The inertia of the symmetric approximation for low rank matrices

Marta Casanellas *[1], Jesús Fernández-Sánchez †[1], and Marina Garrote-López ‡[2]

[1]Dept. Matemàtiques, Universitat Politècnica de Catalunya.
Av. Diagonal, 647. 08028 Barcelona, Spain

July 28, 2017

Abstract

In many areas of applied linear algebra, it is necessary to work with matrix approximations. A usual situation occurs when a matrix obtained from experimental or simulated data is needed to be approximated by a matrix that lies in a corresponding statistical model and satisfies some specific properties. In this short note, we focus on symmetric and positive-semidefinite approximations and we show that the positive and negative indices of inertia of the symmetric approximation and the rank of the positive-semidefinite approximation are always bounded from above by the rank of the original matrix.

keywords: symmetric matrices, positive definiteness, rank approximation, inertia indices.
MSC 15A18, 15B57

1 Introduction

A common problem in applied linear algebra is finding the nearest matrix X to a given matrix A subject to some specific properties of X. For example, in many applications of different areas such as machine learning and statistics, the matrix A is obtained from empirical or simulated data while X should lie in some statistical model. Among the usual properties required for X we find: having a certain rank or being orthogonal, symmetric, positive-definite... We refer the reader to the paper of Higham [5] for a nice introduction to this kind of applied problems. The quoted paper also contains a survey on theoretical results and computational methods usually applied to nearness problems for fundamental matrix properties like symmetry, positive-definiteness, orthogonality or normality.

In this short note, we focus on the study of the rank and inertia of the symmetric and the symmetric positive-semidefinite (*PSD* for short) approximations (in the Frobenius norm) of a low rank matrix. Namely, our main result (Theorem 3.1) states that the inertia indices of the symmetric approximation of a matrix are upper bounded by its rank.

The PSD approximation is easy to compute in the Frobenius norm (see [4]) and plays an important role in detecting and modifying an indefinite Hessian matrix in Newton methods for optimisation [5]. Some other relevant applications are discussed by Duff, Erisman and Reid in [2], where the authors mention the importance of finding the closest positive-semidefinite matrix on sparsity optimisation as well as in building a large circuit analysis model from measured data. In machine learning, some successful applications that depend on symmetric positive-semidefinite matrices (covariance, correlation or kernel matrices) are multi-camera tracking based on covariance matrices derived from appearance silhouettes, medical diagnostics via diffusion tensor imaging, computational anatomy, robust face recognition and action recognition (see [8]). The particular application that originally motivated this paper is the use of algebraic tools in phylogenetics. More precisely, our study is inspired by a theoretical result that states that a certain matrix has to be PSD of low rank in order to correspond to a distribution arising from a hidden Markov process on a certain phylogenetic tree (see Proposition 4.5 in [1]). If

*marta.casanellas@upc.edu
†jesus.fernandez.sanchez@upc.edu. Corresponding author.
‡marinagarrotelopez@gmail.com

one wants to apply this result to real data, it is crucial to know whether the rank is preserved under the PSD approximation. This is guaranteed by our Corollary 3.3.

The outline of this article is as follows. We start with a Preliminaries section where the basic tools needed to state the main theorem are introduced. In Section 3 we present the main result (Theorem 3.1) and, as a byproduct, we are able to prove in Corollary 3.3 that the nearest symmetric positive-semidefinite matrix to a rank k matrix has also rank less than or equal to k.

2 Preliminaries

We present some known results that will be used throughout the rest of the paper and we refer to the book [7] for the linear algebra background needed for this paper.

Given a real square matrix $A \in \mathcal{M}_n(\mathbb{R})$, the spectrum of A is the set of all its eigenvalues. The *inertia* (or *signature*) of a real symmetric matrix S is the number, counted with multiplicity, of positive, negative and zero eigenvalues. It can be denoted by the triplet (i_+, i_-, i_0) where i_+ and i_- are also known as the *positive* and *negative inertia indices* of S, respectively. By the Spectral Theorem (see page 517 in [7]) the eigenvalues of any real symmetric matrix S are all real and S diagonalizes through an orthonormal basis of eigenvectors.

The problem of finding the closest symmetric matrix S to a real matrix $A \in \mathcal{M}_n(\mathbb{R})$ was solved by Fan and Hoffman in 1955 (see [3]). It is well known that S equals $\dfrac{A + A^T}{2}$ and is unique in the Frobenius norm.

Moreover, Higham proved in [4] that if $S = UH$ is a polar decomposition of S (that is, U is orthogonal and H is symmetric and positive semidefinite), then the PSD approximation X to A in the Frobenius norm is $X = \dfrac{S + H}{2}$.

3 Results

From now on, unless noted otherwise, $A \in \mathcal{M}_n(\mathbb{R})$ is a real matrix of $\operatorname{rank}(A) = k$ and $S = \dfrac{A + A^T}{2}$ stands for its nearest symmetric matrix.

Our main result states that the positive and negative indices of inertia of S are upper bounded by $\operatorname{rank}(A)$.

Theorem 3.1 *Let $A \in \mathcal{M}_n(\mathbb{R})$ be a matrix of $\operatorname{rank}(A) = k$ and S the symmetric approximation of A. Then $i_+(S) \leq k$, $i_-(S) \leq k$ and $i_0(S) \geq \max\{0, n - 2k\}$.*

The following easy lemma is crucial to prove Theorem 3.1.

Lemma 3.2 *For any $v \in \mathbb{R}^n$ we have $v^T A v = v^T S v$.*

Proof. For any $v \in \mathbb{R}^n$, $v^T A v$ is a scalar, so $v^T A v = (v^T A v)^T$. It follows that

$$v^T S v = v^T \left(\frac{A + A^T}{2}\right) v = \frac{1}{2}\left(v^T A v + v^T A^T v\right) = \frac{1}{2}\left(v^T A v + (v^T A v)^T\right) =$$
$$= \frac{1}{2}\left(v^T A v + v^T A v\right) = v^T A v,$$

which proves the result. □

We are now ready to prove Theorem 3.1.

Proof. We proceed by contradiction. Let (i_+, i_-, i_0) denote the inertia of S and suppose $i_+ = N$ with $N > k$.

Let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors of S ordered so that $v_1, \ldots, v_N$ correspond to the positive eigenvalues. That is, Sv_i equals $\lambda_i v_i$ with $\lambda_i > 0$ for all $i \in \{1, \ldots, N\}$, and $\lambda_i \leq 0$ if $i > N$. Let V be the subspace spanned by these eigenvectors $V := \langle v_1, \ldots, v_N \rangle$. Then for all nonzero $v \in V$ it is satisfied that $v^T S v > 0$.

Grassman's formula states that $\dim(E + F) = \dim(E) + \dim(F) - \dim(E \cap F)$ for all E, F vector subspaces of the same vector space (see 4.4.19 in [7]). Note that $\dim(V)$ is N and $\dim(\ker(A))$ is $n - k$ since $\operatorname{rank}(A)$ equals k. Thus we infer

$$\dim(V + \ker(A)) + \dim(V \cap \ker(A)) = \dim(V) + \dim(\ker(A)) = N + n - k. \qquad (1)$$

Since $\dim(V + \ker(A)) \leq n$ expression 1 gives us that $\dim(V \cap \ker(A)) \geq 1$. Consequently, there exists at least one element $w \neq \vec{0}$ such that $w \in V \cap \ker(A)$. By Lemma 3.2 $w^T A w$ equals $w^T S w$, which is positive since

w belongs to V. But at the same time w belongs to $\ker(A)$ and hence $w^T A w = w^T \vec{0}$ which is equal to zero. This leads to a contradiction and we conclude $i_+ \leq k$.

An analogous argument can be used to prove that $i_- \leq k$. The fact that $i_0 \geq \max\{0, n-2k\}$ follows trivially since $i_+ + i_- + i_0 = n$. □

Using this result, it is easy to bound the rank of the PSD approximation of a matrix $A \in \mathcal{M}_n(\mathbb{R})$:

Corollary 3.3 *The rank of the PSD approximation X of a real matrix A is less than or equal to* $\operatorname{rank}(A)$.

PROOF. There exists an orthonormal basis $\{v_1, \ldots, v_n\}$ of eigenvectors of S, the symmetric approximation of A, with respective eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ such that $S = P\Lambda P^T$ where $\Lambda = \operatorname{diag}(\lambda_i)$ and P is the orthogonal matrix with these eigenvectors as columns. In particular, $P^{-1} = P^T$. We define $\bar{\Lambda}$ as the diagonal matrix with entries $\bar{\lambda}_i = \max\{0, \lambda_i\}$. Then, the PSD approximation of A is

$$X = P\bar{\Lambda}P^T. \tag{2}$$

We claim that $P\bar{\Lambda}P^T$ is equal to the PSD approximation of A. In order to show it, first consider $\sigma_i = \frac{\lambda_i}{|\lambda_i|}$ if $\lambda_i \neq 0$ or $\sigma_i = 0$ if $\lambda_i = 0$, and define $\Sigma = \operatorname{diag}(\sigma_i)$ and $|\Lambda| = \operatorname{diag}(|\lambda_i|)$. Then it is easy to see that $S = P\Sigma|\Lambda|P^T$ and since P is orthogonal, $S = UH$ is a polar decomposition of S where $U = P\Sigma P^T$ and $H = P|\Lambda|P^T$.

Therefore, the representation of $X = P\bar{\Lambda}P^T$ follows because $\dfrac{(\Sigma + Id)|\Lambda|}{2}$ is equal to $\bar{\Lambda}$ and

$$\begin{aligned} X &= \frac{S+H}{2} = \frac{(P\Sigma P^T)(P|\Lambda|P^T) + P|\Lambda|P^T}{2} = \frac{(P\Sigma P^T + Id)P|\Lambda|P^T}{2} \\ &= \frac{P(\Sigma + Id)|\Lambda|P^T}{2} = P\bar{\Lambda}P^T. \end{aligned}$$

From this new expression of X in 2 we see that $\operatorname{rank}(X) = \#\{\bar{\lambda}_i | \bar{\lambda}_i \neq 0\}$ ($\bar{\lambda}_i$ counted with multiplicity) which coincides with the positive inertia index of S. Finally since $i_+(S) \leq \operatorname{rank}(A)$ by Theorem 3.1, we obtain $\operatorname{rank}(X) \leq \operatorname{rank}(A)$. □

From now on we focus on the extremal cases of $\operatorname{rank}(X)$. First, we state a known result.

Theorem 3.4 *(Marsaglia [6]) Let $A, B \in \mathcal{M}_n(\mathbb{R})$. Let C_A, C_B be the subspaces generated by their columns and R_A, R_B subspaces generated by their rows. Let $c = \dim(C_A \cap C_B)$ and $r = \dim(R_A \cap R_B)$, then*

$$\operatorname{rank}(A) + \operatorname{rank}(B) - c - r \leq \operatorname{rank}(A+B) \leq \operatorname{rank}(A) + \operatorname{rank}(B) - max\{c, r\}. \tag{3}$$

In particular, $\operatorname{rank}(A+B) = \operatorname{rank}(A) + \operatorname{rank}(B)$ *if and only if* $\dim(C_A \cap C_B) = \dim(R_A \cap R_{A^T}) = \dim(R_A \cap R_B) = 0$.

Back to our case of interest, we take $B = A^T$ and analyse the cases when $dim(C_A \cap C_{A^T}) = dim(C_A \cap R_A)$ attains its minimal and maximal possible values. The following corollary characterises the former case.

Corollary 3.5 *Both the positive and negative indices of inertia of the symmetric approximation S are equal to* $\operatorname{rank}(A)$ *if and only if* $\dim(C_A \cap R_A) = 0$.

PROOF. Using Theorem 3.4 and the fact that $S = \dfrac{A + A^T}{2}$ we have that $\operatorname{rank}(S)$ is $2k$ if and only if $\dim(C_A \cap C_{A^T}) = \dim(R_A \cap R_{A^T}) = \dim(R_A \cap C_A)$ is equal to zero. Moreover, since $\operatorname{rank}(S) = i_+ + i_-$ and $i_+, i_- \leq k$ (by Theorem 3.1), $\operatorname{rank}(S)$ is equal to $2k$ if and only if $i_+ = i_- = k$. In summary, both $i_+(S)$ and $i_-(S)$ are equal to k if and only if $\operatorname{rank}(S)$ is $2k$, which holds if and only if $C_A \cap R_A$ is equal to zero. □

Note that the maximal value for $\dim(C_A \cap R_A)$ is $\operatorname{rank}(A) = k$. In this case, applying Theorem 3.4 to the case $B = A^T$, we derive that $0 \leq \operatorname{rank}(S) \leq k$. Obviously $\operatorname{rank}(S) = 0$ if and only if $S = 0$ (and hence A is antisymmetric and $\operatorname{rank}(X) = 0$). The upper bound is achieved if and only if $i_0(S) = n-k$ and $i_+(S) + i_-(S) = k$. Therefore $R_A = C_A = R_S$, since $\operatorname{rank}(A)$ is equal to $\operatorname{rank}(S)$. This happens when A is symmetric; in this case we have $S = A$ and X has rank equal to the number of positive eigenvalues of A (counted with multiplicity).

Acknowledgement

M.Casanellas and J.Fernández-Sánchez were partially supported by Spanish government MTM2012-38122-C03-01/FEDER and Generalitat de Catalunya 2014 SGR-634.

References

[1] E. S. Allman, J. A. Rhodes, and A. Taylor. A semialgebraic description of the general Markov model on phylogenetic trees. *SIAM J. Discrete Math.*, (28):736–755, 2014.

[2] A. M. Erisman, I. S. Duff, and J. K. Reid. *Direct Methods for Sparse Matrices*. Oxford University Press, Inc., New York, NY, USA, 1986.

[3] K. Fan and A. J. Hoffman. Some metric inequalities in the space of matrices. *Proceedings of the American Mathematical Society*, 6(1):111–116, 1955.

[4] N. J. Higham. Computing a Nearest Symmetric Positive Semidefinite Matrix. *Linear Algebra and its Applications*, 103:103–118, 1988.

[5] N. J. Higham. Matrix Nearness Problems and Applications. *Applications of Matrix Theory*, 22, 1989.

[6] G. Marsaglia. Bounds on the rank of the sum of matrices. In *Trans. Fourth Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes (Prague, 1965)*, pages 455–462. Academia, Prague, 1967.

[7] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[8] S. Sra and A. Cherian. Generalized dictionary learning for symmetric positive definite matrices with application to nearest neighbor retrieval. In *Machine Learning and Knowledge Discovery in Databases*, volume 6913, pages 318–332. Springer, 2011.