MDPI

Article

Measurement Matrix Optimization for Compressed Sensing System with Constructed Dictionary via Takenaka–Malmquist Functions

Qiangrong Xu [1], Zhichao Sheng [1], Yong Fang [1],* and Liming Zhang [2]

1 Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China; hahah@shu.edu.cn (Q.X.); zcsheng@shu.edu.cn (Z.S.)
2 Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau 999078, China; lmzhang@um.edu.mo
* Correspondence: yfang@shu.edu.cn

Abstract: Compressed sensing (CS) has been proposed to improve the efficiency of signal processing by simultaneously sampling and compressing the signal of interest under the assumption that the signal is sparse in a certain domain. This paper aims to improve the CS system performance by constructing a novel sparsifying dictionary and optimizing the measurement matrix. Owing to the adaptability and robustness of the Takenaka–Malmquist (TM) functions in system identification, the use of it as the basis function of a sparsifying dictionary makes the represented signal exhibit a sparser structure than the existing sparsifying dictionaries. To reduce the mutual coherence between the dictionary and the measurement matrix, an equiangular tight frame (ETF) based iterative minimization algorithm is proposed. In our approach, we modify the singular values without changing the properties of the corresponding Gram matrix of the sensing matrix to enhance the independence between the column vectors of the Gram matrix. Simulation results demonstrate the promising performance of the proposed algorithm as well as the superiority of the CS system, designed with the constructed sparsifying dictionary and the optimized measurement matrix, over existing ones in terms of signal recovery accuracy.

Keywords: compressed sensing (CS); sparse representation; the Takenaka–Malmquist (TM) functions; mutual coherence; equiangular tight frame (ETF)

1. Introduction

Sparse representation has become a powerful tool, as it can efficiently model signals to facilitate compressing and processing [1–3]. Due to the redundancy in the majority of signals in nature, transforming the original signal to a sparse or compressed version through a certain domain is meaningful for reducing costs in transportation and storage. In this case, the signal can be expressed as a linear combination of a few given basis functions, which indicates that most of the representation's coefficients are zero or close to zero.

Compressed sensing (CS), as one of the most attractive emerging fields in recent years, has shown promising results in many applications, such as channel or spectrum estimation [4,5], cognitive radio communications [6], color pattern detection [7,8], and people-centric sensing [9]. CS states that signals which have sparse structures on an appropriate domain can be effectively recovered from the underdetermined linear projections. However, it is often the case that signals of interest are not exactly sparse but compressible. Projecting the signal into its proper sparse domain to obtain a sparser structure is essential in CS [10]. Hence, one critical issue in sparse representation is how to choose the dictionary (as well as the transform basis) to reduce the number of measurements required for reconstruction [11]. In most cases, the original signal is sparse in a different transform

domain, e.g., discrete cosine transform (DCT), discrete wavelet transform (DWT), and so on. Despite being simple and having fast computations, these nonadaptive dictionaries are unable to sparsely represent a given signal of interest well [12]. The approximate solution after the sparse representation of signals still have many nonzero elements. In fact, if an N-dimensional signal admits K-sparse representation in a dictionary, one can reconstruct exactly the original signal with large probability with $M = O(Klog(N/K))$ measurements [10,13]. This shows that only a small number of measurements are necessary to accurately recover the original signal when the signal exhibits a sparser representation. In our previous work [14], we presented an effective sparse representation approach for wireless channels based on the modified Takenaka–Malmquist (TM) basis, under which the represented wireless channel exhibits a sparser structure than the existing sparsifying dictionaries. As the results shown, the approach facilitates to improve the performance of CS. In this paper, we construct a sparse dictionary using the TM basis functions. Compared with other sparse bases, the signal of interest can be represented more accurately with only a limited number of TM basis functions.

The recent growing trend involves replacing the conventional random sensing matrices by optimized ones in order to enhance CS performance further [15–19]. Some conditions have been put forth to evaluate the qualities of a sensing matrix to guarantee exact signal recovery. Candès et al. introduced the restricted isometry property (RIP) and proved that if a sensing matrix satisfies RIP, the original signal can be recovered with high probability [13]. However, it is difficult to verify whether a given sensing matrix satisfies RIP. Another relatively simple property is the mutual coherence [15–20]. The coherence of matrix A means the maximum absolute correlation between different columns of A. Gaussian and Bernoulli matrices are widely used due to their entries being generated by an independent identically distributed (i.i.d.) process and satisfying the RIP. However, it shows low recovery accuracy in the CS system when using the class of matrices as a sensing matrix. In order to improve the CS performance, researchers attempt to optimize the sensing matrix with the hope of increasing the reconstruction probability and taking a fewer number of measurements [15–18]. Elad proposed an optimized algorithm that iteratively minimizes the t-averaged mutual coherence using shrinkage operation by singular value decomposition (SVD) [15]. However, Elad's method is time-consuming and creates some large values of coherence. Duarte-Carvajalino and Sapiro proposed a framework for a joint sensing matrix and dictionary design and optimization [16]. The method addressed the problem by making the Gram matrix as close as possible to an identity matrix. Their proposed approach is noniterative and does not have a clear meaning after multiple approximation procedures. In this paper, we proposed a method based on the equiangular tight frame (ETF) design. The optimization objective is to make the Gram matrix of the sensing matrix as close as possible to an ETF. For clarity, the threefold main contributions of this paper are summarized as follows:

- It is revealed that the existing dictionaries are unable to sparsely represent a given signal well. The TM functions have adaptability and robustness in system identification, which invoked us to use it as a basis to build a new dictionary for the sparse representation of a given signal. The simulation results illustrate that compared to other dictionaries, using the constructed dictionary to represent a given signal can obtain a lower mean square error (MSE) under the same number of atoms.
- In order to improve the CS performance further, an ETF-based iterative minimization algorithm is proposed for optimizing the measurement matrix. It is shown that the approach leads to a lower coherence of the equivalent dictionary and significantly improves the signal recovery accuracy than the methods in [15,16].
- The reconstruction quality obtained with our constructed dictionary and the optimal measurement matrix by our approach is superior to that obtained with the existing dictionary and random measurement matrix (such as wavelets or DCT). Moreover, simulation experiments show that after the random measurement matrix optimized by our proposed algorithm, the reconstruction success rate is also improved.

The rest of the paper is organized as follows. In Section 2, we review the basic principles of the CS framework. Section 3 introduces the TM basis function and describes how to construct a dictionary for sparse representation. An ETF-based iterative minimization method is developed in Section 4 for optimizing the measurement matrix. Experimental results are provided in Section 5. Section 6 concludes the paper.

2. Background and Preliminaries

CS states that it can enable exact signal reconstruction from far fewer samples than required by the classical Nyquist–Shannon sampling theorem if the signal admits a sparse representation in a certain domain. Consider a signal $\mathbf{x} \in \mathbb{R}^N$, which is sparse over the transformation basis matrix $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$. Let s be the sparse representation of $\mathbf{x}$ over the basis matrix $\mathbf{\Psi}$. Accordingly, $\mathbf{x}$ can be described as

$$\mathbf{x} = \mathbf{\Psi}s, \tag{1}$$

where the signal $\mathbf{x}$ is said to be K-sparse in the $\mathbf{\Psi}$ domain if $\|s\|_0 = K$ ($K \ll N$). The ℓ_0-norm used here denotes the number of the nonzero elements in s.

The basic model of CS is as follows:

$$\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x}, \tag{2}$$

where the resulting signal vector $\boldsymbol{y} \in \mathbb{R}^{M \times 1}$ is called the measurement vector, $\mathbf{x} \in \mathbb{R}^{N \times 1}$ is the given signal of interest, and $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ (with $M \times N$) is called the measurement matrix. Since the original signal $\mathbf{x}$ is sparse in the $\mathbf{\Psi}$ domain, we can rewrite the measurement model as

$$\boldsymbol{y} = \mathbf{\Phi}\mathbf{\Psi}s = As, \tag{3}$$

where $A = \mathbf{\Phi}\mathbf{\Psi} \in \mathbb{R}^{M \times L}$ ($L \geq N \geq M$) represents the sensing matrix, $\mathbf{\Psi} \in \mathbb{R}^{N \times L}$ is the sparsifying basis, e.g., an orthonormal or overcomplete dictionary.

To recover the high-dimensional original signal from the low-dimensional measurement data, it is necessary to solve the following minimization problem

$$\min_{s}\|s\|_0 \; subject \; to \; \boldsymbol{y} = As. \tag{4}$$

In general, the reconstruction problem in (4) is a nondeterministic polynomial hard (NP-hard) problem. Fortunately, research in [21] and [22] has demonstrated that ℓ_1-norm minimization is similar to sparse solutions with ℓ_0-norm minimization. Therefore, the ℓ_1-norm is used to substitute for the ℓ_0-norm in (4). The solution to the above problem is the same as the one to the ℓ_1-norm minimization below

$$\min_{s}\|s\|_1 \; subject \; to \; \boldsymbol{y} = As \tag{5}$$

where (5) can be solved efficiently using convex relaxation algorithm such as basis pursuit (BP) [23]. Another strategy to address the problem in (4) is to use greedy algorithms, such as the orthogonal matching pursuit (OMP), to find a local optimal solution through multiple iterations [24].

The premise of CS is to ensure that the original signal is sparse or can be represented sparsely. The well-known RIP provides a sufficient condition for an exact or approximate recovery of a sparse representation s from the low-dimensional measurements $\boldsymbol{y}$. An M × L sensing matrix A is said to satisfy the RIP of order K if there exists a restricted isometry constant δ_K ($0 < \delta_K < 1$), such that

$$(1 - \delta_K)\|s\|_2^2 \leq \|s\|_2^2 \leq (1 + \delta_K)\|s\|_2^2, \tag{6}$$

where $\|s\|_0 = K$.

RIP measures the degree to which each submatrix consisting of the most *K* columns of *A* is close to being an isometry. However, the RIP is too strict and difficult to be used for proving whether a sensing matrix *A* satisfies the condition or not. Therefore, the most known constructions of matrices satisfying RIP are random matrices, including the Gaussian matrix, Bernoulli matrix, and almost other matrices with i.i.d. entries. Another relatively simple and feasible property for evaluating the quality of sensing matrices is the mutual coherence. It requires that the rows of measurement matrix $\mathbf{\Phi}$ cannot sparsely represent the columns of the sparse basis $\mathbf{\Psi}$ and vice versa. The mutual coherence $\mu(A)$ is defined as the largest absolute normalized inner product between different columns in sensing matrix *A*. Formally, that is expressed as

$$\mu(A) = \max_{1 \le i,j \le L,\ i \ne j} \frac{\left| a_i^T a_j \right|}{\|a_i\|_2 \cdot \|a_j\|_2}, \tag{7}$$

where a_i is the *i*-th column of *A*.

The mutual coherence can intuitively reflect the quality of a sensing matrix and play important role in the signal reconstruction performance. A smaller value of $\mu(A)$ means that the measurement matrix $\mathbf{\Phi}$ and the sparsifying dictionary $\mathbf{\Psi}$ are less relevant. It means that the two matrices are orthogonal if $\mu(A)$ equals to zero.

An alternative way to measure the mutual coherence is considering the corresponding Gram matrix $G = \widetilde{A}^T \widetilde{A}$, in which $\widetilde{A}$ is obtained by normalizing each column of *A*. The maximum value of the off-diagonal elements in *G* denotes the mutual coherence, that is

$$\mu(A) = \max_{1 \le i,j \le L,\ i \ne j} \left| g_{ij} \right|. \tag{8}$$

Generally, a *K*-sparse signal *s* can be recovered from (3) by (4) or (5) if the following inequality holds [25–27]:

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu(A)} \right). \tag{9}$$

Even though in many situations a small $\mu(A)$ is desired, the following well-known result of Welch indicates that $\mu(A)$ is bounded below:

$$\mu(A) \ge \mu_E = \sqrt{\frac{L - M}{M(L - 1)}}. \tag{10}$$

It should be noted that the Welch bound μ_E is approximately equal to $\frac{1}{\sqrt{M}}$ when $L \ge M$.

3. Construction Procedure of the Dictionary

In this section, we briefly introduce and analyze the Takenaka–Malmquist basis functions. In addition, we show how to construct a new dictionary using the TM basis functions for sparse representation in CS. The preliminary results show the representation advantages of this new approach.

3.1. The Takenaka–Malmquist Functions

The TM basis functions were first widely used in the field of control engineering. Extensive research has examined how the TM basis is suitable for linear time-invariant (LTI) system identification [14,28–30]. For a deterministic signal, we can learn from the idea of system identification and represent it using the TM basis. The TM functions are defined as follows:

$$F_k(z) = \frac{\sqrt{1 - |b_k|^2}}{z - b_k} \prod_{i=1}^{k-1} \frac{1 - \bar{b}_i z}{z - b_i}, \quad k = 1, 2, \cdots \tag{11}$$

where $b_i \in \mathbb{C}$, $|b_i| < 1$, $\mathbb{C}$ represents an open unit circle in the complex plane, and $\overline{b}_i$ denotes the conjugation of b_i.

The TM basis functions are constructed to be orthonormal by using the Gram–Schmidt procedure. According to the Cauchy integral formula, it is easy to prove the orthogonality between TM basis functions by

$$\langle F_k, F_1 \rangle = F_k\left(1/\overline{b}_1\right)\sqrt{1 - |b_1|^2} = 0, \forall k > 1 \tag{12}$$

which forces $F_k(z)$ to have a zero at $1/\overline{b}_1$.

The orthogonality between the TM basis functions allowed it to serve as a dictionary, from which the representation coefficients of a signal could be efficiently selected. Hence, only a limited number of linear combinations of TM basis functions are required to approximate the original signal. This provides us with a theoretical basis and feasibility to build a dictionary next.

3.2. Construction of the Dictionary

The number of nonzero terms K of the sparse representation coefficient s of a signal $\mathbf{x}$ in a sparsifying dictionary $\mathbf{\Psi}$ is a very important factor. The smaller the value of K, the fewer the number of measurements are required for signal recovery. Therefore, the sparsifying dictionary $\mathbf{\Psi}$ should be properly selected so that the signal has a sparser representation structure. In other words, K should be reduced as close to zero as possible. The most commonly used approach is using basis function to construct a sparsifying dictionary, such as DWT, DCT, and so forth. For a given dictionary, a signal is approximately represented as a linear combination of the dictionary atoms:

$$\mathbf{x} \approx \sum_{k=1}^{n} s_k \psi_k. \tag{13}$$

The approximate solution after the sparse representation of signals in a DCT (or DWT) domain still have many nonzero elements. Under the same number of atoms, the approximate representation of the signal has a large error. To further improve the sparse representation performance, we construct a sparsity dictionary by using the TM basis functions. The detailed construction process is as follows.

For a given signal $\mathbf{x} \in \mathbb{R}^N$, N poles are uniformly selected in the unit circle, and then the poles are set as the poles of the TM basis functions, forming an initial sparsifying dictionary $\mathbf{\Psi} = \left[F_1(z), \cdots, F_{\xi}(z), \cdots, F_N(z)\right]$. Then, the TM basis function corresponding to each pole takes $\omega_k = \frac{2\pi(k-1)}{N}, k = 1, 2, \cdots, N$ as the sampling frequency to sample the signal $\mathbf{x}$ at equal intervals. In this way, the column vector corresponding to each TM basis function is set by

$$F_{\xi} = \begin{bmatrix} F_{\xi}\left(e^{j0}\right) \\ \vdots \\ F_{\xi}\left(e^{j\omega_k}\right) \\ \vdots \\ F_{\xi}\left(e^{j\frac{2\pi(N-1)}{N}}\right) \end{bmatrix}. \tag{14}$$

Therefore, the constructed dictionary can be expressed as follows

$$\mathbf{\Psi} = \begin{bmatrix} F_1\left(e^{j0}\right) & \cdots & F_{\xi}\left(e^{j0}\right) & \cdots & F_N\left(e^{j0}\right) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ F_1\left(e^{j\omega_k}\right) & \cdots & F_{\xi}\left(e^{j\omega_k}\right) & \cdots & F_N\left(e^{j\omega_k}\right) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ F_1;\left(e^{j\frac{2\pi(N-1)}{N}}\right) & \cdots & F_{\xi}\left(e^{j\frac{2\pi(N-1)}{N}}\right) & \cdots & F_N\left(e^{j\frac{2\pi(N-1)}{N}}\right) \end{bmatrix}. \tag{15}$$

To illustrate the performance of the constructed dictionary, we provide a comparison of the results in Figure 1.
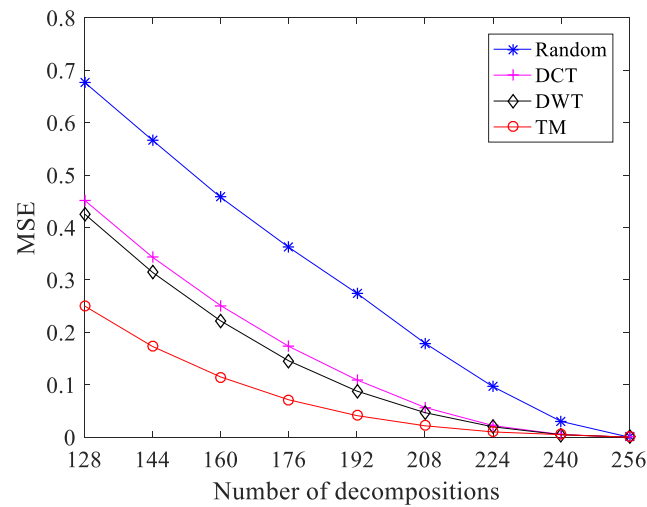

Figure 1. Comparison between the four sparsifying dictionaries under the different number of decompositions using orthogonal matching pursuit (OMP).

As a performance measurement between the original signal and the reconstructed signal, the MSE is defined as follows:

$$MSE = \frac{1}{N} \sum_{K=1}^{N} \|\mathbf{x} - \overline{\mathbf{x}}\|_2^2. \tag{16}$$

Without loss of generality, a Rayleigh-distributed signal is taken here as an example. The length of the signal is 256. A random dictionary of size 256 × 256 is considered, which is drawn from i.i.d. zero mean and unit variance Gaussian distribution. The DCT, DWT, and the constructed TM dictionary of the same dimension are compared. The OMP algorithm is used to compare the MSE of the reconstructed signal after the sparse representation of the original signal under different numbers of decomposed atoms. The results demonstrate that compared with the random, DCT, and DWT dictionaries, using the constructed TM dictionary to represent a given signal can get lower MSE under the same number of decompositions.

4. Optimizing the Measurement Matrix

In this section, we investigate the problem of optimizing the measurement matrix given a sparsifying dictionary to increase CS performance further. To address the problem, an ETF-based iterative minimization algorithm is proposed.

The object is to optimize the measurement matrix **Φ** to minimize the mutual coherence $\mu(\mathbf{A})$. In other words, it is to find a measurement matrix **Φ** making the corresponding Gram matrix as close to an identity matrix as possible. The column-normalized form of the sensing matrix *A* is called an ETF. However, it is often that such an ETF does not exist for any arbitrary dimension [25,31]. Therefore, our optimization goal is to make the sensing matrix *A* as close as possible to an ETF. The minimization problem is formulated as follows:

$$\min_{\mathbf{\Phi}, \mathbf{H} \epsilon \Lambda_{set}} \|\mathbf{A}^T \mathbf{A} - \mathbf{H}\|_F^2. \tag{17}$$

where $\|\cdot\|_F$ is the Frobenius norm. The Λ_{set} denotes a convex set that leads the optimization problem to have solutions. For any arbitrary dimension, Λ_{set} is regularized as

$$\Lambda_{set} = \left\{ \mathbf{H} \in \mathbb{R}^{L \times L} : \mathbf{H} = \mathbf{H}^T, diag\, \mathbf{H} = 1, \max_{i \neq j} |h_{ij}| \leq \mu_E \right\}. \tag{18}$$

As mentioned before, μ_E is the Welch bound and is used here as a prescribed parameter.

The alternate optimization method is usually used to address the minimization problem in (17). The difference between our proposed method and others is that we assign weight values to the Gram matrix of the previous iteration as well as the Gram matrix of the current iteration in the stage of updating the Gram matrix. We modify the singular value without changing the properties of the Gram matrix, so that the minimum singular value of the updated Gram matrix increases and the maximum singular value decreases, and then the linear independence between the column vectors of the Gram matrix is enhanced. In this way, a new diagonal matrix is obtained by using the eigenvalue averaging method for the updated Gram matrix to construct an optimized measurement matrix. The detailed optimization procedure is shown in Algorithm 1.

In this algorithm, the Gram matrix of the normalized sensing matrix is computed. In order to make it close to an ETF, the Gram matrix is projected onto the convex set Λset by the shrinking operation function:

$$g_{ij} = \begin{cases} 1 & i = j \\ g_{ij} & i \neq j, |g_{ij}| \leq \mu_E \\ sign(g_{ij}) \cdot \mu_E & i \neq j, |g_{ij}| \mu_E \end{cases}. \tag{19}$$

Since the Gram matrix is projected onto Λset, its structure changes significantly. In order to retain some of its feature information and approximate the ETF, we give weights to the Gram matrix before and after the update:

$$\mathbf{G}_{(t+1)} = \beta \mathbf{G}_{(t)} + (1 - \beta)\mathbf{G}_{(t-1)}. \tag{20}$$

In general, the aforementioned shrinking operation causes the rank of the Gram matrix to be full. Therefore, the next step is to force a rank M by a singular value decomposition of the obtained Gram matrix, i.e., $\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{G}$, in which $rank(\mathbf{\Lambda}) = M \ll L$. Let $\lambda_1, \lambda_2, \cdots, \lambda_M$ be the eigenvalues of the diagonal matrix $\mathbf{\Lambda}$. According to singular value decomposition, the smaller the maximum singular value, the better the incoherence of the matrix. Therefore, in order to reduce the mutual coherence between the different columns of A, we modify the eigenvalues by averaging all eigenvalues of the obtained Gram matrix and get a new diagonal matrix $\widetilde{\mathbf{\Lambda}}$. Let $\widetilde{\lambda}_i (i = 1, 2, \cdots, M)$ be the diagonal elements of $\widetilde{\mathbf{\Lambda}}$, and $\widetilde{\lambda}_i = \frac{1}{M}\sum_{i=1}^{M} \lambda_i$. Then, we build the squared root of the obtained Gram matrix, $\mathbf{G} = \mathbf{S}^T\mathbf{S}$, where $\mathbf{S} = \sqrt{\widetilde{\mathbf{\Lambda}}}\mathbf{V}^T$. Now, the solution of optimizing the measurement matrix is equivalent to solving

$$\min_{\mathbf{\Phi}} \|\mathbf{S} - \mathbf{\Phi}\mathbf{\Psi}\|_F^2. \tag{21}$$

The solution process is as follows:

$$\begin{aligned} \|\mathbf{S} - \mathbf{\Phi}\mathbf{\Psi}\|_F^2 &= tr\left[(\mathbf{S} - \mathbf{\Phi}\mathbf{\Psi})^T(\mathbf{S} - \mathbf{\Phi}\mathbf{\Psi})\right] \\ &= tr\left(\mathbf{S}^T\mathbf{S}\right) - 2tr\left(\mathbf{\Psi}^T\mathbf{\Phi}^T\mathbf{S}\right) + tr\left(\mathbf{\Psi}^T\mathbf{\Phi}^T\mathbf{\Phi}\mathbf{\Psi}\right). \end{aligned} \tag{22}$$

Here, $tr(\cdot)$ denotes the matrix trace operation.

$$\begin{aligned} \frac{\partial \|\mathbf{S} - \mathbf{\Phi}\mathbf{\Psi}\|_F^2}{\partial \mathbf{\Phi}} &= -2\frac{\partial tr\left(\mathbf{\Psi}^T\mathbf{\Phi}^T\mathbf{S}\right)}{\partial \mathbf{\Phi}} + \frac{\partial tr\left(\mathbf{\Psi}^T\mathbf{\Phi}^T\mathbf{\Phi}\mathbf{\Psi}\right)}{\partial \mathbf{\Phi}} \\ &= -2\frac{\partial tr\left(\mathbf{\Phi}^T\mathbf{S}\mathbf{\Psi}^T\right)}{\partial \mathbf{\Phi}} + \frac{\partial tr\left(\mathbf{\Phi}\mathbf{\Psi}\mathbf{\Psi}^T\mathbf{\Phi}^T\right)}{\partial \mathbf{\Phi}} \\ &= -2\mathbf{S}\mathbf{\Psi}^T + 2\mathbf{\Phi}\mathbf{\Psi}\mathbf{\Psi}^T. \end{aligned} \tag{23}$$

Solved by least squares, it is now easy to find the optimized solutions $\mathbf{\Phi} = \sqrt{\widetilde{\mathbf{\Lambda}}}\mathbf{V}^T\mathbf{\Psi}^\dagger$, where † denotes pseudoinverse.

Algorithm 1: The proposed optimization algorithm.

Input: The given sparsifying dictionary $\mathbf{\Psi} \in \mathbb{R}^{N \times L}$, the number of iterations *Iter*, and threshold μ_E.

Output: The measurement matrix $\mathbf{\Phi}$.

1: **Initialization:** Set $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ to be a random matrix.

2: **for** $l = 1 : Iter$ **do**

3:　　Normalize the columns of the sensing matrix $A = \mathbf{\Phi}\mathbf{\Psi}$ to obtain $\tilde{A}$.

4:　　Compute the Gram matrix $G = \tilde{A}_\mathbf{T}\tilde{A}$.

5:　　Update the Gram matrix by the shrinking operation (19).

6:　　Assign weight values between the previous and current Gram matrix to update the Gram matrix: $\mathbf{G}_{(t+1)} = \beta\mathbf{G}_{(t)} + (1-\beta)\mathbf{G}_{(t-1)}$.

7:　　Apply SVD $\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{G}$ and force the rank of $\mathbf{G}$ to be equal to M.

8:　　Average the eigenvalues to get new diagonal matrix $\widetilde{\mathbf{\Lambda}}$.

9:　　Update the measurement matrix $\mathbf{\Phi} = \sqrt{\widetilde{\mathbf{\Lambda}}}\mathbf{V}^T\mathbf{\Psi}^\dagger$.

10: **end for**

---

5. Experimental Results

In this section, we carry out various numerical simulations to support the effectiveness of the proposed approach. The experiments include a performance evaluation of the proposed algorithm and a performance comparison of different sparsifying dictionaries. In Section 5.1, we present some experiments to illustrate the performance of the proposed algorithm and compare it with other algorithms. In Section 5.2 we demonstrate the superiority of the CS system designed with the constructed sparsifying dictionary over the existing ones under the same measurement matrix. In addition, the signal recovery quality are evaluated to compare the performance of the different sparsifying dictionaries under the optimized measurement matrix using our proposed approach.

5.1. Performance Evaluation of the Proposed Algorithm

In this subsection, the performance of the proposed algorithm in the CS system is evaluated. Several comparison simulations are given to verify the effectiveness of the proposed algorithm in terms of the measures $\mu(A)$ and the signal recovery performance. Here, a random dictionary (every entry is drawn from i.i.d. zero mean and unit variance Gaussian distribution) of size 256×400 was used. We generated a Gaussian matrix with a size of 64×256 as a random measurement matrix and used it as an initial condition to obtain the corresponding optimized measurement matrix by using the proposed algorithm and the algorithm in [15,16], respectively.

The parameter settings in the abovementioned algorithms are shown as follows. For Elad's algorithm, the parameters t, γ are set to 0.2 and 0.55, respectively. The iteration number *Iter* is set to 30 for both Elad's and our algorithm. It is worth noting that the Duarte-Carvajalino and Sapiro algorithm is noniterative.

Since the corresponding Gram matrices with smaller nonzero terms result in good reconstruction performance, the local cumulative coherence values can be used to evaluate the performance of the sensing matrix. The histogram of the absolute off-diagonal entries of the corresponding normalized Gram matrix to each sensing matrix is depicted in Figure 2. As can be seen, both the method proposed in [15,16] and our proposed method reduce the number of large absolute values in the Gram matrix. Compared with Elad's method and Sapiro's method, it easy to find in Figure 2 that the largest absolute values are smaller. Furthermore, the histogram of our method shows that the local cumulative mutual coherence values shift significantly to the left and concentrate within 0.2, which means that the local coherence is also smaller.
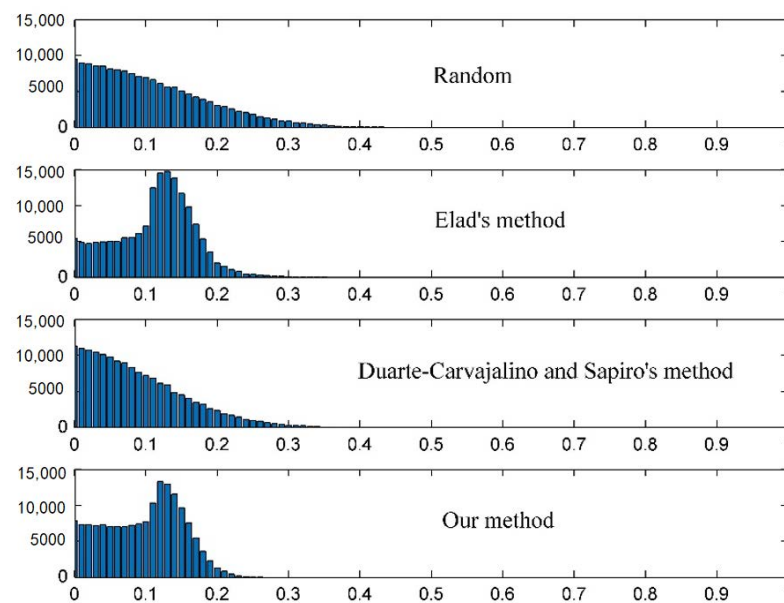
Figure 2. Histogram of the absolute off-diagonal entries of the corresponding normalized Gram matrix to each of the four sensing matrices.

To illustrate the behavior of the proposed algorithm, the evolution of the mutual coherence versus iteration number is shown in Figure 3. Since Sapiro's method is noniterative, its optimized value is shown as a point in the graph. Both Elad's method and our proposed method demonstrate the mutual coherence value decreases with the number of iterations. However, our method shows that the rate of decline is significantly faster. Moreover, the final value of $\mu(A)$ is also smaller than the first two.
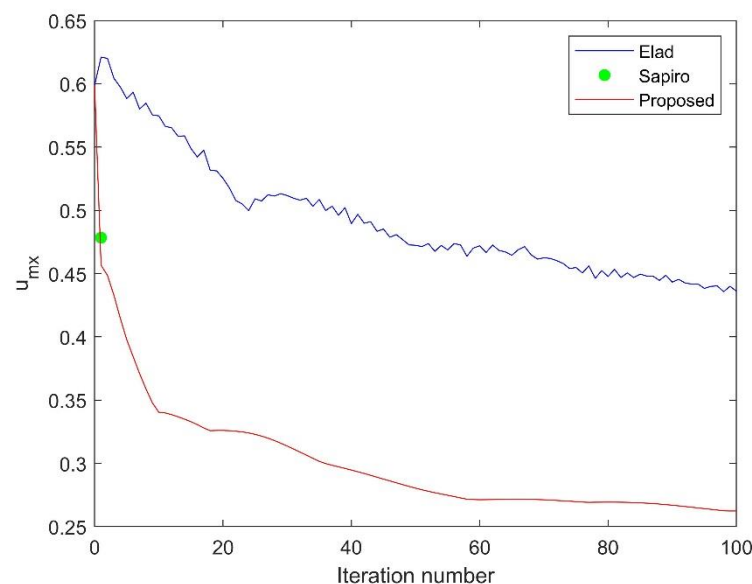
Figure 3. Evolution of the mutual coherence versus iteration number.

Next, the signal reconstruction performance via the abovementioned measurement matrices are further investigated. We use the percentage of successful recovery to evaluate the sparse signal recovery quality. It is considered to be a successful recovery if the reconstruction error $\|x - \hat{x}\|_2 / \|x\|_2 < 10^{-3}$. The known OMP algorithm is used for signal recovery.

In the first experiment, we generate 1000 K-sparse vectors with nonzero entries chosen at random for each sparsity level K. These sparse vectors are used as test signals to evaluate

the CS performance. Figure 4 shows the signal reconstruction performance of the CS system before and after the measurement matrix optimization using different optimization methods, with signal sparsity level *K* varying from 4 to 28 and *M* = 64. As can be seen, all the optimization methods lead to improved CS performance. In addition, the result shown in Figure 4 indicates that the proposed method performs better than both Elad's method and Duarte-Carvajalino and Sapiro's method.
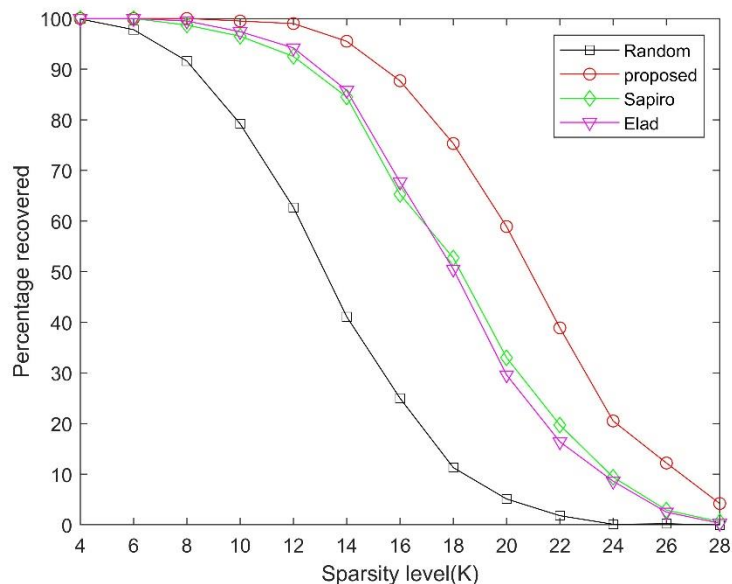

Figure 4. Comparison of the reconstruction performances using various measurement matrices under the same measurements *M* = 64.

In the second experiment, the sparsity level *K* is fixed, and the measurement value *M* is changeable. Figure 5 shows the comparison of the reconstruction performances using various measurement matrices under the same sparsity level *K* = 10. As the number of measurements increases, the percentage of signal reconstruction also increases. As expected, it is shown in Figure 5 once again that the proposed method outperforms the other two methods.
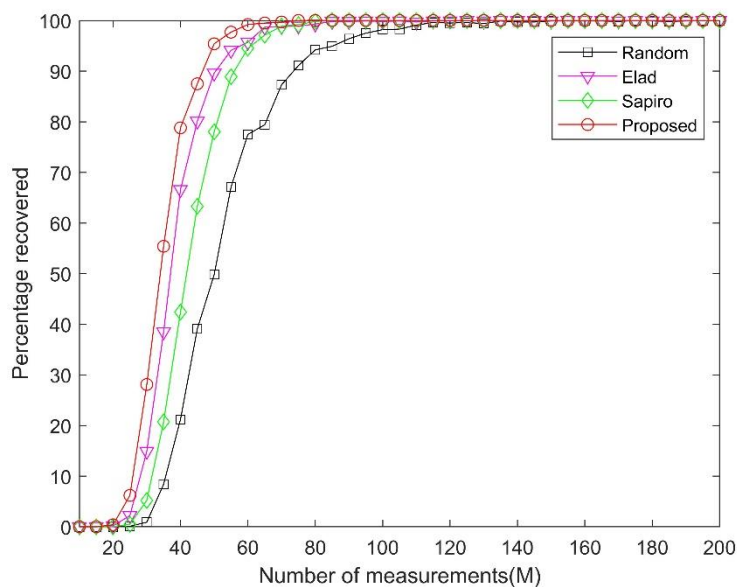

Figure 5. Comparison of the reconstruction performances using various measurement matrices under the same sparsity level *K* = 10.

5.2. Performance Comparison of Different Sparsifying Dictionaries

In this subsection, we investigate the performance of different sparsifying dictionaries in the CS system. As in the previous section, we quantify the performance by measuring the percentage of successful recovery signals. We previously verified that compared to DCT, DWT, and random dictionaries, using the constructed TM dictionary to represent a given signal can get lower MSE under the same number of atoms. Here, we further evaluate the reconstruction performance in CS systems using the constructed TM, DCT, DWT, and random matrices as sparsifying dictionaries.

We set up a new experiment, in which the dimension of the sparsifying dictionaries is set to 256×256. In addition, we generate a measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{64 \times 256}$ with random Gaussian distribution. By multiplying it with the above sparsifying dictionaries, we obtain four sensing matrices. Then, the OMP algorithm is used to recover the signal. In the experiments, we tested 1000 sparse signals in each trial. We evaluated the performance of a sparsifying dictionary in the CS system by the percentage of successful signal recovery. The error threshold between the original signal and the reconstructed signal is the same as in the previous subsection, which is 10^{-3}. Figure 6 shows the comparison of the reconstruction performances using different sparsifying dictionaries under the same random measurement matrix. The results demonstrate that the constructed TM dictionary has a better performance in terms of reconstruction quality compared with the random, DCT, and DWT dictionaries. Using TM as a sparsifying dictionary can achieve a higher reconstruction rate among these dictionaries under the same sparsity level.
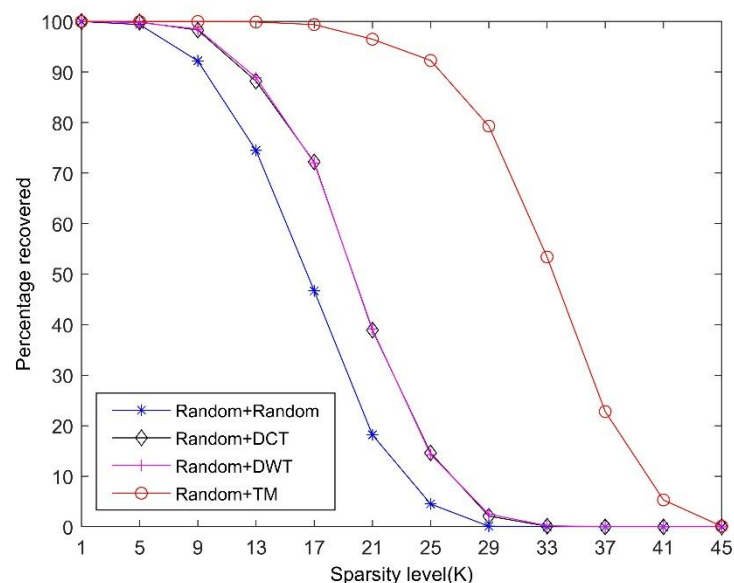

Figure 6. Comparison of the reconstruction performances using different sparsifying dictionaries under the same random measurement matrix ($M = 64$).

In addition, we fixed the sparsity level of the signal and performed comparison experiments by changing the measurement numbers of the measurement matrix. The sparsity level K is set to 25. Figure 7 reveals that by using the TM dictionary as $\mathbf{\Psi}$, one can use a fewer number of measurements to achieve the same performance as that obtained using the DCT, DWT, and random dictionaries.
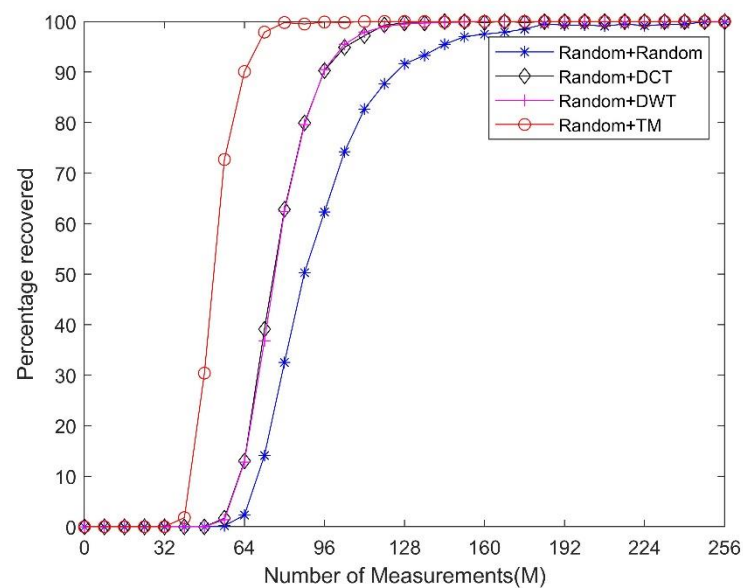
Figure 7. Comparison of the reconstruction performances using different sparsifying dictionaries under the same sparsity level *K* = 25 and the same random measurement matrix.

To further evaluate the performance of the constructed TM sparsifying dictionary, we used the proposed algorithm to optimize the measurement matrix in order to reduce the mutual coherence between the dictionary and the measurement matrix. Different from previous experiments, we compared the performance of different sparsifying dictionaries in CS systems under an optimized measurement matrix. As shown in Figures 8 and 9, after using the proposed algorithm to optimize the measurement matrix, the performance of the signal recovery is improved. It was observed that using the initial random measurement matrix and the constructed TM dictionary in the CS system can allow us to still obtain a better performance than using the optimized measurement matrices with the random, DCT, and DWT dictionaries. This proves that our constructed TM dictionary is more robust and effective.
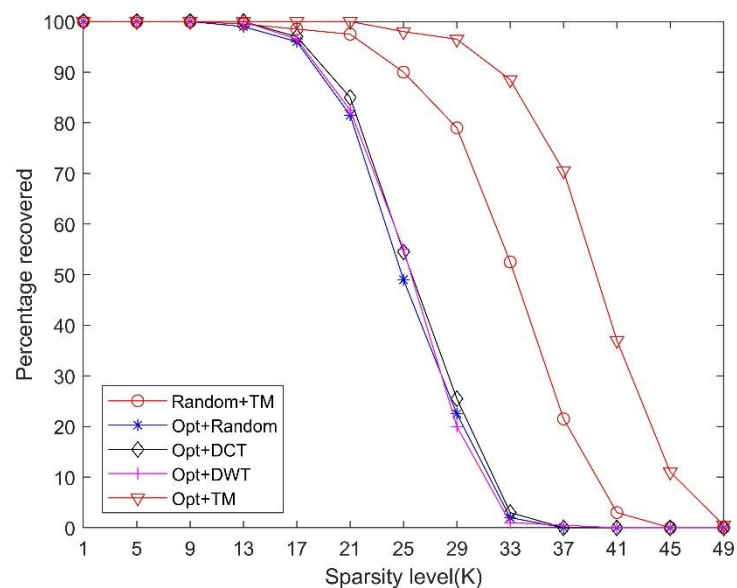

Figure 8. Comparison of the reconstruction performances using different sparsifying dictionaries with the optimized measurement matrix (*M* = 64).
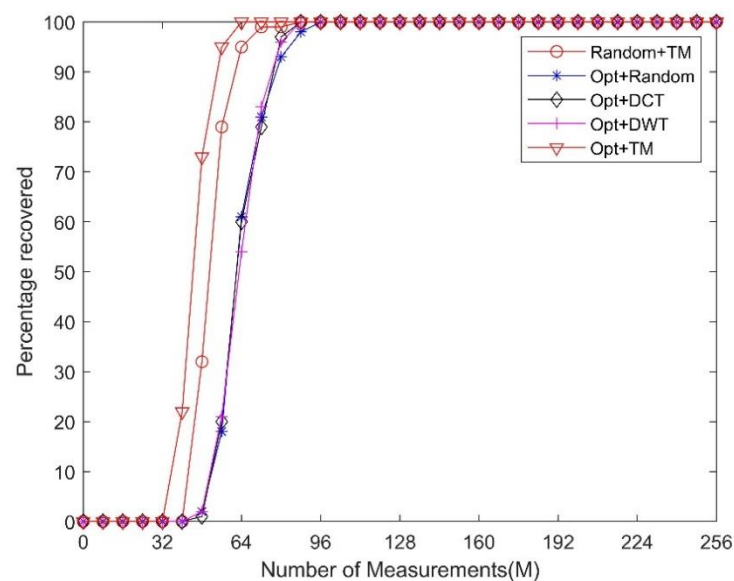
Figure 9. Comparison of the reconstruction performances using different sparsifying dictionaries under the same sparsity level $K = 25$ and the optimized measurement matrix.

6. Conclusions

In this paper, we constructed a novel sparsifying dictionary using the TM functions for the sparse representation of a given signal. The advantages are higher performance in sparsely representing signals of interest at the same degree of atomic decomposition. Our proposed iterative algorithm is based on the equiangular tight frame, for which its related mutual coherence between the measurement matrix and the sparsifying dictionary is reduced. Simulation results demonstrated the promising performance of the algorithm and the superiority of the CS system, designed with the constructed dictionary and the optimized measurement matrix, over the existing ones in terms of signal recovery accuracy.

In the future, we will try to combine the constructed sparsifying dictionary with the measurement matrix to optimize both simultaneously. We will also try to apply our method in practical applications, such as channel estimation. Further investigation of this topic is needed.

Author Contributions: Conceptualization, Q.X. and Y.F.; methodology, Q.X. and Y.F.; validation, Q.X.; writing—original draft preparation, Q.X.; writing—review and editing, Q.X., Y.F., Z.S., and L.Z.; All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Elad, M. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*; Springer: New York, NY, USA, 2010.
2. Qin, Z.; Fan, J.; Liu, Y.; Gao, Y.; Li, G.Y. Sparse Representation for Wireless Communications: A Compressive Sensing Approach. *IEEE Signal Process. Mag.* **2018**, *35*, 40–58. [CrossRef]
3. Xiao, D.; Liang, J.; Ma, Q.; Xiang, Y.; Zhang, Y. High capacity data hiding in encrypted image based on compressive sensing for nonequivalent resources. *Comput. Mater. Contin.* **2019**, *58*, 1–13. [CrossRef]
4. Akbarpour-Kasgari, A.; Ardebilipour, M. Massive MIMO-OFDM Channel Estimation via Distributed Compressed Sensing. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 376–379. [CrossRef]
5. Liu, S.; Zhang, Y.D.; Shan, T.; Tao, R. Structure-Aware Bayesian Compressive Sensing for Frequency-Hopping Spectrum Estimation with Missing Observations. *IEEE Trans. Signal Process.* **2018**, *66*, 2153–2166. [CrossRef]

6.   Sharma, S.K.; Lagunas, E.; Chatzinotas, S.; Ottersten, B. Application of Compressive Sensing in Cognitive Radio Communications: A Survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1838–1860. [CrossRef]

7.   Shi, W.; Jiang, F.; Liu, S.; Zhao, D. Image Compressed Sensing Using Convolutional Neural Network. *IEEE Trans. Image Process.* **2020**, *29*, 375–388. [CrossRef]

8.   Rousseau, S.; Helbert, D. Compressive Color Pattern Detection Using Partial Orthogonal Circulant Sensing Matrix. *IEEE Trans. Image Process.* **2020**, *29*, 670–678. [CrossRef]

9.   Xu, Y.; Li, S.; Zhang, Y. Privacy-Aware Service Subscription in People-Centric Sensing: A Combinatorial Auction Approach. *Comput. Mater. Contin.* **2019**, *61*, 129–139. [CrossRef]

10.   Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [CrossRef]

11.   Rubinstein, R.; Bruckstein, A.M.; Elad, M. Dictionaries for Sparse Representation Modeling. *Proc. IEEE* **2010**, *98*, 1045–1057. [CrossRef]

12.   Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322. [CrossRef]

13.   Candes, E.J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **2006**, *52*, 489–509. [CrossRef]

14.   Lei, Y.; Fang, Y.; Zhang, L.; Weighted, A. K-SVD-Based Double Sparse Representations Approach for Wireless Channels Using the Modified Takenaka-Malmquist Basis. *IEEE Access* **2018**, *6*, 54331–54342. [CrossRef]

15.   Elad, M. Optimized Projections for Compressed Sensing. *IEEE Trans Signal Process.* **2007**, *55*, 5695–5702. [CrossRef]

16.   Duarte-Carvajalino, J.M.; Sapiro, G. Learning to Sense Sparse Signals: Simultaneous Sensing Matrix and Sparsifying Dictionary Optimization. *IEEE Trans. Image Process.* **2009**, *18*, 1395–1408. [CrossRef] [PubMed]

17.   Li, G.; Zhu, Z.; Yang, D.; Chang, L.; Bai, H. On Projection Matrix Optimization for Compressive Sensing Systems. *IEEE Trans. Signal Process.* **2013**, *43*, 129–159. [CrossRef]

18.   Hong, T.; Li, X.; Zhu, Z.; Li, Q. Optimized structured sparse sensing matrices for compressive sensing. *Signal Process.* **2019**, *159*, 119–129. [CrossRef]

19.   Li, B.; Zhang, L.; Kirubarajan, T. Projection matrix design using prior information in compressive sensing. *Signal Process.* **2017**, *135*, 36–47. [CrossRef]

20.   Naidu, R.R.; Murthy, C.R. Construction of Binary Sensing Matrices Using Extremal Set Theory. *IEEE Signal Process. Lett.* **2017**, *24*, 211–215. [CrossRef]

21.   Candes, E.J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* **2005**, *51*, 4203–4215. [CrossRef]

22.   Donoho, D.L. For most large underdetermined systems of linear equations the ℓ1-norm solution is also the sparsest solution. *Commun. Pure Appl. Math* **2006**, *51*, 797–829. [CrossRef]

23.   Chen, S.S.; Donoho, D.L.; Saunderso, M.A. Atomic decomposition by basis pursuit. *SIAM Rev.* **2001**, *43*, 129–159. [CrossRef]

24.   Pati, Y.C.; Rezaiifar, R.; Krishnaprasad, P.S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1–3 November 1993; 1993; 1, pp. 40–44.

25.   Tropp, J.A.; Dhillon, I.S.; Heath, R.W.; Strohmer, T. Designing structured tight frames via an alternating projection method. *IEEE Trans. Inf. Theory* **2005**, *51*, 188–209. [CrossRef]

26.   Gribonval, R.; Nielsen, M. Sparse representations in unions of bases. *IEEE Trans. Inf. Theory* **2003**, *49*, 3320–3325. [CrossRef]

27.   Tropp, J.A. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory* **2004**, *50*, 2231–2242. [CrossRef]

28.   Mi, W.; Qian, T. Frequency-domain identification: An algorithm based on an adaptive rational orthogonal system. *Automatica* **2012**, *48*, 1154–1162. [CrossRef]

29.   Chen, Q.; Mai, W.; Zhang, L.; Mi, W. System identification by discrete rational atoms. *SIAM Rev.* **2015**, *56*, 53–59. [CrossRef]

30.   Qian, T.; Zhang, L.; Li, Z. Algorithm of Adaptive Fourier Decomposition. *IEEE Trans. Signal Process.* **2011**, *59*, 5899–5906. [CrossRef]

31.   Chen, W.; Rodrigues, M.R.D.; Wassell, I.J. Atomic decomposition by basis pursuit. *SIAM Rev.* **2013**, *61*, 2016–2029.