5 Some bases are better than others

Given a vector space V of dimension n over a field $\mathbb{F}$, and a basis $\mathcal{B}$ of V, and a linear transformation $T : V \to V$, we can write the $\mathcal{B}$-matrix of T; its columns are the $\mathcal{B}$-coordinate vectors of the images of the basis elements under T. Obviously, different choices of basis determine different matrices, and the following questions have (long and) interesting answers.

1. Is there a basis with respect to which the matrix of T has a "nice" form (e.g. diagonal, upper or power triangular, etc)

2. Given a pair of $n \times n$ matrices over $\mathbb{F}$, how can we decide whether they represent the same linear transformation or not?

If v is an eigenvector of T, i.e. $T(v) = \lambda v$ for some $\lambda \in \mathbb{F}$, and if $\mathcal{B}$ is a basis of V whose jth element is v, then Column j of the $\mathcal{B}$-matrix of T has λ as its jth entry and is otherwise full of zeros. If *every* element of the basis $\mathcal{B}$ is an eigenvector of T, then the $\mathcal{B}$-matrix of T is a diagonal matrix, and the converse is also true.

Theorem 5.1. *The linear transformation $T : V \to V$ can be represented by a diagonal matrix with entries in V if and only if V has a basis consisting of eigenvectors of T. In this case the diagonal entries are the eigenvalues to which these basis elements correspond as eigenvectors.*

We say that T is *diagonalizable* (over $\mathbb{F}$) in this case.
Two remarks on this theorem:

1. It is possible for T not to be diagonalizable over $\mathbb{F}$ but to be diagonalizable over some extension of $\mathbb{F}$. For example let T be a rotation through θ of $\mathbb{R}^2$, where $\theta \notin \{0, \pi\}$. Then we have seen that T has no eigenvector in $\mathbb{R}^2$ and hence there is no choice of basis of $\mathbb{R}^2$ for which T is represented by a diagonal matrix.

   With respect to the standard basis $\{(1,0),(0,1)\}$ of $\mathbb{R}^2$, the matrix of T is $A_T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. We can extend T to the linear transformation of $\mathbb{C}^2$ that multiplies every column vector with complex entries on the left by A_T. Then we can observe that $\begin{pmatrix} 1 \\ i \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -i \end{pmatrix}$ are eigenvectors, with corresponding eigenvalues $\cos\theta - i\sin\theta$ and $\cos\theta + i\sin\theta$ respectively. Thus over the field of complex numbers, T is represented by the diagonal matrix

$$\begin{pmatrix} \cos\theta - i\sin\theta & 0 \\ 0 & \cos\theta + i\sin\theta \end{pmatrix}.$$

2. It is possible for T not to be diagonalizable over $\mathbb{F}$ or over any extension of $\mathbb{F}$. Let T be the linear transformation of $\mathbb{R}^2$ defined by

$$(x, y) \to (x, x + y).$$

   It is easily observed that T is a linear transformation. If (x, y) is an eigenvector of T corresponding to the eigenvalue λ, then

$$(x, x + y) = \lambda(x, y) \Longrightarrow x = \lambda x,\ x + y = \lambda y.$$

   These equations are simultaneously satisfied only if $x = 0$ and $\lambda = 1$, which means that the eigenvectors of T are the (non-zero) points of the line $x = 0$, they form only a one-dimensional subspace. Thus $\mathbb{R}^2$ does not have a basis consisting of eigenvectors of T and T is not diagonalizable over $\mathbb{R}$. In this case, interpreting T as a linear transformation of $\mathbb{C}^2$ does not help; there is still only a single line of eigenvectors. The formula $(x, y) \to (x, x + y)$ does not define a diagonalizable linear transformation over any field.

Remark: The vector v is an eigenvector of the linear transformation corresponding to the eigenvalue λ if and only if for every basis $\mathcal{B}$ of V the matrix equation

$$A_{\mathcal{B}}\,[v]_{\mathcal{B}} = \lambda[v]_{\mathcal{B}},$$

where $A_{\mathcal{B}}$ denotes the matrix of T with respect to $\mathcal{B}$. In particular, λ is an eigenvalue of $A_{\mathcal{B}}$ for every choice of $\mathcal{B}$, and matrices representing the same linear transformation over different bases have the same eigenvalues.

Theorem 5.2. *Let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of T in $\mathbb{F}$, with corresponding eigenvectors $v_1, \dots, v_k$. Then $\{v_1, \dots, v_k\}$ is a linearly independent set in V.*

Proof. Suppose not, and let

$$a_1v_1 + \cdots + a_kv_k = 0$$

be a non-trvial expression for 0 as a linear combination of the v_i with as few non-zero coefficients as possible. We can assume (after reordering if necessary) that $a_1 \neq 0$. Note that at least one other a_i is also non-zero (since $v_1 \neq 0$). In the following the first line is obtained by multiplying the above expression by λ_1, and the second by applying T. We have

$$\begin{aligned} a_1\lambda_1v_1 + a_2\lambda_1v_2 + \cdots + a_k\lambda_1v_k &= 0 \\ a_1\lambda_1v_1 + a_2\lambda_2v_2 + \cdots + a_k\lambda_kv_k &= 0 \end{aligned}$$

Subtracting the second equation here from the first gives

$$a_2(\lambda_1 - \lambda_2)v_2 + \cdots + a_k(\lambda_1 - \lambda_k)v_k = 0.$$

Since the expressions $\lambda_1 - \lambda_i$ are non-zero for $i > 1$, at least one of the coefficients in this new expression is not zero, and this is a shorter expression for zero as a non-trivial linear combination of the v_i than the original one. This contradiction yields the conclusion that the v_i are linearly independent. □

Notes Two consequences of this theorem:

1. The number of distinct eigenvalues of a linear transformation $T : V \to V$ cannot exceed the dimension n of V (since a set of more than n vectors cannot be linearly independent). Note that we can deduce this without having to think about the characteristic polynomial. The same comment applies to any $n \times n$ matrix, since they represent linear transformations.

2. If V has dimension n and the linear transformation $T : V \to V$ has n distinct eigenvalues, then V has a basis consisting of eigenvectors of T and T is diagonalizable.

We conclude this section by considering the "matrix" meaning of diagonalizability. Let A be the matrix that represents the linear transformation $T : V \to V$ with respect to the basis $\mathcal{B}$ of V, and let $\mathcal{B}'$ be another basis of V. Let P be the matrix whose jth column is the $\mathcal{B}$-coordinate vector of the jth element of the basis $\mathcal{B}'$. Then, using the definition of matrix-vector multiplication we can observe that for any vector in V,

$$[v]_{\mathcal{B}} = P[v]_{\mathcal{B}'}.$$

We refer to P as the *change of basis* matrix from $\mathcal{B}'$ to $\mathcal{B}$. Note that P must be invertible since there are n linearly independent vectors in its columnspace (namely the $\mathcal{B}$-coordinate vectors of the elements of the basis $\mathcal{B}'$). Moreover, rearranging the above equation gives

$$[v]_{\mathcal{B}'} = P^{-1}[v]_{\mathcal{B}},$$

so the change of basis matrices from $\mathcal{B}'$ to $\mathcal{B}$ and from $\mathcal{B}$ to $\mathcal{B}'$ are inverses of each other. We now note how the matrix of T with respect to $\mathcal{B}'$ depends on A and on P. To determine this we must ask, for a vector $v \in V$, by what matrix should we multiply $[v]_{\mathcal{B}'}$ in order to obtain $[T(v)]'_{\mathcal{B}}$. We have

$$\begin{aligned} [v]_{\mathcal{B}} &= P[v]_{\mathcal{B}'} \\ \Longrightarrow [T(v)]_{\mathcal{B}} &= AP[v]_{\mathcal{B}'} \\ \Longrightarrow [T(v)]'_{\mathcal{B}} &= P^{-1}APv_{\mathcal{B}'} \end{aligned}$$

Thus the matrix of T with respect to $\mathcal{B}'$ is $P^{-1}AP$.

Definition 5.3. *Two matrices* A *and* B *in* $M_n(\mathbb{F})$ *are said to be* similar *(over* $\mathbb{F}$*) if there exists a nonsingular matrix* $P \in M_n(\mathbb{F})$ *for which*

$$B = P^{-1}AP.$$

Similarity is a very important relation on the set of square matrices of given size over a field. If two matrices are similar it means that they represent the same linear transformation with respect to different choices of basis. Similarity is an equivalence relation on $M_n(\mathbb{F})$ (something for you to check). A matrix A is said to *diagonalizable* if it is similar to a diagonal matrix, which means that there exists a basis of $\mathbb{F}^n$ consisting of eigenvectors of A, or equivalently that the linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^n$ defined as left multiplication by A is diagonalizable.