On the Design of Linear Projections for Compressive Sensing with Side Information

Meng-Yang Chen, Francesco Renna, Miguel R. D. Rodrigues
Department of Electronic and Electrical Engineering, University College London, U.K.
Email: {meng-yang.chen.14, f.renna, m.rodrigues}@ucl.ac.uk

Abstract—In this paper, we study the problem of projection kernel design for the reconstruction of high-dimensional signals from low-dimensional measurements in the presence of side information, assuming that the signal of interest and the side information signal are described by a joint Gaussian mixture model (GMM). In particular, we consider the case where the projection kernel for the signal of interest is random, whereas the projection kernel associated to the side information is designed. We then derive sufficient conditions on the number of measurements needed to guarantee that the minimum mean-squared error (MMSE) tends to zero in the low-noise regime. Our results demonstrate that the use of a designed kernel to capture side information can lead to substantial gains in relation to a random one, in terms of the number of linear projections required for reliable reconstruction.

Index Terms—Kernel design, side information, Gaussian mixture model (GMM), minimum mean squared error (MMSE)

I. INTRODUCTION

Recently, compressive sensing (CS) has been established as an efficient technique in the field of signal processing in order to perform data acquisition and reconstruction, without any or with minimal loss of information [1], [2]. In particular, CS provides nearly perfect reconstruction with less measurements than conventional schemes provided that the signal is sparse in some orthonormal dictionary or frame [1], [2]. Reconstruction is performed by using tractable ℓ_1 minimization methods [3] or iterative methods, like greedy matching pursuit [4], [5].

The use of additional information – commonly known as side information – can also be leveraged to enhance the performance of compressive sensing [6]–[10] (see also related work in the information-theoretic literature [11], [12]); this is due to the fact that possible correlations between the signal of interest and the side information can be efficiently exploited both during the acquisition phase and the reconstruction phase in order to boost performance. This opens up the possibility to enhance various applications, e.g., video streams are often accompanied with audio tracks, hyperspectral images can often be complemented with standard RGB versions, and high-resolution images can also be complemented with low-resolution counterparts [13].

A number of authors have proposed frameworks for compressive sensing with side information [6]–[10], [13]–[16]. In particular, in CS scenarios based on sparse models, side information is offered in the form of partial information on the support [6] or as a noisy version of the signal of interest [7]–[10]. Such approaches are shown to require less

measurements than traditional CS scheme to guarantee reliable reconstruction. More recently, the impact of side information has also been studied in conjunction with structured signal models, e.g., union-of-subspaces [13]–[15] and Bayesian dictionary learning [16], where it has been shown to improve reconstruction performance.

In particular, references [13]–[15] study conditions on the number of measurements for reliable reconstruction in the scenario where both the signal of interest and the side information obey a joint (correlated) Gaussian Mixture model (GMM). References [13], [14] consider the scenario where the measurements taken from the signal of interest and the side information are random, whereas [15] considers the scenario where the measurements taken from the signal of interest are optimally designed and the side information is not compressed. One of the insights offered by [13]–[15] relates to the fact that optimal acquisition of the signal of interest – via the design of the respective measurement kernel[1] – does not result in an improvement in the minimum number of measurements needed to drive the reconstruction error to zero in the low-noise regime.

These works then motivate the question:

Can we improve the reconstruction performance in compressive sensing with side information applications by designing the projection kernel associated to the side information?

This paper answers this question by considering sufficient and (occasionally) necessary conditions on the number of measurements for the reconstruction minimum mean-squared error (MMSE) to approach zero in the low-noise regime. It also provides the form of the kernel designs which are associated with such conditions.

The paper is organized as follows: Section II defines the signal and the system model. Section III provides the sufficient conditions (and necessary conditions in some cases) on the number of measurements for the reconstruction MMSE to approach zero in the low-noise regime, for the scenario where one designs the side information measurement kernel. Section IV presents a number of numerical results that illustrate how theory aligns with experiments. Conclusions are made in Section V.

We adopt the notation: boldface upper-case letters denote matrices ($\mathbf{X}$) and boldface lower-case letters denote column

---

[1]Throughout the paper, we will use the terms linear projections, kernels and measurements, interchangeably.
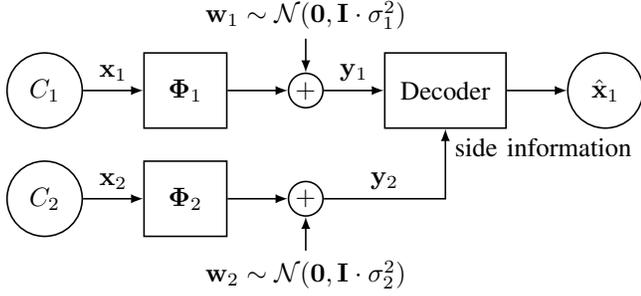
Fig. 1. Compressive sensing model in the presence of side information.

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). $(\cdot)^{\mathrm{T}}$, $\mathrm{rank}(\cdot)$ and $\mathbb{E}[\cdot]$ represent the transpose, rank, and expectation operators, respectively. The Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The Moore-Penrose pseudoinverse of a matrix is donated by $(\cdot)^\dagger$. $\mathrm{Im}(\cdot)$, $\mathrm{Null}(\cdot)$ and $\dim(\cdot)$ denote the (column) image, null space of a matrix and the dimension of a linear space, respectively.

For reasons of space, we relegate the mathematical proofs of our results to an extended version of this work [17].

II. MODEL

Fig. 1 depicts the compressive measurement model under consideration where $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ is the signal of interest and $\mathbf{x}_2 \in \mathbb{R}^{n_2}$ is the side information. Our objective is to reconstruct the signal of interest both from noisy linear measurements $\mathbf{y}_1 \in \mathbb{R}^{m_1}$, with $m_1 \leq n_1$, associated with $\mathbf{x}_1$ given by

$$\mathbf{y}_1 = \boldsymbol{\Phi}_1 \mathbf{x}_1 + \mathbf{w}_1, \tag{1}$$

and noisy linear measurements $\mathbf{y}_2 \in \mathbb{R}^{m_2}$, with $m_2 \leq n_2$, associated with $\mathbf{x}_2$ given by

$$\mathbf{y}_2 = \boldsymbol{\Phi}_2 \mathbf{x}_2 + \mathbf{w}_2, \tag{2}$$

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

We will be assuming that the signal of interest $\mathbf{x}_1$ and the side information signal $\mathbf{x}_2$ are described by a joint Gaussian mixture model. In particular, $\mathbf{x}_1$ and $\mathbf{x}_2$ are associated with underlying class labels $C_1 \in \{1, \ldots, K_1\}$ and $C_2 \in \{1, \ldots, K_2\}$, respectively. 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 underlying class labels $C_1 = i$ and $C_2 = k$ obey the joint probability density function (pdf):

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

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

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

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

Here,

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

where $\boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)}$ 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 to the images of the covariance matrices $\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}$, $\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ associated to different class labels $C_1 = i$ and $C_2 = k$ are independently drawn from a continuous pdf over the corresponding Grassmann manifold.[2]

The decoder produces an estimate $\hat{\mathbf{x}}_1$ of the signal $\mathbf{x}_1$ given the noisy linear measurements $\mathbf{y}_1$ and $\mathbf{y}_2$ by using the optimal conditional mean estimator given by:

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

where $p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2)$ is the *a posteriori* pdf of $\mathbf{x}_1$ given the linear measurements $\mathbf{y}_1$ and $\mathbf{y}_2$. Therefore, the reconstruction MMSE is given by:

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

which is a function of the distribution of $\mathrm{p}(\mathbf{x}_1, \mathbf{x}_2)$, the noise variances σ_1^2, σ_2^2 and projection kernels $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$.

Our goal is to determine sufficient conditions on the number of measurements taken both from the signal of interest and the side information so that the reconstruction error tends to zero in the low-noise regime, i.e.

$$\lim_{\sigma_1^2, \sigma_2^2 \to 0} \mathrm{MMSE}(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) = 0, \tag{8}$$

Such conditions will be shown both for the scenario where the measurement matrices $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ are obtained randomly[3] and for the scenario where the measurement matrix $\boldsymbol{\Phi}_1$ is

---

[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, digits classification, etc. [18]. This assumption will enable us to simplify the statement of some of our results.

[3]In this work, random $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ 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 $p(\boldsymbol{\Theta}\mathbf{M}) = p(\mathbf{M})$, or $p(\mathbf{M}\boldsymbol{\Psi}) = p(\mathbf{M})$, respectively, for any orthogonal matrix $\boldsymbol{\Theta}$ or $\boldsymbol{\Psi}$. 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.

random but the measurement matrix $\mathbf{\Phi}_2$ is optimal. In particular, the optimal measurement matrix $\mathbf{\Phi}_2^\star$ corresponds to the solution to the optimization problem

$$\underset{\mathbf{\Phi}_2}{\text{minimize}} \ \text{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2)$$
$$\text{subject to } \text{tr}(\mathbf{\Phi}_2\mathbf{\Phi}_2^\mathsf{T}) \leq m_2, \tag{9}$$

where the trace constraint in (9) limits the average energy corresponding to the projection kernel. We denote the MMSE associated with a random measurement matrix $\mathbf{\Phi}_1$ and an optimal measurement matrix $\mathbf{\Phi}_2^\star$ by $\text{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2^\star)$.

The challenge associated with the characterization of such sufficient conditions on the number of measurements for reliable reconstruction (i.e., such that (8) holds) is due to the fact that the (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 to the MMSE as a means to characterize conditions for reliable classification.

III. Necessary and sufficient conditions for reliable classification

We focus now on determining sufficient conditions (and necessary conditions in some cases) on the number of measurements for reliable classification (i.e., such that (8) holds). 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 simplified case when $\mathbf{x}_1$ and $\mathbf{x}_2$ are jointly Gaussian distributed with mean $\boldsymbol{\mu}_\mathbf{x}$ and covariance matrix $\mathbf{\Sigma}_\mathbf{x}$, i.e., when $K_1 = K_2 = 1$ in (4).

Theorem 1: Consider the measurement model in (1) and (2), where $\mathbf{x}_1$ and $\mathbf{x}_2$ obey 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 $r_{\mathbf{x}_2} = \text{rank}(\mathbf{\Sigma}_{\mathbf{x}_2})$. Consider also that the linear spaces associated to the images of the covariance matrices $\mathbf{\Sigma}_\mathbf{x}$, $\mathbf{\Sigma}_{\mathbf{x}_1}$ and $\mathbf{\Sigma}_{\mathbf{x}_2}$ are independently drawn from a continuous pdf over the corresponding Grassmann manifold. Assume that $\mathbf{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ is random, 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

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

Proof: See [17]. ∎

Theorem 1 provides 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. In particular, the sufficient conditions on the number of measurements for

reliable reconstruction follow from an upper bound to the MMSE associated with the use of a suboptimal design for $\mathbf{\Phi}_2$ in *lieu* of the optimal design $\mathbf{\Phi}_2^\star$ corresponding to the solutions of (9) (See [17]).

This suboptimal design is obtained from the generalized singular value decomposition (GSVD) [19] 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:

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

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

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

$$\mathbf{\Lambda_B} = \begin{matrix} r_{\mathbf{x}_1}+r_{\mathbf{x}_2}-r_\mathbf{x} \\ r_\mathbf{x}-r_{\mathbf{x}_1} \end{matrix} \begin{bmatrix} \overset{r_{\mathbf{x}_1}+r_{\mathbf{x}_2}-r_\mathbf{x}}{\mathbf{0}} & \overset{r_\mathbf{x}-r_{\mathbf{x}_1}}{\ } \\ & \mathbf{D_B} \end{bmatrix}. \tag{13}$$

Then, the suboptimal design is given by

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

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

In turn, the necessary conditions on the number of measurements for reliable reconstruction follow by proving that there exists no kernel design that can guarantee error free reconstruction in the low-noise regime with a lower number of measurements (See [17]).

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 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.

The result in Theorem 1 together with the result in [13, Theorem 3] also showcase the merit of an optimal measurement matrix $\mathbf{\Phi}_2^\star$ 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 $\mathcal{M}_\mathsf{G}^\mathsf{R}$ be the set of pairs (m_1, m_2) such that $\lim_{\sigma_1^2, \sigma_2^2 \to 0} \text{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = 0$ for random $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$, and let $\mathcal{M}_\mathsf{G}^\mathsf{D}$ be the set of pairs (m_1, m_2) such that $\lim_{\sigma_1^2, \sigma_2^2 \to 0} \text{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2^\star) = 0$ for random $\mathbf{\Phi}_1$ and the optimal $\mathbf{\Phi}_2^\star$. Then, it holds

$$\mathcal{M}_\mathsf{G}^\mathsf{R} \subseteq \mathcal{M}_\mathsf{G}^\mathsf{D}. \tag{15}$$

Proof: See [17]. ∎

B. GMM sources

We now consider the scenario where $\mathbf{x}_1$ and $\mathbf{x}_2$ obey a joint GMM. 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 covariance 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, on defining $s_{\mathbf{x}_1}^{(i,k)} = r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)}$, 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

$$m_1 >$$
$$\begin{cases} r_{\mathbf{x}_1}^{(i,k)} - m_2 & , \text{if } m_2 \leq s_{\mathbf{x}_1}^{(i,k)} \\ \min\{r_{\mathbf{x}}^{(i,k)} - m_2, r_{\mathbf{x}_1}^{(i,k)} - m_2\} & , \text{if } s_{\mathbf{x}_1}^{(i,k)} < m_2 \leq r_{\mathbf{x}_2}^{(i,k)} \\ r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} & , \text{if } m_2 > r_{\mathbf{x}_2}^{(i,k)} \end{cases}$$
$$\qquad (17)$$

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

Proof: See [17]. ∎

The sufficient conditions for reliable reconstruction for GMMs embodied in Theorem 3 are obtained by considering a specific (suboptimal) design for $\boldsymbol{\Phi}_2$ which is inspired by the projection design which achieves the necessary and sufficient conditions for reliable reconstruction for Gaussian sources. This suboptimal design is obtained from the generalized singular value decomposition (GSVD) associated 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:

$$\mathbf{A}^{(i,k)} = \mathbf{U}^{(i,k)}\mathbf{C}^{(i,k)}(\mathbf{X}^{(i,k)})^{\text{T}} \qquad (20)$$
$$\mathbf{B}^{(i,k)} = \mathbf{V}^{(i,k)}\mathbf{S}^{(i,k)}(\mathbf{X}^{(i,k)})^{\text{T}}, \qquad (21)$$

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

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

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

where the diagonal blocks in $\boldsymbol{\Lambda}_{\mathbf{A}}^{(i,k)}$ and $\boldsymbol{\Lambda}_{\mathbf{B}}^{(i,k)}$ are all square.

We then define the matrix

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

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 $\sum_{i,k} m_2^{(i,k)} = m_2$.[4]

In fact, the $m_2^{(i,k)}$ measurements picked from $\bar{\bar{\boldsymbol{\Phi}}}_2^{(i,k)}$ capture the portion of the linear space 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 $C_1 = i, C_2 = k$.

Via Theorem 3 and [13, 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 $\mathcal{M}_{\text{GMM}}^{\text{R}}$ 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 [13, Theorem 4] for random $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$, and let $\mathcal{M}_{\text{GMM}}^{\text{D}}$ 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

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

Proof: See [17]. ∎

We finally note that, for both cases of Gaussian and GMM sources, designed projection kernels $\boldsymbol{\Phi}_2$ 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 $\mathbf{x}_1$.

IV. NUMERICAL RESULTS

We now provide numerical results that showcase the value of using an optimal measurement matrix $\boldsymbol{\Phi}_2^{\star}$ in relation to a random one. We set $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

[4]It is also possible to tighten further the sufficient conditions in Theorem 3 by choosing the values of $m_2^{(i,k)}$ in order to minimize the value of the maximum among the right hand side of (17) for all (i,k), subject to the constraints $m_2^{(i,k)} \leq r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)} - r_{\mathbf{x}}^{(i,k)}$, $\forall (i,k)$ and $\sum_{i,k} m_2^{(i,k)} = m_2$.
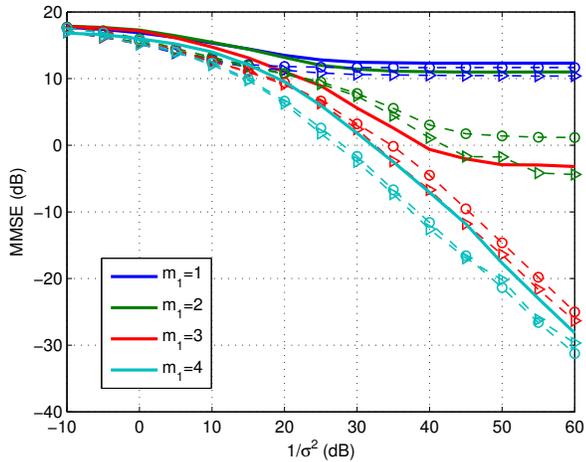
Fig. 2. MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3, 4$ and $m_2 = 2$ for joint GMM sources. Side information with random projection kernel (solid lines). Side information with designed projection kernel $\mathbf{\Phi}_2^\star$ (numerical solution) (triangles) and suboptimal design (based on GSVD) (circles).

We consider the case that $\mathbf{x}_1$ and $\mathbf{x}_2$ are drawn from a joint GMM, with dimensions $n_1 = 14$ and $n_2 = 6$, and $K_1 = K_2 = 2$. The means associated with the various Gaussian distributions are taken to be equal to zero, i.e., $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \mathbf{0}$, and the covariances of the Gaussian distributions are randomly generated such that $r_{\mathbf{x}}^{(i,k)} = 5$, $r_{\mathbf{x}_1}^{(i,k)} = 4$ and $r_{\mathbf{x}_2}^{(i,k)} = 3$, i.e., the images of the input covariance matrices associated to different classes are drawn uniformly at random from the corresponding Grassmann manifold. We fix $m_2 = 2$ and we consider three different cases: i) the case when $\mathbf{\Phi}_2$ is random; 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 (25).

Fig. 2 depicts the MMSE vs. $1/\sigma^2$. The results in [13] show that the MMSE approaches zero in the low-noise regime when $m_1 + m_2 > r_{\mathbf{x}}^{(i,k)} = 5$ for the case of random projection kernels. On the other hand, Theorem 3 shows that the MMSE approaches zero in the low-noise regime when $m_1 > 2$ and $m_2 = 2$. It is apparent that the number of projections that guarantee reliable reconstruction is less for scenarios where one uses designed measurements in comparison to scenarios where one uses random measurements.

Fig. 2 also shows that the numerical results are very well aligned with theoretical ones. In particular, it can be seen that the design in (25) requires the same number of projections as the optimal design in (9) in this case.

V. Conclusion

This work illustrates the impact on reconstruction performance of side information projection kernel design when both the signal of interest and side information are drawn from a joint Gaussian mixture model (GMM).

We have provided sufficient conditions on the number of linear projections that guarantee the reconstruction MMSE to approach zero in the low-noise regime. We have also provided the associated linear projection designs: these capture the linear space associated to components in the side information which are mostly correlated with the signal of interest.

It has been established that the optimization of the linear projections used to capture the signal of interest in the presence of side information does not lead to gains in relation to standard random linear projections, in terms of the number of measurements for the MMSE to approach zero in the low-noise. One surprising element of this current work relates to the fact that the acquisition of side information using appropriate linear projection designs can lead to further gains compared to random linear projections.

References

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

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

[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

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

[5] 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.

[6] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, 2010.

[7] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," *arXiv preprint arXiv:1311.0576*, 2013.

[8] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med.l Phys.*, vol. 35, no. 2, pp. 660–663, 2008.

[9] L. Weizman, Y. Eldar, and D. Bashat, "The application of compressed sensing for longitudinal MRI," *arXiv preprint arXiv:1407.2602*, 2014.

[10] J. Mota, N. Deligiannis, and M. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *arXiv preprint arXiv:1408.5250*, 2014.

[11] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, 1975.

[12] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[13] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M. Rodrigues, "Classification and reconstruction of high-dimensional signals from low-dimensional noisy features in the presence of side information," *arXiv preprint arXiv:1412.0614*, 2014.

[14] ——, "Classification and reconstruction of compressed GMM signals with side information," in *IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2015.

[15] M. Chen, F. Renna, and M. R. D. Rodrigues, "Signal reconstruction in the presence of side information: The impact of projection kernel design," Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Mar. 2016, http://www.ucl.ac.uk/~uceemc2/documents/icassp2016.pdf.

[16] 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, Sept 2015.

[17] M. Chen, F. Renna, and M. R. D. Rodrigues, "On the design of linear projections for compressive sensing with side information," 2016, http://www.ucl.ac.uk/~uceemc2/documents/isit2016.pdf.

[18] M. Chen, W. Carson, M. Rodrigues, R. Calderbank, and L. Carin, "Communications inspired linear discriminant analysis," in *Int. Conf. Machine Learn. (ICML)*, Jun.-Jul. 2012.

[19] 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.