SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Matrix Invariants

av

Anders Johansson

2020 - No K27

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Matrix Invariants

Anders Johansson

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Torbjörn Tambour

2020

Abstract

The set of invertible square matrices in the set of all square matrices is a group under matrix multiplication, and it acts on the set of all square matrices through conjugation. Invariant polynomial functions, or *invariants*, to this group action are polynomials in the elements of the matrix that are constant on the orbits. The main result of this thesis, is that *all* invariants can be expressed in terms of a few special invariants. These special invariants are the coefficients in the characteristic polynomial of the matrix that is subject to the group action. To do this we will look at the properties of permutations and linear transformations, their matrix representations as well as the characteristic polynomial of a matrix and its connection to these special invariants. We will also look at properties of polynomials, mainly symmetric polynomials and use the fundamental theorem of symmetric polynomials to find the invariants for different types of matrices.

Acknowledgements

I want to give special thanks to my mentor Torbjörn Tambour for his excellent help and guidance throughout my work on this thesis, as well as his contribution to enlightening my mathematical perspective.

Preface

In this thesis, I have taken the four books listed in the references by: Springer, Biggs, Nagell and Friedberg et al. as a basis for the thesis. Consequently some of the references are used explicitly, and referred to in the text with a bracketed number, along with a paragraph- or section number [n]: §/sec, while some references are used more generally and thus not explicitly referred to in the text. It is hard to give specific references to Springer and Biggs, since they are used more across-the-board compared to Nagell and Friedberg et al.

Contents

1 Introduction and Summary of results

Let $\mathbb{M}_n$ be the set of all $n \times n$-matrices over $\mathbb{C}$, and let $\mathbb{G}$ be the set of all invertible matrices in $\mathbb{M}_n$. Then $\mathbb{G}$ is a group under matrix multiplication, called *the general linear group over* $\mathbb{C}$, $\mathbb{GL}(n, \mathbb{C})$.

A polynomial function

$$p \colon \mathbb{M}_n \to \mathbb{C}$$

is a function that originates in a polynomial, that is a polynomial in the elements of the matrix. As an example of such a polynomial function, we take the sum of the matrix elements a_{ij} of a matrix $A \in \mathbb{M}_n$, where $i > j$, that is

$$p(A) = \sum_{1 \leq i < j \leq n} a_{ij}$$

which is a polynomial in the elements of A. If we then let the variables of this polynomial assume values in $\mathbb{C}$, the polynomial will evaluate to an element in $\mathbb{C}$.

In this thesis we will investigate invariants to matrices, which are polynomial functions that are *constant on the orbits* of the *group action* of the group $\mathbb{GL}(n, \mathbb{C})$ on the set $\mathbb{M}_n$.

Definition 1.1: Group Action
A *group action* f, of the group $\mathbb{G}$ on the set $\mathbb{A}$, is a function from the Cartesian product $\mathbb{G} \times \mathbb{A}$ to $\mathbb{A}$

$$f \colon \mathbb{G} \times \mathbb{A} \to \mathbb{A}; \;\; \mathbb{G} \times \mathbb{A} = \{(G, A) \colon G \in \mathbb{G}, A \in \mathbb{A}\}$$
$$f \colon (G, A) \mapsto G.A; \;\; G \in \mathbb{G}, \; A \in \mathbb{A}$$

that obeys the two axioms of group actions:

Axiom 1.1.1: One element $G_1 \in \mathbb{G}$ acting on an element $G_2.A \in \mathbb{A}$ has to be equal to the composition of the elements $G_1 G_2 \in \mathbb{A}$ acting on the element $A \in \mathbb{A}$

$$G_1.(G_2.A) = (G_1 G_2).A \;\; ; \forall \, G_1, G_2 \in \mathbb{G}, \; \forall A \in \; \mathbb{A}.$$

Axiom 1.1.2: The identity element in $\mathbb{G}$ has to act on all elements $A \in \mathbb{A}$ to give the same element of $\mathbb{A}$

$$G_{\mathbb{1}}.A = A, \;\; \forall \, A \in \mathbb{A}$$

where $G_{\mathbb{1}}$ is the identity element of $\mathbb{G}$. To avoid confusion with the number 1 and the basis elements e_i, we will denote the identity by the subscript $\mathbb{1}$.

The group $\mathbb{GL}(n, \mathbb{C})$ acts on $\mathbb{M}_n$ through

$$G.A = GAG^{-1}; \;\; \forall \, G \in \mathbb{GL}(n, \mathbb{C}), A \in \mathbb{M}_n, \tag{1.0.1}$$

this will be shown in section 2.2. The *orbit* O_A of the element A in $\mathbb{M}_n$, are all the elements in $\mathbb{M}_n$ of the form $G.A$ or GAG^{-1}, where G is an element in the group acting on an element A in the set subject to the group action. This is expressed as

$$O_A = \{G.A \colon G \in \mathbb{GL}(n, \mathbb{C})\} \tag{1.0.2}$$

and is called the orbit of the element $A \in \mathbb{M}_n$, under the group action of the group $\mathbb{GL}(n, \mathbb{C})$ on the set $\mathbb{M}_n$.

Definition 1.2: Invariant
A function $I \colon \mathbb{M}_n \to \mathbb{C}$ is an invariant to $\mathbb{GL}(n, \mathbb{C})$ if

$$I(G.A) = I(A), \ \ \forall\, G \in \mathbb{GL}(n, \mathbb{C}). \tag{1.0.3}$$

Two well-known invariant polynomial functions of a matrix A, is the *trace* and the *determinant* of A.

If $A = (a_{ij})_{1 \leq i < j \leq n}$, we have that the *trace* of A

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii} \tag{1.0.4}$$

and the *determinant* of A

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)} \tag{1.0.5}$$

are polynomial functions originated in a polynomial in the elements of A. As they are invariants, this means that for the polynomial function p, it needs to be the case that

$$p(G.A) = p(A); \ \ \forall\, G \in \mathbb{GL}(n, \mathbb{C}), \ \forall\, A \in \mathbb{M}_n. \tag{1.0.6}$$

From (1.0.1) and (1.0.2), along with (1.0.6) with $p = \det$ and $p = \operatorname{tr}$ respectively, we have

$$\det(GAG^{-1}) = \det(G(AG^{-1})) = \det((AG^{-1})G) = \det(A(G^{-1})G) = \det(A)$$

$$\operatorname{tr}(GAG^{-1}) = \operatorname{tr}(A)$$

thus (1.0.6) is fulfilled, and the polynomial functions $\det(A)$ and $\operatorname{tr}(A)$ are invariants to $\mathbb{GL}(n, \mathbb{C})$.

The set of polynomial functions on $\mathbb{M}_n$ forms a ring denoted $\mathcal{P}(\mathbb{M}_n)$. The set of invariant polynomial functions $\mathcal{P}(\mathbb{M}_n)^{\mathbb{G}}$ is a subring of $\mathcal{P}(\mathbb{M}_n)$, in which the trace and the determinant are two elements. In this thesis we aim to describe the ring $\mathcal{P}(\mathbb{M}_n)^{\mathbb{G}}$.

To do this we will look at the properties of permutations and linear transformations associated with them, and how they act on vector spaces. The characteristic polynomial of a matrix

$$p_A(t) = t^n - p_1(A)t^{n-1} + p_2(A)t^{n-2} + \ldots + (-1)^{n-1}p_{n-1}(A)t + (-1)^n p_n(A)$$

will be central to this investigation and its connection to these special invariants. We will see that for all n the trace and the determinant are the coefficient of t^{n-1} and the constant term in the characteristic polynomial

$$\mathrm{tr}(A) = p_1(A)$$

$$\det(A) = p_n(A).$$

The other polynomials $p_2(A), p_3(A), \ldots, p_{n-1}(A)$ does not have any specific names, but all polynomials functions $p_i(A)$ will have the property

$$p_i(GAG^{-1}) = p_i(A) \tag{1.0.7}$$

in other words, they are invariants to $\mathbb{GL}(n, \mathbb{C})$.

We will also look at properties of polynomials, mainly symmetric polynomials and use the fundamental theorem of symmetric polynomials to determine all p_i for diagonal matrices. We will then extend that into the diagonalizable matrices, and finally arrive at the main result for all matrices.

We finish by showing that all elements in $\mathcal{P}(\mathbb{M}_n)^{\mathbb{G}}$ can be written as polynomials in the invariants p_i (1.0.7).

2 Permutations and Linear Transformations

Permutations and the linear transformations on vector spaces associated with them, will be a core concept of this thesis. We will therefore begin with some definitions of these, and then proceed with some fundamental properties of them, namely *invertibility* and *group homomorphisms*

Definition 2.1: Permutation
A permutation σ of a non-empty, finite set X, is a bijective map from X to itself

$$\sigma\colon X \to X.$$

A permutation is a mapping that interchanges the positions of the elements in the set being subject to it, we use the standard notation σ to denote an arbitrary permutation.

The group of permutations of the set of natural numbers $\mathbb{N} = \{1, 2, ..., n\}$ is denoted S_n, and is called the symmetric group of degree n.

Let $\sigma \in S_n$ be an arbitrary permutation, and let $\boldsymbol{e} = \{e_1, e_2, ..., e_n\}$ be a basis in an n-dimensional vector space over $\mathbb{C}$, denoted $\mathbb{V}$. We now define a linear transformation f_σ on $\mathbb{V}$, associated with the permutation σ as a linear transformation that permutes the basis elements of $\boldsymbol{e}$ by

$$f_\sigma(e_i) = e_{\sigma(i)} \tag{2.0.1}$$

For an arbitrary vector $\vec{v}$, we then have

$$f_\sigma(a_1e_1 + a_2e_2 + ... + a_ne_n) = a_1e_{\sigma(1)} + a_2e_{\sigma(2)} + ... + a_ne_{\sigma(n)}. \tag{2.0.2}$$

If we in (2.0.2) let σ be the identity permutation, $\sigma_{\mathbb{1}}$, that maps each element onto itself $\sigma_{\mathbb{1}}(1) = 1, \sigma_{\mathbb{1}}(2) = 2, \ldots, \sigma_{\mathbb{1}}(n) = n$, we see that the transformation associated with it, $f_{\sigma_{\mathbb{1}}}$, will map $\vec{v}$ onto itself, since we in (2.0.1) would have $\sigma_{\mathbb{1}}(e_1) = e_1, \sigma_{\mathbb{1}}(e_2) = e_2, \ldots, \sigma_{\mathbb{1}}(e_n) = e_n$. We have

$$\begin{aligned} f_{\sigma_{\mathbb{1}}}(a_1e_1 + a_2e_2 + ... + a_ne_n) = a_1e_{\sigma_{\mathbb{1}}(1)} + a_2e_{\sigma_{\mathbb{1}}(2)} + ... + a_ne_{\sigma_{\mathbb{1}}(n)} = \\ = (a_1e_1 + a_2e_2 + ... + a_ne_n). \end{aligned} \tag{2.0.3}$$

This is the identity transformation associated with the identity permutation, that will map every element of $\mathbb{V}$ onto itself.

With the basis $\boldsymbol{e} = \{e_1, e_2, ..., e_n\}$ for the vector space $\mathbb{V}$, we can express the linear transformation associated with a permutation as a $n \times n$–matrix in that basis, where n is the dimension of the vector space. A permutation matrix is a square matrix of dimension n composed of ones and zeros, it has one 1–entry per row and column and 0–entries everywhere else, we will show this in the next section.

If we then take the identity permutation $\sigma_{\mathbb{1}}$, and express it as a permutation matrix in the basis $\boldsymbol{e}$, it will be the identity matrix of dimension n, denoted E_n, since it has to map every element of the permuted object onto itself. The transformation associated with it $f_{\sigma_{\mathbb{1}}}$, expressed as a matrix in the basis $\boldsymbol{e}$, will then also be the identity matrix.

The matrix representation of (2.0.3) is

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{\boldsymbol{e}} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{n-1} \\ a_n \end{pmatrix}_{\boldsymbol{e}} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{n-1} \\ a_n \end{pmatrix}_{\boldsymbol{e}} .$$

We now look at a non-trivial example of a permutation in S_n. Let for example $\beta = (1\ 2\ 3\ ...\ n)$ be a cyclic permutation of length n, which maps the first element in $\mathbb{N}$ onto the second, the second onto the third and so on. Then

$$\begin{aligned} f_\beta(a_1 e_1 + a_2 e_2 + ... + a_n e_n) &= a_1 e_{\beta(1)} + a_2 e_{\beta(2)} + ... + a_n e_{\beta(n)} = \\ &= (a_1 e_2 + a_2 e_3 + ... + a_{n-1} e_n + a_n e_1) \end{aligned} \tag{2.0.4}$$

and hence the corresponding matrix A_β for this permutation in the basis $\boldsymbol{e}$ is

$$A_\beta = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{\boldsymbol{e}}$$

which gives the matrix representation of (2.0.4) as

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{\boldsymbol{e}} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{n-2} \\ a_{n-1} \\ a_n \end{pmatrix}_{\boldsymbol{e}} = \begin{pmatrix} a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_{n-1} \\ a_n \\ a_1 \end{pmatrix}_{\boldsymbol{e}} .$$

2.1 Properties of the Linear Transformation f_σ

A very fundamental property of f_σ is that if we have two arbitrary permutations τ and π, and let the linear transformations associated with them act on an arbitrary vector $\vec{v} \in \mathbb{V}$, it does not matter if we let them act on $\vec{v}$ separately, or combining the transformations and then act with the combined transformation on $\vec{v}$.

Theorem 2.1. *The composed image of the separate transformations will be the same as the image of the composed transformation*

$$f_\pi f_\tau(\vec{v}) = f_{\pi\tau}(\vec{v}). \tag{2.1.1}$$

Proof. First we show that (2.1.1) holds for a basis element $e_i \in \mathbb{V}$

$$(f_\pi f_\tau)(e_i) = f_{\pi\tau}(e_i).$$

The left hand-side of the equation can be rewritten as

$$(f_\pi f_\tau)(e_i) = (f_\pi)(f_\tau(e_i)) = (f_\pi)(e_{\tau(i)}) = (f_\pi)(e_j) = e_{\pi(j)} = e_{\pi\tau(i)} = f_{\pi\tau}(e_i)$$

since f_σ is a linear transformation, we can generalize to an arbitrary vector $\vec{v} \in \mathbb{V}$

$$(f_\pi f_\tau)(\vec{v}) = f_{\pi\tau}(\vec{v})$$

and the proof is complete. □

Invertibility of f_σ:

We will now prove that f_σ is invertible as a linear transformation, i.e. that there for all linear transformations f_σ, there exists a linear transformation f_σ^{-1} that is the inverse transformation of f_σ. We state this as a theorem.

Theorem 2.2. *For all linear transformations f_σ, there exists a linear transformation f_σ^{-1} that is the inverse transformation of f_σ.*

Proof. The fact that f_σ has the property (2.1.1), then gives us

$$f_\sigma f_{\sigma^{-1}} = f_{\sigma\sigma^{-1}} = f_{\sigma_{\mathbb{1}}}$$

and we have shown that the linear transformation of the inverse permutation σ^{-1}, is in fact the inverse of the linear transformation of the permutation σ

$$f_{\sigma^{-1}} = f_\sigma^{-1}$$

which concludes the proof. □

2.2 Group Homomorphisms

The property (2.1.1) of f_σ tells us that it is a transformation that preserves matrix multiplication. To explain what this means we look at the concept of a *group homomorphism*.

The permutations σ are elements of the symmetric group of degree n, S_n, and the matrix representations of the linear transformations associated with the permutations f_σ are invertible square matrices of dimension n.

Let

$$g\colon \mathbb{GL}(n,\mathbb{C}) \times \mathbb{M}_n \to \mathbb{M}_n \tag{2.2.1}$$

be the function

$$g(G,A) = G.A = GAG^{-1};\ \forall\, G \in \mathbb{GL}(n,\mathbb{C}), A \in \mathbb{M}_n. \tag{2.2.2}$$

Theorem 2.3. *The function g, from the Cartesian product of $\mathbb{GL}(n,\mathbb{C})$ and $\mathbb{M}_n$ to $\mathbb{M}_n$, defines a group action by the group $\mathbb{GL}(n,\mathbb{C})$ on the set $\mathbb{M}_n$.*

Proof. If we look at the definition of a group action, Definition 1.1, we see that for (2.2.2) to be a group action the element GMG^{-1} has to be an element of $\mathbb{M}_n$, that is a $n\times n$–matrix. Since the elements G and G^{-1} in $\mathbb{GL}(n,\mathbb{C})$ are invertible $n\times n$–matrices, and the product of any number of $n\times n$–matrices is also a $n\times n$–matrix, we know that the first part of the definition of a group action is satisfied.

For the first Axiom, Axiom 1.1.1, we see that the composition of the elements G and G^{-1} in the group $\mathbb{GL}(n,\mathbb{C})$ acting on an element $A \in \mathbb{M}_n$, has to be the same as the elements G and G^{-1} in the group $\mathbb{GL}(n,\mathbb{C})$ acting on an element $A \in \mathbb{M}_n$, and then composed in the set element $\mathbb{M}_n$. So we want to show that

$$(G_1G_2).A = G_1.(G_2.A);\quad \forall\, G \in \mathbb{GL}(n,\mathbb{C}), A \in \mathbb{M}_n.$$

We begin with applying the definition of the group action (2.2.2) to the left hand side

$$(G_1G_2).A = (G_1G_2)A(G_1G_2)^{-1}$$

the properties of inverses gives

$$(G_1G_2)A(G_1G_2)^{-1} = (G_1G_2)A(G_2^{-1}G_1^{-1})$$

and by the law of associativity

$$(G_1G_2)A(G_2^{-1}G_1^{-1}) = G_1(G_2AG_2^{-1})G_1^{-1}.$$

We can now see that the expression in the parenthesis is exactly the action of the element G_2 on the element A

$$G_1(G_2AG_2^{-1})G_1^{-1} = G_1(G_2.A)G_1^{-1}$$

and then the right hand side is the action of the element G_1 on the element $G_2.A$, so we have

$$G_1(G_2.A)G_1^{-1} = G_1.(G_2.A)$$

thus we have concluded that

$$(G_1G_2).A = G_1.(G_2.A)$$

and the first Axiom is fulfilled.

The second Axiom, Axiom 1.1.2, states that the identity element in the group $\mathbb{GL}(n, \mathbb{C})$ acting on an element $A \in \mathbb{M}_n$, has to yield the same element $A \in \mathbb{M}_n$ back again. Since the identity element in $\mathbb{GL}(n, \mathbb{C})$ is the identity matrix, and since any matrix A multiplied by the identity matrix E is equal to A, this Axiom is fulfilled.

Then, since the first part of the definition and the two Axioms of a group action is fulfilled by the group action g, the proof of (2.2.2) is complete. □

which in other words means that the group action g is *conjugation* of the element $A \in \mathbb{M}_n$ by the element $G \in \mathbb{GL}(n, \mathbb{C})$.

Definition 2.3: Group Homomorphism φ
A group homomorphism is a function/mapping φ from a group G to another group G' that preserves the algebraic structure of the groups, the property can be expressed as

$$\begin{gathered} \varphi \colon G \to G' \\ \varphi(a \circ b) = \varphi(a) \diamond \varphi(b) \; ; \forall \quad a, b \in G \end{gathered} \tag{2.2.3}$$

where the notation $\circ$ is for composition in the group G, and $\diamond$ is for composition in the group G'. In words this would mean that it does not matter if you compose the elements a and b in the group G and then act with φ on the combined element $a \circ b$, left hand side of (2.2.3), or if you act with φ on each element $a, b \in G$ first and then compose the result in G', right hand side of (2.2.3).

The notation for composition is very often omitted, and since we will be dealing with compositions of permutations and compositions of the linear transformations associated with them from now on, we will not be using it.

Now we look at the specific groups we are interested in, the symmetric group of degree n, S_n and the general linear group of degree n over $\mathbb{C}$, $\mathbb{GL}(n, \mathbb{C})$.

Theorem 2.4. *Let g be the function from S_n to $\mathbb{GL}(n, \mathbb{C})$*

$$\left.\begin{aligned} g &\colon S_n \to \mathbb{GL}(n, \mathbb{C}) \\ g &\colon \sigma \mapsto f_\sigma \end{aligned}\right\} \quad \sigma \in S_n, \quad f_\sigma \in \mathbb{GL}(n, \mathbb{C}) \tag{2.2.4}$$

The function g is a group homomorphism from S_n to $\mathbb{GL}(n, \mathbb{C})$.

Proof. To prove (2.2.4), we need to show that it obeys the criteria of a group homomorphism (2.2.3). With our groups being S_n and $\mathbb{GL}(n, \mathbb{C})$, we want to show that

$$g(\pi\tau) = g(\pi)g(\tau) \ ; \forall \ \ \pi, \tau \in S_n.$$

The function g maps an arbitrary permutation onto the linear transformation associated with it $g(\sigma) = f_\sigma$, we have

$$g(\pi\tau) = f_{\pi\tau}$$
$$g(\pi)g(\tau) = f_\pi f_\tau$$

and (2.1.1) gives us that composition in S_n is the same as composition in $\mathbb{GL}(n, \mathbb{C})$

$$f_{\pi\tau} = f_\pi f_\tau$$

i.e. that g is a group homomorphism, so

$$g(\pi\tau) = g(\pi)g(\tau) \ ; \forall \ \ \pi, \tau \in S_n$$

and the proof is complete. □

3 The Permutation Matrix Properties

In this section we will investigate some of the permutation matrices properties, which are the nature of the permutation matrix *elements*, and the *inverse* and the *transpose* of the matrix. These properties will be important when we look at the group action of $\mathbb{GL}(n, \mathbb{C})$ on $\mathbb{M}_n$.

The Permutation Matrix elements:

In the previous section it was stated that the matrix A_σ that represents the linear transformation f_σ in the basis $\boldsymbol{e}$, has exactly one 1-entry per row and column and 0-entries everywhere else. We are now going to prove that, and we begin with the general form of the matrix-entries. To describe the elements of A_σ we are going to use the *Kronecker delta* function, which is defined as

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{else.} \end{cases}$$

Theorem 3.1. *The element in position* (i, j) *in* A_σ *is*

$$a_{ij} = \delta_{i\sigma(j)},\ \delta_{i\sigma(j)} = \begin{cases} 1 & \textit{if} \quad i = \sigma(j) \\ 0 & \textit{else.} \end{cases} \tag{3.0.1}$$

Proof. If we have a basis $\boldsymbol{e}$ for the vector space $\mathbb{V}$, and a linear transformation T on $\mathbb{V}$, the matrix for the linear transformation in the basis $\boldsymbol{e}$ will have the images of the basis vectors as columns

$$[T]_{\boldsymbol{e}} = \begin{pmatrix} | & | & & | \\ T(e_1) & T(e_2) & \dots & T(e_n) \\ | & | & & | \end{pmatrix}_{\boldsymbol{e}}.$$

All linear transformations on the vector space $\mathbb{V}$ in the basis $\boldsymbol{e}$, associated with permutations σ, will then have images of the basis vectors that are also basis vectors. It will simply permute the *order* of the columns. This means that the matrix of a permutation transformation is a permutation of the columns in the identity matrix, in said basis $\boldsymbol{e}$

$$[E_n]_{\boldsymbol{e}} = \begin{pmatrix} | & | & & | \\ e_1 & e_2 & \dots & e_n \\ | & | & & | \end{pmatrix}_{\boldsymbol{e}} ;\ [A_\sigma]_{\boldsymbol{e}} = \begin{pmatrix} | & | & & | \\ \sigma(e_1) & \sigma(e_2) & \dots & \sigma(e_n) \\ | & | & & | \end{pmatrix}_{\boldsymbol{e}}.$$

The column j in A_σ is the image of the basis vector e_j, which is $e_{\sigma(j)}$. It has an 1-entry in position $\sigma(j)$ and 0-entries everywhere else.

When going through the rows $i = 1, 2, \dots, n$ in A_σ, the 1-entry in that row will be in the column $\sigma^{-1}(i)$, since if an element in A_σ is 1 we have

$$i = \sigma(j) \Leftrightarrow j = \sigma^{-1}(i)$$

then we can, for every row i find a column with an 1-entry, when $i = \sigma(j)$. And every other column in that row will have 0-entries, when $i \neq \sigma(j)$. Which means that we can write the elements in column j as

$$\delta_{i\sigma(j)}.$$

This gives that

$$a_{ij} = \delta_{i\sigma(j)}$$

where a_{ij} is the element of A_σ in position i, j, and we have proved (3.0.1) □

Inverse and Transpose of A_σ

From this follows directly that the permutation matrix is an *orthogonal* matrix, which means that all the columns are *orthonormal* vectors, they are all unit vectors that are perpendicular to each other. All orthogonal matrices are invertible and their *inverse* is equal to the *transpose*, thus for permutation matrices, the inverse is equal to the transpose

$$A_\sigma^{-1} = A_\sigma^t.$$

The elements of $A_\sigma^{-1} = A_\sigma^t$ will look similar to the elements of A_σ, but with the difference that it is the row-index that is being permuted by σ

$$A_\sigma^{-1} = A_\sigma^t,\ a_{ij}^t = a_{ji} = \delta_{j\sigma(i)} \tag{3.0.2}$$

where a_{ij}^t is the element in position ij in the matrix A_σ^t, and they are the position-transposed elements a_{ji} of the elements a_{ij} in the matrix A_σ. Also, since the transpose of a permutation matrix is also a permutation matrix, the inverse will be a permutation matrix as well.

3.1 Group Actions of Permutation Matrices

As we saw in the preceding section, all permutation matrices A_σ are invertible with inverse A_σ^{-1} equal to the transpose A_σ^t. Since the set of all invertible matrices in $\mathbb{M}_n$ is $\mathbb{G}$, and $\mathbb{G}$ is a group under matrix multiplication, $\mathbb{GL}(n, \mathbb{C})$, the set of all permutation matrices A_σ for $\sigma \in S_n$, denoted $\mathbb{A}$, is a subset of $\mathbb{GL}(n, \mathbb{C})$

$$\mathbb{A} \subset \mathbb{GL}(n, \mathbb{C}).$$

Thus we have from (2.2.1), that $\mathbb{A}$ acts on $\mathbb{M}_n$ through

$$A_\sigma.B = A_\sigma B A_\sigma^{-1};\ \ \forall\, A_\sigma \in \mathbb{A}, B \in \mathbb{M}_n \tag{3.1.1}$$

as well as

$$A_\sigma.B = A_\sigma B \tag{3.1.2}$$

$$A_\sigma.B = B A_\sigma^{-1}. \tag{3.1.3}$$

We will now investigate this group action, and look at what happens in the three cases. The first one (3.1.1) is exactly *conjugation* of B by A_σ, we will look at a basic example of this when B is a diagonal 3×3–matrix, to get the structure apparent.

A common notation for diagonal matrices we will use, to avoid writing it out, is to just write the diagonal elements $\text{diag}(d_1, d_2, \dots, d_n)$. Since all other elements are, by definition of a diagonal matrix, zero, we do not lose any information by doing so

$$D = \begin{pmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_n \end{pmatrix} = \text{diag}(d_1, d_2, d_3, \dots, d_n).$$

Example 3.1:

Let:

$$A_\sigma = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \ , \ A_\sigma^{-1} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \ , \ D = \text{diag}(d_1, d_2, d_3).$$

If we conjugate D with A_σ, we get

$$A_\sigma D A_\sigma^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} d_2 & 0 & 0 \\ 0 & d_3 & 0 \\ 0 & 0 & d_1 \end{pmatrix} = \\ = \text{diag}(d_2, d_3, d_1)$$

and

$$A_\sigma^{-1} D A_\sigma = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} d_3 & 0 & 0 \\ 0 & d_1 & 0 \\ 0 & 0 & d_2 \end{pmatrix} = \\ = \text{diag}(d_3, d_1, d_2).$$

We see that conjugation by a permutation matrix will interchange the elements of D. In the first case, $A_\sigma D A_\sigma^{-1}$, we have two row-interchanges and two column-interchanges: rows 1 and 3, followed by rows 1 and 2 and then columns 1 and 3, followed by rows 1 and 2. In the second case, $A_\sigma^{-1} D A_\sigma$, we have the same two row-interchanges, and for the columns we have columns 1 and 2, same as in the first case, but also columns 2 and 3 which is different from the first case.

Now we will proceed with the group actions in equations (3.1.1), (3.1.2) and (3.1.3), acting on arbitrary square matrices B and determine the properties of the matrices $A_\sigma B$, BA_σ^{-1} and $A_\sigma BA_\sigma^{-1}$. But before we get into those in detail, we will formulate a more compact notation for matrix multiplication.

If we then start with the base case of a multiplication of two square matrices A and B, both with dimension n, we have

$$AB = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_{k=1}^{n} a_{1k}b_{k1} & \sum_{k=1}^{n} a_{1k}b_{k2} & \cdots & \sum_{k=1}^{n} a_{1k}b_{kn} \\ \sum_{k=1}^{n} a_{2k}b_{k1} & \sum_{k=1}^{n} a_{2k}b_{k2} & \cdots & \sum_{k=1}^{n} a_{2k}b_{kn} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} a_{nk}b_{k1} & \sum_{k=1}^{n} a_{nk}b_{k2} & \cdots & \sum_{k=1}^{n} a_{nk}b_{kn} \end{pmatrix} \implies (AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

where $(AB)_{ij}$ refers to the element in position ij in the matrix AB. This notation, with a sum, is the more compact notation we are going to use. The case of three matrices multiplied together is calculated analogous to the two-matrix case, the only difference is that we get a double sum since we are performing another matrix multiplication. So the compact notation for the general cases of two matrices A and B and three matrices A, B and C, multiplied together, will have the structure

$$(AB)_{ij} = \sum_{k} a_{ik}b_{kj} \tag{3.1.4}$$

$$(ABC)_{ij} = \sum_{k,l} a_{ik}b_{kl}c_{lj} \tag{3.1.5}$$

where $(AB)_{ij}$ refers to the element in position ij in the matrix AB, and $(ABC)_{ij}$ refers to the element in position ij in the matrix ABC.

Now we will proceed with the group actions above, and see what effect the different cases have on B. Let B be an arbitrary $n \times n$–matrix, and let A_σ be an arbitrary permutation matrix.

$A_\sigma B$:

For the matrix in (3.1.2), we have a permutation matrix in place of the more general matrix A in (3.1.4), and since the element a_{ik} is determined by equation (3.0.2), we get (3.1.4) as

$$(A_\sigma B)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = \sum_{k=1}^{n} \delta_{i\sigma(k)}b_{kj}.$$

The sum over k will only have non-zero terms when $i = \sigma(k)$, and since σ is a permutation with an inverse, the k that gives non-zero terms is

$$i = \sigma(k) \Leftrightarrow k = \sigma^{-1}(i).$$

Putting this into the sum we get

$$\sum_{k=1}^{n} \delta_{i\sigma(k)} b_{kj} = b_{\sigma^{-1}(i)j}$$

since all other terms in the sum where $k \neq \sigma^{-1}(i)$ have an $\delta_{i\sigma(k)} = 0$, we have only the term left where $k = \sigma^{-1}(i)$ which yields $\delta_{i\sigma(k)} = 1$ times the b with $k = \sigma^{-1}(i)$.

$\boldsymbol{BA_{\sigma}^{-1}}$**:**

With equation (3.1.4) the matrix in (3.1.3) will be

$$\left(BA_{\sigma}^{-1}\right)_{ij} = \sum_{k=1}^{n} b_{ik} a_{jk} \tag{3.1.6}$$

which will be calculated with (3.0.2) as we did before

$$\sum_{k=1}^{n} b_{ik} a_{jk} = \sum_{k=1}^{n} b_{ik} \delta_{j\sigma(k)}$$

$$j = \sigma(k) \Leftrightarrow k = \sigma^{-1}(j)$$

that gives us

$$\sum_{k=1}^{n} b_{ik} \delta_{j\sigma(k)} = b_{i\sigma^{-1}(j)}.$$

$\boldsymbol{A_{\sigma}BA_{\sigma}^{-1}}$**:**

For the matrix in (3.1.1), we use (3.1.5). But since C is a another permutation matrix in this case, A_{σ}^{-1}, we have two a-factors in each term of the sum

$$\left(A_{\sigma}BA_{\sigma}^{-1}\right)_{ij} = \sum_{k,l} a_{ik} b_{kl} a_{lj}^{t} \tag{3.1.7}$$

using (3.0.2) for a_{ik} and (3.0.3) for a_{lj}^{t}, we get

$$\sum_{k,l} a_{ik} b_{kl} a_{lj}^{t} = \sum_{k,l} a_{ik} b_{kl} a_{jl} = \sum_{k,l} \delta_{i\sigma(k)} b_{kl} \delta_{j\sigma(l)}.$$

We use the same approach as previously, only this time we will do it twice, once for the sum over k and once for the sum over l. The sum over k,

$$\sum_k \delta_{i\sigma(k)} b_{kl} \delta_{j\sigma(l)},$$

will only have a non-zero term when

$$i = \sigma(k) \Leftrightarrow k = \sigma^{-1}(i).$$

This give us

$$\sum_{k,l} \delta_{i\sigma(k)} b_{kl} \delta_{j\sigma(l)} = \sum_l b_{\sigma^{-1}(i)l} \delta_{j\sigma(l)},$$

and then the sum over l, in the same way, will only have a non-zero term when

$$j = \sigma(l) \Leftrightarrow l = \sigma^{-1}(j),$$

so we get

$$\sum_l b_{\sigma^{-1}(i)l} \delta_{j\sigma(l)} = b_{\sigma^{-1}(i)\sigma^{-1}(j)}.$$

We then have the important expression for the elements in an arbitrary matrix B, conjugated by a permutation matrix A_σ

$$\left(A_\sigma B A_\sigma^{-1}\right)_{ij} = b_{\sigma^{-1}(i)\sigma^{-1}(j)}. \tag{3.1.8}$$

Comparing the order of the indices in (3.1.4) and (3.1.6) as well as in (3.1.5) and (3.1.7), with (3.0.2), we can see that multiplying a square matrix B with a permutation matrix A_σ will interchange rows if we multiply B with A_σ from the left. This is because the indices of the second matrix has swapped places $k \rightleftharpoons j$. We will have interchange of columns if multiplied from the right. Consequently, multiplying from both sides will yield both an interchange of rows and columns.

As we saw previously the sum (3.1.7) will reduce to just one number when we apply equations (3.0.1) and (3.0.2). We will now let the arbitrary matrix conjugated by the permutation matrix be a diagonal matrix D, and show that this will give us a permutation of the non-zero elements in D, that is the diagonal elements.

Theorem 3.2. *Let D be an arbitrary diagonal matrix. The conjugation of D by the permutation matrix A_σ permutes the diagonal elements in D as*

$$A_\sigma D A_\sigma^{-1} = diag(d_{\sigma^{-1}(1)}, d_{\sigma^{-1}(2)}, \ldots, d_{\sigma^{-1}(n)}). \tag{3.1.9}$$

Proof. The only difference between the case with a diagonal matrix and an arbitrary square matrix B is that we only have non-zero entries on the diagonal, we can simply apply (3.1.8) to the matrix in (3.1.9). But with the difference that we will only have non-zero entries when $j = i$, so we can simplify the notation and drop one of the subscripts, as we did earlier for diagonal matrices since they are always the same

$$d_{\sigma^{-1}(i)\sigma^{-1}(j)} = d_{\sigma^{-1}(i)\sigma^{-1}(i)} = d_{\sigma^{-1}(i)}$$

we get

$$\left(A_\sigma D A_\sigma^{-1}\right)_i = d_{\sigma^{-1}(i)}.$$

We have then that for each element in the diagonal matrix D, the conjugation by the permutation matrix A_σ will permute the element by σ^{-1}. Indeed

$$A_\sigma D A_\sigma^{-1} = \mathrm{diag}(d_{\sigma^{-1}(1)}, d_{\sigma^{-1}(2)}, \ldots, d_{\sigma^{-1}(n)})$$

which concludes the proof. □

We then have concluded that the group actions of $\mathbb{A}$ on $\mathbb{M}_n$, will result in

$$\begin{aligned}
&A_\sigma.B = A_\sigma B A_\sigma^{-1} \ \text{ gives } \ \left(A_\sigma B A_\sigma^{-1}\right)_{ij} = b_{\sigma^{-1}(i)\sigma^{-1}(j)} \\
&A_\sigma.B = A_\sigma B \ \text{ gives } \ \left(A_\sigma B\right)_{ij} = b_{\sigma^{-1}(i)\ j} \\
&A_\sigma.B = B A_\sigma^{-1} \ \text{ gives } \ \left(B A_\sigma^{-1}\right)_{ij} = b_{i\ \sigma^{-1}(j)}
\end{aligned}$$

i.e. a permutation of rows and columns in the first case, and permutations of rows and columns in the second and third case respectively. For the subsequent sections however, we will mainly be interested in the first one. The other two were mainly calculated to show the connection of left- and right-side multiplication with conjugation.

4 The Characteristic Polynomial

The characteristic polynomial to a $n \times n$–matrix is given by

$$p_A(t) = \det(tE_n - A) \tag{4.0.1}$$

where E_n is the $n \times n$–identity matrix.

$p_A(t)$ has the form

$$p_A(t) = t^n - p_1(A)t^{n-1} + p_2(A)t^{n-2} + \ldots + (-1)^{n-1}p_{n-1}(A)t + (-1)^n p_n(A) \tag{4.0.2}$$

where $p_i(A)$ is a polynomial of degree i in the elements of A. The polynomials $p_1(A)$ and $p_n(A)$ are the invariants (1.0.4) and (1.0.5) we looked at in section 1, the trace and the determinant of A respectively. We are now going to prove the expressions (1.0.4) and (1.0.5), and we formalize by stating this as a theorem.

Theorem 4.1. *Let A be an arbitrary $n \times n$-matrix over $\mathbb{C}$, the trace and the determinant of A are given by*

$$p_1(A) = \operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii} \tag{4.0.3}$$

$$p_n(A) = \det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)} \tag{4.0.4}$$

respectively.

Proof. (4.0.4) To prove this, we start with the cases of a 2×2–matrix and a 3×3–matrix
2×2–matrix, $n = 2$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad p_A(t) = \det(tE_n - A) = \begin{vmatrix} t - a_{11} & a_{12} \\ a_{21} & t - a_{22} \end{vmatrix} =$$
$$= (t - a_{11})(t - a_{22}) - (a_{12}a_{21}) = a_{11}a_{22} - a_{11}t - a_{22}t + t^2 - a_{12}a_{21} =$$
$$= t^2 - t(a_{11} + a_{22}) + a_{11}a_{22} - a_{12}a_{21}$$

We can identify that $p_1(A) = \operatorname{tr}(A)$ and $p_2(A) = \det(A)$.

3×3–matrix, $3 = 2$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad p_A(t) = \det(tE_n - A) = \begin{vmatrix} t - a_{11} & a_{12} & a_{13} \\ a_{21} & t - a_{22} & a_{23} \\ a_{31} & a_{32} & t - a_{33} \end{vmatrix} =$$
$$= (t - a_{11}) \begin{vmatrix} t - a_{22} & a_{23} \\ a_{32} & t - a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & t - a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & t - a_{22} \\ a_{31} & a_{32} \end{vmatrix} =$$
$$= t^3 - t^2(a_{11} + a_{22} + a_{33}) + t(-a_{11}a_{22} - a_{11}a_{33} + a_{23}a_{32} - a_{22}a_{33} + a_{12}a_{21} + a_{13}a_{31})$$
$$+ a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22}$$

Also in this we have that $p_1(A) = \mathrm{tr}(A)$ and $p_3(A) = \det(A)$

$\boldsymbol{n \times n}$**–matrix,** $\boldsymbol{n = n}$:

The final step then, is to show that $p_1(A) = \mathrm{tr}(A)$ and $p_n(A) = \det(A)$ for any sized square matrix. The proof for the latter, is very straight forward.

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad \det(tE_n - A) = \begin{vmatrix} t - a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & t - a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & t - a_{nn} \end{vmatrix}.$$

Expanding the determinant gives us the characteristic polynomial

$$p_A(t) = t^n - p_1(A)t^{n-1} + p_2(A)t^{n-2} + \ldots + (-1)^{n-1}p_{n-1}(A)t + (-1)^n p_n(A)$$

and since what we are interested in here is the final term $p_n(A)$, which does not have any t-factor, we can just set $t = 0$

$$p_A(0) = 0^n - p_1(A)0^{n-1} + p_2(A)0^{n-2} + \ldots + (-1)^{n-1}p_{n-1}(A)0 + (-1)^n p_n(A) = \\ = (-1)^n p_n(A)$$

then $\det(tE_n - A)$ with $t = 0$ gives us $\det(0E_n - A) = \det(-A)$, and we can factor out n number of (-1)-factors from the determinant, this proves that

$$p_n(A) = \det(A)$$

and the proof is complete. □

Proof. (4.0.3) Next, we are going to prove that $p_1(A) = \mathrm{tr}(A)$ for any sized square matrix. To do this, we will start with expanding the determinant along the first row, and then do a proof by induction.

$$\det(tE_n - A) = \begin{vmatrix} t - a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & t - a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & t - a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & t - a_{nn} \end{vmatrix} =$$

$$= (t-a_{11})\begin{vmatrix} t-a_{22} & 0-a_{23} & \dots & 0-a_{2n} \\ 0-a_{32} & t-a_{33} & \dots & 0-a_{3n} \\ \vdots & \ddots & \ddots & \vdots \\ 0-a_{n2} & 0-a_{n3} & \dots & t-a_{nn} \end{vmatrix} - a_{12}\begin{vmatrix} 0-a_{21} & 0-a_{23} & \dots & 0-a_{2n} \\ 0-a_{31} & t-a_{33} & \dots & 0-a_{3n} \\ \vdots & \ddots & \ddots & \vdots \\ 0-a_{n1} & 0-a_{n3} & \dots & t-a_{nn} \end{vmatrix}$$

$$+ a_{13}\begin{vmatrix} 0-a_{21} & t-a_{22} & \dots & 0-a_{2n} \\ 0-a_{31} & 0-a_{32} & \dots & 0-a_{3n} \\ \vdots & \ddots & \ddots & \vdots \\ 0-a_{n1} & 0-a_{n2} & \dots & t-a_{nn} \end{vmatrix} + \dots$$

$$\dots + (-1)^n a_{1n}\begin{vmatrix} 0-a_{21} & t-a_{22} & \dots & 0-a_{2(n-1)} \\ 0-a_{31} & 0-a_{32} & \dots & 0-a_{3(n-1)} \\ \vdots & \ddots & \ddots & \vdots \\ 0-a_{n1} & 0-a_{n2} & \dots & t-a_{n(n-1)} \end{vmatrix}.$$

We can see that the determinant with the factor $t - a_{11}$ has $n-1$ number of t-factors and that all the ones following that one, the ones with factors a_{12}, $-a_{13} \dots (-1)^n a_{1n}$, has $n-2$ number of t-factors. So the only term in the sum that will contribute to the polynomial $p_1(A)$, is the first one

$$(t-a_{11})\begin{vmatrix} t-a_{22} & 0-a_{23} & \dots & 0-a_{2n} \\ 0-a_{32} & t-a_{33} & \dots & 0-a_{3n} \\ \vdots & \ddots & \ddots & \vdots \\ 0-a_{n2} & 0-a_{n3} & \dots & t-a_{nn} \end{vmatrix} = (t-a_{11})\det(tE_n - A').$$

We denote this determinant by $\det(tE_n - A')$, since we will need a name for it in the polynomial to follow. The matrix A' is the reduction of A by one row and one column

$$A' = \begin{pmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}, \text{ where } a'_{ij} = a_{ij}\ .$$

We will now proceed with the proof by induction and the basis-case is already proven earlier, be it the 2×2–case or the 3×3–case. We now assume that for the n-1× n-1–case, the determinant is

$$\begin{vmatrix} t-a_{22} & 0-a_{23} & \dots & 0-a_{2n} \\ 0-a_{32} & t-a_{33} & \dots & 0-a_{3n} \\ \vdots & \ddots & \ddots & \vdots \\ 0-a_{n2} & 0-a_{n3} & \dots & t-a_{nn} \end{vmatrix} =$$

$$= t^{n-1} - (a_{22} + a_{33} + \dots + a_{nn})t^{n-2} + \dots + (-1)^{n-2}t + (-1)^{n-1}\det(A')$$

and by multiplying the polynomial by the factor we had in the original expression, $t - a_{11}$, we get

$$(t-a_{11})\left(t^{n-1} - (a_{11} + a_{22} + \dots + a_{nn})t^{n-2} + \dots + (-1)^{n-2}t + (-1)^{n-1}\det(A')\right)$$

only the terms $t\left(-(a_{22}+a_{33}+\ldots+a_{nn})t^{n-2}\right)$ and $(-a_{11})(t^{n-1})$ have degree $n-1$ of t, and therefore will contribute to the polynomial $p_1(A)$, thus we can discard all the other factors. The expanded form of the contributing factors then is

$$t\left(-(a_{22}+a_{33}+\ldots+a_{nn})t^{n-2}\right)+(-a_{11})(t^{n-1}) = -(a_{11}+a_{22}+a_{33}+\ldots+a_{nn})t^{n-1}$$

comparing this coefficient with the coefficient for $p_1(A)$ in the characteristic polynomial,

$$p_A(t) = t^n - p_1(A)t^{n-1} + p_2(A)t^{n-2} + \ldots + (-1)^{n-1}p_{n-1}(A)t + (-1)^n p_n(A)$$

shows that

$$p_1(A) = \mathrm{tr}(A)$$

then, by the induction hypothesis, the proof is complete. □

The roots of $p_A(t)$ are the eigenvalues to A, denoted λ_i, so we can express $p_A(t)$ in terms of the variable t and the eigenvalues λ_i as

$$p_A(t) = (t-\lambda_1)(t-\lambda_2)\ldots(t-\lambda_n). \tag{4.0.5}$$

From the facts that $\mathrm{tr}(A) = p_1(A)$ and $\det(A) = p_n(A)$ for $p_A(t)$ expressed in terms of $p_i(A)$ and t. It follows that with (4.0.2) expressed in terms of λ_i and t, the expansion of (4.0.5), we can express the trace and the determinant of A in terms of the eigenvalues.

From the characteristic polynomial (4.0.2) expressed as (4.0.5), and the method used to prove (4.0.4) where we set $t = 0$ and by factoring out n number of (-1)-terms, we have

$$\det(A) = p_A(0) = (0-\lambda_1)(0-\lambda_2)\ldots(0-\lambda_n) = (-1)^n(\lambda_1)(\lambda_2)\ldots(\lambda_n)$$

which give us

$$p_n(A) = \det(A) = \prod_{i=1}^{n} \lambda_i\ . \tag{4.0.6}$$

To get a term of degree $n-1$ in t, we have to factor out t from $n-1$ parenthesis in (4.0.5), and factor out λ from the n:th parenthesis. We can do this in n different ways, depending on what parenthesis we factor out λ from, and consequently do not factor out t from. The end result however, will be the same in that the coefficient will be the the sum of all λ_i

$$p_1(A) = (\lambda_1 + \lambda_2 + \ldots + \lambda_n)$$

which is the sum of the eigenvalues to A

$$p_1(A) = \mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i\ . \tag{4.0.7}$$

This is exactly the relationship between roots and coefficients of a polynomial, sometimes referred to as *Vieta's Formulas*, and we will look more closely at this relationship in the next section.

5 Polynomials and Matrices

In this section we are going to investigate polynomials and the polynomial functions originated in them more closely.

5.1 The Elementary Symmetric Polynomials

Definition 5.1: Symmetric Polynomials
A polynomial $f(x_1, x_2, \ldots, x_n)$ is said to be *symmetric* if a permutation of the variables does not change the polynomial, i.e. if

$$f(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}) = f(x_1, x_2, \ldots, x_n),\ \forall \sigma \in S_n \tag{5.1.1}$$

The relationship between roots of a polynomial and the coefficients of the polynomial is as follows. The roots of a polynomial and their multiplicity are unambiguously determined if the coefficients of the polynomial are known. And vice versa, the coefficients of the polynomial are unambiguously determined if the roots of a polynomial and their multiplicity are known. If the coefficient of the leading term, the one with the highest exponent, is set to 1, the polynomial with roots: $\alpha_1, \alpha_2, \ldots, \alpha_n$ is

$$p(t) = (t - \alpha_1)(t - \alpha_2) \ldots (t - \alpha_n). \tag{5.1.2}$$

If we execute the multiplication of the parentheses in (5.1.2) we get

$$p(t) = t^n - e_1(\boldsymbol{\alpha})t^{n-1} + e_2(\boldsymbol{\alpha})t^{n-2} - \ldots + (-1)^{n-1}e_{n-1}(\boldsymbol{\alpha})t + (-1)^n e_n. \tag{5.1.3}$$

To get a term of degree n we factor out t from n parenthesis, to get a term of degree $n-1$ we factor out t from $n-1$ parenthesis, and so on. We get the following relationship between roots and coefficients

$$\begin{aligned}
e_1(\boldsymbol{\alpha}) &= \sum \alpha_i \\
e_2(\boldsymbol{\alpha}) &= \sum \alpha_i \alpha_j \\
e_3(\boldsymbol{\alpha}) &= \sum \alpha_i \alpha_j \alpha_k \\
&\vdots \\
e_v(\boldsymbol{\alpha}) &= \sum_{1 \le i_1 < i_2 < \ldots < i_v \le n} \alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_v} \\
&\vdots \\
e_n(\boldsymbol{\alpha}) &= \alpha_1 \alpha_2 \ldots \alpha_n
\end{aligned}$$

where the coefficients $e_i(\boldsymbol{\alpha}) = e_i(\alpha_1, \alpha_2, \ldots, \alpha_n)$ are the *elementary symmetric polynomials* in $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Definition 5.2: Elementary Symmetric Polynomials
The elementary symmetric polynomials $e_i(\boldsymbol{x}) = e_i(x_1, x_2, \ldots, x_n)$, are defined by

$$(t - x_1)(t - x_2) \ldots (t - x_n) = t^n - e_1(\boldsymbol{x})t^{n-1} + e_2(\boldsymbol{x})t^{n-2} - \ldots + (-1)^n e_n(\boldsymbol{x}). \tag{5.1.4}$$

In the case of two variables and three variables the elementary symmetric polynomials are given by

$$\boldsymbol{n=2:} \begin{cases} e_1(\boldsymbol{x}) = x_1 + x_2 \\ e_2(\boldsymbol{x}) = x_1x_2 \end{cases} \qquad \boldsymbol{n=3:} \begin{cases} e_1(\boldsymbol{x}) = x_1 + x_2 + x_3 \\ e_2(\boldsymbol{x}) = x_1x_2 + x_1x_3 + x_2x_3 \\ e_3(\boldsymbol{x}) = x_1x_2x_3 \end{cases}$$

in general, e_k has degree k, and we can write it as

$$e_k(\boldsymbol{x}) = \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} x_{i_1} x_{i_2} \ldots x_{i_k} \tag{5.1.5}$$

where all polynomials (5.1.5) obeys (5.1.1).

Another property that a polynomial can possess is that of being homogeneous, which means that all terms in the polynomial is of the same degree. For example, the polynomial $3x^4$ is homogeneous of degree 4 in the variable x, and the polynomial $3x^2y^2 + x^4 - 3xy^3$ is homogeneous of degree 4 in the variables x and y. The general property can be written as

Definition 5.3: Homogeneous Polynomial
The polynomial $f(x_1, x_2, \ldots, x_n)$ is homogeneous of degree k if

$$f(tx_1, tx_2, \ldots, tx_n) = t^k(x_1, x_2, \ldots, x_n).$$

5.2 Polynomial Functions

The characteristic polynomial to A

$$\begin{aligned} p_A(t) = \det(tE_n - A) = \\ t^n - p_1(A)t^{n-1} + p_2(A)t^{n-2} + \ldots + (-1)^{n-1}p_{n-1}(A)t + (-1)^n p_n(A) \end{aligned}$$

consists of the variable t and the polynomials $p_i(A)$. The polynomials $p_i(A)$ are homogeneous polynomials of degree i in the elements of A.

If B is invertible, we can write the characteristic polynomial to BAB^{-1} as

$$\begin{aligned} p_{BAB^{-1}}(t) &= \det(tE_n - BAB^{-1}) = \det(BB^{-1}tE_n - BAB^{-1}) = \\ &= \det(BtE_nB^{-1} - BAB^{-1}) = \det(B(tE_n - A)B^{-1}). \end{aligned}$$

We then have that the properties of the determinant makes

$$\det(B(tE_n - A)B^{-1}) = \det(B)\det(tE_n - A)\det(B^{-1}) =$$
$$= \det(B)\det(B^{-1})\det(tE_n - A) = \det(tE_n - A)$$

and we get that

$$p_A(t) = p_{BAB^{-1}}(t).$$

Expressing this for the polynomials $p_i(A)$, we have

$$p_i(A) = p_i(BAB^{-1}). \tag{5.2.1}$$

Now we have concluded that the polynomials $p_i(A)$, the coefficients in the characteristic polynomial, are invariant under conjugation by matrices in $\mathbb{GL}(n, \mathbb{C})$.

Before we extend our investigation to all polynomials $p(A)$ that are invariant under conjugation by matrices in $\mathbb{GL}(n, \mathbb{C})$, we are going to need some fundamental facts about polynomials in general. A polynomial in the variables $x_1, x_2, \ldots, x_n$ is an expression of the form

$$f(x_1, x_2, \ldots, x_n) = \sum_{k_1, k_2, \ldots, k_n} c_{k_1, k_2, \ldots, k_n} x_1^{k_1} x_2^{k_2} x_n^{k_n}$$

where the sum is finite, it only has an finite amount of non-zero coefficients $c_{k_1, k_2, \ldots, k_n}$. The summation is taken over $k_1 = 0, 1, \ldots, n_1; k_2 = 0, 1, \ldots, n_2$ and so forth to $k_n = 0, 1, \ldots, n_n$.

A polynomial function, $f(a_1, a_2, \ldots, a_n) : \mathbb{C}^n \to \mathbb{C}$, originated in a polynomial $f(x_1, x_2, \ldots, x_n)$, takes scalars as input in place of the variables of the polynomial, with the same structure. With the expression of a polynomial above, we can express a polynomial function in the same notation, with a sum.

The polynomial

$$f(x_1, x_2, \ldots, x_n) = \sum_{k_1, k_2, \ldots, k_n} c_{k_1, k_2, \ldots, k_n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}$$

will generate a polynomial function $\mathbb{C}^n \to \mathbb{C}$ by

$$f(a_1, a_2, \ldots, a_n) = \sum_{k_1, k_2, \ldots, k_n} c_{k_1, k_2, \ldots, k_n} a_1^{k_1} a_2^{k_2} \ldots a_n^{k_n}$$
$$\begin{cases} f : \mathbb{C}^n \to \mathbb{C} \\ f : (a_1, a_2, \ldots, a_n) \mapsto \sum_{k_1, k_2, \ldots, k_n} c_{k_1, k_2, \ldots, k_n} a_1^{k_1} a_2^{k_2} \ldots a_n^{k_n} \end{cases}$$

where x_i are variables in the polynomial and a_i are scalars. If all the coefficients are zero, we call the polynomial the *zero polynomial.*

The following properties of polynomials are going to be necessary later.

Property 5.1: If f is a polynomial in n variables such that $f(a_1, a_2, \ldots, a_n) = 0$ for all $(a_1, a_2, \ldots, a_n) \in \mathbb{C}^n$, then f is the zero polynomial.

Property 5.2: If the product of two polynomials are the zero polynomial, then at least one of the polynomials is the zero polynomial.

I state them here to have them in context with the definitions of polynomials and polynomial functions, but we are not going to use them until we get to the proof for all matrices $\mathbb{M}_n$ in section 6.3.

5.3 The Fundamental Theorem of Symmetric Polynomials

A very important theorem for symmetric polynomials is *The Fundamental Theorem of Symmetric Polynomials*, which states that any symmetric polynomial can be written in terms of the elementary symmetric polynomials.

Theorem 5.1 (**The Fundamental Theorem of Symmetric Polynomials**). [3]*: n:o 56*
All symmetric polynomials

$$f(x_1, x_2, \ldots, x_n) = \sum c_{k_1,k_2,\ldots,k_n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}$$

can be unambiguously be expressed as a polynomial of the elementary symmetric polynomials $e_1, e_2, \ldots, e_n$ *in the variables* $\boldsymbol{x} = x_1, x_2, \ldots, x_n$

$$F(e_1(\boldsymbol{x}), e_2(\boldsymbol{x}), \ldots, e_n(\boldsymbol{x})) = \sum \gamma_{k_1,k_2,\ldots,k_n} e_1^{v_1}(\boldsymbol{x}) e_2^{v_2}(\boldsymbol{x}) \ldots e_n^{v_n}(\boldsymbol{x})$$

where the coefficients γ *are homogeneous linear integer polynomials of the coefficients* c*.*

To prove the theorem we will do a proof by induction, originally given by Cauchy, and we will begin with the base case of $n = 2$. But first we are going to need a lemma

Lemma 5.1. [3]*: n:o 54*
If a symmetric polynomial in the variables $x_1, x_2, \ldots, x_n$ *can be expressed as a polynomial in the elementary symmetric polynomials* $e_1, e_2, \ldots, e_n$ *of the same variables, that is, if*

$$p(x_1, x_2, \ldots, x_n) = p(e_1(x_1, x_2, \ldots, x_n), e_2(x_1, x_2, \ldots, x_n), \ldots, e_n(x_1, x_2, \ldots, x_n))$$

the expression is unambiguous.

An alternate formulation of the lemma is: *the elementary symmetric polynomials of n variables are independent of each other.*

Proof: Theorem 5.1 *We follow the proof in* [3] *n:o 56*

$n = 2$:

For two variables, we then have

$$e_1 = x_1 + x_2$$
$$e_2 = x_1x_2$$

If f is an arbitrary symmetric polynomial of x_1 and x_2, we will have

$$f(x_1, x_2) = f(x_1, e_1 - x_1) = A_0x_1^m + a_1x_1^{m-1} + \ldots + A_m$$

the coefficients A_i are polynomials in a_1 with homogeneous linear integer polynomials of the coefficients in f. If we now set

$$\Phi(z) = A_0z^m + A_1z^{m-1} + \ldots + A_m$$

and divide Φ by $f(z) = (z - x_1)(z - x_2) = z^2 + e_1z + e_2$, we will get

$$\Phi(z) = Q(z)f(z) + A + Bz$$

where A and B are polynomials with coefficients that are homogeneous linear integer polynomials of the coefficients in f.

For $z = x_1$ we then have

$$f(x_1, x_2) = A + Bx_1$$

and since f is symmetric we will also get

$$f(x_1, x_2) = A + Bx_2.$$

Since x_1 and x_2 are, of each other, independent variables, it follows that $B = 0$ and thus

$$f(x_1, x_2) = A$$

and the theorem is proven for $n = 2$.

Induction step:

We now assume that the theorem holds for $n - 1$ variables, and want to prove that it then holds for n variables.

The symmetric polynomial f then is a polynomial of n variables, $f(x_1, x_2, \ldots, x_n)$. We arrange f in order of the exponents as

$$f(x_1, x_2, \ldots, x_n) = f_0x_1^{\mu} + f_1x_1^{\mu-1} + \ldots + f_{\mu-1}x_1 + f_{\mu} \qquad (5.3.1)$$

the coefficients $f_0, f_1, \ldots, f_{\mu}$ are symmetric polynomials of the $n - 1$ variables $x_2, x_3, \ldots, x_n$, with coefficients that are homogeneous linear integer polynomials of the coefficients in f. We denote the elementary symmetric polynomials of these $n-1$ variables by $e'_1, e'_2, \ldots, e'_{n-1}$.

If now

$$f(z) = \prod_{i=1}^{n}(z - x_i) = z^n + e_1 z^{n-1} + \ldots + e_{n-1}z + e_n$$

we will have

$$\frac{f(z)}{z - x_1} = z^{n-1} + e'_1 z^{n-2} + \ldots + e'_{n-2}z + e'_{n-1}$$

thus we get

$$\begin{aligned}
e'_1 &= x_1 + e_1 \\
e'_2 &= x_1^2 + e_1 x_1 + e_2 \\
e'_3 &= x_1^3 + e_1 x_1^2 + e_2 x_1 + e_3 \\
&\vdots \\
e'_{n-1} &= x_1^{n-1} + e_1 x_1^{n-2} + \ldots + e_{n-1}.
\end{aligned}$$

Because the theorem by assumption holds for $n-1$ variables, we can then express the coefficients $f_0, f_1, \ldots, f_\mu$ as polynomials in the elementary symmetric polynomials $e_1, e_2, \ldots, e_n$ and x_1, with coefficients that are homogeneous linear integer polynomials of the coefficients in f.

By inserting the expressions for $e'_1, e'_2, \ldots, e'_{n-1}$ above into (5.3.1), and then again arrange f in order of the exponents of x_1, we get

$$f(x_1) = A_0 x_1^m + A_1 x_1^{m-1} + \ldots + A_{m-1}x_1 + A_m.$$

The coefficients $A_0, A_1, \ldots, A_m$ in $f(x_1)$ are polynomials in $e_1, e_2, \ldots, e_n$, and the coefficients in e_i are homogeneous linear integer polynomials of the coefficients in f.

If we now, similarly to the $n = 2$-case, set

$$\Phi(z) = A_0 z^m + A_1 z^{m-1} + \ldots + A_{m-1}z + A_m$$

and divide by $f(z)$ to get

$$\Phi(z) = Q(z)f(z) + \Psi(z)$$

where

$$\Psi(z) = C_0 z^{n-1} + C_1 z^{n-2} + \ldots + C_{n-2}z + C_{n-1}$$

the coefficients $C_0, C_1, \ldots, C_{n-2}, C_{n-1}$ are polynomials in $e_1, e_2, \ldots, e_n$, with coefficients that are homogeneous linear integer polynomials of the coefficients in f.

If we now set $z = x_1$ in $\Phi(z) = Q(z)f(z) + \Psi(z)$ we will, since $f(x_1) = 0$, have

$$f = \Phi(x_1) = \Psi(x_1)$$

and since f is symmetric in $x_1, x_2, \ldots, x_n$, this equality will hold if x_1 is replaced with any other of the variables x_i. Then it follows that the polynomial

$$C_0 z^{n-1} + C_1 z^{n-2} + \ldots + C_{n-2} z + C_{n-1} - f$$

will vanish for n different values: $z = x_1, z = x_2, \ldots, z = x_n$. It follows then, from the fundamental theorem of algebra, that the polynomial will vanish identically since its degree is at most $n - 1$. In other words we must have $f = C_{n-1}$, i.e. f is expressed as a polynomial in the elementary symmetric polynomials $e_1, e_2, \ldots, e_n$, whose coefficients are homogeneous linear integer polynomials of the coefficients in f.

By Lemma 5.1 the expression is also unambiguous, thus by the induction hypothesis, the theorem is proven. $\square$

6 Invariant polynomial functions

Now we are finally ready to investigate the core concept of this thesis, which are all the polynomial functions p that are invariants to $\mathbb{GL}(n, \mathbb{C})$.

Let

$$p\colon \mathbb{M}_n \to \mathbb{C} \tag{6.0.1}$$

be a polynomial function that obeys (5.2.1) for all square matrices A and invertible matrices B. Then we have

$$p(BAB^{-1}) = p(A); \;\; \forall \;\; A \in \mathbb{M}_n, \;\; \forall \;\; B \in \mathbb{GL}(n, \mathbb{C}). \tag{6.0.2}$$

We have already seen some examples of (6.0.1), namely the trace of A, $\mathrm{tr}(A) = p_1(A)$ and the determinant of A, $\det(A) = p_n(A)$. In order to determine all polynomial functions (6.0.1) on the set of square matrices $\mathbb{M}_n$ that fulfils (6.0.2), we are going to start with the restriction of all square matrices to the diagonal matrices $\mathbb{D}_n$. We will then proceed to diagonalizable matrices $\mathbb{H}_n$, and finally $\mathbb{M}_n$ the set of all $n \times n$–matrices over $\mathbb{C}$. The reason for this approach is rather straight forward, it is hard to start with all $n \times n$–matrices over $\mathbb{C}$, and since

$$\mathbb{D}_n \subset \mathbb{H}_n \subset \mathbb{M}_n$$

we can investigate the smallest subset of $\mathbb{M}_n$, the diagonal matrices $\mathbb{D}_n$, first, and then expand the argument to bigger subsets to finally arrive at the entire set $\mathbb{M}_n$.

6.1 Diagonal Matrices

We can solve the problem for diagonal matrices $D \in \mathbb{D}_n$, with Theorem 5.1. We want to determine all polynomial functions (6.0.1), for the case of the subset consisting of the diagonal matrices in the set of all $n \times n$–matrices over $\mathbb{C}$. So the polynomial functions when we restrict us to $\mathbb{D}_n$ will be

$$p\colon \mathbb{D}_n \to \mathbb{C}$$

and (6.0.2) in this case is

$$p(BDB^{-1}) = p(D); \;\; \forall \;\; D \in \mathbb{D}_n, \;\; \forall \;\; B \in \mathbb{GL}(n, \mathbb{C}).$$

It must then especially hold for all *permutation matrices* A_σ, where the permutation σ is an element in the symmetric group of degree n, S_n

$$p(A_\sigma D A_\sigma^{-1}) = p(D) \; ; \; \forall \; A_\sigma, \; \sigma \in S_n. \tag{6.1.1}$$

Now, since a polynomial function, $p\colon \mathbb{D}_n \to \mathbb{C}$, originated in a polynomial in the diagonal elements of D

$$p(D) = p(\mathrm{diag}(d_1, d_2, \ldots, d_n))$$

and since we have from (3.1.9) that

$$A_\sigma(\mathrm{diag}(d_1, d_2, \ldots, d_n))A_\sigma^{-1} = \mathrm{diag}(d_{\sigma^{-1}(1)}, d_{\sigma^{-1}(2)}, \ldots, d_{\sigma^{-1}(n)})$$

in accordance with (6.1.1), it then has to be the case that

$$p(\mathrm{diag}(d_{\sigma^{-1}(1)}, d_{\sigma^{-1}(2)}, \ldots, d_{\sigma^{-1}(n)})) = p(\mathrm{diag}(d_1, d_2, \ldots, d_n))$$

for all permutations $\sigma \in S_n$, thus $p(\mathrm{diag}(d_1, d_2, \ldots, d_n))$ is a symmetric polynomial in $\mathrm{diag}(d_1, d_2, \ldots, d_n)$. According to Theorem (5.1) there exists a polynomial $P(e_1(d_1, d_2, \ldots, d_n), e_2(d_1, d_2, \ldots, d_n), \ldots, e_n(d_1, d_2, \ldots, d_n))$ such that

$$\begin{aligned} &p(\mathrm{diag}(d_1, d_2, \ldots, d_n) = \\ &\quad = P(e_1(d_1, d_2, \ldots, d_n), e_2(d_1, d_2, \ldots, d_n), \ldots, e_n(d_1, d_2, \ldots, d_n)) \end{aligned}$$

and since Theorem 5.1 state that we can write *any symmetric polynomial* in terms of the *elementary symmetric polynomials*

$$p_i(\mathrm{diag}(d_1, d_2, \ldots, d_n) = e_i(d_1, d_2, \ldots, d_n)$$

we have showed that

$$p(D) = p(\mathrm{diag}(d_1, d_2, \ldots, d_n) = P(p_1(D), p_2(D), \ldots, p_n(D)).\ \square \tag{6.1.2}$$

6.2 Diagonalizable Matrices

We want to determine all polynomial functions (6.0.1), for the case of the subset consisting of the diagonalizable matrices in the set of all $n \times n$–matrices over $\mathbb{C}$.

Let $\mathbb{H}_n$ be the set of diagonalizable matrices in $\mathbb{M}_n$. The polynomial function's condition (6.0.2), will then, similarly to the previous case of diagonal matrices, be restricted to the subset $\mathbb{H}_n$ of $\mathbb{M}_n$.

Let

$$p : \mathbb{H}_n \to \mathbb{C}$$

be the polynomial functions when we restrict us to the subset $\mathbb{H}$ of $\mathbb{M}_n$, then (6.0.2) gives

$$p(BHB^{-1}) = p(H);\ \ \forall\ \ H \in \mathbb{H}_n,\ \ \forall\ \ B \in \mathbb{GL}(n, \mathbb{C}). \tag{6.2.1}$$

We know that all matrices $H \in \mathbb{H}_n$ are diagonalizable, and we know that if a matrix H is diagonalizable, there exists another matrix T such that

$$H = TDT^{-1}$$

where D is a diagonal matrix with the eigenvalues of H as entries, and T has the eigenvectors of H as columns.

Now, the polynomial functions property (6.2.1), for this new representation of H, give us

$$p(B(TDT^{-1})B^{-1}) = p(TDT^{-1}) = p(D)$$

which in turn gives

$$p(H) = p(D)$$

and according to (6.1.2), we then have

$$p(H) = P(p_1(D), p_2(D), \ldots, p_n(D))$$

which gives

$$p(H) = P(p_1(TDT^{-1}), p_2(TDT^{-1}), \ldots, p_n(TDT^{-1}))$$

and then, since $p_i(TDT^{-1}) = p_i(H)$, we have

$$\begin{aligned} p(H) = P(p_1(TDT^{-1}), p_2(TDT^{-1}), \ldots, p_n(TDT^{-1})) = \\ = P(p_1(H), p_2(H), \ldots, p_n(H)) \end{aligned} \tag{6.2.2}$$

and the proof for diagonalizable matrices is complete. □

6.3 All Matrices

Before we proceed with the proof for all square matrices, we are going to go through some more theory about diagonalizability and eigenvalues, the discriminant of a polynomial and their connection to the characteristic polynomial of a matrix.

A polynomial function

$$p \colon \mathbb{M}_n \to \mathbb{C}$$

is determined by its restriction to the subset $\mathbb{H}_n$ of $\mathbb{M}_n$, and in fact restricted to matrices with n distinct eigenvalues. To show this, we will begin with a proof that if all eigenvalues of a matrix A are distinct, the eigenvectors of A are linearly independent and thus A is diagonalizable.

We will formalize by stating this as a theorem

Theorem 6.1. [4]*: Sec. 5.2*
Let A be the matrix representation of a linear operator T on a vector space $\mathbb{V}$, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of A. If $v_1, v_2, \ldots, v_k$ are eigenvectors of A such that λ_i corresponds to v_i, then the set of eigenvectors $\{v_1, v_2, \ldots, v_k\}$ are linearly independent.

Proof. We follow the proof in [4] *Sec. 5.2*
The proof is by induction over k

Base case:
Suppose $k = 1$, then $v_k = v_1 \neq \vec{0}$ since an eigenvector by definition cannot be the zero-vector, and then v_1 is linearly independent.

Induction hypothesis:
We now assume that the theorem holds true for $k-1$ distinct eigenvalues, where $k - 1 \geq 1$, and that we have k eigenvectors $v_1, v_2, \ldots, v_k$ corresponding to the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. We want to show that the set of eigenvectors $\{v_1, v_2, \ldots, v_k\}$ is linearly independent.

Induction step:
Suppose that $a_1, a_2, \ldots a_k$ are scalars such that

$$a_1 v_1 + a_2 v_2 + \ldots + a_k v_k = 0$$

applying $A - \lambda_k E$ on both sides of the equation gives

$$a_1(\lambda_1 - \lambda_k)v_1 + a_2(\lambda_2 - \lambda_k)v_2 + \ldots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = 0$$

then, by the induction hypothesis, the set of eigenvectors $\{v_1, v_2, \ldots, v_k\}$ is linearly independent, and hence

$$a_1(\lambda_1 - \lambda_k) = a_2(\lambda_2 - \lambda_k) = \ldots = a_{k-1}(\lambda_{k-1} - \lambda_k) = 0\ .$$

Since the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct, we know that $\lambda_i - \lambda_k \neq 0$ for $1 \leq i \leq k - 1$. So then it has to be the case that

$$a_1 = a_2 = \ldots = a_{k-1} = 0$$

and then equation $a_1 v_1 + a_2 v_2 + \ldots + a_k v_k = 0$ will reduce as

$$a_1 v_1 + a_2 v_2 + \ldots + a_k v_k = 0 \;\Rightarrow\; a_k v_k = 0$$

and since $v_k \neq \vec{0}$ as previously stated, it must be the case that $a_k = 0$. So then we have $a_1 = a_2 = \ldots = a_{k-1} = a_k = 0$, and from that it follows that the set of eigenvectors $\{v_1, v_2, \ldots, v_k\}$ is linearly independent. □

Then we can formulate a corollary to the theorem that will conclude our desired property of A, that is A being diagonalizable.

Corollary 6.1. [4]*: Sec. 5.2*
Let A be the matrix representation of a linear operator on an n-dimensional vector space $\mathbb{V}$. If A has n distinct eigenvalues, A is diagonalizable.

Proof. We follow the proof in [4] *Sec. 5.2*
Suppose A has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. For each i we choose an eigenvector v_i corresponding to the eigenvalue λ_i, then by the theorem the set of eigenvectors $\{v_1, v_2, \ldots, v_k\}$ is linearly independent, and since $\dim(\mathbb{V}) = n$ this set is a basis for the vector space $\mathbb{V}$. And since a linear operators matrix representation on a finite-dimensional vector space is diagonalizable if and only if there exists an ordered basis for the vector space consisting of eigenvectors of A, A is diagonalizable. □

6.3.1 The Discriminant

A way to find an algebraic condition for diagonalizability of a matrix A is the *discriminant* to said matrix's characteristic polynomial $p_A(t)$.

Definition 6.1: The Discriminant

The *discriminant* to a polynomial $x_1, x_2, \ldots, x_n$ is defined as the polynomial

$$\Delta(x_1, x_2, \ldots, x_n) = \prod_{1 \le i < j \le n} (x_i - x_j)^2. \tag{6.3.1}$$

We will now see that the discriminant is a symmetric polynomial, i.e. that

$$\prod_{1 \le i < j \le n} (x_i - x_j)^2 = \prod_{1 \le i < j \le n} (x_{\sigma(i)} - x_{\sigma(j)})^2 \ .$$

Proof:

We start with the basic cases where we have two variables x_1 and x_2, and three variables, x_1, x_2 and x_3 in the polynomial, since that is a good foundation for the proof.

$\boldsymbol{n = 2}$**:**

The discriminant to a polynomial in the two variables is

$$\Delta(x_1, x_2) = \prod_{1 \le i < j \le 2} (x_i - x_j)^2 = (x_1 - x_2)^2$$

and if we look at the transposition permutation of $\Delta(x_1, x_2)$ we have

$$\tau_{12}\left(\Delta(x_1, x_2)\right) = (x_2 - x_1)^2 = \left(-(x_1 - x_2)\right)^2 = \Delta(x_1, x_2)$$

then $\Delta(x_1, x_2)$ is a symmetric polynomial, and we can then express it in the elementary symmetric polynomials

$$\begin{aligned} \Delta(x_1, x_2) &= (x_2 - x_1)^2 = x_1^2 - 2x_1x_2 + x_2^2 \\ &= x_1^2 + 2x_1x_2 + x_2^2 - 4x_1x_2 \\ &= (x_1 + x_2)^2 - 4x_1x_2 \\ &= e_1(x_1, x_2)^2 - e_2(x_1, x_2). \end{aligned}$$

$\boldsymbol{n = 3}$**:**

The discriminant to a polynomial in the three variables is

$$\Delta(x_1, x_2, x_3) = \prod_{1 \le i < j \le 3} (x_i - x_j)^2 = (x_1 - x_2)^2 (x_1 - x_3)^2 (x_2 - x_3)^2$$

and if we look at the cyclic permutation $\beta = (1\ 2\ 3)$ of $\Delta(x_1, x_2, x_3)$, we have

$$\begin{aligned}\beta\left(\Delta(x_1, x_2, x_3)\right) &= (x_2 - x_3)^2(x_2 - x_1)^2(x_3 - x_1)^2 \\ &= \left(-(x_3 - x_2)\right)^2\left(-(x_1 - x_2)\right)^2\left(-(x_1 - x_3)\right)^2 \\ &= \Delta(x_1, x_2, x_3)\end{aligned}$$

then $\Delta(x_1, x_2, x_3)$ is also a symmetric polynomial. We can then again express it in the elementary symmetric polynomials, we will omit the explicit calculation here, but we would get

$$\Delta(x_1, x_2, x_3) = e_1^2 e_2^2 - 4e_2^3 - 4e_1^3 e_3 - 27e_3^2 + 18e_1 e_2 e_3\ .$$

$\boldsymbol{n = n}$**:**

We now let $\sigma \in S_n$ be an arbitrary permutation and let the polynomial have any number n of variables. The discriminant to the polynomial in n variables is

$$\Delta(x_1, x_2, \ldots, x_n) = \prod_{1 \le i < j \le n} (x_i - x_j)^2\ .$$

The symmetry of $\Delta(x_1, x_2, \ldots, x_n)$ is proven similarly to the cases of $n = 2$ and $n = 3$. For any permutation $\sigma \in S_n$, we will have

$$\sigma\left(\Delta(x_1, x_2, \ldots, x_n)\right) = \prod_{1 \le i < j \le n} (x_{\sigma(i)} - x_{\sigma(j)})^2$$

and because $i \neq j$ always, then $\sigma(i) \neq \sigma(j)$ always also. Thus, we have

$$\prod_{1 \le i < j \le n} (x_{\sigma(i)} - x_{\sigma(j)})^2 = \Delta(x_1, x_2, \ldots, x_n)$$

and the proof that the discriminant is a symmetric polynomial is complete. $\square$

From the fact that $\Delta(x_1, x_2, \ldots, x_n) = \Delta(\boldsymbol{x})$ is a symmetric polynomial, it follows that we can then express $\Delta(x_1, x_2, \ldots, x_n) = \Delta(\boldsymbol{x})$ in the elementary symmetric polynomials. But we are however not going to need the explicit expressions for $\Delta(\boldsymbol{x})$ in the elementary symmetric polynomials, merely the fact that we *can express* them in the elementary symmetric polynomials, in accordance with Theorem 5.1

$$\Delta(\boldsymbol{x}) = P(e_1(\boldsymbol{x})), e_2(\boldsymbol{x}), \ldots, e_n(\boldsymbol{x}))$$

and since the polynomial $\Delta(\boldsymbol{x})$ has integer coefficients, then by Theorem 5.1 the polynomial $P(e_1(\boldsymbol{x})), e_2(\boldsymbol{x}), \ldots, e_n(\boldsymbol{x}))$ will also.

The discriminant and eigenvalues

Let

$$f(x) = x^n - a_1 x^{n-1} + a_2 x^{n-2} - \ldots + (-1)^{n-1} a_{n-1} x + (-1)^n a_n$$

be a polynomial with roots $\alpha_1, \alpha_2, \ldots, \alpha_n$, not necessarily distinct, the factorization representation of $f(x)$ is

$$f(x) = (x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_n) .$$

According to the relationship between roots and coefficients (section 6) we have

$$a_i = e_i(\boldsymbol{\alpha})$$

and the discriminant to $f(x)$ is then defined as

$$\Delta(f) = \Delta(\boldsymbol{\alpha}) = \prod_{1 \le i < j \le n} (\alpha_i - \alpha_j)^2$$

then it must be the case that

$$\Delta(f) = P(e_1(\boldsymbol{\alpha}), e_2(\boldsymbol{\alpha}), \ldots, e_n(\boldsymbol{\alpha})) = P(a_1, a_2, \ldots, a_n)$$

thus the discriminant of $f(x)$ can be expressed as a polynomial with integer coefficients, where the coefficients are the coefficients of the polynomial $f(x)$, as we saw in the example above. So in principle, it is possible to calculate the discriminant of $f(x)$ without knowing the roots of the polynomial $f(x)$.

We are now going to investigate the important situation where $\Delta(f) = 0$, we will see that this will be the case if and only if $\Delta(f)$ has *at least one root of multiplicity two or greater*. This follows from the fact that the discriminant is a product, and for it to be zero we must have at least one of the factors equal to zero. This happens exactly when $\alpha_i = \alpha_j$ for at least one factor $(\alpha_i - \alpha_j)^2$, and it will be non-zero whenever $\alpha_i \neq \alpha_j$ for all i and j

$$\begin{aligned} &(\alpha_i - \alpha_j)^2 = 0, \text{ if and only if } \alpha_i = \alpha_j \\ &(\alpha_i - \alpha_j)^2 \neq 0, \text{ if and only if } \alpha_i \neq \alpha_j, \ \forall \ i, j . \end{aligned}$$

Let $A \in \mathbb{M}_n$, and we again have the characteristic polynomial to A

$$p_A(t) = \det(tE_n - A),$$

the roots of $p_A(t)$ are the eigenvalues of A, $\lambda_1, \lambda_2, \ldots, \lambda_n$. We will then have the discriminant to the characteristic polynomial, $\Delta(p_A(t)) \neq 0$, if and only if *A has n distinct eigenvalues*, which is the same as n eigenvalues of multiplicity 1

$$\Delta(p_A(t)) \neq 0, \text{ if and only if } \lambda_i \neq \lambda_j, \ \forall \ i, j . \tag{6.3.2}$$

6.3.2 Products of Polynomials

Now we are going to make use of the properties 5.1 and 5.2 of polynomials in section 5.

If $f(x)$ is a polynomial, that is not the zero polynomial, in one variable over $\mathbb{C}$ (or any other infinite field like $\mathbb{R}$ or $\mathbb{Q}$), then the set of roots to f is finite or "small", we call this subset ω

$$\omega = \{a \in \mathbb{C} \colon f(a) = 0\}.$$

The subset of $\mathbb{C}$ where f is non-zero, is then infinite or "large", we call this subset Ω

$$\Omega = \{a \in \mathbb{C} \colon f(a) \neq 0\}$$

Now we assume that $g(x)$ is another polynomial such that $g(a) = 0$ for all $a \in \Omega$. Since a polynomial only can have a finite number of roots, and since Ω is infinite, g has to be the zero polynomial. This also holds for a polynomial of many variables, $g(\boldsymbol{a}) = g(a_1, a_2, \ldots, a_n)$.

Theorem 6.2. *Let*

$$f(x_1, x_2, \ldots, x_n)$$

be a polynomial over $\mathbb{C}$, and we now set the subset of $\mathbb{C}$ where f is non-zero to

$$\Omega = \{(a_1, a_2, \ldots, a_n) \in \mathbb{C}^n \colon f(a_1, a_2, \ldots, a_n) \neq 0\}$$

now lets assume that $g(x_1, x_2, \ldots, x_n)$ is a polynomial such that

$$g(a_1, a_2, \ldots, a_n) = 0 \ \ , \forall \ \ (a_1, a_2, \ldots, a_n) \in \Omega, \ \ \textit{then } g \textit{ is the zero polynomial.} \tag{6.3.3}$$

Proof. Let

$$h(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n) g(x_1, x_2, \ldots, x_n)$$

then from Properties 5.1 and 5.2 of polynomials, we have

$$\begin{aligned} \text{if } \ (a_1, a_2, \ldots, a_n) \notin \Omega, \ \text{ then } \ & f(a_1, a_2, \ldots, a_n) = 0 \\ \text{and thus } \ & h(a_1, a_2, \ldots, a_n) = 0 \end{aligned}$$

$$\begin{aligned} \text{if } \ (a_1, a_2, \ldots, a_n) \in \Omega, \ \text{ then } \ & g(a_1, a_2, \ldots, a_n) = 0 \\ \text{and thus } \ & h(a_1, a_2, \ldots, a_n) = 0 \,. \end{aligned}$$

We can conclude that $h(a_1, a_2, \ldots, a_n) = 0$ for all $(a_1, a_2, \ldots, a_n)$ in $\mathbb{C}^n$, and then h has to the the zero polynomial. And since $h = fg$ is a product of two polynomials, it cannot be the zero polynomial unless at least one of the polynomials in the product f or g is the zero polynomial, then because f is not the zero polynomial, g is the zero polynomial. □

6.4 All invariants

From previous section 6.2 on diagonalizable matrices, we now know that the polynomial function

$$p\colon \mathbb{M}_n \to \mathbb{C} \tag{6.4.1}$$

has the properties

$$p(A) = p(BAB^{-1})$$
$$p(A) = P(p_1(A), p_2(A), \ldots, p_n(A))$$

for all matrices A in the subset $\mathbb{H}_n$ of $\mathbb{M}_n$, and all invertible matrices B. We are now going to investigate the polynomial function for all square matrices $\mathbb{M}_n$. With the known properties, Corollary 6.1 then gives us that all matrices with n distinct eigenvalues, $\mathbb{N}_n$, is a subset of $\mathbb{H}_n$, we can use the discriminant to extend the set $\mathbb{H}_n$ to the whole of $\mathbb{M}_n$.

Let

$$G(A) = p(A) - P(p_1(A), p_2(A), \ldots, p_n(A))$$

then G is a polynomial function

$$G\colon \mathbb{M}_n \to \mathbb{C} \tag{6.4.2}$$

that we know is the zero-polynomial for the subset of diagonalizable matrices in the set of all matrices, $\mathbb{H}_n \subset \mathbb{M}_n$. We are now going to show that $G(A)$ is the zero-polynomial for *all* matrices $A \in \mathbb{M}_n$, that is, that

$$p(A) = P(p_1(A), p_2(A), \ldots, p_n(A)), \ \ \forall\, A \in \mathbb{M}_n.$$

To do this we use, as mentioned above, the discriminant to extend $\mathbb{N}_n$ to the whole of $\mathbb{M}_n$ by forming the product of $G(A)$ and $\Delta(p_A)$, and show that the product is in fact zero for all $A \in \mathbb{M}_n$.

So I now claim that

$$G(A)\Delta(p_A) = 0, \ \ \forall\, A \in \mathbb{M}_n \tag{6.4.3}$$

where $\Delta(p_A)$ is the discriminant to the characteristic polynomial $p_A(t)$ of A.

Proof of the claim:

To prove (6.4.3) we look at two cases, when A *has* n distinct eigenvalues and then is diagonalizable, and when A does *not* have n distinct eigenvalues and is not diagonalizable.

According to (6.3.2) the discriminant to the characteristic polynomial of A, is non-zero if and only if A has n distinct eigenvalues

$$\Delta(p_A) \neq 0 \ \text{ if and only if } \ \lambda_i \neq \lambda_j, \ \forall \ i, j\ .$$

If A has n distinct eigenvalues, A is diagonalizable and we have from (6.2.2) that

$$p(A) = P(p_1(A), p_2(A), \ldots, p_n(A))$$

which in turn means that

$$G(A) = p(A) - p(A) = 0 \ .$$

If, on the other hand, A does *not* have n distinct eigenvalues (6.3.2) gives

$$\Delta(p_A) = 0 \ .$$

To summarize, we have that if A is not diagonalizable, $\Delta(p_A) = 0$ and

$$G(A)\Delta(p_A) = 0$$

and if A is diagonalizable, $G(A) = 0$ and

$$G(A)\Delta(p_A) = 0 \ .$$

From the Properties 5.1 and 5.2 of polynomials, we have that if the product $G(A)\Delta(p_A) = 0$ in (6.4.3), one of the polynomials has to be the zero-polynomial. Since $\Delta(p_A)$ is not the zero polynomial, then $G(A)$ has to be the zero polynomial.

From this we get that the polynomial function (6.4.2) fulfils the property (6.4.1), and we have

$$p(A) = P(p_1(A), p_2(A), \ldots, p_n(A)), \quad \forall\, A \in \mathbb{M}_n \tag{6.4.4}$$

which concludes the proof for all matrices $A \in \mathbb{M}_n$. $\square$

Illustrative Example:

An example of (6.4.4), for a 3×3-matrix, is the sum of the squares of the eigenvalues to A

$$p(A) = \sum_{i=1}^{3} \lambda_i^2 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2.$$

Because

$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = (\lambda_1 + \lambda_2 + \lambda_3)^2 - 2(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3) = e_1(\lambda)^2 - 2e_2(\lambda)$$

the polynomial $p(A)$ is

$$p(A) = p_1(A)^2 - 2p_2(A) = P(p_1(A), p_2(A)). \tag{6.4.5}$$

Let

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 2 & 1 \end{pmatrix}$$

then the eigenvalues to A are

$$\lambda_1 = 1,\ \lambda_2 = 1+\sqrt{2},\ \ \lambda_3 = 1-\sqrt{2}.$$

In the elementary symmetric polynomials we can express the square of the sum of the eigenvalues, $1^2 + (1+\sqrt{2})^2 + (1-\sqrt{2})^2$, as

$$(1{+}1{+}\sqrt{2}{+}1{-}\sqrt{2})^2{-}2\left((1+\sqrt{2}) + (1-\sqrt{2}) + (1+\sqrt{2})(1-\sqrt{2})\right) = e_1(\lambda)^2{-}2e_2(\lambda).$$

We then get the polynomial $p(A)$ as

$$\begin{aligned} p(A) = (1+1+\sqrt{2}+1-\sqrt{2})^2 - 2\left((1+\sqrt{2}) + (1-\sqrt{2}) + (1+\sqrt{2})(1-\sqrt{2})\right) = \\ = P(1+1+\sqrt{2}+1-\sqrt{2},\ \ (1+\sqrt{2}) + (1-\sqrt{2}) + (1+\sqrt{2})(1-\sqrt{2})) \end{aligned}$$

which is (6.4.5) with

$$\begin{aligned} p_1(A) &= 1+1+\sqrt{2}+1-\sqrt{2} \\ p_2(A) &= (1+\sqrt{2}) + (1-\sqrt{2}) + (1+\sqrt{2})(1-\sqrt{2}). \end{aligned}$$

7 Conclusion and Discussion

Conclusion:

Let $\mathbb{M}_n$ be the set of all $n \times n$-matrices over $\mathbb{C}$, and let $\mathbb{G}$ be the set of all invertible matrices in $\mathbb{M}_n$. Then $\mathbb{G}$ is a group under matrix multiplication, called *the general linear group over* $\mathbb{C}$, $\mathbb{GL}(n,\mathbb{C})$. The group acts on $\mathbb{M}_n$ through

$$G.A = GAG^{-1}.$$

The linear transformation f_σ acts on an arbitrary vector $\vec{v} \in \mathbb{V}$ as

$$(f_\pi f_\tau)(\vec{v}) = f_{\pi\tau}(\vec{v}).$$

For our specific groups, the symmetric group of degree n, S_n and the general linear group of degree n over $\mathbb{C}$, $\mathbb{GL}(n,\mathbb{C})$.

Let g be the function from S_n to $\mathbb{GL}(n,\mathbb{C})$

$$\left.\begin{aligned} &g\colon S_n \to \mathbb{GL}(n,\mathbb{C}) \\ &g\colon \sigma \mapsto f_\sigma \end{aligned}\right\} \quad \sigma \in S_n, \quad f_\sigma \in \mathbb{GL}(n,\mathbb{C}).$$

The function g is a group homomorphism from S_n to $\mathbb{GL}(n,\mathbb{C})$.

We then have concluded that the group actions of $\mathbb{A}$ on $\mathbb{M}_n$, will result in

$$\begin{aligned} &A_\sigma.B = A_\sigma B A_\sigma^{-1} \text{ gives } \left(A_\sigma B A_\sigma^{-1}\right)_{ij} = b_{\sigma^{-1}(i)\sigma^{-1}(j)} \\ &A_\sigma.B = A_\sigma B \text{ gives } \left(A_\sigma B\right)_{ij} = b_{\sigma^{-1}(i)\ j} \\ &A_\sigma.B = B A_\sigma^{-1} \text{ gives } \left(B A_\sigma^{-1}\right)_{ij} = b_{i\ \sigma^{-1}(j)}\ . \end{aligned}$$

A polynomial function

$$p\colon \mathbb{M}_n \to \mathbb{C}$$

is a function that originates in a polynomial, that is a polynomial in the elements of the matrix. And we have seen that the special invariants to $\mathbb{GL}(n,\mathbb{C})$ that are polynomials in the elements of the matrix $A \in \mathbb{M}_n$, have the property

$$p(BAB^{-1}) = p(A); \ \ \forall\ \ A \in \mathbb{M}_n, \ \ \forall\ \ B \in \mathbb{GL}(n,\mathbb{C})$$

we have showed that all invariants to matrices can be written in terms of the special invariants.

For diagonal matrices $\mathbb{D}_n \subset \mathbb{M}_n$, we showed this in Theorem 5.1, and we got

$$p(D) = p(\text{diag}(d_1, d_2, \ldots, d_n)) = P(p_1(D), p_2(D), \ldots, p_n(D)).$$

We then extended the argument to diagonalizable matrices $\mathbb{H}_n \subset \mathbb{M}_n$, which gave us

$$\begin{aligned} p(H) = P(p_1(TDT^{-1}), p_2(TDT^{-1}), \ldots, p_n(TDT^{-1})) = \\ = P(p_1(H), p_2(H), \ldots, p_n(H)). \end{aligned}$$

For the case of all square matrices, we looked at the set of diagonalizable matrices with n distinct eigenvalues $\mathbb{N}_n \subset \mathbb{M}_n$, and with the discriminant to the characteristic polynomial to A, $p_A(t)$ we could expand $\mathbb{N}_n$ to the whole of $\mathbb{M}_n$. This gave us, with the addition of the polynomial function $G(A)$, gave us the result for all matrices $A \in \mathbb{M}_n$

$$p(A) = P(p_1(A), p_2(A), \ldots, p_n(A)), \ \ \forall\ A \in \mathbb{M}_n \ .$$

Discussion:

This is a remarkable result! That we can describe *all* invariants, the elements of the ring $\mathcal{P}(\mathbb{M}_n)^{\mathbb{G}}$, with the special invariants $p_i(BAB^{-1}) = p_i(A)$.

Though it is outside the scope of this thesis, an interesting continuation of this argument would be to investigate the orbits of the group action. This would give us a way to categorize the orbits or to give specific elements in them, and determine if two matrices in $\mathbb{M}_n$ are elements in the same orbit. The result we have arrived at gives us the first important step needed to try and categorize the orbits.

Another interesting question to look at would be if the invariants does *separate* the orbits, that is if there for two orbits O_1 and O_2 always exists an invariant function f such that

$$f(O_1) \neq f(O_2).$$

If we look at the most basic case of 2×2-matrices, the invariants are

$$p_1(A) = \mathrm{tr}(A)$$
$$p_2(A) = \det(A).$$

They do separate the orbits corresponding to matrices with two distinct eigenvalues. Let for example A and B be two 2×2-matrices with two distinct eigenvalues

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \ B = \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix}$$

then we have

$$p_1(A) = \mathrm{tr}(A) = 1 + 2 = 3$$
$$p_2(A) = \det(A) = 2$$

$$p_1(B) = \mathrm{tr}(B) = 3 + 4 = 7$$
$$p_2(B) = \det(B) = 12$$

and the invariants does separate the orbits of A and B

$$p_1(O_A) \neq p_1(O_B)$$
$$p_2(O_A) \neq p_2(O_B).$$

When we have 2×2-matrices with a double eigenvalue, one eigenvalue of multiplicity two, we can for example see that

$$E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \text{ and } A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

have the same eigenvalue, 1, of multiplicity two. So the invariants will have the same value on the orbits of E_2 and A

$$p_1(A) = \mathrm{tr}(A) = p_1(E_2) = \mathrm{tr}(E_2) = 2$$
$$p_2(A) = \det(A) = p_2(E_2) = \det(E_2) = 1 .$$

But since the identity matrix E_2 forms an orbit on its own, an orbit with only the element E_2

$$O_{E_2} = E_2 ,$$

the matrices E_2 and A are not conjugated, there exists no matrix G such that

$$GE_2G^{-1} = A .$$

This means that even though the invariants evaluates to the same value in both cases

$$f(O_{E_2}) = f(O_A),$$

the orbits are still distinct from one another

$$O_{E_2} \neq O_A$$

and the invariants does *not* separate the orbits in this case.

On the other hand there exists matrices in the orbit of A that lies arbitrarily close to E_2, since A is conjugate to the matrix

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \ \forall x \neq 0.$$

8 References

[1]: T.A Springer: *Invariant Theory: Lecture Notes in Mathematics 585.* Springer 1977
[2]: N.L Biggs: *Discrete Mathematics*, Second Edition. Oxford 2002
[3]: T. Nagell: *Lärobok I Algebra.* Almqvist och Wiksells Akademiska Handböcker 1949
[4]: S. Friedberg, A.Insel, L.Spence: *Linear Algebra*, Fourth Edition. Pearson 2014