SIGNAL RECONSTRUCTION IN THE PRESENCE OF SIDE INFORMATION: THE IMPACT OF PROJECTION KERNEL DESIGN

Meng-Yang Chen, Francesco Renna, and Miguel R. D. Rodrigues

Department of Electronic and Electrical Engineering, University College London
{meng-yang.chen.14, f.renna, m.rodrigues}@ucl.ac.uk

ABSTRACT

This paper investigates the impact of projection design on the reconstruction of high-dimensional signals from low-dimensional measurements in the presence of side information. In particular, we assume that both the signal of interest and the side information are described by a joint Gaussian mixture model (GMM) distribution. Sharp necessary and sufficient conditions on the number of measurements needed to guarantee that the average reconstruction error approaches zero in the low-noise regime are derived, for both cases when the side information is available at the decoder or at the decoder and encoder. Numerical results are also presented to showcase the impact of projection design on applications with real imaging data in the presence of side information.

Index Terms— Kernel design, side information, compressive sensing, Gaussian mixture models, minimum mean-squared error.

I. INTRODUCTION

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

However, in many application scenarios, users are offered with further information about the signal of interest – known as *side information* –in the form of a signal correlated with the signal of interest. For example, video streams are usually presented with corresponding audio tracks, high definition images can be presented together with low-resolutions versions, hyperspectral images can be accompanied by RGB versions of the same subject [5], etc.

Recent advances in CS have developed frameworks and reconstruction schemes that leverage the presence of side information to improve reconstruction performance of sparse signals from linear projection measurements [6]–[11].

Of particular relevance, in [5], [12], the impact of side information for both signal classification and reconstruction is studied for the case when the signal of interest and the side information can be described by a joint Gaussian mixture model (GMM). In particular, [5], [12] provide sufficient conditions on the number of measurements required for reliable signal classification and reconstruction, when assuming that random linear projection are extracted from the signal of interest and the side information.

There are multiple reasons for adopting a GMM representation, which can be seen as a union of (linear or affine) subspaces, where each subspace is associated with the translation of the image of the (possibly low-rank) covariance matrix of each Gaussian component within the GMM. In fact, low-rank GMM priors have been shown to approximate signals in compact manifolds [13] and have been shown to provide state-of-the-art results in practical problems in image processing [14], dictionary learning [13], image classification [15] and video compression [16].

This paper asks the question:

- *What is the impact of projection kernel design on the reconstruction performance of CS of GMM signals with side information?*

In particular, we consider projection designs aimed at minimizing the reconstruction minimum mean-squared error (MMSE) and that leverage the presence of side information. We provide necessary and sufficient conditions on the number of designed projections from the signal of interest in order to guarantee that the reconstruction error approaches zero in the low-noise regime. In this way, in fact we generalize the results in [17] to the case when side information is available to the decoder and in the projection design phase.

In the remainder, we denote matrices with boldface upper-case letters ($\mathbf{X}$) and column vectors with boldface lower-case letters ($\mathbf{x}$). The symbols $\mathbf{I}_n$ and $\mathbf{0}_{m \times n}$ represent the identity matrix of dimension $n \times n$ and the all-zero-entries matrix of dimension $m \times n$, respectively (subscripts will be dropped when the dimensions are clear from the context). $\mathbb{E}[\cdot]$ and rank($\cdot$) represent the expectation and the rank operators, respectively. The Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
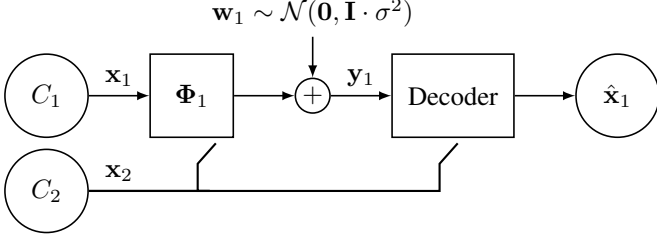
Fig. 1. Compressive sensing with side information model. The side information can be available at the decoder only or else at both the encoder and the decoder.

II. MODEL

We study the problem of the reconstruction of a high-dimensional signal $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ from noisy, linear, compressive measurements $\mathbf{y}_1 \in \mathbb{R}^{m_1}$, with $m_1 \leq n_1$, in the presence of side information $\mathbf{x}_2 \in \mathbb{R}^{n_2}$, which is correlated with the signal of interest $\mathbf{x}_1$. The noisy linear measurements are given by

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

where $\mathbf{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ is the projection kernel and $\mathbf{w}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma^2)$ represents additive white Gaussian noise (AWGN) that models possible errors introduced in the measurement process.

In this work we consider two different scenarios: i) side information is available only *at the decoder*, i.e., $\mathbf{x}_2$ is used only in the reconstruction algorithm; ii) side information is available *at the encoder and the decoder*, i.e., $\mathbf{x}_2$ is used in the reconstruction algorithm as well as in the design of the projection kernel $\mathbf{\Phi}_1$ (See Fig. 1).[1]

The decoder and the encoder have access to the joint probability density function (pdf) of the signal of interest and the side information $p(\mathbf{x}_1, \mathbf{x}_2)$. We will be assuming that $\mathbf{x}_1$ and $\mathbf{x}_2$ are drawn from a joint Gaussian mixture model (GMM). In particular, $\mathbf{x}_1$ and $\mathbf{x}_2$ are characterized by underlying class labels $C_1 \in \{1, \ldots, K_1\}$ and $C_2 \in \{1, \ldots, K_2\}$, respectively, which have joint probability mass function (pmf) $p_{C_1,C_2}(i,k)$. Conditioned on a pair of classes $C_1 = i$ and $C_2 = k$, the joint distribution of $\mathbf{x}_1$ and $\mathbf{x}_2$ is Gaussian, and we have

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

$$= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} P_{C_1,C_2}(i,k) \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}, \boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}), \tag{2}$$

where

$$\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)} \\ \boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)} & \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} \\ \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} & \boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}. \tag{3}$$

[1] In this work we consider the case in which side information is presented at the decoder (and possibly at the encoder) uncompressed and without additive noise.

In other terms, conditioned on class $C_1 = i, C_2 = k$, $\mathbf{x}_1$ is Gaussian distributed with mean $\boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)}$ and covariance $\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)}$, $\mathbf{x}_2$ is Gaussian distributed with mean $\boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)}$ and covariance $\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}$, and the cross-covariance between $\mathbf{x}_1$ and $\mathbf{x}_2$ is given by $\boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}$. The covariance matrices are assumed to be possibly low-rank, and we denote such ranks by $r_{\mathbf{x}}^{(i,k)} = \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)})$, $r_{\mathbf{x}_1}^{(i,k)} = \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})$ and $r_{\mathbf{x}_2}^{(i,k)} = \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})$.

Reconstruction of the signal of interest is performed via the optimal conditional mean estimator

$$\hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{x}_2) = \mathbb{E}[\mathbf{x}_1 | \mathbf{y}_1, \mathbf{x}_2] = \int_{-\infty}^{+\infty} \mathbf{x}_1 p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{x}_2) d\mathbf{x}_1, \tag{4}$$

where $p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{x}_2)$ is the *a posteriori* pdf of $\mathbf{x}_1$ given the compressive measurements $\mathbf{y}_1$ and the side information $\mathbf{x}_2$. Note that the conditional mean estimator in (4) can be also expressed in closed form when $p(\mathbf{x}_1, \mathbf{x}_2)$ is a GMM [5], [13], and it is known to minimize the reconstruction minimum mean-squared error (MMSE)

$$\mathrm{MMSE}(\sigma^2, \mathbf{\Phi}_1) = \mathbb{E}[\|\mathbf{x}_1 - \hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{x}_2)\|^2], \tag{5}$$

which is a function of the joint distribution $p(\mathbf{x}_1, \mathbf{x}_2)$, the projection kernel $\mathbf{\Phi}_1$ and the noise variance σ^2.

We consider both cases of random and designed projection kernels, thus allowing to unveil the impact of projection design on reconstruction performance. In particular, we consider two different scenarios for kernel design, based on the fact that the side information $\mathbf{x}_2$ is offered only at the decoder or both at decoder and encoder:

1) When side information is available at the decoder only, the designed kernel $\mathbf{\Phi}_1^\star$ is obtained as the solution of the optimization problem

$$\begin{aligned} &\underset{\mathbf{\Phi}_1}{\text{minimize}} && \mathrm{MMSE}(\sigma^2, \mathbf{\Phi}_1) \\ &\text{subject to} && \mathrm{tr}(\mathbf{\Phi}_1 \mathbf{\Phi}_1^{\mathrm{T}}) \leq m_1, \end{aligned} \tag{6}$$

where the trace constraint in (6) limits the average energy associated to the projection kernel. We denote the MMSE associated with the optimal kernel design for this case as follows:

$$\mathrm{MMSE}_{\mathrm{d}}^{\mathrm{opt}}(\sigma^2) = \mathrm{MMSE}(\sigma^2, \mathbf{\Phi}_1^\star). \tag{7}$$

2) On the other hand, when $\mathbf{x}_2$ is also available at the encoder, its value is also used to determine the designed kernel that minimizes the MMSE. In particular, for a given value of $\mathbf{x}_2$, the conditional MMSE (conditioned on a specific realization of $\mathbf{x}_2$) is given by

$$\mathrm{MMSE}(\sigma^2, \mathbf{\Phi}_1, \mathbf{x}_2) = \mathbb{E}[\|\mathbf{x}_1 - \hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{x}_2)\|^2 | \mathbf{x}_2], \tag{8}$$

and the designed kernel $\mathbf{\Phi}_1^\star(\mathbf{x}_2)$ is obtained as the solution of the optimization problem

$$\begin{aligned} &\underset{\mathbf{\Phi}_1}{\text{minimize}} && \mathrm{MMSE}(\sigma^2, \mathbf{\Phi}_1, \mathbf{x}_2) \\ &\text{subject to} && \mathrm{tr}(\mathbf{\Phi}_1 \mathbf{\Phi}_1^{\mathrm{T}}) \leq m_1. \end{aligned} \tag{9}$$

We denote the MMSE associated with the optimal kernel design for this case as

$$\text{MMSE}_{\text{ed}}^{\text{opt}}(\sigma^2) = \mathbb{E}[\text{MMSE}(\sigma^2, \mathbf{\Phi}_1^\star(\mathbf{x}_2), \mathbf{x}_2)], \quad (10)$$

where the expectation is taken with respect to $\mathbf{x}_2$.

Note that, in all scenarios, the decoder has perfect knowledge of the projection kernel $\mathbf{\Phi}_1$ adopted to compress $\mathbf{x}_1$.

III. PHASE TRANSITION ANALYSIS

We first determine the impact of projection design on the reconstruction MMSE by analyzing its low-noise *phase transition*. Namely, a phase transition in the MMSE is observed when the number of measurements extracted from the signal of interest m_1 is such that the corresponding MMSE tends to zero when $\sigma^2 \to 0$.

Theorem 1: Consider the measurements model in (1) with side information $\mathbf{x}_2$. Assume $\mathbf{x}_1$ and $\mathbf{x}_2$ are drawn from the joint GMM distribution in (2), which is known at both decoder and encoder. Consider the case when side information $\mathbf{x}_2$ is available at the decoder only, with $\mathbf{\Phi}_1 = \mathbf{\Phi}_1^\star$, where $\mathbf{\Phi}_1^\star$ is the solution of the optimization problem (6). Then,

$$\lim_{\sigma^2 \to 0} \text{MMSE}_{\text{d}}^{\text{opt}}(\sigma^2) = 0 \Rightarrow m_1 \geq \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x_2}}^{(i,k)}, \quad (11)$$

$$m_1 > \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x_2}}^{(i,k)} \Rightarrow \lim_{\sigma^2 \to 0} \text{MMSE}_{\text{d}}^{\text{opt}}(\sigma^2) = 0. \quad (12)$$

Consider now the case in which side information $\mathbf{x}_2$ is available at both encoder and decoder, with $\mathbf{\Phi}_1 = \mathbf{\Phi}_1^\star(\mathbf{x}_2)$, where, for each value of $\mathbf{x}_2$, $\mathbf{\Phi}_1^\star(\mathbf{x}_2)$ is obtained as the solution of the optimization problem (9). Then, it holds

$$\lim_{\sigma^2 \to 0} \text{MMSE}_{\text{ed}}^{\text{opt}}(\sigma^2) = 0 \Rightarrow m_1 \geq \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x_2}}^{(i,k)}, \quad (13)$$

$$m_1 > \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x_2}}^{(i,k)} \Rightarrow \lim_{\sigma^2 \to 0} \text{MMSE}_{\text{ed}}^{\text{opt}}(\sigma^2) = 0. \quad (14)$$

Proof: See the Appendix. ∎

The results in Theorem 1 show that the same necessary and sufficient conditions for the MMSE phase transition hold for both cases when side information is available at the decoder only, or at both decoder and encoder. Moreover, on comparing such conditions with those in [5, Theorem 4 and 5] obtained for the case of random measurements, we can observe that projection design does not guarantee any advantage in terms of MMSE phase transition with respect to using random projections.

On the other hand, projection design has a significant impact on the reconstruction performance when the noise level is not negligible, as testified by the numerical results reported in the next section.
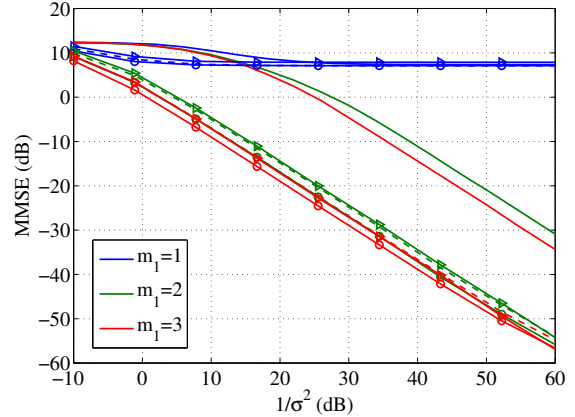
Fig. 2. MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3$ with side information. Side information at the decoder only with random projection kernel (solid lines) and with designed projection kernel (dashed lines). Side information at both the encoder and the decoder with designed projection kernel (circles) and suboptimal design (classification of $\mathbf{x}_2$) (triangles).

IV. NUMERICAL RESULTS

IV-A. Synthetic data

We first provide numerical results on the reconstruction MMSE vs. $1/\sigma^2$ for synthetic data. We consider a joint GMM distribution for $\mathbf{x}_1$ and $\mathbf{x}_2$, with dimensions $n_1 = 10$ and $n_2 = 6$, and $K_1 = K_2 = 2$. All the means in (3) are zero and the covariance matrices in (3) are such that $r_{\mathbf{x}}^{(i,k)} = 5$, $r_{\mathbf{x}_1}^{(i,k)} = 3$ and $r_{\mathbf{x}_2}^{(i,k)} = 3$ for $i = 1, 2$ and $k = 1, 2$.

By using the results in [17], it is possible to predict that phase transition of the MMSE is obtained only when $m_1 > \max_{i,k} r_{\mathbf{x}_1}^{(i,k)}$ for the case when no side information is available. On the other hand, the impact of side information is showcased in Fig. 2, where we report the MMSE values for the case of side information available at the decoder (with random and designed kernels) and at both encoder and decoder. Designed kernels are obtained by approximating numerically the solution of the problems in (6) and (9).

A further projection design scheme for the case of side information at the encoder and the decoder is considered: class labels $\hat{C}_1$ and $\hat{C}_2$ are estimated via a MAP classifier from the side information $\mathbf{x}_2$. Then, $\mathbf{\Phi}_1$ is obtained as the optimal kernel design for Gaussian inputs with distribution $\mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}_1}^{(\hat{C}_1, \hat{C}_2)}, \mathbf{\Sigma}_{\mathbf{x}_1}^{(\hat{C}_1, \hat{C}_2)})$, as described in [17, Theorem 4].

The numerical results are perfectly aligned with the predictions contained in Theorem 1, as we observe that the presence of side information guarantees the MMSE phase transition with $m_1 > \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} = 2$. Moreover, careful design of the projection kernel and the availability of side information at the encoder side do not provide a further advantage in terms of measurements needed for the MMSE phase transition, but they provide significant gains (almost 30 dB) in terms of robustness against noise. We also note

that the suboptimal scheme for the case of side information at both encoder and decoder, though computationally simpler than the design obtained via the solution of the problem in (9), achieves very similar MMSE values.

IV-B. Real data

We now consider a reconstruction example with real imaging data. We use the image "Lena" with resolution 512×512 as the input signal and a low-resolution version (128×128 pixels) of the same image as side information. In this case, the input vectors $\mathbf{x}_1$ represent 8×8 non-overlapping patches extracted from the input image, and the vectors $\mathbf{x}_2$ represent 2×2 patches from low-resolution image. A 20-classes joint GMM distribution describing $\mathbf{x}_1$ and $\mathbf{x}_2$ is trained via the expectation maximization (EM) algorithm [18] using images from the "Caltech 101" dataset [19].

(a) Random kernel, side information at decoder

(b) Designed kernel, side information at decoder

(c) Designed kernel, side information at encoder and decoder

Fig. 3. Reconstruction results of the image "Lena" for $\sigma^2 = -60$ dB. From left to right, the reconstruction PSNR values are 30.7 dB, 36.3 dB, 36.1 dB.

Fig. 3 reports some reconstruction examples obtained from $m_1 = 15$ measurements from each patch and with noise level $\sigma^2 = -60$ dB. From left to right, the recovery images correspond to the following scenarios: (a) random projection with side information at the decoder only; (b) designed projection kernel with side information at the decoder only; (c) designed kernel with side information at both the encoder and the decoder. In this last case, we have considered the design based on the classification of the side information $\mathbf{x}_2$ and on the kernel construction for single Gaussian sources described in [17, Theorem 4]. Then, on comparing case (a) with case (b), we can observe that, although it does not guarantee benefits in terms of phase transition for low-rank GMM sources, kernel design can significantly improve reconstruction quality with respect to random projection kernels. On the other hand, providing the encoder with side information does not guarantee significant advantages. In fact, the suboptimal approach based on the classification of $\mathbf{x}_2$ yields peak signal-to-noise ratio (PSNR) values slightly lower than those obtained with side information at the decoder only.

V. CONCLUSIONS

In this work, we assessed the impact of projection kernel design in signal reconstruction from noisy, linear projec-

tions in the presence of side information. In particular, we have considered the case when both the signal of interest and the side information are jointly described by a GMM distribution, with possibly low-rank, class-conditioned input covariance matrices. We have considered both cases when side information is provided at the decoder and at the encoder and the decoder, i.e., when side information is used in the design of the projection kernel.

Perhaps surprisingly, the main results of this work show that the characterization of the minimum number of measurements needed to drive to zero the reconstruction error in the low-noise regime with designed kernel is the same obtained for random projection kernels. Nevertheless, careful projection kernel design can lead to significant reduction of the reconstruction error for finite noise levels. On the other hand, kernel design schemes that leverage the presence of side information at the encoder and the decoder do not provide significant advantages with respect to kernel designs with side information at the decoder only.

APPENDIX

The sufficient conditions (12) and (14) can be obtained by upper bounding the MMSE associated with designed kernels with the MMSE obtained with random projection kernels and by using the results in [5, Theorem 2].

Consider now the necessary conditions (11) and (13). On adopting the notation $\mathrm{MMSE}(\mathbf{u}|\mathbf{v}) = \mathbb{E}[\|\mathbf{u} - \mathbb{E}[\mathbf{u}|\mathbf{v}]\|^2]$, we can write the MMSE with side information as

$$\mathrm{MMSE}(\sigma^2, \mathbf{\Phi}_1) = \mathbb{E}[\mathrm{MMSE}(\sigma^2, \mathbf{\Phi}_1, \mathbf{x}_2)] \qquad (15)$$
$$= \mathbb{E}[\mathrm{MMSE}(\mathbf{u}|\mathbf{v})], \qquad (16)$$

where the expectation is taken with respect to $\mathbf{x}_2$, and where $\mathbf{u} \sim p(\mathbf{x}_1|\mathbf{x}_2)$ and $\mathbf{v} = \mathbf{\Phi}_1 \mathbf{u} + \mathbf{w}_1$. Then, we observe that the vector $\mathbf{u}$ is distributed according to a GMM with class-conditioned means and covariance matrices given by

$$\boldsymbol{\mu}_{\mathbf{u}}^{(i,k)} = \boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)} + \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})^\dagger (\mathbf{x}_2 - \boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)}) \qquad (17)$$

and

$$\boldsymbol{\Sigma}_{\mathbf{u}}^{(i,k)} = \boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})^\dagger \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}, \qquad (18)$$

respectively, where the symbol $(\cdot)^\dagger$ denotes the Moore-Penrose pseudoinverse [20]. Then, by leveraging [17, Theorem 6], we have that a necessary condition for $\mathrm{MMSE}(\mathbf{u}|\mathbf{v})$ to approach zero in the low-noise regime is given by $m_1 \geq \max_{(i,k)} \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{u}}^{(i,k)})$. Note that such necessary condition holds also when $\mathbf{\Phi}_1$ is designed in order to minimize the MMSE, thus implying that the necessary condition for the phase transition of $\mathrm{MMSE}(\mathbf{u}|\mathbf{v})$ holds for both cases when side information is available at the decoder and at both the encoder and the decoder (i.e., when the designed $\mathbf{\Phi}_1$ is a function of the current realization of $\mathbf{x}_2$). Finally, we note that $\boldsymbol{\Sigma}_{\mathbf{u}}^{(i,k)}$ is the generalized Schur complement of $\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ of the positive semidefinite matrix $\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}$. Then, by using the result in [21], we have that $\mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{u}}^{(i,k)}) = r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)}$, which concludes the proof.

VI. REFERENCES

[1] E.J. 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.L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

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

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

[6] N. Vaswani and Wei 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.F.C. Mota, N. Deligiannis, and M.R.D. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *arXiv preprint arXiv:1408.5250*, 2014.

[11] J.F.C. Mota, N. Deligiannis, A. Sankaranarayanan, V. Cevher, and M.R.D. Rodrigues, "Adaptive-rate sparse signal reconstruction with application in compressive background subtraction," *arXiv preprint arXiv:1503.03231*, 2015.

[12] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M.R.D. Rodrigues, "Classification and reconstruction of compressed GMM signals with side information," in *IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2015.

[13] 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.

[14] 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.

[15] 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.

[16] J. Yang, X. Yuan, X. Liao, P. Llull, D.J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using Gaussian mixture models," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4863–4878, 2014.

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

[18] C.M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[19] Li Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 594–611, 2006.

[20] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

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