PROJECTION DESIGN FOR COMPRESSIVE SOURCE SEPARATION USING MEAN ERRORS AND CROSS-VALIDATION

Dhruv Shah *Ajit Rajwade*

ABSTRACT

This paper addresses the task of projection design for source separation in the compressive domain, where one observes a compressed linear mixture of two source signals with known priors. By positioning that both the sources follow a Gaussian mixture, we formulate an objective that tightly approximates the minimum mean squared error and solve the optimization problem using a gradient-based approach. In the blind setting, where the mixing ratio unknown, we propose a cross-validation approach to independently estimate the mixing ratio. We also provide a number of numerical results on synthetic and real image data that validate our findings. To the best of our knowledge, this is the first effort in projection design for prior-based source separation.

***Index Terms*—** compressed sensing, projection design, source separation, Bayesian estimation, cross-validation

1. INTRODUCTION

Source separation refers to the problem of identifying the constituents of a mixture of signals. It arises in many applications in image source separation [1, 2], audio [3], cancer genetics [4] etc. A large part of literature focuses on the problem of Blind Source Separation (BSS), where one aims to recover unobserved source signals and mixing weights from a linear mixture [5]. The task is intrinsically ill-defined and requires additional assumptions and/or prior knowledge to be solved reliably. Popular methods to address the BSS problem by imposing additional constraints on the source signals include independent component analysis – where the source signals are separated by minimizing mutual information between sources [6], dictionary-based separation by exploiting sparsity priors [7, 8], and informed source separation – where side information is used to assist the separation process [9–11]. [11, 12] consider variants of the source separation problem where MAP estimation is performed on a compressed linear mixture.

Compressed sensing (CS) theory dictates that signals have a fully or approximately sparse representation can be recovered with minimal information loss from certain corrupted lower dimensional linear projections [13, 14]. While CS theory provides guarantees on recovery for randomized projections with entries drawn from Gaussian/Bernoulli distributions, it has been shown that additional information about the signal can be incorporated into the inference task to improve estimation performance [15, 16]. Popular approach to designing projections typically optimize on quality metrics like the mutual coherence, average coherence, estimation mean square error (MSE), entropy of projections etc. under a variety of sensing constraints [12, 16–20].

To incorporate statistical priors efficiently, the Statistical Compressive Sensing (SCS) framework provides performance guarantees for recovery of signals drawn from a Gaussian mixture (GM) prior [21]. GMs are simple yet effective priors on natural image patches degraded by noise, subsampling or linear effects [22, 23] and have been shown to achieve state-of-the-art results in various applications across image and video processing [21–24]. Theoretically, the versatility of GMs comes from the analogy that they can be seen as the Bayesian counterpart of the union-of-subspaces model where each subspace corresponds to the covariance matrix of a component of the GM [25]. While the minimum MSE estimator has analytical form for a linear model with GM priors on signal and noise, the performance measure (MMSE) does not have a closed form. Recent works attempt the projection design problem by optimizing on an analytical upper bound [11, 12] or an analytical lower bound [19].

In this work, we revisit findings on the tightness of these approximations from [26] and formulate a design objective that uses a combination of the two, based on the signal strengths. We pose an optimization problem that capitalizes on a tight approximation of the estimation MSE of *both* constituent signals and solve it using a gradient-based approach. To extend the proposed method to the *blind* separation setting, we also propose a cross-validation (CV) approach to estimate the mixing weights of the linear mixture of the two signals. The remainder of the paper is organized as follows: in Section 2 we describe our imaging model and build the preliminaries for our setup. Section 3 presents the proposed design objective, the optimization problem thus formed and a CV-based scheme for estimating the mixing weights. We validate our approach using experiments on synthetic data and real images in Section 4, highlighting the benefits of the new approach, and conclude in Section 5.

Dhruv Shah (dhruv.shah@iitb.ac.in) is with the Department of Electrical Engineering, Indian Institute of Technology Bombay (IITB). Ajit Rajwade (ajitvr@cse.iitb.ac.in) is with the Department of Computer Science and Engineering, IITB. He recognizes support from IITB Seed Grant 14IRCCSG012.

2. PRELIMINARIES

As alluded above, our goal is to study the source separation problem in settings where we may design the projection matrix $\mathbf{\Phi}$ based on prior information on the signals and noise. We consider the standard linear mixing model:

$$\boldsymbol{y} = \mathbf{\Phi}(\boldsymbol{x} + \lambda \boldsymbol{c}) + \boldsymbol{\eta} \tag{1}$$

where $\{\boldsymbol{x}, \boldsymbol{c}\} \in \mathbb{R}^n$ are the similarly scaled signals composing the linear mixture with mixing ratio λ, and $\boldsymbol{y} \in \mathbb{R}^m$ is the measurement vector corrupted by additive noise $\boldsymbol{\eta}$.

2.1. Acquisition Model

We consider the image acquisition model in [27] (hereby referred to as Block-SPC) which uses a digital micromirror device (DMD) as a spatial light modulator. The scene is divided into *non-overlapping* blocks of a fixed size (say 16×16 when $n = 16$) and sensed independently across blocks. The model can thus be expressed as $\boldsymbol{y}_i = \mathbf{\Phi}(\boldsymbol{x}_i + \lambda \boldsymbol{c}_i) + \boldsymbol{\eta}_i$, where $\{\boldsymbol{x}_i, \boldsymbol{c}_i\}$ are the i^{th} vectorized patches with n resolution elements. The elements of $\mathbf{\Phi}$ are implemented as reflectivity levels of the DMD and hence, a practical sensing matrix faces *optical* constraints. For example, the DMD of Block-SPC is capable of 256 levels of reflectivity, imposing $\Phi_{ij} \in \mathcal{P}$, where $\mathcal{P}$ is the set of $8-$bit uniformly quantized values $\in [0, 1]$.

2.2. Modeling Priors on Images

In this work, we assume that an image is composed of non-overlapping patches, each constituted by a linear mixture of two signals $\{\boldsymbol{x}_i, \boldsymbol{c}_i\}$, drawn independently from known mixture distributions having Q_x and Q_c components, and corrupted by noise $\boldsymbol{\eta}_i$ drawn from a known Gaussian $\mathcal{N}(0, \mathbf{\Sigma}_\eta)$.

$$\begin{aligned} p(\boldsymbol{x}) &= \sum_{j=1}^{Q_x} \pi_{x,j}\, \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu}_{x,j}, \mathbf{\Sigma}_{x,j}), \\ p(\boldsymbol{c}) &= \sum_{j=1}^{Q_c} \pi_{c,j}\, \mathcal{N}(\boldsymbol{c} | \boldsymbol{\mu}_{c,j}, \mathbf{\Sigma}_{c,j}) \end{aligned} \tag{2}$$

Note that the covariance matrices are not assumed to be full-rank. In fact, learned priors on natural image patches have been observed to furnish non-degenerate covariance matrices with fast eigenvalue decay, in analogy to *compressible* signals for conventional CS [21].

2.3. Minimizing the Estimation MSE

For the standard linear setup $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x} + \boldsymbol{\eta}$, the aim of CS is to design a decoder and projection matrix pair that minimizes the mean-squared error (MSE) associated with the recovered signal $\hat{\boldsymbol{x}}$. The projection design problem can be stated as

$$\mathbf{\Phi}^* = \arg\min_{\mathbf{\Phi} \in \mathcal{O}} \min_{\hat{\boldsymbol{x}}_\Phi \in \mathcal{X}} \mathbb{E}_{\boldsymbol{x}, \boldsymbol{\eta}}[\|\boldsymbol{x} - \hat{\boldsymbol{x}}_\Phi(\boldsymbol{y})\|^2] \tag{3}$$

where $\mathcal{O}$ is a (possibly constrained) class of matrices and $\mathcal{X}$ is a (possibly constrained) class of estimation strategies. The nature of constraints imposed by $\mathcal{O}$ governs the nature of the optimization problem posed above. A large body of existing work focuses on *energy-constrained* designs, which impose a constraint on the Frobenius norm of the projection matrix: since each row is a linear operator, this constraint amounts to a constraint on average energy per row of the matrix. While these constraints are well-suited for applications in communications applications and facilitate closed-form optimization using a waterfilling-based approach [17, 29], they do not represent optical constraints levied by a practical compressive imager [27]. In accordance with the acquisition model discussed in Section 2.1, we focus on optically-constrained designs, where $\mathcal{O} = \{\mathbf{\Phi} : \Phi_{ij} \in \mathcal{P}\}$.

3. METHODS

3.1. Optimization Objective for Bayesian Design

We first study the optimization objective for the case of zero-mean Gaussian priors on signals and then extend the idea to GM priors, accordingly. For the Gaussian case with signal covariance matrices $\mathbf{\Sigma}_x, \mathbf{\Sigma}_c$ and noise covariance $\mathbf{\Sigma}_\eta$, the MAP estimates from the compressive mixture (1) are given by

$$\begin{aligned} \hat{\boldsymbol{x}} &= \mathbf{\Sigma}_x \mathbf{\Phi}^T (\mathbf{\Phi} \mathbf{\Sigma}_x \mathbf{\Phi}^T + \mathbf{\Sigma}_{\hat{\eta}_x})^{-1} \boldsymbol{y} \\ \hat{\boldsymbol{c}} &= \lambda \mathbf{\Sigma}_c \mathbf{\Phi}^T (\lambda^2 \mathbf{\Phi} \mathbf{\Sigma}_c \mathbf{\Phi}^T + \mathbf{\Sigma}_{\hat{\eta}_c})^{-1} \boldsymbol{y} \end{aligned} \tag{4}$$

where $\mathbf{\Sigma}_{\hat{\eta}_x} = \lambda^2 \mathbf{\Phi} \mathbf{\Sigma}_c \mathbf{\Phi}^T + \mathbf{\Sigma}_\eta$ and $\mathbf{\Sigma}_{\hat{\eta}_c} = \mathbf{\Phi} \mathbf{\Sigma}_x \mathbf{\Phi}^T + \mathbf{\Sigma}_\eta$ are the effective noise covariance matrices. Estimation theory dictates that the MAP estimate is optimal in the MSE sense [30] and the expected estimation errors $\mathcal{M}_{\mathbf{\Phi},x} = \mathbb{E}[\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2]$ and $\mathcal{M}_{\mathbf{\Phi},c} = \mathbb{E}[\|\boldsymbol{c} - \hat{\boldsymbol{c}}\|_2^2]$ can be computed as traces of the corresponding error covariance matrices

$$\begin{aligned} \mathcal{M}_x &= \text{trace}\Big\{ \overbrace{\mathbf{\Sigma}_x - \mathbf{\Sigma}_x \mathbf{\Phi}^T (\mathbf{\Phi} \mathbf{\Sigma}_x \mathbf{\Phi}^T + \mathbf{\Sigma}_{\hat{\eta}_x})^{-1} \mathbf{\Phi} \mathbf{\Sigma}_x}^{\mathbf{K}_x} \Big\} \\ \mathcal{M}_c &= \text{trace}\Big\{ \underbrace{\mathbf{\Sigma}_c - \mathbf{\Sigma}_c \mathbf{\Phi}^T (\mathbf{\Phi} \mathbf{\Sigma}_c \mathbf{\Phi}^T + \mathbf{\Sigma}_{\hat{\eta}_c})^{-1} \mathbf{\Phi} \mathbf{\Sigma}_c}_{\mathbf{K}_c} \Big\} \end{aligned} \tag{5}$$

We choose an optimization objective of the form $J = \mathcal{M}_x + k \cdot \mathcal{M}_c$, parametrized by $k > 0$ to tune the importance of signal $\boldsymbol{c}$ with respect to $\boldsymbol{x}$. This parameter facilitates flexibility in design objective based on the intended application: choosing $k = 1$ gives equal importance to both signals, whereas $k \ll 1$ would focus on recovery of $\boldsymbol{x}$. For the particular choice of $k = 0$, the objective resembles that of matrix design for compressive clutter removal for Gaussian sources, where the second signal is treated as structured noise.

For the more general scenario of GM priors (2), an analysis of various approximations to the MMSE metric $\mathcal{M}_x$ reveals two viable closed-form estimates: a lower bound based

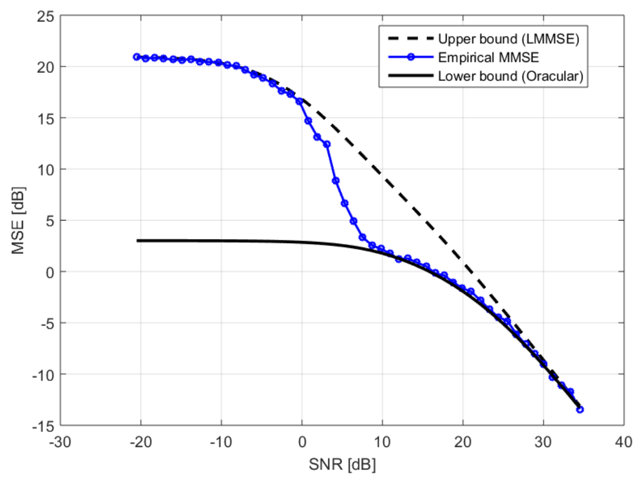


Fig. 1: Empirical MMSE $\mathcal{M}_x$ (blue) and its analytical bounds

on oracular model-selection and an upper bound based on the linear MMSE estimator [26]. We have

$$\mathcal{M}_x^L = \sum_{k,l} \pi_{x,k}\, \pi_{c,l}\, \mathcal{M}_{x|k,l} \;\leq \mathcal{M}_x \leq \mathcal{M}_x^U \tag{6}$$

where $\mathcal{M}_x^L$ is the oracular MMSE, $\mathcal{M}_{x|k,l}$ denotes the Gaussian MMSE corresponding to mixture components k, l in (5) and $\mathcal{M}_x^U$ represents the linear MMSE corresponding to average signal covariances of $\boldsymbol{x}, \boldsymbol{c}$[1]. Figure 1 shows the trends in the empirical MMSE of $\boldsymbol{x}$ and the bounds suggested by (6) as a function of the mixing ratio λ for a simulated linear mixture. The empirical MMSE is computed as the expected MSE of the analytical MMSE estimator over a large number of randomized runs. We note from Fig. 1 that for the regime $\lambda < 0.6$ (SNR > 5), i.e. when signal $\boldsymbol{x}$ is stronger than clutter $\lambda\boldsymbol{c}$, the lower bound $\mathcal{M}_x^L$ is tight and presents a good approximation to the intractable MMSE $\mathcal{M}_x$. Conversely, the upper bound $\mathcal{M}_c^U$ presents a tight approximation to the intractable $\mathcal{M}_c$ over this low SNR regime (with respect to $\boldsymbol{c}$). A suitable objective function for the $\lambda < 0.6$, hence, can be obtained by using appropriate analytical forms for the expected estimation MSEs as follows

$$J = \mathcal{M}_x^L + k\mathcal{M}_c^U \tag{7}$$

Thus using our observations from Fig. 1, we have arrived at a principled criterion for projection design for prior-based source separation.

3.2. Gradient-based Optimization

The optimization objective for projection design thus becomes $\boldsymbol{\Phi}^* = \arg\min_{\boldsymbol{\Phi}\in\mathcal{O}} \mathcal{M}_x^L + k\,\mathcal{M}_c^U$ where $\mathcal{O}$ is as per Section 2.3. The objective function is differentiable and gradients can be computed with respect to elements of $\boldsymbol{\Phi}$

$$\begin{aligned}
\frac{\partial \mathcal{M}_{x|k,l}}{\partial \Phi_{\rho\omega}} &= -\text{trace}\Big\{\mathbf{K}_{x|k,l}^2(\boldsymbol{\Phi}^T\mathbf{J}^{\rho\omega} + \mathbf{J}^{\omega\rho}\boldsymbol{\Phi})\boldsymbol{\Sigma}_{\hat{\eta}_x}^{-1}\Big\} \\
\frac{\partial \mathcal{M}_x^L}{\partial \Phi_{\rho\omega}} &= \sum_{k=1}^{Q_x}\sum_{l=1}^{Q_c} \pi_{x,k}\pi_{c,l} \cdot \frac{\partial \mathcal{M}_{x|k,l}}{\partial \Phi_{\rho\omega}} \\
\frac{\partial \mathcal{M}_c^U}{\partial \Phi_{\rho\omega}} &= -\text{trace}\Big\{\mathbf{K}_c^2(\boldsymbol{\Phi}^T\mathbf{J}^{\rho\omega} + \mathbf{J}^{\omega\rho}\boldsymbol{\Phi})\boldsymbol{\Sigma}_{\hat{\eta}_c}^{-1}\Big\}
\end{aligned} \tag{8}$$

where $\mathbf{K}_c^U, \mathbf{K}_{x|k,l}$ are defined as per (5) for the corresponding distributions, as defined in (6). The constrained minimization problem is solved by projected gradient descent (on the set $\mathcal{O}$) with adaptive step size. A multi-start strategy is adopted to combat the non-convexity of the objective function.

3.3. Cross-validation for Estimating Mixing Ratio

So far, the design strategy assumes knowledge of the mixing ratio to be available a priori. However, a compressive imager may suffer from clutter of unknown strength (usually weaker than the image signal). Instead of imposing a Bayesian hyperprior on the mixing ratio, we use cross-validation (CV), a strategy widely used in CS literature for estimating signal support, noise level, optimizer parameters etc. [8, 31, 32].

For any projection matrix $\boldsymbol{\Phi}$ and compressive measurement $\boldsymbol{y}$ obtained according to (1), we split the measurements into two sets: the estimation set $\boldsymbol{y}_{\text{est}}$ and the validation set $\boldsymbol{y}_{CV}$ (respectively generated by modified projection matrices $\boldsymbol{\Phi}_{\text{est}}$ and $\boldsymbol{\Phi}_{CV}$). We feed the estimation set to the decoder Δ and generate estimates of the signals $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{c}}$. These estimates are then used to evaluate the model parameters (here, λ) over the validation set, i.e., $\boldsymbol{y}_{CV}$

$$\lambda^* = \arg\min_{\lambda\in\mathcal{Y}} \mathbb{E}\left[\|\boldsymbol{y}_{CV} - \boldsymbol{\Phi}_{CV}(\tilde{\boldsymbol{x}} + \lambda\tilde{\boldsymbol{c}})\|^2\right] \tag{9}$$

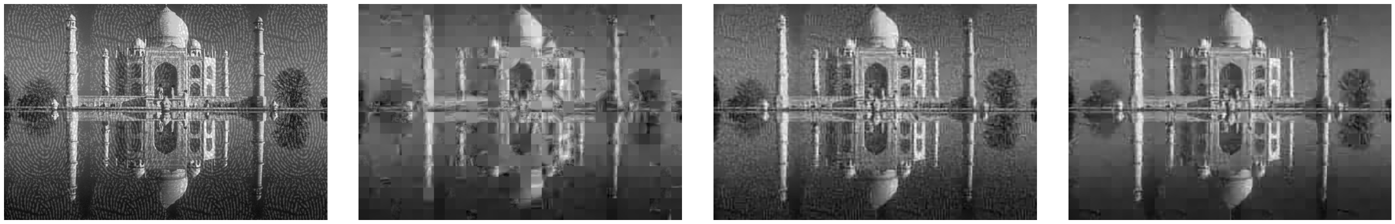
where $\mathcal{Y}$ is the set of candidate mixing ratios and the expectation is taken over all possible partitions.

4. EVALUATION

We evaluate the performance of our proposed method on real and synthetic data of varying complexity. We consider the Block-SPC model with 12.5% compressive measurements.

To demonstrate the fidelity of the proposed objective to the chosen task and its dependence on the knowledge of λ, we conduct a simulation experiment on signals with Gaussian priors and a wide range of mixing ratios λ. Starting with a random $\boldsymbol{\Phi}_0 \in \mathcal{O}$, we design the projection matrix using the proposed method ($k = 0.4$) and compare the recovery performance with matrices designed (a) agnostic to structure of $\boldsymbol{c}$ and (b) using a conservative λ (~ 3). Figure 1 shows the trends in recovery error for both signals as a function of λ. We

[1]the linear estimator fits a Gaussian prior on the source distribution, resulting in the average signal covariance matrix



(a) Mixture ($x + \lambda c$) (b) Random $\mathbf{\Phi}$ (c) $J = \mathcal{M}_x^U$ (d) $J = \mathcal{M}_x^L$

Fig. 2: *Compressive clutter removal*: recovered signals $\hat{\boldsymbol{x}}$ from 12.5% measurements of a binary mixture (a) using: (b) random projections, (c) projections designed using $J = \mathcal{M}_x^U$ (upper bound, LMMSE), and (d) projections designed using $J = \mathcal{M}_x^L$ (lower bound, oracular). PSNR (b-d): 20.01, 20.42, **21.35**. Zoom into electronic version for a better view.

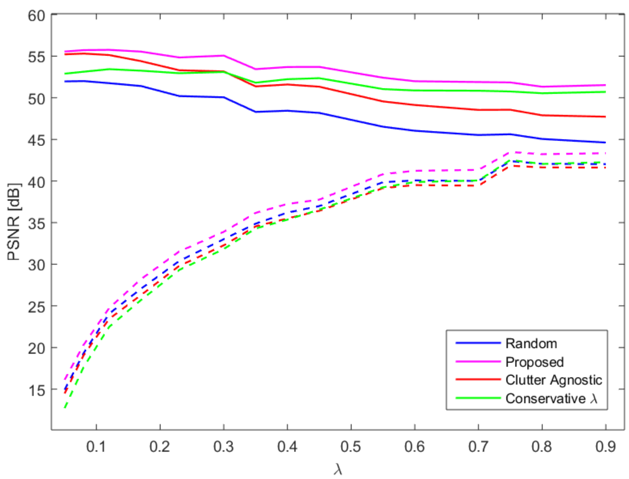


Fig. 3: PSNR trends for compressive source separation of signals $\boldsymbol{x}$ (solid lines) and $\boldsymbol{c}$ (dashed lines) drawn from Gaussian priors using different projections ($k = 0.4$)

observe that the knowledge of λ aids recovery and projections designed with this knowledge outperform the others in terms of PSNR. We also note that the particular choice of $k = 0.4$ lays more emphasis on recovery of $\boldsymbol{x}$: this is evident from the improved reconstruction quality of $\boldsymbol{x}$, when compared to $\boldsymbol{c}$.

To demonstrate the appropriateness of the analytical lower bound, we conduct an experiment with GM priors on both signals with $l_x = l_c = 10$. Each mixture component is zero-mean with fast eigenvalue decay in their covariance matrices. Figure 4 shows the trends in recovery performance of projections designed with the lower bound (proposed, $k = 0$), with the upper bound [12] and random projections. Over the regime $\lambda < 1$, the tightness of the lower bound ensures better recovery of signals with the proposed objective function.

Finally, we show results from the compressive recovery of image signals from a binary mixture ($\lambda = 0.1$). We consider two classes of signals – images of natural scenes and fingerprint images – and independently learn GM priors with 20 components each using MAP-EM on a large set of patches from well-curated datasets. Figure 2 shows a sample mixture and recovered signals using (b) random projections, (c) projections designed with the lower bound, and (d) with the upper bound[2] (proposed, $k = 0$). The proposed method offers better reconstructions, both visually and quantitatively. The reader is directed to the supplemental material [33] for results on a wider set of images.

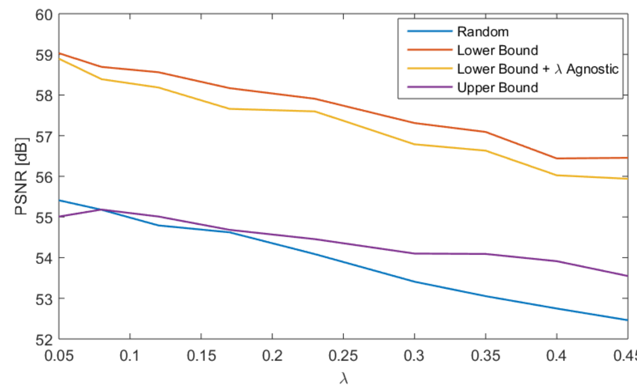


Fig. 4: PSNR trends for compressive source separation of signals $\boldsymbol{x}$ (solid lines) and $\boldsymbol{c}$ (dashed lines) drawn from Gaussian mixture priors using different projections ($k = 0$)

5. CONCLUSION

We investigated the problem of projection design for compressive source separation, with optical constraints and statistical priors on signals. Imposing a GM prior on the constituents of a binary mixture, we propose an optimization objective that closely approximates the MMSE error in closed-form and solve it using a gradient-based approach. We also describe a cross-validation scheme for inferring the mixing ratio λ of the binary mixture, generalizing the estimation process to the blind setting. Matrices designed using the proposed method are superior in terms of visual quality and reconstruction error when compared to random matrices and those designed solely for the upper bound. Experiments on simulated data validate the fidelity of the design scheme and its dependence on the knowledge of λ. A more thorough analysis of this dependence and the sensitivity of the design to stochasticity in λ is deferred to a subsequent effort.

[2]Sensing matrices are appended with a row of 1s to estimate the block mean

References

[1] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Blind Separation of Image Sources via Adaptive Dictionary Learning," *IEEE Trans. on Image Proc.*, vol. 21, 2012.

[2] H. Farid and E. H. Adelson, "Separating Reflections and Lighting Using Independent Components Analysis," in *Proc. CVPR*, 1999.

[3] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, "Underdetermined blind audio source separation using modal decomposition," *EURASIP Journal on Audio, Speech, and Music Proc.*, 2007.

[4] C. Yau, O. Papaspiliopoulos, G. O. Roberts, and C. Holmes, "Bayesian non-parametric hidden markov models with applications in genomics," *Journal of the Royal Statistical Society. Series B*, vol. 73, no. 1, 2011.

[5] P. Grady, B. Pearlmutter, and S. Rickard, "Survey of Sparse and Non-Sparse Methods in Source Separation," *Intl. Journal Imag. Sys. and Tech.*, vol. 15, no. 1, 2005.

[6] A. Hyvärinen and E. Oja, "Independent Component Analysis : Algorithms and Applications," *Neural Networks*, vol. 1, no. 1, 2000.

[7] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[8] P. Boufounos, M. Duarte, and R. Baraniuk, "Sparse Signal Reconstruction from Noisy Compressive Measurements using Cross Validation," in *Proc. SSP*, 2007.

[9] M. Parvaix, L. Girin, and J. Brossier, "A watermarking-based method for informed source separation of audio signals with a single sensor," *IEEE Trans. on Audio, Speech, and Lang. Proc.*, vol. 18, no. 6, 2010.

[10] S. Gorlow and S. Marchand, "Informed separation of spatial images of stereo music recordings using second-order statistics," in *Proc. MLSP*, 2013.

[11] Z. Sabetsarvestani, F. Renna, F. Kiraly, and M. R. D. Rodrigues, "Source Separation in the Presence of Side Information : Necessary and Sufficient Conditions for Reliable De-mixing," in *Proc. GlobalSIP*, 2018.

[12] S. Jain, A. Soni, and J. D. Haupt, "Compressive measurement designs for estimating structured signals in structured clutter: A Bayesian Experimental Design approach," in *Proc. Asilomar Conf. SSC*, 2013.

[13] D. L. Donoho, "Compressed Sensing," *IEEE Trans. on Info. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[14] E. J. Candes, "The Restricted Isometry Property and its Implications for Compressed Sensing," 2008.

[15] M. Elad, "Optimized Projections for Compressed Sensing," *IEEE Trans. on Signal Proc.*, Dec 2007.

[16] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. on Image Proc.*, vol. 18, no. 7, pp. 1395–1408, July 2009.

[17] W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications-Inspired Projection Design with Application to Compressive Sensing," *SIAM Journal on Imaging Sciences*, 2012.

[18] A. Kotwal and A. Rajwade, "Optimizing Matrices For Compressed Sensing Using Existing Goodness Measures," *CoRR*, 2017.

[19] D. Shah, A. Kotwal, and A. Rajwade, "Designing Constrained Projections for Compressed Sensing: Mean Errors and Anomalies with Coherence," in *Proc. GlobalSIP*, 2018.

[20] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *Proc. EUSIPCO*, 2010.

[21] G. Yu and G. Sapiro, "Statistical Compressed Sensing of Gaussian Mixture Models," *IEEE Trans. on Signal Proc.*, vol. 59, no. 12, pp. 5842–5858, Dec 2011.

[22] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. ICCV*, 2011.

[23] G. Yu, G. Sapiro, and S. Mallat, "Solving Inverse Problems with Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity," *CoRR*, 2010.

[24] 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. on Img. Proc.*, 2014.

[25] 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. on Signal Proc.*, 2014.

[26] J. Flam, S. Chatterjee, K. Kansanen, and T. Ekman, "On MMSE Estimation: A Linear Model Under Gaussian Mixture Statistics," *IEEE Trans. on Signal Proc.*, 2012.

[27] R. Kerviche, N. Zhu, and A. Ashok, "Information Optimal Scalable Compressive Imager Demonstrator," in *Proc. IEEE ICIP*, Oct 2014.

[28] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 83–91, March 2008.

[29] L. Wang, M. Chen, M. Rodrigues, D. Wilcox, R. Calderbank, and L. Carin, "Information-Theoretic Compressive Measurement Design," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 39, pp. 1150–64, 2017.

[30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., 1993.

[31] R. Ward, "Compressed Sensing with Cross Validation," *IEEE Trans. on Info. Theory*, vol. 55, no. 12, 2009.

[32] J. Zhang, L. Chen, and Y. Gu, "Cross Validation in Compressive Sensing and its Application OMP-CV Algorithm," *CoRR*, 2016.

[33] Dhruv Shah and Ajit Rajwade, "Supplemental material: Projection design for compressive source separation using mean errors and cross-validation,"

https://www.cse.iitb.ac.in/~ajitvr/Papers/CompressiveSP-ICIP-Supp.pdf.
[34] K. B. Petersen and M. S. Pedersen, "The Matrix Cookbook," 2012.