

Reconstruction of Signals Drawn from a Gaussian Mixture via Noisy Compressive Measurements

Francesco Renna, *Member, IEEE*, Robert Calderbank, *Fellow, IEEE*, Lawrence Carin, *Fellow, IEEE*,
and Miguel R. D. Rodrigues, *Member, IEEE*

Abstract—This paper determines to within a single measurement the minimum number of measurements required to successfully reconstruct a signal drawn from a Gaussian mixture model in the low-noise regime. The method is to develop upper and lower bounds that are a function of the maximum dimension of the linear subspaces spanned by the Gaussian mixture components. The method not only reveals the existence or absence of a minimum mean-squared error (MMSE) error floor (phase transition) but also provides insight into the MMSE decay via multivariate generalizations of the MMSE dimension and the MMSE power offset, which are a function of the interaction between the geometrical properties of the kernel and the Gaussian mixture. These results apply not only to standard linear random Gaussian measurements but also to linear kernels that minimize the MMSE. It is shown that optimal kernels do not change the number of measurements associated with the MMSE phase transition, rather they affect the sensed power required to achieve a target MMSE in the low-noise regime. Overall, our bounds are tighter and sharper than standard bounds on the minimum number of measurements needed to recover sparse signals associated with a union of subspaces model, as they are not asymptotic in the signal dimension or signal sparsity.

Index Terms—Compressive sensing, Gaussian mixtures, reconstruction, classification, MMSE, MMSE decay, MMSE power offset, phase transition, kernel design

I. INTRODUCTION

The foundation of the digital revolution is the Shannon-Nyquist theorem, which provides a theoretical basis for digital processing of analog signals: it states that the sampling rate should be at least twice the Fourier bandwidth of the signal. The discrete-time representation of a continuous-time signal, which lies at the heart of analogue to digital conversion, offers the means to resilient data communication, storage and processing.

This paper was presented in part at the 2013 IEEE Global Conference on Signal and Information Processing.

The work of F. Renna was supported by Fundação para a Ciência e a Tecnologia through the research project CMU-PT/SIA/0026/2009. The work of M. R. D. Rodrigues was supported by the EPSRC through the research grant EP/K503459/1. The work of R. Calderbank and L. Carin was partially supported by the Defense Advanced Research Projects Agency (DARPA) under the KeCom program, and also by the Office of Naval Research (ONR) and by the Air Force Office of Scientific Research under the Complex Networks Program. This work was also supported by the Royal Society International Exchanges Scheme IE120996.

F. Renna is with the Instituto de Telecomunicações and the Departamento de Ciência de Computadores, Faculdade de Ciências da Universidade do Porto, Porto, Portugal (e-mail: frenna@dcc.fc.up.pt).

R. Calderbank and L. Carin are with the Department of Electrical and Computer Engineering, Duke University, Durham NC, USA (e-mail: {robert.calderbank, lcarin}@duke.edu).

M. R. D. Rodrigues is with the Department of E&EE, University College London, London, UK (email: m.rodrigues@ucl.ac.uk).

It has been recognized recently that the so-called Nyquist rate can be excessive in various emerging applications [1]–[3]: this – in addition to representing a burden to analog-to-digital converters [4], [5] – can also lead to a huge number of samples that compromise communications, storage and processing resources. Modern acquisition systems thus adopt a two-step approach that involves both an analog-to-digital conversion operation, whose purpose is to convert the information-bearing signal from the analogue to the digital domain, and a compression operation whose purpose is to offer succinct (near lossless) representations of the data.

There has been a recent emergence of a new sensing modality, emblematically known as Compressive Sensing (CS) [6]–[8], that offers the means to simultaneously sense and compress a signal without any loss of information (under appropriate conditions on the signal model and measurement process). The sensing process is based on the projection of the signal of interest onto a set of vectors, which can be either constituted randomly [6]–[10] or designed [11], [12], and the recovery process is based on the resolution of an inverse problem. It is well known that the reconstruction of an n -dimensional signal that admits an s -sparse representation in some orthonormal basis or frame, via ℓ_0 -pseudonorm minimization algorithms, requires only $s + 1$ noiseless measurements [13], [14]. However, there are no tractable algorithms able to solve such a minimization problem. On the other hand, ℓ_1 minimization methods [15] or iterative methods, like greedy matching pursuit [16]–[18], provide reliable reconstruction with overwhelming probability with only $\mathcal{O}(s \log(n/s))$ linear random measurements or projections [6], [8], [10].

However, even in early CS studies, it has been recognized that it is possible to derive better compression performance, in terms of the minimum number of measurements necessary to achieve perfect or nearly perfect reconstruction, by leveraging the fact that signals often obey models with additional structure beyond conventional sparsity. Some popular models that capture such additional structure include the union of subspaces [19]–[22], wavelet trees [19], [23] or manifolds [24], [25]. Within the union of subspaces model, the source signal is assumed to belong to one out of a collection of K subspaces with dimension less than or equal to s . Recovery of a signal in a union of subspaces is equivalent to the reconstruction of a block-sparse signal, when the individual subspaces in the union of subspaces model are decomposable as the direct sum of a given number of lower dimensional subspaces [21]. The minimum number of measurements required for reliable reconstruction in such scenarios has been shown to be of

the order $\mathcal{O}(s + \log(2K))$ [19] when using mixed ℓ_2/ℓ_1 -norm minimization [21]. On the other hand, tree models, where the non-zero coefficients of the source signal are known to be gathered into a rooted, connected, tree structure, can describe the most relevant wavelet coefficients of piecewise smooth signals or images [26]. In this case, the number of measurements required for reliable reconstruction is of order $\mathcal{O}(s)$ [23] by using a model-based version of the CoSaMP algorithm [27]. Finally, in the case of Riemannian manifolds, the minimum number of random projection measurements needed for reliable reconstruction has been derived to be of the order $\mathcal{O}(s \log(nVR\tau^{-1}))$, where s is the dimension of the manifold, τ^{-1} being its condition number, V being its volume and R being its geodesic covering regularity [24].

Another very useful structured model is the Gaussian mixture model (GMM) [25], [28]–[30]. GMMs are typically used in conjunction with the Bayesian CS formalism that entails the use of statistical descriptions of the source and a statistical description of the measurement system in order to perform reconstruction [31]. One important feature of these models relates to the existence of efficient and optimal inversion procedures, which can be expressed analytically in closed form [25]. The other important feature – in addition to the approximation of any distribution with arbitrary precision [32] – relates to the fact that these models have also been shown to provide state-of-the-art results in various practical problems [29]. These include problems in image processing such as interpolation, zooming, deblurring [28]–[30] and dictionary learning [25].

A GMM also relates to the other well-known structured models in the literature [19]–[22], [24], [25]. For example, the GMM can be seen as a Bayesian counterpart of the union of subspaces model (assuming each GMM mixture component has a near-low-rank covariance matrix). In fact, a signal drawn from a GMM lies in a union of subspaces, where each subspace corresponds to the image of each class-conditioned covariance matrix in the model¹. In addition, a low-rank GMM can also be seen as an approximation to a compact manifold. Compact manifolds can be covered by a finite collection of topological disks that can be represented by high-probability ellipsoids living on the principal hyperplanes corresponding to the different components of a low-rank GMM [25]. However, we emphasize that adopting a GMM *in lieu* of the other structured models has a specific advantage. Reconstruction of a signal drawn from a GMM from compressive linear measurements in Gaussian noise can be very effectively performed via a closed-form inversion formula [25].

As such, and also in view of its practical relevance, this paper studies in detail the behavior of the minimum mean-squared error (MMSE) associated with the reconstruction of a signal drawn from a GMM, based on a set of linear and noisy compressive measurements. We consider the asymptotic regime of low-noise, which is relevant in various signal and image processing scenarios [33], [34]. The emphasis is to understand, as a function of the properties of the linear measurement kernel and the Gaussian mixture, whether the

MMSE converges or does not converge to zero as the noise power converges to zero, i.e. the MMSE phase transition. The main contributions are:

- A bound on the number of linear random measurements that are necessary to reconstruct perfectly a signal drawn from a GMM in the low-noise regime;
- A bound on the number of linear random measurements that are sufficient to reconstruct perfectly a signal drawn from a GMM in the low-noise regime, by analyzing the MMSE performance of a (sub-optimal) classification and reconstruction strategy;
- Generalization of the bounds on the number of measurements that are necessary and/or sufficient to reconstruct a signal drawn from a GMM based on a set of noisy compressive measurements, considering the scenario where the linear measurement kernel is constituted randomly and then extended to the case for which the linear measurement kernel is *designed* to minimize the mean-squared error;
- Characterization, whenever possible, of a more refined behavior of the low-noise asymptotics of the MMSE, that portray the existence or absence of an MMSE error floor (the phase transition) as well as the MMSE decay, as a function of the geometry of the kernel and the geometry of the Gaussian mixture.

Overall, this contribution offers an analysis of the reconstruction performance and associated phase transitions that, and in contrast with other results in the CS literature (e.g. [6], [9], [19], [21]–[24], [35]–[40]), is non-asymptotic in the signal dimension or the signal sparsity. Recent works have also proposed the use of message passing and belief propagation methods to increase the speed of reconstruction algorithms [41]–[45]. However, these approaches have also been studied under the large-system assumption. In particular, message passing methods are proved to be computationally efficient in large-scale applications, while guaranteeing reliable reconstruction with a number of measurements n the same order as ℓ_1 -norm minimization [41], [43]. In fact, message passing algorithms are shown to be equivalent to MMSE estimation in the asymptotic large-system limit, when the projection matrix is sparse and the inputs are independent identically distributed (i.i.d.) with arbitrary distribution [42].

On the other hand, the analysis proposed in this paper is based on models that naturally incorporate memory rather than memoryless models, as opposed to previous contributions in the literature on the information-theoretic characterization of CS [38]–[40]. Memoryless models, however, have been characterized in terms of the reconstruction MMSE in the large-system limit. In particular, [46] shows that, in the large-system limit, the overall mean-squared error (MSE) can be decoupled into the MSE relative to the reconstruction of the individual elements of the sparse signal, and such value admits a single-letter characterization. This analysis is justified on the basis of a heuristic method from statistical physics, the *replica method*, but it can be proved rigorously for the case of sparse measurement matrices. The replica method has been shown to provide bounds that are in agreement with the exact

¹More generally, a signal drawn from a GMM model lies in a union of affine spaces rather than linear subspaces, where each affine space is associated with the mean and covariance of each class in the GMM model.

analysis for the case of sparsity-pattern recovery in the low-noise regime [47], and it has been used also to evaluate the MSE associated to the maximum *a posteriori* (MAP) estimator in the large system limit [48].

The remainder of the paper is organized as follows: Section II introduces the system model and the main performance quantities and definitions. The impact of a random linear measurement kernel on the behavior of the MMSE and its phase transition is investigated first for a signal drawn from a Gaussian distribution in Section III, to develop essential intuition; we then consider a signal drawn from a mixture of Gaussian distributions in Section IV. The impact on the phase transition of designing the measurement kernel is then illustrated in Section V. Section VI exhibits various numerical results both with synthetic and real data, illustrating the main operational features of the problem. Section VII summarizes the main contributions. For completeness, Appendix A collects a number of useful Lemmas that are relevant for the proofs of the main results reported in Appendices B and C.

We use the following notation: boldface upper-case letters denote matrices (\mathbf{X}) and boldface lower-case letters denote column vectors (\mathbf{x}); the context defines whether the quantities are deterministic or random. 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 whenever the dimensions are clear from the context). The expressions $(\cdot)^\dagger$, $\text{tr}(\cdot)$, $\text{rank}(\cdot)$ represent the transpose, trace and the rank operators, respectively. $\text{Im}(\cdot)$ and $\text{Null}(\cdot)$ denote the (column) image and null space of a matrix, respectively, $(\cdot)^\perp$ denotes the orthogonal complement of a linear subspace, and $\text{dim}(\cdot)$ denotes the dimension of a linear subspace. $\mathbb{E}[\cdot]$ represents the expectation operator. The Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

II. SYSTEM MODEL

We study the problem associated with the reconstruction of a source signal $\mathbf{x} \in \mathbb{R}^n$ from a set of $\ell < n$ noisy, linear projections $\mathbf{y} \in \mathbb{R}^\ell$ where

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

and $\boldsymbol{\Phi} \in \mathbb{R}^{\ell \times n}$ is the linear measurement kernel² and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \cdot \mathbf{I}_\ell)$ is zero-mean, additive white Gaussian noise (AWGN). We consider in Sections III and IV random measurement kernel designs, where the entries of $\boldsymbol{\Phi}$ are drawn i.i.d. from a zero-mean, fixed-variance, Gaussian distribution, which is common in the CS literature [6], [8]. However, we also consider in Section V design of the measurement kernel $\boldsymbol{\Phi}$, that aims to minimize the reconstruction error.

In this work we also concentrate on two particular distributions for the source vector: a Gaussian distribution and a GMM distribution (of course, the former is a special case of the latter). For a Gaussian source vector, $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\Sigma}_\mathbf{x})$ where $\boldsymbol{\mu}_\mathbf{x}$ represents the mean and $\boldsymbol{\Sigma}_\mathbf{x}$ represents the (possibly

rank deficient) covariance matrix. We denote the eigenvalue decomposition of the positive semidefinite covariance matrix

$$\boldsymbol{\Sigma}_\mathbf{x} = \mathbf{U}_\mathbf{x} \boldsymbol{\Lambda}_\mathbf{x} \mathbf{U}_\mathbf{x}^\dagger = \mathbf{U}_\mathbf{x} \text{diag}(\lambda_{x_1}, \dots, \lambda_{x_s}, 0, \dots, 0) \mathbf{U}_\mathbf{x}^\dagger, \quad (2)$$

where the orthogonal matrix $\mathbf{U}_\mathbf{x}$ contains the eigenvectors of $\boldsymbol{\Sigma}_\mathbf{x}$, the diagonal matrix $\boldsymbol{\Lambda}_\mathbf{x} = \text{diag}(\lambda_{x_1}, \dots, \lambda_{x_s}, 0, \dots, 0)$ contains the eigenvalues of $\boldsymbol{\Sigma}_\mathbf{x}$, $\lambda_{x_1} \geq \dots \geq \lambda_{x_s} > 0$, and $s = \text{rank}(\boldsymbol{\Sigma}_\mathbf{x})$ represents the rank of $\boldsymbol{\Sigma}_\mathbf{x}$.

For a GMM source, $\mathbf{x} \sim \sum_k p_k \mathcal{N}(\boldsymbol{\mu}_\mathbf{x}^{(k)}, \boldsymbol{\Sigma}_\mathbf{x}^{(k)})$, so that the source vector is assumed to be drawn from one out of K different classes with probability p_k , $k = 1, \dots, K$ where the distribution of the source vector conditioned on the class k is Gaussian with mean $\boldsymbol{\mu}_\mathbf{x}^{(k)}$ and (possibly rank deficient) covariance matrix $\boldsymbol{\Sigma}_\mathbf{x}^{(k)}$. We let $s_k = \text{rank}(\boldsymbol{\Sigma}_\mathbf{x}^{(k)})$, $s_{km} = \text{rank}(\boldsymbol{\Sigma}_\mathbf{x}^{(k)} + \boldsymbol{\Sigma}_\mathbf{x}^{(m)})$ and $s_{\max} = \max_k s_k$.

A low-rank modeling approach, where $s_k < n$ for some or all k , is the basis of the theory developed in Sections III, IV and V. This implies that a realization of the source signal lies on one out of the K affine subspaces associated with the translation by the mean vector $\boldsymbol{\mu}_\mathbf{x}^{(k)}$ of the subspaces corresponding to the images of the class-conditioned covariance matrices $\boldsymbol{\Sigma}_\mathbf{x}^{(k)}$. Therefore, a signal drawn from a GMM model can also be seen to lie on a union of subspaces (or affine subspaces) [21]. However, the fact that natural signals and images are not always exactly low-rank but rather “approximately” low-rank is also discussed in the sequel (see Section VI-B).

We use the MMSE to assess the level of distortion incurred in the reconstruction of the original source vector \mathbf{x} from the projections vector \mathbf{y} in the CS model in (1), which is given by:

$$\text{MMSE} = \mathbb{E} [\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\|^2], \quad (3)$$

where

$$\hat{\mathbf{x}}(\mathbf{y}) = \mathbb{E} [\mathbf{x} | \mathbf{y}]. \quad (4)$$

We focus on the characterization of the behavior of the MMSE in the low-noise regime, i.e., for $\sigma^2 \rightarrow 0$, which represents the regime with most operational relevance in many signal and image processing applications (the noise magnitude typically considered for these applications is $\sigma^2 = -60$ dB [33], [34]). This includes the characterization of an asymptotic expansion of the MMSE as $\sigma^2 \rightarrow 0$ together with the characterization of the phase transition of the MMSE as $\sigma^2 \rightarrow 0$. The MMSE phase transition, in line with other results in the literature [37], [38], corresponds to the minimum number of measurements that guarantee perfect reconstruction in the low-noise regime.

III. GAUSSIAN SOURCES

We first consider the characterization of the MMSE phase transition and a low-noise MMSE expansion associated with a linearly and compressively measured signal drawn from a Gaussian source. Such characterizations, which can be crisply expressed in terms of the geometry of the measurement kernel, the geometry of the source and their interplay, pave the way to the characterization of the MMSE phase transition associated with GMM sources.

²Throughout the paper, we will refer to $\boldsymbol{\Phi}$ as the sensing matrix, measurement matrix and kernel, interchangeably.

For a Gaussian source with mean $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{\Sigma}_x$, in the presence of zero-mean, additive Gaussian noise with covariance matrix $\sigma^2 \mathbf{I}$, the conditional mean estimator can be expressed as follows [49]:

$$\hat{\mathbf{x}}(\mathbf{y}) = \mathcal{W}(\mathbf{y}) = \boldsymbol{\mu}_x + \mathbf{W}(\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\mu}_x), \quad (5)$$

where $\mathbf{W} = \boldsymbol{\Sigma}_x \boldsymbol{\Phi}^\dagger (\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \boldsymbol{\Sigma}_x \boldsymbol{\Phi}^\dagger)^{-1}$ corresponds to the Wiener filter associated with a Gaussian source with mean zero and covariance $\boldsymbol{\Sigma}_x$, and the MMSE can also be expressed in closed form as follows [49]:

$$\text{MMSE}^G(\sigma^2) = \text{tr} \left(\boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x \boldsymbol{\Phi}^\dagger (\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \boldsymbol{\Sigma}_x \boldsymbol{\Phi}^\dagger)^{-1} \boldsymbol{\Phi} \boldsymbol{\Sigma}_x \right), \quad (6)$$

where we explicitly highlight the dependence of the MMSE on the measurement noise variance σ^2 , assuming a fixed Gaussian source with covariance $\boldsymbol{\Sigma}_x$. The following theorem now reveals the asymptotic behavior of the MMSE in the low-noise regime. We define

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_x^{\frac{1}{2}} \boldsymbol{\Phi}^\dagger \boldsymbol{\Phi} \boldsymbol{\Sigma}_x^{\frac{1}{2}} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\dagger, \quad (7)$$

where $\boldsymbol{\Sigma}_x^{\frac{1}{2}}$ is the (positive semidefinite) matrix square root of $\boldsymbol{\Sigma}_x$, \mathbf{U} is an orthogonal matrix that contains the eigenvectors of $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{\ell'}}, 0, \dots, 0)$ is a diagonal matrix that contains the eigenvalues of $\boldsymbol{\Sigma}$, $\lambda_1 \geq \dots \geq \lambda_{\ell'} > 0$ and $\ell' = \text{rank}(\boldsymbol{\Sigma})$. Note also that, when the entries of the measurement kernel $\boldsymbol{\Phi}$ are drawn i.i.d. from a zero-mean, fixed-variance, Gaussian distribution, the rank of $\boldsymbol{\Sigma}$ is equal to $\ell' = \text{rank}(\boldsymbol{\Sigma}) = \min\{s, \ell\}$, with probability 1.

Theorem 1: Consider the linear measurement model in (1) where $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$, $s = \text{rank}(\boldsymbol{\Sigma}_x)$ and ℓ is the number of measurements. The first-order, low-noise expansion of the MMSE is given by:

$$\text{MMSE}^G(\sigma^2) = \mathcal{M}_\infty^G + \mathcal{D}^G \cdot \sigma^2 + o(\sigma^2). \quad (8)$$

The zero-order term in the expansion, which relates to the MMSE floor, is given by:

$$\mathcal{M}_\infty^G = \lim_{\sigma^2 \rightarrow 0} \text{MMSE}^G(\sigma^2) = \sum_{i=\ell'+1}^s \mathbf{u}_i^\dagger \boldsymbol{\Sigma}_x \mathbf{u}_i, \quad (9)$$

where the vectors $\mathbf{u}_{\ell'+1}, \dots, \mathbf{u}_s$ form an orthonormal basis of the linear subspace $\text{Null}(\boldsymbol{\Sigma}) \cap \text{Null}(\boldsymbol{\Sigma}_x)^\perp$, and the coefficient of the first-order term in the expansion is given by:

$$\mathcal{D}^G = \sum_{i=1}^{\ell'} \frac{1}{\lambda_i} \mathbf{u}_i^\dagger \boldsymbol{\Sigma}_x \mathbf{u}_i, \quad (10)$$

where $\mathbf{u}_1, \dots, \mathbf{u}_{\ell'}$ are eigenvectors of $\boldsymbol{\Sigma}$ corresponding to the positive eigenvalues $\lambda_1, \dots, \lambda_{\ell'}$.

Proof: The proof of this Theorem is provided in Appendix B. ■

The value of the zero-order term is clearly zero if $\ell \geq s$ and it is non-zero otherwise with its value dictated by the interaction of the geometry of the source with the geometry of the measurement kernel, as portrayed via (9) and (7). The value of the coefficient of the first-order term is non-zero with its value depending as well on the interaction of the description of the source and of the kernel.

Theorem 1 and (8), (9) and (10) then lead to the conclusions:

- When $\ell < s$ the number of measurements is not sufficient to capture the full range of the source, so that the reconstruction is not perfect ($\mathcal{M}_\infty^G \neq 0$). On the other hand, when $\ell \geq s$ such a number of measurements capture completely the source information leading to perfect reconstruction ($\mathcal{M}_\infty^G = 0$) in the low-noise regime. That is, one requires the number of linear random measurements to be greater than or equal to the dimension of the subspace spanned by the source for the phase transition to occur.
- When $\ell \geq s$, the rate of decay of the MMSE is $\mathcal{O}(\sigma^2)$ as $\sigma^2 \rightarrow 0$, as in the scalar case [50]. On the other hand, the power offset of the MMSE on a $10 \cdot \log_{10} \frac{1}{\sigma^2}$ -scale is dictated by the quantity $10 \cdot \log_{10} \mathcal{D}^G$. In fact, the quantity \mathcal{D}^G , which represents the multivariate Gaussian counterpart of the MMSE dimension put forth in [50], distinguishes MMSE expansions associated with different realizations of the measurement kernel and different source covariances.
- Of particular interest, the presence or absence of a MMSE floor depends only on the relation between the number of measurements ℓ and the rank of the source covariance s . On the other hand, the exact value of the MMSE floor (when $\ell < s$) and the MMSE power offset (when $\ell \geq s$) depends on the relation between the geometry of the random measurement kernel and the geometry of the source.

IV. GMM SOURCES

We are now ready to consider the characterization of MMSE phase transitions associated with a linearly and compressively measured signal drawn from a GMM source. In particular, and in view of the lack of a closed-form tractable MMSE expression, we derive necessary and sufficient conditions for the phase transitions to occur via bounds to the MMSE for GMM sources, which will be denoted by the symbol $\text{MMSE}^{\text{GM}}(\sigma^2)$.

A. Necessary condition

The necessary condition on the number of random linear measurements (components of $\boldsymbol{\Phi}$ constituted at random) for the MMSE phase transition to occur is based on the analysis of a lower bound to the MMSE. We express the lower bound in terms of the Gaussian MMSE associated to the actual class from which each signal realization is drawn. Namely, we have

$$\text{MMSE}^{\text{GM}}(\sigma^2) = \mathbb{E} [\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\|^2] \quad (11)$$

$$= \sum_{k=1}^K p_k \mathbb{E} [\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\|^2 | c = k] \quad (12)$$

$$\geq \sum_k p_k \text{MMSE}_k^G(\sigma^2) = \text{MSE}_{\text{LB}}(\sigma^2), \quad (13)$$

where $\text{MMSE}_k^G(\sigma^2)$ denotes the MMSE associated with the reconstruction of Gaussian signals \mathbf{x} in class $c = k$ from the measurement vector \mathbf{y} . Note that the equality in (12) is due to the total probability formula and the inequality in (13) follows

from the optimality of the MMSE estimator (5) for a single Gaussian source.

Via the analysis of $\text{MSE}_{\text{LB}}(\sigma^2)$, we obtain immediately the following necessary condition on the number of random linear measurements for the true MMSE to approach zero.

Theorem 2: Consider the linear measurement model in (1) where $\mathbf{x} \sim \sum_k p_k \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(k)}, \boldsymbol{\Sigma}_{\mathbf{x}}^{(k)})$, $s_{\text{max}} = \max_k s_k$, with $s_k = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}}^{(k)})$, and ℓ is the number of measurements. Then, with probability 1, it follows that:

$$\lim_{\sigma^2 \rightarrow 0} \text{MMSE}^{\text{GM}}(\sigma^2) = 0 \Rightarrow \ell \geq s_{\text{max}}. \quad (14)$$

Proof: It is evident that if $\text{MMSE}_k^{\text{G}}(\sigma^2) \rightarrow 0$ as $\sigma^2 \rightarrow 0$, for all k , then $\text{MSE}_{\text{LB}}(\sigma^2) \rightarrow 0$ as $\sigma^2 \rightarrow 0$. Theorem 1 proves that a necessary and sufficient condition for $\lim_{\sigma^2 \rightarrow 0} \text{MMSE}_k^{\text{G}}(\sigma^2) = 0$ to hold with probability 1 is that $\ell \geq s_k$; therefore, a necessary condition for $\lim_{\sigma^2 \rightarrow 0} \text{MMSE}^{\text{GM}}(\sigma^2) = 0$ to hold with probability 1 is that $\ell \geq s_{\text{max}} = \max_k s_k$. ■

An immediate corollary is the following first-order, low-noise expansion for the lower-bound of the MMSE for GMM inputs:

$$\text{MSE}_{\text{LB}}(\sigma^2) = \sum_k p_k \mathcal{M}_{\infty_k}^{\text{G}} + \left(\sum_k p_k \mathcal{D}_k^{\text{G}} \right) \sigma^2 + o(\sigma^2), \quad (15)$$

where $\mathcal{M}_{\infty_k}^{\text{G}}$ and \mathcal{D}_k^{G} are the zero and first-order terms of the expansion of the Gaussian MMSE corresponding to class k .

B. Sufficient condition

The sufficient condition on the number of random linear measurements for the MMSE phase transition to occur is based instead on the analysis of an upper bound to the MMSE. We construct such an MMSE upper bound – denoted as $\text{MSE}_{\text{CR}}(\sigma^2)$ – by using a (sub-optimal) *classify and reconstruct* procedure³:

- 1) First, we obtain an estimate of the signal class by the MAP classifier as follows:

$$\hat{c} = \arg \max_k p(\mathbf{y}|c=k)p_k, \quad (16)$$

where the variable \hat{c} represents the estimate of the signal class, the random variable c represents the actual signal class and $p(\mathbf{y}|c=k)$ denotes the conditioned probability density function (pdf) of the measurement vector \mathbf{y} given the signal class k ;

- 2) Then, we reconstruct the source vector \mathbf{x} from the measurement vector \mathbf{y} by using the conditional mean estimator associated with the class estimate \hat{c} as follows:

$$\hat{\mathbf{x}}(\mathbf{y}, c = \hat{c}) = \mathcal{W}_{\hat{c}}(\mathbf{y}) = \boldsymbol{\mu}_{\mathbf{x}}^{(\hat{c})} + \mathbf{W}_{\hat{c}}(\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\mu}_{\mathbf{x}}^{(\hat{c})}). \quad (17)$$

Note that a similar approach has been shown to offer state-of-the-art performance in the reconstruction of signals drawn from GMM sources from compressive measurements [29], [30].

³Note that the classify and reconstruct procedure is presented here as a mathematical tool to determine an upper bound to the number of measurements needed to obtain perfect reconstruction when using the optimal conditional mean estimator.

The optimality of the conditional mean estimator together with the (in general) sub-optimality of the classify and reconstruction approach leads immediately to the fact that:

$$\text{MMSE}^{\text{GM}}(\sigma^2) \leq \text{MSE}_{\text{CR}}(\sigma^2). \quad (18)$$

The analysis of the *classify and reconstruct* MMSE, which is aided by recent results on the characterization of the performance of GMM classification problems from noisy compressive measurements [51], then leads to the following sufficient condition on the number of random linear measurements for the true MMSE to approach zero.

Theorem 3: Consider the linear measurement model in (1) where $\mathbf{x} \sim \sum_k p_k \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(k)}, \boldsymbol{\Sigma}_{\mathbf{x}}^{(k)})$, $s_{\text{max}} = \max_k s_k$, with $s_k = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}}^{(k)})$, and ℓ is the number of measurements. Then, with probability 1, it holds:

$$\ell > s_{\text{max}} \Rightarrow \lim_{\sigma^2 \rightarrow 0} \text{MMSE}^{\text{GM}}(\sigma^2) = 0. \quad (19)$$

Proof: The proof of this Theorem is provided in Appendix C. ■

C. The MMSE phase transition

The characterization of the MMSE phase transition follows by combining the results encapsulated in Theorems 2 and 3. In particular, it is possible to construct a sharp characterization of the transition that is accurate within one measurement where:

- When $\ell < s_{\text{max}}$, the function $\text{MMSE}^{\text{GM}}(\sigma^2)$ converges to an error floor as $\sigma^2 \rightarrow 0$;
- When $\ell > s_{\text{max}}$, the function $\text{MMSE}^{\text{GM}}(\sigma^2)$ converges to zero as $\sigma^2 \rightarrow 0$;
- When $\ell = s_{\text{max}}$, the function $\text{MMSE}^{\text{GM}}(\sigma^2)$ may or may not approach zero as $\sigma^2 \rightarrow 0$, depending on the exact class dependent source covariances (see Section VI).

Note that – akin to the Gaussian result – one requires the number of linear random measurements to be greater than the largest of the dimensions of the subspaces spanned by the class dependent source covariances for the phase transition to occur. This is due to the fact that – as reported in Appendix C – with such a number of measurements one is able to classify perfectly and thereby to reconstruct perfectly in the low-noise regime. Note also that the *classify and reconstruct* procedure is nearly “phase transition” optimal: the number of measurements required by such a procedure differs at most by one measurement from the number of measurements required by the optimal conditional mean estimation strategy. This also provides a rationale for the state-of-the-art results reported in [29], [30], which are based on the use of the class conditioned Wiener filters for reconstruction and the detection of the *a posteriori* most probable class of the signal.

V. FROM RANDOM TO DESIGNED KERNELS

The emphasis of Sections III and IV has been on the derivation of necessary and sufficient conditions on the number of linear random measurements for the MMSE phase transition to occur. However, in view of recent interest on the design of linear measurements in the literature [11], [12], [33], [34],

[52], it is also natural to ask whether designed kernels have an impact on such bounds.

In particular, we seek to characterize the impact on the MMSE phase transition of kernels designed via the following optimization problem:

$$\begin{aligned} & \underset{\Phi}{\text{minimize}} && \text{MMSE}(\sigma^2, \Phi) \\ & \text{subject to} && \text{tr}(\Phi\Phi^\dagger) \leq \ell \end{aligned} \quad (20)$$

where the constraint guarantees that, on average, the rows of the designed kernel have unit ℓ_2 -norm. Note that, for the sake of clarity, we now express explicitly that the MMSE is a function of the linear measurement kernel.

We start by showcasing the optimal linear kernel design for a Gaussian source, where ‘‘optimality’’ is defined by the optimization problem in (20), which follows immediately by leveraging results on the joint optimization of transmitter and receiver for coherent multiple input multiple output (MIMO) transmission [53].

Theorem 4: Consider the linear measurement model in (1) where $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}})$, $s = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}})$ and ℓ is the number of measurements. Then, the measurement kernel Φ^* that solves the optimization problem in (20) can be expressed as follows:

$$\Phi^* = \left[\text{diag} \left(\sqrt{\lambda_{\Phi,1}^*}, \dots, \sqrt{\lambda_{\Phi,\ell}^*} \right) \mathbf{0}_{\ell \times (n-\ell)} \right] \mathbf{U}_{\mathbf{x}}^\dagger, \quad (21)$$

where the squared singular values of Φ^* are obtained through the water-filling principle [54] as follows

$$\lambda_{\Phi,i}^* = \left[\eta - \frac{\sigma^2}{\lambda_{\mathbf{x},i}} \right]^+, \quad (22)$$

and $\eta > 0$ is such that $\sum_{i=1}^{\ell} \lambda_{\Phi,i}^* \leq \ell$ and $[x]^+ = \max\{x, 0\}$.

Proof: The problem of finding the projection matrix which minimizes the reconstruction MMSE for Gaussian input signals can be mapped, with appropriate modifications, to the problem of finding the linear precoder which minimizes the MMSE of a MIMO transmission system. In particular, the object function of this minimization problem is a Schur-concave function of the MMSE matrix and the optimal linear precoder is shown to diagonalize the MIMO channel matrix [53, Theorem 1]. This implies that, in our scenario, the designed kernel right singular vectors correspond to the source covariance eigenvectors, i.e., that the measurement kernel that minimizes the MMSE exposes the modes of the source covariance⁴. Then, the fact that the squared singular values of Φ^* obey the water-filling type of interpretation in (22) follows from the Karush-Kuhn-Tucker (KKT) conditions associated with the optimization problem yielded by taking the kernel right singular vectors to correspond to the source covariance eigenvectors. ■

It is now straightforward to show that kernel design does not impact the phase transition of the MMSE associated with Gaussian sources. However, and despite the fact that kernel design does not affect the number of measurements necessary

⁴In general, when the additive Gaussian noise \mathbf{w} has a non diagonal covariance matrix $\boldsymbol{\Sigma}_{\mathbf{w}}$, the measurement kernel Φ^* can be shown to expose and align the modes of both the input source and the noise, as observed for the measurement kernel which maximizes the mutual information between \mathbf{x} and \mathbf{y} [33].

to observe the phase transition, we also show that there is value in using designed kernels *in lieu* of random ones because one can thus improve reconstruction performance in terms both of a lower error floor (if present) and a lower power offset.

Theorem 5: Consider the linear measurement model in (1) where $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}})$, $s = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}})$, ℓ is the number of measurements and $\Phi = \Phi^*$, where Φ^* solves the optimization problem in (20). Then, the first-order, low-noise expansion of the MMSE is given by:

$$\text{MMSE}^G(\sigma^2, \Phi^*) = \mathcal{M}_\infty^{\text{GD}} + \mathcal{D}^{\text{GD}} \cdot \sigma^2 + o(\sigma^2). \quad (23)$$

where

$$\mathcal{M}_\infty^{\text{GD}} = \sum_{i=\ell'+1}^s \lambda_{\mathbf{x},i}, \quad \mathcal{D}^{\text{GD}} = (\ell')^2 / \ell \quad (24)$$

and $\ell' = \min\{s, \ell\}$.

Proof: On substituting the expression of Φ^* in (21) into (6), it is possible to expand the MMSE associated with the optimal linear kernel design as follows:

$$\text{MMSE}^G(\sigma^2, \Phi^*) = \sum_{i=1}^{\ell'} \frac{\lambda_{\mathbf{x},i}}{1 + \frac{1}{\sigma^2} \lambda_{\mathbf{x},i} \lambda_{\Phi,i}^*} + \sum_{i=\ell'+1}^s \lambda_{\mathbf{x},i}. \quad (25)$$

Observe that in the limit $\sigma^2 \rightarrow 0$, it follows from (22) that $\lambda_{\Phi,i}^* = \ell/\ell'$ for $i = 1, \dots, \ell'$ and $\lambda_{\Phi,i}^* = 0$ for $i = \ell', \dots, \ell$. Moreover, notice also that the second term in (25) is identically equal to zero if and only if $\ell \geq s$. ■

Note that when $\ell < s$ the error floor corresponds to the source power that the kernel fails to capture in view of its compressive nature. On the other hand, when $\ell \geq s$ the MMSE decay associated with an optimal kernel is equal to that of a random one, i.e. $\mathcal{O}(\sigma^2)$, but the MMSE power offset is lower: it is only a function of the number of measurements and the dimension of the subspace spanned by the source, independently of the exact form of the eigenvectors or eigenvalues of the source covariance.

Finally, it is also straightforward to show that kernel design does not impact the phase transition of the MMSE associated with GMM sources. The minimum number of measurements required to perfectly reconstruct the signal in the low-noise regime is also $\ell > s_{\text{max}}$ in this case. However, careful kernel design can increase the system performance by guaranteeing lower error floors and power offsets.

The method leveraged to prove this result is also based on the analysis of lower and upper bounds to the MMSE associated with the reconstruction of signals drawn from a GMM source that are sensed now via the optimal linear kernel design. In particular, we consider a lower bound – which we denote by $\text{MSE}_{\text{LBD}}(\sigma^2)$ – which is expressed in terms of the value of the MMSE corresponding to the actual Gaussian class from which each realization of the input signal is drawn and the corresponding optimal kernel Φ_k^* in (21). Then,

$$\text{MMSE}^{\text{GM}}(\sigma^2, \Phi^*) \geq \text{MSE}_{\text{LBD}}(\sigma^2) \quad (26)$$

$$= \sum_k p_k \text{MMSE}_k^G(\sigma^2, \Phi_k^*). \quad (27)$$

We also consider a trivial upper bound: the MMSE associated

with the optimal kernel design can always be upper bounded by the MMSE associated with a random kernel design.

Theorem 6: Consider the linear measurement model in (1) where $\mathbf{x} \sim \sum_k p_k \mathcal{N}(\boldsymbol{\mu}_x^{(k)}, \boldsymbol{\Sigma}_x^{(k)})$, $s_{\max} = \max_k s_k$, with $s_k = \text{rank}(\boldsymbol{\Sigma}_x^{(k)})$, ℓ is the number of measurements and $\boldsymbol{\Phi} = \boldsymbol{\Phi}^*$, where $\boldsymbol{\Phi}^*$ that solves the optimization problem in (20). Then, it holds

$$\lim_{\sigma^2 \rightarrow 0} \text{MMSE}^{\text{GM}}(\sigma^2, \boldsymbol{\Phi}^*) = 0 \Rightarrow \ell \geq s_{\max}, \quad (28)$$

and

$$\ell > s_{\max} \Rightarrow \lim_{\sigma^2 \rightarrow 0} \text{MMSE}^{\text{GM}}(\sigma^2, \boldsymbol{\Phi}^*) = 0. \quad (29)$$

Proof: It is possible to prove the sufficient condition (29) by observing that

$$\text{MMSE}^{\text{GM}}(\sigma^2, \boldsymbol{\Phi}^*) \leq \text{MMSE}^{\text{GM}}(\sigma^2, \boldsymbol{\Phi}) \quad (30)$$

for all possible measurement matrices $\boldsymbol{\Phi}$ that verify the trace constraint in (20). Among them, we can consider random kernels with Gaussian, zero-mean and fixed-variance, i.i.d. entries and we can obtain the sufficient condition by leveraging directly the result in Theorem 3. On the other hand, in order to prove the necessary condition (28), consider the lower bound in (27) and observe that

$$\text{MMSE}_k^{\text{G}}(\sigma^2, \boldsymbol{\Phi}_k^*) \leq \text{MMSE}_k^{\text{G}}(\sigma^2, \boldsymbol{\Phi}) \quad (31)$$

for all possible measurement matrices $\boldsymbol{\Phi}$ that verify the trace constraint in (20). Also in this case, we can consider random kernels with Gaussian, zero-mean and fixed-variance, i.i.d. entries. Then, it is possible to prove (28) by leveraging the necessary condition embedded in Theorem 5, that is,

$$\lim_{\sigma^2 \rightarrow 0} \text{MMSE}_k^{\text{G}}(\sigma^2, \boldsymbol{\Phi}) = 0 \Rightarrow \ell \geq s_k. \quad (32)$$

In view of (23) and (27), the low-noise expansion of the proposed lower bound is given by:

$$\text{MSE}_{\text{LBD}}(\sigma^2) = \sum_k p_k \mathcal{M}_{\infty_k}^{\text{GD}} + \left(\sum_k p_k \mathcal{D}_k^{\text{GD}} \right) \sigma^2 + o(\sigma^2), \quad (33)$$

where $\mathcal{M}_{\infty_k}^{\text{GD}}$ and $\mathcal{D}_k^{\text{GD}}$ are the zero-order term and the coefficient of the first-order term of the expansion of the Gaussian MMSE corresponding to class k . We also conclude that, though kernel design may not increase the MMSE decay rate beyond $\mathcal{O}(\sigma^2)$, it can have an impact on the MMSE power offset associated with GMM sources.

It is important to note though that this analysis has concentrated on offline designs where the ℓ measurements are designed concurrently [33], [52], rather than online kernel designs that entail a sequential design of the measurements by leveraging information derived from previous measurements [30], [33], [55]. It is possible that such online designs have an impact on the phase transition.

VI. NUMERICAL RESULTS

We now provide results both with synthetic and real data to illustrate the theory. Recovery is based upon the conditional expectation, which is analytic for the case of GMM priors,

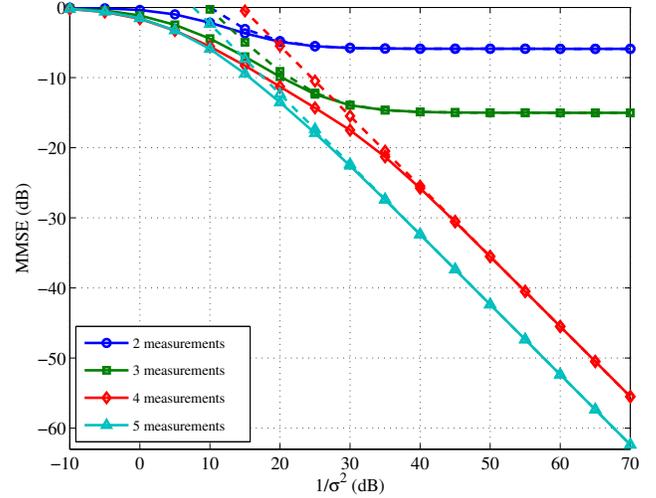


Fig. 1. MMSE vs. $1/\sigma^2$ for different numbers of random measurements $\ell = 2, 3, 4, 5$ for a Gaussian source with $n = 5$ and $s = 4$. Actual MMSE (solid lines) and low-noise, first-order expansions (dashed lines).

and optimal in terms of mean-squared error. In all simulations we use random measurement kernels, where the entries of the kernel are realizations of i.i.d. Gaussian random variables with zero mean and unit variance which are subsequently normalized by a scaling factor that guarantees that $\text{tr}(\boldsymbol{\Phi}^\dagger \boldsymbol{\Phi}) \leq \ell$.

A. Synthetic data

Fig. 1 shows the MMSE vs. $1/\sigma^2$ for a Gaussian source with dimension $n = 5$ and rank $s = 4$. We confirm that the MMSE phase transition occurs with $\ell = 4$ measurements. We also confirm that the first-order expansion in (8) captures well the behavior of the MMSE, both in the presence and absence of an error floor, for values of $1/\sigma^2$ larger than 20-30 dB. In fact, such noise amplitudes are already well below $1/\sigma^2 = 60$ dB, which is a noise level with operational significance for various image processing applications [33], [34]. Note also that by taking the number of measurements to be greater than the number of measurements that achieves the phase transition we do not affect the MMSE decay but we only affect the MMSE power offset.

Fig. 2 now shows the values of the MMSE for a 2-classes GMM input with $n = 4$, $p_1 = p_2 = 0.5$, means $\boldsymbol{\mu}_x^{(1)} = \boldsymbol{\mu}_x^{(2)} = \mathbf{0}$, and covariances drawn from a central Wishart distribution [56, p. 84] with dimension 4 and degrees of freedom 2, so that $s_1 = s_2 = 2$. We report the actual value of $\text{MMSE}^{\text{GM}}(\sigma^2)$, the lower bound $\text{MSE}_{\text{LB}}(\sigma^2)$, and the classify and reconstruct upper bound $\text{MSE}_{\text{CR}}(\sigma^2)$. We also report another MMSE upper bound associated with a linear estimator, i.e., the linear minimum mean-squared error (LMMSE) [57]. We notice that when $\ell = 2$ the lower bound converges to zero as $\sigma^2 \rightarrow 0$, whereas the classify and reconstruct upper bound converges to an error floor. However, it appears that the classify and reconstruct upper bound captures better the features of the actual MMSE, which also exhibits an error floor. On the other hand, notice that when $\ell \geq 3$ both the lower

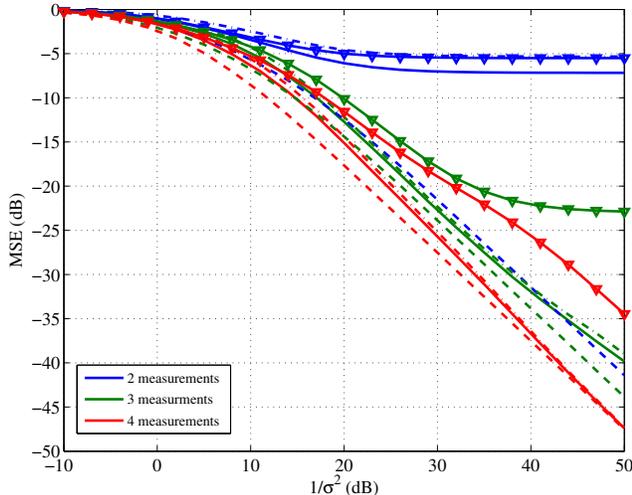


Fig. 2. MMSE vs. $1/\sigma^2$ for different numbers of random measurements $\ell = 2, 3, 4$ for a 2-classes GMM source with $s_1 = s_2 = 2$. Actual MMSE (solid lines), lower bound (dashed lines), CR upper bound (dashed-dotted lines) and LMMSE upper bound (triangles).

and upper bound converge to zero as $\sigma^2 \rightarrow 0$, as expected. It is also interesting to observe that the LMMSE upper bound does not describe crisply the MMSE phase transition: it is in fact possible to show by leveraging the previous machinery that such a sub-optimal estimator requires the number of random measurements to be larger than or equal to the dimension of the direct sum of the subspaces spanned by the different signals in the different classes for a MMSE phase transition to occur.

Another interesting feature relates to the fact that when $\ell = 3$ both the upper bound to the MMSE and the actual MMSE are $\mathcal{O}(\sigma)$ as $\sigma^2 \rightarrow 0$ whereas when $\ell > 3$ the upper and lower bound to the MMSE and the actual MMSE are $\mathcal{O}(\sigma^2)$ as $\sigma^2 \rightarrow 0$. This behavior appears to be related to the fact that when $\ell = 3$ or $\ell > 3$ the misclassification probability of the optimal MAP classifier, which is also used in the classify and reconstruction procedure, approaches zero as $\sigma^2 \rightarrow 0$ with different decays [51]. This behavior stands in contrast to the scalar case [50] and implies that the MMSE dimension corresponding to the mixture of Gaussian vectors is not equal to the mixture of the MMSE dimensions of the individual Gaussian vectors: this is only true when the first-order expansions of lower and the upper bound coincide.

B. Real data

The phase transition phenomena can also be observed in the reconstruction of real imagery data. As an example, we consider a 256×256 cropped version of the image “barbara”. The input \mathbf{x} , in this case, represents 8×8 non-overlapping patches extracted from the image. The source is described by a 20-classes GMM prior that is obtained by training the non-parametric, Bayesian, dictionary learning algorithm described in [25] over 100,000 patches randomly extracted from 500

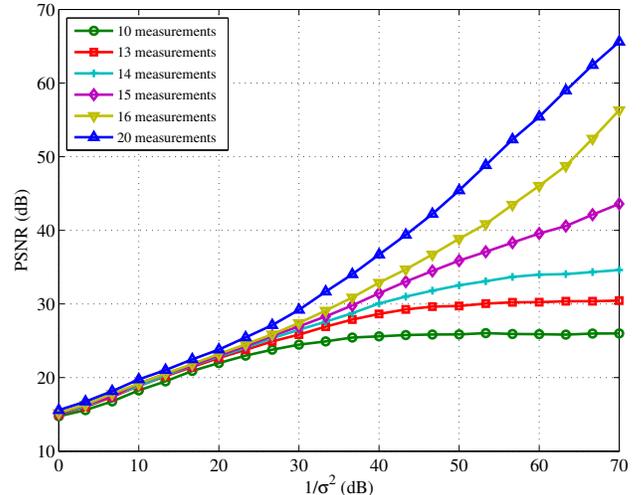


Fig. 3. PSNR vs. $1/\sigma^2$. Image “barbara.png”, projected onto a GMM model with $s_{\max} = 14$.

images in the Berkeley Segmentation Dataset⁵. Note that the image “barbara” is not in the training ensemble. The so obtained GMM prior has full rank, class conditioned input covariance matrices $\Sigma_{\mathbf{x}}^{(k)}$. In order to fit the trained GMM to the low-rank model (or, equivalently, to the model of a union of subspaces), which is the basis of our theory, only the first $s_{\max} = 14$ principal components of each class-conditioned input covariance matrix are retained, and the remaining 50 eigenvalues of each covariance matrix are set to be equal to zero. Moreover, our test image is modified by projecting each patch extracted from the image “barbara” onto the 14-dimensional sub-space corresponding to the low-rank input covariance matrix of the class associated with that particular patch. Note that projecting the image onto the lower dimensional subspaces does not introduce substantial distortion. In fact, the peak signal-to-noise ratio (PSNR) of the projected image with respect to the original ground truth in this case is equal to 77.3 dB⁶. This is a manifestation of the fact that natural images are well represented by “almost low-rank” GMM priors, and the eigenvalues of the corresponding class conditioned input covariance matrices decay rapidly. This underscores that the low-rank GMM representation is a good model for patches extracted from natural imagery, and therefore of significant practical value.

Reconstruction of the vectors \mathbf{x} from the compressive measurements \mathbf{y} is performed by the conditional mean estimator corresponding to the trained GMM prior after taking the 14 principal components, which can be written in closed-form [25].

Fig. 3 shows the PSNR vs. $1/\sigma^2$ for different numbers of compressive measurements. It is possible to clearly observe

⁵<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html>

⁶The PSNR values obtained by choosing $s_{\max} = 13$ and $s_{\max} = 15$ are 76.9 dB and 78 dB, respectively. On the other hand, setting $s_{\max} = 10$ reduces the PSNR to 75.2 dB and it induces visible distortion effects.



Fig. 4. Barbara. The right hand column contains the non-compressed image “barbara.png” after projection onto a GMM model with $s_{\max} = 14$. The three rows correspond to $\ell = 8, 12, 16$ random measurements, from the top to the bottom. The first three columns, from left to right, correspond to the noise levels $1/\sigma^2 = 20, 40, 60$ dB.

the phase transition when $\ell > s_{\max}$ and that, similarly to what has been noted in Fig. 2, the PSNR increases approximately as $\mathcal{O}(1/\sigma)$ when $\ell = 15$, and as $\mathcal{O}(1/\sigma^2)$ when $\ell > 15$. Finally, in Fig. 4 are reported some reconstruction examples. The right hand column contains the image “barbara” projected onto the union of 14-dimensional subspaces that characterize the GMM prior. The three rows correspond to $\ell = 8, 12, 16$ random measurements, from the top to the bottom. The first three columns, from left to right, correspond to the noise levels $1/\sigma^2 = 20, 40, 60$ dB. When the number of measurements is below the phase transition, perfect reconstruction is not possible, even when $\sigma^2 \rightarrow 0$: this is particularly evident by observing the high-frequency content of the image. On the other hand, when $\ell = 16$, a clear phase transition is observable.

It is also relevant to reflect further on the fact that our theory applies only to low-rank rather than the so-called “approximately” low-rank models. Arguably, a GMM model trained with natural images is not exactly low-rank, as real signals do not perfectly lie on the union of low-dimensional subspaces [25]. In fact, it is typical to describe the class conditioned covariance matrices associated with a GMM model as follows:

$$\Sigma_{\mathbf{x}}^{(k)} = \bar{\Sigma}_{\mathbf{x}}^{(k)} + \varepsilon \mathbf{I}_n, \quad (34)$$

where the matrix $\bar{\Sigma}_{\mathbf{x}}^{(k)}$ is exactly low-rank, and the matrix $\varepsilon \mathbf{I}_n$ accounts for model mismatch between real data and their projection onto the principal components that contain the large majority of the information associated to the data. Given the fact that the compressive sensing model in (1) with $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ and full-rank GMM with class conditioned covariances $\Sigma_{\mathbf{x}}^{(k)}$ is mathematically equivalent to the compressive sensing model in (1) with $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \varepsilon \Phi \Phi^\dagger + \sigma^2 \mathbf{I})$ and low-rank GMM with class conditioned covariances $\bar{\Sigma}_{\mathbf{x}}^{(k)}$, then it is possible to appreciate the impact of model mismatch on the theory.

For example, consider our earlier testing model with images that are not projected onto lower-dimensional union of subspaces. It is evident that the reconstruction PSNR would now be upper bounded as $\sigma^2 \rightarrow 0$ for all $\ell < n$, in view of the noise amplification from σ^2 to roughly⁷ $1/(\sigma^2 + \varepsilon)$. This also leads to the conclusions that the performance of the “approximately” low-rank model as $\sigma^2 \rightarrow 0$ is comparable to the performance of the low-rank model with $1/\sigma^2 = 1/\varepsilon$. That is, operating an “approximately” low-rank model at a

⁷In fact, the noise power depends also on the value of $\text{tr}(\Phi \Phi^\dagger)$. However, when the entries of Φ are i.i.d., zero-mean, Gaussian, the matrix $\Phi \Phi^\dagger$ approximates well the identity matrix.

certain σ^2 leads to a performance that is comparable to that of operating a low-rank model at $\sigma^2 + \varepsilon$.

Nonetheless, it is worth noting that natural images projected on low-rank union of subspaces turn out to be very good approximations of the original images, as it was shown for the case of the image ‘‘barbara’’ (see Fig. 4), thus validating the effectiveness of this model in representing real data.

VII. CONCLUSION

The principal contribution is a non-asymptotic (in the number of dimensions) characterization of MMSE phase transitions associated with the reconstruction of GMM signals from noisy compressive measurements. In particular, it has been shown that with either random or optimal kernel designs it is sufficient to take the number of measurements to be strictly greater than the largest dimension of the sub-spaces spanned by the class-conditioned signals in the GMM model – a dual of sparsity – for a phase transition to occur. It has also been shown that additional measurements and/or designed measurements translate only onto (occasionally) an improved MMSE decay or an improved MMSE power offset in the low-noise regime. Another interesting by-product of the contribution is the fact that a (sub-optimal) classify and reconstruction procedure is nearly phase-transition optimal. This also provides a rationale for the state-of-the-art performance of similar reconstruction procedures, as reported in the recent literature [29], [30].

APPENDIX A USEFUL LEMMAS

Lemma 1: Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a positive semidefinite matrix. Then, for all $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{A}\mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{x}^\dagger \mathbf{A}\mathbf{x} = 0. \quad (35)$$

Proof: If $\mathbf{x}^\dagger \mathbf{A}\mathbf{x} = 0$, then, $\|\mathbf{A}^{\frac{1}{2}}\mathbf{x}\|^2 = 0$, in which $\mathbf{A}^{\frac{1}{2}}$ is the positive semidefinite matrix square root of \mathbf{A} . Therefore, $\mathbf{A}^{\frac{1}{2}}\mathbf{x} = \mathbf{0}$, which also implies $\mathbf{A}\mathbf{x} = \mathbf{0}$. The opposite implication in (35) is straightforward. ■

Lemma 2: Given two positive semidefinite matrices, $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{n \times n}$, with ranks s_1 and s_2 , respectively, and $s_{12} = \text{rank}(\mathbf{A}_1 + \mathbf{A}_2)$, then

$$\begin{aligned} \frac{s_1 + s_2}{2} = s_{12} &\Leftrightarrow \text{Null}(\mathbf{A}_1) = \text{Null}(\mathbf{A}_2) \\ &\Leftrightarrow \text{Im}(\mathbf{A}_1) = \text{Im}(\mathbf{A}_2) \end{aligned} \quad (36)$$

Proof: First, it is easy to show that

$$\text{Null}(\mathbf{A}_1 + \mathbf{A}_2) = \text{Null}(\mathbf{A}_1) \cap \text{Null}(\mathbf{A}_2), \quad (37)$$

by leveraging Lemma 1 and considering that $\mathbf{x}^\dagger(\mathbf{A}_1 + \mathbf{A}_2)\mathbf{x} = 0$ if and only if $\mathbf{x}^\dagger \mathbf{A}_1 \mathbf{x} = 0$ and $\mathbf{x}^\dagger \mathbf{A}_2 \mathbf{x} = 0$. Therefore, $\dim \text{Null}(\mathbf{A}_1 + \mathbf{A}_2) \leq \dim \text{Null}(\mathbf{A}_1)$ and $\dim \text{Null}(\mathbf{A}_1 + \mathbf{A}_2) \leq \dim \text{Null}(\mathbf{A}_2)$, which immediately implies $s_{12} \geq \max\{s_1, s_2\}$, and, in our case, $s_1 = s_2 = s_{12}$. As a consequence, we have that

$$\dim \text{Null}(\mathbf{A}_1) = \dim \text{Null}(\mathbf{A}_2) = \dim \text{Null}(\mathbf{A}_1 + \mathbf{A}_2), \quad (38)$$

which can be combined with (37) to give

$$\text{Null}(\mathbf{A}_1) = \text{Null}(\mathbf{A}_2). \quad (39)$$

Moreover, the last part of (36) can be easily obtained by observing that the image of a positive semidefinite matrix is the orthogonal complement of its null space. ■

Lemma 3: Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a positive semidefinite matrix with $s = \text{rank}(\mathbf{A})$ and let $\mathbf{x} \in \mathbb{R}^n$. Then,

$$\text{rank}(\mathbf{A} + \mathbf{x}\mathbf{x}^\dagger) = s + 1 \Leftrightarrow \mathbf{x} \notin \text{Im}(\mathbf{A}). \quad (40)$$

Proof: First, note that $\text{rank}(\mathbf{A} + \mathbf{x}\mathbf{x}^\dagger) \leq s + 1$ [58, §0.4.5.d]. Then, the proof is based on expressing \mathbf{A} in terms of its eigenvalues $\lambda_{\mathbf{A}_i}$ and eigenvectors $\mathbf{u}_{\mathbf{A}_i}$ as $\mathbf{A} = \sum_i \lambda_{\mathbf{A}_i} \mathbf{u}_{\mathbf{A}_i} \mathbf{u}_{\mathbf{A}_i}^\dagger$. ■

APPENDIX B PROOF OF THEOREM 1

We can rewrite the expression of the MMSE for Gaussian input sources in (6) by making use of the cyclic property of the trace and the matrix inversion Lemma [58, §0.7.4]

$$\mathbf{I} - \mathbf{A}^\dagger(\mathbf{A}\mathbf{A}^\dagger + c^{-1}\mathbf{I})^{-1}\mathbf{A} = (c\mathbf{I} + \mathbf{A}^\dagger\mathbf{A})^{-1}, \quad (41)$$

with $\mathbf{A} = \Phi \Sigma_{\mathbf{x}}^{\frac{1}{2}}$, as

$$\text{MMSE}^G(\sigma^2) = \text{tr} \left(\Sigma_{\mathbf{x}} \left(\mathbf{I} + 1/\sigma^2 \Sigma_{\mathbf{x}}^{\frac{1}{2}} \Phi^\dagger \Phi \Sigma_{\mathbf{x}}^{\frac{1}{2}} \right)^{-1} \right) \quad (42)$$

$$= \text{tr} \left(\Sigma_{\mathbf{x}} \mathbf{U} (\mathbf{I} + 1/\sigma^2 \Lambda)^{-1} \mathbf{U}^\dagger \right) \quad (43)$$

$$= \text{tr} \left(\Sigma_{\mathbf{x}} \mathbf{U} \tilde{\Lambda} \mathbf{U}^\dagger \right), \quad (44)$$

where we have used the eigenvalue decomposition of Σ in (7) and introduced the diagonal matrix $\tilde{\Lambda} = \text{diag} \left(\frac{1}{1+\lambda_1/\sigma^2}, \dots, \frac{1}{1+\lambda_{\ell'}/\sigma^2}, 1, \dots, 1 \right)$. Note now that the eigenvalue decomposition of a positive semidefinite matrix can also be written as

$$\mathbf{A} = \mathbf{U}_{\mathbf{A}} \Lambda_{\mathbf{A}} \mathbf{U}_{\mathbf{A}}^\dagger = \sum_i \lambda_{\mathbf{A}_i} \mathbf{u}_{\mathbf{A}_i} \mathbf{u}_{\mathbf{A}_i}^\dagger, \quad (45)$$

in which $\mathbf{u}_{\mathbf{A}_i}$ is the i -th column of $\mathbf{U}_{\mathbf{A}}$. Therefore, the MMSE for Gaussian inputs can be expressed as

$$\text{MMSE}^G(\sigma^2) = \sum_{i=1}^{\ell'} \frac{1}{1 + \lambda_i/\sigma^2} \mathbf{u}_i^\dagger \Sigma_{\mathbf{x}} \mathbf{u}_i + \sum_{i=\ell'+1}^n \mathbf{u}_i^\dagger \Sigma_{\mathbf{x}} \mathbf{u}_i \quad (46)$$

where $\mathbf{u}_1, \dots, \mathbf{u}_{\ell'}$ are the eigenvectors of Σ corresponding to the positive eigenvalues $\lambda_1, \dots, \lambda_{\ell'}$ and the vectors $\mathbf{u}_{\ell'+1}, \dots, \mathbf{u}_n$ can form any orthonormal basis of the null space $\text{Null}(\Sigma)$, which has dimension $\dim(\text{Null}(\Sigma)) = n - \ell'$. The first term in (46) tends to zero when $\sigma^2 \rightarrow 0$, so that the phase transition of the MMSE is determined by the second term, which can be characterized on the basis of the description of the two null spaces $\text{Null}(\Sigma)$ and $\text{Null}(\Sigma_{\mathbf{x}})$. In particular, note that

$$\text{Null}(\Sigma_{\mathbf{x}}) \subseteq \text{Null}(\Sigma), \quad (47)$$

as for each $\mathbf{v} \in \mathbb{R}^n$ such that $\Sigma_{\mathbf{x}} \mathbf{v} = \mathbf{0}$, it holds also $\Sigma \mathbf{v} = \mathbf{0}$, since $\Sigma_{\mathbf{x}}^{\frac{1}{2}} \mathbf{v} = \mathbf{0}$. Therefore, we can choose the basis

$\mathbf{u}_{\ell'+1}, \dots, \mathbf{u}_n$ as follows. The last $n-s$ vectors $\mathbf{u}_{s+1}, \dots, \mathbf{u}_n$ form an orthonormal basis of the space $\text{Null}(\Sigma_{\mathbf{x}})$ and the remaining vectors $\mathbf{u}_{\ell'+1}, \dots, \mathbf{u}_s$ form an orthonormal basis of $\text{Null}(\Sigma) \cap \text{Null}(\Sigma_{\mathbf{x}})^\perp$. Then, we can write (46) as

$$\text{MMSE}^G(\sigma^2) = \sum_{i=1}^{\ell'} \frac{1}{1 + \lambda_i/\sigma^2} \mathbf{u}_i^\dagger \Sigma_{\mathbf{x}} \mathbf{u}_i + \sum_{i=\ell'+1}^s \mathbf{u}_i^\dagger \Sigma_{\mathbf{x}} \mathbf{u}_i. \quad (48)$$

Observe that, when $\ell \geq s$, then $\ell' = \ell \geq s$, and the second term in (48) is identically equal to zero, as the linear space $\text{Null}(\Sigma) \cap \text{Null}(\Sigma_{\mathbf{x}})^\perp$ contains only the zero vector, and the Gaussian MMSE tends to zero when $\sigma^2 \rightarrow 0$. On the other hand, if $\ell < s$, by Lemma 1, $\mathbf{u}_i^\dagger \Sigma_{\mathbf{x}} \mathbf{u}_i > 0$, for $i = \ell'+1, \dots, s$ and the MMSE is characterized by an error floor in the low-noise regime. Specifically, we can expand the Gaussian MMSE as

$$\text{MMSE}^G(\sigma^2) = \mathcal{M}_\infty^G + \mathcal{D}^G \cdot \sigma^2 + o(\sigma^2) \quad (49)$$

where

$$\mathcal{M}_\infty^G = \sum_{i=\ell'+1}^s \mathbf{u}_i^\dagger \Sigma_{\mathbf{x}} \mathbf{u}_i, \quad \mathcal{D}^G = \sum_{i=1}^{\ell'} \frac{1}{\lambda_i} \mathbf{u}_i^\dagger \Sigma_{\mathbf{x}} \mathbf{u}_i. \quad (50)$$

APPENDIX C PROOF OF THEOREM 3

The proof is based on the analysis of the upper bound on $\text{MMSE}^{\text{GM}}(\sigma^2)$ obtained by considering the mean-squared error associated to the classify and reconstruct decoder described in Section IV-B. The value of such upper bound is given by

$$\begin{aligned} \text{MSE}_{\text{CR}}(\sigma^2) &= \sum_k p_k \sum_m p_{\hat{c}|c}(m|k) \\ &\quad \cdot \mathbb{E} [\|\mathbf{x} - \mathcal{W}_m(\mathbf{y})\|^2 | \hat{c} = m, c = k] \quad (51) \\ &= \sum_k p_k p_{\hat{c}|c}(k|k) \mathbb{E} [\|\mathbf{x} - \mathcal{W}_k(\mathbf{y})\|^2 | \hat{c} = c = k] \\ &\quad + \sum_k p_k \sum_{m \neq k} p_{\hat{c}|c}(m|k) \\ &\quad \cdot \mathbb{E} [\|\mathbf{x} - \mathcal{W}_m(\mathbf{y})\|^2 | \hat{c} = m, c = k], \quad (52) \end{aligned}$$

where we have denoted by $p_{\hat{c}|c}(m|k)$ the probability that the MAP classifier yields $\hat{c} = m$ given that the actual input class is $c = k$. The first term, then, can be upper bounded by using the law of total probability, and we obtain

$$\begin{aligned} \text{MSE}_{\text{CR}}(\sigma^2) &\leq \sum_k p_k \mathbb{E} [\|\mathbf{x} - \mathcal{W}_k(\mathbf{y})\|^2 | c = k] \\ &\quad + \sum_k p_k \sum_{m \neq k} p_{\hat{c}|c}(m|k) \\ &\quad \cdot \mathbb{E} [\|\mathbf{x} - \mathcal{W}_m(\mathbf{y})\|^2 | \hat{c} = m, c = k] \quad (53) \\ &= \text{MSE}_{\text{LB}}(\sigma^2) \\ &\quad + \sum_k p_k \sum_{m \neq k} p_{\hat{c}|c}(m|k) \\ &\quad \cdot \mathbb{E} [\|\mathbf{x} - \mathcal{W}_m(\mathbf{y})\|^2 | \hat{c} = m, c = k], \quad (54) \end{aligned}$$

where $\text{MSE}_{\text{LB}}(\sigma^2)$ is the lower bound to the MMSE, which has been shown to approach zero when $\sigma^2 \rightarrow 0$, since we are assuming here that $\ell > s_{\text{max}}$.

Then, we need to show that, when $m \neq k$,

$$\lim_{\sigma^2 \rightarrow 0} p_{\hat{c}|c}(m|k) \mathbb{E} [\|\mathbf{x} - \mathcal{W}_m(\mathbf{y})\|^2 | \hat{c} = m, c = k] = 0, \quad (55)$$

and we consider separately in the remainder of this Appendix the two cases for which the subspaces corresponding to the images of the input covariance matrices of class k and m completely overlap or not. In other terms, we consider separately the cases in which the affine spaces spanned by the signals in class k and class m differ only for a fixed translation or not.

A. *Non-overlapping case:* $\frac{s_k + s_m}{2} < s_{km}$

In this case, given that $\ell > s_{\text{max}}$, by leveraging the results in [51, Theorem 2], we can state that

$$\lim_{\sigma^2 \rightarrow 0} p_{\hat{c}|c}(m|k) = 0. \quad (56)$$

The misclassification probability $p_{\hat{c}|c}(m|k)$ is the measure of the set representing the decision region of the MAP classifier associated with class m with respect to the Gaussian measure corresponding to the Gaussian distribution

$$\mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_k \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{x}}^{(k)} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \cdot \mathbf{I} \end{bmatrix} \right). \quad (57)$$

Moreover it can be shown that, in the limit $\sigma^2 \rightarrow 0$, the product in (55) is upper bounded by the integral of a measurable function over a set with measure zero, which is then equal to zero.

B. *Overlapping case:* $\frac{s_k + s_m}{2} = s_{km}$

Observe that, in this case, Lemma 2 states that $\text{Im}(\Sigma_{\mathbf{x}}^{(k)} + \Sigma_{\mathbf{x}}^{(m)}) = \text{Im}(\Sigma_{\mathbf{x}}^{(k)}) = \text{Im}(\Sigma_{\mathbf{x}}^{(m)})$. Then, we further consider two separate cases: i) the difference of the mean vectors of classes k and m lies on the subspace spanned by the covariance matrices of the two classes and ii) the difference of the mean vectors of classes k and m does not lie on the subspace spanned by the covariance matrices of the two classes. We consider first the case in which

$$\boldsymbol{\mu}_{\mathbf{x}}^{(k)} - \boldsymbol{\mu}_{\mathbf{x}}^{(m)} \notin \text{Im}(\Sigma_{\mathbf{x}}^{(k)} + \Sigma_{\mathbf{x}}^{(m)}). \quad (58)$$

For simplicity of notation, we introduce the symbol $\mathbf{M}_{km} = (\boldsymbol{\mu}_{\mathbf{x}}^{(k)} - \boldsymbol{\mu}_{\mathbf{x}}^{(m)})(\boldsymbol{\mu}_{\mathbf{x}}^{(k)} - \boldsymbol{\mu}_{\mathbf{x}}^{(m)})^\dagger$ and we can show that, with probability 1,

$$\text{rank}(\Phi(\Sigma_{\mathbf{x}}^{(k)} + \Sigma_{\mathbf{x}}^{(m)} + \mathbf{M}_{km})\Phi^\dagger) = s_{km} + 1, \quad (59)$$

as, by Lemma 3, $\text{rank}(\Sigma_{\mathbf{x}}^{(k)} + \Sigma_{\mathbf{x}}^{(m)} + \mathbf{M}_{km}) = s_{km} + 1$ and $\ell > s_{km}$. This implies that, using again Lemma 3, we have

$$\Phi(\boldsymbol{\mu}_{\mathbf{x}}^{(k)} - \boldsymbol{\mu}_{\mathbf{x}}^{(m)}) \notin \text{Im}(\Phi(\Sigma_{\mathbf{x}}^{(k)} + \Sigma_{\mathbf{x}}^{(m)})\Phi^\dagger). \quad (60)$$

Therefore, by using the result in [51, Theorem 3], we can state that $\lim_{\sigma^2 \rightarrow 0} p_{\hat{c}|c}(m|k) = 0$, and, by using a similar proof to that presented in the previous paragraphs, we can show that (55) is satisfied also in this case.

Consider now the case for which

$$\boldsymbol{\mu}_{\mathbf{x}}^{(k)} - \boldsymbol{\mu}_{\mathbf{x}}^{(m)} \in \text{Im}(\Sigma_{\mathbf{x}}^{(k)} + \Sigma_{\mathbf{x}}^{(m)}) = \text{Im}(\Sigma_{\mathbf{x}}^{(m)}). \quad (61)$$

Here, the misclassification probability is not guaranteed to approach zero when $\sigma^2 \rightarrow 0$, and we have to resort to a different proof technique. The rationale behind this proof is that also the mismatched mean-squared error approaches zero in the low-noise regime, provided that the signals in the actual class k and the mismatched class m span the same space.

By using the law of total probability we can write

$$p_{\hat{c}|c}(m|k) \mathbb{E} [\|\mathbf{x} - \mathcal{W}_m(\mathbf{y})\|^2 | \hat{c} = m, c = k] \leq \text{MSE}_{km}^{\text{MIS}}(\sigma^2) \quad (62)$$

where $\text{MSE}_{km}^{\text{MIS}}(\sigma^2) = \mathbb{E} [\|\mathbf{x} - \mathcal{W}_m(\mathbf{y})\|^2 | c = k]$ is the mismatched mean-squared error incurred when estimating signals drawn from the Gaussian class k with the conditional mean estimator associated with class m .

Observe that, on denoting by $\Sigma_{\mathbf{y}}^{(k)} = \mathbf{I}\sigma^2 + \Phi \Sigma_{\mathbf{x}}^{(k)} \Phi^\dagger$ the covariance matrix of the measurement vector \mathbf{y} conditioned on class k , we can write

$$\begin{aligned} \text{MSE}_{km}^{\text{MIS}}(\sigma^2) &= \mathbb{E} \left[\text{tr} \left((\mathbf{x} - \mathcal{W}_m(\mathbf{y})) (\mathbf{x} - \mathcal{W}_m(\mathbf{y}))^\dagger \right) | c = k \right] \\ &= \text{tr} \left(\Sigma_{\mathbf{x}}^{(k)} \right) - 2\text{tr} \left(\mathbf{W}_k \Sigma_{\mathbf{y}}^{(k)} \mathbf{W}_m^\dagger \right) \\ &\quad \text{tr} \left(\mathbf{W}_m \Sigma_{\mathbf{y}}^{(k)} \mathbf{W}_m^\dagger \right) + \text{tr} \left(\mathbf{M}_{km} \right) \\ &\quad - 2\text{tr} \left(\mathbf{M}_{km} \Phi^\dagger \mathbf{W}_m^\dagger \right) \\ &\quad + \text{tr} \left(\Phi \mathbf{W}_m \mathbf{M}_{km} \Phi^\dagger \mathbf{W}_m^\dagger \right). \end{aligned} \quad (63)$$

In order to prove that $\text{MSE}_{km}^{\text{MIS}}(\sigma^2)$ approaches zero when $\sigma^2 \rightarrow 0$, we can show the following four identities:

$$\lim_{\sigma^2 \rightarrow 0} \text{tr} \left(\mathbf{W}_k \Sigma_{\mathbf{y}}^{(k)} \mathbf{W}_m^\dagger \right) = \text{tr} \left(\Sigma_{\mathbf{x}}^{(k)} \right); \quad (64)$$

$$\lim_{\sigma^2 \rightarrow 0} \text{tr} \left(\mathbf{W}_m \Sigma_{\mathbf{y}}^{(k)} \mathbf{W}_m^\dagger \right) = \text{tr} \left(\Sigma_{\mathbf{x}}^{(k)} \right); \quad (65)$$

$$\lim_{\sigma^2 \rightarrow 0} \text{tr} \left(\mathbf{M}_{km} \mathbf{W}_m^\dagger \right) = \text{tr} \left(\mathbf{M}_{km} \right); \quad (66)$$

$$\lim_{\sigma^2 \rightarrow 0} \text{tr} \left(\mathbf{W}_m \Phi \mathbf{M}_{km} \Phi^\dagger \mathbf{W}_m^\dagger \right) = \text{tr} \left(\mathbf{M}_{km} \right). \quad (67)$$

Specifically, we can leverage the inversion Lemma [58]

$$\mathbf{A}(\mathbf{I}c^{-1} + \mathbf{B}\mathbf{A})^{-1}\mathbf{B} = \mathbf{I} - (\mathbf{I} + c\mathbf{A}\mathbf{B})^{-1}, \quad (68)$$

with $\mathbf{A} = \Phi^\dagger$ and $\mathbf{B} = \Phi \Sigma_{\mathbf{x}}^{(m)}$, and write

$$\begin{aligned} \text{tr} \left(\mathbf{W}_k \Sigma_{\mathbf{y}}^{(k)} \mathbf{W}_m^\dagger \right) &= \text{tr} \left(\Sigma_{\mathbf{x}}^{(k)} \Phi^\dagger (\mathbf{I}\sigma^2 + \Phi \Sigma_{\mathbf{x}}^{(m)} \Phi^\dagger)^{-1} \right. \\ &\quad \left. \Phi \Sigma_{\mathbf{x}}^{(m)} \right) \quad (69) \\ &= \text{tr} \left(\Sigma_{\mathbf{x}}^{(k)} \right) \\ &\quad - \text{tr} \left(\Sigma_{\mathbf{x}}^{(k)} \left(\mathbf{I} + \frac{1}{\sigma^2} \Phi^\dagger \Phi \Sigma_{\mathbf{x}}^{(m)} \right)^{-1} \right). \end{aligned} \quad (70)$$

Then, by noting that the matrix $\Phi^\dagger \Phi \Sigma_{\mathbf{x}}^{(m)}$ is diagonalizable with probability 1, and by following steps similar to those adopted in the proof of Theorem 1, we are able to prove (64). Finally, also (65), (66) and (67) are proved by following a completely similar approach.

REFERENCES

[1] D. Healy and D. Brady, "Compression at the physical interface," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 67–71, 2008.

[2] M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed sensing MRI," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 72–82, 2008.

[3] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

[4] "A/D converters," *Analog Devices Corp.*, 2009. [Online]. Available: <http://www.analog.com/en/analog-to-digital-converters/ad-converters/products/index.html>.

[5] "Data converters," *Texas Instruments Corp.*, 2009. [Online]. Available: <http://focus.ti.com/analog/docs/dataconvertershome.tsp>.

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

[7] 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, 2006.

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

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

[10] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, pp. 253–263, Dec. 2008.

[11] A. Ashok, P. Baheti, and M. Neifeld, "Compressive imaging system design using task-specific information," *Applied Optics*, vol. 47, no. 25, pp. 4457–4471, 2008.

[12] P. Baheti and M. Neifeld, "Recognition using information-optimal adaptive feature-specific imaging," *J. Opt. Soc. Amer. A*, vol. 26, no. 4, pp. 1055–1070, 2009.

[13] R. Venkataramani and Y. Bresler, "Further results on spectrum blind sampling of 2d signals," in *IEEE Int. Conf. Image Proc. (ICIP)*, vol. 2, 1998, pp. 752–756.

[14] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," *Available online at: http://dsp.rice.edu/cs/*, 2005.

[15] 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.

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

[17] 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.

[18] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[19] 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, 2009.

[20] 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, 2009.

[21] 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, 2009.

[22] 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, 2010.

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

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

[25] M. Chen, J. Silva, J. Paisley, D. Dunson, and L. Carin, "Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6140–6155, 2010.

[26] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, 1998.

[27] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Computat. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[28] G. Yu and G. Sapiro, "Statistical compressed sensing of Gaussian mixture models," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5842–5858, 2011.

- [29] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, 2012.
- [30] J. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Task-driven adaptive statistical compressive sensing of Gaussian mixture models," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 585–600, 2013.
- [31] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [32] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.
- [33] W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with applications to compressive sensing," *SIAM J. Imag. Sciences*, vol. 5, no. 4, pp. 1185–1212, 2012.
- [34] M. Chen, W. Carson, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications inspired linear discriminant analysis," in *Int. Conf. Machine Learn. (ICML)*, Jun.-Jul. 2012.
- [35] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n ," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.
- [36] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," *Phil. Trans. Roy. Soc. A*, vol. 367, no. 1906, pp. 4273–4293, 2009.
- [37] D. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6920–6941, 2011.
- [38] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6241–6263, 2012.
- [39] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3065–3092, 2012.
- [40] A. Tulino, G. Caire, S. Shamai, and S. Verdú, "Support recovery with sparsely sampled free random matrices," in *IEEE Int. Symp. Information Theory (ISIT)*, Jul.-Aug. 2011.
- [41] D. Baron, S. Sarvotham, and R. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, 2010.
- [42] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *IEEE Inf. Theory Workshop*, Oct. 2006.
- [43] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.
- [44] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, 2011.
- [45] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE Int. Symp. Information Theory (ISIT)*, Jul.-Aug. 2011.
- [46] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *47th Annual Allerton Conf. Commun., Control, and Comput.*, Sep. 2009.
- [47] G. Reeves and M. Gastpar, "Compressed sensing phase transitions: Rigorous bounds versus replica predictions," in *Conf. Inf. Sci. Syst. (CISS)*, Mar. 2012.
- [48] S. Rangan, A. Fletcher, and V. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1902–1923, 2012.
- [49] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ: Prentice Hall, 2000.
- [50] Y. Wu and S. Verdú, "MMSE dimension," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4857–4879, 2011.
- [51] H. Reboredo, F. Renna, R. Calderbank, and M. R. D. Rodrigues, "Compressive classification," in *IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2013.
- [52] J. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Adapted statistical compressive sensing: Learning to sense gaussian mixture models," in *IEEE Int. Conf. on Acoustics, Speech and Sig. Process. (ICASSP)*, Mar. 2012.
- [53] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, 2003.
- [54] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley, 1991.
- [55] J. Duarte-Carvajalino, G. Sapiro, G. Yu, and L. Carin, "Online adaptive statistical compressed sensing of gaussian mixture models," *arXiv preprint arXiv:1112.5895*, 2011.
- [56] M. S. Srivastava and C. G. Khatri, *An Introduction to Multivariate Statistics*. Amsterdam: North-Holland, 1979.
- [57] J. T. Flåm, S. Chatterjee, K. Kansanen, and T. Ekman, "Minimum Mean Square Error Estimation Under Gaussian Mixture Statistics," *arXiv preprint arXiv:1108.3410*, 2011.
- [58] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.



Francesco Renna (S'09, M'11) received his Laurea Specialistica Degree in Telecommunication Engineering and Ph.D. degree in Information Engineering, both from University of Padova, in 2006 and 2011, respectively. In 2007 he was an intern at Infineon Technology AG, in Villach, Austria. From 2009 to 2014 he held visiting research appointments with the Princeton University, the Georgia Institute of Technology (Lorraine campus), Supelec, the Duke University and the University College London. In 2011 he was at the Department of Information Engineering of University of Padova as a Research Fellow. Since 2012 he is at the Instituto de Telecomunicações, University of Porto, as a Research Fellow. His research interests focus on compressive sensing, physical layer security, multicarrier transmissions and quantum key distribution.



Robert Calderbank (M'89, SM'97, F'98) received the BSc degree in 1975 from Warwick University, England, the MSc degree in 1976 from Oxford University, England, and the PhD degree in 1980 from the California Institute of Technology, all in mathematics.

Dr. Calderbank is Professor of Electrical Engineering at Duke University where he now directs the Information Initiative at Duke (iiD) after serving as Dean of Natural Sciences (2010-2013). Dr. Calderbank was previously Professor of Electrical Engineering and Mathematics at Princeton University where he directed the Program in Applied and Computational Mathematics. Prior to joining Princeton in 2004, he was Vice President for Research at AT&T, responsible for directing the first industrial research lab in the world where the primary focus is data at scale. At the start of his career at Bell Labs, innovations by Dr. Calderbank were incorporated in a progression of voiceband modem standards that moved communications practice close to the Shannon limit. Together with Peter Shor and colleagues at AT&T Labs he showed that good quantum error correcting codes exist and developed the group theoretic framework for quantum error correction. He is a co-inventor of space-time codes for wireless communication, where correlation of signals across different transmit antennas is the key to reliable transmission.

Dr. Calderbank served as Editor in Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY from 1995 to 1998, and as Associate Editor for Coding Techniques from 1986 to 1989. He was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996 and from 2006 to 2008. Dr. Calderbank was honored by the IEEE Information Theory Prize Paper Award in 1995 for his work on the Z4 linearity of Kerdock and Preparata Codes (joint with A.R. Hammons Jr., P.V. Kumar, N.J.A. Sloane, and P. Sole), and again in 1999 for the invention of space-time codes (joint with V. Tarokh and N. Seshadri). He has received the 2006 IEEE Donald G. Fink Prize Paper Award, the IEEE Millennium Medal, the 2013 IEEE Richard W. Hamming Medal, and he was elected to the US National Academy of Engineering in 2005.



Lawrence Carin (SM'96, F'01) earned the BS, MS, and PhD degrees in electrical engineering from the University of Maryland, College Park, in 1985, 1986, and 1989, respectively. In 1989 he joined the Electrical Engineering Department at Polytechnic University (Brooklyn) as an Assistant Professor, and became an Associate Professor there in 1994. In September 1995 he joined the Electrical Engineering Department at Duke University, where he is now professor and Chairman of the Electrical and Computer Engineering Department. He held the William

H. Younger Professorship from 2003-2013 (voluntarily relinquished). He is a co-founder of Signal Innovations Group, Inc. (SIG), a small business. His current research interests include signal processing, sensing, and machine learning. He has published over 250 peer-reviewed papers, and he is a member of the Tau Beta Pi and Eta Kappa Nu honor societies.



Miguel R. D. Rodrigues (S'98, A'02, M'03) received the Licenciatura degree in Electrical Engineering from the University of Porto, Portugal in 1998 and the Ph.D. degree in Electronic and Electrical Engineering from University College London, U.K. in 2002.

He is currently a Senior Lecturer with the Department of Electronic and Electrical Engineering, University College London, UK. He was previously with the Department of Computer Science, University of Porto, Portugal, rising through the ranks from Assistant to Associate Professor, where he also led the Information Theory and Communications Research Group at Instituto de Telecomunicações - Porto. He has held postdoctoral or visiting appointments with various Institutions worldwide including University College London, Cambridge University, Princeton University, and Duke University in the period 2003-2013.

His research work, which lies in the general areas of information theory, communications and signal processing, has led to over 100 papers in journals and conferences to date. Dr. Rodrigues was also honored by the IEEE Communications and Information Theory Societies Joint Paper Award in 2011 for his work on Wireless Information-Theoretic Security (joint with M. Bloch, J. Barros and S. M. McLaughlin).