An iterative algorithm for singular value decomposition on noisy incomplete matrices

KyungHyun Cho and Nima Reyhani
Department of Information and Computer Science,
Aalto University School of Science
Email: firstname.lastname@aalto.fi

***Abstract*—In this paper, we propose a simple iterative algorithm, called iSVD, for estimating the singular value decomposition (SVD) of a noisy incomplete given matrix. The iSVD relies on first order optimization over orthogonal manifolds and automatically estimates the rank of the SVD. The main goal here is to estimate the singular vectors through optimization in the right space, which is the space of the orthogonal matrix manifolds. The rank estimation is based on the ratio between estimated large singular values and the sum of all singular values. We empirically evaluate the iSVD on synthetic matrices and image reconstruction tasks. The evaluation shows that the iSVD is comparable to the recently introduced methods for matrix completion such as singular value thresholding (SVT) and fixed-point iteration with approximate SVD (FPCA).**

I. Singular Value Decomposition, Missing Values and Noisy Observation

Singular value decomposition (SVD) has many applications in engineering and data analysis such as dimension reduction and low-rank approximation of the given matrix, e.g. [1].

There are a number of efficient algorithms for computing SVD, e.g. [1], [2]. These algorithms assume all elements of the given matrix is provided. In applications there are situations where the matrix elements are contaminated with additional noise and some of the values are missing. The additional noise likely makes the decomposition harder by decreasing the condition number. In such cases, it is of significant importance to estimate singular vectors which are not modeling the additional noise.

Using SVD one can approximates a given matrix $\mathbf{X}$ by

$$\mathbf{X}_{m\times n} \approx \mathbf{U}_{m\times k}\mathbf{S}_{k\times k}\mathbf{V}_{n\times k}^{\top},$$

where $A^{\top}$ is the transpose of the matrix A, $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices of which each column corresponds to left and right singular vectors, respectively. $\mathbf{S}$ is a diagonal matrix whose diagonal elements are non-negative real singular values. This gives a low-rank approximation of the given matrix $\mathbf{X}$ with a desired rank k.

In this paper, we assume that the given matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ is available and some elements of $\mathbf{X}$ are unknown with probability $1-p$. A set Ω is a set of indices of the observed element (i,j) of $\mathbf{X}$. The goal of a SVD algorithm in this case is to; (1) complete all missing elements of $\mathbf{X}$, (2) estimate orthogonal matrices $\mathbf{U} \in \mathbb{R}^{m\times k}$, $\mathbf{V} \in \mathbb{R}^{n\times k}$ and a diagonal $\mathbf{S} \in \mathbb{R}_{+}^{k\times k}$. In this paper, we also require that the algorithm estimates the rank k, which we refer to this by (3).

Estimating a low rank approximation of incomplete and noisy observations is a new line of research and found many applications in, e.g., collaborative filtering [3, and references therein][4] and Bioinformatics [5], [6]. Under some technical conditions, it is shown that exact recovery is possible as usually the intrinsic rank is small [7].

Here, we shortly introduce two methods proposed recently for solving the incomplete SVD problem. In [7] singular value thresholding (SVT) was introduced that is formulated as follows:

$$\begin{aligned} &\text{minimize} \quad \tau\|\mathbf{Y}\|_{*} + \frac{1}{2}\|\mathbf{Y}\|_F^2 \\ &\text{subject to} \quad [\mathbf{Y}]_{i,j} = [\mathbf{X}]_{i,j}, \forall(i,j)\in\Omega, \end{aligned} \tag{1}$$

where $\|\cdot\|_{*}$ denotes a nuclear-norm, i.e. sum of singular values, and $[\mathbf{A}]_{i,j}$ denotes the (i,j) entries of the matrix $\mathbf{A}$. The algorithm proposed in [7], solves eq. (1) iteratively performing SVD with thresholded singular values. The other approach is called fixed-point continuation with approximate SVD (FPCA) which optimizes the following problem:

$$\begin{aligned} &\text{minimize } \tau\|\mathbf{Y}\|_{*} \\ &\text{subject to } ([\mathbf{Y}]_{i,j} - [\mathbf{X}]_{i,j})^2 \le \theta,\ \forall(i,j)\in\Omega, \end{aligned} \tag{2}$$

where both τ and θ are parameters. The FPCA also mainly relies on the thresholding operator. Additionally, there have been many other work addressing this problem of performing SVD on a noisy, partially-observed matrix (see, e.g., [8]).

In this paper, a new iterative algorithm that simultaneously completes the input matrix, computes SVD parameters $\mathbf{U}$ and $\mathbf{V}$ and determines the rank of the input will be introduced and empirically evaluated. The first part of the evaluation follows some of the experimental procedure given in [7], and the remaining tries reconstructing occluded single-frame images.

II. New cost function and an iterative algorithm

Here we formulate the singular value decomposition with noisy incomplete matrix as follows:

$$\begin{aligned} \min \quad & J := \frac{\lambda}{2}\sum_{(i,j)\in\Omega}([\mathbf{Y}]_{ij} - [\mathbf{X}]_{ij})^2 + \frac{1}{2}\left\|\mathbf{Y} - \mathbf{U}\mathbf{S}\mathbf{V}^{\top}\right\|_{\mathrm{F}}^2 \\ \text{w.r.t.} \quad & \mathbf{U}\in\mathbb{R}^{m\times k}, \mathbf{V}\in\mathbb{R}^{n\times k}, \mathbf{Y}\in\mathbb{R}^{m\times n}, \mathbf{S}\in\mathbb{R}^{k\times k} \\ \text{s.t.} \quad & \mathbf{U}^{\top}\mathbf{U} = \mathbf{I}_k, \mathbf{V}^{\top}\mathbf{V} = \mathbf{I}_k \\ & \mathbf{S} = \mathrm{diag}(s_1,\ldots,s_k), s_i \ge 0\ \forall i = 1,\ldots,k, \end{aligned} \tag{3}$$

where λ is regularization parameter and the operator diag returns a diagonal matrix. In this section, we assume the parameter k fixed. Later, we introduce a procedure for finding parameter k automatically. In the above minimization we are looking for a surrogate matrix $\mathbf{Y}$ which is close, in proper metric, to the given matrix $\mathbf{X}$. The first term in the objective function J is to reduce the additional noise and the second term estimates a representation in terms of $\mathbf{U}\mathbf{S}\mathbf{V}^\top$ for complete matrix $\mathbf{Y}$. The extra constraints $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^\top\mathbf{V} = \mathbf{I}$ encodes the singular value decomposition of the surrogate $\mathbf{Y}$.

Let us assume the diagonal entries of the matrix $\mathbf{S}$ are fixed to an arbitrary non-negative decreasing sequence, then (3) is equivalent to the cost function of the FPCA (2), except for that (3) has a fixed constant $k \leq K$. By replacing the first term of the objective function in (3) with equality constraint for $[\mathbf{Y}]_{ij}, (i,j) \in \Omega$, the cost function reduces to the SVT (1).

The optimization problem in (3) is an optimization over product Steifel manifolds [1]. Here, we use the first order optimization procedure suggested by [10]. In summary, we first compute the gradient of the objective function and then project it onto the tangent space of the Steifel manifold. The first order procedure then consists of moving along the gradient vector on the tangent space using the retraction operator. This procedure is repeated until some convergence criterion is satisfied. The retraction operator is mapping between the tangent space and manifold. Illustrations include the exponential mapping on Riemannian manifolds. The optimization with respect to $\mathbf{S}$ and $\mathbf{Y}$ follows gradient-descent algorithm.

A. *Optimizing* $\mathbf{U}$ *and* $\mathbf{V}$

Matrices $\mathbf{U}$ and $\mathbf{V}$ are on the (orthogonal) Stiefel manifolds $\mathrm{St}(k,m)$ and $\mathrm{St}(k,n)$, where

$$\mathrm{St}(l,p) := \{\mathbf{X} \in \mathbb{R}^{l\times p} : \mathbf{X}^\top\mathbf{X} = \mathbf{I}_p\}.$$

$\mathrm{St}(l,p)$ is an embedded submanifold of $\mathbb{R}^{l\times p}$. The tangent space of $\mathrm{St}(k,m)$ is characterized by

$$T_{\mathbf{U}}\mathrm{St}(k,m) = \{\mathbf{U}\Sigma + \mathbf{U}_\perp\mathbf{Z} : \Sigma^\top = -\Sigma, \mathbf{Z} \in \mathbb{R}^{(m-k)\times k}\},$$

where $\mathbf{Z}_\perp$ is a $m \times (m-k)$ matrix such that $\mathrm{span}(\mathbf{Z}_\perp)$ is the orthogonal complement of $\mathrm{span}(\mathbf{Z})$.

In details, for $\mathbf{U}$ or $\mathbf{V}$, we compute the gradient of J, see (3), in the embedding Euclidean space, and project it to the tangent space of the Stiefel manifold at $\mathbf{U} \in \mathrm{St}(k,m)$, denoted by $T_{\mathbf{U}}\mathrm{St}(k,m)$, with the projection operator $P_{\mathbf{U}}(\nabla)$:

$$P_{\mathbf{U}}(\nabla) = \nabla - \frac{1}{2}\mathbf{U}(\mathbf{U}^\top\nabla + \nabla^\top\mathbf{U}). \tag{4}$$

We then update the current point (which is an orthogonal matrix) along the gradient vector on the tangent space using the retraction operator:

$$R_{\mathbf{U}}(\xi) = \mathrm{qf}(\mathbf{X} + \xi), \tag{5}$$

[1]In [9] a similar approach based on Grassman manifolds has been proposed for completing incomplete matrices using SVD.

where qf and ξ are a Q-factor of the QR decomposition and a projected update direction. For details on the retraction concept please see [10]. In summary we have the following proposition for the first order update rules followed by the retraction (5).

Proposition 1. *The update rules for* $\mathbf{U}$ *and* $\mathbf{V}$ *are*

$$\mathbf{U} \leftarrow \mathbf{U} + \eta(\mathbf{Y}\mathbf{V} + \mathbf{U}\mathbf{V}^\top\mathbf{Y}^\top\mathbf{U})\mathbf{S}, \tag{6}$$

$$\mathbf{V} \leftarrow \mathbf{V} + \eta(\mathbf{Y}^\top\mathbf{U} + \mathbf{V}\mathbf{U}^\top\mathbf{Y}\mathbf{V})\mathbf{S} \tag{7}$$

Proof: See Appendix A. ■

B. *Optimizing* $\mathbf{S}$

The matrix $\mathbf{S}$ is a diagonal matrix with positive entries. It is straightforward to derive update rule for $\mathbf{S}$. By setting the partial-derivative of (3) with respect to s_k to 0 we obtain

$$\mathbf{S} \leftarrow \mathbf{U}^\top\mathbf{Y}\mathbf{V}. \tag{8}$$

However, the above update rule may not satisfy that a singular value must be non-negative. This condition can easily be satisfied by multiplying either $\mathbf{U}$ or $\mathbf{V}$ with $\mathrm{sgn}(\mathbf{S})$ and taking the absolute values of $\mathbf{S}$ as a new $\mathbf{S}$.

C. *Optimizing* $\mathbf{Y}$

We use the gradient-descent (9) for optimizing the objective function J with respect to $\mathbf{Y}$.

Proposition 2. *The update rule for* $\mathbf{Y}$ *is*

$$\mathbf{Y} \leftarrow \mathbf{Y} + \eta\left(\mathbf{U}\mathbf{S}\mathbf{V}^\top - \mathbf{Y} + \lambda Q_\Omega\left(\mathbf{X} - \mathbf{Y}\right)\right), \tag{9}$$

where $Q_\Omega(\mathbf{Z})$ *is a matrix such that its* (i,j)*-th element is* $[\mathbf{Z}]_{i,j}$ *if* $(i,j) \in \Omega$ *and* 0 *otherwise.*

The parameter λ restricts the amount of the distance between completed matrix $\mathbf{Y}$ and the original matrix (using given entires only). The penalization parameter λ should be determined based on the portion of the size of missing values. Intuitively, it can be considered as allowing the SVD parameters to capture more regularities in the whole input matrix $\mathbf{X}$ including the missing values completed by the optimization rather than focusing only on reconstructing the observed values of $\mathbf{X}$.

D. *Automatic Rank Determination*

The update rules that are derived in previous section assumes that the rank k is fixed. Now we propose a procedure to estimate the rank on-the-fly.

One interpretation of a singular value of $\mathbf{X}$ is that it corresponds to the square-root of an eigenvalue of either $\mathbf{X}^\top\mathbf{X}$ or $\mathbf{X}\mathbf{X}^\top$, and the left and right singular vectors correspond to the eigenvectors of them, respectively. We follow the same principle behind the principal component analysis (PCA), which states that most of the variation in the data is explained in the subspace characterized by a set of eigenvectors corresponding to the largest eigenvalues. So, we estimate the rank while updating the parameters such that the lower-bound of the ratio between the sum of singular values estimated so far

Algorithm 1 The iterative algorithm for computing SVD.

Input an input matrix $\mathbf{X}_{m\times n}$, a set of observed indices Ω, a rank-update interval n_r
Initialize $\mathbf{U}_{m\times r}$, $\mathbf{V}_{n\times r}$ and $\mathbf{Y}_{m\times n}$ to some small random values where $r < \min(m, n)$ is a small integer.
Orthogonalize $\mathbf{U}$ and $\mathbf{V}$ according to (5)
Compute $\mathbf{S}$ according to (8)
Set $n = 0$
repeat
  Update $\mathbf{U}$ and $\mathbf{V}$ according to (6) and (7)
  Update $\mathbf{S}$ according to (8)
  Make all s_{ii} positive according to Section II-B
  Update $\mathbf{Y}$ according to (9).
  if $\mod(n, n_r) = 0$ and (10) is *false* **then**
    Append a set of random column vectors to $\mathbf{U}$ and $\mathbf{V}$
    Orthogonalize $\mathbf{U}$ and $\mathbf{V}$
    Update $\mathbf{S}$
    Make all s_{ii} positive
  end if
until Stopping criterion is reached

and the total sum of singular values is larger than a predefined value $\tau \in [0\ 1]$, i.e.

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{k} \lambda_i + (K-k)\lambda_k} > \tau, \qquad (10)$$

where $(\lambda_1, \lambda_2, \cdots, \lambda_K)$ is a non-increasing singular values of a matrix. We can check that the ratio of the sum of the first singular values over the sum of all of them is larger than τ even when we only know the first k singular values out of all K singular values. The algorithm, hence, keeps increasing the rank until (3) hold, which indirectly suggests that the desired amount of the variance, or the standard-deviation, can be explained with the current rank k.

It is important to note that the condition (10) holds when the J is optimized with respect to SVD representation. The algorithm, thus, increases the rank k only after every few predefined updates. At each predefined interval the algorithm appends a new set of singular values and singular vectors to $\mathbf{S}$, $\mathbf{U}$ and $\mathbf{V}$ if (10) does not hold.

The proposed algorithm, called iSVD, is listed in Algorithm 1. One major difference between the iSVD and other similar algorithms is that they rely on the thresholding operator whereas the iSVD utilizes its own rank determination method, which is basically in the same line as the nuclear norm minimization.

III. Experiment

A. Evaluation Criteria

We use the following approximation error measure for evaluating the performance of the algorithms. The error of completing missing values is defined by

$$E_{\text{missing}} = \frac{\left\| Q_{\overline{\Omega}}(\tilde{\mathbf{X}} - \mathbf{Y}) \right\|_F^2}{\left\| Q_{\overline{\Omega}}(\tilde{\mathbf{X}}) \right\|_F^2},$$

where $\tilde{\mathbf{X}}$ is a matrix of true values of the input matrix $\mathbf{X}$ and $\overline{\Omega}$ is a complement of Ω.

We also check the difference between the original matrix with true values and the completed matrix $\mathbf{Y}$, or the reconstruction using the SVD parameters $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{S}$:

$$E_{\text{svd}} = \frac{\left\| \tilde{\mathbf{X}} - \mathbf{U}\mathbf{S}\mathbf{V}^\top \right\|_F^2}{\left\| \tilde{\mathbf{X}} \right\|_F^2},$$

For computing E_{svd} for noisy input matrix, $\tilde{x}_{ij}$ should be replaced with an original value without noise. It is denoted as a relative error in [7].

The other measure is to check the estimated rank. As the rank is directly related to the computational expense as well as the memory requirement, it is important to have as low rank as possible while completing the matrix well enough.

The performance measurements should be computed after the iSVD terminated. The stopping condition checks the

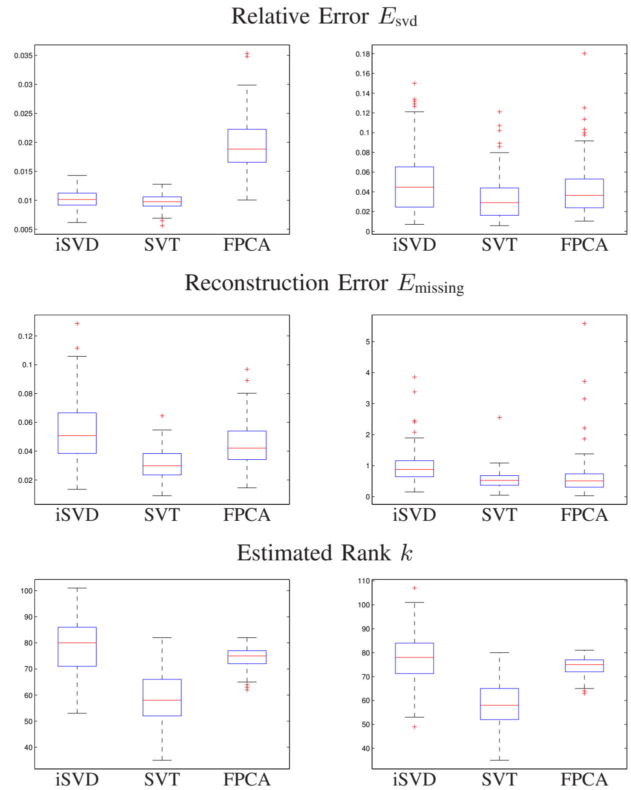

Fig. 1. Relative error, reconstruction error and estimated rank of images corrupted with randomly missing pixels (left) and a missing square (right).

Size	p	K	$\tilde{K}$ iSVD	SVT	FPCA	E_{svd} $(\times 10^{-3})$ iSVD	SVT	FPCA	E_{missing} $(\times 10^{-3})$ iSVD	SVT	FPCA
1000	0.12	10	22	10	7	2.9	0.0017	3.9	3.2	0.0018	4.0
1000	0.39	50	6	50	1	1.8	0.0057	2.1	2.0	0.0077	2.2
1000	0.57	100	4	100	1	1.0	0.0044	1.1	1.1	0.0077	1.1
5000	0.024	10	69	10	1	26.8	0.0020	9.5	27.4	0.0020	9.5
5000	0.10	50	18	50	1	1.6	0.0053	2.2	1.7	0.0056	2.2
5000	0.158	100	10	100	1	1.1	0.0050	1.1	1.1	0.0056	1.1

TABLE I

PERFORMANCE OF ISVD ON NOISELESS SYNTHETIC MATRICES WITH MISSING VALUES. ALL MATRICES WERE SQUARE AND THE SIZE INDICATES THE NUMBER OF ROWS.

changes in the reconstruction error of the observed elements, which is defined by

$$E_{\text{obs}} = \frac{\|Q_\Omega(\mathbf{X} - \mathbf{Y})\|_{\text{F}}^2}{\|Q_\Omega(\mathbf{X})\|_{\text{F}}^2},$$

The iSVD stops when $|E_{\text{obs}}^{(t)} - E_{\text{obs}}^{(t-1)}| < \epsilon$ for more than a predefined number.

B. Synthetic Matrix

In this section iSVD is evaluated on synthetic matrices constructed randomly with a specific predefined rank. The construction was based on [7] except for that uniform random variables $U(0,1)$ was used instead of standard Normal ones for constructing the left and right matrices.

An input matrix $\mathbf{X}$ with a rank K and a size $m \times n$ was constructed by $\mathbf{X}_K = \mathbf{M}_r \mathbf{M}_l^\top$ where $\mathbf{M}_r$ and $\mathbf{M}_l$ are matrices of size $m \times K$ and $n \times K$. Each element of $\mathbf{M}_r$ and $\mathbf{M}_l$ was sampled from $U(0,1)$. Then, each element of X was randomly chosen to be either observed or missing with a probability p. White Gaussian noise with a standard deviation σ was added to each observed element afterward.

As a comparison we tried the SVT and the FPCA as well as the iSVD. We used codes provided by the authors of the corresponding papers for both the SVT and FPCA.

Firstly, all three methods were evaluated on two sets of three matrices of different ranks $K = 10$, 50 and 100. The sets differ by the size of the member matrices which are 1000×1000 and 5000×5000. Each element of the matrices of size 1000×1000 was observed with probabilities $p = 0.12$, 0.39 and 0.57, respectively. They were 0.024, 0.10 and 0.158, respectively, in the case of the matrices of size 5000×5000. We used $\lambda = 0.1$ and $\tau = 0.9$ for the iSVD in this experiment.

In Table I, all three evaluation criteria are shown for each input matrix. It is quite clear that all three methods perform comparably to each other. The SVT performed exceptionally better compared to the others, while the iSVD was slightly better than the FPCA. The iSVD, however, performed poorly when the observation probability was as low as 0.024.

Next, white Gaussian noise was added to each observed element of the matrices of size 1000×1000 constructed with the same parameters as the first set of experiments. In order to test the robustness of the iSVD and the other algorithms to the level of noise, we tested three different standard deviations for the noise which are $\sigma = 0.01$, 0.1 and 1.

In Table II, all three evaluation criteria are shown for each input matrix. The resulting errors are quite similar for all the three algorithms, when there are sufficiently many observed elements with moderate level of noisy.

As was the case with the matrices without any observation noise, when the observation probability is low and the level of noisy is high, the iSVD was not able to perform well compared to the others.

The estimated ranks are, on the other hand, significantly different among the algorithms. Especially, in most cases we can see that the estimated ranks of the SVT are significantly higher than the rest.

C. Image Reconstruction

Images of side-views of cars from Caltech-101 dataset [11] were extracted and made gray-scale images to be experimented on.

Two separate copies of each image were obtained by removing a random square patch of size 50×50 and removing randomly chosen 2500 pixels. These were tested to see how iSVD performed on occluded images and images with salt-and-pepper noise. Additionally, white Gaussian noise with a standard deviation 0.1 was added to each observed element.

A single corrupted image is given as a matrix $\mathbf{X}$ with missing pixels to an algorithm. Each algorithm then returns a complete matrix $\mathbf{Y}$. We consider $\mathbf{Y}$ as a de-noised image with missing pixels reconstructed.

All the three methods were applied to each single corrupted image to complete the whole image including the missing part. In order to avoid overfitting to the noisy observed pixels, we used $\lambda = 0.1$, a larger initial number of singular values of 5, and a smaller $\tau = 0.5$ for iSVD.

Top two rows of Figure 1 show the box plots of relative error of the completed images and reconstruction error of the missing pixels. The relative error values were computed against the original image without noise added. The estimated ranks by the three algorithms are present in the bottom row of Figure 1.

Figure 2 display an original image, the corresponding corrupted images and the reconstructed images according to all three algorithms. We can see that all the algorithms were able to remove the added white noise effectively, while they have difficulty reconstructing the missing squares.

It is clear that all three algorithms perform comparably to each other. The SVT was able to perform well with lower

σ	p	K	$\tilde{K}$ iSVD	SVT	FPCA	E_{svd} ($\times 10^{-3}$) iSVD	SVT	FPCA	E_{missing} ($\times 10^{-3}$) iSVD	SVT	FPCA
0.01	0.12	10	22	10	6	2.8	2.0	4.3	3.2	2.1	4.3
0.01	0.39	50	8	50	1	1.7	0.57	2.1	1.9	0.76	2.2
0.01	0.57	100	4	100	1	1.0	0.43	1.1	1.1	0.76	1.1
0.1	0.12	10	27	58	8	3.3	2.1	2.4	3.6	1.9	2.5
0.1	0.39	50	6	50	1	1.9	2.4	2.2	2.0	2.7	1.1
0.1	0.57	100	4	100	1	1.0	1.0	1.1	1.1	1.5	1.1
1	0.12	10	86	2	29	298.3	18.2	23.9	318.5	18.2	22.9
1	0.39	50	17	470	15	2.3	13.7	2.3	2.5	7.5	2.3
1	0.57	100	5	564	1	1.0	1.5	1.1	1.1	1.4	1.1

TABLE II
PERFORMANCE OF iSVD ON NOISY SYNTHETIC MATRICES WITH MISSING VALUES.

ranks than the other algorithms.

IV. CONCLUSION

In this paper a new iterative algorithm for singular value decomposition (SVD) on a matrix with missing values is proposed. The new algorithm called the iterative SVD (iSVD) is able to complete missing values (as well as adjusting observed values, in case of noisy observation), estimate SVD, and automatically determine the rank.

The iSVD relies on a first order method on a product manifold of the Stiefel manifolds and gradient-descent. The projection of the gradient to the tangent space of the Stiefel manifold and the retraction of the updated parameters enables the algorithm to adapt the rank of the approximation while simultaneously completing the missing values and estimating the SVD parameters.

The empirical evaluation of the iSVD on the synthetic matrices and the single-frame image reconstruction revealed that the iSVD performs comparably, or in some cases favorably, to the recently introduced SVD algorithms based on the thresholding operator such as the SVT and the FPCA. However, the iSVD showed some weakness in some cases, including where the observation probability is low and the level of noise is high at the same time.

There are some possibilities for improving the iSVD. One possibility is to employ the second-order optimization approach, which is left as one of the important future work. In addition, adding penalization terms on the singular values such as trace-norm is another direction for extension of the current paper. It will also be interesting to investigate different norms for penalizing the difference between $\mathbf{Y}$ and $\mathbf{X}$.

REFERENCES

[1] G. H. Golub and C. F. van Van Loan, *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition)*, 3rd ed. The Johns Hopkins University Press, Oct. 1996.
[2] D. Lay, *Linear Algebra and Its Applications*. Pearson Education, 2002.
[3] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Math. Program.*, vol. 128, pp. 321–353, June 2011.
[4] R. Salakhutdinov and A. Mnih, "Collaborative Filtering in a Non-Uniform World: Learning with the Weighted Trace Norm," in *Advances in Neural Information Processing Systems*, vol. 23, 2011.
[5] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–534, 2009.
[6] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for dna microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.
[7] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," 2008. [Online]. Available: http://arxiv.org/abs/0810.3286
[8] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: ACM, 2009, pp. 457–464.
[9] R. H. Keshavan and S. Oh, "A Gradient Descent Algorithm on the Grassman Manifold for Matrix Completion," *CoRR*, vol. abs/0910.5260, 2009.
[10] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press, 2008.
[11] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories." in *Workshop on Generative-Model Based Vision*, 2004.

APPENDIX

A. Proof of Proposition 1

Proof: The steepest-descent update direction for $\mathbf{U}$ in the embedding Euclidean space of $\mathrm{St}(m, k)$ is a gradient of (3) with respect to $\mathbf{U}$ which is

$$\nabla_{\mathbf{U}} J = -\mathbf{YVS}.$$

As the tangent space of the Euclidean space is itself, we need to project it to the tangent space of $\mathrm{St}(m, k)$ which can be done by applying (4).

$$\begin{aligned}
P_{\mathbf{U}}(\mathbf{YVS}) &= \mathbf{YVS} - \frac{1}{2}\mathbf{U}(\mathbf{U}^\top\mathbf{YVS} - \mathbf{SV}^\top\mathbf{Y}^\top\mathbf{U}) \\
&= \mathbf{YVS} - \frac{1}{2}\mathbf{YVS} + \frac{1}{2}\mathbf{USV}^\top\mathbf{Y}^\top\mathbf{U}) \\
&= \frac{1}{2}\mathbf{YVS} + \frac{1}{2}\mathbf{USV}^\top\mathbf{Y}^\top\mathbf{U} \\
&= \frac{1}{2}(\mathbf{YV} + \mathbf{UV}^\top\mathbf{Y}^\top\mathbf{U})\mathbf{S}.
\end{aligned}$$

The factor $\frac{1}{2}$ can be absorbed into a step size η. Then, we get (6).

The same derivation can be done for $\mathbf{V}$ and it is easy to obtain (7). ■

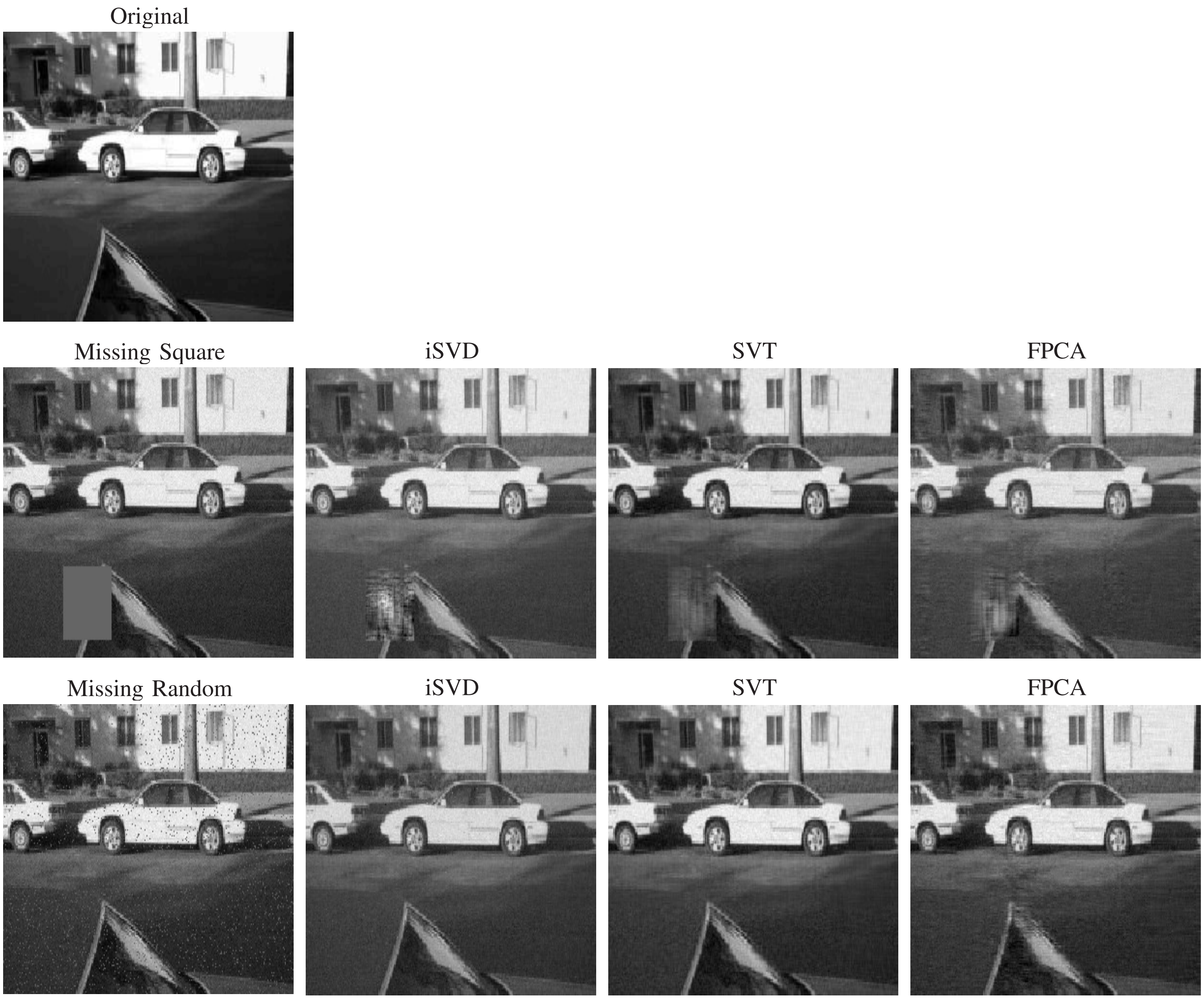

Fig. 2. Reconstruction of a missing square block of a corrupted image using the iSVD, the SVT and the FPCA.