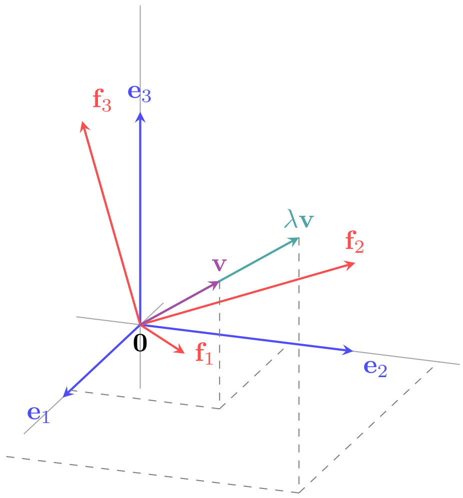
Generalized eigenvalues and the Weierstrass Canonical Form

Gloria, Bodhi, Branko, Albert



Abstract The Weierstrass Canonical Form is the definitive representation of all regular matrix pairs. The Jordan canonical form is part of all standard introductions to linear algbera, this article leverages this familairty with the Jordan form to provide a proof and therein exposition of the Weierstrass form.

Contents

Introduction

Almost all students studying in a STEM discipline are at some point taught fundamental techniques of linear algebra. Even within mathematics curricula these introductions to linear algebra terminate at most around the Jordan Canonical Form. This unfortunately fails to make even the enthusiastic students aware of the much larger world of intermediate to advanced linear algebra.

The Weierstrass Canonical Form (WCF) can present a stable stepping stone into the world of further linear algebra. The reason for this twofold. Firstly it is of interest to those studying applied mathematics, engineering and systems theory due to its practical applications in modeling linear systems. Secondly it should be of interest to students of pure mathematics thanks to its moderately deep geometric foundations and the algebraic techniques used in its proof.

This article, inspired majorly by [2], hopes to pick up where elementary knowledge of linear algebra left off - provide an intuitive form for Jordan Canonical form and provide an exposition of the WCF.

Disclaimer

All proofs presented herein, except when explicitly stated otherwise, are the authors creation and the result of a decent level of trial, tribulation and elation. Any similarity to existing proofs arise, by perhaps the most beautiful property of mathematics, from the structure of the objects being studied. However the question framework was provided by Stephan Trenn and help with completing the text in a timely manner and in a presentable form from Qianyi. Naturally praise should be directed to them and complaints to `lingchenyu2288@gmail.com`.

0 Preliminaries

0.1 Notation

For some linear operator $T : \mathcal{V} \to \mathcal{V}$, denote:

A finite dimensional vector space over the complex field	$\mathcal{V}$
Extended kernel of T	$\operatorname{extker} T = \bigcup_{i=0}^{\infty} \ker T^i$
The cardinality of the spectrum of a linear map T	$\sigma = \lvert \operatorname{spec} T \rvert$
The set of endomorphisms from $\mathcal{V}$ to $\mathcal{V}$	$\mathcal{L}(\mathcal{V})$
Determinant of a matrix representation A of a linear map T	$\lvert T \rvert$
Restriction of a map T on a subset S of its domain	$T\vert_S$
Composition of linear maps T and L	TL
Image of a vector $\mathbf{v}$ under a linear map T	$T\mathbf{v}$

0.2 Some definitions

Definition 0.1. We call m-many subspaces $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_m \subset \mathcal{V}$ independent if and only if the equation

$$0 = \mathbf{s}_1 + \mathbf{s}_2 + \dots + \mathbf{s}_m, \; \mathbf{s}_i \in \mathcal{S}_i$$

has only the solution $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m = \mathbf{0}$.

Definition 0.2. For subspaces $\mathcal{S}, \mathcal{S}' \subset \mathcal{V}$ we call $\mathcal{S}$ and $\mathcal{S}'$ disjoint if and only if $\mathcal{S} \cap \mathcal{S}' = \{\mathbf{0}\}$

Definition 0.3. For disjoint subspaces $\mathcal{S}, \mathcal{S}' \subset \mathcal{V}$ we define

$$\mathcal{S} \oplus \mathcal{S}' = \{\mathbf{v} + \mathbf{u} \mid \mathbf{v} \in \mathcal{S}, \mathbf{u} \in \mathcal{S}\}$$

which is also a subspace of $\mathcal{V}$.

1 The Jordan form

The existance of the Jordan canonical form (JCF) is one of the more central results of elementary linear algebra. This is not only because of its applications within systems theory, but also purely algebraically as a sort of 'invariant' of endomorphisms on an abstract vector space under isomorphism to $\mathbb{C}^n$.

We hope to convince the reader that the JCF can be considered a special case of the Weierstrass canonical form (WCF). Or conversely that the WCF is a generalisation of the JCF. In order to achieve this we will first present a proof of the JCF. And then use notions and techniques developed there to prove the Weierstrass form. Our intention is that during this process the entanglement of the two forms will emerge in a satisfying way. Furthermore, as the authors are algebraically predisposed, we will work through the lens of linear maps in need of a basis under which the maps' character is obvious. Quick transitions into the world of matrices will reveal the famous JCF and WCF matrix structures, these will be saved for conclusionary sections.

1.1 Problem statement

It is well known that all finite dimensional vector spaces are isomorphic to $\mathbb{C}^n$ for some n, for a proof see [1] Theorem 3.70. From this one can choose a basis for the vector space and by observing what any linear operator does to this basis arrive at a so called basis representation of that map, which is a matrix whose columns are the coefficients of the basis vectors under the map. This culminates in the fact that, for a fixed basis, the set of linear operators on a vector space with dimension n is isomorphic to the square matrices of that size $\mathbb{C}^{n\times n}$, again see [1] Theorem 3.71 for a proof. From here matrix theory and useful applications thereof abound.

However a question arises, since there are many bases for the vector space there must be many matrices who actually represent the same linear map. Given a matrix can we perform a systemic change of basis such that if any two matrices derive from the same linear operator, regardless of a choice of basis, we will arrive at the same matrix representation? Would such a 'canonical form' of a matrix reveal invariants of the underlying map under changes of basis ? Intuitively such a form should implicitly contain all the information contained in the map itself.

Formally we might re-state these questions as follows:

Let $T \in \mathcal{L}(\mathcal{V})$, can we construct a basis β of $\mathcal{V}$ such that for any bases γ and γ' we have ${}_\beta[T]_\gamma = {}_\beta[T]_{\gamma'}$?

In the following section we hope to present an intuitive yet rigorous anwser to this question. With the goal of the proof evolving naturally from the question.

1.2 Where to start: Invariant subspaces

Consider a linear operator $T \in \mathcal{L}(\mathcal{V})$, where $\mathcal{V}$ is a finite dimensional vector space. For any subspace $\mathcal{S} \subset \mathcal{V}$ the image of T over $\mathcal{S}$ is, in general, some other arbitrary subspace. It seems reasonable then to want to consider subspaces which preserve their structure under T. The following definition captures this idea.

Definition 1.1. A subspace $\mathcal{S}$ is called *invariant* under the linear map T if

$$\mathbf{v} \in \mathcal{S} \implies T\mathbf{v} \in \mathcal{S}$$

It is best to give some examples of such subspaces before investigating them further:

1.2.1 Examples and properties

Example 1.2. The so called eigenspace of T at eigenvalue λ

$$E_\lambda = \{\mathbf{v} : T\mathbf{v} = \lambda\mathbf{v}\}$$

is invariant - each element is scaled under T, which is still in E_λ.

Example 1.3. The image

$$\operatorname{Im} T = \{T\mathbf{v} : \mathbf{v} \in \mathcal{V}\}$$

is invariant by definition:

$$\mathbf{v} \in T(\mathcal{V}) \implies T\mathbf{v} \in T(\mathcal{V})$$

Example 1.4. The generalized kernel

$$K = \{\mathbf{v} : \exists k \in \mathbb{N} \text{ s.t } T^k\mathbf{v} = \mathbf{0}\}$$

is invariant under T. This is indeed a subspace, and the invariance here again follows from its definition.

$$\mathbf{v} \in K \implies \exists k,\ T^k\mathbf{v} = \mathbf{0} \implies T^{k-1}T\mathbf{v} = \mathbf{0} \implies T\mathbf{v} \in K$$

This set has a stronger structure than being just invariant. It can be shown that

$$\mathbf{v} \notin K \implies T\mathbf{v} \notin K$$

which is not true in general for invariant subspaces.

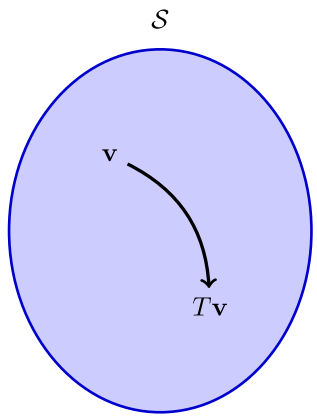


Figure 1: A subspace $\mathcal{S}$ invariant under T

Because of their behavior we could try to 'build' $\mathcal{V}$ out of invariant subspaces - similar to how primes can be used to build the integers. By this we really mean finding some number of subspaces, say k-many, such that if β_i is a basis for the ith subspace $\mathcal{S}_i$ we have:

$$\beta = \bigcup_{i=1}^{k} \beta_i \quad \text{is a basis for } \mathcal{V}$$

$$\text{or equivalently} \qquad \bigoplus_{i=1}^{k} \mathcal{S}_i = \mathcal{V}$$

If one succeeds in this, then the matrix representation of T would take the form

$$[T]_\beta = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{pmatrix}$$

This structure is strictly due to the invariance of the subspaces. If $\beta_i = \{\beta_{i1}, \beta_{i2}, ..., \beta_{im}\}$, for some $m \in \mathbb{N}$, then

$$T(\beta_{ij}) \in \mathcal{S}_i \quad \forall j \leq k,$$

and therefore may be expressed only in terms of $\beta_{i1}, \beta_{i2}, ..., \beta_{im}$ which implies the coordinates of this vector in the $\beta_{l \neq i}$ bases must all be zero. Therefore this vector, when expressed in β, is only nonzero in the $i1$ to im rows. So the column vectors of $[T]_\beta$ from different β_i can never be nonzero in the same row. Therefore the above.

Our goal is to achieve this decomposition whilst choosing the best possible invariant subspaces - so that the above A_i blocks are as simple as possible.

Before continuing some observations of invariant subspaces are in order.

Lemma 1.5. Invariant subspaces (except $\{\mathbf{0}\}$) $\mathcal{S}$ under T all contain an eigenvector of T.

Proof. This is because the linear map

$$T^* : \mathcal{S} \to \mathcal{S},\ T^*\mathbf{v} = T\mathbf{v}$$

is well defined due to the invariance. Then we just pick the eigenvector of T^* which must always exist.[1] $\square$

Lemma 1.6. If $\mathcal{S}$ is invariant under T and L, then $\mathcal{S}$ is invariant under $T + xL$ for any $x \in \mathbb{C}$.

Proof. If $\mathbf{v} \in \mathcal{S}$, then

$$(T + xL)\mathbf{v} = T\mathbf{v} + xL\mathbf{v} \in \mathcal{S}$$

$\square$

Lemma 1.7. If $\mathcal{S}$ is invariant under T then $\mathcal{S}$ is also invariant under T^n for any $n \in \mathbb{N}$.

Proof. The statement is true for $n = 0$, as

$$T^0(\mathcal{S}) = I(\mathcal{S}) = \mathcal{S} \subseteq \mathcal{S}$$

If the statement is true for some n,

$$T^{n+1}(\mathcal{S}) = T(T^n(\mathcal{S})) \subseteq T(\mathcal{S}) \subseteq \mathcal{S}$$

By induction we are done. $\square$

Corollary 1.8. If $\mathcal{S}$ is invariant under T, then it is also invariant under any polynomial of T.

---

[1] Turns out this is due to a property of the complex field called *algebraic closure*. See [4] page 104 for a more detailed discussion of this.

1.3 Why diagonalization is not enough

We start with considering the simplest invariant subspaces - the eigenspaces. For some $T \in \mathcal{L}(\mathcal{V})$, let the spectrum of T be $\{\lambda_1, \lambda_2, ..., \lambda_\sigma\}$, and define the corresponding eigenspaces

$$E_{\lambda_i} = \{\mathbf{v} \in \mathcal{V} : T\mathbf{v} = \lambda_i \mathbf{v}\}$$

It is important to note that

$$E_{\lambda_i} \cap E_{\lambda_j} = \{\mathbf{0}\} \text{ , for all } i \neq j \leq \sigma,$$

and that for an arbitrary basis β_i for E_{λ_i}, the set

$$\bigcup_{i=1}^{\sigma} \beta_i$$

is linearly independent.

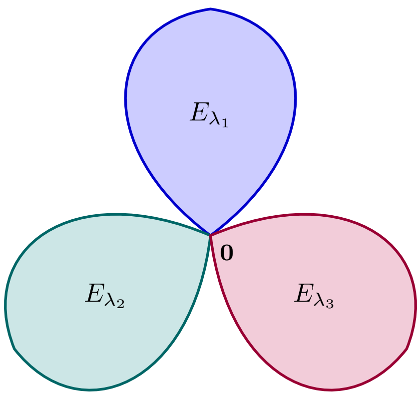


Figure 2: Three eigenspaces of some map T

If the direct sum of the eigenspaces is $\mathcal{V}$ then the above union could be out basis.

Sadly, this is often not the case. An example is the matrix

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

If the direct sum of the eigensapces is not enough to span $\mathcal{V}$, we could try to somehow expand each eigenspace into a bigger subspace such that they do span $\mathcal{V}$. However we must be careful as preserving the invariance and disjointness of the subspaces is essential.

1.4 Generalising the eigenspaces

For each E_{λ_i}, denote

$$G^k_{\lambda_i} = \ker(T - \lambda_i I)^k$$

Note that $G^1_{\lambda_i} = E_{\lambda_i}$ and $G^k_{\lambda_i} \subseteq G^{k+1}_{\lambda_i}$.

We will define

$$\boxed{G_{\lambda_i} = \bigcup_{k=1}^{\infty} G^k_{\lambda_i}}$$

Intuitively this can be thought of as extending the subspace to include vectors that are 'close' to becoming eigenvectors. In the sense that they are a finite number of iterations of $T - \lambda_i I$ away from $\mathbf{0}$.

We still have to show that this is a valid choice. More specifically, we need to show that each G_{λ_i} is invariant under T, and

$$\bigoplus_{i=1}^{\sigma} G_{\lambda_i} = \mathcal{V}$$

1.5 Properties of the generalised eigenspaces

The definition of the extended eigenspace uses infinity as an upperbound. However, infinity is tricky to work with sometimes, so we prove the following lemma:

Lemma 1.9. For any $\lambda_i \in \operatorname{spec} T$, there exists $\kappa_i \in \mathbb{N}$ s.t. for any $\kappa_i \leq k \in \mathbb{N}$, we have $G_{\lambda_i} = G_{\lambda_i}^k = G_{\lambda_i}^{\kappa_i}$.

Proof. Since $G_{\lambda_i}^k$ is a subspace of $\mathcal{V}$ and $G_{\lambda_i}^k \subseteq G_{\lambda_i}^{k+1}$, $(\dim G_{\lambda_i}^k)_{k\in\mathbb{N}}$ is a monotone integer sequence bounded by $\dim \mathcal{V}$. Hence, there exists some $\kappa_i \in \mathbb{N}$ s.t. $\dim G_{\lambda_i}^k = \dim G_{\lambda_i}^{\kappa_i}$ for all $k \geq \kappa_i$.

Since for all $k \geq \kappa_i$, $G_{\lambda_i}^k \supseteq G_{\lambda_i}^{\kappa_i}$, we have $G_{\lambda_i}^k = G_{\lambda_i}^{\kappa_i}$.

$$G_{\lambda_i} = \bigcup_{k=1}^{\infty} G_{\lambda_i}^k = G_{\lambda_i}^{\kappa_i} \cup \bigcup_{k=1}^{\kappa_i - 1} = G_{\lambda_i}^{\kappa_i}$$

□

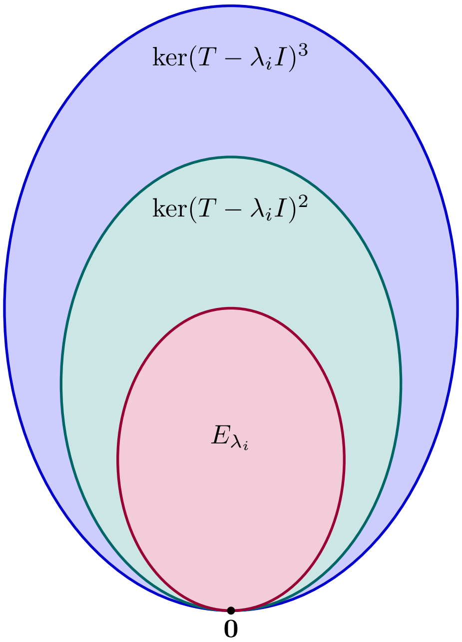


Figure 3: The generalised eigenspace G_{λ_i} of the eigenvalue λ_i. In this case G_{λ_i} stops growing at $\ker(T - \lambda_i I)^3$.

With this, we can show:

Lemma 1.10. G_{λ_i} is invariant under any polynominal of T.

Proof. G_{λ_i} is the extended kernel of $T - \lambda_i I$. From example 1.4, we have G_{λ_i} is invariant under $T - \lambda_i I$.

Applying lemma 1.8, we have G_{λ_i} is invariant under $T = T - \lambda_i I + \lambda_i I$ and thus, any polynominal of T. □

Corollary 1.11. Notably, G_{λ_i} is invariant under $T - xI$ for any choice of $x \in \mathbb{C}$.

However, we can go one step further. Since G_{λ_i} is invariant under $T - xI$, $T - xI$ is a well-defined function on G_{λ_i}. We claim

Lemma 1.12. For all $x \neq \lambda_i \in \mathbb{C}$, $T - xI$ is a bijection on G_{λ_i}

Proof. From lemma 1.9, there exists

$$\begin{aligned}\ker_{G_{\lambda_i}}(T - xI) &= \{\mathbf{v} \in G_{\lambda_i} : (T - xI)\mathbf{v} = \mathbf{0}\} \\ &= \{\mathbf{v} \in \mathcal{V} : T\mathbf{v} = x\mathbf{v} \wedge (T - \lambda_i I)^{\kappa_i}\mathbf{v} = \mathbf{0}\}\end{aligned}$$

Note that since $T\mathbf{v} = x\mathbf{v}$,

$$(T - \lambda_i I)\mathbf{v} = T\mathbf{v} - \lambda_i \mathbf{v} = (x - \lambda_i)\mathbf{v}$$

where $x - \lambda_i \neq 0$ from the assumption.
Hence,

$$\begin{aligned}\ker_{G_{\lambda_i}}(T - xI) &= \{\mathbf{v} \in \mathcal{V} : T\mathbf{v} = x\mathbf{v} \wedge (x - \lambda_i)^{\kappa_i}\mathbf{v} = \mathbf{0}\} \\ &\subseteq \{\mathbf{v} \in \mathcal{V} : (x - \lambda_i)^{\kappa_i}\mathbf{v} = \mathbf{0}\} \\ &= \{\mathbf{v} \in \mathcal{V} : \mathbf{v} = \mathbf{0}\} \\ &= \{\mathbf{0}\}\end{aligned}$$

Since G_{λ_i} has finite dimension, $T - xI$ is a bijection on G_{λ_i}. □

Lemma 1.13.

$$\ker(T - \lambda_i I)^{\kappa_i} \oplus \operatorname{Im}(T - \lambda_i I)^{\kappa_i} = \mathcal{V}$$

where κ_i is as defined in Lemma 1.9.

Proof. First we will show that the two vector spaces are disjoint.

Suppose $\mathbf{v} \in \operatorname{Im}(T - \lambda_i I)^{\kappa_i}$, then there exist $\mathbf{u} \in \mathcal{V}$ s.t. $\mathbf{v} = (T - \lambda_i I)^{\kappa_i}\mathbf{u}$. Since $\mathbf{v} \in \ker(T - \lambda_i I)^{\kappa_i}$,

$$\begin{aligned}(T - \lambda_i I)^{2\kappa_i}\mathbf{u} &= (T - \lambda_i I)^{\kappa_i}\left((T - \lambda_i I)^{\kappa_i}\mathbf{u}\right) \\ &= (T - \lambda_i I)^{\kappa_i}\mathbf{v} \\ &= \mathbf{0}\end{aligned}$$

Hence, $\mathbf{u} \in \operatorname{extker}(T - \lambda_i I) = \ker(T - \lambda_i I)^{\kappa_i}$.

$$\mathbf{v} = (T - \lambda_i I)^{\kappa_i}\mathbf{u} = \mathbf{0}$$

Therefore, $\ker(T - \lambda_i I)^{\kappa_i} \cap \operatorname{Im}(T - \lambda_i I)^{\kappa_i} = \{\mathbf{0}\}$.

Since the two vector spaces are disjoint they are also independent.

By the dimension theorem:

$$\begin{aligned}
&\dim\left(\ker(T - \lambda_i I)^{\kappa_i} \oplus \operatorname{Im}(T - \lambda_i I)^{\kappa_i}\right) \\
=&\dim\ker(T - \lambda_i I)^{\kappa_i} + \dim\operatorname{Im}(T - \lambda_i I)^{\kappa_i} \\
=&\dim\mathcal{V}
\end{aligned}$$

Hence $\ker(T - \lambda_i I)^{\kappa_i} \oplus \operatorname{Im}(T - \lambda_i I)^{\kappa_i} = \mathcal{V}$. □

Theorem 1.14. $G_{\lambda_1} \oplus G_{\lambda_2} \oplus \cdots \oplus G_{\lambda_\sigma} = \mathcal{V}$

Proof. We will show this by induction on σ.

Note that $\{\mathbf{0}\} \oplus G = G$. So for $\sigma = 0$, we will define $G_{\lambda_1} \oplus \cdots \oplus G_{\lambda_\sigma} = \{\mathbf{0}\}$.

Base case:
Recall since for any n-dimensional vector space $\mathcal{V}$ where $n > 0$ there exists eigenvalues for T. If $\sigma = 0$ then we have $n = 0$. Thus, $\mathcal{V} = \{\mathbf{0}\}$ and the theorem holds.

Suppose the theorem is true for some $\sigma - 1 \in \mathbb{N}$.
Consider the subspace $\mathcal{V}' = \operatorname{Im}(T - \lambda_\sigma I)^{\kappa_\sigma}$. Note that from Lemma 1.13, we have $\mathcal{V} = G_{\lambda_\sigma} \oplus \mathcal{V}'$. As $\mathcal{V}'$ is an image of $T - \lambda_\sigma$ on some subspace, we can apply Example 1.3 to show $\mathcal{V}'$ is invariant under T.

Since $\mathcal{V}' \cap G_{\lambda_\sigma} = \{\mathbf{0}\}$, the invariance of T $\operatorname{spec}_{\mathcal{V}'} T \subseteq \operatorname{spec}_{\mathcal{V}} T \setminus \{\lambda_\sigma\}$.
$\forall i < \sigma$, since $T - \lambda_\sigma I$ is a bijection on G_i (Lemma 1.12),

$$G_{\lambda_i} = (T - \lambda_{k-1} I)^{\kappa_\sigma}(G_{\lambda_i}) \subseteq \operatorname{Im}(T - \lambda_{k-1} I)^{\kappa_\sigma} = \mathcal{V}'$$

Hence, $\operatorname{spec}_{\mathcal{V}'} T = \operatorname{spec}_{\mathcal{V}} T \setminus \{\lambda_\sigma\}$, and the G_{λ_i} in $\mathcal{V}'$ is the same as in $\mathcal{V}$. We can apply induction hypothesis on $\mathcal{V}'$ to get

$$\mathcal{V}' = G_{\lambda_1} \oplus G_{\lambda_2} \oplus \cdots \oplus G_{\lambda_{\sigma-1}}$$

Therefore,

$$\mathcal{V} = \mathcal{V}' \oplus G_{\lambda_\sigma} = G_{\lambda_1} \oplus G_{\lambda_2} \oplus \cdots \oplus G_{\lambda_{\sigma-1}} \oplus G_{\lambda_\sigma}$$

□

1.6 Introducing the null-chains

The definition of G_{λ_i} hints at what might be an appropriate choice of basis. Recall that each vector in G_{λ_i} eventually gets mapped to the zero vector under $T - \lambda_i I$ i.e. the map is nilpotent on G_{λ_i}. So it is natural to define the following sets.

Definition 1.15. A *null-chain* under linear map L is a set of (one or more) **non-zero** vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..\mathbf{v}_n\}$ such that

$$\begin{aligned} L(\mathbf{v}_1) &= 0 \\ L(\mathbf{v}_{i+1}) &= \mathbf{v}_i \quad \text{for} \quad i \leq n-1 \end{aligned}$$

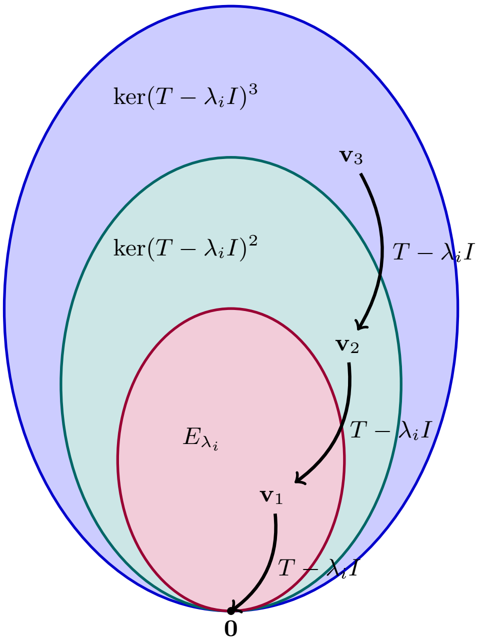


Figure 4: A null-chain $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ under $T - \lambda_i I$ in some generalised eigenspace

This definition immediately yields the following lemma.

Lemma 1.16. A null-chain under a linear map L forms a linearly independent set.

Proof. Consider a null-chain

$$\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$$

and let

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n = \mathbf{0}. \tag{1}$$

Applying L^{n-1} on both sides and considering the definition of a null-chain gives

$$c_1L^{n-1}(\mathbf{v}_1) + ... + c_nL^{n-1}(\mathbf{v}_n) = \mathbf{0} \implies c_n\mathbf{v}_1 = \mathbf{0} \implies c_n = 0.$$

Substituting back in (1) and applying L^{n-2} on both sides gives

$$c_{n-1}\mathbf{v}_1 = \mathbf{0} \implies c_{n-1} = 0.$$

Continuing this inductive process we obtain

$$c_1 = c_2 = ... = c_n = 0$$

□

Corollary 1.17. A collection of null-chains forms a linearly independent set.

We should check weather a null-chain in G_{λ_i} under $T - \lambda_i I$ is a basis for G_{λ_i}. A quick counter example shows in general this is not the case. Consider the matrix

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

Note that $2 \in \operatorname{spec} A$ with $\dim G_2 = 3$. However the longest null-chain for $A - 2I$ is of length 2 so one null-chain cannot span G_3.

The idea is to pick multiple such chains. Then it can be shown:

Theorem 1.18. *A nonzero finite-dimensional vector space $\mathcal{V}$ with a nilpotent linear map L always has a basis composed of disjoint null-chains.*

Proof. We perform strong induction on the dimension of $\mathcal{V}$. Let $\mathcal{V}$ be a vector space with $\dim \mathcal{V} = 1$ and L a nilpotent linear operator on V.

Any $\mathbf{v} \neq \mathbf{0} \in \mathcal{V}$ is a basis of $\mathcal{V}$ and since $L\mathbf{v} = \mathbf{0}$ the null-chain $\{\mathbf{v}\}$ is a basis.

Now that we have our base case let $\dim \mathcal{V} = k$.

Hypothesizing that:

Every vector space with dimension $\leq k-1$ which is nilpotent under L has a basis of disjoint null-chains.

The map L is not invertible on $\mathcal{V}$ by its nilpotency. Therefore

$$\dim L(\mathcal{V}) < \dim \mathcal{V} \tag{2}$$

The dimension theorem gives

$$\dim \mathcal{V} = \dim L(\mathcal{V}) + \dim \ker L \tag{3}$$

Note that $L(\mathcal{V})$ is a subspace of $\mathcal{V}$ therefore it is also nilpotent under L. The induction hypothesis and (2) guarantee there is a basis of $L(\mathcal{V})$ of the form

$$\begin{aligned}
C_1 &= \{\beta_{11}, \beta_{12}, \ldots, \beta_{1k_1}\},\\
C_2 &= \{\beta_{21}, \beta_{22}, \ldots, \beta_{2k_2}\},\\
&\vdots\\
C_l &= \{\beta_{l1}, \beta_{l2}, \ldots, \beta_{lk_l}\}.
\end{aligned}$$

where the C_i's are ℓ-many disjoint nullchains such that

$$\begin{aligned}
L(\beta_{i1}) &= 0\\
L(\beta_{i(k+1)}) &= \beta_{ik}
\end{aligned}$$

We have that

$$|\bigcup_{i=1}^{\ell} C_i| = \dim L(\mathcal{V}).$$

Note that the first vector in each chain is in $\ker(L)$ and therefore

$$\ell \leq \dim\ker L \tag{4}$$

because of the linear independence of the β_{i1}'s.
Consider the last vector β_{ik_i} in each chain. It is in $L(\mathcal{V})$ giving

$$\forall i \leq l,\ \exists \mathbf{u}_i\ :\ L(\mathbf{u}_i) = \beta_{ik_i} \tag{5}$$

Thus we can upgrade our initial chains by adding one more vector

$$C_i \cup \{\mathbf{u}_i\},\ i \leq \sigma \tag{6}$$

and due to their independence due to Corollary 1.17 we have a linearly independent set of size $\dim L(\mathcal{V}) + \ell$.

If $\ell = \dim\ker L$ we are done by the dimension theorem and the collection of newly constructed nullchains at (6) is the desired basis.

Otherwise, extend the linearly independent set $\{\beta_{11}, \beta_{21}, ..., \beta_{l1}\}$into a basis of $\ker L$:

$$\beta_{11}, ..., \beta_{l1}, \mathbf{w}_{l+1}, \mathbf{w}_{l+2} ..., \mathbf{w}_{\dim\ker L}$$

We have constructed $\dim\ker L - l$ more chains all of length one.

Collecting everything we have:

$$\begin{aligned}
&\{\mathbf{w}_{\dim\ker L}\},\\
&\{\mathbf{w}_{\dim\ker L-1}\},\\
&\vdots\\
&\{\mathbf{w}_{l+1}\}\\
&C_1 \cup \{\mathbf{u}_1\}\\
&C_2 \cup \{\mathbf{u}_2\}\\
&\vdots\\
&C_l \cup \{\mathbf{u}_l\}.
\end{aligned}$$

The union of everything above is a linearly independent set due to Corollary 1.17. And has cardinality

$$\begin{aligned}
&\dim\ker L - l + \dim T(\mathcal{V}) + l\\
&= \dim\ker L + \dim T(\mathcal{V})\\
&= \dim \mathcal{V},
\end{aligned}$$

□

1.7 Picking a basis

The map $T - \lambda_i I$ is nilpotent on the generalized eigenspace G_{λ_i} and therefore by Theorem 1.18 we may construct an appropriate basis β_i composed of nullchains for G_{λ_i}.

Consider the matrix representation of T for the basis β_i of G_{λ_i}. A single null-chain N of length $m+1$ in this basis has the following structure:

$$\begin{aligned}
(T - \lambda_i I)\mathbf{v}_1 &= \mathbf{0} \\
(T - \lambda_i I)\mathbf{v}_2 &= \mathbf{v}_1 \\
&\vdots \\
(T - \lambda_i I)\mathbf{v}_m &= \mathbf{v}_{m-1}
\end{aligned}$$

which in terms of T gives

$$\begin{aligned}
T\mathbf{v}_1 &= \lambda_i \mathbf{v}_1 \\
T\mathbf{v}_2 &= \lambda_i \mathbf{v}_2 + \mathbf{v}_1 \\
&\vdots \\
T\mathbf{v}_n &= \lambda_i \mathbf{v}_m + \mathbf{v}_{m-1}
\end{aligned}$$

Computing the matrix representation of T with respect to the nullchain above yields

$$J_{\lambda_i,1} = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & & 1 \\ 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix} \in \mathbb{C}^{m \times m} \tag{7}$$

This structure is the so-called Jordan block. Of course this does not really make sense as the nullchain N is not a basis for any meaningful subspace we have studied. However since the basis β_i is composed of multiple disjoint null-chains we can use (7) to find the matrix representation of the restriction of T onto G_{λ_i}

$$[T|_{G_{\lambda_i}}]_{\beta_i} = J_{\lambda_i} = \begin{pmatrix} J_{\lambda_i,1} & 0 & \cdots & 0 \\ 0 & J_{\lambda_i,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{\lambda_i,k} \end{pmatrix} \tag{8}$$

where each $J_{i\leq k}$ is of form (7).

Theorem 1.14 and theorem 1.18 immediately give $\beta = \bigcup_{i=1}^{\sigma} \beta_i$ is a basis for V. Furthermore all G_{λ_i} are disjoint and T-invariant. So computing the basis representation of T using β gives

$$[T]_\beta = \begin{pmatrix} J_{\lambda_1} & 0 & \cdots & 0 \\ 0 & J_{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{\lambda_\sigma} \end{pmatrix} \tag{9}$$

where each block J_{λ_i} is of form (8). Ultimately this is the desired Jordan canonical form of the linear operator T.

2 The Weierstrass form

2.1 Introduction

The Weierstrass form is a canonical form of two matrices A, B acting on the same space. It is often referred to as generalized Jordan form. The motivation for Weierstrass form goes further than just generalizing Jordan and finding an appropriate basis - it plays crucial role in understanding linear systems of the form

$$B\dot{x} = Ax$$

The relevance of such systems will be discussed in Chapter 5.

Let $A, B \in \mathcal{L}(V)$.

Recalling the introduction of this paper, canonical forms are obtained by decomposing the space into invariant subspaces on which the operators have particularly simple matrix representations.

Assume that we are able to find k subspaces $S_1, S_2, ..., S_k$ such that each S_i is T and L-invariant, and

$$\bigoplus_{i=1}^{k} S_k = V.$$

Picking a basis β_i for S_i and composing $\bigcup \beta_i$ gives us a block-diagonal matrix representation of A *and* B, respectively, *under the same basis.*

The problem is: such subspaces rarely exist. The maps A and B might be too independent of each other and therefore a subspace might act nice under one map, but arbitrarily under the other. For example, take

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

The line $W = \text{span}(e_1)$ is $A-$invariant: $w \in W \implies Aw = w$.

But, $w \in W \implies Bw \in \text{span}(e_2)$, so W is not $B-$invariant.

In fact, for a pair A, B there might be no non-trivial commonly invariant subspace.

Therefore, instead of attempting to describe the maps A and B via a specific basis, we do so with two different bases β and γ such that the matrices

$$_\beta[A]_\gamma,\ \ _\beta[B]_\gamma \tag{10}$$

are of particularly nice type. Prioritizing simplification, Weierstrass canonical form pursues this goal by having to put a restriction on A and B.

We will define a property of the pairs (A, B) referred to as *regular*.

Definition 2.1. A pair of maps (A, B) is called *regular* iff

$$\exists \lambda \in \mathbb{C} \ s.t \ |A + \lambda B| \neq 0 \tag{11}$$

2.2 Statement

We claim that:

For a regular pair of maps (A, B), there exist bases β, γ such that

$$_\beta[A]_\gamma = \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix} \tag{12}$$

$$_\beta[B]_\gamma = \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix}. \tag{13}$$

For the case where A and B are matrices, the statement is equivalent to:

There exists invertible P and Q s.t.

$$(PAQ, PBQ) = \left(\begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix}, \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix}, \right) \tag{14}$$

Where I is the indentity matrix, J is a matrix in Jordan form and N is a nil-potent Jordan matrix.

2.3 Matrix pencils

A useful tool for studying the pairs (A, B) where $A, B \in \mathcal{L}(\mathcal{V})$ is the so-called matrix pencil [2] induced by the pair:

Definition 2.2. The function

$$P(x) = A + Bx, \ x \in \mathbb{C}$$

is called the *linear matrix pencil of the pair* (A, B)

We say that a matrix pencil of the pair (A, B) is *regular* if and only if (A, B) is regular. It is called *singular* otherwise.

The condition (11) enables us to select a non-singular element of the matrix pencil which will help us derive a canonical form of (10).

[2] For a rigorous analysis matrix pencils see [3] page Chapter 12, page 24.

Before continuing, we will first derive some properties regarding the regularity of a matrix pencil.

Lemma 2.3. A matrix pencil $A + Bx$ where $A, B \in \mathcal{L}(\mathcal{V})$ is singular if and only if for all $x \in \mathbb{C}$, exists nonzero $\mathbf{v} \in \mathcal{V}$ such that $(A + Bx)\mathbf{v} = \mathbf{0}$.

Proof.

$$\begin{aligned} |A + Bx| = 0 &\iff \forall x \in \mathbb{C},\ A + Bx = 0 \\ &\iff \forall x \in \mathbb{C},\ A + Bx \text{ is singular} \\ &\iff \forall x \in \mathbb{C},\ \exists \mathbf{0} \neq \mathbf{v} \in \mathcal{V},\ (A + Bx)\mathbf{v} = \mathbf{0} \end{aligned}$$

□

Lemma 2.4. $A + Bx$ is regular if and only if $B + Ax$ is regular.

Proof. Note that the forward direction is equivalent to the backward direction.

If $A + Bx$ is regular, then there exists $x \in \mathbb{C}$ s.t. $|A + Bx| \neq 0$.

If $x = 0$, then A is invertible.

$$|Ax + B| = |A(xI + A^{-1}B)| = |A||xI + A^{-1}B|$$

Pick $x \notin \operatorname{spec} -A^{-1}B$, we have $|Ax + B| \neq 0$.

Otherwise,

$$|Ax^{-1} + B| = |x^{-1}(A + Bx)| = x^{-n}|A + Bx| \neq 0$$

□

Lemma 2.5. Let $A + Bx$ be a regular matrix pencil on $\mathcal{V}$, if $B^{-1}(A(\mathcal{S})) = \mathcal{S}$ for some subspace $\mathcal{S} \subseteq \mathcal{V}$, then $\ker_{\mathcal{S}} A = \{\mathbf{0}\}$.

Proof. Suppose $\ker_{\mathcal{S}} A \neq \{\mathbf{0}\}$.

Since $B^{-1}(A(\mathcal{S})) = \mathcal{S}$,

$$\operatorname{Im}_{\mathcal{S}} B = B(\mathcal{S}) = B(B^{-1}(A(\mathcal{S}))) \subseteq A(\mathcal{S})$$

Hence, for any $x \in \mathbb{C}$, we have

$$\operatorname{Im}_{\mathcal{S}}(A + Bx) \subseteq \operatorname{Im}_{\mathcal{S}} A + \operatorname{Im}_{\mathcal{S}} Bx = \operatorname{Im}_{\mathcal{S}} A$$

And

$$\dim \ker_{\mathcal{S}}(A + Bx) = \dim \mathcal{S} - \dim \operatorname{Im}_{\mathcal{S}}(A + Bx) \geq \dim \mathcal{S} - \dim \operatorname{Im}_{\mathcal{S}} A = \dim \ker_{\mathcal{S}} A > 0$$

Therefore $A + Bx$ is singular for all $x \in \mathbb{C}$. This contradicts the assumption for the matrix pencil to be regular.

By contradiction, $\ker_{\mathcal{S}} A = \{\mathbf{0}\}$ □

2.4 Method

Let $\mathcal{V}$ be a n-dimensional vector spaces.

Let $A + Bx$ be a regular matrix pencil, where A, B are linear maps from $\mathcal{V}$ to $\mathcal{V}$.

Define

$$\mathcal{S}_0 = \{\mathbf{0}\} \tag{15}$$
$$\mathcal{S}_{m+1} = B^{-1}(A(\mathcal{S}_m)) \tag{16}$$
$$\mathcal{S} = \bigcup_{i \in \mathbb{N}} \mathcal{S}_i \tag{17}$$
$$\tag{18}$$

$$\mathcal{W}_0 = \mathcal{V} \tag{19}$$
$$\mathcal{W}_{m+1} = A^{-1}(B(\mathcal{W}_m)) \tag{20}$$
$$\mathcal{W} = \bigcap_{i \in \mathbb{N}} \mathcal{W}_i \tag{21}$$
$$\tag{22}$$

Note that here, A, B and A^{-1}, B^{-1} denote the image and pre-image, respectively.

Observation 2.6. For any $m \in \mathbb{N}$, we have $\mathcal{S}_m \subseteq \mathcal{S}_{m+1}$ and $\mathcal{W}_m \supseteq \mathcal{W}_{m+1}$

Due to the finite-dimensionality of $\mathcal{V}$, we conclude the following.

Corollary 2.7. There exists $\kappa, \kappa' \in \mathbb{N}$ such that for all $k \geq \kappa, k' \geq \kappa'$, we have

$$\mathcal{S} = \mathcal{S}_k$$
$$\mathcal{W} = \mathcal{W}_{k'}$$

As a consequence of this, we have the following lemma:

Lemma 2.8. A is injective on $\mathcal{S}$ and B is injective on $\mathcal{W}$.

Proof. From Corollary 2.7, we have

$$\mathcal{S} = \mathcal{S}_{\kappa+1} = B^{-1}(A(\mathcal{S}_\kappa)) = B^{-1}(A(\mathcal{S}))$$

With this we may directly apply Lemma 2.5, and conclude that A is injective on $\mathcal{S}$.

For the other case, recall that (A, B) is regular if and only if (B, A) is regular (Lemma 2.4). Therefore, we similarly conclude that B is injective on W

□

Recall the definitions for $\mathcal{S}$ and $\mathcal{W}$. $\mathcal{S}_{m+1} = B^{-1}(A(\mathcal{S}_m))$,

$$\mathbf{v} \in \mathcal{S}_{m+1} \iff \exists \mathbf{u} \in \mathcal{S}_m, B\mathbf{v} = A\mathbf{u}$$

Hence, an alternative definition for $\mathcal{S}_m$ is

$$\mathcal{S}_m = \{\mathbf{v}_m \in \mathcal{V} : \exists \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{m-1} \in \mathcal{V},\ \mathbf{v}_0 = \mathbf{0} \wedge \forall i \in [m],\ B\mathbf{v}_{i+1} = A\mathbf{v}i\} \tag{23}$$

Similarly, we have the alternative defination for $\mathcal{W}_m$ as

$$\mathcal{W}_m = \{\mathbf{v}_0 \in \mathcal{V} : \exists \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in \mathcal{V},\ \forall i \in [m],\ B\mathbf{v}_{i+1} = A\mathbf{v}_i\} \tag{24}$$

We have redefined $\mathcal{S}$ and $\mathcal{W}$ sequentially due to having a more descrete view. Note that due to the injectiveness of A on $\mathcal{S}$, we can make the following observation:

Observation 2.9. For all $\mathbf{v} \in \mathcal{S}$, there exists unique $\mathbf{u} \in \mathcal{S}$ s.t. $B\mathbf{v} = A\mathbf{u}$.

Hence, if for some vectors $\mathbf{v} \in \mathcal{S}_{m+1}$ and $\mathbf{u} \in \mathcal{S}$, $B\mathbf{v} = A\mathbf{u}$, then we have $\mathbf{u} \in \mathcal{S}_m$

Since $A^{-1}(B(\mathcal{W})) = \mathcal{W}$, B is injective on $\mathcal{W}$. So $\forall \mathbf{v} \in \mathcal{W}$, $\exists! \mathbf{u} \in \mathcal{W}$ s.t. $A\mathbf{v} = B\mathbf{u}$.

2.5 Characterising $\mathcal{S}$ and $\mathcal{W}$

To show that $\mathcal{S}$ and $\mathcal{W}$ are a good division of $\mathcal{V}$, we need to show $\mathcal{S} \oplus \mathcal{W} = \mathcal{V}$.

To do so, we might want to take advantage of Lemma 1.13 by writing $\mathcal{S}$ and $\mathcal{W}$ as the extended kernel and image of some linear map X.

In fact, Equation 15 is explicitly the definition of the extended kernel of the map $A^{-1}B$, given that A^{-1} exists. Similarly, Equation 19 is the generalized image of $A^{-1}B$. The only thing that stands in our way is that A might not be invertible.

For this purpose, we will introduce an equivalent definition for $\mathcal{S}_m$ and $\mathcal{W}_m$.

Lemma 2.10. For any $\lambda \in \mathbb{C}$, we have

$$\begin{aligned}\mathcal{S}_{m+1} &= B^{-1}((A + \lambda B)(\mathcal{S}_m)) \\ \mathcal{W}_{m+1} &= (A + \lambda B)^{-1}(B(\mathcal{W}_m))\end{aligned}$$

Proof. Recall that

$$\begin{aligned}\mathcal{S}_m &= \{\mathbf{v}_m : \exists \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{m-1},\, \mathbf{v}_0 = \mathbf{0} \wedge \forall i \in [m],\, B\mathbf{v}_{i+1} = A\mathbf{v}_i\} \\ \mathcal{W}_m &= \{\mathbf{v}_0 : \exists \mathbf{v}_1, \ldots, \mathbf{v}_m,\, \forall i \in [m],\, B\mathbf{v}_{i+1} = A\mathbf{v}_i\}\end{aligned}$$

Given sequence $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m$, for any $\lambda \in \mathbb{C}$, one can construct $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_m$ defined as:

$$\mathbf{u}_i = \sum_{j=0}^{i} \binom{i}{j} \lambda^{i-j} \mathbf{v}_j$$

We have

$$
\begin{aligned}
B\mathbf{u}_{i+1} &= B\sum_{j=0}^{i+1}\binom{i+1}{j}\lambda^{i+1-j}\mathbf{v}_j \\
&= B\binom{i+1}{0}\lambda^{i+1}\mathbf{v}_0 + B\sum_{j=1}^{i}\binom{i+1}{j}\lambda^{i+1-j}\mathbf{v}_j + B\binom{i+1}{i+1}\lambda^0\mathbf{v}_{i+1} \\
&= \lambda B\binom{i}{0}\lambda^{i}\mathbf{v}_0 + B\sum_{j=1}^{i}\left(\binom{i}{j}+\binom{i}{j-1}\right)\lambda^{i+1-j}\mathbf{v}_j + \binom{i}{i}\lambda^0 A\mathbf{v}_i \\
&= \lambda B\binom{i}{0}\lambda^{i+1}\mathbf{v}_0 + B\sum_{j=1}^{i}\binom{i}{j}\lambda^{i+1-j}\mathbf{v}_j + B\sum_{j=1}^{i}\binom{i}{j-1}\lambda^{i+1-j}\mathbf{v}_j + \binom{i}{i}\lambda^0 A\mathbf{v}_i \\
&= \lambda B\binom{i}{0}\lambda^{i+1}\mathbf{v}_0 + B\sum_{j=1}^{i}\binom{i}{j}\lambda^{i+1-j}\mathbf{v}_j + \sum_{j=1}^{i}\binom{i}{j-1}\lambda^{i+1-j}A\mathbf{v}_{j-1} + \binom{i}{i}\lambda^0 A\mathbf{v}_i \\
&= \lambda B\binom{i}{0}\lambda^{i+1}\mathbf{v}_0 + B\sum_{j=1}^{i}\binom{i}{j}\lambda^{i+1-j}\mathbf{v}_j + \sum_{j=0}^{i-1}\binom{i}{j}\lambda^{i-j}A\mathbf{v}_j + \binom{i}{i}\lambda^0 A\mathbf{v}_i \\
&= \lambda B\sum_{j=0}^{i}\binom{i}{j}\lambda^{i-j}\mathbf{v}_j + \sum_{j=0}^{i}\binom{i}{j}\lambda^{i-j}A\mathbf{v}_j \\
&= (A+\lambda B)\sum_{j=0}^{i}\binom{i}{j}\lambda^{i-j}\mathbf{v}_j \\
&= (A+\lambda B)\mathbf{u}_i
\end{aligned}
$$

This process can also be reversed using $\lambda' = -\lambda$. Hence, there's a one-to-one relation between the sequence $\mathbf{v}_i$ and $\mathbf{u}_i$.

Hence,

$$
\begin{aligned}
\mathcal{S}_m &= \{\mathbf{v}_m : \exists \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{m-1},\ \mathbf{v}_0 = \mathbf{0} \wedge \forall i \in [m],\ B\mathbf{v}_{i+1} = (A+\lambda B)\mathbf{v}_i\} \\
\mathcal{W}_m &= \{\mathbf{v}_0 : \exists \mathbf{v}_1, \ldots, \mathbf{v}_m,\ \forall i \in [m],\ B\mathbf{v}_{i+1} = (A+\lambda B)\mathbf{v}_i\}
\end{aligned}
$$

or equivalently,

$$
\begin{aligned}
\mathcal{S}_{m+1} &= B^{-1}((A+\lambda B)(\mathcal{S}_m)) \\
\mathcal{W}_{m+1} &= (A+\lambda B)^{-1}(B(\mathcal{S}_m))
\end{aligned}
$$

□

With that in mind, due to the regularity of the pair (A, B) We can choose some $\lambda \in \mathbb{C}$ s.t. $A + \lambda B$

is invertible and therefore we will consequently provide an explicit matrix X such that

$$\mathcal{S} = \ker X, \ \mathcal{W} = imX$$

Lemma 2.11. $\mathcal{S} \oplus \mathcal{W} = \mathcal{V}$

Proof. Note that since $A + Bx$ is regular, there exist $\lambda \in \mathbb{C}$ such that $A + \lambda B$ is non-singular.

Then, we can show

$$\begin{aligned}\mathcal{S}_m &= \ker((A + \lambda B)^{-1}B)^m \\ \mathcal{W}_m &= \mathrm{Im}((A + \lambda B)^{-1}B)^m\end{aligned}$$

by induction.

For the base case, note that

$$\begin{aligned}\mathcal{S}_0 &= \{\mathbf{0}\} = \ker((A + \lambda B)^{-1}B)^0 \\ \mathcal{W}_0 &= \mathcal{V} = \mathrm{Im}((A + \lambda B)^{-1}B)^0\end{aligned}$$

Suppose this is true for some $m \in \mathbb{N}$. We have

$$\begin{aligned}\mathcal{S}_{m+1} &= B^{-1}((A + \lambda B)(\mathcal{S}_m)) \\ &= \{\mathbf{v} \in \mathcal{V} : Bv \in (A + \lambda B)(\mathcal{S}_m)\} \\ &= \{\mathbf{v} \in \mathcal{V} : (A + \lambda B)^{-1}Bv \in \mathcal{S}_m\} \\ &= \{\mathbf{v} \in \mathcal{V} : ((A + \lambda B)^{-1}B)^{m+1}v = \mathbf{0}\} \\ &= \ker((A + \lambda B)^{-1}B)^{m+1}\end{aligned}$$

and

$$\begin{aligned}\mathcal{W}_{m+1} &= ((A + \lambda B)^{-1}(B(\mathcal{S}_m))) \\ &= ((A + \lambda B)^{-1}B)(\mathcal{S}_m) \\ &= ((A + \lambda B)^{-1}B)^{m+1}(\mathcal{V}) \\ &= \mathrm{Im}((A + \lambda B)^{-1}B)^{m+1}\end{aligned}$$

Therefore, we 1.13, we have

$$\mathcal{V} = \mathcal{S} \oplus \mathcal{W}$$

□

Having the decomposition $\mathcal{V} = \mathcal{S} \oplus \mathcal{W}$, we may construct a basis for $\mathcal{V}$ by uniting bases for $\mathcal{S}$ and $\mathcal{W}$,respectively.

The final form given at (27) indicates that we should be searching for an additional basis for $\mathcal{V}$. The following lemma indicates how we might do that.

Lemma 2.12. $A(\mathcal{S}) \oplus B(\mathcal{W}) = \mathcal{V}$.

y

Proof. First, we will show $A(\mathcal{S})$ and $B(\mathcal{W})$ are disjoint:

Suppose $\mathbf{v} \in A(\mathcal{S}) \cap B(\mathcal{W})$.

Then, there exists $\mathbf{u} \in \mathcal{S}$, $\mathbf{u}' \in \mathcal{W}$ s.t. $\mathbf{v} = A\mathbf{u}$, $\mathbf{v} = B\mathbf{u}'$.

Hence, $A\mathbf{u} = B\mathbf{u}'$, $\mathbf{u} \in A^{-1}(B\mathbf{u}') \subseteq \mathcal{W}$.

Since $\mathcal{S}$ and $\mathcal{W}$ are disjoint, $\mathbf{u} \in \mathcal{S} \cap \mathcal{W} = \{\mathbf{0}\}$.

$$\mathbf{v} = A\mathbf{u} = A\mathbf{0} = \mathbf{0}$$

This implies $A(\mathcal{S}) \cap B(\mathcal{W}) = \{\mathbf{0}\}$

Now, note that A is injective on $\mathcal{S}$ and B is injective on $\mathcal{W}$. Hence, we have

$$\dim(A(\mathcal{S}) \oplus B(\mathcal{W})) = \dim A(\mathcal{S}) + \dim B(\mathcal{W}) = \dim \mathcal{S} + \dim \mathcal{W} = n$$

So $A(\mathcal{S}) + B(\mathcal{W}) = \mathcal{V}$ □

2.6 Bases for the subspaces

Now that we have characterised the subspaces, we are ready to choose appropriate bases. Before we begin, note that the maps A and B are well defined even on the constrained subspaces.

$$B(\mathcal{S}) = B(\mathcal{S}_{\kappa+1}) = B(B^{-1}(A(\mathcal{S}_\kappa))) \subseteq A(\mathcal{S}_\kappa) = A(\mathcal{S})$$

and

$$A(\mathcal{W}) = A(\mathcal{W}_{\kappa'+1}) = A(A^{-1}(B(\mathcal{W}_{\kappa'}))) \subseteq B(\mathcal{W}_{\kappa'}) = B(\mathcal{W})$$

Hence, we can divide the domain into $\mathcal{S}$ and $\mathcal{W}$, and codomain into $A(\mathcal{S})$ and $B(\mathcal{W})$. For each of those divisions, we can choose a suitable basis for A and B

2.6.1 Basis for $\mathcal{S}$ and $A(\mathcal{S})$

First, we will only look at the domain of $\mathcal{S}$ and the codomain of $A(\mathcal{S})$.

Now, we will try to choose a basis for $\mathcal{S}$ and $A(\mathcal{S})$.

Note that after choosing a basis $\beta_{\mathcal{S}}$ for $\mathcal{S}$, the obvious choice of basis for $A(\mathcal{S})$ is $\gamma_{\mathcal{S}} = A(\beta_{\mathcal{S}})$. This way, the matrix representation of A on $\mathcal{S}$ will be the identity matrix I.

We would like to also have a nice matrix representation for B.

Note that since A is invertible and ${}_{\beta_{\mathcal{S}}}[A^{-1}]_{\gamma_{\mathcal{S}}} = I$,

$${}_{\gamma_{\mathcal{S}}}[B]_{\beta_{\mathcal{S}}} = {}_{\beta_{\mathcal{S}}}[A^{-1}]_{\gamma_{\mathcal{S}}}[B]_{\beta_{\mathcal{S}}} = {}_{\beta_{\mathcal{S}}}[A^{-1}B]_{\beta_{\mathcal{S}}}$$

Hence, we can choose $\beta_{\{}\mathcal{S}\}$ to be the basis where $A^{-1}B$ is the simplest. The obvious choice is the same method as we did jordan canonical form.

We claim $A^{-1}B$ is nilpotent.

Proof. Let $\mathbf{v}_k \in \mathcal{S} = \mathcal{S}_k$.

Define $\mathbf{v}_i = A^{-1}B\mathbf{v}_{i+1}$ for all $i \in [k]$. Note that

$$\begin{aligned} \mathbf{v}_i &= A^{-1}B\mathbf{v}_{i+1} \\ A\mathbf{v}_i &= B\mathbf{v}_{i+1} \end{aligned}$$

So $\forall i \in [k]$, $\mathbf{v}_i \in \mathcal{S}_i$. Notably, $\mathbf{v}_0 \in \mathcal{S}_0 = \{\mathbf{0}\}$.

Hence, $(A^{-1}B)^k\mathbf{v}_k = \mathbf{v}_0 = \mathbf{0}$.

Since this is true for any $\mathbf{v}_k \in \mathcal{S}$, $(A^{-1}B)^k = 0$. Hence, $A^{-1}B$ is nilpotent. □

Now, we can apply similar method as we did for Jordan canonical form. For the subspace $\mathcal{S}$ and map $A^{-1}B$, we can find a basis composed of null-chains. Let

$$\{\mathbf{v}_{1,1}, \mathbf{v}_{1,2}, \dots, \mathbf{v}_{1,m_1}\}, \{\mathbf{v}_{2,1}, \mathbf{v}_{2,2}, \dots, \mathbf{v}_{2,m_2}\}, \dots, \{\mathbf{v}_{t,1}, \mathbf{v}_{t,2}, \dots, \mathbf{v}_{t,m_t}\}$$

be such a basis. We have

$$\mathbf{v}_{i,j} = A^{-1}B\mathbf{v}_{i,j+1}$$

.

Let $\mathbf{u}_{i,j} = A\mathbf{v}_{i,j}$. $\mathbf{u}$ forms a basis for $A(\mathcal{S})$.

We have

$$B\mathbf{v}_{i,j+1} = AA^{-1}B\mathbf{v}_{i,j+1} = A\mathbf{v}_{i,j} = \mathbf{u}_{i,j}$$

Hence, B can be written in the form

$$
{}_{\gamma_{\mathcal{S}}}[B]_{\beta_{\mathcal{S}}} =
\begin{bmatrix}
\overbrace{\begin{matrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{matrix}}^{m_1} & & & \text{\Huge 0} \\
& \overbrace{\begin{matrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{matrix}}^{m_2} & & \\
& & \ddots & \\
\text{\Huge 0} & & & \overbrace{\begin{matrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{matrix}}^{m_t}
\end{bmatrix}
\tag{25}
$$

2.6.2 Basis for $\mathcal{W}$ and $B(\mathcal{W})$

Now we'll look at A and B as function from $\mathcal{W}$ to $B(\mathcal{W})$.

We can apply the same method as above. However, $B^{-1}A$ is no longer nilpotent.

However, we can use the conclusion from the proof for jordan canonical form.

For each eigenvalue of $B^{-1}A$, we can look at the generalized eigenspace of $B^{-1}A$. The direct sum of the eigenspaces is equal to $\mathcal{W}$.

For each eigenvalue λ_i, we can find a basis

$$\beta_{\mathcal{W}} = \{\mathbf{v}_{1,1}, \mathbf{v}_{1,2}, \dots, \mathbf{v}_{1,m_1}\}, \{\mathbf{v}_{2,1}, \mathbf{v}_{2,2}, \dots, \mathbf{v}_{2,m_2}\}, \dots, \{\mathbf{v}_{t,1}, \mathbf{v}_{t,2}, \dots, \mathbf{v}_{t,m_t}\}$$

where

$$\mathbf{v}_{i,j} = (B^{-1}A - \lambda_i I)\mathbf{v}_{i,j+1}$$

.

Let $\mathbf{u}_{i,j} = B\mathbf{v}_{i,j}$. $\mathbf{u}$ form a basis $\gamma_{\mathcal{W}}$ for $B(\mathcal{W})$.

We have

$$A\mathbf{v}_{i,j+1} = BB^{-1}A\mathbf{v}_{i,j+1} = B\mathbf{v}_{i,j} + \lambda_i B\mathbf{v}_{i,j+1} = \mathbf{u}_{i,j} + \lambda_i \mathbf{u}_{i,j+1}$$

$$ {}_{\gamma_{\mathcal{W}}}[A]_{\beta_{\mathcal{W}}} = \begin{bmatrix} \overbrace{\begin{matrix}\delta_0 & 1 & & \\ & \delta_0 & \ddots & \\ & & \ddots & 1 \\ & & & \delta_0\end{matrix}}^{k_0} & & & \text{\huge 0} \\ & \overbrace{\begin{matrix}\delta_1 & 1 & & \\ & \delta_1 & \ddots & \\ & & \ddots & 1 \\ & & & \delta_1\end{matrix}}^{k_1} & & \\ & & \ddots & \\ \text{\huge 0} & & & \overbrace{\begin{matrix}\delta_l & 1 & & \\ & \delta_l & \ddots & \\ & & \ddots & 1 \\ & & & \delta_l\end{matrix}}^{k_l} \end{bmatrix} \tag{26}$$

Therefore, choosing $\beta = \beta_{\mathcal{S}} \cup \beta_{\mathcal{W}}$ as basis for domain and $\gamma = \gamma_{\mathcal{S}} \cup \gamma_{\mathcal{W}}$ as basis for co-domain, We have

$$\begin{aligned} {}_{\gamma}[A]_{\beta} &= \begin{pmatrix} J & 0 \\ 0 & I \end{pmatrix}, \\ {}_{\gamma}[B]_{\beta} &= \begin{pmatrix} I & 0 \\ 0 & N \end{pmatrix}. \end{aligned} \tag{27}$$

3 Applications and Examples

As we have stated before, the main application of the Weierstrass canonical form is to help solve differential algebraic equations (DAEs) of the form $B\dot{x} = Ax$. Given that B may not be an invertible matrix, we must find another way to simplify such DAEs. This leads to a separate application of Weierstrass form, which will be useful in finding a solutions : the Drazin inverse of a matrix.

3.1 The Drazin inverses

Definition: For $A \in \mathbb{R}^{n\times n}$, there exists a *drazin inverse* $A^D \in \mathbb{R}^{n\times n}$ such that it has these properties:

- $AA^D = A^D A$
- $A^D A A^D = A^D$
- $\exists\ k \in \mathbb{N}$ such that $A^D A^{k+1} = A^k$

The existence of such an inverse for any matrix and its uniqueness can be proven and it is shown in *Theorem 2.19* in [6]
From [2]: Given a regular matrix pencil $A - B\partial$, with A and B such that $AB = BA$, then the drazin inverses of A and B are as follows:

$$B^D = [W\ S]\begin{bmatrix} I_{n_1} & 0 \\ 0 & 0 \end{bmatrix}[BW\ AS]^{-1} \text{ and } A^D = [S\ W]\begin{bmatrix} J^D & 0 \\ 0 & I_{n_2} \end{bmatrix}[BW\ AS]^{-1}$$

The proof that E^D and A^D follow the three properties of the drazin inverses appears in [2] at *Proposition 2.15.* The forms the drazin inverses take are very similar to applying the Weierstrass canonical form transformation on each of B and A.

3.2 Solving the DAE

3.2.1 Using the drazin inverse

Using the generalized inverses that we have found in the previous section, we manage to find the general solution to the initial value problem $B\dot{x} = Ax$; $x(0) = x_0$, for the case where A and B commute. It takes the form:

$$x(t) = e^{B^D A t} B^D B x_0$$

It follows from *Remark 2.21* in [2], but the system $B\dot{x} = Ax \; + \; f$ is replaced by a simpler version where $f = 0$.

This way of finding the solution of the DAE is long and computational (given the fact that B^D must be computed), with the additional restriction that $AB = BA$. A better way to reach the solution would be through the use of the generalized eigenvectors.

3.2.2 Using the generalized eigenvectors

As in section *3.2 Differential algebraic equation - revisited* of [2], the generalized eigenvectors of the regular pencil $A \; - B\partial$ constitute a basis of solutions for $\ker(A - B\frac{d}{dt})$ (i.e. the space of solutions for the system of differential equations $B\dot{x} = Ax$). The proposition in [2] goes as follows:
Consider a regular pencil $A \; - B\partial \in \mathbb{C}^{n\times n}[\partial]$. Then $(v_1, v_2, ..., v_k)$ for λ a generalized eigenvalue of the pencil is a *chain of generalized eigenvectors* if and only if the following functions are linearly independent solutions for $B\dot{x} = Ax$:

$$x_i : \mathbb{R} \to \mathbb{C}^n \quad x_i(t) = e^{\lambda t} \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \begin{bmatrix} \frac{t^{i-1}}{(i-1)!} \\ \vdots \\ t \\ 1 \end{bmatrix}, \text{ for } i \in \{1, 2, \dots, k\}$$

3.2.3 By directly applying the transformation

If we apply the transformation matrices that turn a pencil into the Weierstrass canonical form on the DAE, we manage to directly decouple the system into two independent part as for example [5]. We can deal with each of them separately since they each represent the two subspaces that partition the vectorspace. Consider the homogenous DAE

$$B\dot{x} = Ax \tag{28}$$

Let P and Q be square matrices such that

$$PAQ = \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix} \quad \text{and} \quad PBQ = \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix}$$

Then some algebra yields

$$\begin{aligned} & B\dot{x} = Ax \\ & P \cdot | \quad B\dot{x} = Ax \\ & PB\dot{x} = PAx \end{aligned}$$

Now we want A and B to turn into the WCF, so we are missing Q on the right side of each. We define $\tilde{x} := Q^{-1}x$ as our new variable. The system becomes:

$$PBQ\dot{\tilde{x}} = PAQ\tilde{x} \iff \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix} \dot{\tilde{x}} = \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix} \tilde{x}$$

If we let $\tilde{x} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}$ the system above can be expressed as the following

$$\begin{cases} \dot{\tilde{x}}_1 = J\tilde{x}_1 \\ N\dot{\tilde{x}}_2 = \tilde{x}_2 \end{cases}$$

The first equation is a homogeneous linear system so we may recall its solution is of the form $\tilde{x}_1 = Qce^{Jt}$ for some constant $c \in \mathbb{C}$.[3] And due to the structure of $N \in \mathbb{C}^{m\times m}$ shown in Section **??** for we get $\tilde{x}_{2,1} = \tilde{x}_{2,m} = 0$ and $\tilde{x}_{2,i} = \dot{\tilde{x}}_{2,i}$ if and only if $N_{i+1,i} \neq 0$ and zero otherwise. Lastly we apply Q to retrieve x.

Hence we have fully characterised all solutions of (28).

3.3 Example: Computing the Weierstrass Canonical Form

We are trying to find two square matrices P and Q such that $PAQ = \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix}$ and $PBQ = \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix}$ where $A = \begin{bmatrix} 3 & 0 \\ -1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 1 \\ -4 & -2 \end{bmatrix}$.

For the Weierstrass form to exist as we have defined in the proofs above, we need to make sure that the pencil $A - B\partial$ is regular. The determinant of $A - \lambda B$ for $\lambda := 1$ is 6 which is nonzero, so the pencil is indeed regular.

$$\det(A - \lambda B) = \begin{vmatrix} 3 - 2\lambda & -\lambda \\ -1 + 4\lambda & 1 + 2\lambda \end{vmatrix} = 3\lambda + 3$$

The polynomial for the pencil has degree only one, when it should have degree two. This tells us that there must be a hidden root: an infinite eigenvalue. The eigenvalues with which we are working are $\lambda_1 = -1$ and λ_∞. As in the proof, we split the basis of $\mathbb{R}^2$ into two subspaces: one that deals with the finite eigenvalues and another that tackles the infinite ones.

[3] c Is determined by initial conditions of the system.

First, we look for an eigenvector for λ_1. As defined in (**??**), we look for $v_1 \in \mathbb{R}^2$ s.t. $Av_1 = \lambda_1 B v_1$. We get $v_1 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$ as the first column for the matrix Q.

Next, using (**??**), we get $Bv_2 = 0$, for the first vector in the chain that is contained in the "infinite" vector subspace. Solving it, we get $v_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

Collecting the two vectors, we get $Q = \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -5 & 2 \end{bmatrix}$

Now we need to find $P = \begin{bmatrix} Bv_1 & Av_2 \end{bmatrix}^{-1}$, so we can easily reach $P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} \end{bmatrix}$

To check that everything is correct:

$$PAQ = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -5 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\implies PAQ = \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix}, \text{ where } J = \lambda_1 \text{ and } I \text{ is one-dimensional.}$$

and

$$PBQ = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -4 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -5 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\implies PBQ = \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix}, \text{ where } N = 0 \text{ since it is a nilpotent Jordan matrix of dimension one.}$$

Conclusion

In summary, whilst proving the JCF, many interesting characteristics of linear maps emerge and are put to use. This is seen in how the eigenspaces divide the vector space into manageable subspaces and how the map remains invariant over these spaces.

During the process we realized that the vector space can be partitioned into subspaces that are built from the eigenspaces. Iterating maps were the key to expanding the eigenspaces, and proved to be a more useful concept than initially thought. The idea of iterating the maps lead us to the definition of the null chains. Because of their many properties, they play a major role as a basis of the vectorspace.

After concluding the JCF proof, we turned our attention to the more general problem of proving the WCF. We needed new definitions for eigenvalues and eigenvectors that comply to our new algebraic structure: matrix pencils. In defining the generalized eigenvalues, we found that some of them do not appear in the polynomial characterized by the determinant of the matrix pencil. We call them infinite eigenvalues, as they are the zero eigenvalue of the reverse matrix pair. In order to account for both types of eigenvalues, we looked for two bases for the vectorspace, each divided into two parts. We pick the bases as a collection of chains again, with a definition that incorporates the behaviour of both linear transformations. Using the two bases we can create their corresponding transformation matrices which when applied to the matrix pair puts it into WCF. Geometrically these transformations are independent changes of basis for the domain and codomain of the matrices that define the pencil.

Clearly some ideas pervade both these proofs. The theme of vector-chains and decomposition into invariant subspaces reoccur in central roles. In fact the WCF can be proven without the use of the JCF, in which case the JCF follows immediately as a corollary for the case where the second matrix in the pair is the identity.

The importance and interconnection of both the WCF and JCF should now be evident.

References

[1] Sheldon Axler. *Linear algebra done right.* Springer, 2024.

[2] Thomas Berger, Achim Ilchmann, and Stephan Trenn. The quasi-weierstraß form for regular matrix pencils. *Linear Algebra and its Applications*, 436(10):4052–4069, 2012.

[3] FR Gantmacher. *The Theory of Matrices, Volume 2*, volume 133. American Mathematical Society, 2026.

[4] Paul Richard Halmos, Paul Richard Halmos, Paul Richard Halmos, Hongrie Mathématicien, Paul Richard Halmos, and Hungary Mathematician. *Finite-dimensional vector spaces*, volume 11. Springer, 1958.

[5] Jan Heiland. Linear daes with constant coefficients, 2026. `https://www.janheiland.de/script-daes/III.html` [Accessed: 2026-06-19].

[6] Peter Kunkel and Volker Mehrmann. *Differential-Algebraic Equations: Analysis and Numerical Solution.* EMS Publishing House, Zürich, Switzerland, 2006.