Math 350: Linear Algebra

Wolmer V. Vasconcelos

Set 6

Fall 2010

Last Class... and Today ...

- Determinants
- Some of its Applications
- ◦ Invariant subspaces
- ◦ Eigenvectors and Eigenvalues
- ◦ Diagonalization

Outline

Consider the following differential equations (or systems of)

$$
\begin{aligned}
y' &= ay, \quad a \in \mathbb{R} \\
y'' + ay' + by &= 0, \quad a, b \in \mathbb{R} \\
\begin{bmatrix} y_1' \\ y_2' \end{bmatrix} &= \begin{bmatrix} 10y_1 + 3y_2 \\ 3y_1 + 2y_2 \end{bmatrix}
\end{aligned}
$$

Question: What are their resemblances? Which ones can we solve directly?
They are equations, or systems, of linear differential equations with constant coefficients.

The first equation, $y' = ay$, is the easiest to deal with: $y = ce^{at}$ is the general solution.

We will argue that the others, with a formulation using vectors and matrices, have the same kind of solution. Let us do the last one first:

$$\begin{bmatrix} y_1' \\ y_2' \end{bmatrix} = \begin{bmatrix} 10y_1 + 3y_2 \\ 3y_1 + 2y_2 \end{bmatrix}$$

Set

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \mathbf{Y}' = \begin{bmatrix} y_1' \\ y_2' \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}$$

Now observe:

$$\mathbf{Y}' = \mathbf{AY}.$$

$$\mathbf{Y}' = \mathbf{A}\mathbf{Y}.$$

Question: This looks like $y' = ay$, which has $y = ce^{at}$ for solution. You should be tempted to expect the solution to be

$$\mathbf{Y} = \mathbf{C}e^{t\mathbf{A}}.$$

What is $e^{t\mathbf{A}}$, the **exponential** of the matrix $t\mathbf{A}$? What could it be?

Let us turn to the second order differential equation

$$y'' + ay' + by = 0$$

If we set $z_1 = y$ and $z_2 = y' = z_1'$, $z_2' = y'' = -ay' - by = -bz_1 - az_2$ which can be written in matrix formulation as

$$\mathbf{Z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad \mathbf{Z}' = \begin{bmatrix} z_1' \\ z_2' \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0 & -b \\ 1 & -a \end{bmatrix}$$

We get

$$\mathbf{Z}' = \mathbf{A}\mathbf{Z},$$

as above $\mathbf{Z} = \mathbf{C}e^{t\mathbf{A}}$ if we could make sense of then exponential of a matrix.

We return to this–promise–for the moment just think the possibility:

The function e^x has a power series expansion

$$e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!} + \cdots$$

If we replace x by the square matrix $\mathbf{A}$ (and 1 by $\mathbf{I}$), we get

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2} + \cdots + \frac{\mathbf{A}^n}{n!} + \cdots ,$$

We just must make sure that a theory of series of makes sense. The answer will be sure. Think about the adjustments to be made.

Just for fun let us calculate the exponential of $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

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

$$\begin{bmatrix} 1 + 1 + 1/2 + \cdots + 1/n! + \cdots & 1 + \underbrace{1 + 2 \cdot 1/2 + \cdots + n \cdot 1/n! + \cdots}_{=e} \\ 0 & 1 + 1/2 + \cdots + 1/n! + \cdots \end{bmatrix}$$

$$e^{\mathbf{A}} = \begin{bmatrix} e & e \\ 0 & e \end{bmatrix}$$

Convergence of $e^{\mathbf{A}}$

That

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2} + \cdots + \frac{\mathbf{A}^n}{n!} + \cdots$$

makes sense is due to the power of $n!$:
Suppose $\mathbf{A} = [a_{ij}]$ is $m \times m$ and that the absolute value of its entries $|a_{ij}| \leq r$.This implies that the entries of $\mathbf{A}^2$

$$|\sum_{k=1}^{m} a_{ik} a_{kj}| \leq mr^2$$

Similarly one finds that the entries of $\mathbf{A}^n$ are bounded by

$$m^{n-1} r^n$$

This implies that the series in any entry of $e^{\mathbf{A}}$ is bounded by the series

$$\sum_{n=0}^{\infty} \frac{m^{n-1} r^n}{n!}$$

that is convergent [e.g. use **ratio test**].

This proves $e^{\mathbf{A}}$ makes sense since the series in each of its entries is absolutely convergent.

Let us show a long application:

$$\det(e^{\mathbf{A}}) = e^{\mathrm{Trace}(\mathbf{A})}$$

This is obvious if **A** is a diagonal matrix,

$$\mathbf{A} = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}, \quad e^{\mathbf{A}} = \begin{bmatrix} e^a & 0 & 0 \\ 0 & e^b & 0 \\ 0 & 0 & e^c \end{bmatrix}, \quad \det(e^{\mathbf{A}}) = e^{a+b+c},$$

but in general...

Outline

Sweet representation of a linear transformation

Let **V** be a finite dimensional vector space and

$$\mathbf{T} : \mathbf{V} \to \mathbf{V}$$

a linear transformation.
Question: Is there a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of **V** so that the matrix representation

$$[\mathbf{T}]_{\mathcal{B}}$$

is as 'simple' [e.g. with plenty of 0's] as possible?
Answer: Well... but for the most 'interesting' matrices the answer is YES.

Invariant subspace

Let **V** be a finite dimensional vector space and

$$\mathbf{T} : \mathbf{V} \to \mathbf{V}$$

a linear transformation.
If $\mathbf{W} \subset \mathbf{V}$ is a subspace, it is of interest to know whether for $w \in \mathbf{W}$ its image $\mathbf{T}(w) \in \mathbf{W}$. Clearly this will not happen often.

Definition

W is a **T-invariant subspace** if $\mathbf{T}(\mathbf{W}) \subset \mathbf{W}$. That is, the restriction of (the function) **T** to **W** is a linear transformation of it. We denote the restriction of **T** to **W** by $\mathbf{T}_{\mathbf{W}}$.

Let us see what this implies for the matrix representation of **T.** Let $\mathcal{B} = \{w_1, \ldots, w_r\}$ be a basis of **W**, and complete it to a basis of **V**

$$\mathcal{A} = \{w_1, \ldots, w_r, v_{r+1}, \ldots, v_n\}.$$

Since $\mathbf{T}(w_i) \in \mathbf{W}$, it is a linear combination of the first r vectors, the first r columns of the matrix is

$$[\mathbf{T}]_{\mathcal{A}} = \begin{bmatrix} \boxed{[\mathbf{T}_{\mathbf{W}}]_{\mathcal{B}}} & * & \cdots & * \\ O_{(n-r)\times r} & * & \cdots & * \end{bmatrix}$$

$$[\mathbf{T}]_{\mathcal{A}} = \begin{bmatrix} a & b & * & \cdots & * \\ c & d & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ 0 & 0 & * & \cdots & * \end{bmatrix}$$

Blocks

Suppose **T** is a L.T. of vector space **V** with a basis $\mathcal{A} = v_1, \ldots, v_r, v_{r+1}, \ldots, v_n$. Suppose $\mathbf{T}(v_i)$ for $i \leq r$, is a linear combination of the first r basis vectors, and $\mathbf{T}(v_i)$ for $i > r$, is a linear combination of the last $n - r$ basis vectors.

Claim: The matrix representation has the block format

$$[\mathbf{T}]_{\mathcal{A}} = \begin{bmatrix} \boxed{r \times r} & O \\ O & \boxed{(n-r) \times (n-r)} \end{bmatrix}$$

This can be refined to more than two blocks. The extreme case is when all blocks are 1×1. The representation is then said to be diagonal.

Eigenvector

The extreme case of an invariant subspace is one of the top 5 notions of L.A.:

Definition

An **eigenvector** of the linear transformation **T** is a **nonzero** vector v such that

$$\mathbf{T}(v) = \lambda \cdot v.$$

The scalar λ is called the (corresponding) **eigenvalue.**

Means: The line $\mathbf{F}v$ is an invariant subspace of **T.** Note that v must be **nonzero**, but that λ could be zero. Observe who comes first: **eigenvector** $\rightarrow$ **eigenvalue.**

To keep in mind:

$$v \neq O, \quad \mathbf{T}(v) = \lambda v$$

Note: Any nonzero multiple of v is also an eigenvector [with the same eigenvalue]

$$av \neq 0 \quad \mathbf{T}(av) = a\mathbf{T}(v) = a\lambda v = \lambda(av)$$

The subspace spanned by v is **invariant** under **T**

Examples

- One of the most important L.T. of Mathematics is $\mathbf{T} := \frac{d}{dt}$. (On the appropriate V.S.) Its eigenvectors are

$$\frac{d}{dt}(f(t)) = \lambda \cdot f(t),$$

that is $f(t) = e^{\lambda t}$ and its nonzero scalar multiples $ce^{\lambda t}$.

- Let $\mathbf{T}$ be the identity L.T. $\mathbf{I}$. Then any nonzero vector is a eigenvector. Same property for the [null] O mapping.

- For an angle $0 < \alpha < \pi$, let

$$\mathbf{T}(x, y) = (x \cos\alpha + y \sin\alpha, -x \sin\alpha + y \cos\alpha)$$

This is a rotation in the plane by α degrees. Clearly there is no nonzero vector v in the real plane $\mathbb{R}^2$ that is aligned with $\mathbf{T}(v)$.

- Let **T** be the L.T.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Its eigenvectors are (and their nonzero multiples)

$$\mathbf{T}(i) = 1 \cdot i, \quad \mathbf{T}(j) = 2 \cdot j, \quad \mathbf{T}(k) = 0 \cdot k$$

If $\mathbf{T}$ is a linear transformation of $\mathbf{F}^2$ with a matrix representation

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

we know that

$$\mathbf{A}^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Thus, if

$$\mathbf{A}(v) = \lambda v, \quad v \neq 0$$

$$\mathbf{A}(\mathbf{A}(v)) = \mathbf{A}^2(v) = O = \mathbf{A}(\lambda v) = \lambda(\mathbf{A}(v)) = \lambda^2 v$$

so $\lambda = 0$ since $v \neq O$.

Let **V** be the vector space of all $n \times n$ real matrices, and let **T** be the transformation

$$\mathbf{T}(\mathbf{A}) = \mathbf{A}^t$$

T is a linear transformation. If $\mathbf{A} \neq O$ is one of its eigenvectors,

$$\mathbf{A}^t = \lambda \mathbf{A}$$

So, transposing again we get

$$\mathbf{A} = (\mathbf{A}^t)^t = \lambda \mathbf{A}^t = \lambda^2 \mathbf{A}$$

$$(\lambda^2 - 1)\mathbf{A} = O$$

This means that $\lambda = \pm 1$
If $\lambda = 1$, **A** is symmetric
If $\lambda = -1$, **A** is skew-symmetric

Question:

Given a n-by-n matrix $\mathbf{A}$ [usually representing some linear transformation $\mathbf{T}$], how are the **eigenvectors** to be found? Although the **eigenvalues** come after the **eigenvectors**, in some approaches they will appear first. Look at the following analysis: $\mathbf{A}v = \lambda v$, for $v \neq O$ means that

$$(\mathbf{A} - \lambda \mathbf{I}_n)v = O,$$

Conclusion: v is a nonzero vector of the **nullspace** of $\mathbf{A} - \lambda \mathbf{I}_n$ and therefore $\operatorname{rank}(\mathbf{A} - \lambda \mathbf{I}_n) < n$. This in turn means that

$$\det(\mathbf{A} - \lambda \mathbf{I}_n) = 0.$$

Characteristic polynomial of a matrix

Definition

The **characteristic polynomial** of the n-by-n matrix $\mathbf{A} = [a_{ij}]$ is the polynomial

$$p(x) = \det(\mathbf{A} - x\mathbf{I}_n) = \det \begin{bmatrix} a_{11} - x & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} - x \end{bmatrix}.$$

The equation $p(x) = 0$ is called the **characteristic equation.**

Observe that $\det(\mathbf{A} - x\mathbf{I}_n)$ is a polynomial of degree n,

$$\det(\mathbf{A} - x\mathbf{I}_n) = (-1)^n x^n + c_{n-1}x^{n-1} + \cdots + c_0.$$

The characteristic polynomial of $\mathbf{A} = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}$ is

$$\det \begin{bmatrix} 10 - x & 3 \\ 3 & 2 - x \end{bmatrix} = (10 - x)(2 - x) - 9 = x^2 - 12x + 11$$

Its roots are

$$\lambda = \frac{12 \pm \sqrt{12^2 - 4 \times 11}}{2} = 6 \pm 5$$

With the eigenvalues in hand we solve for the eigenvectors.

$\lambda = 11$: Will determine the nullspace of $\mathbf{A} - 11\mathbf{I}_2$

$$\left[\begin{array}{cc|c} 10-11 & 3 & 0 \\ 3 & 2-11 & 0 \end{array}\right] \to \left[\begin{array}{cc|c} -1 & 3 & 0 \\ 0 & 0 & 0 \end{array}\right] \quad v_1 = \left[\begin{array}{c} 3 \\ 1 \end{array}\right]$$

$\lambda = 1$: Will determine the nullspace of $\mathbf{A} - \mathbf{I}_2$

$$\left[\begin{array}{cc|c} 10-1 & 3 & 0 \\ 3 & 2-1 & 0 \end{array}\right] \to \left[\begin{array}{cc|c} 3 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \quad v_2 = \left[\begin{array}{c} 1 \\ -3 \end{array}\right]$$

Let us Verify that it will work out for any real symmetric matrix
$\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$
The characteristic polynomial is

$$\det \begin{bmatrix} a - x & b \\ b & c - x \end{bmatrix} = (a - x)(c - x) - b^2 = x^2 - (a + c)x + ac - b^2,$$

whose roots are

$$\lambda = \frac{a + c \pm \sqrt{(a + c)^2 - 4(ac - b^2)}}{2}$$

Incredibly (?) the quantity under the sign is $(a - c)^2 + 4b^2 \geq 0$, so either there are two distinct real roots or $a = c$, $b = 0$. In both cases the matrix is diagonalizable.

A different kind is the rotation $\mathbf{R}_\alpha$ by α degrees in the plane $\mathbb{R}^2$: $\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$. Its characteristic polynomial is

$$\det \begin{bmatrix} \cos\alpha - x & -\sin\alpha \\ \sin\alpha & \cos\alpha - x \end{bmatrix} = (\cos\alpha - x)^2 + \sin^2\alpha = x^2 - (2\cos\alpha)x + 1.$$

Its roots are

$$\lambda = \frac{2\cos\alpha \pm \sqrt{4\cos^2\alpha - 4}}{2},$$

which is not real unless $\alpha = 0, \pi$.

We already know that rotations $0 < \alpha < \pi$ have no real eigenvalues.
Let us try $\alpha = \pi/2$ anyway: $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. The characteristic polynomial is $x^2 + 1$, so the (complex) eigenvalues are $\lambda = \pm i$.

$\lambda = i$: Will determine the nullspace of $\mathbf{A} - i\mathbf{I}_2$

$$\left[\begin{array}{cc|c} -i & 1 & 0 \\ -1 & -i & 0 \end{array}\right] \to \left[\begin{array}{cc|c} -i & 1 & 0 \\ 0 & 0 & 0 \end{array}\right], \quad v_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}$$

$\lambda = -i$: Will determine the nullspace of $\mathbf{A} + i\mathbf{I}_2$

$$\left[\begin{array}{cc|c} i & 1 & 0 \\ -1 & i & 0 \end{array}\right] \to \left[\begin{array}{cc|c} i & 1 & 0 \\ 0 & 0 & 0 \end{array}\right], \quad v_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$$

Proposition

Let $\mathbf{A}$ *be a n-by-n matrix over the field* $\mathbf{F}$*. A scalar* $\lambda \in \mathbf{F}$ *is an eigenvalue for some eigenvector* $v \in \mathbf{F}^n$ *iff* λ *is a root of the polynomial* $\det(\mathbf{A} - x\mathbf{I}_n)$*.*

Proof.

We have already observed that if $\mathbf{A}v = \lambda v$, $v \neq 0$, then λ is a root of the char polynomial. Conversely, if $\det(\mathbf{A} - \lambda \mathbf{I}_n) = 0$, then $\mathrm{rank}(\mathbf{A} - \lambda \mathbf{I}_n) < n$. This implies, by the dimension formula, that the nullspace of $\mathbf{A} - \lambda \mathbf{I}_n \neq O$. Any nonzero vector in this nullspace will satisfy

$$\mathbf{A}v = \lambda v.$$

□

Corollary

The number of distinct eigenvalues of the n-by-n matrix $\mathbf{A}$ *is at most n. (The set of eigenvalues of a matrix–or of a linear transformation is called its* **spectrum***).*

Characteristic polynomial of a linear transformation

It seems that we have only defined the characteristic polynomial for matrices. Suppose **T** is a L.T. If we have two bases $\mathcal{A}$, $\mathcal{B}$ of the vector space, we have two representations

$$\mathbf{A} = [\mathbf{T}]_{\mathcal{A}}, \quad \mathbf{B} = [\mathbf{T}]_{\mathcal{B}}$$

and therefore we have, apparently, two possibly different polynomials

$$\det(\mathbf{A} - x\mathbf{I}_n), \quad \det(\mathbf{B} - x\mathbf{I}_n).$$

But we proved that **A** and **B** are related: There is an invertible matrix **P** such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$. Now observe

$$\begin{aligned}
\det(\mathbf{B} - x\mathbf{I}_n) &= \det(\mathbf{P}^{-1}\mathbf{A}\mathbf{P} - x\mathbf{I}_n) = \det(\mathbf{P}^{-1}\mathbf{A}\mathbf{P} - \mathbf{P}^{-1}x\mathbf{I}_n\mathbf{P}) \\
&= \det(\mathbf{P}^{-1}(\mathbf{A} - x\mathbf{I}_n)\mathbf{P}) \\
&= \det(\mathbf{P}^{-1})\det(\mathbf{A} - x\mathbf{I}_n)\det(\mathbf{P}) \\
&= \det(\mathbf{A} - x\mathbf{I}_n)
\end{aligned}$$

Conclusion: The characteristic polynomial is the same for all representations of **T.**

Eigenspaces

Definition

If λ is an eigenvalue of $\mathbf{A}$, the nullspace of $\mathbf{A} - \lambda\mathbf{I}_n$, denoted by E_λ, is called the **eigenspace** associated to λ.

Observe that E_λ is invariant under $\mathbf{A}$: If $v \in E_\lambda$ then $\mathbf{A}v \in E_\lambda$.

Polynomials and their roots

If $f(x) = a_n x^n + \cdots + a_0$ is a polynomial of degree n, with coefficients in the field **F** a root is a scalar r such that $f(r) = 0$. It is a hard problem to find r.

Proposition

If $f(x)$ and $g(x)$ are two polynomials, then there exist polynomials $q(x)$ and $r(x)$ where

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

where $r(x) = 0$ or degree $r(x) <$ degree $g(x)$.

$q(x)$ is called the **quotient**, and $r(x)$ the **remainder** of the division of $f(x)$ by $g(x)$. They are found by the **long division algorithm**.

Corollary

If r is a root of the nonzero polynomial $f(x)$, then $f(x) = (x - r)q(x)$, where $\deg q(x) = \deg f(x) - 1$. As a consequence, a polynomial $f(x)$ of degree n has at most n roots.

Algebraic multiplicity of a root

If $f(x) = a_n x^n + \cdots + a_0$ is a nonzero polynomial and r is one of its roots,

$$f(x) = (x - r)g(x).$$

It may occur that r is a root of $g(x)$, $g(x) = (x - r)h(x)$. As the degrees of the quotients decrease, we eventually have

$$f(x) = (x - r)^s q(x), \quad q(r) \neq 0.$$

Definition

We say that r is a root of $f(x)$ of **order** or **multiplicity** s.

Multiplicities of an eigenvalue

Let λ be an eigenvalue of the matrix **A.** There are two notions of multiplicity associated to λ:

- If λ is a root of order s of the characteristic polynomial $\det(\mathbf{A} - x\mathbf{I}_n)$, we say that λ has **algebraic multiplicity** s**.**
- If the eigenspace E_λ has dimension t, we say that λ has **geometric multiplicity** t.

Proposition

For any eigenvalue λ of a matrix $\mathbf{A}$*,*

$$\textbf{algebraic multiplicity} \geq \textbf{geometric multiplicity}.$$

Proof.

Assume $v_1, \ldots, v_t$ is a basis of E_λ, and we use it as the beginning of a basis for the whole vector space, the representation of the L.T. has the block format

$$\begin{bmatrix} \lambda \mathbf{I}_t & \mathbf{B} \\ O & \mathbf{C} \end{bmatrix}, \qquad \det(\mathbf{A} - x\mathbf{I}_n) = (\lambda - x)^t \det(\mathbf{C} - x\mathbf{I}_{n-t}).$$

□

Properties of eigenvalues

Let $\mathbf{A}$ be a square matrix.

1. If λ is an eigenvalue of $\mathbf{A}$, then λ^2 is an eigenvalue of $\mathbf{A}^2$:

$$\mathbf{A}^2(v) = \mathbf{A}(\mathbf{A}(v)) = \mathbf{A}(\lambda v) = \lambda \mathbf{A}(v) = \lambda\lambda v = \lambda^2 v.$$

2. More generally, if $g(x)$ is a polynomial (e.g. $x^2 - 2x + 1$) then

$$g(\mathbf{A})(v) = g(\lambda)v, \quad (\mathbf{A}^2 - 2\mathbf{A} + \mathbf{I})(v) = (\lambda^2 - 2\lambda + 1)(v).$$

3. If $\mathbf{A}$ is invertible, $\mathbf{A}^{-1}(v) = \frac{1}{\lambda} v$.

1. If $p(x) = \det(\mathbf{A} - x\mathbf{I}_n) = (-1)^n x^n + \cdots + a_0$ is the characteristic polynomial of $\mathbf{A}$, then $a_0 = \det(\mathbf{A})$. Plug in $x = 0$ in $p(x)$.
2. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{A}$, then

$$\det(\mathbf{A}) = \lambda_1 \cdot \lambda_2 \cdots \lambda_n.$$

In the decomposition of $p(x)$,

$$p(x) = (-1)^n (x - \lambda_1) \cdots (x - \lambda_n),$$

plug in $x = 0$ and use the observation above.

Complex Numbers

1. If the field is the complex number filed $\mathbb{C}$, any polynomial $f(x) \in \mathbb{C}[x]$ factors completely

$$f(x) = a_n(x - r_1)\cdots(x - r_n)$$

   As a consequence, the eigenvalues of a complex matrix always exist in the field.

2. If **A** is a real matrix, its characteristic polynomial $p(x) = \det(\mathbf{A} - x\mathbf{I}_n)$ is a real polynomial and always have a full set $\lambda_1, \ldots, \lambda_n$ of complex eigenvalues, some of which may be real.

1. If $\lambda = a + bi$, is a complex root of $f(x)$, $f(\lambda) = 0$, observe that

$$f(a + bi) = 0 \Rightarrow f(a - bi) = 0,$$

because all coefficients of $f(x)$ are real.Let us explain: Say

$$7(a + bi)^3 - 2(a + bi)^2 + 117(a + bi) + \pi = 0.$$

Complex conjugation, $a + bi \to \overline{a + bi} = a - bi$ has the property: $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$. But if z_1, say, is real (like the coefficients of the polynomial), $\overline{z_1} = z_1$, so they are not affected by changing all $a + bi$ into $a - bi$. So if one is a root, so will be the other.

2. Thus the complex conjugate $a - bi$ of an eigenvalue $a + bi$ is also an eigenvalue: So complex eigenvalues of a real matrix occur in pairs.

Outline

Linear independence of eigenvectors

Let **T** be a L.T. (or matrix). Suppose there is a basis made up of eigenvectors, say $\mathcal{B} = \{v_1, \ldots, v_n\}$, $\mathbf{T}(v_i) = \lambda_i v_i$. The corresponding matrix representation is

$$[\mathbf{T}]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

This is not always possible: Let $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ whose characteristic polynomial is x^2. There is just one eigenvalue, $\lambda = 0$. But the corresponding eigenspace E_0 has for basis $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. We do not have a basis of eigenvectors, so **A** is not diagonalizable.

Let us explore what is needed to have a basis of eigenvectors.

Proposition

Let **T** *be a linear transformation and let* $v_1, \ldots, v_r$ *be a set of eigenvectors of* **T***, associated to distinct eigenvalues* $\lambda_1, \ldots, \lambda_r$*. Then the* v_i *are linearly independent.*

Proof. Suppose $c_1 v_1 + \cdots + c_r v_r = O$. Using induction on r, we are going to show that all $c_i = 0$. We are going to multiply the equation by λ_1 and apply $\mathbf{T}$ to it to obtain the following two equations:

$$\begin{aligned} \lambda_1(c_1 v_1 + \cdots + c_r v_r) &= \lambda_1 c_1 v_1 + \cdots + \lambda_1 c_r v_r = 0 \\ \mathbf{T}(c_1 v_1 + \cdots + c_r v_r) &= \lambda_1 c_1 v_1 + \cdots + \lambda_r c_r v_r = 0 \end{aligned}$$

If we subtract one from the other we get the shorter equation,

$$\underbrace{(\lambda_2 - \lambda_1)c_2}\, v_2 + \cdots + \underbrace{(\lambda_r - \lambda_1)c_r}\, v_r = 0$$

By the induction hypothesis, all $c_i(\lambda_i - \lambda_1) = 0$, for $i > 1$. Since $\lambda_i \neq \lambda_1$, this means $c_i = 0$ for $i > 1$. Finally, since $v_1 \neq 0$ this will imply $c_1 = 0$ as well.

Let $\lambda_1, \ldots, \lambda_r$ be the set of eigenvalues of **T**, and let $E_{\lambda_1}, \ldots, E_{\lambda_r}$ be the corresponding set of eigenspaces. For each of these we pick a basis $\mathcal{B}_i$. For simplicity, take 3 eigenvalues and assume the bases chosen for the 3 eigenspaces are

$$\{u_1, u_2, u_3\}, \{v_1, v_2\}, \{w_1, w_2\}$$

Claim: These 7 vectors are linearly independent. Suppose

$$\underbrace{a_1 u_1 + a_2 u_2 + a_3 u_3}_{u} + \underbrace{b_1 v_1 + b_2 v_2}_{v} + \underbrace{c_1 w_1 + c_2 w_2}_{w} = 0,$$

which we write as $1 \cdot u + 1 \cdot v + 1 \cdot w = 0$. Note that if $u \neq 0$ it is an eigenvector (and v and w as well), by the Proposition, $u = v = w = 0$, and then that $a_1 = \cdots = c_2 = 0$, by the linear independence of the respective bases.

Theorem

Let **A** *be a n-by-n matrix with n eigenvalues (maybe repeated). Then* **A** *is diagonalizable iff for every eigenvalue its geometric multiplicity is equal to its algebraic multiplicity.*

Proof. Let $\lambda_1, \ldots, \lambda_r$ be the set of DISTINCT eigenvalues of **A**, and let $E_{\lambda_1}, \ldots, E_{\lambda_r}$ be the corresponding set of eigenspaces. We have the equalities

$$\begin{aligned}\sum_i \textbf{geom. mult. of } \lambda_i &= \sum_i \dim E_{\lambda_i} \\ \sum_i \textbf{alg. mult. of } \lambda_i &= n.\end{aligned}$$

Since **alg. mult. of** $\lambda_i \geq$ **geom. mult. of** λ_i, if equality for each i holds, the previous discussion shows that we can have a basis of eigenvectors by collecting bases in the E_{λ_i}. The converse is clear.

Corollary

Let $\mathbf{A}$ *be a n-by-n matrix with n distinct eigenvalues. Then* $\mathbf{A}$ *is diagonalizable.*

Theorem

Let $\mathbf{A}$ *be a n-by-n matrix.* $\mathbf{A}$ *is invertible iff* $\lambda = 0$ *is not an eigenvalue.*

Proof.

$\mathbf{A}$ is invertible iff it is one-one: $\mathbf{A}(v) \neq 0 \cdot v$ if $v \neq O$. □

Let **A** be a n-by-n matrix and assume $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a basis made up of its eigenvectors, $\mathbf{A}(v_i) = \lambda_i v_i$. The matrix

$$\mathbf{P} = [v_1 | \cdots | v_n]$$

is invertible since the v_i form a basis. **Claim:**

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

To prove we apply **D** to the standard basis $e_1, \ldots, e_n$. Note that $\mathbf{P}(e_1) = v_1$. For instance

$$\mathbf{D}(e_1) = \mathbf{P}^{-1}(\mathbf{A}(\mathbf{P}(e_1))) = \mathbf{P}^{-1}(\mathbf{A}(v_1)) = \mathbf{P}^{-1}(\lambda_1 v_1) = \lambda_1 \mathbf{P}^{-1}(v_1) = \lambda_1 e_1$$

Note that if **A** is diagonalizable, that is there is an invertible matrix **P** such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$ (= diagonal), a host of related matrices are also diagonalizable:

1. Any power of **A** is diagonalizable (let us do square):

$$\mathbf{D}^2 = (\mathbf{P}^{-1}\mathbf{A}\mathbf{P})(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}) = \mathbf{P}^{-1}\mathbf{A}\underbrace{\mathbf{P}\mathbf{P}^{-1}}_{\mathbf{I}}\mathbf{A}\mathbf{P} = \mathbf{P}^{-1}\mathbf{A}^2\mathbf{P}$$

   and certainly $\mathbf{D}^2$ is diagonal.

2. If **A** is invertible [and diagonalizable!] its inverse $\mathbf{A}^{-1}$ is also diagonalizable:

$$\mathbf{D}^{-1} = (\mathbf{P}^{-1}\mathbf{A}\mathbf{P})^{-1} = \mathbf{P}^{-1}\mathbf{A}^{-1}\underbrace{(\mathbf{P}^{-1})^{-1}}_{} = \mathbf{P}^{-1}\mathbf{A}^{-1}\mathbf{P}$$

3. If $g(x)$ is any polynomial and **A** is diagonalizable, then $g(\mathbf{A})$ is diagonalizable (check).

Diagonalization Summary

Let **A** be a n-by-n matrix for which we want to find a possible diagonalization.

1. Find the characteristic polynomial $p(x) = \det(\mathbf{A} - x\mathbf{I}_n)$. Rating: **Routine**, if at times long.
2. Decompose $p(x)$ and collect factors

$$p(x) = (-1)^n (x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r}$$

   Rating: **Very Hard**
3. For each λ_i find dim E_{λ_i} and check it is m_i. Rating: **Gaussian elim**

Comment: This is kind of vague. We need predictions. That is: Guarantees that certain kinds of matrices are diagonalizable.

Examples

Example: Let **A** be the real matrix

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & c \end{bmatrix},$$

where c is some number.
(a) What are the eigenvalues of A?
(b) If $c \neq 1, 2$, why is A diagonalizable? What happens when $c = 1$ or $c = 2$?

Answer: (a) The characteristic polynomial is

$$\det(\mathbf{A} - x\mathbf{I}_3) = (2 - x)(1 - x)(c - x),$$

whose roots are the eigenvalues: $1, 2, c$.

(b) If $c \neq 1, 2$, there are [automatically] 3 independent eigenvectors and therefore the matrix is diagonalizable.

If $c = 1$ or $c = 2$, it may go either way [diagonalizable or not] so we must check further to see whether the geometric multiplicities are equal or not to the algebraic multiplicities. For $c = 1$: The nullspace of $\mathbf{A} - \mathbf{I}_3$

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

is generated by

$$\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

and **A** is not diagonalizable.
Doing likewise for $c = 2$ will again show that **A** is not diagonalizable.

Example:

Given the real matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 2 & 0 \\ 3 & 0 & 5 \end{bmatrix} \quad \mathbf{A} - x\mathbf{I}_3 = \begin{bmatrix} 2-x & 0 & 3 \\ 0 & 2-x & 0 \\ 3 & 0 & 5-x \end{bmatrix}$$

(a) Find its characteristic polynomial.
(b) Find its eigenvalues.
(c) Explain why **A** is diagonalizable. [You do not have to find the eigenvectors to answer.]

Answer: (a) To find $\det(\mathbf{A} - x\mathbf{I}_3)$, we **expand along the second column**

$$\det(\mathbf{A} - x\mathbf{I}_3) = (2-x)((2-x)(5-x)-9) = (2-x)(x^2-7x+1).$$

(b) Use the quadratic formula to find the roots of the factor $x^2 - 7x + 1$:

$$\frac{7 \pm \sqrt{49 - 4}}{2} = \frac{7 \pm 3\sqrt{5}}{2}$$

Together with 2 these roots are the eigenvalues.

(c) Since the 3 eigenvalues are distinct, we have a basis of eigenvectors for $\mathbb{R}^3$ and **A** is diagonalizable.

Chaos

Let λ be an eigenvalue of the matrix $\mathbf{A}$: $\mathbf{A}v = \lambda v$. To find $v \neq 0$ we find the nullspace of $\mathbf{A} - \lambda \mathbf{I}_n$.

Suppose a mistake was made and instead of λ we have $\lambda + \epsilon$. If this value is not an eigenvalue the nullspace of

$$\mathbf{A} - (\lambda + \epsilon)\mathbf{I}_n$$

is O, not a vector 'close' to v. What to do?

Some stability

Question: Assume **A** admits a basis of eigenvectors. How can we find one, or more eigenvectors, if we cannot solve the characteristic equation? Here is a popular technique. Let $u \in \mathbb{R}^n$ picked at random [?]. We know that

$$u = u_1 + u_2 + \cdots + u_r, \quad \mathbf{A}u_i = \lambda_i u_i$$

where the u_i belong to different eigenspaces. Of course, the right hand of this equality is invisible to us. Let us assume $|\lambda_1| > |\lambda_i|, \quad i > 1$. Observe what happens when we apply **A** repeatedly to u:

$$\mathbf{A}^n(u) = \underbrace{\lambda_1^n u_1} + \lambda_2^n u_2 + \cdots + \lambda_r^n u_r$$

The growth in the coordinates of $\mathbf{A}^n(u)$ is coming from $\lambda_1^n u_1$.

If we compare the two vectors

$$\mathbf{A}^n(u) = \underbrace{\lambda_1^n u_1} + \lambda_2^n u_2 + \cdots + \lambda_r^n u_r$$

$$\mathbf{A}^{n+1}(u) = \underbrace{\lambda_1^{n+1} u_1} + \lambda_2^{n+1} u_2 + \cdots + \lambda_r^{n+1} u_r$$

It will follow that

$$\lim_n \frac{||\mathbf{A}^{n+1}(u)||}{||\mathbf{A}^n(u)||} = |\lambda_1|,$$

more precisely: If we set $v_n = \frac{\mathbf{A}^n(u)}{||\mathbf{A}^n(u)||}$, then

$$\mathbf{A}(v_n) \simeq \lambda_1 v_n, \quad n \gg 0.$$

Let us re-visit a problem and solve it in two different ways: It is the system of differential equations

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \mathbf{Y}' = \begin{bmatrix} y_1' \\ y_2' \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}, \quad \mathbf{Y}' = \mathbf{AY}.$$

Earlier we found the eigenvalues and bases for the eigenspaces:

$$\lambda = 11: \quad v_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \lambda = 1: \quad v_2 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

If we change the coordinates

$$\mathbf{Z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad \mathbf{Y} = \underbrace{\begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix}}_{\mathbf{P}} \mathbf{Z}$$

Now observe:

$$\mathbf{Z}' = \mathbf{P}^{-1}\mathbf{Y}' = \mathbf{P}^{-1}\mathbf{AY} = (\mathbf{P}^{-1}\mathbf{AP})\mathbf{Z} = \begin{bmatrix} 11 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{Z}.$$

This is a system that is easy to solve

$$\begin{aligned} z_1' &= 11z_1 \to z_1 = c_1 e^{11x} \\ z_2' &= z_2 \to z_2 = c_2 e^{x} \end{aligned}$$

From which we get the solution

$$\mathbf{Y} = \begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} c_1 e^{11x} \\ c_2 e^{x} \end{bmatrix}$$

Another solution

Let $\mathbf{Y}' = \mathbf{AY}$ be a system of differential equations in the variable t. If it is just $y' = ay$, the solution would be $y = ce^{at}$:

$$y = ce^{ta} = c(1 + ta + t^2\frac{a^2}{2} + \cdots + t^n\frac{a^n}{n!} + \cdots)$$

Let us try the same with a matrix. If we replace a by the square matrix $\mathbf{A}$ (and 1 by $\mathbf{I}$), we get

$$e^{t\mathbf{A}} = \mathbf{I} + t\mathbf{A} + t^2\frac{\mathbf{A}^2}{2} + \cdots + \underbrace{t^n\frac{\mathbf{A}^n}{n!}} + \cdots$$

Note that the derivative of the nth term is $nt^{n-1}\frac{\mathbf{A}^n}{n!} = \mathbf{A}(t^{n-1}\frac{\mathbf{A}^{n-1}}{(n-1)!})$, and thus if $\mathbf{Y} = e^{t\mathbf{A}}$ then $\mathbf{Y}' = \mathbf{AY}$.

We just must make sure that a theory of series makes sense and taking derivatives of these expressions makes sense.

At the end we will also put in a constant: $\mathbf{Y} = e^{t\mathbf{A}}\mathbf{Y}_0$.

The expression we wrote above for $e^{t\mathbf{A}}$ is actually a set of 2^2 series, one for each cell (i,j) of the 2-by-2 matrix. That is, when we consider the sum of the terms

$$t^n \frac{\mathbf{A}^n}{n!}$$

we observe that convergence, for one, comes from the fact that the $n!$ factor grows much faster than the entries $\mathbf{A}^n_{(i,j)}$. Let us give an example. Suppose $\mathbf{A}$ is a 2-by-2 diagonal matrix with 11 and 1 on the diagonal. $\mathbf{A}^n$ is also diagonal with entries 11^n nd 1^n. Adding the series would give the matrix

$$\begin{bmatrix} e^{11t} & 0 \\ 0 & e^t \end{bmatrix} = \begin{bmatrix} 1 + 11t + 1/2(11t)^2 + \cdots & 0 \\ 0 & 1 + t + 1/2t^2 + \cdots \end{bmatrix}$$

Not only this is a nice computation, but tells us the same would work whenever $\mathbf{A}$ is a diagonal matrix. Let us show how it would work when $\mathbf{A}$ diagonalizable.

Let us show how compute $e^{t\mathbf{A}}$ if $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, with $\mathbf{D}$ diagonal.

Noting that

$$\mathbf{A}^n = \mathbf{P}\mathbf{D}^n\mathbf{P}^{-1},$$

we have

$$\begin{aligned} e^{t\mathbf{A}} = \sum \frac{t^n}{n!}\mathbf{A}^n &= \sum \frac{t^n}{n!}\mathbf{P}\mathbf{D}^n\mathbf{P}^{-1} \\ &= \mathbf{P}(\sum \frac{t^n}{n!}\mathbf{D}^n)\mathbf{P}^{-1} \\ &= \mathbf{P}e^{t\mathbf{D}}\mathbf{P}^{-1} \end{aligned}$$

Exercise: $\det e^A = e^{\mathrm{Trace}\,(\mathbf{A})}$. (This is beautiful because while we have a great deal of trouble with $e^{\mathbf{A}}$, its determinant is easy!)

Theorem

The solution of the differential equation $\mathbf{Y}' = \mathbf{AY}$ *is*

$$\mathbf{Y} = e^{t\mathbf{A}}\mathbf{C},$$

for some constant vector $\mathbf{C}$*.*

Observe where the constant goes. If you set $t = 0$, $\mathbf{Y}_0 = \mathbf{C}$, that is the components of $\mathbf{C}$ are the initial condition: $y_1(0), y_2(0)$.

Clearly the method will work for matrices of any size.

If $\mathbf{A}$ is diagonalizable we know how to compute $e^{t\mathbf{A}}$. If not ... also!

Outline

Homework

1. Section 5.1: 2d, 3c, 4h, 14, 16, 17, 21
2. Section 5.2: 2g, 8, 12, 13, 19
3. Prove that for any real $n \times n$ matrix $\mathbf{A}$, $\det(e^{\mathbf{A}}) = e^{trace(\mathbf{A})}$: First prove for $\mathbf{A}$ upper triangular, and then use the fact that there are complex matrices P and $\mathbf{B}$ such that $P^{-1}\mathbf{A}P = \mathbf{B}$, where $\mathbf{B}$ is upper triangular.

Outline

HomeQuiz #6

1. Section 5.1: 2d, 3c, 18a
2. Section 5.2: 2g, 12, 19
3. Prove that for any real $n \times n$ matrix $\mathbf{A}$, $\det(e^{\mathbf{A}}) = e^{trace(\mathbf{A})}$: First prove for $\mathbf{A}$ upper triangular, and then use the fact that there are complex matrices P and $\mathbf{B}$ such that $P^{-1}\mathbf{A}P = \mathbf{B}$, where $\mathbf{B}$ is upper triangular.

Last 2 Classes... and Today ...

- Eigenvectors and Eigenvalues
- Characteristic Polynomials
- Eigenspaces & Multiplicities
- Diagonalization
- ◦ Invariant subspaces
- ◦ Cyclic subspaces
- ◦ Cayley-Hamilton Theorem

Outline

Let $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ be a linear transformation. A **T-invariant** subspace is a subspace $\mathbf{W}$ such that $\mathbf{T}(\mathbf{W}) \subset \mathbf{W}$. This means that when we restrict the function $\mathbf{T}$ to the subspace we still get a L.T. but on a smaller space

$$\mathbf{T}_{\mathbf{W}} : \mathbf{W} \to \mathbf{W}.$$

1. the nullspace $\mathbf{W} = N(\mathbf{T})$: $\mathbf{T}(N(\mathbf{T})) = (O) \subset N(\mathbf{T})$
2. the range $\mathbf{W} = R(\mathbf{T})$: $\mathbf{T}(R(\mathbf{T})) \subset R(\mathbf{T})$
3. for any eigenvector v the line $\mathbf{W} = \mathbf{F}v$: $\mathbf{T}(v) = \lambda v \in \mathbf{W}$

It is easier to study $\mathbf{T}_{\mathbf{W}}$, and from there study $\mathbf{T}$. Let us clarify this.

Usefulness

Pick a basis $w_1, \ldots, w_r$ of **W** and enlarge it to the basis $\mathcal{A} = \{w_1, \ldots, w_r, v_{r+1}, \ldots, v_n\}$ of **V**. The matrix representation of **T**,

$$[\mathbf{T}(w_1) \ldots, \mathbf{T}(w_r), \mathbf{T}(v_{r+1}), \ldots, \mathbf{T}(v_n)]$$

has the block format

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline O & \mathbf{C} \end{array}\right]$$

where **A** is the matrix representation of $\mathbf{T_W}$.

Note how this gives that the characteristic polynomial of **T** is the product

$$\det\left[\begin{array}{c|c} \mathbf{A} - x\mathbf{I}_r & \mathbf{B} \\ \hline O & \mathbf{C} - x\mathbf{I}_{n-r} \end{array}\right] = \det(\mathbf{A} - x\mathbf{I}_r)\det(\mathbf{C} - x\mathbf{I}_{n-r})$$

Theorem

The characteristic polynomial of the restriction $\mathbf{T_W}$ *divides the characteristic polynomial of* **T**.

Cyclic invariant subspace

An effective method to find invariant subspaces for a L.T. **T** is the following: Pick a nonzero vector $w \in \mathbf{V}$ and consider the sequence

$$w, \mathbf{T}(w), \mathbf{T}^2(w), \ldots, \mathbf{T}^m(w), \ldots$$

If **V** is finite-dimensional, these vectors cannot be linearly independent, so for some m ($\leq \dim \mathbf{V}$)

$$\mathbf{T}^m(w) = c_0 w + c_1 \mathbf{T}(w) + c_2 \mathbf{T}^2(w) + \cdots + c_{m-1} \mathbf{T}^{m-1}(w).$$

Example

Let $\mathbf{T} : \mathbf{F}^3 \to \mathbf{F}^3$ be defined by

$$\mathbf{T}(a, b, c) = (-b + c, a + c, 3c), \quad w = e_1 = (1, 0, 0)$$

$$\begin{aligned} \mathbf{T}(e_1) &= (0, 1, 0) = e_2 \\ \mathbf{T}^2(e_1) = \mathbf{T}(e_2) &= (-1, 0, 0) = -e_1 \end{aligned}$$

$$(w, \mathbf{T}(w), \mathbf{T}^2(w), \ldots) = (e_1, e_2).$$

$$\mathbf{T}^2(w) = -w$$

Proposition

Let m be the smallest integer such that

$$\mathbf{T}^m(w) = c_0 w + c_1 \mathbf{T}(w) + c_2 \mathbf{T}^2(w) + \cdots + c_{m-1}\mathbf{T}^{m-1}(w).$$

Then $w, \mathbf{T}(w), \ldots, \mathbf{T}^{m-1}(w)$ *are linearly independent and span a* $\mathbf{T}$*-invariant subspace* $\mathbf{W}$*. Moreover, the characteristic polynomial of* $\mathbf{T}_\mathbf{W}$ *is* $(-1)^m(x^m - c_{m-1}x^{m-1} - \cdots - c_0)$*.*

Proof: That $w, \ldots, \mathbf{T}^{m-1}(w)$ are lin. ind., follows from the choice of m. To prove that $\mathbf{W}$ is invariant, note

$$w \to \mathbf{T}(w) \to \mathbf{T}^2(w) \to \cdots \to \mathbf{T}^{m-1}(w)$$

that the image of $\mathbf{T}^{m-1}(w)$ is $\mathbf{T}^m(w)$, which is a linear combination of $\mathbf{T}^i(w)$, $i < m$.

Now we write the matrix representation of $\mathbf{T_W}$:

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & c_0 \\ 1 & 0 & \cdots & 0 & c_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & c_{m-1} \end{bmatrix}$$

For $m = 4$ the matrix is

$$\begin{bmatrix} 0 & 0 & 0 & c_0 \\ 1 & 0 & 0 & c_1 \\ 0 & 1 & 0 & c_2 \\ 0 & 0 & 1 & c_3 \end{bmatrix}$$

whose characteristic polynomial is [expanding along the last column

$$\begin{aligned} \det \begin{bmatrix} -x & 0 & 0 & c_0 \\ 1 & -x & 0 & c_1 \\ 0 & 1 & -x & c_2 \\ 0 & 0 & 1 & c_3 - x \end{bmatrix} &= -c_0 - c_1 x + c_2(-x^2) - (c_3 - x)x^3 \\ &= x^4 - c_3 x^3 - c_2 x^2 - c_1 x - c_0. \end{aligned}$$

1. The invariant subspace $\mathbf{W} = (w, \mathbf{T}(w), \ldots, \mathbf{T}^{m-1}(w))$ is called the **cyclic** subspace generated by w.
2. The characteristic polynomial of $\mathbf{A} = \mathbf{T}_{\mathbf{W}}$ is $p(x) = (-1)^m(x^m - c_{m-1}x^{m-1} - \cdots - c_0)$. One of its properties is

$$p(\mathbf{A}) = O,$$

that is, the matrix $\mathbf{A}$ is a 'zero' of the polynomial $p(x)$.

To verify, we check that $p(\mathbf{A})(v) = O$ for every vector in $\mathbf{W}$. Since v is a lin. comb. of the $\mathbf{A}^i(w)$, $i < m$, ETS $p(\mathbf{A})(\mathbf{A}^i(w)) = O$. For $i = 0$, $p(\mathbf{A})(w) = 0$, by the choice of m. For $i > 0$,

$$p(\mathbf{A})(\mathbf{A}^i(w)) = \mathbf{A}^i(p(\mathbf{A})(w)) = O.$$

Outline

Cayley-Hamilton Theorem

Let $\mathbf{A}$ be an $n \times n$ matrix. If we consider a set of powers of $\mathbf{A}$, including $\mathbf{I}$,

$$\mathbf{I}, \mathbf{A}, \mathbf{A}^2, \ldots, \mathbf{A}^m$$

and m is large enough, say $m \geq n^2$, they cannot be linearly independent:

Reason: The list has $> n^2$ vectors of the space of $n \times n$ matrices which is of dimension n^2. This leads to the linear relation

$$c_0\mathbf{I} + c_1\mathbf{A} + \cdots + c_m\mathbf{A}^m = 0,$$

where not all c_i are zero. In other words, the nonzero polynomial

$$\mathbf{f}(x) = c_0 + c_1 x + \cdots + c_m x^m$$

has the property

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

Remark: Suppose $c_0 \neq 0$. Then from

$$c_0\mathbf{I} + c_1\mathbf{A} + \cdots + c_m\mathbf{A}^m = 0$$

we get

$$\mathbf{A}(-\frac{c_1}{c_0}\mathbf{I} - \cdots - \frac{c_m}{c_0}\mathbf{A}^{m-1}) = \mathbf{I}$$

that is

$$\mathbf{A}^{-1} = -(\frac{c_1}{c_0}\mathbf{I} + \cdots + \frac{c_m}{c_0}\mathbf{A}^{m-1})$$

The next theorem is a classic. What the **Pythagorean** is for triangles, it is for matrices.

Theorem (Cayley-Hamilton)

For a matrix $\mathbf{A}$ *of characteristic polynomial* $p(x) = \det(\mathbf{A} - x\mathbf{I})$*,* $p(\mathbf{A}) = O$*.*

This means that for any vector v, $p(\mathbf{A})(v) = O$. The proof is now easy: For $v \neq O$, consider the cyclic subspace

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

If $g(x)$ is the characteristic polynomial of the restriction of $\mathbf{A}$ on $\mathbf{W}$, we proved that

1. $g(x)$ divides $p(x)$: $p(x) = q(x)g(x)$
2. $g(\mathbf{A})(v) = O$
3. It follows that

$$p(\mathbf{A})(v) = q(\mathbf{A})(g(\mathbf{A})(v)) = q(\mathbf{A})(O) = O$$

   to prove the assertion.

What is wrong with the 'proof': Plug $x = \mathbf{A}$ in

$$\det(\mathbf{A} - x\mathbf{I}_n) = (-1)^n(x^n - a_{n-1}x^{n-1} + \cdots + (-1)^a_0)$$

$$\det \begin{bmatrix} a_{11} - x & a_{12} \\ a_{21} & a_{22} - x \end{bmatrix} = x^2 - (a_{11} + a_{22})x + a_{11}a_{22} - a_{12}a_{21}$$