

Supplementary Materials:

Bilinear Factor Matrix Norm Minimization for Robust PCA: Algorithms and Applications

Fanhua Shang, *Member, IEEE*, James Cheng, Yuanyuan Liu, Zhi-Quan Luo, *Fellow, IEEE*,
and Zhouchen Lin, *Senior Member, IEEE*



In this supplementary material, we give the detailed proofs of some theorems, lemmas and properties. We also provide the stopping criterion of our algorithms and the details of Algorithm 2. In addition, we present two new ADMM algorithms for image recovery application and their pseudo-codes, and some additional experimental results for both synthetic and real-world datasets.

NOTATIONS

\mathbb{R}^l denotes the l -dimensional Euclidean space, and the set of all $m \times n$ matrices¹ with real entries is denoted by $\mathbb{R}^{m \times n}$. $\text{Tr}(X^T Y) = \sum_{ij} X_{ij} Y_{ij}$, where $\text{Tr}(\cdot)$ denotes the trace of a matrix. We assume the singular values of $X \in \mathbb{R}^{m \times n}$ are ordered as $\sigma_1(X) \geq \sigma_2(X) \geq \dots \geq \sigma_r(X) > \sigma_{r+1}(X) = \dots = \sigma_n(X) = 0$, where $r = \text{rank}(X)$. Then the SVD of X is denoted by $X = U \Sigma V^T$, where $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$. I_n denotes an identity matrix of size $n \times n$.

Definition 5. For any vector $x \in \mathbb{R}^l$, its ℓ_p -norm for $0 < p < \infty$ is defined as

$$\|x\|_{\ell_p} = \left(\sum_{i=1}^l |x_i|^p \right)^{1/p}$$

where x_i is the i -th element of x . When $p = 1$, the ℓ_1 -norm of x is $\|x\|_{\ell_1} = \sum_i |x_i|$ (which is convex), while the ℓ_p -norm of x is a quasi-norm when $0 < p < 1$, which is non-convex and violates the triangle inequality. In addition, the ℓ_2 -norm of x is $\|x\|_{\ell_2} = \sqrt{\sum_i x_i^2}$.

The above definition can be naturally extended from vectors to matrices by the following form

$$\|S\|_{\ell_p} = \left(\sum_{i,j} |s_{i,j}|^p \right)^{1/p}.$$

Definition 6. The Schatten- p norm ($0 < p < \infty$) of a matrix $X \in \mathbb{R}^{m \times n}$ is defined as follows:

$$\|X\|_{S_p} = \left(\sum_{i=1}^n \sigma_i^p(X) \right)^{1/p}$$

where $\sigma_i(X)$ denotes the i -th largest singular value of X .

-
- F. Shang, J. Cheng and Y. Liu are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. E-mail: {fhshang, jcheng, yyliu}@cse.cuhk.edu.hk.
 - Z.-Q. Luo is with Shenzhen Research Institute of Big Data, the Chinese University of Hong Kong, Shenzhen, China and Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA. E-mail: luoqz@cuhk.edu.cn and luoqz@umn.edu.
 - Z. Lin is with Key Laboratory of Machine Perception (MOE), School of EECS, Peking University, Beijing 100871, P.R. China, and the Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, Shanghai 200240, P.R. China. E-mail: zlin@pku.edu.cn.

1. Without loss of generality, we assume $m \geq n$ in this paper.

In the following, we will list some special cases of the Schatten- p norm ($0 < p < \infty$).

- When $0 < p < 1$, the Schatten- p norm is a quasi-norm, and it is non-convex and violates the triangle inequality.
- When $p = 1$, the Schatten-1 norm (also known as the nuclear norm or trace norm) of X is defined as

$$\|X\|_* = \sum_{i=1}^n \sigma_i(X).$$

- When $p = 2$, the Schatten-2 norm is more commonly called the Frobenius norm² defined as

$$\|X\|_F = \sqrt{\sum_{i=1}^n \sigma_i^2(X)} = \sqrt{\sum_{i,j} X_{i,j}^2}.$$

APPENDIX A: PROOF OF LEMMA 2

To prove Lemma 2, we first define the doubly stochastic matrix, and give the following lemma.

Definition 7. A square matrix is doubly stochastic if its elements are non-negative real numbers, and the sum of elements of each row or column is equal to 1.

Lemma 7. Let $P \in \mathbb{R}^{n \times n}$ be a doubly stochastic matrix, and if

$$0 \leq x_1 \leq x_2 \leq \dots \leq x_n, \quad y_1 \geq y_2 \geq \dots \geq y_n \geq 0, \quad (36)$$

then

$$\sum_{i,j=1}^n p_{ij} x_i y_j \geq \sum_{k=1}^n x_k y_k.$$

The proof of Lemma 7 is essentially similar to that of the lemma in [1], thus we give the following proof sketch for this lemma.

Proof: Using (36), there exist non-negative numbers α_i and β_j for all $1 \leq i, j \leq n$ such that

$$x_k = \sum_{1 \leq i \leq k} \alpha_i, \quad y_k = \sum_{k \leq j \leq n} \beta_j \quad \text{for all } k = 1, \dots, n.$$

Let δ_{ij} denote the Kronecker delta (i.e., $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise), we have

$$\begin{aligned} \sum_{k=1}^n x_k y_k - \sum_{i,j=1}^n p_{ij} x_i y_j &= \sum_{i,j=1}^n (\delta_{ij} - p_{ij}) x_i y_j \\ &= \sum_{1 \leq i,j \leq n} (\delta_{ij} - p_{ij}) \sum_{1 \leq r \leq i} \alpha_r \sum_{j \leq s \leq n} \beta_s \\ &= \sum_{1 \leq r,s \leq n} \alpha_r \beta_s \sum_{r \leq i \leq n, 1 \leq j \leq s} (\delta_{ij} - p_{ij}). \end{aligned}$$

If $r \leq s$, by the lemma in [1] we know that

$$\begin{aligned} \sum_{1 \leq i < r, 1 \leq j \leq s} (\delta_{ij} - p_{ij}) &\geq 0, \quad \text{and} \\ \sum_{r \leq i \leq n, 1 \leq j \leq s} (\delta_{ij} - p_{ij}) + \sum_{1 \leq i < r, 1 \leq j \leq s} (\delta_{ij} - p_{ij}) &= 0. \end{aligned}$$

Therefore, we have

$$\sum_{r \leq i \leq n, 1 \leq j \leq s} (\delta_{ij} - p_{ij}) \leq 0.$$

The same result can be obtained in a similar way for $r \geq s$. □

Proof of Lemma 2:

Proof: Using the properties of the trace, we know that

$$\text{Tr}(X^T Y) = \sum_{i,j=1}^n (U_X)_{ij}^2 \lambda_i \tau_j.$$

Note that U_X is a unitary matrix, i.e., $U_X^T U_X = U_X U_X^T = I_n$, which implies that $((U_X)_{ij}^2)$ is a doubly stochastic matrix. By Lemma 7, we have $\text{Tr}(X^T Y) \geq \sum_{i=1}^n \lambda_i \tau_i$. □

2. Note that the Frobenius norm is the induced norm of the ℓ_2 -norm on matrices.

APPENDIX B: PROOFS OF THEOREMS 1 AND 2

Proof of Theorem 1:

Proof: Using Lemma 3, for any factor matrices $U \in \mathbb{R}^{m \times d}$ and $V \in \mathbb{R}^{n \times d}$ with the constraint $X = UV^T$, we have

$$\left(\frac{\|U\|_* + \|V\|_*}{2} \right)^2 \geq \|X\|_{S_{1/2}}.$$

On the other hand, let $U^* = L_X \Sigma_X^{1/2}$ and $V^* = R_X \Sigma_X^{1/2}$, where $X = L_X \Sigma_X R_X^T$ is the SVD of X as in Lemma 3, then we have

$$X = U^*(V^*)^T \text{ and } \|X\|_{S_{1/2}} = (\|U^*\|_* + \|V^*\|_*)^2/4.$$

Therefore, we have

$$\min_{U, V: X=UV^T} \left(\frac{\|U\|_* + \|V\|_*}{2} \right)^2 = \|X\|_{S_{1/2}}.$$

This completes the proof. \square

Proof of Theorem 2:

Proof: Using Lemma 4, for any $U \in \mathbb{R}^{m \times d}$ and $V \in \mathbb{R}^{n \times d}$ with the constraint $X = UV^T$, we have

$$\left(\frac{\|U\|_F^2 + 2\|V\|_*}{3} \right)^{3/2} \geq \|X\|_{S_{2/3}}.$$

On the other hand, let $U^* = L_X \Sigma_X^{1/3}$ and $V^* = R_X \Sigma_X^{2/3}$, where $X = L_X \Sigma_X R_X^T$ is the SVD of X as in Lemma 3, then we have

$$X = U^*(V^*)^T \text{ and } \|X\|_{S_{2/3}} = [(\|U^*\|_F^2 + 2\|V^*\|_*)/3]^{3/2}.$$

Thus, we have

$$\min_{U, V: X=UV^T} \left(\frac{\|U\|_F^2 + 2\|V\|_*}{3} \right)^{3/2} = \|X\|_{S_{2/3}}.$$

This completes the proof. \square

APPENDIX C: PROOF OF PROPERTY 4

Proof: The proof of Property 4 involves some properties of the ℓ_p -norm, which we recall as follows. For any vector x in \mathbb{R}^n and $0 < p_2 \leq p_1 \leq 1$, the following inequalities hold:

$$\|x\|_{\ell_1} \leq \|x\|_{\ell_{p_1}}, \quad \|x\|_{\ell_{p_1}} \leq \|x\|_{\ell_{p_2}} \leq n^{\frac{1}{p_2} - \frac{1}{p_1}} \|x\|_{\ell_{p_1}}.$$

Let X be an $m \times n$ matrix of rank r , and denote its compact SVD by $X = U_{m \times r} \Sigma_{r \times r} V_{n \times r}^T$. By Theorems 1 and 2, and the properties of the ℓ_p -norm mentioned above, we have

$$\|X\|_* = \|\text{diag}(\Sigma_{r \times r})\|_{\ell_1} \leq \|\text{diag}(\Sigma_{r \times r})\|_{\ell_{\frac{1}{2}}} = \|X\|_{\text{D-N}} \leq \text{rank}(X) \|X\|_*,$$

$$\|X\|_* = \|\text{diag}(\Sigma_{r \times r})\|_{\ell_1} \leq \|\text{diag}(\Sigma_{r \times r})\|_{\ell_{\frac{2}{3}}} = \|X\|_{\text{F-N}} \leq \sqrt{\text{rank}(X)} \|X\|_*.$$

In addition,

$$\|X\|_{\text{F-N}} = \|\text{diag}(\Sigma_{r \times r})\|_{\ell_{\frac{2}{3}}} \leq \|\text{diag}(\Sigma_{r \times r})\|_{\ell_{\frac{1}{2}}} = \|X\|_{\text{D-N}}.$$

Therefore, we have

$$\|X\|_* \leq \|X\|_{\text{F-N}} \leq \|X\|_{\text{D-N}} \leq \text{rank}(X) \|X\|_*.$$

This completes the proof. \square

APPENDIX D: SOLVING (18) VIA ADMM

Similar to Algorithm 1, we also propose an efficient algorithm based on the alternating direction method of multipliers (ADMM) to solve (18), whose augmented Lagrangian function is given by

$$\begin{aligned} \mathcal{L}_\mu(U, V, L, S, \hat{V}, Y_1, Y_2, Y_3) = & \frac{\lambda}{3} \left(\|U\|_F^2 + 2\|\hat{V}\|_* \right) + \|\mathcal{P}_\Omega(S)\|_{\ell_{2/3}}^{2/3} + \langle Y_1, \hat{V} - V \rangle + \langle Y_2, UV^T - L \rangle + \langle Y_3, L + S - D \rangle \\ & + \frac{\mu}{2} \left(\|UV^T - L\|_F^2 + \|\hat{V} - V\|_F^2 + \|L + S - D\|_F^2 \right) \end{aligned}$$

where $Y_1 \in \mathbb{R}^{n \times d}$, $Y_2 \in \mathbb{R}^{m \times n}$ and $Y_3 \in \mathbb{R}^{m \times n}$ are the matrices of Lagrange multipliers.

Algorithm 2 ADMM for solving (S+L)_{2/3} problem (18)**Input:** $D \in \mathbb{R}^{m \times n}$, the given rank d and λ .**Initialize:** $\mu_0, \rho > 1, k=0$, and ϵ .

```

1: while not converged do
2:   while not converged do
3:     Update  $U_{k+1}$  and  $V_{k+1}$  by (61) and (62), respectively.
4:     Update  $\hat{V}_{k+1}$  by  $\hat{V}_{k+1} = \mathcal{D}_{2\lambda/3\mu_k}(V_{k+1} - (Y_2)_k/\mu_k)$ .
5:     Update  $L_{k+1}$  and  $S_{k+1}$  by (43) and (33) in this paper, respectively.
6:   end while // Inner loop
7:   Update the multipliers  $Y_1^{k+1}, Y_2^{k+1}$  and  $Y_3^{k+1}$  by
      $Y_1^{k+1} = Y_1^k + \mu_k(\hat{V}_{k+1} - V_{k+1}), Y_2^{k+1} = Y_2^k + \mu_k(U_{k+1}V_{k+1}^T - L_{k+1})$ , and  $Y_3^{k+1} = Y_3^k + \mu_k(L_{k+1} + S_{k+1} - D)$ .
8:   Update  $\mu_{k+1}$  by  $\mu_{k+1} = \rho\mu_k$ .
9:    $k \leftarrow k + 1$ .
10: end while // Outer loop
Output:  $U_{k+1}$  and  $V_{k+1}$ .

```

Update of U_{k+1} and V_{k+1} :For updating U_{k+1} and V_{k+1} , we consider the following optimization problems:

$$\min_U \frac{\lambda}{3} \|U\|_F^2 + \frac{\mu_k}{2} \|UV_k^T - L_k + \mu_k^{-1}Y_2^k\|_F^2, \quad (37)$$

$$\min_V \|\hat{V}_k - V + \mu_k^{-1}Y_1^k\|_F^2 + \|U_{k+1}V^T - L_k + \mu_k^{-1}Y_2^k\|_F^2, \quad (38)$$

and their optimal solutions can be given by

$$U_{k+1} = \mu_k P_k V_k \left(\frac{2\lambda}{3} I_d + \mu_k V_k^T V_k \right)^{-1}, \quad (39)$$

$$V_{k+1} = \left[\hat{V}_k + \mu_k^{-1}Y_1^k + P_k^T U_{k+1} \right] \left(I_d + U_{k+1}^T U_{k+1} \right)^{-1}, \quad (40)$$

where $P_k = L_k - \mu_k^{-1}Y_2^k$.**Update of \hat{V}_{k+1} :**To update \hat{V}_{k+1} , we fix the other variables and solve the following optimization problem

$$\min_{\hat{V}} \frac{2\lambda}{3} \|\hat{V}\|_* + \frac{\mu_k}{2} \|\hat{V} - V_{k+1} + Y_1^k/\mu_k\|_F^2. \quad (41)$$

Similar to (23) and (24), the closed-form solution of (41) can also be obtained by the SVT operator [2] defined as follows.

Definition 8. Let Y be a matrix of size $m \times n$ ($m \geq n$), and $U_Y \Sigma_Y V_Y^T$ be its SVD. Then the singular value thresholding (SVT) operator \mathcal{D}_τ is defined as [2]:

$$\mathcal{D}_\tau(Y) = U_Y \mathcal{S}_\tau(\Sigma_Y) V_Y^T,$$

where $\mathcal{S}_\tau(x) = \max(|x| - \tau, 0) \cdot \text{sgn}(x)$ is the soft shrinkage operator [3], [4], [5].**Update of L_{k+1} :**For updating L_{k+1} , we consider the following optimization problem:

$$\min_L \|U_{k+1}V_{k+1}^T - L + \mu_k^{-1}Y_2^k\|_F^2 + \|L + S_k - D + \mu_k^{-1}Y_3^k\|_F^2. \quad (42)$$

Since (42) is a least squares problem, and thus its closed-form solution is given by

$$L_{k+1} = \frac{1}{2} \left(U_{k+1}V_{k+1}^T + \mu_k^{-1}Y_2^k - S_k + D - \mu_k^{-1}Y_3^k \right). \quad (43)$$

Together with the update scheme of S_{k+1} , as stated in (33) in this paper, we develop an efficient ADMM algorithm to solve the Frobenius/nuclear hybrid norm penalized RPCA problem (18), as outlined in **Algorithm 2**.

APPENDIX E: PROOF OF THEOREM 3

In this part, we first prove the boundedness of multipliers and some variables of Algorithm 1, and then we analyze the convergence of Algorithm 1. To prove the boundedness, we first give the following lemma.

Lemma 8 ([6]). Let \mathcal{H} be a real Hilbert space endowed with an inner product $\langle \cdot, \cdot \rangle$ and a corresponding norm $\|\cdot\|$, and $y \in \partial\|x\|$, where $\partial f(x)$ denotes the subgradient of $f(x)$. Then $\|y\|^* = 1$ if $x \neq 0$, and $\|y\|^* \leq 1$ if $x = 0$, where $\|\cdot\|^*$ is the dual norm of $\|\cdot\|$. For instance, the dual norm of the nuclear norm is the spectral norm, $\|\cdot\|_2$, i.e., the largest singular value.

Lemma 9 (Boundedness). Let $Y_1^{k+1} = Y_1^k + \mu_k(\widehat{U}_{k+1} - U_{k+1})$, $Y_2^{k+1} = Y_2^k + \mu_k(\widehat{V}_{k+1} - V_{k+1})$, $Y_3^{k+1} = Y_3^k + \mu_k(U_{k+1}V_{k+1}^T - L_{k+1})$ and $Y_4^{k+1} = Y_4^k + \mu_k(L_{k+1} + S_{k+1} - D)$. Suppose that μ_k is non-decreasing and $\sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k} < \infty$, then the sequences $\{(U_k, V_k)\}$, $\{(\widehat{U}_k, \widehat{V}_k)\}$, $\{(Y_1^k, Y_2^k, Y_3^k, Y_4^k)\}$, $\{L_k\}$ and $\{S_k\}$ produced by Algorithm 1 are all bounded.

Proof: Let $\mathcal{U}_k := (U_k, V_k)$, $\mathcal{V}_k := (\widehat{U}_k, \widehat{V}_k)$ and $\mathcal{Y}_k := (Y_1^k, Y_2^k, Y_3^k, Y_4^k)$. By the first-order optimality conditions of the augmented Lagrangian function of (17) with respect to \widehat{U} and \widehat{V} (i.e., Problems (23) and (24)), we have

$$0 \in \partial_{\widehat{U}} \mathcal{L}_{\mu_k}(\mathcal{U}_{k+1}, \mathcal{V}_{k+1}, L_{k+1}, S_{k+1}, \mathcal{Y}_k) \quad \text{and} \quad 0 \in \partial_{\widehat{V}} \mathcal{L}_{\mu_k}(\mathcal{U}_{k+1}, \mathcal{V}_{k+1}, L_{k+1}, S_{k+1}, \mathcal{Y}_k),$$

i.e., $-Y_1^{k+1} \in \frac{\lambda}{2} \partial \|\widehat{U}_{k+1}\|_*$ and $-Y_2^{k+1} \in \frac{\lambda}{2} \partial \|\widehat{V}_{k+1}\|_*$, where $Y_1^{k+1} = \mu_k(\widehat{U}_{k+1} - U_{k+1} + Y_1^k/\mu_k)$ and $Y_2^{k+1} = \mu_k(\widehat{V}_{k+1} - V_{k+1} + Y_2^k/\mu_k)$. By Lemma 8, we obtain

$$\|Y_1^{k+1}\|_2 \leq \frac{\lambda}{2} \quad \text{and} \quad \|Y_2^{k+1}\|_2 \leq \frac{\lambda}{2},$$

which implies that the sequences $\{Y_1^k\}$ and $\{Y_2^k\}$ are bounded.

Next we prove the boundedness of $\{Y_4^k/\sqrt[3]{\mu_{k-1}}\}$. Using (27), it is easy to show that $\mathcal{P}_{\Omega}^{\perp}(Y_4^{k+1}) = 0$, which implies that the sequence $\{\mathcal{P}_{\Omega}^{\perp}(Y_4^{k+1})\}$ is bounded. Let $A = D - L_{k+1} - \mu_k^{-1}Y_4^k$, and $\Phi(S) := \|\mathcal{P}_{\Omega}(S)\|_{\ell_1/2}^{1/2}$. To prove the boundedness of $\{Y_4^k/\sqrt[3]{\mu_{k-1}}\}$, we consider the following two cases.

Case 1: $|a_{ij}| > \frac{3}{2\sqrt[3]{\mu_k^2}}$, $(i, j) \in \Omega$.

If $|a_{ij}| > \frac{3}{2\sqrt[3]{\mu_k^2}}$, and using Proposition 1, then $|[S_{k+1}]_{ij}| > 0$. The first-order optimality condition of (27) implies that

$$[\partial\Phi(S_{k+1})]_{ij} + [Y_4^{k+1}]_{ij} = 0,$$

where $[\partial\Phi(S_{k+1})]_{ij}$ denotes the gradient of the penalty $\Phi(S)$ at $[S_{k+1}]_{ij}$. Since $\partial\Phi(s_{ij}) = \text{sign}(s_{ij})/(2\sqrt{|s_{ij}|})$, then

$$|[Y_4^{k+1}]_{ij}| = \left| \frac{1}{2\sqrt{|[S_{k+1}]_{ij}|}} \right|.$$

Using Proposition 1, we have

$$\left| \frac{[Y_4^{k+1}]_{ij}}{\sqrt[3]{\mu_k}} \right| = \left| \frac{1}{2\sqrt[3]{\mu_k}\sqrt{|[S_{k+1}]_{ij}|}} \right| = \frac{\sqrt{3}}{2\sqrt{2|a_{ij}|(1 + \cos(\frac{2\pi - 2\phi_{\gamma}(a_{ij})}{3}))}\sqrt[3]{\mu_k}} \leq \frac{1}{\sqrt{2}},$$

where $\gamma = 2/\mu_k$.

Case 2: $|a_{ij}| \leq \frac{3}{2\sqrt[3]{\mu_k^2}}$, $(i, j) \in \Omega$.

If $|a_{ij}| \leq \frac{3}{2\sqrt[3]{\mu_k^2}}$, and using Proposition 1, we have $[S_{k+1}]_{ij} = 0$. Since

$$|[Y_4^{k+1}/\mu_k]_{ij}| = |[L_{k+1} + \mu_k^{-1}Y_4^k - D + S_{k+1}]_{ij}| = |[L_{k+1} + \mu_k^{-1}Y_4^k - D]_{ij}| = |a_{ij}| \leq \frac{3}{2\sqrt[3]{\mu_k^2}},$$

then

$$\left| \frac{[Y_4^{k+1}]_{ij}}{\sqrt[3]{\mu_k}} \right| \leq \frac{3}{2}, \quad \text{and} \quad \|Y_4^{k+1}\|_F / \sqrt[3]{\mu_k} = \|\mathcal{P}_{\Omega}(Y_4^{k+1})\|_F / \sqrt[3]{\mu_k} \leq \frac{3\|\Omega\|}{2}.$$

Therefore, $\{Y_4^k/\sqrt[3]{\mu_{k-1}}\}$ is bounded.

By the iterative scheme of Algorithm 1, we have

$$\begin{aligned} \mathcal{L}_{\mu_k}(\mathcal{U}_{k+1}, \mathcal{V}_{k+1}, L_{k+1}, S_{k+1}, \mathcal{Y}_k) &\leq \mathcal{L}_{\mu_k}(\mathcal{U}_{k+1}, \mathcal{V}_{k+1}, L_k, S_k, \mathcal{Y}_k) \\ &\leq \mathcal{L}_{\mu_k}(\mathcal{U}_k, \mathcal{V}_k, L_k, S_k, \mathcal{Y}_k) \\ &= \mathcal{L}_{\mu_{k-1}}(\mathcal{U}_k, \mathcal{V}_k, L_k, S_k, \mathcal{Y}_{k-1}) + \alpha_k \sum_{j=1}^3 \|Y_j^k - Y_j^{k-1}\|_F^2 + \beta_k \frac{\|Y_4^k - Y_4^{k-1}\|_F^2}{\sqrt[3]{\mu_{k-1}^2}}, \end{aligned}$$

where $\alpha_k = \frac{\mu_{k-1} + \mu_k}{2(\mu_{k-1})^2}$, $\beta_k = \frac{\mu_{k-1} + \mu_k}{2(\mu_{k-1})^{4/3}}$, and the above equality holds due to the definition of the augmented Lagrangian function $\mathcal{L}_\mu(U, V, L, S, \hat{U}, \hat{V}, \{Y_i\})$. Since μ_k is non-decreasing, and $\sum_{k=1}^{\infty} (\mu_k / \mu_{k-1}^{4/3}) < \infty$, then

$$\begin{aligned} \sum_{k=1}^{\infty} \beta_k &= \sum_{k=1}^{\infty} \frac{\mu_{k-1} + \mu_k}{2\mu_{k-1}^{4/3}} \leq \sum_{k=1}^{\infty} \frac{\mu_k}{\mu_{k-1}^{4/3}} < \infty, \\ \sum_{k=1}^{\infty} \alpha_k &= \sum_{k=1}^{\infty} \frac{\mu_{k-1} + \mu_k}{2\mu_{k-1}^2} \leq \sum_{k=1}^{\infty} \frac{\mu_k}{\mu_{k-1}^2} < \sum_{k=1}^{\infty} \frac{\mu_k}{\mu_{k-1}^{4/3}} < \infty. \end{aligned}$$

Since $\{\|Y_4^k\|_F / \sqrt[3]{\mu_{k-1}}\}$ is bounded, and $\mu_k = \rho\mu_{k-1}$ and $\rho > 1$, then $\{\|Y_4^{k-1}\|_F / \sqrt[3]{\mu_{k-1}}\}$ is also bounded, which implies that $\{\|Y_4^k - Y_4^{k-1}\|_F^2 / \sqrt[3]{\mu_{k-1}^2}\}$ is bounded. Then $\{\mathcal{L}_{\mu_k}(\mathcal{U}_{k+1}, \mathcal{V}_{k+1}, L_{k+1}, S_{k+1}, \mathcal{B}_k)\}$ is upper-bounded due to the boundedness of the sequences of all Lagrange multipliers, i.e., $\{Y_1^k\}$, $\{Y_2^k\}$, $\{Y_3^k\}$ and $\{\|Y_4^k - Y_4^{k-1}\|_F^2 / \sqrt[3]{\mu_k^2}\}$.

$$\frac{\lambda}{2}(\|\hat{U}\|_* + \|\hat{V}\|_*) + \|\mathcal{P}_\Omega(S_k)\|_{\ell_{1/2}}^{1/2} = \mathcal{L}_{\mu_{k-1}}(\mathcal{U}_k, \mathcal{V}_k, L_k, S_k, \mathcal{B}_{k-1}) - \frac{1}{2} \sum_{i=1}^4 \frac{\|Y_i^k\|_F^2 - \|Y_i^{k-1}\|_F^2}{\mu_{k-1}}$$

is upper-bounded (note that the above equality holds due to the definition of $\mathcal{L}_\mu(U, V, L, S, \hat{U}, \hat{V}, \{Y_i\})$), thus $\{S_k\}$, $\{\hat{U}_k\}$ and $\{\hat{V}_k\}$ are all bounded.

Similarly, by $U_k = \hat{U}_k - [Y_1^k - Y_1^{k-1}] / \mu_{k-1}$, $V_k = \hat{V}_k - [Y_2^k - Y_2^{k-1}] / \mu_{k-1}$, $L_k = U_k V_k^T - [Y_3^k - Y_3^{k-1}] / \mu_{k-1}$ and the boundedness of $\{\hat{U}_k\}$, $\{\hat{V}_k\}$, $\{Y_i^k\}$ ($i=1, 2, 3$), and $\{Y_4^k / \sqrt[3]{\mu_{k-1}}\}$, thus $\{U_k\}$, $\{V_k\}$ and $\{L_k\}$ are also bounded. This means that each bounded sequence must have a convergent subsequence due to the Bolzano-Weierstrass theorem. \square

Proof of Theorem 3:

Proof: (I) $\hat{U}_{k+1} - U_{k+1} = [Y_1^{k+1} - Y_1^k] / \mu_k$, $\hat{V}_{k+1} - V_{k+1} = [Y_2^{k+1} - Y_2^k] / \mu_k$, $U_{k+1} V_{k+1}^T - L_{k+1} = [Y_3^{k+1} - Y_3^k] / \mu_k$ and $L_{k+1} + S_{k+1} - D = [Y_4^{k+1} - Y_4^k] / \mu_k$. Due to the boundedness of $\{Y_1^k\}$, $\{Y_2^k\}$, $\{Y_3^k\}$ and $\{Y_4^k / \sqrt[3]{\mu_{k-1}}\}$, the non-decreasing property of $\{\mu_k\}$, and $\sum_{k=0}^{\infty} (\mu_{k+1} / \mu_k^{4/3}) < \infty$, we have

$$\begin{aligned} \sum_{k=0}^{\infty} \|\hat{U}_{k+1} - U_{k+1}\|_F &\leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} \|Y_1^{k+1} - Y_1^k\|_F < \infty, \quad \sum_{k=0}^{\infty} \|\hat{V}_{k+1} - V_{k+1}\|_F \leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} \|Y_2^{k+1} - Y_2^k\|_F < \infty, \\ \sum_{k=0}^{\infty} \|L_{k+1} - U_{k+1} V_{k+1}^T\|_F &\leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} \|Y_3^{k+1} - Y_3^k\|_F < \infty, \quad \sum_{k=0}^{\infty} \|L_{k+1} + S_{k+1} - D\|_F \leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^{5/3}} \frac{\|Y_4^{k+1} - Y_4^k\|_F}{\mu_k^{1/3}} < \infty, \end{aligned}$$

which implies that

$$\begin{aligned} \lim_{k \rightarrow \infty} \|\hat{U}_{k+1} - U_{k+1}\|_F &= 0, \quad \lim_{k \rightarrow \infty} \|\hat{V}_{k+1} - V_{k+1}\|_F = 0, \quad \lim_{k \rightarrow \infty} \|L_{k+1} - U_{k+1} V_{k+1}^T\|_F = 0, \\ \text{and } \lim_{k \rightarrow \infty} \|L_{k+1} + S_{k+1} - D\|_F &= 0. \end{aligned}$$

Hence, $\{(U_k, V_k, \hat{U}_k, \hat{V}_k, L_k, S_k)\}$ approaches to a feasible solution. In the following, we will prove that the sequences $\{U_k\}$ and $\{V_k\}$ are Cauchy sequences.

Using $Y_1^k = Y_1^{k-1} + \mu_{k-1}(\hat{U}_k - U_k)$, $Y_2^k = Y_2^{k-1} + \mu_{k-1}(\hat{V}_k - V_k)$ and $Y_3^k = Y_3^{k-1} + \mu_{k-1}(U_k V_k^T - L_k)$, then the first-order optimality conditions of (19) and (20) with respect to U and V are written as follows:

$$\begin{aligned} &\left(U_{k+1} V_k^T - L_k + \frac{Y_3^k}{\mu_k} \right) V_k + \left(U_{k+1} - \hat{U}_k - \frac{Y_1^k}{\mu_k} \right) \\ &= \left(U_{k+1} V_k^T - U_k V_k^T - \frac{Y_3^{k-1}}{\mu_{k-1}} + \frac{Y_3^k}{\mu_{k-1}} + \frac{Y_3^k}{\mu_k} \right) V_k + U_{k+1} - U_k + U_k - \hat{U}_k - \frac{Y_1^k}{\mu_k} \\ &= (U_{k+1} - U_k)(V_k^T V_k + I_d) + \left(\frac{Y_3^k - Y_3^{k-1}}{\mu_{k-1}} + \frac{Y_3^k}{\mu_k} \right) V_k + \frac{Y_1^{k-1} - Y_1^k}{\mu_{k-1}} - \frac{Y_1^k}{\mu_k} \\ &= 0, \end{aligned} \tag{44}$$

$$\begin{aligned}
& \left(V_{k+1} U_{k+1}^T - L_k^T + \frac{(Y_3^k)^T}{\mu_k} \right) U_{k+1} + \left(V_{k+1} - \hat{V}_k - \frac{Y_2^k}{\mu_k} \right) \\
&= \left(V_{k+1} U_{k+1}^T - V_k U_k^T - \frac{(Y_3^{k-1})^T}{\mu_{k-1}} + \frac{(Y_3^k)^T}{\mu_{k-1}} + \frac{(Y_3^k)^T}{\mu_k} \right) U_{k+1} + V_{k+1} - V_k + V_k - \hat{V}_k - \frac{Y_2^k}{\mu_k} \\
&= (V_{k+1} - V_k)(U_{k+1}^T U_{k+1} + I_d) + V_k(U_{k+1}^T - U_k^T)U_{k+1} + \left(\frac{(Y_3^k)^T - (Y_3^{k-1})^T}{\mu_{k-1}} + \frac{(Y_3^k)^T}{\mu_k} \right) U_{k+1} + \frac{Y_2^{k-1} - Y_2^k}{\mu_{k-1}} - \frac{Y_2^k}{\mu_k} \\
&= 0.
\end{aligned} \tag{45}$$

By (44) and (45), we obtain

$$\begin{aligned}
& U_{k+1} - U_k \\
&= \left[\left(\frac{Y_3^{k-1} - Y_3^k}{\mu_{k-1}} - \frac{Y_3^k}{\mu_k} \right) V_k + \frac{Y_1^k - Y_1^{k-1}}{\mu_{k-1}} + \frac{Y_1^k}{\mu_k} \right] (V_k^T V_k + I_d)^{-1}, \\
& V_{k+1} - V_k \\
&= \left[V_k(U_k^T - U_{k+1}^T)U_{k+1} + \left(\frac{(Y_3^{k-1})^T - (Y_3^k)^T}{\mu_{k-1}} - \frac{(Y_3^k)^T}{\mu_k} \right) U_{k+1} + \frac{Y_2^k - Y_2^{k-1}}{\mu_{k-1}} + \frac{Y_2^k}{\mu_k} \right] (U_{k+1}^T U_{k+1} + I_d)^{-1}.
\end{aligned}$$

Recall that

$$\sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^{4/3}} < \infty,$$

we have

$$\sum_{k=0}^{\infty} \|U_{k+1} - U_k\|_F \leq \sum_{k=0}^{\infty} \mu_k^{-1} \vartheta_1 \leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} \vartheta_1 \leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^{4/3}} \vartheta_1 < \infty,$$

where the constant ϑ_1 is defined as

$$\vartheta_1 = \max \left\{ \left[\left(\rho \|Y_3^{k-1} - Y_3^k\|_F + \|Y_3^k\|_F \right) \|V_k\|_F + \rho \|Y_1^k - Y_1^{k-1}\|_F + \|Y_1^k\|_F \right] \|(V_k^T V_k + I_d)^{-1}\|_F, k = 1, 2, \dots \right\}.$$

In addition,

$$\begin{aligned}
& \sum_{k=0}^{\infty} \|V_{k+1} - V_k\|_F \\
& \leq \sum_{k=0}^{\infty} \frac{1}{\mu_k} \left[\left(\rho \|Y_3^{k-1} - Y_3^k\|_F + \|Y_3^k\|_F \right) \|U_{k+1}\|_F + \rho \|Y_2^k - Y_2^{k-1}\|_F + \|Y_2^k\|_F \right] \|(U_{k+1}^T U_{k+1} + I_d)^{-1}\|_F \\
& \quad + \sum_{k=0}^{\infty} \left(\|V_k\|_F \|U_{k+1}\|_F \|(U_{k+1}^T U_{k+1} + I_d)^{-1}\|_F \right) \|U_{k+1} - U_k\|_F \\
& \leq \sum_{k=0}^{\infty} \vartheta_2 \|U_{k+1} - U_k\|_F + \sum_{k=0}^{\infty} \frac{1}{\mu_k} \vartheta_3 \leq \sum_{k=0}^{\infty} \vartheta_2 \|U_{k+1} - U_k\|_F + \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} \vartheta_3 < \infty,
\end{aligned}$$

where the constants ϑ_2 and ϑ_3 are defined as

$$\begin{aligned}
\vartheta_2 &= \max \left\{ \|V_k\|_F \|U_{k+1}\|_F \|(U_{k+1}^T U_{k+1} + I_d)^{-1}\|_F, k = 1, 2, \dots \right\}, \\
\vartheta_3 &= \max \left\{ \left[\left(\rho \|Y_3^{k-1} - Y_3^k\|_F + \|Y_3^k\|_F \right) \|U_{k+1}\|_F + \rho \|Y_2^k - Y_2^{k-1}\|_F + \|Y_2^k\|_F \right] \|(U_{k+1}^T U_{k+1} + I_d)^{-1}\|_F, k = 1, 2, \dots \right\}.
\end{aligned}$$

Consequently, both $\{U_k\}$ and $\{V_k\}$ are convergent sequences. Moreover, it is not difficult to verify that $\{U_k\}$ and $\{V_k\}$ are both Cauchy sequences.

Similarly, $\{\hat{U}_k\}$, $\{\hat{V}_k\}$, $\{S_k\}$ and $\{L_k\}$ are also Cauchy sequences. Practically, the stopping criterion of Algorithm 1 is satisfied within a finite number of iterations.

(III) Let $(U_*, V_*, \hat{U}_*, \hat{V}_*, L_*, S_*)$ be a critical point of (17), and $\Phi(S) = \|\mathcal{P}_\Omega(S)\|_{\ell_{1/2}}^{1/2}$. Applying the Fermat's rule as in [7] to the subproblem (27), we then obtain

$$\begin{aligned}
0 &\in \frac{\lambda}{2} \partial \|\hat{U}_*\|_* + (Y_1^*) \text{ and } 0 \in \frac{\lambda}{2} \partial \|\hat{V}_*\|_* + (Y_2^*), \\
0 &\in \partial \Phi(S_*) + \mathcal{P}_\Omega(Y_4^*) \text{ and } \mathcal{P}_{\Omega^c}(Y_4^*) = \mathbf{0}, \\
L_* &= U_* V_*^T, \hat{U}_* = U_*, \hat{V}_* = V_*, L_* + S_* = D.
\end{aligned}$$

Applying the Fermat's rule to (27), we have

$$0 \in \partial\Phi(S_{k+1}) + \mathcal{P}_\Omega(Y_4^{k+1}), \quad \mathcal{P}_{\Omega^c}(Y_4^{k+1}) = 0. \quad (46)$$

In addition, the first-order optimal conditions for (23) and (24) are given by

$$0 \in \lambda\partial\|\widehat{U}_{k+1}\|_* + 2Y_1^{k+1}, \quad 0 \in \lambda\partial\|\widehat{V}_{k+1}\|_* + 2Y_2^{k+1}. \quad (47)$$

Since $\{U_k\}$, $\{V_k\}$, $\{\widehat{U}_k\}$, $\{\widehat{V}_k\}$, $\{L_k\}$ and $\{S_k\}$ are Cauchy sequences, then

$$\begin{aligned} \lim_{k \rightarrow \infty} \|U_{k+1} - U_k\| &= 0, \quad \lim_{k \rightarrow \infty} \|V_{k+1} - V_k\| = 0, \quad \lim_{k \rightarrow \infty} \|\widehat{U}_{k+1} - \widehat{U}_k\| = 0, \\ \lim_{k \rightarrow \infty} \|\widehat{V}_{k+1} - \widehat{V}_k\| &= 0, \quad \lim_{k \rightarrow \infty} \|L_{k+1} - L_k\| = 0, \quad \lim_{k \rightarrow \infty} \|S_{k+1} - S_k\| = 0. \end{aligned}$$

Let U_∞ , V_∞ , \widehat{U}_∞ , \widehat{V}_∞ , S_∞ and L_∞ be their limit points, respectively. Together with the results in **(I)**, then we have that $U_\infty = \widehat{U}_\infty$, $V_\infty = \widehat{V}_\infty$, $L_\infty = U_\infty V_\infty^T$ and $L_\infty + S_\infty = D$. Using (46) and (47), the following holds

$$\begin{aligned} 0 &\in \frac{\lambda}{2}\partial\|\widehat{U}_\infty\|_* + (Y_1)_\infty \quad \text{and} \quad 0 \in \frac{\lambda}{2}\partial\|\widehat{V}_\infty\|_* + (Y_2)_\infty, \\ 0 &\in \partial\Phi(S_\infty) + \mathcal{P}_\Omega((Y_4)_\infty) \quad \text{and} \quad \mathcal{P}_{\Omega^c}((Y_4)_\infty) = \mathbf{0}, \\ L_\infty &= U_\infty V_\infty^T, \quad \widehat{U}_\infty = U_\infty, \quad \widehat{V}_\infty = V_\infty, \quad L_\infty + S_\infty = D. \end{aligned}$$

Therefore, any accumulation point $\{(U_\infty, V_\infty, \widehat{U}_\infty, \widehat{V}_\infty, L_\infty, S_\infty)\}$ of the sequence $\{(U_k, V_k, \widehat{U}_k, \widehat{V}_k, L_k, S_k)\}$ generated by Algorithm 1 satisfies the KKT conditions for the problem (17). That is, the sequence asymptotically satisfies the KKT conditions of (17). In particular, whenever the sequence $\{(U_k, V_k, L_k, S_k)\}$ converges, it converges to a critical point of (15). This completes the proof. \square

APPENDIX F: STOPPING CRITERION

For the problem (15), the KKT conditions are

$$\begin{aligned} 0 &\in \frac{\lambda}{2}\partial\|U^*\|_* + Y^*V^* \quad \text{and} \quad 0 \in \frac{\lambda}{2}\partial\|V^*\|_* + (Y^*)^T U^*, \\ 0 &\in \partial\Phi(S^*) + \mathcal{P}_\Omega(\widehat{Y}^*) \quad \text{and} \quad \mathcal{P}_{\Omega^c}(\widehat{Y}^*) = \mathbf{0}, \\ L^* &= U^*(V^*)^T, \quad \mathcal{P}_\Omega(L^*) + \mathcal{P}_\Omega(S^*) = \mathcal{P}_\Omega(D). \end{aligned} \quad (48)$$

Using (48), we have

$$-Y^* \in \frac{\lambda}{2}\partial\|U^*\|_*(V^*)^\dagger \quad \text{and} \quad -Y^* \in [(U^*)^T]^\dagger(\frac{\lambda}{2}\partial\|V^*\|_*)^T \quad (49)$$

where $(V^*)^\dagger$ is the pseudo-inverse of V^* . The two conditions hold if and only if

$$\partial\|U^*\|_*(V^*)^\dagger \cap [(U^*)^T]^\dagger(\partial\|V^*\|_*)^T \neq \emptyset.$$

Recalling the equivalence relationship between (15) and (17), the KKT conditions for (17) are given by

$$\begin{aligned} 0 &\in \frac{\lambda}{2}\partial\|\widehat{U}^*\|_* + Y_1^* \quad \text{and} \quad 0 \in \frac{\lambda}{2}\partial\|\widehat{V}^*\|_* + Y_2^*, \\ 0 &\in \partial\Phi(S^*) + \mathcal{P}_\Omega(\widehat{Y}^*) \quad \text{and} \quad \mathcal{P}_{\Omega^c}(\widehat{Y}^*) = \mathbf{0}, \\ L^* &= U^*(V^*)^T, \quad \widehat{U}^* = U^*, \quad \widehat{V}^* = V^*, \quad L^* + S^* = D. \end{aligned} \quad (50)$$

Hence, we use the following conditions as the stopping criteria for Algorithm 1:

$$\max\{\epsilon_1/\|D\|_F, \epsilon_2\} < \epsilon$$

where $\epsilon_1 = \max\{\|U_k V_k^T - L_k\|_F, \|L_k + S_k - D\|_F, \|Y_1^k(\widehat{V}_k)^\dagger - (\widehat{U}_k^T)^\dagger(Y_2^k)^T\|_F\}$ and $\epsilon_2 = \max\{\|\widehat{U}_k - U_k\|_F/\|U_k\|_F, \|\widehat{V}_k - V_k\|_F/\|V_k\|_F\}$.

Algorithm 3 Solving image recovery problem (51) via ADMM**Input:** $\mathcal{P}_\Omega(D)$, the given rank d , and λ .**Initialize:** $\mu_0 = 10^{-4}$, $\mu_{\max} = 10^{20}$, $\rho > 1$, $k = 0$, and ϵ .1: **while** not converged **do**

2: $U_{k+1} = \left[(L_k + Y_3^k / \mu_k) V_k + \hat{U}_k - Y_1^k / \mu_k \right] (V_k^T V_k + I)^{-1}.$

3: $V_{k+1} = \left[(L_k + Y_3^k / \mu_k)^T U_{k+1} + \hat{V}_k - Y_2^k / \mu_k \right] (U_{k+1}^T U_{k+1} + I)^{-1}.$

4: $\hat{U}_{k+1} = \mathcal{D}_{\lambda/(2\mu_k)}(U_{k+1} + Y_1^k / \mu_k)$, and $\hat{V}_{k+1} = \mathcal{D}_{\lambda/(2\mu_k)}(V_{k+1} + Y_2^k / \mu_k).$

5: $L_{k+1} = \mathcal{P}_\Omega \left(\frac{D + \mu_k U_{k+1} V_{k+1}^T - Y_3^k}{1 + \mu_k} \right) + \mathcal{P}_\Omega^\perp (U_{k+1} V_{k+1}^T - Y_3^k / \mu_k).$

6: $Y_1^{k+1} = Y_1^k + \mu_k (U_{k+1} - \hat{U}_{k+1})$, $Y_2^{k+1} = Y_2^k + \mu_k (V_{k+1} - \hat{V}_{k+1})$, and $Y_3^{k+1} = Y_3^k + \mu_k (L_{k+1} - U_{k+1} V_{k+1}^T).$

7: $\mu_{k+1} = \min(\rho \mu_k, \mu_{\max}).$

8: $k \leftarrow k + 1.$

9: **end while****Output:** U_{k+1} and V_{k+1} .**APPENDIX G: ALGORITHMS FOR IMAGE RECOVERY**

In this part, we propose two efficient ADMM algorithms (as outlined in **Algorithms 3** and **4**) to solve the following D-N/F-N penalty regularized least squares problems for matrix completion:

$$\min_{U, V, L} \frac{\lambda}{2} (\|U\|_* + \|V\|_*) + \frac{1}{2} \|\mathcal{P}_\Omega(L) - \mathcal{P}_\Omega(D)\|_F^2, \text{ s.t., } L = UV^T, \quad (51)$$

$$\min_{U, V, L} \frac{\lambda}{3} (\|U\|_F^2 + 2\|V\|_*) + \frac{1}{2} \|\mathcal{P}_\Omega(L) - \mathcal{P}_\Omega(D)\|_F^2, \text{ s.t., } L = UV^T. \quad (52)$$

Similar to (17) and (18), we also introduce the matrices \hat{U} and \hat{V} as auxiliary variables to (51) (i.e., (35) in this paper), and obtain the following equivalent formulation,

$$\begin{aligned} \min_{U, V, \hat{U}, \hat{V}, L} \quad & \frac{\lambda}{2} (\|\hat{U}\|_* + \|\hat{V}\|_*) + \frac{1}{2} \|\mathcal{P}_\Omega(L) - \mathcal{P}_\Omega(D)\|_F^2, \\ \text{s.t., } \quad & L = UV^T, U = \hat{U}, V = \hat{V}. \end{aligned} \quad (53)$$

The augmented Lagrangian function of (53) is

$$\begin{aligned} \mathcal{L}_\mu = \quad & \frac{\lambda}{2} (\|\hat{U}\|_* + \|\hat{V}\|_*) + \frac{1}{2} \|\mathcal{P}_\Omega(L) - \mathcal{P}_\Omega(D)\|_F^2 + \langle Y_1, U - \hat{U} \rangle + \langle Y_2, V - \hat{V} \rangle + \langle Y_3, L - UV^T \rangle \\ & + \frac{\mu}{2} (\|U - \hat{U}\|_F^2 + \|V - \hat{V}\|_F^2 + \|L - UV^T\|_F^2) \end{aligned}$$

where Y_i ($i = 1, 2, 3$) are the matrices of Lagrange multipliers.

Updating U_{k+1} and V_{k+1} :

By fixing the other variables at their latest values, and removing the terms that do not depend on U and V and adding some proper terms that do not depend on U and V , the optimization problems with respect to U and V are formulated as follows:

$$\|U - \hat{U}_k + Y_1^k / \mu_k\|_F^2 + \|L_k - UV_k^T + Y_3^k / \mu_k\|_F^2, \quad (54)$$

$$\|V - \hat{V}_k + Y_2^k / \mu_k\|_F^2 + \|L_k - U_{k+1} V^T + Y_3^k / \mu_k\|_F^2. \quad (55)$$

Since both (54) and (55) are smooth convex optimization problems, their closed-form solutions are given by

$$U_{k+1} = \left[(L_k + Y_3^k / \mu_k) V_k + \hat{U}_k - Y_1^k / \mu_k \right] (V_k^T V_k + I)^{-1}, \quad (56)$$

$$V_{k+1} = \left[(L_k + Y_3^k / \mu_k)^T U_{k+1} + \hat{V}_k - Y_2^k / \mu_k \right] (U_{k+1}^T U_{k+1} + I)^{-1}. \quad (57)$$

Algorithm 4 Solving image recovery problem (52) via ADMM**Input:** $\mathcal{P}_\Omega(D)$, the given rank d , and λ .**Initialize:** $\mu_0 = 10^{-4}$, $\mu_{\max} = 10^{20}$, $\rho > 1$, $k = 0$, and ϵ .1: **while** not converged **do**

2: $U_{k+1} = [(\mu_k L_k + Y_2^k) V_k] (\mu_k V_k^T V_k + (2\lambda/3)I)^{-1}.$

3: $V_{k+1} = \left[(L_k + Y_2^k / \mu_k)^T U_{k+1} + \hat{V}_k - Y_1^k / \mu_k \right] (U_{k+1}^T U_{k+1} + I)^{-1}.$

4: $\hat{V}_{k+1} = \mathcal{D}_{2\lambda/(3\mu_k)}(V_{k+1} + Y_1^k / \mu_k).$

5: $L_{k+1} = \mathcal{P}_\Omega \left(\frac{D + \mu_k U_{k+1} V_{k+1}^T - Y_2^k}{1 + \mu_k} \right) + \mathcal{P}_\Omega^\perp (U_{k+1} V_{k+1}^T - Y_2^k / \mu_k).$

6: $Y_1^{k+1} = Y_1^k + \mu_k (V_{k+1} - \hat{V}_{k+1})$ and $Y_2^{k+1} = Y_2^k + \mu_k (L_{k+1} - U_{k+1} V_{k+1}^T).$

7: $\mu_{k+1} = \min(\rho \mu_k, \mu_{\max}).$

8: $k \leftarrow k + 1.$

9: **end while****Output:** U_{k+1} and V_{k+1} .**Updating \hat{U}_{k+1} and \hat{V}_{k+1} :**By keeping all other variables fixed, \hat{U}_{k+1} is updated by solving the following problem:

$$\frac{\lambda}{2} \|\hat{U}\|_* + \frac{\mu_k}{2} \|U_{k+1} - \hat{U} + Y_1^k / \mu_k\|_F^2. \quad (58)$$

To solve (58), the SVT operator [2] is considered as follows:

$$\hat{U}_{k+1} = \mathcal{D}_{\lambda/(2\mu_k)}(U_{k+1} + Y_1^k / \mu_k). \quad (59)$$

Similarly, \hat{V}_{k+1} is given by

$$\hat{V}_{k+1} = \mathcal{D}_{\lambda/(2\mu_k)}(V_{k+1} + Y_2^k / \mu_k). \quad (60)$$

Updating L_{k+1} :By fixing all other variables, the optimal L_{k+1} is the solution to the following problem:

$$\frac{1}{2} \|\mathcal{P}_\Omega(L) - \mathcal{P}_\Omega(D)\|_F^2 + \frac{\mu_k}{2} \|L - U_{k+1} V_{k+1}^T + \frac{Y_3^k}{\mu_k}\|_F^2. \quad (61)$$

Since (61) is a smooth convex optimization problem, it is easy to show that the optimal solution to (61) is

$$L_{k+1} = \mathcal{P}_\Omega \left(\frac{D + \mu_k U_{k+1} V_{k+1}^T - Y_3^k}{1 + \mu_k} \right) + \mathcal{P}_\Omega^\perp (U_{k+1} V_{k+1}^T - Y_3^k / \mu_k) \quad (62)$$

where \mathcal{P}_Ω^\perp is the complementary operator of \mathcal{P}_Ω , i.e., $\mathcal{P}_\Omega^\perp(D)_{ij} = 0$ if $(i, j) \in \Omega$, and $\mathcal{P}_\Omega^\perp(D)_{ij} = D_{ij}$ otherwise.Based on the description above, we develop an efficient ADMM algorithm for solving the double nuclear norm minimization problem (51), as outlined in **Algorithm 3**. Similarly, we also present an efficient ADMM algorithm to solve (52), as outlined in **Algorithm 4**.**APPENDIX H: MORE EXPERIMENTAL RESULTS**

In this paper, we compared both our methods with the state-of-the-art methods, such as LMaFit³ [8], RegL1⁴ [9], Unifying [10], factEN⁵ [11], RPCA⁶ [6], PSVT⁷ [12], WNNM⁸ [13], and LpSq⁹ [14]. The Matlab code of the proposed methods can be downloaded from the link¹⁰.

3. <http://lmafit.blogs.rice.edu/>4. <https://sites.google.com/site/yinqiangzheng/>5. <http://cpslab.snu.ac.kr/people/eunwoo-kim>6. <http://www.cis.pku.edu.cn/faculty/vision/zlin/zlin.htm>7. <http://thohkaistackr.wix.com/page>8. <http://www4.comp.polyu.edu.hk/~cslzhang/>9. <https://sites.google.com/site/feipingnie/>10. https://www.dropbox.com/s/9ah4oezv1b1x5jm/Code_DFMNM.zip?dl=0

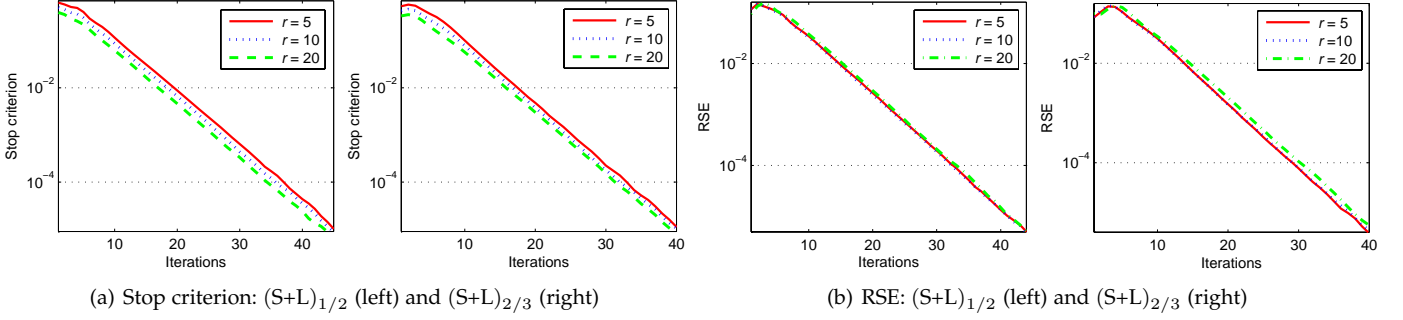


Fig. 1. Convergence behavior of our $(S+L)_{1/2}$ and $(S+L)_{2/3}$ methods for the cases of the matrix rank 5, 10 and 20.

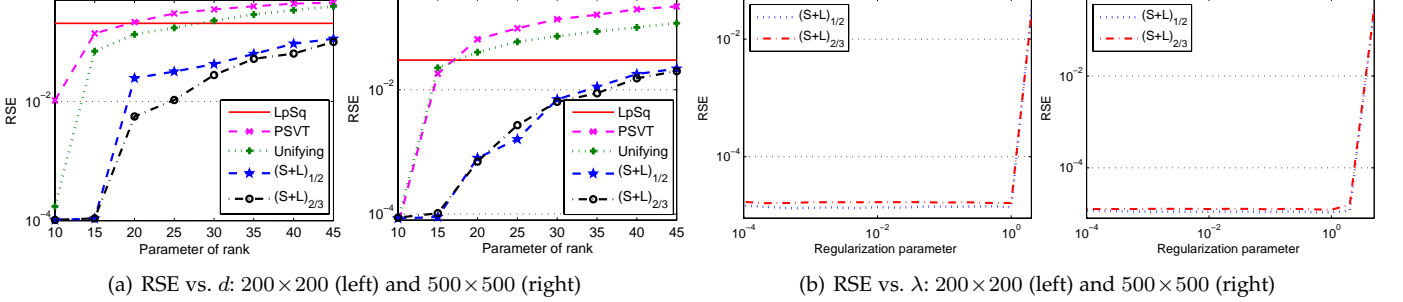


Fig. 2. Comparison of RSE results on corrupted matrices with different rank parameters (a) and regularization parameters (b).

Convergence Behavior

Fig. 1 illustrates the evolution of the relative squared error (RSE), i.e. $\|L - \bar{L}\|_F / \|\bar{L}\|_F$, and stop criterion over the iterations on corrupted matrices of size $1,000 \times 1,000$ with outlier ratio 5%, respectively. From the results, it is clear that both the stopping tolerance and RSE values of our two methods decrease fast, and they converge within only a small number of iterations, usually within 50 iterations.

Robustness

Like the other non-convex methods such as PSVT and Unifying, the most important parameter of our methods is the rank parameter d . To verify the robustness of our methods with respect to d , we report the RSE results of PSVT, Unifying and our methods on corrupted matrices with outlier ratio 10% in Fig. 2(a), in which we also present the results of the baseline method, LpSq [14]. It is clear that both our methods perform much more robust than PSVT and Unifying, and consistently yield much better solutions than the other methods in all settings.

To verify the robustness of both our methods with respect to another important parameter (i.e. the regularization parameter λ), we also report the RSE results of our methods on corrupted matrices with outlier ratio 10% in Fig. 2(b). Note that the rank parameter of both our methods is computed by our rank estimation procedure. From the results, one can see that both our methods demonstrate very robust performance over a wide range of the regularization parameter, e.g. from 10^{-4} to 10^0 .

Text Removal

We report the text removal results of our methods with varying rank parameters (from 10 to 40), as shown in Fig. 3, where the rank of the original image is 10. We also present the results of the baseline method, LpSq [14]. The results show that our methods significantly outperform the other methods in terms of RSE and F-measure, and they perform much more robust than Unifying with respect to the rank parameter.

Moving Object Detection

We present the detailed descriptions for five surveillance video sequences: Bootstrap, Hall, Lobby, Mall and WaterSurface data sets, as shown in Table 1. Moreover, Fig. 4 illustrates the foreground and background separation results on the Hall, Mall, Lobby and WaterSurface data sets.

TABLE 1
Information of the surveillance video sequences used in our experiments.

| Datasets | Bootstrap | Hall | Lobby | Mall | WaterSurface |
|-----------------------|--------------------------|--------------------------|--------------------------|-------------------------|-------------------------|
| Size \times #frames | $[120, 160] \times 1000$ | $[144, 176] \times 1000$ | $[128, 160] \times 1000$ | $[256, 320] \times 600$ | $[128, 160] \times 500$ |
| Description | Crowded scene | Crowded scene | Dynamic foreground | Crowded scene | Dynamic background |

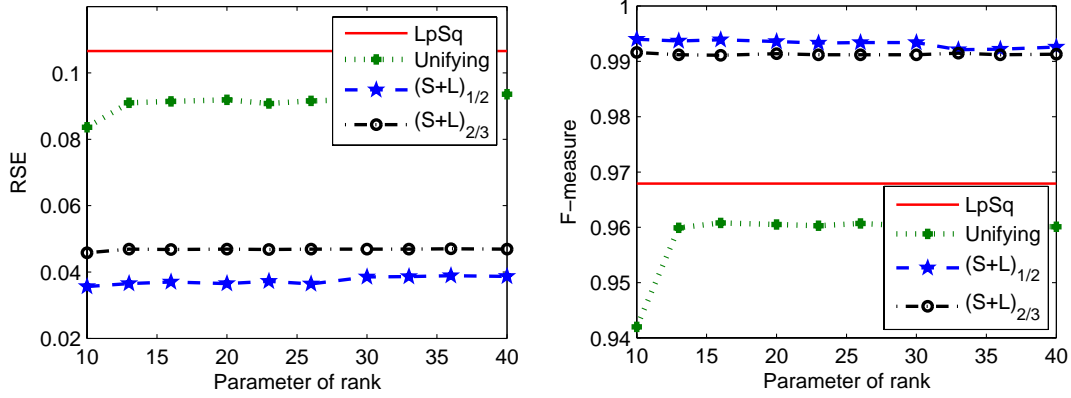


Fig. 3. The text removal results (including RSE and F-measure) of LpSq [14], Unifying [10] and our methods with different rank parameters.

Image Alignment

We also conducted some image alignment experiments on the Dummy images used in [15]. Figs. 5(b) and 5(c) show the results of alignment, low-rank and sparse estimations by both our methods, some of which are shown in Fig. 5(a). We can observe that both our methods are able to align the facial images nicely despite illumination variations and occlusions, and correctly detect and remove them.

Image Inpainting

In this part, we first reported the average PSNR results of two proposed methods (i.e., D-N and F-N) with different ratios of random missing pixels from 95% to 80%, as shown in Fig. 6. Since the methods in [16], [17] are very slow, we only present the average inpainting results of APGL¹¹ [18] and WNNM¹² [13], both of which use the fast SVD strategy and need to compute only partial SVD instead of the full one. Thus, APGL and WNNM are usually much faster than the methods [16], [17]. Considering that only a small fraction of pixels are randomly selected, thus we conducted 50 independent runs and report the average PSNR and standard deviation (std). The results show that both our methods consistently outperform APGL [18] and WNNM [13] in all the settings. This experiment actually shows that both our methods have even greater advantage over existing methods in the cases when the number of observed pixels is very limited, e.g., 5% observed pixels.

As suggested in [17], we set $d=9$ for our two methods and TNNR¹³ [17]. To evaluate the robustness of our two methods with respect to their rank parameter, we report the average PSNR and standard deviation of two proposed methods with varying rank parameter d from 7 to 15, as shown in Fig. 7. Moreover, we also present the average inpainting results of TNNR [17] over 50 independent runs. It is clear that two proposed methods perform much more robust than TNNR with respect to the rank parameter.

REFERENCES

- [1] L. Mirsky, "A trace inequality of John von Neumann," *Monatsh. Math.*, vol. 79, pp. 303–306, 1975.
- [2] J. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [3] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pur. Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [4] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [11] <http://www.math.nus.edu.sg/~mattohkc/NNLS.html>
- [12] <http://www4.comp.polyu.edu.hk/~cslzhang/>
- [13] <http://sites.google.com/site/zjuyaohu/>

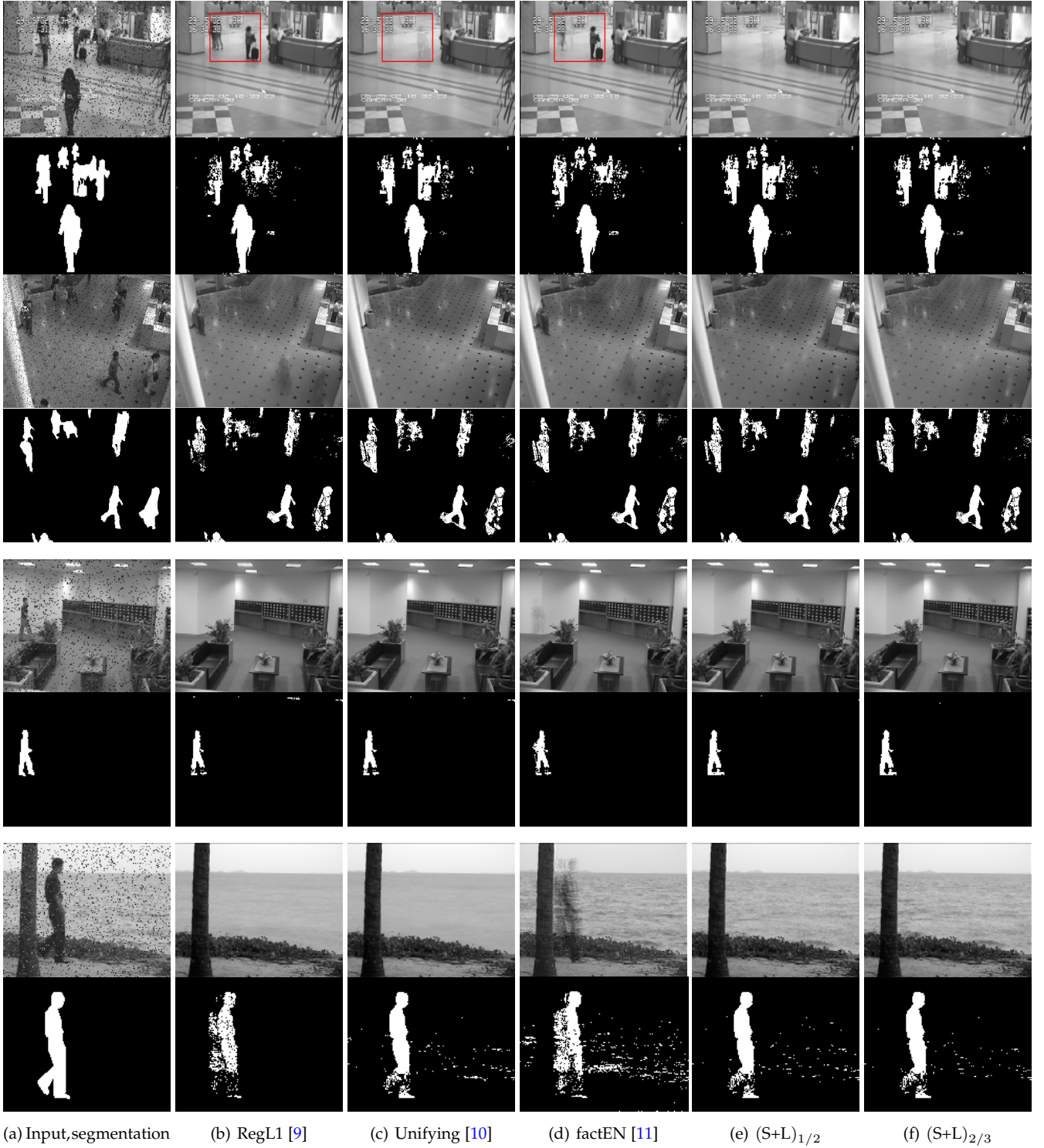


Fig. 4. Background and foreground separation results of different algorithms on the Hall, Mall, Lobby and WaterSurface data sets. The one frame with missing data of each sequence (top) and its manual segmentation (bottom) are shown in (a). The results by different algorithms are presented from (b) to (f), respectively. The top panel is the recovered background and the bottom panel is the segmentation.

- [5] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1107–1130, 2008.
- [6] Z. Lin, M. Chen, and L. Wu, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Univ. Illinois, Urbana-Champaign, Tech. Rep., Mar. 2009.
- [7] F. Wang, W. Cao, and Z. Xu, "Convergence of multi-block Bregman ADMM for nonconvex composite problems," *arXiv:1505.03063*, 2015.
- [8] Y. Shen, Z. Wen, and Y. Zhang, "Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization," *Optim. Method Softw.*, vol. 29, no. 2, pp. 239–263, 2014.
- [9] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi, "Practical low-rank matrix approximation under robust L_1 -norm," in *Proc. IEEE*



(a) Original occluded images

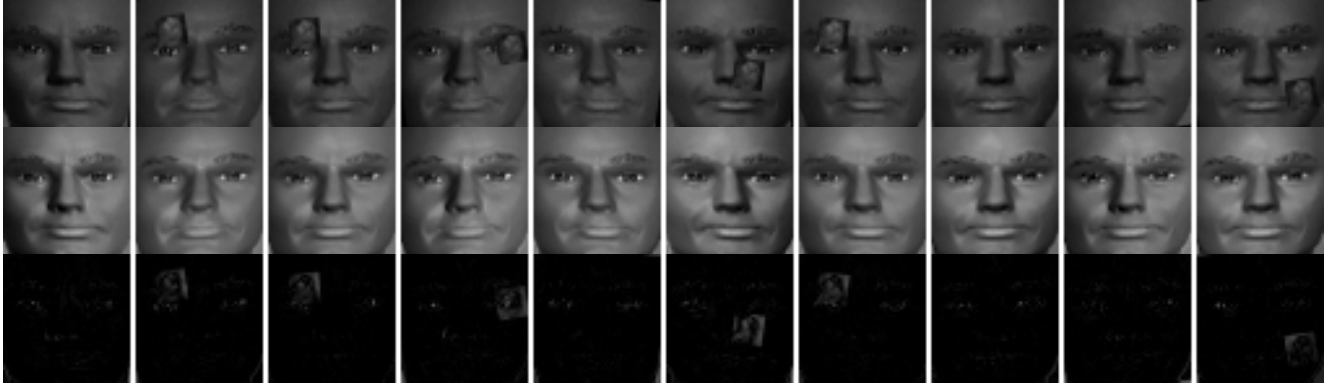
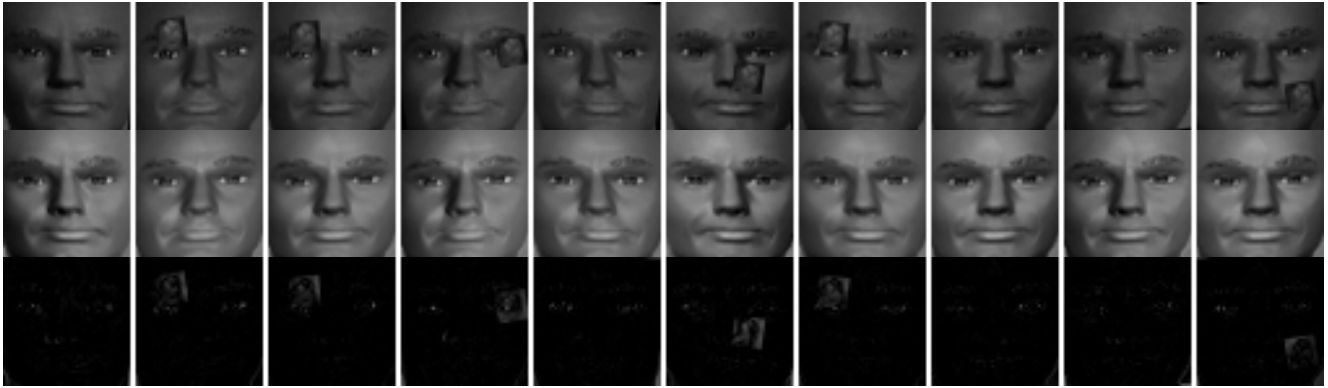
(b) Alignment results by our $(S+L)_{1/2}$ method(c) Alignment results by our $(S+L)_{2/3}$ method

Fig. 5. Aligning the Dummy images used in [15]. (a) Ten out of the 100 images. (b) and (c): From top to bottom show the aligned, low-rank and sparse results using our $(S+L)_{1/2}$ and $(S+L)_{2/3}$ methods, respectively.

Conf. Comput. Vis. Pattern Recognit., 2012, pp. 1410–1417.

- [10] R. Cabral, F. De la Torre, J. Costeira, and A. Bernardino, “Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2488–2495.
- [11] E. Kim, M. Lee, and S. Oh, “Elastic-net regularization of singular values for robust subspace learning,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 915–923.
- [12] T. Oh, Y. Tai, J. Bazin, H. Kim, and I. Kweon, “Partial sum minimization of singular values in robust PCA: Algorithm and applications,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 744–758, Apr. 2016.
- [13] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, “Weighted nuclear norm minimization and its applications to low level vision,” *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, 2017.
- [14] F. Nie, H. Wang, X. Cai, H. Huang, and C. Ding, “Robust matrix completion via joint Schatten p -norm and L_p -norm minimization,” in *Proc. IEEE Int. Conf. Data Mining*, 2012, pp. 566–574.
- [15] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2233–2246, Nov. 2012.
- [16] C. Lu, J. Tang, S. Yan, and Z. Lin, “Generalized nonconvex nonsmooth low-rank minimization,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 4130–4137.
- [17] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, “Fast and accurate matrix completion via truncated nuclear norm regularization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, Sep. 2013.
- [18] K.-C. Toh and S. Yun, “An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems,” *Pac. J. Optim.*, vol. 6, pp. 615–640, 2010.

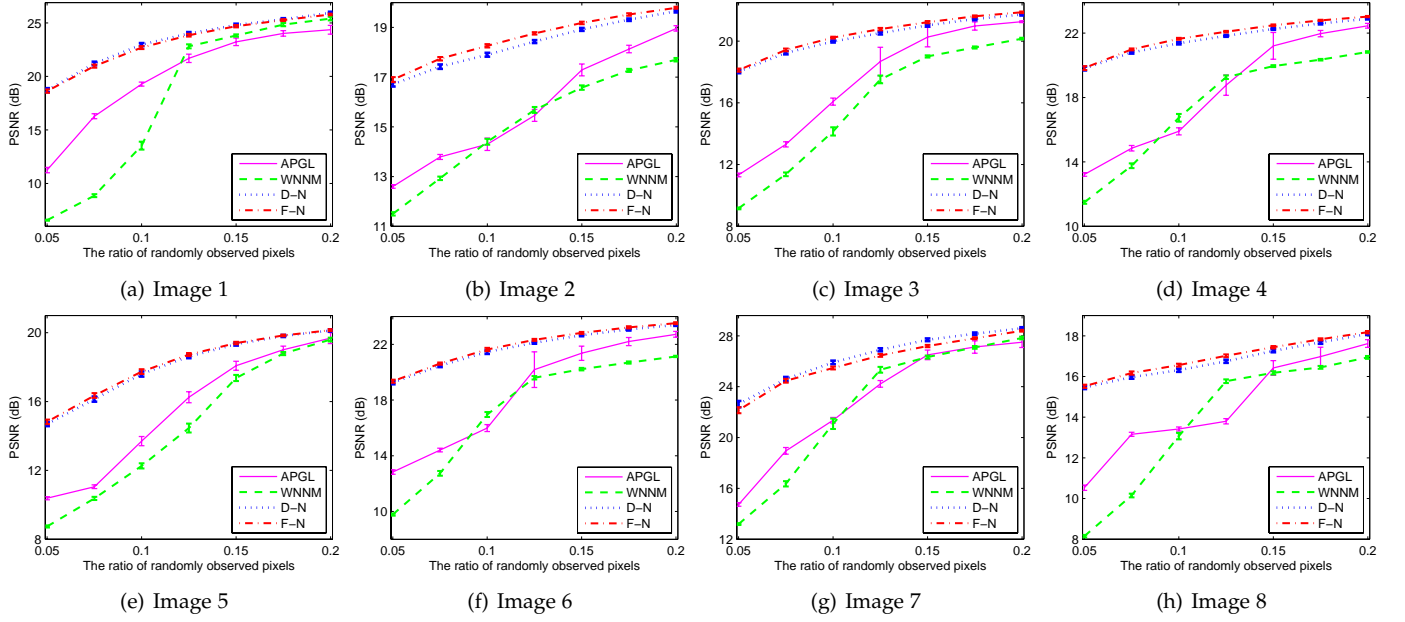


Fig. 6. The average PSNR and standard deviation of APGL [18], WNNM [13] and both our methods for image inpainting vs. fraction of observed pixels (best viewed in colors).

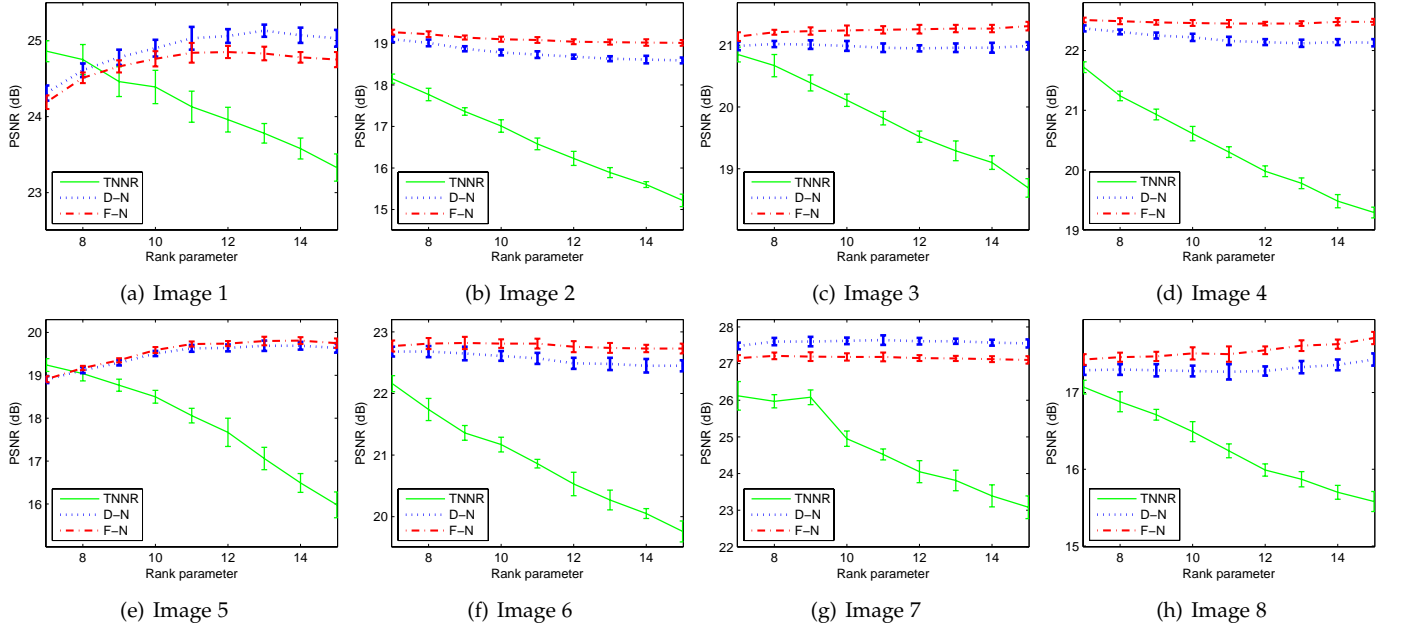


Fig. 7. The average PSNR and standard deviation of TNNR [17] and both our methods for image inpainting vs. the rank parameter (best viewed in colors).