DESIGNING CONSTRAINED PROJECTIONS FOR COMPRESSED SENSING: MEAN ERRORS AND ANOMALIES WITH COHERENCE

Dhruv Shah[*]   *Alankar Kotwal*[*]   *Ajit Rajwade*[✉]

ABSTRACT

Most existing work in designing sensing matrices for compressive recovery is based on optimizing some quality factor, such as mutual coherence, average coherence or the restricted isometry constant (RIC), of the sensing matrix. In this paper, we report anomalous results that show that such a design is not always guaranteed to improve reconstruction results. We also present a design method based on the minimum mean squared error (MMSE) criterion, imposing priors on signal and noise for natural images, and show that it yields results superior to results from coherence-based methods while taking into account physical constraints on the sensing matrix.

***Index Terms*—** projection design, average coherence, Bayesian estimation, structured sparsity, compressed sensing

1. INTRODUCTION

Compressed sensing theory states that signals having a fully or approximately sparse representation in a dictionary $\mathbf{\Psi}$ can be recovered with zero or minimal information loss from certain linear projections lower in dimension than those suggested by the Nyquist-Shannon sampling theorem [1, 2]. For optimal recovery, the theory imposes constraints on the projection $\mathbf{\Phi}$. It can be shown that matrices drawn from Gaussian or Bernoulli distributions satisfy one such constraint, the restricted isometry (RIP) [3], with overwhelming probability, making them well-suited to compressive recovery.

It has, however, also been shown [4–10] that optimizing $\mathbf{\Phi}$ leads to better-than-random recovery. Due to the exponential computational complexity of the RIP, early efforts towards matrix design turned to bounds like [11], arguing that minimizing the mutual coherence of the dictionary $\mathbf{A} = \mathbf{\Phi\Psi}$ improves a worst-case bound on recovery error. The mutual coherence $\mathbf{A}$ is defined as $\mu_{\max} = \|\hat{\mathbf{A}}^T\hat{\mathbf{A}} - \mathbf{I}\|_\infty$, where $\hat{\mathbf{A}}$ is the column-normalized $\mathbf{A}$. To relax the l_∞ norm, these methods try to minimize some function of off-diagonal entries of the Gram matrix $\mathbf{G} = \hat{\mathbf{A}}^T\hat{\mathbf{A}}$, for example, the average coherence $\mu_{\text{avg}} = \|\mathbf{G} - \mathbf{I}\|_F$ [5–7]. This approach leads us to the anomalous behaviour of the mutual coherence and the RIC, an interesting negative result that we report in the paper.

[*]These authors contributed equally. Dhruv Shah (dhruv.shah@iitb.ac.in) is with the Department of Electrical Engineering, Indian Institute of Technology Bombay (IITB). Alankar Kotwal (aloo@cmu.edu) is with the Robotics Institute, Carnegie Mellon University. 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.

In contrast, some schemes use concepts like the Rényi entropy of the projections [12, 13] or estimation mean-squared error (MSE) [14–17]. [12] proposes a novel communications-inspired design, drawing parallels to precoder design for multiple input, multiple output systems. Bayesian experimental design principles are used in [14] to optimize the estimation MSE with priors on signal, noise and clutter subject to an energy constraint, to which [16] adds information about signal support. [18] designs energy-constrained matrices based on the minimum mean square error (MMSE) in the low-noise regime. In this paper, we replace energy constraints in MMSE design by optical constraints imposed by sensing hardware.

To deal with statistical priors better, [19, 20] formulate a Bayesian framework for signal estimation and projection optimization. Unlike in sparse modeling, where reconstruction involves nonlinear optimization, a piecewise-linear estimator (PLE) with a Gaussian Mixture Model (GMM) prior was introduced in [21], which was extended to the Statistical Compressive Sensing (SCS) framework [22]. GMMs are simple yet effective priors on natural image patches degraded by noise, subsampling or linear effects [21, 23]. Projections for SCS were designed in [7] by minimizing coherence similar to [5], [24] by using information-theoretic metrics to design for reconstruction and classification jointly, and [25], by deriving a closed-form solution minimizing oracle MSE with energy constraints. Our approach brings together the benefits of adapted sensing in the Bayesian framework and the versatility of the PLE and the GMM prior. In this work, we present:

1. An average coherence-based design for the chosen architecture and anomalous behavior in mutual coherences and RICs of designed matrices, an interesting negative result;
2. A novel approach to sensing matrix design, using Bayesian $\mathcal{A}$-optimality within the SCS framework [22] subject to a learned GMM prior on natural image patches and *optical constraints* levied by the acquisition model;
3. A comparison between matrices designed based on average coherence and the proposed MMSE-based design, showing the superiority of the latter approach.

Following a description of acquisition and optical constraints in Section 2, we evaluate a coherence-based design algorithm and report anomalous results in Section 3. Our main contribution, a novel algorithm for optimizing projections of natural images using GMMs, is presented in Section 4. We validate and compare with other methods in Section 5, highlighting the benefits of the new approach, and conclude in Section 6.

2. ACQUISITION MODEL

We consider the acquisition model in [17] (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 measurement can be written as $\boldsymbol{y}_i = \boldsymbol{\Phi}\boldsymbol{x}_i + \boldsymbol{\eta}_i$, where $\boldsymbol{x}_i$ is the i^{th} vectorized patch with n resolution elements, $\boldsymbol{\Phi}$ is the $m \times n$ sensing matrix, and $\boldsymbol{y}_i$ is the $m-$dimensional measurement vector, corrupted by noise $\boldsymbol{\eta}_i$. The elements of the sensing matrix $\boldsymbol{\Phi}$ 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 $\in [0, 1]$. These constraints are more relevant in making Block-SPC sensing matrices realizable than previously considered energy constraints, which are applicable to communications design [12–14].

3. COHERENCE-BASED DESIGN

Following [11], consider solving the compressed sensing problem $\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{\eta}$ with the basis pursuit (BP) solver

$$\hat{\boldsymbol{x}} = \arg\min_{\tilde{\boldsymbol{x}}} \|\tilde{\boldsymbol{x}}\|_1 \text{ such that } \|\boldsymbol{y} - \mathbf{A}\tilde{\boldsymbol{x}}\|_2 \leq \epsilon. \tag{1}$$

Defining $\boldsymbol{x}_s$ to be the best s-sparse approximation to $\boldsymbol{x}$ and if $\|\boldsymbol{\eta}\|_2 \leq \tau$ and $s \leq \frac{1}{2}\left(1 + 1/\mu_{\max}\right)$:

$$\|\boldsymbol{h}\|_2 = \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq C_1(\epsilon + \tau) + C_2\|\boldsymbol{x} - \boldsymbol{x}_s\|_1. \tag{2}$$

C_1 and C_2 are increasing functions of $\mu_{\max}$. Hence, optimizing on $\mu_{\max}$ improves a *worst-case* bound on the recovery error $\|\boldsymbol{h}\|_2$. Similar worst-case RIC-based bounds exist [3].

3.1. Optimizing on Average Coherence

Following prior work [5–7], we use $\mu_{\text{avg}} = \|\hat{\mathbf{A}}^T\hat{\mathbf{A}} - \mathbf{I}\|_{\text{F}}$ instead of $\mu_{\max}$. This simplifies the minimization problem and accounts for the maximum as well as all other off-diagonal elements of $\mathbf{G}$, encouraging incoherence between all $\mathbf{A}$ column pairs, not just the most incoherent one. We impose the optical constraint by solving the constrained minimization using multi-start projected gradient descent with adaptive step size:

$$\hat{\boldsymbol{\Phi}} = \arg\min_{\Phi_{ij} \in \mathcal{P}} \left\|\hat{\mathbf{A}}^T\hat{\mathbf{A}} - \mathbf{I}\right\|_F^2. \tag{3}$$

3.2. Anomalous Observations

For evaluating the algorithm, we fix $\boldsymbol{\Psi}$ as the 2D-DCT dictionary and draw a [0, 1]-uniform random $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$. We run the optimization routine seeded with this random $\boldsymbol{\Phi}$ for different values of m with $n = 16 \times 16$. The corresponding values of coherence $\mu_{\max}$ and restricted isometry constants δ_3 and δ_4 for the original and optimized dictionaries, for a small set of designed matrices, can be found in Table 1.

m	**96**	**128**	**150**	**175**	**200**	**250**
μ_{avg}	0.082	0.070	0.065	0.060	0.056	0.051
	0.078	0.067	0.063	0.057	0.053	0.050
$\mu_{\max}$	0.409	**0.339**	0.338	**0.310**	0.270	0.256
	0.394	**0.371**	0.326	**0.315**	0.268	0.253
δ_3	**0.614**	0.577	0.567	0.495	0.447	**0.386**
	0.701	0.546	0.509	0.466	0.423	**0.405**
δ_4	$\varnothing$	0.718	0.719	0.615	0.575	**0.519**
		0.688	0.644	0.576	0.552	**0.525**

Table 1: Simulation results of various matrix descriptors for $\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Psi} \in \mathbb{R}^{m \times 256}$ for a uniform random seed. For each descriptor, the first row contains values for the initial $\mathbf{A}_0 = \boldsymbol{\Phi}_0\boldsymbol{\Psi}$ and the second row contains values for $\mathbf{A}$ optimized according to (3). $\varnothing$ means that the corresponding quantity could not be computed. Anomalous behavior boldfaced.

$m \rightarrow$	**96**		**128**		**250**	
ID $\downarrow$	$\boldsymbol{\Phi}_0$	$\tilde{\boldsymbol{\Phi}}$	$\boldsymbol{\Phi}_0$	$\tilde{\boldsymbol{\Phi}}$	$\boldsymbol{\Phi}_0$	$\tilde{\boldsymbol{\Phi}}$
1	19.85	20.36	19.99	20.41	20.50	20.56
2	25.95	26.45	26.02	26.49	26.42	26.46
3	19.33	20.03	19.47	20.10	20.06	20.19
4	21.42	22.34	21.55	22.47	22.28	22.33
5	18.44	18.77	18.53	18.86	18.92	18.97
6	20.33	20.57	20.40	20.63	20.48	20.60
7	26.33	27.91	26.42	28.06	26.60	28.71
8	23.20	23.86	23.48	23.94	23.98	23.98

Table 2: PSNR values from reconstruction of eight images from BSDS500 at three different measurement levels, using random ($\boldsymbol{\Phi}_0$) and coherence-optimized sensing ($\tilde{\boldsymbol{\Phi}}$) matrices.

We observe that contrary to the expected behaviour [4–7], the minimization (3) may lead to an increase in mutual coherence or RIC values (see Table 1 for instances). These observations are frequent – for a set of 200 random seeds, 49% of the optimized matrices show anomalies in $\mu_{\max}$ and 40% in δ_s. However, since descending on μ_{avg} even in the above anomalous cases offers better reconstruction (see Table 2), we demonstrate examples where a decrease in $\mu_{\max}$ or δ_s does not guarantee better reconstruction errors, and hence these cannot be a reliable metric for our setup.

Relaxing the ℓ_∞ norm is a heuristic because (2) refers only to $\mu_{\max}$, not μ_{avg}. To the best of our knowledge, there are no theoretical performance bounds on CS with μ_{avg}. Even though using $\mu_{\max}$ or RIC directly (as in [9, 10, 27–29]) is theoretically sound, the anomalous results say that minimizing these quantities may not yield an improvement in terms of reconstruction error. The worst-case bound (2) fails to capture the general behaviour of recovery errors as the sensing matrix changes. A way to capture this behaviour is minimizing *average-case* errors, which is the subject of the next section.

4. MMSE-BASED DESIGN

In this section, we describe our design method considering (a) a statistical model representing natural image patches and (b) statistical properties of noise in compressive measurements.

4.1. Modeling Natural Images

[23] establishes that natural images are well-represented by Gaussian mixtures on small patches. We assume that images are composed of non-overlapping patches $\boldsymbol{x}_i \in \mathbb{R}^n$ drawn from a Gaussian mixture having c components. Each mixture component is parameterized by its weight π_j, $n \times 1$ mean vector $\boldsymbol{\mu}_j$ and $n \times n$ covariance matrix $\boldsymbol{\Sigma}_j$ for $j = 1 \dots c$: $p(\boldsymbol{x}) = \sum_{j=1}^{c} \pi_j \, \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ where $\mathcal{N}(.)$ denotes the multivariate normal distribution. The mixture components can be learned offline using Expectation Maximization (EM). Note that unlike [14, 22], which use the *average* covariance matrix or force zero-mean components, we learn the prior with unconstrained means and full covariance matrices.

4.2. Piecewise-Linear Estimation and Matrix Design

We exploit prior information on the signal (structured sparsity) using the PLE, leveraging strong performance bounds on estimation with Gaussian mixtures using SCS [21, 22]. The GMM-based SCS decoder [1] estimates $\hat{\boldsymbol{x}}$ maximizing the log posterior: $\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x},j} f(\boldsymbol{x}|\boldsymbol{y}, \pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, by first computing component-wise linear MAP estimates (Wiener filter):

$$\hat{\boldsymbol{x}}_j = \boldsymbol{\Sigma}_j \boldsymbol{\Phi}^T (\boldsymbol{\Phi}\boldsymbol{\Sigma}_j\boldsymbol{\Phi}^T + \boldsymbol{\Sigma}_{\boldsymbol{\eta}})^{-1}(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\mu}_j) + \boldsymbol{\mu}_j \quad (4)$$

and then selecting the best (MAP) model $\hat{j}$:

$$\hat{j} = \arg\min_j \|\boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}_j\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\eta}}} + \left\|\hat{\boldsymbol{x}}_j - \boldsymbol{\mu}_j\right\|^2_{\boldsymbol{\Sigma}_j} + \log|\boldsymbol{\Sigma}_j| \quad (5)$$

where $\|.\|_{\mathbf{A}}$ is the weighed ℓ_2 norm with kernel $\mathbf{A}^{-1}$. We can express our design as a Bayesian $\mathcal{A}$-optimality problem [31] to minimize MMSE, which is the average prediction variance, over the design region. For a Gaussian $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, estimation theory [32] says that the MAP decoder (4) is optimal in the MMSE sense, and the expected error $\mathcal{M}_{\boldsymbol{\Phi}} = \mathbb{E}[\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2]$ is the trace of the error covariance matrix $\mathbf{K}$:

$$\mathcal{M}_{\boldsymbol{\Phi}} = \text{trace}\Big\{ \underbrace{\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^T + \boldsymbol{\Sigma}_{\boldsymbol{\eta}})^{-1}\boldsymbol{\Phi}\boldsymbol{\Sigma}}_{\mathbf{K}} \Big\} \quad (6)$$

For a GMM prior with $\mathcal{M}_{\boldsymbol{\Phi},j}$ defined for each component,

$$\mathcal{M}_{\boldsymbol{\Phi}} = \sum_{j=1}^{c} \pi_j \cdot \mathbb{E}\Big[\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2 \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j \Big] = \sum_{j=1}^{c} \pi_j \mathcal{M}_{\boldsymbol{\Phi},j} \quad (7)$$

Our minimization problem thus becomes

$$\hat{\boldsymbol{\Phi}} = \min_{\Phi_{ij} \in \mathcal{P}} \sum_{j=1}^{c} \pi_j \mathcal{M}_{\boldsymbol{\Phi},j}. \quad (8)$$

[1] Alternatively, the approximate MAP decoder may also be used [23, 30]

Notice that our constraints in $\mathcal{P}$ are on the individual elements of $\boldsymbol{\Phi}$, not energy constraints as imposed in [16–18, 25].

4.3. Proposed Design Approach

Similar to Section 3.1, we perform multi-start projected gradient descent with adaptive step size. Computing the gradients of $\mathcal{M}$ with respect to elements of $\boldsymbol{\Phi}$ gives

$$\begin{aligned} \frac{\partial \mathcal{M}_{\boldsymbol{\Phi},j}}{\partial \Phi_{\rho\omega}} &= -\frac{1}{\sigma^2_{\boldsymbol{\eta}}} \text{trace}\Big\{ \mathbf{K}_j^2 (\boldsymbol{\Phi}^T \mathbf{J}^{\rho\omega} + \mathbf{J}^{\omega\rho}\boldsymbol{\Phi}) \Big\} \\ \frac{\partial \mathcal{M}_{\boldsymbol{\Phi}}}{\partial \Phi_{\rho\omega}} &= \sum_{j=1}^{c} \pi_j \cdot \frac{\partial \mathcal{M}_{\boldsymbol{\Phi},j}}{\partial \Phi_{\rho\omega}} \end{aligned} \quad (9)$$

where $\mathbf{K}_j$ is the error covariance matrix of the j^{th} component, (ρ, ω) are indices in $\boldsymbol{\Phi}$ and $\mathbf{J}^{\rho\omega}$ is a matrix with indices (χ, ζ). $(\mathbf{J}^{\rho\omega})_{\chi\zeta} = \boldsymbol{\delta}_{\rho-\chi}\boldsymbol{\delta}_{\omega-\zeta}$ and $\boldsymbol{\delta}_{(.)}$ is the Kronecker delta[2].

5. EVALUATION

We evaluate the performance of our method on natural image datasets. We learn a GMM prior with 100 mixture components using MAP-EM[3] on a set of 2×10^4 16×16 patches from the Berkeley segmentation training data set[4] (BSDS500). For a range of values for m, we seed our algorithm with a [0, 1]-uniform random $\boldsymbol{\Phi}_0$, and run it for 100 iterations.

We test on two different datasets – *unseen* BSDS500 test data and the INRIA Holidays dataset[5] containing natural images with a similar resolution to BSDS500 and a wide range of scene complexities. Compressive measurements are then synthetically made using random as well as designed $\boldsymbol{\Phi}$, with the noise level set to 1% of the measurements.

We compare our results with results from popular algorithms – coherence-optimized design in [6] ($\boldsymbol{\Phi}^{\ell_1}_{\text{opt}}$) and uniform random $\boldsymbol{\Phi}$ with recovery using the PLE as in [22] ($\boldsymbol{\Phi}^{\text{PLE}}_{\text{rand}}$) – as well as baseline CS recovery with a uniform random matrix ($\boldsymbol{\Phi}^{\ell_1}_{\text{rand}}$), imposing optical constraints. For ℓ_1 recovery, we solve the BP problem (1) with the 2D-DCT basis as the sparsifying dictionary using the SPGL1 solver[6].

Figure 1 shows the reconstruction results from 12.5% measurements. The proposed method performs better in terms of peak signal to noise ratio (PSNR) values, and more significantly so in terms of visual quality. Reconstruction using the proposed method also demonstrates significantly lesser *block-seam artifacts* (see the zoomed-in rectangle in red in the mid and bottom images), which is common in most block-based models. This is consolidated by the PSNR values from a diverse set of unseen images at 12.5% and 25% measurements, respectively (see tables 3 and 4)[7].

[2] Please refer to the supplemental material for a proof [34]
[3] Unoptimized MATLAB code sourced from http://prml.github.io/
[4] BSDS500: eecs.berkeley.edu/Research/Projects/CS/vision/bsds/
[5] INRIA Holidays Dataset: lear.inrialpes.fr/ jegou/data.php
[6] SPGL1 Solver: http://www.cs.ubc.ca/$\sim$mpf/spgl1
[7] Superior results using 25 mixture components can be found in [34]

(a) $\mathbf{\Phi}^{\ell_1}_{rand}$ (b) $\mathbf{\Phi}^{\ell_1}_{opt}$ [6] (c) $\mathbf{\Phi}^{\mathrm{PLE}}_{rand}$ [22] (d) Proposed Method

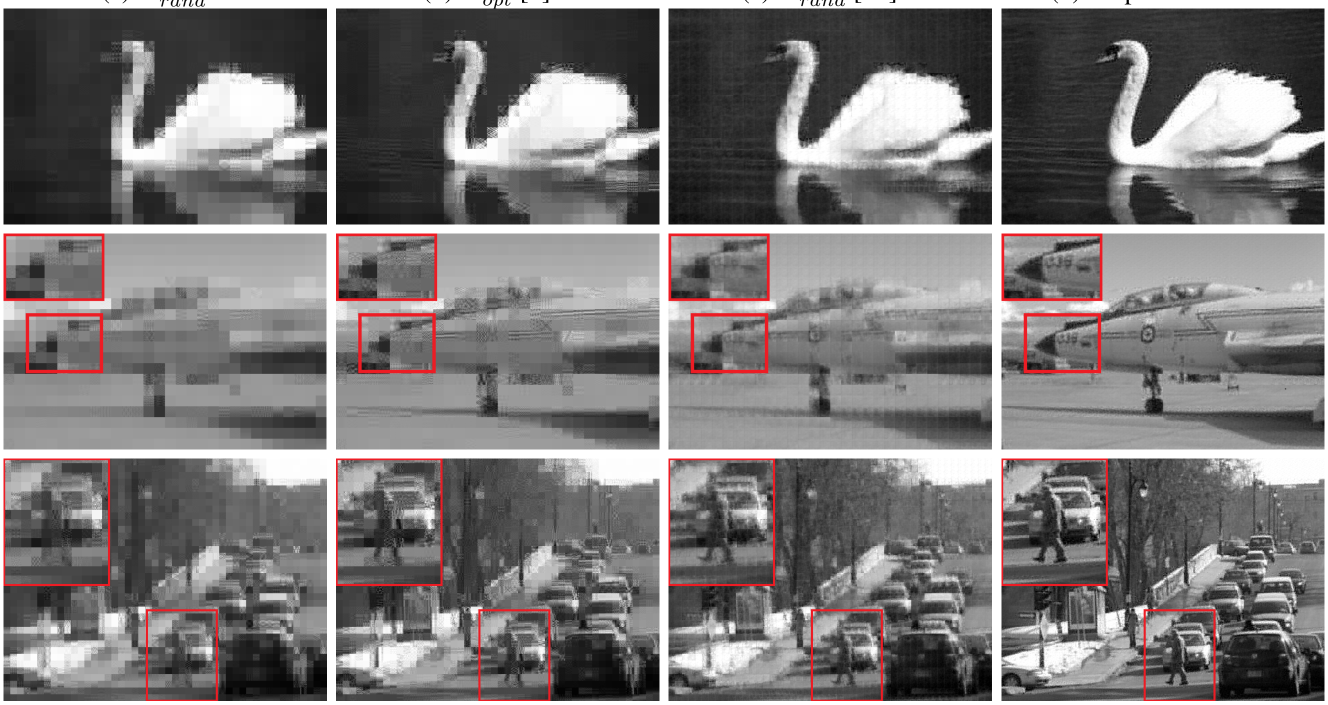

Fig. 1: `bsds500/test/8068` (top), `bsds500/test/10081` (middle) & `inria_holidays/pippin_city66` (bottom) dataset images reconstructed using 12.5% compressive measurements ($m = 32$). Left to right PSNR: $20.269, 21.089, 22.548, 23.844$ (top), $21.57, 22.496, 23.412, 24.844$ (middle) and $18.659, 19.289, 20.447, 21.664$ (bottom). Zoom into electronic version for a better view. Refer to the supplemental material for more results.

Table 3: PSNR values from reconstruction of eight images from BSDS500 at 12.5% measurements ($m = 32$).

Image #	$\mathbf{\Phi}^{\ell_1}_{\mathrm{rand}}$	$\mathbf{\Phi}^{\ell_1}_{\mathrm{opt}}$ [6]	$\mathbf{\Phi}^{\mathrm{PLE}}_{\mathrm{rand}}$ [22]	Proposed
1	18.1798	18.9733	20.1748	**21.1772**
2	25.1973	25.335	26.5923	**27.2622**
3	18.3463	18.6617	19.8185	**20.9138**
4	19.8323	21.0297	21.8075	**23.0925**
5	17.6444	17.6018	18.7195	**19.6204**
6	19.9804	19.8052	20.8265	**21.5922**
7	26.1505	26.6871	27.7679	**29.0994**
8	21.8317	22.0654	23.4717	**24.5041**

Table 4: PSNR values from reconstruction of eight images from BSDS500 at 25% measurements ($m = 64$).

Image #	$\mathbf{\Phi}^{\ell_1}_{\mathrm{rand}}$	$\mathbf{\Phi}^{\ell_1}_{\mathrm{opt}}$ [6]	$\mathbf{\Phi}^{\mathrm{PLE}}_{\mathrm{rand}}$ [22]	Proposed
1	19.0832	19.9328	20.6108	**21.2366**
2	25.4462	26.0214	26.8775	**27.3121**
3	18.7209	19.6227	20.2133	**21.0092**
4	20.768	21.9571	22.296	**23.1672**
5	17.9932	18.3577	19.0837	**19.7392**
6	20.0471	20.2565	21.0641	**21.6476**
7	26.2	27.382	28.1836	**29.223**
8	22.5409	23.2168	23.9792	**24.6183**

6. CONCLUSION

We investigated the problem of projection design for compressive sensing with optical constraints. Having seen cases where average coherence optimization improves recovery even when mutual coherence or RIC increases, we are convinced that that these may not always be reliable metrics for sensing matrix design. We then turn to an average-case error-based design method. We model image patches as drawn from a Gaussian mixture, a good prior for natural images. Matrices designed using the proposed method are superior in terms of recovery error and visual clarity when compared to random matrices and those designed with average coherence.

We plan to extend this work to other signal priors, for example using a Poisson noise model with a Gamma mixture prior. Using this setup in source separation and estimation with clutter is being explored. Applications of this method in other domains like designing optimal trajectories in $k-$space for MR acquisition are also being looked into.

Supplemental material: For the authors' implementation and additional results, refer to the supplemental material [34].

References

[1] D. L. Donoho, “Compressed ensing,” *IEEE Trans. on Info. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[2] E. J. Candes and M. B. Wakin, “An Introduction To Compressive Sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[3] E. J. Candes, “The Restricted Isometry Property and its Implications for Compressed Sensing,” 2008.

[4] M. Elad, “Optimized Projections for Compressed Sensing,” *IEEE Trans. on Signal Proc.*, Dec 2007.

[5] 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.

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

[7] J. M. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, “Adapted statistical compressive sensing: Learning to sense gaussian mixture models,” in *Proc. IEEE ICASSP*, March 2012, pp. 3653–3656.

[8] M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, “Adaptive matrix design for boosting compressed sensing,” *IEEE Trans. on Circuits and Systems*, March 2018.

[9] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, “Compressive Coded Aperture Spectral Imaging: An Introduction,” *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 105–115, Jan 2014.

[10] I. A. Gkioulekas and T. Zickler, “Dimensionality Reduction Using the Sparse Linear Model,” in *Advances in Neural Information Processing Systems (NIPS)*. 2011.

[11] T. T. Cai, L. Wang, and G. Xu, “Stable recovery of sparse signals and an oracle inequality,” *IEEE Trans. on Info. Theory*, vol. 56, no. 7, pp. 3516–22, July 2010.

[12] 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.

[13] 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.

[14] 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.

[15] A. Kotwal and A. Rajwade, “Optimizing matrices for compressed sensing using existing goodness measures: Negative results, and an alternative,” *arXiv*, 2017.

[16] B. Li, L. Zhang, T. Kirubarajan, and S. Rajan, “A projection matrix design method for MSE reduction in adaptive compressive sensing,” *Signal Processing*, vol. 141, pp. 16 – 27, 2017.

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

[18] 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.*, May 2014.

[19] S. Ji, Y. Xue, and L. Carin, “Bayesian Compressive Sensing,” *IEEE Trans. on Signal Proc.*, 2008.

[20] MW. Seeger and H. Nickisch, “Compressed sensing and bayesian experimental design,” in *Proc. ICML*, 2008.

[21] G. Yu, G. Sapiro, and S. Mallat, “Solving Inverse Problems with Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity,” *ArXiv*, 2010.

[22] 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.

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

[24] J. M. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, “Task-Driven Adaptive Statistical Compressive Sensing of Gaussian Mixture Models,” *IEEE Trans. on Signal Proc.*, vol. 61, no. 3, pp. 585–600, Feb 2013.

[25] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, “Projection design for statistical compressive sensing: A tight frame based approach,” *IEEE Trans. on Signal Proc.*, vol. 61, no. 8, pp. 2016–2029, 2013.

[26] 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.

[27] A. Kotwal and A. Rajwade, “Optimizing Codes for Source Separation in Compressed Video Recovery and Color Image Demosaicing,” *arXiv*, 2016.

[28] R. Obermeier and J. A. Martinez-Lorenzo, “Sensing Matrix Design via Mutual Coherence Minimization for Electromagnetic Compressive Imaging Applications,” *IEEE Trans. on Comp. Imag.*, pp. 217–229, June 2017.

[29] D. Bryant, C. J. Colbourn, D. Horsley, and P. Cathin, “Compressed Sensing With Combinatorial Designs: Theory and Simulations,” *IEEE Trans. on Info. Theory*, vol. 63, no. 8, pp. 4850–59, Aug 2017.

[30] M. A. Carreira-Perpinan, “Mode-finding for mixtures of gaussian distributions,” *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 22, no. 11, pp. 1318–23, 2000.

[31] K. Chaloner and I. Verdinelli, “Bayesian experimental design: A review,” *Statistical Science*, 1995.

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

[33] K. B. Petersen and M. S. Pedersen, “The Matrix Cookbook,” 2012.

[34] D. Shah, “Supplemental Material,” prieuredesion.github.io/ constrained-projections/, 2018.