

# P-Tensor Product in Compressed Sensing

Haipeng Peng, Yaqi Mi, Lixiang Li<sup>1</sup>, H. Eugene Stanley, and Yixian Yang

**Abstract**—The dimension matching is a tough problem in the vector and matrix computations. In the traditional mode, there is only one way to calculate the angle between the 1-D plane and the 3-D vector, it is the projection. However, there are a number of lines on the plane, and taking only the projection to represent the plane is kind of a narrow choice. Furthermore, in the matrix multiplication, the dimension restriction is strict. In order to solve these problems, this paper defines a new model called *P*-tensor product (PTP), which cannot only define the inner product of two vectors with unmatched dimensions but also give a new way to solve the problems in the matrix operations. Aiming at decreasing the large storage space of the random matrix in compressed sensing (CS), the PTP can reconstruct a high-dimensional matrix by using a matrix, which can be chosen as any kind of matrix. Similar with the traditional CS, we analyze some reconstruction conditions of PTP-CS such as, the spark, the coherence, and the restricted isometry property. The theorems proposed in this paper have a broad sense, and they possess a good universality for various tensor product CS methods. The experimental results demonstrate that our PTP-CS model can not only give more choices to the types of Kronecker matrix and decrease the storage space of the traditional CS but also maintain the considerable recovery performance. Besides, the proposed PTP-CS model can improve the signal transmission efficiency in the Internet of Things.

**Index Terms**—Compressed sensing (CS), Internet of Things (IoT), matrix, tensor product, vector.

## I. INTRODUCTION

IN THE conventional data sampling system, the classical Nyquist–Shannon sampling theorem states that if the signal is of limited bandwidth and its sampling frequency is twice as the highest frequency, then the original continuous signal can be completely reconstructed from its samples. As a new signal processing technology, the compressed sensing (CS) breaks through the limitation of the Nyquist–Shannon theorem and it has attracted much attention [1]–[3]. In view of the sparsity of the signal, the CS can obtain the valid information through a nonadaptive sampling method and reconstruct the original

Manuscript received September 11, 2018; revised October 15, 2018 and November 21, 2018; accepted November 30, 2018. Date of publication December 14, 2018; date of current version May 8, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0800602 and in part by the National Natural Science Foundation of China under Grant 61771071 and Grant 61573067. (Corresponding author: Lixiang Li.)

H. Peng, Y. Mi, L. Li, and Y. Yang are with the Information Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the National Engineering Laboratory for Disaster Backup and Recovery, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: penghaipeng@bupt.edu.cn; miyaqi007@bupt.edu.cn; lixiang@bupt.edu.cn; yxyang@bupt.edu.cn).

H. E. Stanley is with the Department of Physics, Boston University, Boston, MA 02215 USA (e-mail: hes@bu.edu).

Digital Object Identifier 10.1109/JIOT.2018.2886841

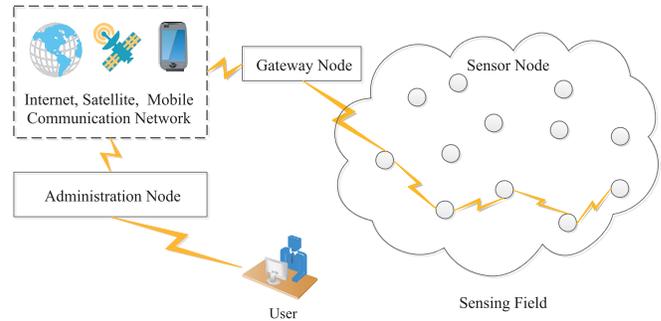


Fig. 1. Structure of WSN. The user can send the data information to the administration node by the computer. Then the administration node can send the data to the Internet, the satellite, or the mobile communication network. Subsequently, the data may be sent to the sensor nodes in the sensing field through the gateway node.

signal by a nonlinear method. For the signal processing, the CS can remove the redundancy, save the vital information of the image and further reduce the dimension of the data [4]. The CS provides a useful way in signal encryption, which is necessary for virtual property security. Furthermore, the CS can be used in many fields, such as, radar imaging [5], magnetic resonance imaging [6], and multileaks identification [7].

As a significant concept in recent years, the Internet of Things (IoT) has attracted quiet a lot of attention. The IoT can collect large numbers of information on time during the connection procedure. It includes the sensing layer, the network layer and the application layer. The wireless sensor network (WSN) is a part of sensing layer, and it can collect a large amount of data and transmit them. However, the sensors in the WSN have limited energy. Therefore, transmitting a large amount of data in an efficient way is a problem that requires an urgent solution. The CS can solve this problem well, i.e., it can distort the redundant data and transmit the signal fast [8]. As is shown in Fig. 1, the data information from the user are sent to the sensor nodes in the sensing field. During this procedure, the data should be transmitted to the administration node and some networks, the satellite, the gateway node, the Internet, and the mobile communication network. If the data needed to be transmitted is extremely large, then it would cost a lot of time and energy. Especially for the sensor with limited energy, a large amount of data is likely to be transmitted unsuccessfully. In the traditional CS, although the signal can be compressed before transmission, the storage space needed by the measurement matrix is quiet big. Hence, reducing the storage space of the measurement matrix in the CS is an important topic in the field of WSN.

In the CS, the design of the measurement matrix also gaining popularity. The key procedure of CS is to “compress”

the original signal to a lower-dimensional space by multiplying it with the measurement matrix, which should ensure less data collection, higher adaptability, easier implementation, and better optimization. In view of the above requirements, Do *et al.* [9] proposed the structurally random matrix (SRM). Although the SRM can improve the quality of the compression, it is hard to be implemented because of its high complexity. Gan [10] proposed a block CS model called Bayesian CS, which divided the image into  $B \times B$  nonoverlapping blocks. Then each block was observed by an orthonormal and independent identically distributed (i.i.d) Gaussian matrix [11]. In this way, the storage space of the measurement matrix is reduced and the speed of CS is improved. Peng *et al.* [12] proposed an effective method based on Chaotic CS (CCS), which only stored the matrix generation parameters so as to save the storage space. It is found that the random matrix is highly adaptable, and is easily obtained when it is chosen as the measurement matrix. However, the size of the original signal is comparatively large. Furthermore, the restriction of the traditional matrix multiplication is severely significant. For instance, the traditional multiplication among the matrices should satisfy the condition of dimension matching or complete the operation by an intermediate matrix. That is, the random measurement matrix requires a large storage space. Therefore, it is important to break the limitation of the matrix multiplication.

In view of the dimension matching problem, Zheng *et al.* [13] proposed a new framework called shift-invariant dictionary learning, which mainly focused on global patterns rather than shift-invariant local patterns. In 2008, Cheng *et al.* [14]–[16] proposed a matrix product called the semi-tensor product (STP), which has been widely used in various fields, such as Boolean networks [17] and linear algebra [18]. Especially, Xie *et al.* [19] applied the STP into the CS called semi-tensor product CS (STP-CS), which had broken through the restriction of the traditional CS on matrix dimension matching and obtained good results in the decrement of the matrix storage space. When STP-CS was applied to IoT, Peng *et al.* [20] obtained a good performance in the encryption and saved the storage space of the secret key. However, the STP has an evident limitation in the calculation of the inner product between two vectors with different dimensions, i.e., the definition of the inner product between two vectors is a vector rather than a certain number. Furthermore, it does not define the angle between them. Therefore, there is an urgent requirement to develop new definitions of these concepts for the vectors with different dimensions, i.e., it is important that the concept or the model of STP should be extended to a more general case.

In this paper, we propose a new matrix operation called  $P$ -tensor product (PTP), which is more flexible and more general in solving the problem of the dimension matching between two matrices. In the definition of PTP, the matrix with smaller size can be enlarged to a larger size by making the Kronecker product with matrix  $P$ , which can be chosen as any kind of matrix rather than only the identity matrix. Moreover, we extend the PTP to other vector operations, such as the inner product and the included angle between two vectors with unmatched dimensions. It breaks through the traditional

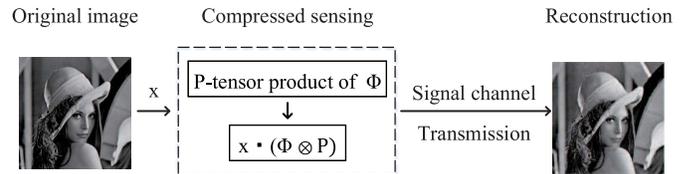


Fig. 2. CS using the PTP. The original image  $x$  is compressed by the measurement matrix, which is a product of two matrices. In the CS, By making the PTP of the matrix  $P$ , the original low-dimensional measurement matrix  $\Phi$  can be enlarged to a high-dimensional matrix, so that the storage space which is required by the measurement matrix is reduced. After transmitting in the signal channel, the image is reconstructed.

concept of the angle between two vectors under different dimensional spaces, and we define a new inner product and an angle between two vectors with different dimensions in  $P$ -transform (i.e., the transform uses the matrix or vector  $P$  to make two matrices or vectors that have the matched dimensions and then continue the next step, such as the multiplication between the matrices and the calculation of the inner product).

When PTP is applied into CS, the proposed model PTP-CS provides a method to observe the high-dimensional original signal by the low-dimensional random matrix. Hence, the storage space of the random matrix is reduced significantly. As is shown in Fig. 2, the original image  $x$  will multiply the  $(\Phi \otimes P)$ . Here, the  $(\Phi \otimes P)$  is the PTP of the original measurement matrix  $\Phi$ . In this way, the PTP-CS has more choices of the matrix  $P$  than the STP-CS and it has the opportunity to adjust the properties of the final measurement matrix. The storage spaces of PTP-CS and STP-CS have little difference. After the compression by the measurement matrix, the image will be transmitted on the signal channel to the terminal. Then the signal can be reconstructed by the proposed recovery algorithm based on the iteratively reweighted least squares (IRLSs). As a method of signal processing, similar with [20], the PTP-CS can be used in IoT and it provides a more efficient way for the communication theory. Compared to the existing CS schemes, the PTP-CS has a lower requirement of the storage space and it can get a similar as well as more stable recovery result. Furthermore, the conditions of the measurement matrix are analyzed in a broad sense, and we give three theorems of PTP-CS, which can be used to analyze other existing CS models, for example, the STP-CS. The advantages of PTP and PTP-CS are listed as follows.

- 1) *New Tensor Product Model Overhead*: We define a new inner product operation and a new angle between two vectors with unmatched dimensions in  $P$ -transform. Meanwhile, the PTP can break through the restrictions of the traditional matrix multiplication and the STP operation.
- 2) *Good Universality and Low Storage*: The theorems of the PTP-CS can be used in a broad sense. They can be used to analyze other CS models and optimize the compression method. Moreover, through the PTP-CS, the low-dimensional measurement matrix can be extended to the high-dimensional matrix and then can be used to compress the signal.
- 3) *New Reconstruction Model*: In view of the PTP-CS, we propose a new model based on the IRLSs. It is found

that the proposed model has a good performance in the signal reconstruction.

## II. FUNDAMENTAL KNOWLEDGE

### A. Compressive Sensing

Suppose  $x \in R^{n \times 1}$  is a  $k$ -sparse discrete signal in the orthogonal basis  $\Psi$ ,  $\Phi \in R^{m \times n}$  is the *measurement matrix*, and it is dependent with  $\Psi$ , the normalized model of CS can be denoted as

$$y = \Phi x \quad (1)$$

where  $m < n$ ,  $y \in R^{m \times 1}$ ,  $x$  is sparse or sparse on an orthogonal basis, that is,

$$x = \Psi s \quad (2)$$

where  $\Psi \in R^{n \times n}$  is also called the *sparse matrix*. Obviously, for  $x$  in the time or space domain and  $s$  in the  $\Psi$  domain,  $x$  and  $s$  can represent the signal equivalently [21]. The signal  $x$  is  $k$ -sparse, that is, there are  $k$  nonzero rows in  $s$  ( $k \ll n$ ). From (1) and (2), we get

$$y = \Phi x = \Phi \Psi s = \Theta s \quad (3)$$

where  $\Theta = \Phi \Psi$  is called the *sensing matrix*. CS includes three parts, i.e., the signal sparse representation, the design of measurement matrix, and the reconstruction algorithm. Now we begin to illustrate these three aspects.

1) *Signal Sparse Representation*: The sparsity is that there are many zero elements in the signal so that the signal can be sampled and compressed effectively in the transform domain. Suppose the signal  $x$  is  $k$ -sparse, then we can construct the original signal from  $k$  values of these  $m$  measurements, where  $x \in R^n$  and  $k \ll m < n$ . Due to the sparsity in a certain orthogonal basis, the discrete cosine transform (DCT) matrices and the discrete wavelet transform (DWT) matrices are often used as the sparsity matrix. In 2008, Rauhut *et al.* [22] decomposed the sparse signals by constructing the redundant dictionaries, and the CS was generalized from the orthogonal basis space to the redundant dictionary.

In the practice, DCT and DWT are easy to be implemented. In this paper, we use the DWT matrix as the sparsity matrix.

2) *Conditions of Measurement Matrix*: The design of measurement matrix has attracted much attention in recent years. The measurement matrix  $\Phi \in R^{m \times n}$  ( $m < n$ ) measures the original signal  $x$  to obtain the vector  $y$ . For a  $k$ -sparse signal  $x$ , in order to ensure the accuracy of the recovery, the matrix  $\Phi$  must satisfy the following restricted isometry property (RIP) [23] denoted by  $\delta_k^\Phi$ , i.e.,

$$(1 - \delta_k^\Phi) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_k^\Phi) \|x\|_2^2 \quad (4)$$

where  $\delta_k^\Phi \in (0, 1)$ . The sufficient condition to recover the signal is  $\delta_{2k} < \sqrt{2} - 1$  [24]. Cai *et al.* [25] gave its exact value  $\delta_k < 0.307$ . The commonly used measurement matrices are Gaussian matrix [26], Bernoulli matrix, Toeplitz matrix [27], the chaotic matrix [28], etc. In addition, the random matrix, such as Gaussian matrix and Bernoulli matrix, can satisfy RIP with a high probability [3]. Candes and Tao [29] proved that the i.i.d Gaussian random matrix was a universal  $\Phi$  in CS whereas it was uncertain and wasted the storage space.

Therefore, how to reduce the required space of CS is an important problem to be solved.

Due to the difficulty of RIP verification, the performance of the  $\Phi$  is usually evaluated by the spark property, namely, the solution of the minimum of linear correlation vectors in the matrix. As for the  $k$ -sparse signal  $x$ , if and only if  $\text{spark}(\Phi) > 2k$  can we obtain the exact approximation of the signal with a minimization  $l_0$ -norm optimization problem. Another important property is the coherence [30]. The coherence coefficient  $\mu(\Phi)$  shows the redundancy of the information in  $\Phi$ . The coherence coefficient is denoted as

$$\mu(\Phi) = \max_{1 \leq i \neq j \leq n} \frac{|\langle \varphi_i, \varphi_j \rangle|}{\|\varphi_i\|_2 \|\varphi_j\|_2} \quad (5)$$

where  $\varphi_i$  is the  $i$ th column of  $\Phi$ . Evidently,  $\|\varphi_i\|_2 = 1$ ,  $\mu \in [\sqrt{(n-m)/(m(n-1))}, 1]$ , and the lower boundary of the coherence is called the *Welch bound* [3], [31]–[33].  $\mu$  can measure the similarity between two column vectors. If  $\mu$  is large, at least two column vectors are similar to each other. On the contrary, if  $\mu$  is small, the columns of  $\Phi$  are almost orthogonal. Other low-rank approaches (for example, the non-negative method) are widely studied in recent years [34]–[38].

The conditions of  $\Phi$  include three aspects, i.e., simple to achieve, low storage and good adaptability. In this paper, we propose a new model PTP-CS, which can reduce the memory space required by  $\Phi$  and is easier to be implemented. In addition, it ensures a good observational performance and has a good adaptability by our simulation results.

3) *Reconstruction*: The reconstruction algorithms include the greedy iterative algorithm and the algorithms based on Bayesian framework as well as convex optimization. The greedy iterative algorithm aims to find each nonzero coefficient through local optimization in each iteration, including matching pursuit (MP) [39], orthogonal MP (OMP) [40], stagewise OMP (StOMP) [41], and other various improved algorithms of OMP [42]. Although these algorithms have fast speeds to reconstruct the signal, they have a dramatic demand of the measurement matrix. Meanwhile, they may be unsuitable for some CS models, i.e., they cannot give stable recovery performances. The algorithms based on Bayesian framework mainly consider about the time correlation of the signal and they have high accuracies, such as BCS [43], expectation-maximization [44], sparse Bayesian learning [45], and so on. The convex optimization algorithm turns the nonconvex problem into the convex problem to recover the original signal. The typical method is basis pursuit, which is based on the minimal  $l_1$ -norm. Chen *et al.* [46] gave the solving process in detail. Moreover, there is another method called FOCUSS [47], [48] based on  $l_\rho$ -norm ( $0 < \rho < 1$ ) [49]. In this paper, we mainly focus on the reconstruction algorithms based on the convex optimization.

For the reconstruction algorithms based on the convex optimization, the problem of recovering the  $k$ -sparse signal  $x$  from the observed value  $y$ , namely, the solution of the nonzero sparse minimization problem, can be transformed into the following  $l_0$ -minimization:

$$x^* = \arg \min_x \|x\|_0 \quad \text{subject to } y = \Phi x. \quad (6)$$

The  $l_0$ -norm is NP-hard while the  $l_1$ -norm is convex. Therefore, the solution of (6) is converted to obtain the following  $l_1$ -minimization when  $\Phi$  satisfies RIP:

$$x^* = \arg \min_x \|x\|_1 \quad \text{subject to } y = \Phi x. \quad (7)$$

In the STP-CS model, Wang *et al.* [50] used the IRLSs minimization [51], [52] to reconstruct the signal and obtained a good result. Different from the STP-CS reconstruction model in [50], in this paper, we propose a new reconstruction model based on  $l_\rho$ -norm ( $0 < \rho < 1$ ) which combines the IRLS. It can be seen that the proposed reconstruction model can be applied well to PTP-CS.

### B. Kronecker and Tensor Products

The Kronecker product is a special type of tensor product.

*Definition 1* [53]: Let the matrices  $\Phi \in R^{m \times n}$ ,  $P \in R^{p \times q}$ ,  $\Phi = [\varphi_1, \dots, \varphi_n]$ , then the Kronecker product of  $\Phi$  and  $P$  is defined as

$$\Phi \otimes P = (\varphi_1 P, \dots, \varphi_n P) = \begin{pmatrix} \varphi_{11}P & \varphi_{12}P & \dots & \varphi_{1n}P \\ \varphi_{21}P & \varphi_{22}P & \dots & \varphi_{2n}P \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m1}P & \varphi_{m2}P & \dots & \varphi_{mn}P \end{pmatrix}. \quad (8)$$

Thus,  $\Phi \otimes P$  is a matrix with size  $mp \times nq$ , i.e.,  $R^{m \times n} \times R^{p \times q} \rightarrow R^{mp \times nq}$ .

*Property 1*: Let  $\Phi \in R^{m \times n}$ ,  $P \in R^{p \times q}$ , then

$$\Phi \otimes P = (\Phi \otimes I_p) \cdot (I_n \otimes P). \quad (9)$$

*Property 2*: If  $\Phi$  and  $P$  are both invertible matrices, then

$$(\Phi \otimes P)^{-1} = \Phi^{-1} \otimes P^{-1}. \quad (10)$$

*Property 3*: The transposed matrix of the Kronecker product is

$$(\Phi \otimes P)^T = \Phi^T \otimes P^T. \quad (11)$$

The tensor product is always used in the high-dimensional space, and it is a multilinear mapping for  $n$  vector spaces, namely

$$\mathcal{T} : \mathcal{T}^{I_1} \times \mathcal{T}^{I_2} \times \dots \times \mathcal{T}^{I_n} \rightarrow \mathcal{T}^{I_1 \times I_2 \times \dots \times I_n}. \quad (12)$$

The tensor product can be used for Tucker decomposition [54], [55], which is a high-dimensional principle component analysis. Actually, the Kronecker product is also called tensor product most of the time due to a little difference between them. Aiming at the first-order tensor and the second-order tensor, we mainly focus on the vector and the matrix tensor products in this paper and consider the Kronecker product and the tensor product, on the whole, are same.

### C. Generalized Permutation Matrix

*Definition 2*: If there is only one nonzero element 1 in each row and each column, then the square matrix is called the permutation matrix.

The permutation matrix  $D$  has the following properties [56].

- 1)  $(D_{u \times v})^T = D_{v \times u}$ .
- 2)  $D^T D = D D^T = I$ .
- 3)  $D^T = D^{-1}$ .

From the above properties 2) and 3) of the permutation matrix  $D$ , we can find that the permutation matrix is the orthogonal matrix.

*Definition 3*: If there is only one nonzero element in each row and each column, then the square matrix is called the generalized permutation matrix which is denoted by  $g$ .

Obviously, it has the same properties with the permutation matrix. Furthermore, a square matrix is  $g$  if and only if it can be represented as a product of a permutation matrix and a nonsingular diagonal matrix, that is,

$$g = D \cdot \Lambda \quad (13)$$

where  $\Lambda$  is a nonsingular diagonal matrix.

## III. P-TENSOR PRODUCT COMPRESSED SENSING

### A. P-Tensor Product Model

The concept of PTP is a new mathematical matrix operation, which can break through the limitation of the dimensions of two vectors or matrices. Above all, it gives a more accurate definition of the inner product of two vectors with unmatched dimensions, and then we define the angle between them under the  $P$ -transform. Here, we give the standard definition of the PTP.

*Definition 4*: Let  $\alpha = [a_1, a_2, \dots, a_n]$  be a  $n$ -dimensional row vector, and let  $\lambda = [b_1, b_2, \dots, b_p]^T$  be a  $p$ -dimensional column vector. If  $n$  is a factor of  $p$ , i.e.,  $p = t \times n$ ,  $P \in R^{t \times t}$ , then we have

$$\alpha \times^P \lambda = \sum_{k=1}^n (a_k P) \lambda^k \in R^{t \times 1} \quad (14)$$

where  $\alpha \times^P \lambda$  is called the PTP of the vectors  $\alpha$  and  $\lambda$ . Here,  $\lambda = (\lambda^1, \lambda^2, \dots, \lambda^n)$ ,  $\lambda^i \in R^{t \times 1}$ ,  $i = 1, 2, \dots, n$ . Otherwise, if  $p$  is a factor of  $n$ , i.e.,  $n = t \times p$ , then we have

$$\lambda \times^P \alpha = \sum_{k=1}^p \alpha^k (b_k P) \in R^{1 \times t}. \quad (15)$$

Similarly,  $\lambda \times^P \alpha$  is also the PTP of  $\alpha$  and  $\lambda$ . Here,  $\alpha = (\alpha^1, \alpha^2, \dots, \alpha^p)$ ,  $\alpha^i \in R^{1 \times t}$ ,  $i = 1, 2, \dots, p$ .

*Definition 5*: Let  $\alpha = [a_1, a_2, \dots, a_n]$  be a  $n$ -dimensional row vector, and let  $\lambda = [b_1, b_2, \dots, b_p]^T$  be a  $p$ -dimensional column vector. If  $n$  is a factor of  $p$ , i.e.,  $p = t \times n$ ,  $P \in R^{1 \times t}$ , then we have

$$\langle \alpha, \lambda \rangle_P := \sum_{k=1}^n (a_k P) \lambda^k \quad (16)$$

where  $\langle \alpha, \lambda \rangle_P$  is called the  $P$ -inner product of the vectors  $\alpha$  and  $\lambda$ .

*Definition 6*: The cosine value of the angle between the vectors  $\alpha$  and  $\lambda$  is

$$\cos(\alpha, \lambda)_P = \frac{\langle \alpha, \lambda \rangle_P}{\|P\| \cdot \|\alpha\| \cdot \|\lambda\|} \quad (17)$$

where  $\cos(\alpha, \lambda)_P$  is called the cosine value of the vectors  $\alpha$  and  $\lambda$  in the  $P$ -transform, i.e., this transform uses the matrix or vector  $P$  to make two matrices or vectors having

the matched dimensions and then continue toward the next step, such as the multiplication between the matrices and the calculation of the inner product.

Equation (17) gives a new way to calculate the angle of two vectors with different dimensions. For example, let  $\alpha = (1 \ 1)$ ,  $\lambda = (1 \ 0 \ 0 \ 1)^T$ ,  $P = (1 \ 0)$ , then  $\langle \alpha, \lambda \rangle_P = 1$  and  $\cos(\alpha, \lambda)_P = 1/2$ .  $\langle \alpha, \lambda \rangle_P = 1$  is called the *inner product* of the vectors  $\alpha$  and  $\lambda$  under the  $P$ -transform, and the  $\cos(\alpha, \lambda)_P = 1/2$  is called the *cosine value* of the vectors  $\alpha$  and  $\lambda$  under the  $P$ -transform.

By the above new definitions of the inner product and the angle between two vectors, we can further get the included angle between two vectors with different dimensions. For example, the traditional angle between a line and a plane can only be obtained by the vertical projection. Let a plane  $\gamma$  be  $x = 1$ , the vector  $\vec{OA} = (1 \ 1 \ 1)$  is a ray from the origin to the point  $A$ , then the angle between the  $\vec{OA}$  and the plane  $\gamma$  requires to be changed to the angle between  $\vec{OA}$  and the projection line, which is projected to  $\gamma$ . However, by the  $P$ -transform, we can extend the dimension of  $x = 1$  and get a line on the plane  $\gamma$  by  $x \otimes P$ , where  $P = (0 \ 1 \ 2)^T$ . By (17), we can get the cosine value of  $\vec{OA}$  and  $x = 1$  in the  $P$ -transform. As is shown in Fig. 3, the line  $l_1$  is the traditional projection of the line  $l$  in plane  $\gamma$ . Meanwhile, we can define other lines on plane  $\gamma$  by a proper  $P$  to get the angle between the plane and the line. The  $P$ -angle is a new definition for two vectors which have unmatched dimensions. If  $n = p$ , it is the inner product of the traditional vector.

Evidently, the inner product defined in the STP is different from the traditional inner product. As is mentioned in [19], the inner product of two vectors is also a vector rather than a certain value. So it cannot be used to calculate the cosine value of the vectors, which is not defined in the STP. It seems like a special inner product with the vector form and has no relationship with the calculation of the angle between two vectors. Therefore, the PTP is a good way to calculate the inner product, the cosine value and the angle between two vectors with unmatched dimensions.

**Definition 7:** Let  $\Phi \in R^{m \times n}$ ,  $x \in R^{p \times q}$ , the least common multiple of  $n$  and  $p$  is  $t$ , i.e.,  $t = \text{lcm}\{n, p\}$ . Note that  $y = \Phi \times_P x$  is the PTP of  $\Phi$  and  $x$ , so we have

$$y^{ij} = \langle \Phi^i, x^j \rangle_P \quad (18)$$

where  $\Phi^i$  is the  $i$ th row of  $\Phi$  with dimension  $n$ ,  $x^j$  is the  $j$ th column of  $x$  with dimension  $p$ ,  $i = 1, 2, \dots, m, j = 1, 2, \dots, q$ .

Evidently,  $y$  is a matrix made up by  $m \times q$  blocks. If  $P = I_{1 \times 1}$ , then it is the traditional multiplication. Furthermore, the PTP of two matrices can be defined in another way as

$$\Phi \times_P x = (\Phi \otimes P_{l \times t/n}) \cdot (x \otimes P_{t/p \times h}) \quad (19)$$

where  $l$  and  $h$  are random positive integers. Now we take  $P$  as a square matrix for convenience. Remark that  $\Phi \succ_t x$  is  $n = tp$  and  $\Phi \prec_t x$  is  $p = tn$ , where  $t$  is a positive integer. So the PTP of two matrices can be written as

$$\Phi \times_P x = \begin{cases} \Phi(x \otimes P), & \Phi \succ_t x \\ (\Phi \otimes P)x, & \Phi \prec_t x. \end{cases} \quad (20)$$

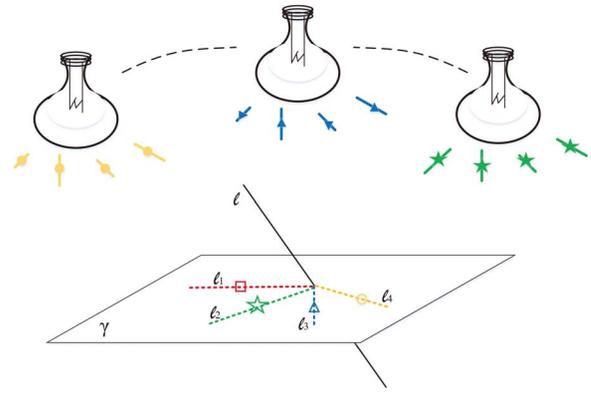


Fig. 3.  $P$ -angle of the vectors with different dimensions. The red line  $l_1$  with the mark square ( $\square$ ) is the traditional projection of the black line  $l$  under the plane  $\gamma$ . Three different lamp bulbs represent different lights, including the green line  $l_2$  with the mark star ( $\star$ ), the blue line  $l_3$  with the mark triangle ( $\Delta$ ), and the yellow line  $l_4$  with the mark circle ( $\circ$ ). These dotted lines with different marks on the plane  $\gamma$  are the projections of three lights with different colors. All of the projections can be defined by our method.

Suppose  $F$ ,  $G$ , and  $Q$  are matrices with proper dimensions, some rules among them can be given as follows.

**Property 4:** The distributive rule is given as

$$\begin{cases} F \times_P (lG \pm hQ) = lF \times_P G \pm hF \times_P Q, \\ (lF \pm hG) \times_P Q = lF \times_P Q \pm hG \times_P Q. \end{cases} \quad l, h \in \mathbb{R} \quad (21)$$

**Property 5:** The associative rule is given as

$$(F \times_P G) \times_P Q = F \times_P (G \times_P Q). \quad (22)$$

**Property 6:**  $F \times_P G$  and  $G \times_P F$  have the same eigenvalues, and  $\text{tr}(F \times_P G) = \text{tr}(G \times_P F)$ .

**Property 7:** If  $P^T = P$ , then we have

$$(F \times_P G)^T = G^T \times_P F^T. \quad (23)$$

**Property 8:** If  $F$ ,  $G$ , and  $P$  are invertible, then we have

$$(F \times_P G)^{-1} = G^{-1} \times_P F^{-1}. \quad (24)$$

In the above analyses, we only focus on the left matching of the matrices. Similar with (20), the right form of PTP can be defined as

$$\Phi \times_P x = \begin{cases} \Phi(P \otimes x), & \Phi \succ_t x \\ (P \otimes \Phi)x, & \Phi \prec_t x. \end{cases} \quad (25)$$

In order to distinguish these two types of PTP, we denote  $\Phi \times_P x$  as the *right PTP* and we denote  $\Phi \times_P x$  as the *left PTP*. The right PTP can satisfy Properties 4–8 as mentioned above. Furthermore, let  $\Phi \in R^{m \times n}$ ,  $x \in R^{p \times q}$ , if  $\Phi \succ_t x$ , then we can get the relationships of these two types of PTP as follows:

$$\Phi \times_P x = \Phi \times W_{[p,t]} \times_P x \times W_{[t,q]} \quad (26)$$

and

$$\Phi \times_P x = \Phi \times W_{[t,p]} \times_P x \times W_{[q,t]}. \quad (27)$$

Oppositely, if  $\Phi \prec_r x$ , the corresponding properties can be obtained similarly. In (26) and (27),  $W_{[p,t]}$  is the transposition matrix, and it can be defined as follows.

*Definition 8:* Let  $W_{[p,t]} \in R^{pt \times pt}$ , each row and each column of the matrix  $W$  are labeled by the double indices  $(i, j)$ , the columns are arranged by  $\text{Id}(i, j; p, t)$  and the rows are arranged by  $\text{Id}(J, I; t, p)$ . The element in the  $[(I, J), (i, j)]$  is

$$W_{[(I,J),(i,j)]} = \begin{cases} 1, & I = i \text{ and } J = j \\ 0, & \text{else.} \end{cases} \quad (28)$$

Evidently,  $W_{[p,t]}^T = W_{[p,t]}^{-1} = W_{[p,t]}$ .  
If  $P = I$ , it is the STP.

Furthermore, when we take the product operation of the traditional product and Kronecker product or PTP, Kronecker product cannot satisfy the associative rules, which is inconvenient for the theoretical studies, i.e., for the matrices  $F, G$ , and  $Q$  with proper dimensions, we have

$$(F \otimes G) \cdot Q \neq F \otimes (G \cdot Q) \quad (29)$$

while the PTP can satisfy the associative rules, i.e.,

$$\left( F \underset{\times}{\otimes} G \right) \cdot Q = F \underset{\times}{\otimes} (G \cdot Q). \quad (30)$$

This property of PTP is convenient for the applications, and PTP can overcome the insufficiencies of Kronecker product. When PTP is applied to the PTP-CS, we can get the definition of PTP-CS as follows:

$$y = \Phi \underset{\times}{\otimes} x \quad (31)$$

where  $\Phi$  is the original measurement matrix and  $x$  is a  $k$ -sparse signal. The matrix  $P$  can be any kind of invertible matrices. Here, we adopt the matrix  $P$  as the Gaussian random matrix and the generalized permutation matrix. The Gaussian random matrix is subject to normal distribution with mean 0 and variance 1.

### B. Spark

As is mentioned in Section II, the spark is an important property in the CS. So we give a corollary of the spark in the PTP-CS, and further give a theorem with a detailed proof.

*Corollary 1:* If  $P$  is an invertible matrix and  $\text{spark}(\Phi) > 2[kn/q]$ , then for each measurement  $y \in R^{mq/n}$ , there is at most one  $k$ -sparse signal  $x \in R^q$  so that  $y = \Phi \underset{\times}{\otimes} x$ , where  $[\cdot]$  is a floor function.

Here, the floor function is an integer-valued function in which the fractions are rounded down. So we can get the maximal integer that is not bigger than the input value.

*Theorem 1:* Let  $P$  is an invertible matrix, for each measurement  $y \in R^{(mq/n)}$ , if there is at most one signal  $x \in \Sigma_k$  so that  $y = \Phi \underset{\times}{\otimes} x$ , then  $\text{spark}(\Phi) > 2[kn/q]$ .

*Proof:* For (31), we have

$$y_{\frac{mq}{n} \times 1} = \left( \Phi_{m \times n} \otimes P_{\frac{q}{n} \times \frac{q}{n}} \right) \cdot x_{q \times 1}. \quad (32)$$

By the definition of the tensor product, (32) can be converted as

$$y_{\frac{mq}{n} \times 1} = \left[ \left( \Phi_{m \times n} \otimes I_{\frac{q}{n}} \right) \cdot \left( I_n \otimes P_{\frac{q}{n}} \right) \right] \cdot x_{q \times 1}. \quad (33)$$

Note that  $P'_{q \times q} = I_n \otimes P_{(q/n)}$ . Equation (33) can be written as

$$y = \left( \Phi \otimes I_{\frac{q}{n}} \right) \cdot P' \cdot x. \quad (34)$$

If  $P$  is invertible, by Property 2 we obtain

$$P'^{-1} = I^{-1} \otimes P^{-1}. \quad (35)$$

Evidently,  $P'$  is also invertible, thus

$$P'^{-1}y = P'^{-1} \cdot \left( \Phi \otimes I_{\frac{q}{n}} \right) \cdot P' \cdot x. \quad (36)$$

Note that  $T = P'^{-1} \cdot (\Phi \otimes I_{[q/n]}) \cdot P'$ ,  $Y = \Phi \otimes I_{[q/n]}$ , we can get  $T \sim Y$ .

Because  $T \sim Y$ ,  $T$  and  $Y$  has the same linear correlation. For the right part of (36), similar with [19], we can prove  $\text{spark}(\Phi) > 2[kn/q]$  by contradiction. Note that  $\tau = 2[kn/q]$ , then  $\text{spark}(\Phi) \leq \tau$ .  $\Phi = (\varphi_1, \varphi_2, \dots, \varphi_m)$ , where  $\varphi_i$  is the column vector with dimension  $n$ . Hence, there is a nonzero vector  $x \in \Sigma_\tau$  such that  $\sum_{i=1}^{\tau} \varphi_i x_i = 0$  by using the definition of the sparsity.

Note

$$\Phi \otimes I_{\frac{q}{n}} = \left( \beta_1^0, \beta_1^1, \dots, \beta_1^{\frac{q}{n}-1}, \dots, \beta_n^0, \dots, \beta_n^{\frac{q}{n}-1} \right) \quad (37)$$

where  $\beta_i$  is a column vector with dimension  $mq/n$ . By inserting  $((q/n) - 1)$  zero vectors into  $\varphi_i$ , we can get

$$\beta_i = \left( \underbrace{\varphi_{i1}, 0, \dots, 0}_{(q/n)}, \underbrace{\varphi_{i2}, 0, \dots, 0}_{(q/n)}, \dots, \underbrace{\varphi_{im}, 0, \dots, 0}_{(q/n)} \right)^T. \quad (38)$$

$\beta_i^j$  is obtained by moving the first element to the  $(j+1)$ th element of each bracketed part, where  $j = 0, 1, 2, \dots, (q/n) - 1$ . Obviously, for any  $j$  we have

$$\sum_{i=1}^{\tau} \beta_i^j x_i = 0. \quad (39)$$

By (37) and (39), we obtain

$$\left( \Phi \otimes I_{\frac{q}{n}} \right) \cdot (x_1, \dots, x_1, \dots, x_\tau, \dots, x_\tau, 0, \dots, 0)^T = 0 \quad (40)$$

where  $x_i = (x_1, \dots, x_\tau)$ ,  $i = 1, 2, \dots, \tau$ , and the number of each element is  $q/n$ .

Let  $X = (x_1, \dots, x_1, \dots, x_\tau, \dots, x_\tau, 0, \dots, 0)^T$ , then  $X \in \Sigma_{(\tau q/n)}$ . There exist  $X_1, X_2 \in \Sigma_{(\tau q/2n)}$  so that  $X = X_1 - X_2$ . Because  $\tau$  is an even integer, we have  $(\Phi \otimes I_{(q/n)}) \cdot (X_1 - X_2) = 0$  and  $\Phi \otimes X_1 = \Phi \otimes X_2$ . In addition,  $(\tau q/2n) = (kn/q)(q/n) \leq k$ ,  $X_1, X_2 \in \Sigma_k$ . It is obvious that this conclusion contradicts the assumption, so we have  $\text{spark}(\Phi) > 2(kn/q)$ . As a result, we prove the theorem that if  $P$  is invertible and there is at most one signal  $x \in \Sigma_k$  such that  $y = \Phi \underset{\times}{\otimes} x$ , then  $\text{spark}(\Phi) > 2(kn/q)$ . ■

### C. Coherence

The coherence is another important condition of the measurement matrix. Here, we give a theorem of the coherence in the PTP-CS, and further we give a detailed mathematical proof.

*Theorem 2:*  $\mu(\Phi \otimes P_{\frac{q}{n}}) = \max\{\mu(\Phi), \mu(P_{(q/n)})\}$ .

*Proof:* Let  $\Phi \otimes P_{(q/n)} = C$ , by (5) we have

$$\mu(C) = \max_{1 \leq i \neq j \leq n} \frac{|c_i, c_j|}{\|c_i\|_2 \|c_j\|_2}. \quad (41)$$

By (8), we have

$$(\Phi \otimes P_{\frac{q}{n}}) = \begin{pmatrix} \varphi_{11}P & \varphi_{12}P & \cdots & \varphi_{1n}P \\ \varphi_{21}P & \varphi_{22}P & \cdots & \varphi_{2n}P \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m1}P & \varphi_{m2}P & \cdots & \varphi_{mn}P \end{pmatrix}. \quad (42)$$

Note that  $P_{(q/n)} = (p_1, p_2, \dots, p_{(q/n)})$ , where  $p_i$  is a column vector with dimension  $(q/n)$  and  $i = 1, 2, \dots, (q/n)$ , then (42) can be expanded as

$$\begin{aligned} & \Phi \otimes P_{\frac{q}{n}} \\ &= \begin{pmatrix} \varphi_{11}p_1 & \cdots & \varphi_{11}p_{\frac{q}{n}} & \varphi_{12}p_1 & \cdots & \varphi_{12}p_{\frac{q}{n}} & \cdots & \varphi_{1n}p_1 & \cdots & \varphi_{1n}p_{\frac{q}{n}} \\ \varphi_{21}p_1 & \cdots & \varphi_{21}p_{\frac{q}{n}} & \varphi_{22}p_1 & \cdots & \varphi_{22}p_{\frac{q}{n}} & \cdots & \varphi_{2n}p_1 & \cdots & \varphi_{2n}p_{\frac{q}{n}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \varphi_{m1}p_1 & \cdots & \varphi_{m1}p_{\frac{q}{n}} & \varphi_{m2}p_1 & \cdots & \varphi_{m2}p_{\frac{q}{n}} & \cdots & \varphi_{mn}p_1 & \cdots & \varphi_{mn}p_{\frac{q}{n}} \end{pmatrix}. \end{aligned} \quad (43)$$

Select two different column vectors from  $\Phi \otimes P_{(q/n)}$  randomly and calculate their inner product. There are  $q$  columns in the  $\Phi \otimes P_{(q/n)}$ , so the selection can be denoted as  $C_q^2$  by the theory of permutation and combination, i.e., there exists  $([q(q-1)]/2)$  cases in the selection and we can obtain  $([q(q-1)]/2)$  values of inner products. Finally, we can find the maximum of the  $([q(q-1)]/2)$  values of inner products. Here, suppose  $c_i = (\varphi_{1e}p_f, \varphi_{2e}p_f, \dots, \varphi_{me}p_f)$  and  $c_j = (\varphi_{1d}p_r, \varphi_{2d}p_r, \dots, \varphi_{md}p_r)^T$ , the inner product of the column vectors  $c_i$  and  $c_j$  is the maximum of all  $([q(q-1)]/2)$  inner products, and their inner product is defined as

$$\begin{aligned} \langle c_i, c_j \rangle &= \varphi_{1e}p_f \cdot \varphi_{1d}p_r^T + \varphi_{2e}p_f \cdot \varphi_{2d}p_r^T \\ &+ \cdots + \varphi_{me}p_f \cdot \varphi_{md}p_r^T. \end{aligned} \quad (44)$$

Let  $\Phi_{m \times n} = (\varphi_1, \varphi_2, \dots, \varphi_n)$  where  $\varphi_j$  is a column vector with dimension  $m$  and  $j = 1, 2, \dots, n$ . Extracting the common factor, we have

$$\langle c_i, c_j \rangle = \langle \varphi_e, \varphi_d \rangle \cdot \langle p_f, p_r \rangle. \quad (45)$$

By (45), (41) can be converted as

$$\mu(C) = \max_{1 \leq (e,d) \neq (f,r) \leq n} \frac{|\langle \varphi_e, \varphi_d \rangle \cdot \langle p_f, p_r \rangle|}{\|\varphi_e\|_2 \cdot \|\varphi_d\|_2 \cdot \|p_f\|_2 \cdot \|p_r\|_2}. \quad (46)$$

According to the definition of the coherence, we should discuss the following three cases of (46).

1)  $e \neq d, f \neq r$

$$\mu(C) = \mu(\Phi) \cdot \mu(P). \quad (47)$$

2)  $e = d, f \neq r$ .

We can easily get  $(|\langle \varphi_e, \varphi_d \rangle| / [\|\varphi_e\|_2 \cdot \|\varphi_d\|_2]) = 1$ , so

$$\mu(C) = \max_{1 \leq f \neq r \leq n} \frac{|\langle p_f, p_r \rangle|}{\|p_f\|_2 \cdot \|p_r\|_2}. \quad (48)$$

3)  $e \neq d, f = r$ .

Similarly, we have

$$\mu(C) = \max_{1 \leq e \neq d \leq n} \frac{|\langle \varphi_e, \varphi_d \rangle|}{\|\varphi_e\|_2 \cdot \|\varphi_d\|_2}. \quad (49)$$

Because the coherence of any matrix satisfies  $\mu \in [\sqrt{[n-m]/[m(n-1)]}, 1]$ , we should eliminate case 1)  $e \neq d, f \neq r$ , which is defined by (47). In a word,  $\mu(\Phi \otimes P_{(q/n)}) = \max\{\mu(\Phi), \mu(P_{(q/n)})\}$ . ■

In [19], the coherence is also analyzed, which is

$$\mu(\Phi \otimes I_{\frac{q}{n}}) = \mu(\Phi). \quad (50)$$

In the PTP-CS, if the matrix  $P$  is the identity matrix, namely,  $P = I$ , then  $\mu(\Phi \otimes P_{(q/n)})$  can be transformed as

$$\begin{aligned} \mu(\Phi \otimes P_{\frac{q}{n}}) &= \mu(\Phi \otimes I_{\frac{q}{n}}) \\ &= \max\{\mu(\Phi), \mu(I_{\frac{q}{n}})\} \\ &= \mu(\Phi). \end{aligned} \quad (51)$$

Equation (51) implies that the coherence of the measurement matrix in STP-CS is a special case of the coherence in PTP-CS. It proves that the PTP-CS is a generalization of the STP-CS once again, and the PTP-CS provides more choices for the measurement matrix. So it has a better universality in the CS.

By [57], we have the relation between the spark and the coherence of an arbitrary matrix  $\Phi$  as  $\text{spark}(\Phi) \geq 1 + (1/\mu(\Phi))$ . So we can get the following corollary by Theorem 2.

*Corollary 2:* If  $k < (1/2)(1 + [1/\mu(\Phi)])$  and  $k < (1/2)(1 + [1/\mu(P)])$ , then for each measurement vector  $y \in R^m$  there exists at most one signal  $x \in \Sigma_k$  such that  $y = \Phi \otimes x$ .

#### D. RIP

Here, we will present a theorem of the RIP in the PTP-CS.

*Lemma 1* [58]: If the matrices  $X_1, X_2, \dots, X_M$  satisfy the RIP of the sequence  $k$ , which is denoted as  $\delta_k(X_1), \delta_k(X_2), \dots, \delta_k(X_M)$ , then

$$\delta_k(X_1 \otimes X_2 \otimes \cdots \otimes X_M) \leq \prod_{i=1}^M (1 + \delta_k(X_i)) - 1. \quad (52)$$

*Theorem 3:* Suppose the measurement matrix  $\Phi$  satisfies the RIP of a sequence  $k$ , which is denoted as  $\delta_k^\Phi$  ( $0 < \delta_k^\Phi < 1$ ), meanwhile, the matrix  $P$  satisfies  $\delta_k^P$  ( $0 < \delta_k^P < 1$ ). Then  $\Phi \otimes P_{q/n}$  satisfies  $\delta_k^{\Phi \otimes P_{q/n}}$ , and we have

$$\delta_k^{\Phi \otimes P_{q/n}} \leq (1 + \delta_k^\Phi) \cdot (1 + \delta_k^P) - 1 \quad (53)$$

namely,

$$\delta_k^{\Phi \otimes P_{q/n}} \leq \delta_k^\Phi \cdot \delta_k^P + \delta_k^\Phi + \delta_k^P. \quad (54)$$

#### E. Reconstruction

We will give a new algorithm by the IRLSs [46] based on  $l_\rho$ -minimization. Because of the discontinuity of  $l_0$ -norm [59], it is an NP-hard problem to get the  $l_0$ -minimization directly. As for the signal reconstruction, the authors in [60]–[62] recovered the signal by getting the  $l_1$ -minimization and using the

IRLSs minimization. In the iteration process, we get the vector weight of next iteration by convex optimization, and it runs until it satisfies the stopping criterion. Compared with the  $l_1$ -norm, the  $l_\rho$ -norm can obtain a more accurate result with less measurements. Also, if there is some noise in the signal, the  $l_\rho$ -norm can keep a good performance in the reconstruction [63].

Here, we begin to present more details about the proposed reconstruction algorithm based on the IRLS in the sense of PTP. Let the measurement matrix  $\Phi \in R^{m \times n}$  ( $m < n$ ). Obviously,  $\Phi$  is *row full rank* denoted as  $\text{rank}(\Phi) = m$ .  $\|x\|_0$  is the  $l_0$ -norm of  $x \in R^q$ , then  $\|x\|_0 < k$ . We obtain the following equation of the linear system:

$$y = \Phi \times x \quad (55)$$

where  $y \in R^{(mn/q)}$ . Because  $m < n$ , (55) has infinitely many solutions for any  $y$ . Let the set of all solutions be  $f(y) = \Phi^{-1}(y)$ .

Based on the  $l_\rho$ -minimization with  $0 < \rho < 1$ , we optimize the above problem, and let  $\|x\|_\rho$  be the  $l_\rho$ -norm of  $x$ . Then

$$\arg \min_x \|x\|_\rho = \left( \sum_{i=1}^q |x_i|^\rho \right)^{\frac{1}{\rho}}. \quad (56)$$

The approximate solution of  $l_\rho$ -minimization can be considered as

$$\arg \min_x \|x\|_\rho = \frac{1}{\rho} \left( \sum_{i=1}^q x_i^2 + \sigma^{1+\rho} \right)^{\frac{\rho}{2}}. \quad (57)$$

In the iteration process, we define the function for the  $l_\rho$ -minimization as

$$\mathcal{L}_\rho(x, w, \sigma) = \frac{\rho}{2} \left[ \sum_{i=1}^q x_i^2 w_i + \sum_{i=1}^q \left( \sigma^2 w_i + \frac{2-\rho}{\rho} w_i^{\frac{\rho}{\rho-2}} \right) \right]. \quad (58)$$

In (58), note  $w$  is the weight vector and  $w \in R^q$ . Let the initial value of  $w$  be  $w^{(0)} = (1, \dots, 1)$ ,  $\sigma_0 = 1$ . The update equation of  $w$  is

$$w_i^{(n)} = \left( (x_i^{(n)})^2 + \sigma_n^{1+\rho} \right)^{\frac{2-\rho}{\rho}}. \quad (59)$$

So we can obtain

$$x^{(n+1)} = \arg \min_{x \in f(y)} \mathcal{L}_\rho(x^{(n)}, w^{(n)}, \sigma_n). \quad (60)$$

In the iteration process,  $\sigma_n$  can be updated as

$$\sigma_{n+1} = \min \left( \sigma_n, \frac{r(x^{(n+1)})_{k+1}}{q} \right) \quad (61)$$

where  $k$  is an RIP sequence that the signal  $x$  can satisfy, in other words, it can be considered as the sparsity of the signal  $x$ .  $r(x)$  is the absolute value of each component in the vector  $x$ , and it obeys the descending order, namely,  $r(x)_1 \geq r(x)_2 \geq \dots \geq r(x)_q \geq 0$ .  $r(x^{(n+1)})_{k+1}$  is the  $(k+1)$ th component of  $x^{(n+1)}$  and it is sorted by the decreasing order. The terminal

---

### Algorithm 1 PTP-CS Reconstruction Algorithm

---

**Input:** Signal  $x$ .

**Initialize:**  $w^{(0)} = (1, \dots, 1)$ ,  $\sigma_0 = 1$ ,  $x^{(0)} = (1, \dots, 1)$

**Output:** The original signal  $x$

- 1) **Set**  $x$  as  $k$ -sparse,  $x \in R^q$ ,  $y \in R^{mn/q}$  as the measurement vector
  - 2) **for** each column  $i$  in  $y$
  - 3)     **while**  $\sigma$  satisfies the condition, we set
  - 4)         update  $w_i^{(n)} \leftarrow ((x_i^{(n)})^2 + \sigma_n^{1+\rho})^{\frac{2-\rho}{\rho}}$ ;
  - 5)         get  $H_n \leftarrow \frac{1}{w_i^{(n)}}$ ;
  - 6)         update  $x^{(n+1)} \leftarrow (\Phi \times H^T)^T \cdot [\Phi \times (\Phi \times H^T)^T]^{-1} \cdot y$ ;
  - 7)         update  $\sigma_{n+1} \leftarrow \min(\sigma_n, \frac{r(x^{(n+1)})_{k+1}}{q})$ ;
  - 8)     **end while**
  - 9)     **return**  $x^{(n+1)}$
  - 10) **end for**
- 

condition of the iteration is  $\sigma_n = 0$ , and we can get the sparse solution.

Let  $H_n \in R^q$  be a diagonal matrix, and we have

$$H_{i,i} = \frac{1}{w_i^{(n)}} \quad (62)$$

where  $H_{i,i}$  represents the  $i$ th diagonal element of matrix  $H$ , and  $i = 1, 2, \dots, q$ . Equations (56)–(62) are the main process of IRLS cited in [51].

In conclusion, we can get the solution of (55) as follows:

$$x^{(n+1)} = H \cdot (\Phi \otimes P)^T \cdot [(\Phi \otimes P) \cdot H \cdot (\Phi \otimes P)^T]^{-1} \cdot y. \quad (63)$$

Based on Properties 2 and 3, (63) can be converted as

$$x^{n+1} = \left\{ [H \cdot (\Phi \otimes P)^T]^T \right\} \cdot \left\{ (\Phi \otimes P) \cdot [H \cdot (\Phi \otimes P)^T]^T \right\}^{-1} \cdot y \quad (64)$$

$$x^{n+1} = [(\Phi \otimes P) \cdot H^T]^T \cdot \left\{ (\Phi \otimes P) \cdot [(\Phi \otimes P) \cdot H^T]^T \right\}^{-1} \cdot y \quad (65)$$

$$x^{(n+1)} = \left( \Phi \times H^T \right)^T \cdot \left[ \Phi \times (\Phi \times H^T)^T \right]^{-1} \cdot y. \quad (66)$$

In order to reconstruct the original signal  $x$ , the key point is to get the matrix  $H_n$  in the IRLS. According to above derivation process, we find that the  $H_n$  can be obtained by using (59), (61), (62), (66). As a result, we summarize the main process of the reconstruction method in Algorithm 1.

According to [51], for the  $n$ th iteration in Algorithm 1 with  $n \geq 0$ , we have

$$\begin{aligned} & \mathcal{L}_\rho(x^{(n)}, w^{(n)}, \sigma_n) \\ &= \frac{\rho}{2} \left[ \sum_{i=1}^q (x_i^{(n)})^2 w_i^{(n)} + \sum_{i=1}^q \left( \sigma^2 w_i^{(n)} + \frac{2-\rho}{\rho} (w_i^{(n)})^{\frac{\rho}{\rho-2}} \right) \right] \end{aligned} \quad (67)$$

where  $w_i^{(n)} > 0$ ,  $(x_i^{(n)})^2 \geq 0$ ,  $\sigma_n > 0$ . Similarly, we can get the monotonicity of  $\mathcal{L}_\rho(x^{(n)}, w^{(n)}, \sigma_n)$  by [51] as follows:

$$\begin{aligned} & \mathcal{L}_\rho(x^{(n+1)}, w^{(n+1)}, \sigma_{n+1}) \\ & \leq \mathcal{L}_\rho(x^{(n+1)}, w^{(n)}, \sigma_{n+1}) \\ & \leq \mathcal{L}_\rho(x^{(n+1)}, w^{(n)}, \sigma_n) \\ & \leq \mathcal{L}_\rho(x^{(n)}, w^{(n)}, \sigma_n) \end{aligned}$$

where  $0 < \rho < 1$ , and the algorithm is converged, namely

$$\lim_{n \rightarrow \infty} (x^{(n)} - x^{(n+1)}) = 0. \quad (68)$$

#### F. Compressed Sensing of High-Dimensional Signal

The high-dimensional signal, for example, the video signal, is different from the traditional image signal. The video adds the time dimension to the signal flow. In brief, the video signal is composed in a manner of frame-by-frame. So when it comes to compressing the video signal, we should focus on compressing every frame image in the video. There are some methods to process the video signal, for instance, the Kronecker product in CS can be suitable for various types of signal structures [58]. Moreover, the generalized tensor compressive sensing (GTCS) gave a further study of the Kronecker product in CS [64], [65], and Friedland *et al.* proposed a unified framework for CS with higher order tensors. The GTCS not only keeps the intrinsic structure of the tensor data, but also has a lower computational complexity for the signal reconstruction.

The Kronecker product [58] can expand the sparse basis to a high-dimensional one as

$$\begin{aligned} \Psi &= \Psi_1 \otimes \Psi_2 \otimes \cdots \otimes \Psi_N \\ &= \{\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_N, \psi_i \in \Psi_N, 1 \leq i \leq N\} \end{aligned} \quad (69)$$

where  $\psi_i$  is a sparse basis of the video signal. Suppose that the video signal is  $X$ , whose size is  $N_1 \times N_2 \times N_3$ . According to the form of the sparse basis defined in (69), we can expand the video signal  $X$  as

$$\text{vec}(X) = (\text{vec}(U_1)^T, \dots, \text{vec}(U_{N_3})^T)^T \in R^{N_1 \cdot N_2 \cdot N_3} \quad (70)$$

where  $\text{vec}(U_i)$  is the expansion of the  $i$ th frame image. Therefore, the Kronecker product in CS can be expressed as

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_3} \end{pmatrix} = \begin{pmatrix} \Psi_{N_3} & 0 & \cdots & 0 \\ 0 & \Psi_{N_3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Psi_{N_3} \end{pmatrix} \cdot \text{vec}(X). \quad (71)$$

*Lemma 2* [58]: Let  $\Phi_i, \Psi_i$  be the bases or frames for  $R^{N_i}$  where  $i = 1, \dots, N$ , then

$$\mu(\Phi_1 \otimes \cdots \otimes \Phi_N, \Psi_1 \otimes \cdots \otimes \Psi_N) = \prod_{i=1}^N \mu(\Phi_i, \Psi_i). \quad (72)$$

By (72), we can see that the Kronecker product only focuses on compressing the whole video signal rather than every frame image. For each frame of the video signal, the storage space of the measurement matrix is very large. So for the high-dimensional signal, the PTP-CS provides a method to save the storage of each measurement matrix. In the Kronecker CS, the measurement matrix  $\Phi_i$  ( $i = 1, 2, \dots, N$ ) can be expressed as  $(\Phi'_i \otimes P_i)$  ( $i = 1, 2, \dots, N$ ), where  $\Phi'_i$  can be enlarged by making the tensor product with the matrix  $P$ . Then by using the PTP-CS, similarly with Lemma 2, we give the following theorem.

*Theorem 4*: Let  $\Phi_i, \Psi_i$  be the bases or frames for  $R^{N_i}$ , and  $\Phi_i = (\Phi'_i \otimes P_i)$ ,  $i = 1, \dots, N$ , then  $\mu\{(\Phi'_1 \otimes P_1) \otimes \cdots \otimes$

$(\Phi'_N \otimes P_N), \Psi_1 \otimes \cdots \otimes \Psi_N\} =$

$$\prod_{i=1}^N \mu(\Phi'_i \otimes P_i, \Psi_i). \quad (73)$$

*Proof*: The proof is omitted since it is similar with that of Lemma 2. ■

## IV. EXPERIMENTAL RESULTS

In order to evaluate the performance of PTP-CS in signal recovery, we compare the PTP-CS with the traditional CS, i.e., the measurement matrix  $\Phi$  has the matching dimension with the dimension of the signal, and the recovery method is OMP. Furthermore, we pay more attention to compare the difference between PTP-CS and STP-CS. Meanwhile, we analyze the time and the storage complexity of the PTP-CS model. In addition, we choose the original matrix  $\Phi$  of PTP-CS as Gaussian matrix in the following experiments. We will analyze the proposed model by visual comparison and give more recovery performances of the signals. In the following experiments, we denote the *image compression ratio*  $m/n$  as  $\theta$  and we define the *dimension reduction multiples* of the measurement matrix  $\Phi$  as  $\eta$ .

For image signals, we choose different types of pictures with different sizes to test the performance of the proposed model. The recovery performance of the image signal is evaluated by peak signal to noise ratio (PSNR), which is defined as

$$\text{PSNR} = 10 \lg \left[ \frac{(2^\xi - 1)^2}{\text{MSE}} \right] \quad (74)$$

where  $\xi$  is the bit of each pixel and it is selected as 8 commonly, that is, the gray scale is 255. In this experiment, we set  $\xi = 8$ . The mean square error (MSE) is used to measure the error between the original image and the recovered image. Its mathematical definition is

$$\text{MSE} = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (O(i, j) - R(i, j))^2 \quad (75)$$

where  $m, n$  denote the width and the height of the image,  $O(i, j)$  is the gray value of the original image, and  $R(i, j)$  is the gray value of the recovered image. For 1-D signal, we will analyze the relationship between the recovery performance and the dimension of the measurement matrix.

The *recovery performance* of the signal can be measured as the following relative error:

$$\epsilon = \frac{\|x' - x\|_2}{\|x\|_2} \quad (76)$$

where  $x$  is the original signal,  $x'$  is the recovered signal and  $\|\cdot\|_2$  is the  $l_2$ -norm of vector.

In terms of signal recovery, the IRLS based on  $l_\rho$ -minimization has a good reconstruction performance when  $\rho$  is fairly large. Therefore, we set  $\rho = 0.8$  in the experiments.



Fig. 4. Recovered images by different CS models. The original image is the image *Lena* which is the same as that in Fig. 2. The size of the original image is  $256 \times 256$ . (a) PTP-CS method, where the original measurement matrix  $\Phi$  and the matrix  $P$  are adopted as Gaussian matrix. (b) PTP-CS, where the original measurement matrix  $\Phi$  is selected as Gaussian matrix and the matrix  $P$  is adopted as the generalized permutation matrix. (c)–(f) Traditional CS method. The sparse images observed by the Gaussian matrix, the Bernoulli matrix, the Toeplitz matrix, and the chaotic matrix, respectively. The sizes of these different types of measurement matrices are  $192 \times 256$ , and the recovery method is OMP.

#### A. Visual Comparison

Suppose that the compression ratio  $\theta$  equals to 0.75, i.e., for a picture with size  $256 \times 256$ , the dimension of the measurement matrix is  $192 \times 256$ .  $\eta$  (the dimension reduction multiples of the measurement matrix  $\Phi$ ) equals to 4, i.e., for a  $192 \times 256$ -dimensional matrix, the dimension of this matrix is reduced to be  $48 \times 64$ . That is, the storage space needed by the measurement matrix is reduced dramatically.

In the experiment, first, we use the DWT to make the original image be sparse. Then for the traditional CS, the sparse image is observed by Gaussian matrix [26], Bernoulli matrix, Toeplitz matrix [27], and the chaotic matrix [28] in turn and it is reconstructed by OMP at last. In the PTP-CS, we select the original measurement matrix  $\Phi$  as Gaussian random matrix, then enlarge its dimension by  $\Phi \otimes P$ , and reconstruct it by our model finally. Fig. 4 shows the visual results of our model and the traditional four CS models.

Evidently, in aspect of visual results in Fig. 4, the PTP-CS has the same performance as the traditional CS when  $P$  is selected as different types of matrices. For each type of the CS model, we run the experiment for 50 times to obtain the mean values of PSNR. And we get the mean values of PSNR that are 36.9440, 37.0238, 30.8170, 30.8917, 31.3666, and 30.9748. In which the first two values belong to the PTP-CS, where  $P$  is a Gaussian matrix and  $P$  is a generalized permutation matrix, respectively. Moreover, the last four values belong to the traditional CS in which the measurement matrix is a Gaussian matrix, a Bernoulli matrix, a Toeplitz matrix, and a chaotic matrix, respectively. The sizes of these different types of measurement matrices are  $192 \times 256$ . In conclusion, the PTP-CS can ensure a similar performance with a PSNR value higher than 30 dB.

TABLE I  
PSNRs OF IMAGE RECOVERY FOR DIFFERENT TYPES OF  
MATRICES WITH DIFFERENT  $\eta$

$\eta$	$\theta$	maximum(dB)		minimum(dB)		mean(dB)	
		Gaussian	$g$	Gaussian	$g$	Gaussian	$g$
16	0.25	22.5787	22.5169	6.3698	6.1050	16.3619	15.9039
	0.5	28.7600	30.2680	22.3381	23.2418	26.4795	26.7151
	0.75	37.4620	36.3172	31.8318	32.1375	34.9089	34.8229
8	0.25	21.6755	22.6516	8.0949	8.6762	13.9530	16.6849
	0.5	28.5110	28.4920	26.2418	26.2664	27.4114	27.8872
	0.75	36.5978	36.4131	35.2088	34.3530	35.7890	35.7383
4	0.25	19.4140	21.5296	9.2974	9.0465	14.3978	15.6644
	0.5	29.4151	29.5317	27.2450	27.5953	28.8031	28.7376
	0.75	37.0238	36.9440	35.6981	35.3792	36.2517	36.3167
2	0.25	19.8594	20.6405	11.4373	9.7635	15.2074	15.8026
	0.5	29.5485	29.5041	29.1387	28.5285	29.3115	29.1774
	0.75	36.7877	36.8371	36.4490	36.2658	36.6033	36.4800
1	0.25	21.7260	21.4394	8.8370	10.8076	14.3978	15.2267
	0.5	29.7367	29.6226	27.2450	28.9099	28.8031	29.3201
	0.75	37.0238	36.8933	35.1563	36.3452	36.2517	36.5710

#### B. Quantitative Analysis of 2-D Image

In order to verify the effect of  $\eta$  and the choice of matrix  $P$ , we set  $\theta$  as 0.25, 0.5, 0.75 for the image with size  $256 \times 256$ , and we select  $\eta$  as 1, 2, 4, 8, 16. If  $\eta = 1$ , it converts to the traditional CS. In every experiment,  $\Phi$  is Gaussian matrix and  $P$  is Gaussian matrix or  $g$  (i.e., the generalized permutation matrix). We also take the image *Lena* as the experimental image, whose size is  $256 \times 256$ , and we repeat the experiment for 50 times. In order to investigate the performance of the PTP-CS method, Fig. 5 shows the PSNRs of two types of  $P$  with different  $\eta$  and  $\theta$ ,  $\eta$  in Fig. 5(a)–(e) equals to 1, 2, 4, 8, and 16, respectively, the matrix  $P$  in the left five pictures of Fig. 5 is Gaussian matrix and in the right five pictures of Fig. 5 it is  $g$ . Obviously, we can see from Fig. 5 that the PSNR is unstable with the increment of  $\eta$ , and the types of matrix  $P$  have little influence on the recovery performance. Its reason is that, as we have inferenced in Section II, the coherence coefficient  $\mu$  depends on the matrix with bigger size for these two types of matrices in the tensor product. Since  $\mu(g)$  is 0 calculated by (5),  $\mu(\Phi \otimes g)$  only depends on  $\mu(\Phi)$  of the Gaussian matrix which is subjected to the normal distribution with mean 0 and variance 1. Meanwhile, it is clear that smaller  $\theta$  makes the result of the PSNR less stable.

Table I shows the maximum, the minimum, the mean values of PSNR for 50 experiments under different values of  $\eta$  and  $\theta$ , where  $g$  is Gaussian matrix and the generalized permutation matrix, respectively. The method is PTP-CS. The first column in Table I is the values of  $\eta$ , which are selected as 16, 8, 4, 2, 1, respectively. The second column in Table I is the values of  $\theta$ , which are set as 0.25, 0.5, 0.75 for each value of  $\eta$ , respectively. The remaining six columns of Table I list the maximum, the minimum, the mean values of PSNR corresponding to the values of  $\eta$  and  $\theta$  when the matrix  $P$  is Gaussian random matrix and  $g$ . Although in Fig. 5 the recovery results of the image fluctuate significantly as  $\eta$  is big, we can see from Table I that the minimum of  $\eta$  has a slightly different result, and its big difference is about 5 dB.

In order to compare with the traditional semi-tensor CS [19], in the PTP-CS, we set  $P$  as the identity matrix and we set  $\Phi$  as the Gaussian matrix, then we take the image *Lena* ( $256 \times 256$ )

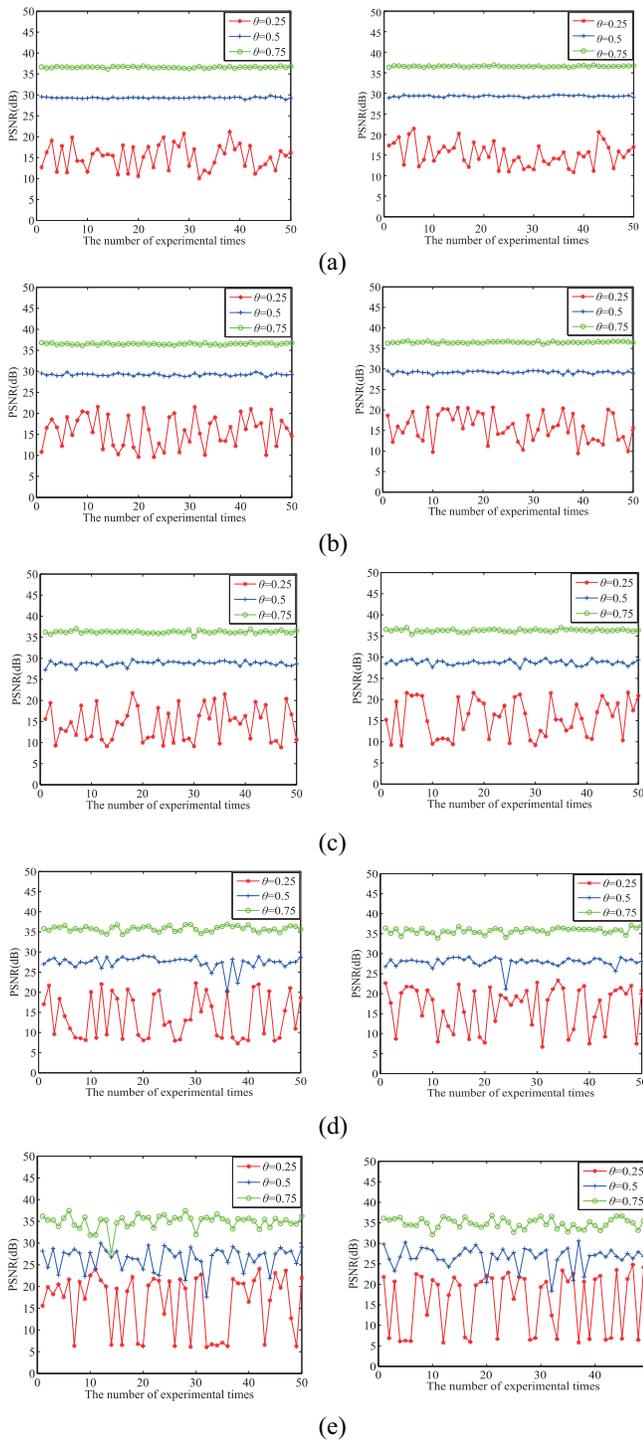


Fig. 5. Value of PSNR varies with  $\eta$  and  $\theta$  as well as two choices of  $P$  in the 50 experiments. The method is PTP-CS. The original image is the image *Lena* ( $256 \times 256$ ). The left five pictures are the results of  $P$  as the Gaussian matrix, and the right five pictures are the results when  $P$  is selected as the generalized permutation matrix. (a)  $\eta = 1$ , (b)  $\eta = 2$ , (c)  $\eta = 4$ , (d)  $\eta = 8$ , and (e)  $\eta = 16$ . Three curves in each picture correspond to the values of  $\theta$  with 0.25, 0.5, and 0.75, respectively. The x-coordinate in each picture represents the number of the experimental times. The y-coordinate represents the value of PSNR in each experiment.

for the experiment. Here, we also take  $\eta$  as 1, 2, 4, 8, 16 and we select  $\theta$  as 0.25, 0.5, and 0.75, respectively, which is a similar process with the above experiment. The experiment has been

TABLE II  
VARIANCES OF THREE CS MODELS WITH DIFFERENT  $\eta$  AND  $\theta$

$\eta$	$\theta$	Gaussian	$g$	STP
16	0.25	51.7343	44.1460	34.9061
	0.5	6.1643	6.6790	2.6539
	0.75	1.6100	2.7210	1.3853
8	0.25	28.8479	29.0261	28.0550
	0.5	1.5103	2.5074	3.4768
	0.75	0.4970	0.4473	0.5179
4	0.25	19.2769	16.1878	20.6556
	0.5	0.3027	0.2799	0.2628
	0.75	0.0822	0.0938	0.1333
2	0.25	11.7415	8.9553	15.3943
	0.5	0.0822	0.0292	0.0415
	0.75	0.0320	0.0268	0.0337
1	0.25	8.1302	16.1878	8.5201
	0.5	0.0354	0.2799	0.0188
	0.75	0.0177	0.0938	0.0161

repeated for 50 times. We also record the values of PSNR for the STP-CS method during 50 experiments. Then we calculate the variances of STP-CS and PTP-CS in different values for  $\eta$  and  $\theta$ , which is shown in Table II. The first column of Table II is the values of  $\eta$  (the dimension reduction multiples of the measurement matrix  $\Phi$ ) which are selected as 16, 8, 4, 2, 1, respectively. The second column of Table II is the values of  $\theta$  (the compression ratio of the image signal) which are set as 0.25, 0.5, 0.75 for each value of  $\eta$ , respectively. The remaining three columns of Table II list the variances of PTP-CS as  $P$  is Gaussian matrix and  $g$ , and the variances of STP-CS under different values of  $\eta$  and  $\theta$ .

From Table II, we can find that if the values of  $\eta$  are selected as 2, 4, 8 in the PTP-CS method, STP-CS has a little bigger variance than PTP-CS, i.e., the PSNR values obtained from STP-CS method have higher dispersion degree than those obtained from the PTP-CS method, since the variance value measures how far a set of (random) numbers are spread out from their average value. That is, the randomness of the STP-CS method is higher than that of the PTP-CS method. In order to make the difference more clear, we record the PSNRs for 50 times in the experiments of STP-CS and PTP-CS. As is shown in Fig. 6(d) and (e), we can easily find that the PTP-CS method has a better performance in stability than the STP-CS method.

Meanwhile, in Fig. 6, we can compare the recovery performances of the PTP-CS and the STP-CS from the visual aspect. Fig. 6(a) is the original image *Pepper* with size  $256 \times 256$ , Fig. 6(b) is the recovery image of PTP-CS and Fig. 6(c) is the recovery image of STP-CS.  $\theta$  is 0.75 in PTP-CS and STP-CS. It seems that the PTP-CS has a better performance.

Actually, the reason why the PTP-CS has a better performance than STP-CS can be explained by the theorems given in Section II. The Gaussian matrix is a kind of random matrix, which is different in each generation. The STP-CS only expands the dimension of the original measurement matrix without changing the properties of the matrix, including the spark, the coherence, and so on. Therefore, the properties of the final measurement matrix for the STP-CS only depend on one random matrix. There is no doubt that the randomness of STP-CS is stronger than that of PTP-CS. Unlike in the STP-CS, both the original measurement matrix and  $P$  are Gaussian

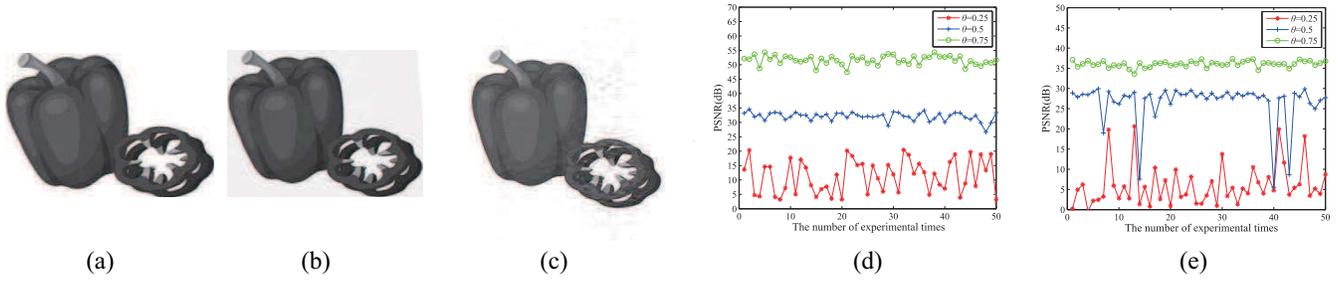


Fig. 6. Recovery results of the image *Pepper* ( $256 \times 256$ ) as  $\eta = 4$ , and two models are PTP-CS and STP-CS. In the PTP-CS,  $P$  is the Gaussian matrix. (a) original image, (b) recovery image of PTP-CS when  $\theta$  is 0.75, (c) recovery image of STP-CS when  $\theta$  is 0.75, (d) three curves correspond to different  $\theta$  whose values are 0.25, 0.5, and 0.75, respectively, and each curve shows one value of PSNR in one experiment, and (e) three curves of PSNR in STP-CS, and these curves correspond to different  $\theta$  whose values are 0.25, 0.5, and 0.75, respectively.

matrices in the PTP-CS, and the coherence is impacted by two matrices. So the randomness of PTP-CS is reduced to some extent compared with STP-CS.

Moreover, we want to reduce the difference or the gap between the original signal  $x$  and the reconstructed signal  $y$ , in STP-CS, it can be denoted as

$$x_{\text{STPCS}}^* = \operatorname{argmin} \|(\Phi \otimes I)x - y\|_2. \quad (77)$$

And in PTP-CS, it can be denoted as

$$x_{\text{PTPCS}}^* = \operatorname{argmin} \|(\Phi \otimes P)x - y\|_2. \quad (78)$$

When the original measurement matrix  $\Phi$  is known, the  $(\Phi \otimes I)$  is known in STP-CS, whereas in PTP-CS the final measurement matrix, i.e., the  $(\Phi \otimes P)$  is still unknown. So the PTP-CS has the opportunity to adjust the coherence value of the final measurement matrix. By some optimization algorithms, the PTP-CS can find a better  $P$  to make its column correlation smaller. In the near future, we will study its optimization method to find better  $P$ .

Through Theorem 2 and the above experiments, the tensor product of the measurement matrix can be generated by many different kinds of matrices. The smaller the coherence  $\mu$  is, the better recovery results we can get. Hence, if we can calculate the  $\mu$  of two tensor matrices, we can predict the approximate recovery result.

For various types of images, besides the image *Pepper*, we choose the image *Vampire* with size  $256 \times 256$  to simulate additionally 50 times, for the sake of adding variation to the analysis. Suppose  $\eta$  (i.e., the dimension reduction multiples of the measurement matrix  $\Phi$ ) is 1 and 4, respectively. When  $\eta$  is 1, the PTP-CS degenerates to the traditional CS. In this experiment, let the original measurement matrix and  $P$  be the Gaussian matrices. Fig. 7 shows the recovery images of the image *Vampire*, where  $\eta$  is 4 and  $\theta$  is 0.75, the method is PTP-CS, and Fig. 7(c) also shows the fluctuations of PSNR, respectively. It can be seen from Fig. 7 that from the visual aspect, the PTP-CS performs well on various types of images. For other types of images, we simulated for many times and obtained similar results. Table III lists the maximum, the minimum and the mean values of PSNR for the image *Vampire* in 50 simulations in detail, where  $\eta$  has two values, i.e., 1 and 4. We can see from Table III that although it is unstable when  $\eta$  is 4, the maximum, the minimum, and the mean values of PSNR are similar with those of the traditional CS.

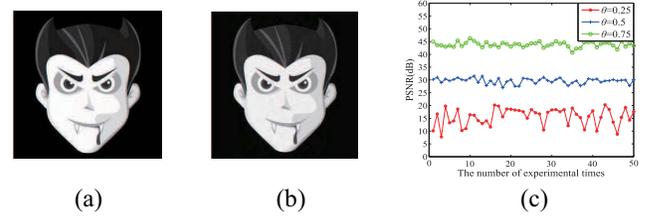


Fig. 7. Recovery results of the image *Vampire* ( $256 \times 256$ ) as  $\eta = 4$ , where  $P$  is the Gaussian matrix. (a) original image, (b) recovery image when  $\theta$  (i.e., the compression ratio) is 0.75, and (c) three curves of PSNR, which correspond to different  $\theta$  whose values are 0.25, 0.5, and 0.75, respectively. The x-coordinate in the (c) represents the number of the experimental times, and the y-coordinate in the (c) represents the value of PSNR in each experiment.

TABLE III  
PSNRs OF IMAGE *Vampire* WHEN  $\theta$  AND  $\eta$  HAVE DIFFERENT VALUES

Name	$\theta$	maximum(dB)		minimum(dB)		mean(dB)	
		$\eta=4$	$\eta=1$	$\eta=4$	$\eta=1$	$\eta=4$	$\eta=1$
Vampire	0.25	20.3184	19.6010	7.7257	12.5732	15.6285	16.6170
	0.5	31.5262	32.0433	26.9700	29.5698	29.6222	30.8814
	0.75	46.2541	46.4838	40.6949	44.4215	43.8152	45.5455

Meanwhile, we choose the images *Cameraman*, *Baboo*, *Barbara*, and *House* ( $256 \times 256$ ) for the performance test. Here, the  $P$  is  $g$  (i.e., the generalized permutation matrix),  $\theta$  is 0.75 and  $\eta$  is 4. The experiment repeats 50 times for each image. As is shown in Fig. 8, four images in the top row are the original images, four images in the middle row are the recovery images with  $\theta$  0.75, and four pictures in the bottom row show the curves of PSNR for each type of image under different values of  $\theta$ . Since the experiment is repeated 50 times for each pictures, three curves in Fig. 8 (i)–(l) contain 50 points, which are the values of PSNR in every experiment. In a word, no matter what type or size the images are, the PTP-CS can recover the image signal to a good extent compared with the traditional CS.

It is worth testing the performance of PTP-CS on various types of images with different sizes. Therefore, we adopt another image *Lena* with size  $512 \times 512$ . The original measurement matrix and  $P$  are selected as the Gaussian matrix for a case of study. Table IV lists the maximum, the minimum and the mean values of PSNR of the image *Lena* ( $512 \times 512$ ) with different  $\theta$  and  $\eta$ . Compared with the traditional CS as  $\eta = 1$ , we can see from Table IV that the maximum and

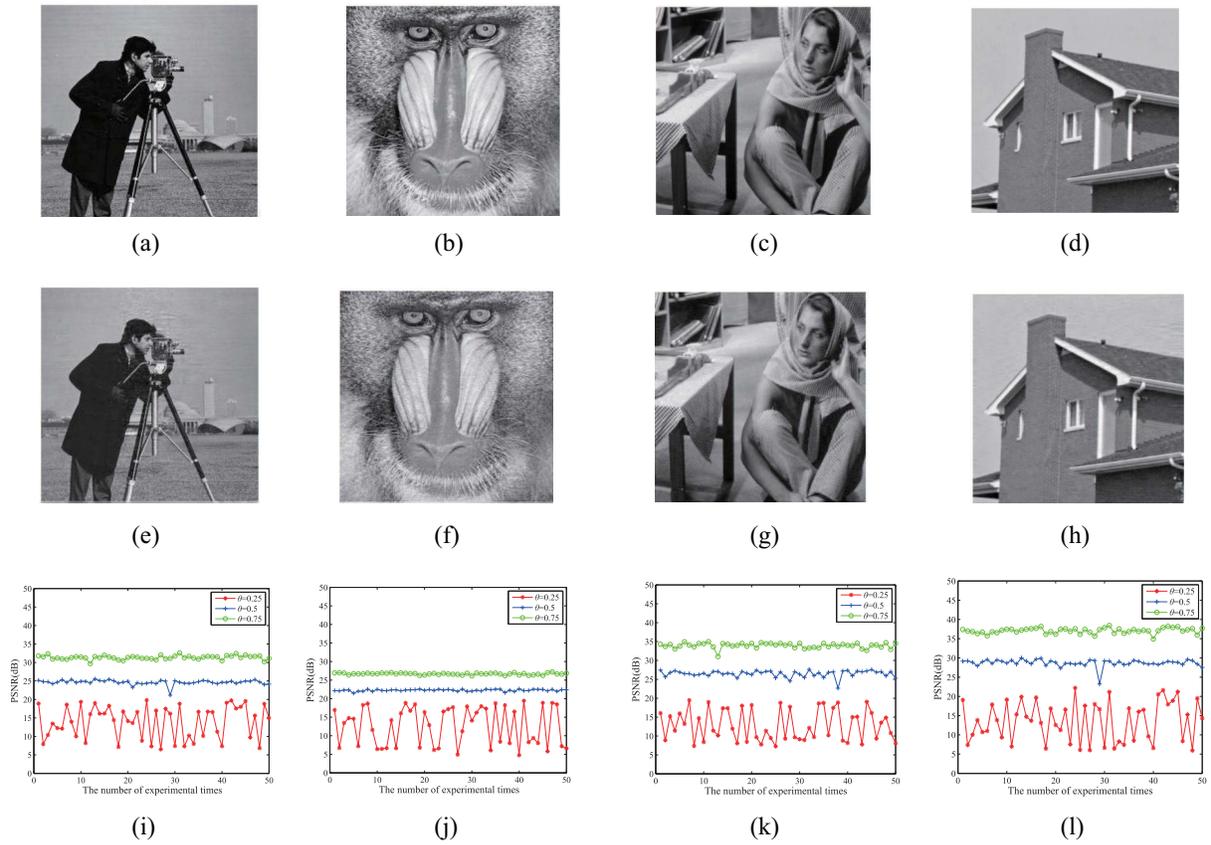


Fig. 8. Recovery results of the images *Cameraman*, *Baboo*, *Barbara*, and *House* ( $256 \times 256$ ) as  $\eta = 4$ , and  $P$  is the generalized permutation matrix. Four original images (a)–(d) with sizes  $256 \times 256$  are placed in the top row. (a) *Cameraman*, (b) *Baboo*, (c) *Barbara*, (d) *House*. The middle four images (e)–(h) are the recovered images of these four original images in the first row. The value of  $\theta$  in the images (e)–(h) is 0.75. The bottom four pictures (i)–(l) display the values of PSNR under three values of  $\theta$  for each type of image. There are three curves of PSNR in each picture, which correspond to different  $\theta$  whose values are selected as 0.25, 0.5, and 0.75, respectively. The  $x$ -coordinate in the (i)–(l) represents the number of the experimental times, and the  $y$ -coordinate in the (i)–(l) represents the value of PSNR in each experiment. (i) PSNR curves of the image *Cameraman*, (j) PSNR curves of the image *Baboo*, (k) PSNR curves of the image *Barbara*, and (l) PSNR curves of the image *House*.

TABLE IV  
PSNRs OF THE IMAGE *Lena* ( $512 \times 512$ ) WITH DIFFERENT  $\theta$  AND  $\eta$

$\eta$	$\theta$	maximum(dB)	minimum(dB)	mean(dB)
16	0.25	22.5169	6.1050	15.9039
	0.5	30.2680	23.2418	26.7151
	0.75	36.3172	32.1375	34.8229
8	0.25	22.6516	8.6762	16.6849
	0.5	28.4920	26.2664	27.8872
	0.75	36.4131	34.3530	35.7383
4	0.25	21.5296	9.0465	15.6644
	0.5	29.5317	27.5953	28.7376
	0.75	36.9440	35.3792	36.3167
2	0.25	20.6405	9.7635	15.8026
	0.5	29.5041	28.5285	29.1774
	0.75	36.8371	36.2658	36.4800
1	0.25	21.4394	10.8076	15.2267
	0.5	29.6226	28.9099	29.3201
	0.75	36.8933	36.3452	36.5710

the mean values are quiet the same whereas the value of the minimum has a small gap. Fig. 9 shows the recovery curves for repeating 50 times as  $\eta = 8$ , and these three curves in Fig. 9 correspond to three values of  $\theta$ . As is shown in Fig. 9, when the image has a bigger size, the PSNR curve is similar with those of Fig. 5 when the dimension of the measurement matrix is reduced. In conclusion, the PTP-CS performs well for the images with different sizes.

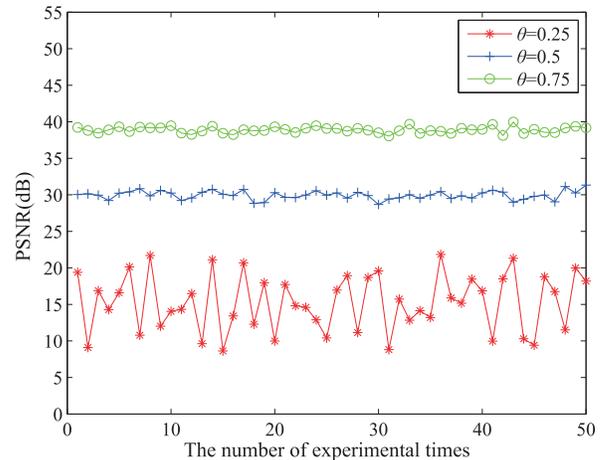


Fig. 9. PSNR curves of the image *Lena* ( $512 \times 512$ ) as  $\eta = 8$ . Here,  $P$  is the Gaussian matrix. From the top to the bottom, three values of  $\theta$  are 0.25, 0.5, and 0.75, respectively. Each curve shows one value of PSNR in 50 experiments.

### C. More Analyses for Reconstruction

In the CS, the recovery results may be diverting with different kinds of measurement matrices and different values of  $\theta$

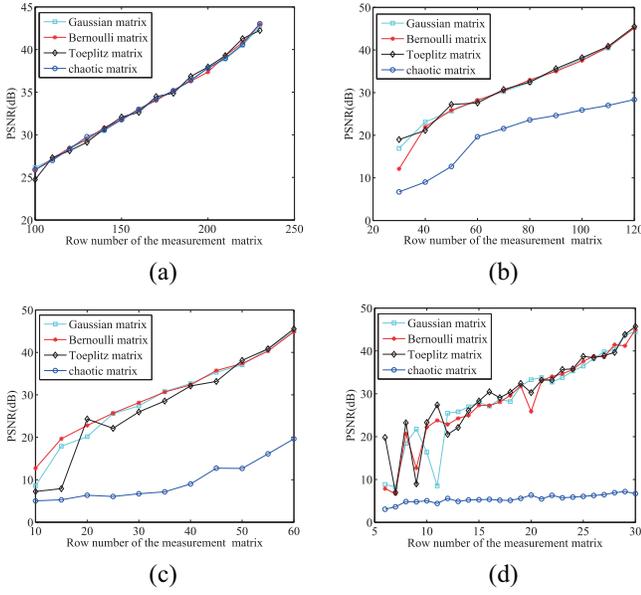


Fig. 10. Effect of the row number of the measurement matrix on the image recovery for the PTP-CS and the traditional CS with four values of  $\eta$ . We take the image *Lena* for the experiment. The PSNR changes with the row number of measurement matrices which are Gaussian, Bernoulli, Toeplitz, and chaotic matrices. (a)  $\eta$  is 1, (b)  $\eta$  is 2, (c)  $\eta$  is 4, and (d)  $\eta$  is 8.

(i.e., the compression ratio). The compression ratio is related with the row of the measurement matrix. Therefore, we adopt four types of matrices and different  $\theta$  to investigate the recovery performance of PTP-CS, and we choose the image *Lena* with size  $256 \times 256$  for the experiment. We set  $P$  as the Gaussian matrix, and we set the original measurement matrix  $\Phi$  as the Gaussian matrix, the Bernoulli matrix, the Toeplitz matrix, and the chaotic matrix in turns. Fig. 10 gives the graph of the relationship between PSNR and the row number of measurement matrix. With the increment of the row number, the compression ratio gradually becomes bigger. The  $x$ -coordinate calibration of these four subfigures in Fig. 10 are different, because the values of  $\eta$  as well as the column numbers in the original measurement matrices are different, and the ranges of the row numbers of the original measurement matrices are different. In order to show the variation of PSNR with the row number of the measurement matrix clearly, the  $x$ -coordinate calibration in Fig. 10(a) ranges from 100 to 250, in Fig. 10(b) it ranges from 20 to 120, in Fig. 10(c) it ranges from 10 to 60, and in Fig. 10(d) it ranges from 5 to 30.

When the value of  $\eta$  (i.e., the dimension reduction multiples of the measurement matrix  $\Phi$ ) is set as 1, as is shown in Fig. 10(a), it degenerates to the traditional CS, four kinds of matrices perform almost to be similar with the increment of the row number. When  $\eta$  is 2, as is shown in Fig. 10(b), these four matrices have different results when the row number is less than 60, and the results of the Gaussian matrix are the most stable. The result is the same if the row number is more than 60, while the result of the chaotic matrix remains unsatisfactory. When  $\eta$  is 4, as is shown in Fig. 10(c), the result of the chaotic matrix is still worse than those of other matrices, and the results of Gaussian and Bernoulli matrices are more stable than that of Toeplitz matrix. When  $\eta$  is 8, as is shown in

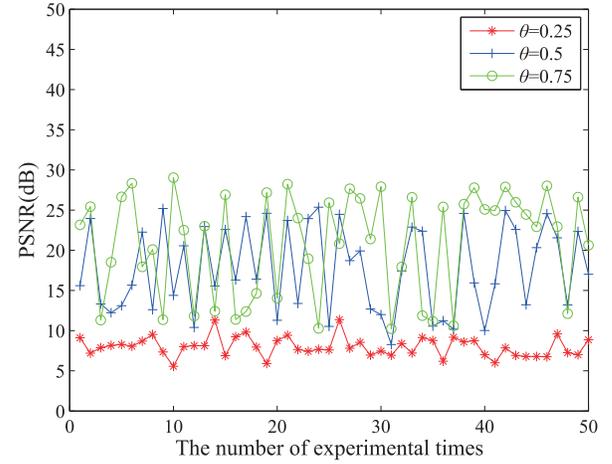


Fig. 11. Recovery result by OMP. The original image is the image *Lena* with  $256 \times 256$  and  $\eta$  is 4. These three curves are the PSNR curves with different  $\theta$ , which correspond to different  $\theta$  whose values are set as 0.25, 0.5, and 0.75, respectively. The  $x$ -coordinate in the picture represents the number of the experimental times, and the  $y$ -coordinate represents the value of PSNR in each experiment.

Fig. 10(d), the chaotic matrix has a lower PSNR curve whereas this curve is relatively flat. The performances of Gaussian, Bernoulli, and Toeplitz matrices fluctuate greatly when their row numbers are less than 15. When the row number is more than 15, the result of the Gaussian matrix performs well. In a word, the Gaussian matrix is the best matrix in comparison to other three kinds of matrices. It yields a stable result in terms of image recovery.

In order to investigate the performance of the reconstruction method,  $P$  is selected as the Gaussian matrix with size  $4 \times 4$ , i.e.,  $\eta$  is 4, the size of the measurement matrix  $\Phi$  is set as  $32 \times 64$ , and we use OMP to reconstruct the image *Lena* ( $256 \times 256$ ). The experiment has been repeated for 50 times, and Fig. 11 shows these 50 values of PSNR in 50 experiments. We can conclude from Fig. 11 that the OMP is not suitable for PTP-CS since the OMP produces a worse result and a more violent fluctuation. Hence, it is necessary to propose a new reconstruction model for PTP-CS.

The CPU time of PTP-CS is shown in Fig. 12. We record the CPU time of the whole process, including making the original image sparse, obtaining the final measurement matrix by PTP, compressing the original image and reconstructing the image. Five pictures in each row of Fig. 12 have different  $\eta$  (the dimension reduction multiples of the measurement matrix  $\Phi$ ), which is selected as 1, 2, 4, 8, 16 from the left to the right, respectively. We record the CPU times when  $P$  is Gaussian matrix or  $g$  (i.e., the generalized permutation matrix). The upper five pictures in Fig. 12 are the curves of the CPU time as  $P$  is Gaussian matrix, and the bottom five pictures in Fig. 12 are the curves of the CPU time as  $P$  is the generalized permutation matrix  $g$ . Three curves in each picture represents the variation of PSNR under three values of  $\theta$ , i.e., 0.25, 0.5, 0.75. Because of the randomness of the original measurement matrix, the CPU time fluctuates for different  $\eta$ . Therefore, we will continue to study new reconstruction algorithm for further optimization.

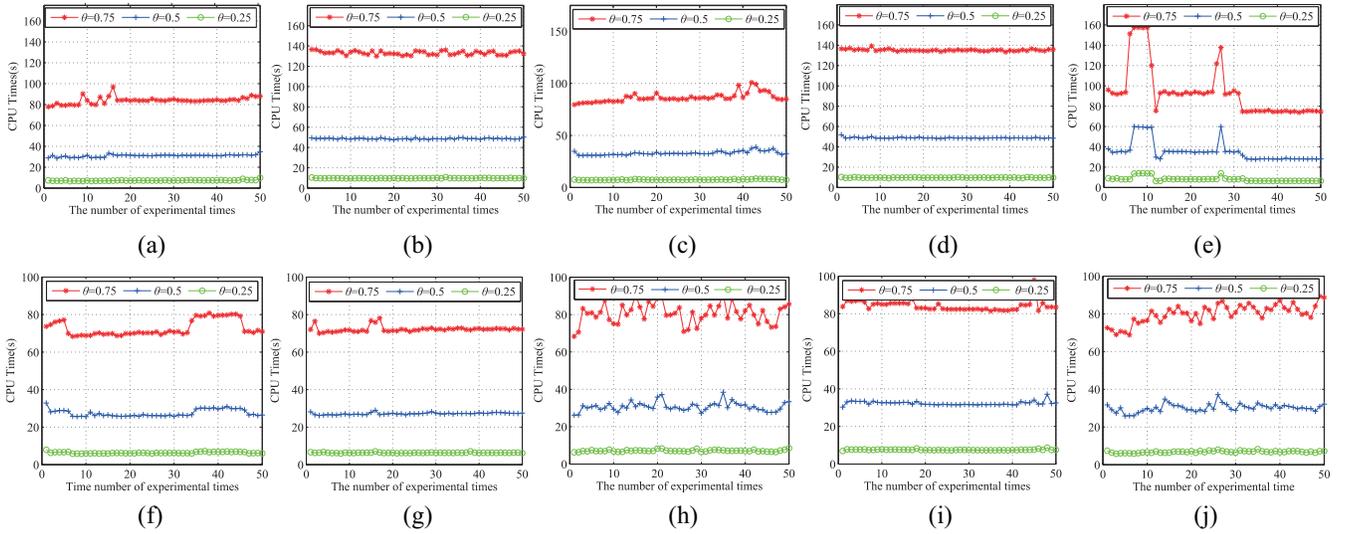


Fig. 12. CPU time of PTP-CS. The original signal is the image *Lena* ( $256 \times 256$ ), and this experiment is performed under different values of  $\eta$  and  $\theta$ , as well as two choices of  $P$ . The above five pictures (a)–(e) are the results of  $P$  selected as the Gaussian matrix, and the bottom five pictures (f)–(j) are the results when  $P$  is selected as the generalized permutation matrix. Three curves in each picture correspond to different values of  $\theta$ , whose values are 0.25, 0.5, and 0.75, respectively. The x-coordinate in each picture represents the number of the experimental times, and the y-coordinate represents the CPU time. (a) and (f)  $\eta = 1$ . (b) and (g)  $\eta = 2$ . (c) and (h)  $\eta = 4$ . (d) and (i)  $\eta = 8$ . (e) and (j)  $\eta = 16$ .

#### D. Quantitative Analysis of Video Frames

As is mentioned in Section III, we can compress the video signal by frames. We download a video and extract one frame from each pair of frames in the video, and then compress them by PTP-CS. Actually, two continuous frames are similar to each other, and the gap matrix between them is sparse. So for the video CS, first, we should obtain the gap matrix between two frames. Second, dispose the first frame as the regular process of CS, including making the image sparse and measuring the image. Third, in view of the sparsity of the gap matrix, we can measure the gap matrix directly without the sparse process. After transmitting in the signal channel, we can recover the first frame and the gap matrix, then the second frame can be obtained by the first frame and the gap matrix.

In the experiment, the size of two frames is  $256 \times 256$ ,  $\eta$  is 4, the matrix  $P$  is  $g$  (i.e., the generalized permutation matrix), and we repeat the experiment for 50 times. As is shown in Fig. 13, Fig. 13(a) is the original image of the first frame, Fig. 13(b) is its recovery image with  $\theta$  0.75, and Fig. 13(c) is the PSNR curves for 50 experiments under different  $\theta$ , whose values are 0.25, 0.5, 0.75, respectively. Similarly, in Fig. 14, Fig. 14(a) is the original image of the second frame, and Fig. 14(b) is its recovery image with  $\theta$  0.75. For  $\theta$  of the second frame, for example, if  $\theta$  is 0.75, then it means that the compression ratio of the first frame and the gap matrix is 0.75 since they are adjacent or continuous frames. In fact, both of these frames are highly similar to each other. Compared Fig. 13(c) with Fig. 14(c), it is easy to notice that the PSNRs of these two adjacent frames have little difference. Without loss of generality, we can compress the whole video based on the previous method. In this way, we just focus on compressing gap matrices between the continual frames rather than compressing and recovering every frame. Given that the gap

matrices are sparse, the time of making the gap matrices sparse can be reduced. We can deduce that the PTP-CS can recover the video image well.

In the video compression, the storage space of the measurement matrix in PTP-CS is reduced significantly. In the Kronecker product CS model, if the compression ratio  $\theta$  is 0.75, then the size of the measurement matrix  $\Phi$  is

$$\Phi \in R^{192 \times 256} \quad (79)$$

which should match the dimension of the image size. However, in the PTP-CS, the size of the original measurement matrix can be

$$\Phi' \in R^{48 \times 64} \quad (80)$$

and the size of  $P$  is

$$P \in R^{4 \times 4}. \quad (81)$$

From (79) to (81), it is easy to notice that the storage space of PTP-CS is much less than that of the Kronecker product CS.

#### E. Quantitative Analysis of 1-D Signal

Besides the 2-D image signal and high-dimensional video signal, there is another common signal with one-dimension, for example, the ECG signal in our daily life. In order to verify the performance of PTP-CS for 1-D signals, we take a signal with size  $1 \times 192$ , four kinds of measurement matrices are selected as Gaussian matrix, Bernoulli matrix, Toeplitz matrix, and the chaotic matrix. We can reduce the dimensions of the matrices significantly. We set  $(\Phi \otimes P)$ , where  $P$  is adopted as Gaussian matrix and  $g$  (i.e., the generalized permutation matrix), respectively. The percentage of recovery is calculated by (76), and if the reconstruction error is less than 0.002, then it can be considered that the signal is reconstructed successfully. For a

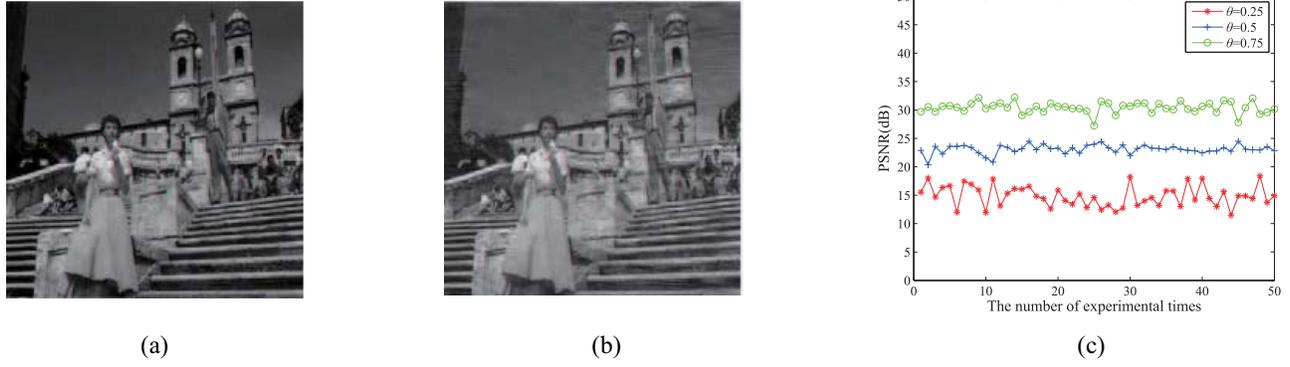


Fig. 13. Recovery results of the first frame ( $256 \times 256$ ), where  $P$  is the generalized permutation matrix. (a) image of the first frame, (b) its recovery frame when  $\theta$  is 0.75, and (c) three curves of PSNR, which correspond to different  $\theta$  whose values are 0.25, 0.5, and 0.75, respectively. The  $x$ -coordinate in (c) represents the number of the experimental times, and the  $y$ -coordinate in the (c) represents the value of PSNR in each experiment.



Fig. 14. Recovery results of the second frame ( $256 \times 256$ ) as  $\eta = 4$ , where  $P$  is the generalized permutation matrix. (a) image of the second frame, (b) its recovery frame when  $\theta$  is 0.75, and (c) three curves of PSNR, which correspond to different  $\theta$  whose values are 0.25, 0.5, and 0.75, respectively. The  $x$ -coordinate in (c) represents the number of the experimental times, and the  $y$ -coordinate in the picture (c) represents the value of PSNR in each experiment.

randomly selected vector  $x \in R^{192}$  with sparsity 25, let the columns of measurement matrices  $n$  be equal to 192, 96, 64, 48, and we can get the relationship between  $m$  and the recovery percentage.

After simulating the PTP-CS and the traditional CS for 1000 times, we draw the effect of  $m$  on the recovery percentage of signal in Fig. 15. If  $n$  equals to 192, it corresponds to the traditional CS. If  $n$  equals 96, 64, 48, it corresponds to the PTP-CS. Evidently, if  $P$  is the Gaussian matrix, as is shown in Fig. 15(a), the recovery percentage can reach 100% when  $n$  is 192, and  $m$  is almost larger than 126. Meanwhile, when  $n$  equals to 96, 64, 48 and  $m$  equals to 64, 50, 39, respectively, the recovery percentage can reach 100%. If  $P$  is the generalized permutation matrix, as is shown in Fig. 15(b), the recovery percentage can reach 100% when  $n$  is 192, and  $m$  is almost larger than 126. When the recovery percentage can reach to 100%,  $n$  equals to 96, 64, 48 and  $m$  equals to 68, 49, 38, respectively. However, for Bernoulli matrix, the recovery percentage will descend with fluctuation when  $m$  is too large. Therefore, the PTP-CS achieves a similar result with the traditional CS in signal recovery, but it requires less storage space than the traditional CS.

#### F. Comparison of Storage Space

The block CS (BCS) proposed by Gan [10] is an effective method to reduce the storage space of the measurement matrix,

the image is divided into small blocks and the size of each block is  $B \times B$ . According to the previous PTP-CS model, i.e.,

$$y_{\frac{mq}{n} \times 1} = \left( \Phi_{m \times n} \otimes P_n^q \times \frac{q}{n} \right) \cdot x_{q \times 1} \quad (82)$$

suppose  $x_i$  is a vector and it represents the  $i$ th block of the image  $y$ , then we have  $y_i = \Phi_{BCS} \cdot x_i$ , where  $\Phi_{BCS}$  is  $\phi_B \times B^2$  with  $\phi_B = (mB^2/n)$ . In BCS,  $\Phi_{BCS}$  is an orthonormal and i.i.d Gaussian matrix [11]. In this way, the storage space of the measurement matrix is reduced, and the speed of the compression is improved. Equivalently, the measurement matrix  $\Phi_{BCS}$  can be expressed as

$$\Phi_{BCS} = I_n \otimes \Phi_B = \begin{pmatrix} \Phi_B & 0 & \cdots & 0 \\ 0 & \Phi_B & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi_B \end{pmatrix}. \quad (83)$$

Furthermore, for a  $k$ -sparse signal  $x \in R^{q \times 1}$ , the initial measurement matrix  $\Phi_S \in R^{m \times n}$ , and in STP-CS [19], the measurement matrix  $\Phi_{STPCS}$  can be expressed as

$$\Phi_{STPCS} = \Phi_S \otimes I_{q/n} = \begin{pmatrix} \Phi_S & 0 & \cdots & 0 \\ 0 & \Phi_S & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi_S \end{pmatrix} \quad (84)$$

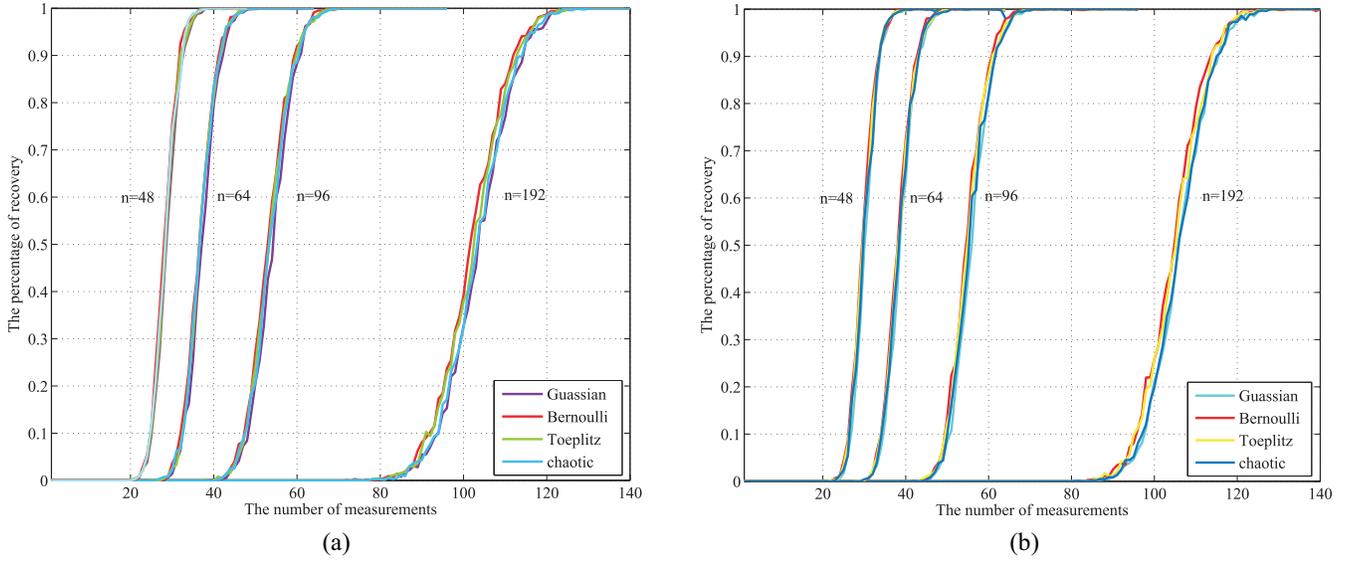


Fig. 15. Effect of  $m$  on the signal recovery for the PTP-CS and the traditional CS. Four kinds of measurement matrices are selected as Gaussian, Bernoulli, Toeplitz, and chaotic matrices, respectively. (a)  $P$  is selected as Gaussian matrix. (b)  $P$  is selected as  $g$ .

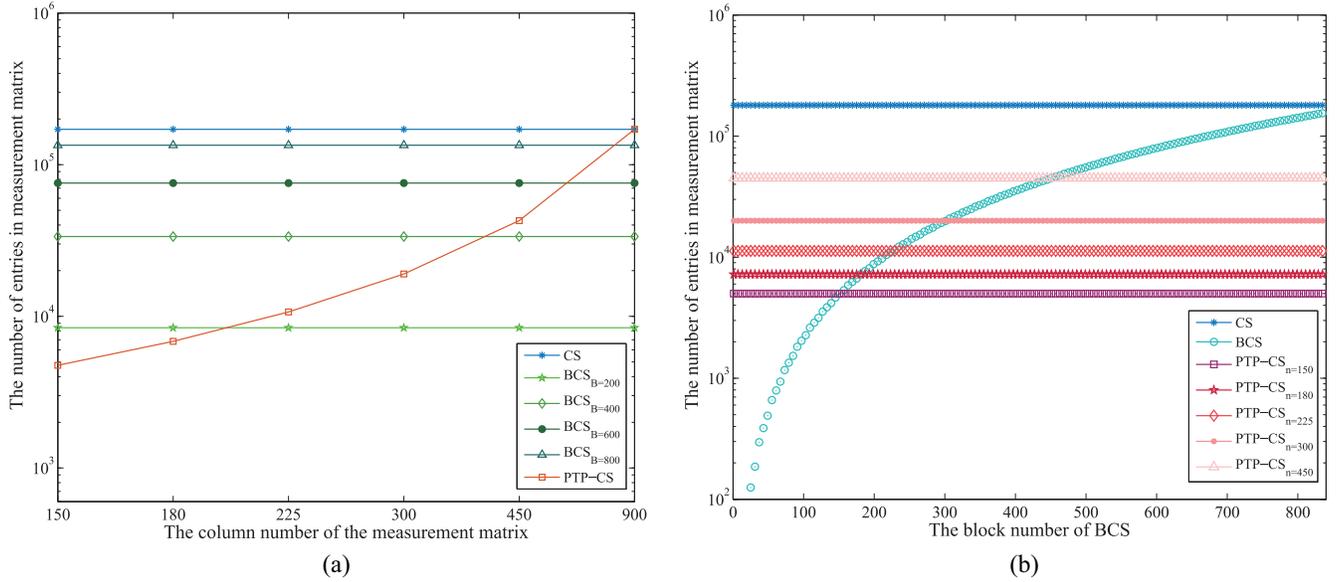


Fig. 16. Storage performance of CS, BCS, and PTP-CS. (a) Variation of measurement matrix entries in traditional CS is shown by the blue line marked by the asterisk (\*). The variation of measurement matrix entries in BCS is shown by the green line marked by four kinds of signs. Here,  $B$  is the size of each block, and the values of  $B$  are selected as 200, 400, 600, and 800, which are drawn by four green levels and marked by the star (\*), the rhombus ( $\diamond$ ), the dot ( $\bullet$ ) and the triangle ( $\Delta$ ), respectively. The variation of measurement matrix entries in PTP-CS is shown by the red line marked by the square ( $\square$ ). (b) Variation of measurement matrix entries in traditional CS is shown by the blue line marked by the asterisk (\*). The variation of measurement matrix entries in BCS is shown by the green line marked by the circle ( $\circ$ ). The variation of measurement matrix entries in PTP-CS is shown by the red line with five kinds of signs. Here,  $n$  is the column number of the measurement matrix, and the values of  $n$  are set as 150, 180, 225, 300, and 450 which are drawn by five red levels and marked with the square ( $\square$ ), the star (\*), the rhombus ( $\diamond$ ), the dot ( $\bullet$ ), and the triangle ( $\Delta$ ), respectively.

where  $n$  is selected as a factor of  $q$ . The PTP CS can be expressed as

$$\Phi_{\text{PTPCS}} = \Phi_P \otimes P_n^q = \begin{pmatrix} \varphi_{11}P & \varphi_{12}P & \cdots & \varphi_{1n}P \\ \varphi_{21}P & \varphi_{22}P & \cdots & \varphi_{2n}P \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m1}P & \varphi_{m2}P & \cdots & \varphi_{mn}P \end{pmatrix} \quad (85)$$

where  $\Phi_{\text{PTPCS}}$  is the measurement matrix, and  $\Phi_P$  is the initial measurement matrix.

From (84) to (85), the STP-CS is similar with the BCS, so does the PTP CS (PTP-CS). However, due to the incommutability of the tensor product operation and various choices of the matrix  $P$ , these three methods are exactly different. Since the PTP-CS and the STP-CS use the same mechanism to reduce the storage space, we decide to compare the storage space reduction among the traditional CS, the BCS, and the PTP-CS. By comparing the entry numbers of measurement

matrices among the traditional CS, the BCS, and the PTP-CS, we can easily find the way of saving the storage space for the measurement matrix. Suppose that the original signal  $x$  is an image with size  $q \times q$ , obviously, for the traditional CS, the size of the measurement matrix is  $m \times q$ . In BCS, the image is divided into small blocks with size  $B \times B$ , and each small block will be compressed by the measurement matrix with size  $\lfloor mB/q \rfloor \times B$ . In PTP-CS and STP-CS, the size of the measurement matrix is  $mn^2/q$ , where  $n$  is a factor of  $q$ . In essence, BCS, STP-CS, PTP-CS can reduce the storage space of the measurement matrix.

In Fig. 16(a) and (b), the original signal is an image with size  $900 \times 900$ , and  $m$  (i.e., the row number of the measurement matrix) is 190, Fig. 16(a) shows the effects of  $n$  (i.e., the column of the measurement matrix) and Fig. 16(b) shows the effects of  $B$  (i.e., the size of the blocks) on the number of entries in different measurement matrices. In Fig. 16(a), we set  $B$  as 200, 400, 600, and 800, respectively. In Fig. 16(b), we set  $n$  as 150, 180, 225, 300, and 450, respectively, which can satisfy that  $n$  is the factor of  $q$ . The experimental results in Fig. 16 shows that the traditional CS has the highest demand of the storage space of the measurement matrix whereas the PTP-CS and the BCS almost have the same performance. Unfortunately, we can find that the entries of the measurement matrix in PTP-CS are not less than those in BCS. Nevertheless, if each of the block is large (i.e., the value of  $B$  is big), the PTP-CS will have a better performance to save the storage space of the measurement matrix. In fact, the size of each block should not be too small. For the size of the block  $B$  and the sparsity of the image signal  $k$ , it should satisfy that  $B \gg k$ . For example, the energy of the image will be focused on one corner after being made sparse, so the image cannot be reconstructed if the block is small. All in all, the PTP-CS seems to be better than other two CS methods.

## V. CONCLUSION

In this paper, we define a new multiplication rule called PTP and apply it to CS. First, we define a new inner product and an included angle between two vectors with different dimensions. Meanwhile, in view of the limitation of the matrix multiplication, we use the matrix  $P$  to change the dimension and give some basic properties of PTP. Second, in the CS, we adopt the  $P$  as the Gaussian random matrix and the generalized permutation matrix in the design of the measurement matrix to reduce the storage space of the measurement matrix. In theoretical aspect, we make the quality analysis by spark, coherence, and RIP in a broad sense, which can be used to analyze other kinds of CS models. For the reconstruction algorithm, we propose a new model based on the IRLSs which is suitable for the proposed PTP-CS model. We can see from the experimental results that the PTP-CS has a good performance on the recoveries of 1-D and 2-D signals, as well as the frames in the video signal. The PTP not only breaks the conventional concept of the vector angle and improves the flexibility of the matrix multiplication, but also significantly reduces the storage space that the measurement matrix needed in CS. By taking the randomness of  $P$  into consideration, in future work, we

will verify the performance of PTP-CS in signal transmission encryption which will mainly focus on  $P$  and study the IoT application. Then, we will complete the concept of the vector relationship in  $P$ -transform. Furthermore, we will optimize the reconstruction algorithm to reduce the CPU time.

## ACKNOWLEDGMENT

The authors would like to thank the Editor and all the reviewers for their efforts and comments on this paper.

## REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [3] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [4] J. Wen, Z. Chen, Y. Han, J. D. Villasenor, and S. Yang, "A compressive sensing image compression algorithm using quantized DCT and noiselet information," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2010, pp. 1294–1297.
- [5] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. IEEE Radar Conf.*, 2007, pp. 128–133.
- [6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [7] B. Zhou, A. Liu, X. Wang, Y. She, and V. Lau, "Compressive sensing-based multiple-leak identification for smart water supply systems," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1228–1241, Apr. 2018.
- [8] Z. C. Li, H. Huang, and S. Misra, "Compressed sensing via dictionary learning and approximate message passing for multimedia Internet of Things," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 505–512, Apr. 2017.
- [9] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 139–154, Jan. 2012.
- [10] L. Gan, "Block compressed sensing of natural images," in *Proc. Int. Conf. Digit. Signal Process.*, 2007, pp. 403–406.
- [11] E. J. Candes and J. K. Romberg, "Signal recovery from random projections," in *Proc. SPIE*, vol. 5674, 2005, pp. 76–86.
- [12] H. P. Peng *et al.*, "Secure and energy-efficient data transmission system based on chaotic compressive sensing in body-to-body networks," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 3, pp. 558–573, Jun. 2017.
- [13] G. Q. Zheng, Y. M. Yang, and J. Carbonell, "Efficient shift-invariant dictionary learning," in *Proc. ACM SIGKDD Int. Conf.*, 2016, pp. 2095–2104.
- [14] D. Z. Cheng, H. Qi, and A. Xue, "A survey on semi-tensor product of matrices," *J. Syst. Sci. Complex.*, vol. 20, no. 2, pp. 304–322, 2007.
- [15] D. Z. Cheng, "Semi-tensor product of matrices and its application to Morgen's problem," *Sci. China*, vol. 44, no. 3, pp. 195–212, 2001.
- [16] D.-Z. Cheng and L.-J. Zhang, "On semi-tensor product of matrices and its applications," *Acta Mathematicae Applicatae Sinica (English Series)*, vol. 19, no. 2, pp. 219–228, 2003.
- [17] D. Laschov and M. Margaliot, "A maximum principle for single-input Boolean control networks," *IEEE Trans. Autom. Control*, vol. 56, no. 4, pp. 913–917, Apr. 2011.
- [18] D. Z. Cheng, "Some applications of semitensor product of matrices in algebra," *Comput. Math. Appl.*, vol. 52, nos. 6–7, pp. 1045–1066, 2006.
- [19] D. Xie, H. P. Peng, L. X. Li, and Y. Yang, "Semi-tensor compressed sensing," *Digit. Signal Process.*, vol. 58, pp. 85–92, Nov. 2016.
- [20] H. P. Peng, Y. Tian, and J. Kurths, "Semitensor product compressive sensing for big data transmission in wireless sensor networks," *Math. Problems Eng.*, vol. 2017, pp. 1–8, Mar. 2017.
- [21] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.
- [22] H. Rauhut, K. Schnass, and P. Vanderheynt, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.

- [23] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathématique*, vol. 346, nos. 9–10, pp. 589–592, 2008.
- [24] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2005.
- [25] T. T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4388–4394, Sep. 2010.
- [26] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, 2011.
- [27] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, Nov. 2010.
- [28] L. Yu, J. P. Barbot, G. Zheng, and H. Sun, "Compressive sensing with chaotic sequence," *IEEE Signal Process. Lett.*, vol. 17, no. 8, pp. 731–734, Aug. 2010.
- [29] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [30] S. Foucart and H. Rauhut, "A mathematical introduction to compressive sensing," *J. Japan Stat. Soc. Jpn. Issue*, pp. 111–131, 2013.
- [31] S. Datta, S. Howard, and D. Cochran, "Geometry of the Welch bounds," *Lin. Algebra Appl.*, vol. 437, no. 10, pp. 2455–2470, 2012.
- [32] T. Strohmer, R. W. Heath, "Grassmannian frames with applications to coding and communication," *Appl. Comput. Harmonic Anal.*, vol. 14, no. 3, pp. 257–275, 2003.
- [33] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 3, pp. 397–399, May 1974.
- [34] X. Luo, M. C. Zhou, S. Li, and M. S. Shang, "An inherently non-negative latent factor model for high-dimensional and sparse matrices from industrial applications," *IEEE Trans. Ind. Informat.*, vol. 14, no. 5, pp. 2011–2022, May 2018.
- [35] X. Luo, J. Sun, Z. Wang, S. Li, and M. Shang, "Symmetric and non-negative latent factor models for undirected, high-dimensional, and sparse networks in industrial applications," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 3098–3107, Dec. 2017.
- [36] X. Luo *et al.*, "A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 579–592, Mar. 2016.
- [37] X. Luo, M. C. Zhou, Y. N. Xia, and Q. Zhu, "An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1273–1284, May 2014.
- [38] X. Luo, M. C. Zhou, Z. D. Wang, Y. Xia, and Q. Zhu, "An effective scheme for QoS estimation via alternating direction method-based matrix factorization," *IEEE Trans. Services Comput.*, to be published.
- [39] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [40] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [41] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.
- [42] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.
- [43] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.
- [44] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "Decoding real-field codes by an iterative expectation-maximization (EM) algorithm," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 3169–3172.
- [45] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
- [46] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
- [47] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
- [48] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
- [49] R. Saab, R. Chartrand, and O. Yilmaz, "Stable sparse approximations via nonconvex optimization," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 3885–3888.
- [50] J. M. Wang, S. P. Ye, R. Yue, and C. Chen, "Low storage space for compressive sensing: Semi-tensor product approach," *EURASIP J. Image Video Process.*, vol. 2017, 2017, p. 51.
- [51] I. Daubechies, R. Devore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.
- [52] M. Chakraborty and R. Nath, "Performance of IRLS algorithm in compressive sensing," in *Proc. Int. Conf. Workshop Emerg. Trends Technol.*, 2011, pp. 320–323.
- [53] R. Bellman, *Introduction to Matrix Analysis*, 2nd ed. New York, NY, USA: McGraw-Hill, 1970.
- [54] L. R. Tucker, "Extension of factor analysis to three-dimensional matrices," in *Contributions to Mathematical Psychology*. New York, NY, USA: Rinehart & Winston, 1964.
- [55] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [56] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, Sep. 1978.
- [57] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via  $l_1$  minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [58] M. F. Duarte and R. G. Baraniuk, "Kronecker product matrices for compressive sensing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2010, pp. 3650–3653.
- [59] G. H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "Fast sparse representation based on smoothed  $l_0$  norm," in *Proc. 7th Int. Conf.*, 2007, pp. 389–396.
- [60] G. Prisco and M. D'Urso, "Maximally sparse arrays via sequential convex optimizations," *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 192–195, 2012.
- [61] S. E. Nai, W. Ser, Z. L. Yu, and H. Chen, "Beampattern synthesis for linear and planar arrays with antenna selection by convex optimization," *IEEE Trans. Antennas Propag.*, vol. 58, no. 12, pp. 3923–3930, Dec. 2010.
- [62] B. Fuchs, "Synthesis of sparse arrays with focused or shaped beam-pattern via sequential convex optimizations," *IEEE Trans. Antennas Propag.*, vol. 60, no. 7, pp. 3499–3503, Jul. 2012.
- [63] R. Saab and Ö. Yilmaz, "Sparse recovery by non-convex optimization—Instance optimality," *Appl. Comput. Harmonic Anal.*, vol. 29, no. 1, pp. 30–48, 2010.
- [64] S. Friedland, Q. Li, and D. Schonfeld, "Compressive sensing of sparse tensors," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4438–4447, Oct. 2014.
- [65] Q. Li, D. Schonfeld, and S. Friedland, "Generalized tensor compressive sensing," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2013, pp. 1–6.



**Haipeng Peng** received the M.S. degree in system engineering from the Shenyang University of Technology, Shenyang, China, in 2006, and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications, Beijing, China, in 2010.

He is currently a Professor with the School of Cyberspace Security, Beijing University of Posts and Telecommunications. He has co-authored 60 papers. His current research interests include compressive sensing, information security, network security, and complex networks.



**Yaqi Mi** received the B.S. degree in computer science and technology from the Zhejiang University of Finance and Economics, Hangzhou, China, in 2016. She is currently pursuing the M.S. degree in computer technology at the Beijing University of Posts and Telecommunications, Beijing, China.

Her current research interests include compressed sensing and signal processing.



**Lixiang Li** received the M.S. degree in circuit and system from Yanshan University, Qinhuangdao, China, in 2003, and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications, Beijing, China, in 2006.

She is currently a Professor with the School of Cyberspace Security, Beijing University of Posts and Telecommunications. She has co-authored over 100 papers. She visited the Potsdam Institute for Climate Impact Research, Potsdam, Germany, in 2011. Her

current research interests include compressive sensing, swarm intelligence, and network security.

Dr. Li was a recipient of the National Excellent Doctoral Theses Winner, the New Century Excellent Talents in University, the Hong Kong Scholar Award Winner, and the Beijing Higher Education Program for Young Talents Winner.



**Yixian Yang** received the M.S. degree in applied mathematics and the Ph.D. degree in electronics and communication systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 1986 and 1988, respectively.

He is the Managing Director of the Information Security Center, Beijing University of Posts and Telecommunications. He has authored or co-authored over 300 high-level papers and 20 monographs. His current research interests include coding and cryptography, information and

network security, and signal and information processing.

Dr. Yang was a recipient of the Yangtze River Scholar Program Professor, the National Outstanding Youth Fund Winner, and the National Teaching Masters.



**H. Eugene Stanley** received the Ph.D. degree from Harvard University, Cambridge, MA, USA, in 1967, under the supervision of T. A. Kaplan and Nobelist J. H. Van Vleck.

In 1979, he became a University Professor, and in 2011, a William Fairfield Warren Distinguished Professor with Boston University, Boston, MA, USA, where he is currently a Professor of physics, chemistry, biomedical engineering, and physiology and also the Director of the Center for Polymer Studies. His current research interests include sta-

tistical physics, complex systems, and economics.

Dr. Stanley is a member of the National Academy of Sciences.