Math 34600 L (43597) - Lectures 03

Ethan Akin
Office: Marshak 325
Email: eakin@ccny.cuny.edu

Spring, 2024

Contents

Linear Transformations

Earlier we define a linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ defined by multiplying the $m \times n$ matrix A.

In general, a *linear transformation* or *linear map* is a function $T : V \to W$ between the vector spaces V and W. That is, the inputs and outputs are vectors, and T satisfies *linearity*, which is also called the *superposition property*:

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) \quad \text{and} \quad T(a\mathbf{v}) = aT(\mathbf{v}). \tag{6.1}$$

In particular, $T(\mathbf{0}) = T(0\mathbf{0}) = 0T(\mathbf{0}) = \mathbf{0}$, and $T(-\mathbf{v}) = T((-1)\mathbf{v}) = (-1)T(\mathbf{v}) = -T(\mathbf{v})$.
It follows that T relates linear combinations:

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots\ c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \ldots c_kL(\mathbf{v}_k). \tag{6.2}$$

This property of linearity is very special. It is a standard algebra mistake to apply it to functions like the square root function and sin and cos etc. for which it does not hold. On the other hand, these should be familiar properties from calculus. The operator D associating to a differentiable function f its derivative Df is a most important example of a linear operator.

From a linear map we get an important examples of subspaces.

For a linear map $T : V \to W$, the set of vectors $\{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$ *solution space of the homogeneous equation*, is a subspace of V called the kernel of T, $Ker(T)$. If $\mathbf{r}$ is not $\mathbf{0}$ then the solution space of $T(\mathbf{v}) = \mathbf{r}$ is not a subspace. For example, it does not contain $\mathbf{0}$.

For a linear map $T : V \to W$, the set of vectors $\{\mathbf{w} \in W :$ for some $\mathbf{v} \in V, \ T(\mathbf{v}) \ = \mathbf{w}\}$ is called *the image of L*, denoted $Im(T)$.

Check that $Ker(T) \subset V$ and $Im(T) \subset W$ are subspaces.

For the linear map $T_A : \mathbb{R}^n \to \mathbb{R}^m$ associated with the $m \times n$ matrix A, $Ker(T_A) = Null(A)$ and $Im(T_A) = Col(A)$.

Theorem 6.01: (a) If $T : V \to W$ and $S : W \to U$ are linear maps, then the composition $S \circ T : V \to U$ defined by $S \circ T(\mathbf{v}) = S(T(\mathbf{v}))$ is a linear map.

(b) A linear map $T : V \to W$ is one-to-one if and only if $Ker(T) = \{\mathbf{0}\}$.

(c) A linear map $T : V \to W$ is onto if and only if $Im(T) = W$.

(d) If a linear map $T : V \to W$ is one-to-one and onto, then the inverse map $T^{-1} : W \to V$ defined by

$$T^{-1}(\mathbf{w}) = \mathbf{v} \quad \Leftrightarrow \quad T(\mathbf{v}) = \mathbf{w} \tag{6.3}$$

is a linear map.

A one-to-one, onto linear map is called a *linear isomorphism*.

Proof: (a)
$S(T(c\mathbf{v}_1+\mathbf{v}_2) = S(cT(\mathbf{v}_1)+T(\mathbf{v}_2)) = cS(T(\mathbf{v}_1))+S(T(\mathbf{v}_2))$.

(b) $T(\mathbf{0}) = \mathbf{0}$ and so if T is one-to-one, $Ker(T) = \{\mathbf{0}\}$. Conversely, if $T(\mathbf{v}_1) = T(\mathbf{v}_2)$, then $T(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$. So if $Ker(T) = \{\mathbf{0}\}$, we have $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$ and so $\mathbf{v}_1 = \mathbf{v}_2$.

(c) This is clear from the definition of $Im(T)$.

(d) $T^{-1}(c\mathbf{w}_1+\mathbf{w}_2) = c\mathbf{v}_1+\mathbf{v}_2 \quad \Leftrightarrow \quad c\mathbf{w}_1+\mathbf{w}_2 = T(c\mathbf{v}_1+\mathbf{v}_2)$.
$\square$

Theorem 6.02: Let $T : V \to W$ be a linear map and $D = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a list of vectors in V. Define $T(D) = \{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$

(a) If $T(D)$ is an li list in W, then D is an li list in V.

(b) If D is an li list in V and $Ker(T) = \{\mathbf{0}\}$, then If $T(D)$ is an li list in W.

(c) If D spans V and $Im(T) = W$, then $T(D)$ spans W.

(d) If D is a basis for V and T is a linear isomorphism, then $T(D)$ is a basis for W.

(e) If D spans V and $S : V \to W$ is a linear map with $T(\mathbf{v}_1) = S(\mathbf{v}_2), \dots, T(\mathbf{v}_1) = S(\mathbf{v}_2)$, then $T = S$. That is, $T(\mathbf{v}) = S(\mathbf{v})$ for all $\mathbf{v} \in V$.

Proof: (a) Assume $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$. We must show $c_1 = \cdots = c_n = 0$.

$\mathbf{0} = T(\mathbf{0}) = T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = c_1 T(\mathbf{v}_1) + \cdots + c_n T(\mathbf{v}_n)$. Because the list $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$ is li, it follows that $c_1 = \cdots = c_n = 0$.

(b) Assume $\mathbf{0} = c_1 T(\mathbf{v}_1) + \cdots + c_n T(\mathbf{v}_n) = T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n)$. Because $Ker(T) = \{\mathbf{0}\}$, $\mathbf{0} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Because D is li, $c_1 = \cdots = c_n = 0$.

(c) For $\mathbf{w} \in W$, there exists $\mathbf{v}$ with $T(\mathbf{v}) = \mathbf{w}$ (Why?). There exist coefficients so that $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ (Why?). So $\mathbf{w} = c_1 T(\mathbf{v}_1) + \cdots + c_n T(\mathbf{v}_n)$.

(d) follows from (b) and (c).

(e) Check that $\{\mathbf{v} : T(\mathbf{v}) = S(\mathbf{v})\}$ is a subspace of V because S and T are linear. The subspace contains the spanning set D and so equals all of V.

$\square$

Theorem 6.03: For $T : V \to W$ a linear map,

$$dimV = dimKer(T) + dimIm(T). \tag{6.4}$$

Proof: Every vector of $Im(T)$ is of the form $T(\mathbf{v})$ and so we can choose a basis $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_r)\}$ for $Im(T)$. Let $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ be a basis for $Ker(T)$. We will show that $\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots \mathbf{v}_n\}$ is a basis for V which will show that $n = dimV$.

Assume $\mathbf{0} = c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r + c_{r+1}\mathbf{v}_{r+1} + \ldots c_n\mathbf{v}_n$. We must show $c_1 = \ldots c_r = c_{r+1} \cdots = c_n = 0$.

Apply T and note the $\mathbf{v}_i \in Ker(T)$ for $r < i \leq n$ implies $\mathbf{0} = c_1 T(\mathbf{v}_1) + \cdots + c_r T(\mathbf{v}_r)$. So $c_1 = \cdots = c_r = 0$ (Why?)

This implies that $\mathbf{0} = c_{r+1}\mathbf{v}_{r+1} + \ldots c_n\mathbf{v}_n$. So $c_{r+1} = \cdots = c_n = 0$ (Why?)

Thus, $\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots \mathbf{v}_n\}$ is li.

Now let $\mathbf{v} \in V$. We must find coefficients so that $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r + c_{r+1}\mathbf{v}_{r+1} + \ldots c_n\mathbf{v}_n$.

There exist coefficients so that $T(\mathbf{v}) = c_1 T(\mathbf{v}_1) + \cdots + c_r T(\mathbf{v}_r)$ (Why?)

$$T(\mathbf{v} - c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r) \;=\; T(\mathbf{v}) - (c_1 T(\mathbf{v}_1) + \cdots + c_r T(\mathbf{v}_r)) \;=\; \mathbf{0}$$

and so there exist coefficients such that

$$\mathbf{v} - c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r = c_{r+1}\mathbf{v}_{r+1} + \ldots c_n\mathbf{v}_n.$$

Thus, $\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots \mathbf{v}_n\}$ spans V.
$\square$

Theorem 6.04: Let $D = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a list of vectors in V. The map $T_D : \mathbb{R}^n \to V$ defined by

$$T_D(x_1, \ldots, x_n) \;=\; \sum_{i=1}^{n} x_1\mathbf{v}_i \;=\; x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n \qquad (6.5)$$

is a linear map.

If D is a basis, then T_D is a linear isomorphism and inverse map $T_D^{-1} : V \to \mathbb{R}^n$ is given by $T_D^{-1}(v) \;=\; [\mathbf{v}]_D$, the coordinate vector of $\mathbf{v}$ with respect to the basis D.

Proof: We saw above that the sum of two linear combinations on a list is the linear combination obtained by adding the corresponding coefficients. Similarly,

$T_D(cx_1, \ldots, cx_n) = cT_D(x_1, \ldots, x_n)$. Thus, T_D is linear.

If D is a basis then for any $\mathbf{v} \in V$ the equation $\mathbf{v} \;=\; x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ uniquely defines the coefficients and the list of coefficients is the coordinate vector $[\mathbf{v}]_D$.
$\square$

Corollary 6.05: For finite dimensional vector spaces V and W, there is a linear isomorphism $T : V \to W$ if and only if $dimV = dimW$. In particular, if $dimV = n$, then there is a linear isomorphism $T : \mathbb{R}^n \to V$.

Proof: If $D = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V, then $T_D : \mathbb{R}^n \to V$ is a linear isomorphism and such a basis exists exactly when $dimV = n$.

Thus, if $dimV = dimW = n$ then there exist linear isomorphisms $T : V \to \mathbb{R}^n$ and $S : W \to \mathbb{R}^n$ and so the composition $S^{-1} \circ T : V \to W$ is a linear isomorphism.

On the other hand, if $T : V \to W$ is a linear isomorphism and D is a basis for V, then by Theorem 6.02(d) $T(D)$ is a basis for W. Therefore, $dimV = \#D = \#T(D) = dimW$.
□

Let us look at Exercise 7.2/ 1b, page 385.

Matrix of a Linear Transformation

Recall that for A an $m \times n$ matrix the linear map $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $T_A(X) = AX$. By using bases we can represent every linear map between finite dimensional vector spaces in this way.

If $T : V \to W$ is a linear map and $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}, D = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ are bases for V and W, respectively, then the matrix $[T]_{DB}$ is the $m \times n$ matrix given by

$$[T]_{DB} = [\ [T(\mathbf{v}_1)]_D \ldots [T(\mathbf{v}_n)]_D\ ]. \qquad (6.6)$$

That is, we form the matrix by applying T to each of the domain basis vectors from B in V. We list them in order, thinking of them as a matrix but with vectors in W instead of columns of numbers. We convert each vector to an actual column of numbers by replacing each by its column of D coordinates. Thus, we obtain the $m \times n$ matrix $[T]_{DB}$.

Theorem 6.06: Let $T : V \to W$ be a linear map with $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}, D = \{\mathbf{w}_1, ..., \mathbf{w}_m\}$ bases for V and W. Let $[T]_{DB}$ be the $m \times n$ matrix associated to the linear map by the choice of bases. If $\mathbf{v} \in V$, then

$$[T(\mathbf{v})]_D = [T]_{DB}[\mathbf{v}]_B. \tag{6.7}$$

That is, the D coordinate vector of $\mathbf{w} = T(\mathbf{v})$ in $\mathbb{R}^m$ is

obtained by multiplying the B coordinate vector of $\mathbf{v}$ in $\mathbb{R}^n$ by the $m \times n$ matrix $[T]_{DB}$.

Proof: By definition $[\mathbf{v}]_B = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$ means $\mathbf{v} = x_1\mathbf{v}_1 + \ldots x_n\mathbf{v}_n$, and so $\mathbf{w} = T(\mathbf{v}) = x_1 T(\mathbf{v}_1) + \ldots x_n T(\mathbf{v}_n)$. By Theorem 6.04, the coordinate map $\mathbf{w} \mapsto [\mathbf{w}]_D$ is linear and so

$$[\mathbf{w}]_D \;=\; x_1[T(\mathbf{v}_1)]_D + \ldots x_n[T(\mathbf{v}_n)]_D.$$

This is, the linear combination of the columns of $[T]_{DB}$ with coefficients $x_1, \ldots, x_n$. That is exactly

$$[T]_{DB} \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \;=\; [T]_{DB}[\mathbf{v}]_B.$$

□

Corollary 6.07: Let $T : V \to W$ and $S : W \to U$ be linear maps with B, D, E bases for V, W and U.

$$[S \circ T]_{EB} \;=\; [S]_{ED}[T]_{DB}. \tag{6.8}$$

That is, the matrix of the composed linear map $S \circ T$ is the

product of the matrices of S and T provided that the same basis D is used for W as the range of T and as the domain of S.

Proof: Let $\mathbf{v}$ be an arbitrary vector in V. By Theorem 6.06 applied first to $S \circ T$, then to S and then to T we have

$$\begin{array}{c} [S \circ T]_{EB}[\mathbf{v}]_B \;=\; [(S \circ T)(\mathbf{v})]_E \;=\; \\ [(S(T(\mathbf{v}))]_E \;=\; [S]_{ED}[T(\mathbf{v})]_D \;= [S]_{ED}[T]_{DB}[\mathbf{v}]_B. \end{array} \tag{6.9}$$

The result follows because if A and B are $m \times n$ matrices such that $AX = BX$ for all $X \in \mathbb{R}^n$, then $A = B$. (Hint: let X vary over the columns of I_n, which list is the standard basis for $\mathbb{R}^n$).
$\square$

An important special case lets us change the coordinates from one basis to another. We use the identity map I on the vector space V, so that $I(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v}$ in V, but we use different bases on the domain and range.
Let $B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}, D = \{\mathbf{w}_1, ..., \mathbf{w}_n\}$ be two different bases on a vector space V. They have the same number n of elements when $dimV = n$. The transition matrix from B to D is given by:

$$[I]_{DB} = [\ [\mathbf{v}_1]_D ... [\mathbf{v}_n)]_D\ ]. \tag{6.10}$$

That is, the columns are the D coordinates of the B vectors listed in order.

Corollary 6.08: Let B and D be bases for a vector space V of dimension n.

(a) $[I]_{BB} = I_n$. That is, the transition matrix from a basis to itself is the identity matrix.

(b) $[I]_{BD} = ([I]_{DB})^{-1}$. That is, the transition matrix from D to B is the inverse matrix of the transition matrix from B to D.

(c) For any vector $\mathbf{v} \in V$,

$$[\mathbf{v}]_D = [I]_{DB}[\mathbf{v}]_B. \qquad (6.11)$$

Proof: (a) is easy to check, e.g. $\mathbf{v}_1 = 1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n$. Then (b) follows from Corollary 6.07.

Finally, (c) is a special case of Theorem 6.06.
□

As we have seen, many of the spaces we look at have a standard basis S whose coordinate vectors are easy to read off. If $T : V \to W$ is a linear map with B is a basis for V and S is a standard basis for W, then it is easy to compute $[T]_{SB}$.

For example, if A is an $m \times n$ matrix and $X \in \mathbb{R}^n$, then with respect to the standard bases S_n on $\mathbb{R}^n$ and S_m on R^m, just as the coordinate vector $[X]_{S_n}$ is X itself, so too $[T_A]_{S_m S_n} = A$.

If $T : V \to W$ is a linear map with $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}, D = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ bases for V and W and S is a standard basis for W, then usually $[T]_{SB}$ and $[I]_{SD}$ are easy to read directly. It is then sometimes easiest to use the following application of Corollaries 6.07 and 6.08:

$$[T]_{DB} = [I]_{DS}[T]_{SB} = ([I]_{SD})^{-1}[T]_{SB}. \qquad (6.12)$$

Let us look at Exercises 9.1/1ad, page 501.

Eigenvalues, Eigenvectors

Suppose $T : V \to V$ is a linear map on an n dimensional vector space V. Since the domain and range are the same space, we usually choose the same basis for the domain and range. If B and D are bases for V then from Corollaries 6.07 and 6.08 we have

$$[T]_{DD} = [I]_{DB}[T]_{BB}[I]_{BD} = ([I]_{BD})^{-1}[T]_{BB}[I]_{BD}. \quad (7.1)$$

All of these matrices are square $n \times n$ matrices.

Again if B is a standard basis, then $[T]_{BB}$ and $[I]_{BD}$ are usually easy to compute.

The question arises, if there is a standard basis, why use any other? The answer is that for a particular problem or particular matrix an alternative basis may be more useful.

If you have taken elementary Physics, then one of the first class of problems you saw concerned motion on an inclined plane. To solve these problems you resolved the vectors associated with the weight and the friction force into component parallel and perpendicular (or normal) to the plane. In effect, you replaced the standard basis $\{\mathbf{i}, \mathbf{j}\}$ by $\{\mathbf{e}_P, \mathbf{e}_N\}$ unit vectors parallel to and normal to the inclined plane.

Given a linear map $T : V \to V$ or associated matrix A we will look for a basis of eigenvectors.

A nonzero vector $\mathbf{v}$ is an *eigenvector* for L with *eigenvalue* λ when $L(\mathbf{v}) = \lambda\mathbf{v}$. That is, $L(v)$ is just a multiple of $\mathbf{v}$. Of course, if $\mathbf{v} = \mathbf{0}$, then $L(\mathbf{v}) = \lambda\mathbf{v}$ for any λ, but if $\mathbf{v} \neq \mathbf{0}$, then the eigenvalue is uniquely determined by the eigenvector.

Proof: If $L(\mathbf{v}) = \lambda_1\mathbf{v} = \lambda_2\mathbf{v}$, then $(\lambda_1 - \lambda_2)\mathbf{v} = \mathbf{0}$ and since $\mathbf{v} \neq \mathbf{0}$ this means $\lambda_1 - \lambda_2 = 0$ and so $\lambda_1 = \lambda_2$. $\square$

A nonzero vector $\mathbf{v}$ is an eigenvector with eigenvalue $\lambda = 0$ if and only if $\mathbf{v}$ is in the kernel of L.

For an $n \times n$ matrix A an eigenvector is a nonzero $n \times 1$ column vector X such that $AX = \lambda X$ or, equivalently, $(\lambda I - A)X = \mathbf{0}$. Thus, an eigenvector for the matrix A is exactly an eigenvector for the linear map T_A.

For a linear map L on V or an $n \times n$ matrix A, the *eigenspace* $E_\lambda(L)$ or $E_\lambda(A)$ is the subspace defined by

$$E_\lambda(L) = \{\mathbf{v} \in V : L(\mathbf{v}) = \lambda \mathbf{v}\}.$$

$$E_\lambda(A) = E_\lambda(T_A) = \{X \in \mathbb{R}^n : AX = \lambda X\} = Null(\lambda I - A). \tag{7.2}$$

So $E_\lambda(A)$ consists of the eigenvectors of A with eigenvalue λ (if any) together with the zero vector.

In particular, $E_0(A) = Null(A)$ and $E_0(L)$ is the kernel of L.

You might think that we find the eigenvectors of the matrix A and then for each one multiply by A to get the associated eigenvalue. In fact, we do the reverse finding the eigenvalues first.

For most values of λ the nullspace $Null(\lambda I - A)$ equals $\{\mathbf{0}\}$ and so there are no eigenvectors with eigenvalue λ.

We know exactly when the nullspace is nontrivial. It is when the system $(\lambda I - A)X = \mathbf{0}$ has nontrivial solutions and so when the rank of $\lambda I - A$ is less than n. This occurs exactly when $\lambda I - A$ is singular, i.e. noninvertible, and so when $det(\lambda I - A) = 0$. So λ is an eigenvalue for A when $x = \lambda$ is a root of the *characteristic equation* $c_A(x) = 0$ where $c_A(x)$ is the *characteristic polynomial* given by

$$c_A(x) = det(xI - A). \tag{7.3}$$

Theorem 7.01: For an $n \times n$ matrix A, $c_A(x)$ is a polynomial of degree n with

$$c_A(x) \;=\; x^n - tr(A)x^{n-1} + \cdots + (-1)^n det(A). \qquad (7.4)$$

where the trace of A, $tr(A) = a_{11} + a_{22} + \cdots + a_{nn}$.

We will omit the proof.

The Fundamental Theorem of Algebra says that a polynomial of degree n has n roots (counting multiplicity, so that $x^2 - 2x + 1 = (x-1)^2$ has the root 1 repeated twice because there are two factors of $(x-1)$). However, we are only interested in real roots and there may be none of these.

Let $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, rotation in the plane through the angle θ. $det(xI - R_\theta) = x^2 - (2\cos\theta)x + 1$ with roots the complex conjugate pair $\cos\theta \pm \mathbf{i}\sin\theta$. It is clear that for θ not an integer multiple of π, the rotated vector $R_\theta(x,y)$ is not a real multiple of (x,y) when $(x,y) \neq \mathbf{0}$.

Diagonalization

What we look for is a basis of eigenvectors. When there is a basis of eigenvectors of T then we call T *diagonalizable*.

Theorem 7.03: Let $D = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for V and $T : V \to V$ be a linear map. For $\{\lambda_1, \ldots, \lambda_n\}$ a list of numbers in $\mathbb{R}$ $diag(\lambda_1, \ldots, \lambda_n)$ denotes the diagonal matrix with $diag(\lambda_1, \ldots, \lambda_n)_{ii} = \lambda_i$ and $diag(\lambda_1, \ldots, \lambda_n)_{ij} = 0$ when $i \neq j$. The list D consists of eigenvectors with λ_i the eigenvalue of $\mathbf{v}_i$ for all i if and only if

$$[L]_{DD} \; = \; diag(\lambda_1, \ldots, \lambda_n).$$

Proof: The i^{th} column of $[L]_{DD}$ is the D coordinate vector for $L(\mathbf{v}_i)$. This coordinate vector $[L(\mathbf{v}_i)]_D$ has a λ_i in the i^{th} place and 0's elsewhere if and only if $L(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$. As they are elements of a basis, no $\mathbf{v}_i = \mathbf{0}$.

We call the $n \times n$ matrix A diagonalizable when the linear map $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is diagonalizable.

Recall that with S the standard basis on $\mathbb{R}^n$, $[T_A]_{SS} = A$. If D is a basis of eigenvectors, then $I_{DS}[T_A]_{SS}I_{SD} = [T_A]_{DD})$ says, with $P = [I]_{SD}$,

$$P^{-1}AP \; = \; diag(\lambda_1, \ldots, \lambda_n) \tag{7.5}$$

Now we describe how to get the basis of eigenvectors when it exists and so how to compute the diagonalizing matrix P.

Theorem 7.04: Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. If D_i is a basis for the eigenspace $E_{\lambda_i}(A),\ i = 1, \ldots, k$, then the combined list $D = D_1 \cup \cdots \cup D_k$ is an li list in $\mathbb{R}^n$. The matrix A is diagonalizable if and only if D is a list of n vectors in total.

Proof: We will illustrate the proof by looking at a special case. Suppose that $k = 3$, $D_1 = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, $D_2 = \{\mathbf{v}_4, \mathbf{w}_5\}, D_3 = \{\mathbf{v}_6, \mathbf{v}_7\}$.

Given

$$(1) \quad c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + c_4\mathbf{v}_4 + c_5\mathbf{v}_5 + c_6\mathbf{v}_6 + c_7\mathbf{v}_7 = 0.$$

We must show all the c_i equal 0.
Multiply by the matrix $\lambda_3 I - A$. Because $A\mathbf{v}_i = \lambda_1\mathbf{v}_i$ for $i = 1, 2, 3$, $A\mathbf{v}_i = \lambda_2\mathbf{v}_i$ for $i = 4, 5$ and $A\mathbf{v}_i = \lambda_3\mathbf{v}_i$ for $i = 6, 7$, we get

$$(2) \quad c_1(\lambda_3 - \lambda_1)\mathbf{v}_1 + c_2(\lambda_3 - \lambda_1)\mathbf{v}_2 + c_3(\lambda_3 - \lambda_1)\mathbf{v}_3 + c_4(\lambda_3 - \lambda_2)\mathbf{v}_4 + c_5(\lambda_3 - \lambda_2)\mathbf{v}_5 = 0.$$

Multiply by $\lambda_2 I - A$ to get

$$\text{(3)} \quad c_1(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)\mathbf{v}_1 + c_2(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)\mathbf{v}_2 \\ + c_3(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)\mathbf{v}_3 = 0.$$

Because $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an li list and $(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1) \neq 0$, we have $c_1 = c_2 = c_3 = 0$.

Because $\{\mathbf{v}_4, \mathbf{v}_5\}$ is li, and $(\lambda_3 - \lambda_2) \neq 0$, equation (2) implies $c_4 = c_5 = 0$.

Finally, equation (1) implies $c_6 = c_7 = 0$ because $\{\mathbf{v}_6, \mathbf{v}_7\}$ is li.

Generalizing this argument we get that the list D is li. Furthermore every eigenvector is a linear combination of one of the D_i's since $\{\lambda_1, \ldots, \lambda_k\}$ lists all the eigenvalues. In particular, the span of D contains all of the eigenvectors.

If D contains fewer than n vectors then its span has dimension less than n and so is a proper subspace of $\mathbb{R}^n$. This means there is no basis of eigenvectors.

On the other hand, if the li list D contains $n = dim\mathbb{R}^n$ vectors, then it is a basis by Theorem 5.05.
$\square$

Corollary 7.05: If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable.

Proof: If $\mathbf{v}_i$ is an eigenvector for λ_i, then $D = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a list of eigenvectors which is li by Theorem 7.04.

Since it consists of n vectors, D is a basis.
$\square$

Our procedure to diagonalize A is as follows

- Compute the real roots of the characteristic polynomial $c_A(x) = det(xI - A)$. These are the eigenvalues of A.
- For each eigenvalue λ compute a basis of the solution space for the homogeneous system $(\lambda I - A)X = \mathbf{0}$.
- Put these bases together. If we have a list D of n vectors then it is the required basis of eigenvectors, and the transition matric $P = [I]_{SD}$, with columns the coordinates of the vectors of D, is the transition matrix so that $P^{-1}AP$ is diagonal. If D has fewer than n vectors, then A is not diagonalizable.

Example: Let $A = \begin{pmatrix} -1 & 2 & 2 \\ 0 & 1 & 2 \\ 0 & 8 & 7 \end{pmatrix}$ so that the determinant of $xI - A$ is

$$(x+1)det(\begin{pmatrix} x-1 & -2 \\ -8 & x-7 \end{pmatrix}) = (x+1)(x^2-8x-9) = (x+1)^2(x-9).$$

So the eigenvalues are -1 and 9.

For $\lambda = -1$, $-I - A$ is row equivalent to $\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ with solution $x_3 = r, x_2 = -r, x_1 = s$. So

$$D_{-1} = \{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \}$$

For $\lambda = 9$, $9I - A$ is row equivalent to $\begin{pmatrix} 1 & 0 & -1/4 \\ 0 & 1 & -1/4 \\ 0 & 0 & 0 \end{pmatrix}$ with solution $x_3 = r, x_2 = x_1 = r/4$. Using $r = 4$ we get $D_9 = \{ \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} \}$.

$$P = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 1 & 4 \end{pmatrix}$$

This is the transition matrix such that $P^{-1}AP \ = \ diag(-1, -1, 9)$.

Let us consider what happens when we use $A = \begin{pmatrix} -1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 8 & 7 \end{pmatrix}$ which has the same characteristic polynomial.

Projections

An important example is the following:

Theorem 7.06: For a linear map P on the n dimensional vector space V, the following are equivalent. When they hold, we call P a *projection*.

(i) $P \circ P = P$.

(ii) For all $\mathbf{v}$ in the image of P, $P(\mathbf{v}) = \mathbf{v}$.

(iii) $E_1(P) = Im(P)$.

(iv) $dim\ E_0(P) + dim\ E_1(P) = n$.

(v) P is diagonalizable with each eigenvalue either 0 or 1.

Proof: (i), (ii) and (iii) are all saying the same thing.

For any linear map P on V, Theorem 6.03 says $dim\ E_0(P) + dim\ Im(P) = n$. Clearly, $E_1(P) \subset Im(P)$. So $dim\ E_0(P) + dim\ E_1(P) = n$ if and only if $dim\ E_1(P) = dim\ Im(P)$ and so if and only if $E_1(P) = Im(P)$. Thus, (iii) $\Leftrightarrow$ (iv).

By Theorem 7.04 (iv) is equivalent to (v).
□

Notice that for a projection P

$$(I - P) \circ (I - P) = I - 2P + P \circ P = I - P.$$

Thus, $I - P$ is a projection which we call the *projection complementary to* P.

Systems of Differential Equations

Just as we can represent a system of linear equations using a single matrix equation, we can do the same for a system of linear differential equations:

$$\frac{dX}{dt} = AX. \tag{7.6}$$

Suppose that the coefficient matrix A is diagonalizable, so that $P^{-1}AP = diag(\lambda_1, \ldots, \lambda_n)$ with P the invertible matrix whose columns form a basis of eigenvectors for A.

We change variables, defining $Y = P^{-1}X$ and so $X = PY$. Because P^{-1} is a constant matrix,

$$\frac{dY}{dt} = P^{-1}\frac{dX}{dt} = P^{-1}AX = P^{-1}APY = diag(\lambda_1, \ldots, \lambda_n)Y. \tag{7.7}$$

That is, we have the system of equations:

$$\begin{aligned} \frac{dy_1}{dt} &= \lambda_1 y_1 \\ \frac{dy_2}{dt} &= \lambda_2 y_2 \\ &\cdot \\ \frac{dy_n}{dt} &= \lambda_n y_n \end{aligned} \tag{7.8}$$

The solution of $\frac{dy_i}{dt} = \lambda_i y_i$ is $y_i(0)e^{\lambda_i t}$. So the solution of the system can be written in matrix form as

$$\begin{aligned} Y &= diag(e^{\lambda_1 t}, \ldots, e^{\lambda_n t})Y(0), \\ X = PY &= Pdiag(e^{\lambda_1 t}, \ldots, e^{\lambda_n t})P^{-1}X(0). \end{aligned} \tag{7.9}$$

If $\{v_1, \ldots, v_n\}$ is the basis of eigenvectors for A with eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$, then the columns of P are the vectors $v_1, \ldots, v_n$. That is,

$$\begin{aligned} P &= [v_1 \ldots v_n] \quad \text{and so} \\ Pdiag(e^{\lambda_1 t}, \ldots, e^{\lambda_n t}) &= [e^{\lambda_1 t} v_1 \ldots e^{\lambda_n t} v_n]. \end{aligned} \tag{7.10}$$

The general solution is $X = c_1 e^{\lambda_1 t} v_1 + \ldots c_n e^{\lambda_n t} v_n$ with

$$Y(0) = \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ c_n \end{pmatrix}.$$

If we are given initial conditions $X(0)$ then we solve for the constants $c_1, \ldots, c_n$ using $Y(0) = P^{-1}X(0)$. So we solve:

$$P \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ c_n \end{pmatrix} = \begin{pmatrix} x_1(0) \\ \cdot \\ \cdot \\ x_n(0) \end{pmatrix} \tag{7.11}$$

and write $X = c_1 e^{\lambda_1 t} v_1 + \ldots c_n e^{\lambda_n t} v_n$.

Let us look at Exercise 3.5/ 1b, page 201.

Euclidean Spaces and Orthogonality

A Euclidean Space is a vector space V equipped with an inner product.

A function associating a real number $\mathbf{v} \cdot \mathbf{w}$ to every pair of vectors $\mathbf{v}, \mathbf{w} \in V$ is called an *inner product* when it satisfies the following properties

- Symmetry: $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$.
- Bilinearity: $\mathbf{v} \cdot (c\mathbf{w}_1 + \mathbf{w}_2) = c(\mathbf{v} \cdot \mathbf{w}_1) + \mathbf{v} \cdot \mathbf{w}_2$.
- Positivity: If $\mathbf{v} \neq \mathbf{0}$, then $\mathbf{v} \cdot \mathbf{v} > 0$.

From Bilinearity, we have $\mathbf{v} \cdot \mathbf{0} = 0$ for any vector $\mathbf{v}$ and so, in particular, $\mathbf{0} \cdot \mathbf{0} = 0$.

For $X, Y \in \mathbb{R}^n$,

$$X \cdot Y = X^T Y = \sum_{i=1}^{n} x_i y_i \tag{8.1}$$

is the usual dot product which motivates our definition.

For $A, B \in M_{mn}$ we define

$$A \cdot B = trace(A^T B) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij} \tag{8.2}$$

is an inner product.

For continuous functions $f, g : [0, 1] \to \mathbb{R}$ we can define the inner product

$$f \cdot g = \int_0^1 f(t) g(t)\, dt. \tag{8.3}$$

In a Euclidean space V we define the length of the vector $\mathbf{v}$ by

$$||\mathbf{v}|| \;=\; \sqrt{\mathbf{v}\cdot\mathbf{v}}. \tag{8.4}$$

Thus, any nonzero vector has a positive length.

We call $\mathbf{v}$ a unit vector when it has length 1. If $\mathbf{v}$ is any nonzero vector, then $(1/||\mathbf{v}|)\mathbf{v}$ is a unit vector.

We call two vectors $\mathbf{v}$ and $\mathbf{w}$ *perpendicular* or *orthogonal* when

$$\mathbf{v}\cdot\mathbf{w} \;=\; 0, \tag{8.5}$$

in which case we write $\mathbf{v} \perp \mathbf{w}$.

A list $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ of nonzero vectors is an *orthogonal list*, when $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for $i \neq j$ from 1 to k. It is an *orthonormal list*, when, in addition, $\mathbf{v}_i \cdot \mathbf{v}_i = 1$ for all i. Thus, an orthogonal list consists of mutually perpendicular nonzero vectors and it is orthonormal when all of the vectors are unit vectors.

Theorem 8.01: An orthogonal list of nonzero vectors is linearly independent.

Proof: Assume $c_1\mathbf{v}_1 + \dots c_k\mathbf{v}_k = \mathbf{0}$. Take the dot product with $\mathbf{v}_i$.

From bilinearity and orthogonality we get $c_i(\mathbf{v}_i \cdot \mathbf{v}_i) = \mathbf{v}_i \cdot \mathbf{0} = 0$. Because $\mathbf{v}_i$ is nonzero, $\mathbf{v}_i \cdot \mathbf{v}_i > 0$ and so $c_i = 0$.
$\square$

Theorem 8.02: For an n dimensional Euclidean space, there exists an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$.

Proof: Begin with an arbitrary basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. The Gram-Schmidt procedure constructs an orthogonal list $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ such that for $k = 1, \ldots n$,

$$span(\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}) \; = \; span(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}). \qquad (8.6)$$

To begin with, let $\mathbf{w}_1 \; = \; \mathbf{v}_1$.

Now assume that for some $k < n$ the orthogonal list $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ which satisfies (8.6)has been constructed.

Define

$$\mathbf{w}_{k+1} = \mathbf{v}_{k+1} - \sum_{i=1}^{k} \frac{(\mathbf{v}_{k+1} \cdot \mathbf{w}_i)}{\mathbf{w}_i \cdot \mathbf{w}_i} \mathbf{w}_i. \tag{8.7}$$

(To get rid of fractions, you can multiply $\mathbf{w}_{k+1}$ by any nonzero constant.) Check that $\mathbf{w}_{k+1} \cdot \mathbf{w}_i = 0$ for $i = 1, \ldots, k$. Because $\mathbf{v}_{k+1}$ is not in $span(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}) = span(\{\mathbf{w}_1, \ldots, \mathbf{w}_k\})$, it follows that $\mathbf{w}_{k+1} \neq \mathbf{0}$.

Thus $\{\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{w}_{k+1}\}$ is an orthogonal list.

Since each of the elements of the list is in $span(\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}\})$ it follows that

$span(\{\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{w}_{k+1}\}) \subset span(\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}\})$.

From Theorem 8.01, each of the subspaces has dimension $k + 1$ and so they are equal.

Continue the process to reach $k = n$.

We can then convert each $\mathbf{w}_i$ to the unit vector $\mathbf{u}_i = (1/||\mathbf{w}_i||)\mathbf{w}_i$.

Clearly, for $k = 1, \ldots, n$

$$span(\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}) = span(\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}) = span(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}).$$

Thus, $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis.
$\square$

Let us look at Exercise 8.1/ 1c, page 416.

There are two pieces of geometry which are useful in a Euclidean space.

Theorem 8.03: Let $\mathbf{v}, \mathbf{w}$ be vectors in a Euclidean space.

(a) (Pythagorean Theorem) If $\mathbf{v} \perp \mathbf{w}$, then $||\mathbf{v} - \mathbf{w}||^2 \ = \ ||\mathbf{v}||^2 + ||\mathbf{w}||^2$.

(b) (Cauchy-Schwarz Inequality) $|\mathbf{v} \cdot \mathbf{w}| \leq ||\mathbf{v}|| \ ||\mathbf{w}||$.

Proof: (a) $||\mathbf{v} - \mathbf{w}||^2 \ = \ (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w})$. and from Bilinearity and Symmetry this equals

$$\mathbf{v} \cdot \mathbf{v} \ - \ 2\mathbf{v} \cdot \mathbf{w} \ + \ \mathbf{w} \cdot \mathbf{w} \ = \ ||\mathbf{v}||^2 \ + \ 0 \ + \ ||\mathbf{w}||^2.$$

(b) For all real values of x, $(x\mathbf{v} + \mathbf{w}) \cdot (x\mathbf{v} + \mathbf{w}) \geq 0$. Expand this to get the quadratic in x:

$$(\mathbf{v} \cdot \mathbf{v})x^2 \ + \ 2(\mathbf{v} \cdot \mathbf{w})x \ + \ (\mathbf{w} \cdot \mathbf{w}) \ \geq \ 0.$$

This means that the quadratic cannot have two distinct real roots. Applying the Quadratic Formula to $Ax^2 + Bx + C$, the nonexistence of distinct real roots means that the discriminant $B^2 - 4AC \leq 0$. So

$$4(\mathbf{v} \cdot \mathbf{w})^2 - 4(\mathbf{v} \cdot \mathbf{v})(\mathbf{w} \cdot \mathbf{w}) \leq 0.$$

□

So we can define the angle between the vectors **v** and **w** to be

$$\theta \ = \arccos[(\mathbf{v} \cdot \mathbf{w})/(||\mathbf{v}|| \ ||\mathbf{w}||)].$$

For a subspace U of a Euclidean space V, we define

$$U^{\perp} = \{\mathbf{v} \in V : \mathbf{v} \cdot \mathbf{w} = 0 \quad \text{for all} \quad \mathbf{w} \in U\}. \tag{8.8}$$

You should check following facts:

- $U^{\perp}$ is a subspace of V.
- $\mathbf{v} \in U \cap U^{\perp}$ implies $\mathbf{v} \cdot \mathbf{v} = 0$ and so $\mathbf{v} = \mathbf{0}$.
- If $span(\{\mathbf{v}_1, \dots, \mathbf{v}_k\}) = U$, then $\mathbf{v} \in U^{\perp}$ if and only if $\mathbf{v} \cdot \mathbf{w}_i = 0$ for $i = 1, \dots, k$.

Choose $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ an orthonormal basis for U. Define the linear map P_U on V by

$$P_U(\mathbf{v}) = \sum_{i=1}^{k} (\mathbf{v} \cdot \mathbf{u}_i)\mathbf{u}_i. \tag{8.9}$$

Linearity of P_U follows from Bilinearity of the inner product.

If U is a subspace of $\mathbb{R}^n$, then for $\mathbf{v} = X \in \mathbb{R}^n$, $X \cdot \mathbf{u}_i = \mathbf{u}_i^T X$ and so $(X \cdot \mathbf{u}_i)\mathbf{u}_i = (\mathbf{u}_i \mathbf{u}_i^T)X$.

Thus, $P_U(X) = AX$ with the matrix given by

$$A = \sum_{i=1}^{k} \mathbf{u}_i \mathbf{u}_i^T. \tag{8.10}$$

If U is a subspace of $\mathbb{R}^n$, then for $\mathbf{v} = X \in \mathbb{R}^n$, $X \cdot \mathbf{u}_i = \mathbf{u}_i^T X$ and so $(X \cdot \mathbf{u}_i)\mathbf{u}_i = (\mathbf{u}_i \mathbf{u}_i^T)X$.

Thus, $P_U(X) = AX$ with the matrix given by

$$A = \sum_{i=1}^{k} \mathbf{u}_i \mathbf{u}_i^T. \tag{8.10}$$

Notice that $\mathbf{u}_i$ is an $n \times 1$ matrix and so $\mathbf{u}_i^T$ is $1 \times n$ and $\mathbf{u}_i \mathbf{u}_i^T$ is $n \times n$.

Theorem 8.04: For U a subspace of the Euclidean space V and $P : V \to V$ a map on V, the following are equivalent.

(i) With P_U defined by (8.9), $P = P_U$.

(ii) P is a projection with image U and kernel $U^\perp$.

(iii) P is a linear map such that for all $\mathbf{v} \in V$, $P(v) \in U$ and $v - P(v) \in U^\perp$.

In particular, the *orthogonal projection* P_U does not depend on the choice of orthonormal basis.

Proof: Observe first that $P_U(\mathbf{v}) \in U$ for all $\mathbf{v}$ and so $Im(P_U) \subset U$.

Because $span(\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}) = U$ and the list is li, $P_U(\mathbf{v}) = \mathbf{0}$ if and only if $\mathbf{v} \in U^\perp$. That is, the kernel of P_U is $U^\perp$.

If $\mathbf{v} \in U$, then because $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis for U we can write $\mathbf{v} = \sum_{i=1}^{k} c_i \mathbf{u}_i$. Because the basis is orthonormal we can take the inner product with $\mathbf{u}_i$ to get $\mathbf{v} \cdot \mathbf{u}_i = c_i$. This means that $\mathbf{v} = P_U(\mathbf{v})$. Hence, $Im(P_U) = U$ and $P_U \circ P_U = P_U$. That is, P_U is a projection.

Thus, (i) implies (ii).

(ii) implies (iii): If P is a projection with image U, then $P(\mathbf{v}) \in U$ for all $\mathbf{v}$. For any projection $P(\mathbf{v} - P(\mathbf{v})) = P(\mathbf{v}) - P(\mathbf{v}) = \mathbf{0}$ and so $\mathbf{v} - P(\mathbf{v})$ is in the kernel of P. So if the kernel is $U^\perp$, then $\mathbf{v} - P(\mathbf{v}) \in U^\perp$.

(iii) implies (i): Because P_U has kernel $U^\perp$ ((i) implies (ii)) it follows that $\mathbf{0} = P_U(\mathbf{v} - P(\mathbf{v}))$. That is, $P_U(\mathbf{v}) = P_U(P(\mathbf{v}))$. Because P_U is the identity on U and $P(\mathbf{v}) \in U$, we have $P_U(\mathbf{v}) = P(\mathbf{v})$ for all $\mathbf{v} \in V$.
$\square$

For any vector $\mathbf{v} \in V$, the projection $P_U(\mathbf{v})$ is best approximation of $\mathbf{v}$ by a vector in U. That is,

$$\mathbf{w} \in U, \text{ and } \mathbf{w} \neq P_U(\mathbf{v}) \implies ||\mathbf{v} - \mathbf{w}|| > ||\mathbf{v} - P_U(\mathbf{v})||. \tag{8.11}$$

Proof: $\mathbf{w} - P_U(\mathbf{v}) \in U$ and so is perpendicular to $\mathbf{v} - P_U(\mathbf{v})$.

Since $\mathbf{v} - \mathbf{w} = (\mathbf{v} - P_U(\mathbf{v})) - (\mathbf{w} - P_U(\mathbf{v}))$, the Pythagorean Theorem implies

$$||\mathbf{v} - \mathbf{w}||^2 = ||\mathbf{v} - P_U(\mathbf{v})||^2 + ||\mathbf{w} - P_U(\mathbf{v})||^2.$$

□

Best Approximation of a Solution

Now consider the system $AX = B$ with A an $n \times m$ matrix. The system is inconsistent when it has no solution which means that B is not a linear combination of the columns of A, that is, B is not in the column space $Col(A)$.

We can ask what element of the column space is the best approximation of B. We know the answer already from (8.11) which implies that $B_1 = P_{Col(A)}(B)$ is the unique element of $Col(A)$ which is closest to B.

$B_1 = P_{Col(A)}(B)$ is characterized by the conditions $B_1 \in Col(A)$ and $B - B_1 \in Col(A)^{\perp}$.

The first condition says that there exists a solution Z of $AZ = B_1$. The second says that $B - B_1 = B - AZ$ is perpendicular to every column of A, or, equivalently $A^T(B - AZ) \ = \ \mathbf{0}$. Therefore to obtain the best approximation $B_1 = AZ$ we solve the *normal equation*

$$(A^T A)Z \ = \ A^T B. \tag{8.12}$$

The solution Z of (8.12) always exists but is not unique when the columns are not li. However $B_1 = AZ$ is unique.

To see this directly, suppose that $(A^T A)Z_1 = (A^T A)Z$ and so $(A^T A)(Z - Z_1) = \mathbf{0}$. Our next result shows that this means that $A(Z - Z_1) = \mathbf{0}$ and so $AZ = AZ_1$.

Theorem 8.05: The $m \times n$ matrix A and the $n \times n$ matrix $A^T A$ have the same null space and the same rank.

Proof: If $AX = \mathbf{0}$, then $A^T AX = \mathbf{0}$. If $A^T AX = \mathbf{0}$, then

$$0 \;=\; X^T A^T AX \;=\; (AX) \cdot (AX) \;=\; ||AX||^2.$$

Therefore, $AX = \mathbf{0}$.

For both A and $A^T A$ Theorem 5.06 says that the rank r equals n minus the dimension of the null space.
$\square$

Least Squares Approximation

For an $n \times 1$ column vector X and a function $f : \mathbb{R} \to \mathbb{R}$ we let $f(X)$ be the $n \times 1$ column vector $\begin{pmatrix} f(x_1) \\ \cdot \\ f(x_n) \end{pmatrix}$.

Now suppose we are given n data pairs $(x_1, y_1), \ldots, (x_n, y_n)$ which we can regard as a pair X, Y of $n \times 1$ column vectors. We want to choose coefficients $z_1, \ldots, z_k$ so that with $f(x) = z_1 + z_2 x + \ldots z_k x^{k-1}$, $f(X)$ is the best approximation to Y. That is, we want to choose the coefficients so that $||Y - f(X)||^2$ is as small as possible.

We use the $n \times k$ matrix

$$A = [\mathbf{1} \;\; X \;\; X^2 \ldots X^{k-1}] = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdot & \cdot & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \cdot & \cdot & x_2^{k-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 & \cdot & \cdot & x_n^{k-1} \end{pmatrix}$$

and we solve the normal equation $A^T AZ = A^T Y$.

Let us look at Exercise 5.6/ 1a page 318.

Orthogonal Matrices

Theorem 8.06: For an $n \times n$ matrix U the following conditions are equivalent. When they hold we call U an *orthogonal matrix.*

(i) U is invertible and $U^{-1} = U^T$.

(ii) $U^T U = I_n$.

(iii) The columns of U form an orthonormal list and so provide an orthonormal basis in $\mathbb{R}^n$.

(iv) The rows of U form an orthonormal list and so provide an orthonormal basis in $\mathbb{R}^n$.

If U is orthogonal, then U^T is orthogonal.

Proof: By Theorem 2.01, it suffices to check cancellation on one side and so (i) is equivalent to (ii). Multiplying out we see that (ii) is equivalent to (iii).

If $U^{-1} = U^T$, then $(U^T)^{-1} = (U^{-1})^T = (U^T)^T$ and so U^T is orthogonal. Condition (iii) for U^T is the same as condition (iv) for U.

$\square$

Symmetric Matrices

An $n \times n$ matrix A is called a *symmetric matrix* when $A^T = A$.

It will be our final task to show that any symmetric map has an orthonormal basis of eigenvectors and to apply this result. In Theorem 7.04 and Corollary 7.05 we saw that a list of eigenvectors associated with distinct eigenvalues is necessarily li. For a symmetric matrix we have a stronger result.

Theorem 8.10: If A is a symmetric $n \times n$ matrix with $AX_1 = \lambda_1 X_1$ and $AX_2 = \lambda_2 X_2$ $\lambda_1 \neq \lambda_2$, then the dot product $X_1^T X_2$ equals zero.

Proof: From symmetry we have

$$\lambda_1 X_1^T X_2 \; = \; (AX_1)^T X_2 \; = \; X_1^T A^T X_2 \; = \; X_1^T A X_2 \; = \; \lambda_2 X_1^T X_2.$$

Since $\lambda_1 \neq \lambda_2$, it follows that $X_1^T X_2 = 0$.
$\square$

When we look at rotations in the plane we saw that it is possible to have a linear map with no eigenvectors at all. This occurs when the characteristic polynomial $c_A(x) = det(xI - A)$ of the associated matrix has no real roots.

However, for a symmetric matrix we have

Theorem 8.11: If A is a symmetric matrix, then the roots of the characteristic polynomial $c_A(x)$ are all real numbers. In particular, any complex eigenvalue is in fact real.

We will omit the proof of this. It is given on page 305 of the book and requires a digression using matrices with complex entries.

Theorem 8.12:(*Principal Axis Theorem*) If A is an $n \times n$ matrix, then the following are equivalent.

(i) A has an orthonormal basis of eigenvectors.

(ii) A is orthogonally diagonalizable. That is there exists an orthogonal matrix P and a diagonal matrix D such that $A = P^{-1}DP = P^{T}DP$.

(iii) A is symmetric.

Proof: (i) $\Leftrightarrow$ (ii) and (ii)$\Rightarrow$ (iii) are clear.

For (iii) $\Rightarrow$ (ii) we sketch the argument from page 420 of the book.
Because A has a real eigenvalue, it has a unit eigenvector $\mathbf{x}_1$ with eigenvalue λ_1. We can extend to get a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ which we can take to be orthonormal by using the Gram-Schmidt process.

With $P_1 = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ we have
$P_1^T A P_1 = P_1^{-1} A P_1 = \begin{pmatrix} \lambda_1 & B \\ 0 & A_1 \end{pmatrix}$. This is symmetric and so
$B = 0$ and A_1 is symmetric.

Using induction on n we may assume that A_1 is orthogonally diagonalizable and so there exists an orthgonal $(n-1) \times (n-1)$ matrix Q such that $Q^{-1} A_1 Q$ is diagonal and so with $P_2 = \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}$ we get the orthogonal matrix $P = P_1 P_2$ so that $P^{-1} A P$ is diagonal.
□

Let us look at Exercises 8.2/ 5be, page 425.

Singular Values Decomposition

Throughout we fix an $m \times n$ matrix A. We will study it by looking at the $n \times n$ symmetric matrix $A^T A$.

Theorem 8.13: The null spaces of A and $A^T A$ are the same.
That is, $AX = \mathbf{0}$ if and only if $A^T AX = \mathbf{0}$.
The rank of A equals the rank of $A^T A$.

Proof: If $AX = \mathbf{0}$ then $A^T AX = \mathbf{0}$. Conversely, if $A^T AX = \mathbf{0}$, then $0 = X^T A^T AX = (AX) \cdot (AX) = ||AX||^2$. Since the length $||AX|| = 0$, the vector AX equals $\mathbf{0}$.

For an $m \times n$ or $n \times n$ matrix, the rank equals n minus the dimension of the null space.
$\square$

Because the $n \times n$ matrix $A^T A$ is symmetric, it has an orthonormal basis B_n of eigenvectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ with eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$. For $i, j = 1, \ldots, n$

$$(A\mathbf{u}_i) \cdot (A\mathbf{u}_j) = \mathbf{u}_i^T A^T A \mathbf{u}_j = \lambda_j \mathbf{u}_i^T \mathbf{u}_j = \lambda_j (\mathbf{u}_i \cdot \mathbf{u}_j). \qquad (8.13)$$

Because $B_n = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is orthonormal, we have

$$\begin{aligned} (A\mathbf{u}_i) \cdot (A\mathbf{u}_j) &= 0 \quad \text{if} \quad i \neq j, \\ ||A\mathbf{u}_i||^2 &= \lambda_i ||\mathbf{u}_i||^2 = \lambda_i. \end{aligned} \qquad (8.14)$$

Therefore all the eigenvalues λ_i are non-negative. We define the *singular values*

$$\sigma_i = \sqrt{\lambda_i} = ||A\mathbf{u}_i||. \qquad (8.15)$$

We can choose the order of $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ so that $\lambda_1 \geq \lambda_2 \ldots \lambda_r > 0, \lambda_{r+1} = \cdots = \lambda_n = 0$.

Because $A^T A$ is similar to the diagonal matrix $diag(\lambda_1, \ldots, \lambda_n)$ (and so has the same rank), it follows that the rank of A equals the rank of $A^T A$ equals the number r of positive eigenvalues. For $i = 1, \ldots, r$ define

$$\mathbf{v}_i = (1/\sigma_i) A\mathbf{u}_i. \tag{8.16}$$

From (8.14) we see that $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthonormal list of vectors in $\mathbb{R}^m$. pause Furthermore,

$$\begin{aligned} A\mathbf{u}_i &= \sigma_i \mathbf{v}_i \quad \text{for} \quad i = 1, \ldots, r, \\ A\mathbf{u}_i &= \mathbf{0} \quad \text{for} \quad i = r+1, \ldots, n \end{aligned} \tag{8.17}$$

because $\sigma_i = 0$ for $i > r$.

Extend the list $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ to obtain $D_m = \{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_m\}$ an orthonormal basis for $\mathbb{R}^m$. For the linear map $T_A : \mathbb{R}^n \to \mathbb{R}^m$, the matrix $[T_A]_{D_m B_n}$ is obtained by applying T_A to the vectors of the basis B_n and then computing the D_m coordinates to obtain the columns.

$$[A\mathbf{u}_1 \ldots A\mathbf{u}_r \; A\mathbf{u}_{r+1} \ldots A\mathbf{u}_n] \;=\; [\sigma_1 \mathbf{v}_1 \ldots \sigma_r \mathbf{v}_r \; \mathbf{0} \ldots \mathbf{0}]. \quad (8.18)$$

So in Block form

$$[T_A]_{D_m B_n} \;=\; \Sigma \;=\; \begin{pmatrix} diag & \mathbf{0}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix} \qquad (8.19)$$

with *diag* equal to the $r \times r$ diagonal matrix $diag(\sigma_1, \ldots, \sigma_r)$.

With respect to the standard bases S_n and S_m on $\mathbb{R}^n$ and $\mathbb{R}^m$, the matrix $[T_A]_{S_m S_n} = A$.

Let $Q = [I]_{S_n B_n}$ whose columns are the standard coordinates of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$.

Let $P = [I]_{S_m D_m}$ whose columns are the standard coordinates of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

These are orthogonal matrices and

$$A = [T_A]_{S_m S_n} \; = \; [I]_{S_m D_m}[T_A]_{D_m B_n}[I]_{B_n S_n} \; = \; P\Sigma Q^{-1} \; = \; P\Sigma Q^T. \tag{8.20}$$

This is called the *Singular Values Decomposition* (the SVD) of A.

Let us find the SVD of

$$A = \begin{pmatrix} 1 & -1 \\ 0 & 0 \\ 1 & -1 \end{pmatrix}.$$