An Efficient Optimization Algorithm for Measurement Matrix Based on SVD and Improved Nesterov Accelerated Gradient

Bo ZHANG, Renjie YI, Zhike WANG, Jianfeng PU, Yang SUN

Aviation University of Air Force, Nanhu Road 2222, Changchun City, China

yirenjie17@alumni.nudt.edu.cn, {bo_zhang0106, zhike7343, pujianfeng2019, sunyang87}@126.com

Submitted November 24, 2024 / Accepted February 28, 2025 / Online first April 30, 2025

Abstract. *In compressed sensing, a measurement matrix having low coherence with a specified sparse dictionary has been shown to be advantageous over a Gaussian random matrix in terms of reconstruction performance. In this paper the problem of efficiently designing the measurement matrix is addressed. The measurement matrix is designed by iteratively minimizing the difference between the Gram matrix of the sensing matrix and a target Gram matrix. A new target Gram matrix is designed by applying singular value decomposition to the sensing matrix and utilizing entry shrinking in the Gram matrix, leading to lower mutual coherence indicators. An improved Nesterov accelerated gradient algorithm is derived to update the measurement matrix, which can improve the convergence behavior. An efficient optimization algorithm for measurement matrix is proposed on the basis of alternating minimization. The experimental results and analysis show that the proposed algorithm performs well in terms of both computational complexity and reconstruction performance.*

Keywords

Compressed sensing, equiangular tight frame, singular value decomposition, mutual coherence, Nesterov accelerated gradient

1. Introduction

Compressed sensing (CS) [1] targets Shannon's theory and attempts to recover the sparse signal from a small number of linear measurements. Consider a signal $\boldsymbol{x} \in \mathbb{R}^N$, assuming that $\boldsymbol{x}$ has a sparse representation in a specified domain $\boldsymbol{\Psi} \in \mathbb{R}^{N \times L}$ $(N \le L)$ as $\boldsymbol{x} = \boldsymbol{\Psi s}$ where $\boldsymbol{s} \in \mathbb{R}^L$ is a sparse coefficient vector. $\boldsymbol{\Psi}$ is also called the sparse dictionary. $\boldsymbol{s}$ is called K-sparse if $\|\boldsymbol{s}\|_0 = K$ where $\|\boldsymbol{s}\|_0$ denotes the number of nonzero elements in $\boldsymbol{s}$. Via multiplication with measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ $(M \ll N)$, $\boldsymbol{x}$ is transformed into compressed measurements

$$\boldsymbol{y} = \boldsymbol{\Phi x} = \boldsymbol{Ds}\,. \tag{1}$$

where $\boldsymbol{D} = \boldsymbol{\Phi\Psi}$ is called the sensing matrix.

Recent studies have indicated that a measurement matrix having low coherence with a specified sparse dictionary can significantly improve the reconstruction performance. [2] considered a different mutual coherence [3], referred to as the t-averaged mutual coherence, which reflects average behavior of the sensing matrix $\boldsymbol{D}$. The large off-diagonal entries of the Gram matrix $\boldsymbol{G} = \boldsymbol{D}'\boldsymbol{D}$ were "shrinked" multiplying their values by $0 < \gamma < 1$, resulting in smaller μ_{ave} and higher reconstruction performance than those of Gaussian random matrix. [4] defined the global mutual coherence and utilized eigenvalue averaging to decrease μ_{all}. [5] attempted to find the sensing matrix $\boldsymbol{D}$ such that the Gramm matrix is as close to the identity as possible. The algorithmic concepts of the K-SVD [6] were applied in optimizing the measurement matrix. In [7], a method based on the equiangular tight frame (ETF) [8] design was proposed to minimize the coherence between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. The experimental results demonstrate that the optimized measurement matrix performs better than [2], [4] in terms of reconstruction performance.

Motivated by [7], [9–14] take advantage of the ETF to design the measurement matrix. In [9], [10], a gradient-based technique was applied to update the measurement matrix. In [11–13], analytical solutions of the measurement matrix are derived via singular value decomposition. An iterative method sharing the same concept as K-SVD was proposed in [14], which updates the measurement matrix row by row. The abovementioned algorithms focus on a single coherence metric, leading to less than satisfying progress in reconstruction performance. The work in [15] aims to reduce multiple coherence metrics simultaneously. The experimental results verify the effectiveness of the proposed algorithm in terms of reconstruction performance. However, the eigenvalue decomposition employed in [15] significantly increases the computational complexity. In [16], an iterative optimization algorithm was proposed based on the RIP to reduce mutual coherence. The algorithm performs slightly better than [9], [10] in both reconstruction performance and computational complexity.

DOI: 10.13164/re.2025.0234

Notably, the existing algorithm in [15] improves the reconstruction performance at the cost of computational complexity. To address this problem, an efficient optimization algorithm for the measurement matrix is proposed in this work. The main contributions of this study are as follows:

- To reduce multiple coherence metrics simultaneously, a new target Gram matrix is designed by applying SVD to the sensing matrix and utilizing entry shrinking in the Gram matrix without increasing complexity.
- An improved Nesterov accelerated gradient algorithm is derived to update the measurement matrix. By properly controlling the momentum term, the improved algorithm converges faster than the gradient in [9], [16].
- An iterative alternating minimization algorithm for the measurement matrix is proposed. The algorithm updates the target Gram matrix and measurement matrix alternatively. The experimental results verify the superiority of the proposed algorithm in improving convergence behavior and reconstruction performance.

The remainder of this paper is organized as follows. The problem formulation is introduced in Sec. 2. The proposed methods are presented in Sec. 3, where the approach for designing the target Gram matrix and the improved Nesterov accelerated gradient algorithm are described in detail. The procedure of the proposed algorithm can also be found in Sec. 3. Extensive experiments are carried out to verify the effectiveness of the proposed algorithm in Sec. 4, and finally, Section 5 concludes the paper.

2. Preliminaries and Problem Formulation

As mentioned above, a successful CS requires the measurement matrix $\boldsymbol{\Phi}$ and sparse dictionary $\boldsymbol{\Psi}$ to be incoherent. Good measures of coherence between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ (or, equivalently, columns of the sensing matrix $\boldsymbol{D}$) can be obtained by referring to the definitions of maximal mutual coherence, *t*-averaged mutual coherence and global mutual coherence.

Rewrite $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_L] \in \mathbb{R}^{M \times L}$ where $\boldsymbol{d}_i \in \mathbb{R}^M$ represents the ith column of $\boldsymbol{D}$ and where $\|\boldsymbol{d}_i\|_2 = 1$. Denote g_{ij} as the entry located in the ith row and jth column in $\boldsymbol{G} = \boldsymbol{D}'\boldsymbol{D}$. Clearly, $g_{ij} = \boldsymbol{d}_i'\boldsymbol{d}_j$ denotes the coherence between $\boldsymbol{d}_i$ and $\boldsymbol{d}_j$. Here, we quote these definitions as those presented by Elad [2], and Zhao [4].

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

$$\mu_{\max} = \max_{i \neq j} \left\{ \frac{\left|\boldsymbol{d}_i' \boldsymbol{d}_j\right|}{\left\|\boldsymbol{d}_i\right\|_2 \left\|\boldsymbol{d}_j\right\|_2} \right\} = \max_{i \neq j} \left|g_{ij}\right|. \tag{2}$$

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

$$\mu_{\text{ave}} = \frac{\sum_{i \neq j} \left(\left|g_{ij}\right| \geq t\right) \left|g_{ij}\right|}{\sum_{i \neq j} \left(\left|g_{ij}\right| \geq t\right)} \tag{3}$$

where $\left(\left|g_{ij} \geq t\right|\right)$ is an indicator function.

Definition 3. *For a matrix* $\boldsymbol{D}$*, the global mutual coherence* μ_{all} *is defined as the sum of squares of all off-diagonal entries in the Gram matrix of* $\boldsymbol{D}$*, denoted as*

$$\mu_{\text{all}} = \sum_{i \neq j} g_{ij}^2 . \tag{4}$$

The above indicators measure coherence in different ways. Generally, the lower the values of these indicators are, the better the reconstruction performance. Therefore, existing measurement matrix optimization algorithms attempt to reduce these indicators to improve performance. As mentioned in the previous section, the progress in reconstruction performance resulting from reducing a single indicator is significantly less than that of reducing multiple indicators simultaneously. Inspired by this, we consider reducing $\mu_{\max}$, μ_{ave}, and μ_{all} simultaneously in this work.

According to the above definitions, $\mu_{\max}$, μ_{ave} and μ_{all} are related to the off-diagonal entries of $\boldsymbol{G}$. To reduce these indicators, an intuitive way is to assign minor values to g_{ij} and then solve $\boldsymbol{G} = \boldsymbol{\Psi}'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{\Psi}$. Notably, a change in g_{ij} results in $\boldsymbol{G}$ being full rank in general, i.e., $\text{Rank}(\boldsymbol{G}) = L$, where $\text{Rank}(\cdot)$ returns the rank of the matrix. However, $\text{Rank}(\boldsymbol{\Psi}'\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{\Psi}) = M < L$, indicating that the closed-form solution of $\boldsymbol{\Phi}$ is unachievable. Furthermore, $|g_{ij}|$ cannot be arbitrarily lowered since $\mu_{\max} = \max_{i \neq j} \left|g_{ij}\right|$ is bounded by [8]

$$\mu_{\max} \geq \mu_{\text{welch}} = \sqrt{\frac{L - M}{M(L - 1)}} \tag{5}$$

where the bound is achievable only when $\boldsymbol{D}$ is an ETF.

To address this problem, Xu et al. proposed a framework for optimizing a measurement matrix in [7]. The main idea is to design a 'good' Gram matrix first and then find the solution $\boldsymbol{\Phi}$ such that $\boldsymbol{G} = (\boldsymbol{\Phi}\boldsymbol{\Psi})'\boldsymbol{\Phi}\boldsymbol{\Psi}$ is as close as possible to such a designed Gram matrix. Generally, the designed Gram matrix is called the target Gram matrix and is denoted by $\boldsymbol{G}_{\text{t}}$. In [7], $\boldsymbol{G}_{\text{t}}$ is designed as the Gram matrix of a set of relaxed ETFs, i.e., $\boldsymbol{G}_{\text{t}} \in \Omega = \left\{ \boldsymbol{Z} \in \mathbb{R}^{L \times L} : \boldsymbol{Z} = \boldsymbol{Z}', \boldsymbol{Z}(i,i) = 1, \max_{i \neq j} \left|\boldsymbol{Z}(i,j)\right| \leq \mu_{\text{welch}} \right\}$. The optimal measurement matrix is investigated by solving:

$$\min_{\boldsymbol{G}_t \in \Omega} f\left(\boldsymbol{\Phi}, \boldsymbol{G}_t\right) = \min_{\boldsymbol{G}_t \in \Omega} \left\| \left(\boldsymbol{\Phi}\boldsymbol{\Psi}\right)' \boldsymbol{\Phi}\boldsymbol{\Psi} - \boldsymbol{G}_t \right\|_F^2 \tag{6}$$

The minimization problem can be solved by alternating projection, which iteratively minimizes (6) to find the optimal $\boldsymbol{\Phi}$. The idea is to update $\boldsymbol{G}_t$ and $\boldsymbol{\Phi}$ alternatively and repeat this process for several iterations. Under this framework, designing $\boldsymbol{G}_t$ and updating $\boldsymbol{\Phi}$ have been considered in many works. A shrinkage method has been used for the Gram matrix $\boldsymbol{G}$ to design $\boldsymbol{G}_t$ in [9–15], whereas [15] differs from the others in that the former employed eigenvalue decomposition. With respect to updating $\boldsymbol{\Phi}$, [9], [10] applied gradient-based methods, and [11–15] used matrix decomposition, such as eigenvalue decomposition (ED) and QR decomposition. Notably, [15] employed eigenvalue decomposition to design $\boldsymbol{G}_t$ and update $\boldsymbol{\Phi}$ in the proposed JOAM algorithm, resulting in high complexity. The JOAM is introduced in Tab. 1.

Patel and Vaish [16] proposed a different framework for optimizing a measurement matrix based on restricted isometric property (RIP) [17], which can be described as

$$\min \left\| \boldsymbol{\Phi}\boldsymbol{\Psi} - \boldsymbol{D}_t \right\|_F^2 \tag{7}$$

where $\boldsymbol{D}_t \in \Gamma = \left\{ \boldsymbol{E} \in \mathbb{R}^{M \times L} : \left[\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V}\right] = \mathrm{svd}\left(\boldsymbol{E}\right), \boldsymbol{S}\left(i,i\right) = 1 \right\}$. The authors used singular value decomposition (SVD) to design a 'good' sensing matrix $\boldsymbol{D}_t$ and employed the gradient descent method to update $\boldsymbol{\Phi}$. The minimization problem shown in (7) is simpler than that shown in (6) because the former does not involve the operation of the Gram matrix. However, the proposed algorithm presented in [16] is far inferior to JOAM in terms of reconstruction performance. The reason behind this is that the off-diagonal entries of the Gram matrix $\boldsymbol{G}$ directly reflect the mutual coherence of $\boldsymbol{D}$, and those entries can be reduced when (6) is efficiently solved with a good target Gram matrix $\boldsymbol{G}_t$.

According to these observations, we consider designing a new measurement matrix optimization algorithm, which can yield comparable reconstruction performance to that of JOAM with much lower complexity.

Input: The dictionary matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N \times L}$, the number of iterations $Iter$.

Output: Sparse measurement matrix $\boldsymbol{\Phi}$.

Initialization: Initialize $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ to a random matrix, number of iterations $k = 0$, mutual coherence threshold μ_0, threshold ε.

While $q < Iter$ and $\left|\mu_{\mathrm{ave}}\left(t+1\right) - \mu_{\mathrm{ave}}\left(t\right)\right| > \varepsilon$ do

1. Update $q = q + 1$.
2. Calculate $\boldsymbol{D} = \boldsymbol{\Phi}\boldsymbol{\Psi}$ and normalize the columns of $\boldsymbol{D}$.
3. Calculate $\boldsymbol{G} = \boldsymbol{D}'\boldsymbol{D}$ and Apply ED to $\boldsymbol{G}$ as $\boldsymbol{G} = \boldsymbol{V}\boldsymbol{H}\boldsymbol{V}'$.
4. Set the positive eigenvalues in $\boldsymbol{H}$ as L/M to obtain $\hat{\boldsymbol{H}}$ and further obtain $\boldsymbol{G}_t = \boldsymbol{V}\hat{\boldsymbol{H}}\boldsymbol{V}'$.

While $k < I_{\max}$ and $\mu_{\max} > \mu_0$ do

5. Update $k = k + 1$.
6. Apply entry shrinking to $\boldsymbol{G}_t$ as follows

$$\boldsymbol{G}_t\left(i,j\right) = \begin{cases} \boldsymbol{G}_t\left(i,j\right), & \left|\boldsymbol{G}_t\left(i,j\right)\right| \le \mu_0, i \ne j \\ 1, & i = j \\ \mathrm{sign}\left(\boldsymbol{G}_t\left(i,j\right)\right)\mu_0, & \text{otherwise} \end{cases}$$

7. Apply ED to $\boldsymbol{G}_t$ as $\boldsymbol{G}_t = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$ where the eigenvalues in $\boldsymbol{\Lambda}$ are arranged in descending order.
8. Keep only the M largest positive eigenvalues in $\boldsymbol{\Lambda}$ to obtain $\hat{\boldsymbol{\Lambda}}$, and further obtain $\boldsymbol{G}_t = \boldsymbol{U}\hat{\boldsymbol{\Lambda}}\boldsymbol{U}'$.

end while

9. Calculate the sensing matrix $\boldsymbol{D} = \hat{\boldsymbol{\Lambda}}^{\frac{1}{2}}\boldsymbol{U}'$

end while

Return the measurement matrix according to $\boldsymbol{D} = \boldsymbol{\Phi}\boldsymbol{\Psi}$

Tab. 1. JOAM algorithm.

3. Preliminaries and Problem Formulation

In this section, we target solving the minimization problem shown in (6) on the basis of alternating minimization. As to be seen, a new target Gram matrix is designed first, and then an improved Nesterov accelerated gradient algorithm is derived to update the measurement matrix. Based on the above, we propose an efficient measurement matrix optimization algorithm.

3.1 Designing $\boldsymbol{G}_t$

We attempt to find the nearest solution of $\boldsymbol{G}$ by solving (6). In should be noted that the cost function $f(\boldsymbol{\Phi}, \boldsymbol{G}_t) > 0$ holds for most cases since $\mathrm{Rank}(\boldsymbol{G}) \ne \mathrm{Rank}(\boldsymbol{G}_t)$. When the minimization problem is efficiently solved, i.e., $f(\boldsymbol{\Phi}, \boldsymbol{G}_t)$ is small enough, g_{ij} will be as close as possible to $g_{t_{ij}}$, where $g_{t_{ij}}$ denotes the entry located in the ith row and jth column in $\boldsymbol{G}_t$. Therefore, $\boldsymbol{G}_t$ plays an important role in reducing the mutual coherence metrics. In the following, we design a new target Gram matrix, that can reduce $\mu_{\max}$, μ_{ave}, and μ_{all} simultaneously.

Starting with a given $\boldsymbol{\Phi}$ and a specified dictionary $\boldsymbol{\Psi}$, we normalize the columns of $\boldsymbol{D} = \boldsymbol{\Phi}\boldsymbol{\Psi}$. Assume that the SVD of $\boldsymbol{D}$ can be written as $\boldsymbol{D} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}'$, where $\boldsymbol{U} \in \mathbb{R}^{M \times M}$ and $\boldsymbol{V} \in \mathbb{R}^{L \times L}$ are unitary matrices, and $\boldsymbol{S} = [\boldsymbol{\Sigma}\ \boldsymbol{0}] \in \mathbb{R}^{M \times L}$ with $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_M)$. σ_i denotes the nonzero singular values of $\boldsymbol{D}$. According to Definition 3, we have

$$\mu_{\mathrm{all}} = \sum_{i \ne j} g_{ij}^2 = \sum_{i,j=1}^{L} g_{ij}^2 - \sum_{i=1}^{L} g_{ii}^2. \tag{8}$$

In (8), $\sum_{i,j=1}^{L} g_{ii}^2 = L$ since $g_{ii} = 1$. $\sum_{i,j=1}^{L} g_{ij}^2$ represents the square of the Frobenius norm of $\boldsymbol{G}$. With simple calculation, we have that

$$\sum_{i,j=1}^{L} g_{ij}^2 = \left\|\boldsymbol{G}\right\|_F^2 = \left\|\boldsymbol{V}\boldsymbol{S}'\boldsymbol{S}\boldsymbol{V}'\right\|_F^2 \overset{(a)}{=} \left\|\boldsymbol{S}'\boldsymbol{S}\right\|_F^2 = \sum_{i=1}^{M} \left(\sigma_i^2\right)^2 \tag{9}$$

where quality (a) follows from unitary invariance. Thus, (8) can be rewritten as

$$\mu_{\text{all}} = \sum_{i=1}^{M}\left(\sigma_i^2\right)^2 - L. \tag{10}$$

Recall that a lower μ_{all} can yield a better CS result, and minimizing μ_{all} is highly important. In (10), μ_{all} reaches the minimum when $\sum_{i=1}^{M}\left(\sigma_i^2\right)^2$ takes the minimum value. Note that σ_i^2 represents the eigenvalue of $\boldsymbol{G}$ and $\sum_{i=1}^{M}\sigma_i^2 = \sum_{i=1}^{L} g_{ii} = L$. Then, it follows from the Cauchy-Schwarz inequality that $\sum_{i=1}^{M}\left(\sigma_i^2\right)^2$ reaches the minimum only if $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_M^2 = L/M$.

Based on the above analysis, a new target Gram matrix that focuses on reducing μ_{all} is formulated as follows

$$\boldsymbol{G}_{\text{t}} = \boldsymbol{V}\hat{\boldsymbol{S}}'\hat{\boldsymbol{S}}\boldsymbol{V}' \tag{11}$$

where $\hat{\boldsymbol{S}} = \begin{bmatrix}\hat{\boldsymbol{\Sigma}} & \boldsymbol{0}\end{bmatrix}$ with $\hat{\boldsymbol{\Sigma}} = \text{diag}\left(\underbrace{\sqrt{L/M}, \sqrt{L/M}\cdots\sqrt{L/M}}_{M}\right)$.

Notably, the sum of squares of all entries in $\boldsymbol{G}_{\text{t}}$ reaches the minimum under the assumption that the sum of its diagonal entries is equal to L. However, the sum of squares of all diagonal entries in $\boldsymbol{G}_{\text{t}}$ may not take the fixed value L since diagonal entries have changed with $\hat{\boldsymbol{S}}$. This implies that the sum of squares of all off-diagonal entries in $\boldsymbol{G}_{\text{t}}$ may not achieve the minimum. Considering that the off-diagonal elements are much greater in number than the diagonal elements are, the sum of squares of all off-diagonal entries of $\boldsymbol{G}_{\text{t}}$ will fluctuate around the minimum value $L^2/M - L$.

Recall that our goal is to reduce $\mu_{\max}$, μ_{ave}, and μ_{all} simultaneously. In the following, we utilize a shrinkage method to decrease the large absolute value entries of $\boldsymbol{G}_{\text{t}}$ shown in (11). In [2], large absolute off-diagonal entries are "shrinked" multiplying their values by $0 < \gamma < 1$. Compared with the original measurement matrix, the optimized measurement matrix has obvious advantages in both reducing μ_{ave} and the reconstruction error. In [7], a more efficient shrinkage method was proposed based on ETF. The off-diagonal entries whose absolute value is greater than μ_{welch} are forcibly assigned the value $\pm\mu_{\text{welch}}$. As a result, the absolute value of off-diagonal entries is upper bounded by μ_{welch}, and $\mu_{\max}$ sharply decreases compared with that in [2]. In [9], [11], the authors considered different upper limits for the off-diagonal entries of the target Gram matrix. Although different upper limits yield different CS results, μ_{welch} is empirically a suitable choice in terms of reconstruction performance.

Based on the above analysis, we shrink the off-diagonal entries of the target Gram matrix shown in (11) via the shrinkage method proposed in [7], which can be described as follows

$$g_{\text{t}_{ij}} = \begin{cases} g_{\text{t}_{ij}}, & \left|g_{\text{t}_{ij}}\right| \le \mu_{\text{welch}}, i \ne j \\ 1, & i = j \\ \text{sign}\left(g_{\text{t}_{ij}}\right)\mu_{\text{welch}}, & \text{otherwise} \end{cases} \tag{12}$$

where sign(·) denotes the sign function.

By utilizing the shrinking operation for $\boldsymbol{G}_{\text{t}}$, the off-diagonal entries with large absolute values will be intensively constrained, therefore the average of all absolute off-diagonal entries above t will decrease significantly. Since the off-diagonal entries with absolute values below μ_{welch} remain unchanged, and these entries make up the majority of all off-diagonal entries, there is a certain similarity between the matrices before and after entry shrinking, which means that $\boldsymbol{G}_{\text{t}}$ obtained from (12) maintains the ability to reduce μ_{all}.

According to the above analysis, the target Gram matrix obtained from (11–12), can reduce $\mu_{\max}$, μ_{ave}, and μ_{all} simultaneously. In addition, we perform SVD on the sensing matrix $\boldsymbol{D}$ instead of performing ED on its Gram matrix, resulting in lower computational load. In the following subsection, we keep $\boldsymbol{G}_{\text{t}}$ fixed and update $\boldsymbol{\Phi}$ to minimize (6).

3.2 Updating Φ

The steepest descent is widely used in measurement matrix optimization. Recent works [18], [19] have shown that gradient-based algorithms can solve the minimization problem shown in (6). However, the gradient descent methods used in [8, 9, 16] require a large number of iterations to reach convergence, resulting in a great computational load. To address this problem, we propose an improved NAG algorithm to update the measurement matrix.

We simplify cost function $f(\boldsymbol{\Phi}, \boldsymbol{G}_{\text{t}}) = \left\|(\boldsymbol{\Phi\Psi})'\boldsymbol{\Phi\Psi} - \boldsymbol{G}_{\text{t}}\right\|_{\text{F}}$ as $f(\boldsymbol{\Phi})$. With simple calculations, the gradient of $f(\boldsymbol{\Phi})$ with respect to $\boldsymbol{\Phi}$ can be described as

$$\nabla f(\boldsymbol{\Phi}) = 4\boldsymbol{\Phi\Psi}(\boldsymbol{\Phi\Psi})'\boldsymbol{\Phi\Psi\Psi}' - 4\boldsymbol{\Phi\Psi G}_{\text{t}}\boldsymbol{\Psi}'. \tag{13}$$

The full update equation in the steepest descent algorithm is written as

$$\boldsymbol{\Phi}_{k+1} = \boldsymbol{\Phi}_k - \alpha\nabla f(\boldsymbol{\Phi}_k) \tag{14}$$

where k is the iteration index and where α is the step size. Unlike the steepest descent, NAG [20] adds a momentum term and looks ahead the approximate future position. The full update equations of NAG are formulated as

$$\boldsymbol{V}_{k+1} = \gamma\boldsymbol{V}_k + \alpha\nabla f(\boldsymbol{\Phi}_k - \gamma\boldsymbol{V}_k), \tag{15}$$

$$\boldsymbol{\Phi}_{k+1} = \boldsymbol{\Phi}_k - \boldsymbol{V}_{k+1} \tag{16}$$

where γ is the momentum parameter and is set to approximately 0.9. The momentum term increases updates for dimensions whose gradients point in the same direction.

Computing $\boldsymbol{\Phi}_k - \gamma \boldsymbol{V}_k$ provides an approximation of the next position of the parameters.

Theoretically, the value of the cost function decreases with the increasing iteration index provided that the step size is appropriate. In most existing methods, the step size is set as a fixed value for convenience. However, this approach may lead to an increase in the value of the cost function, which means that $f(\boldsymbol{\Phi}_k) > f(\boldsymbol{\Phi}_{k-1})$ may occur at a specified iteration index.

Assume that $f(\boldsymbol{\Phi})$ is smooth at $\boldsymbol{\Phi}_k$. We consider approximating $f(\boldsymbol{\Phi}_{k-1})$ and $f(\boldsymbol{\Phi}_{k+1})$ with a Taylor series approximation. Notably, higher-order terms of the Taylor series cause greater computational load but contribute less to the approximation. Here, we use a truncated Taylor series to approximate the cost function, which can be described as

$$f(\boldsymbol{\Phi}_{k-1}) \approx f(\boldsymbol{\Phi}_k) + \left[\frac{\partial f(\boldsymbol{\Phi}_k)}{\partial \operatorname{vec}(\boldsymbol{\Phi}_k)}\right]' \operatorname{vec}(\boldsymbol{\Phi}_{k-1} - \boldsymbol{\Phi}_k), \quad (17)$$

$$f(\boldsymbol{\Phi}_{k+1}) \approx f(\boldsymbol{\Phi}_k) + \left[\frac{\partial f(\boldsymbol{\Phi}_k)}{\partial \operatorname{vec}(\boldsymbol{\Phi}_k)}\right]' \operatorname{vec}(\boldsymbol{\Phi}_{k+1} - \boldsymbol{\Phi}_k) \quad (18)$$

where vec(·) denotes the vectorization function.

If $f(\boldsymbol{\Phi}_k) > f(\boldsymbol{\Phi}_{k-1})$ holds, then it follows from (17) that

$$\left[\frac{\partial f(\boldsymbol{\Phi}_k)}{\partial \operatorname{vec}(\boldsymbol{\Phi}_k)}\right]' \operatorname{vec}(\boldsymbol{\Phi}_{k-1} - \boldsymbol{\Phi}_k) < 0 \quad (19)$$

holds approximately. According to (16), we have $\boldsymbol{V}_k = \boldsymbol{\Phi}_{k-1} - \boldsymbol{\Phi}_k$ and then (19) can be rewritten as

$$\left[\frac{\partial f(\boldsymbol{\Phi}_k)}{\partial \operatorname{vec}(\boldsymbol{\Phi}_k)}\right]' \operatorname{vec}(\boldsymbol{V}_k) < 0. \quad (20)$$

Following (15) and (16), we can rewrite (18) as

$$\begin{aligned} f(\boldsymbol{\Phi}_{k+1}) &\approx f(\boldsymbol{\Phi}_k) - \left[\frac{\partial f(\boldsymbol{\Phi}_k)}{\partial \operatorname{vec}(\boldsymbol{\Phi}_k)}\right]' \operatorname{vec}\left(\gamma \boldsymbol{V}_k + \alpha \nabla f(\boldsymbol{\Phi}_k - \gamma \boldsymbol{V}_k)\right) \\ &= f(\boldsymbol{\Phi}_k) - \gamma \left[\frac{\partial f(\boldsymbol{\Phi}_k)}{\partial \operatorname{vec}(\boldsymbol{\Phi}_k)}\right]' \operatorname{vec}(\boldsymbol{V}_k) \\ &\quad - \left[\frac{\partial f(\boldsymbol{\Phi}_k)}{\partial \operatorname{vec}(\boldsymbol{\Phi}_k)}\right]' \operatorname{vec}\left(\alpha \nabla f(\boldsymbol{\Phi}_k - \gamma \boldsymbol{V}_k)\right) \end{aligned} \quad (21)$$

which, together with (20), implies that the momentum term $\boldsymbol{V}_k$ may lead to an increase in the value of the cost function $f(\boldsymbol{\Phi}_{k+1})$. Such an increase is unexpected in solving the minimization problem since more iterations are needed to reach convergence.

However, this increase can be transformed into a decrease through multiplication of the momentum term by an adaptive factor β, which can be formulated as

$$\boldsymbol{V}_{k+1} = \beta_k \gamma \boldsymbol{V}_k + \alpha \nabla f(\boldsymbol{\Phi}_k - \gamma \boldsymbol{V}_k) \quad (22)$$

where the value of β_k is taken as

$$\beta_k = \begin{cases} 1 & f(\boldsymbol{\Phi}_{k-1}) > f(\boldsymbol{\Phi}_k) \\ -1 & f(\boldsymbol{\Phi}_{k-1}) \le f(\boldsymbol{\Phi}_k) \end{cases} \quad (23)$$

If $f(\boldsymbol{\Phi}_k) \ge f(\boldsymbol{\Phi}_{k-1})$ holds, then β_k is set to –1 to prevent the value of the cost function from increasing. Moreover, $\boldsymbol{V}_{k+1}$ is no longer the past time step since the momentum term $\gamma \boldsymbol{V}_k$ is forced to change its direction, which implies that $\boldsymbol{V}_{k+1}$ may cause oscillation in the $(k+2)$th iteration. Therefore, after the $(k+1)$th iteration, we reset $\boldsymbol{V}_{k+1}$ as follows:

$$\boldsymbol{V}_{k+1} = \alpha \nabla f(\boldsymbol{\Phi}_k - \gamma \boldsymbol{V}_k) \quad (24)$$

where the changed momentum term $-\gamma \boldsymbol{V}_k$ is removed. As a result, this improved NAG can immediately correct its course and converge faster than NAG.

Based on alternating minimization, the proposed algorithm is summarized in Tab. 2. The proposed algorithm is an iterative method. The main complexity is located in steps 2, 3, 4, 9 and 11. In those steps, the flops required are $O(MNL + ML)$, $O(ML^2)$, $O(ML^2)$, $O(MNL + 2M^2L + ML^2)$ and $O(ML + ML^2)$. Since $M \ll N \le L$, the complexity of our proposed algorithm is approximately equal to $O(Iter \cdot I_{\max} \cdot ML^2)$. The complexities of JOAM and the meth-

Input: The dictionary matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N \times L}$, the number of iterations $Iter$, $I_{\max}$, the step size α, the momentum parameter γ.

Output: measurement matrix $\boldsymbol{\Phi}$.

Initialization: Initialize $\boldsymbol{\Phi}_0 \in \mathbb{R}^{M \times N}$ to a random matrix, number of iterations $k = q = 0$, adaptive factor $\beta_0 = \beta_1 = 1$, momentum term $\boldsymbol{V}_0 = 0$.

While $q < Iter$ do

1. Update $q = q + 1$.
2. Calculate $\boldsymbol{D} = \boldsymbol{\Phi}\boldsymbol{\Psi}$ and normalize the columns of $\boldsymbol{D}$.
3. Apply SVD to $\boldsymbol{D}$ as $\boldsymbol{D} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}'$, and set the positive singular values in $\boldsymbol{S}$ as $\sqrt{L/M}$ to obtain $\hat{\boldsymbol{S}}$.
4. Calculate the target Gram matrix $\boldsymbol{G}_t = \boldsymbol{V}\hat{\boldsymbol{S}}'\hat{\boldsymbol{S}}\boldsymbol{V}'$.
5. Apply entry shrinking to $\boldsymbol{G}_t$ as follows

$$\boldsymbol{G}_t(i,j) = \begin{cases} \boldsymbol{G}_t(i,j), & |\boldsymbol{G}_t(i,j)| \le \mu_{\text{welch}}, i \ne j \\ 1, & i = j \\ \operatorname{sign}(\boldsymbol{G}_t(i,j))\,\mu_{\text{welch}}, & \text{otherwise} \end{cases}$$

While $k < I_{\max}$ do

6. Update $k = k + 1$.
7. If $k \ge 3$, go to step 9. Otherwise, go to step 10.
8. If $f(\boldsymbol{\Phi}_{k-2}) < f(\boldsymbol{\Phi}_{k-1})$, update $\beta_{k-1} = -1$, else update $\beta_{k-1} = 1$.
9. Calculate the momentum term $\boldsymbol{V}_k = \beta_{k-1}\gamma \boldsymbol{V}_{k-1} + \alpha \nabla f(\boldsymbol{\Phi}_{k-1} - \gamma \boldsymbol{V}_{k-1})$.
10. Update the measurement matrix $\boldsymbol{\Phi}_k = \boldsymbol{\Phi}_{k-1} - \boldsymbol{V}_k$.
11. Calculate the cost function value $f(\boldsymbol{\Phi}_k)$.
12. If $f(\boldsymbol{\Phi}_{k-2}) < f(\boldsymbol{\Phi}_{k-1})$, update $\boldsymbol{V}_k = \alpha \nabla f(\boldsymbol{\Phi}_{k-1} - \gamma \boldsymbol{V}_{k-1})$.

end while

end while

return $\boldsymbol{\Phi}_k$

Tab. 2. The proposed algorithm.

od introduced in [16] are approximately equal to $O(Iter \cdot I_{\max} \cdot L^3)$ and $O(Iter \cdot I_{\max} \cdot MNL)$, respectively. Clearly, the complexity of our algorithm is much less than that of JOAM and slightly greater than that of [16].

4. The Simulation Experiments

In this section, we conduct several experiments to examine the performance of the proposed algorithm. The proposed algorithm is compared with the methods presented in [9], [15], and [16], which is denoted as Abolghasemi, JOAM, and Patel, respectively. The Gaussian random measurement matrix is denoted as Gaussian. In Abolghasemi, Patel and the proposed algorithm, the step size is set to 0.01. The iteration numbers are set as $Iter = 100$ and $I_{\max} = 50$. In the definition of μ_{ave}, t is set to μ_{welch}. In the following experiments (except Subsection 4.3.2), an $N \times L$ matrix with normally distributed elements is used as the dictionary matrix $\boldsymbol{\Psi}$, where $N = L = 120$.

4.1 Comparing the Coherence Metrices

In this subsection, experiments are carried out to verify the effectiveness of the proposed algorithm in reducing mutual coherence.

Figure 1 presents the frequency histogram of g_{ij} $(i \neq j)$ falling into different intervals. The Gaussian and Patel distributions have long tails, with maximal values exceeding 0.6 and 0.48, respectively. A longer tail implies higher $\mu_{\max}$ and μ_{ave}. Moreover, the distributions of Abolghasemi, JOAM, and the proposed algorithm are relatively concentrated in the interval (0.1, 0.2), indicating lower $\mu_{\max}$ and μ_{ave}. In the following, we conduct several experiments by varying M from 20 to 50. The results averaged from 500 independent experiments are recorded in Tab. 3–5.

Table 3 shows that JOAM performs the best in reducing $\mu_{\max}$, followed by the proposed algorithm, Abolghasemi,

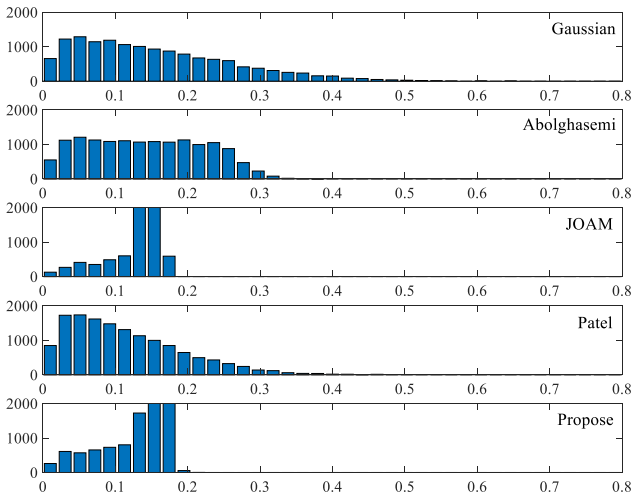


Fig. 1. Histogram of the absolute off-diagonal values of $\boldsymbol{G}$ while $M = 40$.

M	Gaussian	Abolghasemi	JOAM	Patel	Propose	μ_{welch}
20	0.7770	0.5837	**0.3427**	0.7189	0.3620	0.2050
25	0.7251	0.4978	**0.2804**	0.6528	0.2972	0.1787
30	0.6926	0.4355	**0.2332**	0.5945	0.2507	0.1588
35	0.6595	0.3884	**0.1995**	0.5474	0.2154	0.1429
40	0.6328	0.3487	**0.1734**	0.5012	0.1891	0.1296
45	0.6116	0.3197	**0.1526**	0.4630	0.1681	0.1183
50	0.5933	0.2928	**0.1356**	0.4285	0.1511	0.1086

Tab. 3. $\mu_{\max}$ by different algorithms versus M.

M	Gaussian	Abolghasemi	JOAM	Patel	Propose	μ_{welch}
20	0.3332	0.3135	**0.2377**	0.3110	0.2441	0.2050
25	0.2980	0.2739	**0.2031**	0.2719	0.2089	0.1787
30	0.2718	0.2445	**0.1773**	0.2421	0.1830	0.1588
35	0.2512	0.2211	**0.1573**	0.2181	0.1628	0.1429
40	0.2348	0.2020	**0.1412**	0.1982	0.1465	0.1296
45	0.2199	0.1860	**0.1276**	0.1811	0.1329	0.1183
50	0.2085	0.1723	**0.1160**	0.1661	0.1213	0.1086

Tab. 4. μ_{ave} by different algorithms versus M.

M	Gaussian	Abolghasemi	JOAM	Patel	Propose	$L^2/M-L$
20	816.83	757.07	613.97	608.83	**600.33**	600
25	677.10	593.65	467.17	462.92	**456.32**	456
30	583.46	484.74	370.06	365.54	**360.30**	360
35	517.12	405.29	300.59	296.03	**291.73**	291.43
40	470.24	345.03	248.39	243.81	**240.29**	240
45	426.71	298.81	207.67	203.24	**200.28**	200
50	398.67	261.20	174.93	170.69	**168.27**	168

Tab. 5. μ_{all} by different algorithms versus M.

Patel, and Gaussian. Moreover, the values listed below the proposed algorithm are close to those listed by JOAM and significantly lower than those listed by Abolghasemi and Patel. Similar results can be observed by inspecting Tab. 4.

Table 5 shows that the proposed algorithm performs the best in reducing μ_{all}, followed by Patel, JOAM, Abolghasemi, and Gaussian. The values listed below the proposed algorithm are very close to the minimal values listed in the last column, which coincides with our theoretical analysis presented in Sec. 3.1. In summary, the proposed algorithm is advantageous over the other algorithms in reducing μ_{ave}, $\mu_{\max}$ and μ_{all} simultaneously.

4.2 Comparing the Convergence

In this subsection, we conduct an experiment to compare the convergence of the improved NAG algorithm with that of NAG algorithm and the gradient-based algorithm used in [9, 10, 16], which is referred to as the Gradient. Based on the same framework shown in Tab. 2, we utilize the above three algorithms to update the measurement matrix and record the cost function values at different iterations.

Figure 2 illustrates cost function value $f(\boldsymbol{\Phi}_k, \boldsymbol{G}_{t_k})$ as a function of iteration. Overall, all the lines show a down-

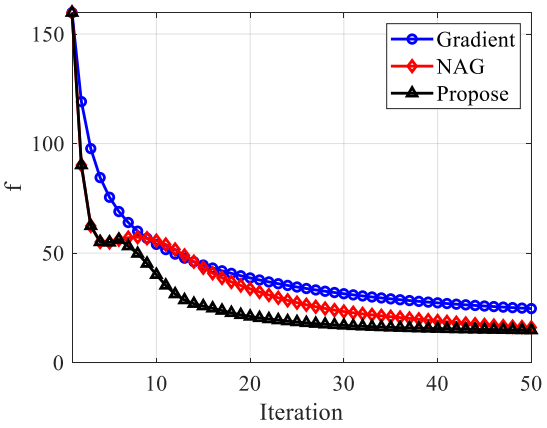


Fig. 2. The change tendency of cost function value with iteration numbers while $M = 40$.

ward trend as the number of iterations increases, which implies that the above three algorithms converge. The NAG and the improved NAG increase at the sixth iteration. However, NAG continues increasing until the eighth iteration whereas the improved NAG immediately decreases at the next iteration. Moreover, the improved NAG leads to a smaller steady state $f(\boldsymbol{\Phi}_k, \boldsymbol{G}_{t_k})$ and faster convergence.

In the proposed algorithm, we repeat steps 6–12 until a fixed number of iterations is reached. It should be noted that the stopping criteria can be utilized. For example, we may continue the algorithm until the cost function value for two successive iterations does not change significantly or until reaching to a desired value. Then, the proposed algorithm can considerably reduce the duration of the optimization run and reduce the computational load.

4.3 Comparing the Reconstruction Performance

4.3.1 Comparison of One-dimensional Signals

In this subsection, we conduct several experiments to verify the superiority of the proposed algorithm over the existing algorithms in terms of reconstruction performance. The original one-dimensional signal has a sparse representation as $\boldsymbol{x} = \boldsymbol{\Psi s}$ where $\boldsymbol{s} \in \mathbb{R}^L$ is K-sparse, and each nonzero entry of $\boldsymbol{s}$ is randomly positioned. The length of $\boldsymbol{s}$ is set as $L = 120$. The orthogonal matching pursuit (OMP) [21] algorithm is employed in reconstruction. Let $\sigma = \|\boldsymbol{s}_e - \boldsymbol{s}\|_2 / \|\boldsymbol{s}\|_2$ denote the reconstruction error, where $\boldsymbol{s}_e$ is the reconstructed signal of $\boldsymbol{s}$. A reconstruction is considered successful if $\sigma \leq 10^{-6}$ holds. P_{suc} denotes the percentage of successful reconstructions. Each experiment is performed for 500 random sparse ensembles.

Case 1: Comparison of P_{suc} in the noiseless case.

In this case, we conduct two separate experiments, first by fixing $K = 8$ and varying the compression ratio (CR) from 16.7% ($M = 20$) to 41.7% ($M = 50$) and second by fixing $M/N = 33.3\%$ ($M = 40$) and varying K from 4 to 20.

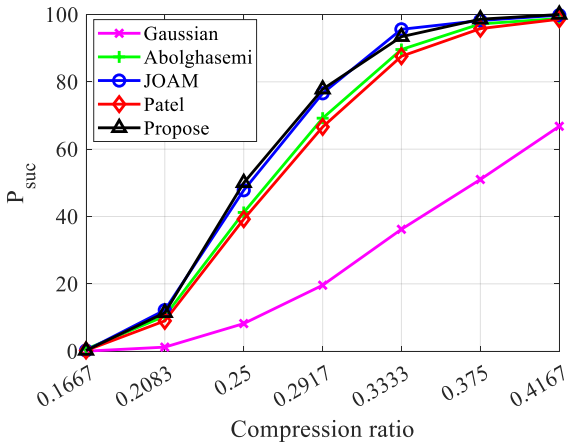


Fig. 3. The change tendency of P_{suc} with CR while $K = 12$.

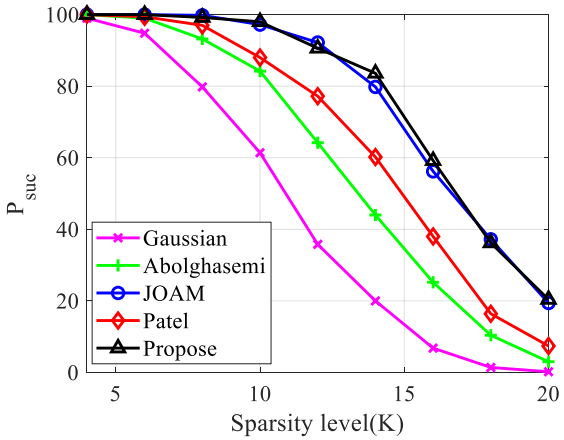


Fig. 4. The change tendency of P_{suc} with K while CR = 33.3%.

The results of the first experiment are shown in Fig. 3. We observe that P_{suc} increases with increasing CR. The achieved P_{suc} when using the proposed algorithm is comparable to those when using JOAM, and significantly higher than those when using Abolghasemi and Patel. Similar superiority can be found by inspecting Fig. 4, where the change tendency of P_{suc} with K is depicted.

Case 2: Comparison of σ in the noisy case.

In the noisy case, the linear model can be written as $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{v}$ where $\boldsymbol{v}$ is the additive Gaussian noise. In this experiment, the parameters are $M = 40$ and $K = 8$. Since successful reconstruction is usually unachievable in the presence of noise, we utilize the reconstruction error to evaluate the performance. The change in σ with varying SNR is depicted in Fig. 5. As seen, the proposed algorithm leads to a comparable σ to JOAM and performs much better than Abolghasemi and Patel in terms of reconstruction accuracy.

Figures 6 and 7 show σ as function of M and K, respectively. From Fig. 3 and Fig. 6, we observe that the proposed algorithm performs better than Abolghasemi and Patel in both noisy and noise-free cases. This shows that the proposed algorithm is robust in noisy cases.

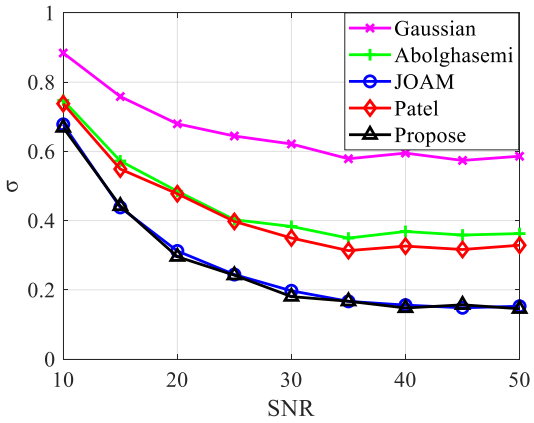


Fig. 5. The change tendency of σ with SNR (dB) while CR = 25% and K = 10.

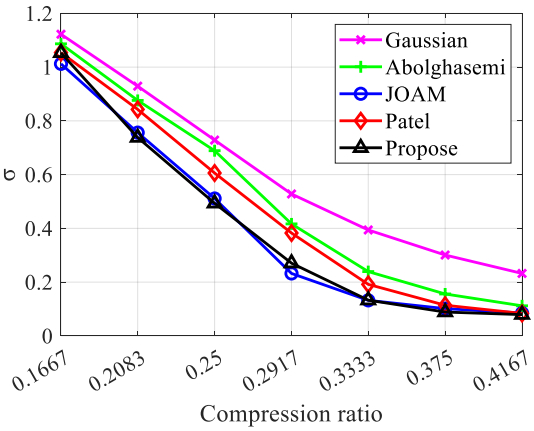


Fig. 6. The change tendency of σ with CR while K = 12 and SNR = 20 dB.

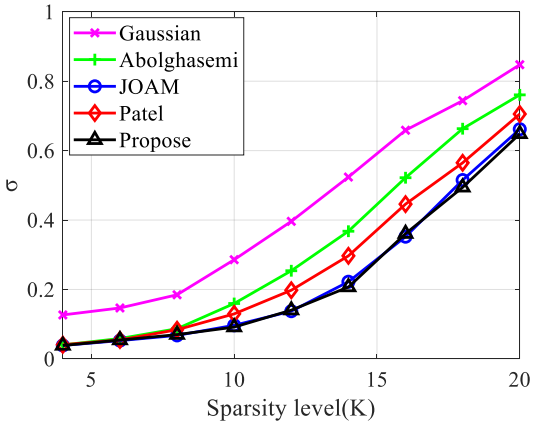
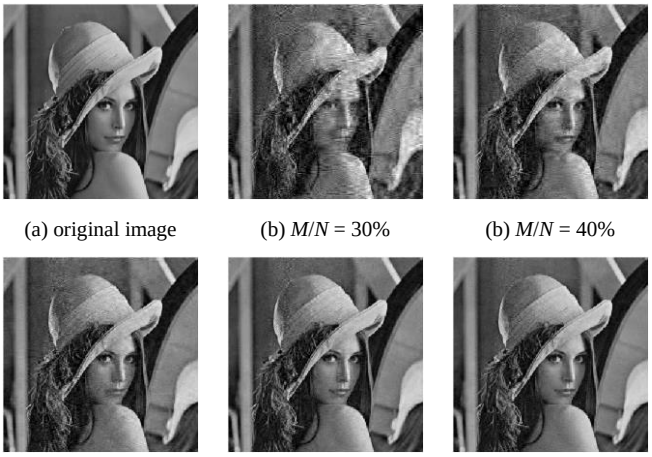


Fig. 7. The change tendency of σ with K while CR = 33.3% and SNR = 20 dB.

4.3.2 Comparison of Two-dimensional Signals

In this subsection, two-dimensional natural images of size 256 × 256 are used as input signals. A wavelet basis of size 256 × 256 is considered the sparse dictionary. The OMP algorithm is employed in reconstruction. Figure 8 displays the visual result of Lena when using the proposed algorithm.



(a) original image (b) M/N = 30% (b) M/N = 40%

(b) M/N = 50% (b) M/N = 60% (b) M/N = 70%

Fig. 8. Lena and its reconstructed images at different compression ratio using the proposed algorithm.

Image	CR	Gaussian	Abolghasemi	JOAM	Patel	Propose
Lena	**0.5**	25.83	26.75	27.24	26.89	27.14
	0.7	30.87	31.61	32.25	31.78	32.21
	0.9	37.26	38.41	38.71	38.45	38.68
Boat	**0.5**	23.52	24.09	24.48	24.13	24.42
	0.7	28.03	28.85	29.29	28.94	29.23
	0.9	35.09	36.75	37.34	36.81	37.25

Tab. 6. PSNR (dB) for Lena and Boat.

Gaussian, Abolghasemi, JOAM, and Patel are used for comparison with the proposed algorithm. The reconstruction performance is evaluated in terms of the peak signal-to-noise ratio (PSNR) given by [22]. A larger PSNR means better signal reconstruction performance. The results of PSNR are shown in Tab. 6.

As shown in Tab. 6, the optimized measurement matrices yield better performance than the Gaussian random matrix does. The proposed algorithm has comparable performance in terms of PSNR with JOAM. The PSNRs of the proposed algorithm are better than those of Patel for the tested images. It is interesting to observe that the difference of PSNR values of the optimization algorithms is not great. The reason is that the images are not strictly sparse in the wavelet basis, and high value of sparsity leads to a decrease in reconstruction accuracy.

5. Conclusions

In this paper, we propose an efficient optimization algorithm for the measurement matrix. A new target Gram matrix is designed to reduce multiple mutual coherence indicators simultaneously. Moreover, an improved NAG algorithm for updating the measurement matrix is derived to accelerate convergence. The experimental results show that the proposed algorithm performs as well as JOAM [15] in terms of reconstruction performance but with sig-

nificantly lower complexity. In addition, at the cost of slightly higher complexity, the proposed algorithm is far more advantageous than the existing method [16] in reconstruction.

References

[1] DONOHO, D. L. Compressed sensing. *IEEE Transactions on Information Theory*, 2006, vol. 52, no. 4, p. 1289–1306. DOI: 10.1109/TIT.2006.871582

[2] ELAD, M. Optimized projections for compressed sensing. *IEEE Transactions on Signal Processing*, 2007, vol. 55, no. 12, p. 5695–5702. DOI: 10.1109/TSP.2007.900760

[3] DONOHO, D. L., ELAD, M. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ1 minimization. *Proceedings of the National Academy of Sciences*, 2003, vol. 100, no. 5, p. 2197–2202. DOI: 10.1073/pnas.0437847100

[4] ZHAO, R., QIN, Z., HU, S. An optimization method for measurement matrix based on eigenvalue decomposition (in Chinese). *Signal Processing (Xinhao Chuli)*, 2012, vol. 28, no. 5, p. 653–658. DOI: 10.3969/j.issn.1003-0530.2012.05.006

[5] DUARTE-CARVAJALINO, J. M., SAPIRO, G. Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Transactions on Image Processing*, 2009, vol. 18, no. 7, p. 1395–1408. DOI: 10.1109/TIP.2009.2022459

[6] AHARON, M., ELAD, M., BRUCKSTEIN A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 2006, vol. 54, no. 11, p. 4311–4322. DOI: 10.1109/TSP.2006.881199

[7] XU, J., PI, Y., CAO, Z. Optimized projection matrix for compressive sensing. *EURASIP Journal on Advances in Signal Processing*, 2010, vol. 2020, p. 1–8. DOI: 10.1155/2010/560349

[8] SUSTIK, M. A., TROPP, J. A., DHILLON, I. S., et al. On the existence of equiangular tight frames. *Linear Algebra and Its Applications*, 2007, vol. 426, no. 2, p. 619–635. DOI: 10.1016/j.laa.2007.05.043

[9] ABOLGHASEMI, V., FERDOWSI, S., SANEI, S. A gradient-based alternating minimization approach for optimization of the measurement matrix in compressive sensing. *Signal Processing*, 2012, vol. 92, no. 4, p. 999–1009. DOI: 10.1016/j.sigpro.2011.10.012

[10] NOUASRIA, H., ET-TOLBA, M. A fast gradient-based sensing matrix optimization approach for compressive sensing. *Signal, Image and Video Processing*, 2022, vol. 16, no. 8, p. 2279–2286. DOI: 10.1007/s11760-022-02193-4

[11] LI, G., ZHU, Z., YANG, D., et al. On projection matrix optimization for compressive sensing systems. *IEEE Transactions on Signal Processing*, 2013, vol. 61, no. 11, p. 2887–2898. DOI: 10.1109/TSP.2013.2253776

[12] ENTEZARI, R., RASHIDI, A. Measurement matrix optimization based on incoherent unit norm tight frame. *AEU-International Journal of Electronics and Communications*, 2017, vol. 82, p. 321–326. DOI: 10.1016/j.aeue.2017.09.015

[13] HADJI, B., AISSA-EL-BEY, A., FERGANI, L., et al. Joint hybrid precoding and combining design based multi-stage compressed sensing approach for mmwave MIMO channel estimation. *IEEE Access*, 2023, vol. 11, p. 12398–112413. DOI: 10.1109/ACCESS.2023.3322658

[14] HONG, T., LI, S., BAI, H., et al. An efficient algorithm for designing projection matrix in compressive sensing based on alternating optimization. *Signal Processing*, 2016, vol. 125, p. 9–20. DOI: 10.1016/j.sigpro.2015.12.015

[15] JIN, S., SUN, W., HUANG, L. Joint optimization methods for Gaussian random measurement matrix based on column coherence in compressed sensing. *Signal Processing*, 2023, vol. 207, p. 1–13. DOI: 10.1016/j.sigpro.2023.108941

[16] PATEL, S., VAISH, A. An efficient optimization of measurement matrix for compressive sensing. *Journal of Visual Communication and Image Representation*, 2023, vol. 95, p. 1–10. DOI: 10.1016/j.jvcir.2023.103904

[17] CANDES, E. J., TAO, T. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 2006, vol. 52, no. 12, p. 5406–5425. DOI: 10.1109/TIT.2006.885507

[18] ZHU, Z., LI, Q., TANG, G., et al. Global optimality in low-rank matrix optimization. *IEEE Transactions on Signal Processing*, 2018, vol. 66, no. 13, p. 3614–3628. DOI: 10.1109/TSP.2018.2835403

[19] LI, Q., ZHU, Z., TANG, G. The nonconvex geometry of low-rank matrix optimization. *Information & Inference: A Journal of the IMA*, 2019, vol. 8, no. 1, p. 51–96. DOI: 10.1093/imaiai/iay003

[20] NESTEROV, Y. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$ (in Russian). *Doklady Akademii Nauk SSSR*, 1983, vol. 269, no. 3, p. 543–547.

[21] TROPP, J. A., GILBERT, A. C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 2007, vol. 53, no. 12, p. 4655–4666. DOI: 10.1109/TIT.2007.909108

[22] ELAD, M., AHARON, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 2006, vol. 15, no. 12, p. 3736–3745. DOI: 10.1109/TIP.2006.881969

About the Authors ...

Bo ZHANG was born in Xian, Shanxi, China, in 1987. He received his M.Sc. and the Ph.D. from the Air Force Engineering University in 2010 and 2016, respectively. His research interests include software radio, signal processing and wireless communication.

Renjie YI (corresponding author) was born in Suizhou, Hubei, China, in 1990. He received his M.Sc. and the Ph.D. from the National University of Defense Technology in 2016 and 2021, respectively. His research interests include the areas of wireless communications, signal processing and sparse recovery.

Zhike WANG was born in Dehui, Jilin, China, in 1982. He received his M.Sc. from the Air Force Engineering University in 2009, and the Ph.D. from Tsinghua University in 2020. His research interests include distributed control in sensor networks and optimization problem.

Jianfeng PU was born in Hailin, Heilongjiang, China, in 1975. He received his M.Sc. from Beihang University. His research interests include radar waveform design and signal processing.

Yang SUN was born in Tonghua, Jilin, China, in 1987. She received her M.Sc. from Jilin University in 2012. Her research interests include signal processing and radio-frequency antennas.