13. Canonical forms: I

In the proof of rank-nullity, the following commutative diagram played a key role:

$$\begin{tikzcd} V \arrow[r, "\phi"] \arrow[d, "f"'] & W \arrow[d, "g"] \\ F^n \arrow[r, "\psi"] & F^m. \end{tikzcd}$$

Here commutative refers to the fact that there are two ways to navigate from the top left hand corner to the bottom right hand corner and either way gives the same function. We could either go down and across or we could go across and then down. Thus commutativity means

$$\psi \circ f = g \circ \phi.$$

As f is invertible, this is equivalent to requiring,

$$\psi = g \circ \phi \circ f^{-1},$$

which is what we had before. Now $\psi \colon F^n \longrightarrow F^m$ is a linear map and so corresponds to a matrix $A \in M_{n,n}(F)$. The interesting thing is that we get to choose both f and g. Presumably as we vary f and g, A might become more complicated or less complicated. So what is the optimal choice of f and g?

Theorem 13.1. *Let $\phi \colon V \longrightarrow W$ be a linear map between two vector spaces of dimension n and m.*

Then we may find two linear isomorphisms $f \colon V \longrightarrow F^n$ and $g \colon W \longrightarrow F^m$ such that the resulting linear map $\psi \colon F^n \longrightarrow F^m$ is given by

$$(r_1, r_2, \dots, r_n) \longrightarrow (r_1, r_2, \dots, r_k, 0, 0, \dots, 0),$$

where k is the rank of ϕ (here there are as many zeroes as it takes to fill up $m - k$ entries).

Example 13.2. *Let $\phi \colon V \longrightarrow W$ be a linear map between two vector spaces of dimension* 5 *and* 3*, of rank* 2*.* (13.1) *says that we can choose f and g so that*

$$\psi \colon F^5 \longrightarrow F^3,$$

is given by

$$\phi(r_1, r_2, r_3, r_4, r_5) = (r_1, r_2, 0).$$

The image is the span of $f_1 = (1, 0, 0)$ and $f_2 = (0, 1, 0)$. The kernel is the span of $e_3 = (0, 0, 1, 0, 0)$, $e_4 = (0, 0, 0, 1, 0)$ and $e_5 = (0, 0, 0, 0, 1)$. So the rank is two (as predicted) and the nullity is three. In fact (13.1) *gives an easy way to prove rank-nullity, since the image is spanned by the first k vectors and the kernel is spanned by the last $n-k$ vectors.*

To give a clue as to how we are going to prove (13.1)*, let us consider the corresponding matrix,* $A \in M_{3,5}(F)$*,*

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

This looks suspiciously like the endproduct of Gauss Jordan elimination.

Proof of (13.1). Pick any two linear isomorphisms $f_1 \colon V \longrightarrow F^n$ and $g_1 \colon W \longrightarrow F^m$. Let

$$\psi_1 = g_1 \circ \phi \circ f_1^{-1} \colon F^n \longrightarrow F^m.$$

Then ψ_1 is linear and so there is a matrix A_1 such that $\psi(v) = A_1 v$. Apply Gauss Jordan elimination to A_1. We end up with a matrix U in reduced row echelon form. As in the proof of the PLU decomposition, we can encode the steps of the elimination in a single matrix E, the product of all the elementary matrices, corresponding to all of the elementary row operations. It follows that $EA_1 = U$. Let $g_2 \colon F^m \longrightarrow F^m$ be the linear map given by E, so that $g_2(w) = Ew$. Let

$$\psi' = g_2 \circ g_1 \circ \phi \circ f_1^{-1} \colon F^n \longrightarrow F^m$$

Then the matrix of ψ' is $EA_1 = U$.

We are not done yet, but we are close. Now we want to apply some elementary column operations to U. One way to say this is that we want to apply elementary row operations to U^t, the transpose of U. If we apply Gauss Jordan elimination to U^t we end up with a matrix U_1 in the correct form. In this way we find a matrix $E' \in M_{n,n}(F)$ (U^t has n rows after all) so that

$$E'U^t = U_1$$

Taking transposes we get

$$A = (U_1)^t = U(E')^t.$$

Now A also has the correct form (meaning an unbroken line of 1's on the 'diagonal' and zeroes everywhere else). The matrix E' is invertible and so is the matrix $(E')^t$. Let $f_2 \colon F^n \longrightarrow F^n$ be the inverse map. Let

$$\psi = g_2 \circ g_1 \circ \phi \circ f_1^{-1} \circ f_2^{-1} \colon F^n \longrightarrow F^m.$$

Then A is the matrix associated to ψ''. So we want the linear isomorphisms

$$f = f_2 \circ f_1 \colon V \longrightarrow F^n \qquad \text{and} \qquad g = g_2 \circ g_1 \colon W \longrightarrow F^m. \qquad \square$$

It is quite instructive to think about the geometric meaning of the proof of (13.1). Suppose we are given a map from $\phi\colon F^n \longrightarrow F^m$. This gives us a matrix A. The image is some random subspace of F^m. We first find a linear isomorphism $g\colon F^m \longrightarrow F^m$ which maps the image to the subspace spanned by the first k standard vectors. Then we look at the kernel of A. We can find a linear isomorphism $f\colon F^n \longrightarrow F^n$ so that the kernel becomes the subspace spanned by the last $n-k$ vectors.

It is still not quite the case that this forces the resulting map to have such a straightforward shape as the one given by (13.1), but it gives an idea of how (13.1) works geometrically.

In fact it is interesting to reformulate (13.1) in terms of matrices:

Corollary 13.3. *Let $A \in M_{m,n}(F)$.*

Then we may find invertible matrices $B \in M_{m,m}(F)$ and $C \in M_{n,n}(F)$ such that $U = BAC \in M_{m,n}(F)$ is a matrix whose only non-zero entries are $u_{ii} = 1$ for $1 \leq i \leq k$.

Proof. Let $\phi\colon F^n \longrightarrow F^m$ be the linear transformation given by A. Then by (13.1) we may find linear isomorphisms $f\colon F^n \longrightarrow F^n$ and $g\colon F^m \longrightarrow F^m$ such that $\psi = g \circ \phi \circ f^{-1}\colon F^n \longrightarrow F^m$ has the standard form given by (13.1). Let U be the matrix corresponding to ψ. Then U has the correct form. Let B and C be the matrices corresponding to g and f^{-1}. As

$$\psi = g \circ \phi \circ f^{-1} \qquad \text{we have} \qquad U = BAC. \qquad \square$$

So this is one possible canonical form for a matrix. However suppose we start with a square matrix. In fact suppose we start with a linear transformation $\phi\colon V \longrightarrow V$. Now could choose two different linear isomorphisms f and g as before. But it is more natural to fix only one, that is to put $g = f$. The relevant commutative diagram is then

$$\begin{tikzcd} V \arrow[r, "\phi"] \arrow[d, "f"'] & V \arrow[d, "f"] \\ F^n \arrow[r, "\psi"] & F^n. \end{tikzcd}$$

Here $\psi = f \circ \phi \circ f^{-1}$. At the level of matrices we have $B = PAP^{-1}$, where A is the matrix corresponding to ϕ (for this to make sense, we have to assume that $V = F^n$, of course), B is the matrix corresponding to ψ and P is the matrix corresponding to f. In other words, at the level of matrices we are back to the old problem of looking at similar matrices. This suggests:

Definition-Lemma 13.4. *Let $\phi\colon V \longrightarrow V$ be a linear function. We say that a non-zero vector $v \in V$ is an* ***eigenvector*** *with* ***eigenvalue*** *λ if $\phi(v) = \lambda v$.*

$$E_\lambda(\phi) = \{\, v \in V \mid \phi(v) = \lambda v \,\},$$

is a subspace of V, called an ***eigenspace****.*

Proof. It suffices to prove that $E_\lambda(A)$ is a subspace of V. But this is immediate, as

$$E_\lambda(\phi) = \operatorname{Ker}(\phi - \lambda I),$$

where I is the identity function $I\colon V \longrightarrow V$. □

Theorem 13.5. *Let $\phi\colon V \longrightarrow V$ be a linear function.*

If ϕ has a basis of eigenvectors $v_1, v_2, \dots, v_n$ then there is a linear isomorphism $f\colon V \longrightarrow F^n$ such that $\psi = f \circ \phi \circ f\colon F^n \longrightarrow F^n$ is the linear map $(r_1, r_2, \dots, r_n) \longrightarrow (\lambda_1 r_1, \lambda_2 r_2, \dots, \lambda_n r_n)$. Here $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues and f is given by the basis $v_1, v_2, \dots, v_n$.

Furthermore ψ is unique up to reordering the eigenvalues.

Proof. Let f be given by the basis of eigenvectors $v_1, v_2, \dots, v_n$. Then v_i is sent to e_i. Therefore

$$\begin{aligned}\psi(e_i) &= f(\phi(v_i)) \\ &= f(\lambda_i v_i) \\ &= \lambda_i f(v_i) \\ &= \lambda_i e_i.\end{aligned}$$

Therefore ψ has the given form.

Let D be the matrix corresponding to ψ. Then D is a diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \dots, \lambda_n$. But the only eigenvalues of a diagonal matrix are the diagonal entries. So the eigenvalues are unique. □

We already know that this is not the complete answer to the canonical form of linear functions. The problem is that there are matrices which are not diagonalisable. The corresponding functions cannot be put into the form given in (13.5). However we do have:

Theorem 13.6. *Let $\phi\colon V \longrightarrow V$ be a linear function and let $v_1, v_2, \dots, v_k$ be eigenvectors with distinct eigenvalues.*

Then $v_1, v_2, \dots, v_k$ are independent. In particular if $k = \dim V$ then ϕ has a basis of eigenvectors.

Proof. There are two obvious ways to prove this. Either choose an isomorphism $f\colon V \longrightarrow V$, let $\psi = f \circ \phi \circ f^{-1}\colon F^n \longrightarrow F^n$ and use the fact that the matrix A corresponding to ψ has k eigenvectors $w_i = f(v_i)$

with the same distinct eigenvalues, or simply run the same argument as before to the function ϕ rather than the matrix A. $\square$