Theoretical Foundations of Linear Algebra

Supplemental Materials

December 30, 2023

Transformations, Vector Spaces, and Subspaces

Prove $C(AB)$ is contained in $C(A)$
$C(AB)$ is defined as the set of b such that $ABx = b$ is consistent for some x. Let $y = Bx$, then $Ay = b$ is consistent, the definition of $C(A)$. As all b in $C(AB)$ satisfy the condition for $C(A)$, $C(AB)$ is contained in $C(A)$.

Prove $V + V^\perp = R^n$
Let $\{v_1, \cdots, v_k\}$ be an orthogonal basis for V, so V has dimension k. Every subspace has an orthogonal basis, and any basis can be converted to an orthogonal basis.
Let $\{u_1, \cdots, u_{n-k}\}$ be an orthogonal basis for $V^\perp$, so $V^\perp$ has dimension $n - k$. We can use the fact that $V + V^\perp$ is undoubtedly a subset of R^n, and if it has the same dimension, the sets must be equal. If n vectors are linearly independent in $V + V^\perp$, we are done.

That is, we need to show $c_1 v_1 + c_2 v_2 + \cdots c_k v_k + d_1 u_1 + d_2 u_2 + \cdots d_{u-kn-k} = 0$ implies all constants are 0.
If we dot with v_i, we get $c_i ||v_i||^2 = 0$, and as $v_i = 0$ would contradict the idea that it was part of a basis, we see $c_i = 0$. Doing this for all c and d, we get that all constants are 0.
Alternatively, we could not use orthogonal bases and rearrange this and get
$c_1 v_1 + \cdots = -(d_1 v_1 \cdots)$. As these two spaces only share $\{0\}$, each side must equal 0, only achievable if all constants are 0 because each side is linearly independent.
As we know that $V + V^\perp$ is the same dimension as R^n, and most certainly a subset, we can utilize an existing proposition to say they are equal.

Prove Nonsingular Matrices have the property that the Column Space equals the Row Space. Any nonsingular matrix will guarantee this because $C(A)$ and $R(A)$ both have dimension n, and thus are equal to R^n because they are both subsets of R^n, and Proposition 4.3 says then they must be equal to R^n. And obviously, $R^n = R^n$.

Prove Symmetric Matrices have the property that C(A) equals $R(A)$.
We know in general $R(A) = C(A^T)$. If A is symmetric, then $A^T = A$. That yields $R(A) = C(A)$.

Prove that the Kernel and Image are subspaces.

Kernel:

Zero: $\ker(T)$ is defined as $\{v \text{ s.t.} T(v) = 0\}$. A linear transformation maps 0 to 0, thus 0 is in Kernel.

Addition: If u, v, are in the kernel, then $T(u) = 0$ and $T(v) = 0$. We want to know if $T(u + v)$ gets mapped to 0. By linearity, we see $T(u + v) = T(u) + T(v) = 0 + 0 = 0$.

Scaling: If v is in the kernel, then $T(v) = 0$. We want to know if $T(cv)$ gets mapped to 0. By linearity, $T(cv) = cT(v) = c_0 = 0$.

Image:

Zero: $im(T)$ is defined as $\{w \text{ s.t. } T(v) = w\}$. We know $w = 0$ is possible because $T(0) = 0$. As 0 is in the output, it is in the image.

Addition: If u,v are in image then $T(a) = u$ and $T(b) = v$. We want to know if $u + v$ is in the output.

We can get this by taking $T(a + b)$ because it will yield $T(a) + T(b) = u + v$. Scaling: If v is in the image then $T(a) = v$. We want to know if cv is in the output. By taking $T(ca)$, by linearity, we get $cT(a)$ or cv.

If $Kernel = \{0\}$, transformation injective.

If $Image = \{\text{all of output space}\}$, transformation surjective.

If both, isomorphism.

Prove $U + W$ is a subspace if U and W are. $U + W$ is $\{u + w$ for all pairs of u and $w\}$.

Zero: 0 is in both U and W as they are subspaces. 0+0 yields 0.

Addition: We want to show for a1 and a_2 in $U + W$ that $a_1 + a_2$ is in $U + W$. a1 is $u_1 + w_1$ and a_2 is $u_2 + w_2$ by definition of how the new subspace was created. So this new element would be $(u_1 + w_1) + (u_2 + w_2)$. Rearranging, $(u_1 + u_2) + (w_1 + w_2)$. We know each of these are in U and W respectively as U and W are subspaces, so as this sum is a sum of an element in U and W, it is in $U + W$ by definition.

Scaling: We want to show that for a1 in $U + W$ that $c(a_1)$ is in $U + W$. Decomposing this, we get $c(u_1 + w_1) -> cu_1 + cw_1$. Apply that U and W under scaling, so $c(u_3)$ and $c(w_3)$ are in U and W respectively.

Prove that U intersection W is a subspace if U and W are.

Zero: 0 is in both by definition of subspaces so 0 is in the intersection subspace.

Addition: If x and y are in the intersection, then x and y are in both subspaces, and as subspaces are closed under addition, then $x + y$ are in both subspaces, and thus in the intersection.

Scaling: If x is in the intersection, then x is in both subspaces, and subspaces are closed under scaling, so cx is in both subspaces, so cv is in the intersection.

Prove $\dim(U + W) = dim(U) + dim(W) - dim(U \text{intersection} W)$.

Let $\{v_1, \cdots, v_k\}$ be a basis for the intersection. Let $\{v_1, \cdots, v_k, u_{k+1}, \cdots, u_m\}$ be a basis for $U(\dim m)$ and $\{v_1, \cdots, v_k, w_{k+1}, \cdots w_n\}$ be a basis for W

Claim: $\{v_1, \cdots, v_k, u_{k+1}, \cdots u_m, w_{k+1}, \cdots, w_n\}$ is a basis for $U + W$. By construction, it spans $U + V$.

These additional vectors that are part of the basis for U or W but not U intersection W are independent from each other because otherwise they would have been in the intersection. $\dim(U + W)$ is $m + n - k$ by counting.

On the right side, we get $m + n - k$-, confirming this relationship.

Prove V intersection V Perp is $\{0\}$.

The intersection is vectors in both V and $V^\perp$. To be in both V and $V^\perp$, you would have to be orthogonal to yourself (along with other criteria). That is, we need $x \times x = 0$, or $||x||^2 = 0$, only possible if $x = 0$.

Prove $C(A)$ is orthogonal to $N(A^T)$.

Let c be in $C(A)$ and l in $N(A^T)$. By definition, $c = Ax$ for some x. We know $A^T \times l = 0$ by definition of l being in the left null space. We want to show $c \times l = 0$, or $Ax \times l = 0$. We know $(Ax) \times l = (Ax)^T \times l$, so we get $x^T A^T \times l$, and $(A^T l)$ is 0, so $x^T 0 = 0$.

Corollary: $R(A)$ is orthogonal to $N(A)$.

By definition, $C(A) = R(A^T)$. Thus we know $R(A^T)$ is orthogonal to $N(A^T)$. If we set $B = A^T$, we would see $R(B)$ is orthogonal to $N(B)$.

Prove $R(A)$ is orthogonal to $N(A)$.

Let r be in $R(A)$, so $r = A^T y$, and n satisfy An $= 0$. We want to show $r * n = 0$, or $A^T y * n = 0$. Using properties of the transpose, we get $y^T (A * n)$ or $y^T \times 0 = 0$.

Prove that the set of $[a\ b\ a + b]$ is a subspace.

Zero: $a = b = 0$ yields the 0 vector Addition: $x_1 + x_2 = [a_1\ b_1\ a_1 + b_1] + [a_2\ b_2 a_2 + b_2] = [a_1 + a_2\ b_1 + b_2\ a_1 + b_1 + a_2 + b_2]$. We want to show this is of the form $[c\ d\ c + d]$. Taking $a_1 + a_2 = c$ and $b_1 + b_2 = d$, by addition properties, we get that $a_1 + b_1 + a_2 + b_2$ is indeed $c + d$ or $a_1 + a_2 + b_1 + b_2$. Scaling: $c[a\ b\ a + b] = [ca\ cb\ c(a + b)] = [ca\ cb\ ca + cb]$, which is indeed of the form $[e\ f\ e + f]$.

Transformation V to W:

If it is injective, ker $= \{0\}$

If it is surjective, $Image = \{$All of $W\}$

A transformation is an isomorphism (bijective) if and only if both ker $= \{0\}$ and $Image = \{$All of $W\}$

Injective: different inputs lead to different outputs. T(a) $\neq$ T(b) unless a=b

Surjective $=$ every element in W can be made from some element in V.

$Image =$ Column Space. $Kernel =$ Null Space.

Suppose A and B are $n \times n$ matrices. Show that if either A or B is nonsingular, then AB and BA are similar.

Similarity in general: $B = P^{-1} A P$

Here, $(AB) = P^{-1}(BA)P$. We need to show that there exists an invertible P such that

3

this is true. Let's consider the following cases:

1. If B is nonsingular, we can choose $P = B$. Since B is invertible, $P^{-1} = B^{-1}$, and we have:
$$(AB) = B^{-1}(BA)B = A^{B^{-1}}BAB^{-1}BA = ABA.$$

Thus, in this case, AB and BA are similar.

2. If A is nonsingular, we can choose $P = A$. Again, since A is invertible, $P^{-1} = A^{-1}$, and we have:
$$(BA) = A^{-1}ABA = B^{A^{-1}}ABA^{-1}AB = BAB.$$

In this case as well, AB and BA are similar.

Therefore, we have shown that if either A or B is nonsingular, then AB and BA are similar.

Independence, Bases, and Dimension

Proposition 4.1 - V is a subspace of R_n. $\{v_1, \cdots, v_k\}$ is a basis for V. If $w_1, \cdots, w_l$ are in V and $l > k$, then $\{w_1, \cdots, w_l\}$ must be linearly dependent.

Proof: We know each w_i can be made by a unique linear combination of $v_1, \cdots, v_k$. We can express this as matrices.
$W = V * A$, where W has column vectors of w_i, V has column vectors of V_i, and A is a coefficient matrix.

If $l > k$, then A cannot have a pivot in every column. There exist free variables, meaning at least one column can be expressed as a linear combination of the other columns. This means there is a non-trivial linear combination of the columns of A, which will result in a non-trivial linear combination of the $w_1, \cdots, w_l$ (Matrix W) that gives the 0 vector. Thus, the set is linearly dependent.

Theorem 4.2 - V is a subspace of R_n. Show that two bases must have the same number of elements.
Proof: Use Proposition 4.1. Let the two bases be $v_1, \cdots, v_x$ and $w1, \cdots, wy$. Assume for contradiction x does not equal y. There are two cases to check. Let $x > y$. Proposition 4.1 would say that if $\{w_1, \cdots, v_y\}$ is a basis (it is by definition) $\{v_1, \cdots, v_x\}$ would have to be linearly dependent.
This contradicts the idea that $\{v_1, \cdots, vx\}$ is a basis. Likewise, if $y > x$, that would make $\{w_1, \cdots, w_y\}$ linearly dependent. Thus, it is proven by extending Proposition 4.1 since we have the luxury of knowing both these are bases.

Proposition 4.3 - V and W are subspaces of R_n and W is contained in V. If dimV=dimW, show that $V = W$. Proof: Let $\{w_1, \cdots, w_k\}$ be a basis for W. We know all w_i are in V and are linearly independent in V. If $\{w_1, \cdots, w_k\}$ does not span V then there exist vectors that can be added to make $\{w_1, \cdots, w_k\}$ span V. However, that would make

4

dimV greater than $\dim W$, so by contradiction, we know that $\{w_1, \cdots, w_k\}$ spans V. They are still linearly independent, so $V = W$.

Proposition 4.4 - V is a k-dimensional subspace. Prove that any k vectors that span V must be linearly independent, and any k linearly independent vectors in V must span V.

Proof 1 (Prop 4.4): Any k vectors that span V are linearly independent.
Suppose you have a set of k linearly dependent vectors that span V. If you took out n dependent vectors, you would have $k - n$ vectors that span V, but by the definition of dimension, the number of basis vectors must be k. Thus, the set must be linearly independent.
Proof 2 (Prop 4.4): Any k linearly independent vectors must span V.
Suppose the k linearly independent vectors did not span V. That would mean there exist some (n) linearly independent vectors that could be added to the set to make it span V. That would mean the dimension of V would be $k + n$, but it is k. Thus, k linearly independent vectors must span V.

Theorem 4.6 - A is an mxn matrix with rank r. Show these are true:

(i) $\dim R(A) = \dim C(A) = r$

(ii) $\dim N(A) = n - r$

(iii) $\dim N(A^T) = m - r$

Proof (i): There are r pivots, r non-zero rows, so there are r vectors in $R(A)$, so the dimension of $R(A)$ is r. The rank of a matrix is the same as its transpose, so this holds for $C(A)$ too as $C(A) = R(A^T)$.
Proof (ii): There are either pivots or non-pivots (free variables) in columns. The dimension of $N(A)$ is equal to the number of non-pivots, which is $n - r$.
Proof (iii): The number of free variables in A^T are the number of zero rows in A, which is $m - r$.
The Nullity-Rank Theorem is a direct consequence of this and states $nullity(A) + rank(A) = n$. Likewise, $left\_nullity(A) + rank(A) = m$.

Proposition 3.2. Suppose $v_1, \ldots, v_k \in \mathbb{R}^n$ form a linearly independent set, and suppose $v \in \mathbb{R}^n$. Then $\{v_1, \ldots, v_k, v\}$ is linearly independent if and only if $v \notin \text{Span}(v_1, \ldots, v_k)$.

Direction 1: If they were linearly independent, then $c_1 v_1 + \ldots + c_k v_k + c_{k+1} v = 0$. If v were in the span, then v could be expressed as a nontrivial combination, allowing you to substitute that combination in the original expression, contradicting linear independence. Thus, it cannot be in the span.

Direction 2: If v is not in the span, then v cannot be made as a linear combination of $v_1, \ldots, v_k$. In $c_1 v_1 + \ldots + c_k v_k + c_{k+1} v = 0$, if c_{k+1} was nonzero, you could rearrange and isolate v_{k+1}, creating a linear combination, but that would contradict the fact that

v cannot be made as one (because it is not in the span). Thus, c_{k+1} is 0 and adding the vector preserves the set's linear independence.

Let A be an $n \times n$ matrix. Prove that if A is nonsingular and $\{v_1, \ldots, v_k\}$ is linearly independent, then $\{Av_1, Av_2, \ldots, Av_k\}$ is likewise linearly independent.

Suppose that $A(c_1v_1 + \ldots + c_kv_k) = 0$ is linearly dependent. Then there exist non-trivial solutions. Factoring out an A and multiplying by A^{-1}, we get $c_1v_1 + \ldots + c_kv_k = 0$, which contradicts the fact that the set is linearly dependent, as $v_1, \ldots, v_k$ are linearly independent. The solution set is preserved through these operations, which is why we can deduce $Av_1, \ldots, Av_k$ is linearly independent. Alternatively, we could start with $c_1v_1 + \ldots + c_kv_k = 0$ implying all constants are 0. If we multiply by A on both sides, we get $c_1Av_1 + \ldots + c_kAv_k = 0$, and the solution set is preserved, so all constants are still 0, so this transformed set must likewise be linearly independent.

Let $V \subset \mathbb{R}^n$ be a subspace, let $\{v_1, \ldots, v_k\}$ be a basis for V, and let $w_1, \ldots, w_m \in V$ be vectors such that $\text{Span}(w_1, \ldots, w) = V$. Prove that $m \geq k$.

V is dimension k. If m was less than k, then V would be a dimension less than k, which isn't possible as then $v_1, \ldots, v_k$ would have to be dependent, contradicting the idea that they are a basis.

Vectors and Solutions to Systems of Equations

Prove $Span(v + cw, w) = Span(v, w)$
Direction 1: A vector in $Span(v + cw, w)$ can be written as a $(v + cw) + bw$, expanding to $av + acw + bw$. $av + (ac + b)w$ is a linear combination of v and w, so every vector in $Span(v + cw, w)$ is in $Span(v, w)$
Direction 2: A vector in $Span(v, w)$ can be written as $dv + ew$. If we take $d = a$ and $e = ac + b$, we get a linear combination of $v + cw$ and w. d and e are still in the real numbers so this substitution was allowed. Thus, $Span(v, w)$ is in $Span(v + cw, w)$.

Prove that if u is a linear combination of $v_1, \cdots, v_k$ that cu also is.
$u = c_1v_1 +, \cdots, c_kv_k$ for some c_i. $cu = cc_1v_1 + \cdots + cc_kv_k$.
$c \times c_i$ is still a scalar, so this is still a linear combination of vectors.

Prove that if u and w are linear combinations of $v1, \cdots, vk$, $u + w$ also is.
$u = c_1v_1 + \cdots + c_kv_k$ for some c_i and $w = u = d_1v_1 + \cdots + d_kv_k$ for some d_i.
Thus, $u + w = (c_1 + d_1)v_1 + \ldots..(c_k + d_k)v_k$ is still a linear combination because $c_i + d_i$ are still scalars.

Consider the line $x = x_0 + rw$ and the plane $x = su + tv$. Show that if they intersect, x_0 is in the plane. That would mean for some r_0, s_0, and t_0, $x_0 + r_0w = s_0u + t_0v$. $x_0 =$

6

$s_0u + t_0v - r_0w$.

For the line to intersect the plane, w must have a directional component that lies in the plane. Thus, w can be expressed as a linear combination of u and v, so x_0 is solely a linear combination of u and v, the vectors that span the plane.

Consider the lines $x = x_0 + tv$ and $x = x_1 + su$. Show that they intersect if and only if $x_0 - x_1$ lies in $Span(u, v)$.

Direction 1: If the lines intersect, then $x_0 - x_{10}$ lies in $Span(u, v)$

That would mean for some t_0 and s_0, $x_0 + t_0v = x_1 + s_0u$. $x_0 - x_1 = s_0u - t_0v$. We see $x_0 - x_1$ is a linear combination of u and v, so it is in the span.

Direction 2: If $x_0 - x_1$ lies in $Span(u, v)$, then the lines intersect

If $x_0 - x_1$ lies in the span, it is equal to $au + bv$ for some a and b.

Rearranging, $x_0 = x_1 + au + bv$

This shows that x_0 (a point on the first line) can be reached from a point x_1 (a point on the second line) via a linear combination of u and v, implying an intersection, as the lines have to connect for this to be possible.

Prove the Cauchy-Schwarz Inequality: $-x \times y| \leq \|x\| \|y\|$.

Additionally, equality holds if and only if one of the vectors is a scalar multiple of the other.

If one of the vectors is the 0 vector, it holds.

First, let's assume x and y are unit vectors. So $\|x\| = \|y\| = 1$.

Consider $x + y$ and $x - y$, components of the parallelogram formed by x and y.

$\|x + y\|^2 = \|x\|^2 + 2x \times y + \|y\|^2 = 2(x \times y + 1)$

$\|x - y\|^2 = \|x\|^2 - 2x \times y + \|y\|^2 = 2(-x \times y + 1)$

These results must be nonnegative because the length of vectors is non-negative.

We get $x \times y + 1 \geq 0$ and $-x \times y + 1 \geq 0$

$-1 \leq x \times y \leq 1$, or equivalently, $|x \times y| \leq 1$

Equality holds if and only if $x + y = 0$ or $x - y = 0$, equivalent to x being $\pm y$.

When we have non-unit vectors, they can be normalized: $\frac{x}{\|x\|}$ and $\frac{y}{\|y\|}$.

We see that $\left(\frac{x}{\|x\|}\right) \times \left(\frac{y}{\|y\|}\right) \leq 1$ by plugging in these normalized vectors.

Multiplying by $\|x\| \|y\|$, we get the inequality.

$|x \times y| \leq \|x\| \|y\|$

Equality holds if and only if $\frac{x}{\|x\|} = \pm \frac{y}{\|y\|}$

If x is orthogonal to each of the vectors $v1, ... vk$, show that x is orthogonal to any linear combination of them.

Dotting $c_1v_1 + \cdots + c_kv_k$ with x: $c_i(v_i \times x) = c_i(0) = 0$. Each term is 0, so $x \times$ (any linear combination) $= 0$, meaning it is orthogonal.

If $x \times y = 0$ for all x, prove that $y = 0$.

Let's look at the specific case of when $x = y$. That makes it $\|y\|^2 = 0$, and for this to

be true, y must equal 0.

Prove $Ax = b$ is consistent precisely when b can be written as a linear combination of the columns.

Direction 1: If $Ax = b$ is consistent for some x, then b is a linear combination of the columns of A, with the x is being the weights.

Direction 2: If b can be written as a linear combination of the columns, then $b = c_1a_1 + \cdots + c_na_n$. If $x = [c_1, \cdots, c_n]$, Ax would yield exactly that, so it would be consistent.

This also proves the $C(A)$ is equivalent to the set of all b s.t. $Ax = b$ is consistent.

Theorem 5.3: Assume the system $Ax = b$ is consistent, and let u_1 be a particular solution. Then all the solutions to the non-homogenous system are of the form $u = u_1 + v$ for some solution v of the associated homogeneous system $Ax = 0$.

Recall, x is a solution if it makes $Ax = b$ consistent for some b (including 0) and a particular solution is a solution to the non-homogenous system.

Direction 1: $u_1 + v$ is a particular solution to the non-homogenous system If we take $A(u_1 + v)$, distributing, we get $Au_1 + Av$, and $Au_1 = b$ and $Av = 0$, so this is $b + 0 = b$.

Direction 2: Showing every solution is of the form $u_1 + v$

We need to show that there exists a v s.t. $u_1 + v$ is a particular solution and that v is a solution to the homogenous solution. That means v is $u - u_1$.

$A(u - u_1) = Au - Au_1$. In direction 1, we showed $u(u = u_1 + v)$ is a solution to $Ax = b$, and know u_1 is as well, so we get $b - b = 0$.

Thus, $u - u_1$ is a solution to $Ax = 0$, meaning this equals v for some v. If $v = u - u_1$, we get that $u = u1 + v$. This direction proved every solution to $Ax = b$ can be expressed as $u_1 + v$ for some v.

Geometric Interpretation: When $Ax = b$ is consistent, its solutions are obtained by translating the set of solutions to the homogenous system by a particular solution u.

Linear Transformations and Matrix Properties

Prove projection is a linear transformation

(i) $T(x + y) = \dfrac{(x + y) \cdot a}{||a||^2}a = \dfrac{x \cdot a}{||a||^2}a + \dfrac{y \cdot a}{||a||^2}a = T(x) + T(y)$

(ii) $T(cx) = \dfrac{(cx) \cdot a}{||a||^2}a = c\dfrac{x \cdot a}{||a||^2} = cT(x)$

Prove every linear transformation has a standard matrix.

We want to show $T(x) = A(x)$ for x in R^n

$x = x_1(\vec{e}_1) + \ldots + x_n(\vec{e}_n)$ $T(x) = T(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n)$

Knowing that this is a linear transformation, we can use properties of it.

$$\begin{bmatrix} | & | & & | \\ T(\vec{e}_1) & T(\vec{e}_2) & \cdots & T(\vec{e}_n) \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

This looks like the result of a matrix-vector product with xi as elements of the vector and $T(\vec{e}_i)$ as the columns of the matrix.

Find the inverse of a 2×2 rotation matrix
$$A_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$
To undo rotation by theta, you must rotate by $-\theta$. Note that $cos(-\theta) = \cos(\theta)$ and $\sin(\theta) = -\sin(\theta)$
$$A_\theta^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Lemma 1. For an mxn matrix A, the following are true:

   (i) A has a right inverse $\Leftrightarrow A$ has rank m and m$\leq$n (iff transformation surjective)

   (ii) A has a left inverse $\Leftrightarrow A$ has rank n and n$\leq$m (iff transformation injective)

   Corollary:

   (iii) A is square and $n \times n$, rank n $\Leftrightarrow$ having a left and right inverse.

Lemma 2. Prove that a matrix has a left inverse and a right inverse, they are equal.
A must be square and rank n to have both a left and a right inverse by property (iii) of Lemma 1.
Suppose $CA = I = AB$.
$C = CI = C(AB) = (CA)(B) = (I)(B) = B$. So $C = B$.

Lemma 3. For a square matrix, the existence of a right inverse implies the existence of a left inverse, and vice versa.
If a square matrix has a right inverse, its rank equals the number of rows, which makes it invertible, so a left inverse exists.
Likewise, if it has a left inverse, its rank equals the number of columns, which makes it invertible, so a right inverse exists.

Proposition 3.4. Suppose A and B are invertible $n \times n$ matrices. Then their product AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.

We only need to check one side. If a square matrix has a right inverse, it also must have a left inverse, and by Lemma 3.1, they are the same.

Check that $(AB)(B^{-1}A^{-1}) = I$

By associativity, $A(BB^{-1})A^{-1} = AA^{-1} = I$

Lemma 5. Prove that each vector in a subspace V has a unique representation as a linear combination of the basis $\{v_1, \ldots, v_k\}$ for V.

Suppose this arbitrary element $u \in V$ has two different representations:

$$u = c_1 v_1 + \ldots + c_k v_k \quad \text{and} \quad u = d_1 v_1 + \ldots + d_k v_k.$$

Then, subtracting these equations, we get

$$(c_1 - d_1)v_1 + \ldots + (c_k - d_k)v_k = 0.$$

Since a basis is linearly independent, $c_i - d_i$ must be 0 for all i. Thus, $c_i = d_i$, and there were never two different representations.

If $v_1, \ldots, v_k$ is an orthonormal basis for a subspace V, then the unique representation of any vector $u \in V$ in terms of this basis is given by $u = \langle u, v_1 \rangle v_1 + \ldots + \langle u, v_k \rangle v_k$, where $\langle \cdot, \cdot \rangle$ denotes an arbitrary inner product.

We know $\mathbf{u} = c_1 \mathbf{v}_1 + \cdots + c_k \mathbf{v}_k$ for some unique constants $c_1, \ldots, c_k$.

Taking the inner product of both sides with an arbitrary $\mathbf{v}_j$ where $1 \leq j \leq k$, we have

$$\langle \mathbf{v}_j, \mathbf{u} \rangle = \langle \mathbf{v}_j, c_1 \mathbf{v}_1 + \cdots + c_k \mathbf{v}_k \rangle.$$

Using the linearity of the inner product, we get

$$\langle \mathbf{v}_j, \mathbf{u} \rangle = \langle \mathbf{v}_j, c_1 \mathbf{v}_1 \rangle + \cdots + \langle \mathbf{v}_j, c_k \mathbf{v}_k \rangle.$$

Using the scaling properties, we obtain

$$\langle \mathbf{v}_j, \mathbf{u} \rangle = c_1 \langle \mathbf{v}_j, \mathbf{v}_1 \rangle + \cdots + c_k \langle \mathbf{v}_j, \mathbf{v}_k \rangle.$$

Note that $\mathbf{v}_j$ is an arbitrary element in our orthonormal basis, so $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for all $i \neq j$. Thus, we are left with

$$\langle \mathbf{v}_j, \mathbf{u} \rangle = c_j \langle \mathbf{v}_j, \mathbf{v}_j \rangle = c_j \|\mathbf{v}_j\|^2.$$

Using the fact that $\|\mathbf{v}_j\|^2 = 1$, we see

$$\langle \mathbf{v}_j, \mathbf{u} \rangle = c_j.$$

Thus, for an arbitrary index j of the basis, $c_j = \langle \mathbf{v}_j, \mathbf{u} \rangle$.

Therefore, we can see the unique representation of $\mathbf{u}$ in V with an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is

$$\langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \cdots + \langle \mathbf{u}, \mathbf{v}_k \rangle \mathbf{v}_k.$$

Proposition 5.1.

(i) $(A^T)^T = A$

(ii) $(cA)^T = cA^T$

(iii) $(A + B)^T = A^T + B^T$

(iv) $(AB)^T = B^T A^T$

(v) When A is invertible, then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$

Proof of (v): Transpose $A^T(A^{-1})^T$. $(A^T(A^{-1})^T)^T = (A^{-1})^{T^T} A^{T^T} = A^{-1}A = I$, so we showed the right inverse of A^T is $(A^{-1})^T$. We have found the right inverse of A^T, which guarantees it has full row rank. To be invertible, A must be square, so A^T must also be square. Seeing as A^T has rank = # rows, and the # rows = # cols, it is nonsingular $\Leftrightarrow$ invertible.

Lemma 4. $x \times y = x^T \times y$
If x and y are the same dimension (say column vectors), $x \times y = x_1 y_1 + x_2 y_2 + \cdots = \sum_{i=1}^{n} x_i y_i$ (scalar multiplication). If x is transposed, it becomes a row vector. If x and y were originally row vectors, then we could use the fact $x \times y = y \times x$ and transpose x here to become a column vector, yielding the same result.
We now have a form of matrix multiplication, a $1 \times n \times n \times 1$ product. This yields $x_1 y_1 + x_2 y_2 + \cdots = \sum_{i=1}^{n} x_i y_i$.

Proposition 5.2. Prove $Ax \times y = x \times A^T y$
$Ax \times y = (Ax)^T \times y$ by Lemma 4
$(Ax)^T \times y = (x^T A^T) \times y = x^T \times (A^T y)$

Determinants

Theorem 1.2. Let A be a square matrix. Then A is nonsingular if and only if $\det A$ is nonzero.
Direction 1: If a matrix is nonsingular, it has a nonzero determinant.
A has an inverse. $1 = \det(I) = \det(AA^{-1}) = \det(A)\det(A^{-1})$. If $\det(A)$ was zero, this product could not be one.
Direction 2: If a matrix has nonzero determinant, it is nonsingular.
(Proving Contrapositive: If a matrix is singular, it has zero determinant.)
If A is singular, it has no inverse. Thus, $1 = \det(I)$ does not equal $\det(AB) = $

$\det(A)\det(B)$ for any B. The only way for $1 = xy$ to not hold for all y is if $x = 0$, so $\det(A) = 0$. As we proved the contrapositive, we also get that any matrix with nonzero determinant is nonsingular.

Proposition 1.3. If A is a triangular matrix, then the determinant is the product of its diagonal entries.

U is an upper triangular matrix. As the transpose of an upper triangular matrix is lower triangular and vice versa, if we prove one, it will hold for the other (the diagonal entries remain the same in A and A^T).

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ 0 & u_{22} & u_{23} & \cdots & u_{2n} \\ 0 & 0 & u_{33} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_{nm} \end{bmatrix}$$

If we perform Cofactors expansion along the first column, only the first term in the sum will matter (all the next terms are multiplied by 0 because the rest of the column is 0). We get $u_{11} \times$ [a $(n-1) \times (n-1)$ matrix with u_{22} at the top left]. To find the determinant of this matrix, we will use Cofactors again along its first column, and once again, the zeros in later entries of the column make the determinant of the $n-1$ matrix $u_{22} \times$ [$n-2$ matrix starting with u_{33}]. Repeating this down to a 1x1 matrix, where the determinant is just the element, we get the product of diagonal entries.

Proposition 1.4. Let E be an elementary matrix, and let A be a square matrix. Then $\det(EA) = \det(E)\det(A)$.

Elementary matrices are one of three types.

If E was a row swap matrix, it would flip the flip the sign of the determinant of A. $\det(E)$ is -1 as it is a swap from the identity matrix, and $-1 \times \det(A)$ is indeed -$\det(A)$.

If E was a scaling matrix, it would scale the determinant of A by c. $\det(E)$ is c as it scales a row by c. Plugging in c for $\det(E)$, we see $c \times det(A)$ is indeed $c \times det(A)$.

If E was a row addition matrix, it would not change the determinant of A. Likewise, as E just added a row from the identity matrix, that operation does not change the determinant, so its determinant is 1. $1\det(A) = det(A)$.

Theorem 1.5. Let A and B be $n \times n$ matrices. Then $\det(AB) = \det(A)\det(B)$.

(A consequence of this theorem is that $\det(AB) = \det(BA)$)

If A is singular, then AB would also be singular (Lemma 0). That would mean $\det(AB)$ is 0, and the $\det(A)\det(B)$ would be 0 because $\det(A)$ is 0.

If A is nonsingular, it can be written as a product of elementary matrices $E_n \cdots E_1$. By

Proposition 1.4, $\det(A) = \det(E_n) \cdots \det(E_1)$.

Now, finding $\det(AB)$ or $\det(E_n \cdots E_1 B)$, we can apply the same property. We get $\det(E_n) \cdots \det(E1) \det(B)$ and already showed that $\det(A) = \det(E_n) \cdots \det(E_1)$.

Lemma 0. If A is singular, then AB is also singular.

Suppose AB is nonsingular. That would mean there is some C s.t. $AB(C) = I$. By associativity, that would mean $A(BC) = I$, but A is singular and thus has no inverse. Thus, AB must be singular.

Corollary 1.6A. If A is nonsingular, then $\det(A^{-1} = \frac{1}{\det(A)}$.

$1 = \det(I) = \det(AA^{-1}) = \det(A) \det(A^{-1})$.

Rearranging $1 = \det(A) \det(A^{-1})$, we get $\det(A^{-1}) = \frac{1}{\det(A)}$.

Corollary 1.6B. Similar matrices have the same determinant.

Let A and B be similar matrices. Then, $B = P^{-1} A P$.

$\det(P^{-1} A P) = \det(P^{-1}) \det(A) \det(P)$. By Corollary 1.6A, $det(A^{-1}) \times \det(A) = 1$, leaving just $\det(A)$.

Proposition 1.7. $\det(A) = \det(A^T)$

Suppose A is singular. Then, so is A^T.

If A is nonsingular, then $\det(A) = \det(E_n \cdots E_1) = \det(E_n) \cdots det(E_1)$.

$det(A^T) = \det((E_1 \cdots E_1)^T) = \det(E_1^T \cdots E_n^T) = \det(E_1^T) \cdots \det(E_n^T)$. By Lemma 1, $\det(E) = \det(E^T)$, so $\det(A^T) = \det(E_1) \cdots \det(E_n)$. Scalar multiplication is commutative so it is true.

Lemma 1: $\det(E) = \det(E^T)$

Three Cases:

Row scaling matrices are symmetric, and thus $E = E^T$.

Adding rows does not change the determinant, and the transpose of that still is a matrix that adds rows, so its determinant does not change.

Transposing a row swap matrix gives a column swap, and column operations have the same effect on the determinate as row operations.

Corollary 1.8:

From Proposition 1.7, column operations work the same as row operations.

Lemma 1.10: If two rows of a matrix A are equal, then $\det(A) = 0$.

This is a singular matrix so $\det(A) = 0$.

Alternatively, we can say that if we swapped those two rows, nothing would change, so we would get the same determinant, but the determinant changes sign when rows are

swapped. $\det(A) = -\det(A)$ implies $\det(A)$ is 0.

Suppose A is an orthogonal matrix. What are the possible values for $\det(A)$?
An orthogonal matrix has the property $A^T A = I$.
$1 = \det(I) = \det(A^T A) = \det(A)\det(A^T)$, and as $\det(A) = \det(A^T)$, $1 = \det(A)^2$ so
$\det(A) = \pm 1$.

Properties of Determinants:

1. If A' exchanges two rows from A, $\det A' = -\det A$

2. If A' multiplies a row by a scalar c, $\det A' = c \det A$

3. If A' adds a multiple of a row to another row, $det A' = \det A$.

4. $\det(I) = 1$

(3 Prime) Suppose the ith row of A is the sum of two vectors A' and A'' and all other rows are the same. Then, $\det A = \det A' + \det''$.

Lemma 1.11. If the determinant satisfies (1) (2) (3 Prime) (4) then it satisfies (1) (2) (3) (4).

The ijth cofactor of a matrix A is defined as $C_{ij} = (-1)^{i+j} \times \det(A_{ij})$, with A being the $(n-1)$ by $(n-1)$ matrix without the ith and jth column.

Proposition 2.1

Let A be an $n \times n$ matrix. Then for fixed i, we have

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij}$$

Utilize Properties (1) (2) (3') and (4)

Corollary 2.2

$$\det A = \sum_{i=1}^{n} a_{ij} C_{ij}$$

Using Proposition 2.1 and $\det(A) = \det(A^T)$, this follows.

Proposition 2.3: Cramer's Rule. If A is nonsingular, the ith coordinate of the vector x that solves $Ax = b$ is $x_i = \frac{\det(B_i)}{\det(A)}$ where B_i is the matrix obtained by replacing the ith

column of A by the vector b.

$$\det B_i = \det \begin{bmatrix} | & | & | & & | & & | \\ a_1 & a_2 & \cdots & x_1 a_1 + \cdots + x_n a_n & \cdots & a_n \\ | & | & | & & | & & | \end{bmatrix}$$

$$= \det \begin{bmatrix} | & | & | & | & | \\ a_1 & a_2 & \cdots & x_i a_i & a_n \\ | & | & | & | & | \end{bmatrix} = x_i \det A$$

We first express b as a linear combination of the columns of A as we know there will be a unique combination as it is nonsingular. Then we can subtract x_j amounts of aj for all i not equal to j from the column with b, leaving just $x_i a_i$. Subtracting a multiple of a column from another column does not change the determinant. The constant can then be pulled out by column operations of the determinant, giving $\det(B_i) = x_i \det(A)$ which can be arranged to $x_i = \frac{\det(B_i)}{\det(A)}$.

Proposition 2.4. If A is nonsingular, then $A^{-1} = \frac{1}{\det(A)} C^T$, where C is the matrix of its cofactors.

Eigenvalues and Eigenvectors

A is diagonalizable if there exists P s.t. $P^{-1}AP$ is diagonal. In words, A is diagonalizable if it is similar to a diagonal matrix.
A transformation is diagonal if $T(v_i) = \lambda_i \times v_i$ for any scalar λ_i.

Proposition 1.1. A linear transformation $T : V \to V$ is diagonalizable if and only if there is a basis for V consisting of the eigenvectors for T.

Lemma 1.2. Let A be $n \times n$ and λ be a scalar. $E(\lambda) = \{x \in \mathbb{R}_n : Ax = \lambda_x\} = N(A - \lambda I)$
Notice that $Ax = \lambda x$ can be rewritten as $Ax - \lambda x = 0$ and $Ax - \lambda[I]x = 0$.
Factoring, $(A - \lambda[I])x = 0$, and the set of all x such that this is true is $N(A - \lambda I)$.

Proposition 1.3. Let A be an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.
λ is an eigenvalue of $A \Leftrightarrow$ there exists a nontrivial x that satisfies $Ax = \lambda x \Leftrightarrow$ there exists a nontrivial x in the null space of $A - \lambda I \Leftrightarrow A - \lambda I$ is singular $\Leftrightarrow \det(A - \lambda I) = 0$.

As these statements are logically equivalent, the other direction also holds.

Equivalence 1: By the definition of an eigenvalue, $T(x) = Ax = \lambda x$ for nonzero x.

Equivalence 2: By Lemma 1.2.

Equivalence 3: Singularity $\Leftrightarrow$ nontrivial null space because rank + nullity = n (number of columns). As rank is less than n (if a square matrix had rank n, it would be nonsingular), then nullity is at least 1, meaning it contains more than the zero vector.

Equivalence 4: Singular matrices have determinant 0

$p(t) = pA(t) = \det(A - tI)$ is the characteristic polynomial of A.

Proposition 1.4. If A and B are similar matrices, then $pA(t) = pB(t)$ (which means they have the same eigenvalues).

Suppose $B = P^{-1}AP$ (or $A = P^{-1}BP$).

Then $pB(t) = \det(B - tI) = \det(P^{-1}AP - tI) = \det(P^{-1}AP - P^{-1}(tI)P)$.

tI still equals $P^{-1}(tI)P$ because you can pull out the scalar t.

Factoring, $\det(P^{-1}(AP - tI)P) = \det(P^{-1})\det(A - tI)\det(P)$.

Recall, $\det(P^{-1}) = \frac{1}{\det(P)}$.

We are left with $\det(A - tI)$, or $pA(t)$.

Proposition 1.5. For any square matrix A, A and A^T have the same set of eigenvalues. That is, $\det(A - tI)$ *equals* $\det(A^T - tI)$.

Proof. We use the properties of determinants and transposes:

$$\begin{aligned}
\det(A - tI) &= \det((A - tI)^T) \quad \text{(since } \det(B) = \det(B^T) \text{ for any } B) \\
&= \det(A^T - (tI)^T) \quad \text{(distributing the transpose)} \\
&= \det(A^T - tI) \quad \text{(since } (tI)^T = tI, I \text{ is symmetric).}
\end{aligned}$$

This shows that $\det(A - tI) = \det(A^T - tI)$ so A and A^T have the same eigenvalues.

Theorem 2.1. Let $T : V \to V$ be a linear transformation. Suppose $v_1, \cdots, v_k$ are eigenvectors of T with distinct corresponding eigenvalues $\lambda_1, \cdots, \lambda_k$. Then $\{v_1, \cdots, v_k\}$ is a linearly independent set of vectors.

Suppose $c_1 v_1 + \cdots + c_k v_k = T(c_1 v_1 + ... + c_k v_k) = c_1 T(v_1) + \cdots + c_k T(v_k) = T(0) = 0$ if it is a linear transformation.

Recall if $Ax = \lambda x$. Ax can be represented as $T(x)$. Thus, $T(v_i) = \lambda_i \times v_i$.

$c_1 \lambda_1 v_1 + \cdots + c_k \lambda_k v_k = 0$

We have a system of equations now.

$c_1 v_1 + c_2 v_2 + \cdots = 0$

$c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \cdots = 0$

Multiply equation 1 by λ_1 and then subtract it from the second equation.

We get $c_2(\lambda_2 - \lambda_1)v_2 + c_3(\lambda_3 - \lambda_1)v_3 + \cdots = 0$.

Iterate this and multiply equation 1 by λ_i and subtract it.

This will eventually give us $c_k(\lambda_k - \lambda_{i-1})v_k = 0$.

None of the $(\lambda_j - \lambda_i)$ for $i \neq j$ are zero because the λs are distinct.

The vectors v_i cannot be 0 as it is an eigenvector, so this constant c_k must be 0. Now that we know this term is zero, backtrack to c_{k-1}. The same argument applies. So all constants must be zero, so the set is linearly independent.

Corollary 2.2. Suppose V is an n-dimensional vector space and $T : V \to V$ has n distinct (real) eigenvalues. Then T is diagonalizable.

Proof. The set of the n corresponding eigenvectors will be linearly independent and will hence give a basis for V. The matrix for T with respect to a basis of eigenvectors is always diagonal.

Algebraic Multiplicity: λ's multiplicity as a root of $p(t)$. Alternatively - the highest power of $t - \lambda$ dividing $p(t)$

Geometric Multiplicity: The dimension of the eigenspace that corresponds to λ. Note that dimension cannot be 0 because an eigenspace would never contain the zero vector.

Proposition 2.3. Let λ be an eigenvalue of algebraic multiplicity m and geometric multiplicity d. Then $1 \leq d \leq m$.

Geometric multiplicity, the dimension of the eigenspace must be at least dimension 1 because it doesn't contain the 0 vector and thus can't be dimension 0 (and dimensions obviously cannot be negative).

Algebraic multiplicity must be at least one because the roots of the characteristic polynomial are the eigenvalues of a matrix, and λ is an eigenvalue, so m is at least 1. The algebraic multiplicity m of λ is defined as the multiplicity of λ as a root of the characteristic polynomial of A. It represents the maximum number of linearly independent eigenvectors associated with λ. Therefore, the geometric multiplicity, which counts the actual number of such independent eigenvectors, cannot exceed the algebraic multiplicity. Thus, $d \leq m$.

Theorem 2.4. Let $T : V \to V$ be a linear transformation. Let its distinct eigenvalues be $\lambda_1, \cdots, \lambda_k$ and assume these are all real numbers. Then T is diagonalizable if and only if the geometric multiplicity, di, of each λ_i equals its algebraic multiplicity, m_i

Direction 1: Suppose $d_i = m_i$ for each λ_i. This implies a total of $m_1 + \cdots + m_k$ linearly independent eigenvectors, forming a basis for V. Hence, T is diagonalizable.

Direction 2: If T is diagonalizable, there exists a basis of V consisting of eigenvectors of T. The total number of these eigenvectors equals n, the dimension of V. As the sum of algebraic multiplicities $m_1 + \cdots + m_k$ also equals n, and these eigenvectors span the eigenspaces, it follows that $d_i = m_i$ for each λ_i.

Theorem 0.1 (Spectral Theorem). *Let A be a symmetric $n \times n$ matrix. Then*

1. *The eigenvalues of A are real.*

2. *There is an orthonormal basis $\{q_1, \ldots, q_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors of A. That is, there is an orthogonal matrix Q so that $Q^T A Q = \Lambda$ is diagonal.*

Theorem 0.2 (Cayley-Hamilton Theorem)**.** *Let A be a square matrix over a commutative ring. Then the characteristic polynomial $p(\lambda)$ of A is such that $p(A) = 0$. In other words, A satisfies its own characteristic equation.*

Theorem 0.3 (Invariance)**.** *Let A be an $n \times n$ matrix, and let λ be any scalar. Then the subspace $V = C(A - \lambda I)$ has the property that whenever $v \in V$, it is the case that $Av \in V$. That is, the subspace V is invariant under μ_A.*

Theorem 0.4 (Jordan Canonical Form)**.** *Suppose the characteristic polynomial of an $n \times n$ complex matrix A is $p(t) = \pm(t - \lambda_1)^{m_1}(t - \lambda_2)^{m_2} \cdots (t - \lambda_k)^{m_k}$. Then there is a basis $\mathcal{B}$ for $\mathbb{C}^n$ with respect to which the matrix representing μ_A is "block diagonal".*

Theorem 0.5. *Let A be a diagonalizable $n \times n$ matrix. The general solution of the initial value problem*

$$(*) \quad \frac{d\mathbf{x}}{dt} = A\mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

is given by $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$.

Theorem 0.6. *Let n be a positive integer. The set of solutions of the nth-order ODE (*) is an n-dimensional subspace of $C^\infty(\mathbb{R})$, the vector space of infinitely differentiable functions defined on $\mathbb{R}$. In particular, the initial value problem*

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_2 y''(t) + a_1 y'(t) + a_0 y(t) = 0$$

with initial conditions

$$y(0) = c_0, \quad y'(0) = c_1, \quad y''(0) = c_2, \quad \ldots, \quad y^{(n-1)}(0) = c_{n-1}$$

has a unique solution.