KNOWLEDGE-ENHANCED COMPRESSIVE MEASUREMENT DESIGNS FOR ESTIMATING SPARSE SIGNALS IN CLUTTER*

SWAYAMBHOO JAIN, AKSHAY SONI, JARVIS HAUPT,
NIKHIL RAO, AND ROBERT NOWAK

Abstract. In statistical estimation tasks arising for example in compressive sensing applications, we are often equipped with prior knowledge about the object we wish to infer (e.g., smoothness, characterized by the presence of only low-frequency components in the Fourier domain; a priori region of interest knowledge; or shared features extracted from sets of training data similar to the signal being acquired). This work investigates an estimation task where the goal is to estimate an unknown signal (vector) from compressive measurements corrupted by additive pre-measurement noise ("clutter") as well as post-measurement noise, but where some (perhaps limited) prior knowledge on the signal, clutter, and noise is available. We pose the overall problem as an optimization, whose goal is to minimize the mean square error incurred in estimating the signal of interest, and derive a tractable convex program for designing knowledge-enhanced compressive measurement operators in such settings. We demonstrate, via simulation, the improvements of our proposed approach relative to traditional CS techniques that obtain measurements using iid random sensing matrices.

Key words. compressive sensing, sparse estimation, experimental design

1. Introduction. This paper examines an experimental design problem arising in statistical estimation tasks. Let $\mathbf{x} \in \mathbb{R}^p$ denote our object of interest, and suppose that we obtain n noisy measurements of $\mathbf{x} \in \mathbb{R}^p$ according to the model

$$\mathbf{y} = \mathbf{A}(\mathbf{x} + \mathbf{c}) + \mathbf{w}. \tag{1.1}$$

Here, we may view $\mathbf{A}$ as an $n \times p$ "sensing matrix", $\mathbf{c}$ as a $p \times 1$ vector of pre-measurement noise or "clutter," and $\mathbf{w}$ as a $n \times 1$ vector that represents additive measurement noise. Investigations of this general model in settings where $n < p$ are at the heart of the literature in compressive sensing (CS), and much work has been done on the development and analysis of sensing and inference procedures that aim to estimate $\mathbf{x}$ from such noisy linear measurements in the case where $\mathbf{x}$ is *sparse* (i.e., when $\mathbf{x}$ has at most $k \ll p$ nonzero or significant entries).

A canonical result in this domain is the following: let $\mathbf{c} = \mathbf{0}$ and $\mathbf{w} \sim \mathcal{N}(0, \mathbf{I}_{n\times n})$, and suppose that $\mathbf{A}$ is a randomly generated matrix whose entries are iid $\mathcal{N}(0, 1/n)$. In this case, sparse $\mathbf{x}$ having no more than k nonzero entries can be accurately estimated from a collection of $n = O(k \log p)$ compressive measurements, in the sense that an estimate $\widehat{\mathbf{x}}$ can be obtained from $\{\mathbf{y}, \mathbf{A}\}$ which satisfies $\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2 \leq \text{const. } k\sigma^2 \log p$ with high probability (see, for example, [1]). Several works in the CS literature have examined the effects of clutter in compressive measurements (i.e., the case $\mathbf{c} \neq \mathbf{0}$), but these investigations have typically been limited to the case where the clutter is modeled as white Gaussian noise [2, 3, 4].

Here we focus on a *knowledge-enhanced* estimation problem associated with the compressive measurements obtained via the model (1.1). Our aim is to estimate $\mathbf{x}$, and we assume that we are equipped with some additional *prior knowledge* about $\mathbf{x}$, $\mathbf{c}$, and $\mathbf{w}$. The prior knowledge about $\mathbf{x}$ could describe, for example, a small collection of possible supports (locations where $\mathbf{x}$ takes its nonzero values) and their relative

*S.J., A.S., and J.H. are with the Dept. of Electrical and Computer Engineering, Univeristy of Minnesota, Minneapolis, MN; email: {jainx174, sonix022, jdhaupt}@umn.edu. N.R. and R.N. are with the Dept. of Electrical and Computer Engineering, Univeristy of Wisconsin – Madison; email: nrao2@wisc.edu, nowak@ece.wisc.edu. The student authors S.J., N.R, and A.S. all contributed equally to this work. This work was supported by DARPA/ONR Award No. N66001-11-1-4090.

frequencies of occurrence, or correlation structure among the nonzeros of $\mathbf{x}$. Likewise, prior knowledge about $\mathbf{c}$ and $\mathbf{w}$ may identify the correlation structures or supports of each. The question we address here is, how should we design the sensing matrix $\mathbf{A}$ to take advantage of this prior knowledge?

The main contribution of this work is to demonstrate that knowledge-enhanced sensing matrix designs can significantly outperform the more "traditional" iid random sensing matrices in noisy compressive sensing estimation tasks. We formulate the sensing matrix design task as an optimization, with the aim to minimize the mean-square error (MSE) of the *best possible* estimator of $\mathbf{x}$. We describe the prior information on the quantities $\mathbf{x}$, $\mathbf{c}$, and $\mathbf{w}$ in terms of distributions with known first- and second-order statistics, and show that, following a few modest simplifications, the overall sensing matrix design task in this setting can be performed by solving a tractable *convex optimization* (eqn. (4.1)) followed by simple matrix factorization. We demonstrate the performance improvements resulting from our approach, relative to traditional CS techniques using iid random sensing matrices, via simulation.

The remainder of this paper is organized as follows. Following a brief discussion of our contribution in the context of existing work (below), we describe our overall problem setting in Section 2. In Section 3 we derive our main result – a convex program for "knowledge-enhanced" sensing matrix design. We provide experimental validation of our approach in Section 5, and briefly discuss extensions in Section 6.

1.1. Connections with Prior Works. The work [5] proposed one of the first approaches to design compressive sensing matrices given some prior signal knowledge. That work considered noise-free settings and assumed knowledge of a dictionary in which the signals being observed were sparse, and proposed a sensing matrix design procedure whose aim is to reduce the coherence between the learned sensing matrix and the known dictionary. Extensions of this idea aimed at designing both the dictionary and the sensing matrix given a collection of training data were examined by [6]. The recent work [7] examined knowledge-enhanced CS design tasks assuming a Gaussian mixture prior on the signal being acquired, and proposed a design criteria based on coherence minimization between the learned sensing matrix and a dictionary composed of eigenvectors of the mixture covariance matrices. Along these lines, the work [8] examined sensing designs based on learned correlations in training data. We note that none of these approaches utilize the statistical estimation theoretic formulation we adopt here, nor do they explicitly treat the separation from clutter, as here.

Most closely related to our formulation here is the body of prior work on optimal designs for space-time linear coding in MIMO applications – see, for example, [9], which examined qualitatively similar estimation problems. Inclusion of the clutter term here makes the overall problem significantly more challenging in the more general settings we consider here (e.g., when signal and clutter cannot be simultaneously decorrelated).

2. Problem Statement. As stated above, our ultimate inference goal is to accurately estimate the vector $\mathbf{x}$ given measurements of the form (1.1). In this section we describe our overall sensing matrix design methodology.

2.1. Quantifying Prior Information. In our approach here we will assume that the vector $\mathbf{x} \in \mathbb{R}^p$ that we wish to estimate is a random quantity drawn from an m_x-component mixture distribution (m_x an integer). We do not assume full knowledge of the distribution, but only that i-th mixture component has known weight $\pi_{x,i}$ and is a p-dimensional zero-mean random vector with known $p \times p$ covariance matrix

$\mathbf{\Sigma}_{x,i}$, for $i = 1, 2, \ldots, m_x$. Vectors drawn from a given mixture component (described by a particular covariance matrix $\mathbf{\Sigma}_{x,i}$) have common supports and share the same correlation structure among their nonzero elements.

We note that the covariance matrices $\mathbf{\Sigma}_{x,i}$ are not assumed here to be full-rank. On the contrary, rank-deficiency in any of the $\mathbf{\Sigma}_{x,i}$ amounts to a form of *sparsity*, as random vectors $\mathbf{x} \in \mathbb{R}^p$ drawn from a distribution with covariance matrix of rank $r < p$ inherently lie on a r-dimensional subspace of $\mathbb{R}^p$. Thus, the formulation described here can model various forms of sparsity and structure that have been studied in the literature, including simple k-sparse vectors (the collection of $\mathbf{\Sigma}_{x,i}$ describe all $m_x = \binom{p}{k}$ unique subsets of $\{1, 2, \ldots, p\}$ of cardinality k, and $\pi_i = 1/m_x$), block sparsity, group sparsity (with potentially overlapping groups), tree sparsity, and so on. It is worth noting that our model does not assume that vectors drawn from different models be orthogonal, though that structure, if present, could easily be captured by this formulation.

We use similar prior distributions on the clutter $\mathbf{c}$ and the additive measurement noise $\mathbf{w}$. In particular, we model the clutter $\mathbf{c}$ as a realization of an m_c-component mixture distribution whose i-th mixture component has weight $\pi_{c,i}$ and is a zero-mean random vector with covariance matrix $\mathbf{\Sigma}_{c,i}$, for $i = 1, 2, \ldots, m_c$. Likewise, we can model the noise $\mathbf{w}$ generally as a realization of an m_w-component mixture distribution whose i-th mixture component has weight $\pi_{w,i}$ and is a zero-mean random vector with covariance matrix $\mathbf{\Sigma}_{w,i}$, for $i = 1, 2, \ldots, m_w$, though in the exposition that follows we will consider $\mathbf{w}$ to be additive uncorrelated zero-mean noises with variance σ^2. We assume that the random quantities $\mathbf{x}$, $\mathbf{c}$, and $\mathbf{w}$ are uncorrelated.

2.2. Minimizing the Estimation MSE. Our aim here is to minimize the mean-square error (MSE) associated with our estimate of the signal $\mathbf{x}$. Let us denote by $\widehat{\mathbf{x}}_{\mathbf{A}} = \widehat{\mathbf{x}}_{\mathbf{A}}(\mathbf{y})$ an estimator of $\mathbf{x}$, which is a function of the measurements $\mathbf{y}$ obtained via (1.1) using a particular $n \times p$ sensing matrix $\mathbf{A}$. The mean-square error of a given $\widehat{\mathbf{x}}$ is $d_{\mathrm{MSE}}(\widehat{\mathbf{x}}_{\mathbf{A}}) \triangleq \mathbb{E}_{\mathbf{x},\mathbf{c},\mathbf{w}}\left[\|\mathbf{x} - \widehat{\mathbf{x}}_{\mathbf{A}}(\mathbf{y})\|^2\right]$, where the subscript denotes that the expectation is with respect to all of the random quantities. The criteria for optimal design of the sensing matrix $\mathbf{A}$ in this case can be stated as an optimization – the optimal choice of $\mathbf{A}$, denoted by $\mathbf{A}^*$, is

$$\mathbf{A}^* = \arg\min_{\mathbf{A}\in\mathcal{A}} \min_{\widehat{\mathbf{x}}_{\mathbf{A}}\in\mathcal{X}} d_{\mathrm{MSE}}(\widehat{\mathbf{x}}_{\mathbf{A}}), \tag{2.1}$$

where $\mathcal{A}$ is a (possibly constrained) class of sensing matrices and $\mathcal{X}$ is the class of possible estimators. In words, $\mathbf{A}^* \in \mathcal{A}$ is the sensing matrix yielding measurements for which the MSE of the *best possible* estimator (from the class $\mathcal{X}$) is minimum.

It is worth noting that the presence of the measurement noise $\mathbf{w}$ is only relevant when the sensing matrix $\mathbf{A}$ is constrained in some way – else, one could simply scale each of the elements of $\mathbf{A}$ toward infinity, making the overall effect on $\mathbf{w}$ negligible in the estimation task. Here our focus will be on *energy-constrained* designs $\mathbf{A}$; in particular, we choose $\mathcal{A}$ in (2.1) as $\mathcal{A} = \{\mathbf{A} : \|\mathbf{A}\|_F = \alpha\}$ for some (specified) $\alpha > 0$, where the notation $\|\cdot\|_F$ denotes the matrix Frobenius norm.

3. Knowledge-enhanced Sensing Matrix Designs to Minimize MSE. It is well-known from statistical estimation theory that, for the minimum MSE task (MMSE) task described above, the optimal estimator of $\mathbf{x}$ is the conditional mean $\mathbf{x}$ given the observations $\mathbf{y}$; that is, $\widehat{\mathbf{x}}_{\mathbf{A},\mathrm{MMSE}}(\mathbf{y}) = \mathbb{E}[\mathbf{x}|\mathbf{y}]$. Here, our prior knowledge is limited to first- and second-order statistics of the signal, clutter, and noise, and without full knowledge of the distributions we are unable to compute this estimator

in closed form. Instead, we consider restricting the class of estimators $\mathcal{X}$ in (2.1) to be the class of *linear* estimators of $\mathbf{x}$, as described below.

Let us define the *average* signal covariance matrix $\mathbf{\Sigma}_x$ as $\mathbf{\Sigma}_x = \sum_{i=1}^{m_x} \pi_{x,i} \mathbf{\Sigma}_{x,i}$, and similarly for $\mathbf{\Sigma}_c$, and we assume that $(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)$ is invertible.[1] Further, as alluded above, we make a standard assumption on the additive noise, that it be uncorrelated, and each element w_i of $\mathbf{w}$ has the same variance; here, this amounts to the choices $m_w = 1$ and $\pi_{w,1} = 1$, thus $\mathbf{\Sigma}_{w,1} = \mathbf{\Sigma}_w = \sigma^2 \mathbf{I}_{n\times n}$. In this case, the linear MMSE estimator is just the Wiener Filter, easily shown here to be $\widehat{\mathbf{x}}_{\mathbf{A},\mathrm{LMMSE}}(\mathbf{y}) = \mathbf{\Sigma}_x \mathbf{A}' \left(\mathbf{A}\left(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\right)\mathbf{A}' + \sigma^2 \mathbf{I}_{n\times n}\right)^{-1} \mathbf{y}$, where $\mathbf{A}'$ denotes the matrix transpose. It follows (after a bit of algebra) that $\mathbb{E}_{\mathbf{x},\mathbf{c},\mathbf{w}}\left[\|\mathbf{x} - \widehat{\mathbf{x}}_{\mathbf{A},\mathrm{LMMSE}}(\mathbf{y})\|^2\right] = \mathrm{tr}\{\mathbf{\Sigma}_x - \mathbf{\Sigma}_x \mathbf{A}' \left(\mathbf{A}\left(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\right)\mathbf{A}' + \sigma^2 \mathbf{I}_{n\times n}\right)^{-1} \mathbf{A}\mathbf{\Sigma}_x\}$, where $\mathrm{tr}\{\cdot\}$ denotes the matrix trace (the sum of the diagonal elements). Thus, we can express our sensing matrix design task as the optimization

$$\mathbf{A}^* = \arg\max_{\mathbf{A}\in\mathcal{A}} \ \mathrm{tr}\left\{\mathbf{\Sigma}_x \mathbf{A}' \left(\mathbf{A}\left(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\right)\mathbf{A}' + \sigma^2 \mathbf{I}_{n\times n}\right)^{-1} \mathbf{A}\mathbf{\Sigma}_x\right\}. \tag{3.1}$$

Now, we approximate the objective function by replacing the inverse term by its linear approximation. That is, we use the fact that

$$\left(\mathbf{A}\left(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\right)\mathbf{A}' + \sigma^2 \mathbf{I}_{n\times n}\right)^{-1} \approx \frac{1}{\sigma^2}\left(\mathbf{I}_{n\times n} - \frac{1}{\sigma^2}\mathbf{A}\left(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\right)\mathbf{A}'\right), \tag{3.2}$$

which follows from truncating the power series expansion of the inverse, which is valid when all eigenvalues of the matrix $\mathbf{A}\left(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\right)\mathbf{A}'/\sigma^2$ are less than 1 in magnitude. Equivalently, since $\mathbf{A}\left(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\right)\mathbf{A}'$ is symmetric and positive semidefinite (PSD) its eigenvalues and singular values coincide, implying that the power series expansion is valid if and only if

$$\|\mathbf{A}\left(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\right)\mathbf{A}'\| < \sigma^2, \tag{3.3}$$

where the notation $\|\cdot\|$ denotes the spectral norm (the largest singular value) of the matrix argument. For the purposes of the derivation here, the condition (3.3) is taken as an *assumption* – it can later be verified for a particular solution $\mathbf{A}^*$. Using this approximation we can "complete the square," and use the assumed invertibility of $(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)$, to obtain that (3.1) can be approximated by the optimization

$$\mathbf{A}^* = \arg\min_{\mathbf{A}\in\mathcal{A}} \left\|(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)^{1/2}\mathbf{A}'\mathbf{A}\mathbf{\Sigma}_x - \frac{\sigma^2}{2}(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)^{-1/2}\mathbf{\Sigma}_x\right\|_F^2. \tag{3.4}$$

4. Sensing Matrix Design by Convex Optimization. Note that the optimization (3.4) is over $\mathbf{A}$, but $\mathbf{A}$ appears only in its quadratic form $\mathbf{A}'\mathbf{A}$ (since, in particular, the constraint $\|\mathbf{A}\|_F = \alpha$ for $\mathbf{A} \in \mathcal{A}$ is equivalent to $\mathrm{tr}\{\mathbf{A}'\mathbf{A}\} = \alpha^2$). This suggests that rather than solve (3.4) directly, we instead solve the optimization over the quadratic form, then "factor" the solution to obtain the optimal $\mathbf{A}^*$. Specifically, we define $\mathbf{M} = \mathbf{A}'\mathbf{A}$, and as a first step to solving (3.4) we solve the following *convex* optimization over the cone of symmetric PSD matrices $\mathbf{M}$:

$$\begin{aligned}\mathbf{M}^* = \arg\min_{\mathbf{M}} \ &\left\|(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)^{1/2}\mathbf{M}\mathbf{\Sigma}_x - \frac{\sigma^2}{2}(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)^{-1/2}\mathbf{\Sigma}_x\right\|_F \\ &\text{s.t. } \mathrm{tr}(\mathbf{M}) = \alpha^2;\ \mathbf{M} \succeq \mathbf{0},\ \text{symmetric.}\end{aligned} \tag{4.1}$$

[1] Invertibility of $(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)$ is a mild assumption, easily enforced in practice. In particular, if $(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)$ happens to be rank deficient (having, say, rank $\widetilde{p} < p$), we can simply restrict the action of the sensing matrix to the $\widetilde{p}-$dimensional subspace on which $(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)$ is full-rank.

Then, we perform an eigendecomposition of the symmetric PSD $\mathbf{M}^*$ to obtain $\mathbf{M}^* = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$, from which it follows that $\mathbf{A}^* = \mathbf{\Lambda}^{1/2}\mathbf{U}'$.

It is worth noting that the matrix $\mathbf{A}^*$ obtained as described above will be a square $(p \times p)$ matrix, though our overall objective was to learn an $n \times p$ sensing matrix with $n < p$. There are a few ways that our approach may be adapted to yield "proper" CS matrices. First, we note that the constraint $\mathrm{tr}(\mathbf{M}) = \alpha^2$ has the effect of controlling the rank of the solution $\mathbf{A}^*$.[2] In the event that α is "sufficiently small" the matrix $\mathbf{A}^*$ obtained as above will, in fact, be low rank; In this case, we can form our sensing matrix using as many rows as the nonzero entries along the diagonal of $\mathbf{\Lambda}^{1/2}$.

Alternatively, we can view the square matrix $\mathbf{A}^*$ as a form of "linear preprocessing" for a "standard" iid random CS matrix, say $\mathbf{C}$. In other words, we may form our $n \times p$ measurement matrix $\check{\mathbf{A}}^*$ as $\check{\mathbf{A}}^* = \mathbf{C}\mathbf{A}^*$. When the entries of $\mathbf{C}$ are drawn from certain zero-mean distributions (e.g., Gaussian, symmetric Bernoulli) and have variance $1/n$ per element, it can be easily shown that $\|\check{\mathbf{A}}\|_F^2 \approx \|\mathbf{A}^*\|_F^2$ with high probability provided $n \geq \text{const.} \cdot \log p$ (this follows, for example, from a simple application of the Johnson-Lindenstrauss Lemma to the columns of $\mathbf{A}^*$). Thus, this construction still satisfies our energy constraint (at least with high probability), and allows for direct control over the number of measurements n regardless of the solution $\mathbf{A}^*$. We take this latter approach in the simulations in the following section.

5. Evaluation. We evaluate the performance of our proposed procedure on a "spikes and sinusoids" problem motivated by early works in sparse separation. We consider signals of size $p = 75$ and select as our signal model a subset of $m_x = 50$ unit-norm columns of a $p \times p$ discrete cosine transform (DCT) matrix. From these, we form a total of 50 rank-one models $\mathbf{\Sigma}_{x,i}$, each being simply the outer product of one of the sinusoid vectors with itself. Similarly, we form the clutter model using 50 unit-norm elements of the canonical basis, and form the rank-one models $\mathbf{\Sigma}_{c,i}$. The selected waveforms are assembled into an $p \times (m_x + m_c)$ dictionary $\mathbf{D} = [\mathbf{D}_x \ \mathbf{D}_c]$.

Then, for a subset of possible values of n we perform 1000 trials of the following experiment. First, we select one model randomly from the set $\{\mathbf{\Sigma}_{x,i}\}_{i=1}^{m_x}$ and generate $\mathbf{x}$ as a zero-mean Gaussian random vector having this covariance matrix, and we generate $\mathbf{c}$ similarly using one model selected randomly from $\{\mathbf{\Sigma}_{c,i}\}_{i=1}^{m_c}$. We then generate two sets of observations – the first set $\mathbf{y}_1$ is obtained using an $n \times p$ measurement matrix $\mathbf{A}_1$ having iid $\mathcal{N}(0, \frac{\alpha^2}{np})$ entries, while the second set $\mathbf{y}_2$ uses a measurement matrix of the form $\mathbf{A}_2 = \mathbf{C}\mathbf{A}^*$ where $\mathbf{C}$ has iid $\mathcal{N}(0, \frac{1}{n})$ elements. The additive noise in each case is $\mathbf{w} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_{n\times n})$, where $\sigma^2 = 0.001$. We consider two SNR settings, corresponding to $\alpha = 0.75$ and $\alpha = 1$. We obtain the estimates in each case as $\widehat{\mathbf{x}}_i = [\mathbf{D}_x \ \mathbf{0}] \ \left[\arg\min_\beta \|\beta\|_1 \text{ s.t. } \|\mathbf{y}_i - \mathbf{A}_i\mathbf{D}\beta\|_2^2 \leq n\sigma^2\right]$, for $i = 1, 2$, and for each value of n compute the empirical average squared error as an approximation to $\mathbb{E}_{\mathbf{x},\mathbf{c},\mathbf{w}}\left[\|\mathbf{x} - \widehat{\mathbf{x}}_i\|_2^2\right]$ for $i = 1, 2$.[3] The empirical results, depicted in Figure 5.1, show that our approach results in reduced estimation error compared to traditional compressive sensing.

6. Conclusions. We have demonstrated that knowledge-enhanced compressive sensing designs can outperform traditional iid random CS sensing matrices in estimation problems where the signal to be estimated is corrupted by clutter prior to the

[2] Techniques that minimize the rank of symmetric PSD matrices using the trace – or more generally, minimizing the rank of matrices using the nuclear norm (sum of singular values) – have become quite common in modern sparse estimation works; see, for example, [10, 11, 12, 13].

[3] All optimizations were performed using the `cvx` MATLAB package (`http://cvxr.com/cvx/`)

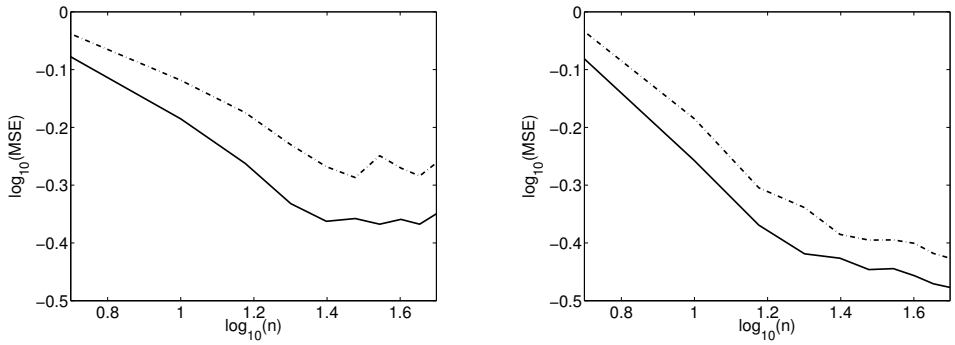

Fig. 5.1: *Estimation MSE comparison between traditional CS (dashed) and the proposed approach (solid). Here,* $\alpha = 0.75$ *in panel (a), and* $\alpha = 1$ *in panel (b).*

compressive measurement operation. Our proposed design approach follows from a few simplifying assumptions, perhaps most notably (3.3), which qualitatively amounts to a "low SNR" assumption.[4] Future directions for this work include investigations of the design problem we consider here in high SNR scenarios, as well as the development of efficient algorithms for solving the optimization (4.1) for larger problem sizes.

REFERENCES

[1] E. J. Candès and T. Tao, *The Dantzig selector: Statistical estimation when p is much larger than n*, Ann. Statist., 35 (2007), pp. 2313-2351.

[2] G. Reeves and M. Gastpar, *Differences between observation and sampling error in sparse signal reconstruction*, Proc. IEEE SSP Conf., (2007), pp. 690-694.

[3] S. Aeron, V. Saligrama, and M. Zhao, *Information theoretic bounds for compressed sensing*, IEEE Trans. Inform. Theory, 56 (2010), pp. 5111-5130.

[4] E. Arias-Castro and Y. Eldar, *Noise folding in compressed sensing*, IEEE Signal Proc. Letters, 18 (2011), pp. 478-481.

[5] M. Elad, *Optimized projections for compressed sensing*, IEEE Trans. Signal Proc., 55 (2007), pp. 5695-5702.

[6] J. M. Duarte-Carvajalino and G. Sapiro, *Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization*, IEEE Trans. Image Proc., 18 (2009), pp. 1395-1408.

[7] J. M. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, *Task-driven adaptive statistical compressive sensing of Gaussian mixture models*, preprint (2007).

[8] N. Rao and R. Nowak, *Correlated Gaussian designs for compressive imaging*, Proc. IEEE Intl. Conf. on Image Proc., (2012).

[9] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, *Optimal designs for space-time linear precoders and decoders*, IEEE Trans. Signal Proc., 50 (2002), pp. 1051-1064.

[10] M. Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Electrical Engineering Department, Stanford University (2002).

[11] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, 9 (2009), pp. 717-772.

[12] A. Chai, M. Moscoso, and G. Papanicolaou, *Array imaging using intensity-only measurements*, Technical report, Stanford University (2010).

[13] E. J. Candès, T. Strohmer, and V. Voroninski, *PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming*, preprint (2011).

[4] Note that $\|\mathbf{A}(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)\mathbf{A}'\| \leq \mathrm{tr}\{\mathbf{A}(\mathbf{\Sigma}_x + \mathbf{\Sigma}_c)\mathbf{A}'\} \leq \|\mathbf{A}'\mathbf{A}\|_F \cdot \|\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\|_F$. Since $\mathrm{tr}\{\mathbf{A}'\mathbf{A}\mathbf{A}'\mathbf{A}\} \leq (\mathrm{tr}\{\mathbf{A}'\mathbf{A}\})^2$, we have that $\|\mathbf{A}'\mathbf{A}\|_F \leq \mathrm{tr}\{\mathbf{A}'\mathbf{A}\} = \alpha$, implying that a sufficient (albeit possibly very loose) condition which ensures that (3.3) holds is $\alpha < \sigma^2/\|\mathbf{\Sigma}_x + \mathbf{\Sigma}_c\|_F$.