



Non-linear matrix completion

Jicong Fan*, Tommy W.S. Chow

Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China



ARTICLE INFO

Article history:

Received 26 April 2017

Revised 4 September 2017

Accepted 6 October 2017

Available online 12 October 2017

Keywords:

Matrix completion

Low-rank

Kernel

Schatten p -norm

Image inpainting

Single-/multi-label classification

Non-linear denoising

ABSTRACT

Conventional matrix completion methods are generally linear because they assume that the given data are from linear transformations of lower-dimensional latent subspace and the matrix is of low-rank. Therefore, these methods are not effective in recovering incomplete matrices when the data are from non-linear transformations of lower-dimensional latent subspace. Matrices consisting of such nonlinear data are always of high-rank or even full-rank. In this paper, a novel method, called non-linear matrix completion (NLMC), is proposed to recover missing entries of data matrices with non-linear structures. NLMC minimizes the rank (approximated by Schatten p -norm) of a matrix in the feature space given by a non-linear mapping of the data (input) space, where kernel trick is used to avoid carrying out the unknown non-linear mapping explicitly. The proposed NLMC is compared with existing methods on a toy example of matrix completion and real problems including image inpainting and single-/multi-label classification. The experimental results verify the effectiveness and superiority of the proposed method. In addition, the idea of NLMC can be extended to a non-linear rank-minimization framework applicable to other problems such as non-linear denoising.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Matrix completion [1–4] is to recover the missing entries of partially observed matrices. The problem of missing data [5–7] and matrix completion occurs in many areas such as machine learning and computer vision. Matrix completion has been applied to many tasks such as collaborative filtering [8], image recovery/inpainting [9–11], and classification [12–14]. A fundamental assumption of matrix completion methods is that the given data matrix is of low-rank. The low-rank property enables us to recover the missing entries through minimizing the matrix rank. Many researchers have provided theoretical proofs that the missing entries of a low-rank matrix could be exactly recovered with high probability under certain constraints of missing rate, matrix rank, and sampling scheme [1,15–17].

There are two major categories of matrix completion methods. The first category is matrix factorization based methods [18,19]. The underlying theory is that an $m \times n$ matrix of rank- r can be factorized into two smaller matrices of size $m \times r$ and $r \times n$, where $r < \min(m, n)$. Hence the missing entries can be recovered through finding such pair-wise matrices [18,20]. Another category of matrix completion methods is rank minimization based methods [1,21,22]. A widely-used convex relaxation of rank is nuclear-norm, the sum of the singular values of matrix. Recently, a few extensions of

nuclear-norm were applied to matrix completion. For example, in [23], Schatten p -norm, defined as the p -root of the sum of singular values' p -power, was used for matrix completion. Nuclear-norm is a special case of Schatten p -norm when $p = 1$. Compared with nuclear-norm, Schatten p -norm with $0 < p < 1$ sometimes provided better performance in matrix completion [23]. In [24], given an incomplete matrix in which the observed entries are corrupted by sparse noises, the Schatten p -norm of the unknown low-rank matrix and the ℓ_p -norm of the approximation errors for the observed entries are minimized jointly. Therefore, the method is robust to sparse noises. In [9], truncated nuclear-norm minimization was proposed for matrix completion. Truncated nuclear-norm [9,25], defined as the nuclear-norm subtracted by the sum of the largest few singular values, is a better approximation than nuclear-norm for matrix rank because the largest few singular values contain important information and should be preserved. In [26], an iteratively reweighted nuclear-norm algorithm was proposed to solve the nonconvex nonsmooth extensions (e.g., Schatten p -norm and Logarithm) of low-rank minimization problem. The experimental results in [26] showed that Schatten p -norm outperformed other nonconvex nonsmooth extensions of rank-minimization. In [27], a self-representation based matrix completion framework was proposed for handling incomplete data drawn from multiple subspaces. It has been shown that sparse self-representation based matrix completion can often outperform least-square and low-rank self-representations based matrix completion.

* Corresponding author.

E-mail address: fanj.c.rick@gmail.com (T.W.S. Chow).

It is worth noting that the aforementioned methods are linear matrix completion (LMC) methods. Here the linearity means that the data points forming the given matrix are from linear transformations of a lower-dimensional subspace. In other words, a given data point $x \in \mathbb{R}^m$ is produced by $x = Pz$, where $z \in \mathbb{R}^r$ ($r < m$) is the data in the latent subspace, variables in z are mutually uncorrelated, and $P \in \mathbb{R}^{m \times r}$ is the linear transformation. Therefore, the matrix $X \in \mathbb{R}^{m \times n}$ ($n > r$) formed by such data points is of low-rank and the missing entries can be recovered through rank-minimization. The linear and low-rank model is the fundamental assumption of matrix completion and principle component analysis (PCA) [28]. However, data with non-linear structures widely exist in many areas such as machine learning and computer vision [4]. The non-linearity means that the observation $x \in \mathbb{R}^m$ is given by a non-linear mapping $f : \mathbb{R}^r \rightarrow \mathbb{R}^m$ of the data $z \in \mathbb{R}^r$ ($r < m$) in the latent subspace, i.e., $x = f(z)$. Numerous non-linear techniques of dimensionality reduction [29] or feature extraction, such as kernel PCA (KPCA) [30], have been proposed to handle data of non-linear structures. The aforementioned matrix completion methods are unable to effectively recover the missing entries of matrices when the data have non-linear structures.

More recently, a few researchers attempted to incorporate non-linear techniques into matrix completion [31–34]. The attempts were often related with the so-called inductive matrix completion [35] that takes advantage of the side information such as (in terms of recommendation systems) the users' age or social network or the movie's genre. In [32], a method called goal-directed inductive matrix completion was proposed. The method assumed that the incomplete matrix X could be approximated as $X = \Phi(A)\Phi(B)^T$, where $\Phi(A)$ and $\Phi(B)$ are the nonlinear mapped features of the side information A and B , and C is a low-rank matrix to learn. In [34], a nonlinear matrix completion method was proposed for classification. The method arranges the nonlinear mapped features and the known/unknown labels into a matrix assumed to be of low-rank. The unknown labels are regarded as missing entries and hence can be recovered by matrix completion. However, these non-linear extensions of matrix completion have considerable limitations. First, in the methods of [31,32], the nonlinear operations are only performed on side information. Thus, the nonlinear operations will be inapplicable when there is no available side information (e.g., in classification tasks) or the side information are incomplete. Second, in the methods of [32,33], the incomplete matrix X is actually factorized into three matrices. Therefore, in terms of X , they are linear methods and cannot be regarded as exact nonlinear methods. Finally, the method in [34] is only applicable to classification task because it requires that all the missing entries can be organized into a sub-matrix. Obviously, it cannot be applied to more general matrix completion problems such as image inpainting and collaborative filtering. Moreover, the method requires three parameters (kernel parameter, regularization parameter, and the number of low-rank components) to be determined in advance, which increases the difficulty of practical application.

In this paper, we study the problem of matrix completion for data of non-linear structures. Hence, we propose a novel method, called non-linear matrix completion (NLNC). As a matter of fact, for complete data of non-linear structures, numerous non-linear techniques such as KPCA, SVM, and artificial neural networks can be used to exploit the data and solve practical tasks such as dimensionality reduction and classification. Unfortunately, those non-linear techniques are inapplicable to incomplete data. NLNC aims at recovering the missing entries of matrices where the data have non-linear structures. Hence, NLNC can also be utilized to preprocess incomplete data of non-linear structures in many tasks such as dimensionality reduction and classification. NLNC takes advantage of kernel methods [30,36,37] to handle non-linearity because kernel method can be a non-parameter

method in which there is no need to learn parameters. Specifically, NLNC maps the given data into some feature space via a non-linear function and minimizes the Schatten p -norm of the matrix formed by the mapped features. The kernel trick is utilized to avoid performing the non-linear mapping function explicitly. As Schatten p -norm with $0 < p \leq 1$ is an approximation for matrix rank, NLNC minimizes the matrix rank of data in the feature space induced by kernel. Hence, the missing entries of the data in the input (observation) space are optimized to fit a low-rank or approximately low-rank model in the feature space. The minimization problem of NLNC can be solved by well-developed non-linear optimization techniques such as non-linear conjugate gradient method and Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) [38] algorithm. We compare NLNC with existing methods [9,20,26,34,39] on synthetic data, image inpainting, and single-/multi-label classification [12,40–43] tasks. In addition, we extend the idea of NLNC to a non-linear rank-minimization framework that can be used to solve other problems such as non-linear denoising.

The contributions of this paper are three-fold. First, the problem of matrix completion for data of non-linear structures is studied, and a new method called NLNC is proposed to solve the problem. Compared with the recent attempts (e.g., [31–34]) in incorporating nonlinear techniques into matrix completion, our NLNC is more flexible and versatile because it absolutely gets rid of the limitations of the methods proposed in [31–34]. Compared with the nonlinear method in [34] that has three parameters to determine, our NLNC method has only one parameter (the kernel parameter), which makes our NLNC easier to apply in practice. It should be pointed out that in terms of correlations among the incomplete data, these previous attempts [31–34] are actually linear methods, but our NLNC is exactly nonlinear. Second, we propose to apply NLNC to single-/multi-label classification. In the classification tasks, NLNC outperforms not only other matrix completion methods but also conventional classifiers. Finally, the idea of NLNC is extended to a non-linear rank-minimization framework that is applicable to noisy NLNC problem and non-linear de-noising problem. For example, a non-linear rank-minimization based denoising (NLRCMDN) method is proposed for image denoising and NLRCMDN outperforms PCA and KPCA (pre-image problem [44–46]) significantly. It is worth noting that, concurrently with our work, Ongie et al. [47] developed a similar solution for the problem of non-linear matrix completion. However, our solution is more complete on account of the following aspects. First, we studied both polynomial kernel and RBF kernel while Ongie et al. studied only polynomial kernel. The optimization in [47] is only applicable to polynomial kernel. Second, we studied both Schatten p -norm and nuclear norm while Ongie et al. studied only nuclear norm. Third, we studied the problems of matrix completion based image inpainting and single-/multi-label classification while Ongie et al. studied the problem of subspace segmentation with missing data. Finally, we extended our proposed method as a nonlinear rank minimization framework for solving other problems such as nonlinear denoising.

The remaining content of this paper are organized as the followings. In Section 2, the related work are introduced. Section 3 details our non-linear matrix completion method as well as its extensions and applications. Section 4 compares the proposed method with existing methods in the tasks of synthetic matrix completion, image inpainting, single-/multi-label classification, and image denoising. Section 5 gives a conclusion for this paper.

2. Related work

In this section, we give more details of a few state-of-the-art matrix completion methods that will be compared with the method proposed in this paper.

Matrix factorization (MF) based matrix completion [18] solves the following problem

$$\begin{aligned} & \min_{X,P,Z} \|P\|_F^2 + \|Z\|_F^2, \\ & \text{s.t. } X = PZ, \quad X_{i,j} = M_{i,j}, \quad (i, j) \in \Omega \end{aligned} \quad (1)$$

where $P \in \mathbb{R}^{m \times r}$, $Z \in \mathbb{R}^{r \times n}$, and r is a parameter that should be determined beforehand. The parameter r should be the rank of X . But in practice, it is difficult to know or estimate r because X is incomplete. In [20], the proposed low-rank matrix fitting (LMFit) algorithm is able to dynamically and adaptively adjust r .

Nuclear norm minimization (NNM) based matrix completion [1] can be formulated as

$$\begin{aligned} & \min_X \|X\|_*, \\ & \text{s.t. } X_{i,j} = M_{i,j}, \quad (i, j) \in \Omega, \end{aligned} \quad (2)$$

where $\|X\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(X)$ is the nuclear norm of X and $\sigma_i(X)$ denotes the i -th singular value of X . NNM based matrix completion can be solved through different approaches such as the singular value thresholding (SVT) algorithm [48], inexact augmented Lagrange multiplier (IALM) method [39], and alternating direction method (ADM) [22].

Truncated nuclear norm minimization (TNNM) based matrix completion [9] solves

$$\min_X \|X\|_r, \quad \text{s.t. } X_{i,j} = M_{i,j}, \quad (i, j) \in \Omega. \quad (3)$$

where $\|X\|_r$ denotes the truncated nuclear norm and $\|X\|_r = \|X\|_* - \sum_{i=1}^r \sigma_i(X) = \sum_{i=r+1}^{\min(m,n)} \sigma_i(X)$. Alternating direction method of multipliers (ADMM) [49] can be applied to the optimization of (3).

General nonconvex and nonsmooth model of matrix completion [26] can be formulated as

$$\min_X \sum_{i=1}^{\min(m,n)} g(\sigma_i(X)) + \frac{1}{2} \| (X - M)_{\Omega} \|_F^2, \quad (4)$$

where the penalty function g could be Logarithm penalty, L_p penalty [23,24], and so on. In [26], (4) was solved by an iteratively reweighted nuclear-norm (IRNN) algorithm.

Classification oriented nonlinear matrix completion (C-NLMC) [34] solves the following problem

$$\begin{aligned} & \min_{U,V} \|U\|_F^2 + \|V\|_F^2, \\ & \text{s.t. } (UV^T - \bar{X})_{\Omega} = 0, \quad V^T V = I, \end{aligned} \quad (5)$$

where

$$\bar{X} = \begin{bmatrix} Y_0 & Y_1 \\ \Phi(X_0) & \Phi(X_1) \end{bmatrix}, \quad (6)$$

Y_0 are the known labels for the training data X_0 , and Y_1 are the unknown labels to be predicted for the testing data X_1 . $\Phi(\cdot)$ is a nonlinear mapping implicitly implemented through kernel trick. The optimization problem can be solved with an interior-point algorithm [34].

3. Non-linear matrix completion (NLMC)

3.1. Theory and model of NLMC

In this paper, the problem of matrix completion for data of non-linear structures is studied. We assume that the data of non-linear structures is given by the following non-linear latent variable model:

$$x_i = f(z_i), \quad i = 1, 2, \dots, n \quad (7)$$

where $x \in \mathbb{R}^m$ is the vector of measurements or observations, $z \in \mathbb{R}^r (r < m)$ is the vector of latent variables, and $f(\cdot)$ is a non-linear function. The matrix formed by x_i is $X \in \mathbb{R}^{m \times n}$ corresponding to the matrix $Z \in \mathbb{R}^{r \times n}$ formed by the variables in the latent space. The latent variable model also requires that the r variables of z are mutually uncorrelated, which indicates that Z is of full-rank. Obviously, the model in (7) is beyond the linear assumption of LMC. Therefore, the matrix completion problem based on (7) cannot be solved by LMC methods. In addition, the condition $r < m$ is essential to recover the missing entries of X . The reason is that when $r < m$ or $r \ll m$, the information of X is redundant or highly redundant, which makes it possible to recover the missing entries of X . Therefore, $r < m$ for data of non-linear structures is in the same sense of low-rankness for data of linear structures. We assume that the observed entries of X are sampled uniformly at random, which is the most widely-used assumption in matrix completion [1]. The locations of the observed entries are denoted as Ω , and for any $(i, j) \in \Omega$, $X_{i,j} = M_{i,j}$. The sampling rate of X is denoted as ρ and then the number of observed entries of X is $|\Omega| = \rho mn$. A successful recovery for the missing entries of X indicates that the observed entries should include all the information of the missing entries. In other words, the observed entries of X should be enough to reconstruct Z so as to produce the missing entries of X . Therefore, a necessary condition to recover the missing entries of X is

$$r \leq \rho \times \min(m, n). \quad (8)$$

Otherwise, Z cannot be determined by the observed entries of X and then the missing entries cannot be produced. In linear matrix completion problem, the missing entries can be recovered through minimizing the rank, which is the dimensionality of the linear latent variables. Therefore, in nonlinear matrix completion problem with model (7), we may recover the missing entries through minimizing the dimensionality of the nonlinear latent variables, i.e.,

$$\begin{aligned} & \min_X \dim(Z), \\ & \text{s.t. } X = f(Z), \quad X_{i,j} = M_{i,j}, \quad (i, j) \in \Omega. \end{aligned} \quad (9)$$

However, problem (9) is intractable because the nonlinear mapping $f(\cdot)$ is not given and it is troublesome to learn an approximation for $f(\cdot)$.

In order to handle the nonlinear problem, we assume that there is a non-linear function $\phi(\cdot)$ that maps x into a feature space \mathcal{F} :

$$\phi(\cdot) : \mathbb{R}^m \rightarrow \mathcal{F}^l, \quad (10)$$

where \mathcal{F} is an inner product space and its dimensionality l could be arbitrarily large, possibly infinite. Then the data in \mathcal{F} can be handled by linear techniques. The matrix formed by the data in the feature space is

$$W = \Phi(X) = [\phi(x_1), \phi(x_2), \dots, \phi(x_n)], \quad (11)$$

which should be of low-rank or approximately low-rank. Then the skinny SVD for W is given by

$$W = U_{r'} \Sigma_{r'} V_{r'}^T, \quad (12)$$

where r' is the rank of W . Hence $w_i = \phi(x_i)$, the i -th column of W , can be written as

$$w_i = f'(z'_i) = Pz'_i, \quad i = 1, 2, \dots, n, \quad (13)$$

where $P = U_{r'} \Sigma_{r'}$ and z'_i is the i -th column of $V_{r'}^T$. The assumption and model in (12) and (13) are the same as that of KPCA [30]. It means that in the high-dimensional feature space, W is from a lower-dimensional subspace and can be represented by a linear transformation on a small number of latent variables. This coincides with that in the input space, X is redundant and can be represented by nonlinear transformation $f(\cdot)$ on a small number of latent variables Z . In other words, with the nonlinear mapping $\phi(\cdot)$,

the nonlinear latent variable model in (7) is transformed into the linear latent variable model in (13). Therefore, minimizing the rank of W is equivalent to minimizing the dimensionality of the nonlinear latent variables of X . The missing entries of X can be recovered by solving the following problem

$$\begin{aligned} & \min_X \text{rank}(W), \\ & \text{s.t. } W = \Phi(X), X_{i,j} = M_{i,j}, (i, j) \in \Omega, \end{aligned} \quad (14)$$

which can be regarded as an approximation for problem (9). Problem (14) is NP-hard but it can be approximated as

$$\begin{aligned} & \min_X \|W\|_*, \\ & \text{s.t. } W = \Phi(X), X_{i,j} = M_{i,j}, (i, j) \in \Omega. \end{aligned} \quad (15)$$

because the nuclear norm is the tightest convex relaxation of matrix rank. Recently, as a nonconvex relaxation of rank, the Schatten p -norm [23,24,50] with $0 < p < 1$ was also applied to rank-minimization problems. The Schatten p -norm is defined as

$$\|W\|_{Sp} = \left(\sum_{i=1}^{\min(n,l)} (\sigma_i(W))^p \right)^{1/p}, \quad (16)$$

where $\sigma_i(W)$ is the i -th singular value of W and $0 \leq p < \infty$. It can be seen that nuclear-norm is a special case of Schatten p -norm with $p = 1$, i.e., $\|W\|_{S_1} = \|W\|_*$. If $p = 0$, $\|W\|_{Sp} = \text{rank}(W)$. Therefore, with $0 < p \leq 1$, problem (14) can be approximated as

$$\begin{aligned} & \min_X \|W\|_{Sp}^p, \\ & \text{s.t. } W = \Phi(X), X_{i,j} = M_{i,j}, (i, j) \in \Omega. \end{aligned} \quad (17)$$

The relations among (14), (15) and (17) are as follows: both (15) and (17) are approximations of (14); (15) is a special case of (17). Substituting (16) into (17), we get

$$\begin{aligned} & \min_X \sum_{i=1}^{\min(n,l)} (\sigma_i(W))^p, \\ & \text{s.t. } W = \Phi(X), X_{i,j} = M_{i,j}, (i, j) \in \Omega. \end{aligned} \quad (18)$$

Denote the optimal solution of X in (18) as X^* and the corresponding W as $W^* = \Phi(X^*)$. Let Δ be a non-zero arbitrary perturbation to X^* and $\Delta_{i,j} = 0$ for all $(i, j) \in \Omega$. The perturbation expansion to W^* is denoted by E and $W^* + E = \Phi(X^* + \Delta)$. As Δ is not from the same nonlinear subspace of X^* , E is not from the same linear subspace of W^* . Therefore, with perturbation E , there will be at least one small (zero or nearly zero) singular value of W^* increases significantly, which leads to

$$\|W^* + E\|_{Sp}^p > \|W^*\|_{Sp}^p, \quad (19)$$

It means that the optimal solution of (17) is exactly the true X .

It is difficult to solve (17) directly because the map $\phi(\cdot)$ is not given or difficult to construct explicitly. As a matter of fact, Schatten p -norm can be written as

$$\|W\|_{Sp} = (\text{Tr}((W^T W)^{p/2}))^{1/p}, \quad (20)$$

where $\text{Tr}(\cdot)$ is matrix trace. Substituting (20) into problem (17), we get

$$\begin{aligned} & \min_X \text{Tr}((\Phi(X)^T \Phi(X))^{p/2}), \\ & \text{s.t. } X_{i,j} = M_{i,j}, (i, j) \in \Omega. \end{aligned} \quad (21)$$

which is further written as

$$\begin{aligned} & \min_X \text{Tr}([\phi(x_i)^T \phi(x_j)]_{n \times n})^{p/2}, \\ & \text{s.t. } X_{i,j} = M_{i,j}, (i, j) \in \Omega. \end{aligned} \quad (22)$$

$$\begin{aligned} & \text{In (22), in order to compute the inner product } \langle \phi(x_i), \phi(x_j) \rangle = \phi(x_i)^T \phi(x_j), \text{ we take advantage of kernel representations [30],} \\ & k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle, \end{aligned} \quad (23)$$

which allow us to obtain the value of inner product in the feature space \mathcal{F} without carrying out the non-linear mapping $\phi(\cdot)$ explicitly. In (23), $k(\cdot, \cdot)$ is a kernel that satisfies Mercer's theorem [30]. The basic definition of kernel is as following. Let \mathcal{X} be a non-empty set sometimes referred to as the index set, and let $k(\cdot, \cdot)$ be a symmetric real-valued (could be complex sometimes) function on $\mathcal{X} \times \mathcal{X}$. Although some positive semi-defined kernels have been used in practice, a kernel $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is often a positive definite function, which should meet the following condition:

$$\sum_{i,j=1}^n c_i c_j k(x_i, x_j) \geq 0 \quad (24)$$

for any $n, x_1, \dots, x_n \in \mathcal{X}$ and $c_1, \dots, c_n \in \mathbb{R}$. Hence, the choice of $k(\cdot, \cdot)$ implicitly determines the mapping $\phi(\cdot)$ and the feature space \mathcal{F} . Furthermore, such a feature space \mathcal{F} with a kernel is also called a reproducing kernel Hilbert spaces (RKHS). One widely-used kernel function is polynomial kernel

$$k(x, y) = (x^T y + b)^d, \quad (25)$$

where d is the degree of polynomial and b is a free parameter to trade off the influence of higher-order versus lower-order terms in the polynomial. The dimensionality of the feature space \mathcal{F} given by a polynomial kernel is quite high (i.e., $l \gg m$) but finite. Another popular kernel function is radial basis function (RBF) kernel

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right), \quad (26)$$

where σ is a free parameter determining the smoothness degree of the kernel. The dimensionality of the feature space \mathcal{F} given by an RBF kernel is infinite.

Now substituting (23) into (22) and denoting $K = [K_{ij}]_{n \times n} = [k(x_i, x_j)]_{n \times n}$, we get the following optimization problem

$$\begin{aligned} & \min_X \text{Tr}(K^{p/2}), \\ & \text{s.t. } K = [k(x_i, x_j)]_{n \times n}; X_{i,j} = M_{i,j}, (i, j) \in \Omega. \end{aligned} \quad (27)$$

Problem (27) is our NLMC formulation for data of non-linear structures given by the non-linear latent variable model defined in (7). As defined by (14) and (18), NLMC aims at minimizing the rank or Schatten p -norm ($0 < p \leq 1$) in the high-dimensional feature space \mathcal{F} . The problems (14) and (18) are finally transformed to problem (27), which is tractable because the non-linear mapping $\phi(\cdot)$ is implicitly carried out by kernel trick. Therefore, by solving (27), the missing entries of X can be recovered. The relation between existing methods and the proposed NLMC is shown in Fig. 1.

In [50], it has been proved that a low-rank matrix can be exactly recovered through minimizing Schatten p -norm ($0 < p \leq 1$) if the constraint is linear and with the restricted isometry property [50]. However, in (18), the constraint on W is nonlinear and both $f(\cdot)$ and $\Phi(\cdot)$ are unknown. Therefore, it is difficult to provide a theoretical proof in the manner of [50] or [1] for that W and the missing entries of X can be recovered with high probability. In this paper, thorough experimental results will show that the proposed method can recover the missing entries of X successfully and outperform existing methods significantly in practice.

NLMC can also benefit from other properties of kernel techniques. For example, as there are numerous kernel functions available, new kernels can also be designed through combining multiple different kernel functions to improve the performance of NLMC. It is worth noting that most of the widely-used kernels, such as polynomial kernel and RBF kernel, are positive definite.

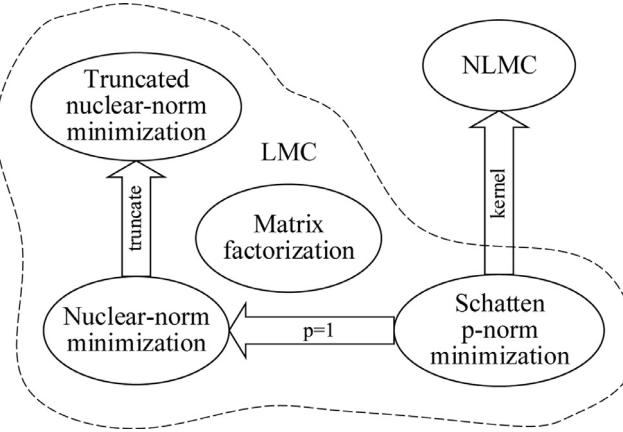


Fig. 1. Relation between existing methods and the proposed NLMC method.

Kernel matrices derived from positive definite kernels are positive definite. It means that the kernel matrix K in NLMC could be non-singular. As $K = W^T W$, $W \in \mathbb{R}^{l \times n}$, and $l \gg n$, then W is non-singular, which indicates that the smallest singular value of W is larger than zero. However, W could be of low-rank approximately, which means a number of singular values of W are nearly zero. As shown in (18), NLMC will make some singular values of W as small as possible so as to recover the missing entries of X .

3.2. Optimization for NLMC

The optimization problem of NLMC in (27) is often nonlinear and nonconvex because Schatten p -norm and kernel functions are often nonconvex. Therefore, NLMC cannot be efficiently solved by a number of convex optimization techniques such as IALM [39] and ADMM [22] that have been widely applied to LMC [9,19]. In addition, NLMC is actually an unconstrained problem because in (27) the constraints can be directly put into the objective function, i.e.,

$$\min_{X_{\Omega}} \text{Tr}(K^{p/2}), \quad (28)$$

where $K = [k(x_i, x_j)]_{n \times n}$ and $\bar{\Omega}$ denotes the positions of unknown entries. In the past decades, numerous of non-linear and unconstrained optimization techniques have been proposed to find local minima or maxima of non-convex problems. Representative non-linear optimization algorithms include non-linear conjugate gradient (NCG), Broyden–Fletcher–Goldfarb–Shanno (BFGS), and Limited-memory BFGS (L-BFGS) [38]. These algorithms usually require the value of objective function and its gradient only.

Denoting the objective function in (28) as $\mathcal{J} = \text{Tr}(K^{p/2})$, its partial derivative w.r.t. the kernel matrix is

$$\frac{\partial \mathcal{J}}{\partial K} = \frac{p}{2} K^{\frac{p-2}{2}}. \quad (29)$$

For any type of kernel function, the following chain rule is true:

$$\frac{\partial \mathcal{J}}{\partial X} = \sum_i^n \sum_j^n \frac{\partial \mathcal{J}}{\partial K_{ij}} \frac{\partial K_{ij}}{\partial X}. \quad (30)$$

For polynomial kernel (25) and RBF kernel (26), (30) can be simplified as

$$\frac{\partial \mathcal{J}}{\partial X_{uv}} = 2 \sum_i^n \frac{\partial \mathcal{J}}{\partial K_{iu}} \frac{\partial K_{iu}}{\partial X_{uv}}, \quad u = 1, \dots, m, \quad v = 1, \dots, n. \quad (31)$$

The partial derivatives of the objective function w.r.t. the unknown entries of X are $\partial \mathcal{J} / \partial X_{uv}$ with $(u, v) \in \bar{\Omega}$. In (29), the computational complexity of the fractional power of K (when $\frac{p-2}{2}$ is not an

integer) is similar with that of SVD. Therefore, the computational complexity of NLMC is comparable with that of rank-minimization based LMC methods.

By inputting the partial derivatives into the optimization algorithms such as L-BFGS and NCG, the missing entries can be obtained. As the problem is non-convex, the initializations of the missing entries affect the values of the local minima and the convergence speed of optimization. We propose to initialize each missing entry as the average of the corresponding row mean and column mean of the known entries. In general, the computational overhead of L-BFGS iteration is larger than that of NCG but L-BFGS usually needs fewer iterations to converge. Therefore, in this paper, we suggest L-BFGS for NLMC because the gradient in NLMC is computationally expensive. For large scale problem, NLMC should be solved by NCG.

3.3. Extension and application of NLMC

3.3.1. Noisy NLMC

The observed entries of an incomplete matrix may be corrupted with noises, i.e., $M_{\Omega} = X_{\Omega} + E_{\Omega}$. The noisy LMC problem has been studied by previous works [51–53]. In this paper, the noisy NLMC problem can be solved by solving the following optimization task:

$$\begin{aligned} & \min_X \text{Tr}(K^{p/2}) + \frac{\lambda}{2} \|H \odot (X - M)\|_F^2, \\ & \text{s.t. } K = [k(x_i, x_j)]_{n \times n}, \end{aligned} \quad (32)$$

where H is a binary matrix that the entries in Ω are 1 and the entries not in Ω are 0; the symbol \odot denotes the Hadamard product; λ is a parameter to regularize the impact of the noises. The partial derivative of the objective function in (32) w.r.t. the kernel matrix is $\partial \mathcal{J} / \partial K = \frac{p}{2} K^{\frac{p-2}{2}} + \lambda H \odot (X - M)$. The optimization of noisy NLMC is similar with that of NLMC.

3.3.2. Non-linear denoising

It is known that PCA can be directly applied to linear denoising problem of complete data, where the observations X are noisy, i.e., $X = \hat{X} + E$. To apply KPCA to non-linear denoising problem, the pre-image problem [44–46,54] of KPCA was studied in the past years. In KPCA pre-image, first, KPCA is used to reduce the dimension of the data (or extract non-linear features); then the data of lower-dimensionality (or extracted features) are mapped back to reconstruct the observed data; the reconstruction of the observed data is regarded as the clean data. As defined in [44], the pre-image problem of KPCA is to find a reconstruction of x , say \hat{x} , that

$$\phi(\hat{x}) \simeq P\phi(x), \quad (33)$$

where $P\phi(x)$ is the projection of $\phi(x)$ on the subspace spanned by the principal eigenvectors of KPCA. Mika et al. [44] proposed to minimize $\|\phi(\hat{x}) - P\phi(x)\|^2$ by a fixed-point iteration method, which is applicable to only RBF kernel. Kwok et al. [45] proposed to find the pre-image of KPCA through locally preserving the distances from feature spaces to input space. The method computes the distances of pair-wise points in the feature space and obtains the input-space distances; then multidimensional scaling is utilized to compute \hat{x} . In this paper, the idea of NLMC can also be extended to the following non-linear denoising problem:

$$\begin{aligned} & \min_{\hat{X}} \text{Tr}(K^{p/2}) + \frac{\lambda}{2} \|\hat{X} - X\|_F^2, \\ & \text{s.t. } K = [k(\hat{x}_i, \hat{x}_j)]_{n \times n}. \end{aligned} \quad (34)$$

We denote problem (34) as non-linear rank-minimization based denoising (NLRMDN). Compared with KPCA pre-image, NLRMDN does not need dimensionality reduction. Moreover, the parameter p

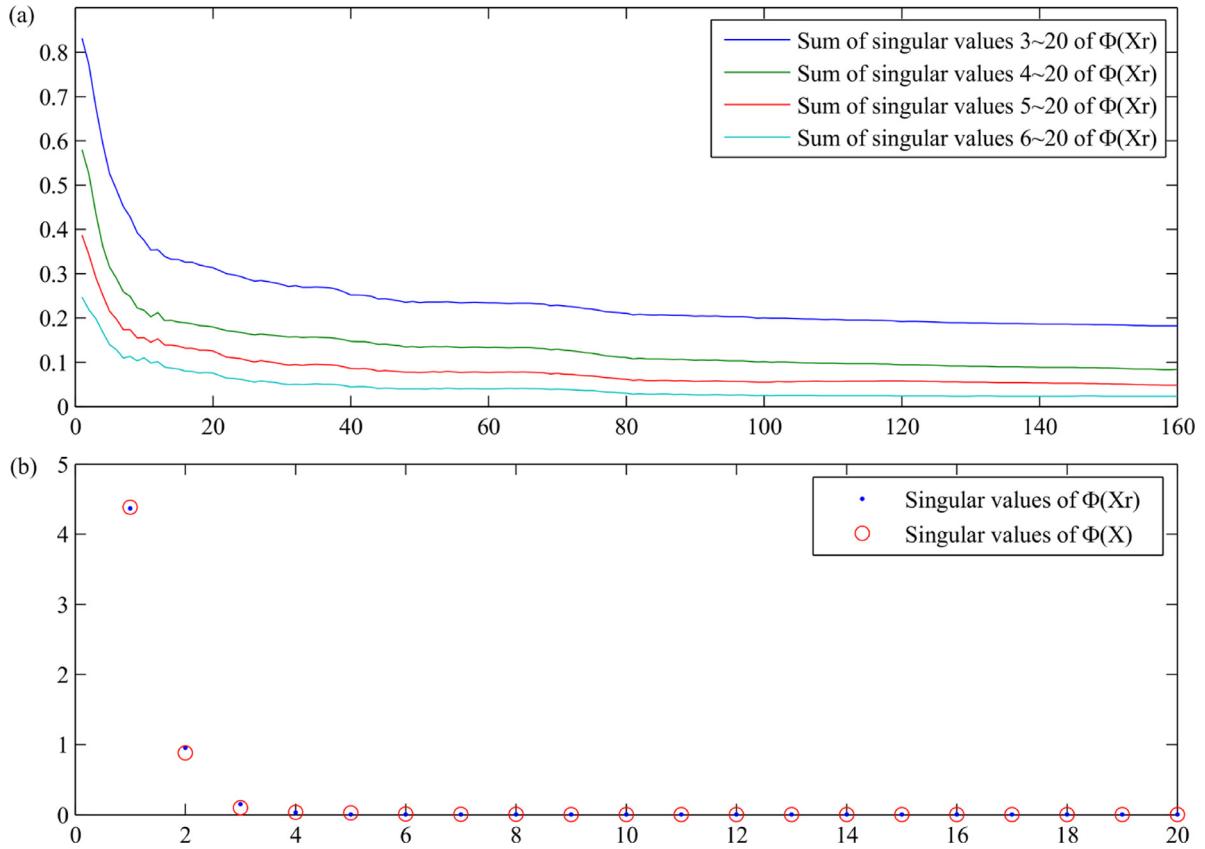


Fig. 2. Singular values of data matrix $\Phi(X_r)$ in feature space given by NLMC with RBF kernel on the toy dataset: (a) the sum of small singular values in 160 iterations; (b) the singular values of $\Phi(X)$ and $\Phi(X_r)$ with final X_r .

of Schatten p -norm makes NLRMDN more flexible than KPCA pre-image. Finally, NLRMDN can handle missing data but KPCA pre-image cannot. Generally, the idea of this paper can be extended to the following non-linear rank-minimization (NLRM) framework:

$$\begin{aligned} & \min_X \text{Tr}(K^{p/2}) + h(X), \\ & \text{s.t. } K = [k(x_i, x_j)]_{n \times n}. \end{aligned} \quad (35)$$

3.3.3. NLMC based classification

In recent years, matrix completion was applied to both single-label classification [12] and multi-label classification [13,55]. In MC based classification, the features and labels are used to form an incomplete matrix where the missing entries are the unknown labels. The unknown labels can be predicted with MC because the label is assumed to be a linear function of the features so that the feature-label matrix is of low-rank. MC can provide higher classification accuracy than conventional classifiers (e.g. SVM) do. In addition, compared with conventional classifiers, MC based classifiers are much more flexible because they can handle missing data. Because existing MC methods are linear methods, we denote the existing MC based classification as LMC based classification. LMC classifiers consider the following problem

$$\begin{aligned} & \min_{Y_{\text{test}}} \|\bar{X}\|_*, \\ & \text{s.t. } \bar{X} = \begin{bmatrix} Y_{\text{train}} & Y_{\text{test}} \\ X_{\text{train}} & X_{\text{test}} \\ 1 & 1 \end{bmatrix} \end{aligned} \quad (36)$$

by assuming that $y = Wx + b$ (a linear classifier), where X_{train} , X_{test} , Y_{train} , and Y_{test} are the matrices of training features, testing features, training labels, and unknown labels respectively. The labels are usually in binary form of (0,1) for single-label data and (-1,1)

for multi-label data. However, a non-linear classifier, i.e., $y = g(x)$ may be better than a linear classifier. Therefore, compared with LMC classifiers, the following NLMC classifier is expected to perform better in classification tasks,

$$\begin{aligned} & \min_{Y_{\text{test}}} \text{Tr}(K^{p/2}), \\ & \text{s.t. } K = [k(\bar{x}_i, \bar{x}_j)]_{n \times n}, \quad \bar{X} = \begin{bmatrix} Y_{\text{train}} & Y_{\text{test}} \\ X_{\text{train}} & X_{\text{test}} \end{bmatrix}. \end{aligned} \quad (37)$$

Because NLMC is non-convex, the initialization of Y_{test} is important to the classification. For single-label data, y_{test} should be initialized as $[1/c, 1/c, \dots, 1/c]^T$ where c is the number of classes. For multi-label data, y_{test} should be initialized as $[0, 0, \dots, 0]^T$ where c is also the number of classes.

4. Experiments

In this paper, the proposed NLMC with second-order polynomial kernel and RBF kernel will be compared with LMaFit, NNM, TNNM, IRNN (Lp penalty), and C-NLMC. The comparative studies are implemented on an example of matrix completion for nonlinear toy data, image inpainting task, and matrix completion based single-/multi-label classification tasks. Particularly, in the tasks of classification, several other classifiers such as SVM, convolutional neural network (CNN) and multi-label kNN [41] will also be compared with NLMC. The proposed NLRMDN will be compared with PCA and two KPCA pre-image methods in denoising task. The optimizations of NLMC and NLRMDN are solved by L-BFGS algorithm of the MATLAB function *minFunc* given by [56]. The matrix completion performance is measured by the relative error of matrix completion, *MC error* = $\|(X_r - X)\|_F / \|X_r\|_F$, in which X_r is the re-

Table 1

MC error (%) on the toy dataset.

δ	LMaFit	NNM	TNNM	IRNN	NLMC-Poly	NLMC-RBF
20%	57.94	40.51	19.63	40.28	3.56	1.62
40%	48.56	50.45	25.42	50.43	8.04	3.91
60%	78.57	62.55	69.64	62.16	20.8	11.28
80%	100	100	100	100	43.84	34.82

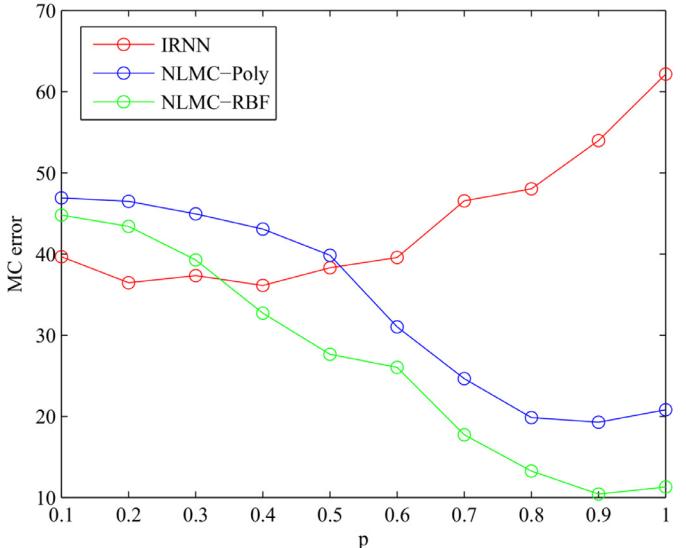
covered matrix, X is the true matrix, and $\bar{\Omega}$ denotes the indices of the unknown entries.

4.1. A non-linear toy example

Considering the non-linear latent variable model defined in (7), we sample the latent variable z from a one-dimensional standard uniform distribution on the open interval (0,1). A five-dimensional measurements x are obtained by the following transformations:

$$\begin{aligned} x_1 &= z, \quad x_2 = z^2, \quad x_3 = z^3, \\ x_4 &= \exp(z), \quad x_5 = \sin(z). \end{aligned} \quad (38)$$

By sampling 20 data points for z , a 5×20 matrix is obtained. We randomly remove 1, 2, 3, and 4 entries of each column of the matrix and get incomplete matrices with missing rates (δ) of 20%, 40%, 60%, and 80% respectively. LMC and NLMC ($p=1$) are applied to recover the missing entries of these matrices. In LMaFit, the minimal rank and maximal rank are set as 1 and 3 respectively; in TNNM, the parameter r is set as 1; in NLMC, the parameter b of polynomial kernel is set as 100 and the parameter σ^2 of RBF kernel is set as 10. The experiments are repeated for 100 times and the average MC errors are reported in Table 1, where the best result in each case is highlighted with bold type. It can be found that, in all cases, NLMC methods outperform all LMC methods significantly, and NLMC with RBF kernel gives the lowest MC errors. To analyse the mechanism of NLMC, fixing $\delta = 60\%$, we recorded the singular values (descending order) of $\Phi(X)$ and $\Phi(X_r)$ in feature space, where X is the original complete matrix and X_r is the recovered matrix given by NLMC with RBF kernel. Fig. 2(a) gives the sums of the small singular values ($3 \sim 20$, $4 \sim 20$, $5 \sim 20$, and $6 \sim 20$) in the 160 iterations. As can be seen, these smaller singular values are minimized successfully. As shown in Fig. 2(b), the singular values of $\Phi(X_r)$ with final X_r are almost the same as that of $\Phi(X)$, and most of the small singular values are nearly zero. These two aspects in Fig. 2 verify how NLMC works and why it can recover the missing entries. To investigate the influence of parameter p , in Fig. 3 ($\delta = 60\%$), NLMC are compared with IRNN for $0.1 \leq p \leq 1$. It

**Fig. 3.** MC errors (%) of IRNN and NLMC with different p for the toy dataset.

is found that NLMC-Poly and NLMC-RBF give the lowest MC errors when $p = 0.9$; IRNN achieves lowest MC error when $p = 0.4$. The results for this toy dataset demonstrate that NLMC is much more effective than LMC in handling data of non-linear structures.

4.2. Image inpainting

We compare NLMC with LMC on image inpainting problem [9,26]. A popular image inpainting method, higher-order total variation (HOTV) [57,58], was also compared in this study. As shown in Fig. 4, we consider eight color images, six of which are from [9,25] and one of which is from [10]. Fig. 5 shows five masks for produce incomplete images: random masks that randomly remove 50% and 80% pixels, text masks with small and large font sizes, and big block mask.

In LMaFit, the rank parameter is chosen from {1, 2, 3, 5, 10, 15, 20, 25, 30}, and the better result of fixed rank and dynamically adjusted rank is reported. In TNNM, the parameter r is chosen from {1, 2, 3, 5, 10, 15, 20}. In NLMC, the parameter b of polynomial kernel is chosen from $\{10^2, 10^3, 10^4\}$; the parameter σ^2 of RBF kernel is chosen from {1, 5, 10, 50, 100}. For the random masks, the average results of 20 repeated trials are recorded. The MC errors for all images with different masks are reported in Table 2. It can

**Fig. 4.** Eight color images for inpainting.

Table 2
MC error (%) for image inpainting.

Image	Mask	HOTV	LMaFit	NNM	TNNM	IRNN	NLMC-Poly	NLMC-RBF
1	random(50%)	9.25	6.88	5.49	5.39	5.43	5.31	5.22
	random(80%)	12.78	10.47	9.27	8.91	9.23	9.08	8.78
	text(small)	11.26	8.07	6.83	6.61	6.81	6.68	6.54
	text(large)	16.54	10.16	8.43	8.25	8.51	8.32	8.19
	block	25.05	9.65	10.66	8.27	10.82	9.14	8.36
2	random(50%)	13.24	8.66	7.89	7.62	7.85	7.63	7.41
	random(80%)	16.37	11.37	11.05	10.53	10.95	10.61	9.95
	text(small)	16.85	9.56	8.98	8.73	9.01	8.82	8.63
	text(large)	21.33	12.29	12.08	10.36	12.24	11.34	10.13
	block	45.19	12.54	12.87	11.24	12.92	11.31	9.77
3	random(50%)	8.59	10.58	9.68	9.38	9.69	9.31	8.37
	random(80%)	10.46	16.05	14.34	13.27	14.35	13.45	11.02
	text(small)	9.95	12.76	11.38	10.64	11.46	11.12	9.78
	text(large)	11.34	14.82	13.51	12.19	13.61	12.83	11.18
	block	46.53	34.32	43.83	24.18	44.79	29.26	19.69
4	random(50%)	4.17	5.66	4.68	4.52	4.64	4.43	4.24
	random(80%)	6.42	8.62	7.81	7.35	7.83	7.41	6.55
	text(small)	5.76	8.13	6.72	6.38	6.71	6.56	5.92
	text(large)	9.23	10.64	9.69	8.93	9.75	9.04	7.84
	block	51.64	20.82	20.69	19.27	20.58	20.19	19.04
5	random(50%)	5.57	5.99	4.73	4.42	4.77	4.31	3.31
	random(80%)	6.84	9.59	9.67	8.99	9.73	8.91	6.04
	text(small)	6.61	7.75	6.24	5.94	6.44	6.08	4.63
	text(large)	10.18	12.19	11.09	10.61	11.22	10.66	7.96
	block	51.73	27.11	26.33	22.87	26.96	23.14	20.73
6	random(50%)	12.98	16.19	13.82	13.26	13.83	13.37	12.53
	random(80%)	15.24	18.74	18.03	17.54	17.92	17.66	16.14
	text(small)	18.33	20.56	18.27	17.77	18.25	17.92	17.06
	text(large)	21.72	23.45	21.92	21.75	21.94	21.58	20.42
	block	58.96	28.08	28.36	28.25	28.37	28.15	27.23
7	random(50%)	13.09	17.81	14.58	14.36	14.64	14.14	12.17
	random(80%)	16.23	24.01	20.46	20.18	20.49	20.12	18.08
	text(small)	14.16	19.43	16.11	15.94	16.19	15.85	13.78
	text(large)	17.18	20.69	17.29	16.85	17.38	16.79	14.42
	block	52.80	16.82	19.43	15.82	19.85	16.67	14.35
8	random(50%)	6.71	6.98	5.68	5.51	5.61	5.59	4.79
	random(80%)	7.56	9.05	7.97	7.73	8.01	7.63	6.26
	text(small)	7.37	7.94	6.56	6.39	6.61	6.46	5.82
	text(large)	7.34	8.42	7.23	6.82	7.19	6.85	6.31
	block	51.48	36.49	41.25	33.64	41.89	36.13	28.15

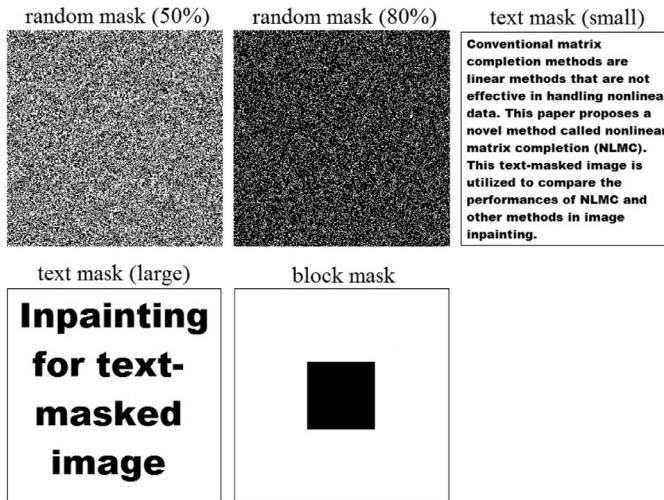


Fig. 5. Five masks for inpainting.

be found that the MC errors given by HOTV for all images with block mask are extremely high. The reason is that HOTV is a local method and cannot effectively inpaint big holes of images [10]. For images 1 and 2, because the rank of the matrices are very low, NLMC methods slightly outperform LMC methods and HOTV performs the worst. However, for images 3, 4, 6, and 7 with random

(50% and 80%) masks and text (small) mask, HOTV gives significant lower errors than all LMC methods do; for images 3, 4, 6, and 7 with random (80%) mask, HOTV outperforms NLMC-RBF slightly. The reason is that the ranks of these images are high and the texture in each image is complex and various, which is challenging for LMC methods, but can be handled by HOTV as it is a local method and powerful in inpainting small holes of images. In general, NLMC-RBF can significantly outperform HOTV in most cases and outperform other matrix completion methods in all cases except image 1 with block mask.

Figs. 6–10 give a few examples of the inpainting results. In Fig. 6, HOTV fails absolutely and NLMC-RBF slightly outperforms other methods. In Figs. 7–10, the recovery performances of NLMC-RBF are significantly better than that of other methods. To study the influence of p (Schatten p -norm) in IRNN and NLMC, Fig. 11 shows the MC errors of IRNN and NLMC with different p ($0.2 \leq p \leq 1$) for image 7 with text (large) mask and image 8 with random (80%) mask (the MC errors of other methods are shown by dashed lines). Considering the results in Figs. 3 and 11, we deduce that, in IRNN and NLMC, an appropriate p smaller than 1 may outperform $p = 1$ but the best value of p depends on the case.

We also consider three benchmark images for object removal [10,11] and a classical exemplar [10] based method is also compared. The inpainting results are shown by Figs. 12–14. In Fig. 12, as the image matrix is of low-rank, matrix completion methods give much better results than HOTV and Exemplar do. In Fig. 13, Exemplar and NLMC-RBF outperform other methods. In Fig. 14,

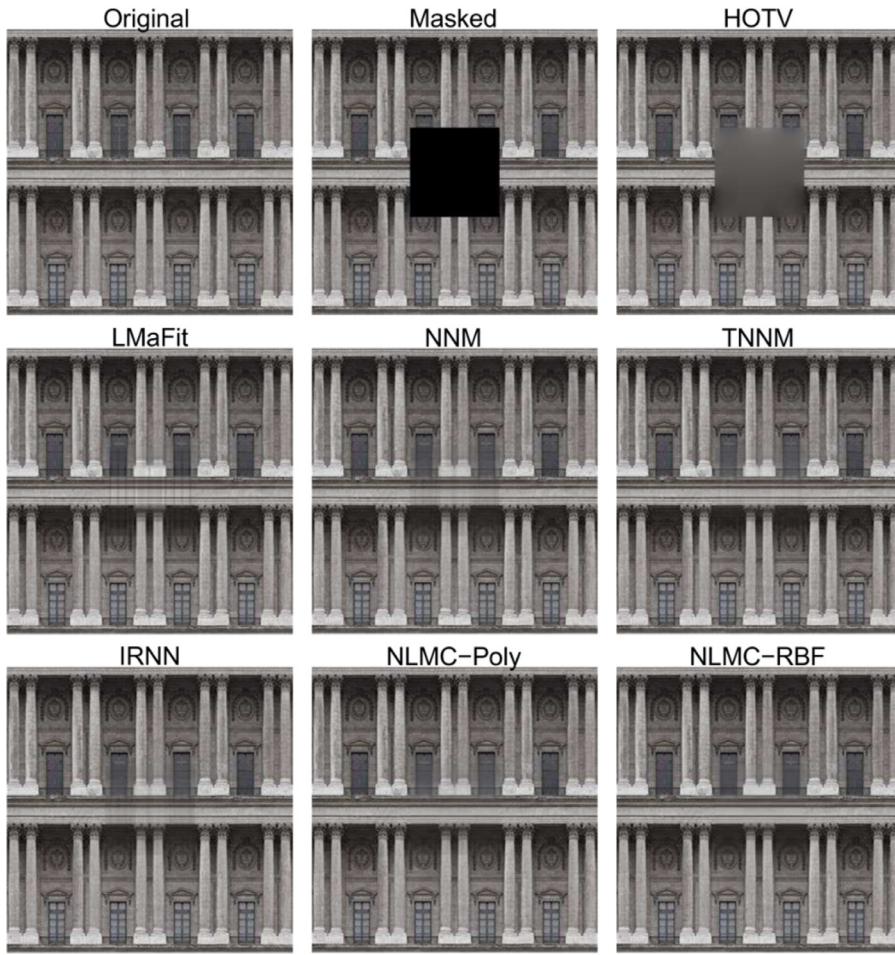


Fig. 6. Inpainting performance for image 2 with block mask.

Table 3
Classification errors (%) on the MNIST digit dataset.

θ	LMaFit	NNM	TNNM	IRNN	C-NLMC	NLMC-Poly	NLMC-RBF	SVM-poly	SVM-RBF	CNN
30%	24.83	20.81	20.71	20.76	16.13	12.21	12.29	14.11	12.56	12.38
50%	23.34	18.51	18.48	18.42	14.24	9.54	9.61	11.84	9.82	11.72
70%	21.33	17.92	18.03	17.93	13.06	8.73	8.59	10.4	8.71	8.83

only NLMC-RBF can remove the object and reconstruct the background successfully.

4.3. Single-/multi-label classification

First, we consider the MC-based single-label classification problem [12] on the MNIST handwritten digit database [59], which has 70,000 examples of handwritten digits (0 – 9). To reduce the computational costs for all related methods, we randomly choose 100 images for each digits and resize the images to 16×16 . It is known that deep learning methods such as convolutional neural network (CNN) can give nearly 100% classification accuracies results on the MNIST dataset under the condition that the number of training samples is huge. If the training set is very small, the performance of deep learning will deteriorate significantly. In this study, NLMC is also compared with SVM and CNN (LeNet). Different percentages (i.e., $\theta = 30\%$, 50% , and 70%) of the MNIST data are used as training data and the remaining data are testing data. The parameters of all methods are adjusted to give their best performances. The average classification errors of 50 repeated trials are reported in Tables 3. As can be seen, the proposed NLMC outperformed all

other methods. There is no doubt that deep learning often give the best performance if the training set is large enough. But for small training set, NLMC could be a choice.

MC-based single-label classification is also conducted on the Extended Yale Face Database B [60], which consists of the face images of 38 subjects. There are about 64 images of various illuminations for each subject. We choose the images of the first 10 subjects and resize the images to 32×32 . In [61] and [62], sparse representation based classification (SRC) was able to provide nearly 98% recognition rate on the Extended Yale B data. Hence, in this study, NLMC is also compared with SRC. We randomly choose 50% of the images of each subject as training data to classify the remaining images. We consider five different cases. In the first case, similar with [62], we use 300-dimension eigenface as features. In the second case, we resize the images to 16×16 and use the pixels as features. In cases 3, 4, and 5, we resize the images to 16×16 and randomly remove 10%, 30%, and 50% of the pixels. Therefore, cases 3, 4, and 5 are incomplete-data classification tasks. For SVM and SRC, the missing entries are filled with zeros. Similar with [62], the features are further normalized to have unit ℓ_2 -norm. In matrix completion based classification, to balance the labels and

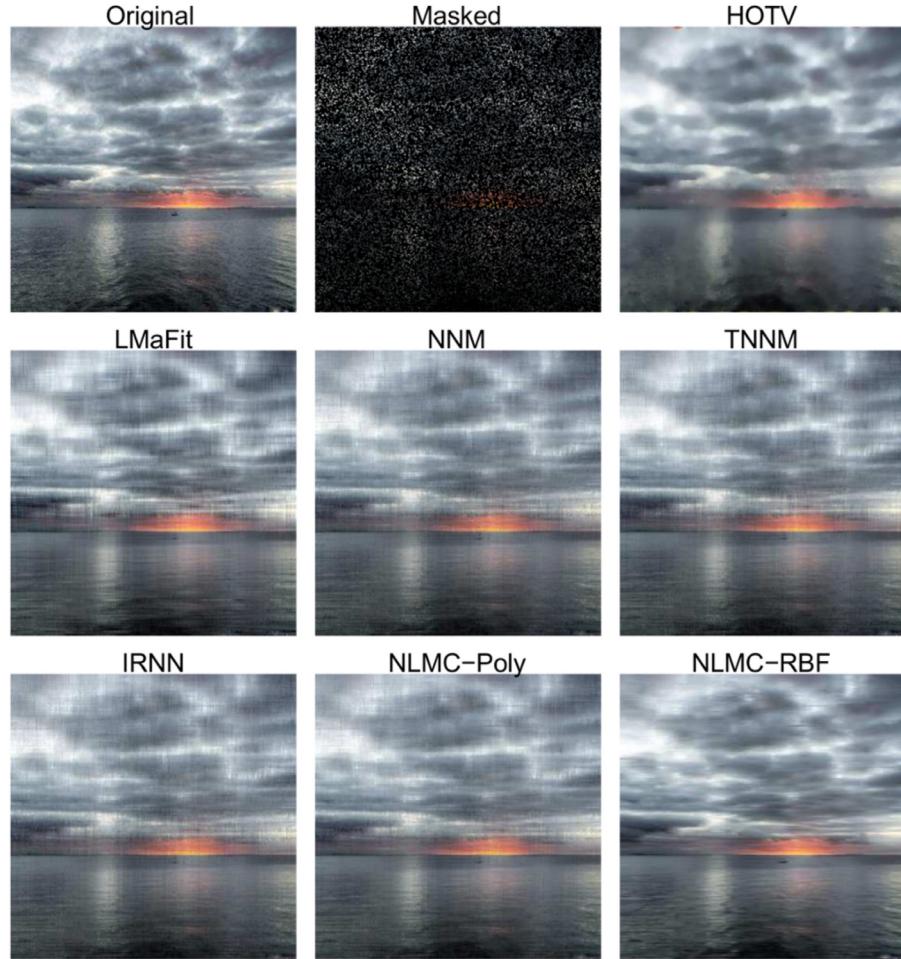


Fig. 7. Inpainting performance for image 5 with random (80%) mask.

Table 4

Classification errors (%) on the Extended Yale Face B dataset with different features (EigenFaces, pixels, and incomplete pixels).

Feature	LMaFit	NNM	TNNM	IRNN	C-NLMC	NLMC-Poly	NLMC-RBF	SVM-poly	SVM-RBF	SRC
EigenFace 300	1.73	1.52	1.49	1.58	3.16	1.39	1.34	2.02	2.09	1.67
Pixel 16×16 (complete)	8.06	4.06	4.02	4.08	10.75	2.42	2.42	3.79	3.84	2.46
Pixel 16×16 (10%missing)	12.35	4.09	4.05	4.12	18.34	3.12	3.19	11.25	12.03	14.32
Pixel 16×16 (30%missing)	19.51	7.94	7.62	8.01	62.92	4.96	4.88	45.94	45.64	50.25
Pixel 16×16 (50%missing)	28.67	10.13	9.86	10.25	76.34	5.81	5.56	68.33	68.02	69.37

features in the incomplete matrices, the features are multiplied by 10. The parameters of all methods are adjusted to give their best performances. The average classification errors of 50 repeated trials are reported in Table 4. NLMC is much better than LMC, SVM, and SRC in all cases. Particularly, for incomplete data, SVM and SRC result in significant errors even if the missing rate is 10%. When the missing rate is 50%, NLMC still give very small errors (< 6%). The main purpose of this single-label classification study is not to challenge the advanced classifiers (e.g. deep learning and SRC) but to verify that NLMC is much more effective than LMC and C-NLMC and considering nonlinearity in matrix completion is really important and useful.

We compare NLMC with LMC and C-NLMC on multi-label classification problem [13,55]. Two conventional multi-label classifiers, ML-kNN [41] and ML-RBF [63], are also considered in this study. It has been reported that these two multi-label classifiers outperformed several other methods such as rank-SVM [41,63]. The first multi-label dataset utilized in this study is the Yeast gene functional data. The Yeast data given by [63] contains 2417 genes and

each gene has a 103-dimensional feature vector. The number of the possible class labels is 14 and the average number of labels for each gene is 4.24 ± 1.57 . The second multi-label dataset considered in this study is a natural scene database consisting of 2000 natural scene images, where all the possible classes are desert, mountains, sea, sunset and trees. The number of images belonging to more than one class is about 22% of the data set and the average number of labels for each image is 1.24 ± 0.44 . Each image is represented by a 294-dimensional feature vector [63]. We use two popular evaluation metrics of multi-label classification: Hamming loss and one error [41]. The Hamming loss measures the times an instance-label pair is misclassified, and the one error measures how many times the top-ranked label is not in the set of proper labels of the instance [41]. For the two metrics, the smaller the better. For each dataset, we randomly choose a subset consisting of 1000 samples and use 30%, 50%, and 70% of the samples to predict the labels of the other samples. The parameters of all methods are adjusted to the best values as possible. The experiments are repeated for 20 times and the average results of the two metrics

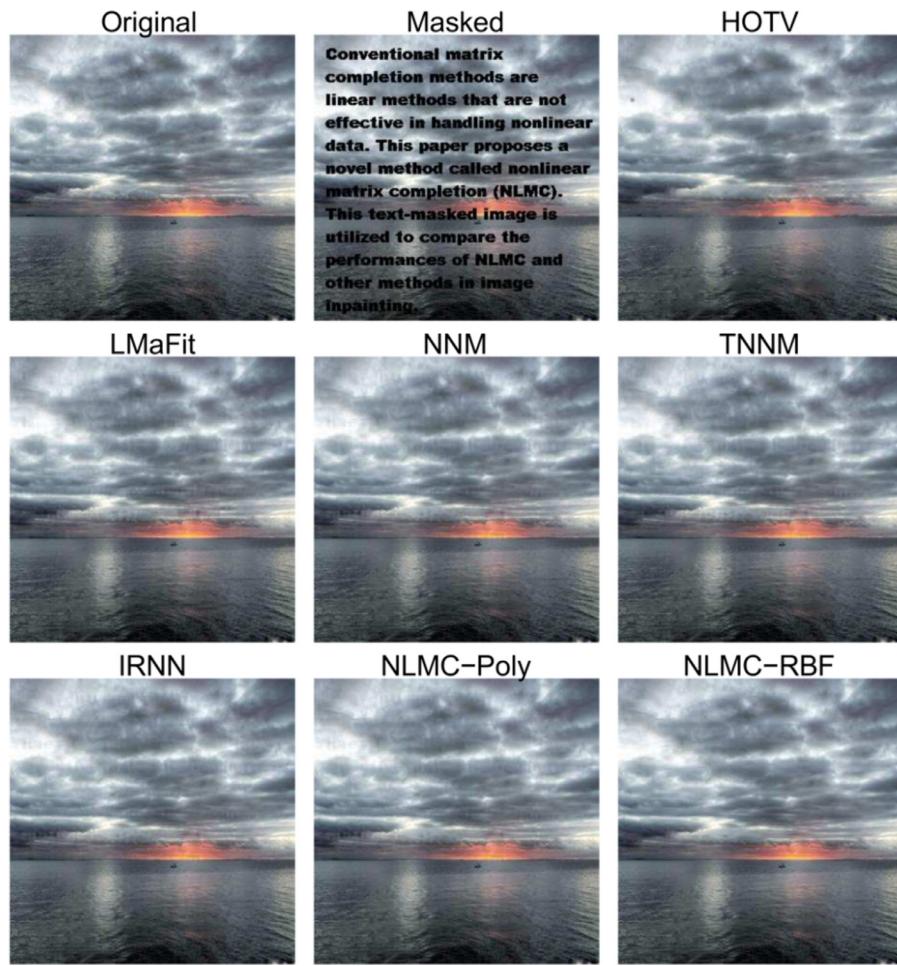


Fig. 8. Inpainting performance for image 5 with text (small) mask.

Table 5
Multi-label classification results for the Yeast dataset.

θ	metric	LMaFit	NNM	TNNM	IRNN	C-NLMC	NLMC-Poly	NLMC-RBF	ML-kNN	ML-RBF
30%	Hamming loss	0.372	0.321	0.325	0.318	0.242	0.206	0.207	0.214	0.253
	One error	0.112	0.091	0.089	0.091	0.176	0.125	0.152	0.136	0.134
50%	Hamming loss	0.316	0.269	0.272	0.270	0.236	0.196	0.198	0.210	0.249
	One error	0.128	0.095	0.094	0.095	0.197	0.109	0.134	0.131	0.132
70%	Hamming loss	0.257	0.234	0.237	0.234	0.224	0.192	0.195	0.203	0.238
	One error	0.121	0.102	0.099	0.107	0.183	0.123	0.143	0.139	0.135

Table 6
Multi-label classification results for the Scene dataset.

θ	metric	LMaFit	NNM	TNNM	IRNN	C-NLMC	NLMC-Poly	NLMC-RBF	ML-kNN	ML-RBF
30%	Hamming loss	0.202	0.197	0.197	0.198	0.270	0.182	0.175	0.188	0.172
	One error	0.131	0.122	0.126	0.122	0.294	0.085	0.059	0.166	0.183
50%	Hamming loss	0.202	0.189	0.190	0.192	0.182	0.172	0.162	0.181	0.162
	One error	0.119	0.113	0.115	0.114	0.209	0.082	0.072	0.183	0.162
70%	Hamming loss	0.198	0.187	0.188	0.187	0.170	0.160	0.150	0.171	0.152
	One error	0.104	0.105	0.107	0.105	0.188	0.072	0.074	0.167	0.144

are reported in [Table 5](#) and [Table 6](#). As can be seen, NLMC-Poly and NLMC-RBF significantly outperform other methods in most cases.

The above single-/multi-label classification results have verified that the proposed NLMC methods are more effective than LMC methods in classification tasks. The reason is that the relationship between features and labels are often nonlinear and NLMC meth-

ods can implicitly construct nonlinear classifiers while LMC methods implicitly construct linear classifiers.

4.4. Image denoising

In this subsection, the proposed NLRMDN will be compared with PCA and KPCA in an image de-noising task. For KPCA, two



Fig. 9. Inpainting performance for image 7 with text (large) mask.

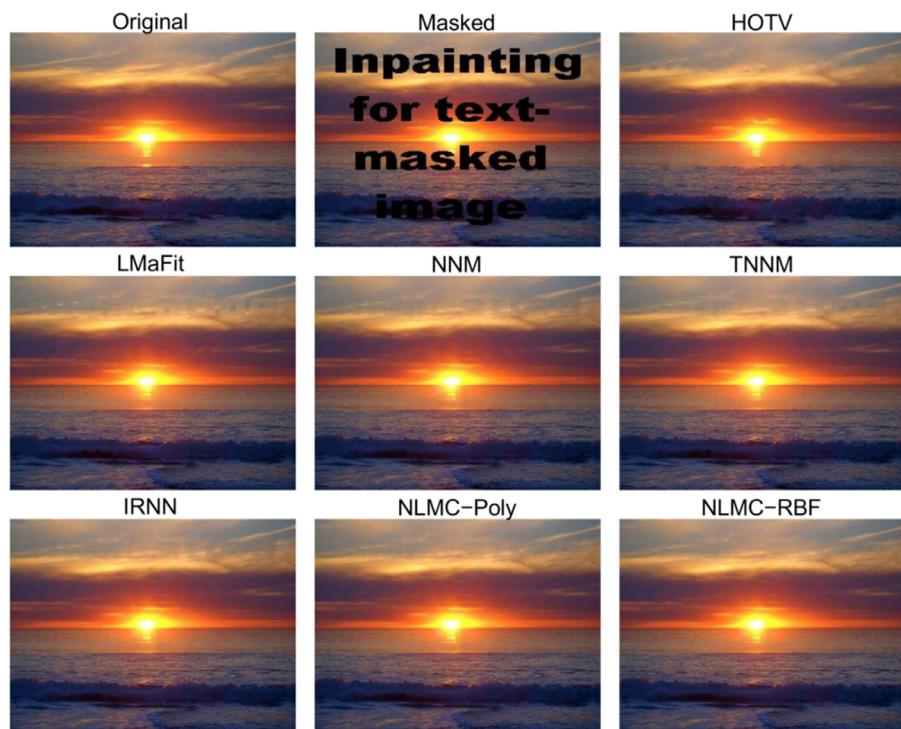


Fig. 10. Inpainting performance for image 8 with text (large) mask.

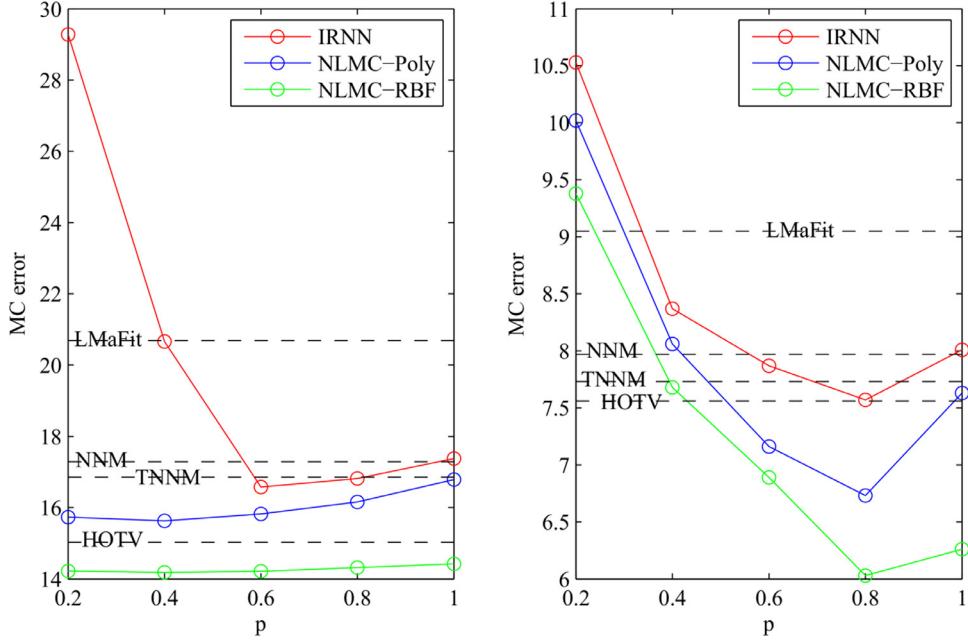


Fig. 11. MC errors (%) of IRNN and NLNC with different p (left: image 7 with text (large) mask; right: image 8 with random (80%) mask).

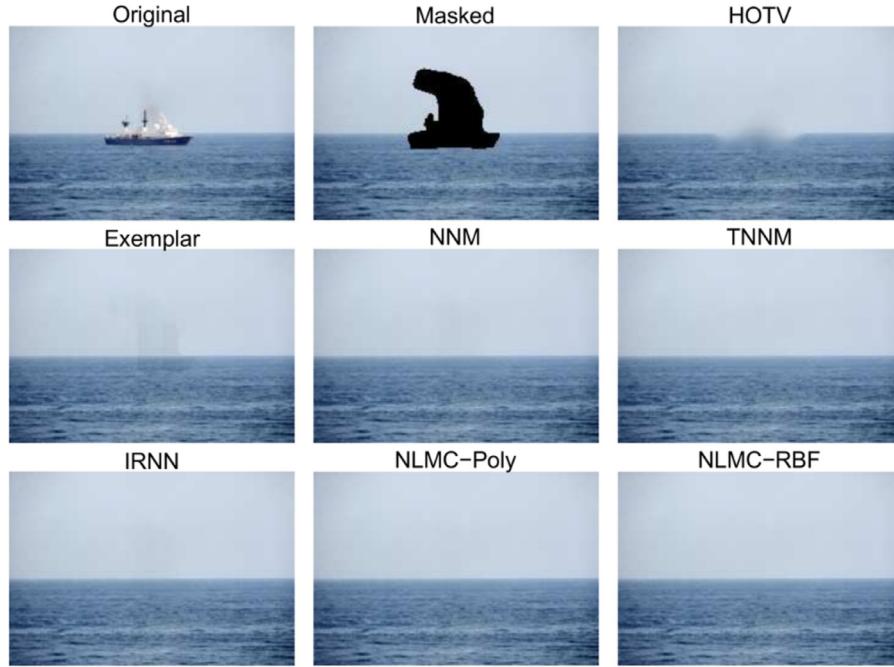


Fig. 12. Inpainting performance for the image with ship masked.

pre-image methods are considered. The first one is the method of Mika et al.[44] with only RBF kernel because the fixed-point iteration method is inapplicable to polynomial kernel. Another one is the method of Kwok et al.[45] with polynomial kernel and RBF kernel. Be similar with the experiments in [44,45], we also utilize the USPS database that consists of 16×16 handwritten digits (0–9). We randomly choose 100 examples of each digit and add a Gaussian noise with zero mean and 0.25 variance to each image. NLRMDN, PCA, and KPCA are performed on each digit separately. The numbers of principal components in PCA and KPCA are chosen from $\{2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6\}$ and the numbers perform best are reported finally. In NLRMDR, for polynomial kernel,

the parameter b is set as 10^4 and λ is set as 20; for RBF kernel, the parameter σ^2 is set as 150 and λ is set as 0.01. The number of neighbors in the KPCA pre-image of Kwok et al. is set as 15. The kernel parameters for KPCA are also carefully set to provide best performances. Fig. 15 shows examples of the original digits, noisy digits, and denoised digits given by the six methods. The experiments are repeated for 50 times and the average results in terms of SNR are shown in Table 7, where the largest SNR for each case is highlighted with bold type. As can be seen, for digits 0,1,6,7,8,9, NLRMDN with RBF kernel provides highest SNRs, while for digits 2,3,4,5,8, NLRMDN with polynomial kernel gives highest SNRs.

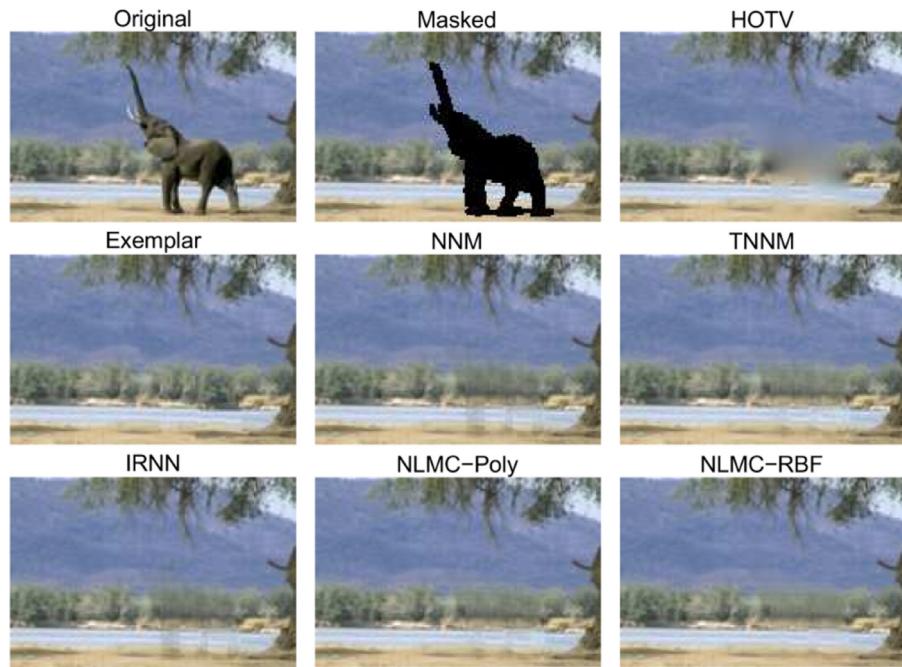


Fig. 13. Inpainting performance for the image with *elephant* masked.

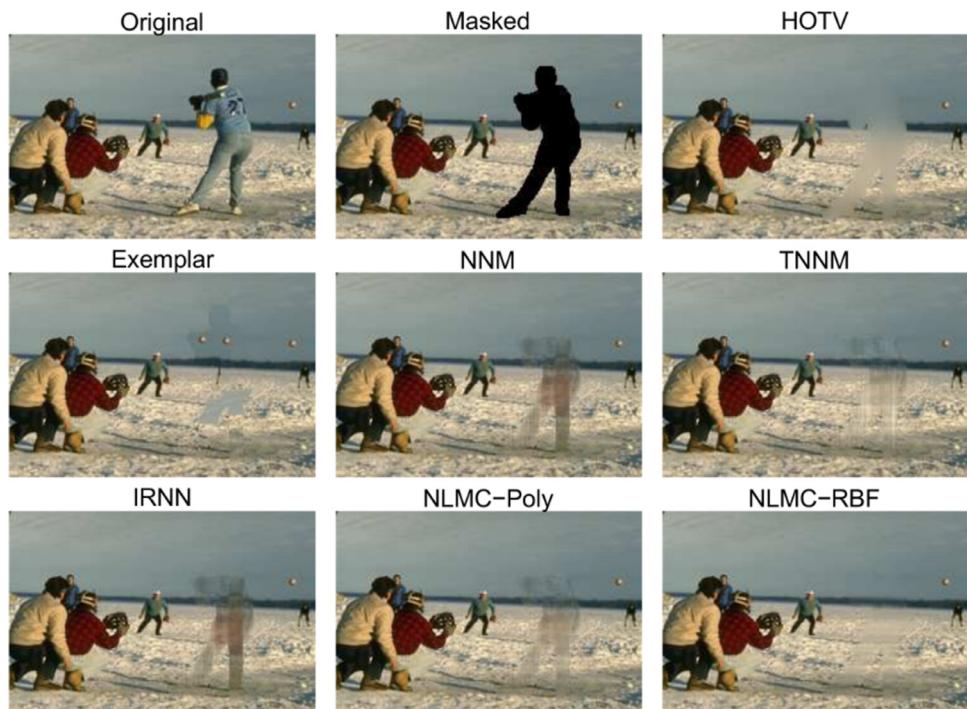


Fig. 14. Inpainting performance for the image with *man* masked.



Fig. 15. Denoising for USPS samples (row 1: original images; row 2: noisy images; rows 3–8: denoised images given by PCA, KPCA-RBF pre-image of Mika et al., KPCA-Poly and KPCA-RBF pre-image of Kwok et al., NLRMDN-Poly, and NLRMDN-RBF respectively)

Table 7

SNRs (dB) of noisy and denoised images of USPS dataset.

	0	1	2	3	4	5	6	7	8	9	average
Noisy image	1.64	2.47	1.77	1.74	2.02	1.75	1.91	2.12	1.78	2.01	1.92
PCA	4.48	7.78	3.78	4.14	4.50	3.93	4.86	5.28	4.25	5.24	4.83
KPCA-RBF pre-image (Mika et al.)	4.66	7.61	4.29	4.52	4.88	4.42	5.09	5.58	4.61	5.50	5.12
KPCA-Poly pre-image (Kwok et al.)	3.95	6.62	3.38	3.74	4.10	3.51	4.42	4.84	3.85	4.77	4.32
KPCA-RBF pre-image (Kwok et al.)	4.43	7.16	4.02	4.19	4.60	3.98	4.93	5.39	4.34	5.32	4.84
NLRMDN-Poly	4.70	7.10	4.33	4.61	4.91	4.45	5.13	5.51	4.70	5.44	5.09
NLRMDN-RBF	4.73	7.85	4.21	4.57	4.90	4.36	5.22	5.65	4.70	5.60	5.18

5. Conclusion

This paper proposed a new method NLMC to solve the matrix completion problem for data of non-linear structures, which is beyond the scope of conventional matrix completion methods because they are linear methods in general. The comparative studies on a nonlinear toy dataset, image inpainting, and single-/multi-label classification validated the effectiveness and superiority of NLMC. Particularly, in image inpainting, the performances of NLMC were much better than that of LMC methods, HOTV, and Exemplar based inpainting methods. In MC-based classification, NLMC outperformed all LMC methods, C-NLMC, SVM, SRC, CNN, ML-kNN, and ML-RBF significantly. The idea of NLMC was extended to a general non-linear rank-minimization framework that can be used to solve other problems such as non-linear denoising. The proposed NLRMDN outperformed PCA and KPCA pre-image methods in image denoising.

References

- [1] E.J. Candes, B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.* 9 (6) (2009) 717–772, doi: [10.1007/s10208-009-9045-5](https://doi.org/10.1007/s10208-009-9045-5).
- [2] Y. Liu, L. Jiao, F. Shang, A fast tri-factorization method for low-rank matrix recovery and completion, *Pattern Recognit.* 46 (1) (2013) 163–173, doi: [10.1016/j.patcog.2012.07.003](https://doi.org/10.1016/j.patcog.2012.07.003).
- [3] Y. Cao, Y. Xie, Poisson matrix recovery and completion, *IEEE Trans. Signal Process.* 64 (6) (2016) 1609–1620, doi: [10.1109/TSP.2015.2500192](https://doi.org/10.1109/TSP.2015.2500192).
- [4] J. Fan, T. Chow, Deep learning based matrix completion, *Neurocomputing* 266 (Supplement C) (2017) 540–549, doi: [10.1016/j.neucom.2017.05.074](https://doi.org/10.1016/j.neucom.2017.05.074).
- [5] B. Wu, S. Lyu, B.-G. Hu, Q. Ji, Multi-label learning with missing labels for image annotation and facial action unit recognition, *Pattern Recognit.* 48 (7) (2015) 2279–2289, doi: [10.1016/j.patcog.2015.01.022](https://doi.org/10.1016/j.patcog.2015.01.022).
- [6] Z. ga Liu, Q. Pan, J. Dezert, A. Martin, Adaptive imputation of missing values for incomplete pattern classification, *Pattern Recognit.* 52 (2016) 85–95, doi: [10.1016/j.patcog.2015.10.001](https://doi.org/10.1016/j.patcog.2015.10.001).
- [7] J. Fan, T.W. Chow, Sparse subspace clustering for data with missing entries and high-rank matrix completion, *Neural Netw.* 93 (2017) 36–44, doi: [10.1016/j.neunet.2017.04.005](https://doi.org/10.1016/j.neunet.2017.04.005).
- [8] X. Su, T.M. Khoshgoftaar, A survey of collaborative filtering techniques, *Adv. Artif. Intell.* 2009 (2009) 4:2–4:2, doi: [10.1155/2009/421425](https://doi.org/10.1155/2009/421425).
- [9] Y. Hu, D. Zhang, J. Ye, X. Li, X. He, Fast and accurate matrix completion via truncated nuclear norm regularization, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (9) (2013) 2117–2130, doi: [10.1109/TPAMI.2012.271](https://doi.org/10.1109/TPAMI.2012.271).
- [10] C. Guillemot, O.L. Meur, Image inpainting : overview and recent advances, *IEEE Signal Process. Mag.* 31 (1) (2014) 127–144, doi: [10.1109/MSP.2013.2273004](https://doi.org/10.1109/MSP.2013.2273004).
- [11] K.H. Jin, J.C. Ye, Annihilating filter-based low-rank Hankel matrix approach for image inpainting, *IEEE Trans. Image Process.* 24 (11) (2015) 3498–3511, doi: [10.1109/TIP.2015.2446943](https://doi.org/10.1109/TIP.2015.2446943).
- [12] A. Goldberg, B. Recht, J. Xu, R. Nowak, X. Zhu, *Transduction with matrix completion: three birds with one stone*, in: *Advances in Neural Information Processing Systems 23*, Curran Associates, Inc., 2010, pp. 757–765.
- [13] Y. Luo, T. Liu, D. Tao, C. Xu, Multiview matrix completion for multilabel image classification, *IEEE Trans. Image Process.* 24 (8) (2015) 2355–2368, doi: [10.1109/TIP.2015.2421309](https://doi.org/10.1109/TIP.2015.2421309).
- [14] J. Ma, Z. Tian, H. Zhang, T.W. Chow, Multi-label low-dimensional embedding with missing labels, *Knowl. Based Syst.* (2017), doi: [10.1016/j.knosys.2017.09.005](https://doi.org/10.1016/j.knosys.2017.09.005).
- [15] Y. Chen, S. Bhojanapalli, S. Sanghavi, R. Ward, Coherent matrix completion, in: *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, JMLR Workshop and Conference Proceedings, 2014, pp. 674–682.
- [16] S. Bhojanapalli, P. Jain, Universal matrix completion, in: *Proceedings of The 31st International Conference on Machine Learning, JMLR, 2014*, pp. 1881–1889.
- [17] G. Liu, P. Li, Low-rank matrix completion in the presence of high coherence, *IEEE Trans. Signal Process.* PP (99) (2016), doi: [10.1109/TSP.2016.2586753](https://doi.org/10.1109/TSP.2016.2586753). 1–1
- [18] N. Srebro, *Learning with Matrix Factorizations*, 2004 Ph.D. thesis. Cambridge, MA, USA.
- [19] Y. Shen, Z. Wen, Y. Zhang, Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization, *Optim. Methods Softw.* 29 (2) (2014) 239–263, doi: [10.1080/10556788.2012.700713](https://doi.org/10.1080/10556788.2012.700713).
- [20] Z. Wen, W. Yin, Y. Zhang, Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm, *Math. Program. Comput.* 4 (4) (2012) 333–361, doi: [10.1007/s12532-012-0044-1](https://doi.org/10.1007/s12532-012-0044-1).
- [21] K.-C. Toh, S. Yun, An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems, *Pacific J. Optim.* (2010).
- [22] C. Chen, B. He, X. Yuan, Matrix completion via an alternating direction method, *IMA J. Numer. Anal.* (2011), doi: [10.1093/imanum/drq039](https://doi.org/10.1093/imanum/drq039).
- [23] F. Nie, H. Huang, C. Ding, Low-rank matrix recovery via efficient Schatten p-norm minimization, in: *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, in: *AAAI'12*, AAAI Press, 2012, pp. 655–661.
- [24] F. Nie, H. Wang, H. Huang, C. Ding, Joint Schatten p-norm and lp-norm robust matrix completion for missing value recovery, *Knowl. Inf. Syst.* 42 (3) (2015) 525–544, doi: [10.1007/s10115-013-0713-z](https://doi.org/10.1007/s10115-013-0713-z).
- [25] Q. Liu, Z. Lai, Z. Zhou, F. Kuang, Z. Jin, A truncated nuclear norm regularization method based on weighted residual error for matrix completion, *IEEE Trans. Image Process.* 25 (1) (2016) 316–330.
- [26] C. Lu, J. Tang, S. Yan, Z. Lin, Generalized nonconvex nonsmooth low-rank minimization, in: *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 4130–4137, doi: [10.1109/CVPR.2014.526](https://doi.org/10.1109/CVPR.2014.526).
- [27] J. Fan, T.W. Chow, Matrix completion by least-square, low-rank, and sparse self-representations, *Pattern Recognit.* 71 (2017) 290–305, doi: [10.1016/j.patcog.2017.05.013](https://doi.org/10.1016/j.patcog.2017.05.013).
- [28] E.J. Candes, X. Li, Y. Ma, J. Wright, Robust principal component analysis? *J. ACM* 58 (3) (2011) 1–37, doi: [10.1145/1970392.1970395](https://doi.org/10.1145/1970392.1970395).
- [29] J. Fan, T.W.S. Chow, M. Zhao, J.K.L. Ho, Nonlinear dimensionality reduction for data with disconnected neighborhood graph, *Neural Process. Lett.* (2017), doi: [10.1007/s11063-017-9676-5](https://doi.org/10.1007/s11063-017-9676-5).
- [30] B. Schlkopf, A. Smola, K.-R. Mller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.* 10 (5) (1998) 1299–1319, doi: [10.1162/089976698300017467](https://doi.org/10.1162/089976698300017467).
- [31] T. Zhou, H. Shan, A. Banerjee, G. Sapiro, Kernelized Probabilistic Matrix Factorization: Exploiting Graphs and Side Information, pp. 403–414.
- [32] S. Si, K.-Y. Chiang, C.-J. Hsieh, N. Rao, I.S. Dhillon, Goal-directed inductive matrix completion, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in: *KDD '16*, ACM, New York, NY, USA, 2016, pp. 1165–1174, doi: [10.1145/2939672.2939809](https://doi.org/10.1145/2939672.2939809).
- [33] X. Liu, C. Aggarwal, Y.-F. Li, X. Kong, X. Sun, S. Sathe, Kernelized matrix factorization for collaborative filtering, in: *SIAM Conference on Data Mining*, 2016, pp. 399–416.
- [34] X. Alameda-Pineda, E. Ricci, Y. Yan, N. Sebe, Recognizing emotions from abstract paintings using non-linear matrix completion, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5240–5248.
- [35] P. Jain, I.S. Dhillon, Provable inductive matrix completion, *arXiv preprint arXiv:1306.0626* (2013).
- [36] A. Papaiannou, S. Zafeiriou, Principal component analysis with complex kernel: the widely linear model, *IEEE Trans. Neural Netw. Learn. Syst.* 25 (9) (2014) 1719–1726, doi: [10.1109/TNNLS.2013.2285783](https://doi.org/10.1109/TNNLS.2013.2285783).
- [37] J. Fan, S.J. Qin, Y. Wang, Online monitoring of nonlinear multivariate industrial processes using filtering KICAPCA, *Control Eng. Pract.* 22 (2014) 205–216, doi: [10.1016/j.conengprac.2013.06.017](https://doi.org/10.1016/j.conengprac.2013.06.017).
- [38] D.C. Liu, J. Nocedal, On the limited memory BFGS method for large scale optimization, *Math. Program.* 45 (1) (1989) 503–528, doi: [10.1007/BF01589116](https://doi.org/10.1007/BF01589116).
- [39] Z. Lin, M. Chen, Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, *arXiv:1009.5055v3 [math.OC]* (2010).
- [40] Y. Li, B. Wu, B. Ghahem, Y. Zhao, H. Yao, Q. Ji, Facial action unit recognition under incomplete data based on multi-label learning with missing labels, *Pattern Recognit.* 60 (2016) 890–900, doi: [10.1016/j.patcog.2016.07.009](https://doi.org/10.1016/j.patcog.2016.07.009).
- [41] M.-L. Zhang, Z.-H. Zhou, ML-knn: a lazy learning approach to multi-label learning, *Pattern Recognit.* 40 (7) (2007) 2038–2048, doi: [10.1016/j.patcog.2006.12.019](https://doi.org/10.1016/j.patcog.2006.12.019).
- [42] Y. Luo, D. Tao, C. Xu, C. Xu, H. Liu, Y. Wen, Multiview vector-valued manifold regularization for multilabel image classification, *IEEE Trans. Neural Netw. Learn. Syst.* 24 (5) (2013) 709–722, doi: [10.1109/TNNLS.2013.2238682](https://doi.org/10.1109/TNNLS.2013.2238682).

- [43] J. Ma, T.W. Chow, Robust non-negative sparse graph for semi-supervised multi-label learning with missing labels, *Inf. Sci. (Ny)* 422 (Supplement C) (2018) 336–351, doi:[10.1016/j.ins.2017.08.061](https://doi.org/10.1016/j.ins.2017.08.061).
- [44] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Schölkopf, G. Ratsch, Kernel PCA and de-noising in feature spaces, in: *Advances in Neural Information Processing Systems 11*, MIT Press, 1999, pp. 536–542.
- [45] J.T.Y. Kwok, I.W.H. Tsang, The pre-image problem in kernel methods, *IEEE Trans. Neural Netw.* 15 (6) (2004) 1517–1525, doi:[10.1109/TNN.2004.837781](https://doi.org/10.1109/TNN.2004.837781).
- [46] M. Kallas, P. Honeine, C. Richard, C. Francis, H. Amoud, Non-negativity constraints on the pre-image for pattern recognition with kernel machines, *Pattern Recognit.* 46 (11) (2013) 3066–3080, doi:[10.1016/j.patcog.2013.03.021](https://doi.org/10.1016/j.patcog.2013.03.021).
- [47] G. Ongie, R. Willett, R.D. Nowak, L. Balzano, Algebraic variety models for high-rank matrix completion, *arXiv preprint arXiv:1703.09631*(2017).
- [48] J.-F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, *SIAM J. Optim.* 20 (4) (2010) 1956–1982, doi:[10.1137/080738970](https://doi.org/10.1137/080738970).
- [49] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Mach. Learn.* 3 (1) (2011) 1–122, doi:[10.1561/2200000016](https://doi.org/10.1561/2200000016).
- [50] L. Liu, W. Huang, D.-R. Chen, Exact minimum rank approximation via Schatten p-norm minimization, *J. Comput. Appl. Math.* 267 (2014) 218–227, doi:[10.1016/j.cam.2014.02.015](https://doi.org/10.1016/j.cam.2014.02.015).
- [51] E.J. Candès, Y. Plan, Matrix completion with noise, *Proc. IEEE* 98 (6) (2010) 925–936, doi:[10.1109/JPROC.2009.2035722](https://doi.org/10.1109/JPROC.2009.2035722).
- [52] Y. Chen, H. Xu, C. Caramanis, S. Sanghavi, Matrix completion with column manipulation: near-optimal sample-robustness-rank tradeoffs, *ArXiv e-prints* (2011).
- [53] O. Klopp, K. Lounici, A.B. Tsybakov, Robust matrix completion, *ArXiv e-prints* (2014).
- [54] W.S. Zheng, J. Lai, P.C. Yuen, Penalized preimage learning in kernel principal component analysis, *IEEE Trans. Neural Netw.* 21 (4) (2010) 551–570, doi:[10.1109/TNN.2009.2039647](https://doi.org/10.1109/TNN.2009.2039647).
- [55] R. Cabral, F.D.I. Torre, J.P. Costeira, A. Bernardino, Matrix completion for weakly-supervised multi-label image classification, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (1) (2015) 121–135, doi:[10.1109/TPAMI.2014.2343234](https://doi.org/10.1109/TPAMI.2014.2343234).
- [56] M. Schmidt, minFunc: unconstrained differentiable multivariate optimization in Matlab, 2005, (<http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html>).
- [57] C.-B. Schönlieb, A. Bertozzi, Unconditionally stable schemes for higher order inpainting, *Commun. Math. Sci.* 9 (2) (2011) 413–457.
- [58] M. Benning, C. Brune, M. Burger, J. Müller, Higher-order TV method—enhancement via Bregman iteration, *J. Sci. Comput.* 54 (2) (2013) 269–310, doi:[10.1007/s10915-012-9650-3](https://doi.org/10.1007/s10915-012-9650-3).
- [59] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324, doi:[10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [60] L. Kuang-Chih, J. Ho, D.J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (5) (2005) 684–698, doi:[10.1109/TPAMI.2005.92](https://doi.org/10.1109/TPAMI.2005.92).
- [61] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
- [62] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition? in: *Computer Vision (ICCV), 2011 IEEE International Conference on*, IEEE, 2011, pp. 471–478.
- [63] M.-L. Zhang, MI-RBF: RBF neural networks for multi-label learning, *Neural Process. Lett.* 29 (2) (2009) 61–74, doi:[10.1007/s11063-009-9095-3](https://doi.org/10.1007/s11063-009-9095-3).

Jicong Fan received his B.E and M.E degrees in Automation and Control Science & Engineering, from Beijing University of Chemical Technology, Beijing, P.R., China, in 2010 and 2013, respectively. From 2013 to 2015, he was a research assistant at the University of Hong Kong and focused on signal processing and data analysis for neuroscience. Currently, he is working toward the PhD degree at the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong S.A.R. His research interests include data mining and machine learning.

Tommy W. S. Chow received the B.Sc. (1st Hons) degree and the Ph.D. degree from the Department of Electrical and Electronic Engineering, University of Sunderland, U.K. He is currently a Professor in the Department of Electronic Engineering at the City University of Hong Kong. His main research areas include neural networks, machine learning, pattern recognition, and fault diagnosis. He received the Best Paper Award in 2002 IEEE Industrial Electronics Society Annual meeting in Seville, Spain. He is an author and co-author of over 170 technical Journal articles related to his research, 5 book chapters, and 1 book. He now serves the IEEE Transactions on Industrial Informatics and Neural Processing letters as Associate editor.