ON THE JORDAN CANONICAL FORM

Our goal in these notes is to prove that if A is an $n \times n$ matrix over a field such that the characteristic polynomial of A contains n roots (up to multiplicity), then A is similar to a block diagonal matrix of the form

$$J = \left(\begin{array}{c|c|c|c} J_1 & & & \\ \hline & J_2 & & \\ \hline & & \ddots & \\ \hline & & & J_t \end{array}\right)$$

where each block J_i is of the form

$$J_i = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix},$$

λ being an eigenvalue of A. Matrices of the form J_i are called *Jordan blocks*. For example, A would be diagonalisable if and only if each J_i had size 1×1.

The diagonal decomposition of a matrix into Jordan blocks has to do with decomposing a vector space into direct sums of what are called invariant subspaces. Let $T\colon V \to V$ be a linear transformation. A subspace U of V is said to be T*-invariant* if $T(U) \subseteq U$. In this situation T may be viewed as a linear transformation $U \to U$. We give our first glimpse of a structure theorem for matrices.

Lemma 1. *Let $T\colon V \to V$ be a linear transformation and suppose that $V = U \oplus W$, where U and W are T-invariant subspaces of V. If $u_1, \dots, u_m$ and $w_1, \dots, w_r$ are bases of U and W, respectively, then the matrix that represents T with respect to the basis*

$$\mathcal{B} = \{u_1, \dots, u_m, w_1, \dots, w_r\}$$

of V has the form

$$M_{\mathcal{B}}^{\mathcal{B}}(T) = \left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right),$$

where A is an $m \times m$ matrix, B is an $r \times r$ matrix and 0 is the zero matrix. (Here the top 0 denotes an $m \times r$ matrix, the bottom 0 an $r \times m$ matrix.)

Proof. This is a routine exercise. $\square$

For a linear transformation $T\colon V \to V$ and an integer $i \geq 0$, we define T^i to be the i-fold composition of T with itself, that is

$$T^i = \underbrace{T \circ T \circ \cdots \circ T}_{i \text{ times}} \colon V \longrightarrow V.$$

By convention, we set $T^0 = \mathrm{id}_V$. The first two results are routine exercises.

Lemma 2. *There are nested sequences of subspaces*

$$\{0_V\} = \operatorname{Ker}(T^0) \subseteq \operatorname{Ker}(T^1) \subseteq \operatorname{Ker}(T^2) \subseteq \cdots \tag{1}$$

and

$$V = \operatorname{Im}(T^0) \supseteq \operatorname{Im}(T^1) \supseteq \operatorname{Im}(T^2) \supseteq \cdots .$$

Lemma 3. *The subspaces* $\operatorname{Ker}(T^i)$ *and* $\operatorname{Im}(T^i)$ *are* T*-invariant.*

We now focus our attention on the above sequence of kernels.

Lemma 4. *In the sequence (1), if* $\operatorname{Ker}(T^i) = \operatorname{Ker}(T^{i+1})$ *for some* $i \geq 0$, *then*

$$\operatorname{Ker}(T^{i+j}) = \operatorname{Ker}(T^i) \qquad \text{for all } j \geq 1.$$

In other words, (1) will be of the form

$$\{0_V\} \subsetneq \operatorname{Ker}(T^1) \subsetneq \cdots \subsetneq \operatorname{Ker}(T^i) = \operatorname{Ker}(T^{i+1}) = \operatorname{Ker}(T^{i+2}) = \cdots .$$

Proof. Let i be as described. We already have the inclusion $\operatorname{Ker}(T^i) \subseteq \operatorname{Ker}(T^{i+j})$. To prove the reverse inclusion we proceed by induction on j. For $j = 1$ there is nothing to prove, so let $j > 1$ and assume that $\operatorname{Ker}(T^{i+j-1}) = \operatorname{Ker}(T^i)$. Let $v \in \operatorname{Ker}(T^{i+j})$. Then $T^{i+j}(v) = 0_V$ so that $T^{i+j-1}(T(v)) = 0_V$. This shows that

$$T(v) \in \operatorname{Ker}(T^{i+j-1}(v)) = \operatorname{Ker}(T^i).$$

This means that $T^i(T(v)) = 0_V$, that is, $T^{i+1}(v) = 0_V$. In other words,

$$v \in \operatorname{Ker}(T^{i+1}) = \operatorname{Ker}(T^i)$$

as required. □

From now on we shall restrict to the case where V is finite dimensional. In this situation the sequence (1) stabilises after at most $\dim V$ steps.

Lemma 5. *If* V *is* n*-dimensional, then*

$$\operatorname{Ker}(T^n) = \operatorname{Ker}(T^{n+1}) = \operatorname{Ker}(T^{n+2}) = \cdots .$$

Proof. We could not have a sequence of proper inclusions

$$\{0_V\} \subsetneq \operatorname{Ker}(T^1) \subsetneq \cdots \subsetneq \operatorname{Ker}(T^n) \subsetneq \operatorname{Ker}(T^{n+1}),$$

because after each such inclusion, the dimension of the subspace would increase by at least 1, i.e., we would have $\dim \operatorname{Ker}(T^{n+1}) \geq n+1$, contradicting the fact that a subspace of V has dimension at most n. □

Before stating the next result, we need to introduce the very important concept of nilpotence. A linear transformation $T\colon V \to V$ is *nilpotent* if there exists an integer $N \geq 0$ such that $T^N = 0$, the zero transformation that assigns each vector in V to 0_V.

Lemma 6. *If V is n-dimensional and $T\colon V \to V$ is a linear transformation, then* $\operatorname{Ker}(T^n)$ *and* $\operatorname{Im}(T^n)$ *are T-invariant and* $V = \operatorname{Ker}(T^n) \oplus \operatorname{Im}(T^n)$. *Moreover, T acts nilpotently on* $\operatorname{Ker}(T^n)$ *and as an isomorphism on* $\operatorname{Im}(T^n)$.

Proof. By Lemma 5 we have

$$\operatorname{Ker}(T^n) = \operatorname{Ker}(T^{n+1}) = \operatorname{Ker}(T^{n+2}) = \cdots .$$

Let $v \in \operatorname{Ker}(T^n) \cap \operatorname{Im}(T^n)$. Then $T^n(v) = 0_V$ and there is some $u \in V$ such that $T^n(u) = v$. We then have $T^{2n}(u) = T^n(v) = 0_V$ so that $u \in \operatorname{Ker}(T^{2n}) = \operatorname{Ker}(T^n)$. It follows that

$$v = T^n(u) = 0_V,$$

hence $\operatorname{Ker}(T^n) \cap \operatorname{Im}(T^n) = \{0_V\}$. By the formula for the dimension of a sum of subspaces, we deduce that

$$\dim \big(\operatorname{Ker}(T^n) + \operatorname{Im}(T^n)\big) = \dim \operatorname{Ker}(T^n) + \dim \operatorname{Im}(T^n).$$

On the other hand, applying the rank-nullity theorem to T^n gives us

$$\dim \operatorname{Ker}(T^n) + \dim \operatorname{Im}(T^n) = \dim V.$$

We must therefore have $\operatorname{Ker}(T^n) + \operatorname{Im}(T^n) = V$, and then $V = \operatorname{Ker}(T^n) \oplus \operatorname{Im}(T^n)$ because $\operatorname{Ker}(T^n) \cap \operatorname{Im}(T^n) = \{0_V\}$.

The fact that $T^n\big(\operatorname{Ker}(T^n)\big) = \{0_V\}$ tells us straight away that T is nilpotent on $\operatorname{Ker}(T^n)$. Now suppose that $v \in \operatorname{Im}(T^n)$ satisfies $T(v) = 0_V$. Write $v = T^n(u)$ for some $u \in V$. We then have $T^{n+1}(u) = 0_V$ so that $u \in \operatorname{Ker}(T^{n+1}) = \operatorname{Ker}(T^n)$, hence $v = T^n(u) = 0_V$. This shows that the kernel of $T\colon \operatorname{Im}(T^n) \to \operatorname{Im}(T^n)$ is zero. Since $\operatorname{Im}(T^n)$ is finite dimensional, that map must be an isomorphism. □

Now suppose that the characteristic polynomial $\chi_T(x)$ of $T\colon V \to V$ has n linear factors, where $n = \dim V$. Let $\lambda_1, \ldots, \lambda_r$ be the distinct eigenvalues of T and let $m_1, \ldots, m_r$ be their respective algebraic multiplicities. We then have

$$m_1 + \cdots + m_r = n.$$

Motivated by the previous result, if λ is any eigenvalue of T, we define the *generalised eigenspace* of λ to be the collection

$$\{v \in V \mid \text{there exists } i \geq 0 \text{ such that } (T - \lambda \operatorname{id}_V)^i(v) = 0_V\}.$$

In other words, this is the subspace of elements in V that are annihilated by some power of $T - \lambda \operatorname{id}_V$. In particular, the eigenspace

$$E_\lambda = \{v \in V \mid (T - \lambda \operatorname{id}_V)(v) = 0_V\}$$

is a subspace of the generalised eigenspace of λ. Studying Lemma 5 reveals that the generalised eigenspace of λ is none other than

$$\operatorname{Ker}\big((T - \lambda \operatorname{id}_V)^n\big).$$

The elements of $\operatorname{Ker}\big((T - \lambda \operatorname{id}_V)^n\big)$ are called *generalised eigenvectors* associated to λ.

If $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues of T, we adopt the notation

$$U_j = \operatorname{Ker}\big((T - \lambda_j \operatorname{id}_V)^n\big).$$

The following result will imply that non-zero generalised eigenvectors associated to distinct eigenvalues are linearly independent. This gives a new proof that eigenvectors associated to distinct eigenvalues are always linearly independent.

Lemma 7. *If $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues of T, then the sum of their generalised eigenspaces is direct. In other words,*

$$U_1 + \cdots + U_r = U_1 \oplus \cdots \oplus U_r.$$

Proof. For each $1 \leq j \leq r$ we need to prove that

$$U_j \cap (U_1 + \cdots + U_{j-1} + U_{j+1} + \cdots + U_r) = \{0_V\}.$$

Without loss of generality it suffices to show that

$$U_1 \cap (U_2 + \cdots + U_r) = \{0_V\}.$$

This will be true if

$$U_1 \cap (U_2 + \cdots + U_j) = \{0_V\} \qquad \text{for all } 2 \leq j \leq r.$$

To prove that, we proceed by induction on j. The base case is $j = 2$, so pick $v \in U_1 \cap U_2$. For the sake of contradiction, suppose that $v \neq 0_V$. Since $v \in U_2 = \operatorname{Ker}\big((T - \lambda_2 \operatorname{id}_V)^n\big)$, there is some integer $m \geq 0$ such that $(T - \lambda_2 \operatorname{id}_V)^m(v) \neq 0_V$ and $(T - \lambda_2 \operatorname{id}_V)^{m+1}(v) = 0_V$. Because $T - \lambda_2 \operatorname{id}_V$ annihilates the vector $w = (T - \lambda_2 \operatorname{id}_V)^m(v)$, the latter is an eigenvector associated to λ_2, that is, $T(w) = \lambda_2 w$. Note that the linear transformations $T - \lambda_1 \operatorname{id}_V$ and $T - \lambda_2 \operatorname{id}_V$ commute with one another. Since $v \in U_1$, it follows that $(T - \lambda_1 \operatorname{id}_V)^n(w) = 0_V$. But T acts on w as scalar multiplication by λ_2, thus

$$(\lambda_2 - \lambda_1)^n w = 0_V.$$

Because λ_1 and λ_2 are distinct, $(\lambda_2 - \lambda_1)^n \neq 0$ in F. This forces $w = 0_V$, contradicting the fact that w is non-zero. We therefore have $U_1 \cap U_2 = \{0_V\}$.

For the induction step, let $j > 2$ and assume that

$$U_1 \cap (U_2 + \cdots + U_{j-1}) = \{0_V\}.$$

Let $v \in U_1 \cap (U_2 + \cdots + U_j)$. We may then write $v = v_1 + \cdots + v_j$ with $v_k \in U_k$ for each $2 \leq k \leq j$. We then have $(T - \lambda_j \operatorname{id}_V)^n(v_j) = 0_V$, so

$$(T - \lambda_j \operatorname{id}_V)^n(v) = (T - \lambda_j \operatorname{id}_V)^n(v_2) + \cdots + (T - \lambda_j \operatorname{id}_V)^n(v_{j-1}). \tag{2}$$

Note that since each U_k is T-invariant, it is also $(T - \lambda_j \operatorname{id}_V)$-invariant. The left hand side of (2) therefore lies in U_1 and the right hand side in $U_2 + \cdots + U_{j-1}$, hence both sides lie in $U_1 \cap (U_2 + \cdots + U_{j-1}) = \{0_V\}$. The fact that $(T - \lambda_j \operatorname{id}_V)^n(v) = 0_V$ implies that $v \in U_j$, so $v \in U_1 \cap U_j$. The base case (letting U_j play the role of U_2) tells us that this intersection is the zero subspace, so $v = 0_V$ as required. □

We are now ready to prove one of the more important theorems in matrix theory.

Theorem 8. *Let $T \colon V \to V$ be a linear transformation with distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ having respective algebraic multiplicities $m_1, \ldots, m_r$. For each $1 \leq j \leq r$, let U_j denote the generalised eigenspace* $\operatorname{Ker}\big((T - \lambda_j \operatorname{id}_V)^n\big)$.

1. *There is a direct sum decomposition $V = U_1 \oplus \cdots \oplus U_r$.*
2. *Each subspace U_j has dimension m_j.*
3. *Each subspace U_j is $(T - \lambda_j \operatorname{id}_V)$-invariant, and $T - \lambda_j \operatorname{id}_V$ is nilpotent on U_j.*

Proof. For each $1 \leq j \leq r$, write $T_j = T - \lambda_j \operatorname{id}_V$ so that $U_j = \operatorname{Ker}(T_j^n)$. Set $W_j = \operatorname{Im}(T_j^n)$. Lemma 6 gives us a decomposition of T_j-invariant subspaces

$$V = U_j \oplus W_j,$$

where T_j acts nilpotently on U_j and as an isomorphism on W_j. Note that $T = T_j + \lambda_j \operatorname{id}_V$, so U_j and W_j are also T-invariant. The characteristic polynomial of T on V is therefore the product of its characteristic polynomial on U_j and its characteristic polynomial on W_j. The fact that T_j acts isomorphically on W_j tells us that λ_j cannot be an eigenvalue of T on W_j. This means that the characteristic polynomial of T on U_j contains all factors $x - \lambda_j$ in the characteristic polynomial of T on V. On the other hand, Lemma 7 tells us that U_j does not contain any of the eigenvectors associated with λ_k for $k \neq j$. It follows that the only linear factors of the characteristic polynomial of T on U_j are of the form $x - \lambda_j$. Putting this all together, we conclude that the characteristic polynomial of T on U_j is $(x - \lambda_j)^{m_j}$. But the degree of the characteristic polynomial of T on U_j is equal to the dimension of U_j. In other words, $\dim U_j = m_j$. This proves (2).

It now follows by Lemma 7 that

$$\dim(U_1 \oplus \cdots \oplus U_r) = m_1 + \cdots + m_r = \dim V.$$

We must therefore have $V = U_1 \oplus \cdots \oplus U_r$, establishing (1).

Part (3) was part of Lemma 6. □

This is actually quite interesting! We know that a vector space V is the direct sum of the eigenspaces for T if and only if any matrix representing T is diagonalisable. Theorem 8 tells us that if we replace 'eigenspace' with 'generalised eigenspace', then the same is true of any linear transformation.

Corollary 9. *Let $T\colon V \to V$ be a linear transformation with distinct eigenvalues $\lambda_1, \dots, \lambda_r$ having respective algebraic multiplicities $m_1, \dots, m_r$. There is a basis $\mathcal{B}$ of V with respect to which T is represented by a block diagonal matrix*

$$M^{\mathcal{B}}_{\mathcal{B}}(T) = \left(\begin{array}{c|c|c|c} A_1 & & & \\ \hline & A_2 & & \\ \hline & & \ddots & \\ \hline & & & A_r \end{array}\right)$$

where $A_j - \lambda_j I$ is a nilpotent matrix. (Here, I denotes an $m_j \times m_j$ identity matrix.)

Proof. For each generalised eigenspace U_j pick a basis $\mathcal{B}_j$. Since $V = U_1 \oplus \cdots \oplus U_r$, Lemma 1 tells us that the matrix representing T with respect to the basis

$$\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$$

of V will be represented by a matrix of the above form. The statement about nilpotence is immediate from part (3) of Theorem 8. □

Theorem 10. *If $T\colon V \to V$ is a nilpotent linear transformation, then there is a collection of vectors $v_1, \dots, v_s$ in V with each $v_i \notin \operatorname{Im}(T)$ such that*

$$\begin{aligned} &v_1,\ T(v_1),\ T^2(v_1),\ \dots,\ T^{M_1}(v_1),\\ &v_2,\ T(v_2),\ T^2(v_2),\ \dots,\ T^{M_2}(v_2),\\ &\vdots\\ &v_s,\ T(v_s),\ T^2(v_s),\ \dots,\ T^{M_s}(v_s) \end{aligned}$$

forms a basis of V. Here, for each $1 \le i \le s$, $M_i \ge 0$ is the integer satisfying $T^{M_i}(v_i) \ne 0_V$ and $T^{M_i+1}(v_i) = 0_V$.

Proof. Since T is nilpotent, there exists an integer $N \ge 0$ such that $T^N = 0$ but $T^{N-1} \ne 0$. This is called the *index of nilpotence* of T. We proceed by induction on N. If $N = 1$ then T is the zero map, so any basis $v_1, \dots, v_s$ of V will be of the above form (with each $N_i = 0$).

Now let $N > 1$ and assume that if $S\colon W \to W$ is a linear transformation and $S^{N-1} = 0$, then W has a basis of the described form. Observe that $T|_{\mathrm{Im}(T)}\colon \mathrm{Im}(T) \to \mathrm{Im}(T)$ satisfies $T|_{\mathrm{Im}(T)}^{N-1} = 0$. By the inductive hypothesis $\mathrm{Im}(T)$ has such a basis

$$
\begin{aligned}
&w_1,\ T(w_1),\ T^2(w_1),\ \ldots,\ T^{N_1}(w_1),\\
&w_2,\ T(w_2),\ T^2(w_2),\ \ldots,\ T^{N_2}(w_2),\\
&\vdots\\
&w_t,\ T(w_t),\ T^2(w_t),\ \ldots,\ T^{N_t}(w_t).
\end{aligned} \tag{3}
$$

Since each $w_j \in \mathrm{Im}(T)$, there exist elements $v_1, \ldots, v_t$ in V such that $T(v_j) = w_j$. Adjoining these vectors to the collection (3) yields the collection

$$
\begin{aligned}
&v_1,\ T(v_1),\ T^2(v_1),\ \ldots,\ T^{N_1+1}(v_1),\\
&v_2,\ T(v_2),\ T^2(v_2),\ \ldots,\ T^{N_2+1}(v_2),\\
&\vdots\\
&v_t,\ T(v_t),\ T^2(v_t),\ \ldots,\ T^{N_t+1}(v_t).
\end{aligned}
$$

One can check that these are linearly independent, but they might not span V. To find the missing vectors, we note that

$$T^{N_1+1}(v_1),\ T^{N_2+1}(v_2),\ \ldots,\ T^{N_t+1}(v_t)$$

form a basis of $\mathrm{Ker}(T) \cap \mathrm{Im}(T)$. Extend this to a basis

$$u_1,\ \ldots,\ u_m,\ T^{N_1+1}(v_1),\ T^{N_2+1}(v_2),\ \ldots,\ T^{N_t+1}(v_t)$$

of $\mathrm{Ker}(T)$. Adjoining the u_i to the above collection gives us

$$
\begin{aligned}
&u_1,\\
&\vdots\\
&u_m,\\
&v_1,\ T(v_1),\ T^2(v_1),\ \ldots,\ T^{N_1+1}(v_1),\\
&\vdots\\
&v_t,\ T(v_t),\ T^2(v_t),\ \ldots,\ T^{N_t+1}(v_t).
\end{aligned}
$$

One may show (although the details are not enlightening) that this final collection is a basis of V having the required form. □

The previous theorem seems mysterious, and its proof was rather boring, I have to admit. So what does it buy us? Notice that each subspace

$$\mathrm{Span}\left\{T^{M_i}(v_i), T^{M_i-1}(v_i), \ldots, T(v_i), v_i\right\}$$

of V is T-invariant. What does the matrix representing T on this subspace look like? Well, T maps the first basis element to zero, it maps $T^{M_i-1}(v_i)$ to $T^{M_i}(v_i)$, it maps $T^{M_i-2}(v_i)$ to $T^{M_i-1}(v_i)$, etc., until we get to T mapping v_i to $T(v_i)$. In other words, T will be represented on this subspace by the $(M_i + 1) \times (M_i + 1)$ matrix

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

The action of T on V with respect to the basis given in Theorem 10 (with each row written in reverse order) will be a block diagonal matrix, where each block is of the above form. In other words, every *nilpotent* linear transformation has a Jordan canonical form (with zeros along the diagonal).

We are now in a position to explain how one can (abstractly) deduce the existence of a Jordan canonical form for any linear transformation $T \colon V \to V$. We do this in steps.

1. Identify the distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of T and form the respective generalised eigenspaces $U_1, \ldots, U_r$.
2. For each $1 \leq j \leq r$, the linear transformation $T_j = T - \lambda_j \operatorname{id}_V$ acts nilpotently on U_j. By the remarks following Theorem 10, we may find a basis $\mathcal{B}_j$ of U_j with respect to which T_j is represented by a block diagonal matrix whose blocks are of the form

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

   It follows that, with respect to the basis $\mathcal{B}_j$, the linear transformation $T = T_j + \lambda_j \operatorname{id}_V$ will be represented by a block diagonal matrix A_j whose blocks are of the form

$$\begin{pmatrix} \lambda_j & 1 & & & \\ & \lambda_j & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_j & 1 \\ & & & & \lambda_j \end{pmatrix}.$$

   In other words, the matrix A_j will be in Jordan form.

3. By Corollary 9, the matrix representing T with respect to the basis $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$ will be

$$\left(\begin{array}{c|c|c|c} A_1 & & & \\ \hline & A_2 & & \\ \hline & & \ddots & \\ \hline & & & A_r \end{array}\right).$$

By the previous step, this will be in Jordan form.

Applying this to matrices. Now suppose that we are given an $n \times n$ matrix with entries in a field F. We know that the function

$$T \colon F^n \longrightarrow F^n, \qquad \mathbf{v} \mapsto A\mathbf{v}$$

is a linear transformation. With respect to the standard basis

$$\mathcal{S} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$$

of F^n, T is represented by none other than A. Our goal will be to find a basis

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$$

of F^n with respect to which T is represented by a Jordan canonical form matrix J. If such a basis can be found (and one always exists by the above theory), then the change of basis matrix from $\mathcal{S}$ to $\mathcal{B}$ will be

$$P = \left(\begin{array}{cccc} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{array} \right),$$

and we will have

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

by our discussion of change of basis and similarity. The basis $\mathcal{B}$ is called a *Jordan basis*. To find a Jordan basis for a matrix, it will be helpful to better understand what properties its basis elements must satisfy.

Suppose that we are given a 17×17 matrix A whose Jordan canonical form is

$$
J = \left(\begin{array}{ccc|cc|cc|cc|c|c|ccc|c|cc}
\lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & \mu & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \nu & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \nu & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \nu & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \nu & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \nu & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \nu
\end{array}\right).
$$

In this example we are implicitly assuming that A has three distinct eigenvalues λ, μ and ν. By similarity, the characteristic polynomial of A will be equal to that of J, and since J is upper triangular, one readily computes this to be

$$\chi_A(x) = \chi_J(x) = -(x-\lambda)^5(x-\mu)^6(x-\nu)^6.$$

This illustrates our first key fact.

> **The algebraic multiplicity of an eigenvalue is equal to the number of times it appears on the main diagonal of the Jordan canonical form.**

Now suppose that J represents T with respect to the Jordan basis

$$\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{17}\}.$$

By the theory presented earlier, we deduce that $\mathbf{v}_1, \ldots, \mathbf{v}_5$ form a basis of the generalised eigenspace of λ, $\mathbf{v}_6, \ldots, \mathbf{v}_{11}$ form basis of the generalised eigenspace of μ and $\mathbf{v}_{12}, \ldots, \mathbf{v}_{17}$ form a basis of the generalised eigenspace of ν.

The first and second Jordan blocks in J tell us the following information.

$$T(\mathbf{v}_1) = \lambda\mathbf{v}_1, \qquad T(\mathbf{v}_2) = \mathbf{v}_1 + \lambda\mathbf{v}_2, \qquad T(\mathbf{v}_3) = \mathbf{v}_2 + \lambda\mathbf{v}_3,$$

$$T(\mathbf{v}_4) = \lambda\mathbf{v}_4, \qquad T(\mathbf{v}_5) = \mathbf{v}_4 + \lambda\mathbf{v}_5$$

From this we see that $\mathbf{v}_1$ and $\mathbf{v}_4$ are eigenvectors associated to λ, but that the other vectors in $\mathcal{B}$ are not. This suggests the second general fact.

> **The geometric multiplicity of an eigenvalue is equal to the number of Jordan blocks that contain it. In particular, the leftmost columns in a Jordan block correspond to eigenvectors in a Jordan basis.**

In the example $\mathbf{v}_1, \mathbf{v}_4$ form a basis of E_λ, $\mathbf{v}_6, \mathbf{v}_8, \mathbf{v}_{10}, \mathbf{v}_{11}$ form a basis of E_μ and $\mathbf{v}_{12}, \mathbf{v}_{15}, \mathbf{v}_{16}$ form a basis of E_ν. In fact, we can see just by counting Jordan blocks that λ has geometric multiplicity 2, μ has geometric multiplicity 4 and ν has geometric multiplicity 3.

Taking into account that T is multiplication by A and rearranging the above equations a bit, we obtain

$$(A - \lambda I)\mathbf{v}_1 = \mathbf{0}, \qquad (A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1, \qquad (A - \lambda I)\mathbf{v}_3 = \mathbf{v}_2,$$

$$(A - \lambda I)\mathbf{v}_4 = \mathbf{0}, \qquad (A - \lambda I)\mathbf{v}_5 = \mathbf{v}_4.$$

The Jordan basis elements associated to blocks containing λ therefore satisfy

$$\begin{aligned} A - \lambda I \colon \mathbf{v}_3 &\longmapsto \mathbf{v}_2 \longmapsto \mathbf{v}_1 \longmapsto \mathbf{0}, \\ \mathbf{v}_5 &\longmapsto \mathbf{v}_4 \longmapsto \mathbf{0}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} A - \mu I \colon \mathbf{v}_7 &\longmapsto \mathbf{v}_6 \longmapsto \mathbf{0}, \\ \mathbf{v}_9 &\longmapsto \mathbf{v}_8 \longmapsto \mathbf{0}, \\ \mathbf{v}_{10} &\longmapsto \mathbf{0}, \\ \mathbf{v}_{11} &\longmapsto \mathbf{0} \end{aligned}$$

and

$$\begin{aligned} A - \nu I \colon \mathbf{v}_{14} &\longmapsto \mathbf{v}_{13} \longmapsto \mathbf{v}_{12} \longmapsto \mathbf{0}, \\ \mathbf{v}_9 &\longmapsto \mathbf{v}_8 \longmapsto \mathbf{0}, \\ \mathbf{v}_{15} &\longmapsto \mathbf{0}, \\ \mathbf{v}_{17} &\longmapsto \mathbf{v}_{16} \longmapsto \mathbf{0}. \end{aligned}$$

All of this speaks to how we interpret a Jordan matrix. Actually computing a Jordan basis is a different matter all together, but we should keep in mind as we go along that we are searching for generalised eigenvectors that exhibit the above sort of behaviour.

Computing Jordan forms and Jordan bases. Let A be an $n \times n$ matrix and

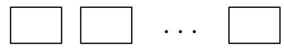
$$T \colon F^n \longrightarrow F^n, \qquad \mathbf{v} \mapsto A\mathbf{v}$$

the usual linear transformation. If $\lambda \in F$ is an eigenvalue of T, we compute the number and sizes of the Jordan blocks containing λ in the following way.

1. Compute the integer

$$d_1 = \dim \operatorname{Ker}(T - \lambda \operatorname{id}_{F^n})$$

using the usual techniques. Draw d_1 boxes in one row, e.g.,

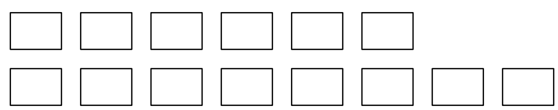


2. Compute the integer

$$d_2 = \dim \operatorname{Ker}\big((T - \lambda \operatorname{id}_{F^n})^2\big).$$

Draw $d_2 - d_1$ boxes in a row above the boxes you have already drawn. Some boxes in the bottom row might not get boxes above them in the top row. If this happens, do not cover such a box in any of the subsequent steps.

As a running example, if $d_1 = 8$ and $d_2 = 14$, then the picture would become

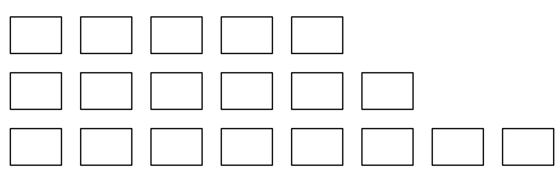


3. Compute the integer

$$d_3 = \dim \operatorname{Ker}\big((T - \lambda \operatorname{id}_{F^n})^3\big).$$

Draw $d_3 - d_2$ boxes in a row above the boxes you have already drawn.

In our example, if $d_3 = 19$, then the picture would become



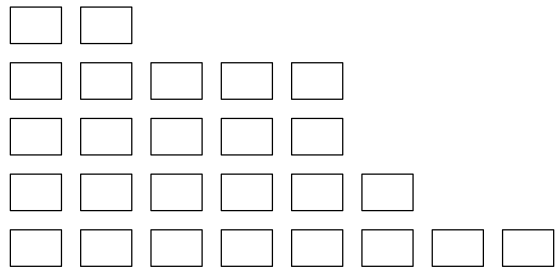
4. In general, keep computing the integers

$$d_i = \dim \operatorname{Ker}\big((T - \lambda \operatorname{id}_{F^n})^i\big).$$

each time placing $d_i - d_{i-1}$ boxes in each row. The process will automatically terminate after the step in which d_i becomes the algebraic multiplicity m of λ since that is the dimension of the generalised eigenspace

$$\dim \operatorname{Ker}\big((T - \lambda \operatorname{id}_{F^n})^n\big).$$

For example, if we had $d_4 = 24$ and $d_5 = m = 26$, the our picture would become

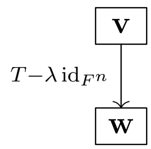


Believe it or not, this tells us the sizes of all of the Jordan blocks containing λ. Each column corresponds to a Jordan block, which will have size the number of boxes in that column. In our running example, λ would have two Jordan blocks of size 5, three Jordan blocks of size 4, one Jordan block of size 2 and two Jordan blocks of size 1. We would write these in the Jordan canonical form as

$$\left(\begin{array}{c|c|c|c|c|c|c|c|c}
\begin{matrix}\lambda & 1 & 0 & 0 & 0\\ 0 & \lambda & 1 & 0 & 0\\ 0 & 0 & \lambda & 1 & 0\\ 0 & 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & 0 & \lambda\end{matrix} & & & & & & & & \\
\hline
 & \begin{matrix}\lambda & 1 & 0 & 0 & 0\\ 0 & \lambda & 1 & 0 & 0\\ 0 & 0 & \lambda & 1 & 0\\ 0 & 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & 0 & \lambda\end{matrix} & & & & & & & \\
\hline
 & & \begin{matrix}\lambda & 1 & 0 & 0\\ 0 & \lambda & 1 & 0\\ 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & \lambda\end{matrix} & & & & & & \\
\hline
 & & & \begin{matrix}\lambda & 1 & 0 & 0\\ 0 & \lambda & 1 & 0\\ 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & \lambda\end{matrix} & & & & & \\
\hline
 & & & & \begin{matrix}\lambda & 1 & 0 & 0\\ 0 & \lambda & 1 & 0\\ 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & \lambda\end{matrix} & & & & \\
\hline
 & & & & & \begin{matrix}\lambda & 1\\ 0 & \lambda\end{matrix} & & & \\
\hline
 & & & & & & \lambda & & \\
\hline
 & & & & & & & \lambda & \\
\hline
 & & & & & & & & \ddots
\end{array}\right),$$

and then continue with the same computations for the other eigenvalues of T.

The question remains as to how one computes a Jordan basis for T. Again, we do this one eigenvalue at a time. To find vectors that correspond to columns in the Jordan blocks containing an eigenvalue λ, we refer back to the square diagrams constructed in the previous steps. Each square will contain a Jordan basis element. If a basis element $\mathbf{w}$ lies below the square containing a vector $\mathbf{v}$, this indicates that $T - \lambda\,\mathrm{id}_{F^n}$ maps $\mathbf{v}$ to $\mathbf{w}$, i.e.,



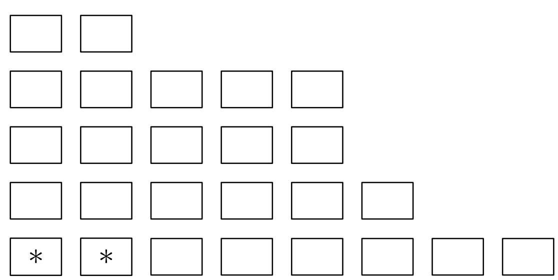
If a vector lies in a square in the bottom row, then the vector inside it lies in the kernel of $T - \lambda\,\mathrm{id}_{F^n}$. Our goal then is to fill in the squares with linearly independent elements that satisfy the described properties.

Warning: The following algorithm is *very* long and would be *very* tedious to implement by hand. The idea is that you should have all of the skills in hand to perform each step if necessary. Ideally you would ask a computer to do all of this, but the point here is for you to imagine what a computer would be doing if such an algorithm were programmed.

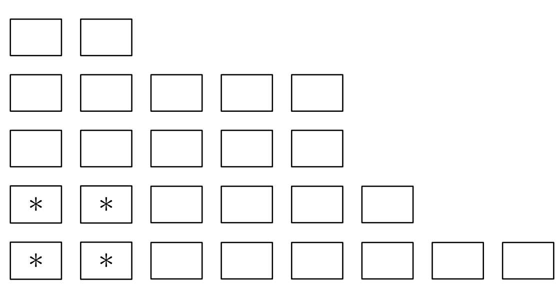
1. Take the size r of the largest Jordan block containing λ, then compute a basis of

$$\mathrm{Ker}(T - \lambda\,\mathrm{id}_{F^n}) \cap \mathrm{Im}\big((T - \lambda\,\mathrm{id}_{F^n})^{r-1}\big).$$

   Fill in the bottom squares of the longest columns with these basis elements. In our running example we have $r = 5$. This step would fill in the squares



2. Find preimages of the elements you found in step 1 under $T - \lambda\,\mathrm{id}_{F^n}$. The diagram for our example will then be filled in as follows.



Note that

$$\operatorname{Ker}(T - \lambda \operatorname{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda \operatorname{id}_{F^n})^{r-1}\big) \subseteq \operatorname{Ker}(T - \lambda \operatorname{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda \operatorname{id}_{F^n})^{r-2}\big).$$

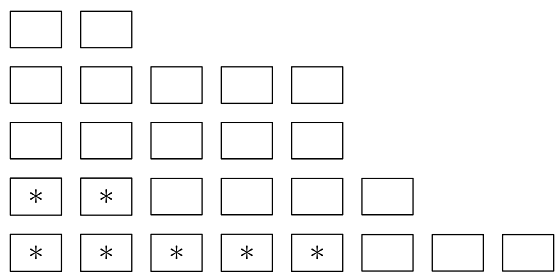
Use the proof of the replacement theorem to extend the basis of

$$\operatorname{Ker}(T - \lambda \operatorname{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda \operatorname{id}_{F^n})^{r-1}\big)$$

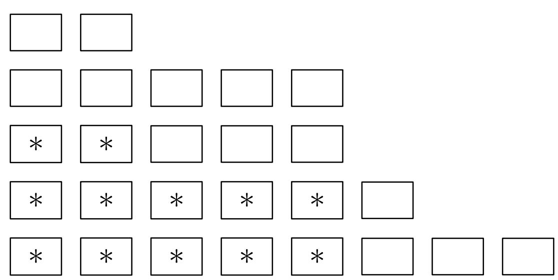
found in step 1 to a basis of

$$\operatorname{Ker}(T - \lambda \operatorname{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda \operatorname{id}_{F^n})^{r-2}\big).$$

Use the basis vectors you added onto the the basis of $\operatorname{Im}\big((T - \lambda \operatorname{id}_{F^n})^{r-1}\big)$ to fill in squares to the right of the boxes you have already filled in the bottom row. In our example this would produce



3. Find preimages of the elements you found in step 2 under $T - \lambda \operatorname{id}_{F^n}$. The diagram will then be filled in as follows.



Note that

$$\operatorname{Ker}(T - \lambda \operatorname{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda \operatorname{id}_{F^n})^{r-2}\big) \subseteq \operatorname{Ker}(T - \lambda \operatorname{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda \operatorname{id}_{F^n})^{r-3}\big).$$

Use the proof of the replacement theorem to extend a basis of

$$\operatorname{Ker}(T - \lambda \operatorname{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda \operatorname{id}_{F^n})^{r-2}\big)$$

(the vectors you have already found for the bottom squares will do) to a basis of

$$\operatorname{Ker}(T - \lambda \operatorname{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda \operatorname{id}_{F^n})^{r-3}\big).$$

Use the basis vectors you added on in this step to fill in squares to the right of the boxes you have already filled in the bottom row. In our example we have

$$\operatorname{Ker}(T - \lambda\,\mathrm{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda\,\mathrm{id}_{F^n})^3\big) = \operatorname{Ker}(T - \lambda\,\mathrm{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda\,\mathrm{id}_{F^n})^2\big),$$

so the previous picture would not change.

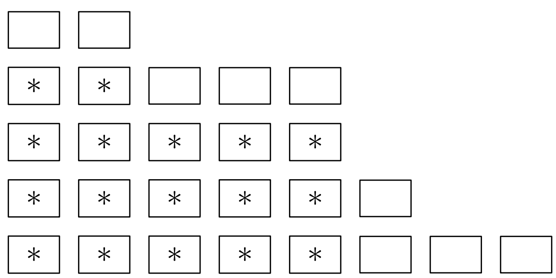
4. Continue in this way. At each step i, find preimages of the elements you found in step $i-1$. Then extend a basis of

$$\operatorname{Ker}(T - \lambda\,\mathrm{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda\,\mathrm{id}_{F^n})^{r-(i-1)}\big)$$

to a basis of

$$\operatorname{Ker}(T - \lambda\,\mathrm{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda\,\mathrm{id}_{F^n})^{r-i}\big).$$

Fill in the boxes in the bottom row with the vectors that this procedure produces. In our example, step 4 would begin by choosing preimages of the vectors found in step 3, giving us



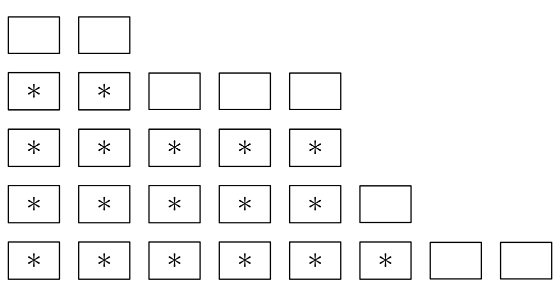
Extending a basis of

$$\operatorname{Ker}(T - \lambda\,\mathrm{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda\,\mathrm{id}_{F^n})^2\big)$$

to a basis of

$$\operatorname{Ker}(T - \lambda\,\mathrm{id}_{F^n}) \cap \operatorname{Im}\big((T - \lambda\,\mathrm{id}_{F^n})\big)$$

would allow us to fill in one box.



Finding preimages of the vectors just found would produce the picture

*	*						
*	*	*	*	*			
*	*	*	*	*			
*	*	*	*	*	*		
*	*	*	*	*	*		

and extending a basis of

$$\operatorname{Ker}(T - \lambda \operatorname{id}_{F^n}) \cap \operatorname{Im}(T - \lambda \operatorname{id}_{F^n})$$

to a basis of

$$\operatorname{Ker}(T - \lambda \operatorname{id}_{F^n})$$

would allow us to fill in the remaining two boxes.

Parts of the above algorithm are actually useful in by-hand computations. Consider the matrix

$$A = \begin{pmatrix} 2 & 3 & 3 & -6 \\ 0 & -3 & 1 & 1 \\ -1 & -3 & -1 & 3 \\ 1 & -1 & 2 & -2 \end{pmatrix}.$$

One computes that the characteristic polynomial of A is

$$\chi_A(x) = (x+1)^4.$$

We see that $\lambda = -1$ is the only eigenvalue of A, necessarily having algebraic multiplicity $m = 4$. In particular, its generalised eigenspace is

$$\operatorname{Nul}(A+I)^4 = \mathbb{R}^4.$$

To determine the sizes of the Jordan blocks, we compute

$$d_1 = \dim \operatorname{Nul}(A+I) = \dim \operatorname{Nul} \begin{pmatrix} 3 & 3 & 3 & -6 \\ 0 & -2 & 1 & 1 \\ -1 & -3 & 0 & 3 \\ 1 & -1 & 2 & -1 \end{pmatrix}.$$

The reduced echelon form of $\left(\begin{array}{c|c} A+I & \mathbf{0} \end{array} \right)$ is

$$\begin{pmatrix} 1 & 0 & \frac{3}{2} & -\frac{3}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

showing that $d_1 = 2$. Next, we compute that

$$d_2 = \dim \operatorname{Nul}(A+I)^2 = \dim \operatorname{Nul} 0 = \dim \mathbb{R}^4 = 4 = m.$$

The square diagram for $\lambda = -1$ is therefore

$$\begin{array}{cc} \boxed{\phantom{\mathbf{e}_1}} & \boxed{\phantom{\mathbf{e}_2}} \\ \boxed{\phantom{\mathbf{a}}} & \boxed{\phantom{\mathbf{b}}} \end{array}$$

This already tells us that the Jordan canonical form of A is

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

To compute a Jordan basis, we note that the largest Jordan block has size $r = 2$. The first step is to compute a basis of

$$\operatorname{Nul}(A+I) \cap \operatorname{Col}(A+I)^{r-1} = \operatorname{Nul}(A+I) \cap \operatorname{Col}(A+I).$$

The latter is precisely the image of the generalised eigenspace $\operatorname{Nul}(A+I)^4 = \mathbb{R}^4$ under left multiplication by $A+I$. In other words, we are searching for a basis of $\operatorname{Col}(A+I)$. From our computation for the null space of $A+I$, we see that the columns

$$\mathbf{a} = \begin{pmatrix} 3 \\ 0 \\ -1 \\ 1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{b} = \begin{pmatrix} 3 \\ -2 \\ -3 \\ -1 \end{pmatrix}$$

in $A+I$ form a basis of $\operatorname{Col}(A+I)$. (They correspond to the basic variables in the system of linear equations $(A+I)\mathbf{x} = \mathbf{0}$.) Since $A+I$ represents the linear transformation

$$T + \mathrm{id}_{\mathbb{R}^4} \colon \mathbb{R}^4 \longrightarrow \mathbb{R}^4, \qquad \mathbf{v} \mapsto (A+I)\mathbf{v}$$

with respect to the standard basis, we see that $(A+I)\mathbf{e}_1 = \mathbf{a}$ and $(A+I)\mathbf{e}_2 = \mathbf{b}$. We may therefore fill in the Jordan basis diagram with

$$\begin{array}{cc} \boxed{\mathbf{e}_1} & \boxed{\mathbf{e}_2} \\ \boxed{\mathbf{a}} & \boxed{\mathbf{b}} \end{array}$$

It follows that

$$\mathcal{B} = \{\mathbf{a}, \mathbf{e}_1, \mathbf{b}, \mathbf{e}_2\}$$

is a Jordan basis. Consequently, the matrix

$$P = \begin{pmatrix} 3 & 1 & 3 & 0 \\ 0 & 0 & -2 & 1 \\ -1 & 0 & -3 & 0 \\ 1 & 0 & -1 & 0 \end{pmatrix}$$

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