Math 451: Abstract Algebra I

Wolmer V. Vasconcelos

Linear Algebra Support

Fall 2009

Outline

Matrix Representation

We first discuss how to represent some [look at the caveat] linear transformations $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ by matrices. Think of $\mathbf{V}$ and $\mathbf{W}$ as $\mathbb{R}^n$ or $\mathbb{C}^n$. It is a process akin to representing vectors by coordinates. Recall that if $v \in \mathbf{V}$ and $\mathcal{B} = v_1, \ldots, v_n$ is a basis of $\mathbf{V}$, we have a unique expression

$$v = x_1 v_1 + \cdots + x_n v_n.$$

We say that the x_i are the coordinates of v with respect to $\mathcal{B}$. We write as

$$[v]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

If $\mathcal{C} = \{w_1, \ldots, w_m\}$ is a basis of $\mathbf{W}$, we would like to find the coordinates of $\mathbf{T}(v)$ in the basis $\mathcal{C}$

$$[\mathbf{T}(v)]_{\mathcal{C}} = \begin{bmatrix} ? \end{bmatrix}.$$

Matrix Representation

In other words, if $v = x_1 v_1 + \cdots + x_n v_n$,

$$\mathbf{T}(v) = y_1 w_1 + \cdots + y_m w_m,$$

we want to describe the y_i in terms of the x_j. The process will be called a matrix representation. It comes about as follows:

$$\sum y_i w_i = T(\sum x_j v_j) = \sum x_j \mathbf{T}(v_j)$$

Thus if we have the coordinates of the $\mathbf{T}(v_j)$,

$$\mathbf{T}(v_j) = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$

we have

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \sum x_j \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$

More pictorially

$$[\mathbf{T}(v)]_{\mathcal{C}} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [\mathbf{T}]_{\mathcal{B}}^{\mathcal{C}} \cdot [v]_{\mathcal{B}}$$

The $n \times m$ matrix

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

is called the matrix representation of **T** relative to the bases $\mathcal{B}$ of **V** and $\mathcal{C}$ of **W**.

Quickly: Once bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ have been chosen, $\mathbf{T}$ is represented by

$$\left[\ a_{ij}\ \right]$$

where the entries come from

$$\mathbf{T}(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

Example

Recall the transpose operation on a square matrix $\mathbf{A}$: if a_{ij} is the (i,j)-entry of $\mathbf{A}$, the (i,j)-entry of $\mathbf{A}^t$ is a_{ji}. This is a linear transformation $\mathbf{T}$ on the space $\mathbf{M}_n(\mathbf{F})$:

$$(\mathbf{A}+\mathbf{B})^t = \mathbf{A}^t + \mathbf{B}^t, \quad (c\mathbf{A})^t = c\mathbf{A}^t.$$

Let us find its matrix representation on $\mathbf{M}_2(\mathbf{F})$. This space has the basis

$$v_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, v_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, v_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Since

$$\mathbf{T}(v_1) = v_1, \quad \mathbf{T}(v_2) = v_3, \quad \mathbf{T}(v_3) = v_2, \quad \mathbf{T}(v_4) = v_4,$$

the matrix representation of transposing is

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

Let $\mathbb{R}_3[x]$ be the space of real polynomials of degree at most 3 and **T** the differentiation operator.

A basis here are the polynomials $1, x, x^2, x^3$. The corresponding matrix representation is

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Eigenvalues: Motivation

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. 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:

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

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 D.E.

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

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

$$\mathbf{Z} = \left[\begin{array}{c} z_1 \\ z_2 \end{array}\right], \quad \mathbf{Z}' = \left[\begin{array}{c} z_1' \\ z_2' \end{array}\right], \quad \mathbf{A} = \left[\begin{array}{cc} 0 & -b \\ 1 & -a \end{array}\right]$$

We get

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

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...

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 $\mathbf{V}$ 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

$\mathbf{W}$ is a **T-invariant subspace** if $\mathbf{T}(\mathbf{W}) \subset \mathbf{W}$. That is, the restriction of (the function) $\mathbf{T}$ to $\mathbf{W}$ is a linear transformation of it. We denote the restriction of $\mathbf{T}$ to $\mathbf{W}$ 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 $\mathbf{T}$ 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 $\mathbf{T}$. Note that v must be **nonzero**, but that λ could be zero. Observe who cames 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 $\mathbf{T}$

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 $\mathbf{T}$ 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 $\mathbf{V}$ be the vector space of all $n \times n$ real matrices, and let $\mathbf{T}$ be the transformation

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

$\mathbf{T}$ 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 $\lambda = 1$, $\mathbf{A}$ is symmetric
If $\lambda = -1$, $\mathbf{A}$ 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 $\mathrm{rank}(\mathbf{A} - \lambda \mathbf{I}_n) < n$. 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 $\mathbf{A} - 11\mathbf{I}_2$

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

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

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

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

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

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

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

$\square$

Corollary

The number of distinct eigenvalues of the n-by-n matrix **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 $\mathbf{T}$ 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 $\mathbf{A}$ and $\mathbf{B}$ are related: There is an invertible matrix $\mathbf{P}$ such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. 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 $\mathbf{F}$ 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.

Proof.

Any other root s of $f(x)$ satisfies

$$f(s) = q(s)(s - r) = 0,$$

so $q(s) = 0$ since $s - r \neq 0$. □

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

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

$\square$

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 $\mathbf{A}$ 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, $a + bi \rightarrow \overline{a+bi} = a - bi$ 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.

Groups

Let **G** be a finite group. There are many injective homomorphisms

$$\varphi : \mathbf{G} \to GL_n(\mathbb{C})$$

Thus we have many ways to view **G** as a group of linear transformations. It helps a lot to know

Theorem

Every $\mathbf{T} \in \mathbf{G}$ *is diagonalizable.*

You should ask how come, when being diagonalizable is kind of dicey.

Linear independence of eigenvectors

Let $\mathbf{T}$ be a L.T. (or matrix). Suppose there is a basis made up of eigenvectors, say $\mathcal{B} = \{v_1, \dots, v_n\}$, $\mathbf{T}(v_i) = \lambda_i v_i$. 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 $\mathbf{A}$ is not diagonalizable.

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

Proposition

Let $\mathbf{T}$ *be a linear transformation and let* $v_1, \ldots, v_r$ *be a set of eigenvectors of* $\mathbf{T}$*, 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 **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 $\mathbf{T}$, 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 **A** *be a n-by-n matrix with n distinct eigenvalues. Then* **A** *is diagonalizable.*

Theorem

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

Proof.

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

Let $\mathbf{A}$ 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 $\mathbf{D}$ to the standard basis $e_1, \ldots, e_n$. Note that $\mathbf{P}(e_1) = v_1$. For instance

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

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

1. Any power of $\mathbf{A}$ 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 $\mathbf{A}$ is invertible [and diagonalizable!] its inverse $\mathbf{A}^{-1}$ is also diagonalizable:

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

3. If $g(x)$ is any polynomial and $\mathbf{A}$ 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 $\dim E_{\lambda_i}$ 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 $\mathbf{A}$ is not diagonalizable.
Doing likewise for $c = 2$ will again show that $\mathbf{A}$ 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 $\mathbf{A}$ 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

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

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 $\mathbf{A} = \mathbf{PDP}^{-1}$, with $\mathbf{D}$ diagonal.

Noting that

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

we have

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

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 $\mathbf{Y}' = \mathbf{A}\mathbf{Y}$ *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!

Homework #6

1. Let $\mathbf{A}$ be a 3×3 real matrix with entries $0, \pm 1$. Determine how large $\det \mathbf{A}$ can be. Care to consider the 4×4 version?
2. 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.

Metric properties of vector spaces

Let **V** be a vector space over the field **F**. We want to develop a geometry for **V**. For that, it is helpful to have a notion of **distance**, or **length**. We will transport and then extend numerous constructions of ordinary geometry and their calculus.

We will restrict ourselves to the cases of **F** $=\mathbb{R}$, or **F** $=\mathbb{C}$. In the case of $\mathbb{C}$, we use the standard notation for the **complex conjugate** of the complex number $z=a+bi$

$$\overline{z}=a-bi.$$

Some of its properties are:

$$\begin{aligned} z\overline{z} &= a^2+b^2 \\ \overline{z_1+z_2} &= \overline{z_1}+\overline{z_2} \\ \overline{z_1\cdot z_2} &= \overline{z_1}\cdot\overline{z_2} \end{aligned}$$

For certain operations, like solving polynomial equations, the polar representation of complex numbers

$$a + bi = r(\cos\theta + i\sin\theta), \quad r = \sqrt{a^2 + b^2}, \quad \tan\theta = \frac{a}{b}$$

is useful.For instance,

$$\sqrt{i} = \pm(\cos\pi/2 + i\sin\pi/2)^{1/2} = \pm(\cos\pi/4 + i\sin\pi/4) = \pm\frac{\sqrt{2}}{2}(1+i).$$

Inner product space

An **inner product vector space V** is a V.S. over $\mathbb{R}$ or $\mathbb{C}$ with a mapping

$$\mathbf{V} \times \mathbf{V} \to \mathbf{F}, \quad (u, v) \to \langle u, v \rangle = u \cdot v \in \mathbf{F}$$

satisfying certain conditions. Let us give an example to guide us in what is needed. Let $\mathbf{V} = \mathbb{R}^n$ and define

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

Note the properties: **bi-additive** ; $v \cdot v$ is a non-negative real number, so we can use $\sqrt{v \cdot v}$ to define the **magnitude** of v.
Question: Could we use the same formula to define an inner product for $\mathbb{C}^n$? Well... $(i) \cdot (i)$ would be -1. Of course the formula still defines a nice bilinear mapping but would not meet our need.

Dot product

Definition

An inner product vector space is a vector space with a mapping

$$\mathbf{V} \times \mathbf{V} \to \mathbf{F}, \quad (u, v) \to u \cdot v \in \mathbf{F}$$

satisfying:

1. $(u_1 + u_2) \cdot v = u_1 \cdot v + u_2 \cdot v$
2. $(cu) \cdot v = c(u \cdot v)$
3. $\overline{u \cdot v} = v \cdot u$
4. $u \cdot u > 0$ if $u \neq O$

The better notation for this product is

$$u \cdot v = \langle u, v \rangle$$

Examples

Of course, the example above of $\mathbb{R}^n$ is the grandmother of all examples. Let us modify it a bit to get an example for $\mathbb{C}^n$:

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1\overline{b_1} + \cdots + a_n\overline{b_n} = \sum_{i=1}^{n} a_i\overline{b_i}.$$

Note the properties: **additive** ; $v \cdot v$ is a non-negative real number

$$v \cdot v = \sum_{i=1}^{n} a_i\overline{a_i}$$

so we can use $\sqrt{v \cdot v}$ to define the **magnitude** of v. Note the lack of full symmetry.

Example of Function Space

Let us give an example from left field: Let **V** be the vector space of all real continuous functions on the interval $[a, b]$, and define for $f(t), g(t) \in \mathbf{V}$,

$$\langle f(t), g(t) \rangle = f(t) \cdot g(t) = \int_a^b f(t)g(t)dt.$$

An important case: If m, n are integers,

$$\begin{aligned}
\langle \sin nt, \cos mt \rangle &= \int_0^{2\pi} \sin nt \cos mt \; dt = 0 \\
\langle \sin nt, \sin mt \rangle &= \int_0^{2\pi} \sin nt \sin mt \; dt = 0, \;\; m \neq n \\
\langle \cos nt, \cos mt \rangle &= \int_0^{2\pi} \cos nt \cos mt \; dt = 0, \;\; m \neq n \\
\langle \sin nt, \sin nt \rangle &= \int_0^{2\pi} \sin^2 nt \; dt = \pi, \;\; n \neq 0
\end{aligned}$$

Example: $\mathbf{M}_n(\mathbf{F})$

Let $\mathbf{V} = \mathbf{M}_n(\mathbf{F})$ be the V.S. of all n-by-n matrices. For any such matrix $\mathbf{A} = [a_{ij}]$ define the **adjoint** of $\mathbf{A}$ (unfortunately we have already used the word for a very different notion!) to be the matrix

$$\mathbf{A}^* = [\overline{a_{ji}}],$$

that is, we transpose $\mathbf{A}$ and take the complex conjugate of each entry. Define the product (Frobenius product)

$$\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{trace}(\mathbf{A}\mathbf{B}^*) = \sum_i (\mathbf{A}\mathbf{B}^*)_{ii}.$$

It is clear that this product has the properties of an inner product. We just check the positivity condition:

$$\begin{aligned} \langle \mathbf{A}, \mathbf{A} \rangle &= \mathrm{trace}(\mathbf{A}\mathbf{A}^*) = \sum_i (\mathbf{A}\mathbf{A}^*)_{ii} \\ &= \sum \sum \cdots \overline{\cdots} = \sum |\cdots|^2 \geq 0 \end{aligned}$$

Proposition

If $\mathbf{V}$ *is an inner product space, the following hold:*

1. $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$
2. $\langle u, cv\rangle = \overline{c}\langle u, v\rangle$
3. $\langle u, O\rangle = \langle O, v\rangle = 0$
4. $\langle u, u\rangle = 0 \text{ iff } u = O$
5. $\langle u, v\rangle = \langle u, w\rangle$ *for all* $u \in \mathbf{V}$ *then* $v = w$

Proof of 1: Note

$$\begin{aligned}\langle u, v + w\rangle &= \overline{\langle v + w, u\rangle} = \overline{\langle v, u\rangle + \langle w, u\rangle} \\ &= \overline{\langle v, u\rangle} + \overline{\langle w, u\rangle} = \langle u, v\rangle + \langle u, w\rangle\end{aligned}$$

Length of a vector

Definition

Let $\mathbf{V}, \langle \cdot, \cdot \rangle$ be an inner product space. If $v \in \mathbf{V}$, the **length** or **norm** of v is the real number $||v|| = \sqrt{\langle v, v \rangle}$.

If $\mathbf{V} = \mathbb{C}^n$, $v = (a_, \ldots, a_n)$,

$$||v|| = \left[\sum_{i=1}^{n} |a_i|^2 \right]^{1/2}$$

If $\mathbf{V}$ is the space of real continuous functions on $[0, 1]$ and inner product is that we defined previously,

$$||f(t)||^2 = \int_0^1 f(t)^2 dt.$$

Framework for Geometry

The following assertions permits the construction of 'recognizable' objects in any inner product space:

Theorem

If $\mathbf{V}$ is an inner product space, then for all $u, v \in \mathbf{V}$

1. *[Cauchy-Schwarz Inequality]*

$$|\langle u, v\rangle| \leq ||u|| \cdot ||v||$$

2. *[Triangle Inequality]*

$$||u+v|| \leq ||u|| + ||v||.$$

The Cauchy-Schwarz Inequality will allow the introduction of **angles** and its **trigonometry** in **V**, while the Triangle Inequality will lead to many constructions extending those we are familiar with in 2- and 3-space.

Proofs of CSI and Δ-Inequality

To prove Cauchy-Schwarz Inequality: Note that for ANY $c \in \mathbf{F}$, $v \neq O$

$$\begin{aligned} 0 \leq ||u - cv||^2 &= \langle u - cv, u - cv \rangle = \langle u, u - cu \rangle - c\langle v, u - cv \rangle \\ &= \langle u, u \rangle - \overline{c}\langle u, v \rangle - c\langle v, u \rangle + c\overline{c}\langle v, v \rangle \end{aligned}$$

If we set $c = \frac{\langle u,v \rangle}{\langle v,v \rangle}$ the inequality becomes

$$0 \leq \langle u, u \rangle - \frac{|\langle u, v \rangle|^2}{||v||^2},$$

which proves the assertion.

For the Δ-inequality: Consider

$$
\begin{aligned}
||u+v||^2 &= \langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\
&= ||u||^2 + (\langle u,v\rangle + \overline{\langle u,v\rangle}) + ||v||^2 = ||u||^2 + 2\Re\langle u,v\rangle + ||v
\end{aligned}
$$

$$
\begin{aligned}
&\le ||u||^2 + 2|\langle u,v\rangle| + ||v||^2 \\
&\le ||u||^2 + 2||u||\cdot||v|| + ||v||^2 \quad \text{by C-S inequality} \\
&= (||u||+||v||)^2.
\end{aligned}
$$

We used that for any complex number $z = a + bi$, its real part $\Re z = a \le |z| = \sqrt{a^2+b^2}$.

Example

To illustrate the power of the axiomatic method, compare the proof above [which holds for ALL examples] with the work needed to check the inequalities just the case of the following example:

$$\left|\sum_{i=1}^{n} a_i\overline{b_i}\right| \leq \left[\sum_{i=1}^{n} |a_i|^2\right]^{1/2} \left[\sum_{i=1}^{n} |b_i|^2\right]^{1/2}$$

$$\left[\sum_{i=1}^{n} |a_i + b_i|^2\right]^{1/2} \leq \left[\sum_{i=1}^{n} |a_i|^2\right]^{1/2} + \left[\sum_{i=1}^{n} |b_i|^2\right]^{1/2}$$

Angles and Distances

Equipped with these results, we can define angles and distances, with many of the usual properties, in any inner product space. For example, for a real inner product space, the Cauchy-Schwarz inequality says that for any two [will assume nonzero] vectors u, v,

$$\langle u, v\rangle \leq ||u|| \cdot ||v||,$$

that is

$$-1 \leq \frac{\langle u, v\rangle}{||u|| \cdot ||v||} \leq 1$$

This means that the ratio can be identified to the cosine, $\cos\alpha$, of a unique angle $0 \leq \alpha \leq \pi$: So we can write

$$\langle u, v\rangle = ||u|| \cdot ||v|| \cos\alpha$$

and say that α is the angle between the vectors u and v.

An important relationship between two vectors u, v is when $\langle u, v\rangle = 0$: We then say that u and v are orthogonal or perpendicular. One notation for this situation is:

$$u \perp v$$

The **distance** between the vectors u, v is defined by

$$\mathrm{dist}(u, v) = ||u - v|| = \langle u - v, u - v\rangle^{1/2}$$

One of its properties follow from the triangle inequality: If u, v, w are three vectors

$$\mathrm{dist}(u, w) \leq \mathrm{dist}(u, v) + \mathrm{dist}(v, w).$$

Properties

These notions have numerous consequences. Let us begin with:

Proposition

Let $v_1, \ldots, v_n$ be nonzero vectors of the inner product space $\mathbf{V}$. If $v_i \perp v_j$ for $i \neq j$, then these vectors are linearly independent.

Proof.

Suppose we have a linear combination

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = O.$$

We claim all $c_i = 0$. To prove, say $c_1 = 0$, take the inner product of the linear combination with v_1:

$$c_1 \underbrace{\langle v_1, v_1 \rangle} + c_2 \underbrace{\langle v_2, v_1 \rangle} + \cdots + c_n \underbrace{\langle v_n, v_1 \rangle} = \langle O, v_1 \rangle = 0.$$

A vector v of length $||v|| = 1$ is called a **unit** vector. They are easy to find: given a nonzero vector u, $v = \frac{u}{||u||}$ is a unit vector.

A set of vectors $v_1, \ldots, v_n$ is said to be **orthonormal** if $v_i \perp v_j$, for $i \neq j$ and $||v_i|| = 1$ for any i. Of course, a good example are the ordinary coordinate vectors of 3-space.

Proposition

Let $\mathbf{V}$ *be an inner product space with an orthonormal basis* $v_1, \ldots, v_n$. *Then for any* $v \in \mathbf{V}$,

$$v = c_1 v_1 + \cdots + c_n v_n,$$

where $c_i = \langle v, v_i \rangle$. *The* c_i *are called the Fourier coefficients of* v *relative to the basis.*

Proof.

To get c_i, it suffices to form the inner product of v with v_i:

$$\langle v, v_i \rangle = c_i \langle v_i, v_i \rangle = c_i,$$

since $\langle v_i, v_i \rangle = 1$ and all other $\langle v_j, v_i \rangle = 0$. $\square$

Matrix representation

Orthonormal bases are also useful in finding the matrix representation of a L.T. $\mathbf{T} : \mathbf{V} \to \mathbf{V}$:

Let $\mathcal{A} = \{v_1, \dots, v_n\}$ be such a basis. Then $[\mathbf{T}]_{\mathcal{A}} = [a_{ij}]$ where a_{ij} are the coefficients in the expression

$$\mathbf{T}(v_j) = a_{1j} v_1 + \cdots + a_{ij} v_i + \cdots + a_{nj} v_n$$

To select a_{ij} it suffices to 'dot' with v_i

$$\langle \mathbf{T}(v_j), v_i \rangle = a_{1j} \underbrace{\langle v_1, v_i \rangle}_{=0} + \cdots + a_{ij} \underbrace{\langle v_i, v_i \rangle}_{=1} + \cdots + a_{nj} \underbrace{\langle v_n, v_i \rangle}_{=0}$$

$$[\mathbf{T}]_{\mathcal{A}} = [\langle \mathbf{T}(v_j), v_i \rangle]$$

Parallelogram Law

Exercise: If u, v are vectors of an inner product space $\mathbf{V}$, verify the parallelogram law:

$$||u+v||^2 + ||u-v||^2 = 2(||u||^2 + ||v||^2).$$

Draw a picture to illustrate this equality.

Things to come

1. We will prove that every finite-dimensional vector space $\mathbf{W}$ of an inner product space $\mathbf{V}$ has an orthonormal basis.
2. This will allow us to express the distance from a vector $v \in \mathbf{V}$ to the subspace $\mathbf{W}$. For instance, if

$$\mathbf{Ax} = \mathbf{b}$$

is a consistent system of linear equations, that is, if there is some solution $\mathbf{Ax}_0 = \mathbf{b}$, we know that the solution set is the set

$$\mathbf{x}_0 + N(\mathbf{A}),$$

where $N(\mathbf{A})$ is the nullspace of $\mathbf{A}$. Now we will be able to find the solution of smallest length, if need be.

Let us show how to obtain an orthonormal basis of a vector space from an arbitrary basis $\mathcal{A} = \{u_1, \ldots, u_n\}$.

If $n = 1$, $w_1 = \frac{u_1}{||u_1||}$ is the answer.

Assume now that we have a basis of two vectors u_1, u_2. We need to find two nonzero vectors v_1, v_2 in the span of u_1, u_2 so that $v_1 \perp v_2$. We use a projection trick: we set $v_1 = u_1$ and look for c so that

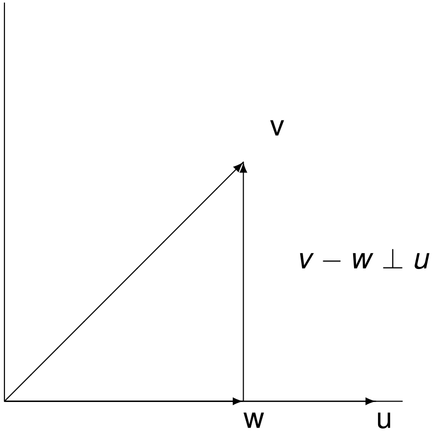
$$v_2 = u_2 - cu_1 \perp v_1,$$

that is

$$\langle v_2, v_1 \rangle = \langle u_2, v_1 \rangle - c\langle u_1, v_1 \rangle = 0$$

$$c = \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle}$$

Observe that v_1, v_2 have same span as u_1, u_2. Now replace v_i by $v_i/||v_i||$.



$w =$ Projection of v along u

Projection formula

If **L** is a line defined by the vector $u \neq O$ and v is another vector,

$$w = \frac{\langle v, u \rangle}{\langle u, u \rangle} u$$

is the **projection** of v along **L** or u.

Proposition

$v - w$ is perpendicular to **L** *and the smallest distance from v to any vector of* **L** *is $||v - w||$.*

Proof.

We have already seen that $v - w \perp v$. If cu is a vector of **L**, the square distance from v to cu is ($v - w \perp$ **L**, so will use Pythagorean Theorem)

$$||v - cu||^2 = ||(v - w) + (w + cu)||^2 = ||v - w||^2 + \underbrace{||w + cu||^2}.$$

Gram-Schmidt Algorithm

The routine to obtain a basis that is orthogonal from another basis [Gram–Schmidt process]:

1. Input: $\mathcal{A} = \{u_1, \ldots, u_n\}$ given basis
2. Set $v_1 = u_1$
3. Compute $v_2, \ldots, v_n$ successively, one at a time, by

$$v_i = \underbrace{u_i - \left(\frac{u_i \cdot v_1}{v_1 \cdot v_1}\right) v_1 - \left(\frac{u_i \cdot v_2}{v_2 \cdot v_2}\right) v_2 - \cdots - \left(\frac{u_i \cdot v_{i-1}}{v_{i-1} \cdot v_{i-1}}\right) v_{i-1}}$$

4. Set $w_i = \frac{v_i}{||v_i||}$
5. Output: $\mathcal{B} = \{w_1, \ldots, w_n\}$ is an orthonormal basis.

Hadamard's Inequality

Let **A** be a matrix whose columns form a basis $\{u_1, u_2, \ldots, u_n\}$ of $\mathbb{R}^n$ (put $n = 3$ for simplicity)

$$\mathbf{A} = [u_1 \mid u_2 \mid u_3]$$

Now consider the matrix

$$\mathbf{B} = [v_1 \mid v_2 \mid v_3] = [u_1 \mid u_2 - a_1 u_1 \mid u_3 - b_1 u_1 - b_2 u_2]$$

where the coefficients are chosen for that the $v_i's$ are perpendicular to one another. Note that **B** is obtained from **A** by adding scalar multiples of columns to another, so

$$\det(\mathbf{A}) = \det(\mathbf{B}).$$

Furthermore, for each i

$$||v_i|| \leq ||u_i||$$

by the projection formula.

Let us calculate $\det(\mathbf{A})^2$:

$$\begin{aligned}
\det(\mathbf{A})^2 &= \det(\mathbf{B})^2 = \det(\mathbf{B})\det(\mathbf{B}^t) \\
&= \det[v_1 \mid v_2 \mid v_3]\det[v_1 \mid v_2 \mid v_3]^t \\
&= \begin{bmatrix} \langle v_1, v_1 \rangle & 0 & 0 \\ 0 & \langle v_2, v_2 \rangle & 0 \\ 0 & 0 & \langle v_3, v_3 \rangle \end{bmatrix} \\
&= \prod \langle v_i, v_i \rangle
\end{aligned}$$

Theorem (Hadamard)

For any square real matrix $\mathbf{A} = [u_1, \ldots, u_n]$,

$$|\det(\mathbf{A})|^2 \leq \prod_{i=1}^{n} \langle u_i, u_i \rangle.$$

For instance, if $\mathbf{A}$ is a 4×4 whose entries are $0, 1, -1$, its column vectors have length at most 2, so that $\det(\mathbf{A}) \leq 16$. According to Joe, there is a such a matrix.

General Projection Formula

Proposition

Let $\mathbf{W}$ *be a subspace with an orthonormal basis* $\mathcal{A} = \{u_1, \ldots, u_n\}$*. For any vector* v*, the vector of* $\mathbf{W}$

$$w = \mathrm{proj}_{\mathbf{W}}(v) = \langle v, u_1 \rangle u_1 \cdots + \langle v, u_n \rangle u_n$$

is the **projection** *of* v *onto* $\mathbf{W}$*. It has the following properties*

1. $v - w$ *is perpendicular to any vector of* $\mathbf{W}$*. (We say that it is perpendicular to* $\mathbf{W}$*)*
2. $||v - w||$ *is the shortest distance from* v *to* $\mathbf{W}$*.*

The proof is like above.

Orthogonal Complement

If $\mathbf{W}$ is a subspace of an inner product space $\mathbf{V}$, its **orthogonal complement** $\mathbf{W}^{\perp}$ is the set of all vectors v that are perpendicular to each vector w of $\mathbf{W}$. In ordinary 3-space $\mathbb{R}^3$, the z-axis is the orthogonal complement of the xy-plane.

Proposition

$\mathbf{W}^{\perp}$ *is a subspace of* $\mathbf{V}$*.*

Proof.

Clearly $O \in \mathbf{W}^{\perp}$. If $v_1, v_2 \in \mathbf{W}^{\perp}$, for any vector $w \in \mathbf{W}$

$$\langle c_1 v_1 + c_2 v_2, w\rangle = c_1\langle v_1, w\rangle + c_2\langle v_2, w\rangle = O,$$

so $\mathbf{W}^{\perp}$ passes the subspace test. $\square$

Example

Let $\mathbf{A}$ be an $m \times n$ real matrix. The nullspace of $\mathbf{A}$ is the set of all n-tuples $\mathbf{x}$ such that

$$\mathbf{Ax} = 0.$$

This means that the nullspace is the orthogonal complement of the row space of $\mathbf{A}$:

$$N(\mathbf{A}) = \text{row space}^{\perp}.$$

Similarly, the **left** nullspace of $\mathbf{A}$, $\text{left } N(\mathbf{A})$, are the m-tuples $\mathbf{y}$ such that

$$\mathbf{yA} = O$$

that is the orthogonal complement of the column space of $\mathbf{A}$.

These observations suggest several properties of the $\perp$ operation:

1. Let $\mathbf{V}$ be a vector space with a basis $e_1, \ldots, e_n$. If $\mathbf{W}$ is spanned by $u_1, \ldots, u_m$, $\mathbf{W}^\perp$ is the set of all vectors $x_1 e_1 + \cdots + x_n e_n$ such that

$$x_1 \langle e_1, u_i \rangle + \cdots + x_n \langle e_n, u_i \rangle = 0, \quad i = 1, \ldots, m.$$

   Thus we find $\mathbf{W}$ by solving a system of linear equations.
2. $\mathbf{W} \cap \mathbf{W}^\perp = (O)$.
3. $\underbrace{\dim \mathbf{W} + \dim \mathbf{W}^\perp = \dim \mathbf{V}}$
4. $(\mathbf{W}^\perp)^\perp = \mathbf{W}$

Proposition

$\dim \mathbf{W} + \dim \mathbf{W}^{\perp} = \dim \mathbf{V}$.

Proof.

Let $u_1, \ldots, u_m$ be an orthonormal basis of $\mathbf{W}$. We define a mapping $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ as follows

$$\mathbf{T}(v) = \langle v, u_1 \rangle u_1 + \cdots + \langle v, u_m \rangle u_m.$$

$\mathbf{T}$ is clearly a linear transformation: This is the orthogonal projection of $\mathbf{V}$ onto $\mathbf{W}$. Its range $R(\mathbf{T})$ is $\mathbf{W}$. Its nullspace $N(\mathbf{T})$ is the set of vectors v such that $\langle v, u_i \rangle = 0$ for each u_i. This is precisely $\mathbf{W}^{\perp}$. From the dimension formula

$$\dim \mathbf{V} = \dim R(\mathbf{T}) + \dim N(\mathbf{T}) = \dim \mathbf{W} + \dim \mathbf{W}^{\perp}.$$

$\square$

Homework #7

1. Let **G** be a finite subgroup of $GL_n(\mathbb{C})$. Prove that every $\mathbf{T} \in \mathbf{G}$ is diagonalizable.
2.

If $\mathbf{V}$ is a vector space over the field $\mathbf{F}$, a linear functional is a linear transformation

$$\mathbf{f} : \mathbf{V} \longrightarrow \mathbf{F}.$$

For example, if $\mathbf{V} = \mathbf{F}^n$ and $\mathbf{a} = [a_1, \dots, a_n]$ is a matrix, then for every column vector $v \in \mathbf{F}^n$, the function

$$v \longrightarrow \mathbf{a} \cdot v$$

is a linear functional. In fact, every linear functional $\mathbf{f}$ has this description.

Inner product spaces, finite/infinite dimensional have a natural method to define linear functionals. Let us exploit it.

Let $\mathbf{V}$ be an inner product space. If $u \in \mathbf{V}$, the mapping

$$\mathbf{f} : \mathbf{V} \to \mathbf{F}, \quad \mathbf{f}(v) = \langle v, u \rangle$$

is a linear functional. Observe that if $\langle v, u \rangle = \langle v, w \rangle$, for all v, then $\langle v, u - w \rangle = 0$ and therefore $u = w$.

Proposition

If $\mathbf{V}$ is a finite-dimensional inner product space, for every linear functional $\mathbf{f}$ on $\mathbf{V}$, there is a unique vector u such that $\mathbf{f}(v) = \langle v, u \rangle$ for all $v \in \mathbf{V}$.

Proof.

Let $v_1, \ldots, v_n$ be an orthonormal basis of $\mathbf{V}$, and let

$$u = \overline{\mathbf{f}(v_1)} v_1 + \cdots + \overline{\mathbf{f}(v_n)} v_n.$$

Note that for each v_j, $\langle v_j, u \rangle = \overline{\overline{\mathbf{f}(v_j)}} = \mathbf{f}(v_j)$, so the functionals defined by u and $\mathbf{f}$ agree on each basis vector, so are

Adjoint of a Linear Transformation

Let $\mathbf{T}$ be a L.T. of the inner product space $\mathbf{V}$. We are going to build another L.T. associated to $\mathbf{T}$, which will be called the **adjoint** of $\mathbf{T}$. It is the parent [or child] of the transpose!

Fix the vector $u \in \mathbf{V}$. Consider the mapping $v \to \langle \mathbf{T}(v), u \rangle$. This is a linear functional. According to the previous Proposition, there is a unique w such that

$$\langle \mathbf{T}(v), u \rangle = \langle v, w \rangle, \qquad \forall v \in \mathbf{V}.$$

We set $w = \mathbf{S}(u)$. This gives a function $\mathbf{S} : \mathbf{V} \to \mathbf{V}$. It is routine to check that if $w_1 = \mathbf{S}(u_1)$ and $w_2 = \mathbf{S}(u_2)$, then $\mathbf{S}(u_1 + u_2) = w_1 + w_2$, and also $\mathbf{S}(cu) = c\mathbf{S}(u)$. This L.T. is denoted $\mathbf{T}^*$ and termed the adjoint of $\mathbf{T}$.

Proposition

Let $\mathbf{T}$ *be a L.T. and let* $\mathbf{A} = [a_{ij}]$ *be its matrix representation relative to the orthonormal basis* $v_1, \ldots, v_n$*. Then the matrix representation of the adjoint* $\mathbf{T}^*$ *is* $\overline{\mathbf{A}}^t = [\overline{a_{ji}}]$*, the conjugate transpose of* $\mathbf{A}$*.*

Proof.

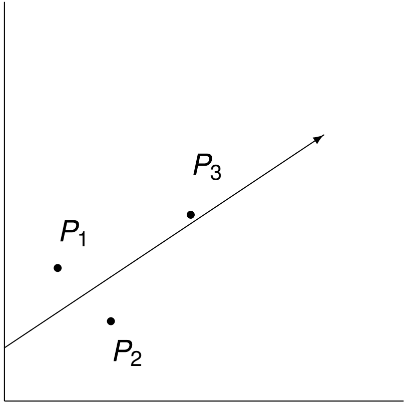
To find the matrix representation $[b_{ij}]$ of $\mathbf{T}^*$ we write $\mathbf{T}^*(v_j) = \sum_i b_{ij} v_i$, so that

$$\overline{b_{ij}} = \langle v_i, \mathbf{T}^*(v_j)\rangle = \langle \mathbf{T}(v_i), v_j\rangle = a_{ji},$$

as desired. $\square$

Problem

Given 3 (or more) points $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, $P_3 = (x_3, y_3)$ in $\mathbb{R}^2$, find the best fit line (what does this mean?):



$$Y = at + b, \quad Y_i = at_i + b, \quad \text{error} \; = |Y_i - y_i|$$

t	y	Y
t_1	y_1	Y_1
$\vdots$	$\vdots$	$\vdots$
t_n	y_n	Y_n

$$\mathbf{E} = \text{Square Error} \; = \sum_{i=1}^{n} |Y_i - y_i|^2 = \sum_{i=1}^{n} |at_i + b - y_i|^2$$

Problem: Find a and b so that the square error is as small as possible. To answer, we first write the problem in vector notation.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\mathbf{E} = ||\mathbf{y} - \mathbf{A}\mathbf{x}||^2$$

We are going to do much better: Given a $m \times n$ matrix $\mathbf{A}$ and a vector $\mathbf{y} \in \mathbf{F}^m$, we are going to find a vector $\mathbf{x}_0 \in \mathbf{F}^n$ such that

$$||\mathbf{y} - \mathbf{A}\mathbf{x}_0||^2 \leq ||\mathbf{y} - \mathbf{A}\mathbf{x}||^2$$

for all $\mathbf{x} \in \mathbf{F}^n$

We know that the answer to this will be affirmative: Let $\mathbf{W}$ be the range of $\mathbf{A}$, that is the set of all vectors $\mathbf{Ax}$, for $\mathbf{x} \in \mathbf{F}^n$. There is a vector $w \in \mathbf{W}$, that is $w = \mathbf{Ax}_0$ such that

$$||\mathbf{y} - \mathbf{Ax}_0||^2 \leq ||\mathbf{y} - \mathbf{Ax}||^2.$$

The issue is how to find $\mathbf{x}_0$ more explicitly. For this we use the notion of the adjoint of a linear transformation:

$$\mathbf{T} : \mathbf{F}^n \to \mathbf{F}^m, \quad \mathbf{T}^* : \mathbf{F}^m \to \mathbf{F}^n$$

$$\langle \mathbf{T}(u), v \rangle_m = \langle u, \mathbf{T}^*(v) \rangle_n$$

To derive the desired formula (known as the projection formula) we need two properties of $\mathbf{T}^*$.

Proposition

Let $\mathbf{A}$ be an $m \times n$ complex matrix and $\mathbf{A}^*$ its adjoint (conjugate transpose). Then

1. $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^*\mathbf{A})$.
2. *If* $\operatorname{rank}(\mathbf{A}) = n$ *then* $\mathbf{A}^*\mathbf{A}$ *is invertible.*

Proof.

It will suffice to show that $\mathbf{A}$ and $\mathbf{A}^*\mathbf{A}$ have the same nullspace. Why?

If $\mathbf{A}^*\mathbf{A}(\mathbf{x}) = 0$, then for all $\mathbf{z} \in \mathbf{F}^n$

$$0 = \langle \mathbf{A}^*\mathbf{A}(\mathbf{x}), \mathbf{z} \rangle_n = \langle \mathbf{A}\mathbf{x}, (\mathbf{A}^*)^*\mathbf{z} \rangle_m = \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{z} \rangle_m =$$

so $\mathbf{A}\mathbf{x} = O$ by choosing $\mathbf{z} = \mathbf{x}$.

The second assertion now follows: Since $\mathbf{A}^*\mathbf{A}$ is an $n \times n$ matrix of rank n, it is invertible. $\square$

Projection Formula

Theorem

Let $\mathbf{A}$ *be an* $m \times n$ *complex matrix and let* $\mathbf{y} \in \mathbf{F}^m$*. Then there exists* $\mathbf{x}_0 \in \mathbf{F}^n$ *such that* $\mathbf{A}^*\mathbf{A}(\mathbf{x}_0) = \mathbf{A}^*\mathbf{y}$ *and* $||\mathbf{A}\mathbf{x}_0 - \mathbf{y}|| \leq ||\mathbf{A}\mathbf{x} - \mathbf{y}||$ *for all* $\mathbf{x} \in \mathbf{F}^n$*. If* $\mathbf{A}$ *has rank* n *then*

$$\mathbf{x}_0 = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{y}.$$

Proof.

Since $\mathbf{A}\mathbf{x}_0 - \mathbf{y}$ is perpendicular to the range of $\mathbf{A}$,

$$0 = \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}_0 - \mathbf{y} \rangle_m = \langle \mathbf{x}, \mathbf{A}^*(\mathbf{A}\mathbf{x}_0 - \mathbf{y}) \rangle = \langle \mathbf{x}, ((\mathbf{A}^*\mathbf{A})\mathbf{x}_0 - \mathbf{A}^*\mathbf{y}) \rangle$$

for all $\mathbf{x} \in \mathbf{F}^n$. Thus $(\mathbf{A}^*\mathbf{A})\mathbf{x}_0 - \mathbf{A}^*\mathbf{y} = 0$ and therefore

$$\mathbf{x}_0 = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{y},$$

Illustration

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}, \quad \operatorname{rank}(\mathbf{A}) = 2, \quad \mathbf{y} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \end{bmatrix}$$

$$\mathbf{A}^*\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix}$$

$$(\mathbf{A}^*\mathbf{A})^{-1} = \frac{1}{20} \begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix}$$

$$\mathbf{x}_0 = \begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 1.7 \\ 0 \end{bmatrix}$$

Answer: The least squares line is

$$y = 1.7t$$

The error is

$$\mathbf{E} = ||\mathbf{A}\mathbf{x}_0 - \mathbf{y}||^2 = 0.3$$

The method is very general: Suppose we are given a number of points and we want to fit a quadratic polynomial

$$Y = at^2 + bt + c$$

to the data.

$$\mathbf{A} = \begin{bmatrix} t_1^2 & t_1 & 1 \\ \vdots & \vdots & \vdots \\ t_n^2 & t_n & 1 \end{bmatrix} \quad \mathbf{x}_0 = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Now $\mathrm{rank}(\mathbf{A}) = 3$ if there are 3 distinct values of t.

Shortest solution

We are going to find the **shortest** solution of a consistent system of equations ($m \times n$)

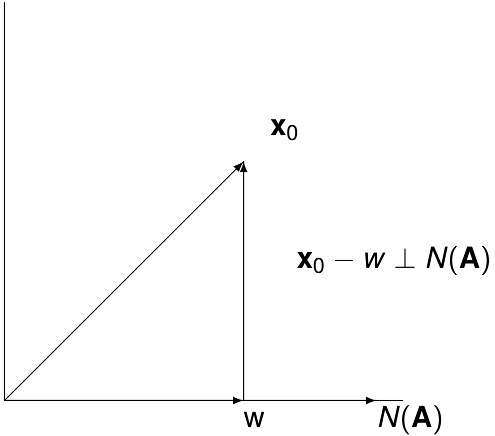
$$\mathbf{A}\mathbf{x} = \mathbf{b}.$$

This will be a solution u such that $||u||$ is minimal. The argument will also show that u is unique.

Let $\mathbf{x}_0$ be a special solution and denote by $N(\mathbf{A})$ the **nullspace** of $\mathbf{A}$. The solution set is

$$\mathbf{x}_0 + N(\mathbf{A}) = \{\mathbf{x}_o + v, \quad v \in N(\mathbf{A})\}.$$

To pick out of this set the vector $\mathbf{x}_0 + v$ of smallest length, note that $||\mathbf{x}_0 + v||$ is the distance from $\mathbf{x}_0$ to $-v$. So we have our answer: Pick for $-v$ the projection w of $\mathbf{x}_0$ into $N(\mathbf{A})$. Then $s = \mathbf{x}_0 - w$ is the desired solution:



$w =$ Projection of $\mathbf{x}_0$ along $N(\mathbf{A})$

One algorithm for the shortest solution

1. Find an orthonormal basis $u_1, \ldots, u_r$ for $N(\mathbf{A})$
2. Determine the projection w of $\mathbf{x}_0$ onto $N(\mathbf{A})$:

$$w = \sum_{i=1}^{r} \langle \mathbf{x}_0, u_i \rangle u_i$$

3. $\mathbf{x}_0 - w$ is the shortest solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$

This solution requires the calculation of the projection of $\mathbf{x}_0$ into $N(\mathbf{A})$. Let us discuss another, more direct, approach. If $v \in N(\mathbf{A})$, $\mathbf{A}(v) = O$,

$$0 = \langle \mathbf{x}, \mathbf{A}(v) \rangle = \langle \mathbf{A}^*(\mathbf{x}), u \rangle$$

which means $v \perp \mathbf{A}^*(\mathbf{x}) = 0$ for all $\mathbf{x}$. This means that the range of $\mathbf{A}^*$, $R(\mathbf{A}^*)$, is contained in the orthogonal complement $N(\mathbf{A})^\perp$ of $N(\mathbf{A})$. By the dimension formula we have $N(\mathbf{A})^\perp = R(\mathbf{A}^*)$.

Summary: The minimal vector s satisfies

$$\mathbf{A}s = \mathbf{b}, \quad s \in R(\mathbf{A}^*)$$

That is, pick any solution of

$$\mathbf{A}\mathbf{A}^*\mathbf{y} = \mathbf{b},$$

and set

$$s = \mathbf{A}^*\mathbf{y}.$$

Homework #9

Section 6.3: 3a, 6, 10, 13, 18, 22a, 23

Today

1. Normal Operators ($\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$): real symmetric/skew symmetric
2. Hermitian Operator
3. Unitary Operator ($\mathbf{T}\mathbf{T}^* = \mathbf{I} = \mathbf{T}^*\mathbf{T}$): Orthogonal
4. Spectral Theorem
5. Goodies: Applications

Interesting diagonalizable operators

We are going to show a class of linear transformations that are diagonalizable. It will include the class represented by real symmetric matrices.

Let $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ be a L.T. of a complex inner product space. We have defined the **adjoint** $\mathbf{T}^*$ of $\mathbf{T}$ as the L.T. with the property

$$\langle \mathbf{T}(u), v \rangle = \langle u, \mathbf{T}^*(v) \rangle, \quad \forall u, v \in \mathbf{V}.$$

Let us compare the eigenvalues and eigenvectors of $\mathbf{T}$ and $\mathbf{T}^*$:

Proposition

If λ is an eigenvalue of $\mathbf{T}$ then $\overline{\lambda}$ is an eigenvalue of $\mathbf{T}^*$.

Proof: Suppose $\mathbf{T}(u) = \lambda u$, $u \neq O$. Then for any $v \in \mathbf{V}$,

$$\begin{aligned} 0 = \langle O, v \rangle = \langle (\mathbf{T} - \lambda \mathbf{I})(u), v \rangle &= \langle u, (\mathbf{T} - \lambda \mathbf{I})^*(v) \rangle \\ &= \langle u, (\mathbf{T}^* - \overline{\lambda} \mathbf{I})(v) \rangle \end{aligned}$$

This says that $O \neq u \perp \text{range}(\mathbf{T}^* - \overline{\lambda}\mathbf{I})$, so the range of $\mathbf{T}^* - \overline{\lambda}\mathbf{I}$ is not the whole of $\mathbf{V}$, which implies nullspace of $\mathbf{T}^* - \overline{\lambda}\mathbf{I} \neq O$. This means that $\overline{\lambda}$ is an eigenvalue of $\mathbf{T}^*$.

Let us use this result to decide when a L.T. **T** of an inner product space **V** admits a basis $\mathcal{A}$ such that

$$[\mathbf{T}]_{\mathcal{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix},$$

that is, **T** admits a matrix representation that is upper triangular.

Note that the characteristic polynomial has all of its roots in the field

$$\det(\mathbf{T} - x\mathbf{I}) = (a_{11} - x)(a_{22} - x) \cdots (a_{nn} - x),$$

that is the characteristic polynomial splits. Recall that this is always the case when the field is $\mathbb{C}$.

Theorem (Schur)

Let $\mathbf{T}$ be a L.T. of the inner product space $\mathbf{V}$. If the characteristic polynomial of $\mathbf{T}$ splits, then $\mathbf{V}$ admits an orthonormal basis $\mathcal{A}$ such that $[\mathbf{T}]_{\mathcal{A}}$ is upper triangular.

Proof: We will argue by induction on $\dim \mathbf{V} = n$. If $n = 1$, the assertion is obvious. Let us assume that the assertion holds for dimension $n - 1$. By the Proposition above, we know that $\mathbf{T}^*$ has one eigenvalue λ. Let u be a unit vector so that $\mathbf{T}^*(u) = \lambda u$, and set $\mathbf{W}$ for the subspace spanned by u. We claim that $\mathbf{W}^\perp$ is $\mathbf{T}$-invariant: If $v \in \mathbf{W}^\perp$

$$\begin{aligned} \langle \mathbf{T}(v), u \rangle &= \langle v, \mathbf{T}^*(u) \rangle = \langle v, \lambda u \rangle \\ &= \overline{\lambda} \langle v, u \rangle = 0 \end{aligned}$$

So $\mathbf{T}(v) \in \mathbf{W}^\perp$.

We also have $\dim W + \dim \mathbf{W}^\perp = \dim \mathbf{V} = n$, so $\dim \mathbf{W}^\perp = n - 1$. Now we apply the induction hypothesis to the restriction of $\mathbf{T}$ to $\mathbf{W}^\perp$: Let $v_1, \ldots, v_{n-1}$ be an orthonormal basis of $\mathbf{W}^\perp$ for which the restriction of $\mathbf{T}$ is upper triangular. If we add to the v_i the vector u, we get the orthonormal basis $\mathcal{A} = v_1, \ldots, v_{n-1}, u$. The matrix representation

$$[\mathbf{T}]_\mathcal{A} = \begin{bmatrix} & & & a_{1n} \\ & [\mathbf{T}]_{\mathbf{W}^\perp} & & \vdots \\ & & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix},$$

which has the desired form.

Normal operator

Observe that if there is an orthonormal basis $\mathcal{A}$ of eigenvectors of $\mathbf{T}$, $[\mathbf{T}]_\mathcal{A}$ is a diagonal matrix, and since $[\mathbf{T}^*]_\mathcal{A} = [\mathbf{T}]^*_\mathcal{A}$, this matrix is also diagonal. Since diagonal matrices commute, we have $\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$.

Definition

A linear transformation $\mathbf{T}$ of an inner product space is **normal** if $\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$.

Example: If $\mathbf{A}$ is a symmetric real matrix, $\mathbf{A}^* = \mathbf{A}^t = \mathbf{A}$, so $\mathbf{A}$ commutes with itself! Skew-symmetric real matrices, $\mathbf{A}^* = -\mathbf{A}$, are also normal.

Theorem

If $\mathbf{T}$ *is a normal operator* ($\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$) *of a complex inner vector space* $\mathbf{V}$*, then there is an orthonormal basis of eigenvectors of* $\mathbf{T}$*. (The converse was proved already so this is a characterization of normal operators.)*

This is an important result, it has many useful consequences. To prove it we shall need some properties of normal operators.

Proposition

Let $\mathbf{T}$ *be a normal operator* ($\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$) *of the inner vector space* $\mathbf{V}$. *Then:*

1. $||\mathbf{T}(u)|| = ||\mathbf{T}^*(u)||$ *for every* $u \in \mathbf{V}$.
2. $\mathbf{T} - c\mathbf{I}$ *is normal for every* $c \in \mathbf{F}$.
3. *If* $\mathbf{T}(u) = \lambda u$ *then* $\mathbf{T}^*(u) = \overline{\lambda} u$.
4. *If* λ_1 *and* λ_2 *are distinct eigenvalues of* $\mathbf{T}$ *with corresponding eigenvectors* u_1 *and* u_2, *then* $u_1 \perp u_2$.

Proof: 1. For any vector $u \in \mathbf{V}$,

$$\begin{aligned} ||\mathbf{T}(u)||^2 &= \langle \mathbf{T}(u), \mathbf{T}(u)\rangle = \langle \mathbf{T}^*\mathbf{T}(u), u\rangle = \langle \mathbf{T}\mathbf{T}^*(u), u\rangle \\ &= \langle \mathbf{T}^*(u), \mathbf{T}^*(u)\rangle = ||\mathbf{T}^*(u)||^2 \end{aligned}$$

2. $(\mathbf{T} - c\mathbf{I})(\mathbf{T}^* - \overline{c}\mathbf{I}) = (\mathbf{T}^* - \overline{c}\mathbf{I})(\mathbf{T} - c\mathbf{I})$: check

3. Suppose $\mathbf{T}(u) = \lambda u$. Let $\mathbf{U} = \mathbf{T} - \lambda \mathbf{I}$. Then $\mathbf{U}(u) = 0$ so by 2. $\mathbf{U}$ is normal and by 1. $\mathbf{U}^*(u) = 0$. That is $\mathbf{T}^*(u) = \overline{\lambda} u$.

4. Let λ_1 and λ_2 be distinct eigenvalues of $\mathbf{T}$ with corresponding eigenvectors u_1 and u_2. Then by 3.

$$\lambda_1 \langle u_1, u_2 \rangle = \langle \lambda_1 u_1, u_2 \rangle = \langle \mathbf{T}(u_1), u_2 \rangle = \langle u_1, \mathbf{T}^*(u_2) \rangle$$

$$= \langle u_1, \overline{\lambda_2} u_2 \rangle = \lambda_2 \langle u_1, u_2 \rangle.$$

Since $\lambda_1 \neq \lambda_2$, $\langle u_1, u_2 \rangle = 0$.

We are now in position to prove that a normal operator $\mathbf{T}$ admits an orthonormal basis $v_1, v_2, \ldots, v_n$ of eigenvectors. We already know, by Schur theorem, that there is an orthonormal basis for which the matrix representation is upper triangular

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

We want to show that the off-diagonal elements are 0, that is, all the v_i are eigenvectors. [For simplicity we take $n = 3$] Note that $\mathbf{T}(v_1) = a_{11}v_1$, so v_1 is an eigenvector. To show v_2 is an eigenvector notice that

$$\mathbf{T}(v_2) = a_{12}v_1 + a_{22}v_2$$

We must show $a_{12} = 0$.

$$\mathbf{T}(v_2) = a_{12}v_1 + a_{22}v_2$$

We must show $a_{12} = 0$:

$$a_{12} = \langle \mathbf{T}(v_2), v_1 \rangle = \langle v_2, \mathbf{T}^*(v_1) \rangle = \langle v_2, \overline{a_{11}} v_1 \rangle = a_{11} \langle v_2, v_1 \rangle = 0$$

as desired. Now with v_1, v_2 eigenvectors, we show that $a_{13} = a_{23} = 0$. We consider

$$\mathbf{T}(v_3) = a_{13}v_1 + a_{23}v_2 + a_{33}v_3$$

The proof is similar: For instance

$$a_{23} = \langle \mathbf{T}(v_3), v_2 \rangle = \langle v_3, \mathbf{T}^*(v_2) \rangle = \langle v_3, \overline{a_{22}} v_2 \rangle = a_{22} \langle v_3, v_2 \rangle = 0$$

We have already remarked that real symmetric matrices, $\mathbf{A} = \mathbf{A}^t$, are normal. It turns out that **complex** symmetric matrices are not always normal. Truly the complex cousins of real symmetric matrices are called:

Definition

Let $\mathbf{T}$ be a linear operator of the inner product space $\mathbf{V}$. $\mathbf{T}$ is called **self-adjoint** (**Hermitian**) if $\mathbf{T} = \mathbf{T}^*$.

$$\mathbf{A} = \begin{bmatrix} 2 & 3+5i \\ 3-5i & 6 \end{bmatrix}$$

Lemma

Let **T** *be a self-adjoint linear operator of the inner product space* **V***. Then*

1. *Every eigenvalue is real.*
2. *If* **V** *is a real vector space then the characteristic polynomial splits.*

Proof: 1. Suppose $\mathbf{T}(u) = \lambda u$, $u \neq O$. By a previous result, $\mathbf{T}^*(u) = \overline{\lambda} u$. Since $\mathbf{T} = \mathbf{T}^*$, λ is real.

2. Let $n = \dim \mathbf{V}$, $\mathcal{B}$ an orthonormal basis of **V** and $\mathbf{A} = [\mathbf{T}]_{\mathcal{B}}$. Then **A** is self-adjoint. Let $\mathbf{T_A}$ be the linear operator of $\mathbb{C}^n$ defined by $\mathbf{T_A}(u) = \mathbf{A}u$ for all $u \in \mathbb{C}^n$.

Note that $\mathbf{T_A}$ is self-adjoint because $[\mathbf{T_A}]_{\mathcal{C}} = \mathbf{A}$, where $\mathcal{C}$ is the standard (orthonormal) basis of $\mathbb{C}^n$. So the eigenvalues of $\mathbf{T_A}$ are real. Since the characteristic polynomial of $\mathbf{T_A}$ is equal to the characteristic polynomial of $\mathbf{A}$, which is equal to the characteristic of $\mathbf{T}$, the characteristic polynomial of $\mathbf{T}$ splits. What we are saying is the following: Let $\mathbf{A}$ be a $n \times n$ symmetric real matrix and employ it to define a L.T. of the **complex** vector space $\mathbb{C}^n$

$$\mathbf{T} = \mathbf{T_A} : \mathbb{C}^n \to \mathbb{C}^n, \quad \mathbf{T}(u) = \mathbf{A}(u).$$

Note $\det(\mathbf{T} - x\mathbf{I}) = \det(\mathbf{A} - x\mathbf{I})$.

First Main Theorem of the Course

Theorem

Let $\mathbf{T}$ *be a linear operator on the finite-dimensional inner product space* $\mathbf{V}$. *Then* $\mathbf{T}$ *is self-adjoint if and only if there exists an orthonormal basis of* $\mathbf{V}$ *consisting of eigenvectors of* $\mathbf{T}$.

Unitary Operators

Definition

A linear operator **T** of the inner product space **V** is called **unitary** if $\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T} = \mathbf{I}$. If **V** is a real inner product space, **T** is called **orthogonal**.

The rotation operator

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

is a major example.

If **A** is a complex n-by-n matrix and $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A} = \mathbf{I}$, the column vectors of **A** form an orthonormal basis of $\mathbb{C}^n$.
We now develop quickly some basic properties of these operators.

Theorem

Let $\mathbf{T}$ be a linear operator of the finite-dimensional inner product space $\mathbf{V}$. TFAE:

1. $\mathbf{T}$ *is an unitary operator:* $\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T} = \mathbf{I}$.
2. $\langle \mathbf{T}(u), \mathbf{T}(v) \rangle = \langle u, v \rangle$ *for all* $u, v \in \mathbf{V}$.
3. *For every orthonormal basis* $\mathcal{B} = v_1, \ldots, v_n$ *of* $\mathbf{V}$, $\mathbf{T}(v_1), \ldots, \mathbf{T}(v_n)$ *is also an orthonormal basis of* $\mathbf{V}$.
4. *For some orthonormal basis* $\mathcal{B} = v_1, \ldots, v_n$ *of* $\mathbf{V}$, $\mathbf{T}(v_1), \ldots, \mathbf{T}(v_n)$ *is also an orthonormal basis of* $\mathbf{V}$.
5. $||\mathbf{T}(u)|| = ||u||$ *for every* $u \in \mathbf{V}$.

Proof. $1 \Rightarrow 2, 3, 4, 5$: (Other $\Rightarrow$ LTR)

$$\langle u, v \rangle = \langle \mathbf{T}^*\mathbf{T}(u), v \rangle = \langle \mathbf{T}(u), (\mathbf{T}^*)^*(v) \rangle = \langle \mathbf{T}(u), \mathbf{T}(v) \rangle.$$

$$\delta_{ij} = \langle v_i, v_j \rangle = \langle \mathbf{T}(v_i), \mathbf{T}(v_j) \rangle.$$

Properties of unitary operators

Let **T** be an unitary operator of the inner product space **V**.

1. The eigenvalues of **T** have length 1: If $\mathbf{T}(u) = \lambda u$,

$$\langle u, u \rangle = \langle \mathbf{T}(u), \mathbf{T}(u) \rangle = \langle \lambda u, \lambda u \rangle = \overline{\lambda}\lambda \langle u, u \rangle$$

   and thus $\overline{\lambda}\lambda = 1$.
2. If **A** is a matrix representation of **T**, $|\det(\mathbf{A})| = 1$:$\det(\mathbf{A})\det(\mathbf{A}^*) = 1$
3. If **T** is orthogonal, $\det(\mathbf{A}) = \pm 1$.
4. If **T** and **U** are unitary operators, then $\mathbf{T}^*$ and $\mathbf{T} \circ \mathbf{U}$ are also unitary operators.

Orthogonal operators of $\mathbb{R}^2$

We have already mentioned rotations, R_α. Let us analyze the possibilities. Let

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = [v_1|v_2] \quad ||v_1|| = ||v_2|| = 1, \quad v_1 \perp v_2$$

be an orthogonal matrix. This means

$$a^2 + c^2 = 1, \quad b^2 + d^2 = 1, \quad ab + cd = 0$$

We can set $a = \cos\alpha, c = \sin\alpha$ and $b = \cos\beta, d = \sin\beta$ so that

$$ab + cd = \cos\alpha\cos\beta + \sin\alpha\sin\beta = \cos(\alpha - \beta) = 0.$$

This means that $\alpha - \beta = \pm\pi/2$. The two possibilities lead to

$$R_\alpha = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{bmatrix}$$

To analyze

$$\mathbf{T} = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{bmatrix}$$

we look at its eigenvalues:

$$\det(\mathbf{T} - x\mathbf{I}) = \begin{bmatrix} \cos\beta - x & \sin\beta \\ \sin\beta & -\cos\beta - x \end{bmatrix} = x^2 - 1$$

So $\lambda = \pm 1$. This means we have an orthonormal basis v_1, v_2, and $\mathbf{T}(v_1) = v_1$, $\mathbf{T}(v_2) = v_2$.
Thus the line $\mathbb{R}v_1$ is fixed under $\mathbf{T}$, and the perpendicular line $\mathbb{R}v_2$ is flipped about $\mathbb{R}v_1$. These transformations are called **reflections**.

Summary: If $\mathbf{A}$ is an orthogonal 2-by-2 matrix, then if $\det\mathbf{A} = 1$, it is a rotation, and if $\det\mathbf{A} = -1$, it is a reflection.

Matrix product and dot product

Let u and v be two vectors of $\mathbb{R}^n$. Their **dot product**

$$u \cdot v = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

can be expressed as a **matrix product**

$$u^t v = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Keep in mind

$$u^t v = u \cdot v$$

Spectral Decomposition

Let **A** be a n-by-n symmetric real matrix, $\mathbf{P} = [v_1|\cdots|v_n]$ a matrix whose columns form an orthonormal basis of eigenvectors of **A**:

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^t = [v_1|\cdots|v_n] \cdot \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \cdot \begin{bmatrix} v_1^t \\ \hline \vdots \\ \hline v_n^t \end{bmatrix}$$

Instead of this representation of **A** as a product of 3 matrices, we are going to express **A** as a **sum** of simple matrices of rank 1.

Expanding we get

$$
\begin{aligned}
\mathbf{A} &= \mathbf{PDP}^t = [v_1|\cdots|v_n]\cdot\begin{bmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{bmatrix}\cdot\begin{bmatrix}v_1^t\\ \hline \vdots\\ \hline v_n^t\end{bmatrix}\\
&= [\lambda_1 v_1|\cdots|\lambda_n v_n]\cdot\begin{bmatrix}v_1^t\\ \hline \vdots\\ \hline v_n^t\end{bmatrix}\\
&= \lambda_1 v_1 v_1^t+\cdots+\lambda_n v_n v_n^t\\
&= \sum\lambda_i\mathbf{P}_i,\quad \mathbf{P}_i = v_i v_i^t.
\end{aligned}
$$

Let us examine the matrices $\mathbf{P}_i$.

1. $\mathbf{P}_i$ has rank 1 and is symmetric

$$\mathbf{P}_i = v_i v_i^t, \quad \mathbf{P}_i^t = (v_i v_i^t)^t = (v_i^t)^t v_i^t = \mathbf{P}_i$$

2. $\mathbf{P}_i$ is a projection

$$\mathbf{P}_i \mathbf{P}_i = (v_i v_i^t)(v_i v_i^t) = v_i (v_i^t v_i) v_i^t = v_i v_i^t = \mathbf{P}_i$$

   since $v_i^t v_i = \langle v_i, v_i \rangle = 1$

3. $\mathbf{P}_i \mathbf{P}_j = O$ for $i \neq j$

$$\mathbf{P}_i \mathbf{P}_j = (v_i v_i^t)(v_j v_j^t) = v_i (v_i^t v_j) v_j^t = O$$

   since $v_i^t v_j = \langle v_i, v_j \rangle = 0$

The equality

$$\mathbf{A} = \sum \lambda_i \mathbf{P}_i, \mathbf{P}_i = v_i v_i^t$$

is called the **spectral decomposition** of **A**.

Example: Let $\mathbf{A} = \begin{bmatrix} 3 & -4 \\ -4 & -3 \end{bmatrix}$

The eigenvalues are 5 and -5, with corresponding [normalized] eigenvectors

$$v_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\mathbf{P}_1 = v_1 v_1^t = \begin{bmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{bmatrix}, \quad \mathbf{P}_2 = v_2 v_2^t = \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix}$$

Exercise:

Let $\mathbf{A}$ be a real symmetric matrix. Prove that there is a symmetric matrix $\mathbf{B}$ such that $\mathbf{B}^3 = \mathbf{A}$.

We know that there is an orthonormal basis $v_1, \ldots, v_n$ of eigenvectors of $\mathbf{A}$. The matrix $\mathbf{P} = [v_1|\cdots|v_n]$ is orthogonal [i.e. $\mathbf{P}^{-1} = \mathbf{P}^t$] and

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

is a real diagonal matrix. Let $\mathbf{E}$ be a real 'cubic root' of $\mathbf{D}$ (if a diagonal entry of $\mathbf{D}$ is d_{ii}, the corresponding entry of $\mathbf{E}$ is the real root ${d_{ii}}^{1/3}$).
Set $\mathbf{B} = \mathbf{P}^{-1}\mathbf{E}\mathbf{P}$. Note

$$\mathbf{B}^t = (\mathbf{P}^{-1}\mathbf{E}\mathbf{P})^t = \mathbf{P}^t\mathbf{E}^t(\mathbf{P}^{-1})^t = \mathbf{P}^{-1}\mathbf{E}\mathbf{P} = \mathbf{B}, \quad \mathbf{B}^3 = \mathbf{P}^{-1}\mathbf{E}^3\mathbf{P} = \mathbf{A}.$$

Exercise: Let $\mathbf{A}$ be skew-symmetric matrix. Prove that $\det \mathbf{A} \geq 0$. *Hint:* Recall that $\mathbf{A}$ is normal, then pair up the complex eigenvalues of $\mathbf{A}$. Moreover, show that if $\mathbf{A}$ has integer entries, then $\det \mathbf{A}$ is the square of an integer.

Real quadratic forms

A real **quadratic form** in n variables is a polynomial

$$\mathbf{q}(\mathbf{x}) = \sum_{i,j} a_{ij} x_i x_j.$$

They occur in the elementary theory of conic sections–e.g. what is $10x^2 + 6xy + 2y^2 = 5$, an ellipse, a parabola, or a hyperbola?– but also in the theory of max and min of functions $\mathbf{f}(x_1, \ldots, x_n)$ of several variables. In both endeavors, a solution arises after an appropriate change of variables, $\mathbf{x} = \mathbf{P}(\mathbf{y})$,

$$\mathbf{q}(\mathbf{x}) = \mathbf{q}(\mathbf{P}(\mathbf{y})) = \sum_i d_i y_i^2.$$

Let us see how this comes about:

Let us begin with $Ax^2 + Bxy + Cy^2$, which we write as $ax^2 + 2bxy + cy^2$. (For general fields this would require $2 \neq 0$.) Now look:

$$\begin{aligned} ax^2 + 2bxy + cy^2 &= x(ax + by) + y(bx + cy) \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \mathbf{x}^t \mathbf{Q} \mathbf{x} \end{aligned}$$

where $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\mathbf{Q}$ is a symmetric matrix.

It is routine to verify that every quadratic form $\mathbf{q}(\mathbf{x})$ has such a representation,

$$\mathbf{q}(\mathbf{x}) = \mathbf{x}^t \mathbf{Q} \mathbf{x}, \quad \mathbf{Q} = \mathbf{Q}^t$$

Now we can apply to $\mathbf{Q}$ the spectral theorem we have developed.

Since $\mathbf{Q}$ is (orthogonally) diagonalizable, there is an orthogonal matrix $\mathbf{P}$ (formed by an orthonormal basis of eigenvectors of $\mathbf{Q}$) such that

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

This means that in $\mathbf{q}(\mathbf{x}) = \mathbf{x}^t\mathbf{Q}\mathbf{x}$, if we change the variables by the rule $\mathbf{x} = \mathbf{P}\mathbf{y}$,

$$\mathbf{q}(\mathbf{x}) = \mathbf{x}^t\mathbf{Q}\mathbf{x} = \mathbf{y}^t\mathbf{P}^{-1}\mathbf{Q}\mathbf{P}\mathbf{y} = \mathbf{y}^t\mathbf{D}\mathbf{y} = \sum_i \lambda_i y_i^2.$$

Some applications

Among the potential applications, we mentioned the identification of conics. For example, $10x_1^2 + 6x_1x_2 + 2x_2^2 = 5$: The matrix

$$\mathbf{Q} = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}$$

has for eigenvalues $11, 1$ with

$$\mathbf{P} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}$$

The change of variables $\mathbf{x} = \mathbf{P}\mathbf{y}$ gives

$$11y_1^2 + y_2^2 = 5,$$

the equation of an ellipse.

Another application, to the theory of max and min appears as follows: If $\mathbf{a}$ is a critical point of the function $\mathbf{f}(\mathbf{x})$–that is all the partial derivatives vanish at $\mathbf{x} = \mathbf{a}$, $\frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{a}) = 0$, Taylor's expansion of $\mathbf{f}$ in a neighborhood of $\mathbf{a}$ gives

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \mathbf{q}(\mathbf{h}) + \mathrm{error}$$

where $\mathbf{q}$ is a quadratic polynomial on the vector $\mathbf{h} = \mathbf{x} - \mathbf{a}$.The corresponding symmetric matrix is

$$\mathbf{Q} = \left[\frac{\partial^2 \mathbf{f}(\mathbf{x})}{\partial x_i \partial x_j}(\mathbf{a})\right]$$

If all the eigenvalues of $\mathbf{Q}$ are positive [negative], $\mathbf{q}(\mathbf{h}) \geq 0$ Then $\mathbf{f}(\mathbf{x}) \geq \mathbf{f}(\mathbf{a})$ in a neighborhood of $\mathbf{a}$: local max [local min] . The other cases are saddle points [the higher dimensional analogues of inflection points]

Rigid Motion

A **rigid motion** on the inner product space **V** is a mapping

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

with the property

$$||\mathbf{T}(u) - \mathbf{T}(v)|| = ||u - v||, \quad \forall u, v\mathbf{V}.$$

That is, **T** preserves distance of the images. A simple example is a translation: If **a** is a fixed vector, the function

$$\mathbf{T}(v) := \mathbf{a} + v$$

is obviously a rigid motion. What else? We have seen that orthogonal transformations **S**, $\mathbf{S}\mathbf{S}^t = \mathbf{I}$, preserve distances. Another such motion is obtained by composition: following a translation with an orthogonal mapping. What else? That is it!

Theorem

Any rigid motion $\mathbf{T}$ *of* $\mathbf{V}$ *decomposes into* $\mathbf{T} = \mathbf{S} \circ \mathbf{U}$*, where* $\mathbf{S}$ *is an orthogonal transformation and* $\mathbf{U}$ *is a translation.*

Proof: Set $\mathbf{a} = \mathbf{T}(O)$. Then the function $\mathbf{F}(u) = \mathbf{T}(u) - \mathbf{a}$ is a rigid motion and $\mathbf{F}(O) = O$. It is enough to prove that $\mathbf{F}$ is orthogonal. Note that

$$||\mathbf{F}(u) - \mathbf{F}(O)|| = ||u - O||,$$

so $\mathbf{F}$ preserves lengths, which is the key property of orthogonal transformations. BUT we are NOT assuming that $\mathbf{F}$ is linear, we must prove it.
We first prove that $\mathbf{F}$ preserves dot products:
$\langle \mathbf{F}(u), \mathbf{F}(v) \rangle = \langle u, v \rangle$: We start from the equality and expand both sides

$$\begin{aligned} ||\mathbf{F}(u) - \mathbf{F}(v)||^2 &= ||u - v||^2 \\ (\mathbf{F}(u) - \mathbf{F}(v)) \cdot (\mathbf{F}(u) - \mathbf{F}(v)) &= (u - v) \cdot (u - v) \\ \underbrace{||\mathbf{F}(u)||^2}_{*} - 2\langle \mathbf{F}(u), \mathbf{F}(v)\rangle + \underbrace{||\mathbf{F}(v)||^2}_{**} &= \underbrace{||u||^2}_{*} - 2\langle u, v\rangle + \underbrace{||v||^2}_{**} \end{aligned}$$

Thus proving

$$\langle \mathbf{F}(u), \mathbf{F}(v)\rangle = \langle u, v\rangle.$$

Now we are going to prove that **F** is a linear function by first showing that it is additive:

$$
\begin{aligned}
||\mathbf{F}(u+v)-\mathbf{F}(u)-\mathbf{F}(v)||^2 &\stackrel{?}{=} 0 \\
||\mathbf{F}(u+v)||^2+||\mathbf{F}(u)||^2+||\mathbf{F}(v)||^2- &= ||u+v||^2+||u||^2+||v||^2- \\
2\langle\mathbf{F}(u+v),\mathbf{F}(u)\rangle-2\langle\mathbf{F}(u+v),\mathbf{F}(v)\rangle &= 2\langle(u+v),u\rangle-2\langle(u+v),v\rangle \\
+2\langle\mathbf{F}(u),\mathbf{F}(v)\rangle &= +2\langle u,v\rangle \\
&= ||(u+v)-u-v||^2=0.
\end{aligned}
$$

Scaling, that $\mathbf{F}(cu)=c\mathbf{F}(u)$ for any $c\in\mathbb{R}$, has a similar proof: Expand

$$||\mathbf{F}(cu)-c\mathbf{F}(u)||^2$$

Homework #10

Section 6.4: 2f, 4, 6, 12, 13, 15

Section 6.5: 6, 10, 11, 17, 27a

Quiz #11

1. Section 6.5, Problem 27d
2. Let **A** be a 3×3 orthogonal matrix. Prove that **A** is similar to a matrix of the form

$$\begin{bmatrix} \mathbf{R} & O \\ O & \pm 1 \end{bmatrix}$$

   where **R** is a 2×2 orthogonal matrix.
3. Section 6.3, Problem 22c
4. Let **A** be a skew-symmetric real matrix. If **A** diagonalizable, prove that $\mathbf{A} = O$.