Jordan canonical form

Easy generalizations of the Hermitian spectral theorem

Definition: $[\underline{A},\underline{B}] \equiv \underline{A}\,\underline{B} - \underline{B}\,\underline{A} \equiv$ *commutator* of $\underline{A}$ and $\underline{B}$.

NOTE: $[\underline{A},\underline{B}] = -[\underline{B},\underline{A}]$.

Theorem. *If $\underline{A}$ and $\underline{B}$ are Hermitian and $\underline{A}\,\underline{B} = \underline{B}\,\underline{A}$ (or, $[\underline{A},\underline{B}] = \underline{0}$), then $\underline{A}$ and $\underline{B}$ can be* **simultaneously** *diagonalized (by a unitary matrix). That is, $\exists$ an ON basis whose elements are eigenvectors of both $\underline{A}$ and $\underline{B}$.*

PROOF: Let $\mathcal{V} = \mathcal{V}(\lambda_1) \oplus \cdots \oplus \mathcal{V}(\lambda_L)$ be the decomposition into eigenspaces of $\underline{A}$. Given $\vec{u} \in \mathcal{V}(\lambda_j)$, we have $\underline{A}\,\underline{B}\vec{u} = \underline{B}\,\underline{A}\vec{u} = \lambda_j \underline{B}\vec{u}$, so $\underline{B}\vec{u} \in \mathcal{V}(\lambda_j)$. Thus this decomposition reduces $\underline{B}$ as well as $\underline{A}$. (In other words, $\underline{B}$ has a block-diagonal form relative to any eigenbasis for $\underline{A}$, with the sizes of the blocks equal to the $\dim \mathcal{V}(\lambda_j)$'s.) If $\underline{A}$ has simple eigenvalues ($\dim \mathcal{V}(\lambda_j) = 1,\ \forall j$), we are done. Otherwise, note that $\underline{B}$ defines a mapping $\underline{B}_j \colon \mathcal{V}(\lambda_j) \to \mathcal{V}(\lambda_j)$, which is *still Hermitian*. (This is most easily seen by inspection of the block-diagonal matrix with respect to an ON basis.) $\underline{B}_j$ can be diagonalized by passing to an ON basis of eigenvectors in $\mathcal{V}(\lambda_j)$. These new basis vectors still satisfy $\underline{A}\vec{v} = \lambda_j \vec{v}$, so both $\underline{A}$ and $\underline{B}$ are now diagonalized.

We've seen that for an Hermitian operator, (a) eigenvalues are real, (b) eigenvectors are (or can be chosen) ON. Many operators are diagonalizable without satisfying these conditions. Let's consider relaxing them in turn.

If there is a *basis* of *ON* eigenvectors but not all eigenvalues are real, then

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \end{pmatrix} = \begin{pmatrix} \operatorname{Re}\lambda_1 & & 0 \\ & \operatorname{Re}\lambda_2 & \\ 0 & & \ddots \end{pmatrix} + i \begin{pmatrix} \operatorname{Im}\lambda_1 & & 0 \\ & \operatorname{Im}\lambda_2 & \\ 0 & & \ddots \end{pmatrix}$$

shows that $\underline{A} = \underline{B} + i\underline{C}$, where $\underline{B}$ and $\underline{C}$ are Hermitian. Thus $\underline{A}^* = \underline{B} - i\underline{C}$. Note also that $[\underline{B},\underline{C}] = \underline{0}$; hence $[\underline{A},\underline{A}^*] = \underline{0}$. Conversely, if $[\underline{A},\underline{A}^*] = \underline{0}$, then

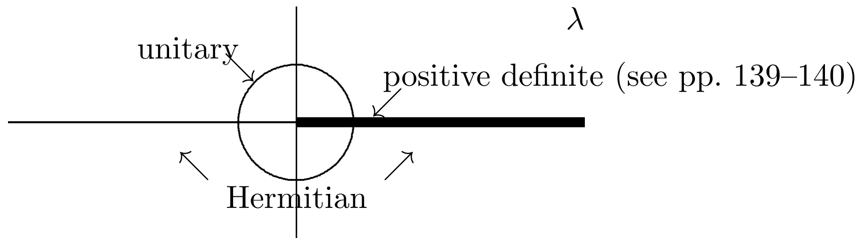
$$\underline{B} \equiv \frac{\underline{A} + \underline{A}^*}{2}, \qquad \underline{C} \equiv \frac{\underline{A} - \underline{A}^*}{2i}$$

are Hermitian and satisfy $[\underline{B},\underline{C}] = \underline{0}$. Therefore, they can be simultaneously diagonalized by a unitary matrix, and we recover the situation we considered first. To make this discussion into a theorem, all we need is a definition to start out from:

Definition: $\underline{A}$ is *normal* if $[\underline{A},\underline{A}^*] = \underline{0}$.

Theorem. $\underline{A}$ *normal* $\iff$ $\underline{A}$ *has an ON basis of eigenvectors (with eigenvalues not necessarily real).*

Special case: $\underline{A}$ *unitary* $\Rightarrow$ $\underline{A}$ *normal;* $\underline{A}$ *unitary* $\iff$ $\underline{A}$ *has an ON basis of eigenvectors and all eigenvalues of* $\underline{A}$ *satisfy* $|\lambda_j| = 1$.



Now consider relaxing (b): The following are easily seen to be equivalent:

1. $\underline{A}$ has a basis of eigenvectors, which *cannot* be made orthonormal (or even orthogonal).

2. $\underline{A}$ can be diagonalized, but by a necessarily *nonunitary* matrix.

3. $\underline{A} = \sum_{\nu=1}^{L} \lambda_\nu \underline{P}_\nu$, where the $\underline{P}_\nu$ are projection operators ($\underline{P}_\nu{}^2 = \underline{P}_\nu$) satisfying the conditions associated with a direct sum,

$$\sum_{\nu=1}^{L} \underline{P}_\nu = \underline{1}, \qquad \underline{P}_\nu \underline{P}_\mu = \underline{0} \text{ if } \mu \neq \nu,$$

but *not* the orthogonality condition, $\underline{P}_\nu^* = \underline{P}_\nu$.

Constructing eigenprojections

This is a reasonable spot to point out how to write down an eigenprojection matrix in terms of the corresponding eigenvector(s).

Suppose first that $\underline{A}$ is Hermitian and that λ_ν is a simple (nondegenerate) eigenvalue. Let $\vec{u}_\nu$ be a normalized eigenvector. Thus $\underline{P}_\nu \vec{v}$ is a multiple of $\vec{u}_\nu$, for any $\vec{v} \in \mathcal{V}$. The coefficient of proportionality is simply $\vec{v} \cdot \vec{u}_\nu$. (If this is not clear, reread the section of your notes on orthogonality, the projection theorem, the Gram–Schmidt theorem, etc., and also the sections on projection operators and on orthogonal projections.) Therefore, if $\vec{u}_\nu$ is represented with respect to a basis by $(u^1, u^2, \ldots)$, then the matrix representing $\underline{P}_\nu$ is P,

$$P^j{}_k \equiv u^j \, \overline{u^k}.$$

Example: $\qquad \vec{u}_\nu = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ i \end{pmatrix}, \qquad P = \frac{1}{5}\begin{pmatrix} 4 & -2i \\ 2i & 1 \end{pmatrix}.$

(Cf. (28.16) of Bowen & Wang.)

The generalization to the case of a multiple (degenerate) eigenvalue is this: Let $\{\vec{u}_1, \vec{u}_2, \ldots\}$ be an orthonormal basis for $\mathcal{V}(\lambda_\nu)$. Then $\underline{P}_\nu$ is the sum of the one-dimensional projections onto the individual $\vec{u}$'s, so its matrix is

$$P^j{}_k \equiv \sum_{l=1}^{\dim \mathcal{V}(\lambda_\nu)} u_l{}^j \, \overline{u_l{}^k}. \tag{$*$}$$

(Again, this should be clear from our treatment of orthogonality-related matters, together with the definition of $\underline{P}_\nu$ as the projection onto $\mathcal{V}(\lambda_\nu)$ along its orthogonal complement.)

This sort of construction of an orthogonal projection is used much more often in practice than "Sylvester's formula", emphasized in the textbook (28.18). Still another method is to take the diagonal matrix representing the projection with respect to an eigenbasis — each of its diagonal elements is either 0 or 1 — and apply to it the same similarity transformation that converts D to A:

$$P = U \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 0 \end{pmatrix} U^{-1}, \tag{$\dagger$}$$

say. It's easy to see that (†) is equivalent to ($*$).

Now let's return to the case of an operator whose eigenvectors are not orthogonal. In this case the eigenprojections are not orthogonal, and so the construction ($*$) must be wrong. (The other two methods remain correct.) Recall that to construct $\underline{P}_\nu$ we need to know not only the basis vectors in $\mathcal{V}(\lambda_\nu)$ [i.e., ran $\underline{P}_\nu$] but also the *other* basis vectors [i.e., ker $\underline{P}_\nu$]. Only when we know *a priori* that $\underline{P}_\nu^* = \underline{P}_\nu$ does one of these determine the other. In this nonorthogonal case we need to replace the complex-conjugated components of the eigenvectors in our formula ($*$) by the complex-conjugated components of the corresponding elements of the *reciprocal basis*. The latter was the subject of a section of the textbook which we skipped (Sec. 14); we'll get there eventually, when we talk about linear functionals. But from (†) we can see already that the vectors in the reciprocal basis are the complex conjugates of the rows of the matrix U^{-1}. (In the orthogonal case, U is unitary, so these are the same as the columns of U; the reciprocal basis is then equal to the original ON basis.)

The Jordan theorem

The only remaining case is that of a matrix which is not diagonalizable at all! So far we know that if the roots of $\det(\underline{A} - \lambda) = 0$ are not all distinct, then an eigenbasis *may* fail to exist. The example $\underline{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ was pointed out earlier. It turns out that this example is typical of the worst that can happen. That is, the closest we can come to diagonalization for the most general $\underline{A}$ is:

Theorem (Jordan canonical form). *Given* $\underline{A}\colon \mathcal{V} \to \mathcal{V}$, *there exists a basis for* $\mathcal{V}$ *(not necessarily ON) with respect to which the matrix of* $\underline{A}$ *is of the form exemplified by*

$$A = \begin{pmatrix} 3 & \leftarrow & 1 & & 0 & & 0 \\ & & \downarrow & & & & \\ 0 & & 3 & \leftarrow & 0 & & 0 \\ & & & & \downarrow & & \\ 0 & & 0 & & 3 & & 0 \\ 0 & & 0 & & 0 & & 2 \end{pmatrix}.$$

Precisely, all off-diagonal elements of A *are* 0*, except that an element on the diagonal above the main diagonal* **may** *be* 1 *if the two adjacent main-diagonal elements are equal. In other words,* $\mathcal{V} = \mathcal{U}_1 \oplus \cdots \oplus \mathcal{U}_L$ *, where each* $\mathcal{U}_\nu$ *has a basis consisting of vectors on which* $\underline{A}$ *acts in one of the following ways:*

(a) $\underline{A}\vec{v} = \lambda_\nu \vec{v}$ *(i.e.,* $\vec{v} \in \mathcal{V}_\nu$ *, the* λ_ν*-eigenvectors); or*

(b) $\underline{A}\vec{v} = \lambda_\nu \vec{v} + \vec{u}$*, where* $\vec{u}$ *is the preceding element in the basis for* $\mathcal{U}_\nu$ *.*

Let $\mathcal{W}_\nu$ be the span of the basis elements of type (b) (which are called λ_ν-*associated* vectors). Then $\mathcal{U}_\nu = \mathcal{V}_\nu \oplus \mathcal{W}_\nu$. All the vectors belonging to a given *Jordan block* in the block-diagonal structure of A are called a *Jordan chain*, since they are linked to each other by $\underline{A}$. Note that there may be several chains belonging to one λ_ν .

Proofs of the Jordan theorem

1. Bowen & Wang's proof (the standard one). It uses
   - Cayley–Hamilton theorem
   - something called the *minimal polynomial* of $\underline{A}$ (see the book)
   - a lemma on the structure of *nilpotent operators* (operators $\underline{N}$ such that there is a p for which $\underline{N}^p = \underline{0}$)
   - factor spaces.
2. R. Fletcher and D. C. Sorensen, "An algorithmic derivation of the Jordan canonical form", Am. Math. Monthly **90**, 12–16 (1983). This proof has a comparatively computational flavor; it concentrates on *matrices*, rather than *operators*.

3. A. Galperin and Z. Waksman, "An elementary approach to Jordan theory", Am. Math. Monthly **87**, 728–732 (1980). This is the proof I shall follow in class.

A related article: B. R. Gelbaum, "An algorithm for the minimal polynomial of a matrix", Am. Math. Monthly **90**, 43–44 (1983).