Missing Entries Matrix Approximation and Completion

Gil Shabat
School of Electrical Engineering
Tel Aviv University
gil@eng.tau.ac.il

Yaniv Shmueli
School of Computer Science
Tel Aviv University
yaniv.shmueli@cs.tau.ac.il

Amir Averbuch
School of Computer Science
Tel Aviv University
amir@math.tau.ac.il

***Abstract*—We describe several algorithms for matrix completion and matrix approximation when only some of its entries are known. The approximation constraint can be any whose approximated solution is known for the full matrix. For low rank approximations, similar algorithms appear recently in the literature under different names. In this work, we introduce new theorems for matrix approximation and show that these algorithms can be extended to handle different constraints such as nuclear norm, spectral norm, orthogonality constraints and more that are different than low rank approximations. As the algorithms can be viewed from an optimization point of view, we discuss their convergence to global solution for the convex case. We also discuss the optimal step size and show that it is fixed in each iteration. In addition, the derived matrix completion flow is robust and does not require any parameters. This matrix completion flow is applicable to different spectral minimizations and can be applied to physics, mathematics and electrical engineering problems such as data reconstruction of images and data coming from PDEs such as Helmholtz's equation used for electromagnetic waves.**

I. Introduction

Matrix completion and matrix approximation are important problems in a variety of fields such as statistics [1], biology [2], statistical machine learning [3], signal processing and computer vision/image processing [4]. Rank reduction by matrix approximation is important, for example, in compression where low rank indicates the existence of redundant information and matrix completion is important in collaborative filtering, such as the Netflix problem and different reconstruction problems. Usually, the matrix completion problem, is defined as finding a matrix, with smallest possible rank, that satisfy the existence of certain entries.

$$\begin{aligned} &\text{minimize rank } (\mathbf{X}) \\ &\text{subject to } X_{i,j} = M_{i,j}, \quad (i,j) \in \Omega. \end{aligned} \tag{I.1}$$

Since Eq. I.1 is an NP-hard problem, some relaxations methods have been proposed. The most popular relaxation is one that replaces the rank by the nuclear norm:

$$\begin{aligned} &\text{minimize } \|\mathbf{X}\|_* \\ &\text{subject to } X_{i,j} = M_{i,j}, \quad (i,j) \in \Omega, \end{aligned} \tag{I.2}$$

where $\|\mathbf{X}\|_*$ denotes the nuclear norm of $\mathbf{X}$ that is equal to the sum of the singular values of $\mathbf{X}$. A small value of $\|\mathbf{X}\|_*$ is related to the property of having a low rank [5]. An iterative solution, which is based on a singular value thresholding, is given in [6]. A completion algorithm, based on the local information of the matrix, is proposed in [7]. In this work, a more robust and simple approach for solving a variety of matrix approximation of certain entries by approximating the full matrix is discussed. We approximate problems of the form

$$\begin{aligned} &\text{minimize } \|\mathcal{P}_\Omega \mathbf{X} - \mathcal{P}_\Omega \mathbf{M}\|_F \\ &\text{subject to } f(\mathbf{X}) \leq 0, \end{aligned} \tag{I.3}$$

given that the solution for

$$\begin{aligned} &\text{minimize } \|\mathbf{X} - \mathbf{M}\|_F \\ &\text{subject to } f(\mathbf{X}) \leq 0 \end{aligned} \tag{I.4}$$

is known. Here, $\{\mathcal{P}_\Omega \mathbf{X}\}_{i,j} = X_{i,j}$ if $(i,j) \in \Omega$ and 0 otherwise. If $f(\mathbf{X})$ is convex and satisfies some condition (which is explained in the next sections), the algorithm finds the global solution. Nevertheless, convergence is guaranteed, but to a local solution. Then, we show how this algorithm can be used for solving a variety of matrix completion problems as well, such as spectral norm completion:

$$\begin{aligned} &\text{minimize } \|\mathbf{X}\|_2 \\ &\text{subject to } X_{i,j} = M_{i,j}, \quad (i,j) \in \Omega, \end{aligned} \tag{I.5}$$

Ky-Fan norm completion:

$$\begin{aligned} &\text{minimize } \|\mathbf{X}\|_{(k)} \\ &\text{subject to } X_{i,j} = M_{i,j}, \quad (i,j) \in \Omega, \end{aligned} \tag{I.6}$$

where $\|\mathbf{X}\|_{(k)} = \sum_{i=1}^{k} \sigma_i$ (sum of largest k singular values). Note that the spectral norm and the nuclear norm are a special case of the Ky-Fan norm. We also discuss approximation problems such as:

$$\begin{aligned} &\text{minimize } \|\mathcal{P}_\Omega \mathbf{X} - \mathcal{P}_\Omega \mathbf{M}\|_F \\ &\text{subject to } \mathbf{X}^{\mathbf{T}}\mathbf{X} = \mathbf{I}. \end{aligned} \tag{I.7}$$

II. Theorems on full matrix approximation

The algorithm that approximates a matrix at certain points requires from us to be able to approximate the matrix when taking into account all its entries. Therefore, we review some theorems on full matrix approximation theorems in addition to the well known Eckart-Young theorem mentioned in the introduction. The low rank approximation problem can be modified to approximate a matrix under the Frobenius norm

while having the Frobenius norm as a constraint as well instead of having low rank. Formally,

$$\begin{aligned} &\text{minimize } \|\mathbf{X}-\mathbf{M}\|_F \\ &\text{subject to } \|\mathbf{X}\|_F \leq \lambda. \end{aligned} \tag{II.1}$$

A solution for Eq. II.1 is given by $\mathbf{X} = \frac{\mathbf{M}}{\|\mathbf{M}\|_\mathbf{F}} \min(\|\mathbf{M}\|_\mathbf{F}, \lambda)$.

Proof: The expression $\|\mathbf{X}\|_F^2 \leq \lambda^2$ can be thought of as an $m \times n$ dimensional ball with radius λ centered at the origin. $\mathbf{M}$ is an $m \times n$ dimensional point. We are looking for a point $\mathbf{X}$ on the ball $\|\mathbf{X}\|_F^2 = \lambda^2$ that has a minimal Euclidean distance (Frobenius norm) from $\mathbf{M}$. If $\|\mathbf{M}\|_F \leq \lambda$ then $\mathbf{X} = \mathbf{M}$ and it is inside the ball having a distance of zero. If $\|\mathbf{M}\|_F > \lambda$, then the shortest distance is given by the line going from the origin to $\mathbf{M}$ whose intersection with the sphere $\|\mathbf{X}\|_F^2 \leq \lambda^2$ is the closest point to $\mathbf{M}$. This point is given by $\mathbf{X} = \frac{\mathbf{M}}{\|\mathbf{M}\|_\mathbf{F}}\lambda$. ■

An alternative approach uses the Lagrange multiplier in a brute-force manner. This leads to a non-linear system of equations, which are difficult to solve. Note that this problem can be easily extended to the general case

$$\begin{aligned} &\text{minimize } \|\mathcal{P}\mathbf{X}-\mathcal{P}\mathbf{M}\|_F \\ &\text{subject to } \|\mathbf{X}\|_F \leq \lambda. \end{aligned} \tag{II.2}$$

Proof: The proof is similar to the previous one but here we are looking for a point $\mathbf{X}$ on the sphere that is the closest to a line whose points $\mathbf{X}' \in \mathcal{H}$ satisfy $\mathcal{P}\mathbf{X}' = \mathcal{P}\mathbf{M}$. By geometrical considerations, this point is given by $\mathbf{X} = \frac{\mathcal{P}\mathbf{M}}{\|\mathcal{P}\mathbf{M}\|_F}\lambda$. ■

Hence, we showed a closed form solution for the problem in Eq. II.2.

Another example is the solution to the problem:

$$\begin{aligned} &\text{minimize } \|\mathbf{X}-\mathbf{M}\|_F \\ &\text{subject to } \mathbf{X}^T\mathbf{X} = \mathbf{I}. \end{aligned} \tag{II.3}$$

This is known as the orthogonal Procrustes problem ([8]) and the solution is given by $\mathbf{X} = \mathbf{U}\mathbf{V}^*$, where the SVD of $\mathbf{M}$ is given by $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$. The solution can be extended to a matrix $\mathbf{X}$ satisfying $\mathbf{X^T X} = \mathbf{D^2}$, where $\mathbf{D}$ is a known or unknown diagonal matrix. When $\mathbf{D}$ is unknown, the solution is the best possible orthogonal matrix. When $\mathbf{D}$ is known, the problem can be converted to become the orthonormal case (Eq. II.3) by substituting $\mathbf{X} = \mathbf{V}\mathbf{D}$ where $\mathbf{V^T V} = \mathbf{I}$. When $\mathbf{D}$ is unknown, the problem can be solved by applying an iterative algorithm that is described in [9].

We now examine the following problem:

$$\begin{aligned} &\text{minimize } \|\mathbf{X}-\mathbf{M}\|_F \\ &\text{subject to } \|\mathbf{X}\|_2 \leq \lambda. \end{aligned} \tag{II.4}$$

A solution to this problem uses the Pinching theorem ([10]):

Lemma II.1 (Pinching theorem)**.** *For every matrix* $\mathbf{A}$ *and a unitary matrix* $\mathbf{U}$ *and for any norm satisfying* $\|\mathbf{U}\mathbf{A}\mathbf{U}^*\| = \|\mathbf{A}\|$ *then* $\|diag(\mathbf{X})\| \leq \|\mathbf{X}\|$.

A proof is given in [12]. An alternative proof is given in [14].

Lemma II.2 (Minimization of the Frobenius norm under the spectral norm constraint)**.** *Assume the SVD of* $\mathbf{M}$ *is given by* $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ *where* $\mathbf{\Sigma} = diag(\sigma_1, .., \sigma_n)$*. Then, the matrix* $\mathbf{X}$*, which minimizes* $\|\mathbf{X}-\mathbf{M}\|_F$ *such that* $\|\mathbf{X}\|_2 \leq \lambda$*, is given by* $\mathbf{X} = \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^*$ *where* $\tilde{\sigma}_i$ *are the singular values of* $\tilde{\Sigma}$ *and* $\tilde{\sigma}_i = \min(\sigma_i, \lambda), i = 1, \dots k,\ k \leq n$.

Proof: $\|\mathbf{X}-\mathbf{M}\|_F = \|\mathbf{X}-\mathbf{U}\mathbf{\Sigma}\mathbf{V}^*\|_F = \|\mathbf{U}^*\mathbf{X}\mathbf{V}-\mathbf{\Sigma}\|_F$. Since $\mathbf{\Sigma}$ is diagonal, $\|\text{diag}(\mathbf{U}^*\mathbf{X}\mathbf{V})-\mathbf{\Sigma}\|_F \leq \|\mathbf{U}^*\mathbf{X}\mathbf{V}-\mathbf{\Sigma}\|_F$. From Lemma II.1 we know that $\|\text{diag}(\mathbf{U}^*\mathbf{X}\mathbf{V})\|_2 \leq \|\mathbf{U}^*\mathbf{X}\mathbf{V}\|_2$. Therefore, $\mathbf{U}^*\mathbf{X}\mathbf{V}$ has to be diagonal and the best minimizer under the spectral norm constraint is achieved by minimizing each element separately yielding $\mathbf{U}^*\mathbf{X}\mathbf{V} = \text{diag}(\min(\sigma_i, \lambda)),\ i = 1, \dots k, k \leq n$. Hence, $\mathbf{X} = \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^*$. ■

The same argument that states that $\mathbf{U}^*\mathbf{X}\mathbf{V}$ has to be diagonal, can also be applied when the constraint is given by the nuclear norm. Define $\tilde{\mathbf{\Sigma}} = \mathbf{U}^*\mathbf{X}\mathbf{V}$. We wish to minimize $\|\tilde{\mathbf{\Sigma}}-\mathbf{\Sigma}\|_F = \sum_i (\tilde{\sigma}_i - \sigma_i)^2$ s.t. $\|\mathbf{X}\|_* = \|\tilde{\mathbf{\Sigma}}\|_* = \sum_i |\tilde{\sigma_i}| \leq \lambda, i = 1, \dots k, k \leq n$. Note that $\tilde{\sigma}_i$ has to be nonnegative otherwise it will increase the Frobenius norm but will not change the nuclear norm. Hence, the problem can now be formulated as:

$$\begin{aligned} &\text{minimize } \textstyle\sum_i (\tilde{\sigma}_i - \sigma_i)^2 \\ &\text{subject to } \textstyle\sum_i \tilde{\sigma}_i \leq \lambda, \\ &\qquad \tilde{\sigma}_i \geq 0. \end{aligned} \tag{II.5}$$

This is a standard convex optimization problem that can be solved by methods such as semidefinite programming [11]. The exact same can be done to the Ky-Fan norm.

III. Approximation of certain entries

Suppose we wish to approximate only certain entries of the matrix, under different constraints, i.e. we are interested in solving Eq. I.3, given that the solution of Eq. I.4 is known and given by $\mathcal{D}\mathbf{M}$, where $\mathcal{D}$ is the solution operator. For example, if the constraint is $\text{rank}(\mathbf{X}) \leq k$ $\mathcal{D}\mathbf{X}$ is the truncated SVD of $\mathbf{X}$ containing the first k singular values. Note that $\mathcal{D}$ is not necessarily convex. We examine the following iterative algorithm:

$$\mathbf{X}_{n+1} = \mathcal{D}(\mathbf{X}_n - \mathcal{P}(\mathbf{X}_n - \mathbf{M})). \tag{III.1}$$

Eq. III.1 can be considered as a projected gradient algorithm with unit step size, where the projection is given by $\mathbf{D}$.

Theorem III.1 (Local Convergence)**.** *: Let* $\epsilon(\mathbf{X}_n) = \|\mathcal{P}\mathbf{X}_n - \mathcal{P}\mathbf{M}\|_F$ *be the error at the* n*th iteration, then* $\epsilon(\mathbf{X}_n)$ *is monotonically decreasing, and because it is bounded the algorithm converges.*

The proof for Theorem III.1 is given in [14]. Theorem III.1 does not say anything about convergence to the global solution. However, when the projection $\mathcal{D}$ is convex and self adjoint ($\mathcal{D} = \mathcal{D}^*$) and the algorithm is modified to have adaptive step size, that is:

$$\mathbf{X}_{n+1} = \mathcal{D}(\mathbf{X}_n - \mu_n\mathcal{P}(\mathbf{X}_n - \mathbf{M})), \tag{III.2}$$

and $\mu_n = \tilde{\mu}2^{-l[n]}$ is computed by Armijo rule in a greedy form, minimizing the error in every iteration:

$$\begin{aligned} l[n] &= \min\{j \in \mathbb{Z}_{\geq 0} : f(\mathbf{X}_{n,j}) \\ &\qquad \leq f(\mathbf{X}_n) - \sigma \text{trace}(\nabla f(\mathbf{X}_n)^T(\mathbf{X}_n - \mathbf{Z}_{n,j}))\}, \\ \text{and } \mathbf{Z}_{n,j} &= \mathcal{D}(\mathbf{X}_n - \tilde{\mu}2^{-j}\nabla f(\mathbf{X}^n)), \end{aligned} \tag{III.3}$$

where $f(X) = \frac{1}{2}\|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\|_F^2$, $\tilde{\mu} > 0$ and $\sigma \in (0, 1)$, Then the algorithm is guarantee to achieve the global solution [13]. This approach has two major problems:

- For the cases of interest, the operators for truncating the nuclear and spectral norm, are not self-adjoint ($\mathcal{D} \neq \mathcal{D}^*$)
- This approach requires applying the Armijo rule in every iteration. This means several applications of the operator $\mathcal{D}$ in each iteration which is usually computationally expensive.

As for the first point, requiring the projection $\mathcal{D}$ to be self-adjoint can be slightly more than needed for the global convergence proof in [13]. This requirement is needed in order to satisfy $\langle X - Y, \mathcal{D}X - X\rangle \geq 0$ for $Y = \mathcal{D}Y$, which always holds when $\mathcal{D} = \mathcal{D}^*$, but also when $\mathcal{D}$ is as we defined in Lemma II.2 and Eq. II.5.

Theorem III.2. *Let $\mathcal{D}$ be the following projection (defined as in Lemma II.2): Given the SVD of X is $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^*$, we define $\mathcal{D}_\lambda X = U\tilde{S}V^*$ where $\tilde{s}_i = min(s_i, \lambda)$. Then, for every matrices $\mathbf{X}$ and $\mathbf{Y}$ such that $Y = \mathcal{D}\mathbf{Y}$, $\langle \mathbf{X} - \mathbf{Y}, \mathcal{D}\mathbf{X} - \mathbf{X}\rangle \geq 0$*

Proof: The condition $\langle \mathbf{X} - \mathbf{Y}, \mathcal{D}\mathbf{X} - \mathbf{X}\rangle \geq 0$ can be reformulated as

$$\langle \mathbf{X}, \mathbf{X} - \mathcal{D}\mathbf{X}\rangle \geq \langle \mathbf{Y}, \mathbf{X} - \mathcal{D}\mathbf{X}\rangle, \tag{III.4}$$

where $\|Y\|_2 \leq \lambda$.

First, note that the value of the right hand side is maximal when $\mathbf{Y}$ and $\mathbf{X} - \mathcal{D}\mathbf{X}$ have the same angle (Cauchy-Schwartz inequality). Hence, we define: $\mathbf{X} = \mathbf{U}\mathbf{S}_X\mathbf{V}^*$, $\mathbf{Y} = \mathbf{U}\tilde{\mathbf{S}}_Y\mathbf{V}^*$ and $\mathcal{D}\mathbf{X} = \mathbf{U}\tilde{\mathbf{S}}_X\mathbf{V}^*$. The tilde is for indicating that the singular values of $\tilde{\mathbf{S}}$ are smaller or equal to λ.

We start by evaluating the left side of Eq.III.4:

$$\langle \mathbf{X}, \mathbf{X} - \mathcal{D}\mathbf{X}\rangle = \text{trace}[\mathbf{S}_X(\mathbf{S}_X - \tilde{\mathbf{S}_X})] = \sum_i s_{x_i}(s_{x_i} - \tilde{s}_{x_i}). \tag{III.5}$$

Now, for $s_{x_i} \leq \lambda$ we get $(s_{x_i} - \tilde{s}_{x_i}) = 0$. Hence, only when $s_{x_i} > \lambda$ the sum grows and the expression can be rewritten as: $\langle \mathbf{X}, \mathbf{X} - \mathcal{D}\mathbf{X}\rangle = \sum_{s_{x_i} > \lambda} s_{x_i}(s_{x_i} - \tilde{s}_{x_i})$

We now observe the right side of Eq. III.4:

$$\langle \mathbf{Y}, \mathbf{X} - \mathcal{D}\mathbf{X}\rangle = \text{trace}[\tilde{\mathbf{S}}_Y(\mathbf{S}_X - \tilde{\mathbf{S}}_X)] = \sum_i \tilde{s}_{y_i}(s_{x_i} - \tilde{s}_{x_i}). \tag{III.6}$$

Again, the elements that contribute to the sum are those for which $s_{x_i} > \lambda$. Hence, on the right side we obtained: $\langle \mathbf{Y}, \mathbf{X} - \mathcal{D}\mathbf{X}\rangle = \sum_{s_{x_i} > \lambda} \tilde{s}_{y_i}(s_{x_i} - \tilde{s}_{x_i})$.

Both expressions can be thought of as a sum of the positive elements $(s_{x_i} - \tilde{s}_{x_i})$ with different coefficients. Both series have the same length ($s_{x_i} > \lambda$) but the coefficient on the left side is s_{x_i} for i's that give $s_{x_i} > \lambda$ and the right hand series coefficients are by definition (since $\|\mathbf{Y}\|_2 \leq \lambda$) smaller than λ. Therefore, the sum of the left side is bigger than the sum of the right side. This completes the proof. ■

This means that for the spectral norm, the algorithm converges to the global solution. The exact same proof can be done for the nuclear norm and Ky-Fan norm as well, showing the algorithm converges to global solution.

Theorem III.3 (Optimal step size)**.** *For the matrix approximation problem (Eq. I.3) with convex $\mathcal{D}$, the optimal step size is given by $\mu_n = 1$.*

The proof of Theorem III.3 is given in [14]. Note that this holds for any case of projected gradient involving orthogonal axes. Theorem III.3 states that in our case, when having a convex constraint and projection, then Eq. III.1 converges to the global solution. This means, that now we can solve a variety of matrix approximation problem with reasonable computation rate. Note, that we have shown that in some cases, global solution is achieved even when the projection is not self-adjoint (orthogonal). The next section shows, how this very simple algorithm, can be applied to matrix completion problems as well.

IV. Matrix Completion

Matrix completion is an important problem that has been investigated extensively. The matrix completion problem differs from the matrix approximation problem by the fact that the known entries must remain fixed while changing their role from the objective function to be minimized to the constraint part. A well investigated matrix completion problem appears in the introduction as the rank minimization problem. Because rank minimization is not convex and NP-hard, it is usually relaxed for the nuclear norm minimization. Since for the convex case, we have seen that Eq. III.1 converges to the global solution, matrix completion can be achieved simply by using binary search. The advantage of this approach over other different approaches, which minimize the nuclear norm for example, is that it is general and can be applied to other problems that were not addressed such as minimizing the spectral norm. Moreover, some algorithms such as the Singular Value Thresholding (SVT) [6] require additional parameters τ and δ that affect the convergence and the final result, where in this approach no external parameters are required (except for tolerance for determining convergence).

This approach is detailed in Algorithm IV.1, which is robust and does not require any tuning, other than tolerance threshold for determining convergence. Algorithm IV.1 can be used for a matrix completion under a variety of constraints.

Fig. IV shows Algorithm IV.1 results over a corrupted image. In the corrupted image, squares of size 3×3 were randomly removed from the image, destroying 18% of it. The reconstruction is more difficult, since the damage is in squares and not just irregular points. The original image nuclear norm is $51,625$, the corrupted nuclear norm is $96,500$ and the norm of the completed matrix is $50,418$. Minimizing nuclear norm for image reconstructing is a well known method, as images

Algorithm IV.1: Matrix Completion using Nuclear Norm / Spectral Norm Minimization

Input: $\mathbf{M}$ - matrix to complete, $\mathcal{P}$ - projection operator that specifies the important entries, tol - admissible approximation error, λ_{tol} - admissible constraint accuracy
Output: $\mathbf{X}$ - Completed matrix

$\mathbf{M} \leftarrow \mathcal{P}\mathbf{M}$
$\lambda_{min} \leftarrow 0$
$\lambda_{max} \leftarrow \|\mathbf{M}\|_*$ (or $\|\mathbf{M}\|_2$ for the spectral norm)
$\lambda \leftarrow 0$
repeat
  $\lambda_{prev} \leftarrow \lambda$
  $\lambda \leftarrow (\lambda_{min} + \lambda_{max})/2$
  $\mathbf{X} \leftarrow$ Approximate $\mathcal{P}\mathbf{M}$ s.t. $\|\mathbf{X}\|_* \leq \lambda$ (or $\|\mathbf{X}\|_2 \leq \lambda$ for the spectral norm case)
  $error \leftarrow \|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\|_F$
  if $error > tol$ **then**
    $\lambda_{min} \leftarrow \lambda$
  else
    $\lambda_{max} \leftarrow \lambda$
  end if
until $error < tol$ and $|\lambda - \lambda_{prev}| < \lambda_{tol}$
return $\mathbf{X}$

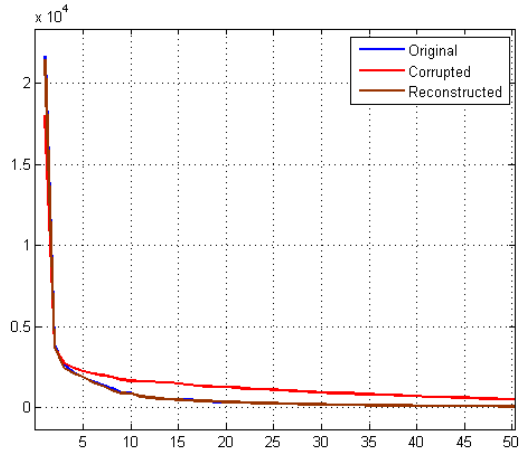

Fig. IV.1. Singular values comparison between the different images.

usually have a low numerical rank as the singular values decay very fast. It can be seen in Fig. IV that the singular values of the reconstructed image, are almost identical to the original.

ACKNOWLEDGMENT

This research was partially supported by the Israel Science Foundation (Grant No. 1041/10) and by the Israeli Ministry of Science & Technology 3-9096.

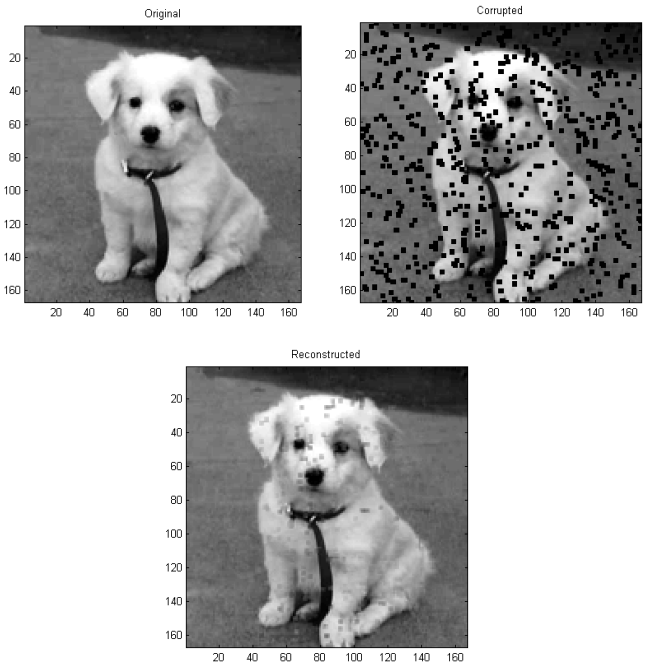

Fig. IV.2. Corrupted dog image and the reconstructed image.

REFERENCES

[1] T.A. Louis, *Finding the observed information matrix when using the EM algorithm*, Journal of the Royal Statistical Society, Series B. (Methodological), Vol. 44, No. 2, pp. 226-233, 1982.
[2] T. Hastie, R. Tibshirani, G. Sherlock, M. Eisen, P. Brown, and D. Botstein, *Imputing missing data for gene expression arrays*. Technical report; Division of Biostatistics, Stanford University, 1999.
[3] N. Srebro and T. Jaakkola, *Weighted low-rank approximations*, Preceeding of the 20th International Conference on Machine Learning (ICML-2003), Washington DC, 2003.
[4] J. Mairal, M. Elad, G. Sapiro, *Sparse representation for color image restoration*, IEEE Transactions on Image Processing, Vol. 17, No. 1, pp.53-69, 2008.
[5] M. Fazel, *Matrix Rank Minimization with Applications. PhD thesis*, Stanford University, 2002.
[6] J.F. Cai, E.J. Candes and Z. Shen, *Singular Value Thresholding Algorithm for Matrix Completion*, SIAM Journal on Optimization, Vol. 20, No. 4, pp. 1956-1982, 2010.
[7] Feng Nan, *Low Rank Matrix Completion*, Master thesis, Massachusetts Institute of Technology, 2009.
[8] P. H. Schonemann, *A generalized solution of the orthogonal procrustes problem*, Psychometrika, Vol. 31, No. 1, pp. 1-10, 1966.
[9] R. Everson, *Orthogonal but not orthonormal Procrustes problem*, 1997.
[10] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics, Springer 1996.
[11] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
[12] , I.C. Gohberg, M.G. Krein, *Introduction to the theory of linear and selfadjoint operators*, Translations of Mathematical Monographs, Vol. 18, pp. 94-95, 1969.
[13] A.N. Iusem, *On the convergence properties of the projected gradient method for convex optimization*, Computational and Applied Mathematics, Vol. 22, No. 1, pp. 37-52, 2003.
[14] G. Shabat, A. Averbuch, *Interest Zone Matrix Approximation*, Electronic Journal of Linear Algebra, Vol. 23, pp. 678-702, 2012.