From Rank Estimation to Rank Approximation: Rank Residual Constraint for Image Restoration

Zhiyuan Zha, Member, IEEE, Xin Yuan, Senior Member, IEEE, Bihan Wen, Member, IEEE, Jiantao Zhou, Member, IEEE, Jiachao Zhang, Member, IEEE and Ce Zhu, Fellow, IEEE

Abstract—In this paper, we propose a novel approach to the rank minimization problem, termed rank residual constraint (RRC) model. Different from existing low-rank based approaches, such as the well-known nuclear norm minimization (NNM) and the weighted nuclear norm minimization (WNNM), which estimate the underlying low-rank matrix directly from the corrupted observations, we progressively approximate the underlying low-rank matrix via minimizing the rank residual. Through integrating the image nonlocal self-similarity (NSS) prior with the proposed RRC model, we apply it to image restoration tasks, including image denoising and image compression artifacts reduction. Towards this end, we first obtain a good reference of the original image groups by using the image NSS prior, and then the rank residual of the image groups between this reference and the degraded image is minimized to achieve a better estimate to the desired image. In this manner, both the reference and the estimated image are updated gradually and jointly in each iteration. Based on the group-based sparse representation model, we further provide an analytical investigation on the feasibility of the proposed RRC model. Experimental results demonstrate that the proposed RRC method outperforms many state-of-the-art schemes in both the objective and perceptual quality.

Index Terms—Low-rank, rank residual constraint, nuclear norm minimization, nonlocal self-similarity, group-based sparse representation, image restoration.

I. INTRODUCTION

LOW-rank matrix estimation has attracted increasing attention due to its wide applications. In particular, it has been successfully applied in various machine learning and computer vision tasks [1–24]. For instance, the Netflix customer data matrix is treated as a low-rank one for the reason that the customers’ choices are largely dependent on a few common factors [1, 2]. The foreground and background in a video are also modeled as being sparse and low-rank, respectively [3–5]. As the matrix formed by nonlocal similar patches in a natural image is of low-rank, various low-rank models for image completion problems have been proposed, such as collaborative filtering [6], image alignment [4], image/video denoising [5, 7, 16], shadow removal [8] and reconstruction of occluded/corrupted face images [1, 9, 24].

One typical low-rank matrix estimation method is the low-rank matrix factorization (LRMF) [9–14], which factorizes an observed matrix \( Y \) into a product of two matrices that can be used to reconstruct the desired matrix with certain fidelity terms. A series of LRMF methods have been developed, such as the classical singular value decomposition (SVD) under \( \ell_2 \)-norm [10], robust LRMF methods under \( \ell_1 \)-norm [11, 12] and other probabilistic methods [13, 14].

Another parallel research is the rank minimization [15–24], with the nuclear norm minimization (NNM) [15] being the most representative one. The nuclear norm of a matrix \( X \), denoted by \( \|X\|_* \), is the summation of its singular values, i.e., \( \|X\|_* = \sum_i \sigma_i(X) \), with \( \sigma_i(X) \) representing the \( i \)-th singular value of \( X \). NNM aims to recover the underlying low-rank matrix \( X \) from its degraded observation matrix \( Y \), while minimizing \( \|X\|_* \). However, NNM tends to over-shrink the rank components, and therefore limits its capability and flexibility. To enforce the rank minimization efficiently, inspired by the success of the well-known nuclear norm minimization (NNM) [15–24], with the nuclear norm minimization (NNM) [15] being the most representative one. The nuclear norm of a matrix \( X \), denoted by \( \|X\|_* \), is the summation of its singular values, i.e., \( \|X\|_* = \sum_i \sigma_i(X) \), with \( \sigma_i(X) \) representing the \( i \)-th singular value of \( X \). NNM aims to recover the underlying low-rank matrix \( X \) from its degraded observation matrix \( Y \), while minimizing \( \|X\|_* \). However, NNM tends to over-shrink the rank components, and therefore limits its capability and flexibility. To enforce the rank minimization efficiently, inspired by the success of the well-known nuclear norm minimization (NNM) [15], we propose a weighted nuclear norm minimization (WNNM) model, which heuristically sets the weights being inverse to the singular values. Compared with NNM, WNNM assigns different weights to different singular values such that the matrix rank estimation becomes more rigid. Similar case also exists in the truncated nuclear norm minimization [18] and the partial sum minimization [19].

However, one common practice in the aforementioned low-rank models is only to estimate the low-rank matrix directly from the corrupted observations, which may lead to a defective result in real applications, such as image inverse problems. In this paper, we propose a novel approach, dubbed rank residual constraint (RRC) model, for the rank minimization problem.
Different from existing low-rank based methods, such as the well-known NNM and WNNM, we progressively approximate the underlying low-rank matrix via minimizing the rank residual. By integrating the image nonlocal self-similarity (NSS) prior [5, 16] with the proposed RRC model, we apply it to image restoration tasks, including image denoising and image compression artifacts reduction. In a nutshell, given a corrupted image \( y \), in each iteration, we construct a reference low-rank matrix \( X' \) (for each image group) by exploiting the image NSS prior, and approximate our recovered matrix \( \hat{X} \) to this reference matrix \( X' \) via the proposed RRC model. Fig. 1 depicts that the reconstructed image from our proposed algorithm can progressively approximate the ground truth, by taking the widely used image House as an example, which is corrupted by zero-mean Gaussian noise with standard deviation \( \sigma = 100 \). It can be observed that the singular values of the reference matrix \( X' \) approaches the singular values of the ground truth progressively and so does the recovered image (Fig. 1 (f-h)).

It is worth noting that the significant difference between the proposed RRC and the existing low-rank based methods (e.g., NNM and WNNM) is that we analyze the rank minimization problem from a different perspective. To be concrete, traditional low-rank based methods estimated the low-rank matrix directly from the corrupted observations. By contrast, in our RRC model, we analyze the rank minimization problem from the point of approximation theory [25], namely, minimizing the rank residual; the singular values of the recovered matrix progressively approaches the singular values of the reference matrix. Note that the reference matrix and the recovered matrix in our RRC model are both updated gradually and jointly in each iteration. Moreover, we provide an analytical investigation on the feasibility of the proposed RRC model from the perspective of group-based sparse representation [26–29].

The rest of this paper is organized as follows. Section II describes the RRC model based on the rank minimization scenario. Section III presents how to use the proposed RRC model for image denoising. Section IV develops the algorithm for image compression artifacts reduction exploiting the proposed RRC model. Section V provides an analysis investigation for the proposed RRC model in terms of group-based sparse representation. Extensive results for image restoration are presented in Section VI and Section VII concludes the paper.

II. RANK MINIMIZATION VIA RANK RESIDUAL CONSTRAINT

In this section, we first analyze the weakness of the traditional NNM model and then propose the RRC model to improve the performance of rank estimation.

A. Nuclear Norm Minimization

According to [6, 15, 17], NNM is the tightest convex relaxation of the original rank minimization problem. Given a data matrix \( Y \in \mathbb{R}^{d \times m} \), the goal of NNM is to find a low-rank matrix \( X \in \mathbb{R}^{d \times m} \) of rank \( r \ll \min(d, m) \), satisfying the following objective function,

\[
\hat{X} = \arg\min_X \left( \frac{1}{2} \| Y - X \|_F^2 + \lambda \| X \|_* \right),
\]
and the singular values of the original matrix $X^*$ are as close as possible. Explicitly, we define the rank residual by

$$\gamma^* \overset{\text{def}}{=} \sigma - \psi^*,$$

where $\sigma$ and $\psi^*$ are the singular values of $X$ and $X^*$, respectively. It can be seen that the rank estimation of the matrix $X$ largely depends on the level of this rank residual $\gamma^*$.

However, in real applications, the original matrix $X^*$ is unavailable, and thus we desire a good estimate of it, denoted by $X'$. Via introducing this $X'$ and defining $\gamma \overset{\text{def}}{=} \sigma - \psi$ with $\psi$ being the singular values of $X'$, we propose the RRC model below,

$$\hat{X} = \arg\min_X \left( \frac{1}{2} \|Y - X\|_F^2 + \lambda \|\gamma\|_S \right),$$

where $S_p$ denotes some type of norm. We will describe how to estimate $X'$ and solve Eq. (3) below. Specifically, in the following sections, we apply the proposed RRC model to image restoration tasks, including image denoising and image compression artifacts reduction.

III. IMAGE DENOISING VIA RANK RESIDENTIAL CONSTRAINT

In this section, we firstly employ the proposed RRC model in image denoising. It is well-known that image denoising [29–31] is not only an important problem in image processing, but also an ideal test bench to measure different statistical image models. Mathematically, image denoising aims to recover the latent clean image $x$ from its noisy observation $y$, which can be modeled as

$$y = x + n,$$

where $n$ is usually assumed to be zero-mean Gaussian with standard deviation $\sigma_n$. Owing to the ill-posed nature of image denoising, it is critical to exploit the prior knowledge that characterizes the statistical features of the image.

The well-known NSS prior [32], which investigates the repetitiveness of textures and structures of natural images within nonlocal regions, implies that many similar patches can be searched given a reference patch. To be concrete, a noisy (vectorized) image $y \in \mathbb{R}^N$ is divided into $n$ overlapping patches of size $\sqrt{d} \times \sqrt{d}$, and each patch is denoted by a vector $y_i \in \mathbb{R}^d$, $i = 1, 2, \ldots, n$. For the $i^{th}$ patch $y_i$, its $m$ similar patches are selected from a surrounding (searching) window with $L \times L$ pixels to form a set $S_i$. Note that the K-Nearest Neighbour (KNN) algorithm [33] is used to search similar patches. Following this, these patches in $S_i$ are stacked into a matrix $Y_i \in \mathbb{R}^{d \times m}$, i.e., $Y_i = \{y_{i,1}, y_{i,2}, \ldots, y_{i,m}\}$. This matrix $Y_i$ consisting of patches with similar structures is thus called a group [26, 27], where $\{y_{i,j}\}_{j=1}^m$ denotes the $j^{th}$ patch in the $i^{th}$ group. Then we have $Y_i = X_i + N_i$, where $X_i$ and $N_i$ are the corresponding group matrices of the original image and noise, respectively. Since all patches in each data matrix have similar structures, the constructed data matrix $Y_i$ is of low-rank. By adopting the proposed RRC model in Eq. (3), the
A. Determine $S_p$

Let us come back to Eq. (5). Obviously, one important issue of our RRC based image denoising is the determination of $S_p$. Hereby, we perform some experiments to investigate the statistical property of $\gamma$, where $\gamma$ denotes the set of $\gamma_i = \sigma_i - \psi_i$, with $\sigma_i$ and $\psi_i$ representing the singular values of $X_i$ and $X'_i$, respectively. $X'_i \in \mathbb{R}^{d \times m}$ is a good estimate of the original image group $X_i$. In order to achieve a high performance for image denoising, we hope that the rank residual $\gamma_i$ of each group is small enough.

B. Estimate $X'$

In Eq. (5), after determining $S_p$, we also need to estimate $X'_i$, as the original image is unavailable in real applications. There are a variety of methods to estimate $X'_i$, which depends on the prior knowledge of the original image $x$. For example, if there exist many example images that are similar enough to the original image $x$, similar patches could be searched to construct the matrix $X'_i$ from the example image set [34, 35]. However, under many practical situations, the example image set is inaccessible. In this paper, inspired by the fact that natural images often contain repetitive structures [26–30, 32], we search nonlocal similar patches in the degraded image to the given patch and use the image NSS prior to estimate $X'_i$. Specifically, a good estimation of each reference patch $x'_{i,j}$ in $X'_i$ can be computed by the weighted average of the patches $\{\hat{x}_{i,k}\}$ in $\hat{X}_i$, associated with each group including $m$ nonlocal similar patches in each iteration. Note that the initialization of $\hat{X}_i$ is the corresponding noisy group $Y_i$. Then we have,

$$x'_{i,j} = \sum_{k=1}^{m} w_{i,k} \hat{x}_{i,k}.$$  

where $x'_{i,j}$ and $\hat{x}_{i,k}$ represent the $j$-th and $k$-th patch of $X'_i$ and $\hat{X}_i$, respectively. $m$ is the total number of similar patches and $w_{i,k}$ is the weight, which is inversely proportional to the distance between patches $\hat{x}_i$ and $\hat{x}_{i,k}$, i.e., $w_{i,k} = \frac{1}{h} \exp(-||x_i - \hat{x}_{i,k}||^2_2 / h)$, where $h$ is a predefined constant and $W$ is a normalization factor. It is worth noting that Eq. (7) is based on nonlocal means filtering [36]. Again, the recovered matrix $\hat{X}_i$ and the reference matrix $X'_i$ are updated gradually and jointly in each iteration.

C. Iterative Shrinkage Algorithm to Solve the Proposed RRC Model for Image Denoising

We now develop an efficient algorithm to solve the optimization in Eq. (6). In order to do so, we first introduce the following Lemma and Theorem.

**Lemma 1.** The minimization problem

$$\hat{x} = \arg\min_x \left( \frac{1}{2} \|a - x\|_2^2 + \tau \|x - b\|_1 \right),$$

has a closed-form solution,

$$\hat{x} = \text{soft}(a - b, \tau) + b.$$  

where $\text{soft}(a, \tau) = \text{sgn}(a) \circ \max(\text{abs}(a) - \tau, 0)$; $\circ$ denotes the element-wise (Hadamard) product, and $a, b, x$ are vectors of the same dimension.

**Proof.** See [37].

**Theorem 1.** (von Neumann) For any two matrices $A, B \in \mathbb{R}^{m \times n}$, $\text{Tr}(A^T B) \leq \text{Tr}(\sigma(A)^T \sigma(B))$, where $\text{Tr}$ calculates the trace of the ensured matrix; $\sigma(A)$ and $\sigma(B)$ are the ordered singular value matrices of $A$ and $B$ with the same order, respectively.

**Proof.** See [38].

We now provide the solution of Eq. (6) by the following Theorem.

**Theorem 2.** Let $Y_i = U_i \Delta_i V_i^T$ be the SVD of $Y_i \in \mathbb{R}^{d \times m}$ with $\Delta_i = \text{diag}(\delta_{i,1}, \ldots, \delta_{i,j})$, $j = \min(d, m)$, $X'_i = R_i \Lambda_i Q_i^T$ be the SVD of $X'_i \in \mathbb{R}^{d \times m}$ with $\Lambda_i = \text{diag}(\psi_{i,1}, \ldots, \psi_{i,j})$. The optimal solution $X_i$ to the problem in Eq. (6) is $U_i \Sigma_i V_i^T$, where $\Sigma_i = \text{diag}(\sigma_{i,1}, \ldots, \sigma_{i,j})$ and the $k$th diagonal element $\sigma_{i,k}$ is solved by

$$\min_{\sigma_{i,k} \geq 0} \left( \frac{1}{2} (\delta_{i,k} - \sigma_{i,k})^2 + \lambda |\sigma_{i,k} - \psi_{i,k}| \right), \quad \forall k = 1, \ldots, j.$$

**Proof.** See Appendix A.

Thereby, the minimization problem in Eq. (6) can be simplified by minimizing the problem in Eq. (10). For fixed $\delta_{i,k},$
where \( \mu \) to update \( t \) procedure several iterations to achieve better results. In the \( \lambda \) iteration, and inspired by [40], the group matrices \( \mathbf{X} \).

The final denoised image \( \mathbf{x} \), based on Lemma 1, the closed-form solution of Eq. (10) is

\[
\sigma_{i,k} = \text{soft}(\delta_{i,k} - \psi_{i,k}, \lambda) + \psi_{i,k}, \quad \text{(11)}
\]

With the solution of \( \mathbf{X} \) in Eq. (11) provided, the estimated group matrix \( \mathbf{X} \) can be reconstructed by \( \mathbf{X} = \mathbf{U} \Sigma \mathbf{V}^T \). Then the denoised image \( \mathbf{x} \) can be reconstructed by aggregating all the group matrices \( \{\mathbf{X}_i\}_i \).

In image denoising, we perform the below denoising procedure several iterations to achieve better results. In the \( t \)th iteration, the iterative regularization strategy [39] is used to update \( y \) by

\[
y^t = x^{t-1} + \mu(y - x^{t-1}),
\]

where \( \mu \) represents the step-size. The standard deviation of the noise in the \( t \)th iteration is adjusted by [5, 27]:

\[
\sigma_n^2 = \rho \sqrt{(\sigma_n^2 - \|y - x^{t-1}\|_2^2)},
\]

where \( \rho \) is a constant.

The parameter \( \lambda \) balances the fidelity term and the regularization term can also be adaptively determined in each iteration, and inspired by [40], \( \lambda \) in each iteration is set to

\[
\lambda = \frac{c \sqrt{2 \sigma_n^2}}{\varphi_i + \epsilon}, \quad \text{(13)}
\]

where \( \varphi_i \) denotes the estimated standard variance of \( \gamma_i \), and \( c, \epsilon \) are the constants.

The complete description of the proposed RRC based image denoising approach to solve the problem in Eq. (6) is presented in Algorithm 1.

### IV. IMAGE COMPRESSION ARTIFACTS REDUCTION VIA RANK RESIDUAL CONSTRAINT

With the rapid development of social network and mobile Internet, billions of image and video resources have been spread through miscellaneous ways on the Internet everyday. To save the limited bandwidth and storage space, lossy compression scheme (e.g., JPEG [41], WebP [32] and HEVC-MSP [43]) has been widely used to compress images and videos.

![The distribution of \( \gamma \) Laplacian](a)

![The distribution of \( \gamma \) Laplacian](b)

Fig. 4. The distributions of the rank residual \( \gamma \). (a) image Lena is compressed by JPEG with QF = 10; (b) image House is compressed by JPEG with QF = 20.

However, these lossy compression techniques often give rise to visually annoying compression artifacts, i.e., sacrificing the image quality to satisfy the bit-budget, which severely degrade the user experience. Furthermore, the performance of many other computer vision tasks (e.g., image recognition [44] and object detection [45]) largely depends on the quality of input images, and therefore it is desired to recover visually pleasing artifact-free images from these compressed images. It is well-known that JPEG is the most popular lossy compression method [41], and therefore in this work we focus on JPEG compression artifacts reduction.

Specifically, we apply the proposed RRC model to image compression artifacts reduction. Similar to the above procedure in Section III, given a JPEG compressed (vectorized) image \( x \in \mathbb{R}^{N^2} \), we extract each patch \( x_k \in \mathbb{R}^{\sqrt{d} \times \sqrt{d}} \) from it, and search for its similar patches to generate \( K \) groups, where each group \( X_k \in \mathbb{R}^{d \times N_k} \), i.e., \( X_k = \{x_{k,1}, x_{k,2}, \ldots, x_{k,m}\} \). In our application of image compression artifacts reduction, the observed data \( X_k \) includes the quantization error, which can be modeled as noise. We thus model the data matrix by \( X_k = Z_k + N_k \), where \( Z_k \) and \( N_k \) are the corresponding group matrices of the original (desired clean) image and the noise assumed to be zero-mean Gaussian noise here [46–49]. Since all patches in each data matrix have similar structures, the constructed data matrix \( X_k \) is of low-rank. By invoking the proposed RRC model in Eq. (3), the low-rank matrix \( Z_k \) can be estimated by solving the following optimization problem,

\[
\hat{Z}_k = \arg \min_{Z_k} \left( \frac{1}{2\sigma_e^2} \|X_k - Z_k\|_F^2 + \lambda \|\gamma_k\|_{S_P} \right), \quad \text{(14)}
\]

where \( \sigma_e^2 \) denotes the variance of additive Gaussian noise, \( \gamma_k = \psi_k - \delta_k \), with \( \psi_k \) and \( \delta_k \) representing the singular values of \( Z_k \) and \( Z_k' \) respectively, and \( Z_k' \in \mathbb{R}^{d \times m} \) is a good estimate of the original image group \( Z_k' \). In order to achieve a high quality recovered image, we also expect that the rank residual \( \gamma_k \) of each group is as small as possible.

According to Eq. (14), one can observe that \( Z_k' \) and \( S_P \) are unknown in our proposed RRC model for image compression artifacts reduction, and thus we need to estimate them. Similar to subsection III-B, we exploit the image NSS prior to estimate \( Z_k' \) from the recovered image \( \hat{z} \) in each iteration, i.e.,

\[
z'_{k,i} = \sum_{i=1}^{m-j+1} w_{k,i} \hat{z}_{k,i}, \quad \text{(15)}
\]

where \( z_{k,j}' \) and \( \hat{z}_{k,i} \) represent the \( j \)-th and \( i \)-th patch of \( Z_k' \) and \( \hat{Z}_k \), respectively. Note that the initialization of \( \hat{z} \) is the JPEG.
compressed image.

For $S_p$, we perform some experiments to investigate the statistical property of $\gamma$, where $\gamma$ denotes the set of $\gamma_k = \psi_k - \hat{\delta}_k$ and we use the original image to construct $Z'$. Two typical images Lena and House are compressed by JPEG with quality factors (QF) = 10 and QF = 20 respectively, to generate the degraded images. Fig. 4 shows the fitting results of empirical distributions of the rank residual $\gamma$ on these two images. One can also observe that both empirical distributions can be well approximated by a Laplacian distribution, which is usually modeled by an $\ell_1$-norm. Therefore, Eq. (14) can be rewritten as

$$\hat{Z}_k = \arg\min_{Z_k} \left( \frac{1}{2\sigma_x^2} \|X_k - Z_k\|^2_F + \lambda \|\gamma_k\|_1 \right). \quad (16)$$

### A. Quantization Constraint for Image Compression Artifacts Reduction

This subsection is presented by following [49], to make the work self-contained and accessible to readers as much as possible. We now formulate image compression artifacts reduction as an image inverse problem. Specifically, given the JPEG compressed image $y$, the original image $x$ can be obtained by solving $\hat{x} = \arg\max_x p(x|y)$. Taking account of the prior $p(x)$ and exploiting the Bayesian rule, we have

$$\hat{x} = \arg\max_x \log(p(y|x)) + \log(p(x)), \quad (17)$$

where the first term is the data fidelity and the second term depends on the employed image priors. In the following, we will introduce how to design each term of Eq. (17).

1) Quantization Noise Model: In our task, the observed JPEG-based compression image is usually modeled as the corrupted image with quantization noise,

$$y = x + s, \quad (18)$$

where $y$ is the JPEG-based compression image with blocking artifacts and $x, s$ represent the original image and the quantization noise, respectively. Many sophisticated models of the quantization noise have been developed [46–52], and in particular a Gaussian model has been widely used owing to its simplicity and effectiveness, which has achieved state-of-the-art results [46–49]. Therefore, the Gaussian model is adopted in our work and specifically we employ the approach proposed in [48] to estimate the noise variance $\sigma_x^2$,

$$\sigma_x^2 = 1.195 \left( \bar{e} \right)^{0.6194} + 0.9693, \quad \bar{e} = \frac{1}{9} \sum_{i,j=1}^{3} M^q_{i,j}, \quad (19)$$

where $M^q$ is the $8 \times 8$ quantization matrix with QF of $q$, $\bar{e}$ is the mean value of the nine upper-left entries in $M^q$, corresponding to lowest-frequency DCT harmonics, and we use $M^q_{i,j}$ to denote the $(i,j)^{th}$ element in $M^q$. It is worth noting that the noise variance $\sigma_x^2$ obtained by Eq. (19) is only the variance of the hypothetical Gaussian noise, which determines the level of adaptive smoothing that is able to reduce compression artifacts generated by the quantization step with $M^q$ [46].

Following this, the first data-fidelity term in Eq. (17) can be formulated by

$$\log(p(y|x)) = -\frac{1}{2\sigma_y^2} \|y - x\|^2_2 , \quad (20)$$

where $\sigma_y^2$ is adaptively determined by Eq. (19) and we have discarded the terms unrelated to $x$.

2) Quantization Constraint Prior: We adopt the quantization constraint (QC) [47–49] prior in the proposed model, to tackle the quantization issue in the JPEG scheme. Specifically, to simplify the first two stages of JPEG compression, let us define an $N \times N$ block-wise DCT operator $A$, which transforms each non-overlapped $8 \times 8$ block of the input image into their frequency domain [48, 49, 53]. Correspondingly, the matrix operator $A^{-1}$ denotes the inverse DCT process. Therefore, the frequency-domain image of $x$ is obtained by the following transform,

$$\hat{x} = Ax, \quad (21)$$

Recall that $M^q$ denotes the $8 \times 8$ block-level quantization matrix, which is determined by the QF in the range of $[1, 100]$. Based on the JPEG compression scheme, we have the following forward process,

$$\hat{y}_{[(k-1)\times N+l]} = \text{round} \left( \frac{\bar{x}_{[(k-1)\times N+l]}}{M_{[k,l]}} \right) M_{[k,l]}, \quad (22)$$

where $1 \leq k, l \leq N$ are the pixel coordinates in the image, and round($\cdot$) denotes the rounding operator mapping the input to its nearest integer. Here $M$ is an $N \times N$ image-level quantization matrix, and the values of its elements $M_{[k,l]} = M^q_{[k,l]}$, where the block-level coordinates $k = \text{mod}(k, 8); \ell = \text{mod}(l, 8)$.

We define the lower and upper bound vectors $\bar{t}$ and $\tilde{u}$ as

$$\bar{t}_{[(k-1)\times N+l]} = (\hat{y}_{[(k-1)\times N+l]} - w) \ast M_{[k,l]}, \quad \tilde{u}_{[(k-1)\times N+l]} = (\hat{y}_{[(k-1)\times N+l]} + w) \ast M_{[k,l]}, \quad (23)$$

where $w$ is a constant which is typically set to be not greater than 0.5 [54]. The frequency-domain coefficients of the image should satisfy the bound (i.e., QC) [53]:

$$\bar{t} \preceq \hat{x} \preceq \tilde{u}, \quad (24)$$

where $\preceq$ denotes the element-wise $\preceq$ operator. Based on (24), the solution space of $x$, which is denoted as $\Omega$, is defined as

$$\Omega = \{ x | \bar{t} \preceq Ax \preceq \tilde{u} \}, \quad (25)$$

which holds for any image coded by JPEG. In this work, we explicitly incorporate the feasible solution space $\Omega$ into Eq. (17).

### B. Joint Model and Algorithm for Image Compression Artifacts Reduction

Now, we incorporate the quantization noise model in Eq. (20) and two image priors (i.e., the proposed RRC in Eq. (16) and QC in Eq. (25)) into Eq. (17), and achieve the joint image compression artifacts reduction as follows:

$$\begin{align*}
\hat{x}, \hat{Z}_k &= \arg\min_{x, Z_k} \frac{1}{2\sigma_x^2} \|y - x\|^2_2 \\
&\quad + \frac{\rho}{2\sigma_x^2} \sum_{k=1}^{K} \|R_k x - Z_k\|^2_F + \lambda \sum_{k=1}^{K} \|\gamma_k\|_1, \quad (26)
\end{align*}$$

s.t. $x \in \Omega$, $\hat{x} = Ax = \{x_{k,1}, x_{k,2}, \ldots, x_{k,m}\}$ is the operator that extracts the group $X_k$ from $x$ and $\{\rho, \lambda\}$ are positive constants.
Here, we introduce $\rho$ to make the solution of Eq. (26) more feasible.

In the following, we develop an alternating minimizing strategy to solve the large scale non-convex optimization problem in Eq. (26). Specifically, the minimization of Eq. (26) is divided into two sub-problems, i.e., $x$ and $Z_k$, and we will show that there is an efficient solution to each of them.

1) $Z_k$ Sub-problem: Given $x$, each $Z_k$ sub-problem can be expressed as:

$$Z_k = \arg \min_{Z_k} \frac{\rho}{2\sigma_k^2} \|R_k x - Z_k\|_F^2 + \lambda \|\gamma_k\|_1,$$  \hspace{1cm} (27)

Let $\mu = \frac{\lambda \sigma_k^2}{\rho}$, we can rewrite Eq. (27) as

$$Z_k = \arg \min_{Z_k} \frac{1}{2} \|X_k - Z_k\|_F^2 + \mu \|\gamma_k\|_1.$$  \hspace{1cm} (28)

Then, based on Theorem 2, the minimization problem in Eq. (28) can be simplified to minimize the following problem,

$$\min_{\psi_{k,i}} \left( \frac{1}{2} \|\sigma_{k,i} - \psi_{k,i}\|_2^2 + \mu \|\psi_{k,i} - \delta_{k,i}\|_2^2 \right), \hspace{1cm} \forall i = 1, \ldots, j,$$  \hspace{1cm} (29)

where $X_k = U_k \Sigma_k V_k^T$ is the SVD of $X_k \in \mathbb{R}^{d \times m}$ with $\Sigma_k = \text{diag}(\sigma_{k,1}, \ldots, \sigma_{k,j})$, and $j = \min(d, m)$. $Z_k' = P_k \Delta_k Q_k^T$ is the SVD of $Z_k' \in \mathbb{R}^{d \times m}$ with $\Delta_k = \text{diag}(\delta_{k,1}, \ldots, \delta_{k,j})$.

For fixed $\sigma_{k,i}, \delta_{k,i}$ and $\mu$, based on Lemma 1, the closed-form solution of Eq. (29) is

$$\psi_{k,i} = \text{soft}(\sigma_{k,i} - \delta_{k,i}, \mu) + \delta_{k,i},$$  \hspace{1cm} (30)

where $\Lambda_k = \text{diag}(\psi_{k,1}, \ldots, \psi_{k,j})$. With the solution of $\psi_{k}$ in Eq. (30) achieved, the estimated matrix $Z_k$ can be reconstructed by $Z_k = U_k \Lambda_k V_k^T$.

2) $x$ Sub-problem: After obtaining each $Z_k$, the desired image $x$ can be reconstructed by solving the following constrained quadratic minimization problem.

$$\hat{x} = \arg \min_{x} \frac{1}{2 \sigma_x^2} \|y - x\|_2^2 + \frac{\rho}{2 \sigma_x^2} \sum_{k=1}^{K} \|R_k x - Z_k\|_F^2,$$  \hspace{1cm} (31)

s.t. $x \in \Omega$.

We first obtain the solution of the unconstrained quadratic minimization of Eq. (31), and later project the solution to $\Omega$. Specifically, without considering the constraint of $\Omega$, Eq. (31) can be rewritten as

$$\hat{x} = \arg \min_{x} \frac{1}{2 \sigma_x^2} \|y - x\|_2^2 + \frac{\rho}{2 \sigma_x^2} \sum_{k=1}^{K} \|R_k x - Z_k\|_F^2,$$  \hspace{1cm} (32)

which has a closed-form solution, i.e.,

$$\hat{x} = \left( I + \frac{\sigma_x^2 \rho}{\sigma_x^2} \sum_{k=1}^{K} R_k^T R_k \right)^{-1} \left( y + \frac{\sigma_x^2 \rho}{\sigma_x^2} \sum_{k=1}^{K} R_k^T Z_k \right),$$  \hspace{1cm} (33)

where $I$ represents an identity matrix, $R_k^T Z_k = \sum_{t=1}^{m} R_k^T Z_{k,t}$, and $R_k^T R_k = \sum_{t=1}^{m} R_k^T R_k, i$. Since $\left( I + \frac{\sigma_x^2 \rho}{\sigma_x^2} \sum_{k=1}^{K} R_k^T R_k \right)$ is actually a diagonal matrix, Eq. (33) can be solved efficiently by element-wise division.

Next, we exploit the projection operator to obtain the solution of Eq. (31) on $\Omega$, that is,

$$\hat{x} = A^{-1} \mathcal{P}(A(\hat{x}), \hat{I}, \hat{u}),$$  \hspace{1cm} (34)

**Algorithm 2** The Proposed RRC for Image Compression Artifacts Reduction.

**Require:** JPEG Compressed bit-stream.

1. Get $y$, $M^2$ from the compressed bit-stream.
2. Set parameters $d$, $m$, $L$, $h$, $c$, $\rho$, $\epsilon$ and $\tau$.
3. Initialize $\hat{x}^0 = y$, $Z_k^0 = X_k^0$, $\forall k = 1, \ldots, K$.
4. Calculate $\sigma_x$ by Eq. (19).
5. Determine the solution space $\Omega$ by Eq. (25).
6. for $t = 1$ to MaxIter ($T$) do
   7. Calculate $\sigma_x$ by Eq. (36).
   8. for Each patch $x_k$ in $x$ do
      9. Find similar patches to construct matrix $X_k$.
      10. Perform $[U_k, \Sigma_k, V_k] = SVD(X_k)$.
      11. Estimate the reference matrix $Z_k'$ by Eq. (15).
      12. Perform $[P_k, \Delta_k, Q_k] = SVD(Z_k')$.
      13. Update $\lambda$ by Eq. (13).
      14. Calculate $\mu = \lambda \sigma_x^2 / \rho$.
      15. Determine $\Lambda_k$ by Eq. (30).
      16. Get the estimation: $Z_k = U_k \Lambda_k V_k^T$.
   end for
   17. Update $\hat{x}$ by Eq. (33).
   18. Update $\hat{x}$ by Eq. (34).
   end for
21. **Output:** The final reconstructed image $\hat{x}$.

where $A$ is the matrix operator defined in Eq. (21), and $v = \mathcal{P}(x, l, u)$ is a projection operator defined by

$$v[k] = \begin{cases} I[k], & \text{if } x[k] < l[k], \\ x[k], & \text{if } l[k] \leq x[k] \leq u[k], \\ u[k], & \text{if } x[k] > u[k]. \end{cases}$$  \hspace{1cm} (35)

From Eqs. (33) and (34), we can obtain the solution of Eq. (31). Regarding the two noise parameters $\sigma_x$ and $\sigma_x$, $\sigma_x$ is adaptively calculated by Eq. (19). Since each $\sigma_x$ in Eq. (27) varies as the iteration number increases, its setting can be adaptively adjusted in each iteration. Inspired by [48, 55], $\sigma_x$ in the $t$th iteration is set to

$$\sigma_x^{(t)} = \eta \sqrt{(\sigma_x^2 - \|z^{(t-1)} - y\|_2^2)},$$  \hspace{1cm} (36)

where $\eta$ is a constant and the image $z^{(t-1)}$ is reconstructed by all the $Z_k^{(t-1)}$ at the $(t-1)^{th}$ iteration. After solving the two sub-problems, we summarize the complete algorithm to solve Eq. (26) in Algorithm 2.

V. CONNECTION TO GROUP SPARSE REPRESENTATION

In this section, we provide an analytical investigation on the connection between the proposed RRC model and the popular group-based sparse representation (GSR) model [26–29]. More specifically, we show the equivalence of the proposed RRC and the GSR using a specific method to construct the group-wise dictionaries, i.e., the group sparsity residual constraint (GSRC) model [58, 63–66].

A. Group-based Sparse Representation

Different from the patch-based sparse representation, e.g., K-SVD [67], the GSR [26–29] models $n$ groups of similar
TABLE I
PSNR (dB) COMPARISON OF NNM, BM3D [30], LSSC [29], EPLL [56], PLOW [57], NCSR [58], GID [59], PGPD [27], LINC [60], AGMM [61], OGLR [62], WNMM [5] AND RRC FOR IMAGE DENOISING.

| Image | NNM | BM3D | LSSC | EPLL | PLOW | NCSR | GID | PGPD | LINC | AGMM | OGLR | WNMM | RRC |
|-------|-----|------|------|------|------|------|-----|------|------|------|------|------|-----|
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
|        |      |      |      |      |      |      |     |      |      |      |      |      |     |
| Average |      |      |      |      |      |      |     |      |      |      |      |      |     |

patches extracted from the image $x$, and represent each group using a data matrix $X^*_i \in \mathbb{R}^{d \times m}$. GSR models each data matrix $X^*_i$ using a group-wise sparse representation as

$$\hat{B}^*_i = \arg \min_{B^*_i} \left( \frac{1}{2} \| X^*_i - D_i B^*_i \|_F^2 + \lambda \| B^*_i \|_1 \right) \ \forall i, \quad (37)$$

where each $B^*_i$ is the group sparse coefficient for $X^*_i$ and $D_i$ represents the dictionary, which is usually learned from each group [26, 58, 66]. The $\ell_1$-norm is imposed on each column of $B^*_i$, and here extended to be the $\ell_1$-norm on matrix.

For image restoration tasks, the GSR model can be applied to recover the group data matrices $\{X^*_i\}$ from their degraded measurements $\{Y_i\}$ by solving the following problem,

$$\hat{A}_i = \arg \min_{A_i} \left( \frac{1}{2} \| Y_i - D_i A_i \|_F^2 + \lambda \| A_i \|_1 \right) \ \forall i. \quad (38)$$

where $A_i$ denotes the group sparse coefficient of each group $Y_i$. Once all group sparse codes $\{A_i\}_{i=1}^p$ are obtained, the underlying image $x$ can be reconstructed as $x = D \circ A$, where $D$ is the global dictionary to sparsely represent all the groups of the degraded image $y$. $A$ denotes a set of $\{A_i\}_{i=1}^p$ and $\circ$ is an operator [26].

In practice, one would like to approximate the oracle group sparse coefficient $B^*_i$ using the group sparse coefficient $A_i$ based on the corrupted measurement $Y_i$. The quality of the approximate heavily depends on the difference between $B^*_i$ and $A_i$, i.e., the group sparsity residual, which we define as the following,

$$R^*_i = A_i - B^*_i, \quad (39)$$

Similar to the proposed RRC model, in real applications, $X^*_i$ is inaccessible and we thus employ an estimate of it, denoted by $X'_i$. Given $X'_i$ and the dictionary $D_i$, the group sparse coefficient $B_i$ for each group $X'_i$ is solved by

$$\hat{B}_i = \arg \min_{B_i} \left( \frac{1}{2} \| X'_i - D_i B_i \|_F^2 + \lambda \| B_i \|_1 \right), \quad (40)$$

As both $A_i$ and $B_i$ are to approximate the underlying oracle $B^*_i$, the difference between them needs to be minimized to achieve accurate approximation [58]. Therefore, the enhanced sparse representation problem based on the group sparse residual constraint (GSRC) model [58, 63–66] is the following,

$$\hat{A}_i = \arg \min_{A_i} \left( \frac{1}{2} \| Y_i - D_i A_i \|_F^2 + \lambda \| A_i - B_i \|_1 \right). \quad (41)$$

B. Dictionary Learning based on SVD

To achieve the equivalence of RRC and GSRC, we introduce a specific approach to learn the group-wise dictionaries. For each group $X_i \in \mathbb{R}^{d \times m}$, the corresponding dictionary is constructed using the SVD of the corrupted measurement $Y_i \in \mathbb{R}^{d \times m}$, which is

$$Y_i = U_i \Delta_i V^T_i = \sum_{k, l} \delta_{i, k} u_{i, k} v^T_{i, k}, \quad (42)$$

where $\delta_i = [\delta_{i, 1}, \ldots, \delta_{i, j}]$ and $j = \min(d, m)$; $\Delta_i = \text{diag}(\delta_i)$ is a diagonal matrix whose non-zero elements are represented by $\delta_i$; $u_{i, k}, v_{i, k}$ are the columns of $U_i$ and $V_i$. 

IEEE TRANSACTION ON IMAGE PROCESSING, 2020
respectively. Each dictionary atom $d_{i,k}$ of the group-wise dictionary $D_i$ is constructed as

$$d_{i,k} = u_{i,k} v_{i,k}^T, \quad \forall k = 1, \ldots, j. \tag{43}$$

Therefore, the adaptive dictionary for each data group is formed as $D_i = [d_{i,1}, d_{i,2}, \ldots, d_{i,j}]$ using the corresponding corrupted measurement $Y_i$.

C. The Equivalence of RRC and GSRC

We show the equivalence of the RRC and GSRC models by showing that the problem Eq. (6) and Eq. (41) are equivalent, provided that the group-wise dictionaries are constructed using Eq. (43). We first prove the Lemma 2.

Lemma 2. Let $Y_i = D_i K_i$, $X_i = D_i A_i$, and $D_i$ is constructed by Eq. (43). We have

$$\|Y_i - X_i\|_F^2 = \|K_i - A_i\|_F^2. \tag{44}$$

Proof. See Appendix B.

Based on Lemma 1, Lemma 2 and Theorem 2, we have the equivalence final result as the Theorem 3.

Theorem 3. Under the condition of the adaptive dictionary $D_i$ shown in Eq. (43), the proposed RRC model in Eq. (6) is equivalent to the GSRC model in Eq. (41).

Proof. See Appendix C.

VI. EXPERIMENTAL RESULTS

In this section, we conduct extensive experiments on image denoising and image compression artifacts reduction to verify the effectiveness of the proposed RRC model. To evaluate the
quality of the recovered images, both PSNR and structural similarity (SSIM) [69] metrics are used. The source codes of all competing methods are obtained from the original authors and we used the default parameter settings. For color images, we only focus on the restoration of luminance channel (in YCrCb space). Due to limit space, please enlarge the tables and figures on the screen for better comparison. We choose the following stop criterion for the proposed RRC based image restoration algorithms,
\[ \frac{\|x^t - x^{t-1}\|_2^2}{\|x^{t-1}\|_2^2} < \tau, \]
where \( \tau \) is a small constant. The source code of the proposed RRC method is available at: https://drive.google.com/open?id=18JSSnuF_3x0AvVYq0ElowErnEynTY8cR.

A. Image Denoising

In image denoising, to validate the denoising performance of the proposed RRC model, we compare it with leading denoising methods, including NNM, BM3D [50], LSSC [29], EPLL [56], Plow [57], NCSR [58], GID [59], PGPD [27], LINC [60], aGMM [61] and OGLR [62]. Note that NNM is the baseline algorithm, and the image nonlocal redundancies are used in all competing methods. The parameter settings of the proposed RRC model for image denoising are as follows. The size of each patch \( \sqrt{d} \times \sqrt{d} \) is set to \( 6 \times 6, 7 \times 7, 8 \times 8 \) and \( 9 \times 9 \) for \( \sigma_n \leq 20, 20 < \sigma_n \leq 50, 50 < \sigma_n \leq 75 \) and \( 75 < \sigma_n \leq 100 \), respectively. The parameters \( (\mu, \rho, c, m, \tau) \) are set to \((0.1, 0.9, 0.9, 60, 0.0001), (0.1, 0.8, 0.9, 60, 0.001), (0.1, 0.8, 0.9, 70, 0.0006), (0.1, 0.8, 1.0, 80, 0.0008), (0.1, 0.8, 1.0, 90, 0.0005) \) and \((0.1, 0.8, 1, 100, 0.002) \) for \( \sigma_n \leq 20, 20 < \sigma_n \leq 30, 30 < \sigma_n \leq 40, 40 < \sigma_n \leq 50, 50 < \sigma_n \leq 75 \) and \( 75 < \sigma_n \leq 100 \), respectively. The searching window for similar patches is set to \( L = 25; \epsilon = 0.2, h = 40 \).

We first evaluate the competing methods on 16 widely used test images, whose scenes are displayed in Fig. 5. Zero mean additive white Gaussian noise is added to these test images to generate the noisy observations. We present the denoising results on six noise levels, i.e., \( \sigma_n = \{ 20, 30, 40, 50, 75 \) and 100\}. The PSNR and SSIM results under these noise levels for all methods are shown in Table I and Table II (red indicates the best and blue is the second best performance), respectively. It is obvious that the proposed RRC significantly outperforms the baseline rank minimization method, i.e., NNM. Meanwhile, one can observe that the proposed RRC outperforms other competing methods in most cases in terms of PSNR. The average gains of the proposed RRC over NNM, BM3D, LSSC, EPLL, Plow, NCSR, GID, PGPD, LINC, aGMM and OGLR methods are as much as \( 1.43\text{dB}, 0.24\text{dB}, 0.29\text{dB}, 0.70\text{dB}, 0.72\text{dB}, 0.40\text{dB}, 1.49\text{dB}, 0.14\text{dB}, 0.25\text{dB}, 0.47\text{dB} \) and \( 0.46\text{dB} \), respectively. In terms of SSIM, it can be seen that the proposed RRC also achieves higher results than other competing methods. The only exception is when \( \sigma_n = 20 \) for which LINC is slightly higher than the proposed RRC method. Nonetheless, under high noise level \( \sigma_n = 100 \), the proposed RRC consistently outperforms other competing methods for all test images (the only exception is the images Flower, J. Bean and Lin for which BM3D, NCSR and LINC respectively, are slightly better than the proposed RRC method).

The visual comparisons in the case of \( \sigma_n = 100 \) for images Monarch and Leaves are shown in Fig. 6 and Fig. 7, respectively. It can be found that NNM, EPLL, Plow, NCSR, GID, PGPD, aGMM and OGLR still suffer from some undesirable visual artifacts, and BM3D, LSSC and LINC tend to oversmooth the image. The proposed RRC approach not only removes most of the visual artifacts, but also preserves large scale sharp edges and small-scale image details.

Now, we compare the proposed RRC with WNNM method [5], which is a well-known rank minimization method that delivers state-of-the-art denoising results. The PSNR and SSIM results of WNNM method on 16 widely used test images are shown in Fig. 8 and Fig. 9, respectively. It can be found that the proposed RRC significantly outperforms WNNM in most cases in terms of PSNR and SSIM. The average gains of the proposed RRC over WNNM are as much as \( 1.89\text{dB}, 0.34\text{dB}, 0.76\text{dB}, 1.04\text{dB}, 0.22\text{dB}, 0.49\text{dB}, 0.67\text{dB}, 0.72\text{dB}, 0.89\text{dB}, 0.94\text{dB}, 1.03\text{dB}, 1.42\text{dB}, 0.89\text{dB}, 0.98\text{dB}\) and \( 0.95\text{dB} \), respectively. In terms of SSIM, it can be seen that the proposed RRC also achieves higher results than WNNM method. The only exception is when \( \sigma_n = 20 \) for which WNNM is slightly higher than the proposed RRC method. Nonetheless, under high noise level \( \sigma_n = 100 \), the proposed RRC consistently outperforms WNNM for all test images (the only exception is the images Flower, J. Bean and Lin for which BM3D, NCSR and LINC respectively, are slightly better than the proposed RRC method).

The visual comparisons in the case of \( \sigma_n = 100 \) for images Monarch and Leaves are shown in Fig. 6 and Fig. 7, respectively. It can be found that NNM, EPLL, Plow, NCSR, GID, PGPD, aGMM and OGLR still suffer from some undesirable visual artifacts, and BM3D, LSSC and LINC tend to oversmooth the image. The proposed RRC approach not only removes most of the visual artifacts, but also preserves large scale sharp edges and small-scale image details.

Now, we compare the proposed RRC with WNNM method [5], which is a well-known rank minimization method that delivers state-of-the-art denoising results. The PSNR and SSIM results of WNNM method on 16 widely used test images are
shown in the penultimate column of Table I and Table II, respectively. It can be seen that though the average PSNR results of RRC are slightly lower (< 0.11dB) than WNNM, the SSIM results of the proposed RRC are higher than WNNM when the noise level $\sigma_n > 30$. It is well-known that SSIM [69] often considers the human visual system and leads to more accurate results. The visual comparison of RRC and WNNM with two exemplar images are shown in Fig. 8 and Fig. 9, where we can observe that although PSNR results of the proposed RRC are lower than WNNM, more details are recovered by RRC than WNNM. This phenomenon has been explained in [71].

We also compare the proposed RRC with a representative deep learning method: DN-CNN [72]. On average, the proposed RRC cannot achieve better performance than DN-CNN. Nonetheless, if the images have many similar structures proposed RRC cannot achieve better performance than DN-CNN [72]. On average, the explained in [71].

Fig. 8. Denoising results of Parrot with $\sigma_n = 100$. (a) Original Parrot image; (b) Noisy image; (c) WNNM [5] (PSNR = 24.85dB, SSIM = 0.7533); (d) RRC (PSNR = 24.83dB, SSIM = 0.7729).

Fig. 9. Denoising results of House with $\sigma_n = 100$. (a) Original House image; (b) Noisy image; (c) WNNM [5] (PSNR = 26.79dB, SSIM = 0.7531); (d) RRC (PSNR = 26.38dB, SSIM = 0.7655).

seen in the last two columns of Table I and Table II, respectively. It can be seen that though the average PSNR results of RRC are slightly lower (< 0.11dB) than WNNM, the SSIM results of the proposed RRC are higher than WNNM when the noise level $\sigma_n > 30$. It is well-known that SSIM [69] often considers the human visual system and leads to more accurate results. The visual comparison of RRC and WNNM with two exemplar images are shown in Fig. 8 and Fig. 9, where we can observe that although PSNR results of the proposed RRC are lower than WNNM, more details are recovered by RRC than WNNM. This phenomenon has been explained in [71].

We also compare the proposed RRC with a representative deep learning method: DN-CNN [72]. On average, the proposed RRC cannot achieve better performance than DN-CNN. Nonetheless, if the images have many similar structures and features, such as images Barabara, Fence, Foreman and House, our proposed RRC can outperform DN-CNN, because the nonlocal redundancies are used [32]. An additional advantage of the proposed RRC is that it enjoys the advantage of training free. Due to the page limits, we don’t show the detailed comparison results here.

Furthermore, we comprehensively evaluate our proposed RRC method on 200 test images from the Berkeley Segmentation dataset (BSD) [70]. Table III lists the average PSNR and SSIM comparison results among thirteen competing methods at six noise levels ($\sigma_n = 20, 30, 40, 50, 75$ and 100). One can observe that our proposed RRC approach achieves the second best PSNR result, which is only falling behind WNNM by less than 0.14dB. Nonetheless, our proposed RRC method obtains the best SSIM results on average.

Finally, we consider the proposed RRC model to real image denoising. Due to the fact that the noise level of real noisy images is unknown, and thus the noise level in the image is required to estimate through some noise estimation methods. In this paper, we adopt the scheme proposed in [73] to estimate the noise level. Fig. 10 shows the denoising results of two real images with more complex noise by our proposed RRC method. It can be seen that the proposed RRC method not only recovers visual pleasant results, but also preserves fine image details. Therefore, these results demonstrate the feasibility of our proposed RRC method for practical image denoising applications.

It is well-known that image denoising is an ideal test bench to measure different statistical image models. Obviously, these experiments have verified that the proposed RRC is a promising image model.

The proposed RRC is a traditional model-based optimization algorithm. Here we evaluate the average running time of the proposed RRC model for image denoising by using 16 widely used datasets with different noise levels. The proposed RRC requires about 5–6 minutes for an image on an Intel (R) Core (TM) i3-4150 with 3.56Hz CPU and 4GB memory PC under the Matlab 2015b environment. The running time of the proposed RRC for image denoising is faster than LSSC, QC and LERaG also exploited low-rank priors and achieved the state-of-the-art image deblocking or denoising results. The parameters of our proposed algorithm for image compression artifacts reduction are as follows. The size of each patch $\sqrt{d} \times \sqrt{d}$ is set to $7 \times 7$. The searching window for similar patches is set to $L = 25$. The maximum iteration number is

1The denoising results of our proposed RRC method for BSD 200 dataset [70] is available at: https://drive.google.com/open?id=lqiuqPEAza1mF--9nR5ahXFls8Z2khUc6.

2The compression artifacts reduction results of our proposed RRC method for BSD 100 dataset [70] is available at: https://drive.google.com/open?id=15hUBFvEHO0Xx35y9-Cc90poHTHB0h2.
| QF | JPEG | SA-DCT | PC-LRM | ANCE | DiTvC | BM3D | WNNM | CON COLOR | SSR-QC | LERaG | RRC |
|----|------|--------|--------|------|-------|------|------|-----------|--------|-------|-----|
| 10 | 27.59 | 28.48  | 28.49  | 28.51 | 28.17 | 28.46 | 28.48 | 28.54     | 28.47  | 28.48 | 28.59|
|    | 0.7688 | 0.7896 | 0.7847 | 0.7916 | 0.7825 | 0.7924 | 0.7828 | 0.7907     | 0.7874 | 0.7939 | 0.8015|
| 20 | 29.97 | 30.73  | 30.78  | 30.80 | 30.63 | 30.75 | 30.79 | 30.86     | 30.78  | 30.86 | 30.98|
|    | 0.8530 | 0.8650 | 0.8620 | 0.8663 | 0.8574 | 0.8681 | 0.8611 | 0.8647     | 0.8633 | 0.8661 | 0.8745|
| 30 | 31.37 | 32.07  | 32.16  | 32.18 | 31.82 | 32.09 | 32.17 | 32.29     | 32.18  | 32.27 | 32.41|
|    | 0.8886 | 0.8984 | 0.8987 | 0.8980 | 0.8949 | 0.8963 | 0.8975 | 0.8975     | 0.8975 | 0.8985 | 0.9064|
| 40 | 32.36 | 33.01  | 33.12  | 33.16 | 32.74 | 33.04 | 33.13 | 33.31     | 33.15  | 33.26 | 33.44|
|    | 0.9084 | 0.9168 | 0.9159 | 0.9174 | 0.8995 | 0.9182 | 0.9158 | 0.9172     | 0.9149 | 0.9168 | 0.9231|
| 50 | 33.17 | 33.79  | 33.91  | 33.97 | 33.39 | 33.82 | 33.93 | 34.16     | 33.98  | 34.08 | 34.28|
|    | 0.9223 | 0.9296 | 0.9292 | 0.9303 | 0.9098 | 0.9307 | 0.9291 | 0.9304     | 0.9287 | 0.9296 | 0.9352|
| 60 | 34.02 | 34.60  | 34.73  | 34.83 | 34.32 | 34.63 | 34.75 | 35.02     | 34.83  | 34.92 | 35.15|
|    | 0.9342 | 0.9406 | 0.9404 | 0.9414 | 0.9202 | 0.9415 | 0.9405 | 0.9418     | 0.9404 | 0.9407 | 0.9455|
| 70 | 35.19 | 35.73  | 35.87  | 36.02 | 35.05 | 35.76 | 35.88 | 36.23     | 36.00  | 36.06 | 36.35|
|    | 0.9477 | 0.9529 | 0.9530 | 0.9541 | 0.9319 | 0.9535 | 0.9531 | 0.9546     | 0.9535 | 0.9532 | 0.9572|
| 80 | 36.97 | 37.46  | 37.60  | 37.81 | 36.29 | 37.50 | 37.62 | 38.05     | 37.75  | 37.76 | 38.16|
|    | 0.9827 | 0.9667 | 0.9670 | 0.9680 | 0.9464 | 0.9671 | 0.9671 | 0.9695     | 0.9676 | 0.9699 | 0.9702|
| 90 | 40.61 | 41.00  | 41.15  | 41.45 | 38.11 | 41.07 | 41.16 | 41.69     | 41.27  | 41.11 | 41.60|
|    | 0.9816 | 0.9836 | 0.9838 | 0.9848 | 0.9753 | 0.9838 | 0.9850 | 0.9863     | 0.9836 | 0.9836 | 0.9857|
| Average | 33.47 | 34.10  | 34.20  | 34.30 | 33.37 | 34.12 | 34.31 | 34.46     | 34.37  | 34.32 | 34.57|
| 50 | 32.93 | 33.71  | 33.86  | 34.00 | 33.50 | 33.69 | 33.86 | 34.27     | 34.13  | 34.02 | 34.31|
|    | 0.9181 | 0.9291 | 0.9295 | 0.9301 | 0.9154 | 0.9298 | 0.9293 | 0.9321     | 0.9313 | 0.9302 | 0.9353|
| 60 | 33.77 | 34.50  | 34.66  | 34.86 | 34.25 | 34.50 | 34.67 | 35.17     | 34.97  | 34.88 | 35.20|
|    | 0.9304 | 0.9398 | 0.9402 | 0.9410 | 0.9255 | 0.9403 | 0.9401 | 0.9434     | 0.9425 | 0.9412 | 0.9459|
| 70 | 34.95 | 35.61  | 35.77  | 35.99 | 35.17 | 35.61 | 35.78 | 36.38     | 36.12  | 36.04 | 36.42|
|    | 0.9447 | 0.9517 | 0.9521 | 0.9531 | 0.9365 | 0.9521 | 0.9520 | 0.9557     | 0.9547 | 0.9531 | 0.9575|
| 80 | 36.70 | 37.30  | 37.47  | 37.75 | 36.43 | 37.31 | 37.47 | 38.19     | 37.81  | 37.76 | 38.21|
|    | 0.9602 | 0.9651 | 0.9655 | 0.9668 | 0.9488 | 0.9654 | 0.9655 | 0.9690     | 0.9678 | 0.9666 | 0.9700|
| 90 | 40.33 | 40.79  | 40.95  | 41.25 | 38.23 | 40.82 | 40.96 | 41.71     | 41.19  | 41.01 | 41.76|
|    | 0.9798 | 0.9820 | 0.9822 | 0.9830 | 0.9697 | 0.9822 | 0.9823 | 0.9842     | 0.9834 | 0.9824 | 0.9848|
| Average | 33.28 | 34.06  | 34.20  | 34.33 | 33.49 | 34.06 | 34.20 | 34.62     | 34.42  | 34.31 | 34.66|
| 50 | 0.9080 | 0.9182 | 0.9182 | 0.9192 | 0.9060 | 0.9193 | 0.9178 | 0.9214     | 0.9203 | 0.9199 | 0.9251|
| Average | 0.9084 | 0.9104 | 0.9098 | 0.9114 | 0.8968 | 0.9112 | 0.9095 | 0.9126     | 0.9112 | 0.9149 | 0.9167|
set to $T = 20$; $h = 40$, $\epsilon = 0.2$, $m = 60$, $\rho = 5$ and $w = 0.2$. The parameters $(\eta, c, \tau)$ are set to $(0.3, 0.9, 0.0007)$, $(0.2, 1.3, 0.0005)$, $(0.2, 1.3, 0.0003)$ and $(0.2, 1.5, 0.0003)$ for $QF \leq 10$, $10 < QF \leq 20$, $20 < QF \leq 30$ and $QF > 30$, respectively.

We comprehensively evaluate all competitive methods on these three datasets at each QF. Nine JPEG qualities are evaluated, i.e., QF = [10, 20, 30, 40, 50, 60, 70, 80 and 90]. The average PSNR and SSIM comparisons are provided in Table IV, with the best results highlighted in bold. It can be seen that our proposed RRC consistently outperforms all competing methods on different JPEG QFs. The only exception is the LIVE1 dataset at QF = 10 and 20 for which LERA$G$ is slightly higher than our proposed method. Based on these three datasets, the average gains of our RRC over JPEG, SA-DCT, PC-LRM, ANCE, DiCTV, BM3D, WNNM, CONCOLOR, SSR-QC and LERA$G$ methods are [1.16dB, 0.50dB, 0.40dB, 0.30dB, 1.09dB, 0.50dB, 0.39dB, 0.07dB, 0.27dB and 0.25dB] in PSNR and [0.0170, 0.0065, 0.0071, 0.0054, 0.0207, 0.0054, 0.0074, 0.0043, 0.0057 and 0.0042] in SSIM, respectively.

The visual comparisons of images 85048 and 24077 on BSD 100 dataset with QF = 10 are shown in Fig. 11 and Fig. 12, respectively, where we compare five typical low-rank based methods (i.e., PC-LRM, WNMM, COCOLOR, SSR-QC and LERA$G$) and two well-known image compression reduction methods (i.e., SA-DCT and ANCE) with the proposed RRC method. One can observe that the blocking artifacts are obvious in the image decoded directly by the standard JPEG. For image 85048, it can be observed that some blocking artifacts are still visible in SA-DCT, ANCE and LERA$G$ methods, while PC-LRM, WNMM, COCOLOR and SSR-QC methods generate over-smooth effect. For image 24077, though SA-DCT, PC-LRM, ANCE, WNMM, COCOLOR, SSR-QC and LERA$G$ methods can suppress the blocking artifacts effectively, they often over-smooth the image. Our proposed RRC method not only removes blocking or ringing artifacts across the image, but also preserves large-scale sharp edges and small-scale fine image details. Obviously, compared with these typical low-rank based methods, the proposed RRC method can achieve higher performance. Therefore, these results demonstrate the effectiveness and superiority of our proposed RRC model.

Recently, deep learning based techniques for image compression artifacts reduction have attracted significant attentions due to their impressive performance. We also compare the proposed RRC with the AR-CNN [78] method, which is deemed as the benchmark of CNN-based compression artifacts reduction algorithms. As shown in Table V, the proposed algorithm outperforms the AR-CNN method across all cases on the BSD 100 dataset. The average PSNR and SSIM gains is up to 0.23dB and 0.0069, respectively. The visual comparisons of image 210088 and 78004 on BSD 100 dataset with QF = 10 are shown in Fig. 13 and Fig. 14, respectively. It can be

---

**Fig. 11.** Visual comparison results of image 85048 at QF = 10. (a) Original image; (b) JPEG compressed image (PSNR = 28.00dB, SSIM = 0.7551); (c) SA-DCT [46] (PSNR = 28.97dB, SSIM = 0.7789); (d) PC-LRM [74] (PSNR = 28.97dB, SSIM = 0.7739); (e) ANCE [75] (PSNR = 29.05dB, SSIM = 0.7852); (f) WNNM [5] (PSNR = 28.90dB, SSIM = 0.7684); (g) CONCOLOR [48] (PSNR = 28.97dB, SSIM = 0.7833); (h) SSR-QC [49] (PSNR = 28.92dB, SSIM = 0.7810); (i) LERA$G$ [76] (PSNR = 28.97dB, SSIM = 0.7865); (j) RRC (PSNR = 29.14dB, SSIM = 0.7997).

**Fig. 12.** Visual comparison results of image 24077 at QF = 10. (a) Original image; (b) JPEG compressed image (PSNR = 27.83dB, SSIM = 0.8897); (c) SA-DCT [46] (PSNR = 28.99dB, SSIM = 0.9029); (d) PC-LRM [74] (PSNR = 29.26dB, SSIM = 0.9064); (e) ANCE [75] (PSNR = 28.98dB, SSIM = 0.9022); (f) WNNM [5] (PSNR = 29.30dB, SSIM = 0.9068); (g) CONCOLOR [48] (PSNR = 29.23dB, SSIM = 0.9100); (h) SSR-QC [49] (PSNR = 29.17dB, SSIM = 0.9070); (i) LERA$G$ [76] (PSNR = 29.15dB, SSIM = 0.9060); (j) RRC (PSNR = 29.43dB, SSIM = 0.9119).

**Fig. 13.** Visual comparison results of image 210088 at QF = 10. (a) Original image; (b) JPEG compressed image (PSNR = 30.33dB, SSIM = 0.8342); (c) AR-CNN [78] (PSNR = 33.21dB, SSIM = 0.9147); (d) RRC (PSNR = 33.37dB, SSIM = 0.9201).

**Fig. 14.** Visual comparison results of image 78004 at QF = 10. (a) Original image; (b) JPEG compressed image (PSNR = 27.72dB, SSIM = 0.7854); (c) AR-CNN [78] (PSNR = 28.86dB, SSIM = 0.8206); (d) RRC (PSNR = 29.57dB, SSIM = 0.8361).

**Table V**

| Methods | QF = 10 | QF = 20 | QF = 30 | QF = 40 | Average |
|---------|---------|---------|---------|---------|---------|
| AR-CNN  | 28.35   | 30.75   | 32.14   | 33.08   | 31.15   |
| RRC     | 28.59   | 30.98   | 32.41   | 33.44   | 31.36   |

| Methods | QF = 10 | QF = 20 | QF = 30 | QF = 40 | Average |
|---------|---------|---------|---------|---------|---------|
| AR-CNN  | 0.0170  | 0.0065  | 0.0071  | 0.0054  | 0.0043  |
| RRC     | 0.8015  | 0.8745  | 0.9057  | 0.9231  | 0.8762  |
observed that AR-CNN still suffers from undesirable artifacts and over-smooth effect. The proposed RRC not only preserves sharp edges and fine details, but also eliminates the blocking artifacts more effective than AR-CNN method. These results further verify the superiority of the proposed RRC model.

Next, we will discuss how to select the best quantization constraint parameter $w$ for the performance of the proposed RRC algorithm. In this paper, we adopt the narrow quantization constraint (NQC) $[54]$ by setting $w = 0.2$ rather than traditional $w = 0.5$. Specifically, we report the performance comparisons with different $w$ for BSD 100 dataset in Table VI. It is quite clear that, the setting of $w = 0.2$ achieves better results than that of $w = 0.5$ and obtains about 0.70dB improvement on average, which verifies the effectiveness of NQC.

| QF | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|----|-----|-----|-----|-----|-----|
| 10 | 28.37 | 28.59 | 28.52 | 28.25 | 27.89 |
| 20 | 30.75 | 30.98 | 30.91 | 30.64 | 30.29 |
| Average | 29.56 | 29.79 | 29.72 | 29.45 | 29.09 |

APPENDIX A
PROOF OF THE THEOREM 2

Proof. Supposing that the SVD of $X_i, Y_i, X'_i$ are $X_i = U_i \Sigma_i V_i^T$, $Y_i = P_i \Delta_i S_i^T$ and $X'_i = R_i \Lambda_i Q_i^T$, respectively, where $\Sigma_i$, $\Delta_i$ and $\Lambda_i$ are ordered singular value matrices with the same order. Recalling Eq. (6) and from Theorem 1, we have

$$
\|Y_i - X_i\|_F^2 = \|P_i \Delta_i S_i^T - U_i \Sigma_i V_i^T\|_F^2
= \text{Tr}(\Delta_i \Delta_i^T) + \text{Tr}(\Sigma_i \Sigma_i^T) - 2\text{Tr}(X_i^T Y_i),
$$

(46)

where the equality holds only when $P_i = U_i$ and $S_i = V_i$. Therefore, Eq. (6) is minimized when $P_i = U_i$ and $S_i = V_i$, and the optimal solution of $\Sigma_i$ is obtained by solving

$$
\min_{\Sigma_i \geq 0} \frac{1}{2} \|\Delta_i \Sigma_i - \Delta_i\|_F^2 + \lambda \sum_{k=1}^j |\gamma_{ik,k}|
= \min_{\sigma_{i,k} \geq 0} \sum_{k=1}^j \left( \frac{1}{2} (\delta_{i,k} - \sigma_{i,k})^2 + \lambda |\delta_{i,k} - \psi_{i,k}| \right),
$$

(47)

where $\sigma_{i,k}, \delta_{i,k}$ and $\psi_{i,k}$ are the $k^{th}$ singular value of $X_i, Y_i$ and $X'_i$, respectively.

VII. CONCLUSION

We have proposed a novel rank minimization model, dubbed rank residual constraint (RRC), to reinterpret the rank minimization problem from the perspective of matrix approximation. Different from existing low-rank based methods, which estimated the underlying low-rank matrix directly from the corrupted observations, we progressively approximate the underlying low-rank matrix by minimizing the rank residual. Based on the group-based sparse representation model, an analytical investigation on the feasibility of the RRC model has been provided. We have developed the high performance low-rank matrix estimation based image restoration algorithms via minimizing the rank residual. Specifically, by exploiting the image NSS prior, we have applied the proposed RRC model to image restoration tasks, including image denoising and image compression artifacts reduction. Experimental results have demonstrated that the proposed RRC not only leads to visible quantitative improvements over many state-of-the-art methods, but also preserves the image local structures and suppresses undesirable artifacts.

ACKNOWLEDGE

The authors would like to appreciate the associate editor for coordinating the review of the manuscript, and appreciate the anonymous reviewers for their constructive suggestions to improve the manuscript. The authors would like to appreciate Prof. Jian Zhang of Peking University, Prof. Tao Yue and Prof. Zhan Ma of Nanjing University for their help and appreciate the authors of [5, 27, 29, 30, 46, 48, 49, 53, 56–61, 72, 74–76] for providing their source codes or experimental results.

APPENDIX B
PROOF OF THE LEMMA 2

Proof. From $D_i$ in Eq. (43) and the unitary property of $U_i$ and $V_i$,

$$
\|Y_i - X_i\|_F^2 = \|D_i (K_i - A_i)\|_F^2 = \|U_i \text{diag}(K_i - A_i) V_i^T\|_F^2
= \text{Tr}(U_i \text{diag}(K_i - A_i) V_i^T V_i \text{diag}(K_i - A_i) U_i^T)
= \text{Tr}(U_i \text{diag}(K_i - A_i) \text{diag}(K_i - A_i) U_i^T)
= \text{Tr}(\text{diag}(K_i - A_i) U_i^T U_i \text{diag}(K_i - A_i))
= \text{Tr}(\text{diag}(K_i - A_i) \text{diag}(K_i - A_i))
= \|K_i - A_i\|_F^2.
$$

(48)

APPENDIX C
PROOF OF THE THEOREM 3

Proof. On the basis of Lemma 2, we have

$$
\hat{A}_i = \arg \min_{A_i} \left( \frac{1}{2} \|Y_i - D_i A_i\|_F^2 + \lambda \|A_i - B_i\|_1 \right)
= \arg \min_{A_i} \left( \frac{1}{2} \|K_i - A_i\|_F^2 + \lambda \|A_i - B_i\|_1 \right)
= \arg \min_{\alpha_i} \left( \frac{1}{2} \|\alpha_i - \alpha_i\|_2^2 + \lambda \|\alpha_i - \beta_i\|_1 \right),
$$

(49)
where \( X_i = D_i A_i \) and \( Y_i = D_i K_i \). \( \alpha_i, \beta_i, \) and \( \kappa_i \) denote the vectorization of the matrix \( A_i, B_i \), and \( K_i \), respectively.

Following this, based on Lemma 1, we have

\[
\alpha_i = \text{soft}(\kappa_i - \beta_i, \lambda) + \beta_i, \tag{50}
\]

Then the closed-form solution of the \( k \)-th element \( \alpha_{i,k} \) of \( \alpha_i \) in Eq. (50) is solved by the following problem,

\[
\alpha_{i,k} = \text{soft}(\kappa_{i,k} - \beta_{i,k}, \lambda) + \beta_{i,k}. \tag{51}
\]

Then, based on the adaptive dictionary \( D_i \) in Eq. (43) and Theorem 2, we have proved that Eq. (51) is equivalent to Eq. (11). Note that we assume the PCA space of \( X_i \) and \( X'_i \) are equivalent here. We have thus that RRC is equivalent to GSRC, i.e.,

\[
\hat{X}_i = \arg \min_{A_i} \frac{1}{2} \| Y_i - D_i A_i \|_F^2 + \lambda \| \gamma_{i,k} \|_1 \tag{52}
\]

\[
\hat{A}_i = \arg \min_{A_i} \frac{1}{2} \| Y_i - D_i A_i \|_F^2 + \lambda \| A_i - B_i \|_1 \tag{53}
\]

**REFERENCES**

[1] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi, “Practical low-rank matrix approximation under robust l 1-norm,” in IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012, pp. 1410–1417.

[2] Y. Srebro and R. R. Salakhutdinov, “Collaborative filtering in a non-uniform world: Learning with the weighted trace norm,” in Advances in Neural Information Processing Systems, 2010, pp. 2056–2064.

[3] Y. Mu, J. Dong, X. Yuan, and S. Yan, “Accelerated low-rank visual recovery by random projection,” in IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2011, pp. 2609–2616.

[4] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, “Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images,” IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 11, pp. 2233–2246, 2012.

[5] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, “Weighted nuclear norm minimization and its applications to low level vision,” International Journal of Computer Vision, vol. 121, no. 2, pp. 183–208, 2017.

[6] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” SIAM Journal on Optimization, vol. 20, no. 4, pp. 1956–1982, 2010.

[7] H. Ji, C. Liu, Z. Shen, and Y. Xu, “Robust video denoising using low rank matrix completion,” in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 2010, pp. 1791–1798.

[8] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” Journal of the ACM, vol. 58, no. 3, p. 11, 2011.

[9] G. Liu, Z. Lin, and Y. Yu, “Robust subspace segmentation by low-rank representation,” in International Conference on Machine Learning, 2010, pp. 663–670.

[10] N. Srebro and T. Jaakkola, “Weighted low-rank approximations,” in International Conference on Machine Learning, 2003, pp. 720–727.

[11] A. Eriksson and A. van den Hengel, “Efficient computation of robust low-rank matrix approximations in the presence of missing data using the 11 norm,” in IEEE Conference on Computer Vision and Pattern Recognition, June 2010, pp. 771–778.

[12] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and Y. Yan, “\( 1 \)-norm low-rank matrix factorization by variational bayesian method,” IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 4, pp. 825–839, April 2015.

[13] A. Mnih and R. R. Salakhutdinov, “Probabilistic matrix factorization,” in Advances in neural information processing systems, 2008, pp. 1257–1264.

[14] Y. Kong, M. Shao, K. Li, and Y. Fu, “Probabilistic low-rank multitask learning,” IEEE transactions on neural networks and learning systems, vol. 29, no. 3, pp. 670–680, 2018.
