Jordan Normal Form

Chapter 1

Jordan Normal Form

1.1 The main theorem

Let A be a $n \times n$ matrix. There exists an invertible matrix S such that

$$S^{-1}AS = J \tag{1.1}$$

where J is a block matrix,

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_m \end{pmatrix} \tag{1.2}$$

and each block J_i is a square matrix of the form

$$J_j = \lambda_j \mathbb{1} + N = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix} \tag{1.3}$$

where λ is an eigenvalue of A.

In the proof it is convenient to work with linear operators on vector spaces rather their matrix representations in some basis. Throughout the rest of the notes we shall therefore assume that V is an $n-$dimensional complex vector space and that A is a linear operator on V.

The kernel (or null space) of A is defined by

$$\ker A = \{x \in V : A\mathbf{x} = 0\} \tag{1.4}$$

and the range

$$\text{range } A = \{A\mathbf{x} :\in V\} \tag{1.5}$$

The kernel and the range are both linear subspaces of V and the dimension theorem says that

$$\dim \ker A + \dim \text{range } A = n \tag{1.6}$$

The subspace $\ker(A - \lambda \mathbb{1})$ of V, that is, the subspace spanned by the eigenvectors belonging to λ, is called the eigenspace corresponding to λ. The number $\dim \ker(A - \lambda \mathbb{1})$ is called the geometric multiplicity of λ. Recall that $\lambda \in \mathbb{C}$ is an eigenvalue if and only if it is a root of the characteristic polynomial

$$p(z) = (-1)^n (z - \lambda_1)^{a_1} (z - \lambda_2)^{a_2} \cdots (z - \lambda_1)^{a_k} \tag{1.7}$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of A. The positive integer a_j is called the algebraic multiplicity of the eigenvalue λ_j. The corresponding geometric multiplicity will be denoted g_j.

1.2 Decomposition into invariant subspaces

We begin with some definitions. Let $V_1, \ldots, V_k$ be subspaces of V. We say that V is the direct sum of $V_1, \ldots, V_k$ if each vector $\mathbf{x} \in V$ can be written in a unique way as

$$\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k, \quad \text{where} \quad \mathbf{x}_j \in V_j,\ j = 1, \ldots, k. \tag{1.8}$$

We use the notation

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k. \tag{1.9}$$

We say that a subspace W of V is invariant under A if

$$\mathbf{x} \in W \quad \text{implies} \quad A\mathbf{x} \in W. \tag{1.10}$$

Example: Suppose that A has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. It then follows that the vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ are linearly independent and thus form a basis for V. Let

$$\ker(A - \lambda_k \mathbb{1}) = \{z\mathbf{u}_k : z \in \mathbb{C}\}, \quad k = 1, \ldots, n,$$

be the corresponding eigenspaces. Each eigenspace is invariant under A since

$$(A - \lambda_k \mathbb{1})\mathbf{u} = \mathbf{0} \quad \text{implies} \quad (A - \lambda_k \mathbb{1})A\mathbf{u} = A(A - \lambda_k \mathbb{1})\mathbf{u} = \mathbf{0}.$$

Moreover,

$$V = (A - \lambda_1 \mathbb{1}) \oplus (A - \lambda_2 \mathbb{1}) \oplus \cdots \oplus (A - \lambda_n \mathbb{1})$$

by the definition of a basis.

□

More generally, suppose that A has k distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ and that the geometric multiplicity g_j of each λ_j equals the algebraic multiplicity a_j. Let $\ker(A - \lambda_j \mathbb{1})$, $j = 1, \ldots, k$, be the corresponding eigenspaces. We can then find a basis for each eigenspace consisting of g_j eigenvectors. The union of these bases consists of $g_1 + \cdots + g_k = a_1 + \cdots + a_k = n$ elements and is linearly independent, since eigenvectors belonging to different eigenvalues are linearly independent. We thus obtain a basis for V and it follows that

$$V = (A - \lambda_1 \mathbb{1}) \oplus (A - \lambda_2 \mathbb{1}) \oplus \cdots \oplus (A - \lambda_k \mathbb{1}).$$

In this basis, A has the matrix

$$D = \begin{pmatrix} \lambda_1 \mathbb{1}_1 & & \\ & \ddots & \\ & & \lambda_k \mathbb{1}_k \end{pmatrix}$$

where $\mathbb{1}_j$ is the $g_j \times g_j$ unit matrix. In other words, D is a diagonal matrix with the eigenvalues on the diagonal, each repeated g_j times. One says that A is diagonalised in the new basis.

□

Not all matrices can be diagonalised.

1.3 Minimal polynomial

There exists a non-zero polynomial p such that $p(\mathbf{A}) = \mathbf{0}$.

Proof:

Here it is convenient to identify A with its matrix in some basis. Note that $\mathbb{C}^{n\times n}$ is an n^2-dimensional vector space. It follows that the n^2+1 matrices $\mathbb{1}, A, A^2, \ldots, A^{n^2}$ are linearly dependent. But this means that there exists numbers $\alpha_0, \ldots, \alpha_{n^2}$, not all zero, such that

$$\alpha_{n^2} A^{n^2} + \alpha_{n^2-1} A^{n^2-1} + \cdots + \alpha_1 A + \alpha_0 \mathbb{1} = 0,$$

that is, $p(A) = 0$, where $p(z) = \alpha_{n^2} z^{n^2} + \alpha_{n^2-1} z^{n^2-1} + \cdots + \alpha_1 z + \alpha_0$.

□

Let $p_{min}(z)$ be a monic polynomial (a monic polynomial is a polynomial with leading coefficient 1) of minimal degree such that $p_{min}(A) = 0$. If $p(z)$ is any polynomial such that $p(A) = 0$ it follows that $p(z) = q(z)p_{min}(z)$ for some polynomial q. To see this, use the division alorithm on p and p_{min}:

$$p(z) = q(z)p_{min}(z) + r(z), \quad \text{where } r = 0 \text{ or } \deg(r) < \deg(p_{min}).$$

Thus $r(A) = p(A) - q(A)p_{min}(A) = 0$. But this implies that $r(z) = 0$, since p_{min} has minimal degree. This also shows that the polynomial p_{min} is unique.

□

It is called the minimal poynomial for A.

By the fundamental theorem of algebra, we can write the minimal polynomial as a product

$$p_{min}(z) = (z-\lambda_1)^{m_1}(z-\lambda_2)^{m_2}\cdots(z-\lambda_k)^{m_k} \tag{1.11}$$

where the numbers λ_j and distinct and $m_j \geq 1$. It will be shown later that the roots λ_j coincide with the eigenvalues of A.

Suppose that $p(z) = p_1(z)p_2(z)$ where p_1 and p_2 are relatively prime. If $p(A) = \mathbf{0}$ we have that

$$V = \ker p_1(A) \oplus \ker p_2(A)$$

and each subspace $\ker p_j(A)$ is invariant under A.

Proof:

The invariance follows from $p_j(A)A\mathbf{x} = Ap_j(A)\mathbf{x} = \mathbf{0}$ for $\mathbf{x} \in \ker p_j(A)$. Since p_1 and p_2 are relatively prime it follows from Bezout's identity (see appendix) that there exists polynmials q_1, q_2 such that

$$p_1(z)q_1(z) + p_2(z)q_2(z) = 1.$$

Thus

$$p_1(A)q_1(A) + p_2(A)q_2(A) = I.$$

Applying this identity to the vector $\mathbf{x} \in V$, we obtain

$$\mathbf{x} = \underbrace{p_1(A)q_1(A)\mathbf{x}}_{\mathbf{x}_2} + \underbrace{p_2(A)q_2(A)\mathbf{x}}_{\mathbf{x}_1},$$

where

$$p_2(A)\mathbf{x}_2 = p_2(A)p_1(A)q_1(A)\mathbf{x} = p(A)q_1(A)\mathbf{x} = \mathbf{0},$$

so that $\mathbf{x}_2 \in \ker p_2(A)$. Similarly, $\mathbf{x}_1 \in p_1(A)q_1(A)$. Thus $V = \ker p_1(A) + \ker p_2(A)$.

Next we prove the uniqueness of $\mathbf{x}_1$ and $\mathbf{x}_2$. If

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 = \mathbf{x}_1' + \mathbf{x}_2', \quad \mathbf{x}_j, \mathbf{x}_j' \in \ker p_j(A),\ j = 1, 2,$$

we obtain that

$$\mathbf{u} = \mathbf{x}_1 - \mathbf{x}_1' = \mathbf{x}_2' - \mathbf{x}_2 \in \ker p_1(A) \cap \ker p_2(A),$$

so that

$$\mathbf{u} = p_1(A)q_1(A)\mathbf{u} + p_2(A)q_2(A)\mathbf{u} = \mathbf{0}.$$

$\square$

With $\lambda_1, \ldots, \lambda_k$ and $m_1, \ldots, m_k$ as in (1.11) we have

$$V = \ker(A - \lambda_1 I)^{m_1} \oplus \cdots \oplus \ker(A - \lambda_k I)^{m_k},$$

where each $\ker(A - \lambda_k I)^{m_j}$ is invariant under A. The numbers $\lambda_1, \ldots, \lambda_k$ are eigenvalues of A.

Proof:

We begin by noting that the polynomials $(z - \lambda_j)^{m_j}$, $j = 1, \ldots, k$, are relatively prime. Repeated of the previous result shows that

$$V = \ker(A - \lambda_1 I)^{m_1} \oplus \cdots \oplus \ker(A - \lambda_k I)^{m_k},$$

with each $\ker(A - \lambda_j I)^{m_j}$ invariant.

Consider the restriction of linear operator A to the subspace $\ker(A - \lambda_j I)^{m_j}$. Obvious, $\ker(A - \lambda_j I)^{m_j} \neq \{\mathbf{0}\}$, for otherwise p_{min} would not be minimal. Since every linear operator on a (non-trivial) finite dimensional complex vector space has an eigenvalue, it follows that there is some non-zero element $\mathbf{u} \in \ker(A - \lambda_j I)^{m_j}$ with $A\mathbf{u} = \lambda \mathbf{u}$, $\lambda \in \mathbb{C}$. But then

$$\mathbf{0} = (A - \lambda_j I)^{m_j} \mathbf{u} = (\lambda - \lambda_j)^{m_j} \mathbf{u},$$

so $\lambda = \lambda_j$. This proves that the roots λ_j of the minima; polynomial are eigenvalues of A. On the other hand if $\mathbf{u}$ is an eigenvector of A corresponding to the eigenvalue λ, we have

$$\mathbf{0} = p_{min}(A)\mathbf{u} = (A - \lambda_1 I)^{m_1} \cdots (A - \lambda_k I)^{m_k} \mathbf{u} = (\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_k)^{m_k} \mathbf{u},$$

so $\lambda = \lambda_j$ for some j, that is, every eigenvalue is a root of the minimal polynomial.

$\square$

The subspace $\ker(A - \lambda_j I)^{m_j}$ is called the generalised eigenspace corresponding to λ_j and a non-zero vector $\mathbf{x} \in \ker(A - \lambda_j I)^{m_j}$ is called a generalised eigenvector. The number m_j is the smallest exponent m such that $(A - \lambda_j I)^m$ vanishes on $\ker(A - \lambda_j I)^{m_j}$. Writing $\mathbf{x} \in V$ as $\mathbf{x} = \mathbf{x}_1 + \tilde{\mathbf{x}}$ according to the decomposition

$$V = (A - \lambda_1 I)^{m_1} \oplus \ker \tilde{p}(A),$$

where $\tilde{p}(z) = (z - \lambda_2)^{m_2} \cdots (z - \lambda_k)^{m_k}$, we would then obtain that

$$(A - \lambda_1 I)^{m_1 - 1} \tilde{p}(A)\mathbf{x} = \tilde{p}(A)(A - \lambda_1 I)^{m_1 - 1}\mathbf{x}_1 + (A - \lambda_1 I)^{m_1 - 1}\tilde{p}(A)\tilde{\mathbf{x}} = \mathbf{0},$$

contradicting the definition of the minimal polynomial.

□

If we can select a basis $\{\mathbf{u}_{j,1}, \dots, \mathbf{u}_{j,n_j}\}$ for each generalized eigenspace, then the union

$$\{\mathbf{u}_{1,1}, \dots, \mathbf{u}_{1,n_1}, \mathbf{u}_{2,1}, \dots, \mathbf{u}_{2,n_2}, \dots, \mathbf{u}_{k,n_k}, \mathbf{u}_{k,1}, \dots, \mathbf{u}_{k,n_k}\}$$

will be a basis for V. Since each generalized eigenspace is invariant under the linear operator A, the matrix for A in this basis will have the block form

$$\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_k \end{pmatrix},$$

where each A_j is a $n_j \times n_j$ square matrix, as

$$\begin{aligned} & A\{\mathbf{u}_{1,1}, \dots, \mathbf{u}_{1,n_1}, \dots, \mathbf{u}_{1,1}, \dots, \mathbf{u}_{k,n_k}\} \\ &= \{\mathbf{u}_{1,1}, \dots, \mathbf{u}_{1,n_1}, \dots, \mathbf{u}_{1,1}, \dots, \mathbf{u}_{k,n_k}\} \begin{pmatrix} & \vdots & & & & \\ \cdots & A_1 & \cdots & & & \\ & \vdots & & & & \\ & & & \ddots & & \\ & & & & \vdots & \\ & & & \cdots & A_k & \cdots \\ & & & & \vdots & \end{pmatrix}, \end{aligned}$$

What remains in order to prove the main theorem is to show that we can select a basis for each generalised eigenspace so that each block A_j takes the form (1.3).

1.4 Proof of the main theorem

Denote the restriction of A to a generalised eigenspace $G(\lambda_j, A) = \ker(A - \lambda_j I)^{m_j}$ by $A_{|G(\lambda_j,A)}$. We have that $A_{|G(\lambda_j,A)}$ has only one eigenvalue. Set $N = A_{|G(\lambda_j,A)} - \lambda_j I$ and then $_j$ is the smallest integer such that $N^{m_j} = 0$. A linear operator N with the property that $N^m = \mathbf{0}$ for some m is called nilpotent.

There is some vector $\mathbf{u} \in \ker(A - \lambda_j I)^{m_j}$ such that $N^{m_j-1}\mathbf{u} \neq \mathbf{0}$. It follows that the vectors $\mathbf{u}, N\mathbf{u}, \dots, N^{m_j-1}\mathbf{u}$ are linearly independent. Indeed, suppose that

$$\alpha_1 \mathbf{u} + \alpha_2 N\mathbf{u} + \cdots + \alpha_n N^{m_j-1}\mathbf{u} = \mathbf{0}.$$

Applying N^{m_j-1} to both sides of this equation we obtain that $\alpha_1 N^{m_j-1}\mathbf{u} = \mathbf{0}$ implies $\alpha_1 = 0$. Then applying N^{m_j-2} to

$$\alpha_2 N\mathbf{u} + \cdots + \alpha_n N^{m_j-1}\mathbf{u} = \mathbf{0}$$

implies that $\alpha_2 = 0$. Proceeding this way we obtain $\alpha_j = 0$ for each j. Thus $\{N^{m_j-1}\mathbf{u}, \ldots, N\mathbf{u}, \mathbf{u}\}$ is a basis for $\ker(A - \lambda_j I)^{m_j}$. The matrix for N in this basis is

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

as

$$\begin{aligned} N\{N^{m_j-1}\mathbf{u}, N^{m_j-2}\mathbf{u}, \ldots, N\mathbf{u}, \mathbf{u}\} &= \{\mathbf{0}, N^{n-1}\mathbf{u}, \ldots, N^2\mathbf{u}, N\mathbf{u}\} \\ &= \{N^{m_j-1}\mathbf{u}, N^{m_j-2}\mathbf{u}, \ldots, N\mathbf{u}, \mathbf{u}\} \begin{pmatrix} 0 & 1 & & & 0 \\ 0 & 0 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & 0 & & & 0 \end{pmatrix}. \end{aligned}$$

In general, a set of non-zero vectors $\mathbf{u}, \ldots, N^{l-1}\mathbf{u}$ with $N^l\mathbf{u} = \mathbf{0}$ is called a Jordan chain. We prove the theorem in general by showing that there is a basis for V consisting of Jordan chains.

Suppose N is nilpotent. Then there exists vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$ and non-negative integers $m_1, \ldots, m_k$ such that

(i) $N^{m_1}\mathbf{u}_1, \ldots, N\mathbf{u}_1, \mathbf{u}_1, \ldots, N^{m_k}\mathbf{u}_k, \ldots, N\mathbf{u}_k, \mathbf{u}_k$ is a basis for V;

(ii) $N^{m_1+1}\mathbf{u}_1 = \cdots = N^{m_k+1}\mathbf{u}_k = \mathbf{0}$.

Proof:

We prove the theorem by induction on the dimension of V. Clearly the theorem holds when the dimension is 1 (in that case, the only nilpotent operator is the 0 operator, take $\mathbf{u}_1$ to be any non-zero vector and $m_1 = 0$).

Suppose now that the theorem holds for all complex vector spaces of dimension less than n, where $n \geq 2$, and assume that $\dim V = n$. Since N is nilpotent it is not one-to-one and therefore $\dim \ \text{range}\ N < n$. By the inductive hypothesis, we can therefore find a basis of Jordan chains

$$N^{m_i-1}\mathbf{u}_i, \ldots, N\mathbf{u}_i, \mathbf{u}_i, \quad i = 1, \ldots, k, \tag{1.12}$$

for range N. For each $\mathbf{u}_i$ we can find a $\mathbf{v}_i \in V$ such that $N\mathbf{v}_i = \mathbf{u}_i$ (since $\mathbf{u}_i \in$ range N). That is, each Jordan chain in the basis for range N can be extended by one element. We claim that the vectors

$$N^{m_i}\mathbf{v}_i, \ldots, N\mathbf{v}_i, \mathbf{v}_i, \quad i = 1, \ldots, k, \tag{1.13}$$

are linearly independent. Indeed, suppose that

$$\sum_{i=1}^{k}\sum_{j=0}^{m_i} \alpha_{i,j} N^j \mathbf{v}_i = \mathbf{0}. \tag{1.14}$$

Applying N to both sides of this, we find that

$$\sum_{i=1}^{k}\sum_{j=0}^{m_i-1} \alpha_{i,j} N^j \mathbf{u}_i = \sum_{i=1}^{k}\sum_{j=0}^{m_i} \alpha_{i,j} N^{j+1} \mathbf{u}_i = \mathbf{0}, \tag{1.15}$$

which, by hypothesis implies that $\alpha_{i,j} = 0$ for $1 \leq i \leq k$, $0 \leq j \leq l_i - 1$. Substituting this into (1.14) gives that

$$\sum_{i=1}^{k} \alpha_{i,m_i} N^{l_i} \mathbf{u}_i = \sum_{i=1}^{k} \alpha_{i,l_i} N^{m_i} \mathbf{u}_i = \mathbf{0}, \tag{1.16}$$

which again implies that $\alpha_{i,m_i} = 0$ for $1 \leq i \leq k$, by our inductive hypothesis.

Extend the vectors in (1.13) to a basis for V by adding vectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$:

$$N^{m_1}\mathbf{v}_1, \ldots, N\mathbf{v}_1, \mathbf{v}_1, \ldots, N^{m_k}\mathbf{v}_k, \ldots, N\mathbf{v}_k, \mathbf{v}_k, \mathbf{w}_1, \ldots, \mathbf{w}_p, \tag{1.17}$$

Each $N\mathbf{w}_j$ is in the range of N and hence we can find an element $\mathbf{x}_j$ in the span of (1.13) such that $N\mathbf{w}_j = N\mathbf{x}_j$. Now let

$$\mathbf{v}_{n+j} = \mathbf{w}_j - \mathbf{x}_j$$

Then $N\mathbf{v}_{n+j} = \mathbf{0}$. Furthermore

$$N^{m_1-1}\mathbf{v}_1, \ldots, N\mathbf{v}_1, \mathbf{v}_1, \ldots, N^{m_k-1}\mathbf{v}_k, \ldots, N\mathbf{v}_k, \mathbf{v}_k, \mathbf{v}_{n+1}, \ldots, \mathbf{v}_{n+p} \tag{1.18}$$

spans V because its span contains each $\mathbf{x}_j$ and each $\mathbf{v}_{n+j}$ and hence $\mathbf{w}_j$ (and because (1.17) spans V). The elements $\mathbf{v}_{n+j}$ are chains of length 1.

$\square$

Let A be a $n \times n$ matrix. There exists an invertible matrix S such that

$$S^{-1}AS = J \tag{1.19}$$

where J is a block matrix,

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_m \end{pmatrix} \tag{1.20}$$

and each block J_i is a square matrix of the form

$$J_j = \lambda_j \mathbb{1} + N = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix} \tag{1.21}$$

The λ_j need not be distinct. Also, J_j may be a 1×1 matrix (λ_j) containing just the eigenvalue of A.

1.5 Uniqueness of the Jordan normal form up to permutation of Jordan blocks

The matrix J is not completely unique, since in that case we can change the order of the Jordan blocks, this is the only thing which is unique. In other words, the number of blocks and their sizes are uniquely determined.

Let β be the total number of blocks and $\beta(k)$ the number of blocks of size $\beta \times \beta$.

$$\begin{aligned}
\dim\ker N &= \beta, \\
\dim\ker N^2 &= \dim\ker N + \beta - \beta(1), \\
&\vdots \\
\dim\ker N^{k+1} &= \dim\ker N^k + \beta - \beta(1) - \cdots - \beta(k).
\end{aligned}$$

$\square$

1.6 Jordan Form for real matrices

1.6.1 Jordan Form for real matrices with compex eigenvalues

Example: Two complex and on real eigenvalue

Consider

$$A\{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,1}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} = \{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,1}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} \begin{pmatrix} \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & \\ 0 & 0 & \overline{\lambda}_1 & 1 & 0 \\ 0 & 0 & 0 & \overline{\lambda}_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \tag{1.22}$$

First define the matrix Q:

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{with inverse} \quad Q^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.23}$$

This simply permutes the basis $\{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,1}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\}$ to the basis $\{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,1}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\}$.

Employing these matrices in (1.22)

$$
\begin{aligned}
&A\{\mathbf{u}_{1,1}, \overline{\mathbf{u}}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} \\
= \ &\{\mathbf{u}_{1,1}, \overline{\mathbf{u}}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & \overline{\lambda}_1 & 1 & 0 \\ 0 & 0 & 0 & \overline{\lambda}_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\
= \ &\{\mathbf{u}_{1,1}, \overline{\mathbf{u}}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} \begin{pmatrix} \lambda_1 & 0 & 1 & 0 & 0 \\ 0 & \overline{\lambda}_1 & 0 & 1 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \overline{\lambda}_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix}
\end{aligned} \tag{1.24}
$$

Define the matrix X:

$$
X = \begin{pmatrix} \frac{1}{2} & -i\frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & i\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -i\frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & i\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{with inverse} \quad X^{-1} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ i & -i & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & i & -i & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.25}
$$

This takes us from the basis

$$
\{\mathbf{u}_{1,1}, \overline{\mathbf{u}}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\}
$$

to the basis

$$
\{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\}
$$

where:

$$
\Re(\mathbf{u}_{1,1}) = \frac{1}{2}(\mathbf{u}_{1,1} + \overline{\mathbf{u}}_{1,1}), \quad \Im(\mathbf{u}_{1,1}) = \frac{1}{2i}(\mathbf{u}_{1,1} - \overline{\mathbf{u}}_{1,1}), \quad \text{etc.} \tag{1.26}
$$

Emplying this in (1.24)

$$
\begin{aligned}
&A\{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\} \\
&= \{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\} \times \\
&\times \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ i & -i & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & i & -i & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 1 & 0 & 0 \\ 0 & \overline{\lambda}_1 & 0 & 1 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \overline{\lambda}_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} 1 & -i & 0 & 0 & 0 \\ 1 & i & 0 & 0 & 0 \\ 0 & 0 & 1 & -i & 0 \\ 0 & 0 & 1 & i & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\
&\{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\} \begin{pmatrix} a_1 & b_1 & 1 & 0 & 0 \\ -b_1 & a_1 & 0 & 1 & 0 \\ 0 & 0 & a_1 & b_1 & 0 \\ 0 & 0 & -b_1 & a_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \\
&\{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\} B
\end{aligned} \tag{1.27}
$$

where $\lambda_1 = a_1 + ib_1$.

General case

Thus easily generalises. Let

$$
\Lambda = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathcal{O} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{1.28}
$$

The Jordan block matrix of dimension $2m \times 2m$ is given by the formula

$$
B = \begin{pmatrix} \Lambda & \mathcal{I} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \Lambda & \mathcal{I} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \Lambda \end{pmatrix} \tag{1.29}
$$

The matrices Q and X are easily generalised. We go from the equation

$$
S^{-1}AS = J
$$

to

$$
(SQX)^{-1}A(SQX) = \mathcal{P}^{-1}A\mathcal{P} = B. \tag{1.30}
$$

Appendix A

Dividing polynomials and Bezout's identity

A.1 Dividing polynomials

We divide the polynomial $p(x) = x^3 - 4x^2 + 2x - 3$ by $q(x) = x + 2$. We take the obvious steps:

(1a) $x^3 - 4x^2 + 2x - 3 = (x^3 + 2x^2) - 6x^2 + 2x - 3$
(1b) $x^3 - 4x^2 + 2x - 3 = x^2(x + 2) - 6x^2 + 2x - 3$
(2a) $-6x^2 + 2x - 3 = (-6x^2 - 12x) + 14x - 3$
(2b) $-6x^2 + 2x - 3 = (-6x)(x + 2) + 14x - 3$
(3a) $14x - 3 = 14x + 28 - 31$
(3b) $14x - 3 = 14(x + 2) - 31$

So that

$$\frac{x^3 - 4x^2 + 2x - 3}{x + 2} = x^2 - 6x + 14 - \frac{31}{x + 2}.$$

It should be obvious that if we divide $p(x)$ by $m(x)$

$$p(x) = m(x)q(x) + r(x) \qquad (0 \leq \deg r < \deg m)$$

$\square$

A.2 Bezout's identity

(Bezout's identity). Let f, g be polynomials. Then there exists a greatest common divisor which is unique up to multiplication by constants. Moreover, for the greatest common divisor h, there exists polynomials p and q such that

$$pf + qg = h.$$

Proof:

Consider the set $S = \{pf + qg : p, q \text{ polynomials}\}$. Let $h = p_h f + q_h g$ be a non-zero polynomial in S with minimal degree. We claim that h divides any polynomial in S.

Suppose not. Then there exists some p_d and q_d such that

$$p_d f + q_d g = sh + r$$

for some polynomials s and r, $r \neq 0$ and $\deg(r) < \deg(h)$. But then

$$r = (p_d - p_h s)f + (q_d - p_d s)g$$

is a non-zero element in S with smaller degree than h, which is a contradiction.

We now claim that h is a greatest common divisor of f and g (note that h is a divisor of f because $f = 1 \cdot f + 0 \cdot g \in S$. Also, h is a divisor of g because $g = 0 \cdot f + 1 \cdot g \in S$). Suppose c is a common divisor of f and g, so that $f = cp_c$ and $g = cq_c$ for some polynomials p_c and q_c. Then

$$h = p_h f + q_h g = p_h p_c c + q_h q_c c = (p_h p_c + q_h q_c)c.$$

Hence, c divides h. In other words, any other common divisor of f and g must divide h. This shows that h is a greasest common divisor. The uniqueness of the greatest common divisor up to multiplicative constant follows from the fact that $\deg(c) = \deg(h)$ only if $\deg(p_h p_c + q_h q_c) = 0$. That is, if $p_h p_c + q_h q_c$ is a constant.

□

If two polynomials f and g have no roots in common then the greatest common divisor is a constant number. The polynomials f and g are said to be relatively prime.