Advanced Murder Tools for Eigenvectors

Aditya Sengupta

December 15, 2019

1 Introduction

Diagonalization is nice! Matrices that are diagonal commute, it's really easy to exponentiate them or raise them to powers, and they have nice interpretations in fields like quantum mechanics (the basis they're associated with is the basis of eigenstates for an operator, and the eigenvalues are quantities like energy that can actually be measured.) However, it doesn't always work, and in this note, we'll talk about why, and we'll develop a generalization that does always work.

2 The process of diagonalizing a matrix

Intuitively, diagonalization is a change of basis from whatever basis you're already in to one in which all the basis elements are eigenvectors. Suppose V is an $n-$dimensional vector space over a field F (which we'll restrict to just $\mathbb{R}$ or $\mathbb{C}$). Let's say we have a matrix A encoding some transformation $T : V \to V$ using some basis β. Then, diagonalization means translating β to some β', and correspondingly translating A to some D, such that for every basis element $\vec{v_i}$,

$$D\vec{v_i} = \lambda_i \vec{v_i}$$

To diagonalize a matrix, we need to find these eigenvalues λ_i and the corresponding eigenvectors $\vec{v_i}$. We do this by solving the characteristic polynomial of the matrix,

$$f_A(t) = \det(A - tI)$$

and for each root of this polynomial, λ_i, we get the eigenvectors by finding a basis for $Nul(A - \lambda_i I)$.

After that, we can populate the change-of-basis matrix and the diagonal transformation matrix:

$$A = PDP^{-1}$$
$$P = \begin{bmatrix} \vec{v_1} & \vec{v_2} & \dots & \vec{v_n} \end{bmatrix}$$
$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

But this only works if the eigenvectors actually give a basis for the space V: what if you can't construct enough eigenvectors? Then you wouldn't be able to make the matrix P and a diagonal representation wouldn't exist. How do we check this?

3 Conditions for diagonalizability

The first thing we do when we try to diagonalize a matrix is compute the characeristic polyomial, so let's look at that. We know it goes to 0 on the eigenvalues, so for each eigenvalue λ_i there should be a term $(t - \lambda_i)$ in the characteristic polynomial. Therefore, we can hopefully write the characteristic polynomial as

$$f_A(t) = c(t - \lambda_1)^{m_1}(t - \lambda_2)^{m_2} \dots (t - \lambda_k)^{m_k}$$

If we can write $f_A(t)$ this way, we say it "splits". We call each of the m_is the "algebraic multiplicity" of the eigenvalue λ_i; they account for the cases where you repeat an eigenvalue. If we have an $m_i > 1$, we need more than one eigenvector to capture all the possibilities for what might have caused the eigenvalue λ_i. We formalize this by defining a subspace $E_{\lambda_i} \subset F^n$, the λ_i-eigenspace for short:

$$E_{\lambda_i} = \{\vec{v} \in V \mid A\vec{v} = \lambda_i \vec{v}, \vec{v} \in F^n\}$$

We can show that $\dim E_{\lambda_i} \leq m_i$: we'll never have more eigenvectors than we need. For now, let's assume that we have equality, i.e. $\dim E_{\lambda_i} = m_i$ for all i. This means each eigenspace has exactly as many linearly independent eigenvectors as we need to match the multiplicity of its eigenvalue. Therefore, we're done: the new basis we want is just a union of the bases from each eigenspace. Let's prove that diagonalizability is equivalent to this property.

Proof. If $\dim E_{\lambda_i} = m_i$ for all i, then a basis of E_{λ_i} has m_i elements. If you combine all of these bases, you get a collection of linearly independent elements. This is true because each individual eigenbasis is linearly independent (it's a basis for its own subspace), and the subspaces are disjoint (otherwise you could have an eigenvector with more than one possible eigenvalue). Also, since we're assuming the characteristic polynomial splits perfectly, the sum of the multiplicities $\sum_i m_i = n$. This means our collection of linearly independent elements has n elements, so it's a basis for V! This means there exists a diagonal representation for the matrix A.

In the other direction, if A is diagonalizable, there exists a basis γ that we can split up into disjoint subsets $\gamma_1, \dots, \gamma_k$. Here, each element in subset γ_i is in E_{λ_i}. Since γ is linearly independent, the number of such elements in γ_i is at most $\dim E_{\lambda_i}$, which we know in turn is less than or equal to the multiplicity m_i. Therefore

$$n = \sum_{i=1}^{k} |\gamma_i| \leq \sum_{i=1}^{k} \dim E_{\lambda_i} \leq \sum_{i=1}^{k} m_i = n$$

Since this inequality starts and ends with n, it has to be equal to n at every step. Further, since $\dim E_{\lambda_i} \leq m_i$ for all i, they must each individually be equal; otherwise, if we had some $\dim E_{\lambda_i} < m_i$ we would need some other $\dim E_{\lambda_j} > m_j$ to match. So $\dim E_{\lambda_i} = m_i$ for all i. □

4 Generalizing to Jordan form

We made two assumptions in our work above: that $f_A(t)$ splits over F, and that $\dim E_{\lambda_i} = m_i$. What if they don't hold?

The first assumption turns out not to be a big deal. Not every polynomial splits over $\mathbb{R}$, e.g. $t^2 + 1$ doesn't. But every polynomial splits over $\mathbb{C}$. So if we have problems with this, we can just split the polynomial over $\mathbb{C}$.

The second assumption is a bigger problem. It's easy to construct examples in which it doesn't hold. For example, suppose we have a matrix

$$J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

The only eigenvalue is λ, and all its eigenvectors are in $\text{Span}\{\begin{bmatrix} 1 & 0 \end{bmatrix}^\intercal\}$. Therefore, $\dim E_\lambda = 1$. But the characteristic polynomial is

$$\det(J - tI) = \begin{vmatrix} \lambda - t & 1 \\ 0 & \lambda - t \end{vmatrix} = (\lambda - t)^2$$

So λ has multiplicity 2, meaning $\dim E_\lambda < m$.

We can prove that it's actually impossible to diagonalize this matrix: in fact, we just did. In the previous section, we showed that diagonalizability was equivalent to $\dim E_{\lambda_i} = m_i$ always holding, so if we show it doesn't hold, we know there's no way to diagonalize the matrix.

However, J is actually really close to a diagonal matrix: if you ignore the 1 in the top-right, it's diagonal. We can read its eigenvalues off the diagonal, just like we want. It's almost there!

We formalize a matrix being "almost there" by saying it's in <u>Jordan canonical form</u>. It looks like this:

$$\begin{bmatrix} \lambda_1 & 1 & 0 & & & & & & \\ 0 & \lambda_1 & 1 & & & & & & \\ 0 & 0 & \lambda_1 & & & & & & \\ & & & \lambda_2 & 1 & & & & \\ & & & 0 & \lambda_2 & & & & \\ & & & & \lambda_3 & & & & \\ & & & & & \ddots & & & \\ & & & & & & & \lambda_k & 1 \\ & & & & & & & 0 & \lambda_k \end{bmatrix}$$

Jordan form actually <u>extends</u> diagonalization: if a matrix is diagonalizable, its Jordan form is its diagonal form.

5 Constructing a Jordan form matrix

Recall that we need Jordan form because diagonalization can't account for cases where we don't have enough eigenvectors. So what we want to do is construct more vectors that are kind of like eigenvectors. To get these other vectors, let's look at a nice property of eigenvectors: BEING MURDERED.

$$T\vec{v} = \lambda\vec{v} \implies (T - \lambda I)\vec{v} = 0$$

The operator $(T - \lambda I)$ murders $\vec{v}$. (This is one of the rare cases in which the technical term is actually cooler: technically, $(T - \lambda I)$ annihilates $\vec{v}$.) In matrix form, suppose we've got the following matrix:

$$A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$\vec{v_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\vec{v_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are both eigenvectors of this matrix, with eigenvalues λ_1 and λ_2. So we can construct annihilators for both:

$$(T - \lambda_1 I)\vec{v_1} = \left(\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} - \lambda_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \vec{v_1} = \begin{bmatrix} 0 & 0 \\ 0 & \lambda_2 - \lambda_1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$(T - \lambda_2 I)\vec{v_2} = \left(\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} - \lambda_2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \vec{v_2} = \begin{bmatrix} \lambda_2 - \lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Not being able to find enough eigenvectors to make a basis of V is caused by not being able to find enough "murder weapons" that are effective enough: all the murder weapons you can find don't really get the job done, because the dimension associated to their eigenspace is too low. So, let's improvise another way of murdering a vector: stab/slash/shoot it again!

Suppose our matrix is a 2×2 Jordan block with eigenvalue λ, that is, it's J from before:

$$J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

$\begin{bmatrix} 1 & 0 \end{bmatrix}^\intercal$ is an eigenvector, so it'd be nice if we could finish our basis by augmenting it with $\begin{bmatrix} 0 & 1 \end{bmatrix}^\intercal$. Sadly, this isn't an eigenvector, but if we tried to kill it like an eigenvector, we notice something:

$$(T - \lambda I) \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \left(\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

It's an eigenvector, which we know how to kill! So, we can kill $\begin{bmatrix} 0 & 1 \end{bmatrix}^\intercal$ by applying $(T - \lambda I)$ twice (two stabs with a normal knife), or equivalently, applying $(T - \lambda I)^2$ once (a sharpened double-bladed sword).

If we can find vectors that can be brought to an eigenvector by applying the annihilator, we can make a basis out of them, and we'll show in a minute that the transformation in this basis is a Jordan-form matrix. These vectors, which we'll call generalized eigenvectors, form a basis for a space. Generalized eigenvectors are different in that you can successively apply the annihilator to one of them to get a cycle of generalized eigenvectors: in this example, we had

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \xrightarrow{T-\lambda I} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \xrightarrow{T-\lambda I} \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The length of the cycle will match the multiplicity of the eigenvalue. The last element of the cycle (before the zero element) will always be an eigenvector.

In analogy to the eigenspaces from before, we'll call this a generalized eigenspace:

$$K_\lambda = \{v \in V \mid (A - \lambda I)^p v = 0 \text{ for some } p \in \{1, 2, \dots\}\}$$

This completes the analogy (between diagonalization and Jordan form, not between diagonalization and murder): we can make a union of cycles of generalized eigenvectors, and that's the basis of V that makes A the simplest matrix possible!

6 Proving the generality of Jordan form

We can prove that every matrix can be put into Jordan form. This has been done rigorously at `https://math.berkeley.edu/~frenkel/math110/jordan.pdf`. One level of detail that was omitted is demonstrating that a transformation, under a basis that's a union of cycles of the transformation's generalized eigenvectors, has the Jordan form as its matrix representation. I'll do this by considering a single Jordan block: each block is independent, so the argument for stacking multiple blocks together follows similarly.

Suppose we have a basis of generalized eigenvectors $\{x_1, x_2, x_3\}$ with eigenvalue λ such that

$$x_3 \xrightarrow{T-\lambda I} x_2 \xrightarrow{T-\lambda I} x_1 \xrightarrow{T-\lambda I} 0$$

To write out the matrix representation, let's think about what T does to each of these elements. x_1 is an eigenvector with eigenvalue λ, so the column associated with x_1 just scales it:

$$\begin{bmatrix} \lambda & & \\ 0 & & \\ 0 & & \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

x_2 yields x_1 under the annihilator, so $(T - \lambda I)x_2 = x_1 \implies Tx_2 = x_1 + \lambda x_2$:

$$\begin{bmatrix} \lambda & 1 & \\ 0 & \lambda & \\ 0 & 0 & \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda \\ 0 \end{bmatrix}$$

Finally, x_3 yields x_2 under the annihilator, so $(T - \lambda I)x_3 = x_2 \implies Tx_3 = x_2 + \lambda x_3$:

$$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ \lambda \end{bmatrix}$$

So just from this structure of generalized eigenvectors and wanting to commit murder, we've built the Jordan form!

7 An example

Suppose we wanted to put the following matrix into Jordan form:

$$\begin{bmatrix} 11 & -4 & -5 \\ 21 & -8 & -11 \\ 3 & -1 & 0 \end{bmatrix}$$

We solve $\det(A - \lambda I) = 0$ and get the polynomial $-\lambda^3 + 3\lambda^2 - 4 = 0$. This can be factored (by trying roots, and by long division) as $(\lambda+1)(\lambda-2)^2$. So we know we have to look at the -1- and 2-eigenspaces, and that ideally we want one eigenvector from the first one and two from the second.

We can get the first decently easily:

$$\text{Nul}(A + I) = \text{Nul} \begin{bmatrix} 12 & -4 & -5 \\ 21 & -7 & -11 \\ 3 & -1 & 1 \end{bmatrix} = \text{Span} \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$$

Notice that the first and second columns are proportional to each other, so there's definitely a degree of linear dependence. Probably the easiest way to proceed, other than "see the answer by magic", is row reduction:

$$\begin{bmatrix} 12 & -4 & -5 \\ 21 & -7 & -11 \\ 3 & -1 & 1 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 3 \\ 21 & -7 & -11 \\ 3 & -1 & 1 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 1 \\ 21 & -7 & 0 \\ 3 & -1 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 3 & -1 & 0 \end{bmatrix}$$

This matches the multiplicity, so we're done! Now, let's look at the 2-eigenspace.

$$\text{Nul}(A - 2I) = \text{Nul} \begin{bmatrix} 9 & -4 & -5 \\ 21 & -10 & -11 \\ 3 & -1 & 2 \end{bmatrix}$$

Here, you might notice that all the rows sum to 0, which gives us one eigenvector, $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^{\mathsf{T}}$. Sadly, though, this is the <u>only</u> eigenvector: although you can construct one of the column vectors as a linear combination of the other two (most easily $C_1 = -C_2 - C_3$), there's no other linear dependence meaning you can't get a second linearly independent vector in the null space. (Rank-nullity means if you have a column space of dimension 2, your null space has dimension 1.) We haven't matched the multiplicity with just regular eigenvectors, so let's try and make a generallized one! We want to find a vector satisfying the following:

$$\begin{bmatrix} 9 & -4 & -5 \\ 21 & -10 & -11 \\ 3 & -1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Row reduction and a bit of normal algebra gives us the solution $x = 1, y = 2, z = 0$. Therefore, we're done if we just construct the change-of-basis matrix like in the previous section:

$$M = SJS^{-1}, S = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 1 & 2 \\ 0 & 1 & 0 \end{bmatrix}, J = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$