Spectral Factorization of Matrices

by

FRANS OTTO GAOSEB

submitted in accordance with the requirements
for the degree of

MASTER OF SCIENCE

in the subject

Mathematics

at the

UNIVERSITY OF SOUTH AFRICA

Supervisor: Professor J.D. Botha

June 2020

0.1 Abstract and Statement of the Research Problem

The research will analyze and compare the current research on the spectral factorization of non-singular and singular matrices. We show that a non-singular non-scalar matrix A can be written as a product $A = BC$ where the eigenvalues of B and C are arbitrarily prescribed subject to the condition that the product of the eigenvalues of B and C must be equal to the determinant of A. Further, B and C can be simultaneously triangularised as a lower and upper triangular matrix respectively. Singular matrices will be factorized in terms of nilpotent matrices and otherwise over an arbitrary or complex field in order to present an integrated and detailed report on the current state of research in this area.

Applications related to unipotent, positive-definite, commutator, involutory and Hermitian factorization are studied for non-singular matrices, while applications related to positive-semidefinite matrices are investigated for singular matrices.

We will consider the theorems found in Sourour [24] and Laffey [17] to show that a non-singular non-scalar matrix can be factorized spectrally. The same two articles will be used to show applications to unipotent, positive-definite and commutator factorization. Applications related to Hermitian factorization will be considered in [26]. Laffey [18] shows that a non-singular matrix A with $\det A = \pm 1$ is a product of four involutions with certain conditions on the arbitrary field. To aid with this conclusion a thorough study is made of Hoffman [13], who shows that an invertible linear transformation T of a finite dimensional vector space over a field is a product of two involutions if and only if T is similar to T^{-1}. Sourour shows in [24] that if A is an $n \times n$ matrix over an arbitrary field containing at least $n + 2$ elements and if $\det A = \pm 1$, then A is the product of at most four involutions.

We will review the work of Wu [29] and show that a singular matrix A of order $n \geq 2$ over the complex field can be expressed as a product of two nilpotent matrices, where the rank of each of the factors is the same as A, except when A is a 2×2 nilpotent matrix of rank one.

Nilpotent factorization of singular matrices over an arbitrary field will also be investigated. Laffey [17] shows that the result of Wu, which he established over the complex field, is also valid over an arbitrary field by making use of a special matrix factorization involving similarity to an LU factorization. His proof is based on an application of Fitting's Lemma to express, up to similarity, a singular matrix as a direct sum of a non-singular and nilpotent

matrix, and then to write the non-singular component as a product of a lower and upper triangular matrix using a matrix factorization theorem of Sourour [24].

The main theorem by Sourour and Tang [26] will be investigated to highlight the necessary and sufficient conditions for a singular matrix to be written as a product of two matrices with prescribed eigenvalues. This result is used to prove applications related to positive-semidefinite matrices for singular matrices.

Keywords: Spectral factorization; Matrix factorization; Singular Matrices; Non-singular matrices; Involutions; Commutators; Unipotent matrices; Positive-definite matrices; Hermitian factorization; Scalar matrices; Nilpotent factorization

This work is based on research supported in part by the National Research Foundation of South Africa (Grant Number 103932).

0.2 Declaration

Student number: 32974302

I declare that Spectral Factorization of Matrices is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

I further declare that I submitted the dissertation to originality checking software and that it falls within the accepted requirements for originality.

I further declare that I have not previously submitted this work, or part of it, for examination at Unisa for another qualification or at any higher education institution.

________________ 11 July 2020

Signature Date

(Mr Frans Gaoseb)

0.3 Acknowledgement

This Dissertation can be compared to a journey taken through the wilderness. It was not always easy to persevere, but with the assistance and guidance of my supervisor Professor Johan Botha and the support of my family I was able to succeed. I would like to say thank you to Professor Botha for all the communications we had over the years that was mostly via email. Even though we only met twice (2 days in total), the guidelines given were very clear and aided me to master the Dissertation topic.

My family was fabulous in encouraging me to complete this Dissertation. They sacrificed a lot in order for me to complete this Dissertation. I appreciated those nice discussions on mathematics with my wife and children on braai (barbeque) nights even though they were not always keen to do so.

Lastly I would like to give a shout out to my hometown Swakopmund in Namibia and to the many schools in the Erongo Region where I conducted mathematics tutorials. Interactions with these communities encouraged me to complete my studies.

Contents

Chapter 1

Introduction and Motivation for the Research

1.1 Research Objectives

Through this research the following research objectives should be met:

* Present spectral factorization of invertible non-scalar matrices ([24] and [17]) in order to place the current investigation concerning factorization in a broader context.
* Present a unified and coherent treatment of the factorization of singular matrices as contained in the works by Wu [29], Laffey [17] and Sourour [26].
* Investigate applications of spectral factorization to invertible matrices i.e. unipotent, positive-definite, commutator, involutory and Hermitian factorization as found in [24], [17], [18] and [26].
* Investigate the conditions under which an invertible matrix can be expressed as a product of two involutions as found in [13].
* Investigate applications of spectral factorization on singular matrices i.e. positive-semidefinite factorization as found in [26].

1.2 Notation

F	Arbitrary Field
F^n	n dimensional vector space over a field F
$\det A$	Determinant of the matrix A
$\operatorname{rank} A$	Rank of the matrix A
$\mathrm{N}(A)$	Nullspace of the matrix A
$\operatorname{diag}(A)$	Diagonal entries of a matrix
$\operatorname{Diag}(A)$	Matrix with the diagonal entries of matrix A and zero elsewhere
$\operatorname{trace}(A)$	Trace of A
$\operatorname{null}(A)$	Nullity of A
$GL(n, F)$	Set of $n \times n$ invertible matrices
$SL(n, F)$	Set of $n \times n$ matrices with determinant 1
$M_n(F)$	$n \times n$ matrices with entries from F
$M_{n\times m}(F)$	$n \times m$ matrices with entries from F
F^n	All $n \times 1$ matrices over F
$C(f_i(x))$	Companion matrix of $f_i(x)$
$\operatorname{Span}\{x\}$	Subspace spanned by x
$\operatorname{char}(F)$	Characteristic of F
$J_n(\lambda)$	$n \times n$ Jordan canonical matrix with λ on the diagonal, one on the subdiagonal and zero elsewhere
A^T	Transpose of matrix A
A^*	Conjugate transpose of matrix A
$\operatorname{Eig}(A)$	Set of eigenvalues of matrix A
Left-null(A)	Set of vectors x such that $x^T A = 0^T$
$\oplus$	Direct sum of matrices
GCD	Greatest common divisor of two or more numbers

1.3 Significance of the Research

The significance of the research is to contribute to a better understanding of the current body of research on factorization of matrices including nilpotent factorization by presenting the main results in a unified and coherent way. Today, matrix factorization is an important and useful tool in many areas of scientific research, including scientific computing. The first matrix factorization result was proved by G. Frobenius [8] in 1910 which stated that a square matrix over an arbitrary field can be expressed as a product of two symmetric matrices.

From the algebraic origins of matrix theory, and matrix factorization in particular, the advent of computers had a significant impact on the further development of matrix factorization because of the needs of scientific computing. The person who is regarded as the father of computer science, Alan Turing, made significant contributions to the design of modern day computers and the factorization of matrices as required by computers. He was highly influential in the development of computer science, giving a formalization of concepts of algorithm and computation with the Turing machine, which is considered a model of a general purpose computer. The Turing machine was built on the principle that it should be able to compute anything that is computable i.e. be programmable. A machine that is programmable is a central concept to this day of the modern computer. Turing worked on the design of the Automatic Computing Engine (ACE) from 1945 to 1947 and in 1946 presented a paper which was the first detailed design of a stored program computer. Turing invented the LU Decomposition method in 1948, which is used today for solving matrix equations [32].

Matrix factorizations like the LU factorization, the QR factorization, and the Singular Value Decomposition (also known as the Eckart-Young factorization) today form part of many scientific subroutine libraries like the IMSL. Matrix decomposition is essentially needed for two reasons, namely for computational convenience and for analytical simplicity. Data matrices representing some numerical observations such as the proximity matrix or the correlation matrix are often huge to analyze and therefore to decompose the data matrices into some lower order or lower rank canonical forms will reveal the inherent characteristics and structure of the matrices and help to interpret their meaning readily. Matrix decomposition methods simplify computations and preserve certain properties such as the determinant, rank or inverse, so that these quantities can be calculated after applying the matrix transformation. L. Hubert, J. Meulman and W. Heiser [14] highlight two purposes of matrix factorization, namely in Numerical Linear Algebra and Applied Statistics/Psychometrics.

Numerical Linear Algebra Given $Ax = b$ for $A \in M_n(\mathrm{R})$ (the set of all $n \times n$ matrices over real numbers) and x, $b \in \mathrm{R}^n$, A is factorized into LU where L is a lower triangular matrix and U is an upper triangular matrix (a square matrix A has an LU Decomposition if it can be reduced to row echelon form without using any row interchanges). The equation $Ax = b$ is replaced by two systems. Find y so that $Ly = b$ and also x so that $Ux = y$. Expressing A as LU is computationally convenient for obtaining a solution to $Ax = b$.

Applied Statistics/Psychometrics A matrix $A \in M_{n\times p}(\mathrm{R})$ represents a data matrix, containing numerical observations on n objects (subjects) over p attributes (variables) and a matrix $B \in M_p(\mathrm{R})$ measures the proximity between attributes i.e. the correlation between columns of A. The purpose of matrix factorization would be to obtain some lower-rank approximation to A or B so as to better understand the relationship within objects and within attributes and how objects relate to attributes.

1.4 Methodology

The following information resources were used to gather information on articles published on factorization of nilpotent matrices.

* The UNISA library catalogue
* E-journals
* MathSciNet
* arXiv (http://arxiv.org)
* Wikipedia
* JSTOR
* Cambridge Journal Online
* ScienceDirect
* Taylor and Francis Online
* SIAM Online Journal Archive 1952 - 1996

Research strategies that were employed during the dissertation include the following:

* Complete a background knowledge course on Abstract Algebra based on Fraleigh [7].

* Analyze all the proofs on products of nilpotent matrices over the complex and arbitrary fields in order to present an integrated and coherent treatment of this research (This refers to the works of P. Y. Wu [29] and T. J. Laffey [17] as contained in the Core Literature under the Literature Review section)
* Investigate if the proof by P.Y. Wu [29] on products of nilpotent matrices over the complex field can be extended to an arbitrary field.
* Study the existing literature ([24], [17] and [26]) on spectral factorization of matrices (singular and non-singular) in order to gain a better understanding of the context of the current research.
* Consult additional sources in order to gain a proper understanding of the existing body of literature related to the research topic, including [5], [7], [9], [10], [21] and [27].

1.5 Literature Review

The literature review has been divided into 5 major categories which include articles on the history of matrix factorization, the Jordan canonical form, spectral factorization, nilpotent factorization, and related research.

History and Significance of Matrix Factorization
To gain some insight into the history of linear algebra and the practical applications of the factorization of matrices the following sources were consulted: [14], [28], [30], [32] and [33].

The Jordan Canonical Form
To come to an accurate understanding of the Jordan canonical form of a matrix the following articles were reviewed: [5], [9], [10], [21] and [27].

Spectral Factorization of Matrices
In order to comprehend the necessary and sufficient conditions for a matrix to be written as a product of two matrices, nilpotent or otherwise with prescribed eigenvalues the following articles will be studied: [24], [17] and [26].

Nilpotent Factorization
The core literature found on the factorization of a singular matrix into nilpotent matrices is listed below along with the main ideas contained in the research:

Literature	Main Relevant Idea
P. Y. Wu, Products of nilpotent matrices [29]	Shows that every $n \times n$ singular matrix over a complex field, with the exception of a 2×2 nonzero nilpotent can be expressed as a product of two nilpotent matrices.
T. J. Laffey, Products of matrices, in Generators and Relations in Groups and Geometries [17]	Shows that a singular $n \times n$ matrix over any arbitrary field can be expressed as a product of two nilpotent matrices except when $n = 2$.

Table 1.1: Core Literature on Nilpotent Factorization of Matrices

1.6 Related Research

Although there are so many interesting topics on the factorization of matrices it is not possible to cover all the topics in the dissertation. This section highlights the points that the research will not tackle and they are as follows:

* Characterization of operators on a Hilbert space that are products of k commuting square zero operators for $k \geq 2$ [4].
* Characterization of operators on an infinite dimensional vector space that are products of two commuting nilpotent operators [4].
* The conditions under which an $n \times n$ matrix can be expressed as a product of two or more square zero matrices in the complex, separable infinite dimensional Hilbert spaces or Calkin Algebras [20].

Nilpotent factorization of matrices over division rings or skew fields will also not be investigated. Some of the research in this area is briefly discussed below.

* N. Jizhu and Z. Yongzheng [15] shows that an $n \times n$ singular matrix T over an arbitrary skew field is a product of two nilpotent matrices B and C with $\operatorname{rank} B = \operatorname{rank} C = \operatorname{rank} T$, except when T is a 2×2 nilpotent matrix of rank 1.
* Similar to the characterisations in [4], A. Mohammadian characterised $n \times n$ singular matrices that can be expressed as a product of two commuting square zero matrices in [19]. The only difference is that Mohammadian considers matrices over a division ring instead of a field. The conclusion of his research is that a singular matrix should be square zero and have a rank of less than or equal to $\frac{n}{4}$ in order for it to be written as a product of two commuting square zero matrices.

Chapter 2

Spectral Factorization of Non-Singular Matrices

In this chapter we discuss spectral factorization of matrices. By this we mean the factorization of a non-singular non-scalar matrix A in the form BC where the eigenvalues of B and C are arbitrary prescribed, subject to the condition that the product of the eigenvalues of B and C must be equal to the determinant of A. This chapter is divided into five sections. Section 2.1 is a detailed presentation of Sourour's Spectral Factorization Theorem [24] and Section 2.2 is a detailed presentation of Laffey's proof which combines portions appearing in two of his papers, namely [17] and [18]. We will focus on [17] and provide additional details omitted in [18]. Section 2.3 will discuss the applications to unipotent, positive-definite and commutators as given firstly in Sourour's presentation and then in Laffey's. Section 2.4 and 2.5 considers products of two involutions and applications to involutory factorizations respectively.

2.1 A.R. Sourour, A Factorization Theorem for Matrices

Suppose $A \in M_n(F)$ and suppose we consider the factorization of A as a product of two matrices such that $A = BC$, where B, $C \in M_n(F)$. Is there a relationship between the eigenvalues of A and the eigenvalues that B and C have? To put it differently. Given A, how much freedom is there in the eigenvalues that B and C can have? In this section we will only focus on invertible matrices A. Obviously no eigenvalue of B and C can be zero, otherwise $\det A = 0$ and A is not invertible, which is a contradiction. All of the above can be stated formally in a theorem, but before we do so it would be appropriate to first consider an important remark as well as a lemma.

Remark: The proof of Sourour is based on two crucial observations. First if A is a non-scalar, invertible matrix, then A is similar to a matrix of the form:

$$A_1 = \begin{pmatrix} \beta_1\gamma_1 & y^T \\ x & R \end{pmatrix} \tag{2.1}$$

where x is nonzero. The proof of this is based on an idea that appeared in [6], namely if A is a non-scalar matrix then there exists a nonzero vector $v \in F^n$ such that v and Av are linearly independent. If each nonzero vector is an eigenvector of A then A would be a scalar matrix as shown in Lemma 2.1.

Lemma 2.1

If each nonzero vector is an eigenvector of matrix A, then A is a scalar matrix.

Proof. Suppose every nonzero vector is an eigenvector of A. We wish to show that all eigenvalues of A are equal and hence A is scalar. Suppose there exists eigenvectors v_1 and v_2 corresponding to distinct eigenvalues λ_1 and λ_2 respectively. By assumption $v_1 + v_2$ is also an eigenvector corresponding to λ_3, say. We find that $Av_1 = \lambda_1 v_1$, $Av_2 = \lambda_2 v_2$ and

$$A(v_1 + v_2) = \lambda_3(v_1 + v_2) = \lambda_3 v_1 + \lambda_3 v_2.$$

Also

$$A(v_1 + v_2) = Av_1 + Av_2 = \lambda_1 v_1 + \lambda_2 v_2.$$

We notice that $\lambda_1 = \lambda_2 = \lambda_3$, which is a contradiction since λ_1 and λ_2 are distinct.

We conclude that if A is non-scalar then not every nonzero vector is an eigenvector of A, and we have the situation whereby $Av \neq \lambda v$, i.e. v and Av are linearly independent. □

Theorem 2.2 [24, The Main Theorem]

Let A be a non-scalar invertible $n \times n$ matrix over a field F and let β_j and γ_j $(1 \leq j \leq n)$ be elements of F such that

$$\prod_{j=1}^{n} \beta_j \gamma_j = \det A$$

There exist $n \times n$ matrices B and C with eigenvalues $\beta_1, \beta_2, \ldots, \beta_n$ and $\gamma_1, \gamma_2, \ldots, \gamma_n$ respectively such that $A = BC$. Furthermore B and C can be chosen so that B is lower triangularizable and C is simultaneously upper triangularizable.

Proof. Suppose we want to investigate whether B can have the eigenvalues $\beta_1, \beta_2, \ldots, \beta_n$ in F and C can have the eigenvalues $\gamma_1, \gamma_2, \ldots, \gamma_n$ in F. An obvious requirement would be that

$$\det A = \beta_1 \cdots \beta_n \gamma_1 \cdots \gamma_n, \tag{2.2}$$

since $A = BC$ implies that $\det A = \det BC = \det B \cdot \det C$.

In general the determinant of a matrix is equal to the product of its eigenvalues and therefore $\det B = \beta_1 \cdots \beta_n$ and $\det C = \gamma_1 \cdots \gamma_n$ and the result for $\det A$ follows. Is (2.2) the only requirement that the eigenvalues must satisfy? Suppose $A = I_n$. In this case $\det A = \beta_1 \cdots \beta_n \cdot \gamma_1 \cdots \gamma_n = 1$, $BC = I$, $B = C^{-1}$ and thus $\{\beta_1, \ldots, \beta_n\} = \{\gamma_1^{-1}, \ldots, \gamma_n^{-1}\}$. If $A = cI$ we find that

$$cI = BC \Rightarrow I = c^{-1}BC \Rightarrow c^{-1}B = C^{-1}$$

which implies that

$$\{c^{-1}\beta_1, \ldots, c^{-1}\beta_n\} = \{\gamma_1^{-1}, \ldots, \gamma_n^{-1}\}$$

and the eigenvalues cannot be chosen independently subject only to (2.2).

Now suppose A is non-scalar and invertible. Sourour proved the amazing result that in this case the β_i and γ_i can be completely arbitrary subject only to condition (2.2). The proof is not complicated and the theorem has some useful applications, simplifying some factorization theorems that used to be difficult to prove, like Ballantine's theorem on Products of positive-definite matrices.

To show that $A \approx A_1$ (2.1) we consider the following:
Let $\tilde{A}$ be the linear transformation on F^n given by $\tilde{A}(x) = Ax$. Let e_1 be a nonzero vector that is not an eigenvector of $A - \beta_1\gamma_1 I$ and $e_2 = (A - \beta_1\gamma_1 I)e_1$.

We choose an ordered basis D of F^n whose first two members are e_1 and e_2 (Such a basis exists since F^n is finite dimensional). Let the columns of A_1 be the coordinates of $\tilde{A}(e_1), \ldots, \tilde{A}(e_n)$ with respect to $e_1, e_2, \ldots, e_n$.

We find that

$$\tilde{A}(e_1) = Ae_1 = \beta_1\gamma_1 e_1 + e_2 + 0 \cdot e_3 + \ldots + 0 \cdot e_n$$

and therefore the first column of A_1 is $(\beta_1\gamma_1, 1, 0, \ldots, 0)^t$. It follows that $A \approx A_1$, where

$$A_1 = \begin{pmatrix} \beta_1\gamma_1 & y^T \\ x & R \end{pmatrix},$$

x is a nonzero column vector, y^T a row vector and $R \in M_{n-1}(F)$.

In the case $n = 1$, $A = [a]$, whereby $a \in F$. A is a scalar matrix and the Theorem does not apply in this case.

In the case $n = 2$ we have that x, y^T and $R = r$ are elements of F. Using the fact that $\det A_1 = \beta_1\beta_2\gamma_1\gamma_2$ and also $\det A_1 = \beta_1\gamma_1 r - xy^T$, we find that

$$\beta_1\gamma_1 r - xy^T = \beta_1\beta_2\gamma_1\gamma_2 \Rightarrow \beta_1\gamma_1 r = \beta_1\beta_2\gamma_1\gamma_2 + xy^T \Rightarrow r = \beta_2\gamma_2 + \gamma_1^{-1}\beta_1^{-1}xy^T,$$

and therefore

$$A_1 = \begin{pmatrix} \beta_1\gamma_1 & y^T \\ x & r \end{pmatrix} = \begin{pmatrix} \beta_1\gamma_1 & y^T \\ x & \beta_2\gamma_2 + \gamma_1^{-1}\beta_1^{-1}xy^T \end{pmatrix}.$$

Furthermore A_1 can be decomposed into two factors B and C with determinants $\beta_1\beta_2$ and $\gamma_1\gamma_2$ respectively such that

$$\begin{aligned} A_1 &= \begin{pmatrix} \beta_1\gamma_1 & y^T \\ x & r \end{pmatrix} = \begin{pmatrix} \beta_1\gamma_1 & y^T \\ x & \beta_2\gamma_2 + \gamma_1^{-1}\beta_1^{-1}xy^T \end{pmatrix} \\ &= \begin{pmatrix} \beta_1 & 0 \\ \gamma_1^{-1}x & \beta_2 \end{pmatrix} \begin{pmatrix} \gamma_1 & \beta_1^{-1}y^T \\ 0 & \gamma_2 \end{pmatrix} \\ &= BC. \end{aligned}$$

This proves the conclusion of the theorem for $n = 2$.

We now assume that $n \geq 3$. A second step is needed to make the induction step work, since if $R - \beta_1^{-1}\gamma_1^{-1}xy^T$ is a scalar matrix, then the induction step cannot be done (the motivation for using the expression $R - \beta_1^{-1}\gamma_1^{-1}xy^T$ is more clearly shown in (2.3)). The induction step is only true for matrices that are invertible, non-scalar and of size less than n.

If $R - \beta_1^{-1}\gamma_1^{-1}xy^T$ is found to be a scalar matrix, a simple transformation is made on A_1 which adds a rank one matrix to R to yield a non-scalar matrix S and leaves $\beta_1\gamma_1$ and x fixed. When A_1 is transformed in this way it yields a matrix A_2 which is similar to A_1, whereby $A_2 = P^{-1}A_1P$ and

$$P = \begin{pmatrix} 1 & w^T \\ 0 & I \end{pmatrix}.$$

Since $\operatorname{rank} A > 2$, the linear span of the columns of R is not contained in the linear span of $\{x\}$. If $\operatorname{Span}\{R\} \subseteq \operatorname{Span}\{x\}$ then $\operatorname{rank} R \leq 1$ since the rank of $\{x\}$ is one. It follows that $\operatorname{rank} A_1$ is at most 2, since column operations on A_1 in which multiples of x is subtracted from the columns of R can be performed to reduce R to the zero matrix. This yields a matrix equivalent to A_1 with nonzero elements only in the first row and column. Thus the rank of A_1, and therefore also A, is at most two which contradicts the fact that A is of rank at least three.

The existence of w^T follows from the fact that $\dim(\text{Left-null}(A)) + \operatorname{rank} A = m$ for any $m \times n$ matrix A. The left null space of matrix A is the same as the kernel of A^T and it follows that:

$$\begin{aligned}[x] &: \dim(\text{Left-null}[x]) + \operatorname{rank}[x] = n-1 \\ [x \quad R] &: \dim(\text{Left-null}[x \quad R]) + \operatorname{rank}[x \quad R] = n-1\end{aligned}$$

Since $\operatorname{rank}[x] < \operatorname{rank}[x \quad R]$ we deduce that

$$\dim(\text{Left-null}[x \quad R]) < \dim(\text{Left-null}[x])$$

and thus there exists a vector w which is in Left-null$[x]$, but which is not in Left-null$[x \quad R]$ and thus $w^T x = 0$ but $w^T R \neq 0$.

Considering A_2 we see that

$$A_2 = \begin{pmatrix} \beta_1\gamma_1 & z^T \\ x & S \end{pmatrix}$$

where $z^T = \beta_1\gamma_1 w^T + (y^T - w^T R)$ and $S = xw^T + R$ and so

$$\begin{aligned} S - \beta_1^{-1}\gamma_1^{-1}xz^T &= xw^T + R - \beta_1^{-1}\gamma_1^{-1}x[\beta_1\gamma_1 w^T + (y^T - w^T R)] \\ &= xw^T + R - xw^T - \beta_1^{-1}\gamma_1^{-1}xy^T + \beta_1^{-1}\gamma_1^{-1}xw^T R \qquad (2.3) \\ &= (R - \beta_1^{-1}\gamma_1^{-1}xy^T) + \beta_1^{-1}\gamma_1^{-1}xw^T R. \end{aligned}$$

So even though $R - \beta_1^{-1}\gamma_1^{-1}xy^T = \alpha I$, i.e. scalar, $S - \beta_1^{-1}\gamma_1^{-1}xz^T \neq \alpha I$ i.e. non-scalar due to the way in which it is defined in (2.3) ($x \neq 0$, $w^T R \neq 0$ and the matrix $\beta_1^{-1}\gamma_1^{-1}xw^T R$ is of rank one).

We apply the induction hypothesis to the $(n-1)\times(n-1)$ non-scalar matrix $S - \beta_1^{-1}\gamma_1^{-1}xz^T$ and $S - \beta_1^{-1}\gamma_1^{-1}xz^T = B_0C_0$ where B_0, $C_0 \in M_{n-1}(F)$.

Since

$$A_2 = \begin{pmatrix} \beta_1\gamma_1 & z^T \\ x & S \end{pmatrix} = \begin{pmatrix} \beta_1\gamma_1 & 0 \\ x & S - \beta_1^{-1}\gamma_1^{-1}xz^T \end{pmatrix} \begin{pmatrix} 1 & \beta_1^{-1}\gamma_1^{-1}z^T \\ 0 & I \end{pmatrix}$$

we have that

$$\det A_2 = \beta_1\gamma_1 \det(S - \beta_1^{-1}\gamma_1^{-1}xz^T) \cdot \det I,$$

hence

$$\begin{aligned}\det(S - \beta_1^{-1}\gamma_1^{-1}xz^T) &= \beta_1^{-1}\gamma_1^{-1}\det A_2 \\ &= \beta_1^{-1}\gamma_1^{-1}\beta_1\gamma_1 \cdots \beta_n\gamma_n = \beta_2\gamma_2 \cdots \beta_n\gamma_n \\ &= \prod_{j=2}^{n} \beta_j\gamma_j.\end{aligned}$$

From the induction hypothesis $\beta_2, \ldots, \beta_n$ and $\gamma_2, \ldots, \gamma_n$ are the eigenvalues of B_0 and C_0 and it follows that

$$A_2 = \begin{pmatrix} \beta_1 & 0 \\ \gamma_1^{-1}x & B_0 \end{pmatrix} \begin{pmatrix} \gamma_1 & \beta_1^{-1}z^T \\ 0 & C_0 \end{pmatrix}.$$

By the induction hypothesis B_0 and C_0 are lower and upper triangularizable and there exists an invertible matrix $Q_0 \in M_{n-1}(F)$ such that $B_1 = Q_0^{-1}B_0Q_0$ is lower triangular and $C_1 = Q_0^{-1}C_0Q_0$ is upper triangular. If

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & Q_0 \end{pmatrix}$$

then we are able to find a matrix A_3 which is similar to A_2 and thus to A, such that

$$A_3 = Q^{-1}A_2Q = \begin{pmatrix} \beta_1 & 0 \\ \xi & B_1 \end{pmatrix} \begin{pmatrix} \gamma_1 & \eta^T \\ 0 & C_1 \end{pmatrix}$$

where ξ is a column vector and η^T is a row vector. This concludes the proof. □

Theorem 2.3 - Additive version

If A is non-scalar it can be expressed as a sum $A = B + C$ where B is lower triangularizable and C is upper triangularizable with diagonals $(\beta_1, \ldots, \beta_n)$ and $(\gamma_1, \ldots, \gamma_n)$ respectively. The eigenvalues $\beta_1, \ldots, \beta_n$ of B and $\gamma_1, \ldots, \gamma_n$ of C are arbitrary, provided $\beta_1 + \cdots + \beta_n + \gamma_1 + \cdots + \gamma_n = \text{trace}(A)$.

Proof. Choose $\lambda_i = \beta_i + \gamma_i$ for $i = 1, \ldots, n$. By Fillmore [6, Theorem 2], there exists an invertible matrix P such that $A = P^{-1}EP$ where $\text{diag}(E) = (\lambda_1, \ldots, \lambda_n)$.

Let

$$E = \begin{pmatrix} \beta_1 & & & \\ & \beta_2 & & \\ & & \ddots & \\ & * & & \beta_n \end{pmatrix} + \begin{pmatrix} \gamma_1 & & * & \\ & \gamma_2 & & \\ & & \ddots & \\ & & & \gamma_n \end{pmatrix}$$

$$= \begin{pmatrix} \beta_1+\gamma_1 & & * & \\ & \beta_2+\gamma_2 & & \\ & & \ddots & \\ & * & & \beta_n+\gamma_n \end{pmatrix}.$$

It follows that

$$A = P^{-1}EP = P^{-1}\left\{\begin{pmatrix} \beta_1 & & & \\ & \beta_2 & & \\ & & \ddots & \\ & * & & \beta_n \end{pmatrix} + \begin{pmatrix} \gamma_1 & & * & \\ & \gamma_2 & & \\ & & \ddots & \\ & & & \gamma_n \end{pmatrix}\right\}P$$

$$= \mathrm{P}^{-1}\begin{pmatrix} \beta_1 & & & \\ & \beta_2 & & \\ & & \ddots & \\ & * & & \beta_n \end{pmatrix}P + P^{-1}\begin{pmatrix} \gamma_1 & & * & \\ & \gamma_2 & & \\ & & \ddots & \\ & & & \gamma_n \end{pmatrix}P$$

$$= \mathrm{B} + \mathrm{C}$$

where B is lower triangularizable and C is simultaneously upper tiangularizable with diagonals $(\beta_1, \ldots, \beta_n)$ and $(\gamma_1, \ldots, \gamma_n)$ respectively (When P is an invertible diagonal matrix this condition will be satisfied). Furthermore since $A \approx E$,

$$\mathrm{trace}(A) = \mathrm{trace}(E) = \sum_{i=1}^{n} \beta_i + \sum_{i=1}^{n} \gamma_i = \sum_{i=1}^{n} \lambda_i.$$

□

2.2 Spectral Factorization of Invertible Matrices - Laffey's Treatment

This section gives a complete and detailed presentation of Laffey's proof for spectral factorization of invertible matrices which combines portions appearing in two separate articles namely [17] and [18]. Laffey wishes to show that a non-scalar invertible matrix A is similar over F to a product LU where L is lower triangular and U is upper triangular, subject to

$\det A = x_1 \cdot x_2 \cdots x_n \cdot y_1 \cdots y_n$ with $x_1, x_2, \ldots, x_n, y_1, \ldots, y_n \in F$, $\text{diag}(L) = (x_1, \ldots, x_n)$ and $\text{diag}(U) = (y_1, \ldots, y_n)$.

If A is scalar then

$$A = cI \Rightarrow cI = LU \Rightarrow c^{-1}L = U^{-1}$$

which implies that

$$\{c^{-1}x_1, \ldots, c^{-1}x_n\} = \{y_1^{-1}, \ldots, y_n^{-1}\}.$$

The elements $x_1, \ldots, x_n, y_1, \ldots, y_n$ of F can be chosen independently subject only to $\det A = x_1 \cdots x_n \cdot y_1 \cdots y_n$.

Before considering the proof of Laffey's Theorem, we present the following result on a non-scalar invertible matrix.

Lemma 2.4

Let $A \in GL(n, F)$, $n \geq 3$ be a non-scalar matrix. Then A is similar to a matrix of the form $A_1 = \begin{pmatrix} z_1 & x^T \\ y & B_1 \end{pmatrix}$ for any given nonzero $z_1 \in F$, where x, $y \in F^{n-1}$ and $B_1 \in M_{n-1}(F)$ is non-scalar.

Proof. We use an adaptation of [18, Lemma 5.4], where the $(1,1)$ position is changed from 1 to z_1. To show that $A \approx A_1$ we will look at 3 cases, namely if the minimal polynomial is of degree at least 3, quadratic and of degree 1.

(I) The minimal polynomial is of degree at least 3. We may choose $v \in F^n$ such that v, Av, A^2v are linearly independent. To demonstrate that such a v exists, note that since the minimal polynomial of A is of degree at least 3 there exists a companion matrix C in the rational canonical form of A of order at least 3; say C is represented with respect to the linearly independent vectors $u_1, u_2, \ldots, u_k$ where $k \geq 3$. Then $Au_1 = u_2$ and $A^2u_1 = Au_2 = u_3$ and therefore u_1, $u_2 = Au_1$, $u_3 =$A$^2u_1$ are linearly independent as required.

Having now established the existence of $v \in F^n$, we proceed by defining $v_1 = v$, $v_2 = Av - z_1v_1$, $v_3 = A^2v$ which are independent and extend these vectors to a basis $v_1, v_2, \ldots, v_n$ of F^n. The columns of A_1 are the coordinates of $Av_1, Av_2, \ldots, Av_n$ with respect to $v_1, v_2, \ldots, v_n$ It follows that $Av_1 = z_1v_1 + 1 \cdot v_2 + 0 \cdot v_3 + \ldots + 0 \cdot v_n$ and thus the first column of A_1 is $(z_1, 1, 0, \ldots, 0)^T$. Also

$$\begin{aligned} Av_2 &= A(Av - z_1v_1) = A^2v - Az_1v_1 \\ &= v_3 - z_1(Av_1) = v_3 - z_1^2v_1 - z_1v_2 \\ &= -z_1^2v_1 - z_1v_2 + 1 \cdot v_3 + 0 \cdot v_4 + \ldots + 0 \cdot v_n \end{aligned}$$

and the second column of A_1 is $(-z_1^2, -z_1, 1, 0, \ldots, 0)^T$ and continuing in this way we see that

$$A \approx A_1 = \begin{pmatrix} z_1 & -z_1^2 & b_{13} & \ldots & b_{1n} \\ 1 & -z_1 & b_{23} & \ldots & b_{2n} \\ 0 & 1 & b_{33} & \ldots & b_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & b_{n3} & \ldots & b_{nn} \end{pmatrix} = \begin{pmatrix} z_1 & x^T \\ y & B_1 \end{pmatrix}.$$

(II) The minimal polynomial of A is quadratic. The rational canonical form of A contains a companion matrix $C(f_2(x))$ of order 2, since the minimal polynomial of A is quadratic. It contains another companion matrix $C(f_1(x))$ of order 1 or 2, such that $f_1 \mid f_2$. By re-ordering the blocks in rational canonical form of A we assume that A is similar to $C(f_2(x)) \oplus C(f_1(x)) \oplus \ldots$. The result for the general case follows by considering the first two companion matrices and assuming that A is one of the following forms:

(i) $n = 3$ and

$$A = \text{Diag}(a, a, b), a \neq b$$

(ii) $n = 3$ and

$$A = \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix} \oplus (a)$$

(iii) $n = 4$ and

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

(iv) $n = 4$ and

$$A = B \oplus B,$$

where

$$B = \begin{pmatrix} 0 & a \\ 1 & b \end{pmatrix}$$

The characteristic polynomial of B is irreducible over F.

We will spend some time to explain the reason for the existence of Cases (i) - (iv). Every $A \in M_n(F)$ is similar to a matrix of the form

$$\text{Diag}[C(f_1(x)), C(f_2(x)), \ldots, C(f_t(x))],$$

where $f_i(x)$ are the nonconstant invariant factors of $xI - A$. We know that $f_t(x)$ is the minimal polynomial of A and in this case it is quadratic. Since $f_i \mid f_{i+1}$ for each i it follows that the canonical form of A which is $\text{Diag}[C(f_1(x)), C(f_2(x)), \ldots, C(f_t(x))]$ consists of 1×1 and or 2×2 blocks of a particular form.

For $n = 3$ we have that $A \approx \text{Diag}[C(f_1(x)), C(f_2(x))]$ where $f_1 = x - a$ and

(a) $f_2(x) = (x - a)(x - b)$ with a and b distinct, or

(b) $f_2(x) = (x - a)(x - a)$

In the first case $C[f_2(x)] = C[(x - a)(x - b)]$ is similar to $\text{Diag}(a, b)$ and in the second case $C[f_2(x)] = C[(x - a)(x - a)]$ is similar to

$$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}.$$

This yields the two cases listed.

For $n = 4$ we have that $A \approx \text{Diag}[C(f_1(x)), \ldots, C(f_3(x))]$ where $f_3(x) = (x - a)(x - b)$ with a and b distinct, or $f_3(x) = (x - a)(x - a)$ or $f_3(x) = [x(x - b) - a]$.

The possible rational canonical forms for $n = 4$:

(a) $f_1 = (x - a)$, $f_2 = (x - a)$ and $f_3(x) = (x - a)(x - b)$

(b) $f_1 = (x - a)(x - b)$ and $f_3(x) = (x - a)(x - b)$

(c) $f_1 = (x - a)$, $f_2 = (x - a)$ and $f_3(x) = (x - a)(x - a)$

(d) $f_1 = (x - a)(x - a)$ and $f_3(x) = (x - a)(x - a)$

(e) $f_1 = [x(x - b) - a]$ and $f_3(x) = [x(x - b) - a]$ (irreducible)

The results for (a), (b) and (c) can be deduced from the case $n = 3$ and we only consider (d) and (e). In (d), $C(f_3(x)) \approx \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$ and in (e), $C(f_3(x)) = C[x(x - b) - a] \approx \begin{pmatrix} 0 & a \\ 1 & b \end{pmatrix}$. This yields the two cases for $n = 4$ We will show that the matrices (i) - (iv) are similar to a matrix of the form

$$\begin{pmatrix} z_1 & x^T \\ y & B_1 \end{pmatrix}$$

where B_1 is non-scalar.

Case (i)

Let

$$A^{'} = \begin{pmatrix} z_1 & -(a-z_1) & 0 \\ b-z_1 & a+b-z_1 & 0 \\ 0 & 0 & a \end{pmatrix}.$$

Since

$$C = \begin{pmatrix} z_1 & -(a-z_1) \\ b-z_1 & a+b-z_1 \end{pmatrix}$$

has distinct eigenvalues a and b it is similar to the diagonal matrix with a and b on its diagonal, hence A and $A^{'}$ are similar. The matrix $A^{'}$ has the desired form if $z_1 \neq b$. If $z_1 = b$ then B_1 is scalar. We find that

$$X \begin{pmatrix} b & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} X^{-1} = \begin{pmatrix} b & 0 & a-b \\ a-b & a & b-a \\ 0 & 0 & a \end{pmatrix}$$

is of the required form, where

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

Case (ii)

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

We find that

$$XAX^{-1} = \begin{pmatrix} z_1 & 1 & 0 \\ -(a-z_1)^2 & 2a-z_1 & 0 \\ z_1-a & 1 & a \end{pmatrix}$$

which is of the required form, where

$$X = \begin{pmatrix} 1 & 0 & 0 \\ a-z_1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Case (iii)

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

and is similar to

$$XAX^{-1} = \begin{pmatrix} z_1 & 1 & 0 & 0 \\ -a^2 + 2az_1 - z_1^2 & 2a - z_1 & 0 & 0 \\ 0 & 0 & z_1 & 1 \\ 0 & 0 & -a^2 + 2az_1 - z_1^2 & 2a - z_1 \end{pmatrix}$$

which is of the required form, where

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a - z_1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & a - z_1 & 1 \end{pmatrix}.$$

Case (iv)

$A = B \oplus B$, where

$$B = \begin{pmatrix} 0 & a \\ 1 & b \end{pmatrix},$$

a, $b \neq 0$. A is similar to

$$XAX^{-1} = \begin{pmatrix} z_1 & -z_1^2 + a + bz_1 & 0 & 0 \\ 1 & b - z_1 & 0 & 0 \\ 0 & 0 & 0 & a \\ 0 & 0 & 1 & b \end{pmatrix}$$

which is of the required form, where

$$X = \begin{pmatrix} 1 & z_1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(III) The minimal polynomial of A is of degree 1. If the minimal polynomial of A is of degree 1, i.e. of the form $x - \alpha$, then $A - \alpha I = 0$ and hence $A = \alpha I$ is scalar. This means that the theorem does not apply to A.

□

Theorem 2.5 [17, Theorem 1.1]

Let $A \in GL(n,F)$, $n \geq 2$ be non-scalar. Let $x_1, x_2, \ldots, x_n, y_1, \ldots, y_n$ be arbitrary elements of F subject to $\det A = x_1 \cdot x_2 \cdots x_n \cdot y_1 \cdots y_n$. Then A is similar over F to a product LU where L is lower triangular, U is upper triangular and $\operatorname{diag}(L) = (x_1, x_2, \ldots, x_n)$, $\operatorname{diag}(U) = (y_1, y_2, \ldots, y_n)$.

Proof. Let $z_i = x_i y_i$, $(i = 1, 2, \ldots, n)$. The proof is done by induction on n.

In the case where $n = 2$ we have that the minimal polynomial of A is quadratic, provided that A is not scalar.

The matrix A is of the form $\begin{pmatrix} c & d \\ e & f \end{pmatrix}$, where $cf - de \neq 0$ and $c \neq f$ if $d = e = 0$. Since A is non-scalar its minimal polynomial is quadratic and the same as its characteristic polynomial, namely

$$f_2(x) = (x-c)(x-f) - de = x^2 - [(de - cf) + (c+f)x].$$

So

$$A \approx C(f_2(x)) = \begin{pmatrix} 0 & (de-cf) \\ 1 & (c+f) \end{pmatrix}$$

and $C(f_2(x))$ is similar to a matrix of the form $\begin{pmatrix} z_1 & x \\ y & b \end{pmatrix}$ where x, y, b are elements of F. Similarity holds due to the fact that $X^{-1}C(f_2(x))X$ has the required form, where $X = \begin{pmatrix} 1 & -z_1 \\ 0 & 1 \end{pmatrix}$. Since A is similar to $C(f_2(x))$ it also is similar to a matrix of the form $\begin{pmatrix} z_1 & x \\ y & b \end{pmatrix}$.

We have that:

$$\begin{aligned} A \approx \begin{pmatrix} z_1 & x \\ y & b \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ yz_1^{-1} & 1 \end{pmatrix} \cdot \begin{pmatrix} z_1 & x \\ 0 & b - xyz_1^{-1} \end{pmatrix} \\ &= \begin{pmatrix} x_1 & 0 \\ yy_1^{-1} & x_2 \end{pmatrix} \cdot \begin{pmatrix} y_1 & x_1^{-1}x \\ 0 & y_2 \end{pmatrix} \end{aligned}$$

since $b = yy_1^{-1}x_1^{-1}x + x_2y_2$. The matrix A is similar to a product of a lower triangular and an upper triangular matrix. Note that $\det A = 1 \cdot z_1(b - xyz_1^{-1})$ and $\det A = x_1x_2y_1y_2 = z_1z_2$. Therefore $z_1(b - xyz_1^{-1}) = z_1z_2$ and $z_2 = b - xyz_1^{-1}$.

Assume that $n \geq 3$ and that the result holds for $n-1$. The matrix A is similar to

$$A_1 = \begin{pmatrix} z_1 & x^T \\ y & B_1 \end{pmatrix}$$

where B_1 is non-scalar. We write

$$A_1 = \begin{pmatrix} 1 & 0 \\ yz_1^{-1} & I \end{pmatrix} \cdot \begin{pmatrix} z_1 & x^T \\ 0 & B_2 \end{pmatrix}$$

where $B_2 = B_1 - yx^T z_1^{-1}$. In order to apply the induction step we need B_2 to be non-scalar. It would thus be necessary to find another matrix A_2 which is similar to A_1 if B_2 were to be scalar.

If B_2 were to be scalar, say $B_2 = bI_n$, then $y \neq 0$ since B_1 is non-scalar in the equation $B_2 = B_1 - yx^T z_1^{-1}$. Since $n \geq 3$, there exists $x_0 \in F^{n-1}$ with $x_0^T y = 0$, $yx_0^T \neq 0$. We see that x_0 and y are so chosen such that if $x_0^T y = 0$ then z_1 remains fixed in A_2 ($A_2 \approx A_1$) and if $yx_0^T \neq 0$ then $B_3 = B_1 - yx_0^T$ and $B_3 \neq B_1$. Now A is similar to

$$\begin{aligned} A_2 &= \begin{pmatrix} 1 & x_0^T \\ 0 & I \end{pmatrix} \cdot A_1 \cdot \begin{pmatrix} 1 & -x_0^T \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} z_1 & x_1^T \\ y & B_3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ yz_1^{-1} & I \end{pmatrix} \cdot \begin{pmatrix} z_1 & x_1^T \\ 0 & B_4 \end{pmatrix} \end{aligned}$$

where $x_1^T = -z_1 x_0^T + x^T + x_0^T B_1$ from $B_3 = B_1 - yx_0^T$. We find that:

$$\begin{aligned} B_4 &= B_3 - yx_1^T z_1^{-1} = (B_1 - yx_0^T) - yx_1^T z_1^{-1} \text{ from } B_3 = B_1 - yx_0^T \\ &= (B_2 + yx^T z_1^{-1} - yx_0^T) - yx_1^T z_1^{-1} \text{ from } B_2 = B_1 - yx^T z_1^{-1}. \end{aligned}$$

Furthermore

$$\begin{aligned} B_4 &= bI + yx^T z_1^{-1} - yx_0^T - y(-z_1 x_0^T + x^T + x_0^T B_1) z_1^{-1} \\ &= bI + yx^T z_1^{-1} - yx_0^T + yx_0^T - yx^T z_1^{-1} - yx_0^T B_1 z_1^{-1} = bI - yx_0^T B_1 z_1^{-1} \end{aligned}$$

from $x_1^T = -z_1 x_0^T + x^T + x_0^T B_1$. It follows that B_4 is non-scalar since $yx_0^T B_1 z_1^{-1}$ has rank 1.

Using induction on n and using B_2 (if it is not scalar) and B_4 (if it is scalar) we find that $A \approx L_1 U_1$ where L_1 is lower triangular and $\operatorname{diag}(L_1) = (1, 1, \ldots, 1)$. The matrix U_1 is upper triangular and $\operatorname{diag}(U_1) = (z_1, z_2, \ldots, z_n)$.

The conclusion of the theorem holds with $L = L_1 D_1$ and $U = D_1^{-1} U_1$, where $D_1 = \operatorname{Diag}(x_1, x_2, \ldots, x_n)$. We find that matrix A obeys

$$A \approx L_1 U_1 \approx L_1 D_1 U_1 D_1^{-1} = LU$$

and matrix $A \approx LU$. The determinant $\det LU = x_1 \cdot x_2 \cdots x_n \cdot y_1 \cdot y_2 \cdots y_n$, since L and U are lower and upper triangular matrices respectively with $\operatorname{diag}(L) = (x_1, x_2, \ldots, x_n)$ and $\operatorname{diag}(U) = (y_1, y_2, \ldots, y_n)$. Since $A \approx LU$, we find that $\det A = x_1 \cdot x_2 \cdots x_n \cdot y_1 \cdot y_2 \cdots y_n$. □

2.3 Applications to Unipotent, Positive-Definite and Commutator Factorization

We will first give Sourour's presentation [24] and then Laffey's as it appears in [17]. We will also consider Laffey's articles as it appears in [18], where necessary.

2.3.1 Products of Positive-Definite Matrices

Sourour

Various authors, such as [22], have shown in their proofs of theorems that if A is similar to a product of 2 or 4 positive-definite matrices then it is also equal to a product of 2 or 4 positive-definite matrices. This fact will be used in the proofs to follow. Sourour studied conditions under which a matrix A (real or complex) can be expressed as a product of 4 positive-definite matrices and under what conditions it can be expressed as a product of 5 positive-definite matrices. This result appears in Ballantine ([1] and [2]). With the use of [24, Theorem 1], it is possible to give a short proof of this result which demonstrates its usefulness. Sourour also contains other partial matrix factorization results with the use of [24, Theorem 1] and because of this it is considered as a first attempt at unifying a number of seemingly unrelated matrix factorization results in one theory. The results here are obtained by manipulating the spectral properties of the matrices involved. More formally the theorem is stated as follows:

Theorem 2.6 [24, Theorem 2]

Let A be a real or complex $n \times n$ matrix. Then

(a) A is a product of four positive-definite matrices if and only if $\det A > 0$ and A is not a scalar αI where α is not positive.

(b) A is a product of five positive-definite matrices if and only if $\det A > 0$.

Proof.

(a) If A is non-scalar, we know from Theorem 2.2 that it can be expressed as $A = BC$ where $\det A = \prod_{i=1}^{n} \beta_i \gamma_i$ and β_i the arbitrarily chosen positive eigenvalues of B and γ_i the arbitrarily chosen positive eigenvalues of C. We choose the β_i's and γ_i's to be distinct. Matrices B and C are thus diagonalizable and there exists a positive diagonal matrix D such that $B = R^{-1}DR$ and R is invertible. Matrix D is positive since all the β_i's are chosen to be positive. Furthermore

$$\begin{aligned} B &= R^{-1}DR = R^{-1}{R^*}^{-1}R^*DR \\ &= (R^{-1}{R^*}^{-1})(R^*DR) = (R^*R)^{-1}(R^*DR) \end{aligned}$$

a product of two positive-definite matrices. Similarly C can be written as a product of two positive-definite matrices and hence A is a product of 4 positive-definite matrices.

The matrix $A = \alpha I$ is trivial if $\alpha > 0$, since then A is already positive-definite, hence we can write $A = A \cdot I \cdot I \cdot I$.

Next we present the converse. First note that if A is a product of 4 positive-definite matrices such that $A = P_1 P_2 P_3 P_4$ then $\det A > 0$. The determinant $\det A$ is positive since all the P_i's have positive eigenvalues and $\det A$ is a product of the eigenvalues of the P_i's.

We want to show that if $A = \alpha I = P_1 P_2 P_3 P_4$ and the P_i's are positive-definite then $\det A > 0$ and $\alpha > 0$. The fact that $\det A > 0$ was explained previously. It follows that $\alpha I = P_1 P_2 P_3 P_4$ and $P_1 P_2 = \alpha P_4^{-1} P_3^{-1}$. We notice that $P_1 P_2 = C(P_1^{\frac{1}{2}} P_2 P_1^{\frac{1}{2}}) C^{-1}$, where $C = P_1^{\frac{1}{2}}$ the unique positive-definite square root of P_1. The matrix $P_1^{\frac{1}{2}} P_2 P_1^{\frac{1}{2}}$ is positive-definite, since

$$x^* P_1^{\frac{1}{2}} P_2 P_1^{\frac{1}{2}} x = (P_1^{\frac{1}{2}} x)^* P_2 (P_1^{\frac{1}{2}} x) > 0$$

where x is a non-zero column vector. The equality follows due to the fact that P_1 is selfadjoint and the inequality due to the fact that P_2 is a positive-definite matrix. Since $P_1 P_2$ is similar to a positive-definite matrix $P_1^{\frac{1}{2}} P_2 P_1^{\frac{1}{2}}$ it has a positive spectrum. Similarly $P_4^{-1} P_3^{-1}$ has a positive spectrum. We know that $y^*(P_4^{-1} P_3^{-1}) y > 0$ for any non-zero column vector y and thus

$$\alpha y^*(P_4^{-1} P_3^{-1}) y = y^*(\alpha P_4^{-1} P_3^{-1}) y = y^* P_1 P_2 y > 0$$

if and only if $\alpha > 0$.

(b) ($\Leftarrow$) Assume that $\det A > 0$ and that $A = \alpha I$ (We can choose A to be any real or complex $n \times n$ matrix). The matrix A can be expressed as $A = \alpha P^{-1} P$ where P is a non-scalar positive-definite matrix. The inverse of a positive-definite matrix is also positive-definite and therefore P^{-1} is a positive-definite matrix. Since $\det A > 0$ it implies that $\det(\alpha P^{-1})$det P > 0. Furthermore $\det P > 0$ (P is positive-definite) and therefore $\det(\alpha P^{-1}) > 0$. Also the matrix αP^{-1} is not a scalar matrix, since P is a non-scalar positive-definite matrix. Applying the result of (a) to αP^{-1}, we see that $A = \alpha I = (\alpha P^{-1}) P$ is a product of 5 positive-definite matrices.

(⇒) Assume that A is a product of five positive-definite matrices. Each of the positive-definite matrices have a positive determinant by definition. The determinant of A is a product of the determinants of the five positive-definite matrices and thus $\det A > 0$. □

Laffey

Laffey also has a theorem [17, Theorem 2.3] that deals with positive-definite matrices in the real case only. Although his results follow directly from Sourour's theorem (Theorem 2.6), we present here the proof as given by Laffey. The theorem can be stated as follows:

Theorem 2.7 [17, Theorem 2.3]

Let A be a non-singular real $n \times n$ matrix with $\det A$ positive and suppose A is not a negative scalar. Then $A = P_1P_2P_3P_4$ where each P_i is a positive-definite real symmetric matrix.

Proof. The case where A is a positive scalar matrix follows trivially. If A is a non-scalar it follows by Theorem 2.5, that A is similar to LU where L is lower triangular with $\mathrm{diag}(L) = d(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n})$, $d = (\det A)^{\frac{1}{n}}$ and

$$U = \begin{pmatrix} 1 & & & \\ & 2 & * & \\ & & \ddots & \\ & & & n \end{pmatrix}.$$

Since $L \approx \mathrm{Diag}(L)$ and

$$\begin{aligned} \mathrm{Diag}(L) &= \begin{pmatrix} (\det A)^{\frac{1}{n}} & & & \\ & \frac{1}{2}(\det A)^{\frac{1}{n}} & & \\ & & \ddots & \\ & & & \frac{1}{n}(\det A)^{\frac{1}{n}} \end{pmatrix} \\ &= \begin{pmatrix} 1 & & & \\ & \frac{1}{2} & & \\ & & \ddots & \\ & & & \frac{1}{n} \end{pmatrix} \begin{pmatrix} (\det A)^{\frac{1}{n}} & & & \\ & (\det A)^{\frac{1}{n}} & & \\ & & \ddots & \\ & & & (\det A)^{\frac{1}{n}} \end{pmatrix} \end{aligned}$$

a product of two real symmetric matrices with positive eigenvalues i.e. a product of two positive-definite matrices.

Every square matrix over an arbitrary field can be expressed as a product of two symmetric matrices [8], therefore $L = S_1S_2$ where S_1 and S_2 are symmetric matrices. Furthermore

$$L = D^{-1}\text{Diag}(L)D = D^{-1}(P_1P_2)D = (D^{-1}P_1(D^{-1})^T)(D^TP_2D) = S_1S_2$$

for some invertible matrix D. It follows that $S_1 = (D^{-1}P_1(D^{-1})^T)$ and $S_2 = (D^TP_2D)$. Both symmetry and eigenvalues are preserved under orthogonal similarity and therefore S_1 and S_2 are positive-definite. It follows that L is a product of two positive-definite matrices. Similarly,

$$\begin{aligned}\text{Diag}(U) &= \begin{pmatrix} 1 & & & \\ & 2 & & \\ & & \ddots & \\ & & & n \end{pmatrix} \\ &= \begin{pmatrix} 1 & & & \\ & 2^{-n} & & \\ & & \ddots & \\ & & & n^{-n} \end{pmatrix} \begin{pmatrix} 1 & & & \\ & 2^{n+1} & & \\ & & \ddots & \\ & & & n^{n+1} \end{pmatrix}\end{aligned}$$

and hence U is a product of two real symmetric matrices with positive eigenvalues i.e. a product of two positive-definite matrices by a similar argument as for L.

Let $LU = (R_1R_2)(R_3R_4)$ where R_i, $1 \leq i \leq 4$ is a positive-definite matrix then

$$\begin{aligned}A &= S^{-1}(LU)S = S^{-1}(R_1R_2R_3R_4)S \\ &= S^{-1}R_1(S^T)^{-1}S^TR_2SS^{-1}R_3(S^T)^{-1}S^TR_4S \\ &= (S^{-1}R_1(S^T)^{-1})(S^TR_2S)(S^{-1}R_3(S^T)^{-1})(S^TR_4S) \\ &= (S^{-1}R_1(S^{-1})^T)(S^TR_2S)(S^{-1}R_3(S^{-1})^T)(S^TR_4S)\end{aligned}$$

for some invertible matrix S. Since each of the products of A are orthogonally similar to a positive-definite matrix they too are positive-definite and it follows that A is a product of 4 positive-definite matrices. □

2.3.2 Applications to Commutators

Sourour

The set of commutators in $GL(n, F)$ is described by $\{BCB^{-1}C^{-1} : B,C \in GL(n, F)\}$. The problem considered by Sourour can be restated as follows: If F has a certain cardinality, is it possible to express the $n \times n$ invertible matrix A, with $\det A = 1$ as a commutator of matrices? The commutator theorem to follow is a special case due to Shoda and Thompson. Sourour's Theorem [24, Theorem 1] alone cannot be used to prove the commutator theorem, since in order to apply it we require the underlying scalar field F to be of a certain order. More formally the partial version of the commutator theorem is stated as follows:

Theorem 2.8 - Shoda-Thompson [24, Theorem 3]

Let $A \in SL(n, F)$.

(a) If F has at least $n + 1$ elements, then A is a commutator of matrices in $GL(n, F)$.

(b) If F has at least $n + 2$ elements, and A is non-scalar, then A is a commutator of matrices in $SL(n, F)$.

(c) If F has at least $n+3$ elements and A is non-scalar, then A is a commutator of matrices with arbitrary prescribed nonzero determinants.

Proof.

(a) ***A* non-scalar**

Sourour makes use of Theorem 2.2 to show that $A = BD$, where B is similar to D^{-1}. The distinct eigenvalues of B and D can be prescribed as $\{\beta_1, \beta_2, \ldots, \beta_n\}$ and $\{\beta_1^{-1}, \beta_2^{-1}, \ldots, \beta_n^{-1}\}$ respectively. The eigenvalues $\{\beta_1, \beta_2, \ldots, \beta_n\}$ can be chosen to be distinct, since F has at least $n+1$ elements. Since the eigenvalues of B and D are distinct the matrices B and D are diagonalizable i.e. similar to a diagonal matrix. Due to the way in which the eigenvalues of D are prescribed we notice that $D \approx B^{-1}$ and therefore $D = CB^{-1}C^{-1}$ for some invertible matrix C. It follows that $A = BD = BCB^{-1}C^{-1}$ and is a commutator of matrices in $GL(n, F)$.

A scalar

Let $B = \mathrm{Diag}\{\alpha, \alpha^2, \ldots, \alpha^n\}$ and $D = \mathrm{Diag}\{1, \alpha^{-1}, \alpha^{-2}, \ldots, \alpha^{1-n}\}$. Since $1 = \alpha^n$ we find that B is similar to the inverse of D and the argument proceeds as before to yield the desired result.

(b) If F has at least $n + 2$ elements, we show that $\beta_1, \beta_2, \ldots, \beta_n$ may be chosen to satisfy $\beta_1 \cdots \beta_n = 1$ in addition to being distinct.

n odd

Let

$$\{\beta_1, \ldots, \beta_n\} = \left\{1, \gamma_1, \gamma_1^{-1}, \ldots, \gamma_{\frac{n-1}{2}}, \gamma_{\frac{n-1}{2}}^{-1}\right\}$$

where $(\gamma_i, \gamma_i^{-1})$ are distinct pairs with $\gamma_i \neq \pm 1, 0$.

n even

If $|F| = n+2$, the characteristic of F is 2 and $1 = -1$ and therefore essentially only two scalars are excluded from F. The elements that are excluded from F (for the β_i's) are 0 and ± 1. We take $\frac{n}{2}$ distinct pairs of the form (β_i, β_i^{-1}) with $\beta_i \neq \pm 1$, 0. All the β_i's and β_i^{-1}'s are taken from F and are distinct. It follows that $B = XGX^{-1}$ where $G \in M_{n\times n}(F)$ is diagonal and non-singular. Let $E = X\tilde{G}X^{-1}$ where $\tilde{G} \in M_{n\times n}(F)$ is an arbitrary non-singular diagonal matrix. Then

$$\begin{aligned} BE &= (XGX^{-1})(X\tilde{G}X^{-1}) = XG\tilde{G}X^{-1} = X\tilde{G}GX^{-1} \\ &= (X\tilde{G}X^{-1})(XGX^{-1}) = EB. \end{aligned}$$

Thus a matrix E of arbitrary non-zero determinant commuting with B exists. Secondly it follows from Theorem 2.2 as was done in (a) that

$$\begin{aligned} B(CE)B^{-1}(CE)^{-1} &= BCEB^{-1}E^{-1}C^{-1} = BCEE^{-1}B^{-1}C^{-1} \\ &= BCB^{-1}C^{-1} = A \end{aligned}$$

where $\det B = 1$ and E can be chosen such that $\det CE = 1$, which completes the proof of (b).

(c) The Proof is similar to (b). If F has at least $n+3$ elements, we show that $\beta_1, \ldots, \beta_n$ may be chosen to satisfy $\beta_1 \cdot \beta_2 \cdot \ldots \cdot \beta_n = \alpha$, where α is arbitrary non-zero and the β_i's are distinct. If F has at least $n+3$ elements and A is non-scalar, then irrespective of whether n is even or odd, F has at least k distinct pairs $\{\beta_1, \beta_1^{-1}\}, \ldots, \{\beta_k, \beta_k^{-1}\}$, where k is the smallest integer such that $k \geq \frac{n}{2}$. This can be demonstrated as follows.

n even

F consists of at least $\frac{n}{2}$ distinct pairs of the form $\{\beta_i, \beta_i^{-1}\}$ where $\beta_i \neq \pm 1, 0$.

(i) $\alpha = -1$ or $+1$. If $\alpha = +1$, choose $\frac{n}{2}$ pairs of the form $\{\beta_i, \beta_i^{-1}\}$ with $\beta_i \neq \pm 1, 0$ from F. If $\alpha = -1 \neq +1$, we choose a pair of the form $\{1, -1\}$ and then $\frac{n-2}{2}$ distinct pairs of the form $\{\beta_i, \beta_i^{-1}\}$ where $\beta_i \neq \pm 1, 0$.

(ii) $\alpha \neq \pm 1$. Take one distinct pair of the form $\{\alpha, 1\}$ and then $\frac{n-2}{2}$ distinct pairs of the form $\{\beta_i, \beta_i^{-1}\}$ where $\beta_i \neq \pm 1, 0, \alpha, \alpha^{-1}$.

n odd

If $|F| > n+3$ we are able to find at least $\frac{(n+1)}{2}$ distinct pairs of the form $\{\beta_i, \beta_i^{-1}\}$ with $\beta_i \neq \pm 1, 0$ in F. If $\alpha \neq \pm 1$, one of the distinct pairs should have the form $\{\alpha, \alpha^{-1}\}$. The remaining elements in F should at least contain $\{\pm 1, 0\}$. We choose $\frac{(n-1)}{2}$ distinct pairs of the form $\{\beta_i, \beta_i^{-1}\}$ with $\beta_i \neq \pm 1, 0, \alpha$ and select $\beta_n = \alpha$. In the case where $|F| = n+3$ we choose $\frac{(n-1)}{2}$ distinct pairs of the form $\{\beta_i, \beta_i^{-1}\}$ with $\beta_i \neq \pm 1, 0, \alpha$ and select β_n which can only be one of the following namely 1 or -1. It follows that α in this case can only be equal to 1 or -1. The rest of the proof of (c) is similar to (b). $\square$

Laffey

Theorem 2.9 [17, Theorem 1.2]

Let F be a field with at least $n+3$ elements and let $A \in SL(n, F)$. There exists X,$Y \in GL(n, F)$ with $A = X^{-1}Y^{-1}XY$.

Proof. **A non-scalar**

Let $a_1, a_1^{-1}, a_2, a_2^{-1}, \ldots, a_k, a_k^{-1}$ $(k = \left[\frac{n}{2}\right])$ be distinct elements of F. This choice is possible since $|F| \geq n+3$. Also the a_i are chosen such that

$$\begin{aligned}\det A &= 1\\ &= x_1 \cdot x_2 \cdots x_n \cdot y_1 \cdot y_2 \cdots y_n\\ &= (a_1 \cdot a_2 \cdots a_k \cdot a_1^{-1} \cdot a_2^{-1} \cdots a_k^{-1} \cdot 1) \cdot (a_1^{-1} \cdot a_2^{-1} \cdots a_k^{-1} \cdot a_1 \cdot a_2 \cdots a_k \cdot 1),\end{aligned}$$

when n is odd or

$$(a_1 \cdot a_2 \cdots a_k \cdot a_1^{-1} \cdot a_2^{-1} \cdots a_k^{-1}) \cdot (a_1^{-1} \cdot a_2^{-1} \cdots a_k^{-1} \cdot a_1 \cdot a_2 \cdots a_k) = 1$$

if n is even. Here,

$$x_i = \begin{cases} a_i & \text{if } 1 \leq i \leq k, \\ a_i^{-1} & \text{if } k+1 \leq i \leq 2k, \\ 1 & \text{if i} = n \text{ and } n \text{ is odd} \end{cases}$$

and $y_i = x_i^{-1}$, $1 \leq i \leq n$. It follows that $A \approx LU$ from Theorem 2.5, where $L \approx \text{Diag}(x_1, \ldots, x_n)$ and $U \approx \text{Diag}(x_1^{-1}, \ldots, x_n^{-1})$. Note that $\det L = 1$,

$U = S^{-1}L^{-1}S$ and $T^{-1}AT = LU$. We define $X = TL^{-1}T^{-1}$ and $Y = TST^{-1}$. The matrix A is such that:

$$\begin{aligned} A &= TLUT^{-1} = (TLT^{-1})(TS^{-1}T^{-1})(TL^{-1}T^{-1})(TST^{-1}) \\ &= (TL^{-1}T^{-1})^{-1}(TST^{-1})^{-1}(TL^{-1}T^{-1})(TST^{-1}) = X^{-1}Y^{-1}XY \end{aligned}$$

and A is a multiplicative commutator.

Laffey claims that the role the matrix E plays that was introduced in Theorem 2.8 (b) can also be achieved by matrices in $F[L]$ (Matrices that commute automatically with L). The algebra generated by L contains for each given diagonal matrix D, an element W with $\text{Diag}(W) = D$. The matrix $W \in F[L]$ of arbitrary diagonal can be generated in the following manner: By subtracting multiples of I from L we obtain lower triangular matrices with 0 in the i-th diagonal entry and non-zero elements elsewhere on the diagonal. Multiplying these matrices yield elements of $F[L]$ with 1 in the j-th diagonal entry and zeros elsewhere on the diagonal. Linear combinations of these matrices yield elements in $F[L]$ with arbitrary diagonal. We are thus able to construct W such that $\text{Diag}(W) = \text{Diag}(b_1, b_2, \ldots, b_n)$ where $b_1 \cdot b_2 \ldots \cdot b_n = \frac{a}{b}$, a is arbitrary non-zero and b is fixed. Assume that $\det S \neq a$ and $\det S = b$. We can change S to $S^{'}$ by means of W such that $S^{'} = WS$ and $\det S^{'} = \det WS = \det W \cdot \det S = \frac{a}{b} \cdot b = a$. The matrix X remains unchanged while Y is changed to $Y^{'}$. The matrix $Y^{'} = TS^{'}T^{-1}$ and since $Y^{'} \approx S^{'}$, $\det Y^{'} = \det S^{'} = a$. Furthermore

$$\begin{aligned} X^{-1}Y^{'^{-1}}XY^{'} &= (TL^{-1}T^{-1})^{-1}(TS^{'}T^{-1})^{-1}(TL^{-1}T^{-1})(TS^{'}T^{-1}) \\ &= TLS^{'^{-1}}L^{-1}S^{'}T^{-1} = TL(WS)^{-1}L^{-1}(WS)T^{-1} \\ &= TLS^{-1}W^{-1}L^{-1}WST^{-1} = TLS^{-1}L^{-1}W^{-1}WST^{-1} \\ &= TLS^{-1}L^{-1}ST^{-1} = TLUT^{-1} = A \end{aligned}$$

(W commutes with L). The commutator formed by X and $Y^{'}$ is still equal to A. The matrix S (and therefore also Y) can be chosen to have $\det S = a$ (If this is not possible S will be transformed to $S^{'}$ as was shown previously), where a is any nonzero element of F. Note that $\det X = 1$. The matrix S can be chosen in $SL(n, F)$ and thus X and Y can also be chosen in $SL(n, F)$. Since $Y = TST^{-1}$ and $\det Y = \det S$.

<u>A **scalar**</u>

Let $A = \alpha I$. We take X as a companion matrix of $x^n + (-1)^n$:

$$X = \begin{pmatrix} 0 & 0 & \dots & 0 & (-1)^{n+1} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

It follows that the characteristic polynomial of X is $x^n + (-1)^n$. We will show that $Y^{-1}(\alpha X)Y = X$, where

$$Y = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \alpha & 0 & \dots & 0 \\ 0 & 0 & \alpha^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \alpha^{n-1} \end{pmatrix}$$

and $\alpha^n = 1$. We find that

$$Y^{-1}(\alpha X)Y = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \alpha^{-1} & 0 & \dots & 0 \\ 0 & 0 & \alpha^{-2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \alpha^{1-n} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & \dots & \pm\alpha \\ \alpha & 0 & 0 & \dots & 0 \\ 0 & \alpha & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

$$\times \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \alpha & 0 & \dots & 0 \\ 0 & 0 & \alpha^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \alpha^{n-1} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & \dots & 0 & (-1)^{n+1} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix} = X$$

and thus $X \approx \alpha X$. Since $Y^{-1}(\alpha X)Y = X \Rightarrow \alpha X = YXY^{-1}$ and there exists a matrix T such that $T = Y^{-1}$ and $\alpha X = T^{-1}XT$. The matrix A is such that

$$A = \alpha I = \alpha X^{-1}X = X^{-1}\alpha X = X^{-1}T^{-1}XT$$

and A is a multiplicative commutator. The result of Laffey's Theorem can also be stated as a corollary (Corollary 2.10). No proof will be presented as the results follow from Theorem 2.9. $\square$

Remark: Laffey's proof in Theorem 2.9 actually proves all three results in Theorem 2.8 in the case where F has at least $n+3$ elements i.e. Laffey's proof shows that,

(i) If F has at least $n+3$ elements, then A is a commutator of matrices in $GL(n, F)$.

(ii) If F has at least $n+3$ elements and A is non-scalar then A is a commutator of matrices in $SL(n, F)$.

(iii) If F has at least $n+3$ elements and A is non-scalar, then A is a commutator of matrices with arbitrary prescribed non-zero determinants.

Corollary 2.10 [18, Theorem 5.7]

Let $A \in SL(n, F)$ and suppose F has at least $n+3$ elements. Then A can be written in the form $X^{-1}Y^{-1}XY$ for some $X \in SL(n, F)$ similar to X^{-1} and $Y \in GL(n, F)$.

2.3.3 Applications to Unipotent Matrices

For this sub-section only the unipotent applications as found in [24, Corollary] will be considered.

A matrix D is said to be unipotent if $D - I$ is a nilpotent matrix. Since $D - I$ is nilpotent, $(D - I)^m = 0$ where m is the index of nilpotency of $D - I$. Hence the minimal polynomial of D is $(x-1)^m$ and therefore all its eigenvalues are equal to 1.

Sourour

Theorem 2.11 [24, Corollary]

Let $A \in M_n(F)$ and assume that $\det A = 1$. Then A is a product of three unipotent matrices. If A is non-scalar, then it is a product of two unipotent matrices.

Proof. Assume that A is non-scalar and $\det A = 1$. We choose all the β's and γ's equal to one and apply Theorem 2.2 to express A as $A = BC$ where the eigenvalues of B and C are all one. From Theorem 2.2, B and C can be chosen so that B is lower triangularizable and C is simultaneously upper triangularizable. It follows that $A \approx B^{'}C^{'}$, where $B \approx B^{'}$ and $C \approx C^{'}$ and $B^{'}$ is lower triangular and $C^{'}$ is upper triangular. Both $B^{'}$ and $C^{'}$ have diagonals (which are also the eigenvalues) equal to one. The matrices $B^{'} - I$ and $C^{'} - I$ are nilpotent (properties of strictly triangular matrices), with an index of nilpotency of l and m respectively. The matrices $B^{'}$ and $C^{'}$ are unipotent. The minimal polynomial of $B^{'}$ is $(x-1)^l$ and that of $C^{'}$ is $(x-1)^m$. Since $B \approx B^{'}$ and $C \approx C^{'}$, B and C also have the minimal polynomials of $(x-1)^l$ and $(x-1)^m$ respectively and so $B-I$ and $C-I$ are nilpotent. The matrices B and C are thus unipotent and A is a product of two unipotent matrices.

If $A = \alpha I$, write $A = (\alpha U)U^{-1}$, where U is a non-scalar unipotent matrix. The matrix U^{-1} is also unipotent, since the inverse of a unipotent matrix is unipotent. Since $\det A = 1$ and $\det U^{-1}U = 1$, (U^{-1} is unipotent) it follows that $\det \alpha U = 1$ as $\det A = \det \alpha U \cdot \det U^{-1}$. The product of a scalar and a non-scalar matrix produces a non-scalar matrix and thus αU is non-scalar. Using the first part of this proof we see that αU is a product of two unipotent matrices and thus $A = (\alpha U)(U^{-1})$ is a product of three unipotent matrices. □

2.4 Products of Two Involutions

This section is based on [13]. An involution is a square matrix whose square is the identity. The result of this paper is that an invertible linear transformation T of a finite dimensional vector space over a field is the product of two involutions if and only if $T \approx T^{-1}$. We will work through Lemma 2.12 to Lemma 2.16 before proving Theorem 2.17.

Lemma 2.12

Suppose f and g are monic polynomials over a field F with non-zero constant terms. The polynomial f has degree m with constant term a and polynomial g has degree n with constant term b. We define the $\tilde{}$ operator as $\tilde{f}(x) = \left(\frac{x^m}{a}\right) f\left(\frac{1}{x}\right)$, where a is the constant term of the polynomial f. Then

1. $\widetilde{fg} = \tilde{f}\tilde{g}$
2. g is irreducible if and only if $\tilde{g}$ is irreducible
3. $C_g^{-1} \approx C_{\tilde{g}}$

Proof. Property 1 is easy to show as

$$\tilde{f}(x)\tilde{g}(x) = \left(\frac{x^m}{a}\right) f\left(\frac{1}{x}\right) \cdot \left(\frac{x^n}{b}\right) g\left(\frac{1}{x}\right) = \left(\frac{x^{m+n}}{ab}\right) (fg) \left(\frac{1}{x}\right) = \widetilde{fg}(x)$$

We next prove property 2.
(⇒) Assume that g is reducible and let $g(x) = h(x)j(x)$, where $h(x)$and $j(x)$ have degrees less than deg g $= n$ and are not constant terms. It follows

that $g = hj \Rightarrow \tilde{g} = \widetilde{hj} = \tilde{h}\tilde{j}$ from property 1. Since deg h = deg $\tilde{h}$ it follows that $\tilde{g}$ is reducible.

($\Leftarrow$) Assume that $\tilde{g}$ is reducible and let $\tilde{g}(x) = r(x)s(x)$, where $r(x)$ and $s(x)$ are non-constant polynomials of degree less than n. We have that $\tilde{g} = rs \Rightarrow \tilde{\tilde{g}} = \widetilde{rs} \Rightarrow g = \tilde{r}\tilde{s}$ and since deg r = deg $\tilde{r}$ it follows that g is reducible.

To show property 3 we introduce the matrix P such that

$$P = P^{-1} = \begin{pmatrix} & & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & 1 & & & \\ 1 & & & & \end{pmatrix}$$

Let

$$\begin{aligned} g &= x^m + a_{m-1}x^{m-1} + a_{m-2}x^{m-2} + \ldots + a_1 x + a \\ &= x^m - \left(-a - a_1 x - \ldots - a_{m-1}x^{m-1}\right). \end{aligned}$$

Also

$$\begin{aligned} \tilde{g} &= \left(\frac{x^m}{a}\right)\left[\frac{1}{x^m} + a_{m-1}\frac{1}{x^{m-1}} + \ldots + \frac{a_1}{x} + a\right] \\ &= \frac{1}{a} + \frac{a_{m-1}}{a}x + \ldots + \frac{a_1}{a}x^{m-1} + x^m \\ &= x^m - \left(-\frac{1}{a} - \frac{a_{m-1}}{a}x - \ldots - \frac{a_1}{a}x^{m-1}\right). \end{aligned}$$

We find that

$$C_g = \begin{pmatrix} 0 & 0 & \ldots & 0 & -a \\ 1 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & \ldots & 1 & -a_{m-1} \end{pmatrix}$$

and

$$C_{\tilde{g}} = \begin{pmatrix} 0 & 0 & \ldots & 0 & -\frac{1}{a} \\ 1 & 0 & \ldots & 0 & -\frac{a_{m-1}}{a} \\ 0 & 1 & \ldots & 0 & -\frac{a_{m-2}}{a} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & \ldots & 1 & -\frac{a_1}{a} \end{pmatrix}.$$

The inverse C_g^{-1}:

$$\mathrm{C}_g^{-1} = \begin{pmatrix} -\frac{a_1}{a} & 1 & 0 & \dots & 0 \\ -\frac{a_2}{a} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -\frac{a_{m-1}}{a} & 0 & 0 & \dots & 1 \\ -\frac{1}{a} & 0 & 0 & \dots & 0 \end{pmatrix}$$

Furthermore

$$\begin{aligned} P^{-1}C_g^{-1}P &= \begin{pmatrix} & & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & 1 & & & \\ 1 & & & & \end{pmatrix} \begin{pmatrix} -\frac{a_1}{a} & 1 & 0 & \dots & 0 \\ -\frac{a_2}{a} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -\frac{a_{m-1}}{a} & 0 & 0 & \dots & 1 \\ -\frac{1}{a} & 0 & 0 & \dots & 0 \end{pmatrix} \\ &\times \begin{pmatrix} & & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & 1 & & & \\ 1 & & & & \end{pmatrix} \\ &= \begin{pmatrix} 0 & \dots & 0 & 0 & -\frac{1}{a} \\ 1 & \dots & 0 & 0 & \frac{a_{m-1}}{a} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & -\frac{a_2}{a} \\ 0 & \dots & 0 & 1 & \frac{a_1}{a} \end{pmatrix} \\ &= C_{\tilde{g}} \end{aligned}$$

and $C_{\tilde{g}} \approx C_g^{-1}$.

□

Remark: A monic polynomial $g(x)$ is symmetric if $g(x) = \tilde{g}(x)$. Also if $g(x)$ is symmetric then $P^{-1}C_g^{-1}P = C_{\tilde{g}} = C_g$.

Lemma 2.13 [13, Lemma 1]

Let A be an invertible $n \times n$ matrix over a field F. If A is similar to A^{-1}, then A is similar to a direct sum, $B_1 \oplus B_2 \oplus \ldots \oplus B_r \oplus D_1 \oplus \ldots \oplus D_s$, where $B_i = C_{f_i}$,

$$D_j = \begin{pmatrix} C_{g_j} & 0 \\ 0 & C_{g_j}^{-1} \end{pmatrix}$$

with each $f_i(x) = (p_i(x))^{n_i}$ for some symmetric irreducible $p_i(x)$; and each $g_i(x)$ a power of an irreducible polynomial.

Proof. We know that $A \approx C_{h_1} \oplus C_{h_2} \oplus \ldots \oplus C_{h_t}$. Since we are dealing with A up to similarity we assume that $A = C_{h_1} \oplus C_{h_2} \oplus \ldots \oplus C_{h_t}$, where the polynomials $h_1(x) \mid h_2(x) \mid \ldots \mid h_t(x)$ are non-trivial invariant factors of A. The reason for the choice of A is that if A is similar to B and also A is similar to A^{-1} then B is similar to B^{-1}, and hence it suffices to prove the result for B.

The matrix A^{-1} is such that

$$\begin{aligned} A^{-1} &= (C_{h_1} \oplus C_{h_2} \oplus \ldots \oplus C_{h_t})^{-1} \\ &= C_{h_1}^{-1} \oplus C_{h_2}^{-1} \oplus \ldots \oplus C_{h_t}^{-1} \end{aligned}$$

Using Property 3 of Lemma 2.12, we realise that $C_{h_j}^{-1} \approx C_{\tilde{h_j}}$ and it follows that $A^{-1} = C_{\tilde{h_1}} \oplus C_{\tilde{h_2}} \oplus \ldots \oplus C_{\tilde{h_t}}$. The invariant factors of A^{-1} are thus $\tilde{h}_1(x) \mid \tilde{h}_2(x) \mid \ldots \mid \tilde{h}_t(x)$, since $h_i(x) \mid h_{i+1}(x)$ if and only if $\tilde{h}_i(x) \mid \tilde{h}_{i+1}(x)$. Since $A \approx A^{-1}$ it follows immediately that $\tilde{h}_j(x) = h_j(x)$, $j = 1, 2, \ldots, t$.

It then will suffice to prove the Lemma for A of the form $C_{m(x)}$, where $\tilde{m}(x) = m(x)$. We write $m(x)$ as the product $\prod((p_i(x))^{n_i})$ of powers of distinct monic irreducible polynomials, $p_i(x)$. Also $\tilde{m}(x) = \prod((p_i(x))^{n_i})$. Since $m(x) = \tilde{m}(x)$ it follows from the uniqueness of such factorizations that the factor $f(x)^n$ in $m(x)$ is either equal to the corresponding $\tilde{f}(x)^n$ in $\tilde{m}(x)$ or the factors $f(x)^n$ and $\tilde{f}(x)^n$ are distinct. The symmetric factors are denoted by $(p_i(x))^{n_i}$ and those that form distinct pairs are denoted by $((q_i(x))^{m_i}, (\tilde{q}_i(x))^{m_i})$. Thus we can write

$$m(x) = (p_1(x))^{n_1} \ldots (p_r(x))^{n_r} \cdot (q_1(x))^{m_1} (\tilde{q}_1(x))^{m_1} \ldots (q_s(x))^{m_s} (\tilde{q}_s(x))^{m_s}$$

where the $p_i(x)$, $q_i(x)$, $\tilde{q}_i(x)$ are distinct irreducible monic polynomials with the p_i's symmetric.

By the Primary Decomposition Theorem A is similar to the direct sum

$$C_{p_1^{n_1}} \oplus C_{p_2^{n_2}} \oplus \ldots \oplus C_{p_r^{n_r}} \oplus C_{q_1^{m_1}} \oplus C_{\tilde{q_1}^{m_1}} \oplus \ldots \oplus C_{q_s^{m_s}} \oplus C_{\tilde{q_s}^{m_s}}.$$

Using the fact that $C_{g_i}^{-1} \approx C_(\tilde{g}_i)$ where $g_i = q_i^{m_i}$ we find that,

$$\begin{aligned} A &\approx C_{p_1^{n_1}} \oplus C_{p_2^{n_2}} \oplus \ldots_{p_r^{n_r}} \oplus C_{q_1^{m_1}} \oplus (C_{q_1^{m_1}})^{-1} \oplus \ldots \oplus C_{q_s^{m_s}} \oplus (C_{q_s^{m_s}})^{-1} \\ &= C_{p_1^{n_1}} \oplus C_{p_2^{n_2}} \oplus \ldots \oplus C_{p_r^{n_r}} \oplus C_{g_1} \oplus C_{g_1}^{-1} \oplus \ldots \oplus C_{g_s} \oplus C_{g_s}^{-1}. \end{aligned}$$

□

Lemma 2.14 [13, Lemma 2]

Let A be a matrix over a field F. Then A is the product of two involutions if and only if there is an involution P with $PAP = A^{-1}$.

Proof. (⇒) Let $A = ST$ where S and T are involutions, that is $S^2 = T^2 = I$ so that $S = S^{-1}$ and $T = T^{-1}$. We take P to be equal to T. It follows that

$$PAP = T(ST)T = TST^2 = TS = T^{-1}S^{-1} = (ST)^{-1} = A^{-1}.$$

(⇐) Let $PAP = A^{-1}$, with P being an involution, i.e. $P^2 = I$. Then $A = I \cdot A = P^2 A = P(PA)$ and both P and PA are involutions . The matrix PA is an involution, since

$$(PA)^2 = PAPA = (PAP)A = A^{-1}A = I.$$

□

Lemma 2.15 [13, Lemma 3]

Let A be a $n \times n$ matrix over a field F, of the form $\begin{pmatrix} B & 0 \\ 0 & B^{-1} \end{pmatrix}$, where B is $\frac{n}{2} \times \frac{n}{2}$ and n is even. Then A is the product of two involutions.

Proof. Let

$$P = \begin{pmatrix} 0 & I_{\frac{n}{2}} \\ I_{\frac{n}{2}} & 0 \end{pmatrix}$$

then

$$P^2 = \begin{pmatrix} 0 & I_{\frac{n}{2}} \\ I_{\frac{n}{2}} & 0 \end{pmatrix} \begin{pmatrix} 0 & I_{\frac{n}{2}} \\ I_{\frac{n}{2}} & 0 \end{pmatrix} = I_n$$

and P is an involution. The matrix

$$PAP = \begin{pmatrix} 0 & I_{\frac{n}{2}} \\ I_{\frac{n}{2}} & 0 \end{pmatrix} \begin{pmatrix} B & 0 \\ 0 & B^{-1} \end{pmatrix} \begin{pmatrix} 0 & I_{\frac{n}{2}} \\ I_{\frac{n}{2}} & 0 \end{pmatrix} = \begin{pmatrix} B^{-1} & 0 \\ 0 & B \end{pmatrix} = A^{-1}.$$

Since $PAP = A^{-1}$ and P is an involution, it follows from Lemma 2.14 that A is a product of two involutions. □

Lemma 2.16 [13, Lemma 4]

If A is the companion matrix for the monic symmetric polynomial $f(x) = [p(x)]^k$ with $p(x)$ irreducible over F, then A is the product of two involutions.

Proof. As was shown in Lemma 2.14 it is sufficient to find a matrix P, satisfying $P^2 = I$ and $PAP = A^{-1}$. We take P to be the involution,

$$P = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Let $f(x) = [p(x)]^k$ with $p(x)$ irreducible over F and $f(x)$ symmetric and monic. If

$$f(x) = x^n - a_{n-1}x^{n-1} - a_{n-2}x^{n-2} - \dots - a_2x^2 - a_1x + a$$

then

$$\tilde{f}(x) = \left(\frac{x^n}{a}\right) f\left(\frac{1}{x}\right) = x^n - \frac{a_1}{a}x^{n-1} - \frac{a_2}{a}x^{n-2} - \dots - \frac{a_{n-1}}{a}x + \frac{1}{a}.$$

The equality $f(x) = \tilde{f}(x)$ implies that $\frac{1}{a} = a$ i.e. $a^2 = 1$ and $\frac{a_{n-k}}{a} = a_k$. When n is even and $a = -1$ then $a_{\frac{n}{2}} = \frac{a_{\frac{n}{2}}}{a}$ and $a_{\frac{n}{2}} = 0$. If $a = 1$ no condition is imposed upon $a_{\frac{n}{2}}$.

Using the fact that

$$f(x) = x^n - \left(-a + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1}\right)$$

and $a_{n-k} = a \cdot a_k$, we wish to establish a special structure for A that will ensure that $PAP = A^{-1}$. The matrix A is such that $A = C_{\tilde{f}} = C_f$ and

$$A = \begin{pmatrix} 0 & 0 & \dots & 0 & -a \\ 1 & 0 & \dots & 0 & a_1 \\ 0 & 1 & \dots & 0 & a_2 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & a_{n-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 0 & -a \\ 1 & 0 & \dots & 0 & \frac{a_{n-1}}{a} \\ 0 & 1 & \dots & 0 & \frac{a_{n-2}}{a} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & a \cdot a_1 \end{pmatrix}.$$

Considering the last column of A, we notice that

$$a_{i,n} = \begin{cases} -a & \text{if } i = 1, \\ \frac{a_{n-i+1}}{a} & \text{if } 2 \le i \le \left[\frac{n}{2}\right], \\ a \cdot a_{n-i+1} & \text{if } i > \left[\frac{n}{2}\right]. \end{cases}$$

The inverse of A is given by:

$$A^{-1} = \begin{pmatrix} a_{n-1} & 1 & 0 & \dots & 0 \\ a_{n-2} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_1 & 0 & 0 & \dots & 1 \\ -a & 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} a \cdot a_1 & 1 & 0 & \dots & 0 \\ a \cdot a_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \frac{a_{n-1}}{a} & 0 & 0 & \dots & 1 \\ -a & 0 & 0 & \dots & 0 \end{pmatrix}$$

Considering the first column of A^{-1}, we notice that

$$a_{i,1} = \begin{cases} a \cdot a_i & \text{if } i \leq \left[\frac{n}{2}\right], \\ \frac{a_i}{a} & \text{if } \left[\frac{n}{2}+1\right] \leq i \leq n-1, \\ -a & \text{if } i = n. \end{cases}$$

We find that

$$\begin{aligned} PAP &= \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \dots & 0 & 0 & -a \\ 1 & \dots & 0 & 0 & a_1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & a_{n-2} \\ 0 & \dots & 0 & 1 & a_{n-1} \end{pmatrix} \\ &\times \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} a_{n-1} & 1 & 0 & \dots & 0 \\ a_{n-2} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_1 & 0 & 0 & \dots & 1 \\ -a & 0 & 0 & \dots & 0 \end{pmatrix} \\ &= A^{-1} \end{aligned}$$

and since P is an involution, it follows from Lemma 2.14 that A is a product of two involutions. This concludes the proof. □

Theorem 2.17 [13, Theorem 1]

Let A be an invertible $n \times n$ matrix ($n > 1$) over a field F; then A is the product of two involutions if and only if A is similar to A^{-1}.

Proof. ($\Leftarrow$) Assume that $A \approx A^{-1}$, then from Lemma 2.13,

$$A \approx B_1 \oplus B_2 \oplus \ldots \oplus B_r \oplus D_1 \oplus \ldots \oplus D_s$$

where $B_i = C_{f_i}$,

$$D_i = \begin{pmatrix} C_{g_i} & 0 \\ 0 & C_{g_i}^{-1} \end{pmatrix}$$

with each $f_i(x) = (p_i(x))^{n_i}$ for some symmetric irreducible $p_i(x)$; each $g_i(x)$ a power of an irreducible polynomial. According to Lemma 2.15 the D_i's have the form

$$\begin{pmatrix} B & 0 \\ 0 & B^{-1} \end{pmatrix}$$

and is a product of two involutions. All the B_i's are companion matrices of the form C_{f_i} where $f_i(x) = (p_i(x))^{n_i}$ and the $f_i(x)$'s are symmetric since the $p_i(x)$'s are symmetric (Lemma 2.13). It thus follows from Lemma 2.16 that all the B_i's are products of two involutions. Since all the B_i's and D_i's are products of two involutions it follows that

$$A \approx B_1 \oplus B_2 \oplus \ldots \oplus B_r \oplus D_1 \oplus \ldots \oplus D_s$$

is a product of two involutions.

($\Rightarrow$) Let A be a product of two involutions. Then by Lemma 2.14 there is an involution P such that $PAP = A^{-1}$ and since P is an involution, $P = P^{-1}$. Thus $PAP^{-1} = A^{-1}$ and $A \approx A^{-1}$. □

Lemma 2.18

The following matrices are products of two involutions:

(i) The permutation matrix P of an n-cycle, i.e. the companion matrix of $x^n - 1$

(ii) The companion matrix of $x^n + (-1)^n$

(iii) αP where $\alpha^n = 1$

(iv) A unipotent matrix

Proof.

(i) The companion matrix of $x^n - 1$ is

$$P = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

and

$$P^{-1} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & \ldots & 0 \end{pmatrix}.$$

We choose a matrix R such that

$$R^{-1} = R = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ & & \iddots & & \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}$$

and

$$\begin{aligned} RPR &= \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ & & \iddots & & \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \\ &\times \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ & & \iddots & & 0 \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix} \\ &= P^{-1}. \end{aligned}$$

The matrix P is similar to P^{-1} and it follows from Theorem 2.17 that P is a product of two involutions.

(ii) The companion matrix of $x^n + (-1)^n$ is

$$A = \begin{pmatrix} 0 & 0 & \dots & 0 & (-1)^{n+1} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

Using matrix R as in (i), we find that

$$RAR^{-1} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ (-1)^{n+1} & 0 & 0 & \dots & 0 \end{pmatrix} = A^{-1}.$$

The matrix A is similar to A^{-1} and it follows from Theorem 2.17 that A is a product of two involutions.

(iii) The matrix αP is such that

$$\alpha P = \begin{pmatrix} 0 & 0 & \dots & 0 & \alpha \\ \alpha & 0 & \dots & 0 & 0 \\ 0 & \alpha & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \alpha & 0 \end{pmatrix}.$$

The characteristic polynomial of αP is

$$\begin{aligned} |xI - \alpha P| &= \det \begin{pmatrix} x & 0 & \dots & 0 & -\alpha \\ -\alpha & x & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & -\alpha & x & 0 \\ 0 & 0 & \dots & -\alpha & x \end{pmatrix} \\ &= x \cdot \det \begin{pmatrix} x & 0 & \dots & 0 \\ -\alpha & x & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -\alpha & x \end{pmatrix} \\ &+ (-1)^{n+1}(-\alpha)\det \begin{pmatrix} -\alpha & x & \dots & 0 \\ 0 & -\alpha & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\alpha \end{pmatrix} \\ &= x \cdot x^{n-1} + (-1)^{n+1}(-\alpha)(-\alpha)^{n-1} \\ &= x^n + (-1)^{2n+1}\alpha^n = x^n - \alpha^n. \end{aligned}$$

The $(n-1) \times (n-1)$ matrices $\begin{pmatrix} x & 0 & \dots & 0 \\ -\alpha & x & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -\alpha & x \end{pmatrix}$ and

$\begin{pmatrix} -\alpha & x & \dots & 0 \\ 0 & -\alpha & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\alpha \end{pmatrix}$ are lower and upper triangular matrices respectively and we find that their determinants are just the products of their diagonals i.e. x^{n-1} and $(-\alpha)^{n-1}$. Furthermore $|xI - \alpha P| = x^n - 1$, since $\alpha^n = 1$. The matrix αP is similar to P, the companion matrix of $x^n - 1$, since there exists a matrix X such that:

$$X = \begin{pmatrix} 0 & \alpha & 0 & \dots & 0 & 0 \\ 0 & 0 & \alpha^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \alpha^{n-2} & 0 \\ 0 & 0 & 0 & \dots & 0 & \alpha^{n-1} \\ \alpha^n & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

The inverse

$$X^{-1} = \begin{pmatrix} 0 & 0 & \dots & 0 & \frac{1}{\alpha^n} \\ \frac{1}{\alpha} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{\alpha^2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1}{\alpha^{n-1}} & 0 \end{pmatrix},$$

$\alpha^n = 1$ and $X^{-1}(\alpha P)X = P$. We have shown in (i) that the companion matrix of $x^n - 1$ is a product of two involutions and thus αP is also a product of two involutions.

An alternative proof is through the use of Smith canonical matrices. According to [5, Theorem 6.12] every non-zero square matrix is equivalent to a Smith canonical matrix which has the form:

$$S(x) = \text{Diag}[f_1(x), f_2(x), \dots, f_r(x), 0],$$

where r is the rank of the original non-zero square matrix, each $f_i(x)$ is monic, and $f_i(x)$ divides $f_{i+1}(x)$, $i = 1, 2, \dots, r-1$. Since $(xI - \alpha P)$ has rank n it follows from [5, Theorem 6.17] that the Smith Canonical matrix equivalent to $(xI - \alpha P)$ is of the form:

$$S(x) = \text{Diag}[f_1(x), f_2(x), \dots, f_n(x)].$$

From the proof of [5, Theorem 6.17], $f_n(x) = \frac{\det(xI - \alpha P)}{d_{n-1}(x)}$, where $d_{n-1}(x)$ is equal to GCD$\{n-1 \times n-1$ subdeterminant of $xI - \alpha P\}$. The determinant of a submatrix is called a subdeterminant.

We can easily see that $d_{n-1} = 1$, since we are able to find at least two sub-matrices of $xI - \alpha P$ whose determinants have a greatest common divisor of 1. It follows from [5, Theorem 6.17] that the minimal polynomial $m(x)$ of $xI - \alpha P$ is such that:

$$\begin{aligned} m(x) = f_n &= \frac{\det(xI - \alpha P)}{d_{n-1}(x)} \\ &= \frac{\det(xI - \alpha P)}{1} = x^n - 1. \end{aligned}$$

Since the minimal polynomial of αP is identical to its characteristic polynomial it follows that αP is similar to the companion matrix of its characteristic polynomial $x^n - 1$. We have shown in (i) that the companion matrix of $x^n - 1$ is a product of two involutions and thus αP is also a product of two involutions.

(iv) A unipotent matrix A is such that its characteristic polynomial is a power of $(x - 1)$. Since A is an $n \times n$ matrix, its characteristic polynomial is $(x - 1)^n$ and $m \leq n$ is minimal with respect to $(A - I)^m = 0$. The largest invariant factor of A which is also the minimal polynomial is $(x-1)^m$. The other invariant factors of A divides $(x-1)^m$ and since $x - 1$ is irreducible it has the form $(x - 1)^i$ for $i \leq m$. The matrix A is similar to a direct sum of the companion matrices of its invariant factors.

Any polynomial of the form $p(x) = (x - 1)^s$ is symmetric since

$$\begin{aligned} \tilde{p}(x) &= \frac{x^s}{(-1)^s} p\left(\frac{1}{x}\right) = \frac{x^s}{(-1)^s}\left(\frac{1}{x} - 1\right)^s \\ &= (-1)^s \left[x\left(\frac{1}{x} - 1\right)\right]^s = (-1)^s(1 - x)^s \\ &= [-1(1 - x)]^s = (x - 1)^s = p(x). \end{aligned}$$

Furthermore since $x - 1$ is irreducible and a polynomial of the form $(x - 1)^s$ is symmetric it follows from Lemma 2.16, that the companion matrices of the invariant factors of A are each a product of two involutions and thus A is also a product of two involutions.

□

2.5 Applications to Involutory Factorization

2.5.1 Sourour

For an invertible $n \times n$ matrix A over F, where F has at least $n+2$ elements and $\det A = \pm 1$, into how many factors of involutions can A be expressed? The answer is at most four and this is how Sourour shows it.

Theorem 2.19 [24, Theorem 5]

Let A be a $n \times n$ matrix over the field F containing at least $n+2$ elements. If $\det A = \pm 1$ then A is a product of at most four involutions.

Proof. **A non-scalar**

In Sourour's proof of involutions he uses some of the arguments in the Shoda-Thompson theorem. He assumes implicitly that A is non-scalar since he applies his Factorization theorem.

det $A = 1$

As in the proof of Theorem 2.8 (b), we may write A as a product BC. If n is even, B and C have distinct eigenvalues of the form $\{\beta_1, \beta_1^{-1}, \ldots, \beta_m, \beta_m^{-1}\}$ and if n is odd B and C have eigenvalues of the form $\{1, \beta_1, \beta_1^{-1}, \ldots, \beta_m, \beta_m^{-1}\}$ where $m = \left[\frac{n}{2}\right]$. The notation $[x]$ denotes a floor function that returns the greatest integer less than or equal to the real number x. The choice of the eigenvalues for B and C follows from Theorem 2.8 (b) for $|F| \geq n+2$ and $\det A = 1$.

Matrices B and C are diagonalizable since they have distinct eigenvalues and it suffices to show that $\begin{pmatrix} \beta & 0 \\ 0 & \beta^{-1} \end{pmatrix}$ is a product of 2 involutions. The matrix $\begin{pmatrix} \beta & 0 \\ 0 & \beta^{-1} \end{pmatrix}$ is indeed a product of 2 involutions, since

$$\begin{pmatrix} \beta & 0 \\ 0 & \beta^{-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & \beta^{-1} \\ \beta & 0 \end{pmatrix}.$$

det $A = -1$

Assume that $-1 \neq 1$.

<u>n is odd</u>

It follows that $\det -A = 1$. Using the first part we see that $-A = BC$ where B and C are each a product of 2 involutions and so the result follows since the negative of an involution is again an involution.

<u>n is even</u>

$\det A = -1$ and $1 \neq -1$. It is possible to choose distinct eigenvalues

$$\{1, -1, \beta_2, \beta_2^{-1}, \ldots, \beta_{\frac{n}{2}}, \beta_{\frac{n}{2}}^{-1}\}$$

for B. The eigenvalues of C remains

$$\{\beta_1, \beta_1^{-1}, \beta_2, \beta_2^{-1}, \ldots, \beta_{\frac{n}{2}}, \beta_{\frac{n}{2}}^{-1}\}.$$

The distinct choices of β_i with the exclusion of $\beta_i \neq 0, \pm 1$ are possible for $|F| \geq n+2$. If $|F| = n+2$, it follows that F is an extension of the $Z/2Z$ field and has characteristic 2. This means that $1+1=0$ or $1=-1$, but in this case we specified that $1 \neq -1$ and thus the distinct choices of β_i are only possible with $|F|$ at least $n+3$. The eigenvalues 1 and -1 of matrix B are responsible for the direct summand

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

which is itself an involution.

<u>A **scalar**</u>

n **even**

<u>$\det A = -1$</u>

Let $B = \text{Diag}(\alpha, \alpha^3, \alpha^5, \ldots, \alpha^{n-1})$ and $E = \text{Diag}(-1, \alpha^{n-2}, \alpha^{n-4}, \ldots, \alpha^2)$, then $A = (B \oplus -B)(-E \oplus E)$. We notice that $\det(B \oplus -B) = 1$ and that $\det(-E \oplus E) = -1$.

$$B \oplus -B = \begin{pmatrix} \alpha & & & & & & & \\ & \alpha^3 & & & & & & \\ & & \ddots & & & & & \\ & & & \alpha^{n-1} & & & & \\ & & & & -\alpha & & & \\ & & & & & -\alpha^3 & & \\ & & & & & & \ddots & \\ & & & & & & & -\alpha^{n-1} \end{pmatrix}.$$

The matrix $B \oplus -B$ has $\frac{n}{2}$ pairs $(-\alpha^{2k+1}, \alpha^{n-(2k+1)})$, $0 \leq k \leq \frac{n}{2} - 1$. The eigenvalues of $B \oplus -B$ are the same as its diagonal entries i.e.

$$\text{Eig}(B \oplus -B) = \{-\alpha, \alpha^{n-1}, -\alpha^3, \alpha^{n-3}, \ldots, -\alpha^{n-1}, \alpha\}.$$

The matrix $B \oplus -B$ is similar to a matrix F such that

$$F = \begin{pmatrix} \alpha & & & & & & & \\ & \alpha^3 & & & & & & \\ & & \ddots & & & & & \\ & & & \alpha^{n-1} & & & & \\ & & & & -\alpha^{n-3} & & & \\ & & & & & -\alpha^{n-5} & & \\ & & & & & & \ddots & \\ & & & & & & & -\alpha \end{pmatrix}$$

and F has the form $\begin{pmatrix} X & 0 \\ 0 & X^{-1} \end{pmatrix}$, which was shown previously to be a product of two involutions.

It follows that

$$\begin{aligned} B \oplus -B \approx F &= \begin{pmatrix} \alpha & & & & & & & \\ & \alpha^3 & & & & & & \\ & & \ddots & & & & & \\ & & & \alpha^{n-1} & & & & \\ & & & & -\alpha^{n-3} & & & \\ & & & & & -\alpha^{n-5} & & \\ & & & & & & \ddots & \\ & & & & & & & -\alpha \end{pmatrix} \\ &= GH \end{aligned}$$

where

$$G = \begin{pmatrix} 0 & I_{\frac{n}{2}} \\ I_{\frac{n}{2}} & 0 \end{pmatrix}$$

and

$$H = \begin{pmatrix} & & & & -\alpha^{n-1} & & & \\ & & & & & -\alpha^{n-3} & & \\ & & & & & & \ddots & \\ & & & & & & & -\alpha \\ \alpha & & & & & & & \\ & \alpha^3 & & & & & & \\ & & \ddots & & & & & \\ & & & \alpha^{n-1} & & & & \end{pmatrix}.$$

Both G and H are involutions, since G^2 and H^2 are equal to I_n.

The eigenvalues of $-E \oplus E$ are

$$\mathrm{Eig}(-E \oplus E) = \{1, -1, -\alpha^{n-2}, \ldots, -\alpha^2, \alpha^2, \ldots, \alpha^{n-2}\}.$$

The matrix $-E \oplus E$ is similar to a matrix of the form $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \oplus R$, where R is $(n-2) \times (n-2)$ and

$$R = \begin{pmatrix} -\alpha^{n-2} & & & & & & & \\ & -\alpha^{n-4} & & & & & & \\ & & \ddots & & & & & \\ & & & -\alpha^4 & & & & \\ & & & & -\alpha^2 & & & \\ & & & & & \alpha^2 & & \\ & & & & & & \ddots & \\ & & & & & & & \alpha^{n-2} \end{pmatrix}.$$

The matrix R has the form $\begin{pmatrix} X & 0 \\ 0 & X^{-1} \end{pmatrix}$ which is a product of two involutions.

It follows that

$$\begin{aligned} -E \oplus E &\approx \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \oplus R \\ &= IJ \end{aligned}$$

where

$$I = I_2 \oplus \begin{pmatrix} 0 & I_{\frac{n-2}{2}} \\ I_{\frac{n-2}{2}} & 0 \end{pmatrix}$$

and

$$J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \oplus \begin{pmatrix} & & & & \alpha^2 & & & \\ & & & & & \alpha^4 & & \\ & & & & & & \ddots & \\ & & & & & & & \alpha^{n-2} \\ -\alpha^{n-2} & & & & & & & \\ & -\alpha^{n-4} & & & & & & \\ & & \ddots & & & & & \\ & & & -\alpha^2 & & & & \end{pmatrix}.$$

Both I and J are involutions since I^2 and J^2 are equal to I_n. The matrix $-E \oplus E$ is a product of two involutions and thus $A = \alpha I$, n even and

$\det A = -1$ is a product of four involutions.

$\underline{\det A = \alpha^n = 1}$

The matrix A can be expressed in the form $A = BC$ where

$$B = \begin{pmatrix} \alpha & & & \\ & \alpha^2 & & \\ & & \ddots & \\ & & & \alpha^n \end{pmatrix}$$

and

$$C = \begin{pmatrix} \alpha^n & & & \\ & \alpha^{n-1} & & \\ & & \ddots & \\ & & & \alpha \end{pmatrix}.$$

For B and C we have $\frac{n}{2} - 1$ pairs of the form $(\alpha^i, \alpha^{n-i}) = (\alpha^i, \alpha^{-i})$ for $1 \leq i < \frac{n}{2}$. The remaining diagonal entries of B or C are $\alpha^{\frac{n}{2}} = \pm 1$ and $\alpha^n = 1$. Matrices B and C are similar to a matrix of the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha^{\frac{n}{2}} & 0 & 0 \\ 0 & 0 & X & 0 \\ 0 & 0 & 0 & X^{-1} \end{pmatrix}$$

where $X = \mathrm{Diag}(\alpha, \alpha^2, \ldots, \alpha^{\frac{n}{2}-1})$ and $X^{-1} = \mathrm{Diag}(\alpha^{n-1}, \alpha^{n-2}, \ldots, \alpha^{\frac{n}{2}+1})$. We find that

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha^{\frac{n}{2}} & 0 & 0 \\ 0 & 0 & X & 0 \\ 0 & 0 & 0 & X^{-1} \end{pmatrix} = \left\{ I_2 \oplus \begin{pmatrix} 0 & I_{\frac{n-2}{2}} \\ I_{\frac{n-2}{2}} & 0 \end{pmatrix} \right\} \cdot \left\{ \begin{pmatrix} 1 & 0 \\ 0 & \alpha^{\frac{n}{2}} \end{pmatrix} \oplus \begin{pmatrix} 0 & X^{-1} \\ X & 0 \end{pmatrix} \right\}$$

a product of two involutions. It follows that A is a product of four involutions.

n **odd**

$\underline{\det A = \alpha^n = 1}$

The matrix A can be expressed in the form $A = BC$, where

$$B = \begin{pmatrix} \alpha & & & \\ & \alpha^2 & & \\ & & \ddots & \\ & & & \alpha^n \end{pmatrix}$$

and

$$C = \begin{pmatrix} \alpha^n & & & \\ & \alpha^{n-1} & & \\ & & \ddots & \\ & & & \alpha \end{pmatrix}.$$

For both B and C we have $\frac{n-1}{2}$ pairs $(\alpha^i, \alpha^{n-i}) = (\alpha^i, \alpha^{-i})$, $1 \leq i \leq \frac{n-1}{2}$ and one entry $\alpha^n = 1$. Both matrices B and C are similar to a matrix of the form $1 \oplus \begin{pmatrix} X & 0 \\ 0 & X^{-1} \end{pmatrix}$ and

$$1 \oplus \begin{pmatrix} X & 0 \\ 0 & X^{-1} \end{pmatrix} = \left\{ 1 \oplus \begin{pmatrix} 0 & I_{\frac{n-1}{2}} \\ I_{\frac{n-1}{2}} & 0 \end{pmatrix} \right\} \cdot \left\{ 1 \oplus \begin{pmatrix} 0 & X^{-1} \\ X & 0 \end{pmatrix} \right\}$$

a product of two involutions. The matrix $\begin{pmatrix} X & 0 \\ 0 & X^{-1} \end{pmatrix}$ can be chosen such that

$$\begin{pmatrix} X & 0 \\ 0 & X^{-1} \end{pmatrix} = \begin{pmatrix} \alpha & & & & & & \\ & \alpha^2 & & & & & \\ & & \ddots & & & & \\ & & & \alpha^{\frac{n-1}{2}} & & & \\ & & & & \alpha^{n-1} & & \\ & & & & & \ddots & \\ & & & & & & \alpha^{\frac{n+1}{2}} \end{pmatrix}.$$

It follows that if $A = \alpha I$, n is odd and $\det A = 1$, then A is the product of at most four involutions.

<u>$\det A = -1$</u>

The result follows from $A = \alpha I$ where $\det A = 1$, by considering the matrix $-A$. The negative of an involution is again an involution and thus A is the product of four involutions. □

2.5.2 Laffey

The main result of this section is to show that if $A \in GL(n,F)$ with $\det A = \pm 1$, then A is the product of four involutions. We also show that if $A \in SL(n,F)$ with $\text{char}(F) \neq 2$, then A is the product of four involutions all similar to $\tilde{J}$. The matrix $\tilde{J}$ or $\tilde{J}_n$ denotes the $n \times n$ matrix, $\text{Diag}(I_k, -I_{n-k})$ where $k = \left[\frac{n+1}{2}\right]$.

Lemma 2.20 [18, Lemma 5.2]

Let $A \in SL(n,F)$ with $\text{char}(F) \neq 2$ be the product PQ where P, Q are involutions and suppose A does not have an eigenvalue equal to ± 1. Then n is even and P, Q are similar to $\tilde{J}$.

Proof. Since $\det A = 1$ and matrix A does not have eigenvalues equal to ± 1, it follows that the eigenvalues of A occur in pairs of the form (a, a^{-1}), where a, $a^{-1} \neq \pm 1$ and $a \cdot a^{-1} = 1$. This conclusion about the eigenvalues are drawn from Theorem 2.5, whereby A can be written as a product of a lower triangular matrix and an upper triangular matrix such that $A = FG$ and $G \approx F^{-1}$.

Assume that P or Q are not similar to $\tilde{J}_n$.

The minimal polynomial of an involution is $x^2 - 1 = (x+1)(x-1)$ and since the minimal polynomial is a product of distinct linear factors it follows that P and Q are diagonalizable. Since P and Q are diagonalizable with eigenvalues ± 1, it follows that

$$F^n = E_1(P) \oplus E_{-1}(P) = E_1(Q) \oplus E_{-1}(Q)$$

where $E_\lambda(P)$ denotes the eigenspace P corresponding to λ. Thus

$$n = \dim(E_1(P)) + \dim(E_{-1}(P)) = \dim(E_1(Q)) + \dim(E_{-1}(Q)).$$

If $\dim(E_1(P)) > \frac{n}{2}$ (when P is not similar to $\tilde{J}_n$) it therefore follows that $E_1(P)$ must intersect one of $E_1(Q)$ and $E_{-1}(Q)$. Matrix P and Q would have a common eigenvector, which leads to a contradiction as explained in the Remark that is to follow. A similar argument holds if $\dim(E_{-1}(P)) > \frac{n}{2}$. Thus

$$\dim(E_1(P)) = \dim(E_{-1}(P)) = \dim(E_1(Q)) = \dim(E_{-1}(Q)) = \frac{n}{2},$$

and it follows that n is even and $P \approx Q \approx \tilde{J}_n$. $\square$

Remark: If λ is an eigenvalue of P, and μ an eigenvalue of Q, both corresponding to the same eigenvector v, then

$$Av = (PQ)v = P(Qv) = P(\mu v) = \mu(Pv) = \mu\lambda v$$

and $\mu\lambda$ is an eigenvalue of A. The eigenvalues of P and Q are 1 and -1 and therefore A can have an eigenvalue equal to ± 1, which is a contradiction.

The following is an adaptation of the results by Laffey due to Botha [3].

Lemma 2.21

$J_n(1) \in M_n(F)$, F a field with $\mathrm{char}(F) \neq 2$, can be expressed as a product $J_n(1) = P_n Q_n$ of involutions both similar to $\tilde{J}_n$.

Proof. Let $P_1 = Q_1 = [1]$ and for $k \geq 1$ define P_{k+1} and Q_{k+1} inductively as

$$P_{k+1} = \begin{pmatrix} 1 & 0 \\ e_1 & -Q_k \end{pmatrix}$$

and

$$Q_{k+1} = P_{k+1} J_{k+1}(1)$$

where e_1 denotes the first canonical basis vector in F^k. We begin by proving by induction that for all $k \geq 1$,

$$P_k^2 = Q_k^2 = I_k, P_k Q_k = J_k(1), Q_k e_1 = e_1 \tag{2.4}$$

and the diagonal entries of P_k and Q_k are both of the form $(1, -1, 1, -1, \ldots)$. The result is true for $k = 1$. Suppose the result holds for $k \geq 1$, then

$$P_{k+1}^2 = \begin{pmatrix} 1 & 0 \\ e_1 - Q_k e_1 & Q_k^2 \end{pmatrix} = I_{k+1}$$

since $Q_k^2 = I_k$ and $Q_k e_1 = e_1$ and hence

$$P_{k+1} Q_{k+1} = P_{k+1}^2 J_{k+1}(1) = J_{k+1}(1).$$

Further

$$Q_{k+1} = P_{k+1} J_{k+1}(1) = \begin{pmatrix} 1 & 0 \\ e_1 & -Q_k \end{pmatrix} \begin{pmatrix} 1 & 0 \\ e_1 & J_k(1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -Q_k J_k(1) \end{pmatrix}$$

hence $Q_{k+1} e_1 = e_1$ and

$$Q_{k+1}^2 = \begin{pmatrix} 1 & 0 \\ 0 & Q_k J_k(1) Q_k J_k(1) \end{pmatrix}.$$

Since

$$Q_k = P_k J_k(1) \Rightarrow P_k = J_k(1) Q_k$$

it follows that

$$Q_k J_k(1) Q_k J_k(1) = Q_k P_k J_k(1) = Q_k^2 = I_k$$

therefore $Q_{k+1}^2 = I_{k+1}$. It follows from the definition of P_{k+1} that its diagonal entries are of the required form and since

$$(P_{k+1})_{ii}(Q_{k+1})_{ii} = (J_{k+1}(1))_{ii} = 1$$

for each $1 \leq i \leq k+1$ the same applies to Q_{k+1}. Hence (2.4) holds for all $k \geq 1$.

P_{k+1} and Q_{k+1} are diagonalizable since they are involutions and since they are also lower triangular it follows that they are similar to their diagonals, hence

$$P_{k+1} \approx Q_{k+1} \approx \tilde{J}_{k+1}.$$

Laffey's proof of the next theorem as it appears in [18, Theorem 5.1] has been adapted for the more general case and is due to Botha [3]. □

Theorem 2.22 [18, Theorem 5.1]

Let $A \in SL(n, F)$ with $\text{char}(F) \neq 2$ be similar to its inverse. Then $A = PQ$ where P,Q are involutions both similar to $\tilde{J}_n$.

Proof. Using a similarity we may assume that

$$A = (\oplus_{i=1}^{k} J_{k_i}(1)) \oplus (\oplus_{i=1}^{l} J_{l_i}(-1)) \oplus B$$

where $k \geq 0$, $l \geq 0$ and B has no eigenvalue ± 1. As in the proof of Laffey it follows that the order m of B is even and that $B = P_B Q_B$ where P_B and Q_B are both similar to $\tilde{J}_m$ (i.e. with equal number of diagonal entries 1 and -1).

Note that $J_{l_i}(-1) \approx -J_{l_i}(1)$, hence $J_{l_i}(-1) = (-P_{l_i})Q_{l_i}$ where $P_{l_i} \approx Q_{l_i} \approx \tilde{J}_{l_i}$. If l_i is even, then $-P_{l_i} \approx P_{l_i}$ since $\tilde{J}_{l_i}$ has equal number of entries 1 and -1 which yields the required factorization in this case. Since $\det A = 1$, the number of odd l_i must be even. Expressing every second odd order summand as $J_{l_i}(-1) = P_{l_i}(-Q_{l_i})$, it follows that the combination of these factorizations yield the required factorization for $(\oplus_{i=1}^{l} J_{l_i}(-1))$ and that $(\oplus_{i=1}^{l} J_{l_i}(-1))$ has even order. Hence $(\oplus_{i=1}^{l} J_{l_i}(-1)) \oplus B$ can be expressed as a product of two involutions each similar to $\tilde{J}$ (and with an equal number of diagonal entries 1 and -1).

Finally for k_i even, use the factorization in the lemma. For odd orders k_i, express every second occurance as $J_{k_i}(1) = (-P_{k_i})(-Q_{k_i})$. This yields the required factorization for $(\oplus_{i=1}^{k} J_{k_i}(1))$ and A as a whole. □

Theorem 2.23 [18, Theorem 5.6]

(a) If $A \in GL(n, F)$ with $\det A = \pm 1$, then A is a product of four involutions.

(b) If $A \in SL(n, F)$ and F does not have characteristic 2, then A is a product of four involutions, all similar to $\tilde{J}$.

Proof.

(a) **A non-scalar**

If A is non-scalar, [18, Theorem 5.3] implies that $A \approx LU$ where $\det L = \pm 1$ and $\det U = 1$. Both L and L^{-1} have the same diagonal entries (eigenvalues) and eigenvectors and are thus similar. From Lemma 2.14 it follows that L is a product of two involutions.

The matrix U is an upper triangular matrix with all its diagonal entries equal to 1 and is a unipotent matrix. In Lemma 2.18 we showed that a unipotent matrix is a product of two involutions and therefore U is a product of two involutions. It follows that A is a product of four involutions.

A scalar

Let $A = \alpha I$. Then $A = (\alpha P)P^{-1}$ where P is the permutation matrix of the n-cycle. It was shown in Lemma 2.18 that both αP and P^{-1} are similar to their inverses and can be expressed as a product of 2 involutions, giving the result.

(b) **A non-scalar**

By [18, Theorem 5.3], $A \approx LU$ where $\det L = 1$ and $\det U = 1$. All the diagonal entries of L and U are equal to 1 and L and U are unipotent matrices. Both L and U are products of two involutions according to Lemma 2.18 and it follows that A is a product of four involutions.

A scalar

n odd

It follows from the argument in (a) that $A = (\alpha P)P^{-1}$ where both αP and P^{-1} are similar to their inverses and a product of two involutions. Matrix A is thus a product of four involutions.

n even

We note that

$$A = [\alpha \text{Diag}(-1, 1, \dots, 1)P][\text{Diag}(-1, 1, \dots, 1)P]^{-1}$$

which is equivalent to replacing the 1 in the first row of P with a -1. Here,

$$\det(\text{Diag}(-1, 1, \dots, 1)P) = \det(\text{Diag}(-1, 1, \dots, 1)P)^{-1} = 1.$$

This ensures that $\det P = \det P^{-1} = 1$. If this change is not made we will have a situation whereby

$$\det \alpha P = \det \alpha I \cdot \det P = \det P = \det P^{-1} = -1$$

and it would not be possible to have αP, $P^{-1} \approx RQ$, where R,$Q \approx \tilde{J}$ and $\det RQ = 1$ as is the requirement in Theorem 2.22. The matrix P changed in this way is still similar to its inverse, since $S^{-1}PS = P^{-1}$ and

$$S = \begin{pmatrix} 0 & I_{\frac{n}{2}} \\ I_{\frac{n}{2}} & 0 \end{pmatrix}.$$

By the same argument as n odd, A is a product of 4 involutions. □

Chapter 3

Spectral Factorization of Singular Matrices

3.1 Products of Two Nilpotent Matrices

In this section we will use the paper by Wu [29] to prove the result when the matrix to be factored is itself nilpotent and then use Laffey's paper [17] to prove the general result which is based on spectral factorization. Wu assumes in his paper [29] that all matrices are complex. Laffey claims that Wu's factorization of nilpotent matrices in terms of nilpotent factors in [29, Lemma 1 - 3] are valid over any field. The reason is that since the characteristic polynomial of the nilpotent matrix A is x^n, which trivially splits over any field, a nilpotent matrix A is similar to its Jordan Canonical form J over any field and the factorizations of J that Wu uses only involve the elements 0 and ± 1, which are elements of all fields.

3.1.1 Wu

We will first consider the 3 lemmas that are important in proving the main theorem.

Lemma 3.1 [29, Lemma 1]

The 2×2 matrix $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ is not a product of two nilpotent matrices.

Proof. Any 2×2 nilpotent matrix A is one of the following forms:

$a\begin{pmatrix} 1 & x \\ \frac{-1}{x} & -1 \end{pmatrix}$ $(x \neq 0)$, $b\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $c\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Any product of two of these matrices gives a matrix other than $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and therefore $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$

is not factorizable into two nilpotent matrices. The three forms are derived as follows:

The determinant and trace of a 2×2 nilpotent matrix

$$A = \begin{pmatrix} w & r \\ s & z \end{pmatrix}$$

is zero. From the determinant it follows that $wz - rs = 0$ and from the trace that $w + z = 0$ or $w = -z = a$. A 2×2 nilpotent is thus of the form:

$$A = \begin{pmatrix} a & r \\ s & -a \end{pmatrix}$$

(i) $a \neq 0$.

Since $\det A = -a^2 - rs = 0$ it follows that $rs = -a^2 \neq 0$, hence $r \neq 0$ and $s \neq 0$. Choose $r = ax$ with $x \neq 0$. Then $s = \frac{-a}{x}$ and therefore A is of the form $a \begin{pmatrix} 1 & x \\ \frac{-1}{x} & -1 \end{pmatrix}$.

(ii) $a = 0$.

Then $\det A = -rs = 0$, so $r = 0$ or $s = 0$, hence

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

or

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

□

Lemma 3.2 [29, Lemma 2]

For any $n \neq 2$, the $n \times n$ matrix

$$J = \begin{pmatrix} 0 & \dots & \dots & 0 \\ 1 & \ddots & \ddots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix}$$

is the product of two nilpotent matrices with ranks equal to rank J.

Proof. For odd n we have that

$$J = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}.$$

Alternatively J can also be expressed as

$$J = (E_{3,1} + E_{4,2} + \dots + E_{n,n-2} + E_{2,n})(E_{1,2} + E_{2,3} + \dots + E_{n-2,n-1} + E_{n,1}).$$

The previous approach for n odd does not work for n even. If for example $n = 6$ we have that

$$J_6 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The second factor is nilpotent, but the first factor is not and has a characteristic polynomial of $x^6 - x^3$.

Instead, Wu uses the following approach for n even:

$$J = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 \\ 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}.$$

To understand the factorization when n is odd it is instructive to note that the n-cycle $\begin{pmatrix} 1 & 2 & 3 & \dots & n \end{pmatrix}$ is the product of two cycles namely

$$\begin{pmatrix} 1 & 3 & 5 & \dots & n & 2 & 4 & \dots & n-1 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & n & n-1 & \dots & 3 & 2 \end{pmatrix}.$$

The product of the permutation matrices of the n-cycles can be described as follows:

$$BC = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

where

$$B = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \end{pmatrix}$$

and

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}.$$

By changing 1 in position $(1, n-1)$ to 0 in matrix B the permutation matrix of

$$\begin{pmatrix} 1 & 3 & 5 & \dots & n & 2 & 4 & \dots & n-1 \end{pmatrix}$$

and by changing 1 in position $(n-1, n)$ to 0 in matrix C the permutation matrix of $\begin{pmatrix} 1 & n & n-1 & \dots & 3 & 2 \end{pmatrix}$ the desired factors of J are obtained. The two modified permutation matrices both have a characteristic polynomial equal to x^n, which shows that all their eigenvalues are zero and hence they are nilpotent. □

Lemma 3.3 [29, Lemma 3]

Any $n \times n$ ($n \neq 2$) nilpotent matrix A is the product of two nilpotent matrices with ranks equal to rank A.

Proof. Any square complex matrix including a complex nilpotent matrix is similar to a matrix in Jordan canonical form [31, Theorem 3.12]. This is also true for a nilpotent matrix in any field since its characteristic polynomial is x^n, which spilts trivially over any field [5, Theorem 5.12]. All the eigenvalues of a nilpotent matrix are equal to zero and its Jordan canonical form is a strictly lower triangular matrix and thus also nilpotent. Subsequently we need only consider in view of Lemma 3.2 the factorization of $J_k \oplus J_2$ $(k \geq 2)$ and $J_2 \oplus J_2 \oplus J_2$. The proof will be complete if we can factor each of these two matrices into two nilpotent factors with ranks equal to the rank of these matrices. The reason for selecting these two matrices is due to the fact that it includes the exceptional case J_2 which is not a product of two nilpotent matrices according to (Lemma 3.1). It would be easy to show that A (nilpotent and $n \neq 2$) is a product of two nilpotent matrices if A is similar to a direct sum of Jordan canonical matrices not containing J_2. In this case we assume that $A \approx J_{k1} \oplus J_{k2} \oplus \ldots \oplus J_{kl}$, where $J_{ki} \neq J_2$ for $1 \leq i \leq l$. From Lemma 3.2 we notice that $J_{ki} = M_i N_i$ where M_i, N_i are nilpotent matrices. It follows that

$$A \approx J_{k1} \oplus J_{k2} \oplus \ldots \oplus J_{kl} = (M_1 \oplus M_2 \oplus \ldots \oplus M_l)(N_1 \oplus N_2 \oplus \ldots \oplus N_l).$$

The matrix A defined in this way would thus be a product of two nilpotent matrices. We now continue with the factorization of $J_k \oplus J_2$ $(k \geq 1)$ and $J_2 \oplus J_2 \oplus J_2$.

Case 1. A is similar to $J_k \oplus J_2$

If k is even

$$\begin{pmatrix} J_k & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{cc|c} & 0 & J_k \\ \hline 0 & 0 & \\ 0 & 1 & 0 \end{array}\right) \left(\begin{array}{cccc|c} & & 0 & & J_2 \\ \hline 1 & 0 & \ldots & 0 & \\ \vdots & \ddots & \vdots & \vdots & \\ 0 & \ldots & 1 & 0 & 0 \\ 0 & 0 & \ldots & 0 & \end{array}\right).$$

The rank of each of the nilpotent factors of $J_k \oplus J_2$ is $k + 2 - 2 = k$, which is the same as the rank of $J_k \oplus J_2$. The characteristic polynomial of each of the factors is x^n and therefore it follows that the factors are nilpotent. The first factor is similar to a Jordan matrix of the form $0 \oplus J_{k+1}(0)$ and the second factor is similar to a Jordan matrix of the form $J_{k+1}(0) \oplus 0$.

If k is odd

We first note that the previous approach for k even does not work for k odd.

For example for $k = 5$ we find that

$$\begin{pmatrix} J_5 & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{cc|c} \multicolumn{2}{c|}{0} & J_5 \\ \hline 0 & 0 & \\ 0 & 1 & 0 \end{array}\right) \left(\begin{array}{ccccc|cc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The first factor is nilpotent, but the second factor is not since it has a characteristic polynomial of $x^4 - x^7$.

However as stated in [4], the approach suggested by Wu for k odd is also not valid in general. As an example, it is stated in [11] that for $k = 7$ the second factor in the proof of Wu is not nilpotent, since it's characteristic polynomial is $x^5 - x^9$. The following correction of the proof of Wu is due to [11, Lemma 2].

For the case where $k = 1$ it is easy to see that:

$$\begin{pmatrix} J_k & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array}\right) \left(\begin{array}{c|cc} 0 & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

The factors are nilpotent since they are triangular matrices with zero entries on the diagonals and are of rank 1.
Suppose $k \geq 3$. Then

$$\begin{pmatrix} J_k & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{c|c} 0 & A_1 \\ \hline J_2 & 0 \end{array}\right) \left(\begin{array}{c|cc} 0 & 1 & 0 \\ & 0 & 0 \\ \hline A_2 & 0 & \end{array}\right) = N_1 N_2$$

where

$$A_1 = [e_2, 0, e_4, e_3, e_6, e_5, \cdots, e_{k-1}, e_{k-2}, e_k]$$

and

$$A_2 = [e_1, e_4, e_3, e_6, e_5, \cdots, e_{k-1}, e_{k-2}, e_k, 0].$$

In the case $k = 3$ we have that $A_1 = [e_2, 0, e_3]$ and $A_2 = [e_1, e_3, 0]$.

We now illustrate Hattingh's factorization for the case $k = 5$ and $k = 7$.

<u>$k = 5$</u>

$$\begin{pmatrix} J_5 & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{c|c} 0 & A_1 \\ \hline J_2 & 0 \end{array}\right) \left(\begin{array}{c|cc} 0 & 1 & 0 \\ & 0 & 0 \\ \hline A_2 & 0 & \end{array}\right) = N_1 N_2$$

$$= \left(\begin{array}{cc|ccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \left(\begin{array}{ccccc|cc} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right).$$

Both the factors have the same rank as $\begin{pmatrix} J_5 & 0 \\ 0 & J_2 \end{pmatrix}$ i.e. 5 and a characteristic polynomial of $-x^7$.

<u>$k = 7$</u>

$$\begin{pmatrix} J_7 & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{c|c} 0 & A_1 \\ \hline J_2 & 0 \end{array}\right) \left(\begin{array}{c|cc} 0 & 1 & 0 \\ & 0 & 0 \\ \hline A_2 & 0 & \end{array}\right) = N_1 N_2$$

where

$$N_1 = \left(\begin{array}{cc|ccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

and

$$N_2 = \left(\begin{array}{ccccccc|cc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right).$$

Both the factors have the same rank as $\begin{pmatrix} J_7 & 0 \\ 0 & J_2 \end{pmatrix}$ i.e. 7 and a characteristic polynomial of $-x^9$.

Case 2. A is similar to $J_2 \oplus J_2 \oplus J_2$

In this case we have that:

$$\begin{pmatrix} J_2 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_2 \end{pmatrix} = \left(\begin{array}{cc|cc|cc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right) \left(\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$
$$= N_1 N_2$$

The rank of each of the nilpotent factors of $J_2 \oplus J_2 \oplus J_2$ is 3 which is the same as the rank of $J_2 \oplus J_2 \oplus J_2$. Furthermore, the characteristic polynomial of each of the factors is x^6 and therefore the two factors are both nilpotent matrices.

We notice that $N_1 = SRS^{-1}$, where

$$S = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and $R = J_3(0) \oplus J_2(0) \oplus J_1(0)$, which is a nilpotent matrix.

Also $N_2 = UWU^{-1}$, where

$$U = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and $W = J_1(0) \oplus J_1(0) \oplus J_4(0)$, which is a nilpotent matrix. □

3.1.2 Laffey

Laffey shows in [17] that Wu's theorem on nilpotent factorization in [29] is true over an arbitrary field F, whereas Wu only proved the theorem for the complex field.

Lemma 3.4

A nilpotent matrix A_0 that is similar to a product N_1N_2, where N_1, N_2 are nilpotent, is still similar to a product $N_1^{'}N_2^{'}$, where $N_1^{'}$, $N_2^{'}$ are nilpotent and the last column of $N_1^{'}$ is zero.

Proof. Since a nilpotent matrix A_0 is similar to a nilpotent Jordan canonical matrix we will only consider the nilpotent Jordan canonical matrices as identified in Lemma 3.3, namely $J_n(0)$, $J_k(0)\oplus J_2(0)$ and $J_2(0)\oplus J_2(0)\oplus J_2(0)$.

Case 1. $J_n(0)$

By definition

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

n odd

According to Lemma 3.2,

$$J = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 \end{pmatrix} = N_1N_2.$$

We can find a matrix $J^{'}$ which is similar to J such that $J^{'} = S_1^{-1}JS_1$ where $S_1 = [e_n, e_1, \cdots, e_{n-1}]$ and

$$J^{'} = N_1^{'}N_2^{'} = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

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

The nilpotent matrix $N_1^{'}$ can be created from N_1 by swopping the one in position (2, n) with the zero in position (1, n - 1). The nilpotent matrix $N_2^{'}$ can be created from N_2 by swopping the one in position (n, 1) with the zero in position (n - 1, n).

n even

The proof of this part is due to [11, Lemma 3]. In Lemma 3.2, $J_n(0)$ is factorized as follows:

$$\begin{aligned} J_n(0) &= \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 \\ 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 \end{pmatrix} \\ &= N_1 N_2 \end{aligned}$$

Let S_2 be the change of basis matrix:

$$S_2 = [e_1, e_3, \cdots, e_{n-1}, e_2, e_4 - e_3, e_6 - e_5, \cdots, e_n - e_{n-1}]$$

and the first factor in

$$S_2^{-1} J_n(0) S_2 = (S_2^{-1} N_1 S_2)(S_2^{-1} N_2 S_2)$$

is of the desired form.

As an example we illustrate this method for $n = 4$, and

$$\begin{aligned} S_2^{-1} N_1 S_2 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \end{aligned}$$

Case 2. $J_k(0) \oplus J_2(0)$

k even

Due to Lemma 3.3 we have the following factorization:

$$\begin{pmatrix} J_k & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{cc|c} & 0 & J_k \\ \hline 0 & 0 & \\ 0 & 1 & 0 \end{array}\right) \left(\begin{array}{cccc|c} & & 0 & & J_2 \\ \hline 1 & 0 & \dots & 0 & \\ \vdots & \ddots & \vdots & \vdots & \\ 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & \end{array}\right)$$

and the first factor has a zero last column.

k odd

The proof of this case is due to [11]. In [11, Lemma 2] we see that for $k \geq 3$

$$\begin{pmatrix} J_k & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{c|c} 0 & A_1 \\ \hline J_2 & 0 \end{array}\right) \left(\begin{array}{c|cc} 0 & 1 & 0 \\ & 0 & 0 \\ \hline A_2 & 0 & \end{array}\right) = N_1 N_2$$

where

$$A_1 = [e_2, 0, e_4, e_3, e_6, e_5, \dots, e_{k-1}, e_{k-2}, e_k]$$

and

$$A_2 = [e_1, e_4, e_3, e_6, e_5, \dots, e_{k-1}, e_{k-2}, e_k, 0].$$

By using the change of basis matrix S_3:

$\mathrm{S}_3 = [e_1, e_{k+2}, e_k, e_{k-1}, e_{k-4}, e_{k-5}, e_{k-8}, \dots, e_{3-(-1)^{\frac{(k-3)}{2}}}, e_{k+1}, e_{k-2}, e_{k-3},$
$\dots, e_{3+(-1)^{\frac{(k-3)}{2}}}]$ we are able to find a matrix similar to $J_k \oplus J_2$ such that:

$$J_k \oplus J_2 \approx S_3^{-1}[J_k \oplus J_2]S_3 = (S_3^{-1} N_1 S_3)(S_3^{-1} N_2 S_3)$$

and the first factor has a zero last column.

For the case $k = 1$:

$$\begin{pmatrix} J_k & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array}\right) \left(\begin{array}{c|cc} 0 & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

the first factor is of the desired form. No similarity transformation is required.

Hattingh' s method can be illustrated with an example where $k = 5$. In this case

$$\begin{pmatrix} J_5 & 0 \\ 0 & J_2 \end{pmatrix} = \left(\begin{array}{c|c} 0 & A_1 \\ \hline J_2 & 0 \end{array}\right) \left(\begin{array}{c|cc} 0 & 1 & 0 \\ & 0 & 0 \\ \hline A_2 & 0 & \end{array}\right) = N_1 N_2$$

$$= \left(\begin{array}{cc|ccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \left(\begin{array}{ccccc|cc} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right).$$

We use the change of basis matrix $S_3 = [e_1, e_7, e_5, e_4, e_6, e_3, e_2]$ and

$$S_3^{-1} N_1 S_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

which has the required form.

Case 3. $J_2(0) \oplus J_2(0) \oplus J_2(0)$

In the proof of Lemma 3.3 it was shown that:

$$\begin{pmatrix} J_2 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_2 \end{pmatrix} = \left(\begin{array}{cc|cc|cc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right) \left(\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$= N_1 N_2$$

and that $N_1 = S_4 R S_4^{-1}$, where

$$S_4 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and $R = J_3(0) \oplus J_2(0) \oplus J_1(0)$, which is a nilpotent matrix with a zero last column. □

Corollary 3.5

A nilpotent matrix A_0 that is similar to a product N_1N_2, where N_1, N_2 are nilpotent, is similar to a product $N_1^{'}N_2^{'}$, where $N_1^{'}$, $N_2^{'}$ are nilpotent and the last row of $N_1^{'}$ is zero.

Proof. The proof follows from Lemma 3.4. In all 3 cases, namely $J_n(0)$, $J_k(0) \oplus J_2(0)$ and $J_2(0) \oplus J_2(0) \oplus J_2(0)$ the second factor N_2 has a zero last column and it follows that N_2^T has a zero last row. Since $A_0 \approx N_1N_2 \approx (N_1N_2)^T = N_2^TN_1^T$, whereby N_1^T and N_2^T are nilpotent (the transpose of a nilpotent matrix is nilpotent), the conclusion of the Corollary follows. □

Lemma 3.6 [12, Proposition 5.20]

Let F be an arbitrary field, and let $A \in M_n(F)$, $B \in M_m(F)$. If the characteristic polynomials of A and B are relatively prime, then $\begin{pmatrix} A & D \\ 0 & B \end{pmatrix}$ is similar to $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ for any matrix $D \in M_{n\times m}(F)$.

Proof. We repeat the proof in [12, Proposition 5.20] here for completeness sake. See also the proposition in [34] where the same proof technique is used.

Define $S : M_{n\times m}(F) \to M_{n\times m}(F)$ by $S(X) = AX - XB$. It is easy to verify that S is a linear transformation. Therefore if we can prove that $S(X) = 0$ implies $X = 0$, we will have proved that S is an automorphism of $M_{n\times m}$. Consequently a unique solution exists for the equation $S(X) = D$ where D is any matrix in $M_{n\times m}(F)$, and the result then follows directly by Roth's removal rule [34], since then

$$\begin{aligned}\begin{pmatrix} I & X \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} &= \begin{pmatrix} I & -X \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} A & D \\ 0 & B \end{pmatrix}.\end{aligned}$$

To this end, let $S(X) = 0$, then $AX = XB$. Let the characteristic polynomial of A be $p(x)$ and the characteristic polynomial of B be $q(x)$. If these polynomials are relatively prime there exist polynomials $f(x)$, $g(x)$ such that:

$$f(x)p(x) + g(x)q(x) = 1$$

Now since A satisfies its own characteristic equation $p(x) = 0$ (by the Cayley-Hamilton theorem) we have

$$f(A)p(A) + g(A)q(A) = g(A)q(A) = I.$$

Now it follows that

$$X = IX = g(A)q(A)X = g(A)Xq(B) = 0$$

where the second last step follows from repeated application of $AX = XB$, and the last step is due to the fact that B satisfies its own characteristic equation. □

Theorem 3.7 [17, Theorem 1.3]

Let F be a field and let $A \in M_n(F)$ with $\det A = 0$. Then A is the product of two nilpotent matrices, except when $n = 2$ and A is a nonzero nilpotent matrix.

Proof. Laffey considers the case where A is nilpotent and also the case where A is not nilpotent. In both cases A is assumed to be singular since it is given that $\det A = 0$. This of course is a necessary condition for A to be a product of two nilpotent matrices, but not quite sufficient as can be seen from the statement of the theorem.

A nilpotent

Laffey observes that the proof of [29, Lemma 3] is valid over any field and not only the complex field as established by Wu. It shows that any $n \times n$ nilpotent matrix A, except for a nonzero 2×2 nilpotent matrix, can be expressed as a product of two nilpotent matrices of the same rank as A. The reason is that since the characteristic polynomial of A is x^n, which trivially splits over any field, a nilpotent matrix A is similar to its Jordan canonical form J over any field and the factorizations of J that Wu uses only involve the elements 0 and ± 1.

A not nilpotent

Since the result is invariant under similarity we may assume that $A = A_0 \oplus A_1$ where $A_0 \neq 0$ is nilpotent and $A_1 \neq 0$ is non-singular.

Since A_0 is nilpotent we can use Lemma 3.3 to represent A_0 as a product of two nilpotents, provided A_0 is not similar to $J_2(0)$. Let $A_0 = N_1 N_2$. Without affecting the nilpotency of N_1 we can change the last column (or last row) of N_1 to zero as explained in Lemma 3.4 of this subsection. Using Theorem 2.5 we can assume under similarity that $A_1 = XY$, where X is lower triangular and Y is upper triangular with $X = (x_{ij})$ and $Y = (y_{ij})$. Laffey considers the following separation of cases partly due to the fact that the factorization $A_0 = N_1N_2$ is not valid when A_0 is a 2×2 nonzero nilpotent matrix, namely: (1) $A_0 = (0)$, (2) A_0 is not similar to $J = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, and (3) A_0 is a 2×2 nilpotent similar to J.

Case 1: $A_0 = (0)$ (one-by-one)
The matrix A can be expressed as follows:

$$\begin{aligned} A &= A_0 \oplus A_1 \\ &= \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ x_{11} & 0 & 0 & \dots & 0 \\ x_{21} & x_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mm} & 0 \end{pmatrix} \begin{pmatrix} 0 & y_{11} & y_{12} & \dots & 0 \\ 0 & 0 & y_{22} & \dots & y_{2m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & y_{mm} \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ X & 0 \end{pmatrix} \begin{pmatrix} 0 & Y \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & XY \end{pmatrix} \end{aligned}$$

where $m = n - 1$. Both factors of A are nilpotent, since their characteristic equations are $x^n = 0$, since X is strictly lower triangular and Y is strictly upper triangular.

Case 2: A_0 is $k \times k$ $(k > 1)$ and not similar to $J = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$

We construct a matrix A_2 which is similar to A and which can be expressed as a product of two nilpotent matrices. By Lemma 3.4, $A_0 = N_1N_2$ where $N_1 = (u_{ij})$ and N_2 are nilpotent and we can assume that the last column of N_1 is zero.

Let

$$A_2 = \left(\begin{array}{ccccc|cccc} u_{11} & u_{12} & \dots & u_{1,k-1} & 0 & & & & \\ u_{21} & u_{22} & \dots & u_{2,k-1} & 0 & & 0 & & \\ \vdots & \vdots & \ddots & \vdots & \vdots & & & & \\ u_{k,1} & u_{k,2} & \dots & u_{k,k-1} & 0 & & & & \\ \hline 0 & 0 & \dots & 0 & x_{11} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & x_{21} & x_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & x_{m1} & x_{m2} & \dots & x_{mm} & 0 \end{array}\right)$$

$$\times \left(\begin{array}{c|cccc} & 0 & 0 & \dots & 0 \\ N_2 & \vdots & \vdots & \ddots & \vdots \\ & 0 & 0 & \dots & 0 \\ & y_{11} & y_{12} & \dots & y_{1m} \\ \hline & 0 & y_{22} & \dots & y_{2m} \\ 0 & \vdots & \vdots & \ddots & \vdots \\ & 0 & 0 & \dots & y_{mm} \\ & 0 & 0 & \dots & 0 \end{array}\right)$$

and $m = n - k$. Note that A_2 is of the form $\begin{pmatrix} A_0 & 0 \\ B & A_1 \end{pmatrix}$. Since the eigenvalues of A_0 are all zero and those of A_1 are all nonzero, it follows that the characteristic polynomials of A_0 and A_1 are relatively prime, hence by Lemma 3.6, A_2 is similar to $\begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix}$.

The two factors in the definition of A_2 are both nilpotent since the characteristic polynomial of each one is a power of x. Since $A \approx A_2$, A is also a product of two nilpotent matrices.

Case 3: A_0 is a 2×2 nilpotent matrix similar to $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$

We again assume that $A_1 = XY$, with X and Y lower and upper triangular respectively, and construct a matrix $A_2 \approx A$ such that A_2 is a product of two nilpotent matrices. Writing

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

it follows that:

$$A_2 = \left(\begin{array}{cc|cccc} 1 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \hline 0 & x_{11} & 0 & 0 & \dots & 0 \\ 0 & x_{21} & x_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & x_{m1} & 0 & \dots & x_{mm} & -1 \end{array}\right) \left(\begin{array}{cc|cccc} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & y_{11} & y_{12} & \dots & y_{1m} \\ \hline 0 & 0 & 0 & y_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & y_{mm} \\ 0 & 0 & 0 & 0 & \dots & 0 \end{array}\right)$$

$$= \left(\begin{array}{cc|ccc} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \hline 0 & 0 & & & \\ \vdots & \vdots & & A_1 & \\ 0 & -1 & & & \end{array}\right) = \begin{pmatrix} A_0 & 0 \\ B & A_1 \end{pmatrix} \approx \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix} = A$$

by Lemma 3.6. The second factor of A_2 is nilpotent since it is an upper triangular matrix with zero diagonals. To show that the first factor (M_1) of A_2 is nilpotent we study the characteristic polynomial of this factor. It follows easily that:

$$|xI - M_1| = \det \begin{pmatrix} x-1 & 0 & 0 & \dots & 0 & -1 \\ 0 & x & 0 & \dots & 0 & 0 \\ 0 & -x_{11} & x & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & -x_{m-1,1} & -x_{m-1,2} & \dots & x & 0 \\ 1 & -x_{m1} & -x_{m2} & \dots & -x_{mm} & x+1 \end{pmatrix}$$

$$= (x-1) \cdot \det \begin{pmatrix} x & 0 & \dots & 0 & 0 \\ -x_{11} & x & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -x_{m-1,1} & -x_{m-1,2} & \dots & x & 0 \\ -x_{m1} & -x_{m2} & \dots & -x_{mm} & x+1 \end{pmatrix}$$

$$+ (-1)^{n+1} \cdot \det \begin{pmatrix} 0 & 0 & \dots & 0 & -1 \\ x & 0 & \dots & 0 & 0 \\ -x_{11} & x & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -x_{m-1,1} & -x_{m-1,2} & \dots & x & 0 \end{pmatrix}$$

$$= (x-1) \cdot x^{n-2} \cdot (x+1) + (-1)^{n+1}(-1)(-1)^n \cdot x^{n-2}$$
$$= (x^2-1)x^{n-2} + (-1)^{2n}x^{n-2}$$
$$= x^n$$

and M_1 is nilpotent.

Since $A \approx A_2$ we find that A can also be expressed as a product of two nilpotent matrices. This concludes Laffey's theorem on the nilpotent factorization of singular matrices. □

3.2 Factorization of Singular Matrices

In [26], Sourour and Tang give the necessary and sufficient conditions that a singular square matrix A over an arbitrary field can be written as a product of two matrices with prescribed eigenvalues. Being a scalar matrix is not important in this article, since scalar matrices, apart from the zero matrix, are invertible and we are only considering singular matrices in this section. The main result of this article is Theorem 3.11. Lemmas 3.8 to 3.10 are very useful to prove Theorem 3.11 and will be discussed next before discussing Theorem 3.11.

3.2.1 Preliminary Results

Lemma 3.8 [26, Lemma 1]

If A is a square matrix that is not a scalar multiple of the identity and if $\lambda \in F$, then A is similar to a matrix whose (1,1) entry is λ.

Proof. Since A is non-scalar there exists a vector v_1 which is not an eigenvector of A such that $Av_1 \neq \lambda v_1$ and $(A - \lambda I)v_1 \neq 0$ (Lemma 2.1).

Let $v_2 = (A - \lambda I)v_1$. Suppose v_1 and v_2 are linearly dependent, then there exist scalars a_1 and a_2 not all zero such that $a_1 v_1 + a_2 v_2 = 0$. If only one of the scalars are nonzero then $v_1 = v_2 = 0$, which is a contradiction. If both the scalars are nonzero then $a_2 v_2 = -a_1 v_1$ and $v_2 = \frac{-a_1}{a_2} v_1$ which is only possible if $v_1 = v_2 = 0$ due to way in which v_2 is defined. This is a contradiction since v_1 and v_2 are nonzero. It follows that v_1 and v_2 are linearly independent.

We extend the vectors v_1 and v_2 to a basis of F^n. The columns of matrix A_1 which is similar to A are the coordinates of $Av_1, Av_2, \ldots Av_n$ with respect to $v_1, v_2, \ldots, v_n$.

Since $Av_1 = \lambda v_1 + v_2$, the matrix A_1 has a first column $(\lambda, 1, 0, \ldots, 0)$ with respect to this basis and this concludes the proof. □

Lemma 3.9 [26, Lemma 2]

If

$$T = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

is an $n \times n$ matrix where A and D are square matrices and if A is invertible, then $\text{null}(T) = \text{null}(D - CA^{-1}B)$.

Proof. Let $S = D - CA^{-1}B$, then

$$T = \begin{pmatrix} A & 0 \\ C & I \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & D - CA^{-1}B \end{pmatrix}$$

and since $\begin{pmatrix} A & 0 \\ C & I \end{pmatrix}$ is nonsingular,

$$\operatorname{rank} T = \operatorname{rank} \begin{pmatrix} I & A^{-1}B \\ 0 & D - CA^{-1}B \end{pmatrix} = \operatorname{rank} \begin{pmatrix} I & A^{-1}B \\ 0 & S \end{pmatrix} = n - \text{null}(S).$$

Since also $\operatorname{rank} T = n - \text{null}(T)$ we have that

$$\text{null}(T) = \text{null}(S) = \text{null}(D - CA^{-1}B).$$

□

Lemma 3.10 [26, Lemma 3]

Let D be an invertible $n \times n$ matrix, and Y any $k \times n$ matrix. Then the two $(n + k) \times (n + k)$ matrices $\begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}$ and $\begin{pmatrix} 0 & Y \\ 0 & D \end{pmatrix}$ are similar.

Proof. Similarity holds due to the following relationship:

$$\begin{pmatrix} I & YD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & -YD^{-1} \\ 0 & I \end{pmatrix} = \begin{pmatrix} 0 & Y \\ 0 & D \end{pmatrix}$$

□

3.2.2 Main Result

Theorem 3.11 [26, Theorem 1]

Let A be an $n \times n$ singular matrix over a field F and let β_j and γ_j $(1 \leq j \leq n)$ be elements of F. If A is not a nonzero 2×2 nilpotent, then A can be factored as a product BC with $\text{Eig}(B) = \{\beta_1, \ldots, \beta_n\}$ and $\text{Eig}(C) = \{\gamma_1, \ldots, \gamma_n\}$ if and only if the number of zeros m among $\beta_1, \gamma_1, \ldots, \beta_n, \gamma_n$ is not less than the nullity of A. If A is a nonzero 2×2 nilpotent then A can be factored as above if and only if $1 \leq m \leq 3$.

Proof. We now consider the necessary and sufficient conditions for A an $n \times n$ singular matrix over a field F to be factored as a product BC with $\text{Eig}(B) = \{\beta_1, \ldots, \beta_n\}$ and $\text{Eig}(C) = \{\gamma_1, \ldots, \gamma_n\}$, where β_j and γ_j

$(1 \leq j \leq n)$ are elements of F. Let m be the number of zeros among $\beta_1, \gamma_1, \ldots, \beta_n, \gamma_n$.

Necessity

Assume that $A = BC$ and that the eigenvalues of B and C are as described above. Let m_1 and m_2 denote the number of zeros among $\{\beta_1, \ldots, \beta_n\}$ and $\gamma = \{\gamma_1, \ldots, \gamma_n\}$ respectively.

By definition

$$\text{null}(A) \leq \text{null}(B) + \text{null}(C) \leq m_1 + m_2 = m.$$

This can be shown by using Sylvester's rank inequality as found in [31, Theorem 2.6]:

$$\begin{aligned} \operatorname{rank} B + \operatorname{rank} C - n &\leq \operatorname{rank} A \\ n - \text{null}(B) + n - \text{null}(C) - n &\leq n - \text{null}(A) \\ \text{null}(A) &\leq \text{null}(B) + \text{null}(C) \end{aligned}$$

Lemma 3.1 is used to show that $m \leq 3$. According to Lemma 3.1 a 2×2 nonzero nilpotent matrix cannot be expressed as a product of two nilpotent matrices and therefore we cannot have a situation where $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0$ i.e. $m \neq 4$ and so m ≤ 3. Since $\text{null}(A) \leq m$ and $\text{null}(A) = n - \operatorname{rank} A = 1$ ($\operatorname{rank} A = 1$), it follows that $m \geq 1$ and that $1 \leq m \leq 3$.

Sufficiency

We proceed by induction on the size of a singular matrix A. The result is trivial for a 1×1 matrix. If A is 1×1 then $A = [0] = [0][a] = BC$, where a is arbitrary. The nullity of A is such that $\text{null}(A) = n - \operatorname{rank} A = 1 - 0 = 1$ and the number of zeroes m is such that $m \geq 1$, since the eigenvalues of B and C are 0 and a respectively.

We assume that the theorem is true for matrices of size less than n, and let A be an $n \times n$ singular matrix. Let $\beta_1, \gamma_1, \ldots, \beta_n, \gamma_n$ be elements of F, exactly m of which are zero where m satisfies the condition of the Theorem.

If $m = 2n$ it means that all the eigenvalues of B and C are zero and B and C are nilpotent for A not a nonzero 2×2 nilpotent matrix and this is mentioned in Theorem 3.12. This Theorem is proved for an arbitrary field in Theorem 3.7. The two cases $m < n$ and $n \leq m \leq 2n - 1$ will be considered next.

Case 1: $m < n$.

The strategy employed here is to show that A is similar to A_1, which can be decomposed into two factors satisfying the conditions of the Theorem. The matrix A_1 is such that

$$A_1 = \begin{pmatrix} \beta_1\gamma_1 & y^T \\ x & D \end{pmatrix}$$

where x, $y \in F^{(n-1)}$ and $D \in M_{n-1}(F)$.

Sourour uses Lemma 3.9 to show that

$$\text{null}(D - \beta_1^{-1}\gamma_1^{-1}xy^T) = \text{null}(A_1) \leq m,$$

where m is the number of zeros among $\beta^{'} \cup \gamma^{'}$ where $\beta^{'} = \beta_2, \ldots, \beta_n$ and $\gamma^{'} = \gamma_2, \ldots, \gamma_n$ (Since $m < n$ not all β_i's and γ_i's are zero. Let β_1 and γ_1 be non-zero).

We know that

$$A_1 = \begin{pmatrix} \beta_1\gamma_1 & y^T \\ x & D \end{pmatrix} = \begin{pmatrix} \beta_1\gamma_1 & 0 \\ x & I \end{pmatrix} \begin{pmatrix} I & \gamma_1^{-1}\beta_1^{-1}y^T \\ 0 & D - x\beta_1^{-1}\gamma_1^{-1}y^T \end{pmatrix}$$

and from Lemma 3.9, $\text{null}(D - \beta_1^{-1}\gamma_1^{-1}xy^T) = \text{null}(A_1) \leq m$.

The induction hypothesis is applied to $D - \beta_1^{-1}\gamma_1^{-1}xy^T$, since it is singular and of size $n-1$. The matrix $(D - \beta_1^{-1}\gamma_1^{-1}xy^T)$ is singular since

$$\det A_1 = \beta_1\gamma_1 \det(D - \beta_1^{-1}\gamma_1^{-1}xy^T) = 0$$

and $\beta_1\gamma_1 \neq 0$.

There exist matrices B_0 and C_0, such that

$$D - \beta_1^{-1}\gamma_1^{-1}xy^T = B_0C_0,$$

$\text{Eig}(B_0) = \beta^{'}$ and $\text{Eig}(C_0) = \gamma^{'}$. The exceptional case where $D - \beta_1^{-1}\gamma_1^{-1}xy^T$ is a nonzero 2×2 nilpotent is not covered here, since if A is 3×3 and $\beta_2 = \beta_3 = \gamma_2 = \gamma_3 = 0$, then $m = 4$ and $m > n = 3$, which is a contradiction since m was defined to be less than n in this case.

Now

$$A_1 = \begin{pmatrix} \beta_1 & 0 \\ \gamma_1^{-1}x & B_0 \end{pmatrix} \begin{pmatrix} \gamma_1 & \beta_1^{-1}y^T \\ 0 & C_0 \end{pmatrix} = BC$$

as required.

Case 2: $n \leq m \leq 2n - 1$

If $\beta = 0$, Sourour shows that a solution can easily be found by factorising A^T as $A^T = RS$ with $\text{Eig}(R) = \text{Eig}(R^T) = \gamma$ and $\text{Eig}(S) = \text{Eig}(S^T) = \beta$ (S and S^T are nilpotent matrices) and then $A = S^T R^T$.

The case where $\beta \neq 0$ would be more interesting to consider. We can assume that $\beta_1 \neq 0$. Since $m \geq n$ at least one $\gamma_j = 0$ and we take $\gamma_1 = 0$. It is thus clear that $\beta_1 \gamma_1 = 0$. From Lemma 3.8 we have that A is similar to A_1 and

$$A_1 = \begin{pmatrix} 0 & y^T \\ 0 & D \end{pmatrix}$$

where $y \in F^{(n-1)}$ and $D \in M_{n-1}(F)$. The selection of A_1 above is possible, since A singular is similar to the transpose of its Jordan canonical form which has a zero first column. This construction is possible if the first Jordan canonical block of A_1 is linked to the zero eigenvalue of A. Since A is singular it has at least one eigenvalue equal to zero. The number of zeros among $\beta' \cup \gamma'$ is $m - 1 \geq n - 1 \geq \text{null}(D)$.

Two subcases are given namely when D is singular and when D is non-singular

<u>Case 2(a) D is non-singular</u>

A transformation is made to D to make it singular by subtracting a rank one matrix from D. The matrix R can be chosen in such a way so that R agree with D in one row and have all other rows zero.

In the case where $n = 3$, $D - R$ (2×2 matrix) is not nilpotent and we avoid the exceptional case where $D - R$ is a 2×2 nonzero nilpotent.

The above proposed method does not work all the time, especially if D is a 2×2 matrix with a zero diagonal such that

$$D = \begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix}.$$

If D has this property then $D - R$ is nilpotent and it would not be possible to always express $D - R$ as a product $B_0 C_0$ as given in the Theorem, since the condition $1 \leq m \leq 3$ would not be met. If F is real, we replace D with a diagonal matrix:

$$T = \begin{pmatrix} -\sqrt{b}\sqrt{c} & 0 \\ 0 & \sqrt{b}\sqrt{c} \end{pmatrix}$$

which is similar to D and subtract a matrix R from it as was done previously. The matrix D is diagonalizable since it has two distinct real eigenvalues namely $-\sqrt{b}\sqrt{c}$ and $\sqrt{b}\sqrt{c}$.

For any arbitrary field we choose R to be

$$R = \begin{pmatrix} -c & b \\ 0 & 0 \end{pmatrix}$$

and $D - R$ is singular, but not nilpotent and the induction hypothesis can be applied to $D - R$. The choice of R for an arbitrary field where D is a 2×2 matrix with a zero diagonal was mentioned in [16, Theorem 1.20].

In general the number of zeros among $\beta' \cup \gamma'$ is $m - 1$ and

$$m - 1 \geq n - 1 \geq \text{null}(D - R).$$

By the induction hypothesis $D - R = B_0 C_0$ with $\text{Eig}(B_0) = \beta'$ and $\text{Eig}(C) = \gamma'$.

The matrix R can be written as $R = zw^T$ where z, $w \in F^{n-1}$. The column vector z has a one in the same row as the non-zero elements of R and zeroes elsewhere, while w^T is the non-zero row of R.

Let

$$B = \begin{pmatrix} \beta_1 & 0 \\ z & B_0 \end{pmatrix}$$

and

$$C = \begin{pmatrix} 0 & w^T \\ 0 & C_0 \end{pmatrix}$$

with $\text{Eig}(B) = \beta$ and $\text{Eig}(C) = \gamma$ and

$$BC = \begin{pmatrix} 0 & \beta_1 w^T \\ 0 & D \end{pmatrix} = \begin{pmatrix} \beta_1 & 0 \\ z & B_0 \end{pmatrix} \begin{pmatrix} 0 & w^T \\ 0 & C_0 \end{pmatrix} = \begin{pmatrix} 0 & \beta_1 w^T \\ 0 & D \end{pmatrix}.$$

Both A_1 and BC are similar to

$$A_2 = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}$$

by Lemma 3.10.

Case 2(b) D is singular

Except when A is a 3×3 nonzero nilpotent and $\beta_2 = \gamma_2 = \beta_3 = \gamma_3 = 0$ can we apply the induction hypothesis to D to establish the existence of matrices B_0 and C_0 such that $D = B_0C_0$, $\mathrm{Eig}(B_0) = \beta'$ and $\mathrm{Eig}(C_0) = \gamma^{`}$. The matrix A is similar to A_1 where

$$A_1 = BC = \begin{pmatrix} \beta_1 & 0 \\ 0 & B_0 \end{pmatrix} \begin{pmatrix} 0 & \beta_1^{-1}y^T \\ 0 & C_0 \end{pmatrix},$$

$\mathrm{Eig}(B) = \beta$ and $\mathrm{Eig}(C) = \gamma$. We now consider two cases not covered by the induction hypothesis i.e. when A is a 3×3 nonzero nilpotent with $\beta_2 = \gamma_2 = \beta_3 = \gamma_3 = 0$ and the rank of A is either equal to 1 or 2.

If the rank of A is one, it is similar to a matrix of the form $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and

$$D = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

From the proof of [26, Lemma 4], we see that

$$D = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ b & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ c & 0 \end{pmatrix}$$

and take $b = c = 1$. We further notice that

$$A \approx BC = \begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & \beta_1^{-1} & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The number of zeroes,

$$m = 5 > \mathrm{null}(A) = n - \mathrm{rank}\, A = 3 - 1 = 2$$

satisfying the conditions of Theorem 3.11.

When A is a 3×3 nilpotent of rank 2 and $\gamma_1 = \beta_2 = \gamma_2 = \beta_3 = \gamma_3 = 0$ then A is similar to a matrix

$$A_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$D = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \approx \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Since $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ (and matrix D) is nilpotent it cannot be expressed as a product of two nilpotents according to Lemma 3.1.

The induction hypothesis however would require D to be a product of two nilpotents, since $\beta_2 = \gamma_2 = \beta_3 = \gamma_3 = 0$, but this is a contradiction. It is thus not possible to express D as $D = B_0C_0$ and to apply the induction hypothesis. We can however still show that Theorem 3.11 is true for A with the given properties without using induction. The matrix A can be factored in the following way:

$$A_1 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & \beta_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = BC,$$

where $\mathrm{Eig}(B) = \{\beta_1, 0, 0\}$ and $\mathrm{Eig}(C) = \{0, 0, 0\}$. The number of zeroes

$$m = 5 > \mathrm{null}(A) = n - \mathrm{rank}\, A = 3 - 2 = 1,$$

which satisfy the conditions of Theorem 3.11. □

Theorem 3.12 [26, Theorem 2]

Let A be a singular square matrix over an arbitrary field. Then A is a product of two nilpotent matrices if and only if A is not a nonzero 2×2 nilpotent matrix.

Proof. This proof was already shown in Theorem 3.7 for arbitrary fields. □

3.2.3 Comparison

In both Theorem 3.11 and Theorem 2.2, Sourour relies in the induction step on having to subtract a rank 1 matrix under certain conditions. Where the conditions of the Theorem cannot be applied a transformation is made to the relevant matrix to make the induction step possible. In Theorem 3.11 the induction step is not always possible, but is shown to be true through other means. Theorem 2.2 is primarily based on non-scalar invertible matrices, but was shown to be true for scalar matrices as well. Theorem 3.11 focuses on square singular matrices. In both articles we notice that whether A is non-singular or singular it can be factorized with prescribed eigenvalues under most conditions.

3.3 Applications

Lemma 3.13

A singular matrix A is similar to a direct sum of matrices each of nullity one.

Proof. By the rational canonical form theorem, A is similar to a direct sum of companion matrices

$$A \approx C(g_1) \oplus C(g_2) \oplus \ldots \oplus C(g_t)$$

where $g_1(x)|g_2(x)|\ldots|g_t(x)$ are the invariant factors of A. Since A is singular, $g_k(0) = 0$ for at least one $1 \leq k \leq t$. Assume k to be the smallest such integer, i.e. $g_i(0) \neq 0$ for $i < k$ and $g_k(0) = \ldots = g_t(0) = 0$. Then $A \approx A_k \oplus \ldots \oplus A_t$ with $A_k = C(g_1) \oplus \ldots \oplus C(g_k)$ and $A_i = C(g_i)$ for $k < i \leq t$ satisfies the requirements of the lemma. □

Theorem 3.14 [26, Theorem 3]

Let A be a real or complex square matrix.

(a) If A is singular, then A is a product of four positive-semidefinite matrices, three of which may be taken to be definite.

(b) If $\det A$ is a nonzero real number and A is not a scalar multiple of I, then A is a product of four Hermitian matrices, at least three of which may be taken to be positive-definite.

(c) If $A = \lambda I$ and $\det A$ is real, then A is a product of four Hermitian matrices, none of which can be definite unless λ^2 is real.

Proof. Each of the classes of Hermitian, positive-definite, and positive-semidefinite matrices are invariant under congruence i.e. if $A \approx C^*AC$, for invertible C, then C^*AC is of the same class as A.

If the matrices R_1, R_2, R_3 and R_4 are Hermitian, positive-definite, positive-semidefinite or a combination of these as described in the statement of the theorem, then from the equation

$$T^{-1}R_1R_2R_3R_4T = (T^{-1}R_2T^{*-1})(T^*R_2T)(T^{-1}R_3T^{*-1})(T^*R_4T),$$

we notice that a matrix similar to the product $R_1R_2R_3R_4$ is also a product of the same class of matrices.

(a) Since A is singular it is similar to a matrix which is a direct sum of matrices of nullity one (Lemma 3.13).

Therefore it suffices to prove the conclusion of the theorem, when $\text{null}(A) = 1$.

By Theorem 3.11 there exists matrices B and C such that $A = BC$ where the eigenvalues of B are distinct and positive and those of C distinct and non-negative. (This theorem can be applied since we need to choose only one zero eigenvalue for C since the nullity of A is equal to one). The matrix B is thus similar to a positive-definite matrix P such that $B = T^{-1}PT$ and C is similar to a positive-semidefinite matrix S such that $C = R^{-1}SR$. Both T and R are invertible. The equation

$$A = (T^{-1}PT^{*-1})(T^*T)(R^{-1}R^{*-1})(R^*SR)$$

gives A as a product of three positive-definite matrices and one positive-semidefinite.

The products T^*T and R^*R are positive-definite, since for $v \neq 0$,

$$v^*(T^*T)v = (v^*T^*)(Tv) = (Tv)^*(Tv) > 0$$

and similarly for R^*R. The matrix $R^{-1}R^{*-1}$ is such that $R^{-1}R^{*-1} = (R^*R)^{-1}$ and the inverse of a positive-definite matrix is also positive-definite. The matrix $T^{-1}PT^{*-1}$ is congruent to P and therefore positive-definite. The matrix R^*SR is congruent to S and therefore also positive-semidefinite.

(b) (i) A is real.

Matrix A is a product of two Hermitian matrices as given in [22, Theorem 1]. Frobenius showed that every matrix over a field is a product of two symmetric matrices over the same field. It follows that A is a product of four symmetric matrices or four Hermitian matrices, since for A real, $A^T = A^*$.

(ii) A is complex.

If $\det A$ is a nonzero real number and A is not scalar, we use Theorem 2.2 to write A as a product of matrices BC where B has n distinct positive eigenvalues and C has n distinct real eigenvalues, at most one of which is negative (This is to make provision for the case where $\det A < 0$). As in (a), $B = (T^{-1}PT^{*-1})(T^*T)$, a product of two positive- definite matrices and $C = (R^{-1}R^{*-1})(R^*SR)$ a product of a positive-definite and a Hermitian matrix, where $B = T^{-1}PT$, $C = R^{-1}SR$ with P positive-definite, S Hermitian, and T and R invertible.

Therefore A is a product of four Hermitian matrices, at least three of which can be taken positive-definite.

If $\det A > 0$, then A is a product of four positive-definite matrices as B and C can be chosen to have n distinct positive real eigenvalues. This result is due to Ballantine ([1] and [2]) (see Theorem 2.6).

(c) If $A = \lambda I$, the proof is given in [22, Theorem 2]. We choose a basis $\{e_1, \ldots, e_n\}$ of $F^{(n)}$. Let U be the shift given by $Ue_j = e_{j+1}$ for $1 \leq j \leq n-1$ and $Ue_n = e_1$. The shift U can also be presented as:

$$U = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

and

$$U\lambda U^T = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix}$$

$$\times \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & \ldots & 0 \end{pmatrix} = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix} = A.$$

Let S be a Hermitian symmetry given by $Se_j = e_{n+1-j}$ and K an operator given by $Ke_j = \bar{\lambda}^{n+1-j}\lambda^j e_{n+1-j}$. The matrix representation of S and K can be given as

$$S = \begin{pmatrix} & & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & 1 & & & \\ 1 & & & & \end{pmatrix}$$

and

$$K = \begin{pmatrix} & & & & \bar{\lambda}\lambda^n \\ & & & \bar{\lambda}^2\lambda^{n-1} & \\ & & \iddots & & \\ & \bar{\lambda}^{n-1}\lambda^2 & & & \\ \bar{\lambda}^n\lambda & & & & \end{pmatrix}.$$

We notice that

$$US = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} & & & & 1 \\ & & & 1 & \\ & & \cdot^{\cdot^{\cdot}} & & \\ & 1 & & & \\ 1 & & & & \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \end{pmatrix} = (US)^* = S^*U^*$$

and the product US is Hermitian. The matrices S and K are also Hermitian since $S = S^*$ and $K = K^*$. Since S is Hermitian S^{-1} is also Hermitian.

The matrix $\lambda U^T K^{-1}$ is such that:

$$\lambda U^T K^{-1} = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}$$

$$\times \begin{pmatrix} & & & \bar{\lambda}^{-1}\lambda^{-n} \\ & & \bar{\lambda}^{-2}\lambda^{-(n-1)} & \\ & \cdot^{\cdot^{\cdot}} & & \\ \bar{\lambda}^{-(n-1)}\lambda^{-2} & & & \\ \bar{\lambda}^{-n}\lambda^{-1} & & & \end{pmatrix}$$

$$= \begin{pmatrix} 0 & \dots & \bar{\lambda}^{-2}\lambda^{-n+2} & 0 \\ \vdots & & \vdots & \vdots \\ \bar{\lambda}^{-n} & \dots & 0 & 0 \\ 0 & \dots & 0 & \bar{\lambda}^{-1}\lambda^{-n+1} \end{pmatrix}$$

Also

$$(\lambda U^T K^{-1})^* = (K^{-1})^*(U^T)^*\bar{\lambda}I$$

$$= \begin{pmatrix} 0 & \dots & \lambda^{-n} & 0 \\ \vdots & & \vdots & \vdots \\ \lambda^{-2}\bar{\lambda}^{-n+2} & \dots & 0 & 0 \\ 0 & \dots & 0 & \lambda^{-1}\bar{\lambda}^{-n+1} \end{pmatrix}$$

and since $\det A = \lambda^n$ is real, it follows that $\lambda^n = \bar{\lambda^n}$ and $\lambda^{n-i} = \bar{\lambda^n}\lambda^{-i}$ for $1 \leq i \leq n-1$. As a result of this

$$(\lambda U^T K^{-1}) = (\lambda U^T K^{-1})^*.$$

It follows that $A = (US)S^{-1}(\lambda U^T K^{-1})K$ is a product of four Hermitian matrices.

Furthermore if λI is a product of four Hermitian matrices, at least one of which is positive-definite, then $\lambda I = H_1 H_2 H^{-1} P^{-1}$ or $\lambda PH = H_1 H_2$ for Hermitian H_1,H_2,H and positive-definite P (H^{-1} and P^{-1} are Hermitian and positive-definite respectively, since they have the same class as their inverses).

Now $H_1 H_2 = H_1(H_2 H_1)H_1^{-1}$ and $H_1 H_2 \approx H_2 H_1$ and so

$$\lambda PH = H_1 H_2 \approx H_2 H_1 = (H_1 H_2)^* = (\lambda PH)^* = \bar{\lambda} H^* P^* = \bar{\lambda} HP$$

and λPH is similar to $\bar{\lambda}HP$. The matrix $\bar{\lambda}HP$ is similar to $\bar{\lambda}PH$ since $P(\bar{\lambda}HP)P^{-1} = \bar{\lambda}PH$.

Since λPH is similar to $\bar{\lambda}HP$ and $\bar{\lambda}PH$ is similar to $\bar{\lambda}HP$ it follows that λPH is similar to $\bar{\lambda}PH$. Subsequently $\lambda \cdot \lambda PH = \lambda^2 PH$ is similar to $\lambda \cdot \bar{\lambda} PH$.

The product PH has real eigenvalues, since it is similar to a Hermitian matrix $P^{\frac{1}{2}} H P^{\frac{1}{2}}$ i.e. $PH = P^{\frac{1}{2}}(P^{\frac{1}{2}} H P^{\frac{1}{2}})P^{\frac{-1}{2}}$. Therefore λ^2 is real. □

Remark 1: If λ^2 is real, then λI is either Hermitian or is a product of three Hermitian matrices. This follows from the equation:

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

The matrix $\begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}$ is such that $\lambda = i$ and $\lambda^2 = i^2 = -1$, which is real. Each of the products of $\begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}$ are equal to their complex conjugates and are thus Hermitian.

Remark 2: The position of the three positive-definite matrices in parts (a) and (b) are arbitrary. This follows from the fact that a product PH of a positive-definite matrix P (or positive-semidefinite) and Hermitian H can be written as a product of a Hermitian and a positive-definite matrix (or positive-semidefinite) i.e. in reverse order as $PH = (PHP^*)P^{*-1}$ where (PHP^*) is Hermitian and P^{*-1} is positive-definite (or positive-semidefinite).

Bibliography

[1] C.S. Ballantine, Products of positive-definite matrices III, Journal of Algebra, no. 10 (1968): 174 – 182.

[2] C.S. Ballantine, Products of positive-definite matrices IV, Linear Algebra and its Applications, no. 3 (1970): 79 – 114.

[3] J.D. Botha, Private communication, 22 November 2018.

[4] D.K. Bukovsek, T. Kosir, N. Novak and P. Oblak, Products of commuting nilpotent operators, Electronic Journal of Linear Algebra, no. 16 (2007): 237 – 247.

[5] C.G. Cullen, Matrices and linear transformations, 2nd Edition (New York: Dover Publications, 1990).

[6] P.A. Fillmore, On similarity and the diagonal of a matrix, The American Mathematical Monthly 76, no. 2 (1969): 167 – 169.

[7] J.B. Fraleigh, A first course in abstract algebra, 8th Edition (London: Pearson, 2011).

[8] F.G. Frobenius, Uber die miteiner matrix vertauschbaren matrizen, Sitzungsberichte der Koniglich Preussischen Akademie der Wissenschaften zu Berlin (1910): 3 – 15.

[9] J.S. Golan, The linear algebra a beginning graduate ought to know, 3rd Edition (Dordrecht: Springer, 2012).

[10] J.I. Hall, Another elementary approach to the Jordan form, The American Mathematical Monthly, no. 98(4) (1991): 336 – 340.

[11] C. Hattingh, A note on products of nilpotent matrices, arXiv:1608.04666v3 [math.RA], 6 Sep 2016.

[12] C. Hattingh, Sums and products of square-zero matrices, M.Sc. dissertation, University of South Africa, March 2018.

[13] F. Hoffman and E. Paige, Products of two involutions in the general linear group, Indiana University Mathematics Journal, no. 20 (1971): 1017 – 1020.

[14] L. Hubert, J. Meulman and W. Heiser, Two purposes for matrix factorization: A historical appraisal, SIAM Review, no. 42(1) (2000): 66 – 82.

[15] N. Jizhu and Z. Yongzheng, Products of nilpotent matrices over skew fields, Northeastern Mathematical Journal, no. 12(3) (1996): 299 – 306.

[16] M.C. Khoury, Products of diagonalizable matrices, M.Sc. dissertation, University of South Africa, September 2002.

[17] T.J. Laffey, Products of matrices: Generators and relations in groups and geometries (Dordrecht: NATO ASI Series, 1991).

[18] T.J. Laffey, Current trends in matrix theory: Factorization of matrices involving symmetric matrices and involutions (New York: Editors F Uhlig and R Grone, North Holland, 1987), 175 – 198.

[19] A. Mohammadian, Sums and products of square-zero matrices, Communications in Algebra, no. 40(12) (2012): 4568 – 4574.

[20] N. Novak, Products of square-zero matrices, Journal of Mathematical Analysis and Applications, no. 339 (2008): 10 – 17.

[21] J.M. Ortega, Matrix theory: A second course (New York: Plenum Press, 1986).

[22] H. Radjavi, Products of Hermitian matrices and symmetries, Proceedings of the American Mathematics Society, no. 21 (1969): 369 – 372; no. 26 (1970): 701.

[23] W.E. Roth, The equations $AX - YB = C$ and $AX - XB = C$ in matrices, Proceedings of the American Mathematical Society, no. 3 (1952): 392 – 396.

[24] A.R. Sourour, A factorization theorem for matrices, Linear and Multilinear Algebra, no.19 (1986): 141 – 147.

[25] A.R. Sourour, Nilpotent factorization of matrices, Linear and Multilinear Algebra, no. 31 (1992): 303 – 308.

[26] A.R. Sourour and K. Tang, Factorization of singular matrices, Proceedings of the American Mathematical Society, no. 116(3) (1992): 629 – 634.

[27] G. Strang, Linear algebra and its applications, 3rd Edition (San Diego: Harcourt, Brace, Jovanovich, Publishers, 1988).

[28] A. Tucker, The growing importance of linear algebra in undergraduate mathematics, The College Mathematics Journal, no. 24 (1993): 3 – 9.

[29] P.Y. Wu, Products of nilpotent matrices, Linear Algebra and its Applications, no. 96 (1987): 227 – 232.

[30] M. Yang, Matrix decomposition, Electrical and Computer Engineering, Northwestern University, Accessed on 16 March 2018.
http://users.eecs.northwestern.edu/ mya671/files/Matrix_YM_.pdf.

[31] F. Zhang, Matrix theory: Basic results and techniques, 2nd Edition (New York: Springer, 2011).

[32] Wikipedia, Alan Turing, Accessed on 16 March 2017.
http://en.m.wikipedia.org/wiki/Alan_Turing.

[33] Wikipedia, Matrix (mathematics), Accessed on 16 March 2017.
http://en.m.wikipedia.org/wiki/Matrix_(mathematics).

[34] Wikipedia, Sylvester equation, Accessed on 1 September 2019.
http://en.m.wikipedia.org/wiki/Sylvester_equation.