University of Pennsylvania
ScholarlyCommons

Technical Reports (CIS) | Department of Computer & Information Science

5-2-2008

Logarithms and Square Roots of Real Matrices

Jean H. Gallier
University of Pennsylvania, jean@cis.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/cis_reports

Part of the Electrical and Computer Engineering Commons

Recommended Citation
Gallier, Jean H., "Logarithms and Square Roots of Real Matrices" (2008). *Technical Reports (CIS)*. Paper 876.
http://repository.upenn.edu/cis_reports/876

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-08-12.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/cis_reports/876
For more information, please contact repository@pobox.upenn.edu.

Logarithms and Square Roots of Real Matrices

Abstract
The need for computing logarithms or square roots of real matrices arises in a number of applied problems. A significant class of problems comes from medical imaging. One of these problems is to interpolate and to perform statistics on data represented by certain kinds of matrices (such as symmetric positive definite matrices in DTI). Another important and difficult problem is the registration of medical images. For both of these problems, the ability to compute logarithms of real matrices turns out to be crucial. However, not all real matrices have a real logarithm and thus, it is important to have sufficient conditions for the existence (and possibly the uniqueness) of a real logarithm for a real matrix. Such conditions (involving the eigenvalues of a matrix) are known, both for the logarithm and the
square root.
As far as I know, with the exception of Higham's recent book [18], proofs of the results involving these conditions are scattered in the literature and it is not easy to locate them. Moreover, Higham's excellent book assumes a certain level of background in linear algebra that readers interested in applications to medical imaging may not possess so we feel that
a more elementary presentation might be a valuable supplement to Higham [18]. In this paper, I present a unified exposition of these results, including a proof of the existence of the Real Jordan Form, and give more direct proofs of some of these results using the Real
Jordan Form.

Disciplines
Electrical and Computer Engineering

Comments
University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-08-12.

This technical report is available at ScholarlyCommons: http://repository.upenn.edu/cis_reports/876

Logarithms and Square Roots of Real Matrices Existence, Uniqueness, and Applications in Medical Imaging

Jean Gallier

Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
jean@cis.upenn.edu

November 9, 2013

Abstract. The need for computing logarithms or square roots of real matrices arises in a number of applied problems. A significant class of problems comes from medical imaging. One of these problems is to interpolate and to perform statistics on data represented by certain kinds of matrices (such as symmetric positive definite matrices in DTI). Another important and difficult problem is the registration of medical images. For both of these problems, the ability to compute logarithms of real matrices turns out to be crucial. However, not all real matrices have a *real* logarithm and thus, it is important to have sufficient conditions for the existence (and possibly the uniqueness) of a real logarithm for a real matrix. Such conditions (involving the eigenvalues of a matrix) are known, both for the logarithm and the square root.

As far as I know, with the exception of Higham's recent book [18], proofs of the results involving these conditions are scattered in the literature and it is not easy to locate them. Moreover, Higham's excellent book assumes a certain level of background in linear algebra that readers interested in applications to medical imaging may not possess so we feel that a more elementary presentation might be a valuable supplement to Higham [18]. In this paper, I present a unified exposition of these results, including a proof of the existence of the Real Jordan Form, and give more direct proofs of some of these results using the Real Jordan Form.

1 Introduction and Motivations

Theorems about the conditions for the existence (and uniqueness) of a real logarithm (or a real square root) of a real matrix are the theoretical basis for various numerical methods for exponentiating a matrix or for computing its logarithm using a method known as *scaling and squaring* (resp. *inverse scaling and squaring*). Such methods play an important role in the *log-Euclidean framework* due to Arsigny, Fillard, Pennec and Ayache and its applications to medical imaging [1, 3, 4, 5].

The registration of medical images is an important and difficult problem. The work described in Arsigny, Commowick, Pennec and Ayache [2] (and Arsigny's thesis [1]) makes an orginal and valuable contribution to this problem by describing a method for parametrizing a class of non-rigid deformations with a small number of degrees of freedom. After a global affine alignment, this sort of parametrization allows a finer local registration with very smooth transformations. This type of parametrization is particularly well adpated to the registration of histological slices, see Arsigny, Pennec and Ayache [5].

The goal is to fuse some affine or rigid transformations in such a way that the resulting transformation is invertible and smooth. The direct approach which consists in blending N global affine or rigid transformations, $T_1, \ldots, T_N$ using weights, $w_1, \ldots, w_N$, does not work because the resulting transformation,

$$T = \sum_{i=1}^{N} w_i T_i,$$

is not necessarily invertible. The purpose of the weights is to define the domain of influence in space of each T_i.

The novel key idea is to associate to each rigid (or affine) transformation, T, of $\mathbb{R}^n$, a vector field, V, and to view T as the diffeomorphism, Φ_1^V, corresponding to the time $t = 1$, where Φ_t^V is the global flow associated with V. In other words, T is the result of integrating an ODE

$$X' = V(X, t),$$

starting with some initial condition, X_0, and $T = X(1)$.

Now, it would be highly desirable if the vector field, V, did not depend on the time parameter, and this is indeed possible for a large class of affine transformations, which is one of the nice contributions of the work of Arsigny, Commowick, Pennec and Ayache [2].

Recall that an affine transformation, $X \mapsto LX + v$, (where L is an $n \times n$ matrix and $X, v \in \mathbb{R}^n$) can be conveniently represented as a linear transformation from $\mathbb{R}^{n+1}$ to itself if we write

$$\begin{pmatrix} X \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} L & v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ 1 \end{pmatrix}.$$

Then, the ODE with constant coefficients

$$X' = LX + v,$$

can be written

$$\begin{pmatrix} X' \\ 0 \end{pmatrix} = \begin{pmatrix} L & v \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X \\ 1 \end{pmatrix}$$

and, for every initial condition, $X = X_0$, its unique solution is given by

$$\begin{pmatrix} X(t) \\ 1 \end{pmatrix} = \exp\left(t \begin{pmatrix} L & v \\ 0 & 0 \end{pmatrix}\right) \begin{pmatrix} X_0 \\ 1 \end{pmatrix}.$$

Therefore, if we can find reasonable conditions on matrices, $T = \begin{pmatrix} M & t \\ 0 & 1 \end{pmatrix}$, to ensure that they have a unique real logarithm,

$$\log(T) = \begin{pmatrix} L & v \\ 0 & 0 \end{pmatrix},$$

then we will be able to associate a vector field, $V(X) = LX + v$, to T, in such a way that T is recovered by integrating the ODE, $X' = LX + v$. Furthermore, given N transformations, $T_1, \ldots, T_N$, such that $\log(T_1), \ldots, \log(T_N)$ are uniquely defined, we can fuse $T_1, \ldots, T_N$ at the *infinitesimal level* by defining the ODE obtained by blending the vector fields, $V_1, \ldots, V_N$, associated with $T_1, \ldots, T_N$ (with $V_i(X) = L_i X + v_i$), namely

$$V(X) = \sum_{i=1}^{N} w_i(X)(L_i X + v_i).$$

Then, it is easy to see that the ODE,

$$X' = V(X),$$

has a unique solution for every $X = X_0$ defined for all t, and the fused transformation is just $T = X(1)$. Thus, the fused vector field,

$$V(X) = \sum_{i=1}^{N} w_i(X)(L_i X + v_i),$$

yields a one-parameter group of diffeomorphisms, Φ_t. Each transformation, Φ_t, is smooth and invertible and is called a *Log-Euclidean polyaffine tranformation*, for short, *LEPT*. Of course, we have the equation

$$\Phi_{s+t} = \Phi_s \circ \Phi_t,$$

for all $s, t \in \mathbb{R}$ so, in particular, the inverse of Φ_t is Φ_{-t}. We can also interpret Φ_s as $(\Phi_1)^s$, which will yield a fast method for computing Φ_s. Observe that when the weight are scalars, the one-parameter group is given by

$$\begin{pmatrix} \Phi_t(X) \\ 1 \end{pmatrix} = \exp\left(t \sum_{i=1}^{N} w_i \begin{pmatrix} L_i & v_i \\ 0 & 0 \end{pmatrix} \right) \begin{pmatrix} X \\ 1 \end{pmatrix},$$

which is the Log-Euclidean mean of the affine transformations, T_i's (w.r.t. the weights w_i).

Fortunately, there is a sufficient condition for a real matrix to have a unique real logarithm and this condition is not too restrictive in practice.

Let $\mathcal{S}(n)$ denotes the set of all real matrices whose eigenvalues, $\lambda + i\mu$, lie in the horizontal strip determined by the condition $-\pi < \mu < \pi$. We have the following weaker version of Theorem 3.10:

Theorem 1.1. *The image,* $\exp(\mathcal{S}(n))$*, of* $\mathcal{S}(n)$ *by the exponential map is the set of real invertible matrices with no negative eigenvalues and* $\exp\colon \mathcal{S}(n) \to \exp(\mathcal{S}(n))$ *is a bijection.*

Theorem 1.1 is stated in Kenney and Laub [23] without proof. Instead, Kenney and Laub cite DePrima and Johnson [13] for a proof but this latter paper deals with complex matrices and does not contain a proof of our result either.

It is also known that under the same condition (no negative eigenvalues) every real $n \times n$ matrix, A, has a real square root, that is, there is a real matrix, X, such that $X^2 = A$. Moreover, if the eigenvalues, $\rho\, e^{i\theta}$, of X satisfy the condition $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$, then X is unique (see Theorem 4.8).

Actually, there is a necessary and sufficient condition for a real matrix to have a real logarithm (or a real square root) but it is fairly subtle as it involves the parity of the number of Jordan blocks associated with negative eigenvalues, see Theorem 3.4. The first occurrence of this theorem that we have found in the literature is a paper by Culver [12] published in 1966. We offer a proof using Theorem 2.7, which is more explicit than Culver's proof.

Curiously, complete and direct proofs of the main Theorems, 3.4, 3.10, and 4.8 do not seem to exist and references found in various papers are sometimes incorrect (for more on this, see the beginning of Section 3, the remark after the proof of Theorem 4.4 and the remark after the proof of Theorem 4.8). Versions of these results do appear in Higham's book [18] but one of the theorems involved (Theorem 1.28) is not proved and closer examination reveals that Theorem 1.36 (in Higham's book) is needed to prove Theorem 1.28.

In view of all this, we feel that providing a unifying treatment and giving complete proofs of these results will be of value to the mathematical community.

2 Jordan Decomposition and the Real Jordan Form

The proofs of the results stated in Section 1 make heavy use of the Jordan normal form of a matrix and its cousin, the Jordan decomposition of a matrix into its semisimple part and its nilpotent part. The purpose of this section is to review these concepts rather thoroughly to make sure that the reader has the background necessary to understand the proofs in Section 3 and Section 4. We pay particular attention to the *Real Jordan Form* (Horn and Johnson [20], Chapter 3, Section 4, Theorem 3.4.5, Hirsh and Smale [19] Chapter 6) which, although familiar to experts in linear algebra, is typically missing from "standard" algebra books. We give a complete proof of the Real Jordan Form as such a proof does not seem to be easily found (even Horn and Johnson [20] only give a sketch of the proof, but it is covered in Hirsh and Smale [19], Chapter 6).

Let V be a finite dimensional real vector space. Recall that we can form the *complexification*, $V_{\mathbb{C}}$, of V. The space $V_{\mathbb{C}}$ is the complex vector space, $V \times V$, with the addition operation given by

$$(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2),$$

and the scalar multiplication given by

$$(\lambda + i\mu) \cdot (u, v) = (\lambda u - \mu v, \mu u + \lambda v) \qquad (\lambda, \mu \in \mathbb{R}).$$

Obviously

$$(0, v) = i \cdot (v, 0),$$

so every vector, $(u, v) \in V_{\mathbb{C}}$, can written uniquely as

$$(u, v) = (u, 0) + i \cdot (v, 0).$$

The map from V to $V_{\mathbb{C}}$ given by $u \mapsto (u, 0)$ is obviously an injection and for notational convenience, we write $(u, 0)$ as u, we suppress the symbol ("dot") for scalar multiplication and we write

$$(u, v) = u + iv, \qquad \text{with } u, v \in V.$$

Observe that if $(e_1, \ldots, e_n)$ is a basis of V, then it is also a basis of $V_{\mathbb{C}}$.

Every linear map, $f\colon V \to V$, yields a linear map, $f_{\mathbb{C}}\colon V_{\mathbb{C}} \to V_{\mathbb{C}}$, with

$$f_{\mathbb{C}}(u + iv) = f(u) + if(v), \qquad \text{for all } u, v \in V.$$

Definition 2.1. A linear map, $f\colon V \to V$, is *semisimple* iff $f_{\mathbb{C}}$ can be diagonalized. In terms of matrices, a real matrix, A, is *semisimple* iff there are some matrices D and P with entries in $\mathbb{C}$, with P invertible and D a diagonal matrix, so that $A = PDP^{-1}$. We say that f is *nilpotent* iff $f^r = 0$ for some positive integer, r, and a matrix, A, is *nilpotent* iff $A^r = 0$ for some positive integer, r. We say that f is *unipotent* iff $f - \mathrm{id}$ is nilpotent and a matrix A is *unipotent* iff $A - I$ is nilpotent.

If A is unipotent, then $A = I + N$ where N is nilpotent. If r is the smallest integer so that $N^r = 0$ (the *index of nilpotency* of N), then it is easy to check that

$$I - N + N^2 + \cdots + (-1)^{r-1} N^{r-1}$$

is the inverse of $A = I + N$.

For example, rotation matrices are semisimple, although in general they can't be diagonalized over $\mathbb{R}$, since their eigenvalues are complex numbers of the form $e^{i\theta}$. Every upper-triangular matrix where all the diagonal entries are zero is nilpotent.

Definition 2.2. If $f\colon V \to V$ is a linear map with V a finite vector space over $\mathbb{R}$ or $\mathbb{C}$, a *Jordan decomposition* of f is a pair of linear maps, $f_S, f_N\colon V \to V$, with f_S semisimple and f_N nilpotent, such that

$$f = f_S + f_N \qquad \text{and} \qquad f_S \circ f_N = f_N \circ f_S.$$

The theorem below is a very useful technical tool for dealing with the exponential map. It can be proved from the so-called primary decomposition theorem or from the Jordan form (see Hoffman and Kunze [22], Chapter 6, Section 4 or Bourbaki [8], Chapter VII, §5).

Theorem 2.1. *If V is a finite dimensional vector space over $\mathbb{C}$, then every linear map, $f\colon V \to V$, has a unique Jordan decomposition, $f = f_S + f_N$. Furthermore, f_S and f_N can be expressed as polynomials in f with no constant term.*

Remark: In fact, Theorem 2.1 holds for any finite dimensional vector space over a *perfect field*, K (this means that either K has characteristic zero of that $K^p = K$, where $K^p = \{a^p \mid a \in K\}$ and where $p \geq 2$ is the characteristic of the field K). The proof of this stronger version of Theorem 2.1 is more subtle and involves some elementary Galois theory (see Hoffman and Kunze [22], Chapter 7, Section 4 or, for maximum generality, Bourbaki [8], Chapter VII, §5).

We will need Theorem 2.1 in the case where V is a real vector space. In fact we need a slightly refined version of Theorem 2.1 for $K = \mathbb{R}$ known as the *Real Jordan form*. First, let us review Jordan matrices and real Jordan matrices.

Definition 2.3. A (complex) *Jordan block* is an $r \times r$ matrix, $J_r(\lambda)$, of the form

$$J_r(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix},$$

where $\lambda \in \mathbb{C}$, with $J_1(\lambda) = (\lambda)$ if $r = 1$. A *real Jordan block* is either

(1) a Jordan block as above with $\lambda \in \mathbb{R}$, or

(2) a real $2r \times 2r$ matrix, $J_{2r}(\lambda, \mu)$, of the form

$$J_{2r}(\lambda,\mu) = \begin{pmatrix} L(\lambda,\mu) & I & 0 & \cdots & 0 \\ 0 & L(\lambda,\mu) & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & I \\ 0 & 0 & 0 & \cdots & L(\lambda,\mu) \end{pmatrix},$$

where $L(\lambda, \mu)$ is a 2×2 matrix of the form

$$L(\lambda,\mu) = \begin{pmatrix} \lambda & -\mu \\ \mu & \lambda \end{pmatrix},$$

with $\lambda, \mu \in \mathbb{R}$, $\mu \neq 0$, with I the 2×2 identity matrix and with $J_2(\lambda, \mu) = L(\lambda, \mu)$ when $r = 1$.

A (complex) *Jordan matrix*, J, is an $n \times n$ block diagonal matrix of the form

$$J = \begin{pmatrix} J_{r_1}(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(\lambda_m) \end{pmatrix},$$

where each $J_{r_k}(\lambda_k)$ is a (complex) Jordan block associated with some $\lambda_k \in \mathbb{C}$ and with $r_1 + \cdots + r_m = n$. A *real Jordan matrix*, J, is an $n \times n$ block diagonal matrix of the form

$$J = \begin{pmatrix} J_{s_1}(\alpha_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{s_m}(\alpha_m) \end{pmatrix},$$

where each $J_{s_k}(\alpha_k)$ is a real Jordan block either associated with some $\alpha_k = \lambda_k \in \mathbb{R}$ as in (1) or associated with some $\alpha_k = (\lambda_k, \mu_k) \in \mathbb{R}^2$, with $\mu_k \neq 0$, as in (2), in which case $s_k = 2r_k$.

To simplify notation, we often write $J(\lambda)$ for $J_r(\lambda)$ (or $J(\alpha)$ for $J_s(\alpha)$). Here is an example of a Jordan matrix with four blocks:

$$J = \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mu & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix}.$$

In order to prove properties of the exponential of Jordan blocks, we need to understand the deeper reasons for the existence of the Jordan form. For this, we review the notion of a minimal polynomial.

Recall that a polynomial, $p(X)$, of degree $n \geq 1$ is a *monic polynomial* iff the monomial of highest degree in $p(X)$ is of the form X^n (that is, the coefficient of X^n is equal to 1). As usual, let $\mathbb{C}[X]$ be the ring of polynomials

$$p(X) = a_0 X^n + a_1 X^{n-1} + \cdots + a_{n-1} X + a_n,$$

with complex coefficient, $a_i \in \mathbb{R}$, and let $\mathbb{R}[X]$ be the ring of polynomials with real coefficients, $a_i \in \mathbb{R}$. If V is a finite dimensional complex vector space and $f\colon V \to V$ is a given linear map, every polynomial

$$p(X) = a_0 X^n + a_1 X^{n-1} + \cdots + a_{n-1} X + a_n,$$

yields the linear map denoted $p(f)$, where

$$p(f)(v) = a_0 f^n(v) + a_1 f^{n-1}(v) + \cdots + a_{n-1} f(v) + a_n v, \qquad \text{for every } v \in V,$$

and where $f^k = f \circ \cdots \circ f$ is the composition of f with itself k times. We also write

$$p(f) = a_0 f^n + a_1 f^{n-1} + \cdots + a_{n-1} f + a_n \mathrm{id}.$$

Do not confuse $p(X)$ and $p(f)$. The expression $p(X)$ denotes a polynomial in the "indeterminate" X, whereas $p(f)$ denotes a linear map from V to V.

For example, if $p(X)$ is the polynomial

$$p(X) = X^3 - 2X^2 + 3X - 1,$$

if A is any $n \times n$ matrix, then $p(A)$ is the $n \times n$ matrix

$$p(A) = A^3 - 2A^2 + 3A - I$$

obtained by formally substituting the matrix A for the variable X.

Thus, we can define a "scalar multiplication", $\cdot\colon \mathbb{C}[X] \times V \to V$, by

$$p(X) \cdot v = p(f)(v), \qquad v \in V.$$

We immediately check that

$$\begin{aligned}
p(X) \cdot (u + v) &= p(X) \cdot u + p(X) \cdot v \\
(p(X) + q(X)) \cdot u &= p(X) \cdot u + q(X) \cdot u \\
(p(X) q(X)) \cdot u &= p(X) \cdot (q(X) \cdot u) \\
1 \cdot u &= u,
\end{aligned}$$

for all $u, v \in V$ and all $p(X), q(X) \in \mathbb{C}[X]$, where 1 denotes the polynomial of degree 0 with constant term 1.

It follows that the scalar multiplication, $\cdot\colon \mathbb{C}[X] \times V \to V$, makes V into a $\mathbb{C}[X]$-module that we will denote by V_f. Furthermore, as $\mathbb{C}$ is a subring of $\mathbb{C}[X]$ and as V is finite-dimensional, V is finitely generated over $\mathbb{C}$ and so V_f is finitely generated as a module over $\mathbb{C}[X]$.

Now, because V is finite dimensional, we claim that there is some polynomial, $q(X)$, that *annihilates* V_f, that is, so that

$$q(f)(v) = 0, \qquad \text{for all} \quad v \in V.$$

To prove this fact, observe that if V has dimension n, then the set of linear maps from V to V has dimension n^2. Therefore any $n^2 + 1$ linear maps must be linearly dependent, so

$$\mathrm{id}, f, f^2, \ldots, f^{n^2}$$

are linearly dependent linear maps and there is a nonzero polynomial, $q(X)$, of degree at most n^2 so that $q(f)(v) = 0$ for all $v \in V$. (In fact, by the *Cayley-Hamilton Theorem*, the characteristic polynomial, $q_f(X) = \det(X\,\mathrm{id} - f)$, of f annihilates V_f, so there is some annihilating polynomial of degree at most n.) By abuse of language (and notation), if $q(X)$ annihilates V_f, we also say that $q(X)$ annihilates V.

The set of annihilating polynomials of V forms a principlal ideal in $\mathbb{C}[X]$, which means that there is a unique monic polynomial of minimal degree, p_f, annihilating V and every other polynomial annihilating V is a multiple of p_f. We call this minimal monic polynomial annihilating V the *minimal polynomial* of f.

The fact that V is annihilated by some polynomial in $\mathbb{C}[X]$ makes V_f a *torsion* $\mathbb{C}[X]$-module. Furthermore, the ring $\mathbb{C}[X]$ has the property that every ideal is a *principal ideal domain*, abbreviated PID (this means that every ideal is generated by a single polynomial which can be chosen to be monic and of smallest degree). The ring $\mathbb{R}[X]$ is also a PID. In fact, the ring $k[X]$ is a PID for any field, k. But then, we can apply some powerful results about the structure of finitely generated torsion modules over a PID to V_f and obtain various decompositions of V into subspaces which yield useful normal forms for f, in particular, the Jordan form.

Let us give one more definition before stating our next important theorem: Say that V is a *cyclic module* iff V is generated by a single element as a $\mathbb{C}[X]$-module, which means that there is some $u \in V$ so that $u, f(u), f^2(u), \ldots, f^k(u), \ldots$, generate V.

Theorem 2.2. *let V be a finite-dimensional complex vector space of dimension n. For every linear map, $f\colon V \to V$, there is a direct sum decomposition,*

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_m,$$

where each V_i is a cyclic $\mathbb{C}[X]$-module such that the minimal polynomial of the restriction of f to V_i is of the form $(X-\lambda_i)^{r_i}$. Furthermore, the number, m, of subspaces V_i and the minimal polynomials of the V_i are uniquely determined by f and, for each such polynomial, $(X-\lambda)^r$, the number, m_i, of V_i's that have $(X-\lambda)^r$ as minimal polynomial (that is, if $\lambda=\lambda_i$ and $r=r_i$) is uniquely determined by f.

A proof of Theorem 2.2 can be found in M. Artin [6], Chapter 12, Section 7, Lang [24], Chapter XIV, Section 2, Dummit and Foote [14], Chapter 12, Section 1 and Section 3, or D. Serre [27], Chapter 6, Section 3. A very good exposition is also given in Gantmacher [15], Chapter VII, in particular, see Theorem 8 and Theorem 12. However, in Gantmacher, elementary divisors are defined in a rather cumbersome manner in terms of ratios of determinants of certain minors. This makes, at times, the proof unnecessarily hard to follow.

The minimal polynomials, $(X-\lambda_i)^{r_i}$, associated with the V_i's are called the *elementary divisors* of f. They need not be distinct. To be more precise, if the set of distinct elementary divisors of f is

$$\{(X-\lambda_1)^{r_1},\ldots,(X-\lambda_t)^{r_t}\}$$

then $(X-\lambda_1)^{r_1}$ appears $m_1\geq 1$ times, $(X-\lambda_2)^{r_2}$ appears $m_2\geq 1$ times, ..., $(X-\lambda_t)^{r_t}$ appears $m_t\geq 1$ times, with

$$m_1+m_2+\cdots+m_t=m.$$

The number, m_i, is called the *multiplicity* of $(X-\lambda_i)^{r_i}$. Furthermore, if $(X-\lambda_i)^{r_i}$ and $(X-\lambda_j)^{r_j}$ are two distinct elementary divisors, it is possible that $r_i\neq r_j$ yet $\lambda_i=\lambda_j$.

Observe that $(f-\lambda_i\mathrm{id})^{r_i}$ is nilpotent on V_i with index of nilpotency r_i (which means that $(f-\lambda_i\mathrm{id})^{r_i}=0$ on V_i but $(f-\lambda_i\mathrm{id})^{r_i-1}\neq 0$ on V_i). Also, note that the monomials, $(X-\lambda_i)$, are the irreducible factors of the minimal polynomial of f.

Next, let us take a closer look at the subspaces, V_i. It turns out that we can find a "good" basis of V_i so that in this basis, the restriction of f to V_i is a Jordan block.

Proposition 2.3. *Let V be a finite-dimensional vector space and let $f\colon V\to V$ be a linear map. If V is a cyclic $\mathbb{C}[X]$-module and if $(X-\lambda)^n$ is the minimal polynomial of f, then there is a basis of V of the form*

$$((f-\lambda\mathrm{id})^{n-1}(u),(f-\lambda\mathrm{id})^{n-2}(u),\ldots,(f-\lambda\mathrm{id})(u),u),$$

for some $u\in V$. With respect to this basis, the matrix of f is the Jordan block

$$J_n(\lambda)=\begin{pmatrix}\lambda & 1 & 0 & \cdots & 0\\ 0 & \lambda & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \ddots & \vdots\\ 0 & 0 & 0 & \ddots & 1\\ 0 & 0 & 0 & \cdots & \lambda\end{pmatrix}.$$

Consequently, λ is an eigenvalue of f.

Proof. A proof is given in Section 6. $\square$

Using Theorem 2.2 and Proposition 2.3 we get the Jordan form for complex matrices.

Theorem 2.4. *(Jordan Form) For every complex $n \times n$ matrix, A, there is some invertible matrix, P, and some Jordan matrix, J, so that*

$$A = PJP^{-1}.$$

If $\{\lambda_1, \ldots, \lambda_s\}$ is the set of eigenvalues of A, then the diagonal elements of the Jordan blocks of J are among the λ_i and every λ_i corresponds to one of more Jordan blocks of J. Furthermore, the number, m, of Jordan blocks, the distinct Jordan block, $J_{r_i}(\lambda_i)$, and the number of times, m_i, that each Jordan block, $J_{r_i}(\lambda_i)$, occurs are uniquely determined by A.

The number m_i is called the *multiplicity* of the block $J_{r_i}(\lambda_i)$. Observe that the column vector associated with the first entry of every Jordan block is an eigenvector of A. Thus, the number, m, of Jordan blocks is the number of linearly independent eigenvectors of A.

Beside the references that we cited for the proof of Theorem 2.2, other proofs of Theorem 2.4 can be found in the literature. Often, these proofs do not cover the uniqueness statement. For example, a nice proof is given in Godement [16], Chapter 35. Another interesting proof is given in Strang [28], Appendix B. A more "computational proof" is given in Horn and Johnson, [20], Chapter 3, Sections 1-4.

Observe that Theorem 2.4 implies that the characteristic polynomial, $q_f(X)$, of f is the product of the elementary divisors of f (counted with their multiplicity). But then, $q_f(X)$ must annihilate V. Therefore, we obtain a quick proof of the Cayley Hamilton Theorem (of course, we had to work hard to get Theorem 2.4!). Also, the minimal polynomial of f is the least common multiple (lcm) of the elementary divisors of f.

The following technical result will be needed for finding the logarithm of a real matrix:

Proposition 2.5. *If J is a $2n \times 2n$ complex Jordan matrix consisting of two conjugate blocks $J_n(\lambda + i\mu)$ and $J_n(\lambda - i\mu)$ of dimension n ($\mu \neq 0$), then there is a permutation matrix, P, and matrix, E, so that*

$$J = PEP^{-1},$$

where E is a block matrix of the form

$$E = \begin{pmatrix} D & I & 0 & \cdots & 0 \\ 0 & D & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & I \\ 0 & 0 & 0 & \cdots & D \end{pmatrix},$$

and with D the diagonal 2×2 matrix

$$D = \begin{pmatrix} \lambda + i\mu & 0 \\ 0 & \lambda - i\mu \end{pmatrix}.$$

Furthermore, there is a complex invertible matrix, Q, and a real Jordan matrix, C, so that

$$J = QCQ^{-1},$$

where C is of the form

$$C = \begin{pmatrix} L & I & 0 & \cdots & 0 \\ 0 & L & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & I \\ 0 & 0 & 0 & \cdots & L \end{pmatrix},$$

with

$$L = \begin{pmatrix} \lambda & -\mu \\ \mu & \lambda \end{pmatrix}.$$

Proof. First, consider an example, namely,

$$J = \begin{pmatrix} \lambda + i\mu & 1 & 0 & 0 \\ 0 & \lambda + i\mu & 0 & 0 \\ 0 & 0 & \lambda - i\mu & 1 \\ 0 & 0 & 0 & \lambda - i\mu \end{pmatrix}.$$

If we permute rows 2 and 3, we get

$$\begin{pmatrix} \lambda + i\mu & 1 & 0 & 0 \\ 0 & 0 & \lambda - i\mu & 1 \\ 0 & \lambda + i\mu & 0 & 0 \\ 0 & 0 & 0 & \lambda - i\mu \end{pmatrix}$$

and we permute columns 2 and 3, we get our matrix,

$$E = \begin{pmatrix} \lambda + i\mu & 0 & 1 & 0 \\ 0 & \lambda - i\mu & 0 & 1 \\ 0 & 0 & \lambda + i\mu & 0 \\ 0 & 0 & 0 & \lambda - i\mu \end{pmatrix}.$$

We leave it as an exercise to generalize this method to two $n \times n$ conjugate Jordan blocks to prove that we can find a permutation matrix, P, so that $E = P^{-1}JP$ and thus, $J = PEP^{-1}$.

Next, as $\mu \neq 0$, the matrix L can be diagonalized and one easily checks that

$$D = \begin{pmatrix} \lambda + i\mu & 0 \\ 0 & \lambda - i\mu \end{pmatrix} = \begin{pmatrix} -i & 1 \\ -i & -1 \end{pmatrix} \begin{pmatrix} \lambda & -\mu \\ \mu & \lambda \end{pmatrix} \begin{pmatrix} -i & 1 \\ -i & -1 \end{pmatrix}^{-1}.$$

Therefore, using the block diagonal matrix $S = \mathrm{diag}(S_2, \ldots, S_2)$ consisting of n blocks

$$S_2 = \begin{pmatrix} -i & 1 \\ -i & -1 \end{pmatrix},$$

we see that

$$E = SCS^{-1}$$

and thus,

$$J = PSCS^{-1}P^{-1},$$

which yields our second result with $Q = PS$. $\square$

Proposition 2.5 shows that every (complex) matrix, A, is similar to a real Jordan matrix. Unfortunately, if A is a *real* matrix, there is no guarantee that we can find a *real* invertible matrix, P, so that $A = PJP^{-1}$, with J a real Jordan matrix. This result known as the *Real Jordan Form* is actually true but requires some work to be established. In this section, we state the theorem without proof. A proof based on Theorem 2.2 is given in Section 6.

Theorem 2.6. *(Real Jordan Form) For every real $n \times n$ matrix, A, there is some invertible (real) matrix, P, and some real Jordan matrix, J, so that*

$$A = PJP^{-1}.$$

For every Jordan block, $J_r(\lambda)$, of type (1), λ is some real eigenvalue of A and for every Jordan block, $J_{2r}(\lambda, \mu)$, of type (2), $\lambda + i\mu$ is a complex eigenvalue of A (with $\mu \neq 0$). Every eigenvalue of A corresponds to one of more Jordan blocks of J. Furthermore, the number, m, of Jordan blocks, the distinct Jordan block, $J_{s_i}(\alpha_i)$, and the number of times, m_i, that each Jordan block, $J_{s_i}(\alpha_i)$, occurs are uniquely determined by A.

Let A be a real matrix and let $(X - \alpha_1)^{r_1}, \ldots, (X - \alpha_m)^{m_1}$ be its list of elementary divisors or, equivalently, let $J_{r_1}(\alpha_1), \ldots, J_{r_m}(\alpha_m)$ be its list of Jordan blocks. If, for every r_i and every real eigenvalue $\lambda_i < 0$, the number, m_i, of Jordan blocks identical to $J_{r_i}(\alpha_i)$ is even, then there is a way to rearrange these blocks using the technique of Proposition 2.5 to obtain a version of the real Jordan form that makes it easy to find logarithms (and square roots) of real matrices.

Theorem 2.7. *(Real Jordan Form, Special Version) Let A be a real $n \times n$ matrix and let $(X - \alpha_1)^{r_1}, \ldots, (X - \alpha_m)^{m_1}$ be its list of elementary divisors or, equivalently, let $J_{r_1}(\alpha_1), \ldots,$ $J_{r_m}(\alpha_m)$ be its list of Jordan blocks. If, for every r_i and every real eigenvalue $\alpha_i < 0$, the number, m_i, of Jordan blocks identical to $J_{r_i}(\alpha_i)$ is even, then there is a real invertible matrix, P, and a real Jordan matrix, J', such that $A = PJ'P^{-1}$ and*

(1) Every block, $J_{r_i}(\alpha_i)$, of J for which $\alpha_i \in \mathbb{R}$ and $\alpha_i \geq 0$ is a Jordan block of type (1) of J' (as in Definition 2.3), or

(2) *For every block, $J_{r_i}(\alpha_i)$, of J for which either $\alpha_i \in \mathbb{R}$ and $\alpha_i < 0$ or $\alpha_i = \lambda_i + i\mu_i$ with $\mu_i \neq 0$ ($\lambda_i, \mu_i \in \mathbb{R}$), the corresponding real Jordan block of J' is defined as follows:*

   (a) *If $\mu_i \neq 0$, then J' contains the real Jordan block $J_{2r_i}(\lambda_i, \mu_i)$ of type (2) (as in Definition 2.3), or*

   (b) *If $\alpha_i < 0$ then J' contains the real Jordan block $J_{2r_i}(\alpha_i, 0)$ whose diagonal blocks are of the form*

$$L(\alpha_i, 0) = \begin{pmatrix} \alpha_i & 0 \\ 0 & \alpha_i \end{pmatrix}.$$

Proof. By hypothesis, for every real eigenvalue, $\alpha_i < 0$, for every r_i, the Jordan block, $J_{r_i}(\alpha_i)$, occurs an even number of times say $2t_i$, so by using a permutation, we may assume that we have t_i pairs of identical blocks $(J_{r_i}(\alpha_i), J_{r_i}(\alpha_i))$. But then, for each pair of blocks of this form, we can apply part (1) of Proposition 2.5 (since α_i is its own conjugate), which yields our result. □

Remark: The above result generalizes the fact that when we have a rotation matrix, R, the eigenvalues -1 occurring in the real block diagonal form of R can be paired up.

The following theorem shows that the "structure" of the Jordan form of a matrix is preserved under exponentiation. This is an important result that will be needed to establish the necessity of the criterion for a real matrix to have a real logarithm. Again, in this section, we state the theorem without proof. A proof is given in Section 6.

Theorem 2.8. *For any (real or complex) $n \times n$ matrix, A, if $A = PJP^{-1}$ where J is a Jordan matrix of the form*

$$J = \begin{pmatrix} J_{r_1}(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(\lambda_m) \end{pmatrix},$$

then there is some invertible matrix, Q, so that the Jordan form of e^A is given by

$$e^A = Q\, e(J)\, Q^{-1},$$

where $e(J)$ is the Jordan matrix

$$e(J) = \begin{pmatrix} J_{r_1}(e^{\lambda_1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(e^{\lambda_m}) \end{pmatrix},$$

that is, each $J_{r_k}(e^{\lambda_k})$ is obtained from $J_{r_k}(\lambda_k)$ by replacing all the diagonal entries λ_k by e^{λ_k}. Equivalently, if the list of elementary divisors of A is

$$(X - \lambda_1)^{r_1}, \ldots, (X - \lambda_m)^{r_m},$$

then the list of elementary divisors of e^A is

$$(X - e^{\lambda_1})^{r_1}, \ldots, (X - e^{\lambda_m})^{r_m}.$$

3 Logarithms of Real Matrices; Criteria for Existence and Uniqueness

If A is any (complex) $n \times n$ matrix we say that a matrix, X, is a *logarithm of* A iff $e^X = A$. Our goal is to find conditions for the existence and uniqueness of real logarithms of real matrices. The two main theorems of this section are Theorem 3.4 and Theorem 3.10. These theorems are used in papers presenting methods for computing the logarithm of a matrix, including Cheng, Higham, Kenney and Laub [11] and Kenney and Laub [23].

Reference [11] cites Kenney and Laub [23] for a proof of Theorem 3.10 but in fact, that paper does not give a proof. Kenney and Laub [23] do state Theorem 3.10 as Lemma A2 of Appendix A, but they simply say that "the proof is similar to that of Lemma A1". As to the proof of Lemma A1, Kenney and Laub state without detail that it makes use of the Cauchy integral formula for operators, a method used by DePrima and Johnson [13] to prove a similar theorem for complex matrices (Section 4, Lemma 1) and where uniqueness is also proved. Kenney and Laub point out that the third hypothesis in that lemma is redundant. Theorem 3.10 also appears in Higham's book [18] as Theorem 1.31. Its proof relies on Theorem 1.28 and Theorem 1.18 (both in Higham's book) but Theorem 1.28 is not proved and only part of theorem 1.18 is proved in the text (closer examination reveals that Theorem 1.36 (in Higham's book) is needed to prove Theorem 1.28). Although Higham's Theorem 1.28 implies the injectivity statement of Theorem 3.8 we feel that the proof of Theorem 3.8 is of independent interest. Furthermore, Theorem 3.8 is a stronger result (it shows that exp is a diffeomorphism).

Given this state of affairs where no explicit proof of Theorem 3.10 seems easily available, we provide a complete proof of Theorem 3.10 using our special form of the Real Jordan Form.

First, let us consider the case where A is a complex matrix. Now, we know that if $A = e^X$, then $\det(A) = e^{\mathrm{tr}(X)} \neq 0$, so A must be invertible. It turns out that this condition is also sufficient.

Recall that for every invertible matrix, P, and every matrix, A,

$$e^{PAP^{-1}} = Pe^A P^{-1}$$

and that for every block diagonal matrix,

$$A = \begin{pmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{pmatrix},$$

we have

$$e^A = \begin{pmatrix} e^{A_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{A_m} \end{pmatrix}.$$

Consequenly, the problem of finding the logarithm of a matrix reduces to the problem of finding the logarithm of a Jordan block $J_r(\alpha)$ with $\alpha \neq 0$. However, every such Jordan block, $J_r(\alpha)$, can be written as

$$J_r(\alpha) = \alpha I + H = \alpha I(I + \alpha^{-1}H),$$

where H is the nilpotent matrix of index of nilpotency, r, given by

$$H = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Furthermore, it is obvious that $N = \alpha^{-1}H$ is also nilpotent of index of nilpotency, r, and we have

$$J_r(\alpha) = \alpha I(I + N).$$

Logarithms of the diagonal matrix, αI, are easily found. If we write $\alpha = \rho e^{i\theta}$ where $\rho > 0$, then $\log \alpha = \log \rho + i(\theta + 2\pi h)$, for any $h \in \mathbb{Z}$, and we can pick a logarithm of αI to be

$$S = \begin{pmatrix} \log \rho + i\theta & 0 & \cdots & 0 \\ 0 & \log \rho + i\theta & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \log \rho + i\theta \end{pmatrix}.$$

Observe that if we can find a logarithm, M, of $I + N$, as S commutes with any matrix and as $e^S = \alpha I$ and $e^M = I + N$, we have

$$e^{S+M} = e^S e^M = \alpha I(I + N) = J_r(\alpha),$$

which means that $S + M$ is a logarithm of $J_r(\alpha)$. Therefore, the problem reduces to finding the logarithm of a unipotent matrix, $I + N$. However, this problem always has a solution. To see this, remember that for $|u| < 1$, the power series

$$\log(1 + u) = u - \frac{u^2}{2} + \frac{u^3}{3} + \cdots + (-1)^{n+1}\frac{u^n}{n} + \cdots$$

is normally convergent. It turns out that the above fact can be generalized to matrices in the following way:

Proposition 3.1. *For every $n \times n$ matrix, A, such that $\|A\| < 1$, the series*

$$\log(I + A) = A - \frac{A^2}{2} + \frac{A^3}{3} + \cdots + (-1)^{n+1}\frac{A^n}{n} + \cdots$$

is normally convergent for any matrix norm $\| \, \|$ (a matrix norm satisfies the inequality $\|AB\| \leq \|A\| \, \|B\|$). Furthermore, if $\|A\| < 1$, then

$$e^{\log(I+A)} = I + A.$$

Remark: For any matrix norm $\| \, \|$ and any complex $n \times n$ matrix A, it can be shown that

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i| \leq \|A\| \,,$$

where the λ_i are the eigenvalues of A. Furthermore, the set of (complex) diagonalizable matrices is dense in the set of all complex matrices (see Serre [27]). Using these two facts, it can be shown that if $\|A\| < 1$, then

$$e^{\log(I+A)} = I + A$$

for any matrix norm.

For any given $r \geq 1$, the exponential and the logarithm (of matrices) turn out to give a homeomorphim between the set of nilpotent matrices, N, and the set of unipotent matrices, $I + N$, for which $N^r = 0$. Let $\mathcal{N}il(r)$ denote the set of (real or complex) nilpotent matrices of any dimension $n \geq 1$ such that $N^r = 0$ and $\mathcal{U}ni(r)$ denote the set of unipotent matrices, $U = I + N$, where $N \in \mathcal{N}il(r)$. If $U = I + N \in \mathcal{U}ni(r)$, note that $\log(I + N)$ is well-defined since the power series for $\log(I + N)$ only has $r - 1$ nonzero terms,

$$\log(I + N) = N - \frac{N^2}{2} + \frac{N^3}{3} + \cdots + (-1)^r \frac{N^{r-1}}{r-1}.$$

Proposition 3.2. *The exponential map,* $\exp\colon \mathcal{N}il(r) \to \mathcal{U}ni(r)$*, is a homeomorphism whose inverse is the logarithm.*

Proof. A complete proof can be found in Mmeimné and Testard [26], Chapter 3, Theorem 3.3.3. The idea is to prove that

$$\log(e^N) = N, \text{ for all } N \in \mathcal{N}il(r) \quad \textit{and} \quad e^{\log(U)} = U, \text{ for all } U \in \mathcal{U}ni(r).$$

To prove the first identity, it is enough to show that for any fixed $N \in \mathcal{N}il(r)$, we have

$$\log(e^{tN}) = tN, \qquad \text{for all } t \in \mathbb{R}.$$

To do this, observe that the functions $t \mapsto tN$ and $t \mapsto \log(e^{tN})$ are both equal to 0 for $t = 0$. Thus, it is enough to show that their derivatives are equal, which is left as an exercise.

Next, for any $N \in \mathcal{N}il(r)$, the map

$$t \mapsto e^{\log(I+tN)} - (I + tN), \qquad t \in \mathbb{R}$$

is a polynomial, since $N^r = 0$. Furthermore, for t sufficiently small, $\|tN\| < 1$ and in view of Proposition 3.1, we have $e^{\log(I+tN)} = I + tN$, so the above polynomial vanishes in a neighborhood of 0, which implies that it is identically zero. Therefore, $e^{\log(I+N)} = I + N$, as required. The continuity of exp and log is obvious. $\square$

Proposition 3.2 shows that every unipotent matrix, $I + N$, has the unique logarithm

$$\log(I + N) = N - \frac{N^2}{2} + \frac{N^3}{3} + \cdots + (-1)^r \frac{N^{r-1}}{r-1},$$

where r is the index of nilpotency of N. Therefore, if we let $M = \log(I+N)$, we have finally found a logarithm, $S + M$, for our original matrix, A. As a result of all this, we have proved the following theorem:

Theorem 3.3. *Every $n \times n$ invertible complex matrix, A, has a logarithm, X. To find such a logarithm, we can proceed as follows:*

(1) Compute a Jordan form, $A = PJP^{-1}$, for A and let m be the number of Jordan blocks in J.

(2) For every Jordan block, $J_{r_k}(\alpha_k)$, of J, write $J_{r_k}(\alpha_j) = \alpha_k I(I + N_k)$, where N_k is nilpotent.

(3) If $\alpha_k = \rho_k e^{i\theta_k}$, with $\rho_k > 0$, let

$$S_k = \begin{pmatrix} \log \rho_k + i\theta_k & 0 & \cdots & 0 \\ 0 & \log \rho_k + i\theta_k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \log \rho_k + i\theta_k \end{pmatrix}.$$

We have $\alpha_k I = e^{S_k}$.

(4) For every N_k, let

$$M_k = N_k - \frac{N_k^2}{2} + \frac{N_k^3}{3} + \cdots + (-1)^{r_k} \frac{N^{r_k - 1}}{r_k - 1}.$$

We have $I + N_k = e^{M_k}$.

(5) If $Y_k = S_k + M_k$ and Y is the block diagonal matrix $\mathrm{diag}(Y_1, \ldots, Y_m)$, then

$$X = PYP^{-1}$$

is a logarithm of A.

Let us now assume that A is a real matrix and let us try to find a real logarithm. There is no problem in finding real logarithms of the nilpotent parts but we run into trouble whenever an eigenvalue is complex or real negative. Fortunately, we can circumvent these problems by using the real Jordan form, provided that the condition of Theorem 2.7 holds.

The theorem below gives a necessary and sufficient condition for a real matrix to have a real logarithm. The first occurrence of this theorem that we have found in the literature is a

paper by Culver [12] published in 1966. The proofs in this paper rely heavily on results from Gantmacher [15]. Theorem 3.4 is also stated in Horn and Johnson [21] as Theorem 6.4.15 (Chapter 6), but the proof is left as an exercise. We offer a proof using Theorem 2.7 which is more explicit than Culver's proof.

Theorem 3.4. *Let A be a real $n \times n$ matrix and let $(X - \alpha_1)^{r_1}, \ldots, (X - \alpha_m)^{m_1}$ be its list of elementary divisors or, equivalently, let $J_{r_1}(\alpha_1), \ldots, J_{r_m}(\alpha_m)$ be its list of Jordan blocks. Then, A has a real logarithm iff A is invertible and if, for every r_i and every real eigenvalue $\alpha_i < 0$, the number, m_i, of Jordan blocks identical to $J_{r_i}(\alpha_i)$ is even.*

Proof. First, assume that A satisfies the conditions of Theorem 3.4. Since the matrix A satisfies the condition of Theorem 2.7, there is a real invertible matrix, P, and a real Jordan matrix, J', so that

$$A = PJ'P^{-1},$$

where J' satisfies conditions (1) and (2) of Theorem 2.7. As A is invertible, every block of J' of the form $J_{r_k}(\alpha_k)$ corresponds to a real eigenvalue with $\alpha_k > 0$ and we can write $J_{r_k}(\alpha_j) = \alpha_k I(I + N_k)$, where N_k is nilpotent. As in Theorem 3.3 (4), we can find a real logarithm, M_k, of $I + N_k$ and as $\alpha_k > 0$, the diagonal matrix $\alpha_k I$ has the real logarithm

$$S_k = \begin{pmatrix} \log \alpha_k & 0 & \cdots & 0 \\ 0 & \log \alpha_k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \log \alpha_k \end{pmatrix}.$$

Set $Y_k = S_k + M_k$.

The other real Jordan blocks of J' are of the form $J_{2r_k}(\lambda_k, \mu_k)$, with $\lambda_k, \mu_k \in \mathbb{R}$, not both zero. Consequently, we can write

$$J_{2r_k}(\lambda_k, \mu_k) = D_k + H_k = D_k(I + D_k^{-1} H_k)$$

where

$$D_k = \begin{pmatrix} L(\lambda_k, \mu_k) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & L(\lambda_k, \mu_k) \end{pmatrix}$$

with

$$L(\lambda_k, \mu_k) = \begin{pmatrix} \lambda_k & -\mu_k \\ \mu_k & \lambda_k \end{pmatrix},$$

and H_k is a real nilpotent matrix. If we let $N_k = D_k^{-1} H_k$, then N_k is also nilpotent, $J_{2r_k}(\lambda_k, \mu_k) = D_k(I + N_k)$, and we can find a logarithm, M_k, of $I + N_k$ as in Theorem 3.3 (4). We can write $\lambda_k + i\mu_k = \rho_k e^{i\theta_k}$, with $\rho_k > 0$ and $\theta_k \in [-\pi, \pi)$, and then

$$L(\lambda_k, \mu_k) = \begin{pmatrix} \lambda_k & -\mu_k \\ \mu_k & \lambda_k \end{pmatrix} = \rho_k \begin{pmatrix} \cos \theta_k & -\sin \theta_k \\ \sin \theta_k & \cos \theta_k \end{pmatrix}.$$

If we set

$$S(\rho_k, \theta_k) = \begin{pmatrix} \log \rho_k & -\theta_k \\ \theta_k & \log \rho_k \end{pmatrix},$$

a real matrix, we claim that

$$L(\lambda_k, \mu_k) = e^{S(\rho_k, \theta_k)}.$$

Indeed, $S(\rho_k, \theta_k) = \log \rho_k I + \theta_k E_2$, with

$$E_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

and it is well known that

$$e^{\theta_k E_2} = \begin{pmatrix} \cos \theta_k & -\sin \theta_k \\ \sin \theta_k & \cos \theta_k \end{pmatrix},$$

so, as $\log \rho_k I$ and $\theta_k E_2$ commute, we get

$$e^{S(\rho_k, \theta_k)} = e^{\log \rho_k I + \theta_k E_2} = e^{\log \rho_k I} e^{\theta_k E_2} = \rho_k \begin{pmatrix} \cos \theta_k & -\sin \theta_k \\ \sin \theta_k & \cos \theta_k \end{pmatrix} = L(\lambda_k, \mu_k).$$

If we form the real block diagonal matrix,

$$S_k = \begin{pmatrix} S(\rho_k, \theta_k) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S(\rho_k, \theta_k) \end{pmatrix},$$

we have $D_k = e^{S_k}$. Since S_k and M_k commute (observe that M_k is obtained from adding up powers of N_k and N_k only has 2×2 blocks above a diagonal of 2×2 blocks and so, it commutes with a block diagonal matrix of 2×2 blocks) and

$$e^{S_k + M_k} = e^{S_k} e^{M_k} = D_k(I + N_k) = J_{2r_k}(\lambda_k, \mu_k),$$

the matrix $Y_k = S_k + M_k$ is a logarithm of $J_{2r_k}(\lambda_k, \mu_k)$. Finally, if Y is the block diagonal matrix $\mathrm{diag}(Y_1, \ldots, Y_m)$, then $X = PYP^{-1}$ is a logarithm of A.

Let us now prove that if A has a real logarithm, X, then A satisfies the condition of Theorem 3.4. As we said before, A must be invertible. Since X is a real matrix, we know from the proof of Theorem 2.6 that the Jordan blocks of X associated with complex eigenvalues occur in conjugate pairs, so they are of the form

$$\begin{aligned} &J_{r_k}(\alpha_k), \quad \alpha_k \in \mathbb{R}, \\ &J_{r_k}(\alpha_k) \quad \text{and} \quad J_{r_k}(\overline{\alpha}_k), \quad \alpha_k = \lambda_k + i\mu_k,\ \mu_k \neq 0. \end{aligned}$$

By Theorem 2.8, the Jordan blocks of $A = e^X$ are obtained by replacing each α_k by e^{α_k}, that is, they are of the form

$$\begin{aligned} &J_{r_k}(e^{\alpha_k}), \quad \alpha_k \in \mathbb{R}, \\ &J_{r_k}(e^{\alpha_k}) \quad \text{and} \quad J_{r_k}(e^{\overline{\alpha}_k}), \quad \alpha_k = \lambda_k + i\mu_k,\ \mu_k \neq 0. \end{aligned}$$

If $\alpha_k \in \mathbb{R}$, then $e^{\alpha_k} > 0$, so the negative eigenvalues of A must be of the form e^{α_k} or $e^{\overline{\alpha}_k}$, with α_k complex. This implies that $\alpha_k = \lambda_k + (2h+1)i\pi$, for some $h \in \mathbb{Z}$, but then $\overline{\alpha}_k = \lambda_k - (2h+1)i\pi$ and so

$$e^{\alpha_k} = e^{\overline{\alpha}_k}.$$

Consequently, negative eigenvalues of A are associated with Jordan blocks that occur in pair, as claimed. $\square$

Remark: It can be shown (see Culver [12]) that all the logarithms of a Jordan block, $J_{r_k}(\alpha_k)$, corresponding to a real eigenvalue $\alpha_k > 0$ are obtained by adding the matrices

$$i2\pi h_k I, \quad h_k \in \mathbb{Z},$$

to the solution given by the proof of Theorem 3.4 and that all the logarithms of a Jordan block, $J_{2r_k}(\alpha_k, \beta_k)$, are obtained by adding the matrices

$$i2\pi h_k I + 2\pi l_k E \quad h_k, l_k \in \mathbb{Z},$$

to the solution given by the proof of Theorem 3.4, where

$$E = \begin{pmatrix} E_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & E_2 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

One should be careful no to relax the condition of Theorem 3.4 to the more liberal condition stating that for every Jordan block, $J_{r_k}(\alpha_k)$, for which $\alpha_k < 0$, the dimension r_k is even ($i.e$, α_k occurs an even number of times). For example, the following matrix

$$A = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$$

satisfies the more liberal condition but it does not possess any real logarithm, as the reader will verify. On the other hand, we have the following corollary:

Corollary 3.5. *For every real invertible matrix, A, if A has no negative eigenvalues, then A has a real logarithm.*

More results about the number of real logarithms of real matrices can be found in Culver [12]. In particular, Culver gives a necessary and sufficient condition for a real matrix, A, to have a unique real logarithm. This condition is quite strong. In particular, it requires that all the eigenvalues of A be real and positive.

A different approach is to restrict the domain of real logarithms to obtain a sufficient condition for the uniqueness of a logarithm. We now discuss this approach. First, we state the following property that will be useful later:

Proposition 3.6. *For every (real or complex) invertible matrix, A, there is a semisimple matrix, S, and a unipotent matrix, U, so that*

$$A = SU \qquad \text{and} \quad SU = US.$$

Furthermore, S and U as above are unique.

Proof. Proposition 3.6 follows immediately from Theorem 2.1, the details are left as an exercise. □

The form, SU, of an invertible matrix is often called the *multiplicative Jordan decomposition*.

Definition 3.1. Let $\mathcal{S}(n)$ denote the set of all real matrices whose eigenvalues, $\lambda + i\mu$, lie in the horizontal strip determined by the condition $-\pi < \mu < \pi$.

It is easy to see that $\mathcal{S}(n)$ is star-shaped (which means that if it contains A, then it contains λA for all $\lambda \in [0, 1]$) and open (because the roots of a polynomial are continuous functions of the coefficients of the polynomial). As $\mathcal{S}(n)$ is star-shaped, it is path-connected. Furthermore, if $A \in \mathcal{S}(n)$, then $PAP^{-1} \in \mathcal{S}(n)$ for every invertible matrix, P. The remarkable property of $\mathcal{S}(n)$ is that the restriction of the exponential to $\mathcal{S}(n)$ is a diffeomorphism onto its image. To prove this fact we will need the following proposition:

Proposition 3.7. *For any two real or complex matrices, S_1 and S_2, if the eigenvalues, $\lambda + i\mu$, of S_1 and S_2 satisfy the condition $-\pi < \mu \leq \pi$, if S_1 and S_2 are semisimple and if $e^{S_1} = e^{S_2}$, then $S_1 = S_2$.*

Proof. Since S_1 and S_2 are semisimple, they can be diagonalized over $\mathbb{C}$, so let $(u_1, \ldots, u_n)$ be a basis of eigenvectors of S_1 associated with the (possibly complex) eigenvalues $\lambda_1, \ldots, \lambda_n$ and let $(v_1, \ldots, v_n)$ be a basis of eigenvectors of S_2 associated with the (possibly complex) eigenvalues $\mu_1, \ldots, \mu_n$. We prove that if $e^{S_1} = e^{S_2} = A$, then $S_1(v_i) = S_2(v_i)$ for all v_i, which shows that $S_1 = S_2$.

Pick any eigenvector, v_i, of S_2 and write $v = v_i$ and $\mu = \mu_i$. We have

$$v = \alpha_1 u_1 + \cdots + \alpha_k u_k,$$

for some unique α_j's. We compute $A(v)$ in two different ways. We know that $e^{\mu_1}, \ldots, e^{\mu_n}$ are the eigenvalues of e^{S_2} for the eigenvectors $v_1, \ldots, v_n$, so

$$A(v) = e^{S_2}(v) = e^{\mu} v = \alpha_1 e^{\mu} u_1 + \cdots + \alpha_k e^{\mu} u_k.$$

Similarly, we know that $e^{\lambda_1}, \ldots, e^{\lambda_n}$ are the eigenvalues of e^{S_1} for the eigenvectors $u_1, \ldots, u_n$, so

$$\begin{aligned} A(v) &= A(\alpha_1 u_1 + \cdots + \alpha_k u_k) \\ &= \alpha_1 A(u_1) + \cdots + \alpha_k A(u_k) \\ &= \alpha_1 e^{S_1}(u_1) + \cdots + \alpha_k e^{S_1}(u_k) \\ &= \alpha_1 e^{\lambda_1} u_1 + \cdots + \alpha_k e^{\lambda_n} u_k. \end{aligned}$$

Therefore, we deduce that

$$\alpha_k e^{\mu} = \alpha_k e^{\lambda_k}, \qquad 1 \leq k \leq n.$$

Consequently, if $\alpha_k \neq 0$, then

$$e^{\mu} = e^{\lambda_k},$$

which implies $\mu - \lambda_k = i2\pi h$, for some $h \in \mathbb{Z}$. However, due to the hypothesis on the eigenvalues of S_1 and S_2, μ and λ_i must belong to the horizontal strip determined by the condition $-\pi < \Im(z) \leq \pi$, so we must have $h = 0$ and then $\mu = \lambda_k$.

If we let $I = \{k \mid \lambda_k = \mu\}$ (which is nonempty since $v \neq 0$), then $v = \sum_{k \in I} \alpha_k u_k$ and we have

$$\begin{aligned} S_1(v) &= S_1\left(\sum_{k \in I} \alpha_k u_k\right) \\ &= \sum_{k \in I} \alpha_k S_1(u_k) \\ &= \sum_{k \in I} \alpha_k \lambda_k u_k \\ &= \sum_{k \in I} \alpha_k \mu u_k \\ &= \mu \sum_{k \in I} \alpha_k u_k = \mu v. \end{aligned}$$

Therefore, $S_1(v) = \mu v$. As μ is an eigenvector of S_2 for the eigenvalue μ, we also have $S_2(v) = \mu v$. Therefore,

$$S_1(v_i) = S_2(v_i), \quad i = 1, \ldots, n,$$

which proves that $S_1 = S_2$. □

Obviously, Proposition 3.7 holds for real semisimple matrices, S_1, S_2, in $\mathcal{S}(n)$, since the condition for being in $\mathcal{S}(n)$ is $-\pi < \Im(\alpha) < \pi$ for every eigenvalue, α, of S_1 or S_2.

We can now state our next theorem, an important result. This theorem is a consequence of a more general fact proved in Bourbaki [9] (Chapter III, Section 6.9, Proposition 17, see also Theorem 6).

Theorem 3.8. *The restriction of the exponential map to $\mathcal{S}(n)$ is a diffeomorphism of $\mathcal{S}(n)$ onto its image, $\exp(\mathcal{S}(n))$. If $A \in \exp(\mathcal{S}(n))$, then $PAP^{-1} \in \mathcal{S}(n)$, for every (real) invertible matrix, P. Furthermore, $\exp(\mathcal{S}(n))$ is an open subset of $\mathbf{GL}(n, \mathbb{R})$ containing I and $\exp(\mathcal{S}(n))$ contains the open ball, $B(I, 1) = \{A \in \mathbf{GL}(n, \mathbb{R}) \mid \|A - I\| < 1\}$, for every norm $\| \, \|$ on $n \times n$ matrices satisfying the condition $\|AB\| \leq \|A\| \, \|B\|$.*

Proof. A complete proof is given in Mmeimné and Testard [26], Chapter 3, Theorem 3.8.4. Part of the proof consists in showing that exp is a local diffeomorphism and for this, to prove that $d\exp(X)$ is invertible. This requires finding an explicit formula for the derivative of the exponential and we prefer to omit this computation, which is quite technical. Proving that $B(I,1) \subseteq \mathcal{S}(n)$ is easier but requires a little bit of complex analysis. Once these facts are established, it remains to prove that exp is injective on $\mathcal{S}(n)$, which we will prove.

The trick is to use both the Jordan decomposition and the multiplicative Jordan decomposition! Assume that $X_1, X_2 \in \mathcal{S}(n)$ and that $e^{X_1} = e^{X_2}$. Using Theorem 2.1 we can write $X_1 = S_1 + N_1$ and $X_2 = S_2 + N_2$, where S_1, S_2 are semisimple, N_1, N_2 are nilpotent, $S_1N_1 = N_1S_1$, and $S_2N_2 = N_2S_2$. From $e^{X_1} = e^{X_2}$, we get

$$e^{S_1}e^{N_1} = e^{S_1+N_1} = e^{S_2+N_2} = e^{S_2}e^{N_2}.$$

Now, S_1 and S_2 are semisimple, so e^{S_1} and e^{S_2} are semisimple and N_1 and N_2 are nilpotent so e^{N_1} and e^{N_2} are unipotent. Moreover, as $S_1N_1 = N_1S_1$ and $S_2N_2 = N_2S_2$, we have $e^{S_1}e^{N_1} = e^{N_1}e^{S_1}$ and $e^{S_2}e^{N_2} = e^{N_2}e^{S_2}$. By the uniqueness property of Proposition 3.6, we conclude that

$$e^{S_1} = e^{S_2} \quad \text{and} \quad e^{N_1} = e^{N_2}.$$

Now, as N_1 and N_2 are nilpotent, there is some r so that $N_1^r = N_2^r = 0$ and then, it is clear that $e^{N_1} = I + \widetilde{N}_1$ and $e^{N_2} = I + \widetilde{N}_2$ with $\widetilde{N}_1^r = 0$ and $\widetilde{N}_2^r = 0$. Therefore, we can apply Proposition 3.2 to conclude that

$$N_1 = N_2.$$

As $S_1, S_2 \in \mathcal{S}(n)$ are semisimple and $e^{S_1} = e^{S_2}$, by Proposition 3.7, we conclude that

$$S_1 = S_2.$$

Therefore, we finally proved that $X_1 = X_2$, showing that exp is injective on $\mathcal{S}(n)$. □

Remark: Since proposition 3.7 holds for semisimple matrices, S, such that the condition $-\pi < \mu \leq \pi$ holds for every eigenvalue, $\lambda + i\mu$, of S, the restriction of the exponential to real matrices, X, whose eigenvalues satisfy this condition is injective. Note that the image of these matrices under the exponential contains matrices, $A = e^X$, with negative eigenvalues. Thus, combining Theorem 3.4 and the above injectivity result we could state an existence and uniqueness result for real logarithms of real matrices that is more general than Theorem 3.10 below. However this is not a practical result since it requires a condition on the number of Jordan blocks and such a condition is hard to check. Thus, we will restrict ourselves to real matrices with no negative eigenvalues (see Theorem 3.10).

Since the eigenvalues of a nilpotent matrix are zero and since symmetric matrices have real eigenvalues, Theorem 3.8 has has two interesting corollaries. Denote by $\mathbf{S}(n)$ the vector space of real $n \times n$ matrices and by $\mathbf{SPD}(n)$ the set of $n \times n$ symmetric, positive, definite matrices. It is known that $\exp\colon \mathbf{S}(n) \to \mathbf{SPD}(n)$ is a bijection.

Corollary 3.9. *The exponential map has the following properties:*

(1) The map $\exp\colon \mathcal{N}il(r) \to \mathcal{U}ni(r)$ *is a diffeomorphism.*

(2) The map $\exp\colon \mathbf{S}(n) \to \mathbf{SPD}(n)$ *is a diffeomorphism.*

By combining Theorem 3.4 and Theorem 3.8 we obtain the following result about the existence and uniqueness of logarithms of real matrices:

Theorem 3.10. *(a) If A is any real invertible $n\times n$ matrix and A has no negative eigenvalues, then A has a unique real logarithm, X, with $X \in \mathcal{S}(n)$.*

(b) The image, $\exp(\mathcal{S}(n))$*, of $\mathcal{S}(n)$ by the exponential map is the set of real invertible matrices with no negative eigenvalues and* $\exp\colon \mathcal{S}(n) \to \exp(\mathcal{S}(n))$ *is a diffeomorphism between these two spaces.*

Proof. (a) If we go back to the proof of Theorem 3.4, we see that complex eigenvalues of the logarithm, X, produced by that proof only occur for matrices

$$S(\rho_k, \theta_k) = \begin{pmatrix} \log \rho_k & -\theta_k \\ \theta_k & \log \rho_k \end{pmatrix},$$

associated with eigenvalues $\lambda_k + i\mu_k = \rho_k\, e^{i\theta_k}$. However, the eigenvalues of such matrices are $\log \rho_k \pm i\theta_k$ and since A has no negative eigenvalues, we may assume that $-\pi < \theta_k < \pi$, and so $X \in \mathcal{S}(n)$, as desired. By Theorem 3.8, such a logarithm is unique.

(b) Part (a) proves that the set of real invertible matrices with no negative eigenvalues is contained in $\exp(\mathcal{S}(n))$. However, for any matrix, $X \in \mathcal{S}(n)$, since every eigenvalue of e^X is of the form $e^{\lambda+i\mu} = e^{\lambda}e^{i\mu}$ for some eigenvalue, $\lambda + i\mu$, of X and since $\lambda + i\mu$ satisfies the condition $-\pi < \mu < \pi$, the number, $e^{i\mu}$, is never negative, so e^X has no negative eigenvalues. Then, (b) follows directly from Theorem 3.8. $\square$

Remark: Theorem 3.10 (a) first appeared in Kenney and Laub [23] (Lemma A2, Appendix A) but without proof.

4 Square Roots of Real Matrices; Criteria for Existence and Uniqueness

In this section we investigate the problem of finding a square root of a matrix, A, that is, a matrix, X, such that $X^2 = A$. If A is an invertible (complex) matrix, then it always has a

square root, but singular matrices may fail to have a square root. For example, the nilpotent matrix,

$$H = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

has no square root (ckeck this!). The problem of finding square roots of matrices is thoroughly investigated in Gantmacher [15], Chapter VIII, Sections 6 and 7. For singular matrices, finding a square root reduces to the problem of finding the square root of a nilpotent matrix, which is not always possible. A necessary and sufficient condition for the existence of a square root is given in Horn and Johnson [21], see Chapter 6, Section 4, especially Theorem 6.1.12 and Theorem 6.4.14. This criterion is rather complicated because its deals with non-singular as well as singular matrices. In this paper, we will restrict our attention to invertible matrices. The main two Theorems of this section are Theorem 4.4 and Theorem 4.8. The former theorem appears in Higham [17] (Theorem 5). The first step is to prove a version of Theorem 2.8 for the function $A \mapsto A^2$, where A is invertible. In this section, we state the following theorem without proof. A proof is given in Section 6.

Theorem 4.1. *For any (real or complex) invertible $n \times n$ matrix, A, if $A = PJP^{-1}$ where J is a Jordan matrix of the form*

$$J = \begin{pmatrix} J_{r_1}(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(\lambda_m) \end{pmatrix},$$

then there is some invertible matrix, Q, so that the Jordan form of A^2 is given by

$$e^A = Q\, s(J)\, Q^{-1},$$

where $s(J)$ is the Jordan matrix

$$s(J) = \begin{pmatrix} J_{r_1}(\lambda_1^2) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(\lambda_m^2) \end{pmatrix},$$

that is, each $J_{r_k}(\lambda_k^2)$ is obtained from $J_{r_k}(\lambda_k)$ by replacing all the diagonal enties λ_k by λ_k^2. Equivalently, if the list of elementary divisors of A is

$$(X - \lambda_1)^{r_1}, \ldots, (X - \lambda_m)^{r_m},$$

then the list of elementary divisors of A^2 is

$$(X - \lambda_1^2)^{r_1}, \ldots, (X - \lambda_m^2)^{r_m}.$$

Remark: Theorem 4.1 can be easily generalized to the map $A \mapsto A^p$, for any $p \geq 2$, that is, by replacing A^2 by A^p, provided A is invertible. Thus, if the list of elementary divisors of A is

$$(X - \lambda_1)^{r_1}, \ldots, (X - \lambda_m)^{r_m},$$

then the list of elementary divisors of A^p is

$$(X - \lambda_1^p)^{r_1}, \ldots, (X - \lambda_m^p)^{r_m}.$$

The next step is to find the square root of a Jordan block. Since we are assuming that our matrix is invertible, every Jordan block, $J_{r_k}(\alpha_k)$, can be written as

$$J_{r_k}(\alpha_k) = \alpha_k I \left(I + \frac{H}{\alpha_k} \right),$$

where H is nilpotent. It is easy to find a square root of $\alpha_k I$. If $\alpha_k = \rho_k e^{\theta_k}$, with $\rho_k > 0$, then

$$S_k = \begin{pmatrix} \sqrt{\rho_k}\, e^{i\frac{\theta_k}{2}} & 0 & \cdots & 0 \\ 0 & \sqrt{\rho_k}\, e^{i\frac{\theta_k}{2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\rho_k}\, e^{i\frac{\theta_k}{2}} \end{pmatrix}$$

is a square root of $\alpha_k I$. Therefore, the problem reduces to finding square roots of unipotent matrices. For this, we recall the power series

$$\begin{aligned} (1+x)^{\frac{1}{2}} &= 1 + \frac{1}{2}x + \cdots + \frac{1}{n!}\frac{1}{2}\left(\frac{1}{2} - 1\right)\cdots\left(\frac{1}{2} - n + 1\right)x^n + \cdots \\ &= \sum_{n=0}^{\infty} (-1)^{n-1} \frac{(2n)!}{(2n-1)(n!)^2 2^{2n}}\, x^n, \end{aligned}$$

which is normally convergent for $|x| < 1$. Then, we can define the power series, R, of a matrix variable, A, by

$$R(A) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(2n)!}{(2n-1)(n!)^2 2^{2n}}\, A^n,$$

and this power series converges normally for $\|A\| < 1$. As a formal power series, note that $R(0) = 0$ and $R'(0) = \frac{1}{2} \neq 0$ so, by a theorem about formal power series, R has a unique inverse, S, such that $S(0) = 0$ (see Lang [25] or H. Cartan [10]). But, if we consider the power series, $S(A) = (I + A)^2 - I$, when A is a real number, we have $R(A) = \sqrt{1 + A} - 1$, so we get

$$R \circ S(A) = \sqrt{1 + (1 + A)^2 - 1} - 1 = A,$$

from wich we deduce that S and R are mutual inverses. But, R converges everywhere and S converges for $\|A\| < 1$, so by another theorem about converging power series, if we let $\sqrt{I+A} = R(A) + I$, there is some r, with $0 < r < 1$, so that

$$(\sqrt{I+A})^2 = I + A, \qquad \text{if} \quad \|A\| < r$$

and

$$\sqrt{(I+A)^2} = I + A, \qquad \text{if} \quad \|A\| < r.$$

If A is unipotent, that is, $A = I + N$ with N nilpotent, we see that the series has only finitely many terms. This fact allows us to prove the proposition below.

Proposition 4.2. *The squaring map, $A \mapsto A^2$, is a homeomorphism from $\mathcal{U}ni(r)$ to itself whose inverse is the map $A \mapsto \sqrt{A} = R(A - I) + I$.*

Proof. If $A = I + N$ with $N^r = 0$, as $A^2 = I + 2N + N^2$ it is clear that $(2N + N^2)^r = 0$, so the squaring map is well defined on unipotent matrices. We use the technique of Proposition 3.2. Consider the map

$$t \mapsto (\sqrt{I + tN})^2 - (I + tN), \qquad t \in \mathbb{R}.$$

It is a polynomial since $N^r = 0$. Furthermore, for t sufficiently small, $\|tN\| < 1$ and we have $(\sqrt{I + tN})^2 = I + tN$, so the above polynomial vanishes in a neighborhood of 0, which implies that it is identically zero. Therefore, $(\sqrt{I + N})^2 = I + N$, as required.

Next, consider the map

$$t \mapsto \sqrt{(I + tN)^2} - (I + tN), \qquad t \in \mathbb{R}.$$

It is a polynomial since $N^r = 0$. Furthermore, for t sufficiently small, $\|tN\| < 1$ and we have $\sqrt{(I + tN)^2} = I + tN$, so we conclude as above that the above map is identically zero and that $\sqrt{(I + N)^2} = I + N$. $\square$

Remark: Proposition 4.2 can be easily generalized to the map $A \mapsto A^p$, for any $p \geq 2$, by using the power series

$$(I + A)^{\frac{1}{p}} = I + \frac{1}{p}A + \cdots + \frac{1}{n!}\frac{1}{p}\left(\frac{1}{p} - 1\right)\cdots\left(\frac{1}{p} - n + 1\right)A^n + \cdots.$$

Using proposition 4.2, we can find a square root for the unipotent part of a Jordan block,

$$J_{r_k}(\alpha_k) = \alpha_k I\left(I + \frac{H}{\alpha_k}\right).$$

If $N_k = \frac{H}{\alpha_k}$, then

$$\sqrt{I + N_k} = I + \sum_{j=1}^{r_k - 1}(-1)^{j-1}\frac{(2j)!}{(2j-1)(j!)^2 2^{2j}}\, N_k^j$$

is a square root of $I + N_k$. Therefore, we obtained the following theorem:

Theorem 4.3. *Every (complex) invertible matrix, A, has a square root.*

Remark: Theorem 4.3 can be easily generalized to p^{th} roots, for any $p \geq 2$,

We now consider the problem of finding a real square root of an invertible real matrix. It turns out that the necessary and sufficient condition is exactly the condition for finding a real logarithm of a real matrix.

Theorem 4.4. *Let A be a real invertible $n \times n$ matrix and let $(X-\alpha_1)^{r_1}, \ldots, (X-\alpha_m)^{m_1}$ be its list of elementary divisors or, equivalently, let $J_{r_1}(\alpha_1), \ldots, J_{r_m}(\alpha_m)$ be its list of Jordan blocks. Then, A has a real square root iff for every r_i and every real eigenvalue $\alpha_i < 0$, the number, m_i, of Jordan blocks identical to $J_{r_i}(\alpha_i)$ is even.*

Proof. The proof is very similar to the proof of Theorem 3.4 so we only point out the necessary changes. Let J' be a real Jordan matrix so that

$$A = PJ'P^{-1},$$

where J' satisfies conditions (1) and (2) of Theorem 2.7. As A is invertible, every block of J' of the form $J_{r_k}(\alpha_k)$ corresponds to a real eigenvalue with $\alpha_k > 0$ and we can write $J_{r_k}(\alpha_j) = \alpha_k I(I + N_k)$, where N_k is nilpotent. As in Theorem 4.3, we can find a real square root, M_k, of $I + N_k$ and as $\alpha_k > 0$, the diagonal matrix $\alpha_k I$ has the real square root

$$S_k = \begin{pmatrix} \sqrt{\alpha_k} & 0 & \cdots & 0 \\ 0 & \sqrt{\alpha_k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\alpha_k} \end{pmatrix}.$$

Set $Y_k = S_k M_k$.

The other real Jordan blocks of J' are of the form $J_{2r_k}(\lambda_k, \mu_k)$, with $\lambda_k, \mu_k \in \mathbb{R}$, not both zero. Consequently, we can write

$$J_{2r_k}(\lambda_k, \mu_k) = D_k(I + N_k)$$

where

$$D_k = \begin{pmatrix} L(\lambda_k, \mu_k) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & L(\lambda_k, \mu_k) \end{pmatrix}$$

with

$$L(\lambda_k, \mu_k) = \begin{pmatrix} \lambda_k & -\mu_k \\ \mu_k & \lambda_k \end{pmatrix},$$

and $N_k = D_k^{-1} H_k$ is nilpotent. We can find a square root, M_k, of $I + N_k$ as in Theorem 4.3. If we write $\lambda_k + i\mu_k = \rho_k e^{i\theta_k}$, then

$$L(\lambda_k, \mu_k) = \rho_k \begin{pmatrix} \cos\theta_k & -\sin\theta_k \\ \sin\theta_k & \cos\theta_k \end{pmatrix}.$$

Then, if we set

$$S(\rho_k, \theta_k) = \sqrt{\rho_k} \begin{pmatrix} \cos\left(\frac{\theta_k}{2}\right) & -\sin\left(\frac{\theta_k}{2}\right) \\ \sin\left(\frac{\theta_k}{2}\right) & \cos\left(\frac{\theta_k}{2}\right) \end{pmatrix},$$

a real matrix, we have

$$L(\lambda_k, \mu_k) = S(\rho_k, \theta_k)^2.$$

If we form the real block diagonal matrix,

$$S_k = \begin{pmatrix} S(\rho_k, \theta_k) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S(\rho_k, \theta_k) \end{pmatrix},$$

we have $D_k = S_k^2$ and then the matrix $Y_k = S_k M_k$ is a square root of $J_{2r_k}(\lambda_k, \mu_k)$. Finally, if Y is the block diagonal matrix $\mathrm{diag}(Y_1, \ldots, Y_m)$, then $X = PYP^{-1}$ is a square root of A.

Let us now prove that if A has a real square root, X, then A satisfies the condition of Theorem 4.4. Since X is a real matrix, we know from the proof of Theorem 2.6 that the Jordan blocks of X associated with complex eigenvalues occur in conjugate pairs, so they are of the form

$$\begin{aligned} &J_{r_k}(\alpha_k), \quad \alpha_k \in \mathbb{R}, \\ &J_{r_k}(\alpha_k) \quad \text{and} \quad J_{r_k}(\overline{\alpha}_k), \quad \alpha_k = \lambda_k + i\mu_k,\ \mu_k \neq 0. \end{aligned}$$

By Theorem 4.1, the Jordan blocks of $A = X^2$ are obtained by replacing each α_k by α_k^2, that is, they are of the form

$$\begin{aligned} &J_{r_k}(\alpha_k^2), \quad \alpha_k \in \mathbb{R}, \\ &J_{r_k}(\alpha_k^2) \quad \text{and} \quad J_{r_k}(\overline{\alpha}_k^2), \quad \alpha_k = \lambda_k + i\mu_k,\ \mu_k \neq 0. \end{aligned}$$

If $\alpha_k \in \mathbb{R}$, then $\alpha_k^2 > 0$, so the negative eigenvalues of A must be of the form α_k^2 or $\overline{\alpha}_k^2$, with α_k complex. This implies that $\alpha_k = \sqrt{\rho_k}\, e^{i\frac{\pi}{2}}$, but then $\overline{\alpha}_k = \sqrt{\rho_k}\, e^{-i\frac{\pi}{2}}$ and so

$$\alpha_k^2 = \overline{\alpha}_k^2.$$

Consequently, negative eigenvalues of A are associated with Jordan blocks that occur in pair, as claimed. $\square$

Remark: Theorem 4.4 can be easily generalized to p^{th} roots, for any $p \geq 2$,

Theorem 4.4 appears in Higham [17] as Theorem 5 but no explicit proof is given. Instead, Higham states: "The proof is a straightfoward modification of Theorem 1 in Culver [12] and is omitted." Culver's proof uses results from Gantmacher [15] and does not provide a constructive method for obtaining a square root. We gave a more constructive proof (but perhaps longer).

Corollary 4.5. *For every real invertible matrix, A, if A has no negative eingenvalues, then A has a real square root.*

We will now provide a sufficient condition for the uniqueness of a real square root. For this, we consider the open set, $\mathcal{H}(n)$, consisting of all real $n \times n$ matrices whose eigenvalues, $\alpha = \lambda + i\mu$, have a positive real part, $\lambda > 0$. We express this condition as $\Re(\alpha) > 0$. Obviously, such matrices are invertible and can't have negative eigenvalues. We need a version of Proposition 3.7 for semisimple matrices in $\mathcal{H}(n)$.

Remark: To deal with p^{th} roots, we consider matrices whose eigenvalues, $\rho e^{i\theta}$, satisfy the condition $-\frac{\pi}{p} < \theta < \frac{\pi}{p}$.

Proposition 4.6. *For any two real or complex matrices, S_1 and S_2, if the eigenvalues, $\rho e^{i\theta}$, of S_1 and S_2 satisfy the condition $-\frac{\pi}{2} < \theta \leq \frac{\pi}{2}$, if S_1 and S_2 are semisimple and if $S_1^2 = S_2^2$, then $S_1 = S_2$.*

Proof. The proof is very similar to that of Proposition 3.7 so we only indicate where modifications are needed. We use the fact that if u is an eigenvector of a linear map, A, associated with some eigenvalue, λ, then u is an eigenvector of A^2 associated with the eigenvalue λ^2. We replace every occurrence of e^{λ_i} by λ_i^2 (and e^{μ} by μ^2). As in the proof of Proposition 3.7, we obtain the equation

$$\alpha_1 \mu^2 u_1 + \cdots + \alpha_k \mu^2 u_k = \alpha_1 \lambda_1^2 u_1 + \cdots + \alpha_k \lambda^2 u_k.$$

Therefore, we deduce that

$$\alpha_k \mu^2 = \alpha_k \lambda_k^2, \qquad 1 \leq k \leq n.$$

Consequently, as $\mu, \lambda_k \neq 0$, if $\alpha_k \neq 0$, then

$$\mu^2 = \lambda_k^2,$$

which implies $\mu = \pm\lambda_k$. However, the hypothesis on the eigenvalues of S_1 and S_2 implies that $\mu = \lambda_k$. The end of the proof is identical to that of Proposition 3.7. $\square$

Obviously, Proposition 4.6 holds for real semisimple matrices, S_1, S_2, in $\mathcal{H}(n)$.

Remark: Proposition 4.6 also holds for the map $S \mapsto S^p$, for any $p \geq 2$, under the condition $-\frac{\pi}{p} < \theta \leq \frac{\pi}{p}$.

We have the following analog of Theorem 3.8, but we content ourselves with a weaker result:

Theorem 4.7. *The restriction of the squaring map, $A \mapsto A^2$, to $\mathcal{H}(n)$ is injective.*

Proof. Let $X_1, X_2 \in \mathcal{H}(n)$ and assume that $X_1^2 = X_2^2$. As X_1 and X_2 are invertible, by Proposition 3.6, we can write $X_1 = S_1(I + N_1)$ and $X_2 = S_2(I + N_2)$, where S_1, S_2 are semisimple, N_1, N_2 are nilpotent, $S_1(I+N_1) = (I+N_1)S_1$ and $S_2(I+N_2) = (I+N_2)S_2$. As $X_1^2 = X_2^2$, we get

$$S_1^2(I + N_1)^2 = S_2^2(I + N_2)^2.$$

Now, as S_1 and S_2 are semisimple and invertible, S_1^2 and S_2^2 are semisimple and invertible, and as N_1 and N_2 are nilpotent, $2N_1 + N_1^2$ and $2N_2 + N_2^2$ are nilpotent, so $(I + N_1)^2$ and $(I + N_2)^2$ are unipotent. Moreover, $S_1(I + N_1) = (I + N_1)S_1$ and $S_2(I + N_2) = (I + N_2)S_2$ imply that $S_1^2(I + N_1)^2 = (I + N_1)^2 S_1^2$ and $S_2^2(I + N_2)^2 = (I + N_2)^2 S_2^2$. Therefore, by the uniqueness statement of Proposition 3.6, we get

$$S_1^2 = S_2^2 \quad \text{and} \quad (I + N_1)^2 = (I + N_2)^2.$$

However, as $X_1, X_2 \in \mathcal{H}(n)$ we have $S_1, S_2 \in \mathcal{H}(n)$ and Proposition 4.6 implies that $S_1 = S_2$. Since $I + N_1$ and $I + N_2$ are unipotent, proposition 4.2 implies that $N_1 = N_2$. Therefore, $X_1 = X_2$, as required. □

Remark: Theorem 4.7 also holds for the restriction of the squaring map to real or complex matrices, X, whose eigenvalues, $\rho e^{i\theta}$, satisfy the condition $-\frac{\pi}{2} < \theta \leq \frac{\pi}{2}$. This result is proved in DePrima and Johnson [13] by a different method. However, DePrima and Johnson need an extra condition, see the discussion at the end of this section.

We can now prove the analog of Theorem 3.10 for square roots.

Theorem 4.8. *If A is any real invertible $n \times n$ matrix and A has no negative eigenvalues, then A has a unique real square root, X, with $X \in \mathcal{H}(n)$.*

Proof. If we go back to the proof of Theorem 4.4, we see that complex eigenvalues of the square root, X, produced by that proof only occur for matrices

$$S(\rho_k, \theta_k) = \sqrt{\rho_k} \begin{pmatrix} \cos\left(\frac{\theta_k}{2}\right) & -\sin\left(\frac{\theta_k}{2}\right) \\ \sin\left(\frac{\theta_k}{2}\right) & \cos\left(\frac{\theta_k}{2}\right) \end{pmatrix},$$

associated with eigenvalues $\lambda_k + i\mu_k = \rho_k\, e^{i\theta_k}$. However, the eigenvalues of such matrices are $\sqrt{\rho_k}\, e^{\pm i\frac{\theta_k}{2}}$ and since A has no negative eigenvalues, we may assume that $-\pi < \theta_k < \pi$, and so $-\frac{\pi}{2} < \frac{\theta_k}{2} < \frac{\pi}{2}$, wich means that $X \in \mathcal{H}(n)$, as desired. By Theorem 4.7, such a square root is unique. □

Theorem 4.8 is stated in a number of papers including Bini, Higham and Meini [7], Cheng, Higham, Kenney and Laub [11] and Kenney and Laub [23]. Theorem 4.8 also appears in Higham [18] as Theorem 1.29. Its proof relies on Theorem 1.26 and Theorem 1.18 (both in

Higham's book), whose proof is not given in full (closer examination reveals that Theorem 1.36 (in Higham's book) is needed to prove Theorem 1.26). Although Higham's Theorem 1.26 implies our Theorem 4.7 we feel that the proof of Theorem 4.7 is of independent interest and is more direct.

As we already said in Section 3, Kenney and Laub [23] state Theorem 4.8 as Lemma A1 in Appendix A. The proof is sketched briefly. Existence follows from the Cauchy integral formula for operators, a method used by DePrima and Johnson [13] in which a similar result is proved for complex matrices (Section 4, Lemma 1). Uniqueness is proved in DePrima and Johnson [13] but it uses an extra condition. The hypotheses of Lemma 1 in DePrima and Johnson are that A and X are *complex* invertible matrices and that X satisfies the conditions

(i) $X^2 = A$,

(ii) the eigenvalues, $\rho\, e^{i\theta}$, of X satisfy $-\frac{\pi}{2} < \theta \leq \frac{\pi}{2}$,

(iii) For any matrix, S, if $AS = SA$, then $XS = SX$.

Observe that condition (ii) allows $\theta = \frac{\pi}{2}$, which yields matrices, $A = X^2$, with negative eigenvalues. In this case, A may not have any real square root but DePrima and Johnson are only concerned with *complex* matrices and a complex square root always exists. To guarantee the existence of real logarithms, Kenney and Laub tighten condition (ii) to $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$. They also assert that condition (iii) follows from conditions (i) and (ii). This can be shown as follows: First, recall that we have shown that uniqueness follows from (i) and (ii). Uniqueness under conditions (i) and (ii) can also be shown to be a consequence of Theorem 2 in Higham [17]. Now, assume $X^2 = A$ and $SA = SA$. We may assume that S is invertible since the set of invertible matrices is dense in the set of all matrices. Then, as $SA = AS$, we have

$$(SXS^{-1})^2 = SX^2S^{-1} = SAS^{-1} = A.$$

Thus, SXS^{-1} is a square root of A. Furthermore, X and SXS^{-1} have the same eigenvalues so SXS^{-1} satisfies (i) and (ii) and, by uniqueness, $X = SXS^{-1}$, that is, $XS = SX$.

Since Kenney and Laub only provide a sketch of Theorem A1 and since Higham [18] does not give all the details of the proof either, we felt that the reader would appreciate seeing a complete proof of Theorem 4.8.

5 Conclusion

It is interesting that Theorem 3.10 and Theorem 4.8 are the basis for numerical methods for computing the exponential or the logarithm of a matrix. The key point is that the following identities hold:

$$e^A = (e^{A/2^k})^{2^k} \quad \text{and} \quad \log(A) = 2^k \log(A^{1/2^k}),$$

where in the second case, $A^{1/2^k}$ is the unique kth square root of A whose eigenvalues, $\rho\, e^{i\theta}$, lie in the sector $-\frac{\pi}{2^k} < \theta < \frac{\pi}{2^k}$. The first identity is trivial and the second one can be shown by induction from the identity

$$\log(A) = 2\log(A^{1/2}),$$

where $A^{1/2}$ is the unique square root of A whose eigenvalues, $\rho\, e^{i\theta}$, lie in the sector $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$. Let $\widetilde{X} = A^{1/2}$, whose eigenvalues, $\rho\, e^{i\theta}$, lie in the sector $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$. Then, it is easy to see that the eigenvalues, α, of $\log(\widetilde{X})$ satisfy the condition $-\frac{\pi}{2} < \Im(\alpha) < \frac{\pi}{2}$. Then, $X = 2\log(\widetilde{X}) = 2\log(A^{1/2})$ satisfies

$$e^X = e^{\log(A^{1/2})+\log(A^{1/2})} = e^{\log(A^{1/2})}e^{\log(A^{1/2})} = A^{1/2}A^{1/2} = A,$$

and the eigenvalues, α, of X satisfy the condition $-\pi < \Im(\alpha) < \pi$ so, by the uniqueness part of Theorem 3.10, we must have $\log(A) = 2\log(A^{1/2})$.

The identity $\log(A) = 2^k \log(A^{1/2^k})$ leads to a numerical method for computing the logarithm of a (real) matrix first introduced by Kenney and Laub known as the *inverse scaling and squaring algorithm*, see Kenney and Laub [23] and Cheng, Higham, Kenney and Laub [11]. The idea is that if A is close to the identity, then $\log(A)$ can be computed accurately using either a truncated power series expansion of $\log(A)$ or better, rational approximations know as *Padé approximants*. In order to bring A close to the identity, iterate the operation of taking the square root of A to obtain $A^{1/2^k}$. Then, after having computed $\log(A^{1/2^k})$, scale $\log(A^{1/2^k})$ by the factor 2^k. For details of this method, see Kenney and Laub [23] and Cheng, Higham, Kenney and Laub [11]. The inverse squaring and scaling method plays an important role in the *log-Euclidean framework* introduced by Arsigny, Fillard, Pennec and Ayache, see Arsigny [1], Arsigny, Fillard, Pennec and Ayache [3, 4] and Arsigny, Pennec and Ayache [5].

6 Appendix; Some Proofs Regarding the Jordan Form

Proposition 2.3. *Let V be a finite-dimensional vector space and let $f\colon V \to V$ be a linear map. If V is a cyclic $\mathbb{C}[X]$-module and if $(X-\lambda)^n$ is the minimal polynomial of f, then there is a basis of V of the form*

$$((f-\lambda\mathrm{id})^{n-1}(u), (f-\lambda\mathrm{id})^{n-2}(u), \ldots, (f-\lambda\mathrm{id})(u), u),$$

for some $u \in V$. With respect to this basis, the matrix of f is the Jordan block

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}.$$

Consequently, λ is an eigenvalue of f.

Proof. Since V is a cyclic $\mathbb{C}[X]$-module, there is some $u \in V$ so that V is generated by $u, f(u), f^2(u), \ldots$, which means that every vector in V is of the form $p(f)(u)$, for some polynomial, $p(X)$. We claim that $u, f(u), \ldots, f^{n-2}(u), f^{n-1}(u)$ generate V, which implies that the dimension of V is at most n.

This is because if $p(X)$ is any polynomial of degree at least n, then we can divide $p(X)$ by $(X-\lambda)^n$ obtaining

$$p = (X-\lambda)^n q + r,$$

where $0 \leq \deg(r) < n$ and as $(X-\lambda)^n$ annihilates V, we get

$$p(f)(u) = r(f)(u),$$

which means that every vector of the form $p(f)(u)$ with $p(X)$ of degree $\geq n$ is actually a linear combination of $u, f(u), \ldots, f^{n-2}(u), f^{n-1}(u)$.

We claim that the vectors

$$u, (f-\lambda\mathrm{id})(u), \ldots, (f-\lambda\mathrm{id})^{n-2}(u)(f-\lambda\mathrm{id})^{n-1}(u)$$

are linearly independent. Indeed, if we had a nontrivial linear combination

$$a_0(f-\lambda\mathrm{id})^{n-1}(u) + a_1(f-\lambda\mathrm{id})^{n-2}(u) + \cdots + a_{n-2}(f-\lambda\mathrm{id})(u) + a_{n-1}u = 0,$$

then the polynomial

$$a_0(X-\lambda)^{n-1} + a_1(X-\lambda)^{n-2} + \cdots + a_{n-2}(X-\lambda) + a_{n-1}$$

of degree at most $n-1$ would annihilate V, contradicting the fact that $(X-\lambda)^n$ is the minimal polynomial of f (and thus, of smallest degree). Consequently, as the dimension of V is at most n,

$$((f-\lambda\mathrm{id})^{n-1}(u), (f-\lambda\mathrm{id})^{n-2}(u), \ldots, (f-\lambda\mathrm{id})(u), u),$$

is a basis of V and since $u, f(u), \ldots, f^{n-2}(u), f^{n-1}(u)$ span V,

$$(u, f(u), \ldots, f^{n-2}(u), f^{n-1}(u))$$

is also a basis of V.

Let us see how f acts on the basis

$$((f-\lambda\mathrm{id})^{n-1}(u), (f-\lambda\mathrm{id})^{n-2}(u), \ldots, (f-\lambda\mathrm{id})(u), u).$$

If we write $f = f - \lambda\mathrm{id} + \lambda\mathrm{id}$, as $(f-\lambda\mathrm{id})^n$ annihilates V, we get

$$f((f-\lambda\mathrm{id})^{n-1}(u)) = (f-\lambda\mathrm{id})^n(u) + \lambda(f-\lambda\mathrm{id})^{n-1}(u) = \lambda(f-\lambda\mathrm{id})^{n-1}(u)$$

and

$$f((f-\lambda\mathrm{id})^k(u)) = (f-\lambda\mathrm{id})^{k+1}(u) + \lambda(f-\lambda\mathrm{id})^k(u), \qquad 0 \leq k \leq n-2.$$

But this means precisely that the matrix of f in this basis is the Jordan block $J_n(\lambda)$. $\square$

To the best of our knowledge, a complete proof of the real Jordan form is not easily found. Horn and Johnson state such a result as Theorem 3.4.5 in Chapter 3, Section 4, in [20]. However, they leave the details of the proof that a real P can be found as an exercise. A complete proof is given in Hirsh and Smale [19]. This proof is given in Chapter 6, and relies on results from Chapter 2 and Appendix III.

We found that a proof can be obtained from Theorem 2.2. Since we believe that some of the techniques involved in this proof are of independent interest, we present this proof in full detail. It should be noted that we were inspired by some arguments found in Gantmacher [15], Chapter IX, Section 13.

Theorem 2.6. *(Real Jordan Form) For every real $n \times n$ matrix, A, there is some invertible (real) matrix, P, and some real Jordan matrix, J, so that*

$$A = PJP^{-1}.$$

For every Jordan block, $J_r(\lambda)$, of type (1), λ is some real eigenvalue of A and for every Jordan block, $J_{2r}(\lambda, \mu)$, of type (2), $\lambda + i\mu$ is a complex eigenvalue of A (with $\mu \neq 0$). Every eigenvalue of A corresponds to one of more Jordan blocks of J. Furthermore, the number, m, of Jordan blocks, the distinct Jordan block, $J_{s_i}(\alpha_i)$, and the number of times, m_i, that each Jordan block, $J_{s_i}(\alpha_i)$, occurs are uniquely determined by A.

Proof. Let $f\colon V \to V$ be the linear map defined by A and let $f_{\mathbb{C}}$ be the complexification of f. Then, Theorem 2.2 yields a direct sum decomposition of $V_{\mathbb{C}}$ of the form

$$V_{\mathbb{C}} = V_1 \oplus \cdots \oplus V_m, \tag{$*$}$$

where each V_i is a cyclic $\mathbb{C}[X]$-module (associated with $f_{\mathbb{C}}$) whose minimal polynomial is of the form $(X - \alpha_i)^{r_i}$, where α is some (possibly complex) eigenvalue of f. If W is any subspace of $V_{\mathbb{C}}$, we define the *conjugate*, $\overline{W}$, of W by

$$\overline{W} = \{u - iv \in V_{\mathbb{C}} \mid u + iv \in W\}.$$

It is clear that $\overline{W}$ is a subspace of $V_{\mathbb{C}}$ of the same dimension as W and obviously, $\overline{V_{\mathbb{C}}} = V_{\mathbb{C}}$. Our first goal is to prove the following claim:

Claim 1. For each factor, V_j, the following properties hold:

(1) If $u+iv, f_{\mathbb{C}}(u+iv), \ldots, f_{\mathbb{C}}^{r_j-1}(u+iv)$ span V_j, then $u-iv, f_{\mathbb{C}}(u-iv), \ldots, f_{\mathbb{C}}^{r_j-1}(u-iv)$ span $\overline{V}_i$ and so, $\overline{V}_i$ is cyclic with respect to $f_{\mathbb{C}}$.

(2) If $(X-\alpha_i)^{r_i}$ is the minimal polynomial of V_i, then $(X-\overline{\alpha}_i)^{r_i}$ is the minimal polynomial of $\overline{V}_i$.

Proof of Claim 1. As $f_{\mathbb{C}}(u+iv) = f(u)+if(v)$, we have $f_{\mathbb{C}}(u-iv) = f(u)-if(v)$. It follows that $f_{\mathbb{C}}^k(u+iv) = f^k(u) + if^k(v)$ and $f_{\mathbb{C}}^k(u-iv) = f^k(u) - if^k(v)$, which implies that if V_j

is generated by $u+iv, f_{\mathbb{C}}(u+iv), \ldots, f_{\mathbb{C}}^{r_j}(u+iv)$ then $\overline{V}_j$ is generated by $u-iv, f_{\mathbb{C}}(u-iv), \ldots, f_{\mathbb{C}}^{r_j}(u-iv)$. Therefore, $\overline{V}_j$ is cyclic for $f_{\mathbb{C}}$.

We also prove the following simple fact: If

$$(f_{\mathbb{C}} - (\lambda_j + i\mu_j)\mathrm{id})(u+iv) = x+iy,$$

then

$$(f_{\mathbb{C}} - (\lambda_j - i\mu_j)\mathrm{id})(u-iv) = x-iy.$$

Indeed, we have

$$\begin{aligned} x+iy &= (f_{\mathbb{C}} - (\lambda_j + i\mu_j)\mathrm{id})(u+iv) \\ &= f_{\mathbb{C}}(u+iv) - (\lambda_j + i\mu_j)(u+iv) \\ &= f(u) + if(v) - (\lambda_j + i\mu_j)(u+iv) \end{aligned}$$

and by taking conjugates, we get

$$\begin{aligned} x-iy &= f(u) - if(v) - (\lambda_j - i\mu_j)(u-iv) \\ &= f_{\mathbb{C}}(u-iv) - (\lambda_j - i\mu_j)(u-iv) \\ &= (f_{\mathbb{C}} - (\lambda_j - i\mu_j)\mathrm{id})(u-iv), \end{aligned}$$

as claimed.

From the above, $(f_{\mathbb{C}} - \alpha_j\mathrm{id})^{r_j}(x+iy) = 0$ iff $(f_{\mathbb{C}} - \overline{\alpha}_j\mathrm{id})^{r_j}(x-iy) = 0$. Thus, $(X - \overline{\alpha}_j\mathrm{id})^{r_j}$ annihilates $\overline{V}_j$ and as $\dim \overline{V}_j = \dim V_j$ and $\overline{V}_j$ is cyclic, we conclude that $(X - \overline{\alpha}_j)^{r_j}$ is the minimal polynomial of $\overline{V}_j$. $\square$

Next we prove

Claim 2. For every factor, V_j, in the direct decomposition $(*)$, we have:

(A) If $(X - \lambda_j)^{r_j}$ is the minimal polynomial of V_j, with $\lambda_j \in \mathbb{R}$, then either

(1) $V_j = \overline{V}_j$ and if $u+iv$ generates V_j, then $u-iv$ also generates V_j, or

(2) $V_j \cap \overline{V}_j = (0)$ and

- (a) the cyclic space $\overline{V}_j$ also occurs in the direct sum decomposition $(*)$
- (b) the minimal polynomial of $\overline{V}_j$ is $(X - \lambda_j)^{r_j}$
- (c) the spaces V_j and $\overline{V}_j$ contain only complex vectors (this means that if $x+iy \in V_j$, then $x \neq 0$ and $y \neq 0$ and similarly for $\overline{V}_j$).

(B) If $(X - (\lambda_j + i\mu_j))^{r_j}$ is the minimal polynomial of V_j with $\mu_j \neq 0$, then

(d) $V_j \cap \overline{V}_j = (0)$

(e) the cyclic space $\overline{V}_j$ also occurs in the direct sum decomposition $(*)$

(f) the minimal polynomial of $\overline{V}_j$ is $(X - (\lambda_j - i\mu_j))^{r_j}$

(g) the spaces V_j and $\overline{V}_j$ contain only complex vectors.

Proof of Claim 2. By taking the conjugate of the direct sum decomposition $(*)$ we get

$$V_{\mathbb{C}} = \overline{V}_1 \oplus \cdots \oplus \overline{V}_m.$$

By Claim 1, each $\overline{V}_j$ is a cyclic subspace with respect to $f_{\mathbb{C}}$ of the same dimension as V_j and the minimal polynomial of $\overline{V}_j$ is $(X - \overline{\alpha}_j)^{r_j}$ if the minimal polynomial of V_j is $(X - \alpha_j)^{r_j}$. It follows from the uniqueness assertion of Theorem 2.2 that the list of conjugate minimal polynomials

$$(X - \overline{\alpha}_1)^{r_1}, \ldots, (X - \overline{\alpha}_m)^{r_m}$$

is a permutation the list of minimal polynomials

$$(X - \alpha_1)^{r_1}, \ldots, (X - \alpha_m)^{r_m}$$

and so, every $\overline{V}_j$ is equal to some factor V_k (possibly equal to V_j if α_j is real) in the direct decomposition $(*)$, where V_k and $\overline{V}_j$ have the same minimal polynomial, $(X - \overline{\alpha}_j)^{r_j}$.

Next, assume that $(X - \lambda_j)^{r_j}$ is the minimal polynomial of V_j, with $\lambda_j \in \mathbb{R}$. Consider any generator, $u + iv$, for V_j. If $u - iv \in V_j$, then by Claim 1, $\overline{V}_j \subseteq V_j$ and so $\overline{V}_j = V_j$, as $\dim \overline{V}_j = \dim V_j$. We know that $u + iv, f_{\mathbb{C}}(u + iv), \ldots, f_{\mathbb{C}}^{r_j}(u + iv)$ generate V_j and that $u - iv, f_{\mathbb{C}}(u - iv), \ldots, f_{\mathbb{C}}^{r_j}(u - iv)$ generate $\overline{V}_j = V_j$, which implies (1).

If $u - iv \notin V_j$, then we proved earlier that $\overline{V}_j$ occurs in the direct sum $(*)$ as some V_k and that its minimal polynomial is also $(X - \lambda_j)^{r_j}$. Since $u - iv \notin V_j$ and V_j and $\overline{V}_j$ belong to a direct sum decomposition, $V_j \cap \overline{V}_j = (0)$ and 2(a) and 2(b) hold. If $u \in V_j$ or $iv \in V_j$ for some real $u \in V$ or some real $v \in V$ and $u, v \neq 0$, as V_j is a complex space, then $v \in V_j$ and either $u \in \overline{V}_j$ or $v \in \overline{V}_j$, contradicting $V_j \cap \overline{V}_j = (0)$. Thus, 2(c) holds.

Now, consider the case where $\alpha_j = \lambda_j + i\mu_j$, with $\mu_j \neq 0$. Then, we know that $\overline{V}_j = V_k$ for some V_k whose minimal polynomial is $(X - (\alpha_j - i\mu_j))^{r_j}$ in the direct sum $(*)$. As $\mu_j \neq 0$, the cyclic spaces V_j and $\overline{V}_j$ correspond to distinct minimal polynomials $(X - (\alpha_j + i\mu_j))^{r_j}$ and $(X - (\alpha_j - i\mu_j))^{r_j}$, so $V_j \cap \overline{V}_j = (0)$. It follows that V_j and $\overline{V}_j$ consist of complex vectors as we already observed. Therefore, (d), (e), (f), (g) are proved, which finishes the proof of Claim 2. $\square$

This completes the proof our theorem. $\square$

Theorem 2.8. *For any (real or complex) $n \times n$ matrix, A, if $A = PJP^{-1}$ where J is a Jordan matrix of the form*

$$J = \begin{pmatrix} J_{r_1}(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(\lambda_m) \end{pmatrix},$$

then there is some invertible matrix, Q, so that the Jordan form of e^A is given by

$$e^A = Q\, e(J)\, Q^{-1},$$

where $e(J)$ is the Jordan matrix

$$e(J) = \begin{pmatrix} J_{r_1}(e^{\lambda_1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(e^{\lambda_m}) \end{pmatrix},$$

that is, each $J_{r_k}(e^{\lambda_k})$ is obtained from $J_{r_k}(\lambda_k)$ by replacing all the diagonal entries λ_k by e^{λ_k}. Equivalently, if the list of elementary divisors of A is

$$(X - \lambda_1)^{r_1}, \ldots, (X - \lambda_m)^{r_m},$$

then the list of elementary divisors of e^A is

$$(X - e^{\lambda_1})^{r_1}, \ldots, (X - e^{\lambda_m})^{r_m}.$$

Proof. Theorem 2.8 is a consequence of a general theorem about functions of matrices proved in Gantmacher [15], see Chapter VI, Section 8, Theorem 9. Because a much more general result is proved, the proof in Gantmacher [15] is rather involved. However, it is possible to give a simpler proof exploiting special properties of the exponential map.

Let f be the linear map defined by the matrix A. The strategy of our proof is to go back to the direct sum decomposition given by Theorem 2.2,

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_m,$$

where each V_i is a cyclic $\mathbb{C}[X]$-module such that the minimal polynomial of the restriction of f to V_i is of the form $(X - \lambda_i)^{r_i}$. We will prove that

(1) The vectors

$$u, e^f(u), (e^f)^2(u), \ldots, (e^f)^{r_i - 1}(u)$$

form a basis of V_i (here, $(e^f)^k = e^f \circ \cdots \circ e^f$, the composition of e^f with itself k times).

(2) The polynomial $(X - e^{\lambda_i})^{r_i}$ is the minimal polynomial of the restriction of e^f to V_i.

First, we prove that V_i is invariant under e^f. Let $N = f - \lambda_i \mathrm{id}$. To say that $(X - \lambda_i)^{r_i}$ is the minimal polynomial of the restriction of f to V_i is equivalent to saying that N is nilpotent with index of nilpotency, $r = r_j$. Now, N and $\lambda_i \mathrm{id}$ commute so as $f = N + \lambda_i \mathrm{id}$, we have

$$e^f = e^{N + \lambda_i \mathrm{id}} = e^N e^{\lambda_i \mathrm{id}} = e^{\lambda_i} e^N.$$

Furthermore, as N is nilpotent, we have

$$e^N = \mathrm{id} + N + \frac{N^2}{2!} + \cdots + \frac{N^{r-1}}{(r-1)!},$$

so

$$e^f = e^{\lambda_i}\left(\mathrm{id} + N + \frac{N^2}{2!} + \cdots + \frac{N^{r-1}}{(r-1)!}\right).$$

Now, V_i is invariant under f so V_i is invariant under $N = f - \lambda_i\mathrm{id}$ and this implies that V_i is invariant under e^f. Thus, we can view V_i as a $\mathbb{C}[X]$-module with respect to e^f.

From the formula for e^f we get

$$\begin{aligned} e^f - e^{\lambda_i}\mathrm{id} &= e^{\lambda_i}\left(\mathrm{id} + N + \frac{N^2}{2!} + \cdots + \frac{N^{r-1}}{(r-1)!}\right) - e^{\lambda_i}\mathrm{id} \\ &= e^{\lambda_i}\left(N + \frac{N^2}{2!} + \cdots + \frac{N^{r-1}}{(r-1)!}\right). \end{aligned}$$

If we let

$$\widetilde{N} = N + \frac{N^2}{2!} + \cdots + \frac{N^{r-1}}{(r-1)!},$$

we claim that

$$\widetilde{N}^{r-1} = N^{r-1} \quad \text{and} \quad \widetilde{N}^r = 0.$$

The case $r = 1$ is trivial so we may assume $r \geq 2$. Since $\widetilde{N} = NR$ for some R such that $NR = RN$ and $N^r = 0$, the second property is clear. The first property follows by observing that $\widetilde{N} = N + N^2T$, where N and T commute, so using the binomial formula,

$$\widetilde{N}^{r-1} = \sum_{k=0}^{r-1}\binom{r-1}{k}N^k(N^2T)^{r-1-k} = \sum_{k=0}^{r-1}\binom{r-1}{k}N^{2r-k-2}T^{r-1-k} = N^{r-1},$$

since $2r - k - 2 \geq r$ for $0 \leq k \leq r-2$ and $N^r = 0$.

Recall from Proposition 2.3 that

$$((f - \lambda_i\mathrm{id})^{r_i-1}(u), \ldots, (f - \lambda_i\mathrm{id})(u), u)$$

is a basis of V_i, which implies that $N^{r-1}(u) = (f - \lambda_i\mathrm{id})^{r_i-1}(u) \neq 0$. Since $\widetilde{N}^{r-1} = N^{r-1}$, we have $\widetilde{N}^{r-1}(u) \neq 0$ and as $\widetilde{N}^r = 0$, we have $\widetilde{N}^r(u) = 0$. It is well-known that these two facts imply that

$$u, \widetilde{N}(u), \ldots, \widetilde{N}^{r-1}(u)$$

are linearly independent. Indeed, if we had a linear dependence relation

$$a_0u + a_1\widetilde{N}(u) + \cdots + a_{r-1}\widetilde{N}^{r-1}(u) = 0,$$

by applying $\widetilde{N}^{r-1}$, as $\widetilde{N}^r(u) = 0$ we get $a_0\widetilde{N}^{r-1}(u) = 0$, so, $a_0 = 0$ as $\widetilde{N}^{r-1}(u) \neq 0$; by applying $\widetilde{N}^{r-2}$ we get $a_1\widetilde{N}^{r-1}(u) = 0$, so $a_1 = 0$; using induction, by applying $\widetilde{N}^{r-k-2}$ to

$$a_{k+1}\widetilde{N}^{k+1}(u) + \cdots + a_{r-1}\widetilde{N}^{r-1}(u) = 0,$$

we get $a_{k+1} = 0$ for $k = 0, \ldots, r-2$. Since V_i has dimension r $(= r_i)$, we deduce that

$$(u, \widetilde{N}(u), \ldots, \widetilde{N}^{r-1}(u))$$

is a basis of V_i. But $e^f = e^{\lambda_i}(\mathrm{id} + \widetilde{N})$, so for $k = 0, \ldots, r-1$, each $\widetilde{N}^k(u)$ is a linear combination of the vectors $u, e^f(u), \ldots, (e^f)^{r-1}(u)$ which implies that

$$(u, e^f(u), (e^f)^2(u), \ldots, (e^f)^{r-1}(u))$$

is a basis of V_i. This implies that any annihilating polynomial of V_i has degree no less than r and since $(X - e^{\lambda_i})^r$ annihilates V_i (because $e^f - e^{\lambda_i}\mathrm{id} = e^{\lambda_i}\widetilde{N}$ and $\widetilde{N}^r = 0$), it is the minimal polynomial of V_i.

In summary, we proved that each V_i is a cyclic $\mathbb{C}[X]$-module (with respect to e^f) and that in the direct sum decomposition

$$V = V_1 \oplus \cdots \oplus V_m,$$

the polynomial $(X - e^{\lambda_i})^{r_i}$ is the minimal polynomial of V_i, which is Theorem 2.2 for e^f. Then, Theorem 2.8 follows immediately from Proposition 2.3. $\square$

Theorem 4.1. *For any (real or complex) invertible $n \times n$ matrix, A, if $A = PJP^{-1}$ where J is a Jordan matrix of the form*

$$J = \begin{pmatrix} J_{r_1}(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(\lambda_m) \end{pmatrix},$$

then there is some invertible matrix, Q, so that the Jordan form of A^2 is given by

$$e^A = Q\, s(J)\, Q^{-1},$$

where $s(J)$ is the Jordan matrix

$$s(J) = \begin{pmatrix} J_{r_1}(\lambda_1^2) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(\lambda_m^2) \end{pmatrix},$$

that is, each $J_{r_k}(\lambda_k^2)$ is obtained from $J_{r_k}(\lambda_k)$ by replacing all the diagonal enties λ_k by λ_k^2. Equivalently, if the list of elementary divisors of A is

$$(X - \lambda_1)^{r_1}, \ldots, (X - \lambda_m)^{r_m},$$

then the list of elementary divisors of A^2 is

$$(X - \lambda_1^2)^{r_1}, \ldots, (X - \lambda_m^2)^{r_m}.$$

Proof. Theorem 4.1 is a consequence of a general theorem about functions of matrices proved in Gantmacher [15], see Chapter VI, Section 8, Theorem 9. However, it is possible to give a simpler proof exploiting special properties of the squaring map.

Let f be the linear map defined by the matrix A. The proof is modeled after the proof of Theorem 2.8. Consider the direct sum decomposition given by Theorem 2.2,

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_m,$$

where each V_i is a cyclic $\mathbb{C}[X]$-module such that the minimal polynomial of the restriction of f to V_i is of the form $(X - \lambda_i)^{r_i}$. We will prove that

(1) The vectors

$$u, f^2(u), f^4(u), \ldots, f^{2(r_i-1)}(u)$$

form a basis of V_i.

(2) The polynomial $(X - \lambda_i^2)^{r_i}$ is the minimal polynomial of the restriction of f^2 to V_i.

Since V_i is invariant under f, it is clear that V_i is invariant under $f^2 = f \circ f$. Thus, we can view V_i as a $\mathbb{C}[X]$-module with respect to f^2. Let $N = f - \lambda_i \mathrm{id}$. To say that $(X - \lambda_i)^{r_i}$ is the minimal polynomial of the restriction of f to V_i is equivalent to saying that N is nilpotent with index of nilpotency, $r = r_j$. Now, N and $\lambda_i \mathrm{id}$ commute so as $f = \lambda_i \mathrm{id} + N$, we have

$$f^2 = \lambda_i^2 \mathrm{id} + 2\lambda_i N + N^2$$

and so

$$f^2 - \lambda_i^2 \mathrm{id} = 2\lambda_i N + N^2.$$

Since we are assuming that f is invertible, $\lambda_i \neq 0$, so

$$f^2 - \lambda_i^2 \mathrm{id} = 2\lambda_i \left(N + \frac{N^2}{2\lambda_i} \right).$$

If we let

$$\widetilde{N} = N + \frac{N^2}{2\lambda_i},$$

we claim that

$$\widetilde{N}^{r-1} = N^{r-1} \quad \text{and} \quad \widetilde{N}^r = 0.$$

The proof is identical to the proof given in Theorem 2.8. Again, as in the proof of Theorem 2.8, we deduce that we have $\widetilde{N}^{r-1}(u) \neq 0$ and $\widetilde{N}^r(u) = 0$, from which we infer that

$$(u, \widetilde{N}(u), \ldots, \widetilde{N}^{r-1}(u))$$

is a basis of V_i. But $f^2 - \lambda_i^2 \mathrm{id} = 2\lambda_i \widetilde{N}$, so for $k = 0, \ldots, r-1$, each $\widetilde{N}^k(u)$ is a linear combination of the vectors $u, f^2(u), \ldots, f^{2(r-1)}(0)$ which implies that

$$(u, f^2(u), f^4(u), \ldots, f^{2(r-1)}(u))$$

is a basis of V_i. This implies that any annihilating polynomial of V_i has degree no less than r and since $(X - \lambda_i^2)^r$ annihilates V_i (because $f^2 - \lambda_i^2 \mathrm{id} = 2\lambda_i \widetilde{N}$ and $\widetilde{N}^r = 0$), it is the minimal polynomial of V_i. Theorem 4.1 follows immediately from Proposition 2.3. □

References

[1] Vincent Arsigny. *Processing Data in Lie Groups: An Algebraic Approach. Application to Non-Linear Registration and Diffusion Tensor MRI.* PhD thesis, École Polytechnique, Palaiseau, France, 2006. Thèse de Sciences.

[2] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A fast and log-euclidean polyaffine framework for locally affine registration. Technical report, INRIA, 2004, route des Lucioles, 06902 Sophia Antipolis Cedex, France, 2006. Report No. 5865.

[3] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

[4] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. on Matrix Analysis and Applications*, 29(1):328–347, 2007.

[5] Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Polyrigid and polyaffine transformations: a novel geometrical tool to deal with non-rigid deformations–application to the registration of histological slices. *Medical Image Analysis*, 9(6):507–523, 2005.

[6] Michael Artin. *Algebra.* Prentice Hall, first edition, 1991.

[7] Dario A. Bini, Nicholas J. Higham, and Beatrice Meini. Algorithms for the matrix pth root. *Numerical Algorithms*, 39:349–378, 2005.

[8] Nicolas Bourbaki. *Algèbre, Chapitres 4-7.* Eléments de Mathématiques. Masson, 1981.

[9] Nicolas Bourbaki. *Elements of Mathematics. Lie Groups and Lie Algebras, Chapters 1–3.* Springer, first edition, 1989.

[10] Henri Cartan. *Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes*. Hermann, 1961.

[11] Sheung H. Cheng, Nicholas J. Higham, Charles Kenney, and Alan J. Laub. Approximating the logarithm of a matrix to specified accuracy. *SIAM Journal on Matrix Analysis and Applications*, 22:1112–1125, 2001.

[12] Walter J. Culver. On the existence and uniqueness of the real logarithm of a matrix. *Proc. Amer. Math. Soc.*, 17:1146–1151, 1966.

[13] C. R. DePrima and C. R. Johnson. The range of $A^{-1}A^*$ in $\mathbf{GL}(n, \mathbf{C})$. *Linear Algebra and Its Applications*, 9:209–222, 1974.

[14] David S. Dummit and Richard M. Foote. *Abstract Algebra*. Wiley, second edition, 1999.

[15] F.R. Gantmacher. *The Theory of Matrices, Vol. I*. AMS Chelsea, first edition, 1977.

[16] Roger Godement. *Cours d'Algèbre*. Hermann, first edition, 1963.

[17] Nicholas J. Higham. Computing real square roots of a real matrix. *Linear Algebra and its Applications*, 88/89:405–430, 1987.

[18] Nicholas J. Higham. *Functions of Matrices. Theory and Computation*. SIAM, first edition, 2008.

[19] Morris W. Hirsh and Stephen Smale. *Differential Equations, Dynamical Systems and Linear Algebra*. Academic Press, first edition, 1974.

[20] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, first edition, 1990.

[21] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, first edition, 1994.

[22] Hoffman Kenneth and Kunze Ray. *Linear Algebra*. Prentice Hall, second edition, 1971.

[23] Charles S. Kenney and Alan J. Laub. Condition estimates for matrix functions. *SIAM Journal on Matrix Analysis and Applications*, 10:191–209, 1989.

[24] Serge Lang. *Algebra*. Addison Wesley, third edition, 1993.

[25] Serge Lang. *Complex Analysis*. GTM No. 103. Springer Verlag, fourth edition, 1999.

[26] R. Mneimné and F. Testard. *Introduction à la Théorie des Groupes de Lie Classiques*. Hermann, first edition, 1997.

[27] Denis Serre. *Matrices, Theory and Applications*. GTM No. 216. Springer Verlag, second edition, 2010.

[28] Gilbert Strang. *Linear Algebra and its Applications*. Saunders HBJ, third edition, 1988.