Linear independence of eigenvectors to distinct eigenvalues

Theorem 1 below is a special case of FIS Theorem 5.5 that isolates the most important principle behind the FIS theorem; in some sense Theorem 1 is the more important of these two theorems, even though it is less general. Of the two theorems, Theorem 1 is easier to state and easier to understand, and its proof is easier to follow.

After the proof of Theorem 1 there are some comments and two corollaries. Corollary 1 isolates another simple, important principle; Corollary 2 is the full-blown FIS Theorem 5.5. Although this route to a proof of Theorem 5.5 is longer than the one in the book, it's a bit more informative, and I think you'll find it easier to follow.

Theorem 1 ("Eigenvectors to different eigenvalues are linearly independent") *Let* T *be a linear operator on a vector space* V*. Suppose that* $v_1, \ldots, v_k$ *are eigenvectors of* T *corresponding to* **distinct** *eigenvalues* $\lambda_1, \ldots, \lambda_k$ *respectively.* (Remember that "distinct" means $\lambda_i \neq \lambda_j$ whenever $i \neq j$.) *Then*

$$\text{the list } v_1, \ldots, v_k \text{ is linearly independent.} \tag{1}$$

Proof: We proceed by induction on k.

First suppose $k = 1$. Let v_1 be an eigenvector of T and let λ_1 be the corresponding eigenvalue (the distinctness criterion is met vacuously). By definition, eigenvectors are never the zero vector, so $\{v_1, \ldots, v_k\} = \{v_1\}$ is linearly independent.

Now let $m \in \mathbf{N}$, and assume that the assertion in the theorem holds for $k = m$. [Students: I recommend that while reading the argument below, you carry it out explicitly for the case $m = 1$ on a separate piece of paper.] Suppose that $v_1, \ldots, v_{m+1}$ are eigenvectors of T corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_{m+1}$ respectively. Let $a_1, \ldots, a_{m+1} \in \mathbf{R}$ be such that

$$a_1v_1 + \cdots + \cdots + a_iv_i + \ldots a_mv_m + a_{m+1}v_{m+1} = 0_V \ . \tag{2}$$

Applying T to both sides of equation (2), and using both the linearity of T and the hypothesis that $\mathsf{T}(v_i) = \lambda_iv_i$ for $1 \leq i \leq m+1$, we obtain

$$a_1\lambda_1v_1 + \cdots + a_i\lambda_iv_i + \cdots + a_m\lambda_mv_m + a_{m+1}\lambda_{m+1}v_{m+1} = \mathsf{T}(0_V) = 0_V \ . \tag{3}$$

But multiplying both sides of equation (2) by λ_{m+1} yields

$$a_1\lambda_{m+1}v_1 + \cdots + a_i\lambda_{m+1}v_i + \cdots + a_m\lambda_{m+1}v_m + a_{m+1}\lambda_{m+1}v_{m+1} = \mathsf{T}(0_V) = 0_V \ , \tag{4}$$

so, subtracting equation (4) from equation (3), we find that

$$a_1(\lambda_1 - \lambda_{m+1})v_1 + \cdots + a_i(\lambda_i - \lambda_{m+1})v_i + \cdots + a_m(\lambda_m - \lambda_{m+1})v_m = 0_V \ . \tag{5}$$

Since $v_1, \ldots, v_m$ are eigenvectors of T corresponding to distinct eigenvalues, our inductive hypothesis guarantees that the list $v_1, \ldots, v_m$ is linearly independent. Hence, equation (5) implies that $a_i(\lambda_i - \lambda_{m+1}) = 0$ for each $i \in \{1, \ldots, m\}$. But for each such i, the assumed distinctness of the eigenvalues implies that $\lambda_i - \lambda_{m+1} \neq 0$, and hence that $a_i = 0$.

Therefore $a_i = 0$ for $1 \leq i \leq m$, which simplifies equation (2) to $a_{m+1}v_{m+1} = 0_V$. But $v_{m+1} \neq 0_V$ since v_{m+1} is an eigenvector. Hence $a_{m+1} = 0$.

Thus, equation (2) holds only when $a_i = 0$ each $i \in \{1, \ldots, m+1\}$. Therefore the list $v_1, \ldots, v_{m+1}$ is linearly independent.

By induction, the assertion in the theorem holds for all $k \geq 1$. ■

Some comments:

1. Since no eigenvector can *ever* correspond to two different eigenvalues (not just in the setting of this theorem), the theorem's assumption that $\lambda_i \neq \lambda_j$ for $i \neq j$ guarantees that $v_i \neq v_j$ for $i \neq j$. Thus, the terms of the list $v_1, \ldots, v_k$ in the theorem form a k-element set, and in place of statement (1) I could have said that the *set* $\{v_1, v_2, \ldots, v_k\}$ is linearly independent (as I did on the homework page). The reason I used the "list" wording in this handout's statement of the theorem was to avoid having to say most of the preceding in the proof itself, which would have distracted from the main idea of the proof.

   Something worth keeping in mind is that no matter what the context (i.e. regardless of whatever hypotheses are in effect), saying that a *list* of vectors $v_1, \ldots, v_k$ is linearly independent is *never weaker* than saying that the *set* $\{v_1, \ldots, v_k\}$ is linearly independent, and is sometimes stronger. (It's stronger when the terms of the list are *not* all distinct, as discussed in the "Lists ..." handout.)

2. As a reminder: observe that between equations (2) and (3), equations (3) and (4), and equations (4) and (5), I put *words* telling the reader the *argument.* As discussed in Assignment 0 reading, "equation equation equation equation" is not an *argument.* A proof is a logical *argument.*

   In this proof, I put *more* words between consecutive equations than I really had to. I had the luxury of not being under time-pressure, and, as a teacher, I try to write proofs that every student in the class should be able to follow without a struggle. The bare minimum that a *student* needs to put between equations (or before the first equation) in a proof is an appropriate *logical connector*, a word or phrase that gives the logical relation between equations (or between the first equation and whatever preceded it). E.g. between equations (3) and (4) I could simply have said "Equation (2) implies".

   But, between equations (3) and (4), I could *not* have simply said "*This* implies", since the default meaning of "This" would have been the previous equation, (3) (which does *not* imply equation (4)). **Use pronouns sparingly, if at all, in a proof**; there is too much danger of ambiguity, or of the default meaning of your pronoun (according to standard rules of syntax) being the wrong noun, or of your pronoun having no antecedent at all. Although the *most* dangerous pronoun tends to be "it", almost every pronoun you might be tempted to use in a proof is potentially dangerous. (The one exception is "we". A personal pronoun, "we" has no danger of being misinterpreted to stand for a statement or a mathematical object!) If you look at my proofs, you'll rarely see a pronoun other than "we".

3. Whenever I give you a handout, I assume that you've read all the handouts I've previously assigned you to read—and that if you've forgotten anything that I addressed in a particular handout, you'll know that you should look at that handout again to refresh your memory. In particular, this applies if you're not sure **exactly** what the term "inductive hypothesis" means, or if my proof does not contain certain terminology you expected to see in a proof-by-induction.

Corollary 1 *Let* T *be a linear operator on a vector space* V*. Suppose that* $\lambda_1, \ldots, \lambda_k$ *are* distinct *eigenvalues of* T*, and for each* $i \in \{1, \ldots, k\}$*, let* $v_i \in E_{\lambda_i}(\mathsf{T})$ *(the* λ_i*-eigenspace of* T*). Suppose that* $v_1 + v_2 + \cdots + v_k = 0_V$*. Then* $v_i = 0_V$ *for each* i*.*

Proof: Assume that $v_i \neq 0_V$ for m values of i, where $m \geq 1$. Reordering the v_i if necessary, we may assume that $v_i \neq 0_V$ whenever $i \leq m$ and (if $m < k$) that $v_i = 0_V$ whenever $i > m$. Then, by definition of *eigenspace*, for each $i \in \{1, \ldots, m\}$, the vector v_i is an eigenvector of T with eigenvalue λ_i. But since $v_i = 0_V$ for $i > m$,

$$v_1 + \cdots + v_m = v_1 + \cdots + v_m + \cdots + v_k = 0_V,$$

so the set $\{v_1, \ldots, v_m\}$ is linearly dependent. But by Theorem 1, this is impossible.

Hence $v_i = 0_V$ for each $i \in \{1, \ldots, k\}$. ■

Corollary 2 (FIS Theorem 5.5) *Let* T *be a linear operator on a vector space* V*. Suppose that* $\lambda_1, \ldots, \lambda_k$ *are* distinct *eigenvalues of* T*, and for each* $i \in \{1, \ldots, k\}$ *let* S_i *be a linearly independent set of eigenvectors of* T *with eigenvalue* λ_i*. Then* $S_1 \cup \cdots \cup S_k$ *is linearly independent.*

Proof: For each $i \in \{1, \ldots, k\}$, let $m_i = |S_i|$. We may restrict attention to the case in which $m_i \geq 1$ $(S_i \neq \emptyset)$ for each i.

For each $i \in \{1, \ldots, k\}$, let $m_i = |S_i|$. Since the eigenvalue associated to any eigenvector is unique, $S_i \cap S_j = \emptyset$ for $i \neq j$, so $|S| = m_1 + \cdots + m_k$. Let $N = m_1 + \cdots + m_k$ and let $S = S_1 \cup \cdots \cup S_k$. Enumerate the elements of S as $v_1, \ldots, v_N$, where the first m_1 vectors are the elements of S_1, the next m_2 are the elements of S_2, etc.

Suppose $c_1, \ldots, c_N \in \mathbf{R}$ are such that $\sum_{j=1}^{N} c_j v_j = 0_V$. Let w_1 be the sum of the first m_1 terms of $\sum_{j=1}^{N} c_j v_j$, let w_2 be the sum of the next m_2 terms, etc., with w_k being the sum of the last m_k terms. Then $w_i \in E_{\lambda_i}(\mathsf{T})$ for each $i \in \{1, \ldots, k\}$ (since eigenspaces are subspaces), and $w_1 + \dots w_k = 0_V$. From Corollary 1, it follows that $w_i = 0_V$ for each $i \in \{1, \dots, k\}$. For each such i, since S_i is linearly independent, we must have $c_j = 0$ for each j in the index-range corresponding to the elements of S_i. Hence $c_j = 0$ for each $j \in \{1, \dots, N\}$.

Thus S is linearly independent. $\blacksquare$