Math 350: Linear Algebra

Wolmer V. Vasconcelos

Set 8

Spring 2010

Outline

Past... and Future ...

A high point recently:

Theorem

Let $\mathbf{T}$ *be a linear operator on the finite-dimensional inner product space* $\mathbf{V}$*. Then* $\mathbf{T}$ *is self-adjoint (i.e.* $\mathbf{T} = \mathbf{T}^*$*) if and only if there exists an orthonormal basis of* $\mathbf{V}$ *consisting of eigenvectors of* $\mathbf{T}$*.*
Alternatively, if $\mathbf{T}$ *is self-adjoint, there is an unitary operator* $\mathbf{U}$ *such that*

$$\mathbf{UTU}^* = \text{Diagonal}.$$

This is a powerful result. It requires very specific hypotheses which fortunately occur in many important problems. These are analytic/algebraic theorems, need $\mathbb{R}$ or $\mathbb{C}$.

Now...

Jordan blocks

Let **A** be a 8-by-8 matrix with 3 eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of multiplicities $3, 2, 3$ resp. Here is an example:

$$\mathbf{A} = \begin{bmatrix} \boxed{\mathbf{J}_1} & O & O \\ O & \boxed{\mathbf{J}_2} & O \\ O & O & \boxed{\mathbf{J}_3} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{bmatrix}$$

The color coded blocks are called **Jordan Blocks**

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

$$\det(\mathbf{J} - x\mathbf{I}) = (\lambda - x)^3, \quad \text{eigenvalue: } \lambda \text{ (triple)}$$

$$\text{eigenspace: } N(\mathbf{J} - \lambda\mathbf{I}), \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ (single)}$$

Not diagonalizable: Eigenspace too small. We are going to define another subspace.

$$\mathbb{N} = \mathbf{J} - \lambda\mathbf{I} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Observe:

$$\mathbb{N}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbb{N}^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This show that the nullspace of $(\mathbf{J} - \lambda\mathbf{I})^3$ has dimension 3.

Part of the usefulness of these blocks is that we can define and calculate functions such as $\exp \mathbf{J}$, $\sin \mathbf{J}$, etc and have an analysis based on them.

A big theorem

Our goal is to prove and illustrate the use of

Theorem

Let $\mathbf{T}$ *be a linear operator on the finite-dimensional vector space* $\mathbf{V}$*. If the characteristic polynomial of* $\mathbf{T}$ *splits, then* $\mathbf{T}$ *has a matrix representation made up of Jordan blocks. Moreover, the number and sizes of the blocks for each eigenvalue are unique.*

For example, if **A** is a 8-by-8 matrix with 3 eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of multiplicities $3, 2, 3$ resp. there is an invertible matrix **P** such that

$$\mathbf{PAP}^{-1} = \begin{bmatrix} \boxed{\mathbf{J}_1} & O & O \\ O & \boxed{\mathbf{J}_2} & O \\ O & O & \boxed{\mathbf{J}_3} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{bmatrix}$$

We know very well what is need to achieve diagonalization of a matrix **A**: we need a basis $\mathcal{A} = \{v_1, \ldots, v_n\}$ of eigenvectors. This requires that for each eigenvalue λ,

$$\dim E_\lambda = \text{algebraic multiplicity of } \lambda$$

Diagonalization will not occur if one of the eigenspaces is too small. Let us do something about it.

Generalized eigenspace

Definition

Let $\mathbf{T}$ be a linear operator of the vector space $\mathbf{V}$ and let λ be a scalar. A nonzero vector $v \in \mathbf{V}$ is a **generalized eigenvector** of $\mathbf{T}$ if $(\mathbf{T} - \lambda \mathbf{I})^p(v) = O$ for some positive integer p. We denote this set, together with the vector O, by K_λ.

Note that **eigenvectors** correspond to the case $p = 1$. Consider

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$\lambda = 1$ is the only eigenvalue. e_1 is an eigenvector: $(\mathbf{A} - \mathbf{I})(e_1) = O$

e_2 is not an eigenvector: $(\mathbf{A} - \mathbf{I})(e_2) \neq O$ but $(\mathbf{A} - \mathbf{I})^2(e_2) = O$, so e_2 is a generalized eigenvector.

Proposition

Let $\mathbf{T}$ *be a linear operator of the vector space* $\mathbf{V}$*, and let* λ *be an eigenvalue of* $\mathbf{T}$*.*

1. K_λ *is a* $\mathbf{T}$*-invariant subspace of* $\mathbf{V}$ *containing the eigenspace* E_λ*.*
2. *For any scalar* $\mu \neq \lambda$*, the restriction of* $\mathbf{T} - \mu\mathbf{I}$ *to* K_λ *is one-one.*

Proof: 1. If $u, v \in K_\lambda$,

$$(\mathbf{T} - \lambda\mathbf{I})^p(u) = O, \quad (\mathbf{T} - \lambda\mathbf{I})^q(v) = O,$$

If m is the larger of the integers p, q,

$$(\mathbf{T} - \lambda\mathbf{I})^m(u + v) = (\mathbf{T} - \lambda\mathbf{I})^m(u) + (\mathbf{T} - \lambda\mathbf{I})^m(v) = O + O = O.$$

The scaling property is also clear. As for $\mathbf{T}$-invariant assertion, if $u \in K_\lambda$, $(\mathbf{T} - \lambda\mathbf{I})^p(\mathbf{T}(u)) = \mathbf{T}(\mathbf{T} - \lambda\mathbf{I})^p(u) = O$, therefore $\mathbf{T}(u) \in K_\lambda$.

Part 2. of the Theorem on generalized eigenspaces: That $\mathbf{T} - \mu\mathbf{I}$, $\mu \neq \lambda$, is one-one on K_λ:

If $u \in K_\lambda$, that is $(\mathbf{T} - \lambda\mathbf{I})^p(u) = 0$ and $(\mathbf{T} - \mu\mathbf{I})u = O$ then we claim $u = O$.

Deny. We may assume $v = (\mathbf{T} - \lambda\mathbf{I})^{p-1}(u) \neq 0$. But then

$$(\mathbf{T} - \mu\mathbf{I})v = (\mathbf{T} - \mu\mathbf{I})(\mathbf{T} - \lambda\mathbf{I})^{p-1}u = (\mathbf{T} - \lambda\mathbf{I})^{p-1}(\mathbf{T} - \mu\mathbf{I})u = O$$

$$\text{and} \quad (\mathbf{T} - \lambda\mathbf{I})v = (\mathbf{T} - \lambda\mathbf{I})^p u = O,$$

so $\mathbf{T}(v) = \lambda v$ and $\mathbf{T}(v) = \mu v$, $\mu \neq \lambda$, $v \neq O$, a contradiction.

Corollary

dim $K_\lambda \leq$ algebraic multiplicity *of* λ.

Proof: Let $f(x) = \prod_i (x - \lambda_i)^{m_i}$ be the characteristic polynomial of **T**.

Part 1 of the Proposition says that K_λ is a **T**-invariant subspace, while Part 2 says that the restriction of **T** to K_λ has no eigenvalue $\neq \lambda$. By a previous result, we know that the characteristic polynomial $g(x)$ of this restriction divides $f(x)$.It follows that $g(x) = (x - \lambda)^n$,

$$n \leq m = \text{algebraic muliplicity of } \lambda.$$

Later we will show that $n = m$.

GCD of polynomials

If $f(x)$ and $g(x)$ are polynomials in $\mathbf{F}[x]$, the **greatest common divisor** is the monic polynomial of highest degree $h(x)$ that divides $f(x)$ and $g(x)$

$$\text{gcd}(f(x), g(x)) = h(x)$$

For example,

$$\text{gcd}((x-1)^3(x-2)^2, (x-1)(x-2)^4) = (x-1)(x-2)^2.$$

An elementary, but very useful fact, is that long division provides an effective method to find gcds.

Proposition

A polynomial $f(x) \in \mathbb{R}[x]$ *of degree* $f(x) \geq 1$ *has multiple roots if and only if* $\gcd(f(x), f'(x)) \neq 1$*.*

Thus, while it is hard to find the roots of a polynomial $f(x)$, it is easy to determine whether it has multiple roots!
The explanation is very simple: If $f(x)$ has a root of algebraic multiplicity m,

$$f(x) = (x - a)^m g(x), \quad g(a) \neq 0,$$

its derivative

$$f'(x) = m(x - a)^{m-1} g(x) + (x - a)^m g'(x)$$

has a as a root with multiplicity $m - 1$. This implies that $(x - a)^{m-1}$ is a common factor of $f(x)$ and $f'(x)$, and therefore will be a factor of $\gcd(f(x), f'(x))$.

1. If $\gcd(f(x), f'(x)) = 1$, then $f(x)$ has no repeated (complex) roots.
2. Suppose $f(x)$ is the characteristic polynomial of a 3-by-3 complex matrix $\mathbf{A}$, and we must decide whether it is diagonalizable.What to do?
    1. If $\gcd(f(x), f'(x)) = 1$, by the discussion above the roots are distinct, and we are done: $\mathbf{A}$ is diagonalizable.
    2. If there is a double root a and a single root b, $\gcd(f(x), f'(x)) = (x - a)$. We check the dimension of the eigenspace E_a, if $\dim E_a = 2$, ok, otherwise not diagonalizable.
    3. If a is a triple root, $\gcd(f(x), f'(x)) = (x - a)^2$. Again we check whether $\dim E_a = 3$.

Long division

Recall the long division algorithm for polynomials in $\mathbf{F}[x]$: If $f(x), g(x) \neq 0$ are polynomials, there exist polynomials $q(x), r(x)$ such that

$$f(x) = q(x)g(x) + r(x), \quad r(x) = 0 \text{ or } \deg r(x) < \deg g(x)$$

Look at a consequence:

$$\gcd(f(x), g(x)) = \gcd(g(x), r(x))$$

since any polynomial $p(x)$ that divides (both) $f(x), g(x)$ will divide $g(x), r(x)$, and conversely. Note that the data of $g(x), r(x)$ has lower degrees, so we can turn this into an algorithm:

gcd algorithm

Starting at

$$f(x) = q(x)g(x) + r(x),$$

1. Iterating, if $r(x) \neq 0$ and we divide $g(x) = q_1(x)r(x) + r_1(x)$, then any polynomial $p(x)$ that divides (both) $f(x), g(x)$ will divide $r(x), r_1(x)$, and conversely.
2. Since $\deg g(x) > \deg r(x) > \deg r_1(x) > \cdots$, ultimately we shall have $r_{n-1}(x) = q_{n-1}(x)r_n(x), \quad r_n(x) \neq O$.
3. $r_n(x)$ is (a) largest degree polynomial that divides both $f(x)$ and $g(x)$, and any such polynomial will divide $r_n(x)$.

Theorem

If $r_n(x)$ is the last nonzero remainder in the sequence of long divisions, then $r_n(x)$ divides $f(x)$ and $g(x)$. Moreover, there exist polynomials $a(x), b(x)$ such that

$$r_n(x) = a(x)f(x) + b(x)g(x).$$

$r_n(x)$ is called the (a) **GCD** *of $f(x)$ and $g(x)$.*

Proof: For simplicity suppose $n = 2$, so we have the divisions

$$f = qg + r, \quad g = q_1 r + r_1, \quad r = q_2 r_1 + r_2, \quad , r_1 = q_3 r_2$$

$$\begin{aligned} r_2 &= r - q_2 r_1 = r - q_2(g - q_1 r) = r(1 + q_2 q_1) - q_2 g \\ &= (f - qg)(1 + q_2 q_1) - q_2 g \end{aligned}$$

Now we collect the coefficient of f–it will be $a(x)$–and of g–it will be $b(x)$: $\gcd(f, g) = a(x)f(x) + b(x)g(x)$

Jordan Decomposition

We are now going to apply these observations to the characteristic polynomial $p(x) = \det(\mathbf{A} - x\mathbf{I})$ of a matrix $\mathbf{A}$, whose eigenvalues λ_i exist in the field $\mathbf{F}$. Note for $\mathbf{F} = \mathbb{C}$, this is the case for all matrices. Underlying the following discussion is the assumption that

$$p(x) = \pm \prod_{i=1}^{m} (x - \lambda_i)^{m_i}.$$

1. If $f(x) = (x - \lambda)^m$, $g(x) = (x - \mu)^n$ and $\lambda \neq \mu$ are different scalars, then $\gcd(f(x), g(x)) = 1$, this means that there is a (decomposition) $1 = a(x)f(x) + b(x)g(x)$.
2. Consider now the case of the 3 polynomials,

$$f(x) = (x-\lambda_1)^m(x-\lambda_2)^n, g(x) = (x-\lambda_1)^m(x-\lambda_3)^p, h(x) = (x-\lambda_2)^n(x-\lambda_3)^p$$

where $\lambda_1, \lambda_2, \lambda_3$ are distinct. Note that

$$\begin{aligned} \gcd(f, g) &= (x - \lambda_1)^m \\ \gcd(f, h) &= (x - \lambda_2)^n \\ \gcd(g, h) &= (x - \lambda_3)^p \\ \gcd(f, g, h) &= \gcd((x - \lambda_1)^m, h) = 1 \end{aligned}$$

3. These equations, will imply that we have an equality

$$1 = a(x)f(x) + b(x)g(x) + c(x)h(x).$$

Suppose the characteristic polynomial of **T** has a decomposition

$$\det(x\mathbf{I} - \mathbf{T}) = (x-a)^m(x-b)^n(x-c)^p.$$

The polynomials $\mathbf{f}(x) = (x-b)^n(x-c)^p$, $\mathbf{g}(x) = (x-a)^m(x-c)^p$, $\mathbf{h}(x) = (x-a)^m(x-b)^n$, have gcd $= 1$ as they have no common divisor. According to the observation above, we have an equality

$$1 = A(x)\mathbf{f}(x) + B(x)\mathbf{g}(x) + C(x)\mathbf{h}(x)$$

Evaluating $x \to \mathbf{T}$ gives the equality

$$\mathbf{I} = A(\mathbf{T})\mathbf{f}(\mathbf{T}) + B(\mathbf{T})\mathbf{g}(\mathbf{T}) + C(\mathbf{T})\mathbf{h}(\mathbf{T})$$

Applying to an arbitrary vector **v** we have

$$\begin{aligned}\mathbf{v} = \mathbf{I}(\mathbf{v}) &= \underbrace{A(\mathbf{T})(\mathbf{T} - b\mathbf{I})^n(\mathbf{T} - c\mathbf{I})^p(\mathbf{v})}_{v_1} + \underbrace{B(\mathbf{T})(\mathbf{T} - a\mathbf{I})^m(\mathbf{T} - c\mathbf{I})^p(\mathbf{v})}_{v_2} \\ &+ \underbrace{C(\mathbf{T})(\mathbf{T} - a\mathbf{I})^m(\mathbf{T} - b\mathbf{I})^n(\mathbf{v})}_{v_3}\end{aligned}$$

$$\mathbf{v} = v_1 + v_2 + v_3$$

$$(\mathbf{T} - a\mathbf{I})^m(v_1) = A(\mathbf{T})(\mathbf{T} - a\mathbf{I})^m(v_1) = A(\mathbf{T})(\mathbf{T} - a\mathbf{I})^m(\mathbf{T} - b\mathbf{I})^n(\mathbf{T} - c\mathbf{I})^p(v) = 0$$

by Cayley-Hamilton. This says that every vector **v** is a sum of vectors in K_a, K_b and K_c. It is also easy to see that v_1, v_2, v_3 are linearly independent.

Chinese Remainder Theorem

Theorem

Let $f_1(x), \dots, f_m(x)$ be polynomials of $\mathbf{F}[x]$. If $g(x) = gcd(f_1(x), \dots, f_m(x))$ there are polynomials $a_i(x)$ such that

$$g(x) = a_1(x)f_1(x) + \cdots + a_m(x)f_m(x).$$

Let **T** be a linear operator on the finite-dimensional vector space **V.** Suppose its characteristic polynomial $\det(\mathbf{T} - x\mathbf{I})$ splits:

$$f(x) = \pm \prod_{i=1}^{m} (x - \lambda_i)^{n_i}, \quad \text{distinct } \lambda_i.$$

For each i, setting $f_i(x) = \frac{f(x)}{(x-\lambda_i)^{n_i}}$, gives us a collection $f_1(x), \dots, f_m(x)$ of gcd $= 1$: In

$$1 = a_1(x)f_1(x) + \cdots + a_m(x)f_m(x)$$

replace $x \to \mathbf{T}$

$$\mathbf{I} = a_1(\mathbf{T})f_1(\mathbf{T}) + \cdots + a_m(\mathbf{T})f_m(\mathbf{T})$$

Now we are going to make several observations about this decomposition.

1. The range of $f_i(\mathbf{T})$ is contained in the generalized eigenspace K_{λ_i}:If $u = f_i(\mathbf{T})(v)$,

$$(\mathbf{T} - \lambda_i)^{n_i} f_i(\mathbf{T})(v) = f(\mathbf{T})(v) = 0,$$

   since by the Cayley-Hamilton theorem $f(\mathbf{T}) = 0$.

2. For every $v \in \mathbf{V}$

$$v = \mathbf{I}(v) = \overbrace{a_1(\mathbf{T})f_1(\mathbf{T})(v)}^{\in K_{\lambda_1}} + \cdots + \overbrace{a_m(\mathbf{T})f_m(\mathbf{T})(v)}^{\in K_{\lambda_m}}$$

Last Class... and Today ...

- Jordan blocks are convenient for computation
- Some high school algebra
- Generalized eigenspaces
- ○ Jordan decomposition

Generalized eigenvectors and eigenspaces

Let $\mathbf{T} : \mathbf{F}^n \to \mathbf{F}^n$ be a L.T. with characteristic polynomial

$$\mathbf{f}(\mathbf{x}) = (\mathbf{x} - \lambda_1)^p(\mathbf{x} - \lambda_2)^q(\mathbf{x} - \lambda_3)^r$$

For each λ we have the subspace of **generalized eigenvectors**

$$K_\lambda = \{v : (\mathbf{T} - \lambda\mathbf{I})^i(v) = 0\}$$

Unlike the behavior of eigenspaces, we always have

$$\mathbf{F}^n = K_{\lambda_1} \oplus K_{\lambda_2} \oplus K_{\lambda_3},$$

that is $\mathbf{F}^n$ has a basis of generalized eigenvectors.

Generalized eigenvectors and eigenspaces

- If $\mathbf{T}$ is a linear operator of the vector space $\mathbf{V}$ and λ is a scalar, a nonzero vector $v \in \mathbf{V}$ is a **generalized eigenvector** of $\mathbf{T}$ if $(\mathbf{T} - \lambda\mathbf{I})^p(v) = O$ for some positive integer p. We denote this set, together with the vector O, by K_λ. K_λ is usually bigger than the eigenspace E_λ.
- In fact,

$$\mathbf{V} = \bigoplus_i K_{\lambda_i},$$

in particular, $\mathbf{V}$ has a basis made up of generalized eigenvectors.

This representation says that every vector $v \in \mathbf{V}$ can be written as

$$v = v_1 + \cdots + v_m, \quad v_i \in K_{\lambda_i}$$

Since we already proved that $\dim K_{\lambda_i} \leq n_i$, the algebraic multiplicity of λ_i, this equality proves equality of the dimensions. It can be written as

$$\mathbf{V} = K_{\lambda_1} \oplus \cdots \oplus K_{\lambda_m},$$

and the matrix representation of **T** has the block format (after picking bases of the K_{λ_i}'s)

$$[\mathbf{T}] = \begin{bmatrix} [\mathbf{T}]_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & [\mathbf{T}]_m \end{bmatrix}$$

What this does is to allow us to assume that the characteristic polynomial of **T** has the form $(x-\lambda)^n$. We will argue that such linear operator have a matrix representation made up of Jordan blocks with the same λ. Let us look at one such $p \times p$ block

$$\mathbf{A} = [v_1|\cdots|v_p] = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}$$

$$\underbrace{\mathbf{A}(v_1) = \lambda v_1}_{\text{eigenvector}}, \quad \mathbf{A}(v_2) = v_1 + \lambda v_2, \cdots, \mathbf{A}(v_p) = v_{p-1} + \lambda v_p$$

If we write these equations in the reverse order, we get

Consider the subspace K of the vectors v such that $(\mathbf{A}-\lambda)^3(v)=0$. Assume $(\mathbf{A}-\lambda)^2(v)\neq 0$ for some vector $v\in K$. Fix such a vector and label it v_3.

$$\begin{aligned}(\mathbf{A}-\lambda\mathbf{I})(v_3) &= v_2 \quad \mathbf{A}(v_3)=v_2+\lambda v_3\\ (\mathbf{A}-\lambda\mathbf{I})^2(v_3) &= v_1 \quad \mathbf{A}(v_2)=v_1+\lambda v_2\\ (\mathbf{A}-\lambda\mathbf{I})^3(v_3) &= 0 \quad \mathbf{A}(v_1)=\lambda v_1\end{aligned}$$

Starting on v_3 and applying $\mathbf{U}=\mathbf{A}-\lambda\mathbf{I}$ repeatedly we get all the vectors of the basis

$$v_3\to v_2\to v_1\to O$$

We will say that v_3 is the generator of the basis, and that $\gamma=\{v_1,v_2,v_3\}$ is a cycle of generalized eigenvectors, v_1 is the initial and v_3 the end vectors: They form a so-called **dot diagram**

$$\text{generator}=v_3=\bullet\to\bullet\to\cdots\to\bullet=v_1=\text{eigenvector}$$

$$\begin{aligned}(\mathbf{A}-\lambda\mathbf{I})(v_p) &= v_{p-1}\\ (\mathbf{A}-\lambda\mathbf{I})^2(v_p) &= v_{p-2}\\ &\vdots\\ (\mathbf{A}-\lambda\mathbf{I})^{p-1}(v_p) &= v_1\\ (\mathbf{A}-\lambda\mathbf{I})^{p}(v_p) &= 0\end{aligned}$$

Starting on v_p and applying $\mathbf{U} = \mathbf{A} - \lambda\mathbf{I}$ repeatedly we get all the vectors of the basis

$$v_p \to v_{p-1} \to \cdots \to v_2 \to v_1 \to O$$

We will say that v_p is the generator of the basis, and that $\gamma = \{v_1, v_2, \ldots, v_p\}$ is a cycle of generalized eigenvectors, v_1 is the initial and v_p the end vectors: They form a so-called **dot diagram**

$$\text{generator} = v_p = \bullet \to \bullet \to \cdots \to \bullet = v_1 = \text{eigenvector}$$

Proposition

Let $\mathbf{T}$ *be a linear operator on the vector space* $\mathbf{V}$. *For some scalar* λ *and some integer p, suppose v is a nonzero vector such that*

$$(\mathbf{T}-\lambda\mathbf{I})^p(v)=O, \quad (\mathbf{T}-\lambda\mathbf{I})^{p-1}(v)\neq O.$$

Then the p vectors $(\mathbf{T}-\lambda\mathbf{I})^{p-1}(v),\ldots,(\mathbf{T}-\lambda\mathbf{I})(v), v$ *are linearly independent. They span a* $\mathbf{T}$*-invariant subspace* $\mathbf{W}$ *and the matrix representation of* $[\mathbf{T}]_{\mathbf{W}}$ *with respect to this basis is a Jordan block.*

Proof: Let us denote these vectors by $v_1,\ldots,v_p=v$, respectively. Suppose we have a linear relation $c_1v_1+\cdots+c_pv_p=O$. Let us prove all $c_i=0$. Let us argue just one case as the general case is similar. Suppose $c_p\neq 0$. Apply the operator $(\mathbf{T}-\lambda\mathbf{I})^{p-1}$ to the relation to obtain

$$v_i = (\mathbf{T} - \lambda \mathbf{I})^{p-i}(v)$$

$$c_1(\mathbf{T} - \lambda \mathbf{I})^{p-1}(v_1) + \cdots + c_p \underbrace{(\mathbf{T} - \lambda \mathbf{I})^{p-1}(v_p)}_{=v_1} = O$$

Note that all terms vanish, except for the last. This contradicts $c_p \neq 0$.

The subspace **W** clearly satisfies $\mathbf{T}(\mathbf{W}) \subset \mathbf{W}$. Finally, note that

$$\begin{aligned} \mathbf{T}(v_i) &= \mathbf{T}(\mathbf{T} - \lambda \mathbf{I})^{p-i}(v) \\ &= (\mathbf{T} - \lambda \mathbf{I})^{p-i+1}(v) + \lambda(\mathbf{T} - \lambda \mathbf{I})^{p-i}(v) = v_{i-1} + \lambda v_i, \end{aligned}$$

which shows that the matrix representation is $\begin{bmatrix} \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & \lambda \end{bmatrix}$

We come now to the crux of the problem: Given a linear operator **T** whose characteristic polynomial is $\pm(x-\lambda)^n$, to prove that there is a matrix representation made up of λ-Jordan blocks (same λ)

$$\begin{bmatrix} \mathbf{J}_1 & O & O \\ O & \mathbf{J}_2 & O \\ O & O & \mathbf{J}_3 \end{bmatrix} = \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

We are going to prove the existence of such representation and the uniqueness of the number and sizes of the blocks.

Situation:

$\mathbf{T} : K_\lambda \to K_\lambda$, dim $K_\lambda = n$, characteristic polynomial of $\mathbf{T}$ is $(x - \lambda)^n$. The eigenspace is $E_\lambda \subset K_\lambda$.

Goal: We will show that K_λ has a basis

$$\mathcal{B} = \bigcup_{i=1}^{m} \gamma_i$$

where each γ_i is a cycle of generalized eigenvectors. The Jordan representation comes from the corresponding matrix representation. For example, if $K_\lambda = E_\lambda$, then a basis of E_λ gives the cycles, all of length 1, and the matrix representation is just $\lambda\mathbf{I}_n$.

1. We are going to argue by induction on $n = \dim K_\lambda$. If $n = 1$ (or, more generally, $K_\lambda = E_\lambda$), there is nothing to prove.
2. Let $\mathbf{Z}$ be the range of $\mathbf{T} - \lambda\mathbf{I}$. For simplicity of notation call this map $\mathbf{U} : K_\lambda \to K_\lambda$. Note that E_λ is the nullspace of $\mathbf{U}$, and therefore $\dim E_\lambda + \dim \mathbf{Z} = n$, by the dimension formula.
3. Since $\dim \mathbf{Z} < n$ and the characteristic polynomial of the restriction of $\mathbf{T}$ to $\mathbf{Z}$ divides $(x - \lambda)^n$, the induction hypothesis guarantees a basis for $\mathbf{Z}$:

$$\gamma' : w, (\mathbf{T} - \lambda\mathbf{I})(w), \ldots, (\mathbf{T} - \lambda\mathbf{I})^{p-1}(w)$$

$$\mathcal{B}' = \bigcup_{i=1}^{r} \gamma_i'$$

where each γ_i' is a cycle of generalized eigenvectors of $\mathbf{Z}$. Let us consider one of these cycles γ':

$$\gamma_i' : w, (\mathbf{T} - \lambda\mathbf{I})(w), \ldots, (\mathbf{T} - \lambda\mathbf{I})^{p-1}(w)$$

But w belongs to the range of $(\mathbf{T} - \lambda\mathbf{I})$, that is $w = (\mathbf{T} - \lambda\mathbf{I})(v)$, for some $v \in \mathbf{V}$. This gives a cycle of **V** itself:

$$\gamma_i : v, (\mathbf{T} - \lambda\mathbf{I})(v), \ldots, (\mathbf{T} - \lambda\mathbf{I})^{p}(v)$$

In this manner, for every γ_i' of **Z** we get a longer cycle (by 1 more vector) of **V**.

We recall that vector at the end of the list are the only eigenvectors and that

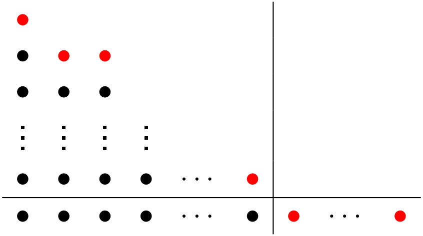
$$\bigcup_{i=1}^{r} \gamma_i$$

contains just r independent eigenvectors, the same set as the basis $\mathcal{B}'$ of **Z**. If these eigenvectors are $u_1, \ldots, u_r$, add (if necessary) $u_{r+1}, \ldots, u_s$ to form a basis of the eigenspace E_λ. Each of these u_i defines a new cycle γ_i of length 1, $i > r$.

Dot Diagrams and Enlarged Cycles

•: vectors in the set $\mathcal{B}'$

•: vectors added.



$\mathbf{T} - \lambda\mathbf{I}$ maps each dot to dot under. Last row is a basis of E_λ: it is mapped to O

Proposition (Very technical, I apologize)

The vectors in the set

$$\mathcal{B} = \bigcup_{i=1}^{s} \gamma_i$$

form a basis of **V**.

Proof: First let us count the number of elements of added to pass from the basis $\mathcal{B}'$ of **Z** to the set $\mathcal{B}$ of **V**:

$$r \text{ (1 for each of the } r \text{ cycles in } \mathcal{B}') + (s - r) = s = \dim E_\lambda$$

$$\text{Therefore} \quad \text{cardinality of } \mathcal{B}' + s = \dim \mathbf{Z} + s = n = \dim \mathbf{V}$$

To prove $\mathcal{B}$ is a basis, ETS that it spans **V**, as they have already the right number of elements for a basis.

Let $u \in \mathbf{V}$ and consider $(\mathbf{T} - \lambda\mathbf{I})(u) \in \mathbf{Z}$. Since every vector in $\mathcal{B}'$ is the image under $\mathbf{T} - \lambda\mathbf{I}$ of some vector in $\mathcal{B}$, we can write

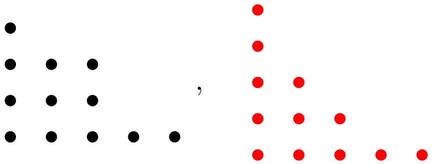
$$(\mathbf{T} - \lambda\mathbf{I})(v) = \text{Linear combination of } (\mathbf{T} - \lambda\mathbf{I})(v_i), \quad v_i \in \mathcal{B}.$$

This implies that

$$(\mathbf{T} - \lambda\mathbf{I})\underbrace{(v - \text{Linear combination of } v_i)}_{=w} = O$$

Thus $w \in E_\lambda$. Since $\mathcal{B}$ contains a basis of E_λ, this implies v lies in the span of $\mathcal{B}$.

To illustrate the uniqueness of Jordan decomposition, suppose **T** gives rise to two different cycle decomposition for K_λ:



Observe that many things match: dim $K_\lambda = 12$ [number of dots, red or black], dim $E_\lambda = 5$ (number of piles, columns). Now we are going to observe things that are off:

$$(\mathbf{T} - \lambda\mathbf{I})^4(\text{any } \bullet) = 0, \quad (\mathbf{T} - \lambda\mathbf{I})^4(\text{top } \bullet) \neq 0$$

This illustrate the argument: The number of dots at level ℓ is the dimension of the subspace of the vectors v of $\mathbf{V}$ such that

$$(\mathbf{T} - \lambda\mathbf{I})^{\ell}(v) = 0$$

Level 1: $5 = 5$
Level 2: $5 + 3 = 5 + 3$
Level 3: $5 + 3 + 3 > 5 + 3 + 2$
Level 4: $5 + 3 + 3 + 1 > 5 + 3 + 2 + 1$

It is clear that the two piles must coincide.

Summary: If the piles are ordered by sizes, they must be identical.

Jordan Decomposition Theorem

Theorem

Any linear operator $\mathbf{T}$ *whose characteristic polynomial* $p(x) = \pm \prod_{i=1}^{m}(x - \lambda_i)^{n_i}$ *splits has a unique matrix representation into blocks*

$$[\mathbf{T}]_{\mathcal{B}} = \begin{bmatrix} \mathbf{A}_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & \mathbf{A}_m \end{bmatrix}$$

where each $\mathbf{A}_i$ *has a representation by Jordan* λ_i*-blocks whose number and sizes are uniquely defined*

$$\begin{bmatrix} \lambda_i & 1 & \cdots & 0 \\ 0 & \lambda_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_i \end{bmatrix}.$$

Exercise

Prove: If $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ is diagonalizable and $\mathbf{W}$ is an invariant subspace then $\mathbf{T_W}$ is also diagonalizable.

- Since $\mathbf{W}$ is $\mathbf{T}$-invariant, picking a basis for $\mathbf{W}$ and extending to a basis for $\mathbf{V}$ we get a matrix representation for $\mathbf{T}$ of the form

$$\begin{bmatrix} \mathbf{T_W} & \mathbf{B} \\ O & \mathbf{C} \end{bmatrix}$$

- It follows that the characteristic polynomial $\mathbf{f}(x)$ of $\mathbf{T}$ is the product of the characteristic polynomials $\mathbf{g}(x)$ and $\mathbf{h}(x)$ of $\mathbf{T_W}$ and $\mathbf{C}$.
- If λ is an eigenvalue and $\mathbf{f}(x) = (x-\lambda)^r F(x)$, $\mathbf{g}(x) = (x-\lambda)^p G(x)$, $\mathbf{h}(x) = (x-\lambda)^q H(x)$ extracting the roots equal to λ, $r = p + q$.

- Since $\mathbf{T}$ is diagonalizable, $\dim E_\lambda = r$, while the corresponding eigenspace of $\mathbf{W}$ is $E'_\lambda = E_\lambda \cap \mathbf{W}$. We must show that $\dim E'_\lambda = p$.
- If $v_1, \ldots, v_n$ is a basis of E'_λ, and completing to a basis of E_λ $v_1, \ldots, v_n, v_{n+1}, \ldots, v_r$, we get the invariant subspace

$$\mathbf{L} = \mathbf{W} + (v_{n+1}, \ldots, v_r) = \mathbf{W} \times (v_{n+1}, \ldots, v_r)$$

- $$\mathbf{T_L} = \begin{bmatrix} \mathbf{T_W} & O \\ O & \lambda \mathbf{I}_{r-n} \end{bmatrix}$$

  a matrix whose characteristic polynomial has the factor $(x-\lambda)^p(x-\lambda)^{r-n}$, and $p + r - n \leq r$, which is not possible if $p > n$. Thus $\dim E'_\lambda = p$.

Minimal Polynomial of a Matrix

Given a square matrix $\mathbf{A}$, there are polynomials $\mathbf{f}(x)$ such that

$$\mathbf{f}(\mathbf{A}) = 0.$$

The best known is $\mathbf{f}(x) = \det(\mathbf{A} - x\mathbf{I})$, the characteristic polynomial: by Cayley-Hamilton:

$$\mathbf{f}(\mathbf{A}) = 0.$$

What else?

Definition

Let $\mathbf{A}$ be a n-by-n matrix. The **minimal polynomial** of $\mathbf{A}$ is the monic polynomial $m(x) = x^m + c_{m-1}x^{m-1} + \cdots + c_0$ of least degree such that

$$m(\mathbf{A}) = \mathbf{A}^m + c_{m-1}\mathbf{A}^{m-1} + \cdots + c_0\mathbf{I} = O.$$

1. If $\mathbf{A} = \mathbf{I}_n$, then $m(x) = x - 1$.
2. If $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $m(x) = x^2$.
3. In the case of the Jordan block $\mathbf{J} = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$, $m(x) = (x - \lambda)^3$. For a block of size n, $m(x) = (x - \lambda)^n$.

$$\mathbf{J} = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix}, \quad \mathbf{U} = \mathbf{J} - \lambda \mathbf{I} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{U}^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{U}^3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{U}^4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$m(x) = (x - \lambda)^4$$

Observe the right drift of the diagonal of 1's until it leaves the matrix!

$$\mathbf{A} = \begin{bmatrix} \boxed{\mathbf{J}_1} & O & O \\ O & \boxed{\mathbf{J}_2} & O \\ O & O & \boxed{\mathbf{J}_3} \end{bmatrix} = \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

Note how $\mathbf{A} - \lambda\mathbf{I}$ is made up of blocks like $\mathbf{J}_1 - \lambda\mathbf{I}$ whose third (or second in the blue case) power is O: $m(x) = (x - \lambda)^3$. (We will make this more precise soon.)

$$\mathbf{A} = \begin{bmatrix} \mathbf{J}_1 & O & O \\ O & \mathbf{J}_2 & O \\ O & O & \mathbf{J}_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{bmatrix}$$

If $\lambda_1 = \lambda_2 \neq \lambda_3$, then $m(x) = (x - \lambda_1)^3 (x - \lambda_3)^3$. This is because $(x - \lambda_1)^3$ works for both λ_1 blocks.

$$\mathbf{A} = \begin{bmatrix} \boxed{\mathbf{J}_1} & O & O \\ O & \boxed{\mathbf{J}_2} & O \\ O & O & \boxed{\mathbf{J}_3} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{bmatrix}$$

If $\lambda_1, \lambda_2, \lambda_3$ are distinct, then $m(x) = (x - \lambda_1)^3 (x - \lambda_2)^2 (x - \lambda_3)^3$.

Proposition

The minimal polynomial $m(x)$ of $\mathbf{A}$ divides every nonzero polynomial $f(x)$ such that $f(\mathbf{A}) = O$. In particular, $m(x)$ divides the characteristic polynomial $p(x) = \det(\mathbf{A} - x\mathbf{I})$ of $\mathbf{A}$.

Proof: To show that $m(x)|f(x)$, use long division to write $f(x) = q(x)m(x) + r(x)$, where $\deg r(x) < \deg m(x)$. We claim that $r(x) = 0$: We have

$$\underbrace{f(\mathbf{A})}_{=0} = q(\mathbf{A})\underbrace{m(\mathbf{A})}_{=0} + r(\mathbf{A}).$$

Thus $r(\mathbf{A}) = 0$. If $r(x) \neq 0$ we can divide it by its leading coefficient and have a monic polynomial $r_0(x)$ of degree smaller than $\deg m(x)$ such that $r_0(\mathbf{A}) = O$, a contradiction. The last assertion is the Cayley-Hamilton theorem.

If we have the Jordan decomposition of **A**, we can be very explicit about its minimal polynomial.

Proposition

Let **A** *be linear operator whose distinct eigenvalues are* $\lambda_1, \ldots, \lambda_m$. *For each eigenvalue* λ_i, *suppose the largest size of its Jordan blocks is* p_i. *Then the minimal polynomial of* **A** *is*

$$m(x) = \prod_{i=1}^{m} (x - \lambda_i)^{p_i}.$$

In particular, if all eigenvalues of **A** *are simple, then the minimal and the characteristic polynomials are equal, except possibly by* $\pm$. *More generally, the minimal and characteristic polynomials coincide (up to sign) if and only for each eigenvalue* λ_i *there is only one Jordan block.*

Let **A** be a n-by-n matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ Here is what they should look like:

$$\mathbf{A} = \begin{bmatrix} \boxed{\mathbf{J}_1} & O & O & O \\ O & \boxed{\mathbf{J}_2} & O & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & 0 & \boxed{\mathbf{J}_m} \end{bmatrix}$$

$\mathbf{J}_i$ is the single Jordan block corresponding to λ_i

Problems

In the following $\mathbf{A}$ is a $n \times n$ matrix.

1. Prove that $\mathbf{A}$ is diagonalizable if and only if its minimal polynomial $m(x)$ has no repeated root. Discuss how to verify that a polynomial over $\mathbb{R}$ has no repeated roots.
2. Prove that the minimal and characteristic polynomial of $\mathbf{A}$ coincide (up to sign) if and only if there is a vector v such that

$$\mathbf{V} = (v, \mathbf{A}(v), \ldots, \mathbf{A}^{n-1}(v)).$$

   These matrices are also known as non-derogatory.
3. Prove that if $\mathbf{A}$ is non-derogatory, any matrix $\mathbf{B}$ that commutes with $\mathbf{A}$ is a linear combination of $\mathbf{I}, \mathbf{A}, \ldots, \mathbf{A}^{n-1}$, that is $\mathbf{B} = f(\mathbf{A})$ a polynomial in $\mathbf{A}$.

Homework

Section 7.1: 2a, 3a, 5, 7e

Quiz #8

1. Section 7.1: Problem 2a
2. Section 7.1: Problem 7e
3. Prove that **A** is diagonalizable if and only if its minimal polynomial $m(x)$ has no repeated root. Discuss how to verify that a polynomial over $\mathbb{R}$ has no repeated roots.
4. Check whether the real matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & -4 \\ 4 & 1 & 2 \end{bmatrix}$$

is diagonalizable by examining the gcd of its characteristic polynomial and its derivative.

Final Orientation

Final will be comprehensive but topics will be emphasized according to the following classification:

- **VITs: Very Important Topics**
- **BITs: Basic Important Topics**
- **LITs: Basic but Less Important Topics**

VITs

- Diagonalization of L.T.'s
- Normal Operators
- Unitary/Orthogonal Operators
- Hermitian/Symmetric Operators
- Spectral Theorems
- Jordan Canonical Forms

BITs

- Eigenvectors, Eigenvalues
- Characteristic polynomials
- Generalized eigenvectors
- Invariant subspaces
- Cayley-Hamilton theorem
- Minimal polynomial of a linear operator

LITs

- Determinants
- Bases and dimension of vector spaces
- Nullspace and range of a L.T.; dimension formula
- Orthogonality of vectors