Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

ELA

INTEREST ZONE MATRIX APPROXIMATION[*]

GIL SHABAT[†] AND AMIR AVERBUCH[‡]

Abstract. An algorithm for matrix approximation, when only some of its entries are taken into consideration, is described. The approximation constraint can be any whose approximated solution is known for the full matrix. For low rank approximations, this type of algorithms appears recently in the literature under different names, where it usually uses the Expectation-Maximization algorithm that maximizes the likelihood for the missing entries. In this paper, the algorithm is extended to different cases other than low rank approximations under Frobenius norm, such as minimizing the Frobenius norm under nuclear norm constraint, spectral norm constraint, orthogonality constraint and more. The geometric interpretation of the proposed approximation process along with its optimality for convex constraints is also discussed. In addition, it is shown how the approximation algorithm can be used for matrix completion as well, under a variety of spectral regularizations. Its applications to physics, electrical engineering and data interpolation problems are also described.

Key words. Matrix approximation, Matrix completion.

AMS subject classifications. 15A83, 15A18.

1. Introduction.

Matrix completion and matrix approximation are important problems in a variety of fields such as statistics [18], biology [13], statistical machine learning [26], signal and computer vision/image processing [19], to name some. Rank reduction by matrix approximation is important for example in compression where low rank indicates the existence of redundant information. Therefore, low rank matrices are better compressed. In statistics, matrix completion can be used for survey completion and in image processing it is used for interpolation needs. In general, low rank matrix completion is an NP-hard problem, therefore, some relaxations methods have been proposed. For example, instead of solving the problem

$$\begin{array}{ll} & \text{minimize rank } (\mathbf{X}) \\ & \text{subject to } X_{i,j} = M_{i,j}, \quad (i,j) \in \Omega \end{array} \tag{1.1}$$

it can be approximated by

$$\begin{array}{ll} & \text{minimize } ||\mathbf{X}||_* \\ & \text{subject to } X_{i,j} = M_{i,j}, \quad (i,j) \in \Omega, \end{array} \tag{1.2}$$

[*] Received by the editors on January 22, 2012. Accepted for publication on August 1, 2012. Handling Editor: James G. Nagy.

[†] School of Electrical Engineering, Tel-Aviv University, Israel (gil@eng.tau.ac.il).

[‡] School of Computer Science, Tel-Aviv University, Israel (amir@math.tau.ac.il).

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 [9]. An iterative solution, which is based on a singular value thresholding, is given in [4]. A completion algorithm, based on the local information of the matrix, is proposed in [22]. This powerful approach enables to divide a large matrix into a set of smaller blocks, which can be processed in parallel and thus it suits large matrices processing.

In this paper, we are interested in a different yet similar problem:

$$\begin{array}{l}\text{minimize } \|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\|_F \\ \text{subject to } f(\mathbf{X}) \leq 0,\end{array} \tag{1.3}$$

where $\|\cdot\|_F$ is the Frobenius norm and $\mathcal{P}$ is a projection operator that indicates the entries we wish to approximate, i.e., $\mathcal{P}\mathbf{X} = \mathbf{B} \odot \mathbf{X}$, where $\mathbf{B}$ is a matrix of zeros and ones, and $\odot$ is a pointwise multiplication. This setup is also called Interest-Zone-Matrix-Approximation (IZMA). We show that a simple iterative algorithm for solving Eq. (1.3) (locally) exists if a solution for the full case matrix in Eq. (1.4)

$$\begin{array}{l}\text{minimize } \|\mathbf{X} - \mathbf{M}\|_F \\ \text{subject to } f(\mathbf{X}) \leq 0\end{array} \tag{1.4}$$

is known, where $f(\mathbf{X})$ is the same as in Eq. (1.3). If $f(\mathbf{X})$ is convex, the problem can be solved globally. A solution of Eq. (1.4) for $f(\mathbf{X}) = rank(\mathbf{X}) - k$ is known as the Eckart-Young Theorem [7] and it is given by the singular value decomposition (SVD) procedure.

However, when only some entries participate in the process, the solution provides more degrees of freedom for the approximation. Hence, there are many possibilities to approximate the matrix and the solution is not unique. When the problem is convex, there is one minimum.

A generalization of the Eckart-Young matrix approximation theorem is given in [12], where the low rank approximation of the matrix keeps a specified set of unchanged columns. An algorithm for solving the Interest-Zone-Matrix-Approximation problem in Eq. (1.3) for the low rank case appears in recent literature under different names such as “SVD-Impute” [28] and “Hard-Impute” [20], where the motivation for the solution method came from maximizing the likelihood over the missing entries by applying the EM algorithm and not from minimizing the mean squared error (MSE). The algorithm is:

$$\mathbf{X}_n = \mathcal{D}_k((\mathcal{I} - \mathcal{P})\mathbf{X}_{n-1} + \mathcal{P}\mathbf{M}), \tag{1.5}$$

where $\mathcal{D}_k\mathbf{X}$, which is the best rank k approximation (in Frobenius norm) for $\mathbf{X}$, keeps the first k singular values of $\mathbf{X}$ while zeroing the rest, i.e. $\mathcal{D}_k\mathbf{X} = \mathbf{U\Sigma_k V^T}$ and diag($\mathbf{\Sigma_k}$) $= (\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$. Therefore, the EM algorithm converges to a local maximum of the likelihood. However, this does not say anything about the MSE that we try to minimize. Along with the "Hard-Imput" algorithm in [20], the "Soft-Impute" is an additional algorithm that is similar to the algorithm in Eq. (1.5). The only difference is that $\mathcal{D}_k$ is replaced by a "softer" operator $\mathcal{B}_\alpha$ which zeros only the singular values of a given matrix that exceed a certain threshold α. The "Soft-Impute" algorithm is the solution for the following problem:

$$\begin{aligned} &\text{minimize } \|\mathbf{X}\|_* \\ &\text{subject to } \|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\|_F \leq \delta. \end{aligned} \tag{1.6}$$

A proof for the convergence of an algorithm for solving Eq. (1.6) where the error is monotonic decreasing is given in [20]. An attempt to extend Eq. (1.5) to weighted low rank approximations, such that the weights are not necessarily zero or one ($\mathcal{P}$ operator), is given in [26] by modifying Eq. (1.5) to become

$$\mathbf{X}_n = \mathcal{D}_k((\mathbf{1} - \mathbf{W}) \odot \mathbf{X}_{n-1} + \mathbf{W} \odot \mathbf{M}), \tag{1.7}$$

where $\mathbf{W}$ is a matrix whose elements satisfy $0 \leq w_{i,j} \leq 1$ and $\odot$ is pointwise multiplication. In this approach, the missing entries are filled iteratively to maximize the likelihood. The EM algorithm converges monotonically to the maximum likelihood but not necessarily to a local minimum of the MSE as can be seen in Appendix A. A correct algorithm with the correct proof for the weighted case is given in [15]. Solution to the case where the constraint is $\|\mathbf{X}\|_* < \lambda$ can be found by other methods, for examples method that involves optimization. A recent approach that uses the simplex approach can be found in [5]. Despite approximation methods for certain entries appear in recent literature, approximation methods under spectral norm, and orthogonality constraint have not been investigated. In this paper, we introduce new theorems that approximate full matrices under other constraints such as spectral and nuclear norm and prove that an algorithm that approximates certain entries using these theorems always converges and finds the global solution when the constraint is convex. The algorithm can be used for cases such as:

$$\begin{aligned} &\text{minimize } \|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\|_F \\ &\text{subject to } \mathbf{X}^T\mathbf{X} \text{ diagonal} \end{aligned} \tag{1.8}$$

or

$$\begin{aligned} &\text{minimize } \|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\|_F \\ &\text{subject to } \|\mathbf{X}\|_2 < \lambda \end{aligned} \tag{1.9}$$

and other cases.

The paper has the following structure: Related theorems on full matrix approximation is given in Section 2. Section 3 describes the new algorithm that approximates

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

a matrix by taking into account some of the matrix entries. The proposed IZMA algorithm was applied to different applications as described in Section 4. Appendix A shows via an example that the algorithm in [26] does not converge to a local minimum. Several inequalities, which are needed in the paper, are proved in Appendix B with additional new theorems.

2. 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,

$$\begin{aligned}&\text{minimize } \|\mathbf{X}-\mathbf{M}\|_F\\ &\text{subject to } \|\mathbf{X}\|_F \le \lambda.\end{aligned} \tag{2.1}$$

A solution for Eq. (2.1) is given by $\mathbf{X} = \frac{\mathbf{M}}{\|\mathbf{M}\|_\mathbf{F}} \min(\|\mathbf{M}\|_\mathbf{F}, \lambda)$.

Proof. The expression $\|\mathbf{X}\|_F^2 \le \lambda^2$ 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 $\|\mathbf{M}\|_F \le \lambda$, 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 $\|\mathbf{X}\|_F^2 \le \lambda^2$ 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

$$\begin{aligned}&\text{minimize } \|\mathcal{P}\mathbf{X}-\mathcal{P}\mathbf{M}\|_F\\ &\text{subject to } \|\mathbf{X}\|_F \le \lambda.\end{aligned} \tag{2.2}$$

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. (2.2).

Another example is the solution to the problem:

$$\begin{aligned}&\text{minimize } \|\mathbf{X}-\mathbf{M}\|_F\\ &\text{subject to } \mathbf{X}^T\mathbf{X} = \mathbf{I}.\end{aligned} \tag{2.3}$$

This is known as the orthogonal Procrustes problem ([25]) 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 $\mathbf{X}^{\mathbf{T}}\mathbf{X} = \mathbf{D}^{\mathbf{2}}$, 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. (2.3)) by subtituting $\mathbf{X} = \mathbf{V}\mathbf{D}$ where $\mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{I}$. When $\mathbf{D}$ is unknown, the problem can be solved by applying an iterative algorithm that is described in [8].

We now examine the following problem:

$$\begin{array}{l} \text{minimize } \|\mathbf{X} - \mathbf{M}\|_F \\ \text{subject to } \|\mathbf{X}\|_2 \leq \lambda. \end{array} \tag{2.4}$$

A solution to this problem uses the Pinching theorem ([2]):

Lemma 2.1 (Pinching theorem). *For every matrix* $\mathbf{A}$ *and unitary matrix* $\mathbf{U}$, *and for any norm satisfying* $\|\mathbf{U}\mathbf{A}\mathbf{U}^*\| = \|\mathbf{A}\|$, *it holds that* $\|\operatorname{diag}(\mathbf{A})\| \leq \|\mathbf{A}\|$.

A proof is given in [10]. An alternative proof is given in Appendix B.4.

Lemma 2.2 (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* $\mathbf{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$. *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* $\tilde{\sigma_i} = \min(\sigma_i, \lambda)$, $i = 1, \ldots, k$, $k \leq n$.

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, $\|\operatorname{diag}(\mathbf{U}^*\mathbf{X}\mathbf{V}) - \mathbf{\Sigma}\|_F \leq \|\mathbf{U}^*\mathbf{X}\mathbf{V} - \mathbf{\Sigma}\|_F$. By Lemma 2.1, $\|\operatorname{diag}(\mathbf{U}^*\mathbf{X}\mathbf{V})\|_2 \leq \|\mathbf{U}^*\mathbf{X}\mathbf{V}\|_2$. 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 $\mathbf{U}^*\mathbf{X}\mathbf{V} = \operatorname{diag}(\min(\sigma_i, \lambda))$, $i = 1, \ldots k$, $k \leq n$. 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. $\|\mathbf{X}\|_* = \|\tilde{\mathbf{\Sigma}}\|_* = \sum_i |\tilde{\sigma_i}| \leq \lambda, i = 1, \ldots k, k \leq n$. 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:

$$\begin{array}{l} \text{minimize } \sum_i (\tilde{\sigma_i} - \sigma_i)^2 \\ \text{subject to } \sum_i \tilde{\sigma_i} \leq \lambda \\ \text{subject to } \tilde{\sigma_i} \geq 0. \end{array} \tag{2.5}$$

This is a standard convex optimization problem that can be solved by methods such as semidefinite programming [3].

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

3. Approximation of certain entries. Suppose that we wish to approximate a matrix by taking into account some of its entries given by some projection operator. In other words, we are looking for the matrix $\mathbf{X}$ that minimizes the error function $\epsilon(\mathbf{X}) \triangleq \|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\|_F$ under certain constraint on $\mathbf{X}$ as in Eq. (1.3)assuming the solution for the full matrix problem $\|\mathbf{X} - \mathbf{M}\|_F$ in Eq. (1.4) is known. As was seen before, the constraint can be the rank constraint, the spectral norm constraint, or the nuclear norm constraint, orthonormality and others. In addition to the projection operator $\mathcal{P}$, we define the operator $\mathcal{D}$, which is a solution for the full matrix problem, i.e., the solution to Eq. (1.4). Another operator, denoted by $\mathcal{W}$, is the entries replacing the operator defined by $\mathcal{W}\mathbf{X} \triangleq (\mathcal{I} - \mathcal{P})\mathbf{X} + \mathcal{P}\mathbf{M}$, where $\mathcal{I}$ is the identity operator ($\mathcal{I}\mathbf{X} = \mathbf{X}$). The matrix $\mathbf{M}$ can be considered as a parameter of the operator which replaces the entries in $\mathbf{X}$ by the entries from $\mathbf{M}$ as indicated by the operator $\mathcal{P}$. $\mathcal{W}$ satisfies the following properties:

1. $\mathcal{P}\mathcal{W}\mathbf{X} = \mathcal{P}\mathbf{M}$;
2. $(\mathcal{I} - \mathcal{P})\mathcal{W}\mathbf{X} = (\mathcal{I} - \mathcal{P})\mathbf{X}$;
3. $\mathbf{X} - \mathcal{W}\mathbf{X} = \mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}$.

For simplicity, we define another operator by $\mathcal{T} \triangleq \mathcal{D}\mathcal{W}$.

THEOREM 3.1. *$\epsilon(\mathcal{T}^{n+1}\mathbf{X}) \leq \epsilon(\mathcal{T}^n\mathbf{X})$ for every $\mathbf{X}$, where $n \geq 1$ and ϵ is the error.*

Proof. Let $\mathcal{H}$ be the Hilbert space of all $m \times l$ matrices equipped with the standard inner product $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{trace}(\mathbf{X}^*\mathbf{Y})$, which induces the standard Frobenius norm $\|\mathbf{X}\|^2 = \text{trace}(\mathbf{X}^*\mathbf{X})$. Assume that $\mathbf{X}$ is an arbitrary matrix in $\mathcal{H}$ and let $\mathbf{M}$ be the matrix whose entries we wish to approximate according to the projection operator $\mathcal{P}$. Since $n \geq 1$, $f(\mathcal{T}^n\mathbf{X}) \leq 0$. Let $\mathbf{Q}$ be the locus of all matrices $\mathbf{Y}$ that satisfy $\mathcal{P}\mathbf{Y} = \mathcal{P}\mathbf{M}$. $\mathbf{Q}$ can be thought as a line parallel to the $\mathcal{I} - \mathcal{P}$ axis and perpendicular to the $\mathcal{P}$ axis - see Fig. 3.1. Note that the error $\epsilon(\mathbf{X})$ is the distance between the matrix point $\mathbf{X}$ and the line $\mathbf{Q}$. Applying $\mathcal{W}$ to $\mathcal{T}^n\mathbf{X}$, denoted by $\mathcal{W}\mathcal{T}^n\mathbf{X}$, which is the zero error matrix and $\mathcal{W}\mathcal{T}^n\mathbf{X}$ on $\mathbf{Q}$ does not necessarily satisfy the constraint. Applying $\mathcal{D}$ to $\mathcal{W}\mathcal{T}^n\mathbf{X}$ produces $\mathcal{T}^{n+1}\mathbf{X}$, which approximates $\mathcal{W}\mathcal{T}^n\mathbf{X}$ best, satisfies the constraint. Therefore, it must be inside a ball that is centered in $\mathcal{W}\mathcal{T}^n\mathbf{X}$ with radius $\|\mathcal{T}^n\mathbf{X} - \mathcal{W}\mathcal{T}^n\mathbf{X}\|$ so that $\|\mathcal{T}^{n+1}\mathbf{X} - \mathcal{W}\mathcal{T}^n\mathbf{X}\| \leq \|\mathcal{T}^n\mathbf{X} - \mathcal{W}\mathcal{T}^n\mathbf{X}\|$ (otherwise $\mathcal{T}^n\mathbf{X}$ approximates $\mathcal{W}\mathcal{T}^n\mathbf{X}$ better which contradicts the best approximation theorem for a full matrix) - see Fig. 3.1. Thus, we obtain:

$$\begin{array}{rcl} \|\mathcal{T}^{n+1}\mathbf{X} - \mathcal{W}\mathcal{T}^n\mathbf{X}\|^2 & = & \|(\mathcal{I} - \mathcal{P})\mathcal{T}^{n+1}\mathbf{X} - (\mathcal{I} - \mathcal{P})\mathcal{W}\mathcal{T}^n\mathbf{X}\|^2 + \\ \|\mathcal{P}\mathcal{T}^{n+1}\mathbf{X} - \mathcal{P}\mathcal{W}\mathcal{T}^n\mathbf{X}\|^2 & \leq & \|\mathcal{T}^n\mathbf{X} - \mathcal{W}\mathcal{T}^n\mathbf{X}\|^2 \\ & = & \|\mathcal{P}\mathcal{T}^n\mathbf{X} - \mathcal{P}\mathbf{M}\|^2 \end{array} \tag{3.1}$$

where in Eq. (3.1) we used the third property of $\mathcal{W}$ and since (according to the first property of $\mathcal{W}$) $\|\mathcal{P}\mathcal{T}^{n+1}\mathbf{X} - \mathcal{P}\mathcal{W}\mathcal{T}^n\mathbf{X}\| = \|\mathcal{P}\mathcal{T}^{n+1}\mathbf{X} - \mathcal{P}\mathbf{M}\|$ we finally obtain

$\|\mathcal{P}\mathcal{T}^{n+1}\mathbf{X} - \mathcal{P}\mathbf{M}\| \leq \|\mathcal{P}\mathcal{T}^{n}\mathbf{X} - \mathcal{P}\mathbf{M}\|$.

Equality holds if and only if $(\mathcal{I} - \mathcal{P})\mathcal{T}^{n+1}\mathbf{X} = (\mathcal{I} - \mathcal{P})\mathcal{W}\mathcal{T}^{n}\mathbf{X} = (\mathcal{I} - \mathcal{P})\mathcal{T}^{n}\mathbf{X}$. □

Geometrically, the algorithm means that in each iteration, our current matrix is projected onto $\mathbf{Q}$. Then, it is approximated by $\mathcal{D}$ to a rank k matrix. The new rank k matrix must be inside a ball centered at the current point in $\mathbf{Q}$ and its radius is the distance to the previous rank k matrix iteration. The new point is projected again onto $\mathbf{Q}$. It continues this way till the radius of each ball is becoming smaller and smaller after each iteration. This is illustrated in Fig. 3.1. This means that the algorithm eventually converges. The convergence speed depends on the convergence value $\kappa_n = \|(\mathcal{I} - \mathcal{P})\mathcal{T}^{n+1}\mathbf{X} - (\mathcal{I} - \mathcal{P})\mathcal{T}^{n}\mathbf{X}\|$. If this value becomes smaller then the algorithm will converge slowly. When $\kappa = 0$, it means that the algorithm reached a convergence point. Different methods for measuring the convergence rate, which originated from the geometry, exist. For example, a good relative measure is $\frac{\text{dist}(\mathcal{P}\mathbf{X_{k-1}},\mathbf{Q})}{\text{dist}(\mathcal{P}\mathbf{X_{k}},\mathbf{Q})}$.

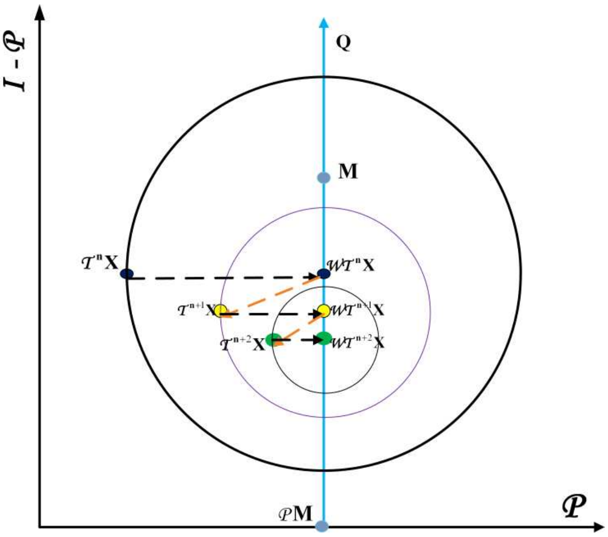


Fig. 3.1: Geometric illustration how the radius of each ball in the proof of Theorem 3.1 is getting smaller and smaller.

Algorithm 1 implements Theorem 3.1.

Theorem 3.1 shows the algorithm converges. However, it does not say anything about its convergence to the global solution even for the case the constraint is convex. Each IZMA iteration can be considered as a projected gradient operation:

$$\mathbf{X}_{n+1} = \mathcal{D}(\mathbf{X}_n - \mu_n \mathcal{P}(\mathbf{X}_n - \mathbf{M})) \tag{3.2}$$

with fixed $\mu_n = 1$. A sequence μ_n such that $\mathbf{X}_n$ converges to the global solution exists when the constraint form a convex set (such as for the cases $\|\mathbf{X}\|_* < \lambda$ and $\|\mathbf{X}\|_2 < \lambda$). Finding the optimal step size μ_n is done by applying Armijo rule while minimizing the solution in each iteration. In this case, the convergence to the global solution is guaranteed and the step size is given by ([14]):

$$\begin{aligned}
&l[n] = \text{argmin}_{j \in \mathcal{Z}_{\geq 0}} : f(\mathbf{X}_{n,j}) \leq f(\mathbf{X}_n) - \sigma \text{trace}(\nabla f(\mathbf{X}_n)^T (\mathbf{X}_n - \mathbf{Z}_{n,j}))\\
&\mathbf{Z}_{n,j} = \mathcal{D}(\mathbf{X}_n - \tilde{\mu} 2^{-j} \nabla f(\mathbf{X}^n))\\
&\mu_n = \tilde{\mu} 2^{-l[n]}
\end{aligned} \tag{3.3}$$

where $f(X) = \frac{1}{2}\|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\|_F^2$, $\tilde{\mu} > 0$ and $\sigma \in (0,1)$. Since convergence is achieved by choosing the best μ_n in each iteration we now show that for our case there is no need to compute the step size in every iteration since the optimal step size is $\mu_n = 1$.

Theorem 3.2. *For the matrix approximation problem defined in Eq. (1.3), the optimal step size in each iteration of Eq. (3.2) is given by* $\mu_n = 1$.

Proof. Let $\mathbf{X}_n$ be a current point in the iterative process that satisfies the constraint (i.e., $n \geq 1$) and let $\mathbf{Q}$ be the geometric region of all the matrices $\mathbf{X}$ satisfies $\|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\| = 0$. The geometric interpretation of an error for a given point $\mathbf{X}$ is the horizontal distance between $\mathbf{X}$ and $\mathbf{Q}$. Let $\mathbf{Y} = \mathbf{X}_n - \mu\mathcal{P}(\mathbf{X}_n - \mathbf{M})$ with $0 < \mu < 1$ and let $\tilde{\mathbf{Y}} = \mathbf{X}_n - \mathcal{P}(\mathbf{X}_n - \mathbf{M})$. Note that the difference between $\mathbf{Y}$ and $\tilde{\mathbf{Y}}$ is strictly on the $\mathcal{P}$ axis and that $\mathbf{Y}$ is between $\mathbf{X}_n$ and $\mathbf{Q}$. $\mathcal{D}$ maps $\mathbf{Y}$ to $\mathcal{D}\mathbf{Y}$ which is the best approximation to $\mathbf{Y}$ satisfies the constraint. This point must be inside ball $\mathcal{B}_1$ centered at $\mathbf{Y}$ with radius $\|\mathbf{Y} - \mathbf{X_n}\|_F$. On the other hand, $\mathcal{D}\tilde{\mathbf{Y}}$ is in ball $\mathcal{B}_2$, centered in $\tilde{\mathbf{Y}}$ whose radius is $\|\tilde{\mathbf{Y}} - \mathbf{X}_n\|_F$ and is the best approximation to $\tilde{\mathbf{Y}}$. Because $\mathcal{D}\mathbf{Y}$ satisfies the constraint then $\mathbf{D}\tilde{\mathbf{Y}}$ must be inside a smaller ball, whose radius is $\|\tilde{\mathbf{Y}} - \mathcal{D}\mathbf{Y}\|_F$. Note that in ball $\mathcal{B}_3$ whose center is $\mathbf{Y}$ and its radius $\|\mathbf{Y} - \mathcal{D}\mathbf{Y}\|_F$ there are no points satisfy the constraint, hence $\mathcal{D}\tilde{\mathbf{Y}} \notin \mathcal{B}_3$. Along with the fact that the line connecting $\mathbf{Y}$ and $\tilde{\mathbf{Y}}$ is parallel to the $\mathcal{P}$ axis we get that $\|\mathcal{P}\mathcal{D}\tilde{\mathbf{Y}} - \mathcal{P}\mathbf{M}\|_F \leq \|\mathcal{P}\mathcal{D}\mathbf{Y} - \mathcal{P}\mathbf{M}\|_F$ which means that in every iteration, choosing $\mu < 1$ will lead to an error greater (or equal) to the error achieved for choosing $\mu = 1$. This completes the proof showing $\mu = 1$ is the best choice. □

Illustration of the proof is given in Figure 3.2. Theorem 3.2 is important in the sense of computational efficiency, since it enables us not to compute the optimal step size in each iteration. Computation of the optimal step size by using equation (3.3) requires applying $\mathcal{D}$ several times in each iteration (which is very often the most computationally heavy part), instead of just once. Though convergence is guaranteed only for cases of convex constraint, there are cases where the convergence of the algorithm is guaranteed when the constraint is rank constraint. Those conditions are related with the restricted isometry property (RIP). For more information the reader is referred to [11] and [21].

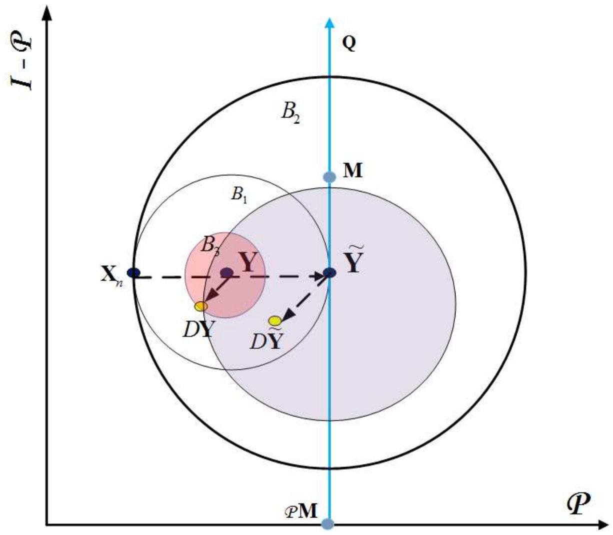


Fig. 3.2: Geometric illustration showing the optimality of $\mu = 1$ (Theorem 3.2).

```
Algorithm 1 Interest Zone Matrix Approximation
Input: $\mathbf{M}$ - matrix to approximate,
$\mathbf{P}$ - projection operator that specifies the important entries,
$\mathbf{B}$ - matrix of 0 and 1,
$\mathbf{X}_0$ - initial guess,
$\mathcal{D}$ - full matrix approximation operator.

Output: $\mathbf{X}$ - Approximated matrix.
Set $\mathbf{X} \leftarrow \mathcal{D}\mathbf{X}_0$ ($\mathbf{X}$ is set by solving (1.4) to be the best approximation of $\mathbf{X}_0$ under the constraint.)
repeat
    $\mathbf{X} \leftarrow \mathcal{W}\mathbf{X}$ (The entries we want to approximate in $\mathbf{X}$ are replaced by the known entries from $\mathbf{M}$ according to $\mathbf{B}$).
    $\mathbf{X} \leftarrow \mathcal{D}\mathbf{X}$ ($\mathbf{X}$ is set by solving (1.4) to be the best approximation of $\mathbf{X}_0$ under the constraint)
until $\|\mathcal{P}\mathbf{X} - \mathcal{P}\mathbf{M}\|_F$ converges
Return $\mathbf{X}$
```

Both the convergence speed and the final matrix that the algorithm converges to depend on the initial matrix $\mathbf{X}_0$. If $\mathcal{D}\mathbf{X}_0$ mainly approximate the values of $(\mathcal{I} - \mathcal{P})\mathbf{X}_0$, then the application of $\mathcal{D}$ will not change $(\mathcal{I} - \mathcal{P})\mathbf{X}_0$ significantly but will change $\mathcal{P}\mathbf{X}_0$.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

To avoid it, the values of $(\mathcal{I}-\mathcal{P})\mathbf{X}_0$ should be at the same order of magnitude as $\mathcal{P}\mathbf{M}$. Application of $\mathcal{W}$ will bring it back very close to the previous iteration. Thus, the algorithm will iterate near two points that are changed very slowly if at all. To avoid from having the algorithm converges to a local minimum, it is suggested to use several initial guesses. As an example, the following numerical example, which shows that the algorithm does not always converge to the global minimum but rather depends on the starting point, is presented. Suppose we wish to approximate by a rank 2 matrix the following full rank 3 × 3 matrix:

$$\mathbf{M}=\begin{bmatrix}1 & 1 & 1\\ 0 & 0.75 & 0.25\\ 0 & 0.25 & 0.75\end{bmatrix},$$

where the interest points are indicated by

$$\mathbf{B}=\begin{bmatrix}0 & 0 & 0\\ 0 & 1 & 1\\ 0 & 1 & 1\end{bmatrix}.$$

If we take as an initial guess $\mathbf{X}_0=\mathbf{M}$, then, after the first iteration, we obtain the matrix

$$\mathbf{X}_1=\begin{bmatrix}1 & 1 & 1\\ 0 & 0.5 & 0.5\\ 0 & 0.5 & 0.5\end{bmatrix}$$

which is a rank 2 matrix that is mapped by $\mathcal{T}$ to itself, i.e., $\kappa=0$ and $\mathbf{X}_i=\mathbf{X}_1$ for $i\geq 1$. In the rank reduction part of the algorithm, the operator $\mathcal{D}$ reduces the rank of $\mathcal{W}\mathbf{X}_1$ but the values of $(\mathcal{I}-\mathcal{P})\mathcal{W}\mathbf{X}_1$ remain unchanged. For example, if we start from a random matrix

$$\mathbf{X}_0=\begin{bmatrix}0.553 & 0.133 & -1.58\\ -0.204 & 1.59 & -0.0787\\ -2.05 & 1.02 & -0.682\end{bmatrix},$$

then eventually we will get the matrix

$$\mathbf{X}_{100}=\begin{bmatrix}0.854 & 0.685 & -1.25\\ -1.32 & 0.75 & 0.25\\ -1.37 & 0.25 & 0.75\end{bmatrix}$$

which has the error $\epsilon(\mathbf{X}_{100})=\|\mathcal{P}\mathbf{X}_{100}-\mathcal{P}\mathbf{M}\|=0$.

4. Applications.

In this section, we show different applications where the IZMA algorithm (Algorithm 1) is utilized.

4.1. Matrix completion. Matrix completion is an important problem that has been investigated extensively recently. 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 researched 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, though other constraints can be used such as spectral norm minimization.

Algorithm 2 Matrix Completion: Nuclear Norm / Spectral Norm Minimization

```
Input:
M - matrix to complete, P - projection operator that specifies the important entries,
B - matrix of 0 and 1. 0 - entry to complete

Output: X - Completed matrix
$M \leftarrow M \odot B$
$\lambda_{min} \leftarrow 0$
$\lambda_{max} \leftarrow \|M\|_*$
repeat
    $\lambda_{prev} \leftarrow \lambda$
    $\lambda \leftarrow (\lambda_{min} + \lambda_{max})/2$
    $\mathbf{X} \leftarrow$ IZMA to approximate $\mathbf{M} \odot \mathbf{B}$ for points $\mathbf{B}$ 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 X
```

Since for the convex case, IZMA converges to the global solution, matrix completion can be achieved 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) [4] require additional parameters τ and δ that affect the convergence and the final result, where in the IZMA algorithm no external parameters are required. The disadvantage is that the computational complexity of one IZMA iter-

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

ation is similar to the computational complexity of the SVT and it has to be applied several times for the binary search to find the correct nuclear norm value.

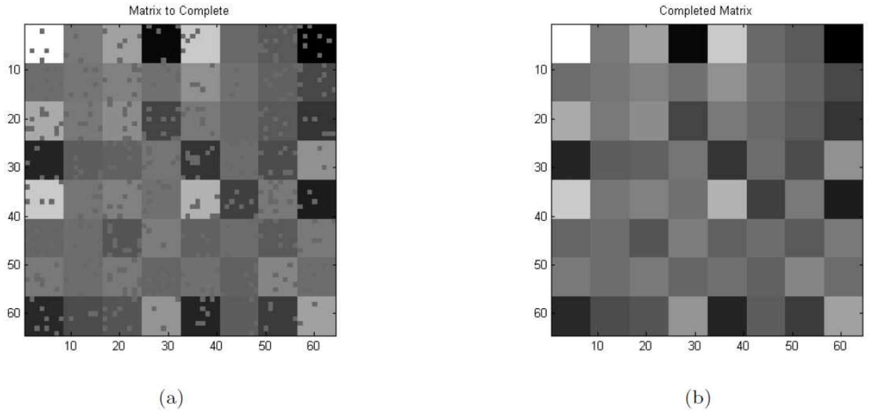


Fig. 4.1: (a) Matrix to complete. (b) Completed matrix with minimal nuclear norm.

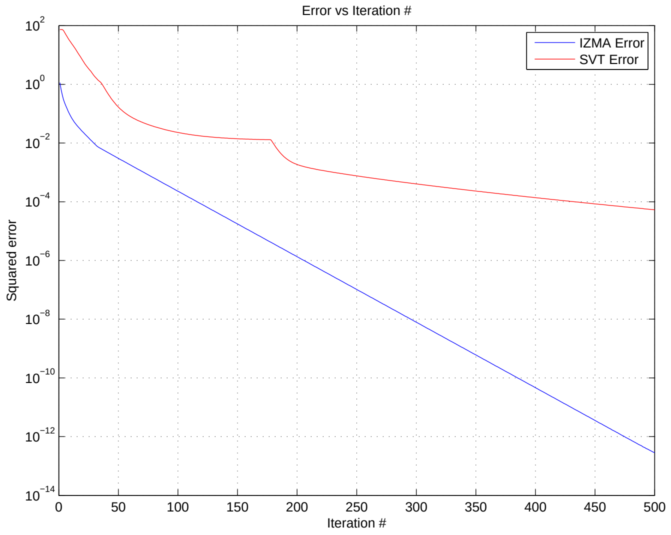


Fig. 4.2: Convergence rate comparison.

Figure 4.1 shows an example for matrix completion. The original nuclear norm

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

of the matrix to complete, with zeros in the location of the missing entries is 26.1. Algorithm 2 was used for the matrix completion, achieving nuclear norm of 15.31. The completed matrix is also low rank, just like the original full matrix. This example also shows why the nuclear norm if often used to approximate rank. Figure 4.2 shows the convergence of the IZMA algorithm assuming the nuclear norm value is known, and compare it to SVT. In reality, IZMA will need to run several times for finding the correct nuclear norm value, so it can outperform SVT when the value of the desired nuclear norm is to be searched on a small interval.

4.2. Image interpolation and approximation. Interpolation reconstructs a discrete function $I[m,n]$ (or a continuous function $I(x,y)$) from a discrete set Ω. Most interpolation methods try to restore the function by assuming it can be spanned by a set of basis functions (called "kernel"). Typical basis functions are splines, polynomials, trigonometric functions, wavelets, radial functions, etc. For example, in order to approximate $I(\mathbf{x}) = I(x,y)$ with a Gaussian radial basis function such as $\phi(r) = \exp(-\beta r^2)$ for some $\beta > 0$, then the approximating function can be written as $Y(\mathbf{x}) = \sum_{i=1}^{N} a_i \phi(\|(\mathbf{x} - c_i)\|_2)$ where $\{c_i\}_{i=1}^{N}$ are the centers in which we lay the radial functions on. $\{a_i\}_{i=1}^{N}$ are the coefficients of the functions, which can be found by solving $\mathbf{a}^* = \text{argmin}\|Y(\mathbf{x}) - I(\mathbf{x})\|_2, \quad \mathbf{x} \in \Omega$. This solves the standard least squares problem on the discrete set Ω.

As was stated above, the same procedure can be repeated for different kernels by minimizing a different metric such as l_1, l_2 or l_∞. It is important to mention that different kernels produce different results. A-priori knowledge about the physical nature of the function we wish to interpolate can be an important input for choosing the interpolation kernel [30]. For example, audio signals are usually spanned well (i.e., they require a small number of coefficients) when trigonometric functions are used, where other signals such as Chirp or Linear FM that are used in radar systems [17] are better adjusted to wavelets or Gabor functions. However, since SVD has the best energy compaction property from all the separable functions, it can be used to find on the fly the appropriate basis functions.

Our approach, which is based on SVD, does not require any a-priori knowledge for the interpolation procedure. It finds it from the available data. A disadvantage of this method is that it is not suitable for sparse data reconstruction. When the data is too sparse, there is insufficient information to extract the most suitable basis functions.

The example in Fig. 4.3 compares between the approximations of missing data through the application of the IZMA algorithm for approximation under rank constraint (though the nuclear norm could also have been used) or the completion, and a standard approximation method that uses the GP interpolation method with Fourier basis functions.

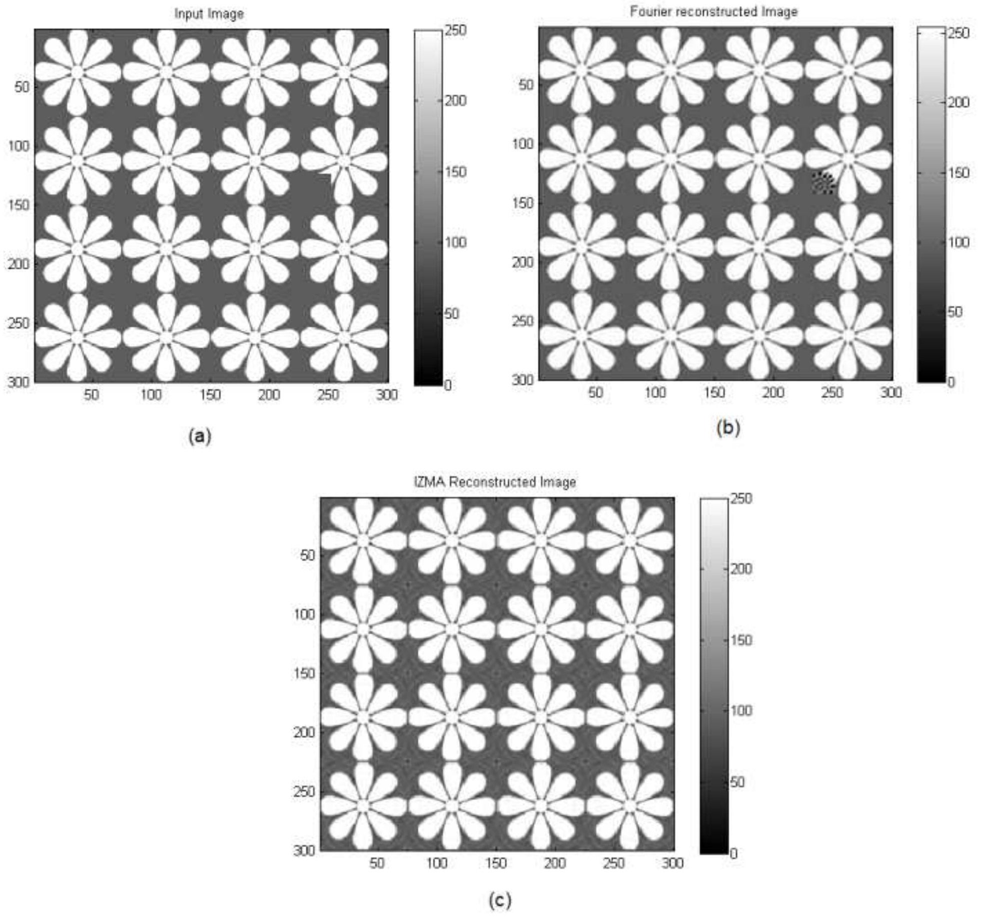
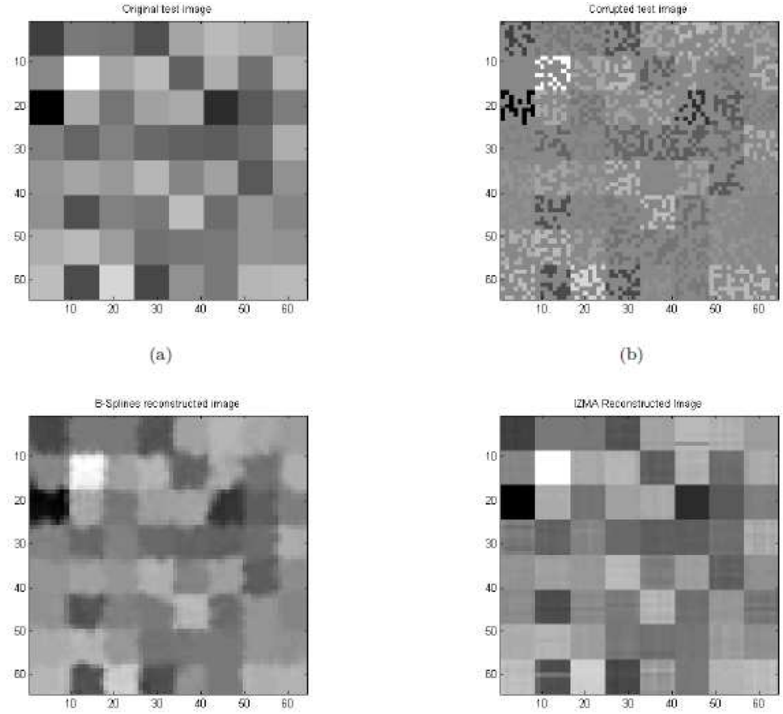


Fig. 4.3: (a) The original input image for the interpolation process. (b) Approximation by GP. (c) Approximation by the IZMA algorithm.

We see from Fig. 4.3 that the IZMA algorithm completed the flower image (of size 300 × 300 pixels) correctly since the basis functions that were used are the flowers components. The Fourier basis functions, on the other hand, failed to reconstruct the flower. The Fourier l_2 error (MSE) is 0.066 (normalized by the number of gray-levels) while the IZMA l_2 error (MSE) is 0.05. Also, from the rank perspective, the Fourier based reconstructed image of rank 131 and the IZMA based algorithm produced a rank 15 matrix. The original image rank was 223.

Another example is illustrated in Fig. 4.4 where the test image was produced from a combination of Haar-wavelet basis functions. 60% of the data was missing. It was restored by the application of the IZMA algorithm to approximate a matrix of rank 7 and the by multilevel B-Splines ([16]). Image size is 64 × 64.



(a) (b)

(c) (d)

Fig. 4.4: (a) The original input image. (b) The image where 60% of the entries are 0. (c) The reconstruction by the application of B-Splines. (d) The reconstruction by the application of the IZMA algorithm.

The RMS error (normalized by the number of gray-levels) after 100 iterations using the IZMA algorithm was 0.016 compared with 0.036 by the multilevel B-splines algorithm. It indicates that the IZMA algorithm found the suitable basis functions and thus achieved a smaller error with a better visual effect.

4.3. Reconstruction of physical signals. A typical family of matrices that have low rank can be originated from PDEs that are solved by separation of variables. In this case, the solution is given as a sum such as $U(x,y) = \sum_{n=1}^{N} X_n(x)Y_n(y)$.

Note that when the solution is stored as a matrix, then the element $X_n(x)Y_n(y)$ is discretized and stored as $\mathbf{XY^T}$ where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors and $\mathbf{XY^T}$ is a matrix of rank 1. After summation, the obtained rank is N since the functions of the solution are linear independent.

As an example, we examine the propagation of an electromagnetic wave inside a cylindrical waveguide of radius R. The electromagnetic waves that travel inside the waveguide are called *modes* and they depend on the input frequency and on the geometry of the waveguide. Usually waveguides are designed to support only one mode. We assume that this is the case. The primary mode and the most important for cylindrical waveguide is the first Transverse Electric mode denoted as TE$_{11}$. TE modes do not have electric field in the z direction but only the magnetic field H_z that is called the “generating field”. The rest of the fields can be derived from it. More information is given in [24]. H_z is found by solving the Hemholtz equation

$$\nabla^2 H_z + k^2 H_z = 0, \quad H_z(R,\theta,z) = 0 \ , \tag{4.1}$$

where ∇^2 is the Laplacian operator in cylindrical coordinates (r,θ,z), $k = \frac{2\pi}{\lambda}$ is the wavenumber and λ is the wavelength. The solution of Eq. (4.1) is known and for TE$_{11}$ it is given by

$$H_z(r,\theta,z) = (A\sin\theta + B\cos\theta)J_1(k_c r)e^{-i\beta z}, \tag{4.2}$$

where $J(x)$ is the Bessel function of the first kind, k_c is the cut-off wavenumber which for TE$_{11}$ is the first zero of $J_1'(x)$ divided by R (in our case $k_c = \frac{1.84}{R}$) and $\beta^2 = k^2 - k_c^2$. For a mode to exist in the waveguide, its cut-off wavenumber k_c must be smaller than k. Hence, λ can be chosen such that only the first mode will excite in the waveguide. The z-axis has only phase accumulation along the waveguide and this is not very interesting. We will investigate the modes as a function of (r,θ).

Assume that the image in Fig. 4.5 is corrupted such that 85% of the data is missing as shown in Fig. 4.6 and it has to be restored . Note that neither information on the geometry of the waveguide nor the wavelength is required. The only required parameter is *the number* of modes, which as we saw earlier, is equal to the rank of the matrix.

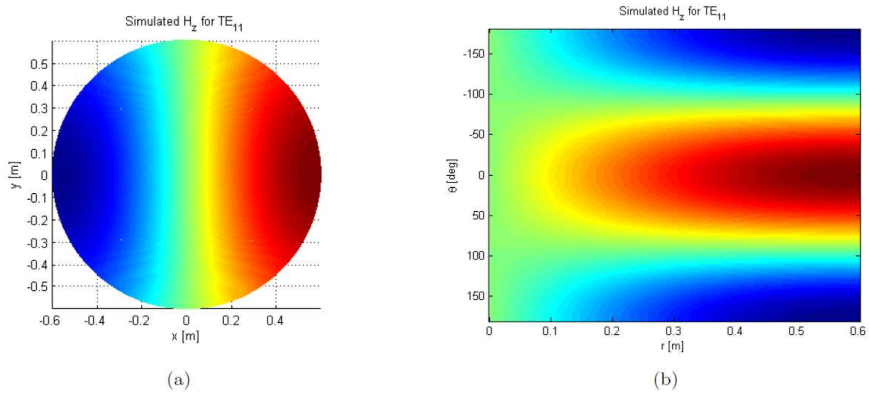


Fig. 4.5: Simulated magnetic field $H_z = \cos(\theta)J_1(k_c r)$, $R = 0.6$m in different coordinate systems. (a) H_z in TE_{11} mode in Cartesian coordinates. (b) H_z in TE_{11} mode in Polar coordinates. Both images are 200×200 pixels.

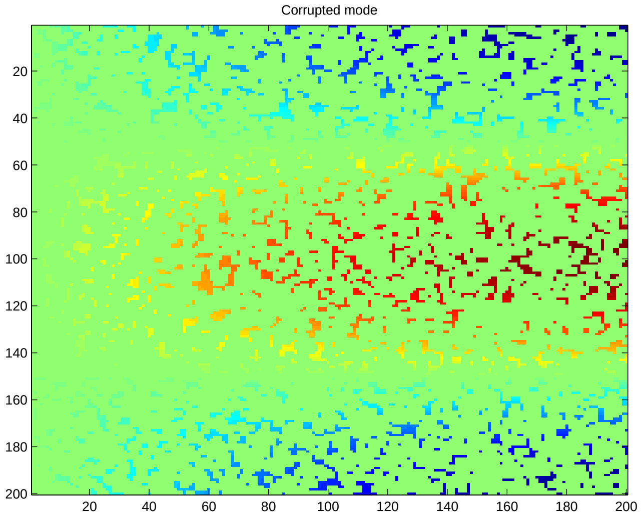


Fig. 4.6: Corrupted TE_{11} mode of Fig. 4.5 in a circular waveguide.

The results from 1000 iterations of the IZMA algorithm is compared with the results from the application of the multilevel B-Splines as shown in Fig. 4.7 and the error is shown in Fig. 4.8.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

ELA

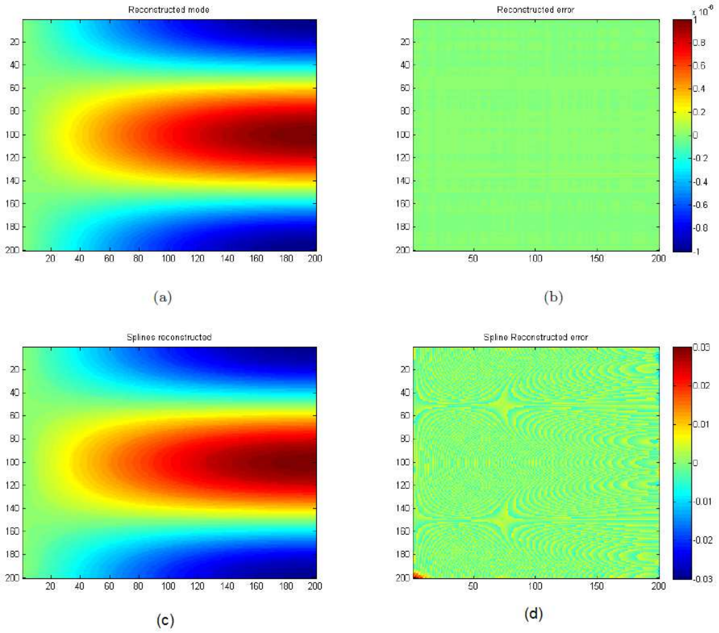


Fig. 4.7: (a) Reconstruction by the application of the IZMA algorithm. (b) The error between the image in (a) and the source image in Fig. 4.5. (c) The reconstructed image from the application of the multilevel B-splines. (d) The error between the image in (c) and the source image in Fig. 4.5.

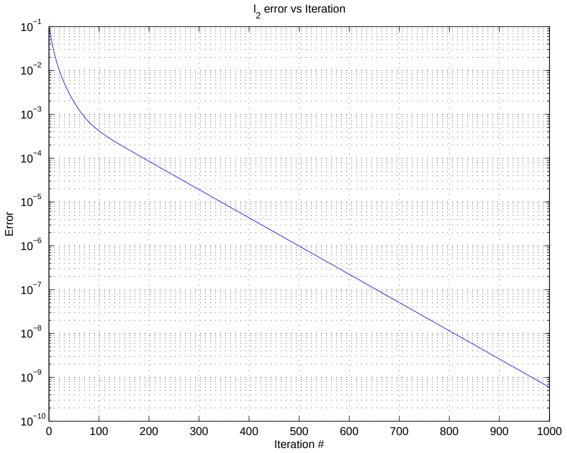


Fig. 4.8: RMS error vs. iteration number.

4.4. Masked SVD. Another useful application of the IZMA algorithm is the SVD calculation of a certain region of a matrix. For example, a matrix can be full rank but may contain a circular region which can be considered as 'rank 1'. The interest zone (or the shape) is defined by the operator $\mathcal{P}$. For example, suppose $\mathbf{M}$ is an $m \times n$ matrix of rank m but there may exist a matrix $\mathbf{X} = \mathbf{U\Sigma V^T}$ of rank $k < m$ such that

$$\mathcal{P}\mathbf{M} = \mathcal{P}\mathbf{X} = \mathcal{P}(\mathbf{U\Sigma V^T}). \tag{4.3}$$

Equation (4.3) can be thought as a way to determine the rank of a sub-region of a matrix and its SVD is calculated when only a certain region is taken into consideration. Note that not always there exists a matrix $\mathbf{X}$ with a lower rank that satisfies Eq. (4.3).

Figure 4.9 shows a 200×200 matrix $\mathbf{M}$ of rank 200 created by a Gaussian noise with zero mean and standard deviation of 1 whose center was replaced by a circle of values one as shown in Fig. 4.9(a). Figure 4.9(c) shows a rank 1 matrix $\mathbf{M}$ that approximates the matrix perfectly within the circle so that $\mathcal{P}\mathbf{M} = \mathcal{P}\mathbf{X}$.

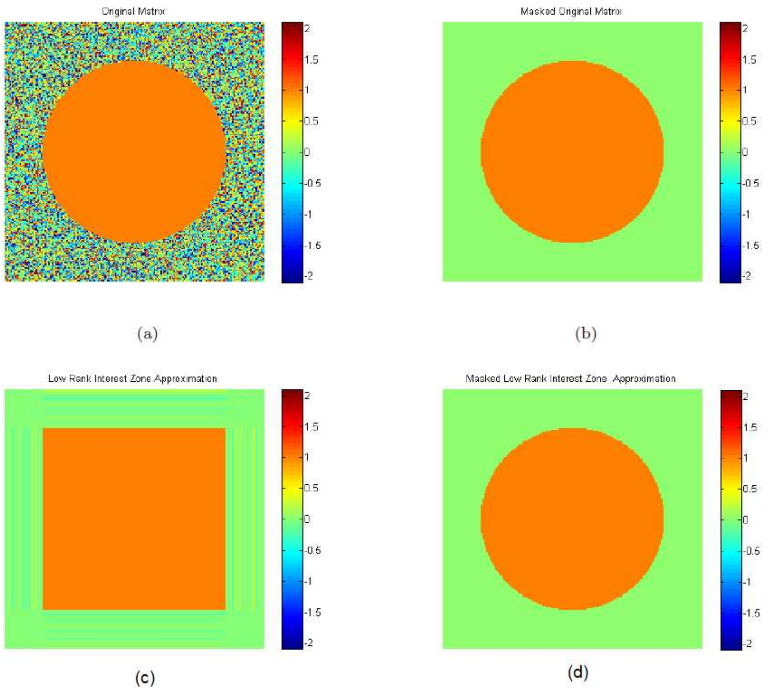


Fig. 4.9: (a) Original matrix $\mathbf{M}$, rank($\mathbf{M}$) = 200. (b) The projected matrix $\mathcal{P}\mathbf{M}$ (zeros outside the circle). (c) The interest zone approximated rank 1 matrix $\mathbf{X}$. (d) $\mathcal{P}\mathbf{X}$ matrix.

5. Conclusions. Theoretical and algorithmic work on matrix approximation and on matrix completion accompanied by several applications such as image interpolation, data reconstruction and data completion are presented in the paper. The full matrix approximation theorems includes approximation under different norms and constraints. The theory is also given a geometrical interpretation. In addition, we proved the convergence of the algorithms to global solution for convex constraints.

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.

Appendix A. Weighted low rank approximation.

Suppose we use the following input for the algorithm:

$$\mathbf{M} = \begin{bmatrix} 0.86 & 0.0892 \\ 0.519 & 0.409 \end{bmatrix}, \mathbf{X_0} = \begin{bmatrix} 0.171 & 0.378 \\ 0.957 & 0.821 \end{bmatrix}, \mathbf{W} = \begin{bmatrix} 0.115 & 0.712 \\ 0.731 & 0.34 \end{bmatrix}.$$

In each step, we get the following error:

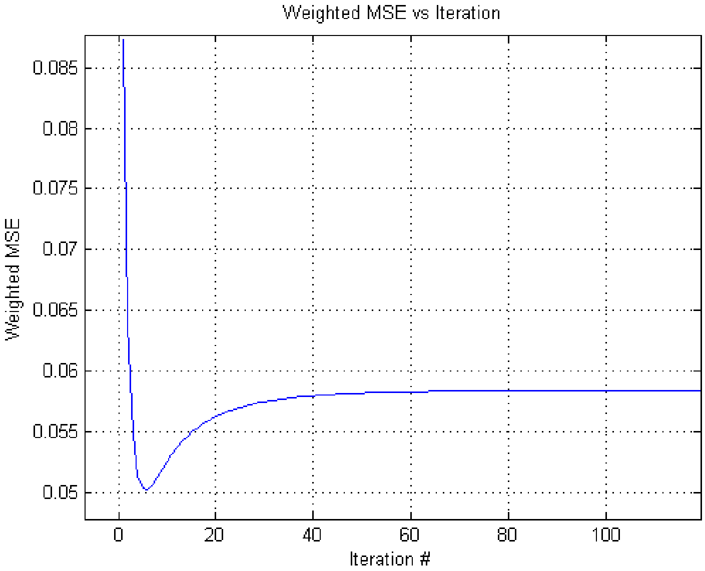


Fig. A.1: A weighted MSE for an arbitrary 2×2 matrix of rank 2 being approximated by a rank 1 matrix.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

Figure A.1 shows that the error decreased and at some point it begins to increase and finally converges to a point whose error is larger. Hence, the algorithm does not converge to a local minimum.

Appendix B. Majority inequalities.

In this appendix, we bring a different proof for the Pinching theory that is based on Jacobi rotations for the Schatten norm. This technique can be used to prove additional theorems that involve eigenvalues and diagonal elements of a matrix. As an additional example, we use this technique to prove the Fischer's inequality and some new inequalities involving exponential and logarithmic functions.

For a matrix $\mathbf{A}_{m\times n}$, the Schatten norm is defined as:

$$\|\mathbf{A}\|_p = \Big(\sum_{i=1}^{\min(m,n)} \sigma_i^p \Big)^{1/p} \tag{B.1}$$

where σ_i are the singular values of $\mathbf{A}$ and $p \in [1, \infty)$. For $p = \infty$, the Schatten norm coincides with the spectral norm and equals to the largest singular value. For $p = 1$, the norm coincides with the nuclear norm and equals to the singular values sum. Note that the Schatten norm is unitary invariant, i.e., $\|\mathbf{UAV}\|_p = \|\mathbf{A}\|_p$, for unitary $\mathbf{U}$ and $\mathbf{V}$.

Jacobi rotations are used to reduce a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ to a diagonal form using rotation matrices. The idea is to reduce the norm of the off-diagonal entries of $\mathbf{A}$ by using the rotation matrix $\mathbf{Q}$. $\mathbf{Q}$ is an $n \times n$ matrix that is equal to the identity matrix except for four entries, given by:

$$\begin{aligned} q_{kk} &= q_{ll} = \cos\theta \\ q_{kl} &= \sin\theta \\ q_{lk} &= -\sin\theta \end{aligned} \tag{B.2}$$

$$\mathbf{B} = \mathbf{Q^T A Q}, \tag{B.3}$$

where θ is chosen to minimize the off-diagonal part of $\mathbf{B}$ that it is given by:

$$\tau \triangleq \cot\theta = \frac{a_{ll} - a_{kk}}{2a_{kl}}, \quad t \triangleq \tan\theta = \frac{\operatorname{sign}(\tau)}{|\tau| + \sqrt{1+\tau^2}}. \tag{B.4}$$

Theorem B.1 (The main theorem). *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a symmetric matrix and let $\mathbf{B} = \mathbf{J^T A J}$ be its Jacobi rotation for the entries (k, l). Assume that $a_{kk} \geq a_{ll}$. Then, $b_{kk} \geq a_{kk}$ and $b_{ll} \leq a_{ll}$. More precisely, $b_{kk} = a_{kk} + \delta$ and $b_{ll} = a_{ll} - \delta$, $(\delta \geq 0)$.*

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

Proof. The proof uses the Jacobi roation. In each application of Jacobi rotation matrix to $\mathbf{A}$, the norm of the off-diagonal part is getting smaller and the diagonal part changes as well. By simple calculations, it is possible to find the following update equations for the new diagonal:

$$\begin{aligned} b_{kk} &= a_{kk} - t a_{kl} \\ b_{ll} &= a_{ll} + t a_{kl}. \end{aligned} \tag{B.5}$$

The sign of t is equal to the sign of τ. The sign of τ depends on the three entries a_{kk}, a_{kl}, a_{ll} as shown from the expression $\tau = \cot\theta = \frac{a_{ll}-a_{kk}}{2a_{kl}}$. We divide it into four cases:

1. $a_{kl} > 0$, $a_{ll} > a_{kk}$: In this case, τ is positive and therefore t is positive. According to Eq. (B.5), $b_{kk} < a_{kk}$ and $b_{ll} > a_{ll}$. The smallest entry a_{kk} becomes even smaller and the largest entry a_{ll} becomes even larger;
2. $a_{kl} > 0$, $a_{ll} < a_{kk}$: Here t is negative and according to Eq. (B.5), a_{kk} is getting larger and a_{ll} is getting smaller ($b_{kk} > a_{kk}$ and $b_{ll} < a_{ll}$);
3. $a_{kl} < 0$, $a_{ll} > a_{kk}$: t is negative, a_{kk} is getting smaller and a_{ll} is getting larger ($b_{ll} > a_{ll}$ and $b_{kk} < a_{kk}$);
4. $a_{kl} < 0$, $a_{ll} < a_{kk}$: t is positive, a_{ll} is getting smaller and a_{kk} is getting larger ($b_{ll} < a_{ll}$ and $b_{kk} > a_{kk}$).

The conclusion from the application of the Jacobi rotation is that the largest diagonal entry becomes even larger and the smallest diagonal entry becomes even smaller, therefore, $\max(|b_{kk}|, |b_{ll}|) \geq \max(|a_{kk}|, |a_{ll}|)$. At convergence, we obtain a diagonal matrix whose entries are the eigenvalues of the initial matrix, but the p-norm of each of the two modified entries increases because of the identity $|x+a|^p+|y-a|^p \geq |x|^p+|y|^p$ for every $x \geq y$ and $a \geq 0$. Hence, the p-norm of the diagonal can only increase between consecutive iterations as long as the off-diagonal part is not zero. □

Lemma B.2. *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric matrix. Then, $\lambda_{\min}(\mathbf{A}) \leq min\ a_{ii}$ and $\lambda_{\max}(\mathbf{A}) \geq max\ a_{ii}$.*

Proof. Since the Jacobi rotations converge to a diagonal matrix whose entries are the eigenvalues, and since in every iteration a pair of entries on the diagonal is changed so that the largest entry is getting even larger and the smallest entry is getting even smaller, then the smallest eigenvalue cannot be larger than the smallest entry on the diagonal. The same argument applies for the largest eigenvalue. □

Theorem B.3 (Pinching for the Schatten norm). *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric matrix. Then, $\|\operatorname{diag}(\mathbf{A})\|_p \leq \|\mathbf{A}\|_p$.*

Proof. We apply the Jacobi rotation to $\mathbf{A}$ such that $\mathbf{B} = \mathbf{J^T A J}$ while operating on entry (k, l). Suppose $a_{kk} \geq a_{ll}$ and $\delta \geq 0$. We examine the expression $|b_{kk}|^p + |b_{ll}|^p{=}|a_{kk} + \delta|^p + |a_{ll} - \delta|^p \geq |a_{kk}|^p + |a_{ll}|^p$. Each iteration increases the l_p norm of the diagonal until it reaches the Schatten norm of $\mathbf{A}$. □ The nuclear norm and the spectral norm are a special case of the more general Schatten norm. In a similar argument it is possible to prove the theorem for the Ky-Fan norm as well.

LEMMA B.4 (Extension to real matrices). *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a square matrix. Then, $\|\operatorname{diag}(\mathbf{A})\|_p \le \|\mathbf{A}\|_p$.*

Proof. From the triangle inequality $\|\mathbf{A}+\mathbf{A}^{\mathbf{T}}\| \le \|\mathbf{A}\| + \|\mathbf{A}^{\mathbf{T}}\| = 2\|\mathbf{A}\|$. Hence, $\|\mathbf{A}\| \ge \frac{1}{2}\|\mathbf{A}+\mathbf{A}^{\mathbf{T}}\|$. Since $\mathbf{A}+\mathbf{A}^{\mathbf{T}}$ is symmetric, we use Theorem B.3 that yields $\|\mathbf{A}\| \ge \frac{1}{2}\|\mathbf{A}+\mathbf{A}^{\mathbf{T}}\| \ge \frac{1}{2}\|\operatorname{diag}(\mathbf{A}+\mathbf{A}^{\mathbf{T}})\| = \|\operatorname{diag}(\mathbf{A})\|$. □

LEMMA B.5 (Extension to complex matrices with real diagonal). *Let $\mathbf{A}$ be a square matrix with real diagonal. Then, $\|\operatorname{diag}(\mathbf{A})\|_p \le \|\mathbf{A}\|_p$.*

Proof. The proof is similar to the proof of Lemma B.4. From the triangle inequality we get $\|\mathbf{A}\| \ge \frac{1}{2}\|\mathbf{A}+\operatorname{conj}(\mathbf{A})\|$. By using Lemma B.4 we get $\frac{1}{2}\|\mathbf{A}+\operatorname{conj}(\mathbf{A})\| \ge \frac{1}{2}\|\operatorname{diag}(\mathbf{A}+\operatorname{conj}(\mathbf{A})\| = \|\operatorname{diag}(\mathbf{A})\|$. Here we used the fact that $\operatorname{diag}(\mathbf{A})$ is real. □

THEOREM B.6 (Extension to complex matrices). *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a square matrix,. Then, $\|\operatorname{diag}(\mathbf{A})\|_p \le \|\mathbf{A}\|_p$.*

Proof. Let $\mathbf{U}$ be a diagonal unitary (square) matrix whose elements are $u_j = e^{-i\theta_j}$ where θ_j is the phase of a_{jj}. Because of the structure of $\mathbf{U}$, $\operatorname{diag}(\mathbf{UA})$ is real. Since $|u_j| = 1$ we get $\|\operatorname{diag}(\mathbf{A})\| = \|\operatorname{diag}(\mathbf{UA})\|$. From Lemma B.5 we get $\|\operatorname{diag}(\mathbf{A})\| = \|\operatorname{diag}(\mathbf{UA})\| \le \|\mathbf{UA}\| = \|\mathbf{A}\|$. □

Extension to rectangular matrices is straightforward: Each rectangular matrix can be zero padded to a square matrix since the singular values of a matrix are invariant to zero padding.

THEOREM B.7 (Fischer's inequality). *Assume that $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric and positive matrix. Then, $\det(A) \le \det(\operatorname{diag}(\mathbf{A}))$.*

Proof. We apply the Jacobi rotation to $\mathbf{A}$ on the entry (k,l). By assuming that $a_{kk} \ge a_{ll}$ we get that the new diagonal entries satisfy $b_{kk}b_{ll} = (a_{kk}+\delta)(a_{ll}-\delta) \le a_{kk}a_{ll}$. Hence, in each iteration the product of the diagonal entries is getting smaller and converges to the product of the eigenvalues (to the determinant), proving the Fischer's theorem. □

THEOREM B.8 (Exponential trace). *Assume that $\mathbf{A} \in \mathbb{C}^{m\times n}$ whose singular values are $\{\sigma_i\}_{i=1}^n$. Then, $\sum_{i=1}^n e^{\sigma_i} \ge \sum_{i=1}^n e^{|a_{ii}|}$.*

Proof. We know from Theorem B.5 that for every integer p, $\|\operatorname{diag}(\mathbf{A})\|_p \le \|\mathbf{A}\|_p$,

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 23, pp. 678-702, August 2012

hence:

$$\begin{aligned}
&1+1+\cdots+1 \le 1+1+\cdots+1 \quad (n \text{ times})\\
&|a_{11}|+|a_{22}|+\cdots+|a_{nn}| \le |\sigma_1|+|\sigma_2|+\cdots+|\sigma_n|\\
&\frac{1}{2}|a_{11}|^2+\frac{1}{2}|a_{22}|^2+\cdots+\frac{1}{2}|a_{nn}|^2 \le \frac{1}{2}|\sigma_1|^2+\frac{1}{2}|\sigma_2|^2+\cdots+\frac{1}{2}|\sigma_n|^2\\
&\qquad\qquad\vdots \qquad\qquad\qquad\qquad\qquad \vdots\\
&\frac{1}{p!}|a_{11}|^p+\frac{1}{p!}|a_{22}|^p+\cdots+\frac{1}{p!}|a_{nn}|^p \le \frac{1}{p!}|\sigma_1|^p+\frac{1}{p!}|\sigma_2|^p+\cdots+\frac{1}{p!}|\sigma_n|^p.
\end{aligned} \tag{B.6}$$

Each term in its identical location across all the expressions (identities) in Eq. (B.6) is summed separately. Pictorially, it sums each column in Eq. (B.6). After summing the equations we get the Taylor expansion of e^x as $p \to \infty$. This completes the proof. □

THEOREM B.9 (Logarithmic product). *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be symmetric and positive whose eigenvalues are* $\{\lambda_i\}_{i=1}^n$. *Then,* $\prod_i \log(1+\lambda_i) \le \prod_i \log(1+a_{ii})$.

Proof. We follow the same argument as in the proof of Theorem $B.8$. Assuming $a_{ii} \ge a_{jj}$, we get after one iteration that $b_{ii} = a_{ii}+\delta$ and $b_{jj} = a_{jj}-\delta$, $\delta \ge 0$. Since $\log(1+a_{ii}+\delta)\log(1+a_{jj}-\delta) \le \log(1+a_{ii})\log(1+a_{jj})$ for $a_{ii} \ge a_{jj}$ and $\delta \ge 0$, the proof is completed. □

REFERENCES

[1] H. Andrews and C. Patterson. Singular value decomposition image coding. *IEEE Transactions on Communications*, 24(4):425–432, 1976.

[2] R. Bhatia. *Matrix Analysis.* Graduate Texts in Mathematics, Springer, 1996.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[4] J.F. Cai, E.J. Candes, and Z. Shen, Singular value thresholding algorithm for matrix completion, *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[5] Y. Chen and X. Ye. *Projection onto a simplex.* Arxiv preprint arXiv:1101.6081, 2011.

[6] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical, Series B (Methodological)*, 39(1):1–38, 1977.

[7] G. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[8] R. Everson. *Orthogonal but not orthonormal Procrustes problem.* Preprint, 1997.

[9] M. Fazel. *Matrix Rank Minimization with Applications.* PhD Thesis, Stanford University, 2002.

[10] I.C. Gohberg and M.G. Krein. *Introduction to the Theory of Linear and Selfadjoint Operators.* Translations of Mathematical Monographs, Vol. 18, 94–95, 1969.

[11] D. Goldfarb and S. Ma. Convergence of fixed point continuation algorithms for matrix rank minization. *Foundations of Computational Mathematics*, 11(2):183–210, 2011.

[12] G.H. Golub, A. Hoffman, and G.W. Stewart. A generalization of the Eckart-Young-Mirsky matrix approximation theorem. *Linear Algebra and its Applications*, 88/89:317–327, 1987.

[13] T. Hastie, et al. *Imputing missing data for gene expression arrays.* Technical Report, Division of Biostatistics, Stanford University, 1999.

[14] A.N. Iusem. On the convergence properties of the projected gradient method for convex optimization. *Computational and Applied Mathematics*, 22(1):37-52, 2003.

[15] H.A.L. Kiers. Weighted least squares fitting using ordinary least squares algorithms. *Psychometrika*, 62(2):215–266, 1997.

[16] S. Lee, G. Wolberg, and S.Y. Shin. Scattered data interpolation with multilevel B-splines, *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–244, 1997.

[17] N. Levanon. *Radar Principles.* Wiley-Interscience, 1988.

[18] T.A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 44(2):226–233, 1982.

[19] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008.

[20] R. Mazumder, T. Hastie, and R. Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11:2287–2322, 2010.

[21] R. Meka, P. Jain, and I.S. Dhillon. *Guaranteed rank minimization via singular value decomposition.* Arxiv preprint arXiv:0909.5457, 2009.

[22] F. Nan. *Low Rank Matrix Completion.* Master Thesis, Massachusetts Institute of Technology, 2009.

[23] A. Papoulis. A new algorithm in spectral analysis and band-limited interpolation. *IEEE Transactions on Circuits and Systems*, 22(9):735–742, 1975.

[24] D.M. Pozar. *Microwave Engineering*, third edition. Wiley, 2004.

[25] P.H. Schonemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

[26] N. Srebro and T. Jaakkola. Weighted low-rank approximations. *Proceedings of the 20th International Conference on Machine Learning (ICML-2003)*, Washington DC, 2003.

[27] M. Tian, S.W. Luo, and L.Z. Liao. An investigation into using singular value decomposition as a method of image compression. *IEEE Proceedings of International Conference on Machine Learning and Cybernetics*, 2005.

[28] O. Troyanskaya, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17(6):520-525, 2001.

[29] P. Waldemar and T.A. Ramstad. Hybrid KLT-SVD image compression. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4:2713–2716, 1997.

[30] L.P. Yaroslavsky, et al. Nonuniform sampling, image recovery from sparse data and the discrete sampling theorem. *Journal of the Optical Society of America A*, 26(3):566–575, 2009.