A New Method of Measurement Matrix Optimization for Compressed Sensing Based on Alternating Minimization

Renjie Yi 1,*, Chen Cui 1, Biao Wu 2 and Yang Gong 1

Abstract: In this paper, a new method of measurement matrix optimization for compressed sensing based on alternating minimization is introduced. The optimal measurement matrix is formulated in terms of minimizing the Frobenius norm of the difference between the Gram matrix of sensing matrix and the target one. The method considers the simultaneous minimization of the mutual coherence indexes including maximum mutual coherence $\mu_{\text{max}}$, 1-averaged mutual coherence $\mu_{\text{ave}}$ and global mutual coherence $\mu_{\text{glob}}$, and solves the problem that minimizing a single index usually results in the deterioration of the others. Firstly, the threshold of the shrinkage function is raised to be higher than the Welch bound and the relaxed Equiangular Tight Frame obtained by applying the new function to the Gram matrix is taken as the initial target Gram matrix, which reduces $\mu_{\text{ave}}$ and solves the problem that $\mu_{\text{max}}$ would be larger caused by the lower threshold in the known shrinkage function. Then a new target Gram matrix is obtained by sequentially applying rank reduction and eigenvalue averaging to the initial one, leading to lower. The analytical solutions of measurement matrix are derived by SVD and an alternating scheme is adopted in the method. Simulation results show that the proposed method simultaneously reduces the above three indexes and outperforms the known algorithms in terms of reconstruction performance.

Keywords: compressed sensing; measurement matrix; Equiangular Tight Frame; mutual coherence

1. Introduction

Compressed sensing (CS) [1] can sample the sparse or compressible signals at a sub-Nyquist rate, which brings great convenience for data storage, transmission, and processing. By adopting the reconstruction algorithms, the signal can be exactly reconstructed from the sampled data. As a marvelous way of signal processing, CS is applied in different fields such as image encryption [2], wideband spectrum sensing [3], wireless sensor network data processing [4], etc.

The original signal $x \in \mathbb{R}^{N \times 1}$ is assumed to have a sparse representation in a known domain $\Psi \in \mathbb{R}_{N \times L}^{N \times L}$ ($N \leq L$) as $x = \Psi s$ where $\Psi$ is the dictionary matrix and $s$ is a $K$-sparse signal. The incomplete measurement $y \in \mathbb{R}^{M \times 1}$ is obtained through the linear model

$$y = \Phi x = Ds$$  \hspace{1cm} (1)

where $\Phi \in \mathbb{R}^{M \times N}$ ($M < N$) is called the measurement matrix and $D = \Phi \Psi$ is the sensing matrix.

Some specific properties of $\Phi$ have great impacts on the reconstruction performance. In [5,6], spark and restricted isometric property (RIP) are respectively proposed as the sufficient conditions on $\Phi$ to recovery guarantee. However, computing the spark of a matrix has combinatorial complexity and certifying RIP for a matrix requires combinatorial search, that is to say, these tasks are NP-hard and difficult to accomplish. To a large extent, the coherence between $\Phi$ and $\Psi$ reflects the performance of meeting the above conditions.
Actually, the coherence is equivalent to the mutual coherence of $D$. Since the mutual coherence can be easily manipulated to provide recovery guarantees, it is commonly used to measure the performance of $\Phi$. The frequently-used mutual coherence indexes include the maximum mutual coherence $\mu_{\text{max}}$ [7], $I$-averaged mutual coherence $\mu_{\text{ave}}$ [8] and global mutual coherence $\mu_{\text{all}}$ [9], which respectively represent the maximum, the average and the sum of squares of the correlation between any distinct pair of columns in $D$.

The first attempt to consider the optimal design of $\Phi$ is given in [8]. The simulation results carried out in [8] show that the optimized $\Phi$ leads to smaller $\mu_{\text{ave}}$ and a substantially better CS reconstruction performance is obtained. Then the optimization of $\Phi$ becomes an important issue in CS. The recent works try to find the optimal $\Phi$ which has excellent performance in reducing the mutual correlation by minimizing the Frobenius norm of the difference between the Gram matrix $G = DD^H$ and the target Gram matrix $G_t$. The main work focuses on designing $G_t$ and finding the best $\Phi$.

In [8], $G_t$ is obtained by shrinking the off-diagonal entries in $G$. The shrinkage technique reduces $\mu_{\text{ave}}$ but it is time-consuming. Furthermore, $\mu_{\text{max}}$ is still large, which ruins the worst-case guarantees of the reconstruction algorithms. A suitable point between the current solution and the one obtained using a new shrinkage function is chosen to design the $G_t$ in [10]. It is of very strong competitiveness in $\mu_{\text{ave}}$ and $\mu_{\text{max}}$. However, the optimal point is hard to determine and the unsuitable point may seriously degrade the algorithm’s performance. In [9], $G_t$ is obtained by averaging the eigenvalues of $G$. Simulation results show that $\mu_{\text{all}}$ is reduced effectively but reducing $\mu_{\text{ave}}$ and $\mu_{\text{max}}$ is hard to be guaranteed, which means $\mu_{\text{ave}}, \mu_{\text{max}}$ and $\mu_{\text{all}}$ may maintain large values. Duarte-Carajalino and Sapiro [11] set $G_t$ as an identity matrix. Since $D$ is overcomplete and $G$ cannot be an identity matrix, simply minimizing the difference between $G$ and $G_t$ does not imply low $\mu_{\text{max}}$ [12].

In [13–17], $G_t$ is chosen from a set of relaxed Equiangular Tight Frames (ETF) [18]. The set can be formulated as $S = \left\{ G_t \in \mathbb{R}^{L \times L} : G_t = G_t', \text{diag}G_t = 1, \max_{t \neq j} |G_t(i,j)| \leq \mu_{\text{welch}} \right\}$ where $\mu_{\text{welch}}$ denotes the Welch bound [19] and $G_t(i,j)$ denotes the $(i,j)$th entry of $G_t$. However, the maximum absolute value of off-diagonal entries in $G$ is almost always greater than $\mu_{\text{welch}}$. In this case, the optimization usually implies a solution $D$ with low $\mu_{\text{ave}}$ but high $\mu_{\text{max}}$. In summary, the target Gram matrices mentioned above only focus on a certain mutual coherence index, and fail to take into account $\mu_{\text{ave}}, \mu_{\text{max}}$ and $\mu_{\text{all}}$ simultaneously. When a certain index is targeted, the other indexes may not decrease significantly or even increase. Therefore, $\Phi$ is not ‘good’ enough and the reconstruction performance is well below par.

After designing $G_t$, the next step is to find the ‘best’ $\Phi$ by approaching $G$ to $G_t$. In [8–11,17], $G$ is obtained by applying SVD to $G_t$ primarily and then the square root $D$ of $G$ is built as $D^HD = G$. At last, $\Phi$ is obtained by $\Phi = DD^H$ where $\dagger$ denotes the Moore Penrose pseudoinverse. This kind of method is intuitive, but the generalized pseudoinverse poses problems of calculation accuracy and robustness [15]. In [14,15], gradient algorithm and quasi-Newtonian algorithm are respectively utilized to attain $\Phi$. Firstly, the cost function $F(\Phi)$ with $\Phi$ as the variable is constructed. Then the search direction is determined by the derivation of $F(\Phi)$. Finally, $\Phi$ is obtained with a fixed step size. However, choosing a suitable step size which has a great influence on the accuracy of the solution requires a lot of comparison work. Moreover, the gradient algorithm and quasi-Newtonian algorithm cannot converge until a certain number of iterations is accomplished, resulting in high computational cost. In [11,16], the method for designing $\Phi$ shares the same concept as K-SVD [20], that is to update a matrix row by row. Eigenvalue decomposition is required to find the square root of the maximum eigenvalue for each row, which results in a significant increase in the calculation. For solving this problem, Hong et al. [16] utilize the power method instead of eigenvalue decomposition. However, the eigenvalue obtained by power method is the one with the largest absolute value. When the eigenvalue is negative, eigenvalue decomposition is still necessary.

The primary contributions of this paper are threefold:
The new target Gram matrix $G_t$ targets $\mu_{ave}$, $\mu_{max}$, and $\mu_{all}$ of $D$ simultaneously is designed. Firstly, a new shrinkage function whose threshold exceeds $\mu_{welch}$ is utilized to determine the initial target Gram matrix. Then $G_t$ is obtained by sequentially applying rank reduction and eigenvalue averaging to the initial matrix.

Analytical solutions of the measurement matrix $\Phi$ to minimize the difference between $G = \Psi' \Phi \Phi' \Psi$ and $G_t$ are derived by SVD.

Based on alternating minimization, an iterative method is proposed to optimize the measurement matrix. The simulation results confirm the effectiveness of the proposed method in decreasing the mutual coherence indexes and improving reconstruction performance.

The remainder of this paper is organized as follows. Some basic definitions related to mutual coherence indexes and frames are described in the next section. The main results are presented in Section 3, where the solutions to the $G_t$ design are characterized and a class of the solutions to the optimal $\Phi$ is derived in detail. The procedure of our method and the discussion can be also found in Section 3. In Section 4, simulations are carried out to confirm the effectiveness of the proposed method. In the end, the conclusion is drawn.

2. Mutual Coherence Indexes and ETFs

2.1. Mutual Coherence Indexes

Rewrite $D = [d_1, d_2 \cdots d_L] \in \mathbb{R}^{M \times L}$ where $d_i \in \mathbb{R}^{M \times 1}$ and $\|d_i\|_2 = 1$. Denote $g_{ij} = d_i^d d_j$ the entry at the position of row $i$ and column $j$ in $G$, where $i, j = 1, 2 \cdots L$. Here, we quote the definitions of the mutual coherence indexes as that presented by Donoho [7], Elad [8], and Zhao [9].

Definition 1. For a matrix $D$, the maximum mutual coherence $\mu_{max}$ is defined as the largest absolute and normalized inner product between all columns in $D$ that can be described as

$$\mu_{max} = \max_{i \neq j} |d_i^d d_j| = \max_{i \neq j} |g_{ij}|$$

(2)

Definition 2. For a matrix $D$, the $t$-averaged mutual coherence $\mu_{ave}$ is defined as the average of all absolute and normalized inner products between different columns in $D$ that are above $t$ and can be described as

$$\mu_{ave} = \frac{\sum_{i \neq j} (|g_{ij} \geq t|) |g_{ij}|}{\sum_{i \neq j} (|g_{ij} \geq t|)}$$

(3)

Definition 3. For a matrix $D$, the global mutual coherence $\mu_{all}$ is defined as the sum of squares of normalized inner products between all columns in $D$ that can be described as

$$\mu_{all} = \sum_{i \neq j} g_{ij}^2$$

(4)

As shown in [5], the original signal can be exactly reconstructed as long as $K < (1 + 1/\mu_{max})/2$. The conclusion is true from a worst-case standpoint which means that $\mu_{max}$ does not do justice to the actual behavior of sparse representations. Therefore, Elad considers that an “average” measure of mutual coherence, namely $\mu_{ave}$, is more likely to describe its true behavior. Different from the previous two indexes, $\mu_{all}$ reflects the overall property of $D$.

In fact, the purpose of reducing the mutual coherence indexes of $D$ is to attain $G$ that meets the following requirements: (1) The maximum absolute value of off-diagonal entries in $G$ is sufficiently small; (2) The number of off-diagonal entries with large absolute value
is minimized; (3) The average of off-diagonal entries with large absolute value is as small as possible. However, when a certain mutual coherence index is targeted solely, we cannot guarantee that the obtained $G$ will fully meet the requirements. Therefore, the decrease of a certain index does not always mean better $\Phi$ and improved reconstruction performance. When the three indexes are reduced simultaneously, the requirements are better satisfied and better performance is obtained.

2.2. ETFs

It is shown in [19] that $\mu_{\text{max}}$ of $D \in \mathbb{R}^{M \times L}$ is lower bounded by

$$\mu_{\text{max}} \geq \mu_{\text{welch}} = \sqrt{\frac{L - M}{M(L - 1)}}$$

The bound is achievable for ETF. Here, we recall the definition of ETF [18].

**Definition 4.** Let $F$ be a $M \times L$ matrix whose columns are $f_1, f_2, \cdots, f_L$. The matrix $F$ is called an equiangular tight frame if it satisfies three conditions

1. Each column has a unit norm: $\|f_i\|_2 = 1$ for $i = 1, 2, \cdots, L$.
2. The columns are equiangular. For some nonnegative $\theta$, we have $f_i^T f_j = \theta$ when $i, j = 1, 2, \cdots, L$ and $i \neq j$.
3. The columns form a tight frame. That is, $FF^T = (L/M)I_M$ where $I_M$ is an $M \times M$ identity matrix.

Sustik et al. [18] show that a real $M \times L$ ($1 < M < L - 1$) ETF exists on if $L \leq \min\{M(M + 1)/2, (L - M)(L - M + 1)/2\}$ holds. Furthermore, $\sqrt{M(L - 2M)/(L - M)}$ and $\sqrt{(L - M)(L - 1)/M}$ must be odd integers when $L \neq 2M$, $M$, and $2M - 1$ must be an odd number and the sum of two squares respectively when $L = 2M$. Fickus et al. [21] surveys some known construction of ETFs and tabulates existence for sufficiently small dimensions. The above studies show that $M$ and $L$ must meet some exacting requirements when an ETF is available for $D$. However, it is really difficult to meet the requirements in practice, which means the maximum absolute value of the off-diagonal entries in $G$ is usually significantly larger than $\mu_{\text{welch}}$.

3. The Proposed Method

The off-diagonal entries in $G$ actually are the inner products between different columns in $D$. Reducing those entries is likely to lead to lower mutual coherence indexes and better performance. The most straightforward approach is to replace large off-diagonal values with small ones. However, it is impossible to solve $\Phi$ from a certain $G$ because of the inequality of rank between $\Psi \Phi^T \Phi \Psi$ and $G$ when the approach is adopted. Therefore, a feasible approach is to minimize the difference between $G$ and $G_t$ that can be formulated as

$$\min\|G_t - G\|_F^2$$

where $G = \Psi \Phi^T \Phi \Psi$. This problem can be solved by alternating minimization strategy [14,16], which iteratively minimizes (5) to find the desired $\Phi$. The idea is to update $G_t$ and $\Phi$ alternatively and repeat this proceeding until a stop criterion is reached. In this section, we design $G_t$ firstly and then derive the analytical solutions of $\Phi$. Finally, an iterative method is proposed to optimize the measurement matrix based on alternating minimization.
3.1. The Design of $G_t$

It can be seen from (5) that $G_t$ plays an important role in measurement matrix optimization. In recent works, $G_t$ is frequently set as the relaxed ETF matrix, which is obtained by applying the following shrinkage function

$$
G_t(i,j) = \begin{cases} 
\frac{s_{ij}}{\lambda_{welch}}, & |s_{ij}| \leq \zeta \\
\text{sign}(s_{ij})\lambda_{welch}, & \text{otherwise} 
\end{cases}
$$

where $\zeta = \mu_{welch}$ for $i,j = 1,2,\ldots,L$ and $i \neq j$. Such a scheme in designing $G_t$ guarantees that the off-diagonal entries with large value of $G$ will be intensively constrained, which means lower $\mu_{ave}$ and $\mu_{max}$.

Recall from Section 2.2 that the Welch bound is not achievable for $G$ in most cases. As shown in [14,16], different $\zeta$ yields different results and $\mu_{welch}$ is not the optimal value. Li [22] et al. found that a smaller $\mu_{max}$ is available when $\zeta$ is slightly larger than $\mu_{welch}$. Inspired by [22], we propose an improved shrinkage function which divides the entries in $G$ into three segments through two thresholds. One of the thresholds is $\mu_{welch}$ and the other is larger than $\mu_{welch}$. The function is as follows

$$
G_t(i,j) = \begin{cases} 
\frac{\text{sign}(s_{ij})\lambda_{welch}}{\text{Thr}}, & |s_{ij}| \geq \text{Thr} \\
\text{sign}(s_{ij})\lambda_{welch}, & \mu_{welch} \leq |s_{ij}| < \text{Thr} \\
s_{ij}, & |s_{ij}| < \mu_{welch} 
\end{cases}
$$

where $\text{Thr} = \mu_{welch} + c$ and $0 < c < \mu_{welch}$. As can be seen from Equation (7), the maximum absolute value of off-diagonal entries in $G_t$ is raised from $\mu_{welch}$ to $\text{Thr}$. According to the previous analysis, the new function is likely to lead to a further reduction in $\mu_{max}$ while maintaining the advantage of Equation (6) with respect to $\mu_{ave}$.

After shrinkage, $G_t$ becomes full rank generally [8], that is $\text{rank}(G_t) = L$. However, the rank of $G$ is identically equal to $M$. Thus, we consider mending this by forcing a rank $M$. A new target Gram matrix, denoted as $G_{t,M}$, is obtained by solving

$$
\min \|G_t - G_{t,M}\|_F^2, \text{ s.t. } \text{rank}(G_{t,M}) = M
$$

The solutions to this problem are given by Theorem 1 below.

**Theorem 1.** Let $G_t \in R^{L \times L}$ be the matrix obtained by applying the shrinkage operation shown as Equation (7) to $G$ and $G_t = \mathbf{P}\mathbf{A}\mathbf{P}'$ be the eigendecomposition of $G_t$. $\mathbf{P}$ is orthonormal with dimension $L$ and $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_L)$ with $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_L|$. The solutions of the minimization problem defined by (8) are characterized by

$$
G_{t,M} = \mathbf{P}\mathbf{A}\mathbf{P}'
$$

where $\mathbf{A} = \left[ \begin{array}{cc} I_M & 0 \\ 0 & 0 \end{array} \right] \in R^{L \times L}$.

**Proof.** Denote $G_{t,M} = XX'$ where $X \in R^{M \times L}$. Let $X = U_X \left[ \begin{array}{cc} \Sigma_X & 0 \end{array} \right] V_X$ be an SVD of $X$, where $U_X \in R^{M \times M}$ and $V_X \in R^{L \times L}$ are unitary. Then $G_{t,M}$ can be rewritten as

$$
G_{t,M} = V_X \left[ \begin{array}{cc} \Sigma_X^2 & 0 \\ 0 & 0 \end{array} \right] V_X
$$
Denote \( f = \| G_I - G_{l,M} \|_F^2 \). By substituting \( G_{l,M} \) with \( G_{l,M} = X'X \), can be rewritten as a function of matrix \( X \). Let \( \partial f / \partial X \) be the derivative of \( f \) with respect to \( X \). The optimal \( X \) should satisfy \( \partial f / \partial X = 0 \). Equivalently, we have

\[
XX'X = XG_I
\]  

(11)

It then follows from \( X = U_X [ \Sigma_X \ 0 ] V_X \) that

\[
\begin{bmatrix}
\Sigma_X^2 & 0 \\
0 & 0
\end{bmatrix} = AV'_X G_I V_X
\]

(12)

Substituting Equation (12) into Equation (10), we obtain

\[
G_{l,M} = V_X AV'_X G_I
\]

(13)

It turns out from the unitary invariance with Equation (13) that

\[
f = \left\| \left( I_L - V_X A V'_X \right) G_I \right\|_F^2
\]

(14)

With a few manipulations, we conclude that the solution of (8) is equivalent to solving

\[
\text{max} \text{tr} \left( G_i' V_X A V'_X G_i \right)
\]

(15)

where \( \text{tr}() \) denotes the matrix trace operation. Noting that \( A = A' A \) and \( G_i = P A P' \), the problem in (15) is equivalent to

\[
\text{max} \left\| A V'_X P A \right\|_F^2
\]

(16)

Denote \( B = V'_X P \) and rewrite \( B \) as \( B = [b_1, b_2 \cdots b_L] \) where \( b_i \in R^{L \times 1} \) for \( i = 1,2 \cdots L \). Rewrite \( A \) as \( A = [e_1, e_2 \cdots e_M, 0,0 \cdots 0]' \), where \( e_i \in R^{L \times 1} \) denotes a unit vector with the \( i \)th entry is equal to 1 for \( j = 1,2 \cdots M \). Then, it is easy to obtain that

\[
A V'_X P A =
\begin{bmatrix}
\lambda_1 e'_1 b_1 & \lambda_2 e'_1 b_2 & \cdots & \lambda_L e'_1 b_L \\
\lambda_1 e'_2 b_1 & \lambda_2 e'_2 b_2 & \cdots & \lambda_L e'_2 b_L \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_1 e'_M b_1 & \lambda_2 e'_M b_2 & \cdots & \lambda_L e'_M b_L \\
0 & 0 & \cdots & 0
\end{bmatrix}
\]

Let \( b_{ij} = e'_i b_j \) be the \( j \)th entry of \( b_i \). It can be shown with some manipulations that the problem in (16) is equivalent to

\[
\text{max} \sum_{i=1}^{L} \sum_{j=1}^{M} \lambda_i^2 b_j^2
\]

(17)

With \( |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_L| \) and \( \sum_{i=1}^{L} \sum_{j=1}^{M} b_{ij}^2 = M \), it is straightforward that \( \sum_{i=1}^{L} \sum_{j=1}^{M} \lambda_i^2 b_j^2 \) reaches the maximum value \( \lambda_1^2 + \lambda_2^2 + \cdots + \lambda_M^2 \) only if \( \sum_{j=1}^{M} b_{ij}^2 = 1 \) for \( i = 1, 2 \cdots M \).

In this case, \( \sum_{j=1}^{M} b_{ij}^2 = 0 \) and \( \sum_{j=1}^{M} b_{ij}^2 = 0 \) hold for \( i = M+1, M+2 \cdots L \). Rewrite \( B \) as
When minimizing the difference between $\mathbf{G}$ with $\mathbf{B} = \text{unitary}$. Similarly, let $\Psi$ the difference between $\Phi$ find the best matrix which leads to better $\mu$ minimizing $\mu$ of good competitiveness in $\mu$ to a smaller sum of squares of all off-diagonal values of $\mathbf{G}$. Noting that $\mathbf{B}$ is most similar to $\mathbf{3}$. The Analytical Solutions of $\Phi$ $\lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_M)$, where $\hat{\lambda} = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_M, 0, \cdots, 0)$, a new target Gram matrix denoted as $\mathbf{G}_{t, \text{opt}}$ with $M$ equal all non-zero eigenvalues is given by $\mathbf{G}_{t, \text{opt}} = \mathbf{P} \hat{\lambda} \mathbf{P}'$.

Recall that $\mathbf{G}_{t, M}$ is most similar to $\mathbf{G}_t$ in terms of Frobenius norm, it means that $\mathbf{G}_{t, M}$ is of good competitiveness in $\mu_{\text{ave}}$ and $\mu_{\text{max}}$. Furthermore, as a variant of $\mathbf{G}_{t, M}$, $\mathbf{G}_{t, \text{opt}}$ reduces the sum of squares of all off-diagonal values of $\mathbf{G}_{t, M}$, leading to a better performance in minimizing $\mu_{\text{all}}$. Therefore, $\mathbf{G}_{t, \text{opt}}$ is more likely to be an ideal solution of target Gram matrix which leads to better $\mu_{\text{ave}}$, $\mu_{\text{max}}$, and $\mu_{\text{all}}$ simultaneously.

3.2. The Analytical Solutions of $\Phi$

After obtaining the target Gram matrix $\mathbf{G}_{t, \text{opt}}$, the next step of the optimization is to find the best $\Phi$. To handle the problem, we try to find the optimal solution by minimizing the difference between $\Psi$ and $\Phi$ and $\mathbf{G}_{t, \text{opt}}$ as

$$\min \| \mathbf{G}_{t, \text{opt}} - \Phi \|_F$$

Let $\mathbf{D} = \mathbf{U}_D \Sigma_D \mathbf{V}_D'$ be an SVD of $\mathbf{D}$, where $\mathbf{U}_D \in \mathbb{R}^{M \times M}$ and $\mathbf{V}_D \in \mathbb{R}^{L \times L}$ are unitary. Similarly, let $\Psi = \mathbf{U}_\Psi \Sigma_\Psi \mathbf{V}_\Psi'$ be an SVD of $\Psi$, where $\mathbf{U}_\Psi \in \mathbb{R}^{N \times N}$ and $\mathbf{V}_\Psi \in \mathbb{R}^{L \times L}$ are unitary. The solutions to this problem are given by Theorem 2 below.
Theorem 2. Let $G_{l_{\text{opt}}}$ be the matrix shown as Equation (21) and $\Lambda_M \in R^{M \times M}$ be the $M$th principal submatrix of $\hat{\Lambda}$. Then the solutions of the minimization problem defined by (22) are characterized by

$$\Phi_{\text{opt}} = U_Z \left[ \left( \Lambda_M \right)^{-\frac{1}{2}} \begin{bmatrix} 0 \\ \Lambda \end{bmatrix} \right] \left[ \begin{bmatrix} 0 \\ \Sigma^{-1} \end{bmatrix} \right] \left[ \begin{bmatrix} 0 \\ \Sigma \end{bmatrix} \right]^{-\frac{1}{2}} U^T \Psi$$

where $U_Z \in R^{M \times M}$ is an arbitrary unitary matrix.

Proof. Assume that the off-diagonal values of $\Sigma_D$ and $\Sigma$ are non-zero, $\Phi$ can be written as

$$\Phi = U_D \left[ \begin{bmatrix} \Sigma_D \\ 0 \end{bmatrix} \right] V_D \left[ \begin{bmatrix} \Sigma^{-1} \end{bmatrix} \right] U^T \Psi$$

By substituting $\Phi$ in (22) with Equation (24), it can be shown with some manipulations that the solutions of the problem in (22) are equivalent to the solutions of the minimization problem defined by (22) are equivalent to the solutions of

$$\min \left\| V_D P \hat{\Lambda}^T V_D - \begin{bmatrix} \Sigma_D^2 \\ 0 \end{bmatrix} \right\|_F^2$$

Let $Z = V_D P \hat{\Lambda}^T V_D$ and $z_i$ be the $i$th diagonal entry of $Z$. Denote $\Lambda_Z = diag(z_1, z_2, \cdots, z_M)$. With further manipulations, we simplify (25) to

$$\min \left\| G_{l_{\text{opt}}} \right\|_F^2 + \left\| \Lambda_Z - \Sigma_D^2 \right\|_F^2 = \left\| \Lambda_Z \right\|_F^2$$

Obviously, the minima are achievable only if $\Lambda_Z = \Sigma_D^2$ holds and $\left\| \Lambda_Z \right\|_F^2$ takes the maximum value. Let $U = V_D P$ and $u_{ij}$ be the $(i, j)$th entry of $U$ where $i, j = 1, 2, \cdots, L$. Noting that the top $M$ diagonal entries of $\hat{\Lambda}$ are all equal to $\hat{\lambda}$, we have

$$z_i = \hat{\lambda} \left( u_{1}^2 + u_{2}^2 + \cdots + u_{M}^2 \right)$$

It is worth noting that $\left\| \Lambda_Z \right\|_F^2 = \sum_{i=1}^{M} z_i^2$ and $\sum_{i=1}^{L} z_i = M \hat{\lambda}$ where $0 \leq z_i \leq \hat{\lambda}$. Hence, it is clear that the maxima of $\frac{M}{i=1} z_i^2$ are reached when $z_i$ takes the maximum value $\hat{\lambda}$ for $i = 1, 2, \cdots, M$. Rewrite $U$ as $U = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{bmatrix}$ where $U_1 \in R^{M \times M}$. Since $z_i = \hat{\lambda}$ holds for $i = 1, 2, \cdots, M$, we have $u_{11}^2 + u_{22}^2 + \cdots + u_{MM}^2 = 1$ and $U_2 = 0, U_3 = 0$ accordingly. As $U_Z$ is a unitary matrix, it is easy to verify that $U_1$ and $U_4$ are both unitary matrices. Then it follows that

$$V_D = \begin{bmatrix} U_1 \\ 0 \\ U_4 \end{bmatrix} p'$$

Substituting Equation (28) into Equation (24), the optimal solution is obtained by

$$\Phi_{\text{opt}} = U_Z \left[ \left( \Lambda_M \right)^{-\frac{1}{2}} \begin{bmatrix} 0 \\ \Lambda \end{bmatrix} \right] \left[ \begin{bmatrix} 0 \\ \Sigma^{-1} \end{bmatrix} \right] \left[ \begin{bmatrix} 0 \\ \Sigma \end{bmatrix} \right]^{-\frac{1}{2}} U^T \Psi$$

The proof is then completed. \square

3.3. Comments

According to Sections 3.1 and 3.2, the procedure for measurement matrix optimization has been summarized in Algorithm 1.
Algorithm 1. The proposed optimization method.

| Step | Description |
|------|-------------|
| 1.   | Compute the sensing matrix $D = \Phi \Psi$ and normalize the columns in $D$. |
| 2.   | Compute Gram matrix $G = DD^\top$. |
| 3.   | Shrink $G$ and obtain $G_1$ by |
|      | $G_1(i,j) = \begin{cases} 
\text{sign}(s_{ij})(\mu_{\text{welch}} + c), & |s_{ij}| \geq \mu_{\text{welch}} + c \\
\text{sign}(s_{ij})\mu_{\text{welch}}, & \mu_{\text{welch}} \leq |s_{ij}| < \mu_{\text{welch}} + c \\
|s_{ij}|, & |s_{ij}| < \mu_{\text{welch}} 
\end{cases}$ |
| 4.   | Apply eigenvalue decomposition to $G_1$ and obtain $G_1 = \mathbf{P}\Lambda\mathbf{P}^\top$. |
| 5.   | Compute the average of the top $M$ diagonal entries in $\Lambda$, denoted as $\hat{\lambda}$. |
| 6.   | Construct $\Lambda_M = R^{M \times M}$ as $\Lambda_M = \text{diag}(\hat{\lambda}, \hat{\lambda}, \ldots, \hat{\lambda})$. |
| 7.   | Update $\Phi$ by $\Phi_t = U_Z \left[ \left( \Lambda_M \right)^{\frac{1}{2}} \begin{bmatrix} P & V \end{bmatrix} \left[ \begin{bmatrix} \Sigma_I \end{bmatrix}^{-1} & 0 \end{bmatrix} \right] U^\top \Psi \right]$. |

end return $\Phi_{\text{Iter}}$

Noting that $G_1$ plays an important role in measurement matrix optimization, Algorithm 1 takes $\mu_{\text{ave}}, \mu_{\text{max}}$, and $\mu_{\text{all}}$ into consideration simultaneously when designing the $G_1$. By minimizing (5), the Gram matrix is most similar to $G_1$ in terms of Frobenius norm, leading to maintain the advantage of $G_1$ in reducing the mutual coherence indexes. Therefore, Algorithm 1 is effective in reducing $\mu_{\text{ave}}, \mu_{\text{max}}$, and $\mu_{\text{all}}$.

In the shrinkage function, a different threshold yields different results. Inspired by [22], we propose a shrinkage function shown as (7) which has a new threshold $\mu_{\text{welch}} + c$. We have not derived the optimal value of $c$ in theory, but setting $c$ to a proper value can also lead to a moderate result.

After averaging the eigenvalues of $G_{1:M}$, the first term on the right part of (20) is minimized. However, the diagonal entries of $G_{1:M}$ change accordingly. Hence, we can’t assure that $\mu_{\text{all}}$ reaches the minima, that is to say, $G_{1:M,\text{opt}}$ may not be the optimal solution in terms of $\mu_{\text{all}}$. Fortunately, we find that the change of $\sum_{i=1}^{L} \lambda_i^2$ is much greater than that of $\sum_{i=1}^{L} \delta_{ij}^2$ in (20), which means our approach is effective in reducing $\mu_{\text{all}}$.

The proposed algorithm is an iterative one. The main complexity of Algorithm 1 for each iteration is located at steps 1, 2, 4, and 7. For those steps, the flops required are $O(MNL), O(ML^2), O(L^3)$, and $O(L^3)$ respectively. Hence, the complexity of Algorithm 1 is approximate to be $O(\text{Iter}L^3)$. Since the complexity for similar algorithms in [8, 16, 17] which apply eigenvalue decomposition or SVD is no less than $O(\text{Iter}L^3)$, the proposed algorithm has not increased the complexity significantly.

4. Simulation Results and Discussion

In this section, we conduct simulations to determine a suitable $c$ firstly. Then, we examine the mutual coherence indexes and reconstruction performance of the proposed method and compare them with the well-established similar algorithms given in [8, 16, 17] by presenting the empirical results. Last, we verify the effectiveness of our method with various measurement matrices and dictionary matrices. The iteration number Iter is set to 100 and $t$ is set to $\mu_{\text{welch}}$. For a given dictionary matrix $\Psi \in R^{120 \times 120}$, $x \in R^{120 \times 1}$ has a sparse representation as $x = \Psi s$ where $s$ is $K$-sparse and each non-zero entry is randomly positioned with a Gaussian distribution of i.i.d. zero-mean and unit variance. Orthogonal Matching Pursuit (OMP) [23] algorithm is employed in signal reconstruction. Denote
\[ \varepsilon = \frac{\| \mathbf{x}_e - \mathbf{x} \|_2}{\| \mathbf{x} \|_2} \] the reconstruction error where \( \mathbf{x}_e \) is the reconstructed signal. The reconstruction is identified as a success, called exact reconstruction, provided \( \varepsilon \leq 10^{-6} \). Denote \( P_{\text{succ}} \) the percentage of successful reconstruction. In Sections 4.1–4.3, \( \Phi_0 \) and \( \Psi \) are both Gaussian random matrices.

4.1. The Choice of \( c \)

Since the analytical solution of \( c \) is extremely difficult, here, we conduct a serious of simulations to find a suitable \( c \). Figure 1 illustrates the change tendency of mutual coherence indexes and \( P_{\text{succ}} \) with argument \( c \). We fix the row number of \( M \) to 28, the sparsity to 8, and varies \( c \) from 0 to 0.16. The experiment is performed for 1000 random sparse ensembles and the results are recorded.

When \( c = 0 \), the shrinkage function shown as (7) is the same as (6). As can be seen from the graphs, when \( c \) increases, \( \mu_{\text{ave}} \) increases, \( \mu_{\text{max}} \) and \( \mu_{\text{all}} \) decrease first and then increase, \( P_{\text{succ}} \) increases firstly and then decreases. \( \mu_{\text{max}} \) and \( \mu_{\text{all}} \) reach the minima when \( c = 0.02 \) and \( c = 0.03 \) respectively. It is worth noting that appropriate increase of \( c \) leads to decrease of \( \mu_{\text{max}} \) and \( \mu_{\text{all}} \) but increase of \( \mu_{\text{ave}} \). When \( c = 0.01 \), better \( \mu_{\text{max}} \) and \( \mu_{\text{all}} \) are obtained and the loss in \( \mu_{\text{ave}} \) is tolerable. Moreover, \( P_{\text{succ}} \) reaches the maxima. Therefore, 0.01 may be a moderate value for \( c \) and \( c \) is set to 0.01 in the simulations in Sections 4.2–4.4.

4.2. Comparing the Mutual Coherence Indexes

This section presents a series of simulations to compare our method with algorithms given in [8,16,17] on the three mutual coherence indexes of \( D \) obtained by \( D = \Phi_0 \Psi \) where \( \Phi_0 \) is the optimized measurement matrix. For convenience, each method is denoted as Propose, Elad, Hong, and Entezari. The down-scaling factor for Elad is set to 0.95. The inner iteration number for Hong is set to 2, which means K-SVD is applied twice in every updating of \( \Phi \). The point is set to 0.5 to update the \( G_t \) in Entezari.

Figure 2 illustrates the change tendency of mutual coherence indexes with iteration number for \( M = 28 \). As can be seen from the figure, the indexes corresponding to different algorithms all change monotonously with the iteration number. When \( \mu_{\text{max}} \) and \( \mu_{\text{ave}} \) converge, the number of iterations required by our method is almost equal to that of Hong and significantly less than that of Elad. When \( \mu_{\text{all}} \) converges, the number of iterations required by our method is equivalent to that of Entezari and significantly less than that of Hong and Elad.

Figure 3 presents the histogram of the absolute off-diagonal values of \( (\Phi_0 \Psi)^\text{T} \Phi_0 \Psi \) for \( M = 28 \). It is seen from the figure that Elad and Entezari have long tails, showing that the number of off-diagonal values that exceed 0.34 is relatively large. The tail of Hong is shorter than that of Elad and Entezari, and reaches the maximum of 0.34. Compared with Hong, our method has a shorter tail which reaches the maximum of 0.32 and has more off-diagonal values below the \( \mu_{\text{welch}} \) (0.1662).

In conclusion, while effectively reducing \( \mu_{\text{max}} \) and \( \mu_{\text{all}} \), our method can maintain a small \( \mu_{\text{ave}} \) at the same time. Additionally, the number of iterations required for the convergence of each index of our method is significantly less than that of Elad. Therefore, from the view of mutual coherence indexes, the measurement matrix obtained by our method has better properties than the other three methods. This coincides with the theoretical results obtained in the Section 3.

4.3. Comparing the Reconstruction Performance

Case 1. Comparison of the \( P_{\text{succ}} \) in the noiseless case.

In this case, we conduct two separate CS experiments, first by fixing \( K = 8 \) and varying \( M \) from 12 to 44 and second by fixing \( M = 28 \) and varying \( K \) from 4 to 20. Each experiment is performed for 1000 random sparse ensembles and the number of exact reconstruction is recorded.
Figures 4 and 5 reveal that the $P_{suc}$ of our method is the highest, which indicates its superiority over the other three methods.

![Graphs showing $\mu_{\text{max}}$, $\mu_{\text{ave}}$, and $\mu_{\text{all}}$ versus $c$](image)

Figure 1. (a) $\mu_{\text{max}}$ and $P_{suc}$ versus $c$; (b) $\mu_{\text{ave}}$ and $P_{suc}$ versus $c$; (c) $\mu_{\text{all}}$ and $P_{suc}$ versus $c$, both with $M = 28$ and $K = 8$. 
Figure 2. The convergence results: (a) evolution of $\mu_{\text{max}}$, (b) $\mu_{\text{ave}}$ and (c) the evolution of $\mu_{\text{all}}$, all versus iteration number, where $M = 28$. 
Figure 3. Histogram of the absolute off-diagonal values of $\langle \Phi_{opt} \Psi \rangle' \Phi_{opt} \Psi$ for $M = 28$.

Table 1. $\mu_{max}$ by Elad, Hong, Entezari, and proposed method versus measurement dimension $M$.

| M   | Elad   | Hong   | Entezari | Propose | $\mu_{welch}$ |
|-----|--------|--------|----------|---------|---------------|
| 12  | 0.8971 | 0.8884 | 0.8465   | 0.8072  | 0.2750        |
| 16  | 0.8157 | 0.6488 | 0.7517   | 0.5211  | 0.2337        |
| 20  | 0.7610 | 0.4625 | 0.6710   | 0.4218  | 0.2050        |
| 24  | 0.6972 | 0.3920 | 0.5874   | 0.3618  | 0.1833        |
| 28  | 0.6186 | 0.3403 | 0.5351   | 0.3223  | 0.1662        |
| 32  | 0.5460 | 0.3069 | 0.4911   | 0.2919  | 0.1520        |
| 36  | 0.4879 | 0.2804 | 0.4472   | 0.2659  | 0.1400        |
| 40  | 0.4454 | 0.2591 | 0.4177   | 0.2487  | 0.1296        |
| 44  | 0.3802 | 0.2404 | 0.3927   | 0.2302  | 0.1205        |

Table 2. $\mu_{ave}$ by Elad, Hong, Entezari, and proposed method versus measurement dimension $M$.

| M   | Elad   | Hong   | Entezari | Propose | $\mu_{welch}$ |
|-----|--------|--------|----------|---------|---------------|
| 12  | 0.4168 | 0.4528 | 0.4141   | 0.3608  | 0.2750        |
| 16  | 0.3507 | 0.2964 | 0.3532   | 0.2979  | 0.2337        |
| 20  | 0.3016 | 0.2566 | 0.3111   | 0.2581  | 0.2050        |
| 24  | 0.2431 | 0.2272 | 0.2772   | 0.2288  | 0.1833        |
| 28  | 0.2212 | 0.2042 | 0.2512   | 0.2060  | 0.1662        |
| 32  | 0.2056 | 0.1854 | 0.2302   | 0.1879  | 0.1520        |
| 36  | 0.1937 | 0.1700 | 0.2118   | 0.1727  | 0.1400        |
| 40  | 0.1841 | 0.1567 | 0.1961   | 0.1600  | 0.1296        |
| 44  | 0.1760 | 0.1454 | 0.1824   | 0.1491  | 0.1205        |
Table 3. $\mu_{all}$ by Elad, Hong, Entezari, and proposed method versus measurement dimension $M$.

| M  | Elad     | Hong     | Entezari | Propose |
|----|----------|----------|----------|---------|
| 12 | 1220.20  | 2638.47  | 1081.18  | 1080.19 |
| 16 | 898.80   | 802.77   | 780.92   | 780.16  |
| 20 | 721.41   | 630.91   | 605.72   | 600.13  |
| 24 | 571.34   | 512.24   | 480.67   | 480.12  |
| 28 | 474.76   | 425.25   | 394.86   | 394.39  |
| 32 | 404.35   | 358.49   | 330.55   | 330.11  |
| 36 | 353.96   | 307.11   | 280.55   | 280.10  |
| 40 | 315.03   | 265.40   | 240.55   | 240.10  |
| 44 | 285.51   | 231.43   | 207.79   | 207.38  |

Figure 4. The change tendency of $P_{suc}$ with $M$ while $K = 8$ in the noiseless case.

Figure 5. The change tendency of $P_{suc}$ with $K$ while $M = 28$ in the noiseless case.
Case 2. Comparison of the \( \varepsilon \) in the noisy case.

To show the robustness of the proposed method in noisy cases we consider the noisy model \( y = \Phi x + v \) where \( v \) is the vector of additive Gaussian noise with zero means. We conduct the experiment by fixing \( M = 28, K = 8 \), and varying SNR from 10 to 50 dB. The experiment is performed for 1000 random sparse ensembles and the average reconstruction error is recorded. From Figure 6, we can see that the reconstruction errors decrease with the increase of SNR, and the error of the proposed method is smaller than that of the others.

![Figure 6](image-url). The change tendency of \( \varepsilon \) with SNR while \( M = 40 \) and \( K = 8 \).

Table 3 presents that \( \mu_{all} \) of the Entezari is slightly larger than that of our method. It is interesting to note that the number of off-diagonal entries with smaller absolute values in the Entezari is significantly larger than that of our method from Figure 3. Moreover, it can be seen from Table 2 that \( \mu_{ave} \) of Hong is slightly lower than that of our method. However, the simulation results show that our method outperforms the others in terms of reconstruction performance. It is also worthy noting that our method reduces \( \mu_{ave}, \mu_{max}, \) and \( \mu_{all} \) simultaneously, leading to better reconstruction performance in CS. This implies that a single mutual coherence index cannot accurately reflect the actual performance of the methods, and verifies the necessity of using multiple indexes simultaneously in measurement matrix optimization.

4.4. Different Kinds of \( \Phi \) and \( \Psi \) Optimized by the Proposed Methods

To analyze the performance of our method with various measurement matrices and dictionary matrices, a serious of simulations are carried out in this section. We choose the measurement matrix as a Gaussian random matrix and a Bernoulli random matrix, and choose the dictionary matrix as a Gaussian random matrix and the DCT matrix, respectively. We compare the mutual coherence indexes and the reconstruction performance before and after optimization. When \( \Psi \) is the Gaussian random matrix, \( \Phi \) belongs to \( \mathbb{R}^{M \times 80} \) and \( \Psi \) belongs to \( \mathbb{R}^{80 \times 120} \). When \( \Psi \) is the DCT matrix, \( \Phi \) belongs to \( \mathbb{R}^{M \times 120} \) and \( \Psi \) belongs to \( \mathbb{R}^{120 \times 120} \). Each experiment is performed for 1000 random sparse ensembles.

The mutual coherence indexes of different measurement matrices \( \Phi \) with different dictionary matrices \( \Psi \) are shown in Figure 7. As seen from the simulations, all the optimized measurement matrices produce smaller \( \mu_{max}, \mu_{ave}, \) and \( \mu_{all} \) than the random ones.
Figure 7. The evolution of (a) $\mu_{\text{max}}$, (b) $\mu_{\text{ave}}$ and (c) $\mu_{\text{all}}$, all versus measurements number with different $D$. 
Figures 8 and 9 present the reconstruction performance of OMP with the optimized measurement matrices and the random ones. It is seen from the graphs in these figures that all the optimized matrices outperform the random ones in terms of the percentage of exact reconstruction.

Figure 8. The change tendency of $P_{\text{Suc}}$ with $M$ while $K = 8$ in the noiseless case.

Figure 9. The change tendency of $P_{\text{Suc}}$ with $K$ while $M = 28$ in the noiseless case.

5. Conclusions

This paper focused on the optimization of measurement matrix for compressed sensing. To decrease $\mu_{\text{max}}$, $\mu_{\text{ave}}$, and $\mu_{\text{all}}$ simultaneously, we designed a new target Gram matrix which was obtained by applying a new shrinkage function to the Gram matrix and updated by performing rank reduction and eigenvalue averaging. Then, we characterized the analytical solutions of the measurement matrix by SVD. Based on alternating minimization, we proposed an iterative method to optimize the measurement matrix. The simulation results show that the proposed method reduces $\mu_{\text{max}}$, $\mu_{\text{ave}}$, and $\mu_{\text{all}}$ simultaneously and outperforms the existing algorithms in terms of reconstruction performance. In addition,
the proposed method is computationally less expensive than some existing algorithms in the literature.

As detailed, we gave the optimal value of $c$ under a fixed matrix scale through simulation. When the scale changes, the value of $c$ in Section 4.1 may no longer be applicable. Therefore, it is meaningful to find the theoretical ‘optimal value’ of $c$. Furthermore, noting that lower mutual coherence indexes mean potentially higher reconstruction performance, further efforts are needed to decrease the indexes simultaneously.

**Author Contributions:** Conceptualization: R.Y.; Methodology: C.C.; Software: R.Y. and Y.G.; Formal analysis: C.C.; Data curation: R.Y.; Writing—original draft preparation: R.Y. and B.W.; Writing—review and editing: C.C. and Y.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** No funding was received for conducting this study.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

**References**

1. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* 2006, 52, 1289–1306. [CrossRef]
2. Xie, Y.; Yu, J.; Guo, S.; Ding, Q.; Wang, E. Image Encryption Scheme with Compressed Sensing Based on New Three-Dimensional Chaotic System. *Entropy* 2019, 21, 819. [CrossRef]
3. Fang, Y.; Li, L.; Li, Y.; Peng, H.; Yang, Y. Low Energy Consumption Compressed Spectrum Sensing Based on Channel Energy Reconstruction in Cognitive Radio Network. *Sensors* 2020, 20, 1264. [CrossRef]
4. Martinez, J.A.; Ruiz, P.M.; Skarmeta, A.F. Evaluation of the Use of Compressed Sensing in Data Harvesting for Vehicular Sensor Networks. *Sensors* 2020, 20, 1434. [CrossRef]
5. Donoho, D.L.; Elad, M. Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization. *Proc. Natl. Acad. Sci. USA* 2003, 100, 2197–2202. [CrossRef]
6. Candes, E.J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* 2005, 51, 4203–4215. [CrossRef]
7. Stark, D.P.B. Uncertainty Principles and Signal Recovery. *SIAM J. Appl. Math.* 1989, 49, 906–931.
8. Elad, M. Optimized Projections for Compressed Sensing. *IEEE Trans. Signal Process.* 2007, 55, 5695–5702. [CrossRef]
9. Shaohai, H.U. An Optimization Method for Measurement Matrix Based on Eigenvalue Decomposition. *Signal Process.* 2012, 151, 45–55. [CrossRef]
10. Yan, W.; Wang, Q.; Shen, Y. Shrinkage-Based Alternating Projection Algorithm for Efficient Measurement Matrix Construction in Compressive Sensing. *IEEE Trans. Instrum. Meas.* 2014, 63, 1073–1084. [CrossRef]
11. 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]
12. Lu, C.; Li, H.; Lin, Z. Optimized Projections for Compressed Sensing via Direct Mutual Coherence Minimization. *Signal Process.* 2018, 151, 45–55. [CrossRef]
13. Xu, J.; Pi, Y.; Cao, Z. Optimized projection matrix for compressive sensing. *EURASIP J. Adv. Signal Process.* 2010, 2010, 560349. [CrossRef]
14. Abolghasemi, V.; Ferdowsi, S.; Sanei, S. A gradient-based alternating minimization approach for optimization of the measurement matrix in compressive sensing. *Signal Process.* 2012, 92, 999–1009. [CrossRef]
15. Zheng, H.; Li, Z.; Huang, Y. An Optimization Method for CS Projection Matrix Based on Quasi-Newton Method. *Acta Electron. Sin.* 2014, 42, 1977–1982.
16. Hong, T.; Bai, H.; Li, S.; Zhu, Z. An efficient algorithm for designing projection matrix in compressive sensing based on alternating optimization. *Signal Process.* 2016, 125, 9–20. [CrossRef]
17. Entezari, R.; Rashidi, A. Measurement matrix optimization based on incoherent unit norm tight frame. *AESI Int. J. Electron. Commun.* 2017, 82, 321–326. [CrossRef]
18. Sustik, M.A.; Tropp, J.A.; Dhillon, I.S.; Heath, R.W. On the existence of equiangular tight frames. *Linear Algebra Its Appl.* 2007, 426, 619–635. [CrossRef]
19. Welch, L. Lower bounds on the maximum cross correlation of signals (Corresp.). *IEEE Trans. Inf. Theory* 1974, 20, 397–399. [CrossRef]
20. Aharon, M.; Elad, M.; Bruckstein, A.M. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Trans. Signal Process.* 2006, 54, 4311–4322. [CrossRef]
21. Fickus, M.; Mixon, D.G. Tables of the existence of equiangular tight frames. *arXiv* 2015, arXiv:1504.00253.
22. Li, G.; Zhu, Z.; Yang, D.; Chang, L.; Bai, H. On Projection Matrix Optimization for Compressive Sensing Systems. *IEEE Trans. Signal Process.* 2013, 61, 2887–2898. [CrossRef]
23. Tropp, J.A.; Gilbert, A.C. Signal Recovery from Random Measurements Via Orthogonal Matching Pursuit. *IEEE Trans. Inf. Theory* 2007, 53, 4655–4666. [CrossRef]