Chapter 8

Linear Transformations

8.1 The Idea of a Linear Transformation

1 A **linear transformation** T takes vectors v to vectors $T(v)$. Linearity requires

$$\boxed{T(c\,v + d\,w) = c\,T(v) + d\,T(w)}$$ Note $T(\mathbf{0}) = \mathbf{0}$ so $T(v) = v + u_0$ is not linear.

2 The input vectors v and outputs $T(v)$ can be in $\mathbf{R}^n$ or matrix space or function space.

3 If A is m by n, $T(x) = Ax$ is linear from the input space $\mathbf{R}^n$ to the output space $\mathbf{R}^m$.

4 The derivative $T(f) = \dfrac{df}{dx}$ is linear. The integral $T^+(f) = \displaystyle\int_0^x f(t)\,dt$ is its pseudoinverse.

5 The product ST of two linear transformations is still linear : $\boxed{(ST)(v) = S(T(v)).}$

When a matrix A multiplies a vector v, it "transforms" v into another vector Av. *In goes v, out comes* $T(v) = Av$. A transformation T follows the same idea as a function. In goes a number x, out comes $f(x)$. For one vector v or one number x, we multiply by the matrix or we evaluate the function. The deeper goal is to see all vectors v at once. We are transforming the whole space $\mathbf{V}$ when we multiply every v by A.

Start again with a matrix A. It transforms v to Av. It transforms w to Aw. Then we *know* what happens to $u = v + w$. There is no doubt about Au, it has to equal $Av + Aw$. Matrix multiplication $T(v) = Av$ gives a *linear transformation* :

> A *transformation* T assigns an output $T(v)$ to each input vector v in $\mathbf{V}$.
>
> The transformation is *linear* if it meets these requirements for all v and w:
>
> (a) $T(v + w) = T(v) + T(w)$ (b) $T(cv) = cT(v)$ for all c.

If the input is $v = 0$, the output must be $T(v) = 0$. We combine rules (a) and (b) into one:

Linear transformation $T(cv + dw)$ ***must equal*** $cT(v) + dT(w)$.

Again I can test matrix multiplication for linearity: $A(cv + dw) = cAv + dAw$ is *true*.

A linear transformation is highly restricted. Suppose T adds u_0 to every vector. Then $T(v) = v + u_0$ and $T(w) = w + u_0$. This isn't good, or at least *it isn't linear*. Applying T to $v + w$ produces $v + w + u_0$. That is not the same as $T(v) + T(w)$:

Shift is not linear $v + w + u_0$ is not $T(v) + T(w) = (v + u_0) + (w + u_0)$.

The exception is when $u_0 = 0$. The transformation reduces to $T(v) = v$. This is the *identity transformation* (nothing moves, as in multiplication by the identity matrix). That is certainly linear. In this case the input space $\mathbf{V}$ is the same as the output space $\mathbf{W}$.

The linear-plus-shift transformation $T(v) = Av + u_0$ is called *"affine"*. Straight lines stay straight although T is not linear. Computer graphics works with affine transformations in Section 10.6, because we must be able to move images.

Example 1   Choose a fixed vector $a = (1, 3, 4)$, and let $T(v)$ be the dot product $a \cdot v$:

The input is $v = (v_1, v_2, v_3)$.   The output is $T(v) = a \cdot v = v_1 + 3v_2 + 4v_3$.

Dot products are linear. The inputs v come from three-dimensional space, so $\mathbf{V} = \mathbf{R}^3$. The outputs are just numbers, so the output space is $\mathbf{W} = \mathbf{R}^1$. We are multiplying by the row matrix $A = \begin{bmatrix} 1 & 3 & 4 \end{bmatrix}$. Then $T(v) = Av$.

You will get good at recognizing which transformations are linear. If the output involves squares or products or lengths, v_1^2 or v_1v_2 or $\|v\|$, then T is not linear.

Example 2   The length $T(v) = \|v\|$ is not linear. Requirement (a) for linearity would be $\|v + w\| = \|v\| + \|w\|$. Requirement (b) would be $\|cv\| = c\|v\|$. Both are false!

Not (a): The sides of a triangle satisfy an *inequality* $\|v + w\| \leq \|v\| + \|w\|$.
Not (b): The length $\| - v\|$ is $\|v\|$ and not $-\|v\|$. For negative c, linearity fails.

Example 3   (Rotation) T is the transformation that *rotates every vector by* $30°$. The *"domain"* of T is the xy plane (all input vectors v). The *"range"* of T is also the xy plane (all rotated vectors $T(v)$). We described T without a matrix: rotate the plane by $30°$.

Is rotation linear? *Yes it is.* We can rotate two vectors and add the results. The sum of rotations $T(v) + T(w)$ is the same as the rotation $T(v + w)$ of the sum. **The whole plane is turning together, in this linear transformation.**

Lines to Lines, Triangles to Triangles, Basis Tells All

Figure 8.1 shows the line from v to w in the input space. It also shows the line from $T(v)$ to $T(w)$ in the output space. Linearity tells us: Every point on the input line goes onto the output line. And more than that: ***Equally spaced points go to equally spaced points***. The middle point $u = \frac{1}{2}v + \frac{1}{2}w$ goes to the middle point $T(u) = \frac{1}{2}T(v) + \frac{1}{2}T(w)$.

The second figure moves up a dimension. Now we have three corners v_1, v_2, v_3. Those inputs have three outputs $T(v_1)$, $T(v_2)$, $T(v_3)$. *The input triangle goes onto the output triangle.* Equally spaced points stay equally spaced (along the edges, and then between the edges). The middle point $u = \frac{1}{3}(v_1 + v_2 + v_3)$ goes to the middle point $T(u) = \frac{1}{3}(T(v_1) + T(v_2) + T(v_3))$.
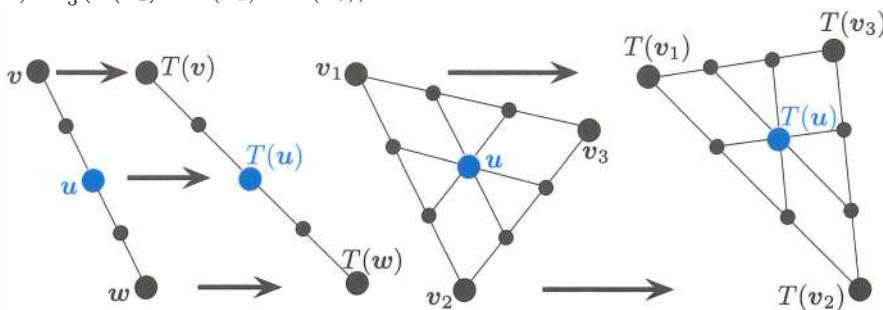

Figure 8.1: Lines to lines, equal spacing to equal spacing, $u = 0$ to $T(u) = 0$.

The rule of linearity extends to combinations of three vectors or n vectors:

Linearity $u = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$ **must transform to**

$$T(u) = c_1 T(v_1) + c_2 T(v_2) + \cdots + c_n T(v_n)$$

(1)

The 2-vector rule starts the 3-vector proof: $T(cu + dv + ew) = T(cu) + T(dv + ew)$. Then linearity applies to both of those parts, to give $cT(u) + dT(v) + eT(w)$.

The n-vector rule (1) leads to the most important fact about linear transformations:

Suppose you know $T(v)$ for all vectors $v_1, \ldots, v_n$ in a basis

Then you know $T(u)$ for every vector u in the space.

You see the reason: Every u in the space is a combination of the basis vectors v_j. Then linearity tells us that $T(u)$ is the same combination of the outputs $T(v_j)$.

Example 4   The transformation T takes the derivative of the input: $T(u) = du/dx$.

How do you find the derivative of $u = 6 - 4x + 3x^2$? You start with the derivatives of 1, x, and x^2. Those are the basis vectors. Their derivatives are 0, 1, and $2x$. Then you use linearity for the derivative of any combination:

$$\frac{du}{dx} = 6 \,(\text{derivative of } 1) - 4 \,(\text{derivative of } x) + 3 \,(\text{derivative of } x^2) = -4 + 6x.$$

All of calculus depends on linearity! Precalculus finds a few key derivatives, for x^n and $\sin x$ and $\cos x$ and e^x. Then linearity applies to all their combinations.

I would say that the only rule special to calculus is the *chain rule*. That produces the derivative of a chain of functions $f(g(x))$.

Nullspace of $T(u) = du/dx$. For the nullspace we solve $T(u) = 0$. The derivative is zero when u *is a constant function*. So the one-dimensional nullspace is a line in function space—all multiples of the special solution $u = 1$.

Column space of $T(u) = du/dx$. In our example the input space contains all quadratics $a + bx + cx^2$. The outputs (the column space) are all linear functions $b + 2cx$. Notice that the **Counting Theorem** is still true: $r + (n - r) = n$.

dimension (**column space**)+dimension (**nullspace**)$=2+1=3=$dimension (**input space**)

What is the matrix for d/dx?   I can't leave derivatives without asking for a matrix. We have a linear transformation $T = d/dx$. We know what T does to the basis functions:

$$v_1, v_2, v_3 = 1, x, x^2 \qquad \frac{dv_1}{dx} = 0 \qquad \frac{dv_2}{dx} = 1 = v_1 \qquad \frac{dv_3}{dx} = 2x = 2v_2. \quad (2)$$

The 3-dimensional input space $\mathbf{V}$ (= quadratics) transforms to the 2-dimensional output space $\mathbf{W}$ (= linear functions). If v_1, v_2, v_3 were vectors, I would know the matrix.

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \text{matrix form of the derivative } T = \frac{d}{dx}. \qquad (3)$$

The linear transformation du/dx is perfectly copied by the matrix multiplication Au.

Input u   **Multiplication $Au = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} b \\ 2c \end{bmatrix}$**   **Output $\dfrac{du}{dx} = b + 2cx$.**
$a + bx + cx^2$

The connection from T to A (we will connect every transformation to a matrix) depended on choosing an input basis $1, x, x^2$ and an output basis $1, x$.

Next we look at integrals. They give the pseudoinverse T^+ of the derivative! I can't write T^{-1} and I can't say "*inverse of T*" when the derivative of 1 is 0.

Example 5 **Integration T^+ is also linear** : $\int_0^x (D + Ex)\, dx = Dx + \frac{1}{2}Ex^2$.

The input basis is now $1, x$. The output basis is $1, x, x^2$. The matrix A^+ for T^+ is 3 by 2:

Input v **Multiplication** $A^+ v = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} D \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ D \\ \frac{1}{2}E \end{bmatrix}$ **Output = Integral of v**
$D + Ex$ $T^+(v) = Dx + \frac{1}{2}Ex^2$

The Fundamental Theorem of Calculus says that integration is the (pseudo)inverse of differentiation. For linear algebra, the matrix A^+ is the (pseudo)inverse of the matrix A:

$$A^+ A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad AA^+ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{4}$$

The derivative of a constant function is zero. That zero is on the diagonal of $A^+ A$. Calculus wouldn't be calculus without that 1-dimensional nullspace of $T = d/dx$.

Examples of Transformations (mostly linear)

Example 6 Project every 3-dimensional vector onto the horizontal plane $z = 1$. The vector $v = (x, y, z)$ is transformed to $T(v) = (x, y, 1)$. This transformation is not linear. Why not? It doesn't even transform $v = \mathbf{0}$ into $T(v) = \mathbf{0}$.

Example 7 Suppose A is an *invertible matrix*. Certainly $T(v + w) = Av + Aw = T(v) + T(w)$. Another linear transformation is multiplication by A^{-1}. This produces the *inverse transformation* T^{-1}, which brings every vector $T(v)$ back to v:

$$T^{-1}(T(v)) = v \quad \text{matches the matrix multiplication} \quad A^{-1}(Av) = v.$$

If $T(v) = Av$ and $S(u) = Bu$, then the product $T(S(u))$ matches the product ABu.

We are reaching an unavoidable question. *Are all linear transformations from $\mathbf{V} = \mathbf{R}^n$ to $\mathbf{W} = \mathbf{R}^m$ produced by matrices?* When a linear T is described as a "rotation" or "projection" or ". . .", is there always a matrix A hiding behind T? Is $T(v)$ always Av?

The answer is *yes*! This is an approach to linear algebra that doesn't start with matrices. We still end up with matrices—*after we choose an input basis and output basis*.

Note Transformations have a language of their own. For a matrix, the column space contains all outputs Av. The nullspace contains all inputs for which $Av = \mathbf{0}$. Translate those words into "*range*" and "*kernel*":

Range of T = set of *all outputs $T(v)$*. Range corresponds to column space.

Kernel of T = set of *all inputs for which $T(v) = \mathbf{0}$*. Kernel corresponds to nullspace.

The range is in the output space $\mathbf{W}$. The kernel is in the input space $\mathbf{V}$. When T is multiplication by a matrix, $T(v) = Av$, range is column space and kernel is nullspace.

Linear Transformations of the Plane

It is more interesting to *see* a transformation than to define it. When a 2 by 2 matrix A multiplies all vectors in $\mathbf{R}^2$, we can watch how it acts. Start with a "house" that has eleven endpoints. Those eleven vectors v are transformed into eleven vectors Av. Straight lines between v's become straight lines between the transformed vectors Av. (The transformation from house to house is linear!) Applying A to a standard house produces a new house—possibly stretched or rotated or otherwise unlivable.

This part of the book is visual, not theoretical. We will show four houses and the matrices that produce them. The columns of H are the eleven corners of the first house. (H is 2 by 12, so **plot2d** in Problem 25 will connect the 11th corner to the first.) A multiplies the 11 points in the house matrix H to produce the corners AH of the other houses.

$$\textbf{House matrix} \quad H = \begin{bmatrix} -6 & -6 & -7 & 0 & 7 & 6 & 6 & -3 & -3 & 0 & 0 & -6 \\ -7 & 2 & 1 & 8 & 1 & 2 & -7 & -7 & -2 & -2 & -7 & -7 \end{bmatrix}.$$
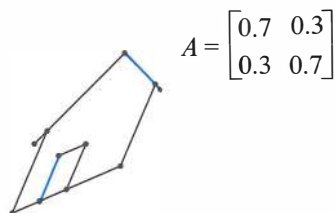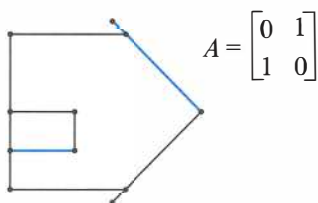

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} \cos 35° & -\sin 35° \\ \sin 35° & \cos 35° \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$
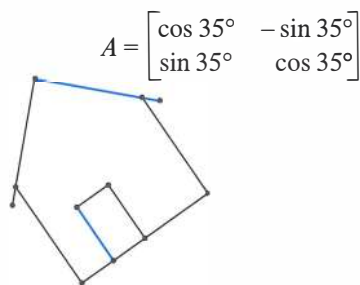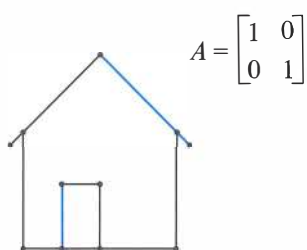
Figure 8.2: Linear transformations of a house drawn by **plot2d**$(A * H)$.

■ REVIEW OF THE KEY IDEAS ■

1. A transformation T takes each v in the input space to $T(v)$ in the output space.

2. T is **linear** if $T(v + w) = T(v) + T(w)$ and $T(cv) = cT(v)$: lines to lines.

3. Combinations to combinations: $T(c_1 v_1 + \cdots + c_n v_n) = c_1 T(v_1) + \cdots + c_n T(v_n)$.

4. $T = $ *derivative* and $T^+ = $ *integral* are linear. So is $T(v) = Av$ from $\mathbf{R}^n$ to $\mathbf{R}^m$.

■ WORKED EXAMPLES ■

8.1 A The elimination matrix $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ gives a *shearing transformation* from (x, y) to $T(x, y) = (x, x + y)$. If the inputs fill a square, draw the transformed square.

Solution The points $(1, 0)$ and $(2, 0)$ on the x axis transform by T to $(1, 1)$ and $(2, 2)$ on the $45°$ line. Points on the y axis are *not moved*: $T(0, y) = (0, y) =$ eigenvectors with $\lambda = 1$.

Vertical lines slide up
This is the shearing
Squares go to parallelograms

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$
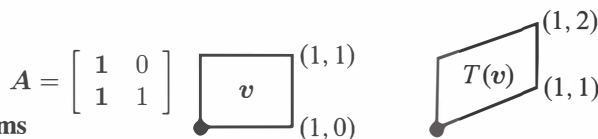
8.1 B A **nonlinear transformation** T is invertible if every b in the output space comes from exactly one x in the input space: $T(x) = b$ always has exactly one solution. Which of these transformations (on real numbers x) is invertible and what is T^{-1}? *None are linear, not even T_3.* When you solve $T(x) = b$, you are inverting T:

$$T_1(x) = x^2 \quad T_2(x) = x^3 \quad T_3(x) = x + 9 \quad T_4(x) = e^x \quad T_5(x) = \frac{1}{x} \text{ for nonzero } x\text{'s}$$

Solution T_1 is not invertible: $x^2 = 1$ has *two* solutions and $x^2 = -1$ has *no* solution. T_4 is not invertible because $e^x = -1$ has no solution. (If the output space changes to *positive* b's then the inverse of $e^x = b$ is $x = \ln b$.)

Notice $T_5^2 =$ identity. But $T_3^2(x) = x + 18$. What are $T_2^2(x)$ and $T_4^2(x)$?

T_2, T_3, T_5 are invertible: $x^3 = b$ and $x + 9 = b$ and $\frac{1}{x} = b$ have one solution x.

$$x = T_2^{-1}(b) = b^{1/3} \qquad x = T_3^{-1}(b) = b - 9 \qquad x = T_5^{-1}(b) = 1/b$$

Problem Set 8.1

1 A linear transformation must leave the zero vector fixed: $T(\mathbf{0}) = \mathbf{0}$. Prove this from $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$ by choosing $\mathbf{w} = $ _____ (and finish the proof). Prove it also from $T(c\mathbf{v}) = cT(\mathbf{v})$ by choosing $c = $ _____ .

2 Requirement (b) gives $T(c\mathbf{v}) = cT(\mathbf{v})$ and also $T(d\mathbf{w}) = dT(\mathbf{w})$. Then by addition, requirement (a) gives $T(\quad) = (\quad)$. What is $T(c\mathbf{v} + d\mathbf{w} + e\mathbf{u})$?

3 Which of these transformations are not linear? The input is $\mathbf{v} = (v_1, v_2)$:

(a) $T(\mathbf{v}) = (v_2, v_1)$ (b) $T(\mathbf{v}) = (v_1, v_1)$ (c) $T(\mathbf{v}) = (0, v_1)$

(d) $T(\mathbf{v}) = (0, 1)$ (e) $T(\mathbf{v}) = v_1 - v_2$ (f) $T(\mathbf{v}) = v_1 v_2$.

4 If S and T are linear transformations, is $T(S(v))$ linear or quadratic?

  (a) (Special case) If $S(v) = v$ and $T(v) = v$, then $T(S(v)) = v$ or v^2?

  (b) (General case) $S(v_1 + v_2) = S(v_1) + S(v_2)$ and $T(v_1 + v_2) = T(v_1) + T(v_2)$ combine into
$$T(S(v_1 + v_2)) = T(\underline{\quad}) = \underline{\quad} + \underline{\quad}.$$

5 Suppose $T(v) = v$ except that $T(0, v_2) = (0, 0)$. Show that this transformation satisfies $T(cv) = cT(v)$ but does not satisfy $T(v + w) = T(v) + T(w)$.

6 Which of these transformations satisfy $T(v + w) = T(v) + T(w)$ and which satisfy $T(cv) = cT(v)$?

  (a) $T(v) = v/\|v\|$ (b) $T(v) = v_1 + v_2 + v_3$ (c) $T(v) = (v_1, 2v_2, 3v_3)$

  (d) $T(v) = $ largest component of v.

7 For these transformations of $V = \mathbf{R}^2$ to $W = \mathbf{R}^2$, find $T(T(v))$. Show that when $T(v)$ is linear, then also $T(T(v))$ is linear.

  (a) $T(v) = -v$ (b) $T(v) = v + (1,1)$

  (c) $T(v) = 90°$ rotation $= (-v_2, v_1)$

  (d) $T(v) = $ projection $= \frac{1}{2}(v_1 + v_2,\ v_1 + v_2)$.

8 Find the range and kernel (like the column space and nullspace) of T:

  (a) $T(v_1, v_2) = (v_1 - v_2, 0)$ (b) $T(v_1, v_2, v_3) = (v_1, v_2)$

  (c) $T(v_1, v_2) = (0, 0)$ (d) $T(v_1, v_2) = (v_1, v_1)$.

9 The "cyclic" transformation T is defined by $T(v_1, v_2, v_3) = (v_2, v_3, v_1)$. What is $T(T(v))$? What is $T^3(v)$? What is $T^{100}(v)$? Apply T a hundred times to v.

10 A linear transformation from V to W has an *inverse* from W to V when the range is all of W and the kernel contains only $v = 0$. Then $T(v) = w$ has one solution v for each w in W. Why are these T's not invertible?

  (a) $T(v_1, v_2) = (v_2, v_2)$ $W = \mathbf{R}^2$

  (b) $T(v_1, v_2) = (v_1, v_2, v_1 + v_2)$ $W = \mathbf{R}^3$

  (c) $T(v_1, v_2) = v_1$ $W = \mathbf{R}^1$

11 If $T(v) = Av$ and A is m by n, then T is "multiplication by A."

  (a) What are the input and output spaces V and W?

  (b) Why is range of T = column space of A?

  (c) Why is kernel of T = null space of A?

12 Suppose a linear T transforms $(1, 1)$ to $(2, 2)$ and $(2, 0)$ to $(0, 0)$. Find $T(v)$:

(a) $v = (2, 2)$ (b) $v = (3, 1)$ (c) $v = (-1, 1)$ (d) $v = (a, b)$.

Problems 13-19 may be harder. The input space V contains all 2 by 2 matrices M.

13 M is any 2 by 2 matrix and $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. The transformation T is defined by $T(M) = AM$. What rules of matrix multiplication show that T is linear?

14 Suppose $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$. Show that the range of T is the whole matrix space **V** and the kernel is the zero matrix:

     (1) If $AM = 0$ prove that M must be the zero matrix.

     (2) Find a solution to $AM = B$ for any 2 by 2 matrix B.

15 Suppose $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$. Show that the identity matrix I is not in the range of T. Find a nonzero matrix M such that $T(M) = AM$ is zero.

16 Suppose T transposes every 2 by 2 matrix M. Try to find a matrix A which gives $AM = M^{\mathrm{T}}$. *Show that no matrix A will do it. To professors:* Is this a linear transformation that doesn't come from a matrix? The matrix should be 4 by 4!

17 The transformation T that transposes every 2 by 2 matrix is definitely linear. Which of these extra properties are true?

     (a) $T^2 =$ identity transformation.

     (b) The kernel of T is the zero matrix.

     (c) Every 2 by 2 matrix is in the range of T.

     (d) $T(M) = -M$ is impossible.

18 Suppose $T(M) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} M \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Find a matrix with $T(M) \neq 0$. Describe all matrices with $T(M) = 0$ (the kernel) and all output matrices $T(M)$ (the range).

19 If A and B are invertible and $T(M) = AMB$, find $T^{-1}(M)$ in the form $(\quad)M(\quad)$.

Questions 20–26 are about house transformations. The output is $T(H) = AH$.

20 How can you tell from the picture of T (house) that A is

     (a) a diagonal matrix?

     (b) a rank-one matrix?

     (c) a lower triangular matrix?

21 Draw a picture of T (house) for these matrices:

$$D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} .7 & .7 \\ .3 & .3 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

22   What are the conditions on $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ to ensure that T (house) will

    (a) sit straight up?

    (b) expand the house by 3 in all directions?

    (c) rotate the house with no change in its shape?

23   Describe T (house) when $T(v) = -v + (1, 0)$. This T is "affine".

24   Change the house matrix H to add a chimney.

25   The standard house is drawn by **plot2d**(H). Circles from o and lines from −:

$$x = H(1, :)'; y = H(2, :)';$$
$$\text{axis}([-10\,10-10\,10]), \text{axis('square')}$$
$$\text{plot}(x, y,' o', x, y,' -');$$

Test **plot2d**(A′∗ H) and **plot2d**(A′∗ A ∗ H) with the matrices in Figure 8.1.

26   Without a computer sketch the houses $A * H$ for these matrices A:

$$\begin{bmatrix} 1 & 0 \\ 0 & .1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} .5 & .5 \\ -.5 & .5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

27   This code creates a vector theta of 50 angles. It draws the unit circle and then it draws T (circle) $=$ ellipse. $T(v) = Av$ **takes circles to ellipses**.

```
A = [2 1;1 2]     % You can change A
theta = [0:2 ∗ pi/50:2 ∗ pi];
circle = [cos(theta); sin(theta)];
ellipse = A ∗ circle;
axis([−4 4 −4 4]); axis('square')
plot(circle(1,:), circle(2,:), ellipse(1,:), ellipse(2,:))
```

28   Add two eyes and a smile to the circle in Problem 27. (If one eye is dark and the other is light, you can tell when the face is reflected across the y axis.) Multiply by matrices A to get new faces.

29   What conditions on det $A = ad - bc$ ensure that the output house AH will

    (a) be squashed onto a line?

    (b) keep its endpoints in clockwise order (not reflected)?

    (c) have the same area as the original house?

30   Why does every linear transformation T from $\mathbf{R}^2$ to $\mathbf{R}^2$ take squares to parallelograms? Rectangles also go to parallelograms (squashed if T is not invertible).

8.2 The Matrix of a Linear Transformation

1 We know all $T(v)$ if we know $T(v_1), \ldots, T(v_n)$ for an input basis $v_1, \ldots, v_n$: use **linearity**.

2 Column j in the "matrix for T" comes from applying T to the input basis vector v_j.

3 Write $T(v_j) = a_{1j}w_1 + \cdots + a_{mj}w_m$ in the output basis of w's. Those a_{ij} go into column j.

4 The matrix for $T(x) = Ax$ is A, *if* the input and output bases = columns of $I_{n \times n}$ and $I_{m \times m}$.

5 When the bases change to v's and w's, the matrix for the same T changes from A to $W^{-1}AV$.

6 Best bases: $V = W$ = eigenvectors and V, W = singular vectors give diagonal Λ and Σ.

The next pages assign a matrix A to every linear transformation T. For ordinary column vectors, the input v is in $\mathbf{V} = \mathbf{R}^n$ and the output $T(v)$ is in $\mathbf{W} = \mathbf{R}^m$. The matrix A for this transformation will be m by n. Our choice of bases in $\mathbf{V}$ and $\mathbf{W}$ will decide A.

The standard basis vectors for $\mathbf{R}^n$ and $\mathbf{R}^m$ are the columns of I. That choice leads to a standard matrix. Then $T(v) = Av$ in the normal way. But these spaces also have other bases, so *the same transformation T is represented by other matrices.* A main theme of linear algebra is to choose the bases that give the best matrix (a diagonal matrix) for T.

All vector spaces $\mathbf{V}$ and $\mathbf{W}$ have bases. Each choice of those bases leads to a matrix for T. When the input basis is different from the output basis, the matrix for $T(v) = v$ will not be the identity I. It will be the "change of basis matrix". Here is the key idea:

> **Suppose we know $T(v)$ for the input basis vectors v_1 to v_n.**
> **Columns 1 to n of the matrix will contain those outputs $T(v_1)$ to $T(v_n)$.**
> A **times** c = **matrix times vector** = **combination of those n columns**.
> Ac **is the correct combination** $c_1 T(v_1) + \cdots + c_n T(v_n) = T(v)$.

Reason Every v is a unique combination $c_1 v_1 + \cdots + c_n v_n$ of the basis vectors v_j. Since T is a linear transformation (here is the moment for linearity), $T(v)$ must be **the same combination** $c_1 T(v_1) + \cdots + c_n T(v_n)$ **of the outputs** $T(v_j)$ **in the columns**.

Our first example gives the matrix A for the standard basis vectors in $\mathbf{R}^2$ and $\mathbf{R}^3$.

Example 1 Suppose T transforms $v_1 = (1, 0)$ to $T(v_1) = (2, 3, 4)$. Suppose the second basis vector $v_2 = (0, 1)$ goes to $T(v_2) = (5, 5, 5)$. If T is linear from $\mathbf{R}^2$ to $\mathbf{R}^3$ then its "standard matrix" is 3 by 2. Those outputs $T(v_1)$ and $T(v_2)$ go into the columns of A:

$$A = \begin{bmatrix} 2 & 5 \\ 3 & 5 \\ 4 & 5 \end{bmatrix} \qquad c_1 = 1 \text{ and } c_2 = 1 \text{ give } T(v_1 + v_2) = \begin{bmatrix} 2 & 5 \\ 3 & 5 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}.$$

Change of Basis

Example 2   Suppose the input space $V = R^2$ is also the output space $W = R^2$. *Suppose that $T(v) = v$ is the identity transformation.* You might expect its matrix to be I, but that only happens when the input basis is the same as the output basis. I will choose different bases to see how the matrix is constructed.

For this special case $T(v) = v$, I will call the matrix B instead of A. We are just changing basis from the v's to the w's. Each v is a combination of w_1 and w_2.

$$\textbf{Input}_{\textbf{basis}} \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 3 & 8 \end{bmatrix} \quad \textbf{Output}_{\textbf{basis}} \begin{bmatrix} w_1 & w_2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix} \quad \begin{array}{l} \textbf{Change } v_1 = 1w_1 + 1w_2 \\ \textbf{of basis } v_2 = 2w_1 + 3w_2 \end{array}$$

Please notice!   I wrote the input basis v_1, v_2 in terms of the output basis w_1, w_2. That is because of our key rule. We apply the identity transformation T to each input basis vector: $T(v_1) = v_1$ and $T(v_2) = v_2$. **Then we write those outputs v_1 and v_2 in the output basis w_1 and w_2.** Those bold numbers $1, 1$ and $2, 3$ tell us column 1 and column 2 of the matrix B (the change of basis matrix): $WB = V$ so $B = W^{-1}V$.

$$\textbf{Matrix } B \textbf{ for}_{\textbf{change of basis}} \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} B \end{bmatrix} = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \text{ is } \begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 3 & 8 \end{bmatrix}. \quad (1)$$

> When the input basis is in the columns of a matrix V, and the output basis is in the columns of W, the change of basis matrix for $T = I$ is $B = W^{-1}V$.

The key   I see a clear way to understand that rule $B = W^{-1}V$. Suppose the same vector u is written in the input basis of v's and the output basis of w's. I will do that three ways:

$$\begin{array}{l} u = c_1 v_1 + \cdots + c_n v_n \\ u = d_1 w_1 + \cdots + d_n w_n \end{array} \text{ is } \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} \text{ and } Vc = Wd.$$

The coefficients d in the new basis of w's are $d = W^{-1}Vc$. Then B is $W^{-1}V$.   (2)

This formula $B = W^{-1}V$ produces one of the world's greatest mysteries: When the *standard basis* $V = I$ is changed to a different basis W, **the change of basis matrix is not W but $B = W^{-1}$.** Larger basis vectors have smaller coefficients!

$$\begin{bmatrix} x \\ y \end{bmatrix} \text{ in the standard basis has coefficients } \begin{bmatrix} w_1 & w_2 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} \text{ in the } w_1, w_2 \text{ basis.}$$

Construction of the Matrix

Now we construct a matrix for any linear transformation. Suppose T transforms the space **V** (n-dimensional) to the space **W** (m-dimensional). We choose a basis $v_1, \ldots, v_n$ for **V** and we choose a basis $w_1, \ldots, w_m$ for **W**. The matrix A will be m by n. To find the first column of A, apply T to the first basis vector v_1. The output $T(v_1)$ is in **W**.

> $T(v_1)$ **is a combination** $\quad a_{11}w_1 + \cdots + a_{m1}w_m \quad$ **of the output basis for W**.

These numbers $a_{11}, \ldots, a_{m1}$ go into the first column of A. Transforming v_1 to $T(v_1)$ matches multiplying $(1, 0, \ldots, 0)$ by A. It yields that first column of the matrix. When T is the derivative and the first basis vector is 1, its derivative is $T(v_1) = \mathbf{0}$. So for the derivative matrix below, the first column of A is all zero.

Example 3 The input basis of v's is $1, x, x^2, x^3$. The output basis of w's is $1, x, x^2$.

Then T **takes the derivative**: $\quad T(v) = \dfrac{dv}{dx} \quad$ and $\quad A \quad = \quad$ "derivative matrix".

If $v = c_1 + c_2 x + c_3 x^2 + c_4 x^3$

then $\dfrac{dv}{dx} = 1c_2 + 2c_3 x + 3c_4 x^2$

$$Ac = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} c_2 \\ 2c_3 \\ 3c_4 \end{bmatrix}$$

> **Key rule:** The jth column of A is found by applying T to the jth basis vector v_j
>
> $$T(v_j) = \text{combination of output basis vectors} = a_{1j}w_1 + \cdots + a_{mj}w_m. \qquad (3)$$

These numbers a_{ij} go into A. *The matrix is constructed to get the basis vectors right.* **Then linearity gets all other vectors right.** Every v is a combination $c_1 v_1 + \cdots + c_n v_n$, and $T(v)$ is a combination of the w's. When A multiplies the vector $c = (c_1, \ldots, c_n)$ in the v combination, Ac produces the coefficients in the $T(v)$ combination. This is because matrix multiplication (combining columns) is linear like T.

The matrix A tells us what T does. Every linear transformation from **V** to **W** can be converted to a matrix. This matrix depends on the bases.

Example 4 For the integral $T^+(v)$, the first basis function is again 1. Its integral is the second basis function x. So the first column of the "integral matrix" A^+ is $(0, 1, 0, 0)$.

The integral of $d_1 + d_2 x + d_3 x^2$

is $\quad d_1 x + \dfrac{1}{2} d_2 x^2 + \dfrac{1}{3} d_3 x^3$

$$A^+ d = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 0 \\ d_1 \\ \frac{1}{2} d_2 \\ \frac{1}{3} d_3 \end{bmatrix}$$

If you integrate a function and then differentiate, you get back to the start. So $AA^+ = I$. But if you differentiate before integrating, the constant term is lost. So A^+A is not I. **The integral of the derivative of** 1 **is zero:**

$$T^+T(1) = \text{integral of zero function } = 0.$$

This matches A^+A, whose first column is all zero. The derivative T has a kernel (the constant functions). Its matrix A has a nullspace. Main idea again: Av copies $T(v)$.

The examples of the derivative and integral made three points. First, linear transformations T are everywhere—in calculus and differential equations and linear algebra. Second, spaces other than $\mathbf{R}^n$ are important—we had functions in $\mathbf{V}$ and $\mathbf{W}$. Third, **if we differentiate and then integrate, we can multiply their matrices** A^+A.

Matrix Products AB Match Transformations TS

We have come to something important—the real reason for the rule to multiply matrices. *At last we discover why*! Two linear transformations T and S are represented by two matrices A and B. Now compare TS with the multiplication AB:

When we apply the transformation T to the output from S, we get TS by this rule: $(TS)(u)$ *is defined to be* $T(S(u))$. The output $S(u)$ becomes the input to T.

When we apply the matrix A to the output from B, we multiply AB by this rule: $(AB)(x)$ is defined to be $A(Bx)$. The output Bx becomes the input to A. *Matrix multiplication gives the correct matrix AB to represent TS.*

The transformation S is from a space $\mathbf{U}$ to $\mathbf{V}$. Its matrix B uses a basis $u_1, \ldots, u_p$ for $\mathbf{U}$ and a basis $v_1, \ldots, v_n$ for $\mathbf{V}$. That matrix is n by p. The transformation T is from $\mathbf{V}$ to $\mathbf{W}$ as before. *Its matrix A must use the same basis $v_1, \ldots, v_n$ for $\mathbf{V}$*—this is the output space for S and the input space for T. **Then the matrix AB matches TS.**

Multiplication The linear transformation TS starts with any vector u in $\mathbf{U}$, goes to $S(u)$ in $\mathbf{V}$ and then to $T(S(u))$ in $\mathbf{W}$. The matrix AB starts with any x in $\mathbf{R}^p$, goes to Bx in $\mathbf{R}^n$ and then to ABx in $\mathbf{R}^m$. **The matrix AB correctly represents TS:**

$$TS: \quad \mathbf{U} \to \mathbf{V} \to \mathbf{W} \qquad AB: \quad (m \text{ by } n)(n \text{ by } p) = (m \text{ by } p).$$

The input is $u = x_1 u_1 + \cdots + x_p u_p$. The output $T(S(u))$ matches the output ABx. *Product of transformations TS matches product of matrices AB.*

The most important cases are when the spaces $\mathbf{U}, \mathbf{V}, \mathbf{W}$ are the same and their bases are the same. With $m = n = p$ we have square matrices that we can multiply.

Example 5 S rotates the plane by θ and T also rotates by θ. Then TS rotates by 2θ. This transformation T^2 corresponds to the rotation matrix A^2 through 2θ :

$$T = S \qquad A = B \qquad T^2 = \text{rotation by } 2\theta \qquad A^2 = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix}. \qquad (4)$$

By matching (transformation)^2 with (matrix)^2, we pick up the formulas for $\cos 2\theta$ and $\sin 2\theta$. Multiply A times A:

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = \begin{bmatrix} \cos^2 \theta - \sin^2 \theta & -2 \sin \theta \cos \theta \\ 2 \sin \theta \cos \theta & \cos^2 \theta - \sin^2 \theta \end{bmatrix}. \quad (5)$$

Comparing (4) with (5) produces $\cos 2\theta = \cos^2 \theta - \sin^2 \theta$ and $\sin 2\theta = 2 \sin \theta \cos \theta$. Trigonometry (the double angle rule) comes from linear algebra.

Example 6 S rotates by the angle θ and T rotates by $-\theta$. Then $TS = I$ leads to $AB = I$. In this case $T(S(\boldsymbol{u}))$ is $\boldsymbol{u}$. We rotate forward and back. For the matrices to match, $AB\boldsymbol{x}$ must be $\boldsymbol{x}$. *The two matrices are inverses.* Check this by putting $\cos(-\theta) = \cos \theta$ and $\sin(-\theta) = -\sin \theta$ into the backward rotation matrix A:

$$AB = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = \begin{bmatrix} \cos^2 \theta + \sin^2 \theta & 0 \\ 0 & \cos^2 \theta + \sin^2 \theta \end{bmatrix} = I.$$

Choosing the Best Bases

Now comes the final step in this section of the book. **Choose bases that diagonalize the matrix.** With the standard basis (the columns of I) our transformation T produces some matrix A—probably not diagonal. That same T is represented by different matrices when we choose different bases. The two great choices are eigenvectors and singular vectors:

> **Eigenvectors** If T transforms $\mathbf{R}^n$ to $\mathbf{R}^n$, its matrix A is square. But using the standard basis, that matrix A is probably not diagonal. If there are n independent eigenvectors, *choose those as the input and output basis.* In this good basis, **the matrix for T is the diagonal eigenvalue matrix Λ.**

Example 7 **The projection matrix** T projects every $\boldsymbol{v} = (x, y)$ in $\mathbf{R}^2$ onto the line $y = -x$. Using the standard basis, $\boldsymbol{v}_1 = (1, 0)$ projects to $T(\boldsymbol{v}_1) = \left(\frac{1}{2}, -\frac{1}{2}\right)$. For $\boldsymbol{v}_2 = (0, 1)$ the projection is $T(\boldsymbol{v}_2) = \left(-\frac{1}{2}, \frac{1}{2}\right)$. Those are the columns of A:

$$\begin{array}{l} \textbf{Projection matrix} \\ \textbf{Standard bases} \\ \textbf{Not diagonal} \end{array} \qquad A = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \text{ has } A^{\mathrm{T}} = A \text{ and } A^2 = A.$$

Now comes the main point of eigenvectors. Make them the basis vectors ! Diagonalize !

When the basis vectors are eigenvectors, the matrix becomes diagonal.

$$v_1 = w_1 = (1, -1) \text{ projects to itself} : T(v_1) = v_1 \text{ and } \lambda_1 = 1$$
$$v_2 = w_2 = (1, 1) \text{ projects to zero} : T(v_2) = 0 \text{ and } \lambda_2 = 0$$

Eigenvector bases
Diagonal matrix
The new matrix is $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \Lambda.$ (6)

Eigenvectors are the perfect basis vectors. They produce the eigenvalue matrix Λ.

What about other choices of *input basis = output basis*? Put those basis vectors into the columns of B. We saw above that the change of basis matrices (between standard basis and new basis) are $B_{\text{in}} = B$ and $B_{\text{out}} = B^{-1}$. The new matrix for T is **similar** to A:

$$A_{\text{new}} = B^{-1}AB \text{ in the new basis of } b\text{'s is similar to } A \text{ in the standard basis}:$$

$$A_{b\text{'s to } b\text{'s}} = B^{-1}_{\text{standard to } b\text{'s}} \quad A_{\text{standard}} \quad B_{b\text{'s to standard}} \quad (7)$$

I used the multiplication rule for the transformation ITI. *The matrices for I, T, I were B^{-1}, A, B.* The matrix B contains the input vectors b in the standard basis.

Finally we allow *different spaces V and W, and different bases v's and w's.* When we know T and we choose bases, we get a matrix A. Probably A is not symmetric or even square. But we can always choose v's and w's that produce a diagonal matrix. This will be the *singular value matrix* $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$ in the decomposition $A = U\Sigma V^{\mathrm{T}}$.

> **Singular vectors** The SVD says that $U^{-1}AV = \Sigma$. The right singular vectors $v_1, \ldots, v_n$ will be the input basis. The left singular vectors $u_1, \ldots, u_m$ will be the output basis. By the rule for matrix multiplication, the matrix for the same transformation in these new bases is $B^{-1}_{\text{out}} AB_{\text{in}} = U^{-1}AV = \Sigma$.

I can't say that Σ is "similar" to A. We are working now with two bases, input and output. But those are *orthonormal bases* and they preserve the lengths of vectors. Following a good suggestion by David Vogan, I propose that we say: Σ **is "isometric" to A.**

Definition $C = Q_1^{-1}AQ_2$ is isometric to A if Q_1 and Q_2 are orthogonal.

Example 8 To construct the matrix A for the transformation $T = \frac{d}{dx}$, we chose the input basis $1, x, x^2, x^3$ and the output basis $1, x, x^2$. The matrix A was simple but unfortunately it wasn't diagonal. But we can take each basis *in the opposite order.*

Now the input basis is $x^3, x^2, x, 1$ and the output basis is $x^2, x, 1$. The change of basis matrices B_{in} and B_{out} are permutations. The matrix for $T(u) = du/dx$ with the new bases is **the diagonal singular value matrix** $B^{-1}_{\text{out}} AB_{\text{in}} = \Sigma$ with σ's $= 3, 2, 1$:

$$B^{-1}_{\text{out}} AB_{\text{in}} = \begin{bmatrix} & & 1 \\ & 1 & \\ 1 & & \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (8)$$

Well, this was a tough section. We found that $x^3, x^2, x, 1$ have derivatives $3x^2, 2x, 1, 0$.

■ REVIEW OF THE KEY IDEAS ■

1. If we know $T(v_1), \ldots, T(v_n)$ for a basis, linearity will determine all other $T(v)$.

2. $\left\{ \begin{array}{l} \text{Linear transformation } T \\ \text{Input basis } v_1, \ldots, v_n \\ \text{Output basis } w_1, \ldots, w_m \end{array} \right\} \rightarrow \begin{array}{l} \text{Matrix } A \ (m \text{ by } n) \\ \quad \text{represents } T \\ \quad \text{in these bases} \end{array}$

3. The change of basis matrix $B = W^{-1}V = B_{\text{out}}^{-1}B_{\text{in}}$ represents the identity $T(v) = v$.

4. If A and B represent T and S, and the output basis for S is the input basis for T, then the matrix AB represents the transformation $T(S(u))$.

5. The best input-output bases are eigenvectors and/or singular vectors of A. Then
$$B^{-1}AB = \Lambda = \text{eigenvalues} \qquad B_{\text{out}}^{-1}AB_{\text{in}} = \Sigma = \text{singular values}.$$

■ WORKED EXAMPLES ■

8.2 A The space of 2 by 2 matrices has these four "vectors" as a basis:

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

T is the linear transformation that *transposes* every 2 by 2 matrix. What is the matrix A that represents T in this basis (output basis = input basis)? What is the inverse matrix A^{-1}? What is the transformation T^{-1} that inverts the transpose operation ?

Solution Transposing those four "basis matrices" just reverses v_2 and v_3:

$$\begin{array}{l} T(v_1) = v_1 \\ T(v_2) = v_3 \\ T(v_3) = v_2 \\ T(v_4) = v_4 \end{array} \quad \text{gives the four columns of} \quad A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The inverse matrix A^{-1} is the same as A. The inverse transformation T^{-1} is the same as T. If we transpose and transpose again, the final matrix equals the original matrix.

Notice that the space of 2 by 2 matrices is 4-dimensional. So the matrix A (for the transpose T) is 4 by 4. The nullspace of A is $\mathbf{Z}$ and the kernel of T is the zero matrix—the only matrix that transposes to zero. The eigenvalues of A are $1, 1, 1, -1$.

Which line of matrices has $T(A) = A^{\text{T}} = -A$ with that eigenvalue $\lambda = -1$?

Problem Set 8.2

Questions 1–4 extend the first derivative example to higher derivatives.

1   The transformation S takes the **second derivative**. Keep $1, x, x^2, x^3$ as the input basis v_1, v_2, v_3, v_4 and also as output basis w_1, w_2, w_3, w_4. Write $S(v_1), S(v_2)$, $S(v_3), S(v_4)$ in terms of the w's. Find the 4 by 4 matrix A_2 for S.

2   What functions have $S(v) = 0$? They are in the kernel of the second derivative S. What vectors are in the nullspace of its matrix A_2 in Problem 1?

3   The second derivative A_2 is not the square of a rectangular first derivative matrix A_1:

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \text{ does not allow } A_1^2 = A_2.$$

Add a zero row 4 to A_1 so that output space = input space. Compare A_1^2 with A_2. Conclusion: We want output basis = _____ basis. Then $m = n$.

4   (a) The product TS of first and second derivatives produces the *third* derivative. Add zeros to make 4 by 4 matrices, then compute $A_1 A_2 = A_3$.

   (b) The matrix A_2^2 corresponds to $S^2 = $ *fourth* derivative. Why is this zero?

Questions 5–9 are about a particular transformation T and its matrix A.

5   With bases v_1, v_2, v_3 and w_1, w_2, w_3, suppose $T(v_1) = w_2$ and $T(v_2) = T(v_3) = w_1 + w_3$. T is a linear transformation. Find the matrix A and multiply by the vector $(1, 1, 1)$. What is the output from T when the input is $v_1 + v_2 + v_3$?

6   Since $T(v_2) = T(v_3)$, the solutions to $T(v) = 0$ are $v = $ _____. What vectors are in the nullspace of A? Find all solutions to $T(v) = w_2$.

7   Find a vector that is not in the column space of A. Find a combination of w's that is not in the range of the transformation T.

8   You don't have enough information to determine T^2. Why is its matrix not necessarily A^2? What more information do you need?

9   Find the *rank* of A. The rank is not the dimension of the whole output space $\mathbf{W}$. It is the dimension of the _____ of T.

Questions 10–13 are about invertible linear transformations.

10   Suppose $T(v_1) = w_1 + w_2 + w_3$ and $T(v_2) = w_2 + w_3$ and $T(v_3) = w_3$. Find the matrix A for T using these basis vectors. What input vector v gives $T(v) = w_1$?

11   Invert the matrix A in Problem 10. Also invert the transformation T—what are $T^{-1}(w_1)$ and $T^{-1}(w_2)$ and $T^{-1}(w_3)$?

12   Which of these are true and why is the other one ridiculous?

   (a) $T^{-1}T = I$ (b) $T^{-1}(T(v_1)) = v_1$ (c) $T^{-1}(T(w_1)) = w_1$.

13   Suppose the spaces V and W have the same basis v_1, v_2.

    (a) Describe a transformation T (not I) that is its own inverse.

    (b) Describe a transformation T (not I) that equals T^2.

    (c) Why can't the same T be used for both (a) and (b)?

Questions 14–19 are about changing the basis.

14   (a) What matrix B transforms $(1, 0)$ into $(2, 5)$ and transforms $(0, 1)$ to $(1, 3)$?

    (b) What matrix C transforms $(2, 5)$ to $(1, 0)$ and $(1, 3)$ to $(0, 1)$?

    (c) Why does no matrix transform $(2, 6)$ to $(1, 0)$ and $(1, 3)$ to $(0, 1)$?

15   (a) What matrix M transforms $(1, 0)$ and $(0, 1)$ to (r, t) and (s, u)?

    (b) What matrix N transforms (a, c) and (b, d) to $(1, 0)$ and $(0, 1)$?

    (c) What condition on a, b, c, d will make part (b) impossible?

16   (a) How do M and N in Problem 15 yield the matrix that transforms (a, c) to (r, t) and (b, d) to (s, u)?

    (b) What matrix transforms $(2, 5)$ to $(1, 1)$ and $(1, 3)$ to $(0, 2)$?

17   If you keep the same basis vectors but put them in a different order, the change of basis matrix B is a _____ matrix. If you keep the basis vectors in order but change their lengths, B is a _____ matrix.

18   The matrix that rotates the axis vectors $(1, 0)$ and $(0, 1)$ through an angle θ is Q. What are the coordinates (a, b) of the original $(1, 0)$ using the new (rotated) axes? This *inverse* can be tricky. Draw a figure or solve for a and b:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \begin{bmatrix} 1 \\ 0 \end{bmatrix} = a \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} + b \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

19   The matrix that transforms $(1, 0)$ and $(0, 1)$ to $(1, 4)$ and $(1, 5)$ is $B = $ _____. The combination $a(1, 4) + b(1, 5)$ that equals $(1, 0)$ has $(a, b) = ($, $)$. How are those new coordinates of $(1, 0)$ related to B or B^{-1}?

Questions 20–23 are about the space of quadratic polynomials $y = A + Bx + Cx^2$.

20   The parabola $w_1 = \frac{1}{2}(x^2 + x)$ equals one at $x = 1$, and zero at $x = 0$ and $x = -1$. Find the parabolas w_2, w_3, and then find $y(x)$ by linearity.

    (a) w_2 equals one at $x = 0$ and zero at $x = 1$ and $x = -1$.

    (b) w_3 equals one at $x = -1$ and zero at $x = 0$ and $x = 1$.

    (c) $y(x)$ equals 4 at $x = 1$ and 5 at $x = 0$ and 6 at $x = -1$. Use w_1, w_2, w_3.

21   One basis for second-degree polynomials is $v_1 = 1$ and $v_2 = x$ and $v_3 = x^2$. Another basis is w_1, w_2, w_3 from Problem 20. Find two change of basis matrices, from the w's to the v's and from the v's to the w's.

22   What are the three equations for A, B, C if the parabola $y = A + Bx + Cx^2$ equals
4 at $x = a$ and 5 at $x = b$ and 6 at $x = c$? Find the determinant of the 3 by 3 matrix.
That matrix transforms values like $4, 5, 6$ to parabolas y—or is it the other way?

23   Under what condition on the numbers $m_1, m_2, \ldots, m_9$ do these three parabolas give
a basis for the space of all parabolas $a + bx + cx^2$?

$$v_1 = m_1 + m_2 x + m_3 x^2, \quad v_2 = m_4 + m_5 x + m_6 x^2, \quad v_3 = m_7 + m_8 x + m_9 x^2.$$

24   The Gram-Schmidt process changes a basis a_1, a_2, a_3 to an orthonormal basis
q_1, q_2, q_3. These are columns in $A = QR$. Show that R is the change of basis
matrix from the a's to the q's (a_2 is what combination of q's when $A = QR$?).

25   Elimination changes the rows of A to the rows of U with $A = LU$. Row 2 of A is
what combination of the rows of U? Writing $A^{\mathrm{T}} = U^{\mathrm{T}} L^{\mathrm{T}}$ to work with columns,
the change of basis matrix is $B = L^{\mathrm{T}}$. We have *bases* if the matrices are _____ .

26   Suppose v_1, v_2, v_3 are **eigenvectors** for T. This means $T(v_i) = \lambda_i v_i$ for $i = 1, 2, 3$.
What is the matrix for T when the input and output bases are the v's?

27   Every invertible linear transformation can have I as its matrix! Choose any input
basis $v_1, \ldots, v_n$. For output basis choose $w_i = T(v_i)$. Why must T be invertible?

28   Using $v_1 = w_1$ and $v_2 = w_2$ find the standard matrix for these T's:

   (a) $T(v_1) = 0$ and $T(v_2) = 3v_1$ (b) $T(v_1) = v_1$ and $T(v_1 + v_2) = v_1$.

29   Suppose T reflects the xy plane across the x axis and S is reflection across the y
axis. If $v = (x, y)$ what is $S(T(v))$? Find a simpler description of the product ST.

30   Suppose T is reflection across the $45°$ line, and S is reflection across the y axis. If
$v = (2, 1)$ then $T(v) = (1, 2)$. Find $S(T(v))$ and $T(S(v))$. Usually $ST \neq TS$.

31   **The product of two reflections is a rotation**. Multiply these reflection matrices to
find the rotation angle:

$$\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{bmatrix}.$$

32   Suppose A is a 3 by 4 matrix of rank $r = 2$, and $T(v) = Av$. Choose input basis
vectors v_1, v_2 from the row space of A and v_3, v_4 from the nullspace. Choose
output basis vectors $w_1 = Av_1$, $w_2 = Av_2$ in the column space and w_3 from the
nullspace of A^{T}. What specially simple matrix represents T in these special bases?

33   The space **M** of 2 by 2 matrices has the basis v_1, v_2, v_3, v_4 in Worked
Example **8.2 A**. Suppose T multiplies each matrix by $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. With w's equal to
v's, what 4 by 4 matrix A represents this transformation T on matrix space?

34   True or False: If we know $T(v)$ for n different nonzero vectors in $\mathbf{R}^n$, then we
know $T(v)$ for every vector v in $\mathbf{R}^n$.

8.3   The Search for a Good Basis

1 With a new input basis B_{in} and output basis B_{out}, every matrix A becomes $B_{\text{out}}^{-1}AB_{\text{in}}$.

2 $B_{\text{in}} = B_{\text{out}} =$ "**generalized eigenvectors of A**" produces the **Jordan form** $J = B^{-1}AB$.

3 The **Fourier matrix** $F = B_{\text{in}} = B_{\text{out}}$ diagonalizes every circulant matrix (use the **FFT**).

4 Sines and cosines, Legendre and Chebyshev: those are great bases for **function space**.

    This is an important section of the book. I am afraid that most readers will skip it—or won't get this far. The first chapters prepared the way by explaining the idea of a **basis**. Chapter 6 introduced the eigenvectors x and Chapter 7 found singular vectors v and u. Those are two winners but many other choices are very valuable.

    First comes the pure algebra from Section 8.2 and then come good bases. The input basis vectors will be the columns of B_{in}. The output basis vectors will be the columns of B_{out}. Always B_{in} and B_{out} are *invertible*—basis vectors are independent!

    Pure algebra   If A is the matrix for a transformation T in the standard basis, then

$$B_{\text{out}}^{-1}AB_{\text{in}} \text{ is the matrix in the new bases.} \tag{1}$$

The standard basis vectors are the *columns of the identity*: $B_{\text{in}} = I_{\text{n}\times\text{n}}$ and $B_{\text{out}} = I_{\text{m}\times\text{m}}$. Now we are choosing special bases to make the matrix clearer and simpler than A. When $B_{\text{in}} = B_{\text{out}} = B$, the square matrix $B^{-1}AB$ is *similar* to A.

    Applied algebra   Applications are all about choosing good bases. Here are four important choices for vectors and three choices for functions. Eigenvectors and singular vectors led to Λ and Σ in Section 8.2. The Jordan form is new.

1 $B_{\text{in}} = B_{\text{out}} =$ **eigenvector matrix** X.   Then $X^{-1}AX =$ **eigenvalues in** Λ.

    This choice requires A to be a square matrix with n independent eigenvectors. "A must be diagonalizable." We get Λ when $B_{\text{in}} = B_{\text{out}}$ is the eigenvector matrix X.

2 $B_{\text{in}} = V$ and $B_{\text{out}} = U$: **singular vectors of A**. Then $U^{-1}AV =$ **diagonal** Σ.

    Σ is the singular value matrix (with $\sigma_1, \ldots, \sigma_r$ on its diagonal) when B_{in} and B_{out} are the singular vector matrices V and U. Recall that those columns of B_{in} and B_{out} are orthonormal eigenvectors of $A^{\text{T}}A$ and AA^{T}. Then $A = U\Sigma V^{\text{T}}$ gives $\Sigma = U^{-1}AV$.

3 $B_{\text{in}} = B_{\text{out}} =$ **generalized eigenvectors of A**. Then $B^{-1}AB =$ **Jordan form J**.

    A is a square matrix but it may only have s independent eigenvectors. (If $s = n$ then B is X and J is Λ.) In all cases Jordan constructed $n - s$ additional "generalized" eigenvectors, aiming to make the Jordan form J *as diagonal as possible*:

      i)   There are s square blocks along the diagonal of J.

      ii)   Each block has one eigenvalue λ, one eigenvector, and 1's above the diagonal.

    The good case has n 1×1 blocks, each containing an eigenvalue. Then J is Λ (diagonal).

Example 1   This Jordan matrix J has eigenvalues $\lambda = 2, 2, 3, 3$ (two double eigenvalues). Those eigenvalues lie along the diagonal because J is triangular. There are two independent eigenvectors for $\lambda = 2$, but there is only *one line of eigenvectors for $\lambda = 3$*. This will be true for every matrix $C = BJB^{-1}$ that is similar to J.

$$\textbf{Jordan matrix}\quad J = \begin{bmatrix} 2 & & & \\ & 2 & & \\ & & 3 & 1 \\ & & 0 & 3 \end{bmatrix} \quad \begin{array}{l} \text{Two 1 by 1 blocks} \\ \text{One 2 by 2 block} \\ \text{Three eigenvectors} \\ \text{Eigenvalues } 2, 2, 3, 3 \end{array}$$

Two eigenvectors for $\lambda = 2$ are $x_1 = (1, 0, 0, 0)$ and $x_2 = (0, 1, 0, 0)$. One eigenvector for $\lambda = 3$ is $x_3 = (0, 0, 1, 0)$. The "generalized eigenvector" for this Jordan matrix is the fourth standard basis vector $x_4 = (0, 0, 0, 1)$. The eigenvectors for J (normal and generalized) are just the columns x_1, x_2, x_3, x_4 of the identity matrix I.

Notice $(J - 3I)x_4 = x_3$. **The generalized eigenvector x_4 connects to the true eigenvector x_3.** A true x_4 would have $(J - 3I)x_4 = 0$, but that doesn't happen here.

Every matrix $C = BJB^{-1}$ that is similar to this J will have true eigenvectors b_1, b_2, b_3 in the first three columns of B. The fourth column of B will be a generalized eigenvector b_4 of C, tied to the true b_3. Here is a quick proof that uses $Bx_3 = b_3$ and $Bx_4 = b_4$ to show: The fourth column b_4 is tied to b_3 by $(C - 3I)b_4 = b_3$.

$$(BJB^{-1} - 3I)\, b_4 = BJ\, x_4 - 3B\, x_4 = B(J - 3I)\, x_4 = B\, x_3 = b_3. \tag{2}$$

The point of Jordan's theorem is that every square matrix A has a complete set of eigenvectors and generalized eigenvectors. When those go into the columns of B, the matrix $B^{-1}AB = J$ is in Jordan form. Based on Example 1, here is a description of J.

The Jordan Form

For every A, we want to choose B so that $B^{-1}AB$ is as *nearly diagonal as possible*. When A has a full set of n eigenvectors, they go into the columns of B. Then $B = X$. The matrix $X^{-1}AX$ is diagonal, period. This is the Jordan form of A—when A can be diagonalized. In the general case, eigenvectors are missing and Λ can't be reached.

Suppose A has s independent eigenvectors. Then it is similar to a Jordan matrix with s blocks. Each block has an *eigenvalue on the diagonal with 1's just above it*. This block accounts for exactly one eigenvector of A. Then B contains generalized eigenvectors as well as ordinary eigenvectors.

When there are n eigenvectors, all n blocks will be 1 by 1. In that case $J = \Lambda$.

The Jordan form solves the differential equation $du/dt = Au$ for **any square matrix** $A = BJB^{-1}$. The solution $e^{At}u(0)$ becomes $u(t) = Be^{Jt}B^{-1}u(0)$. J is triangular and its matrix exponential e^{Jt} involves $e^{\lambda t}$ times powers $1, t, \ldots, t^{s-1}$.

(Jordan form) If A has s independent eigenvectors, it is similar to a matrix J that has s Jordan blocks $J_1 \ldots, J_s$ on its diagonal. Some matrix B puts A into Jordan form:

$$\textbf{Jordan form} \qquad B^{-1}AB = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{bmatrix} = J. \qquad (3)$$

Each block J_i has one eigenvalue λ_i, one eigenvector, and 1's just above the diagonal:

$$\textbf{Jordan block} \qquad J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}. \qquad (4)$$

Matrices are similar if they share the same Jordan form J—not otherwise.

The Jordan form J has an off-diagonal 1 for each missing eigenvector (and the 1's are next to the eigenvalues). In every family of similar matrices, we are picking one outstanding member called J. It is nearly diagonal (or if possible completely diagonal). We can quickly solve $du/dt = Ju$ and take powers J^k. Every other matrix in the family has the form BJB^{-1}.

Jordan's Theorem is proved in my textbook *Linear Algebra and Its Applications*. Please refer to that book (or more advanced books) for the proof. The reasoning is rather intricate and in actual computations the Jordan form is not at all popular—its calculation is not stable. A slight change in A will separate the repeated eigenvalues and remove the off-diagonal 1's—switching Jordan to a diagonal Λ.

Proved or not, you have caught the central idea of similarity—to make A as simple as possible while preserving its essential properties. The best basis B gives $B^{-1}AB = J$.

Question   Find the eigenvalues and all possible Jordan forms if $A^2 = $ zero matrix.

Answer   The eigenvalues must all be zero, because $Ax = \lambda x$ leads to $A^2x = \lambda^2 x = 0x$. The Jordan form of A has $J^2 = 0$ because $J^2 = (B^{-1}AB)(B^{-1}AB) = B^{-1}A^2B = 0$. Every block in J has $\lambda = 0$ on the diagonal. Look at J_k^2 for block sizes $1, 2, 3$:

$$[\,0\,]^2 = [\,0\,] \qquad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Conclusion: If $J^2 = 0$ then all block sizes must be 1 or 2. J^2 is not zero for 3 by 3.

The rank of J (and A) will be the total number of 1's. **The maximum rank is $n/2$.** This happens when there are $n/2$ blocks, each of size 2 and rank 1.

Now come the great bases of applied mathematics. Their discrete forms are vectors in $\mathbf{R}^n$. Their continuous forms are functions in a function space. Since they are chosen once and for all, *without knowing the matrix A*, these bases $B_{\text{in}} = B_{\text{out}}$ probably don't diagonalize A. But for many important matrices A in applied mathematics, the matrices $B^{-1}AB$ are *close to diagonal*.

4 $B_{\text{in}} = B_{\text{out}} =$ Fourier matrix F Then Fx is a Discrete Fourier Transform of x.

Those words are telling us : The Fourier matrix with columns $(1, \lambda, \lambda^2, \lambda^3)$ in equation (6) is important. Those are good basis vectors to work with.

We ask: Which matrices are diagonalized by F? This time we are starting with the eigenvectors $(1, \lambda, \lambda^2, \lambda^3)$ and finding the matrices that have those eigenvectors :

$$\text{If } \lambda^4 = 1 \text{ then } \quad Px = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \lambda^3 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \lambda^3 \end{bmatrix} = \lambda x. \quad (5)$$

P is a permutation matrix. The equation $Px = \lambda x$ says that x is an eigenvector and λ is an eigenvalue of P. Notice how the fourth row of this vector equation is $1 = \lambda^4$. That rule for λ makes everything work.

Does this give four different eigenvalues λ? *Yes.* The four numbers $\lambda = 1, i, -1, -i$ all satisfy $\lambda^4 = 1$. (You know $i^2 = -1$. Squaring both sides gives $i^4 = 1$.) So those four numbers are the eigenvalues of P, each with its eigenvector $x = (1, \lambda, \lambda^2, \lambda^3)$. **The eigenvector matrix F diagonalizes the permutation matrix P :**

$$\begin{matrix} \textbf{Eigenvalue} \\ \textbf{matrix } \Lambda \end{matrix} \quad \begin{bmatrix} 1 & & & \\ & i & & \\ & & -1 & \\ & & & -i \end{bmatrix} \quad \begin{matrix} \textbf{Eigenvector} \\ \textbf{matrix is} \\ \textbf{Fourier} \\ \textbf{matrix } F \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & i^2 & 1 & (-i)^2 \\ 1 & i^3 & -1 & (-i)^3 \end{bmatrix} \quad (6)$$

Those columns of F are orthogonal because they are eigenvectors of P (an orthogonal matrix). Unfortunately this Fourier matrix F is complex (it is the most important complex matrix in the world). Multiplications Fx are done millions of times very quickly, by the Fast Fourier Transform. The FFT comes in Section 9.3.

Key question : What other matrices beyond P have this same eigenvector matrix F ? We know that P^2 and P^3 and P^4 have the same eigenvectors as P. The same matrix F diagonalizes all powers of P. And the eigenvalues of P^2 and P^3 and P^4 are the numbers λ^2 and λ^3 and λ^4. For example $P^2x = \lambda^2 x$:

$$P^2x = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \lambda^3 \end{bmatrix} = \lambda^2 \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \lambda^3 \end{bmatrix} = \lambda^2 x \text{ when } \lambda^4 = 1.$$

The fourth power is special because $P^4 = I$. When we do the "cyclic permutation" four times, $P^4 x$ is the same vector x that we started with. The eigenvalues of $P^4 = I$ are just $1, 1, 1, 1$. And that number 1 agrees with the fourth power of all the eigenvalues of P: $1^4 = 1$ and $i^4 = 1$ and $(-1)^4 = 1$ and $(-i)^4 = 1$.

One more step brings in many more matrices. If P and P^2 and P^3 and $P^4 = I$ have the same eigenvector matrix F, so does any combination $C = c_1 P + c_2 P^2 + c_3 P^3 + c_0 I$:

$$\textbf{Circulant matrix } C = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 \\ c_3 & c_0 & c_1 & c_2 \\ c_2 & c_3 & c_0 & c_1 \\ c_1 & c_2 & c_3 & c_0 \end{bmatrix}$$

has eigenvectors in the Fourier matrix F
has four eigenvalues $c_0 + c_1 \lambda + c_2 \lambda^2 + c_3 \lambda^3$
from the four numbers $\lambda = 1, i, -1, -i$
The eigenvalue from $\lambda = 1$ is $c_0 + c_1 + c_2 + c_3$

That was a big step. We have found all the matrices (circulant matrices C) whose eigenvectors are the Fourier vectors in F. We also know the four eigenvalues of C, but we haven't given them a good formula or a name until now:

The four eigenvalues of C are given by the Fourier transform Fc

$$Fc = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{matrix} c_0 + c_1 + c_2 + c_3 \\ c_0 + ic_1 - c_2 - ic_3 \\ c_0 - c_1 + c_2 - c_3 \\ c_0 - ic_1 - c_2 + ic_3 \end{matrix}$$

Example 2   The same ideas work for a Fourier matrix F and a circulant matrix C of any size. Two by two matrices look trivial but they are very useful. Now eigenvalues of P have $\lambda^2 = 1$ instead of $\lambda^4 = 1$ and the complex number i is not needed: $\boldsymbol{\lambda = \pm 1}$.

Fourier matrix F from eigenvectors of P and C $F = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ Circulant $c_0 I + c_1 P$ $C = \begin{bmatrix} c_0 & c_1 \\ c_1 & c_0 \end{bmatrix}$.

The eigenvalues of C are $c_0 + c_1$ and $c_0 - c_1$. Those are given by the Fourier transform Fc when the vector c is (c_0, c_1). This transform Fc gives the eigenvalues of C for any size n.

Notice that **circulant matrices have constant diagonals**. The same number c_0 goes down the main diagonal. The number c_1 is on the diagonal above, and that diagonal "wraps around" or "circles around" to the southwest corner of C. This explains the name *circulant* and it indicates that these matrices are *periodic* or *cyclic*. Even the powers of λ cycle around because $\lambda^4 = 1$ leads to $\lambda^5, \lambda^6, \lambda^7, \lambda^8 = \lambda, \lambda^2, \lambda^3, \lambda^4$.

Constancy down the diagonals is a crucial property of C. It corresponds to *constant coefficients* in a differential equation. This is exactly when Fourier works perfectly!

The equation $\dfrac{d^2 u}{dt^2} = -u$ is solved by $u = c_0 \cos t + c_1 \sin t$.

The equation $\dfrac{d^2 u}{dt^2} = tu$ cannot be solved by elementary functions.

These equations are linear. The first is the oscillation equation for a simple spring. It is Newton's Law $f = ma$ with mass $m = 1$, $a = d^2 u / dt^2$, and force $f = -u$. Constant coefficients produce the differential equations that you can really solve.

The equation $u'' = tu$ has a variable coefficient t. This is Airy's equation in physics and optics (it was derived to explain a rainbow). The solutions change completely when t passes through zero, and those solutions require infinite series. *We won't go there.*

The point is that equations with constant coefficients have simple solutions like $e^{\lambda t}$. You discover λ by substituting $e^{\lambda t}$ into the differential equation. That number λ is like an eigenvalue. For $u = \cos t$ and $u = \sin t$ the number is $\lambda = i$. Euler's great formula $e^{it} = \cos t + i \sin t$ introduces complex numbers as we saw in the eigenvalues of P and C.

Bases for Function Space

For functions of x, the first basis I would think of contains the powers $1, x, x^2, x^3, \ldots$ Unfortunately this is a terrible basis. Those functions x^n are just barely independent. x^{10} is *almost* a combination of other basis vectors $1, x, \ldots, x^9$. It is virtually impossible to compute with this poor "ill-conditioned" basis.

If we had vectors instead of functions, the test for a good basis would look at $B^T B$. This matrix contains all inner products between the basis vectors (columns of B). *The basis is orthonormal when $B^T B = I$.* That is best possible. But the basis $1, x, x^2, \ldots$ produces the evil **Hilbert matrix**: $B^T B$ has an enormous ratio between its largest and smallest eigenvalues. A large condition number signals an unhappy choice of basis.

Note Now the columns of B are functions instead of vectors. We still use $B^T B$ to test for independence. So we need to know the dot product (inner product is a better name) of two functions—those are the numbers in $B^T B$.

The dot product of vectors is just $x^T y = x_1 y_1 + \cdots + x_n y_n$. The inner product of functions will integrate instead of adding, but the idea is completely parallel:

$$\text{Inner product } (f, g) \quad = \quad \int f(x) g(x) \, dx$$

$$\text{Complex inner product } (f, g) \quad = \quad \int \overline{f(x)} \, g(x) \, dx, \ \ \overline{f} = \text{ complex conjugate}$$

$$\text{Weighted inner product } (f, g)_w \quad = \quad \int w(x) \, \overline{f(x)} \, g(x) \, dx, \ \ w = \text{ weight function}$$

When the integrals go from $x = 0$ to $x = 1$, the inner product of x^i with x^j is

$$\int_0^1 x^i \, x^j \, dx = \frac{x^{i+j+1}}{i+j+1} \Bigg]_{x=0}^{x=1} = \frac{1}{i+j+1} = \text{ entries of Hilbert matrix } B^T B$$

By changing to the symmetric interval from $x = -1$ to $x = 1$, we immediately have *orthogonality between all even functions and all odd functions*:

Interval $[-1, \ 1]$ $\qquad \displaystyle\int_{-1}^1 x^2 \, x^5 \, dx = 0 \qquad \displaystyle\int_{-1}^1 \mathbf{even}(x) \, \mathbf{odd}(x) \, dx = 0.$

This change makes half of the basis functions orthogonal to the other half. It is so simple that we continue using the symmetric interval -1 to 1 (or $-\pi$ to π). But we want a better basis than the powers x^n—hopefully an orthogonal basis.

Orthogonal Bases for Function Space

Here are the three leading even-odd bases for theoretical and numerical computations:

5. The **Fourier basis** $\qquad 1, \sin x, \cos x, \sin 2x, \cos 2x, \ldots$

6. The **Legendre basis** $\qquad 1, x, x^2 - \dfrac{1}{3}, x^3 - \dfrac{3}{5}x, \ldots$

7. The **Chebyshev basis** $\qquad 1, x, 2x^2 - 1, 4x^3 - 3x, \ldots$

The Fourier basis functions (sines and cosines) are all *periodic*. They repeat over every 2π interval because $\cos(x+2\pi) = \cos x$ and $\sin(x+2\pi) = \sin x$. So this basis is especially good for functions $f(x)$ that are themselves periodic: $f(x + 2\pi) = f(x)$.

This basis is also *orthogonal*. Every sine and cosine is orthogonal to every other sine and cosine. Of course we don't expect the basis function $\cos nx$ to be orthogonal to itself.

Most important, the sine-cosine basis is also *excellent for approximation*. If we have a smooth periodic function $f(x)$, then a few sines and cosines (low frequencies) are all we need. Jumps in $f(x)$ and noise in the signal are seen in higher frequencies (larger n). We hope and expect that the signal is not drowned by the noise.

The *Fourier transform* connects $f(x)$ to the coefficients a_k and b_k in its Fourier series:

Fourier series $\qquad f(x) = a_0 + b_1 \sin x + a_1 \cos x + b_2 \sin 2x + a_2 \cos 2x + \cdots$

We see that **function space is infinite-dimensional**. It takes infinitely many basis functions to capture perfectly a typical $f(x)$. But the formula for each coefficient (for example a_3) is just like the formula $b^{\mathsf{T}}a/a^{\mathsf{T}}a$ for projecting a vector b onto the line through a.

Here we are projecting the function $f(x)$ onto the line in function space through $\cos 3x$:

$$\textbf{Fourier coefficient} \quad a_3 = \frac{(f(x), \cos 3x)}{(\cos 3x, \cos 3x)} = \frac{\int f(x) \cos 3x\, dx}{\int \cos 3x \cos 3x\, dx}. \tag{7}$$

Example 3 The double angle formula in trigonometry is $\cos 2x = 2\cos^2 x - 1$. This tells us that $\cos^2 x = \frac{1}{2} + \frac{1}{2}\cos 2x$. A very short Fourier series. So is $\sin^2 x = \frac{1}{2} - \frac{1}{2}\cos 2x$.

Fourier series is just linear algebra in function space. Let me explain that properly as a highlight of Chapter 10 about applications.

Legendre Polynomials and Chebyshev Polynomials

The Legendre polynomials are the result of applying the Gram-Schmidt idea (Section 4.4). The plan is to orthogonalize the powers $1, x, x^2, \ldots$ To start, the odd function x is already orthogonal to the even function 1 over the interval from -1 to 1. Their product $(x)(1) = x$ integrates to zero. But the inner product between x^2 and 1 is $\int x^2 \, dx = 2/3$:

$$\frac{(x^2, 1)}{(1, 1)} = \frac{\int x^2 \, dx}{\int 1 \, dx} = \frac{2/3}{2} = \frac{1}{3} \qquad \text{Gram-Schmidt gives } x^2 - \frac{1}{3} = \textbf{Legendre}$$

Similarly the odd power x^3 has a component $3x/5$ in the direction of the odd function x :

$$\frac{(x^3, x)}{(x, x)} = \frac{\int x^4 \, dx}{\int x^2 \, dx} = \frac{2/5}{2/3} = \frac{3}{5} \qquad \text{Gram-Schmidt gives } x^3 - \frac{3}{5}x = \textbf{Legendre}$$

Continuing Gram-Schmidt for $x^4, x^5, \ldots$ produces every Legendre function—a good basis.

Finally we turn to the Chebyshev polynomials $1, x, 2x^2 - 1, 4x^3 - 3x$. They don't come from Gram-Schmidt. Instead they are connected to $1, \cos\theta, \cos 2\theta, \cos 3\theta$. This gives a giant computational advantage—we can use the Fast Fourier Transform. The connection of Chebyshev to Fourier appears when we set $x = \cos\theta$:

Chebyshev	$2x^2 - 1 = 2(\cos\theta)^2 - 1 = \cos 2\theta$
to Fourier	$4x^3 - 3x = 4(\cos\theta)^3 - 3(\cos\theta) = \cos 3\theta$

The nth degree Chebyshev polynomial $T_n(x)$ converts to Fourier's $\cos n\theta = T_n(\cos\theta)$.

Note   These polynomials are the basis for a big software project called **"chebfun"**. Every function $f(x)$ is replaced by a super-accurate Chebyshev approximation. Then you can integrate $f(x)$, and solve $f(x) = 0$, and find its maximum or minimum. More than that, you can solve differential equations involving $f(x)$—fast and to high accuracy.

When **chebfun** replaces $f(x)$ by a polynomial, you are ready to solve problems.

■ REVIEW OF THE KEY IDEAS ■

1. A basis is good if its matrix B is well-conditioned. Orthogonal bases are best.

2. Also good if $\Lambda = B^{-1}AB$ is diagonal. But the Jordan form J can be very unstable.

3. The Fourier matrix diagonalizes constant-coefficient periodic equations: perfection.

4. The basis $1, x, x^2, \ldots$ leads to $B^T B = $ Hillbert matrix: Terrible for computations.

5. Legendre and Chebyshev polynomials are excellent bases for function space.

Problem Set 8.3

1   In Example 1, what is the rank of $J - 3I$? What is the dimension of its nullspace ? This dimension gives the number of independent eigenvectors for $\lambda = 3$.

The algebraic multiplicity is 2, because $\det(J - \lambda I)$ has the repeated factor $(\lambda - 3)^2$. The geometric multiplicity is 1, because there is only 1 independent eigenvector.

2   These matrices A_1 and A_2 are similar to J. Solve $A_1 B_1 = B_1 J$ and $A_2 B_2 = B_2 J$ to find the basis matrices B_1 and B_2 with $J = B_1^{-1} A_1 B_1$ and $J = B_2^{-1} A_2 B_2$.

$$J = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad A_1 = \begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 4 & -8 \\ 2 & -4 \end{bmatrix}$$

3   This transpose block J^T has the same triple eigenvalue 2 (with only one eigenvector) as J. Find the basis change B so that $J = B^{-1} J^T B$ (which means $BJ = J^T B$):

$$J = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix} \qquad J^T = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

4 J and K are Jordan forms with the same zero eigenvalues and the same rank 2. But show that no invertible B solves $BK = JB$, so K *is not similar to J* :

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

5   If $A^3 = 0$ show that all $\lambda = 0$, and all Jordan blocks with $J^3 = 0$ have size $1, 2$, or 3. It follows that rank $(A) \le 2n/3$. If $A^n = 0$ why is rank $(A) < n$?

6   Show that $u(t) = \begin{bmatrix} te^{\lambda t} \\ e^{\lambda t} \end{bmatrix}$ solves $\dfrac{du}{dt} = Ju$ with $J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$ and $u(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

J is not diagonalizable so $te^{\lambda t}$ enters the solution.

7   Show that the difference equation $v_{k+2} - 2\lambda v_{k+1} + \lambda^2 v_k = 0$ is solved by $v_k = \lambda^k$ and also by $v_k = k\lambda^k$. Those correspond to $e^{\lambda t}$ and $te^{\lambda t}$ in Problem 6.

8   What are the 3 solutions to $\lambda^3 = 1$? They are complex numbers $\lambda = \cos\theta + i\sin\theta = e^{i\theta}$. Then $\lambda^3 = e^{3i\theta} = 1$ when the angle 3θ is 0 or 2π or 4π. Write the 3 by 3 Fourier matrix F with columns $(1, \lambda, \lambda^2)$.

9   Check that any 3 by 3 circulant C has eigenvectors $(1, \lambda, \lambda^2)$ from Problem 8. If the diagonals of your matrix C contain c_0, c_1, c_2 then its eigenvalues are in Fc.

10   Using formula (7) find $a_3 \cos 3x$ in the Fourier series of $f(x) = \begin{cases} 1 \text{ for } -L \le x \le L \\ 0 \text{ for } \quad L \le |x| \le 2\pi \end{cases}$