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 $m \times n$ matrices with real entries is denoted by $\mathbb{R}^{m \times n}$. $\text{Tr}(X^TY) = \sum_{i,j} X_{ij}Y_{ij}$, 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 $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_r(X) > \sigma_{r+1}(X) = \cdots = \sigma_n(X) = 0$, where $r = \text{rank}(X)$. Then the SVD of $X$ is denoted by $X = U\Sigma V^T$, where $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$. $I_n$ denotes an identity matrix of size $n \times n$.

**Definition 5.** For any vector $x \in \mathbb{R}^l$, its $\ell_p$-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 $\ell_1$-norm of $x$ is $\|x\|_{\ell_1} = \sum_i |x_i|$ (which is convex), while the $\ell_p$-norm of $x$ is a quasi-norm when $0 < p < 1$, which is non-convex and violates the triangle inequality. In addition, the $\ell_2$-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: {fshang, jcheng, yyluu}@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: luozq@cuhk.edu.cn and luozq@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 MediNet 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
  \[ \|X\|_1 = \sum_{i=1}^{n} \sigma_i(X). \]
- When $p = 2$, the Schatten-2 norm is more commonly called the Frobenius norm 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
\[ 0 \leq x_1 \leq x_2 \leq \ldots \leq x_n, \ y_1 \geq y_2 \geq \ldots \geq y_n \geq 0, \] (36)
then
\[ \sum_{i,j=1}^{n} p_{ij}x_iy_j \geq \sum_{k=1}^{n} x_ky_k. \]

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 $\alpha_i$ and $\beta_j$ for all $1 \leq i, j \leq n$ such that
\[ x_k = \sum_{1 \leq i \leq k} \alpha_i, \ y_k = \sum_{k \leq j \leq n} \beta_j \text{ for all } k = 1, \ldots, n. \]

Let $\delta_{ij}$ denote the Kronecker delta (i.e., $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise), we have
\[
\sum_{k=1}^{n} x_ky_k - \sum_{i,j=1}^{n} p_{ij}x_iy_j = \sum_{1 \leq i,j \leq n} (\delta_{ij} - p_{ij})x_iy_j = \sum_{1 \leq i,j \leq n} \delta_{ij}x_iy_j - \sum_{i,j=1}^{n} p_{ij}x_iy_j = \sum_{1 \leq i,j \leq n} \alpha_i \beta_j = \sum_{1 \leq r,s \leq n} \alpha_r \beta_s.
\]

If $r \leq s$, by the lemma in [1] we know that
\[
\sum_{1 \leq i < r, 1 \leq j \leq s} (\delta_{ij} - p_{ij}) \geq 0, \quad \text{and} \quad \sum_{r \leq i \leq n, 1 \leq j \leq s} (\delta_{ij} - p_{ij}) + \sum_{1 \leq i < r, 1 \leq j \leq s} (\delta_{ij} - p_{ij}) = 0.
\]

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 $r \geq s$. \hfill \Box

**Proof of Lemma 2:**

**Proof:** Using the properties of the trace, we know that
\[ \text{Tr}(X^TY) = \sum_{i,j=1}^{n} (U_X)_{ij}^2 \lambda_i \tau_j. \]

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 $\text{Tr}(X^TY) \geq \sum_{i=1}^{n} \lambda_i \tau_i$. \hfill \Box

2. Note that the Frobenius norm is the induced norm of the $\ell_2$-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

$$
X = U^* (V^*)^T \quad \text{and} \quad \|X\|_{S_1/2} = \left( \|U^*\|_* + \|V^*\|_* \right)^2 / 4.
$$

Therefore, we have

$$
\min_{U,V:X=UV^T} \left( \frac{\|U\|_* + \|V\|_*}{2} \right)^2 = \|X\|_{S_1/2}.
$$

This completes the proof. \qed

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

$$
X = U^* (V^*)^T \quad \text{and} \quad \|X\|_{S_2/3} = \left( \|U^*\|_F^2 + 2\|V^*\|_*/3 \right)^{3/2}.
$$

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. \qed

APPENDIX C: PROOF OF PROPERTY 4

Proof: The proof of Property 4 involves some properties of the $\ell_p$-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:

$$
\|x\|_{\ell_1} \leq \|x\|_{\ell_{p_1}}, \quad \|x\|_{\ell_{p_2}} \leq \|x\|_{\ell_2} \leq \left( \frac{p_2}{p_1} \right)^{1/2} \|x\|_{\ell_{p_1}}.
$$

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 $\ell_p$-norm mentioned above, we have

$$
\|X\|_* = \|\text{diag}(\Sigma_{r \times r})\|_{\ell_1} \leq \|\text{diag}(\Sigma_{r \times r})\|_{\ell_{p_1}} = \|X\|_{\ell_{p_1}} \leq \|X\|_{D-N} \leq \text{rank}(X)\|X\|_*,
$$

$$
\|X\|_* = \|\text{diag}(\Sigma_{r \times r})\|_{\ell_2} \leq \|\text{diag}(\Sigma_{r \times r})\|_{\ell_{p_2}} = \|X\|_{\ell_{p_2}} \leq \|X\|_{F-N} \leq \sqrt{\text{rank}(X)}\|X\|_*.
$$

In addition,

$$
\|X\|_{F-N} = \|\text{diag}(\Sigma_{r \times r})\|_{\ell_{p_2}} \leq \|\text{diag}(\Sigma_{r \times r})\|_{\ell_{p_1}} = \|X\|_{D-N}.
$$

Therefore, we have

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

This completes the proof. \qed

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

$$
\mathcal{L}_\mu(U,V,L,S,\hat{V},Y_1,Y_2,Y_3) = \frac{\lambda}{3} \left( \|U\|_F^2 + 2\|\hat{V}\|_* \right) + \|P_{\Omega}(S)\|_{S_2/3}^2 + \langle Y_1, \hat{V} - V \rangle + \langle Y_2, UV^T - L \rangle + \langle Y_3, L + S - D \rangle + \frac{\mu}{2} \left( \|UV^T - L\|_F^2 + \|\hat{V} - V\|_F^2 + \|L + S - D\|_F^2 \right)
$$

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 \( \lambda \).

Initialize: \( \mu_0, \rho > 1, k = 0, \) and \( \epsilon \).

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 \( \hat{V}_{k+1} = D_{2\lambda/3\mu_k} (V_{k+1} - (Y_2)/\mu_k) \).
        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
    8:         \( Y_{1}^{k+1} = Y_{1}^{k} + \mu_k (V_{k+1} - V_{k}) \), \( Y_{2}^{k+1} = Y_{2}^{k} + \mu_k (U_{k+1} V_{k}^T - L_{k+1}) \), and \( Y_{3}^{k+1} = Y_{3}^{k} + \mu_k (L_{k+1} + S_{k+1} - D) \).
    9:         Update \( \mu_{k+1} \) by \( \mu_{k+1} = \rho \mu_k \).
    10:     \( k \leftarrow k + 1 \).

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:

\[
\min_U \frac{\lambda}{3} \|U\|_F^2 + \frac{\mu_k}{2} \|UV_k^T - L_k + \mu_k^{-1} Y_{2}^k\|_F^2, \\
\min_V \|\hat{V}_k - V + \mu_k^{-1} Y_{1}^k\|_F^2 + \|U_{k+1} V^T - L_k + \mu_k^{-1} Y_{2}^k\|_F^2, 
\]

and their optimal solutions can be given by

\[
U_{k+1} = \mu_k P_k V_k \left( \frac{2\lambda}{3} I_d + \mu_k V_k^T V_k \right)^{-1}, \\
V_{k+1} = \left[ \hat{V}_k + \mu_k^{-1} Y_{1}^k + P_k^T U_{k+1} \right] \left( I_d + U_{k+1}^T U_{k+1} \right)^{-1},
\]

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

\[
\min_{\hat{V}} \frac{2\lambda}{3} \|\hat{V}\|_F + \frac{\mu_k}{2} \|\hat{V} - V_{k+1} + Y_{2}^k/\mu_k\|_F^2. 
\]

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 \( D_\tau \) is defined as [2]:

\[
D_\tau(Y) = U_Y S_\tau(\Sigma_Y) V_Y^T,
\]

where \( S_\tau(x) = \max(\|x\|_F - \tau, 0) \cdot \text{sgn}(x) \) is the soft shrinkage operator [3], [4], [5].

**Update of** \( L_{k+1} \): 
For updating \( L_{k+1} \), we consider the following optimization problem:

\[
\min_L \|U_{k+1} V_{k+1}^T - L + \mu_k^{-1} Y_{2}^k\|_F^2 + \|L + S_k - D + \mu_k^{-1} Y_{3}^k\|_F^2.
\]

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

\[
L_{k+1} = \frac{1}{2} \left( U_{k+1} V_{k+1}^T + \mu_k^{-1} Y_{2}^k - S_k + D - \mu_k^{-1} Y_{3}^k \right).
\]

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, $\|\cdot\|_\omega$, i.e., the largest singular value.

**Lemma 9** (Boundedness). Let $Y_k^{k+1} = Y_k^4 + \mu_k(\tilde{U}_k + U_{k+1})$, $Y_k^{k+1} = Y_k^2 + \mu_k(\tilde{V}_k - V_{k+1})$, $Y_k^{k+1} = Y_k^3 + \mu_k(U_k + V_k - L_{k+1})$ and $Y_k^{k+1} = Y_k^4 + \mu_k(L_{k+1} - S_{k+1})$. Suppose that $\mu_k$ is non-decreasing and $\sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k} < \infty$, then the sequences $\{(U_k, V_k)\}$, $\{(\tilde{U}_k, \tilde{V}_k)\}$, $\{(Y_k^1, Y_k^2, Y_k^3, Y_k^4)/\sqrt{\mu_k-1}\}$, $\{L_k\}$ and $\{S_k\}$ produced by Algorithm 1 are all bounded.

Proof: Let $\mathcal{Y}_k := (U_k, V_k)$, $\mathcal{Y}_k := (\tilde{U}_k, \tilde{V}_k)$ and $\mathcal{B}_k := (Y_k^1, Y_k^2, Y_k^3, Y_k^4)$. By the first-order optimality conditions of the augmented Lagrangian function of (17) with respect to $\tilde{U}$ and $\tilde{V}$ (i.e., Problems (23) and (24)), we have

$$0 \in \partial_{\mathcal{Y}_k} L_{\mu_k}(\mathcal{Y}_k, \mathcal{B}_k),$$

where $\mathcal{Y}_k$ denotes the gradient of the penalty $\Phi(S)$ at $S_k$. Since $\partial_{\mathcal{Y}_k} = \text{sign}(s_{ij})/(2\sqrt{\|s_{ij}\|})$, then

$$\left| \frac{Y_k^4}{\sqrt{\mu_k}} \right| \leq \frac{1}{2\sqrt{\|S_k\|_{ij}}}.$$

Using Proposition 1, we have

$$\frac{\|Y_k^{k+1}\|_F}{\sqrt{\mu_k}} \leq \frac{3}{2\sqrt{\mu_k}}.$$

where $\gamma = 2/\mu_k$.

**Case 1:** $a_{ij} > \frac{1}{2\sqrt{\mu_k}}$, \text{ (i, j) } \in \Omega.

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

$$\left| \frac{Y_k^{k+1}}{\sqrt{\mu_k}} \right| \leq \left| \frac{\sum_i a_{ij} (1 + \cos(\frac{2\pi - 2\tan^{-1}(a_{ij})}{3}))}{\sqrt{\mu_k}} \right| \leq \frac{1}{\sqrt{2}},$$

where $\gamma = 2/\mu_k$.

**Case 2:** $a_{ij} \leq \frac{1}{2\sqrt{\mu_k}}$, \text{ (i, j) } \in \Omega.

If $a_{ij} \leq \frac{3}{2\sqrt{\mu_k}}$, and using Proposition 1, we have $\|S_k\|_{ij} = 0$. Since

$$\left| \frac{Y_k^{k+1}}{\sqrt{\mu_k}} \right| \leq \frac{3}{2} \text{ and } \left| \frac{Y_k^{k+1}}{\sqrt{\mu_k}} \right| \leq \frac{3}{2\sqrt{\mu_k}}.$$

Then

$$\left| \frac{Y_k^{k+1}}{\sqrt{\mu_k}} \right| \leq \frac{3}{2} \text{ and } \left| \frac{Y_k^{k+1}}{\sqrt{\mu_k}} \right| \leq \frac{3}{2\sqrt{\mu_k}}.$$

Therefore, $\{Y_k^{k+1}/\sqrt{\mu_k} \}$ is bounded.

By the iterative scheme of Algorithm 1, we have

$$L_{\mu_k}(\mathcal{Y}_{k+1}, \mathcal{Y}_{k+1}, L_{k+1}, S_{k+1}, \mathcal{B}_k) \leq L_{\mu_k}(\mathcal{Y}_{k+1}, \mathcal{Y}_{k+1}, L_k, S_k, \mathcal{B}_k) \leq L_{\mu_k}(\mathcal{Y}_{k}, L_k, S_k, \mathcal{B}_k) = L_{\mu_k}(\mathcal{Y}_{k}, L_k, S_k, \mathcal{B}_k) + \alpha_k \sum_{j=1}^{3} \left| \frac{Y_j^k - Y_j^{k-1}}{\sqrt{\mu_k}} \right|.$$
where $\alpha_k = \frac{\mu_k+1}{2(\mu_k+1)^{2}}$, $\beta_k = \frac{\mu_{k+1}}{2(\mu_{k+1})^{3/2}}$, and the above equality holds due to the definition of the augmented Lagrangian function $L_{\mu}(U, V, L, S, \hat{U}, \hat{V}, \{Y_i\})$. Since $\mu_k$ is non-decreasing, and $\sum_{k=1}^{\infty} (\mu_k/\mu_{k-1}) < \infty$, then

$$
\sum_{k=1}^{\infty} \beta_k = \sum_{k=1}^{\infty} \frac{\mu_k-1+\mu_k}{2\mu_k^{1/3}} \leq \sum_{k=1}^{\infty} \frac{\mu_k}{\mu_{k-1}^{4/3}} < \infty,
$$

$$
\sum_{k=1}^{\infty} \alpha_k = \sum_{k=1}^{\infty} \frac{\mu_k}{2\mu_k^{2}} \leq \sum_{k=1}^{\infty} \frac{\mu_k}{\mu_{k-1}^{4/3}} < \infty.
$$

Since $\{\parallel Y_k^{1} F / \sqrt{\mu_{k-1}} \parallel \}$ is bounded, and $\mu_k = \rho \mu_{k-1}$ and $\rho > 1$, then $\{\parallel Y_k^{1} F / \sqrt{\mu_{k-1}} \parallel \}$ is also bounded, which implies that $\{\parallel Y_k - Y_{k-1} \parallel / \sqrt{\mu_{k-1}} \parallel \}$ is bounded. Then $\{L_{\mu_k}(\mathcal{Y}_{k+1}, \mathcal{Y}_{k+1}, L_{k+1}, S_{k+1}, \mathcal{Y}_k)\}$ is upper-bounded due to the boundedness of the sequences of all Lagrange multipliers, i.e., $\{Y_k\}$, $\{Y_k^1\}$, $\{Y_k^2\}$ and $\{Y_k^3\}$.

$$
\frac{\lambda}{2}(\parallel \hat{U}_{k+1} \parallel_{*} + \parallel \hat{V}_{k+1} \parallel_{*}) + \parallel \mathcal{P}_{\Omega}(S_k) \parallel_{1,2}^{1/2} = L_{\mu_k}(\mathcal{U}_k, \mathcal{Y}_k, \mathcal{L}_k, S_k, \mathcal{Y}_{k-1}) - \frac{4}{2} \sum_{i=1}^{\infty} \frac{\parallel Y_k^i \parallel_{F}^{2} - \parallel Y_{k-1}^{i} \parallel_{F}^{2}}{\mu_k} \parallel \hat{U}_{k+1} \parallel_{*}
$$

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

Similarly, by $U_k = \hat{U}_k - [Y_k - Y_{k-1}] / \mu_{k-1}$, $V_k = \hat{V}_k - [Y_k - Y_{k-1}] / \mu_{k-1}$, $L_k = U_k V_k^T - [Y_k - Y_{k-1}] / \mu_{k-1}$ and the boundedness of $\{\hat{U}_k\}$, $\{\hat{V}_k\}$, $\{Y_k\}$ ($i = 1, 2, 3$), and $\{Y_k^3 / \sqrt{\mu_{k-1}}\}$, 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.

**Proof of Theorem 3:**

**Proof:** (I) $\hat{U}_{k+1} - U_{k+1} = [Y_{k+1} - Y_k] / \mu_k$, $\hat{V}_{k+1} - V_{k+1} = [Y_{k+1} - Y_k] / \mu_k$, $U_k V_k^T - L_{k+1} = [Y_{k+1} - Y_k] / \mu_k$ and $L_{k+1} + S_{k+1} - D = \left[ Y_{k+1} - Y_k \right] / \mu_k$. Due to the boundedness of $\{Y_k\}$, $\{Y_k^1\}$, $\{Y_k^2\}$ and $\{Y_k^3 / \sqrt{\mu_{k-1}}\}$, the non-decreasing property of $\{\mu_k\}$, and $\sum_{k=0}^{\infty} (\mu_k / \mu_k^{4/3}) < \infty$, we have

$$
\sum_{k=0}^{\infty} \parallel \hat{U}_{k+1} - U_{k+1} \parallel_{F} \leq \sum_{k=0}^{\infty} \frac{\mu_k}{\mu_k} \parallel Y_{k+1} - Y_k \parallel_{F} < \infty,
$$

$$
\sum_{k=0}^{\infty} \parallel L_{k+1} - U_k V_k^T \parallel_{F} \leq \sum_{k=0}^{\infty} \frac{\mu_k}{\mu_k} \parallel Y_{k+1} - Y_k \parallel_{F} < \infty,
$$

which implies that

$$
\lim_{k \to \infty} \parallel \hat{U}_{k+1} - U_{k+1} \parallel_{F} = 0, \quad \lim_{k \to \infty} \parallel \hat{V}_{k+1} - V_{k+1} \parallel_{F} = 0, \quad \lim_{k \to \infty} \parallel L_{k+1} - U_k V_k^T \parallel_{F} = 0,
$$

and

$$
\lim_{k \to \infty} \parallel L_{k+1} + S_{k+1} - D \parallel_{F} = 0.
$$

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 $Y_k = Y_{k-1} + \mu_{k-1} (\hat{U}_k - U_k)$, $Y_k^1 = Y_{k-1} + \mu_{k-1} (\hat{V}_k - V_k)$ and $Y_k^2 = Y_{k-1} + \mu_{k-1} (U_k V_k^T - L_k)$, then the first-order optimality conditions of (19) and (20) with respect to $U$ and $V$ are written as follows:

$$
(U_{k+1} V_{k+1}^T - L_k + Y_k^1 / \mu_k) V_k + \left( U_{k+1} - \hat{U}_k - Y_k^2 / \mu_k \right)
$$

$$
= (U_{k+1} V_{k+1}^T - U_k V_k^T - Y_k^1 / \mu_k + Y_k^1 / \mu_k) V_k + U_{k+1} - U_k + U_k - \hat{U}_k - Y_k^2 / \mu_k
$$

$$
= (U_{k+1} - U_k) (V_{k+1} V_k + J_d) + \left( Y_k^1 / \mu_k - Y_k^2 / \mu_k \right) V_k + Y_k^1 / \mu_k - Y_k^2 / \mu_k
$$

$$
= 0,
$$

(44)
\[
\begin{align*}
(V_{k+1}U_{k+1}^T - L_k^T + \frac{(Y_3^k)^T}{\mu_k}) U_{k+1} &+ \left(V_{k+1} - \tilde{V}_k - \frac{Y_2^k}{\mu_k}\right) \\
&= (V_{k+1}U_{k+1}^T - V_k U_k^T) \frac{(Y_3^{k-1})^T}{\mu_k} + \frac{(Y_3^k)^T}{\mu_k} U_{k+1} + V_k + V_k - \tilde{V}_k - \frac{Y_2^k}{\mu_k} \\
&= 0.
\end{align*}
\]

By (44) and (45), we obtain
\[
\begin{align*}
U_{k+1} - U_k &= \left[\frac{(Y_3^{k-1} - Y_3^k)}{\mu_k} - \frac{Y_3^k}{\mu_k}\right] V_k + \frac{Y_2^k}{\mu_k} \left(V_k^T V_k + I_d\right)^{-1}, \\
V_{k+1} - V_k &= \left[V_k (U_k^T - U_{k+1}^T) U_{k+1} + \left(\frac{(Y_3^{k-1})^T - (Y_3^k)^T}{\mu_k} - \frac{(Y_3^k)^T}{\mu_k}\right) U_{k+1} + \frac{Y_2^k}{\mu_k} \left(V_k^T V_k + I_d\right)^{-1}\right].
\end{align*}
\]

Recall that
\[
\sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^{4/3}} < \infty,
\]
we have
\[
\sum_{k=0}^{\infty} \|U_{k+1} - U_k\|_F \leq \sum_{k=0}^{\infty} \frac{1}{\mu_k} \vartheta_1 \leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^{2}} \vartheta_1 < \infty,
\]
where the constant \(\vartheta_1\) is defined as
\[
\vartheta_1 = \max \left\{ \left(\rho \|Y_3^{k-1} - Y_3^k\|_F + \|Y_3^k\|_F\right) \|V_k\|_F + \rho \|Y_1^k - Y_1^{k-1}\|_F + \|Y_1^k\|_F \right\} \left(||V_k^T V_k + I_d||_F^{-1}\right)_F, k = 1, 2, \cdots \}.
\]

In addition,
\[
\begin{align*}
&\sum_{k=0}^{\infty} \|V_{k+1} - V_k\|_F \\
&\leq \sum_{k=0}^{\infty} \frac{1}{\mu_k} \left[\left(\rho \|Y_3^{k-1} - Y_3^k\|_F + \|Y_3^k\|_F\right) \|U_{k+1}\|_F + \rho \|Y_2^k - Y_2^{k-1}\|_F + \|Y_2^k\|_F \right] \left(||U_{k+1}^T V_{k+1} + I_d||_F^{-1}\right)_F \\
&\quad + \sum_{k=0}^{\infty} \left(\|V_k\|_F \|U_{k+1}\|_F \left(||U_{k+1}^T V_{k+1} + I_d||_F^{-1}\right)_F\right) \|U_{k+1} - U_k\|_F \\
&\leq \sum_{k=0}^{\infty} \vartheta_2 \|U_{k+1} - U_k\|_F + \sum_{k=0}^{\infty} \frac{1}{\mu_k} \vartheta_3 \leq \sum_{k=0}^{\infty} \vartheta_2 \|U_{k+1} - U_k\|_F + \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^{2}} \vartheta_3 < \infty,
\end{align*}
\]
where the constants \(\vartheta_2\) and \(\vartheta_3\) are defined as
\[
\begin{align*}
\vartheta_2 &= \max \left\{ \|\|V_k\|_F \|U_{k+1}\|_F \left(||U_{k+1}^T V_{k+1} + I_d||_F^{-1}\right)_F, k = 1, 2, \cdots \} \right\}, \\
\vartheta_3 &= \max \left\{ \left(\rho \|Y_3^{k-1} - Y_3^k\|_F + \|Y_3^k\|_F\right) \|U_{k+1}\|_F + \rho \|Y_2^k - Y_2^{k-1}\|_F + \|Y_2^k\|_F \right\} \left(||U_{k+1}^T V_{k+1} + I_d||_F^{-1}\right)_F, k = 1, 2, \cdots \}.
\end{align*}
\]

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, \(\{\tilde{U}_k\}, \{\tilde{V}_k\}, \{S_k\}\) and \(\{L_k\}\) are also Cauchy sequences. Practically, it is not difficult to verify that \(\{U_k\}\) and \(\{V_k\}\) are both Cauchy sequences.

(II) Let \((U_*, V_*, \tilde{U}_*, \tilde{V}_*, L_*, S_*)\) be a critical point of (17), and \(\Phi(S) = \|P_{11}(S)\|_{1/2}\). Applying the Fermat’s rule as in [7] to the subproblem (27), we then obtain
\[
0 \in \lambda \frac{1}{2} \partial ||\tilde{U}_*||_F + (Y_1^*)^* \quad \text{and} \quad 0 \in \lambda \frac{1}{2} \partial ||\tilde{V}_*||_F + (Y_2^*)^*,
\]
\[
0 \in \partial \Phi(S_*) + P_{11}(Y_1^*) \quad \text{and} \quad P_{11}(Y_1^*) = 0,
\]
\[
L_* = U_* V_*^T, \quad \tilde{U}_* = U_*, \quad \tilde{V}_* = V, \quad L_* + S_* = D.
\]
Applying the Fermat’s rule to (27), we have

\[ 0 \in \partial \Phi(S_{k+1}) + P_{\Omega}(Y_{k+1}), \quad P_{\Omega^c}(Y_{k+1}) = 0. \]  

(46)

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

\[ 0 \in \lambda \partial \| \tilde{U}_{k+1} \|_* + 2Y_{1}^{k+1}, \quad 0 \in \lambda \partial \| \tilde{V}_{k+1} \|_* + 2Y_{2}^{k+1}. \]  

(47)

Since \( \{U_k\}, \{V_k\}, \{\tilde{U}_k\}, \{L_k\} \) and \( \{S_k\} \) are Cauchy sequences, then

\[
\begin{align*}
\lim_{k \to \infty} \|U_{k+1} - U_k\| &= 0, \quad \lim_{k \to \infty} \|V_{k+1} - V_k\| = 0, \quad \lim_{k \to \infty} \|\tilde{U}_{k+1} - \tilde{U}_k\| = 0, \\
\lim_{k \to \infty} \|\tilde{V}_{k+1} - \tilde{V}_k\| &= 0, \quad \lim_{k \to \infty} \|L_{k+1} - L_k\| = 0, \quad \lim_{k \to \infty} \|S_{k+1} - S_k\| = 0.
\end{align*}
\]

Let \( U_\infty, V_\infty, \tilde{U}_\infty, \tilde{V}_\infty, S_\infty \) and \( L_\infty \) be their limit points, respectively. Together with the results in (0), then we have that \( U_\infty = \tilde{U}_\infty, V_\infty = \tilde{V}_\infty, L_\infty = U_\infty V_\infty^T \) and \( L_\infty + S_\infty = D \). Using (46) and (47), the following holds

\[
\begin{align*}
0 &\in \frac{\lambda}{2} \partial \|\tilde{U}_\infty\|_* + (Y_1)_\infty \quad \text{and} \quad 0 \in \frac{\lambda}{2} \partial \|\tilde{V}_\infty\|_* + (Y_2)_\infty, \\
0 &\in \partial \Phi(S_\infty) + P_{\Omega^c}(Y_1)_\infty \quad \text{and} \quad P_{\Omega^c}(Y_2)_\infty = 0, \\
L_\infty &\in (U_\infty V_\infty^T)^T \tilde{U}_\infty = U_\infty^T \tilde{V}_\infty = V_\infty, \quad L_\infty + S_\infty = D.
\end{align*}
\]

Therefore, any accumulation point \( \{U_\infty, V_\infty, \tilde{U}_\infty, \tilde{V}_\infty, L_\infty, S_\infty\} \) of the sequence \( \{(U_k, V_k, \tilde{U}_k, \tilde{V}_k, L_k, S_k)\} \) 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

\[ 0 \in \frac{\lambda}{2} \partial \|U_*\|_* + Y_* V_* \quad \text{and} \quad 0 \in \frac{\lambda}{2} \partial \|V_*\|_* + (Y_*)^T U_*, \]  

(48)

\[ 0 \in \partial \Phi(S_*) + P_{\Omega^c}(Y_*) \quad \text{and} \quad P_{\Omega^c}(Y_*) = 0, \]  

\[ L_* = U^*(V_*)^T, \quad P_{\Omega^c}(L_*) + P_{\Omega^c}(S_*) = P_{\Omega^c}(D). \]

Using (48), we have

\[ -Y_* \in \frac{\lambda}{2} \partial \|U_*\|_* (V^*)^T \quad \text{and} \quad -Y_* \in \{(U_*)^T\}^T \left( \frac{\lambda}{2} \partial \|V_*\|_* \right)^T \]  

(49)

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

\[ \partial \|U_*\|_* (V^*)^T \cap \{(U_*)^T\}^T \left( \partial \|V_*\|_* \right)^T \neq \emptyset. \]

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

\[
\begin{align*}
0 &\in \frac{\lambda}{2} \partial \|\tilde{U}_*\|_* + Y_1^* \quad \text{and} \quad 0 \in \frac{\lambda}{2} \partial \|\tilde{V}_*\|_* + Y_2^*, \\
0 &\in \partial \Phi(S_*) + P_{\Omega^c}(Y_*) \quad \text{and} \quad P_{\Omega^c}(Y_*) = 0, \\
L_* &\in U^*(V_*^*)^T, \tilde{U}_* = U^*, \tilde{V}_* = V^*, \quad L_* + S_* = D.
\end{align*}
\]

(50)

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

\[
\max \{\epsilon_1 / \|D\|_r, \epsilon_2\} < \epsilon
\]

where \( \epsilon_1 = \max\{\|U_k V_k^T - L_k\|_F, \|L_k + S_k - D\|_F, \|V_k^T (\tilde{U}_k)^T - (\tilde{V}_k)^T (\tilde{Y}_k)\|_F\} \) and \( \epsilon_2 = \max\{\|\tilde{U}_k - U_k\|_F / \|U_k\|_F, \|\tilde{V}_k - V_k\|_F / \|V_k\|_F\} \).
Algorithm 3 Solving image recovery problem (51) via ADMM

Input: \( \mathcal{P}_{\Omega}(D) \), the given rank \( d \), and \( \lambda \).

Initialize: \( \mu_0 = 10^{-4} \), \( \mu_{\text{max}} = 10^{20} \), \( \rho > 1 \), \( k = 0 \), and \( \epsilon \).

1: while not converged do
2: \( U_{k+1} = (L_k + Y_1^k/\mu_k)U_k + \mathcal{P}_{\Omega} - Y_1^k/\mu_k \) \( V_k = (V_k^T V_k + \lambda I)^{-1} \).
3: \( V_{k+1} = (L_k + Y_1^k/\mu_k)U_{k+1} + \mathcal{P}_{\Omega} - Y_1^k/\mu_k \) \( U_{k+1} = (U_{k+1}^T U_{k+1} + \lambda I)^{-1} \).
4: \( U_{k+1} = 2 \mathcal{D}_{\lambda/(2\mu_k)}(U_{k+1} + Y_1^k/\mu_k), \) and \( V_{k+1} = \mathcal{D}_{\lambda/(2\mu_k)}(V_k + Y_2^k/\mu_k) \).
5: \( L_{k+1} = \mathcal{P}_{\Omega}(U_{k+1}^T V_{k+1} - Y_2^k/\mu_k) \).
6: \( Y_{1,k+1} = Y_1^k + \mu_k(U_{k+1} - \mathcal{P}_{\Omega}) \), \( Y_{2,k+1} = Y_2^k + \mu_k(V_{k+1} - \mathcal{P}_{\Omega}) \), and \( Y_{3,k+1} = Y_3^k + \mu_k(U_{k+1}^T V_{k+1} - \mathcal{P}_{\Omega}) \).
7: \( \mu_{k+1} = \min(\mu_k, \mu_{\text{max}}) \).
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:

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

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

Similar to (17) and (18), we also introduce the matrices \( \mathcal{U} \) and \( \mathcal{V} \) as auxiliary variables to (51) (i.e., (55) in this paper), and obtain the following equivalent formulation,

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

The augmented Lagrangian function of (53) is

\[
\begin{align*}
L_\mu = & \frac{\lambda}{2} (\|U\|_* + \|V\|_*) + \frac{1}{2} \|\mathcal{P}_{\Omega}(L) - \mathcal{P}_{\Omega}(D)\|_F^2 + \langle Y_1, U - \mathcal{U} \rangle + \langle Y_2, V - \mathcal{V} \rangle + \langle Y_3, L - UV^T \rangle \\
& + \frac{\mu}{2} \left( \|U - \mathcal{U}\|_F^2 + \|V - \mathcal{V}\|_F^2 + \|L - UV^T\|_F^2 \right)
\end{align*}
\]

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:

\[
\begin{align*}
\|U - \mathcal{U}_k + Y_1^k/\mu_k\|_F^2 + \|L_k - UV_k^T + Y_3^k/\mu_k\|_F^2, \\
\|V - \mathcal{V}_k + Y_2^k/\mu_k\|_F^2 + \|L_k - U_{k+1}V^T + Y_3^k/\mu_k\|_F^2.
\end{align*}
\]

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

\[
\begin{align*}
U_{k+1} = (L_k + Y_3^k/\mu_k)U_k + \mathcal{P}_{\Omega} - Y_3^k/\mu_k \left(V_k^T V_k + \lambda I^2 \right)^{-1}, \\
V_{k+1} = (L_k + Y_3^k/\mu_k)U_k + \mathcal{P}_{\Omega} - Y_3^k/\mu_k \left(U_{k+1}^T U_{k+1} + \lambda I^2 \right)^{-1}.
\end{align*}
\]
Algorithm 4 Solving image recovery problem (52) via ADMM

Input: $\mathcal{P}_U(D)$, the given rank $d$, and $\lambda$.
Initialize: $\mu_0 = 10^{-4}$, $\mu_{\text{max}} = 10^{20}$, $\rho > 1$, $k = 0$, and $\epsilon$.

1: while not converged do
2: $U_{k+1} = [(\mu_k L_k + Y^k_2)V_k] (\mu_k V^T_k V_k + (2\lambda/3)I)^{-1}$.
3: $V_{k+1} = [(L_k + Y^k_2/\mu_k)^T U_{k+1} + \bar{V}_k - Y^k_1/\mu_k] (U^T_{k+1} U_{k+1} + I)^{-1}$.
4: $\bar{V}_{k+1} = D_{2\lambda/(3\mu_k)}(V_{k+1} + Y^k_1/\mu_k)$.
5: $L_{k+1} = \mathcal{P}_U \left( \frac{D_{\mu_k} U_{k+1} V^T_{k+1} Y^k_2}{1 + \mu_k} \right) + \mathcal{P}_U \left( U_{k+1} V^T_{k+1} - Y^k_2/\mu_k \right)$.
6: $Y^k_1 = Y^k_1 + \mu_k (V_{k+1} - \bar{V}_{k+1})$ and $Y^k_2 = Y^k_2 + \mu_k (L_{k+1} - U_{k+1} V^T_{k+1})$.
7: $\mu_{k+1} = \min (\mu_k, \mu_{\text{max}})$.
8: $k \leftarrow k + 1$.
9: end while
Output: $U_{k+1}$ and $V_{k+1}$.

Updating $\bar{U}_{k+1}$ and $\bar{V}_{k+1}$:
By keeping all other variables fixed, $\bar{U}_{k+1}$ is updated by solving the following problem:

$$\frac{\lambda}{2} \| \bar{U} \|^2 + \frac{\mu_k}{2} \| U_{k+1} - \bar{U} + Y^k_1/\mu_k \|^2.$$  \hfill (58)

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

$$\bar{U}_{k+1} = D_{\lambda/(2\mu_k)} \left( U_{k+1} + Y^k_1/\mu_k \right).$$  \hfill (59)

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

$$\bar{V}_{k+1} = D_{\lambda/(2\mu_k)} \left( V_{k+1} + Y^k_2/\mu_k \right).$$  \hfill (60)

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

$$\frac{1}{2} \| \mathcal{P}_U(L) - \mathcal{P}_U(D) \|^2 + \frac{\mu_k}{2} \| L - U_{k+1} V^T_{k+1} + Y^k_3/\mu_k \|^2.$$  \hfill (61)

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

$$L_{k+1} = \mathcal{P}_U \left( \frac{D + \mu_k U_{k+1} V^T_{k+1} - Y^k_3}{1 + \mu_k} \right) + \mathcal{P}_U \left( U_{k+1} V^T_{k+1} - Y^k_3/\mu_k \right)$$  \hfill (62)

where $\mathcal{P}_U^\perp$ is the complementary operator of $\mathcal{P}_U$, i.e., $\mathcal{P}_U^\perp(D)_{ij} = 0$ if $(i,j) \in \Omega$, and $\mathcal{P}_U^\perp(D)_{ij} = D_{ij}$ 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 LMaFit [8], RegL1 [9], Unifying [10], factEN [11], RPCA [6], PSVT [12], WNNM [13], and LpSq [14]. The Matlab code of the proposed methods can be downloaded from the link [10].

4. https://sites.google.com/site/yinqiangzheng/
5. https://cpslab.snu.ac.kr/people/eunwoo-kim
9. https://sites.google.com/site/feipingjie/
10. https://www.dropbox.com/s/9ah4oezv1b1x5jm/Code_DFNM.zip?dl=0
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 $\lambda$), 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.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Bootstrap</th>
<th>Hall</th>
<th>Lobby</th>
<th>Mall</th>
<th>WaterSurface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Crowded scene</td>
<td>Crowded scene</td>
<td>Dynamic foreground</td>
<td>Crowded scene</td>
<td>Dynamic background</td>
</tr>
</tbody>
</table>

---

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 APGL [11] [18] and WNNM [12] [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 TNNR [13] [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**


13. https://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.


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.


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).