m_1 and m_2 for MMSE phase transition for Gaussian sources.

GMM distribution with the same number of classes as the joint GMM $p(\mathbf{x}_1, \mathbf{x}_2)$, since it holds

$$p(\mathbf{z}) = \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} p(C_1 = i, C_2 = k | \mathbf{y}_2)$$
$$\cdot\, p(\mathbf{x}_1 | \mathbf{y}_2, C_1 = i, C_2 = k) \tag{72}$$

$$= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} p(C_1 = i, C_2 = k | \mathbf{y}_2)$$
$$\cdot\, \frac{p(\mathbf{x}_1, \mathbf{y}_2, C_1 = i, C_2 = k)}{p(\mathbf{y}_2, C_1 = i, C_2 = k)}. \tag{73}$$

Therefore, by leveraging [22, Theorem 3], we can conclude that a sufficient condition on the number of projections m_1 required for $\mathrm{MMSE}(\mathbf{z} | \boldsymbol{\Phi}_1 \mathbf{z} + \mathbf{w}_1) \to 0$ when $\sigma_1^2 \to 0$ is given by

$$m_1 > \max_{(i,k)} \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)}) = \max_{(i,k)} r_{\mathbf{z}}^{(i,k)}, \tag{74}$$

where $\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)}$ is the covariance matrix of $\mathbf{z}$ conditioned on classes $C_1 = i, C_2 = k$. We also note that the covariances $\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)}$ depend only on the statistical description of $\mathbf{x}_1, \mathbf{y}_2$, and not on the particular realization of $\mathbf{y}_2$. Therefore, we can conclude that (74) represents also a sufficient condition for $\mathrm{MMSE}(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2) \to 0$ when $\sigma_1^2 \to 0$. The covariance matrix $\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)}$ can be written as

$$\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)} = \boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} \boldsymbol{\Phi}_2^{\mathrm{T}} (\boldsymbol{\Phi}_2 \boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} \boldsymbol{\Phi}_2^{\mathrm{T}} + \mathbf{I}\sigma_2^2)^{-1} \boldsymbol{\Phi}_2 \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}. \tag{75}$$

Moreover, we are interested in determining the behavior of the MMSE in the low-noise regime, that is when both $\sigma_1^2 \to 0$ and $\sigma_2^2 \to 0$. In particular, when $\sigma_2^2 \to 0$, the matrices $\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)}$ can be expressed as

$$\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)} = \boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} \boldsymbol{\Phi}_2^{\mathrm{T}} (\boldsymbol{\Phi}_2 \boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} \boldsymbol{\Phi}_2^{\mathrm{T}})^{\dagger} \boldsymbol{\Phi}_2 \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}, \tag{76}$$

due to continuity of the Schur component of a matrix with respect to its eigenvalues. Therefore, sufficient conditions for the MMSE to approach zero in the low-noise regime are obtained by expressing the quantities $r_{\mathbf{z}}^{(i,k)} = \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{z}}^{(i,k)})$ as a function of the sources parameters and the number of side information measurements m_2.

In particular, we obtain sufficient conditions for the MMSE phase transition on the number of projections m_1 and m_2 by considering a (possibly suboptimal) measurement matrix design for $\boldsymbol{\Phi}_2$, which is inspired on the design which achieves

the optimal phase transition for Gaussian sources. In particular, we define $\bar{\boldsymbol{\Phi}}_2^{(i,k)}$ as in (25). Then, we define $\bar{\boldsymbol{\Phi}}_2 = [(\bar{\boldsymbol{\Phi}}_2^{(1,1)})^{\mathrm{T}}, \ldots, (\bar{\boldsymbol{\Phi}}_2^{(K_1, K_2)})^{\mathrm{T}}]^{\mathrm{T}}$ as the matrix obtained by stacking all the matrices $\bar{\boldsymbol{\Phi}}_2^{(i,k)}$ for all indexes i and k. Finally, the projection kernel $\boldsymbol{\Phi}_2$ is obtained by m_2 rows from $\bar{\boldsymbol{\Phi}}_2$. More specifically, $\boldsymbol{\Phi}_2$ is obtained by picking $m_2^{(i,k)}$ rows from $\bar{\boldsymbol{\Phi}}_2^{(i,k)}$ and $\sum_{i,k} m_2^{(i,k)} = m_2$.

We will now determine the value of the rank $r_{\mathbf{z}}^{(i,k)}$ when $\boldsymbol{\Phi}_2$ is obtained by the construction previously described. On using (54) and (56), we can write

$$r_{\mathbf{z}}^{(i,k)} = r_{\mathbf{x}_1}^{(i,k)} - \Delta^{(i,k)}(\boldsymbol{\Phi}_2), \tag{77}$$

where

$$\Delta^{(i,k)}(\boldsymbol{\Phi}_2)$$
$$= \mathrm{rank}(\boldsymbol{\Phi}_2 \boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} \boldsymbol{\Phi}_2^{\mathrm{T}})$$
$$\quad - \mathrm{rank}(\boldsymbol{\Phi}_2 (\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger} \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}) \boldsymbol{\Phi}_2^{\mathrm{T}}) \tag{78}$$
$$= \dim(\mathrm{Im}(\boldsymbol{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger} \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}))$$
$$\quad - \dim(\mathrm{Im}(\boldsymbol{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})). \tag{79}$$

Then, by leveraging the expression of the dimension of the intersection of two linear spaces, we can write

$$\dim(\mathrm{Im}(\boldsymbol{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger} \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}))$$
$$= \mathrm{rank}(\boldsymbol{\Phi}_2) + \mathrm{rank}(\mathbf{N}_{1s}^{(i,k)}) - \mathrm{rank}[\boldsymbol{\Phi}_2^{\mathrm{T}} \mathbf{N}_{1s}^{(i,k)}] \tag{80}$$
$$= m_2 + n_2 - (r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)}) - \mathrm{rank}[\boldsymbol{\Phi}_2^{\mathrm{T}} \mathbf{N}_{1s}^{(i,k)}], \tag{81}$$

where the columns of $\mathbf{N}_{1s}^{(i,k)} \in \mathbb{R}^{n_2 \times (n_2 - (r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)}))}$ form a basis of the linear space $\mathrm{Null}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger} \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)})$. Consider now the last term in (81). We can note that

$$\mathrm{rank}[\boldsymbol{\Phi}_2^{\mathrm{T}} \mathbf{N}_{1s}^{(i,k)}] = \mathrm{rank}[(\hat{\boldsymbol{\Phi}}_2^{(i,k)})^{\mathrm{T}} \mathbf{N}_{1s}^{(i,k)}], \tag{82}$$

where $\hat{\boldsymbol{\Phi}}_2^{(i,k)}$ is obtained from $\boldsymbol{\Phi}_2$ by removing the $m_2^{(i,k)}$ rows which are also rows of the matrix $\bar{\boldsymbol{\Phi}}_2^{(i,k)}$, since all rows in $\bar{\boldsymbol{\Phi}}_2^{(i,k)}$ are contained in $\mathrm{Null}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger} \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)})$. Moreover, on leveraging the assumption that the linear spaces associated to the images of the covariance matrices of different Gaussian classes (i, k) are independently drawn from a continuous distribution over the Grassmann manifold, we also have

$$\mathrm{rank}[\boldsymbol{\Phi}_2^{\mathrm{T}} \mathbf{N}_{1s}^{(i,k)}] = \min\{n_2, m_2 - m_2^{(i,k)}$$
$$\quad + n_2 - (r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)})\}. \tag{83}$$

By following similar steps, we can also observe that

$$\dim(\mathrm{Im}(\boldsymbol{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})) \tag{84}$$
$$= m_2 + n_2 - r_{\mathbf{x}_2}^{(i,k)} - \mathrm{rank}[\boldsymbol{\Phi}_2^{\mathrm{T}} \mathbf{N}_2^{(i,k)}], \tag{85}$$

where the column of the matrix $\mathbf{N}_2^{(i,k)} \in \mathbb{R}^{n_2 \times (n_2 - r_{\mathbf{x}_2}^{(i,k)})}$ form a basis of $\mathrm{Null}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})$. Then, from the definition of $\bar{\boldsymbol{\Phi}}_2^{(i,k)}$ in (25), we can observe that the rows of $\bar{\boldsymbol{\Phi}}_2^{(i,k)}$ span a linear space which is contained in $\mathrm{Null}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger} \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)})$ but which

has zero-dimensional intersection with $\text{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)})$. Moreover, on leveraging the assumption that all the linear spaces associated to the images of the covariance matrices of different Gaussian classes (i, k) are randomly drawn from a continuous distribution over the Grassmann manifold, we also have

$$\text{rank}[\mathbf{\Phi}_2^{\mathrm{T}} \mathbf{N}_2^{(i,k)}] = \min\{n_2, m_2 + n_2 - r_{\mathbf{x}_2}^{(i,k)}\}. \tag{86}$$

Finally, on putting together (77), (79), (81), (83), (85), and (86), we can write

$$r_{\mathbf{z}}^{(i,k)} = \min\{r_{\mathbf{x}}^{(i,k)} - m_2, r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)}\}$$
$$+ \max\{m_2 - r_{\mathbf{x}_2}^{(i,k)}, 0\}. \tag{87}$$

Therefore, on considering separately the cases when $m_2 \leq r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)}$, $r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} < m_2 \leq r_{\mathbf{x}_2}^{(i,k)}$, and $m_2 > r_{\mathbf{x}_2}^{(i,k)}$, we have

$$r_{\mathbf{z}}^{(i,k)} =$$
$$\begin{cases} r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)} & , \text{if } m_2 \leq r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} \\ \min\{r_{\mathbf{x}}^{(i,k)} - m_2, \\ \quad r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)}\} & , \text{if } r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} < m_2 \leq r_{\mathbf{x}_2}^{(i,k)} \\ r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} & , \text{if } m_2 > r_{\mathbf{x}_2}^{(i,k)} \end{cases}, \tag{88}$$

which concludes the proof.

B. Proof of Corollary 4

Note that, given any value of m_2, the sufficient conditions for the MMSE to approach zero in the low-noise regime when both $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are random reported in [33, Theorem 4] can be also expressed as

$$m_1 > \begin{cases} r_{\mathbf{x}_1}^{(i,k)} & , \text{if } m_2 \leq r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} \\ r_{\mathbf{x}}^{(i,k)} - m_2 & , \text{if } r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} < m_2 \leq r_{\mathbf{x}_2}^{(i,k)} \\ r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} & , \text{if } m_2 > r_{\mathbf{x}_2}^{(i,k)} \end{cases}, \tag{89}$$

for all (i, k). Then, the proofs of Corollary 4 is obtained by comparing (89) with (17), and recalling that $m_2^{(i,k)} \geq 0, \forall (i, k)$.

REFERENCES

[1] M.-Y. Chen, F. Renna, and M. R. Rodrigues, "On the design of linear projections for compressive sensing with side information," in *Proc. IEEE Int. Symp. Theory*, 2016, pp. 670–674.

[2] E. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[6] 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, Feb. 2006.

[7] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[9] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hedge, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[10] T. Blumensath and M. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1872–1882, Apr. 2009.

[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, Dec. 2010.

[12] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, 2009.

[13] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

[14] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[15] Y. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[16] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[17] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 92–106, Jan. 2009.

[18] 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, May 2012.

[19] W. R. Carson, M. Chen, M. R. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM J. Imag. Sci.*, vol. 5, no. 4, pp. 1185–1212, 2012.

[20] M. Chen, W. Carson, M. Rodrigues, R. Calderbank, and L. Carin, "Communications inspired linear discriminant analysis," in *Proc. Int. Conf. Machine Learn.*, Jun.-Jul. 2012, pp. 919–926.

[21] J. Yang *et al.*, "Video compressive sensing using Gaussian mixture models," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4863–4878, Nov. 2014.

[22] F. Renna, R. Calderbank, L. Carin, and M. Rodrigues, "Reconstruction of signals drawn from a Gaussian mixture via noisy compressive measurements," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2265–2277, May 2014.

[23] C. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag, 2006.

[24] 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, Sep. 2010.

[25] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.,* May 2014, pp. 3330–3334.

[26] L. Weizman, Y. C. Eldar, and D. B. Bashat, "Compressed sensing for longitudinal MRI: An adaptive-weighted approach," *Med. Phys.*, vol. 42, no. 9, pp. 5195–5208, 2015.

[27] 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. Phys.*, vol. 35, no. 2, pp. 660–663, 2008.

[28] J. Mota, N. Deligiannis, and M. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4472–4496, Jul. 2017.

[29] J. Mota, N. Deligiannis, A. Sankaranarayanan, V. Cevher, and M. Rodrigues, "Adaptive-rate sparse signal reconstruction with application in compressive background subtraction," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3651–3666, Jul. 2016.

[30] J. Mota, L. Weizman, N. Deligiannis, Y. Eldar, and M. Rodrigues, "Reference-based compressed sensing: A sample complexity approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* Mar. 2016, pp. 4687–4691.

[31] X. Yuan, T.-H. Tsai, R. Zhu, P. Llull, D. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 964–976, Sep. 2015.

[32] F. Renna *et al.*, "Classification and reconstruction of compressed GMM signals with side information," in *Proc. IEEE Int. Symp. Inf. Theory* Jun. 2015, pp. 994–998.

[33] F. Renna *et al.*, "Classification and reconstruction of high-dimensional signals from low-dimensional features in the presence of side information," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6459–6492, Nov. 2016.

[34] M.-Y. Chen, F. Renna, and M. R. Rodrigues, "Signal reconstruction in the presence of side information: The impact of projection kernel design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 4618–4622.

[35] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," arXiv:0901.3403, 2009.

[36] M. Duarte, M. Wakin, D. Baron, S. Sarvotham, and R. Baraniuk, "Measurement bounds for sparse signal ensembles via graphical models," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4280–4289, Jul. 2013.

[37] P. Song, J. F. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Measurement matrix design for compressive sensing with side information at the encoder," in *Proc. IEEE Statist. Signal Process. Workshop*, 2016, pp. 1–5.

[38] N. Deligiannis, J. Mota, B. Cornelis, M. R. D. Rodrigues, and I. Daubechies, "Multi-modal dictionary learning for image separation with application in art investigation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 751–764, Feb. 2016.

[39] A. Rajwade, D. Kittle, T.-H. Tsai, D. Brady, and L. Carin, "Coded hyperspectral imaging and blind compressive sensing," *SIAM J. Image Sci.*, vol. 6, no. 2, pp. 782–812, 2013.

[40] J. Yang *et al.*, "Video compressive sensing using Gaussian mixture models," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4863–4878, Nov. 2014.

[41] L. Wang, D. E. Carlson, M. Rodrigues, D. Wilcox, R. Calderbank, and L. Carin, "Designed measurements for vector count data," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 1142–1150.

[42] W. Chen, M. R. Rodrigues, and I. Wassell, "Projection design for statistical compressive sensing: A tight frame based approach," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2016–2029, Apr. 2013.

[43] W. Chen and M. R. Rodrigues, "Dictionary learning with optimized projection design for compressive sensing applications," *IEEE Signal Process. Lett.*, vol. 20, no. 10, pp. 992–995, Oct. 2013.

[44] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[45] C. C. Paige and M. A. Saunders, "Towards a generalized singular value decomposition," *SIAM J. Numer. Anal.*, vol. 18, no. 3, pp. 398–405, 1981.

[46] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6241–6263, Oct. 2012.

[47] J. Zhu, D. Baron, and F. Krzakala, "Performance limits for noisy multimeasurement vector problems," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2444–2454, May 2017.

[48] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, Apr. 2006.

[49] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sep. 2013.

[50] C. Poynton, "The rehabilitation of gamma," in *Proc. SPIE Hum. Vis. Electron. Imag. III*, 1998, vol. 3299, pp. 232–249.

[51] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Appl. Opt.*, vol. 49, no. 36, pp. 6824–6833, 2010.

[52] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15, no. 8, pp. 14013–14027, 2007.

[53] A. Wagadarikar, N. Pitsianis, X. Sun, and D. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Opt. Express*, vol. 17, no. 8, pp. 6368–6388, 2009.

[54] P. Song, X. Deng, J. F. Mota, N. Deligiannis, P. L. Dragotti, and M. R. Rodrigues, "Multimodal image super-resolution via joint sparse representations induced by coupled dictionaries," arXiv:1709.08680, 2017.

[55] B.-Y. Wang, X. Zhang, and F. Zhang, "Some inequalities on generalized Schur complements," *Linear Algebra Appl.*, vol. 302, pp. 163–172, 1999.

[56] Y. Tian and G. P. Styan, "Rank equalities for idempotent matrices with applications," *J. Comput. Appl. Math.*, vol. 191, no. 1, pp. 77–97, 2006.

[57] D. V. Ouellette, "Schur complements and statistics," *Linear Algebra Appl.*, vol. 36, pp. 187–295, 1981.

[58] J. Gallier, "The Schur complement and symmetric positive semidefinite (and definite) matrices," vol. 44, pp. 1–12, Dec. 2010. [Online]. Available: ftp://158.130.67.137/pub/cis511/public_html/schur-comp.pdf

Meng-Yang Chen (S'16) received the B.S. degree in electronic engineering from Chung Yuan Christian University, Taoyuan City, Taiwan, in 2007, and the M.S. degree in electrical engineering from the National Sun Yat-Sen University, Kaohsiung, Taiwan, in 2009. Since 2014, he has been working toward the Ph.D. degree at the Department of Electronic and Electrical Engineering, University College London, London, U.K. Between 2009 and 2013, he was a Senior Software Engineer with the Department of Smart Grid, Delta Networks, Inc., Taipei, Taiwan, and the Department of Automotive Electronic Product, Compal Electronics, Inc., Taipei, Taiwan. His research interests include compressive sensing and signal processing.

Francesco Renna (S'09–M'11) received the Laurea Specialistica degree in telecommunication engineering and the Ph.D. degree in information engineering, both from the University of Padova, Padova, Italy, in 2006 and 2011, respectively. Between 2007 and 2017, he held Visiting Researcher and Postdoctoral appointments with the Infineon Technology AG, Princeton University, Georgia Institute of Technology (Lorraine Campus), Supélec, the University of Porto, Duke University, the University College London, and the University of Cambridge. Since 2017, he has been a Postdoctoral Researcher with the Instituto de Telecomunicações and University of Porto, Porto, Portugal. His research interests include high-dimensional information processing and biomedical signal processing.

Dr. Renna was the recipient of a Marie Skłodowska-Curie Individual Fellowship.

Miguel R. D. Rodrigues (SM'15) received the Licenciatura degree in electrical and computer engineering from the University of Porto, Porto, Portugal, and the Ph.D. degree in electronic and electrical engineering from the University College London, London, U.K. He was with the University of Porto rising through the ranks from an Assistant to an Associate Professor, where he was also the Head of the Information Theory and Communications Research Group, Instituto de Telecomunicações Porto. He is currently a Reader in information theory and processing with the Department of Electronic and Electrical Engineering, University College London.

From 2003 to 2016, he held Postdoctoral positions and Visiting appointments with various institutions worldwide, including the University College London, Cambridge University, Princeton University, and Duke University. He is coauthoring a book *Information-Theoretic Methods in Data Science* (Cambridge University Press). He has authored or coauthored more than 150 papers in the leading international journals and conferences. His research interests include the general areas of information theory and processing with a current focus on sensing, analysis, and processing of high-dimensional data.

Dr. Rodrigues is an Associate Editor for the IEEE COMMUNICATIONS LETTERS and a Lead Guest Editor for the Special Issue on Information-Theoretic Methods in Data Acquisition, Analysis, and Processing of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. He was on the organizing and technical committees of various international conferences, including as the Co-Chair of the Technical Program Committee of the IEEE Information Theory Workshop 2016. He was the recipient of the Prestigious IEEE Communications and Information Theory Societies Joint Paper Award 2011. He was also the recipient of fellowships from the Portuguese Foundation for Science and Technology and the Foundation Calouste Gulbenkian.