Projection Design For Statistical Compressive Sensing: A Tight Frame Based Approach

Wei Chen, Miguel R. D. Rodrigues, *Member, IEEE,* and Ian J. Wassell

Abstract

In this paper, we develop a framework to design sensing matrices for compressive sensing applications that lead to good mean squared error (MSE) performance subject to sensing cost constraints. By capitalizing on the MSE of the oracle estimator, whose performance has been shown to act as a benchmark to the performance of standard sparse recovery algorithms, we use the fact that a Parseval tight frame is the closest design - in the Frobenius norm sense - to the solution of a convex relaxation of the optimization problem that relates to the minimization of the MSE of the oracle estimator with respect to the equivalent sensing matrix, subject to sensing energy constraints. Based on this result, we then propose two sensing matrix designs that exhibit two key properties: i) the designs are closed form rather than iterative; ii) the designs exhibit superior performance in relation to other designs in the literature, which is revealed by our numerical investigation in various scenarios with different sparse recovery algorithms including basis pursuit de-noise (BPDN), the Dantzig selector and orthogonal matching pursuit (OMP).

I. INTRODUCTION

The presence of redundancy in most signals in nature offers the means to transform the original signals into a compressed version convenient for storage and transportation. Compressive sensing (CS) is a new sampling paradigm that, instead of conforming to the traditional two-stage process involving signal

sampling followed by signal compression, directly acquires a compressed version of the original signal instead, by leveraging signal sparsity (a form of redundancy) as well as random sensing or measurement. In fact, it has been shown that if an n-dimensional signal admits an s-sparse representation then one can reconstruct exactly the original signal with $m = \mathcal{O}(s \log(n/s))$ measurements [1], [2]; Also, if the original signal admits only a nearly sparse representation (and/or the measurements are corrupted by some noise) then one can still reconstruct the original signal subject to a tolerable distortion [1], [2]. Therefore, CS offers the prospect of a more efficient signal acquisition in relation to traditional Shannon-Nyquist sampling, especially in applications where the sampling process is expensive such as magnetic resonance imaging [3] and data acquisition in wireless sensor networks [4].

A recent growing trend relates to the use of more complex signal models that go beyond the simple sparsity model to further enhance the performance of CS. For example, Baraniuk et al. [5] have introduced model-based compressive sensing, where more realistic signal models such as wavelet trees or block sparsity are leveraged in order to reduce the number of measurements required for reconstruction. In particular, it has been shown that robust signal recovery is possible with $m = \mathcal{O}(s)$ measurements in model-based compressive sensing [5]. Ji et al. [6] introduced Bayesian compressive sensing, where a signal specific statistical model is exploited to reduce the number of measurements needed for reconstruction. In [7], [8], reconstruction methods have been proposed for manifold-based CS, where the signal is assumed to belong to a manifold. Other works that consider various sparsity models that go beyond simple sparsity in order to improve the performance of traditional CS include [9]–[15].

The use of additional signal knowledge also enables one to replace the conventional random sensing matrices by optimized ones in order to further enhance CS performance (e.g., see [16]–[20]). A number of conditions have been put forth to study the impact of the sensing matrices in various recovery algorithms. The null space property represents a necessary and sufficient condition for sparse recovery [2]. However, it is difficult to verify whether or not a certain sensing matrix fulfills this condition. Other more widely used conditions include the restricted isometry property (RIP) [1], which is also difficult to evaluate, and the mutual coherence [21], which is easier to evaluate. However, the fact that these conditions are mainly used to address the worst-case rather than the expected-case performance, renders their use as the basis of sensing matrix designs as too conservative. As such, Calderbank et al. [22] have put forth a weaker version of the RIP, the statistical restricted isometry property (StRIP), where a probability criterion replaces the hard requirement demanded by RIP. StRIP has been also used as the basis of various sensing matrix designs presented in [22].

In this paper, we develop a general framework to design sensing matrices for compressive sensing

applications that lead to good (expected-case) mean squared error (MSE) performance subject to sensing energy constraints, where the expectation is with respect to both the statistical distribution of the signal and the noise. We also leverage additional signal knowledge, by considering a general random signal model where the distinct support patterns of the same sparsity level occur with equal probability in the sparse representation of the original signal, and the autocorrelation matrix of the sparse representation is equal to an identity matrix. Our approach is based on the analysis of the oracle estimator MSE [23], whose performance has been shown to act as a benchmark to the performance of various common sparse recovery algorithms. By showing that good equivalent sensing matrices (that correspond to the product of the sensing matrix and the sparsifying dictionary) ought to be close to a Parseval tight frame, we are then able to put forth two new sensing matrix designs that conform to specific sensing energy constraints. Our experiments reveal that the proposed designs improve signal expected-case reconstruction performance in relation to random designs or other optimized designs [16]–[18]. Another notable advantage of our proposed designs is that they are closed-form whereas the designs in [16]–[18] are iterative.

Our design approach, which is applicable to signals that are sparse in any dictionary, shares some of the elements of the design approach in [24], which is only applicable to signals that are sparse in an orthonormal basis. In particular, this contribution - as does [24] - also capitalizes on the oracle estimator MSE to put forth adequate sensing matrix designs. However, this design approach also departs significantly from that in [24], in view of the fact that it is not clear how to generalize the methodology in [24] from orthonormal to overcomplete dictionaries (namely, Propositions 1 and 2 in [24]).

Therefore, the current generalization is based on two questions that are answered in the article. We first ask:

1) What is the equivalent sensing matrix that leads to the lowest oracle estimator MSE for a certain target signal to noise ratio (SNR) at the input of the oracle estimator?

Further, in view of the fact that a Parseval tight frame is likely to provide a low oracle MSE subject to a target SNR at the input of the oracle, we then ask:

2) What is the sensing matrix that offers the best compromise between "sensing cost" and "closeness" of the equivalent sensing matrix to a Parseval tight frame?

It is this angle-of-attack - which departs from that in [24] - that enables us to generalize the sensing matrix designs for signals that are sparse in arbitrary overcomplete dictionaries. Interestingly, the ensuing designs are shown to reduce to the designs in [24] when the dictionary is orthonormal rather than overcomplete.

The generalization of the work from the orthonormal to overcomplete dictionary case is relevant

not only theoretically but also practically. For example, allowing signals to be sparse in overcomplete dictionaries adds a lot of flexibility and extends the range of applicability for CS [25]–[27]. Of particular relevance, this generalization also leads to further insight about the behavior of random vs. optimized projections: this is also crisply exposed in this contribution.

The rest of this paper is organized as follows. We begin by describing the CS model and assumptions in Section II. Section III provides the rationale for the sensing matrix designs, by highlighting the role of Parseval tight frames in compressive sensing applications. Section IV puts forth our proposed sensing matrix designs, which capitalize on the intuition unveiled in Section III. Section V presents a range of numerical results that highlight the merits of our proposed designs in relation to other designs in the literature. Section VI discusses the MSE performance yielded both by random and optimized projections designs. The main contributions of the article are finally summarized in Section VII.

Throughout this paper, signals are treated as real-valued vectors. Lower-case letters denote scalars, boldface upper-case letters denote matrices, bold face lower-case letters denote column vectors, and calligraphic upper-case letters denote support sets. $\mathbf{0}$ and $\mathbf{1}$ denote a vector with all zeros and all ones, respectively, and $\mathbf{O}_{m \times n}$ denotes an $m \times n$ matrix with all zeros. The superscripts $(\cdot)^T$ and $(\cdot)^{-1}$ denote matrix transpose and matrix inverse, respectively. The ℓ_0 norm, the ℓ_1 norm, and the ℓ_2 norm of vectors, are denoted by $\|\cdot\|_0$, $\|\cdot\|_1$, and $\|\cdot\|_2$, respectively. The Frobenius norm and spectral norm of a matrix $\mathbf{A}$ are denoted by $\|\mathbf{A}\|_F$ and $\|\mathbf{A}\|$, respectively. The rank and trace of a matrix are denoted by $\mathrm{rank}(\cdot)$ and $\mathrm{Tr}(\cdot)$, respectively. The diagonal matrix with diagonal elements given by either vector $\mathbf{a}$ or the diagonal elements of matrix $\mathbf{A}$ is denoted by $\mathrm{Diag}(\mathbf{a})$ or $\mathrm{Diag}(\mathbf{A})$, respectively. The element corresponding to the ith row and jth column of the matrix $\mathbf{A}$ is denoted by $a_{i,j}$, and $\mathbf{a}_i$ denotes the ith column of the matrix $\mathbf{A}$. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, and $\mathbf{J}_n$ denotes the $n \times n$ anti-diagonal matrix (an identity matrix with a reversed order of the columns (or rows)). $\mathbf{E}_{\mathcal{J}}$ denotes the matrix that results from the identity matrix by deleting the set of columns out of the support $\mathcal{J}$. $\mathbb{E}(\cdot)$ denotes the expectation, $\mathbb{E}_{\mathbf{x}}(\cdot)$ and $\mathbb{E}_{\mathcal{J}}(\cdot)$ denote expectation with respect to the distribution of the random vector $\mathbf{x}$, and the random support $\mathcal{J}$, respectively. $\binom{n}{m}$ denotes the number of m combinations from a given set of n elements. $\mathrm{Pr}(\cdot)$ denotes the probability. Finally, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. COMPRESSIVE SENSING MODEL

We consider the standard measurement model given by:

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{f} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the measurement signal vector, $\mathbf{f} \in \mathbb{R}^n$ is the original signal vector, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m) \in \mathbb{R}^m$ is a zero-mean white Gaussian noise vector, and $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ (with $m \leq n$) is the sensing matrix. We assume that the original signal is sparse in some basis, i.e.,

$$\mathbf{f} = \mathbf{\Psi}\mathbf{x}, \tag{2}$$

where $\mathbf{\Psi} \in \mathbb{R}^{n \times \hat{n}}$ ($\hat{n} \geq n$) is a matrix that represents the sparsifying basis, e.g., an orthonormal or overcomplete dictionary, and $\mathbf{x} \in \mathbb{R}^{\hat{n}}$ is a sparse representation of $\mathbf{f} \in \mathbb{R}^n$, i.e., $\|\mathbf{x}\|_0 \leq s \ll \hat{n}$. Then we can rewrite the measurement model as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{x} + \mathbf{n} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{3}$$

where $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi} \in \mathbb{R}^{m \times \hat{n}}$ represents the equivalent sensing matrix. For modeling the sparse sources, we assume i) the distinct support patterns of the same sparsity level occur with equal probability in the sparse representation of the original signal, i.e., $\Pr\left(\mathcal{J}_c^t\right) = P_c$, where $\mathcal{J}_c^t \subset \{1, \ldots, \hat{n}\}$ ($t = 1, \ldots, \binom{\hat{n}}{c}$, $c = 1, \ldots, s$) denotes a signal support with cardinality c and $\sum_{c=1}^{s} \binom{\hat{n}}{c} P_c = 1$; ii) $\mathbb{E}_{\mathbf{x}}(\mathbf{x}\mathbf{x}^T) = \mathbf{I}_{\hat{n}}$. Note that these assumptions can be satisfied by a signal model akin to the widely used Bernoulli-Gaussian model [28]–[34]. In particular, one constrains the cardinality of the support patterns to be less than s, rather than $\hat{n}$; one also constrains the probability of the support patterns to obey $\sum_{c=1}^{s} \binom{\hat{n}}{c} P_c = 1$ rather than a binomial distribution as in the Bernoulli-Gaussian model.

To recover the sparse signal representation $\mathbf{x}$ from the measurement vector $\mathbf{y}$, one can resort to the optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_1 \\ \text{s.t.} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon, \end{aligned} \tag{4}$$

where ϵ is an estimate of the noise level. This program is also known as the basis pursuit de-noise (BPDN) [35].

It has been established in [36] that the now well-known RIP, which has been introduced by Candès and Tao [37], provides a sufficient condition for exact or near exact recovery of a sparse signal representation $\mathbf{x}$ from the measurement vector $\mathbf{y}$ via the ℓ_1 minimization in (4).

Definition 1: A matrix $\mathbf{A} \in \mathbb{R}^{m \times \hat{n}}$ satisfies the RIP of order s with a restricted isometry constant (RIC) $\delta_s \in (0, 1)$ being the smallest number such that

$$(1 - \delta_s)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_s)\|\mathbf{x}\|_2^2 \tag{5}$$

holds for all $\mathbf{x}$ with $\|\mathbf{x}\|_0 \leq s$.

Theorem 1: The solution $\mathbf{x}^*$ of (4) obeys

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq c_1 s^{-1/2}\|\mathbf{x} - \mathbf{x}_s\|_1 + c_2\epsilon, \tag{6}$$

where $c_1 = \frac{2+(2\sqrt{2}-2)\delta_{2s}}{1-(\sqrt{2}+1)\delta_{2s}}$, $c_2 = \frac{4\sqrt{1+\delta_{2s}}}{1-(\sqrt{2}+1)\delta_{2s}}$, $\mathbf{x}_s$ is an approximation of $\mathbf{x}$ with all but the s-largest entries set to zero, and δ_{2s} is the RIC of order $2s$ of matrix $\mathbf{A}$.

This theorem claims that the reconstructed signal representation $\mathbf{x}^*$ is a good approximation to the original signal representation $\mathbf{x}$. In addition, for the noiseless case, any sparse representation $\mathbf{x}$ with support size no larger than s, can be exactly recovered by ℓ_1 minimization if the RIC satisfies $\delta_{2s} < \sqrt{2}-1$. Therefore, it follows that the RIP acts as a proxy to the quality of a sensing matrix. Note that the RIP is a sufficient condition for successful reconstruction but it may be too strict. It has been observed that signals with sparse representations can be reconstructed very well even though the sensing matrices have not been proven to satisfy the RIP [22].

Another way to evaluate a sensing matrix, which is not as computationally intractable as the RIP, is via the mutual coherence of the matrix $\mathbf{A}$, given by [21]:

$$\mu = \max_{1 \leq i,j \leq \hat{n}, i \neq j} |\mathbf{a}_i^T \mathbf{a}_j|. \tag{7}$$

Donoho, Elad and Temlyakov [21] demonstrated that the error of the solution to (4) is bounded if $\mu < \frac{1}{4s-1}$. Therefore, mutual coherence can also be used to measure the quality of a sensing matrix. For example, various sensing matrix design approaches in the literature, such as Elad's method [16], Duarte-Carvajalino and Sapiro's method [17], and Xu et al.'s method [18] are inherently mutual coherence based approaches.

III. DESIGN RATIONALE

We now provide a rationale for the proposed novel sensing matrix designs. The ultimate goal of the sensing matrix designs relates to the minimization of the MSE in estimating $\mathbf{x}$ from $\mathbf{y}$, given by

$$\text{MSE}(\mathbf{\Phi}) = \mathbb{E}_{\mathbf{x},\mathbf{n}}\left(\|\mathcal{F}(\mathbf{\Phi}\mathbf{\Psi}\mathbf{x} + \mathbf{n}) - \mathbf{x}\|_2^2\right), \tag{8}$$

where $\mathcal{F}(\cdot)$ denotes an estimator, subject to appropriate constraints (e.g., sensing energy cost)[1].

The derivation of such a sensing matrix design is very difficult though, because the average MSE in (8) depends upon the actual estimator. Consequently, to avoid the analysis of a single or several practical sparse recovery algorithms such as the BPDN, the Dantzig selector, or the OMP, we capitalize - as in [24] - on the well-known oracle estimator that performs ideal least squares (LS) estimation based on prior knowledge of the sparse vector support $\mathcal{J} \subset \{1, \ldots, \hat{n}\}$ [23]. The rationale of this approach is supported by the fact that the MSE of this oracle LS estimator coincides with the unbiased Cramér-Rao bound (CBD) for exactly s-sparse deterministic vectors [41], so that it represents the best achievable performance for any unbiased estimator. Equally important, this approach is also supported by the fact that the oracle estimator MSE performance acts as a performance benchmark for the key sparse recovery algorithms. For example, Ben-Haim, Eldar and Elad [42] demonstrate both theoretically and numerically that the BPDN, the Dantzig selector, the OMP and thresholding algorithms all achieve performances that are proportional to the oracle estimator MSE.

The oracle estimator MSE incurred in the estimation of a sparse deterministic vector $\mathbf{x}$ in the presence of a standard Gaussian noise vector $\mathbf{n}$, according to the model in (1), is given by [23][2]:

$$
\begin{aligned}
\text{MSE}_{\mathbf{n}}^{\text{oracle}}(\mathbf{A}, \mathbf{x}) =& \mathbb{E}_{\mathbf{n}} \left(\|\mathcal{F}^{\text{oracle}}(\mathbf{A}\mathbf{x} + \mathbf{n}) - \mathbf{x}\|_2^2 \right) \\
=& \sigma^2 \text{Tr} \left(\left(\mathbf{E}_{\mathcal{J}}^T \mathbf{A}^T \mathbf{A} \mathbf{E}_{\mathcal{J}} \right)^{-1} \right).
\end{aligned}
\tag{9}
$$

[1]We would also like to add that one could argue that it is preferable to consider the MSE associated with the estimation of $\mathbf{f}$ (the actual signal) from $\mathbf{y}$ rather than the MSE associated with the estimation of $\mathbf{x}$ (the signal sparse representation) from $\mathbf{y}$. We use the more tractable MSE associated with the estimation of $\mathbf{x}$ from $\mathbf{y}$ because: 1) it can be shown that the MSE performance associated with the (oracle) estimation of $\mathbf{x}$ from $\mathbf{y}$ upper bounds in general the MSE performance associated with the (oracle) estimation of $\mathbf{f}$ from $\mathbf{y}$. In particular, for an orthogonal dictionary, where $\mathbf{\Psi}$ is an orthogonal matrix, $\|\mathbf{f} - \mathbf{f}^*\|_2^2 = \|\mathbf{\Psi}\mathbf{x} - \mathbf{\Psi}\mathbf{x}^*\|_2^2 = \|\mathbf{x} - \mathbf{x}^*\|_2^2$, where $\mathbf{x}^*$ denotes the (oracle) estimate of $\mathbf{x}$ and $\mathbf{f}^* = \mathbf{\Psi}\mathbf{x}^*$ denotes the (oracle) estimate of $\mathbf{f}$; for an overcomplete dictionary, where $\mathbf{\Psi}$ is not an orthogonal matrix, $\|\mathbf{f} - \mathbf{f}^*\|_2^2 = \|\mathbf{\Psi}\mathbf{x} - \mathbf{\Psi}\mathbf{x}^*\|_2^2 \leq \lambda_{\max}^2(\mathbf{\Psi}) \|\mathbf{x} - \mathbf{x}^*\|_2^2$, where $\lambda_{\max}(\mathbf{\Psi})$ is the largest singular value of $\mathbf{\Psi}$; 2) it is also often desirable to manipulate or process the information content of signals in the sparse representation domain rather than the original observation domain, such as in feature extraction, pattern classification and blind source separation [38]–[40]. Therefore, the MSE performance associated with the estimation of $\mathbf{x}$ would be more appropriate than the MSE performance associated with the estimation of $\mathbf{f}$ for such applications.

[2]Note that various works have adopted the oracle minimum MSE (MMSE) estimator in lieu of the oracle LS one in order to obtain a superior MMSE estimate [43]–[45]. The fact that we assume a signal model that does not specify the exact distribution of the sparse signal conditioned on the support - in contrast to [43]–[45] that take the distribution of the sparse signal conditioned on the support to be multi-variate Gaussian - prevents us from exploiting this more powerful estimator. This approach however instils our projections design framework with more generality.

Consequently, the average value of the oracle estimator MSE incurred in the estimation of a sparse random vector $\mathbf{x}$ in the presence of the Gaussian noise vector $\mathbf{n}$ is given by:

$$\text{MSE}^{\text{oracle}}(\mathbf{A}) = \sigma^2 \mathbb{E}_{\mathcal{J}} \left(\text{Tr} \left(\left(\mathbf{E}_{\mathcal{J}}^T \mathbf{A}^T \mathbf{A} \mathbf{E}_{\mathcal{J}} \right)^{-1} \right) \right). \tag{10}$$

We define the coherence matrix of the equivalent sensing matrix as $\mathbf{Q} = \mathbf{A}^T \mathbf{A} = \mathbf{\Psi}^T \mathbf{\Phi}^T \mathbf{\Phi} \mathbf{\Psi}$. We now pose the optimization problem:

$$
\begin{aligned}
\min_{\mathbf{Q}} \quad & \mathbb{E}_{\mathcal{J}} \left(\text{Tr} \left(\left(\mathbf{E}_{\mathcal{J}}^T \mathbf{Q} \mathbf{E}_{\mathcal{J}} \right)^{-1} \right) \right) \\
\text{s.t.} \quad & \mathbf{Q} \succeq 0, \\
& \text{Tr} \left(\mathbf{Q} \right) = m, \\
& rank(\mathbf{Q}) \leq m.
\end{aligned}
\tag{11}
$$

It is relevant to reflect further on the rationale of this optimization problem. This optimization problem defines the coherence matrix of the equivalent sensing matrix - up to a rotation - that minimizes the average value of the oracle MSE subject to appropriate constraints: these include the obvious positive semi-definite and rank constraints on the coherence matrix and - at the heart of the novelty of the approach - a trace constraint on the coherence matrix that acts as a proxy to the sensed energy.

In the noiseless case [16]–[18], it is not common to place a constraint on the sensed energy because recovery is immune to the scaling of the sensing matrix; instead, it is only common to seek sensing matrices that exhibit adequate structure (e.g., [16] uses t-averaged mutual coherence, [17] uses an equivalent sensing matrix whose Gram matrix is similar to an identity matrix, and [18] uses an equivalent sensing matrix which is close to an equiangular tight frame, to seek for sensing matrices with adequate structure).

In contrast, in the noisy case it is important to place a constraint on the sensed energy because recovery is affected both by the sensing matrix structure and immunity to noise. Therefore, the main features of our formulation include:

1) The optimization problem defines equivalent sensing matrices with good structure and immunity to noise.

2) The formulation is such that the sensed energy is directly proportional to the number of measurements. In fact, the sensed energy is given by:

$$
\begin{aligned}
\mathbb{E}_{\mathbf{x}} \left(\text{Tr} \left(\mathbf{\Phi} \mathbf{\Psi} \mathbf{x} \mathbf{x}^T \mathbf{\Psi}^T \mathbf{\Phi}^T \right) \right) &= \text{Tr} \left(\mathbf{\Phi} \mathbf{\Psi} \mathbb{E}_{\mathbf{x}} \left(\mathbf{x} \mathbf{x}^T \right) \mathbf{\Psi}^T \mathbf{\Phi}^T \right) \\
&= \text{Tr} \left(\mathbf{\Phi} \mathbf{\Psi} \mathbf{\Psi}^T \mathbf{\Phi}^T \right) = m,
\end{aligned}
\tag{12}
$$

where we have used the fact that $\mathbb{E}_{\mathbf{x}}\left(\mathbf{x}\mathbf{x}^T\right) = \mathbf{I}_{\hat{n}}$. Note that a modification of the constant of proportionality, which is equal to 1 here, scales only the solution to the optimization problem (11).

3) The formulation is also such that the sensed SNR

$$\frac{\mathbb{E}_{\mathbf{x}}\left(\text{Tr}\left(\mathbf{\Phi}\mathbf{\Psi}\mathbf{x}\mathbf{x}^T\mathbf{\Psi}^T\mathbf{\Phi}^T\right)\right)}{\mathbb{E}_{\mathbf{n}}\left(\text{Tr}\left(\mathbf{n}\mathbf{n}^T\right)\right)} = \frac{1}{\sigma^2}, \tag{13}$$

does not depend on m, n or $\hat{n}$.

We will see that in the presence of noise some of the "noiseless" sensing matrix designs in the literature can yield very poor recovery performance (see Section V). This is due to the fact that upon the normalization of the sensing matrix so that it conforms to a specific sensing cost constraint, the structural properties of the designs are offset by the poor noise immunity of the designs. The optimization problem formulation in (11) aims thus to attain a compromise between the structural and the noise immunity properties of the sensing matrix[3].

The optimization problem (11) is non-convex owing to the rank constraint, and so is very difficult to solve. Therefore, we adopt an approach akin to that in [24]: i) we first consider a convex relaxation of (11) by ignoring the rank constraint; and ii) we then consider the feasible solution that is closest to the solution to the relaxed problem. This procedure produces a sub-optimal equivalent sensing matrix, but extensive simulation results demonstrate that this design outperforms various other designs.

Proposition 1: The solution of the optimization problem:

$$\min_{\mathbf{Q}} \quad \mathbb{E}_{\mathcal{J}}\left(\text{Tr}\left(\left(\mathbf{E}_{\mathcal{J}}^T\mathbf{Q}\mathbf{E}_{\mathcal{J}}\right)^{-1}\right)\right)$$

$$\text{s.t.} \quad \mathbf{Q} \succeq 0, \tag{14}$$

$$\text{Tr}\left(\mathbf{Q}\right) = m,$$

which represents a convex relaxation of the original optimization problem in (11), is the $\hat{n} \times \hat{n}$ matrix $\frac{m}{\hat{n}}\mathbf{I}_{\hat{n}}$.

Proof: See Appendix A. ∎

---

[3]Note that this optimization problem places a cost on the equivalent sensing matrix $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}$, which translates into a constraint on the energy given to the estimator rather than a cost on the sensing matrix $\mathbf{\Phi}$, which translates into a constraint on the sensing energy. We recognize that a sensing energy cost is often more appropriate, but this is difficult to analyze in general. Therefore, our approach when the signal is sparse in a general overcomplete dictionary departs from that when the signal is sparse in an orthonormal dictionary [24]. In particular, we only incorporate the effect of sensing energy constraints into the design framework in Section IV.

It is evident that the solution to the convex relaxation of the original optimization problem is not feasible, because $\texttt{rank}(\frac{m}{\hat{n}}\mathbf{I}_{\hat{n}}) = \hat{n} \geq m$. Therefore, we now propose to determine the $m \times \hat{n}$ matrix $\mathbf{A}$ whose $\hat{n} \times \hat{n}$ coherence matrix $\mathbf{Q} = \mathbf{A}^T\mathbf{A}$ is closest to the $\hat{n} \times \hat{n}$ matrix $\frac{m}{\hat{n}}\mathbf{I}_{\hat{n}}$.

Proposition 2: The solution of the optimization problem:

$$\min_{\mathbf{A}} \quad \left\| \mathbf{A}^T\mathbf{A} - \frac{m}{\hat{n}}\mathbf{I}_{\hat{n}} \right\|_F^2$$
$$\text{s.t.} \quad \text{Tr}\left(\mathbf{A}^T\mathbf{A}\right) = m, \tag{15}$$

is the $m \times \hat{n}$ Parseval tight frame.

Proof: See Appendix B. ∎

A frame in a finite-dimensional real space can be seen as a matrix $\mathbf{A} \in \mathbb{R}^{m \times \hat{n}}$ such that for any vector $\mathbf{z} \in \mathbb{R}^m$,

$$a\|\mathbf{z}\|_2^2 \leq \|\mathbf{A}^T\mathbf{z}\|_2^2 \leq b\|\mathbf{z}\|_2^2, \tag{16}$$

where $a > 0$ and $b > 0$ are known as the frame bounds. Tight frames are a class of frames with equal frame bounds, i.e., $a = b$. A tight frame whose columns have unit ℓ_2 norm is called a unit norm tight frame. A tight frame whose frame bound is equal to 1, is called a Parseval tight frame. Note that any tight frame can be scaled by multiplying by $\frac{1}{\sqrt{a}}$, so that the frame bound becomes equal to 1.

Therefore, the value of the constraint of (15) leads to a frame with a frame bound being equal to 1, and thus results in a Parseval tight frame. By scaling the value of $\text{Tr}\left(\mathbf{A}^T\mathbf{A}\right)$ in the constraint, which in fact alters the target sensing SNR in (13), it is clear that the solution of the optimization problem (15) is still a tight frame. Therefore, we can deduce that the tight frame represents a good equivalent sensing matrix design, in the sense that, among all equivalent sensing matrices that conform to the target sensing SNR, a tight frame is likely to produce a good MSE performance. Appendix C explores another facet of tight frames, including the relationship of a unit-norm tight frame to StRIP.

Note that an alternative way to prove Proposition 2, which has been motivated by the optimization problem put forth by Duarte-Carvajalino and Sapiro [17], is also provided in [19]. The current problem differs from the problems in [17], [19] since our optimization approach is based on a metric with operational significance, the MSE, whereas the optimization approach in [17], [19] is based on mutual coherence.

IV. Novel Sensing Matrix Design Approaches

We now build upon the previous analysis, which suggests that $\mathbf{A} = \mathbf{\Phi\Psi}$ ought to be close to a Parseval tight frame, to propose two sensing matrix designs for the compressive sensing model in (3). In

particular, in view of the fact that it is usual to place a constraint on the sensing energy cost $\|\mathbf{\Phi}\|_F^2 = n$, the design approaches strike a balance between two objectives: i) guaranteeing that the equivalent sensing matrix $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}$ is as close as possible to a Parseval tight frame; and ii) guaranteeing that the sensing cost $\|\mathbf{\Phi}\|_F^2$ is as small as possible. For example, for two different sensing matrices $\mathbf{\Phi}'$ and $\mathbf{\Phi}''$ such that $\mathbf{\Phi}'\mathbf{\Psi} = \mathbf{\Phi}''\mathbf{\Psi}$ is equal or close (e.g., in Frobenius norm sense) to some Parseval tight frame and $\|\mathbf{\Phi}'\|_F^2 < \|\mathbf{\Phi}''\|_F^2$, it may be preferable to use $\mathbf{\Phi}'$ instead of $\mathbf{\Phi}''$ in the compressive sensing model in (3). In fact, the normalization

$$\tilde{\mathbf{\Phi}}' = \frac{\sqrt{n}\mathbf{\Phi}'}{\|\mathbf{\Phi}'\|_F} \tag{17}$$

and

$$\tilde{\mathbf{\Phi}}'' = \frac{\sqrt{n}\mathbf{\Phi}''}{\|\mathbf{\Phi}''\|_F} \tag{18}$$

then ensures that

$$\|\tilde{\mathbf{\Phi}}'\mathbf{\Psi}\|_F^2 > \|\tilde{\mathbf{\Phi}}''\mathbf{\Psi}\|_F^2 \tag{19}$$

and - via the previous analysis - eventually

$$\text{MSE}^{\text{oracle}}(\tilde{\mathbf{\Phi}}'\mathbf{\Psi}) < \text{MSE}^{\text{oracle}}(\tilde{\mathbf{\Phi}}''\mathbf{\Psi}). \tag{20}$$

We note that this design approach, which is applicable to the noisy setting, is fundamentally different from the approaches in [16]–[18], which in contrast apply to the noiseless case. In particular, our design considers the sensing energy cost whereas the designs in [16]–[18] do not. We will reveal the effect of taking into account the sensing energy constraint when we re-normalize the designs in [16]–[18], by showing the radically different performances in the presence of noise.

A. Design Approach 1

We now consider the first sensing matrix design approach, which explicitly performs a balance between the objective of guaranteeing that the equivalent sensing matrix is as close as possible to a Parseval tight frame against the objective of guaranteeing that the sensing energy cost is as small as possible. In particular, we pose the design problem:

$$\min_{\hat{\mathbf{\Phi}}} \quad \left\|\hat{\mathbf{\Phi}}\mathbf{\Psi} - \mathbf{B}\right\|_F^2 + \alpha \left\|\hat{\mathbf{\Phi}}\right\|_F^2, \tag{21}$$

where $\mathbf{B} \in \mathbb{R}^{m \times \hat{n}}$ is a specific target Parseval tight frame and $\alpha \geq 0$ is a specific scalar. The solution to the design problem is:

$$\hat{\mathbf{\Phi}} = \mathbf{B}\mathbf{\Psi}^T \left(\mathbf{\Psi}\mathbf{\Psi}^T + \alpha\mathbf{I}_n\right)^{-1}. \tag{22}$$

In turn, the sensing matrix design, which is consistent with the sensing cost constraint $\|\mathbf{\Phi}\|_F^2 = n$, is:

$$\mathbf{\Phi} = \frac{\sqrt{n}\hat{\mathbf{\Phi}}}{\|\hat{\mathbf{\Phi}}\|_F} = \frac{\sqrt{n}\mathbf{B}\mathbf{\Psi}^T\left(\mathbf{\Psi}\mathbf{\Psi}^T + \alpha\mathbf{I}_n\right)^{-1}}{\|\mathbf{B}\mathbf{\Psi}^T\left(\mathbf{\Psi}\mathbf{\Psi}^T + \alpha\mathbf{I}_n\right)^{-1}\|_F}. \tag{23}$$

We note that the scalar α controls the weight for the energy penalty of the sensing matrix. If the penalty is not considered, i.e., $\alpha = 0$, we have the sensing matrix design $\mathbf{\Phi} = \frac{\sqrt{n}\mathbf{B}\mathbf{\Psi}^T(\mathbf{\Psi}\mathbf{\Psi}^T)^{-1}}{\|\mathbf{B}\mathbf{\Psi}^T(\mathbf{\Psi}\mathbf{\Psi}^T)^{-1}\|_F}$. In contrast, for a very high penalty, i.e., $\alpha \to +\infty$, we have the design $\mathbf{\Phi} = \frac{\sqrt{n}\mathbf{B}\mathbf{\Psi}^T}{\|\mathbf{B}\mathbf{\Psi}^T\|_F}$. In both cases, i.e., $\alpha = 0$ or $\alpha \to +\infty$, the sensing matrix $\mathbf{\Phi}$ turns out to be a unit norm tight frame if the basis $\mathbf{\Psi}$ is an orthonormal matrix and the design target $\mathbf{B}$ is a tight frame with equal column norm, i.e., a scaled unit norm tight frame. We also note that, as will be shown later, the performance gain is greatly affected by the parameter α. In particular, one needs to use some empirical knowledge in order to set a suitable value for α. We next propose a sensing matrix design approach, that does not contain any adjustable parameters.

B. Design Approach 2

We now consider the second sensing matrix design approach, where the objective is to determine the matrix design with the lowest sensing energy cost that is consistent with the fact that the equivalent sensing matrix ought to be a Parseval tight frame. It will be shown that the ensuing design is instilled with operational significance, akin to the design in [20]. We pose the design problem:

$$\begin{aligned} \min_{\hat{\mathbf{\Phi}}} \quad & \left\|\hat{\mathbf{\Phi}}\right\|_F^2 \\ \text{s.t.} \quad & \hat{\mathbf{\Phi}}\mathbf{\Psi}\mathbf{\Psi}^T\hat{\mathbf{\Phi}}^T = \mathbf{I}_m. \end{aligned} \tag{24}$$

The following Proposition defines the solution to this optimization problem. We use the singular value decomposition (SVD) of the dictionary $\mathbf{\Psi} = \mathbf{U}_\Psi\mathbf{\Lambda}_\Psi\mathbf{V}_\Psi^T$, where $\mathbf{U}_\Psi \in \mathbb{R}^{n \times n}$ and $\mathbf{V}_\Psi \in \mathbb{R}^{\hat{n} \times \hat{n}}$ are orthonormal matrices, and $\mathbf{\Lambda}_\Psi \in \mathbb{R}^{n \times \hat{n}}$ is a matrix whose main diagonal entries ($\lambda_1^\Psi \geq \lambda_2^\Psi \geq \ldots \lambda_n^\Psi \geq 0$) are the singular values of $\mathbf{\Psi}$ and the other entries are zeros. We also use the SVD of the sensing matrix $\hat{\mathbf{\Phi}} = \mathbf{U}_{\hat{\Phi}}\mathbf{\Lambda}_{\hat{\Phi}}\mathbf{V}_{\hat{\Phi}}^T$, where $\mathbf{U}_{\hat{\Phi}} \in \mathbb{R}^{m \times m}$ and $\mathbf{V}_{\hat{\Phi}} \in \mathbb{R}^{n \times n}$ are orthonormal matrices, and $\mathbf{\Lambda}_{\hat{\Phi}} \in \mathbb{R}^{m \times n}$ is a matrix whose main diagonal entries ($\lambda_1^{\hat{\Phi}} \geq \lambda_2^{\hat{\Phi}} \geq \ldots \geq \lambda_m^{\hat{\Phi}} \geq 0$) are the singular values of $\hat{\mathbf{\Phi}}$ and the other entries are zeros.

Proposition 3: A sensing matrix that solves the optimization problem in (24) is given by

$$\hat{\mathbf{\Phi}} = \mathbf{U}_{\hat{\Phi}}\mathbf{\Lambda}_{\hat{\Phi}}\mathbf{J}_n\mathbf{U}_\Psi^T, \tag{25}$$

where $\mathbf{U}_{\hat{\Phi}}$ is an arbitrary orthonormal matrix and $\mathbf{\Lambda}_{\hat{\Phi}} = \left[\text{Diag}\left(\frac{1}{\lambda_m^\Psi}, \ldots, \frac{1}{\lambda_1^\Psi}\right) \quad \mathbf{O}_{m \times (n-m)}\right]$.

Proof: Consider the SVD of the dictionary $\boldsymbol{\Psi} = \mathbf{U}_{\boldsymbol{\Psi}}\boldsymbol{\Lambda}_{\boldsymbol{\Psi}}\mathbf{V}_{\boldsymbol{\Psi}}^T$ and the sensing matrix $\hat{\boldsymbol{\Phi}} = \mathbf{U}_{\hat{\boldsymbol{\Phi}}}\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}\mathbf{V}_{\hat{\boldsymbol{\Phi}}}^T$. Then the equivalent sensing matrix can be expressed as:

$$\hat{\boldsymbol{\Phi}}\boldsymbol{\Psi} = \mathbf{U}_{\hat{\boldsymbol{\Phi}}}\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}\mathbf{V}_{\hat{\boldsymbol{\Phi}}}^T\mathbf{U}_{\boldsymbol{\Psi}}\boldsymbol{\Lambda}_{\boldsymbol{\Psi}}\mathbf{V}_{\boldsymbol{\Psi}}^T, \tag{26}$$

and so the Parseval tight frame constraint in (24) can also be expressed as:

$$\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}\mathbf{V}_{\hat{\boldsymbol{\Phi}}}^T\mathbf{U}_{\boldsymbol{\Psi}}\boldsymbol{\Lambda}_{\boldsymbol{\Psi}}\boldsymbol{\Lambda}_{\boldsymbol{\Psi}}^T\mathbf{U}_{\boldsymbol{\Psi}}^T\mathbf{V}_{\hat{\boldsymbol{\Phi}}}\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}^T = \mathbf{I}_m. \tag{27}$$

To satisfy the Parseval tight frame condition in (27), it is clear that m columns of $\mathbf{V}_{\hat{\boldsymbol{\Phi}}}$ have to correspond to m columns of $\mathbf{U}_{\boldsymbol{\Psi}}$. Since the remaining $n - m$ columns of $\mathbf{V}_{\hat{\boldsymbol{\Phi}}}$ do not affect the Parseval tight frame condition at all, then we take without any loss of generality $\mathbf{V}_{\hat{\boldsymbol{\Phi}}} = \mathbf{U}_{\boldsymbol{\Psi}}\boldsymbol{\Pi}$, where $\boldsymbol{\Pi} \in \mathbb{R}^{n \times n}$ is a permutation matrix. Therefore, we can now rewrite the optimization problem as follows:

$$\min_{\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}, \boldsymbol{\Pi}} \quad \left\|\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}\right\|_F^2$$

$$\text{s.t.} \quad \boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}\boldsymbol{\Pi}^T\boldsymbol{\Lambda}_{\boldsymbol{\Psi}}\boldsymbol{\Lambda}_{\boldsymbol{\Psi}}^T\boldsymbol{\Pi}\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}^T = \mathbf{I}_m, \tag{28}$$

$$\boldsymbol{\Pi} \text{ is a permutation matrix.}$$

The solution to this optimization problem is trivially given by:

$$\boldsymbol{\Pi} = \mathbf{J}_n, \tag{29}$$

and

$$\begin{aligned}
\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}} &= \left[\text{Diag}\left(\lambda_1^{\hat{\boldsymbol{\Phi}}}, \lambda_2^{\hat{\boldsymbol{\Phi}}}, \ldots, \lambda_m^{\hat{\boldsymbol{\Phi}}}\right) \; \mathbf{O}_{m \times (n-m)}\right] \\
&= \left[\text{Diag}\left(\frac{1}{\lambda_m^{\boldsymbol{\Psi}}}, \frac{1}{\lambda_{m-1}^{\boldsymbol{\Psi}}}, \ldots, \frac{1}{\lambda_1^{\boldsymbol{\Psi}}}\right) \; \mathbf{O}_{m \times (n-m)}\right].
\end{aligned} \tag{30}$$

∎

Proposition 3 uncovers the key operations performed by this sensing matrix design. In particular, this sensing matrix design i) exposes the modes (singular values) of the dictionary; ii) passes through the m strongest modes and filters out the $n - m$ weakest modes; and iii) weighs the strongest modes. This is accomplished by taking the matrix of right singular vectors of the sensing matrix to correspond to the matrix of left singular vectors of the dictionary and taking the strongest modes of the dictionary.

Proposition 3 leads immediately to the sensing matrix design, which is consistent with the sensing cost constraint $\|\hat{\boldsymbol{\Phi}}\|_F^2 = n$, as follows:

$$\boldsymbol{\Phi} = \frac{\sqrt{n}\hat{\boldsymbol{\Phi}}}{\|\hat{\boldsymbol{\Phi}}\|_F} = \frac{\sqrt{n}\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}\mathbf{J}_n\mathbf{U}_{\boldsymbol{\Psi}}^T}{\|\boldsymbol{\Lambda}_{\hat{\boldsymbol{\Phi}}}\|_F}. \tag{31}$$

Note that the design approach 1 balances the requirements of guaranteeing that the equivalent sensing matrix is as close as possible to a Parseval tight frame against the requirements of guaranteeing that the cost is as small as possible; in the design approach 2, we force the equivalent sensing matrix to be a Parseval tight frame and minimize the sensing energy. Note also that the proposed designs are closed-form whereas other designs in the literature, such as Elad's method [16], Duarte-Carvajalino and Sapiro's method [17], and Xu et al.'s method [18], are iterative.

Finally, it is also interesting to note that design 1 and design 2 reduce to the design in [24], i.e., to a tight frame, when we take the dictionary to be orthonormal rather than overcomplete.

V. PERFORMANCE RESULTS

We now compare the performance of the proposed sensing matrix designs to other designs in the CS setting.

A. Distribution of the off-diagonal entries of the coherence matrix

We first investigate the histogram of the absolute values of the off-diagonal entries of the coherence matrix $\mathbf{\Psi}^T \mathbf{\Phi}^T \mathbf{\Phi} \mathbf{\Psi}$. In this investigation, we use a random dictionary $\mathbf{\Psi} \in \mathbb{R}^{64 \times 80}$ with entries drawn from i.i.d. zero mean and unit variance Gaussian distributions and then normalized to $\|\mathbf{\Psi}\|_F^2 = 80$. We also generate three sensing matrices $\mathbf{\Phi} \in \mathbb{R}^{40 \times 64}$ using the proposed approach 1 with $\alpha = 1$ and $\alpha = 0.1$, and using the proposed approach 2. We compare the performance of the proposed designs with a random Gaussian matrix design and with three iterative designs, namely, Elad's design [16], Xu's design [18] and Sapiro's design [17].

It has been observed that coherence matrices with small off-diagonal entries result in good reconstruction performance in accordance with the mutual coherence reconstruction condition [16]–[18]. Fig. 1 shows that the distributions of the off-diagonal entries in both designs based on approach 1 are better than that for the Gaussian matrix design. In particular, note that the design with $\alpha = 0.1$ has off-diagonal entries with smaller absolute value than does the design with $\alpha = 1$. However, $\|\mathbf{\Phi} \mathbf{\Psi}\|_F^2 = 41.2639$ for the $\alpha = 0.1$ design is lower than $\|\mathbf{\Phi} \mathbf{\Psi}\|_F^2 = 89.1929$ for the $\alpha = 1$ design - owing to the lower penalty used in the optimization problem in (21) - and also lower than $\|\mathbf{\Phi} \mathbf{\Psi}\|_F^2 = 84.5554$ for the Gaussian design. This observation - via the analysis in Section III - ought to lead to poorer MSE performance of the design with $\alpha = 0.1$ in relation to the design with $\alpha = 1$ and also in relation to the Gaussian design. The distribution of the off-diagonal entries in the design based on approach 2 is also better than the Gaussian matrix. In addition, the sensing energy of the equivalent sensing matrix $\|\mathbf{\Phi} \mathbf{\Psi}\|_F^2 = 89.1929$ is not reduced
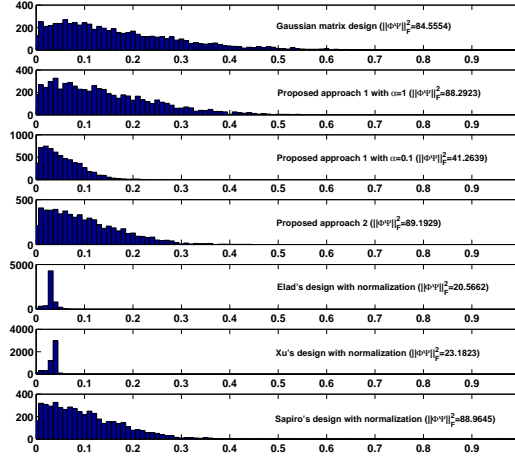
Fig. 1. Histogram of the absolute value of the off-diagonal entries of the coherence matrix.

compared to the Gaussian matrix design. Elad's and Xu's designs, exhibit good mutual coherence but poor sensing energy. The attributes of Sapiro's design are equivalent to those of the design based on approach 2. Yet, our design is non-iterative whereas Sapiro's design follows an iterative procedure.

The reconstruction performance of the proposed designs is further investigated in the following subsections, both in terms of the MSE of the ideal oracle estimator as well as the MSE of practical estimators.

B. The MSE performance using the oracle estimator

In this investigation, we evaluate the MSE performance of various designs using the ideal oracle estimator, which has played a key role in the definition of our designs. The MSE is evaluated by averaging over 1000 trials, where in each trial we generate randomly a sparse vector with s randomly placed ± 1 spikes[4]. The random dictionary $\boldsymbol{\Psi} \in \mathbb{R}^{64 \times 80}$ is generated randomly by drawing its elements from i.i.d. zero mean and unit variance Gaussian distributions and then normalized to $\|\boldsymbol{\Psi}\|_F^2 = 80$. The parameter α is set to be equal to 1 for the design based on approach 1.

Fig. 2 illustrates that the performance of our designs compare very well with that of the best iterative designs. A particularly relevant aspect relates to the sensing matrix normalization of iterative designs.

[4]We have also performed this experiment and the following experiment with sparse vectors where the randomly placed non-zero elements follow a zero-mean unit-variance Gaussian distribution. Such experiments, which are not reported in view of space limitations, also demonstrate that our designs outperform other designs in the literature.
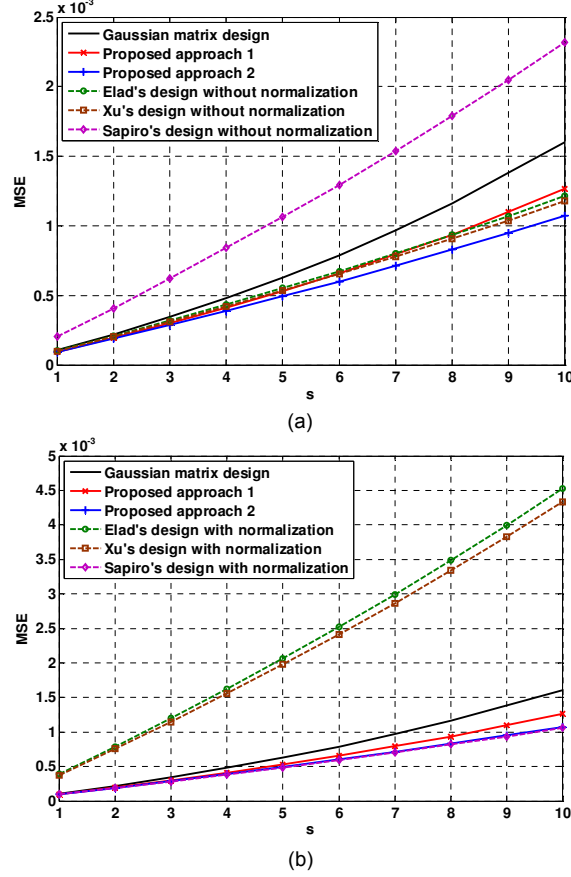
Fig. 2. MSE performance of different sensing matrices for the oracle estimator ($m = 40$, $n = 64$ $\hat{n} = 80$ and $\sigma^2 = 10^{-4}$). (a) Elad's, Xu's and Sapiro's designs are not normalized; (b) Elad's, Xu's and Sapiro's designs are normalized.

Sapiro's design works very well with normalization but Elad's and Xu's designs do not. In fact, the MSE performance of Elad's and Xu's design is worse than that of the random Gaussian design, due to the lower sensing energy (see Fig. 1). The proposed approach 2 has a better MSE performance than approach 1, as the parameter α of approach 1, which is set empirically, affects the performance.

C. The MSE performance using practical estimators

In this investigation, we evaluate the MSE performance of various sensing matrix designs using practical estimators, which include the BPDN, the Dantzig selector and the OMP. As in the previous investigation, the MSE is evaluated by averaging over 1000 trials, where in each trial we generate randomly a sparse vector with s randomly placed ± 1 spikes. The random dictionary $\mathbf{\Psi} \in \mathbb{R}^{64 \times 80}$ is also generated randomly by drawing its elements from i.i.d. zero mean and unit variance Gaussian distributions and then normalized

to $\|\mathbf{\Psi}\|_F^2 = 80$. The parameter α is also set to be equal to 1 for the design based on approach 1.
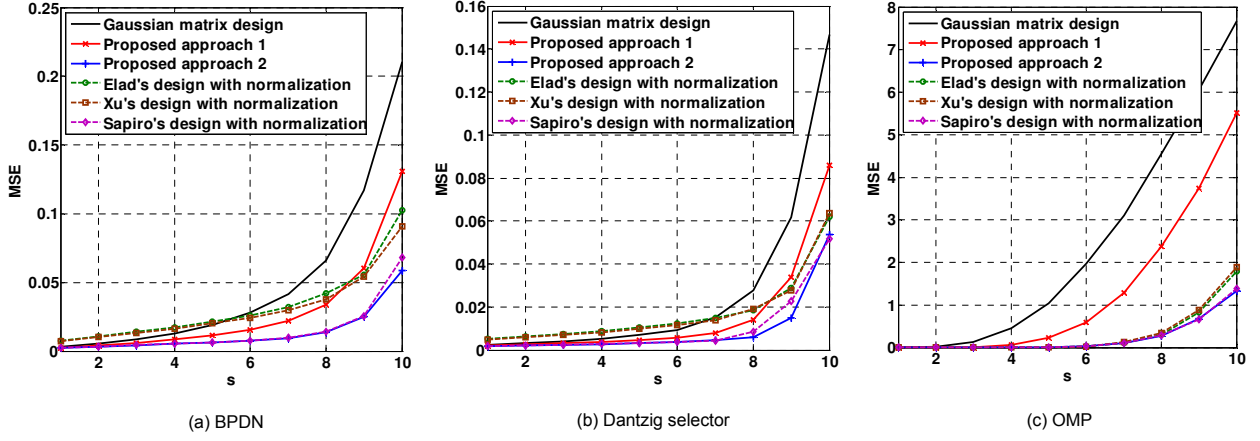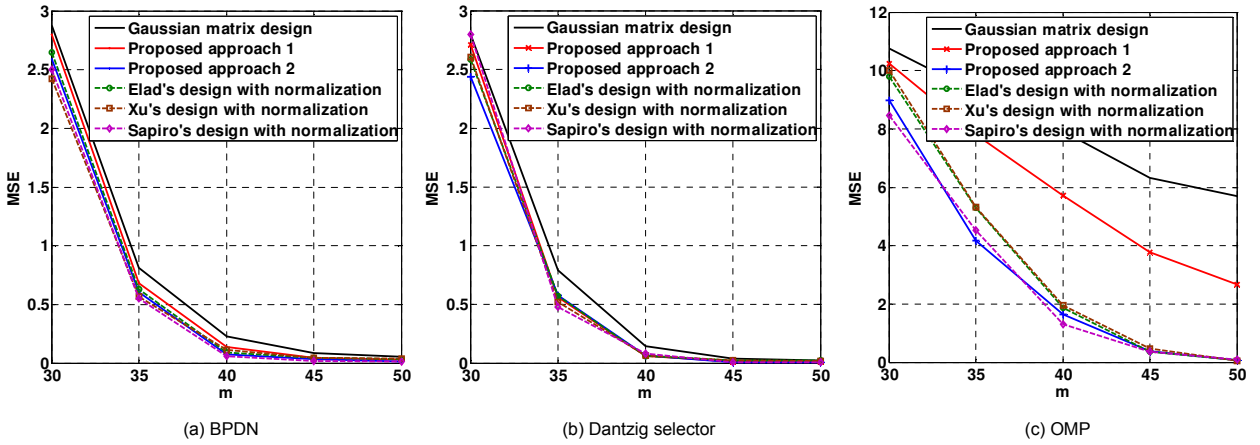

Fig. 3. MSE performance of different sensing matrices for (a) the BPDN, (b) the Dantzig selector, and (c) the OMP ($m = 40$, $n = 64$ $\hat{n} = 80$ and $\sigma^2 = 10^{-4}$).

We first evaluate the MSE performance of various sensing matrix designs for various sparsity levels and for a fixed number of measurements, $m = 40$. Fig. 3 shows that the proposed design approach 1 outperforms the Gaussian matrix design for all the three estimators. In turn, the proposed design approach 2 outperforms all the other designs. In fact, this design is very attractive, due to the low computation cost associated with the generation of the sensing matrix.

Fig. 4. MSE performance of different sensing matrices for (a) the BPDN, (b) the Dantzig selector, and (c) the OMP ($s = 10$, $n = 64$ $\hat{n} = 80$ and $\sigma^2 = 10^{-4}$).

We now evaluate the MSE performance of various sensing matrix designs for various numbers of

measurements and for a fixed sparsity level $s = 10$. Fig. 4 shows once again that the proposed designs outperform the Gaussian matrix design. We note that the proposed designs improve the reconstruction performance for all the three estimators, compared to the Gaussian matrix design. The iterative Elad's design, Sapiro's design, and Xu's design, slightly outperform the proposed designs in some cases, but the computation complexity associated with the generation of these designs is much higher than that associated with the generation of our design.

D. The reconstruction performance for learned dictionaries in CS imaging

We now assess the performance of the proposed designs by considering other practical issues. In particular, we consider real rather than synthetic signals whose representations are typically nearly sparse instead of sparse in some dictionary. We also consider learned dictionaries rather than random ones[5].

In the experiment, we use the cameraman image of size 256×256 pixels, which is partitioned into 1024 nonoverlapping patches of size 8×8 pixels, i.e., $n = 64$. We train a dictionary of size 64×81 for sparsely representing these nonoverlapping patches by using the K-SVD method [46]. The number of measurements for each patch is set to be equal to 40 and the measurements are corrupted by additive zero-mean Gaussian noise with variance $\sigma^2 = 10^{-3}$. We set $\alpha = 1$ for the proposed approach 1. We also use the OMP to reconstruct the image from its noisy measurements owing to its fast execution. We evaluate performance using the reconstructed signal to noise ratio (RSNR):

$$\text{RSNR} = \frac{\|\tilde{\mathbf{f}}\|_2}{\|\mathbf{f} - \tilde{\mathbf{f}}\|_2}, \tag{32}$$

where $\mathbf{f}$ represents the original image and $\tilde{\mathbf{f}}$ represents the reconstructed image.

Fig. 5 demonstrates the higher reconstruction quality and RSNR of our sensing matrix designs in relation to the random Gaussian matrix design. The proposed approach 2 exhibits the best performance. Sapiro's iterative design also exhibits a very good performance but Elad's and Xu's iterative designs with normalized sensing energy exhibit very poor performance, which in fact is worse than that for Gaussian matrix design. Interestingly we recall that the performance of the proposed two designs compare well to that of Gaussian matrix design for random basis and exactly sparse signals as shown in Fig. 3 and 4.

---

[5]We note that the dictionary learning process yields sparse representations that do not necessarily fit the statistical signal model that has been used as a basis of the sensing matrix design procedure. However, the value of the sensing matrix designs is also justified by the fact that it also yields observable gains in this scenario.

(a) Gaussian matrix design
RSNR = 8.5197

(b) Proposed approach 1
RSNR = 9.5386

(c) Proposed approach 2
RSNR = 15.4729

(d) Elad's design with normalization
RSNR = 7.3595

(e) Xu's design with normalization
RSNR = 7.0452

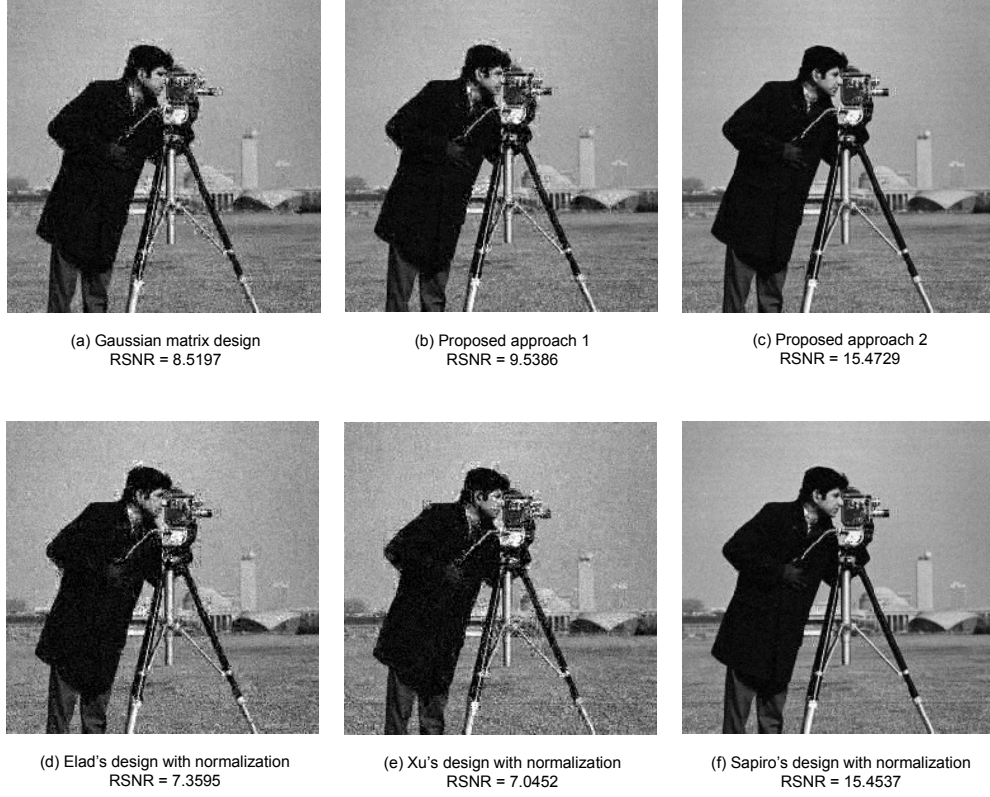(f) Sapiro's design with normalization
RSNR = 15.4537

Fig. 5.   Reconstructed images using a learned basis.

VI. DISCUSSION: RANDOM VS. OPTIMIZED PROJECTIONS

Recent results [47]–[49] have established that - at least asymptotically with the signal ambient dimension - no sensing or reconstruction strategy leads to essentially better performance than random sensing and standard ℓ_1 based reconstruction. In contrast, our results indicate that a tight-frame based sensing matrix design can clearly outperform a random sensing matrix design for low signal ambient dimensions.

It is thus interesting to ask whether our optimized designs can also outperform the random ones with an increase of the signal ambient dimension. This question is also justified by the fact that the recent contributions in the literature concentrate on signals that are sparse in the canonical basis rather than signals that are sparse in an overcomplete dictionary. Interestingly, the numerical analysis reveals that the trends applicable to overcomplete dictionaries can be distinct from those applicable to the canonical dictionary (and also orthonormal ones).

The experiments also consider randomly generated sparse vectors with s randomly placed ± 1 spikes. We consider both a random Gaussian sensing matrix design and an optimized sensing matrix design based
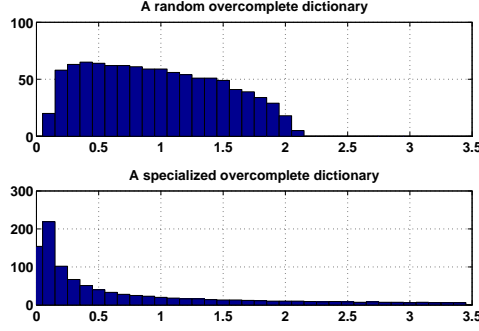
Fig. 6. Histogram of the singular values of a random overcomplete dictionary ($n = 1000$, $\hat{n} = 1200$ and $\|\mathbf{\Psi}\|_F^2 = \hat{n}$) and a specialized overcomplete dictionary ($n = 1000$, $\hat{n} = 1200$, $\|\mathbf{\Psi}\|_F^2 = \hat{n}$ and $\frac{\lambda_{i+1}^{\Psi}}{\lambda_i^{\Psi}} = 0.995$).

on approach 2 due to its low computational cost. The sensing matrix designs are normalized such that $\|\mathbf{\Phi}\|_F^2 = n$. We also consider three distinct dictionaries: i) the canonical basis; ii) a random overcomplete dictionary; and iii) a specified overcomplete dictionary. The random overcomplete dictionary is generated by drawing its elements randomly in accordance with i.i.d. zero-mean unit-variance Gaussian distributions. The specified overcomplete dictionary is generated via its singular value decomposition by taking two randomly generated orthonormal matrices and by taking its positive singular values $\lambda_1^{\Psi} \geq \ldots \geq \lambda_n^{\Psi}$ such that $\lambda_1^{\Psi} = 1$ and $\frac{\lambda_{i+1}^{\Psi}}{\lambda_i^{\Psi}} = 0.995$ ($i = 1, \ldots, n-1$). Both overcomplete dictionaries are also normalized such that $\|\mathbf{\Psi}\|_F^2 = \hat{n}$.

The rationale for considering two different overcomplete dictionaries is because it is not entirely clear how to change the dictionary as the signal ambient dimension is varied[6]. Therefore, two overcomplete dictionaries that exhibit a very different singular value profile as shown in Fig. 6, are chosen that will allow us to articulate different trends in the experiments.

The MSE performance associated with the various sensing matrix designs is also averaged over 1000 trials. We unveil the performance trends by showing how the ratio of the average MSE associated with the optimized sensing matrix design to the average MSE associated with a random sensing matrix design behaves as a function of the signal ambient dimension for various combinations of (m, s), both for the Dantzig selector and the oracle estimator. The signal dimension is restricted to $n = 1000$ due to the long execution time of the simulations.

[6]Note that this issue is not relevant when the signal dimension is fixed as in the previous experiments (or for the canonical basis).
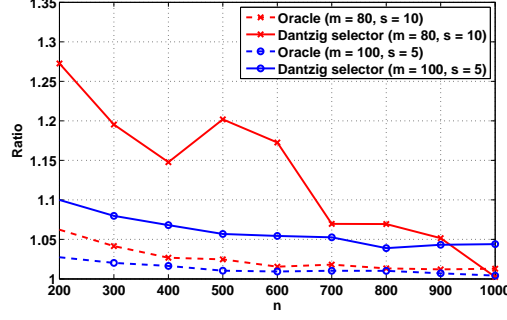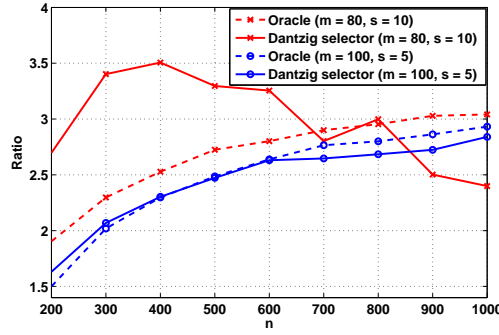
Fig. 7. Ratio of the average MSE associated with an optimized sensing matrix design to that associated with a random Gaussian sensing matrix design for signals that are sparse in the canonical basis ($\sigma^2 = 10^{-4}$).

Fig. 8. Ratio of the average MSE associated with an optimized sensing matrix design to that associated with a random Gaussian sensing matrix design for signals that are sparse in a randomly generated overcomplete dictionarie ($\hat{n} = 1.2n$ and $\sigma^2 = 10^{-4}$).

A. Case I: Signals that are sparse on the canonical basis

Fig. 7 examines how the ratio of the average MSE associated with the optimized sensing matrix design to the average MSE associated with a random Gaussian sensing matrix design - which is a tight frame - behaves as a function of the signal dimension. One observes that the average MSE ratio tends to one with the increase of the signal dimension both for the oracle estimator and the Dantzig selector. This is due to the fact that a random Gaussian matrix tends to a tight frame with the increase of n for a fixed m [50].

It turns out that this result is consistent with the result in [47], where random sensing matrix designs are demonstrated to be near-optimal (asymptotically) for signals that are sparse in the canonical basis.
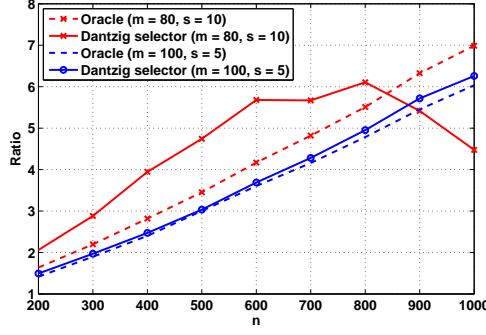
Fig. 9. Ratio of the average MSE associated with an optimized sensing matrix design to that associated with a random Gaussian sensing matrix design for signals that are sparse in a specified overcomplete dictionary ($\hat{n} = 1.2n$ and $\sigma^2 = 10^{-4}$).

B. Case II: Signals that are sparse on an overcomplete dictionary

Figs. 8 and 9 examine how the average MSE ratio behaves as a function of the signal dimension for the random and specified overcomplete dictionaries, respectively. One now observes that - and in sharp contrast to the canonical basis scenario - the average MSE ratio tends to increase with the increase of the signal dimension. This trend is exhibited by the oracle estimator for the pairs ($m = 100$, $s = 5$) and ($m = 80$, $s = 10$). The trend is also exhibited by the Dantzig selector for ($m = 100$, $s = 5$) but not for ($m = 80$, $s = 10$): this exception seems to be due to severe reconstruction errors in view of the fact that one may not be satisfying the requirement $m = \mathcal{O}(s \log(n/s))$ [1], [2].

It is relevant though to point out a major difference in the behavior of the trends for the random and specified overcomplete dictionaries. For the random dictionary, the average MSE ratio appears to saturate with the increase of the signal dimension: this fact can be justified by noting that not only does the optimized design tends to a tight frame with the increase of n for a fixed m - because the m largest singular values of a random dictionary tend to be similar with the increase of n for a fixed m (see also Fig. 6) - but also the random Gaussian matrix design also tends to a tight frame with the increase of n for a fixed m as discussed previously. In contrast, for the specified dictionary the average MSE ratio does not appear to saturate with the increase of the signal dimension.

It turns out that such trends can also be partly reconciled with the arguments of the previous sections. In particular, Figs. 10 and 11 depict how the average sensed energy (i.e., the energy present at the input to the estimator) behaves as a function of the signal dimension for the random and the specified overcomplete dictionaries, respectively. Note that the average sensed energy $\mathbb{E}_{\mathbf{x}} \left(\text{Tr} \left(\mathbf{\Phi} \mathbf{\Psi} \mathbf{x} \mathbf{x}^T \mathbf{\Psi}^T \mathbf{\Phi}^T \right) \right)$ corresponds to the equivalent sensing matrix energy $\|\mathbf{\Phi} \mathbf{\Psi}\|_F^2$ in view of the fact that $\mathbb{E}_{\mathbf{x}}(\mathbf{x}\mathbf{x}^T) = \mathbf{I}_{\hat{n}}$. We
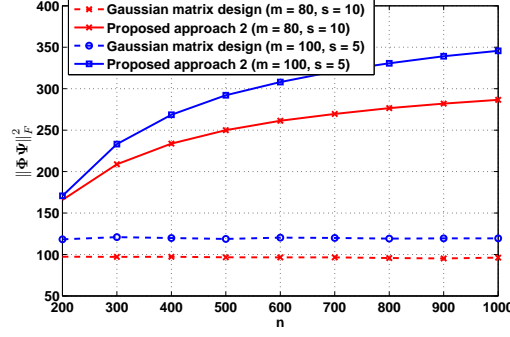
Fig. 10. Average sensed energy $\|\mathbf{\Phi}\mathbf{\Psi}\|_F^2$ for a random overcomplete dictionary ($\hat{n} = 1.2n$).
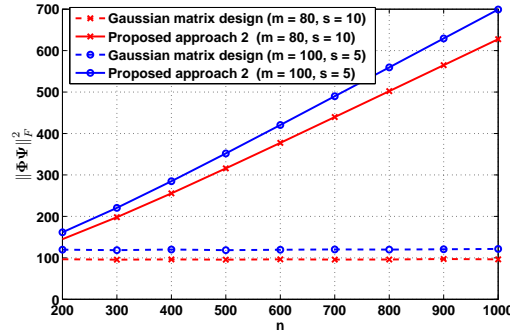

Fig. 11. Average sensed energy $\|\mathbf{\Phi}\mathbf{\Psi}\|_F^2$ for a specialized overcomplete dictionary ($\hat{n} = 1.2n$, $\|\mathbf{\Psi}\|_F^2 = \hat{n}$ and $\frac{\lambda_{i+1}^{\Psi}}{\lambda_i^{\Psi}} = 0.995$).

would like to emphasize that for both Figs. 10 and 11 the sensing matrix designs have been normalized such that $\|\mathbf{\Phi}\|_F^2 = n$.

One observes clearly that the optimized designs have the capability to "sense" higher energy than the random ones in the presence of overcomplete dictionaries (both the random and the specified overcomplete dictionary) and - via the analysis in Section III - potentially have the capability to offer a lower MSE (as confirmed in Figs. 8 and 9). Figs. 10 and 11 also confirm that for the random dictionary the sensed energy tends to saturate with the increase of the signal dimension but for the specified dictionary it does not.

We recognize that this analysis is mainly heuristic: a proper understanding of the advantages of designed projections over random ones in the presence of signals that admit sparse representations in overcomplete dictionaries is beyond the scope of this article. However, the practical relevance of the overall results - independently of whether or not it can be crisply shown that optimized projections clearly outperform

random ones for high ambient dimensions - is also associated with the fact that in some applications it is typical to deal with small dimensions. For example, in certain imaging applications it is standard practice to divide an image into various (possibly overlapping) patches of typically small dimensions [46]. The results then show that there is indeed significant value in using optimized projections in lieu of random ones.

VII. CONCLUSIONS

In this paper, we have considered the design of sensing matrices for CS applications. By showing that one ought to set the equivalent sensing matrix to be equal to a tight frame in order to derive a good MSE performance subject to sensing energy constraints, we have proposed two sensing matrix designs that are instilled with operational significance. Our designs also exhibit various advantages in relation to other designs in the literature. In particular, the proposed designs exhibit MSE performance gains in relation to the conventional random sensing matrix designs as well as other optimized designs. The proposed designs are also closed-form, and as a result easy to generate, whereas other optimized designs in the literature are typically iterative.

ACKNOWLEDGMENT

APPENDIX A
PROOF OF THE PROPOSITION 1

This proof follows the ideas of the proof of Proposition 1 in [24]. Let $\hat{s} \leq s$ be a positive integer. Let also $\mathcal{J}_{\hat{s}}^t \subset \{1, 2, \ldots, \hat{n}\}$ $(t = 1, \ldots, T_{\hat{s}})$ denote a support set with cardinality $\hat{s}$, where $T_{\hat{s}} = \binom{\hat{n}}{\hat{s}} = \frac{\hat{n}!}{\hat{s}!(\hat{n}-\hat{s})!}$. We let $\mathbf{D}_{\mathcal{J}_{\hat{s}}^t} = \mathbf{E}_{\mathcal{J}_{\hat{s}}^t}^T \mathbf{Q} \mathbf{E}_{\mathcal{J}_{\hat{s}}^t}$. We also let $\lambda_{\hat{s}}^{\mathcal{J}_{\hat{s}}^t} \geq \ldots \geq \lambda_1^{\mathcal{J}_{\hat{s}}^t}$ be the eigenvalues of $\mathbf{D}_{\mathcal{J}_{\hat{s}}^t}$. Let $\Pr(\|\mathcal{J}\|_0 = \hat{s})$ denote the probability that the support size of $\mathcal{J}$ is $\hat{s}$.

We now note that

$$
\begin{aligned}
\sum_{t=1}^{T_{\hat{s}}} \mathrm{Tr}\left(\mathbf{D}_{\mathcal{J}_{\hat{s}}^t}\right) &= \mathrm{Tr}\left(\left(\sum_{t=1}^{T_{\hat{s}}} \mathbf{E}_{\mathcal{J}_{\hat{s}}^t} \mathbf{E}_{\mathcal{J}_{\hat{s}}^t}^T\right) \mathbf{Q}\right) \\
&= \mathrm{Tr}\left(\frac{\hat{s} T_{\hat{s}}}{\hat{n}} \mathbf{I}_{\hat{n}} \mathbf{Q}\right) \\
&= \frac{m \hat{s} T_{\hat{s}}}{\hat{n}}.
\end{aligned}
\tag{33}
$$

By the arithmetic mean - harmonic mean inequality, it follows that:

$$\sum_{t=1}^{T_{\hat{s}}}\sum_{\hat{n}=1}^{\hat{s}}\frac{1}{\lambda_{\hat{n}}^{\mathcal{J}_{\hat{s}}^t}} \geq \frac{(\hat{s}T_{\hat{s}})^2}{\sum_{t=1}^{T_{\hat{s}}}\sum_{\hat{n}=1}^{\hat{s}}\lambda_{\hat{n}}^{\mathcal{J}_{\hat{s}}^t}} = \frac{(\hat{s}T_{\hat{s}})^2}{\sum_{t=1}^{T_{\hat{s}}}\text{Tr}\left(\mathbf{D}_{\mathcal{J}_{\hat{s}}^t}\right)} = \frac{\hat{n}\hat{s}T_{\hat{s}}}{m}, \tag{34}$$

where one achieves the lower bound with $\lambda_{\hat{n}}^{\mathcal{J}_{\hat{s}}^t} = \frac{m}{\hat{n}}$ ($\hat{n} = 1,\ldots,\hat{s}$; $t = 1,\ldots,T_{\hat{s}}$). This implies immediately that the matrix $\mathbf{Q} = \frac{m}{\hat{n}}\mathbf{I}_{\hat{n}}$, which is consistent with the constraints, minimizes:

$$\mathbb{E}_{\mathcal{J}_{\hat{s}}}\left(\text{Tr}\left((\mathbf{D}_{\mathcal{J}_{\hat{s}}})^{-1}\right)\right) = \frac{1}{T_{\hat{s}}}\sum_{t=1}^{T_{\hat{s}}}\sum_{\hat{n}=1}^{\hat{s}}\frac{1}{\lambda_{\hat{n}}^{\mathcal{J}_{\hat{s}}^t}}, \tag{35}$$

and hence also minimizes:

$$\mathbb{E}_{\mathcal{J}}\left(\text{Tr}\left((\mathbf{D}_{\mathcal{J}})^{-1}\right)\right) = \sum_{\hat{s}=1}^{s}\text{Pr}(\|\mathcal{J}\|_0 = \hat{s})\mathbb{E}_{\mathcal{J}_{\hat{s}}}\left(\text{Tr}\left((\mathbf{D}_{\mathcal{J}_{\hat{s}}})^{-1}\right)\right). \tag{36}$$

APPENDIX B

PROOF OF THE PROPOSITION 2

This proof follows the ideas of the proof of Proposition 2 in [24]. By using the SVD $\mathbf{A} = \mathbf{U_A}\mathbf{\Lambda_A}\mathbf{V_A}^T$, where $\mathbf{U_A} \in \mathbb{R}^{m\times m}$ and $\mathbf{V_A} \in \mathbb{R}^{\hat{n}\times\hat{n}}$ are orthonormal matrices, and $\mathbf{\Lambda_A} \in \mathbb{R}^{m\times\hat{n}}$ is a matrix whose main diagonal entries ($\lambda_1^A \geq \lambda_2^A \geq \ldots \lambda_m^A \geq 0$) are the singular values of $\mathbf{A}$, and the off-diagonal entries are zeros, we pose the convex optimization problem:

$$\min_{\lambda_1^A,\ldots,\lambda_m^A} \sum_{i=1}^{m}\left(\left(\lambda_i^A\right)^2 - \frac{m}{\hat{n}}\right)^2$$
$$\text{s.t.} \sum_{i=1}^{m}\left(\lambda_i^A\right)^2 = m, \ \lambda_i^A \geq 0 \ (i = 1,2,\ldots,m). \tag{37}$$

which, in view of the fact that $\left\|\mathbf{A}^T\mathbf{A} - \frac{m}{\hat{n}}\mathbf{I}_{\hat{n}}\right\|_F^2 = \sum_{i=1}^{m}\left(\left(\lambda_i^A\right)^2 - \frac{m}{\hat{n}}\right)^2$ and $\text{Tr}\left(\mathbf{A}^T\mathbf{A}\right) = \sum_{i=1}^{m}\left(\lambda_i^A\right)^2$, leads to the solution of (15). Since the solution of (37) is $\lambda_i^A = 1$ ($i = 1,2,\ldots,m$), it follows that any Parseval tight frame is the solution of (15).

APPENDIX C

TIGHT FRAMES AND STRIP

Another benefit of tight frames - more precisely, unit-norm tight frames - is its relation to the weaker version of the RIP, namely the StRIP. The StRIP, which has been proposed by Calderbank et al. [22], can be used to evaluate the expected-case performance of CS, whereas the RIP is a worst-case performance indicator, as is the mutual coherence. The StRIP guarantees successful reconstruction of all but an exponentially small fraction of s sparse signals. The definition of StRIP uses a probability criterion to replace the hard requirement demanded in the definition of the RIP.

Definition 2: A matrix $\mathbf{A} \in \mathbb{R}^{m \times \hat{n}}$ is said to be an (s, δ, η)-StRIP matrix if for s sparse vectors $\mathbf{x} \in \mathbb{R}^{\hat{n}}$ the inequalities

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2 \tag{38}$$

hold with probability exceeding $1 - \eta$ with respect to a uniform distribution of the vectors $\mathbf{x}$ among all s sparse vectors in $\mathbb{R}^{\hat{n}}$ having the same fixed magnitudes.

Calderbank et al. [22] demonstrate that deterministic sensing matrices are StRIP matrices if they satisfy all of the following criteria:

1) The rows of $\mathbf{A}$ are orthogonal and all the row sums are zero, i.e., $\sum_{i=1}^{\hat{n}} \mathbf{a}_i = \mathbf{0}$;

2) The columns of $\mathbf{A}$ form a group under point-wise multiplication;

3) There is one column of $\mathbf{A}$ equal to $\mathbf{1}$, which can be assumed as the first column. For all $i \in \{2, \ldots, \hat{n}\}$, $\|\mathbf{a}_i\|_2^2 \leq m^{2-\beta}$, where $0 < \beta \leq 1$.

A large class of matrices, including discrete chirp sensing matrices, Bose, Chaudhuri, and Hocquenghem (BCH) sensing matrices, Kerdock, Delsarte-Goethals and second order Reed Muller sensing matrices, satisfy these criteria, and thus are StRIP matrices. In [22], they prove that the RIP of these matrices is satisfied with a probability exceeding $1 - \mathcal{O}\left(\exp\left(-\frac{\delta^2 m^\beta}{s}\right)\right)$. However, a unit norm tight frame does not necessarily satisfy these criteria. For example, an orthonormal matrix, which is also a unit norm tight frame with $m = \hat{n}$, does not necessarily satisfy $\sum_{i=1}^{\hat{n}} \mathbf{a}_i = \mathbf{0}$.

The following Proposition demonstrates that a unit-norm tight frame is also a StRIP matrix

Proposition 4: Let $\mathbf{A} \in \mathbb{R}^{m \times \hat{n}}$ be a unit norm tight frame with mutual coherence equal to μ. For any s sparse vectors $\mathbf{x} \in \mathbb{R}^{\hat{n}}$, the RIP holds with probability:

$$\mathbb{P}\left(\left|\|\mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\right| \leq \delta \|\mathbf{x}\|_2^2\right) > 1 - (s/2)^{-\frac{\left(0.3894\delta - \frac{s}{m}\right)^2}{36\mu^2 s \log_e(1+s/2)}}, \tag{39}$$

where $\sqrt{237.42\mu^2 s \log_e(1 + s/2)} + \frac{2.57s}{m} \leq \delta < 1$.

Proof: Let $\mathbf{A}_{\mathcal{J}_s} \in \mathbb{R}^{m \times s}$ be an s-column submatrix of $\mathbf{A} \in \mathbb{R}^{m \times \hat{n}}$ ($s < m \leq \hat{n}$), where $\mathcal{J}_s \subset \{1, \ldots, \hat{n}\}$ denotes a support set with cardinality s. Let $\lambda_1^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} \geq \ldots \geq \lambda_s^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} \geq 0$ be the eigenvalues of the positive semi-definite matrix $\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}$. We have that the maximum eigenvalue $\lambda_1^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}}$ and minimum eigenvalue $\lambda_s^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}}$ of $\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}$ are given by

$$\lambda_1^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} = \max_{\mathbf{z} \neq \mathbf{0}} \frac{\|\mathbf{A}_{\mathcal{J}_s}\mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2}, \tag{40}$$

and

$$\lambda_s^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} = \min_{\mathbf{z} \neq \mathbf{0}} \frac{\|\mathbf{A}_{\mathcal{J}_s}\mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2}, \tag{41}$$

where $\mathbf{z} \in \mathbb{R}^s$. Therefore, we have

$$\lambda_s^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} \|\mathbf{z}\|_2^2 \leq \|\mathbf{A}_{\mathcal{J}_s} \mathbf{z}\|_2^2 \leq \lambda_1^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} \|\mathbf{z}\|_2^2, \tag{42}$$

or

$$(\lambda_s^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} - 1)\|\mathbf{z}\|_2^2 \leq \|\mathbf{A}_{\mathcal{J}_s} \mathbf{z}\|_2^2 - \|\mathbf{z}\|_2^2 \leq (\lambda_1^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} - 1)\|\mathbf{z}\|_2^2, \tag{43}$$

for any $\mathbf{z} \in \mathbb{R}^s$. By defining $\delta_{\mathcal{J}_s} = \max\left(\left|\lambda_1^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} - 1\right|, \left|\lambda_s^{\mathbf{A}_{\mathcal{J}_s}^T \mathbf{A}_{\mathcal{J}_s}} - 1\right|\right)$, it follows that

$$\left|\|\mathbf{A}_{\mathcal{J}_s} \mathbf{z}\|_2^2 - \|\mathbf{z}\|_2^2\right| \leq \delta_{\mathcal{J}_s} \|\mathbf{z}\|_2^2. \tag{44}$$

We can immediately derive that the RIC satisfies $\delta_s = \max_{\mathcal{J}_s} \delta_{\mathcal{J}_s}$ by comparing (44) with (5).

The following theorem, which has been proved by Tropp in [51], defines a probability bound for $\delta_{\mathcal{J}_s}$.

Theorem 2: Theorem Let $\mathbf{A} \in \mathbb{R}^{m \times \hat{n}}$ ($m < \hat{n}$) be a matrix whose columns have unit norm, i.e., $\|\mathbf{a}_i\|_2 = 1$ for all $i \in \{1, \ldots, \hat{n}\}$, $\mathbf{A}_{\mathcal{J}_s} \in \mathbb{R}^{m \times s}$ ($s < m$) be a random s-column submatrix of $\mathbf{A}$ with a support $\mathcal{J}_s \subset \{1, \ldots, \hat{n}\}$ of cardinality s, and μ be the mutual coherence of $\mathbf{A}$. Suppose that

$$\sqrt{144\mu^2 s \log_e(1 + s/2)\rho} + \frac{2s}{\hat{n}}\|\mathbf{A}\|^2 \leq e^{-0.25}\delta, \tag{45}$$

where $\rho \geq 1$ and $0 < \delta < 1$. Then

$$\Pr(\delta_{\mathcal{J}_s} \geq \delta) \leq (s/2)^{-\rho}. \tag{46}$$

We now use the fact that $\mathbf{A}$ is a unit norm tight frame with frame bound equal to $\frac{\hat{n}}{m}$, so that $\|\mathbf{A}\|^2 = \frac{\hat{n}}{m}$. We then rewrite (45) to be

$$1 \leq \rho \leq \frac{\left(e^{-0.25}\delta - \frac{2s}{m}\right)^2}{144\mu^2 s \log_e(1 + s/2)}, \tag{47}$$

where $\sqrt{144e^{0.5}\mu^2 s \log_e(1 + s/2)} + \frac{2e^{0.25}s}{m} \leq \delta < 1$. Since the inequality (46) holds for any ρ satisfying (45), we have that for a random set $\mathcal{J}_s$, the inequality (44) leads to the probability bound

$$\mathbb{P}\left(\left|\|\mathbf{A}_{\mathcal{J}_s} \mathbf{z}\|_2^2 - \|\mathbf{z}\|_2^2\right| \leq \delta\|\mathbf{z}\|_2^2\right) \geq \mathbb{P}(\delta_{\mathcal{J}_s} \leq \delta)$$
$$> 1 - (s/2)^{-\frac{\left(0.3894\delta - \frac{s}{m}\right)^2}{36\mu^2 s \log_e(1 + s/2)}}, \tag{48}$$

when $\sqrt{237.42\mu^2 s \log_e(1 + s/2)} + \frac{2.57s}{m} \leq \delta < 1$. ∎

Remark 1: The mutual coherence μ plays as an important role in this probability bound. The mutual coherence of various unit norm tight frames could be different, and its distribution is unknown. However, the mutual coherence is fixed for some specific unit norm tight frames. For example, the Fourier-Dirac tight frame has mutual coherence $\mu = \frac{1}{\sqrt{m}}$ [51], and the equiangular tight frame has mutual coherence $\mu = \sqrt{\frac{\hat{n} - m}{m(\hat{n} - 1)}}$ [52].

Remark 2: In [22], the authors conclude that the RIP is satisfied with a probability exceeding $1 - \mathcal{O}\left(\exp\left(-\frac{\delta^2 m^\beta}{s}\right)\right)$ $(0 < \beta \leq 1)$ for a large class of matrices, including discrete chirp sensing matrices, Bose, Chaudhuri, and Hocquenghem (BCH) sensing matrices, Kerdock, Delsarte-Goethals and second order Reed Muller sensing matrices. According to Proposition 4, we can conclude that the RIP holds with a probability that exceeds $1 - \mathcal{O}\left(\exp\left(-\frac{\delta^2 m}{s}\right)\right)$ for Fourier-Dirac tight frames and the equiangular tight frame, so that these tight frames exhibit better quality in terms of the StRIP in relation to the sensing matrices in [22].

Remark 3: Proposition 4 requires unit norm tight frames with frame bound $\frac{\hat{n}}{m}$. It turns out that, one can scale a unit norm tight frame via $\mathbf{A} = \frac{\sqrt{m}\mathbf{A}}{\|\mathbf{A}\|_F}$, which leads to a Parseval tight frame with an equal column norm, in order to achieve the frame bound equal to 1 used in the paper. In fact, scaling the unit norm tight frames does not change the matrix structure, only the sensing energy.

REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489 – 509, Feb. 2006.

[2] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289 –1306, Apr. 2006.

[3] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[4] W. Chen, M. Rodrigues, and I. Wassell, "A fréchet mean approach for compressive sensing date acquisition and reconstruction in wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 10, pp. 3598 –3606, Oct. 2012.

[5] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982 –2001, Apr. 2010.

[6] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2346 –2356, Jun. 2008.

[7] C. Hegde and R. Baraniuk, "Signal recovery on incoherent manifolds," *Information Theory, IEEE Transactions on*, vol. 58, no. 12, pp. 7204 –7214, Dec. 2012.

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

[9] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *Image Processing, IEEE Transactions on*, vol. 20, no. 12, pp. 3419 –3430, Dec. 2011.

[10] C. Fu, X. Ji, and Q. Dai, "Adaptive compressed sensing recovery utilizing the property of signal's autocorrelations," *Image Processing, IEEE Transactions on*, vol. 21, no. 5, pp. 2369 –2378, May 2012.

[11] Z. Zhang and B. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 5, pp. 912 –926, Sept. 2011.

[12] T. Peleg, Y. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *Signal Processing, IEEE Transactions on*, vol. 60, no. 5, pp. 2286 –2303, May 2012.

[13] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Image upsampling via spatially adaptive block-matching filtering," in *Proc. of 16th European Signal Processing Conference, EUSIPCO2008*, Aug. 2008.

[14] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *Proc. IEEE International Conference on Image Processing, ICIP 2007.*, Sept. 2007, pp. 549 – 552.

[15] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Spatially adaptive filtering as regularization in inverse imaging: Compressive sensing, super-resolution, and upsampling," *Super-Resolution Imaging (P. Milanfar, ed.), CRC Press / Taylor & Francis*, pp. 123–153, Sept. 2010.

[16] M. Elad, "Optimized projections for compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 55, no. 12, pp. 5695 –5702, Dec. 2007.

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

[18] J. Xu, Y. Pi, and Z. Cao, "Optimized projection matrix for compressive sensing," *EURASIP J. Adv. Signal Process*, vol. 43, pp. 1–8, Feb. 2010.

[19] L. Zelnik-Manor, K. Rosenblum, and Y. Eldar, "Sensing matrix optimization for block-sparse decoding," *Signal Processing, IEEE Transactions on*, vol. 59, no. 9, pp. 4300 –4312, Sept. 2011.

[20] W. Carson, M. Chen, M. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM Journal on Imaging Sciences*, vol. 5, no. 4, pp. 1185–1212, 2012.

[21] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 6 – 18, Jan. 2006.

[22] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 358 –374, Apr. 2010.

[23] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.

[24] W. Chen, M. Rodrigues, and I. Wassell, "On the use of unit-norm tight frames to improve the average mse performance in compressive sensing applications," *Signal Processing Letters, IEEE*, vol. 19, no. 1, pp. 8 –11, Jan. 2012.

[25] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *Information Theory, IEEE Transactions on*, vol. 54, no. 5, pp. 2210–2219, May 2008.

[26] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59 – 73, 2011.

[27] M. Davenport, D. Needell, and M. Wakin, "Signal space cosamp for sparse recovery with redundant dictionaries," *arXiv preprint arXiv:1208.0353*, 2012.

[28] F. Champagnat, Y. Goussard, and J. Idier, "Unsupervised deconvolution of sparse spike trains using stochastic approximation," *Signal Processing, IEEE Transactions on*, vol. 44, no. 12, pp. 2988 –2998, Dec. 1996.

[29] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "An iterative bayesian algorithm for sparse component analysis in presence of noise," *Signal Processing, IEEE Transactions on*, vol. 57, no. 11, pp. 4378 –4390, Nov. 2009.

[30] N. Dobigeon and J.-Y. Tourneret, "Bayesian orthogonal component analysis for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 58, no. 5, pp. 2675 –2685, May 2010.

[31] R. Gribonval and K. Schnass, "Dictionary identification-sparse matrix-factorization via ℓ_1-minimization," *Information Theory, IEEE Transactions on*, vol. 56, no. 7, pp. 3523 –3539, Jul. 2010.

[32] F. Labeau, J.-C. Chiang, M. Kieffer, P. Duhamel, L. Vandendorpe, and B. Macq, "Oversampled filter banks as error correcting codes: theory and impulse noise correction," *Signal Processing, IEEE Transactions on*, vol. 53, no. 12, pp. 4619 – 4630, Dec. 2005.

[33] M. Hyder and K. Mahata, "An improved smoothed ℓ_0 approximation algorithm for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 58, no. 4, pp. 2194–2205, Apr. 2010.

[34] C. Soussen, J. Idier, D. Brie, and J. Duan, "From bernoulli-gaussian deconvolution to sparse signal restoration," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4572 –4584, Oct. 2011.

[35] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[36] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus-Mathématique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[37] E. Candès and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203 – 4215, Dec. 2005.

[38] S. Kim and C. Yoo, "Underdetermined blind source separation based on subspace representation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2604 –2614, Jul. 2009.

[39] V. Bostanov, "Bci competition 2003-data sets ib and iib: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1057 –1061, Jun. 2004.

[40] G. Yen and K.-C. Lin, "Wavelet packet feature extraction for vibration monitoring," *Industrial Electronics, IEEE Transactions on*, vol. 47, no. 3, pp. 650 –667, Jun. 2000.

[41] Z. Ben-Haim and Y. Eldar, "The cramèr-rao bound for estimating a sparse parameter vector," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3384 –3389, Jun. 2010.

[42] Z. Ben-Haim, Y. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *Signal Processing, IEEE Transactions on*, vol. 58, no. 10, pp. 5030 –5043, Oct. 2010.

[43] E. G. Larsson and Y. Selen, "Linear regression with a sparse parameter vector," *Signal Processing, IEEE Transactions on*, vol. 55, no. 2, pp. 451 – 460, Feb. 2007.

[44] P. Schniter, L. Potter, and J. Ziniel, "Fast bayesian matching pursuit: Model uncertainty and parameter estimation for sparse linear models," *OSU ECE Technical Report*, 2009.

[45] M. Elad and I. Yavneh, "A plurality of sparse representations is better than the sparsest one alone," *Information Theory, IEEE Transactions on*, vol. 55, no. 10, pp. 4701 – 4714, Oct. 2009.

[46] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736 –3745, Dec. 2006.

[47] E. J. Candès and M. A. Davenport, "How well can we estimate a sparse vector?" *Applied and Computational Harmonic Analysis (In Press)*, 2012.

[48] F. Ye and C.-H. Zhang, "Rate minimaxity of the lasso and dantzig selector for the ℓ_q loss in ℓ_r balls," *J. Mach. Learn. Res.*, vol. 11, pp. 3519–3540, Dec. 2010.

[49] G. Raskutti, M. Wainwright, and B. Yu, "Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls," *Information Theory, IEEE Transactions on*, vol. 57, no. 10, pp. 6976 – 6994, Oct. 2011.

[50] M. Rudelson and R. Vershynin, "Smallest singular value of a random rectangular matrix," *Communications on Pure and Applied Mathematics*, vol. 62, no. 12, pp. 1707–1739, 2009.

[51] J. A. Tropp, "On the conditioning of random subdictionaries," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 1 – 24, 2008.

[52] T. Strohmer and R. W. H. Jr., "Grassmannian frames with applications to coding and communication," *Applied and Computational Harmonic Analysis*, vol. 14, no. 3, pp. 257 – 275, 2003.