Incoherent and Robust Projection Matrix Design Based on Equiangular Tight Frame

MEENAKSHI†1, (Member, IEEE), AND SESHA N SRIRANGARAJAN‡1,2, (Member, IEEE)

†Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi 110001, India
‡Bharti School of Telecommunication Technology and Management, Indian Institute of Technology Delhi, New Delhi 110001, India

Corresponding author: Seshan Srirangarajan (seshan@ee.iitd.ac.in)

This work was supported in part by the Semiconductor Research Corporation (SRC) under Grant 2019-IR-2924.

ABSTRACT Designing a projection matrix to optimally select the informative samples from high-dimensional data is a challenging task. Several approaches have been proposed for this task, however conventional methods obtain the projection matrix from the corresponding Gram matrix without considering the underlying structure of the equiangular frame. The study propose a framework to optimize the projections based on the equivalent tight frame, which is in turn constructed from the target Gram matrix. The proposed work optimizes the projection matrix by restricting the eigenvalues of the corresponding Gram matrix to ensure reduced pairwise correlation and tightness of the frame. Additionally, an $\ell_2,1$-norm based regularization term and a projection matrix energy constraint are incorporated to reduce the effect of outliers and noisy data. This unified optimization problem results in an incoherent and robust projection matrix. Experiments are performed on synthetic data as well as real images. The performance evaluation is carried out in terms of mutual coherence, signal reconstruction accuracy, and peak signal-to-noise ratio (PSNR). The results show that the sensing error constraint enables the design of optimized projections especially when the signals are noisy and not exactly sparse which is the case in real-world scenarios.

INDEX TERMS Compressed sensing, projection matrix, mutual coherence, equiangular tight frame, sparse encoding error, $\ell_2,1$-norm.

I. INTRODUCTION

In many areas of science and technology, one of the key tasks is to deduce quantities of interest from the measured or sensed data. For instance, in signal and image processing applications, we would like to obtain the signal of interest from the observed measurements. In mathematical terms, the measurement vector $y \in \mathbb{R}^{M \times 1}$ can be related to the signal of interest $x \in \mathbb{R}^{N \times 1}$ via a projection matrix $\Phi \in \mathbb{R}^{M \times N}$ as:

$$
y = \Phi x$$

Conventional techniques suggest that the number of measurements $M$, must be at least as large as the length of the underlying signal. This concept is related to the Nyquist-Shannon sampling theorem. However using compressed sensing (CS) techniques, it is possible to recover signals of length $N$ with the number of measurements being much smaller than $N$, provided the signal is sparse [1]–[6]. In the CS framework, we can reconstruct the signal $x$ from the measurement $y$ assuming the relation in (1) with $M \ll N$.

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchar.

This under-determined system has an infinite number of solutions for $x$ and the sparsity constraint enables us to determine a unique solution.

A signal is said to be sparse if a large number of its values are zero or can be discarded without significant loss of information. Many real signals are compressible and can be approximated using sparse signals either directly or after application of an appropriate transform. If $x$ is a real compressible signal, then it can be expressed as:

$$
x = \sum_{i=1}^{L} \theta_i \psi_i + e = \Psi \theta + e \tag{2}$$

where $\Psi = [\psi_1, \ldots, \psi_L] \in \mathbb{R}^{N \times L}$ is the (sparsifying) transform basis, $\theta \in \mathbb{R}^{L}$ is the vector of sparse coefficients, and $e \in \mathbb{R}^{N}$ is the sparse encoding error (SEE). If $e = 0$, then the signal $x$ is said to be exactly sparse in $\Psi$ with sparsity level $s = \|\theta\|_0$, whereas if $e$ is non-zero but has relatively low energy then $x$ is said to be approximately $s$-sparse with $s = \|\theta\|_0$. Here $\|\theta\|_0$ denotes the number of non-zero elements in $\theta$ and $0$ represents a vector with all the elements being zero.
A well-approximated representation of the signal $x$ requires appropriately designed projections. From (1) and (2), $y = \Phi \Psi \theta + \Phi e$, where $\Phi$ is the projection (or sensing) matrix, $\Phi e$ is the projected noise as a result of the SEE, and $D = \Phi \Psi$ is the equivalent dictionary of the CS system. Let $D \in \mathbb{R}^{M \times L}$ be a matrix with unit norm columns $d_1, d_2, \ldots, d_L$, also known as atoms of the frame $D$. Recovery of the underlying signal $x$ requires the equivalent dictionary $D$ to be as incoherent as possible. The coherence measure $\mu$ is given by:

$$\mu = \max_{1 \leq i \neq j \leq L} |\langle d_i, d_j \rangle|$$

(3)

which refers to the largest (or worst-case) absolute correlation between any two element vectors of $D$ [7]–[9]. The Gram matrix of $D$ is defined as $G = D^T D$, which is a positive semidefinite and symmetric matrix with unit diagonal elements and rank $M$. The off-diagonal elements of $G$ correspond to the inner product of two distinct atoms of $D$. To recover the signal from under-sampled measurements, CS requires sparsity of $x$ and incoherence of $D$. Assuming $y$ and $D$ are known, $x$ can be obtained by solving:

$$\min \| y - D \theta \|_2 \quad \text{s.t.} \quad \| \theta \|_0 \leq s$$

(4)

which is NP-hard. Greedy algorithms, such as orthogonal matching pursuit (OMP) [10], matching pursuit (MP) [11], generalized OMP [12], and others [13] can be employed to solve (4) for $\theta$ under certain conditions (with theoretical guarantees) and recover $x$. For a fixed transform basis $\Psi$, projection matrix is a key determinant in the performance of the recovery algorithms, both convex relaxation-based and greedy algorithms. Optimized projections allow improved signal recovery performance from under-sampled measurements [14].

If $y$ is captured using the frame $D$, the signal $x$ with sparsity $s$ can be recovered by solving (4), provided the following constraint is satisfied:

$$s < \frac{1}{2} \left( 1 + \frac{1}{\mu} \right)$$

(5)

The largest absolute value of the pairwise correlation is lower bounded by the Welch bound ($\mu_w$) [15], and when the signal dimension $N \leq M^2$, is given by:

$$\mu \geq \mu_w = \frac{(L - M)}{M(L - 1)}$$

(6)

The $t$-averaged mutual coherence $\mu_t$ [14], for a given coherence threshold $t$, is an alternative metric for evaluating recovery performance of the projection matrix and given by:

$$\mu_t = \frac{\sum_{1 \leq i \neq j \leq L} (|g_{ij}| \geq t) \cdot |g_{ij}|}{\sum_{1 \leq i \neq j \leq L} (|g_{ij}| \geq t)}$$

(7)

These coherence measures will be used as the performance metrics for evaluating the efficacy of the optimized projections.

Restricted isometry property (RIP) and spark of the projection matrix can be effective in evaluating the effect of optimized projections [3], [6], [9], [16], [16]. The restricted isometry constant (RIC) $\delta_s \in (0, 1)$ is the smallest value of $\delta$ that satisfies [8]:

$$(1 - \delta)\|\theta\|_2^2 \leq \|\Phi \Psi \theta\|_2^2 \leq (1 + \delta)\|\theta\|_2^2$$

(8)

where $\theta \in \mathbb{R}^L$ is an arbitrary $s$-sparse signal. Spark of a matrix is defined as the smallest number of linearly dependent columns. The projection matrix plays two key roles. Firstly, it determines how close the sparsest solution obtained under certain conditions (with theoretical guarantees) and recover $x$. Secondly, projection matrix affects the recovery performance. With RIC $\delta_s < \frac{1}{2\sqrt{s+1}}$ ensures the successful recovery of a sparse signal using OMP when there is no measurement noise [18].

In this paper, we apply concepts from frame theory for designing incoherent tight frames. A frame $D$ in a real or complex $M$-dimensional Hilbert space can be represented as a sequence of $L \geq M$ basis vectors $d_k \in \mathbb{H}^M$, satisfying the Parseval’s condition [19], [20]:

$$\alpha \|v\|_2^2 \leq \|D^Tv\|_2^2 \leq \beta \|v\|_2^2, \quad \forall v \in \mathbb{H}^M$$

(9)

where $\alpha$ and $\beta$ are positive constants. If all the frame vectors have unit $\ell_2$-norm i.e., $\|d_k\|_2 = 1 \forall k$, the frame is known as unit norm frame (UNF). A frame is an equiangular frame if there exists a constant $c \in (0, 1)$ such that $|\langle d_i, d_j \rangle| = c$, $\forall i \neq j$. A frame with the smallest maximum correlation among all UNFs $\in \mathbb{R}^{M \times L}$ in a finite dimension is said to be a Grassmannian frame [21]. The measure of over-completeness or redundancy of a frame $D$ is given by $\rho = L/M$ [21]. If $\alpha = \beta$, then $D$ is said to be an $\alpha$-tight frame which satisfies the relation $\sum_{k=1}^{L} \|d_k\|_2^2 = \alpha M$. In CS, Welch bound for mutual coherence can be achieved with an equiangular tight frame (ETF) as it has low averaged mutual coherence which is essential for the minimal dependency property. An equiangular tight frame is a unit-norm tight frame (UNTF) in which each pair of column vectors has the same absolute inner product [21].

A. OUR CONTRIBUTION

We propose an incoherent and robust projection matrix design (IR-PMD) method. In IR-PMD, an alternating minimization approach is used for constructing an ETF and designing a robust projection matrix using the designed frame based on an $\ell_2,1$-norm regularization and weighted penalty term. Our main contributions are three fold:

- We propose a method for constructing an ETF based on the structural and spectral constraints for reduced
mutual coherence and improved tightness properties by
restricting the eigenvalues of the Gram matrix to achieve
a smaller RIC.

- We formulate the projection matrix design as an
  $\ell_{2,1}$-norm based optimization problem with sensing
  matrix energy constraint. The $\ell_{2,1}$-norm not only
  ensures robustness to outliers and noisy data but also
  promotes row sparsity in order to identify the informa-
  tive samples which to the best of our knowledge is the
  first instance of the $\ell_{2,1}$-norm being employed for
  projection matrix design.

- An adaptive weighted penalty term is introduced to min-
  imize the sensing matrix energy together with the error
term for optimizing the projection matrix.

- We present an alternating minimization algorithm to
  optimize the ETF and design a robust projection matrix
  using the designed ETF.

We will demonstrate the performance of the proposed
approach through a set of experiments on synthetic data and
real images. The visual recovery results will be shown for
some standard images.

B. OUTLINE OF THE PAPER

Section II describes related work in the area of opti-
mized projection matrix design. In Section III, design of
ETF based on the structural and spectral constraints is pre-
sented. Section IV describes the proposed robust projection
matrix design approach. We discuss the sparse encoding error
followed by the proposed framework for designing optimized
projections using the target frame designed based using the
method in Section III. In Section V, we present simulation
results using synthetic data and real images for demonstrating
the performance of the proposed method and comparison
with existing approaches. Concluding remarks are presented
in Section VI.

C. NOTATIONS

Throughout this paper, lowercase letters denote scalars, low-
ercase bold letters denote vectors, and uppercase bold let-
ters denote matrices. $A^T$, $A^{-1}$, and $A^*$ denote the tran-
spose, inverse, conjugate transpose of $A$, respectively, and diag($a$)
denotes a diagonal matrix with the vector $a$ as its diagonal
elements. $\ell_p$-norm is denoted by $\| \cdot \|_p$. For a matrix $A$, $a_{ij}$
denotes the element at the intersection of $i^{th}$ row and $j^{th}$
column, $a_i$ denotes the $i^{th}$ column and $a^*$ denotes the
$i^{th}$ row of matrix $A$. The Frobenius norm of $A$ is $\|A\|_F = \sqrt{\sum \|a_i\|_2^2}$, row-wise $\ell_{2,1}$ norm is given by $\|A\|_{2,1} = \sum_i \sqrt{\sum_j |a_{ij}|^2}$, and the spectral norm is defined as the square root of the largest eigenvalue $\lambda_{\text{max}}$ of the positive
semi-definite matrix $A^T A$. $|N|$ denotes the set $\{1, 2, \ldots, N\}$, card($\mathcal{P}$) represents cardinality of the set $\mathcal{P}$, $A\mathcal{P}$ is the
sub-matrix of $A$ containing the columns indexed by $\mathcal{P}$. The
support of $x \in \mathbb{R}^N$ is the index set of its non-zero elements
i.e., $\text{supp}(x) = \{j \in [N] : x_j \neq 0\}$.

II. RELATED WORK

Next we discuss some of the key frameworks in the literature
for the design of an ETF and optimization of the correspond-
ing projection matrix. Tropp et al. [22] proposed a general
alternating projection method that is capable of solving a
large class of inverse eigenvalue problems which includes the
construction of tight frames. For constructing a tight frame,
they solve a matrix nearness problem based on the applica-
ration requirements. In [23], the tight frames are constructed
based on mutual coherence. The achieved mutual coherence
is found to be close to the lower bound when the frame redu-
dancy is not very high and can be employed for constructing
frame of any dimension. In this paper, in addition to mutual
cohere nce, we focus on restricting the eigenvalues of the
Gram matrix in order to reduce RIC, which in turn ensures
tightness.

Elad [14] initiated the work on projection matrix opti-
mization based on the averaged mutual coherence mea-
ure using a shrinkage scheme. Since mutual coherence (3)
considers worst-case correlation it is not representative of
the average reconstruction performance and the $t$-averaged
mutual coherence is more appropriate. However, this shrink-
age scheme is computationally intensive and introduces few
large off-diagonal elements in the Gram matrix. Due to this
the worst-case guarantees for the recovery algorithms do not
hold. To address the issue of large off-diagonal elements and
to ensure improved tightness of the frame, we apply a shrink-
age function followed by spectral constraints. Xu et al. [24]
proposed a shrinkage function that projects the Gram matrix
onto a non-empty convex set and reduces the off-diagonal
elements towards the Welch bound.

Authors in [25] proposed a method for designing $D$ by
making a subset of its columns orthogonal, which is equiva-
ient to minimizing the difference between the Gram matrix $G$
and the identity matrix (target ETF). The design of the opti-
mal projection matrix for a given transform basis is formul-
ated as:

$$\min_{\Phi} \| \Psi \Psi^T - \Psi \Psi^T \Phi \Psi \Psi^T \|_F^2 \quad \text{or} \quad \min_{\Phi(\Gamma)} \| \Gamma - \Delta^{\Gamma T} \Gamma \Delta \|_F^2$$

(10)

where $V \Delta V^T$ is the eigen decomposition of $\Psi \Psi^T$ and $\Gamma = \Phi V$. From (10), $V \Delta V^T \Phi^T \Phi V \Delta V^T \approx V \Delta V^T$ and thus
$\Delta V^T \Phi^T \Phi V \Delta \approx \Delta$ with the optimization being over $\Phi(\Gamma)$. A major advantage of this method is that it is non-iterative
and offers significant computational improvement compared
to [14]. Similarly, Hong and Zhu [26] optimize the projection
matrix based on the equivalent Gram matrix with a penalty
term to minimize the sensing matrix energy. In [27], an opti-
mized sparse projection matrix is constructed based on the
target Gram matrix with a sensing matrix energy constraint.
However, this results in only a slight reduction in the recon-
struction error as the projection matrix is obtained directly
from the designed Gram matrix.

In our proposed framework, we optimize the frame matrix
based on the target Gram matrix and followed by the
optimization of the projection matrix using the designed tight frame. To ensure robustness, we incorporate $\ell_{2,1}$-norm to reduce the effect of outliers and a weighted penalty term is used for the sensing matrix energy. Furthermore, Zelnik-Manor et al. [28] proposed an optimized projection matrix design based on block-sparse representation which finds application in block-sparse decoding. The objective function is transformed into a weighted surrogate function given by:

$$\|D^TD - I\|_F^2 = \sum_{j=1}^{B} \sum_{i \neq j} \|D[i]^TD[j]\|_F^2 + \sum_{j=1}^{B} \|D[j]^TD[j] - I\|_F^2$$  \hspace{1cm} (11)

where $D$ is represented as a concatenation of $B$ column blocks $D[j]$.

Authors in [29] focus on directly reducing the mutual coherence of the frame, and the optimized projection matrix is designed by solving a non-smooth optimization problem with convergence guarantee. Similar to this, authors in [30], [31], and [32] focus on the optimization of the sensing matrix based on incoherent UNTF. In [31], gradient-based alternating optimization approach is exploited to iteratively optimize the sensing matrix with decrease in the mutual coherence. Similarly, in [32], the goal of optimization is to minimize the mutual coherence between the projection and the sparsifying matrix. An optimal sensing matrix is designed based on incoherent frame vectors by solving a matrix nearness problem with initialization using a partial Fourier basis. Bai et al. [30] optimize the projection matrix such that the corresponding frame matrix $D$ approximates the target ETF. The method includes the sparse representation error to ensure robustness, however it employs the more sensitive Frobenius norm. Similarly, in [33] and [34], authors optimize the projection matrix based on the Gram matrix using alternating optimization. Sparse representation error penalty is included to ensure robustness of the optimized projections. These methods focus on optimizing the projection matrix based on the equivalent Gram matrix rather than using the ETF. Moreover, unlike the proposed approach, they require the training data matrix as well as the sparse encoding error for the optimization of the projection matrix.

In the proposed approach we optimize the sensing matrix with respect to an ETF which provides good interpretability to the sensing matrix. ETF is designed by transforming the eigenvalues to ensure tightness and incoherence. Further, we incorporate the $\ell_{2,1}$-norm in place of the Frobenius norm and an adaptive weighted penalty in the cost function to ensure robustness to noise and outliers while minimizing the sensing matrix energy.

III. PROPOSED ETF DESIGN METHOD

From the preceding discussion on frame design, we note the need for constructing a tight frame is almost equiquadrangular. We formulate an optimization problem that considers the RIP by targeting a smaller value of the RIC and a projection matrix with lower mutual coherence.

Consider the $\ell_1$-coherence function $\mu_1$, a more general concept of coherence, of a frame $D$ defined as:

$$\mu_1(s) = \max_{i \in [L]} \max_{j \in \mathcal{P}} \left\{ \sum_{j \in \mathcal{P}, j \neq i} |\langle d_i, d_j \rangle|, \text{card}(\mathcal{P}) = s, i \notin \mathcal{P} \right\}$$  \hspace{1cm} (12)

where $\mathcal{P} \subset [L]$ and $1 \leq s \leq L - 1$. It can be shown that $\mu \leq \mu_1(s) \leq s\mu$ [20]. Next consider Theorem 5.3 from [20] stated below.

**Theorem 1:** Let $D$ be a frame with $\ell_2$-normalized columns and assume $s \in [L]$. For all $s$-sparse vectors $\theta$:

$$(1 - \mu_1(s - 1)) \|\theta\|_2^2 \leq \|D\theta\|_2^2 \leq (1 + \mu_1(s - 1)) \|\theta\|_2^2$$  \hspace{1cm} (13)

In other words, for each set $\mathcal{P} \subset [L]$ with $\text{card}(\mathcal{P}) \leq s$, the eigenvalues of the Gram matrix $D^T_D D_P$ lie in the interval $[(1 - \mu_1(s - 1)), (1 + \mu_1(s - 1))]$.

**Proof:** Refer [20] for the proof.

In this paper, the primary objective is to construct a frame matrix $D$ which not only satisfies the equiquadrangular property but also exhibits tightness in terms of the spectral constraint. For designing such a frame, we need to constrain the spectral and structural properties of the Gram matrix. Next, we discuss these constraints for frame design and formulate the iterative optimization problem.

A. STRUCTURAL CONSTRAINT

For reduced pairwise mutual correlation, the off-diagonal elements of the Gram matrix should be close to zero and the diagonal elements should be close to unity. The structural constraint set for an ETF with unit norm vectors and pairwise inner product not larger than $\epsilon$ is given by:

$$\mathcal{H}_\epsilon = \left\{ H \in \mathbb{H}^{L \times L}; H = H^T; h_{ii} = 1; \max_{i \neq j} |h_{ij}| \leq \epsilon \right\}$$  \hspace{1cm} (14)

where $\epsilon \in (0, 1)$ controls the search space for the desired ETF. Projecting onto the convex projection set $\mathcal{H}_\epsilon$ bounds the off-diagonal elements, which in turn implies reduced mutual coherence [14], as below:

$$S_{\mathcal{H}_\epsilon}(G) : g_{ij} = \begin{cases} g_{ij} & |g_{ij}| \leq \epsilon \\ \text{sign}(g_{ij}) \cdot \epsilon & \text{otherwise} \end{cases}$$  \hspace{1cm} (15)

B. SPECTRAL CONSTRAINT

The structural constraint helps in reducing the mutual correlation but such frames are far from being tight. The shrinkage function (15) results in a Gram matrix $G$ that is full-rank. To ensure tightness, we use a spectral constraint to reduce the rank of the Gram matrix to $M$ based on Theorem 1. From (13) and mutual coherence measure (3), the eigenvalues of the target Gram matrix should lie in the range $[(1 - \mu_1(s - 1)), (1 + \mu_1(s - 1))]$, where $\mu_1(s) \leq s\mu$. Thus, we design
an ETF based on the sparsity level of the underlying signal and the desired mutual coherence.

Let the eigen decomposition of $G_{\epsilon}$ be $Q\Lambda Q^*$, where the diagonal elements of $\Lambda$ (i.e., the eigenvalues of $G_{\epsilon}$) are sorted in descending order. For the spectral constraints, in terms of restricting the eigenvalues of the Gram matrix, consider a set of Gram matrices corresponding to the set of tight frames given by:

$$\mathcal{H}_T = \left\{ H \in \mathcal{M}_{L \times L}, H = H^T, \text{ and eigenvalues } \Lambda_T \right\}$$

(16)

$$[\Lambda_T]_{ii} = \begin{cases} \Omega_1 & \text{if } \lambda_{ii} < \Omega_1; \quad i \leq M \\ \Omega_2 & \text{if } \lambda_{ii} \geq \Omega_2; \quad i \leq M \end{cases}$$

(17)

where the eigenvalues $\lambda_{ii}$ of the Gram matrix $G_{\epsilon}$ are sorted in descending order and $\Omega_1 = 1 - \mu_1(s - 1)$ and $\Omega_2 = 1 + \mu_1(s - 1)$. The sorted eigenvalues are thresholded as in (17) resulting in a spectrally constrained eigenvalue matrix $\Lambda_T \in \mathcal{M}_{M \times M}$ with $M$ non-zero eigenvalues.

Let $Q_M$ be the $L \times M$ eigenvector matrix formed from the first $M$ columns of $Q$. The desired Gram matrix is given by $G_T^{*} = (Q_M \sqrt{\Lambda_T})(Q_M \sqrt{\Lambda_T})^*$, and the corresponding frame $D_{ETF}$ is obtained as [22]:

$$D_{ETF} = \sqrt{\Lambda_T}Q_M^*$$

(18)

Thus, $D_{ETF}$ is optimized with respect to the Gram matrix by minimizing the off-diagonal elements of the Gram matrices corresponding to the set of tight frames.

**Algorithm 1: Incoherent Frame Design (IFD)**

Require: $D_0$ as initialized frame matrix, $N_{\text{frame}}$

$t = 0$

while $t \leq N_{\text{frame}}$

  Project the Gram matrix $G_t = D_t^T D_t$ onto the structural constraint set $\mathcal{H}_e$ using (15) to obtain the new Gram matrix $G_{\epsilon}$.

  Compute the eigen decomposition of $G_{\epsilon} = Q\Lambda Q^*$.

  Project $G_{\epsilon}$ onto the spectral constraint set $\mathcal{H}_T$ by thresholding the eigenvalues using (17).

  Update the Gram matrix: $G_T^{*} = (Q_M \sqrt{\Lambda_T})(Q_M \sqrt{\Lambda_T})^*$.

  Compute ETF for the Gram matrix $D_{t+1} = \sqrt{\Lambda_T}Q_M^*$.

  $t = t + 1$

end while

Frame matrix $D_{ETF} = D_{t-1}$

**Remarks:**

- The “shrinkage” process (15) is similar to that in [14] as the structural constraint set $\mathcal{H}_e$ limits the pairwise correlation.
- In this work, the tightness constraint for the frame is characterized by restricting the eigenvalues using the set $\mathcal{H}_T$ rather than being defined by a tightness constant.

such as $(L/M)$ [22], [23] or $\left( \frac{M}{M} \sum_{i=1}^{M} \lambda_{ii} \right)$ [30] which limits the search space for the best solution. We define the search space for the frame such that the corresponding Gram matrix has eigenvalues restricted to the range $[\Omega_1, \Omega_2]$ with each eigenvalue being equal to $\Omega_1$ or $\Omega_2$ as given in (17).

- For constructing an ETF, IFD uses two constraint sets: $\mathcal{H}_e$ for the shrinkage process and $\mathcal{H}_T$ for restricting the eigenvalues. In addition, it is required to construct an ETF with Gram matrix whose diagonal elements are unity with rank constrained properties as given below:

  $\mathcal{H}_{\text{diag}} = \{ H \in \mathcal{M}_{L \times L} : H = H^T, h_{ii} = 1, \forall i = 1, \ldots, L \}$

  $\mathcal{H}_{\text{rank}} = \{ H \in \mathcal{M}_{L \times L} : H = H^T, \text{ rank}(H) = M \}$

- Alternating projections do not increase the distance between consecutive iterates. However, it may not result in a solution which is closest to the constraint sets [22] [35]. The solution w.r.t. these constraint sets results in an ETF, and the non-increasing nature of the cost function $\|D_t^T D_t - G_T^{*}\|_F^2$ is guaranteed ensuring convergence of the IFD algorithm. Thus, IFD will converge with a random initialization $D_0$, but it does not guarantee convergence to a global optimal solution.

- There are extensions of the alternating projection method that have considered non-convex constraint sets such as $\mathcal{H}_{\text{rank}}$ and $\mathcal{H}_T$. In [36], alternating projections on manifolds has been studied and convergence is proved for two smooth manifolds which intersect transversally. In [35], authors considered alternating projections on two non-convex sets and the method converges locally to a point of intersection at a linear rate.

- The constraint sets $\mathcal{H}_e$ and $\mathcal{H}_{\text{diag}}$ are convex, and $\mathcal{H}_{\text{rank}}$ and $\mathcal{H}_T$ are smooth manifolds. Hence, the convergence results for the alternating minimization approach cannot be applied directly here [37]. The convergence proof of alternating projections with more than two constraint sets, some of which are non-convex, is still an open problem. Thus, in Section V, we discuss the convergence of the IFD algorithm mainly based on simulations.

**IV. ROBUST PROJECTION MATRIX DESIGN**

In this section, we present the proposed robust projection matrix design algorithm. Using the IFD algorithm discussed in the previous section we can construct an ETF $D_{ETF}$.

**A. PROBLEM FORMULATION**

The underlying signal or data $X$ is represented in terms of the sparsifying basis $\Psi$ by the sparse representation coefficients $\Theta$ with the sparse encoding error (SEE) given by $E = X - \Psi \Theta$. Depending on the sparsity level and its relationship with mutual coherence (5), optimized projections in [14], [24], [25], [28] are designed to reduce pairwise correlation. Projection matrices designed in this manner have very small correlation with the transform basis and the CS system is able to perform better than with a random sensing matrix,
with \( E = 0 \) for exactly sparse signals. In the real world, signals are not exactly sparse and are only compressible. Thus it is not possible to learn the ground truth dictionary and even for a learned transform basis, a non-zero SEE \( E \) exists. In such scenarios, if \( \Phi \) is not designed appropriately, then the recovery performance can be adversely affected when the projection matrix projects \( E \) onto the measurement domain. Therefore, the SEE must be taken into account while optimizing the projection matrix. A robust CS system, where the objective is to reduce the mutual coherence and the sensing matrix energy, can be designed using the formulation below:

\[
\Gamma(\Phi, D) = \min_{\Phi} \|\Phi \Psi - D\|_{2,1} + \gamma \|\Phi E\|^2_F
\]  

(19)

where \( \gamma \) is the weight parameter for the encoding error and \( \|\Phi \Psi - D\|_{2,1} \) is the fidelity term. In conventional methods, the squared \( \ell_2 \)-norm and Frobenius norm are usually employed which tend to amplify the effect of outliers and noisy data. The \( \ell_2,1 \) norm is more robust than the Frobenius norm and squared \( \ell_2 \)-norm. The \( \ell_2,1 \)-norm promotes row sparsity with a few non-zero rows. Thus, (19) optimizes the projection matrix based on the ETF while being robust to noise and outliers.

However, (19) optimizes the projection matrix based on the SEE \( E \), which in turn can be obtained from a large training data \( X \). This would require large memory storage and increase the computational complexity [26]. For a learned transform basis \( \Psi \) which represents the data sparsely, the SEE energy \( \|E\|_F^2 \) is small. However, when projecting the sparse error onto the measurement domain using the projection matrix \( \Phi \), \( \|\Phi E\|_F \) can become very large if the projection matrix is not designed appropriately. This in turn affects the reconstruction accuracy of the CS system adversely. Based on the norm consistency property, we have \( \|\Phi E\|_F \leq \|\Phi\|_F \|E\|_F \). This implies that if \( \Phi \) is optimized appropriately, then a small \( \|\Phi E\|_F \) would result in a smaller projected error \( \|\Phi E\|_F \) [26]. Moreover, when a large training data set is used to represent the class of underlying signals, energy in the corresponding SEE \( E \) is expected to be spread evenly across the columns. Therefore, the resultant SEE \( E \) can be viewed as an additive white Gaussian noise. Based on the above discussion, we can minimize \( \|\Phi\|_F \) while targeting the SEE and construct an optimized projection matrix using the reformulated objective function given below:

\[
\Gamma(\Phi, D) = \min_{\Phi} \|\Phi \Psi - D\|_{2,1} + \gamma \|\Phi\|^2_F
\]  

(20)

Minimizing \( \|\Phi\|_F \) directly not only makes the algorithm independent of the training data \( X \) and the SEE \( E \) but is applicable for any CS system as long as a learned dictionary \( \Psi \) is available. Further, to allow sensing matrix energy to be adaptively minimized w.r.t. the reconstruction fidelity, we introduce a weight matrix \( W \). The proposed incoherent and robust projection matrix design (IR-PMD) is reformulated as

\[
\Gamma(\Phi, D) = \min_{\Phi} \|\Phi \Psi - D\|_{2,1} + \gamma \|\sqrt{W} \Phi\|^2_F
\]  

(21)

where \( \sqrt{W} \) is a diagonal matrix whose elements are adapted based on \( \Phi \), and are given by:

\[
w_{ii} = \frac{1}{2\|\Phi \Psi - D\|_F^2}
\]  

(22)

Here \( \Phi \Psi - D \) is the \( i \)th row of \( \Phi \Psi - D \). The term \( \|\sqrt{W} \Phi\|^2_F \) minimizes the sensing matrix energy weighted by \( \Phi \Psi - D \) and also aids in obtaining a stable optimal solution. A large error term \( \|\Phi \Psi - D\|_F^2 \) results in a small weight \( w_{ii} \) for the corresponding row of \( \Phi \) and vice versa. Hence, the elements of \( \Phi \) with lower fidelity error have a greater weight or impact on the projection matrix optimization.

### B. OPTIMIZATION

We propose an iterative procedure to construct the ETF \( D \) and optimize the projection matrix \( \Phi \) alternately. At each iteration, ETF is constructed using the projection matrix from the previous iteration via the IFD algorithm and then \( \Phi \) is optimized for the designed ETF. Using the frame \( D \) obtained from the IFD algorithm, the solution to (21) gives an optimized projection matrix. For solving (21), the problem as:

\[
\min_{\Phi} \|\sqrt{W} (\Phi \Psi - D)\|^2_F + \gamma \|\sqrt{W} \Phi\|^2_F
\]  

(23)

where \( W \) is a diagonal matrix defined in (22). When \( \|\Phi \Psi - D\|_F^2 = 0 \), we let \( w_{ii} = \frac{1}{2\|\Phi \Psi - D\|_F^2 + \varepsilon} \), where \( \varepsilon \) is a small constant.

Next, let \( \sqrt{W} \Phi = \overline{\Phi} \in \mathbb{R}^{M \times N} \) and \( \sqrt{W} D = \overline{D} \in \mathbb{R}^{M \times L} \). Then, we can rewrite (23) as:

\[
\min_{\overline{\Phi}} \|\overline{\Phi} \Psi - \overline{D}\|^2_F + \gamma \|\overline{\Phi}\|^2_F
\]  

(24)

To update \( \overline{\Phi} \) and eventually \( \Phi \), we differentiate (24) w.r.t. \( \overline{\Phi} \). The derivative w.r.t. \( \overline{\Phi} \) is given as:

\[
\nabla_{\overline{\Phi}}J(\overline{\Phi}) = 2(\overline{\Phi} \Psi - \overline{D}) \Psi^T + 2\gamma \overline{\Phi}
\]  

(25)

We can update \( \overline{\Phi} \) by setting the derivative to zero to obtain the following closed form expression:

\[
\overline{\Phi} = \overline{D} \Psi^T (\Psi \Psi^T + \gamma I_N)^{-1}
\]  

(26)

where \( I_N \in \mathbb{R}^{N \times N} \) is the identity matrix. Finally, the optimized projection matrix \( \Phi \) can be obtained as:

\[
\Phi = (\sqrt{W})^{-1} \overline{\Phi}
\]  

(27)

The main steps of the incoherent and robust projection matrix design (IR-PMD) algorithm are summarized in Algorithm 2.

**Remarks:**

- Different from the approaches in [26], [27] which optimize \( \Phi \) based on the Gram matrix, IR-PMD algorithm optimizes the projection matrix based on the ETF designed via the IFD algorithm to ensure it is close to the target frame. This approach is more intuitive and computationally efficient.
overall cost for updating \( \Phi \) is \( \mathcal{O}(N^3) \). To update \( \Phi \), cost of matrix inversion of \( \sqrt{\mathbf{W}} \) is \( \mathcal{O}(M^3) \). In [30], the computational cost of updating \( \Phi \) with \( J \) signals in the training data set is dominated by eigen decomposition whose complexity is \( \mathcal{O}(\min\{N(L + J)^2, N^2(L + J)\}) \) which is larger than the cost of the proposed approach.

V. SIMULATION RESULTS

In this section, we present the results of experiments we have carried out on both synthetic and real data sets. The first set of experiments are performed on synthetic data to demonstrate the recovery performance of the various CS systems being compared. We generate a set of training signals based on the learned dictionary [38] and white Gaussian noise is added for understanding robustness of the CS systems. We also illustrate the effectiveness of optimized projections in real-world applications by considering image compression.

In the proposed scheme, IFD algorithm allows us to design an ETF which is used as the target frame for optimizing the projection matrix using the IR-PMD algorithm. We will demonstrate the significant properties of the designed ETF such as the absolute and averaged mutual coherence. Next, we demonstrate the performance of incoherent and robust projections through reconstruction results under different conditions. In each CS system, we consider signals that are sparse with respect to a learned transform basis \( \Psi \). The performance of the proposed method will be compared with the techniques presented in [14], [23], [25], [28]–[30].

A. PERFORMANCE EVALUATION ON SYNTHETIC DATA

We generate an \( N \times L \) transform basis \( \Psi \) with Gaussian distributed entries, and initialize the projection matrix to an \( M \times N \) random matrix \( \Phi_0 \). This initial setup is used in each of the following CS systems for optimizing the projection matrix: CS
dCS [25], CSZMRE [28], CSXPC [24], CSCHZ [29], CSBLH [30], CSHZ [26], and CSHELZ [27]. We use \( \Phi_0 \) for CSrand, and for CSElad [14] we use \( \gamma = 0.95 \) to optimize the projections.

First, we compare the performance of frame design using IFD algorithm, CSrand, CSElad, and CSTK [23]. For training and testing, synthetic data is generated using \( \Psi \). Let \( \{\theta_i\}_{i=1}^{2J} \) be \( s \)-sparse vectors of dimension \( L \times 1 \) where each vector is normally distributed with zero mean and unit variance. The signals \( \{x_i\}_{i=1}^{2J} \) are generated as:

\[
x_i = \Psi \theta_i + n_i = x_i^* + n_i, \quad i = 1, 2, \ldots, 2J
\]

where \( n_i \) is zero-mean white Gaussian noise. The corresponding signal-to-noise ratio is denoted by SNR. Let \( X = [x_1, x_2, \ldots, x_{2J}]^T \) be the noisy signal and \( X^* = [x_1^*, x_2^*, \ldots, x_{2J}^*]^T \) be the original noise-free signal. The training and test data sets are obtained by dividing \( X \) into two equal parts i.e., \( X^{\text{test}} \) and \( X^{\text{train}} \) each of size \( N \times J \). Measurement matrix \( Y = \Phi X^{\text{test}} \) is computed using the optimized projection matrix \( \Phi \) obtained via the CS algorithms. In each case, OMP algorithm is used to recover the sparse signals.
from the measurement vectors by solving (4). The normalized mean square reconstruction error (NMSE) is computed as:

$$e_r = \frac{1}{N} \sum_{j=1}^{N} \frac{\|x_j^{\text{test}} - \hat{x}_j^{\text{test}}\|^2_2}{\|x_j^{\text{test}}\|^2_2}$$ (29)

where $\hat{x}_j^{\text{test}}$ is the recovered sparse signal. The mean square error is computed as:

$$\text{MSE} \triangleq \frac{1}{N \times J} \|X^{\text{test}} - \hat{X}^{\text{test}}\|^2_F$$ (30)

$\hat{x}_j^{\text{test}}$ can be expressed in terms of the corresponding sparse coefficients as $\hat{x}_j = \Psi \hat{\theta}_j$ and the sparse coefficients $\hat{\theta}_j$ can be obtained by solving the OMP problem [10]. The peak signal to noise ratio (PSNR) is defined as [38]:

$$\text{PSNR} \triangleq 10 \times \log_{10} \left[ \frac{(2^r - 1)^2}{\text{MSE}} \right]$$ (31)

where $r = 8$ is the number of bits per pixel.

1) CONVERGENCE

For performance evaluation of the IFD algorithm, experiments are carried out with $M = 20$, $L = 120$, and sparsity $s = 4$. We start with an arbitrary $M \times L$ matrix $D$ with full rank, and the shrinkage function followed by the spectral constraint are applied iteratively resulting in a tight frame that is closest to being an incoherent matrix. We consider convergence of the alternating projections in the IFD algorithm based on the numerical results. The distance measure $d_F(j)$ is used to study convergence of the iterative algorithm:

$$d_F(j) = \|G_j - \tilde{G}_T\|^2_F$$ (32)

where $\tilde{G}_T$ is the target optimal Gram matrix that belongs to the sets satisfying structural and spectral constraints and $G_j$ is the designed Gram matrix at the $j$th iteration. Fig. 1(a) shows the evolution of $d_{FS}$ for the IFD algorithm. It is seen that the algorithm eventually results in a frame that lies at the intersection of the constraint sets. Convergence of the $d_{FS}$ indicates that the frame is close to being an incoherent UNTF. For demonstrating convergence of the IR-PMD algorithm we assume $M = 20$, $L = 120$, $N = 60$, and $N_{SE} = 100$ with each experiment being repeated 1000 times. Fig. 1(b) shows the convergence of the objective function $\Gamma(\Phi_j, D_j)$ (21) for different values of $\gamma$ over 10000 iterations. Fig. 2(a) shows the evolution of MSE for different values of $\gamma$. Although, the objective function was seen to converge faster for smaller value of $\gamma$, MSE attains the lowest value for $\gamma = 0.1$. Note that this is the noiseless case and signals are perfectly sparse for the experiment with synthetic data, thus smaller values of $\gamma$ are able to achieve reduced pairwise correlation and improved reconstruction accuracy. Fig. 2(b) shows the evolution of the iterate $\Phi_j$, measured using the expression:

$$\delta_\Phi(t) = \frac{1}{MN} \|\Phi_j - \Phi_{j-1}\|_F$$ (33)

It is seen that $\Phi_j$ converges well for different values of $\gamma$. Next we run a similar experiment using a real data set, namely the Caltech 101 data set. Fig. 3(a) shows the evolution of MSE for different values of $\gamma$. For a real data set, sparse encoding error can be significant as the underlying signal is not exactly sparse. Hence, larger value of $\gamma$ minimizes the sensing matrix energy and leads to reduced MSE. In Fig. 3(b),
both MSE and peak signal-to-noise ratio (PSNR) are shown as a function of $\gamma$. It is noted that $\gamma > 0.2$, results in reduced MSE and increased PSNR. It is observed that for real images with no additional noise the MSE decreases for larger values of $\gamma$ and for exactly sparse synthetic signals $\gamma = 0.1$ gives improved performance. However, when the synthetic signals have a non-zero sparse encoding error, higher values of $\gamma$ minimize the sensing matrix energy with increased PSNR and reduced MSE. Fig. 4 shows the evolution of the averaged mutual coherence for different CS techniques with $\text{SNR} = 10$ dB and $s = 4$. It shows that IFD results in averaged mutual coherence that is closer to the Welch bound compared to the other algorithms. Fig. 5(a) shows the averaged mutual coherence $\mu_t$ (with $t = \epsilon$) for the different frame design methods as a function of the redundancy of the frame. As the number of measurements $M$ increases, all the methods achieve lower mutual coherence. The superiority of IFD algorithm is clearly evident for smaller values of $M$. The reduced mutual coherence achieved by IFD results in a tighter frame, which is then used in the IR-PMD algorithm. Similarly, for projection matrix design approaches, Fig. 5(b) shows the averaged mutual coherence $\mu_t$ as a function of the number of measurements. It is seen that the IR-PMD algorithm is superior compared to the other approaches. In Figs. 6(a) and 6(b), evolution of the averaged mutual coherence $\mu_t$ is shown for different values of $\gamma$ for synthetic and Caltech data sets, respectively. It is seen that smaller value of $\gamma$ results in reduced mutual coherence as $\mu_t$ will be determined by the term $\|\Phi\Psi - \hat{D}\|^2_F$ with $\|\Phi\|^2_F$ being given less weight.

2) RECONSTRUCTION PERFORMANCE

Next we compare the different CS techniques for optimizing the projection matrix in terms of their ability to reconstruct the original signal from under-sampled measurements. We consider a set of synthetic signals generated as earlier. The performance of the algorithms is evaluated in terms of mutual coherence and reconstruction error. The maximum number of iterations is set to 100 for each algorithm except for the non-iterative algorithm $\text{CS}_{\text{ZMRE}}$ which employs the optimized matrix from $\text{CS}_{\text{DCS}}$ as the initial matrix. The parameter values for the different algorithms are selected based on the values mentioned in the corresponding references. If the optimal values were not mentioned, we carried out the parameter tuning to determine the optimal values.

For $\text{CS}_{\text{rand}}$, a normally distributed random sensing matrix is employed and this random matrix is used as the initial sensing matrix for all algorithms except $\text{CS}_{\text{ZMRE}}$ which employs the optimized matrix from $\text{CS}_{\text{DCS}}$ as the initial matrix. The parameter values for the different algorithms are selected based on the values mentioned in the corresponding references. If the optimal values were not mentioned, we carried out the parameter tuning to determine the optimal values.

FIGURE 4. Evolution of averaged mutual coherence with the number of iterations ($M = 20, L = 120, s = 4$, and $\text{SNR} = 10$ dB).

FIGURE 5. Averaged mutual coherence as a function of $M$ for (a) frame design and (b) projection matrix design ($L = 120, \text{SNR} = 10$ dB).

FIGURE 6. Evolution of averaged mutual coherence with the number of iterations ($L = 120, \text{SNR} = 10$ dB).
The MSE over a range of parameter values is computed for the methods and the parameter value resulting in minimum MSE is selected as the optimal value. The parameter values selected are as follows: (i) CS\textsubscript{Elad}: $t = 0.4$, $\gamma = 0.95$, (ii) CS\textsubscript{ZMRE}: $\alpha = 0.99$, (iii) CS\textsubscript{BLH}: $\alpha = 0.8$ [30], (iv) CS\textsubscript{XPC}: $\alpha = 0.7$ [24], (v) CS\textsubscript{CHZ}: $\rho = 0.5$, $\eta = 1.2$, $\alpha = 0.99$, $\beta = 2$, $\text{Iter}_{\text{in}} = 100$ (number of inner iterations), and $\text{Iter}_{\text{out}} = 15$ (number of outer iterations) [29], (vi) CS\textsubscript{HZ}: $\lambda = 0.1$ for synthetic data and $\lambda = 0.9$ for real data sets [26], and (vii) CS\textsubscript{HLZL}: $\lambda = 0.25$, $\kappa = 20$ for synthetic data and $\lambda = 0.5$, $\kappa = 30$ for real data sets [27]. Fig. 7 shows the histogram of the absolute off-diagonal elements of the normalized Gram matrix corresponding to the optimized projection matrix. The proposed IR-PMD algorithm, compared to the other algorithms, results in a histogram that is centered more towards the origin which implies smaller local correlations and thus improved recovery performance.

![FIGURE 7. Histogram of the absolute off-diagonal elements of the optimized Gram matrix ($N = 60$, $L = 120$, and $M = 20$).](image)

Fig. 8 and Fig. 9, show the NMSE ($e_r$) as a function of the sparsity $s$ for frame design and projection matrix design algorithms, respectively. It is seen that the proposed IFD and IR-PMD algorithms result in reduced NMSE compared to the other algorithms. IR-PMD algorithm performs better than the mutual coherence-based methods such as [14], [25], [28], [29]. This indicates that incoherent projections are suitable for exactly sparse signals, however for signals which contain significant amount of noise it is also important to minimize the mutual coherence energy.

![FIGURE 8. Reconstruction error (NMSE, $e_r$) as a function of sparsity s for frame design algorithms (SNR = 10 dB).](image)

![FIGURE 9. Reconstruction error (NMSE, $e_r$) as a function of sparsity s for projection matrix design algorithms (SNR = 10 dB).](image)

In Tables 1 and 2, for each CS technique, we study the performance of the optimized sensing matrix in terms of $\mu$, $\mu_1$, $\|I - G\|^2_F$, and MSE under noiseless and noisy scenarios. For the noiseless case, projection matrix with reduced mutual coherence leads to improved signal reconstruction, however this is not true for the noisy scenario. Even though mutual coherence using IR-PMD algorithm is not significantly reduced, the use of $\ell_{2,1}$-norm results in improved signal reconstruction at low SNR as shown in Table 3. The improved MSE is due to the robustness of $\ell_{2,1}$-norm to outliers. However, for the noiseless case, CS\textsubscript{BLH} has reduced MSE than IR-PMD due to the Frobenius norm used to minimize the mutual correlation. This shows that mutual coherence is not the best measure for estimating signal recovery performance when the signals contain significant amount of noise. For achieving a robust CS system, we consider the encoding error in order to reduce the effect of noise on the signal recovery process.

| TABLE 1. Coherence measures and MSE for various CS systems ($M = 20$, $N = 60$, $L = 120$). |
| --- | --- | --- | --- |
| $\mu$ | $\mu_1$ | $\|I - G\|^2_F$ | MSE |
| CS\textsubscript{Elad} | 0.3179 | 0.7007 | 772.54 | 0.0054 |
| CS\textsubscript{DCS} | 0.3089 | 0.6916 | 602.67 | 0.0026 |
| CS\textsubscript{ZMRE} | 0.3088 | 0.6504 | 603.23 | 0.0029 |
| CS\textsubscript{XPC} | 0.2722 | 0.5601 | 645.05 | 0.0029 |
| CS\textsubscript{CHZ} | 0.3250 | 0.5196 | 754.60 | 0.0052 |
| CS\textsubscript{BLH} | 0.2631 | 0.5846 | 600.33 | 0.0031 |
| CS\textsubscript{HZ} | 0.2791 | 0.5775 | 612.95 | 0.0021 |
| CS\textsubscript{HLZL} | 0.3104 | 0.5205 | 627.35 | 0.0037 |
| CS\textsubscript{IR-PMD} | 0.2650 | 0.4663 | 600.29 | 0.0019 |

| TABLE 2. Coherence measures and MSE for various CS systems ($M = 20$, $N = 60$, $L = 120$). |
| --- | --- | --- | --- |
| SNR = 10 dB |
| $\mu$ | $\mu_1$ | $\|I - G\|^2_F$ | MSE |
| CS\textsubscript{Elad} | 0.3204 | 0.7879 | 794.02 | 0.0042 |
| CS\textsubscript{DCS} | 0.3114 | 0.6758 | 602.75 | 0.0025 |
| CS\textsubscript{ZMRE} | 0.3098 | 0.6857 | 603.99 | 0.0028 |
| CS\textsubscript{XPC} | 0.2761 | 0.6447 | 662.94 | 0.0026 |
| CS\textsubscript{CHZ} | 0.3259 | 0.5147 | 768.65 | 0.0048 |
| CS\textsubscript{BLH} | 0.3098 | 0.6777 | 615.51 | 0.0025 |
| CS\textsubscript{HZ} | 0.2915 | 0.4735 | 620.82 | 0.0028 |
| CS\textsubscript{HLZL} | 0.3133 | 0.7285 | 652.89 | 0.0038 |
| CS\textsubscript{IR-PMD} | 0.2652 | 0.4603 | 600.25 | 0.0015 |
TABLE 3. MSE for various CS systems ($M = 25, N = 60, L = 80$).

|                | $s = 4$  | $s = 8$  | $s = 4$  | $s = 8$  |
|----------------|----------|----------|----------|----------|
| CS_PNL         | 0.0012   | 0.0220   | 0.0013   | 0.0227   |
| CS_PCS         | 2.74e-4  | 0.0130   | 1.87e-4  | 0.0142   |
| CS_SPARSE      | 1.98e-4  | 0.0128   | 2.59e-4  | 0.0138   |
| CS_RLHI        | 6.44e-4  | 0.0211   | 0.0010   | 0.0226   |
| CS_RLHI        | 2.33e-6  | 0.0086   | 4.42e-4  | 0.0234   |
| CS_RLHI        | 4.83e-6  | 0.0084   | 1.63e-4  | 0.0269   |
| CS_RLHI       | 2.77e-4  | 0.0124   | 2.33e-4  | 0.0135   |
| CS_IR-PMD      | 3.33e-6  | 0.0079   | 2.64e-5  | 0.0081   |

In Fig. 10, we study the recovery performance of CS systems as a function of the number of measurements $M$ with $s = 4$, and for two different SNRs. Figs. 9 and 10 clearly demonstrate the superior recovery performance of the proposed IR-PMD algorithm. In Fig. 11, we show the recovery performance as a function of the number of dictionary atoms $L$. For higher SNR, the performance of the mutual coherence based methods and the proposed algorithm are comparable. The $\ell_2,1$-norm based IR-PMD algorithm is found to be more robust against the encoding error and therefore is well suited for applications such as image compression where the sparse encoding error is significant and cannot be ignored.

**B. OPTIMIZED PROJECTIONS FOR IMAGE COMPRESSION**

To demonstrate the practical applicability of the proposed method we perform experiments on standard real images. We investigate the applicability of the proposed CS system for image compression. The objective is to recover the original images from their compressed versions via optimized projections. For sparse representation of the images, we learn the transform basis $\Psi$ using KSVD [38]. It is difficult to learn the ground truth dictionary for a set of real images and there will always be a non-zero encoding error even without any additional noise. Similar to the experiments with synthetic data, we set the parameters as $s = 4, M = 20$, and $L = 100$. We evaluate the recovery performance of the proposed method via two experiments involving standard images and Caltech 101 data set [39].

1) **STANDARD IMAGES**

In this experiment, we use a collection of 40 well-known standard images, including cameraman, Elaine, Lena, and pirate. From these images, we obtain 60000 patches each of size $8 \times 8$ pixels. Of these 50000 patches are used for training and the remaining 10000 are used for testing the recovery performance of the optimized projections and its generalization capability. Each patch is represented as a $64 \times 1$ vector with the set of patches expressed as a matrix of size $64 \times 60000$. We evaluate the recovery performance on the test data by adding white Gaussian noise with the resulting SNR ranging from 5 dB to 45 dB. The recovery performance for image compression is typically measured in terms of the peak signal-to-noise ratio (PSNR) given in (31) and MSE given in (30). The MSE performance of different CS systems is shown in Fig. 12. It is seen that at low SNR even random projections perform better than some of the mutual coherence-based approaches. The proposed IR-PMD algorithm has superior performance and is able to reduce the effect of noise. Fig. 13 shows the PSNR performance as
FIGURE 12. MSE as a function of SNR for optimized projections ($N = 64, L = 100, and M = 20$).

FIGURE 13. PSNR as a function of SNR for optimized projections ($N = 64, L = 100, and M = 20$).

FIGURE 14. PSNR as a function of SNR for optimized projections ($N = 64, L = 100, and M = 20$) for Caltech 101 data set.

FIGURE 15. MSE as a function of SNR for optimized projections ($N = 64, L = 100, and M = 20$) for Caltech 101 data set.

a function of the SNR. Note that IR-PMD outperforms the approaches in [26], [27], [30] which shows the impact of the $\ell_{2,1}$-norm and adaptive weight penalty in achieving improved robustness.

2) CALTECH DATA SET

In this experiment, we use the Caltech 101 data set [39] which contains a total of 9146 images with 101 different object categories and each object category has 40-800 images. As in the previous experiment, these images are converted into training and testing patches. We obtain 60000 patches each of size $8 \times 8$ pixels with 50000 patches used for training and the remaining 10000 are used for testing the recovery performance. The training data is used to train the sparsifying basis. For experimenting with different values of SNR, 1000 samples are randomly selected from $X_{\text{train}}$ for each run. Then, white Gaussian noise is added to each subset of the training samples with SNR varying from 5 dB to 45 dB. The test data set $X_{\text{test}}$ is divided into 10 testing subsets each with 1000 samples which are used to evaluate recovery performance of the different CS systems.

In Fig. 14, the output PSNR is shown for SNR ranging from 5 dB to 30 dB. The results show that CSIR-PMD outperforms the other techniques. The performance of the mutual coherence-based methods is poor even compared to CS$_{\text{rand}}$ when testing on real images with additional noise. The proposed approach outperforms the robust projection matrix design methods such as [26], [27], [30] and is robust even in low SNR conditions as the sensing matrix energy is minimized adaptively based on the reconstruction fidelity. In Fig. 15, MSE is shown as a function of the SNR and CSIR-PMD achieves the best performance among the techniques being compared.

Fig. 16 compares the performance of CS$_{\text{HLZL}}$ and CSIR-PMD algorithms using a projection matrix $\Phi$ trained at SNR = 5 dB. Additional noise is not added to these images as real images already contain noise resulting in a sparse encoding error. As seen in the inset zoom-in views, the edges are much better recovered using CSIR-PMD. The proposed method achieves the best qualitative results in these tests. Using the optimized projections designed by the IR-PMD algorithm, PSNR of the reconstructed images is higher than that of the images reconstructed using the other CS systems. Experimental results on synthetic data, real standard images, and the Caltech 101 data set have demonstrated the effectiveness of the proposed method.

Finally, in order to demonstrate the effectiveness of the $\ell_{2,1}$-norm regularization, we compare the performance of CSIR-PMD with CSWF which employs Frobenius norm and optimizes the objective $\|\Phi \Psi - D_{\text{ETF}}\|_F^2 + \gamma \|\Phi E\|_F^2$. 

VOLUME 9, 2021 

131473
VI. CONCLUSION

In this paper, we have proposed an algorithm for the design of incoherent and robust projection matrix. We have demonstrated that the $\ell_{2,1}$-norm leads to a more robust formulation while effectively minimizing the sensing matrix energy. In practice, the equivalent dictionary should be close to being a tight frame in order to achieve good recovery performance. In this work, we proposed a novel ETF design method based on the structural and spectral constraints on the eigenvalue characteristics. The resulting tight frame is used to design optimized projections. In addition to mutual coherence and tightness constraints, the weighted sensing matrix energy term acts as a penalty to avoid amplification of the sparse encoding error in the measurement domain. A closed form solution set for an optimal projection matrix with weighted

Fig. 17 shows that $\text{CS}_{\text{IR-PMD}}$ based on the $\ell_{2,1}$-norm consistently outperforms the model based on the Frobenius norm.
penalty has been derived. The superiority of the proposed algorithm has been demonstrated with extensive experiments on synthetic data and real images.

REFERENCES

[1] E. J. Candès, “Compressive sampling,” in Proc. Int. Cong. Math., vol. 3, Jan. 2006, pp. 1433–1452.

[2] J. Haupt and R. Nowak, “Compressive sampling vs. conventional imaging,” in Proc. Int. Conf. Image Process., Oct. 2006, pp. 1269–1272.

[3] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] E. J. Candès and T. Tao, “Decoding by linear programming,” IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[5] J. Romberg, “Imaging via compressive sampling,” IEEE Signal Process. Mag., vol. 25, no. 2, pp. 14–20, Mar. 2008.

[6] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[7] R. Gribonval and M. Nielsen, “Sparse representations in unions of bases,” IEEE Trans. Inf. Theory, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[8] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ1 minimization,” Proc. Nat. Acad. Sci. USA, vol. 100, no. 5, pp. 2197–2202, 2003.

[9] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” IEEE Signal Process. Mag., vol. 25, no. 2, pp. 21–30, Mar. 2008.

[10] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[11] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” IEEE Trans. Signal Process., vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[12] J. Wang, S. Kwon, and B. Shim, “Generalized orthogonal matching pursuit,” IEEE Trans. Signal Process., vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[13] G. Chen, D. Li, and J. Zhang, “Iterative gradient projection algorithm for two-dimensional compressive sensing sparse image reconstruction,” Signal Process., vol. 104, pp. 15–26, Nov. 2014.

[14] M. Elad, “Optimized projections for compressed sensing,” IEEE Trans. Signal Process., vol. 55, no. 12, pp. 5695–5702, Dec. 2007.

[15] L. Welch, “Lower bounds on the maximum cross correlation of signals (corresp.),” IEEE Trans. Inf. Theory, vol. IT-20, no. 3, pp. 379–399, May 1974.

[16] E. J. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” IEEE Trans. Inf. Theory, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[17] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” IEEE Trans. Inf. Theory, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[18] L. H. Chang and J.-Y. Wu, “An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 60, no. 9, pp. 5702–5715, Sep. 2014.

[19] S. Waldron, “Generalized Welch bound equality sequences are tight frames,” IEEE Trans. Inf. Theory, vol. 49, no. 9, pp. 2307–2309, Sep. 2003.

[20] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing. Basel, Switzerland: Birkhäuser, 2013.

[21] T. Strohmer and R. W. Heath, Jr., “Grassmannian frames with applications to coding and communication,” Appl. Comput. Harmon. Anal., vol. 14, no. 3, pp. 257–275, May 2003.

[22] J. A. Tropp, I. S. Dhillon, R. W. Heath, Jr., and T. Strohmer, “Designing structured tight frames via an alternating projection method,” IEEE Trans. Inf. Theory, vol. 51, no. 1, pp. 188–209, Jan. 2005.

[23] E. V. Tsiligianni, L. P. Kondi, and A. K. Katsaggelos, “Construction of incoherent unit norm tight frames with consideration of sparse representation error,” IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 5389–5400, Dec. 2015.

[24] A. S. Lewis, D. R. Luke, and J. Malick, “Local linear convergence for alternating and averaged nonconvex projections,” Found. Comput. Math., vol. 9, no. 4, pp. 485–513, Aug. 2009.

[25] A. S. Lewis and J. Malick, “Alternating projections on manifolds,” Math. Oper. Res., vol. 33, no. 1, pp. 216–234, Feb. 2008.

[26] H. H. Bauschke and J. M. Borwein, “On projection algorithms for solving convex feasibility problems,” SIAM Rev., vol. 38, no. 3, pp. 367–426, Sep. 1996.

[27] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[28] L. Fei-Fei, R. Fergus, and P. Perona, “Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories,” in Proc. Conf. Comput. Vis. Pattern Recognit. Workshop, Jun. 2004, p. 178.

T. Hong, H. Bai, S. Li, and Z. Zhu, “An efficient algorithm for designing projection matrix in compressive sensing based on alternating optimization,” Signal Process., vol. 125, pp. 9–20, Aug. 2016.

G. Li, X. Li, S. Li, H. Bai, Q. Jiang, and X. He, “Designing robust sensing matrix for image compression,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 5389–5400, Dec. 2015.

[29] C. Lu, H. Li, and Z. Lin, “Optimized projections for compressed sensing via direct mutual coherence minimization,” Signal Process., vol. 151, pp. 45–55, Oct. 2018.