Elementary Linear Algebra Reference

Nathan Warner

Computer Science
Northern Illinois University
United States

Contents

Vectors

- **Magnitude**: For a vector $x \in \mathbb{R}^n$, $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, the norm (magnitude) of x is

$$\|x\| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$$

- **Triangle inequality**: For vectors $x, y \in \mathbb{R}^n$, we have the inequality

$$\|x + y\| \leqslant \|x\| + \|y\|$$

- **Properties of Vector Operations**: Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in a plane. Let r and s be scalars.

  1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (Commutative property)
  2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (Associative property)
  3. $\mathbf{u} + \mathbf{0} = \mathbf{u}$ (Additive identity property)
  4. $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ (Additive inverse property)
  5. $r(s\mathbf{u}) = (rs)\mathbf{u}$ (Associativity of scalar multiplication)
  6. $(r + s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$ (Distributive property)
  7. $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$ (Distributive property)
  8. $1\mathbf{u} = \mathbf{u}$, $0\mathbf{u} = \mathbf{0}$ (Identity and zero properties)

- **Finding components of a vector given the magnitude and the angle θ**: If $v \in \mathbb{R}^2$, $v = \begin{pmatrix} x \\ y \end{pmatrix}$, then

$$x = \|\vec{v}\| \cos \theta$$
$$y = \|\vec{v}\| \sin \theta.$$

- **Unit vector**: A unit vector is a vector with magnitude 1. For any nonzero vector $\vec{v}$, we can use scalar multiplication to find a unit vector $\vec{u}$ that has the same direction as $\vec{v}$:. To do this, we multiply the vector by the reciprocal of its magnitude:

$$\vec{u} = \frac{1}{\|\vec{v}\|} \vec{v}.$$

- **Properties of the dot product**: Let $\vec{u}$, $\vec{v}$, and $\vec{w}$ be vectors, and let c be a scalar.

  1. Commutative property: $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
  2. Distributive property: $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$
  3. Associative property of scalar multiplication: $(c\vec{u} \cdot \vec{v}) = (c\vec{u}) \cdot \vec{v} = \vec{u} \cdot (c\vec{v})$
  4. Property of magnitude: $\vec{v} \cdot \vec{v} = \|\vec{v}\|^2$

- **Evaluating a dot product**: The dot product of two vectors is the product of the magnitude of each vector and the cosine of the angle between them:

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \cdot \|\vec{v}\| \cdot \cos \theta.$$

- **Find the measure of the angle between two nonzero vectors**:

$$\cos \theta = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \, \|\vec{v}\|}.$$

Note: We are considering $0 \leqslant \theta \leqslant \pi$

- **Vector Projection**: The vector projection of $\mathbf{v}$ onto $\mathbf{u}$ has the same initial point as $\mathbf{u}$ and $\mathbf{v}$ and the same direction as $\mathbf{u}$, and represents the component of $\mathbf{v}$ that acts in the direction of $\mathbf{u}$:.

$$\text{proj}_{\vec{u}} \vec{v} = \frac{\vec{v} \cdot \vec{u}}{\|\vec{u}\|^2} \vec{u}.$$

We say "The vector projection of $\vec{v}$ onto $\vec{u}$"

- **Scalar projection notation**: This is the length of the vector projection and is denoted

$$\|\text{proj}_{\vec{u}} \vec{v}\| = \text{comp}_{\vec{u}} \vec{v} = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|}.$$

- **Decompose some vector $\vec{v}$ into orthogonal components such that one of the component vectors has the same direction as $\vec{u}$**:

    - First, we compute $\vec{p} = \text{proj}_{\vec{u}} \vec{v}$
    - Then, we define $\vec{q} = \vec{v} - \vec{p}$
    - Check that $\vec{q}$ and $\vec{p}$ are orthogonal by finding $\vec{q} \cdot \vec{p}$

- **Two vectors are orthogonal if**:

$$\vec{u} \cdot \vec{v} = 0.$$

- **Two vectors are parallel if**: Two vectors v, u are parallel if there exists some scalar $\alpha \in \mathbb{R}$ such that $\alpha u = v$

    - If $\alpha > 0$, then v points in the same direction as u
    - If $\alpha < 0$, then v points in the opposite direction of u

- **Scalar projection componets of a vector**:

$$\vec{v} = \langle \text{comp}_{\hat{i}} \vec{v}, \text{comp}_{\hat{j}} \vec{v}, \text{comp}_{\hat{k}} \vec{v} \rangle.$$

- **The Cross Product**: produces a vector perpendicular to both vectors involved in the multiplication

Let $\mathbf{u} = \langle u_1, u_2, u_3 \rangle$ and $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$. Then, the cross product $\mathbf{u} \times \mathbf{v}$ is vector

$$\mathbf{u} \times \mathbf{v} = (u_2 v_3 - u_3 v_2)\mathbf{i} - (u_1 v_3 - u_3 v_1)\mathbf{j} + (u_1 v_2 - u_2 v_1)\mathbf{k}$$
$$= \langle u_2 v_3 - u_3 v_2, -(u_1 v_3 - u_3 v_1), u_1 v_2 - u_2 v_1 \rangle..$$

Note: The cross product only works in $\mathbb{R}^3$, additionally, we measure the angle between $\vec{u}$ and $\vec{v}$ in $\vec{u} \times \vec{v}$ from $\vec{u}$ to $\vec{v}$

- **Cross product using matrix and determinant**, suppose we have vectors $\vec{u}$ und $\vec{v}$:. Then we can express them in matrix form as

$$\vec{u} \times \vec{v} = \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{bmatrix}.$$

Then we can find the determinant of this matrix to compute the cross product

$$\vec{u} \times \vec{v} = (u_y v_z - u_z v_y)\hat{i} - (u_x v_z - u_z v_x)\hat{k} + (u_x v_y - u_y v_x)\hat{j}.$$

4

- **Properties of the Cross Product**: Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in space, and let c be a scalar.

  1. Anticommutative property: $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$
  2. Distributive property: $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$
  3. Multiplication by a constant: $c(\mathbf{u} \times \mathbf{v}) = (c\mathbf{u}) \times \mathbf{v} = \mathbf{u} \times (c\mathbf{v})$
  4. Cross product of the zero vector: $\mathbf{u} \times \mathbf{0} = \mathbf{0} \times \mathbf{u} = \mathbf{0}$
  5. Cross product of a vector with itself: $\mathbf{v} \times \mathbf{v} = \mathbf{0}$
  6. Scalar triple product: $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$

- **Magnitude of the Cross Product**: Let $\mathbf{u}$ and $\mathbf{v}$ be vectors, and let θ be the angle between them. Then, $\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\| \cdot \|\mathbf{v}\| \cdot \sin\theta$.

- **Triple Scalar Product**:

  The triple scalar product of vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ is $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$.

  The triple scalar product is the determinant of the 3×3 matrix formed by the components of the vectors

- **triple scalar product identities**:

  (a) $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = -\mathbf{u} \cdot (\mathbf{w} \times \mathbf{v})$
  (b) $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \mathbf{v} \cdot (\mathbf{w} \times \mathbf{u} = \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}))$

- **The zero vector is considered to be parallel to all vectors**:

- **vector equation of a line**:

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}.$$

  Where $\mathbf{v}$ is the direction vector (vector parallel to the line), t is some scalar, and $\mathbf{r}$, $\mathbf{r}_0$ are position vectors

- **Parametric and Symmetric Equations of a Line**: A line L parallel to vector $\mathbf{v} = \langle a, b, c \rangle$ and passing through point $P(x_0, y_0, z_0)$ can be described by the following parametric equations:

$$x = x_0 + ta, \quad y = y_0 + tb, \quad \text{and} \quad z = z_0 + tc.$$

  If the constants a, b, and c are all nonzero, then L can be described by the symmetric equation of the line:
$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

  Note: The parametric equations of a line are not unique. Using a different parallel vector or a different point on the line leads to a different, equivalent representation. Each set of parametric equations leads to a related set of symmetric equations, so it follows that a symmetric equation of a line is not unique either.

- **Vector equation of a line reworked**: Suppose we have some line, with points $P(x_0, y_0, z_0)$, $Q(x_1, y_1, z_1)$. Where $\mathbf{p} = \langle x_0, y_0, z_0 \rangle$ and $\mathbf{Q} = \langle x_1, y_1, z_1 \rangle$ are the corresponding position vectors. Suppose we also have $\mathbf{r} := \langle x, y, z \rangle$. Then our vector equation for a line becomes

$$\mathbf{r} = \mathbf{p} + t\left(\vec{PQ}\right).$$

  By properties of vectors, we get the vector equation of a line passing through points P and Q to be

$$\mathbf{r} = (1 - t)\mathbf{p} + t\mathbf{q}.$$

- **Distance from a Point to a Line**: Let L be a line in space passing through point P with direction vector $\mathbf{v}$. If M is any point not on L, then the distance from M to L is

$$d = \frac{\left\| \overrightarrow{PM} \times \mathbf{v} \right\|}{\|\mathbf{v}\|}$$

- **Vector equation of a plane**: Given a point P and vector $\mathbf{n}$, the set of all points Q satisfying the equation $\mathbf{n} \cdot \overrightarrow{PQ} = 0$ forms a plane. The equation

$$\mathbf{n} \cdot \overrightarrow{PQ} = 0$$

is known as the vector equation of a plane.

- **Scalar equation of a plane**: The scalar equation of a plane containing point $P = (x_0, y_0, z_0)$ with normal vector $\mathbf{n} = \langle a, b, c \rangle$ is

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

- **General form of the equation of a plane**: This equation (the one above) can be expressed as $ax + by + cz + d = 0$, where $d = -ax_0 - by_0 - cz_0$. This form of the equation is sometimes called the general form of the equation of a plane.

Solutions to linear systems

- **Possible solutions to a linear system of two unknowns**: The linear system can have a **unique solution, no solution, or infinitely many solutions**.

- **Does the solution set form a line, plane, hyperplane, or something else?**: The formation of the solution set depends on the number of free variables,

    - **No free variables (one unique solution)**: Intersects at a point
    - **One free variable (Uncountable solutions)**: Solution set is a line (1-dimensional subspace)
    - **Two free variable (Uncountable solutions)**: Solution set forms a plane (2-dimensional subspace)
    - **Three free variable (Uncountable solutions)**: Solution set is a three dimensional subspace (In $\mathbb{R}^3$ it would be the whole space)
    - k **free variables**: Solution set is a k-dimensional subspace in $\mathbb{R}^n$

    Note: A k-dimensional subspace in $\mathbb{R}^n$ means that the solution set spans a k-dimensional space within the n-dimensional ambient space $\mathbb{R}^n$.

- **Determine if three planes intersect at a unique point**: For this, we find all three normal vectors $\vec{\mathbf{n}}_1, \vec{\mathbf{n}}_2$, and $\vec{\mathbf{n}}_3$. Then we find the triple scalar product, that is

$$\vec{\mathbf{n}}_1 \cdot (\vec{\mathbf{n}}_2 \times \vec{\mathbf{n}}_3).$$

If this value is non-zero, we have intersection at a unique point. If the value is zero, we either have no intersection, or intersection at a line.

Linearity

- **The properties of linear equations**: A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ representing a linear equation is linear, meaning it satisfies the following properties for all vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and all scalars $c \in \mathbb{R}$:

    - **Additivity:** $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$
    - **Homogeneity of Degree 1:** $f(c\mathbf{x}) = cf(\mathbf{x})$

        It follows from this that $f(c\mathbf{x})$, when $c = 0$ implies $f(0\mathbf{x}) = 0f(\mathbf{x}) = 0$. Thus, we add the property
    - **Scale by zero**: $f(0) = 0$

These properties define a linear function and imply that the graph of a linear equation is a straight line (in 2D) or a plane (in 3D).

Matrix algebra

- **Laws of matrix addition**:

    - **Addition with the zero matrix**: $0 + A = A$
    - **Communitive law for matrix addition**: $A + B = B + A$
    - **Associativity of matrix addition**: $(A + B) + C = A + (B + C)$

- **Laws of matrix subtraction**:

    - $A - 0 = A$
    - $A - A = 0$
    - $B - A = (-1)(A - B)$

- **Matrix difference (subtraction)**: We can give a definition to the subtraction operator by just defining it as using matrix addition and multiplication by a scalar $A - B = A + (-1B)$

- **Note on matrix multiplication**: Matrix multiplication is general **not** communitive, it can be, but it isn't always. Also, in the real numbers, we know for

$$ab = 0.$$

    Then either a is zero, b is zero, or they are both zero. This is not always the case with matrix multiplication, it is possible to multiply two non-zero matrices and get the zero matrix as a result.

- **Properties of matrix multiplication**:

    1. If A, B, and C are matrices of the appropriate sizes, then

    $$A(BC) = (AB)C.$$

    2. If A, B, and C are matrices of the appropriate sizes, then

    $$(A + B)C = AC + BC.$$

    3. If A, B, and C are matrices of the appropriate sizes, then

    $$C(A + B) = CA + CB.$$

- **Properties of Scalar Multiplication**: If r and s are real numbers and A and B are matrices of the appropriate sizes, then

    1. $r(sA) = (rs)A$
    2. $(r + s)A = rA + sA$
    3. $r(A + B) = rA + rB$
    4. $A(rB) = r(AB) = (rA)B$

- **Note on cancellation**: If a, b, and c are real numbers for which $ab = ac$ and $a \neq 0$, it follows that $b = c$. That is, we can cancel out the nonzero factor a. However, the cancellation law does not hold for matrices.

- **Differences between matrix multiplication and multiplication of real numbers**: We summarize some of the differences between matrix multiplication and the multiplication of real numbers as follows: For matrices A, B, and C of the appropriate sizes,

    1. AB need not equal BA.
    2. AB may be the zero matrix with $A \neq 0$ and $B \neq 0$.
    3. AB may equal AC with $B \neq C$.

Transpose

- **Squared magnitude of a vector**:

$$\|x\|^2 = x^\top x$$

- **Transpose of product of matrices**:

$$(AB)^\top = B^\top A^\top$$

  Consequence:

$$(ABC)^\top = C^\top (AB)^\top = C^\top B^\top A^\top$$

- **Properties of Transpose**: If r is a scalar and A and B are matrices of the appropriate sizes, then

  1. $(A^T)^T = A$
  2. $(A + B)^T = A^T + B^T$
  3. $(AB)^T = B^T A^T$
  4. $(rA)^T = rA^T$

Linear maps

- **Composition**: Let $L : \mathbb{R}^n \to \mathbb{R}^m, \quad v \to L(v)$, and $K : \mathbb{R}^m \to \mathbb{R}^p, \quad L(v) \to K(L(v))$. We see that $L \in \mathbb{R}^{m \times n}$, and $K \in \mathbb{R}^{p \times m}$.

  The composition is

  $$K(L(v)) = (K \circ L)(v) = (KL)(v)$$

- **2D rotation map**:

  $$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

- **3D rotation, but keeping one variable constant, ie rotating about one of the coordinate axis**.

  All there cases below will require a 3×3 matrix

  - **Rotation about the x-axis (rotation in the yz-plane)**:

    $$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin() & \cos(\theta) \end{bmatrix}.$$

  - **Rotation about the y-axis (rotation in the xz-plane)**

    $$R_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}.$$

  - **Rotation about the z-axis (rotation in the xy-plane)**

    $$R_z(\theta) \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

  Notes on consecutive rotations: Two consecutive rotations about different axises is **not** communitive, however if you rotation about the same axis it is.

- **What are the columns of a linear map**: If a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is represented by a matrix A, then the columns of A are exactly the vectors you obtain by applying T to the standard basis vectors $e_1, e_2, \ldots, e_n$. That is,

  $$T(e_j) = Ae_j = \text{column } j \text{ of } A.$$

- **Linear Operations**: A linear map from a vector space to itself, $L : V \to V$ is called a *linear operation*, or just an *operation*

Injective, surjective, bijective

- **Standard Functions**:

  - A function $f : A \to B$ is **surjective (onto)** if for all $b \in B$, there exists an $a \in A$ such that $f(a) = b$. Equivalently, f is surjective if $f(A) = B$.
  - A function $f : A \to B$ is **injective (one-to-one)** if

  $$f(a_1) = f(a_2) \implies a_1 = a_2.$$

    Equivalently, distinct elements of A map to distinct elements of B.

  - A function $f : A \to B$ is **bijective** if it is both injective and surjective. In this case, f is called a **bijection** from A to B, and f has an **inverse function** $f^{-1} : B \to A$.
  - A function $f : A \to B$ is **constant** if there exists $b_0 \in B$ such that $f(a) = b_0$ for all $a \in A$.
  - A function $f : A \to B$ is **identity** on A if $f(a) = a$ for all $a \in A$. This function is usually denoted by id_A.
  - A function $f : A \to B$ is **many-to-one** if two or more distinct elements in A can map to the same element of B.
  - A function $f : A \to B$ is **one-to-one correspondence** if it is bijective. In this case, A and B have the same cardinality.

- **Linear maps**: Consider a linear map $L : \mathbb{R}^n \to \mathbb{R}^k$

  - $\mathbb{R}^n \to \mathbb{R}^k$ where $n < k$ can never be onto, maybe 1-1 though
  - $\mathbb{R}^n \to \mathbb{R}^k$ where $n > k$ can never be 1-1, maybe onto though

  Consider a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$, with $A \in \mathbb{R}^{m \times n}$. By the rank-nullity theorem,

  $$\dim(\mathbb{R}^n) = \dim(\mathrm{Im}(L)) + \dim(\ker(L)).$$

  So, $n = \dim(\mathrm{col}(A)) + \dim(\ker(L))$. If a matrix is wide ($m > n$), then a matrix has full rank if $\mathrm{rank}(A) = n = \dim(\mathrm{Im}(L))$. If A has full rank, then

  $$n = n + \dim(\ker(L)).$$

  So, $\dim(\ker(L)) = 0$, so L is injective. But, $\dim(\mathrm{Im}(L)) = n \neq m$. Since m is the dimension of the codomain, which is different from the dimension of the image, L is not surjective.

  Consider a map where $m < n$. If the matrix representing the map has full rank, then $\mathrm{rank}(A) = m = \dim(\mathrm{Im}(L))$. Thus,

  $$n = m + \dim(\ker(L)).$$

  So, $\dim(\ker(L)) \geqslant 1$, and L is not injective. But, L is surjective since the dimension of the image is equal to the dimension of the codomain, so the image of the map is the same space as the codomain.

- **Surjective and bijective in terms of matrix transformations**:

  If a matrix A represents a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$, the transformation is onto if for every vector $b \in \mathbb{R}^m$ in the codomain, there exists at least one vector $x \in \mathbb{R}^n$ such that $Ax = b$.

Geometric Interpretation: Surjectivity means the matrix transformation "covers" the entire codomain, hitting every possible point. In 2D, this would mean the entire plane is covered by the transformation.

A matrix A is injective if for every x_1 and x_2 in the domain, if $Ax_1 = Ax_2$, then $x_1 = x_2$. This means that no two distinct input vectors can be mapped to the same output vector.

Geometric Interpretation: Injectivity means the transformation doesn't collapse any part of the domain into a lower dimension, so no information is lost.

- **If dimensions are the same**: If $\mathbb{R}^n \to \mathbb{R}^n$, if its onto, its most likely 1-1.

- **Check if a linear map is injective or surjective (Formal)**: To check whether a linear transformation is surjective or injective, we use specific properties of the matrix representing the transformation.

Let $T : V \to W$ be a linear transformation, where V and W are vector spaces and A is the matrix representation of T

  - **Checking Injectivity (One-to-One)**: A linear transformation T is injective (one-to-one) if:

  $$T(\mathbf{v_1}) = T(\mathbf{v_2}) \implies \mathbf{v_1} = \mathbf{v_2}.$$

  Equivalently, T is injective if the only solution to $T(\mathbf{v}) = 0$ (the null space or kernel of T) is $\mathbf{v} = 0$.

  Alternatively, a matrix A is injective if the rank of the matrix (the number of linearly independent columns) is equal to the number of columns of the matrix. This means the matrix has full column rank

  Lastly, we can just check the determinant if the matrix is square.

  - **Check surjectivity**: A linear transformation T is surjective (onto) if for every vector $\mathbf{w} \in W$ (the codomain), there exists a vector $\mathbf{v} \in V$ (the domain) such that:
  $$T(\mathbf{v}) = \mathbf{w}.$$
  This means that T "covers" the entire codomain W, or in other words, the image of T is the entire space W.

  Steps to check surjectivity:
  1. **Image:** The transformation T is surjective if the image (column space or range) of the matrix A spans the entire codomain W. This means

  $$\text{im}(A) = W.$$

  2. **Rank:** For surjectivity, the rank of the matrix must be equal to the dimension of the codomain. If A is an $m \times n$ matrix, the matrix is surjective if its rank is equal to m (the number of rows).

  3. **Determinant for square matrices:** If A is a square matrix, T is surjective if

  $$\det(A) \neq 0,$$

  because a non-zero determinant implies the matrix has full rank, covering the entire codomain.

- **Injective and surjective for square matrices**: For square matrices (say $A \in \mathbb{R}^{n \times n}$), injectivity and surjectivity are actually equivalent, since the number of linearly independent rows equals the number of linearly independent columns which both equal n if the matrix has full rank.

Matrix rank

- **Row-echelon form**: Row echelon form (REF) is a standardized way of arranging the rows of a matrix using elementary row operations. A matrix is in row echelon form if it satisfies these conditions:

  - All zero rows are at the bottom. (If a row is entirely zeros, it appears below any nonzero row.)
  - The first nonzero entry in each nonzero row is 1 (called a leading 1 or pivot).
  - Each leading 1 is to the right of the leading 1 in the row above it. (So the pivots "stair-step" down to the right as you move down the rows.)
  - All entries below a pivot are zero.

  The matrix $A \in \mathbb{R}^{4 \times 4}$

  $$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

  is in row-echelon form

- A non square matrix cannot be invertible.

- **Rank of a square matrix**: For an $n \times n$ square matrix, the **rank** is the dimension of its row space (or column space—they are equal).

  If $\text{rank}(A) = n$, the matrix is **full rank** and invertible.

  - The linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ is:
    * **Injective** (one-to-one): no two distinct inputs map to the same output.
    * **Surjective** (onto): every element of $\mathbb{R}^n$ is hit by some input.
    * Together, injective + surjective = **bijective**, meaning A is invertible.

  Note: If $A \in \mathbb{R}^{n \times n}$ has full rank, then A is nonsingular and $\det(A) \neq 0$

  If $\text{rank}(A) < n$, then:

  - The matrix is singular (non-invertible).
  - The map is neither injective (kernel is nontrivial) nor surjective (image is a proper subspace).

  Note: If A is rank-deficient, then A is singular and $\det(A) = 0$.

- **Rank of a non-square matrix**: For an $m \times n$ matrix A, the **rank** is the dimension of the row space (or column space).

  $$\text{rank}(A) \leqslant \min(m, n).$$

  Consider the map $A : \mathbb{R}^n \to \mathbb{R}^m$:

  - If $m > n$ (tall matrix):
    * The maximum rank is n.

* If $\text{rank}(A) = n$, then the map is **injective** (no two inputs give the same output), but not surjective, since the codomain has higher dimension than the image.
  - If $m < n$ (wide matrix):
    * The maximum rank is m.
    * If $\text{rank}(A) = m$, then the map is **surjective** (fills all of $\mathbb{R}^m$), but not injective, since extra input dimensions collapse.

- **Determine the rank of a square matrix**: We check the determinant. If $\det(A) \neq 0$, the matrix A has full rank and is nonsingular. If $\det(A) = 0$, the matrix A is rank-deficient and is singular.

- **Determine the rank of a non-square matrix**: To determine the rank of a matrix, we need to make sure the $m \times n$ matrix has m linearly independent rows. To do this, we can perform guassian elimination to get the augmented matrix in row echelon form. The rank of the matrix, $\text{rank}(A)$ is the number of non zero rows.

- **Column rank and row rank**: The number of linearly independent columns is called the column rank, and the number of linearly independent rows is called the row rank.

  For any matrix A, the number of linearly independent rows is equal to the number of linearly independent columns. This number is called the rank of A

- **What does the rank tell use about solutions**: If we have full rank, then given a target, there will be a unique solution. If we are rank-deficient, there may be no solution, or infinitely many soluitions.

- **Rank theorem**: Let A be a matrix that represents a linear map L, the rank of the matrix is given by

$$\text{Rank}(A) = \dim(\text{Im}(L)).$$

Furthermore

$$\text{Rank}(A) = \text{Rank}(A^T).$$

- **Solutions of a tall matrix**: Suppose $L : \mathbb{R}^n \to \mathbb{R}^m$ is represented by $A \in \mathbb{R}^{m \times n}$, with $m > n$. If $Ax = b$ has a unique solution for a given b, then two things must be true

  1. $b \in \text{Im}(L)$
  2. $\text{Rank}(A) = n$ (full column rank)

If A has full column rank, then L is injective but not surjective, so $Ax = 0$ implies $x = 0$, and $\text{Im}(L) \subsetneq \mathbb{R}^m$.

Suppose that A was rank deficient. Then, there are multiple members of $\mathbb{R}^n$ that map to zero. If x_0 is a solution to $Ax = b$, then $x_0 + v$, for $v \in \text{Ker}(L)$ also solves $Ax = b$, since

$$A(x_0 + v) = b \implies Ax_0 + Av = b \implies Ax_0 + 0 = b \implies Ax_0 = b.$$

Thus, infinitely many solutions, since $\text{rank}(A) < n$ implies $\dim(\text{ker})(L) \geqslant 1$, so there are infinitely many vectors in the Kernel.

- **Solutions of a wide matrix**: Suppose that $L : \mathbb{R}^n \to \mathbb{R}^m$ is represented by a matrix $A \in \mathbb{R}^{m \times n}$, where $m < n$. In this case, L can be surjective (if $\text{rank}(A) = m$), but not injective. Thus, the only possibilities are no solution, or infinitely many solutions.

If A is rank deficient, then L is not surjective, so $\text{Im}(L) \subsetneq \mathbb{R}^m$. In this case, if $b \notin \text{Im}(L)$, no solution exists. If A has full rank, then $\text{Im}(L) = \mathbb{R}^m$. So, for any $b \in \mathbb{R}^m$, there are infinitely many solutions. Suppose that x_0 solves $Ax = b$. Then, $x_0 + v$, for $v \in \ker(L)$ also solves $Ax = b$, since

$$A(x_0 + v) = b \implies Ax_0 + Av = b \implies Ax_0 + 0 = b \implies Ax_0 = b.$$

- **Solutions of a square matrix**: Suppose that $L : \mathbb{R}^n \to \mathbb{R}^n$ is represented by a matrix $A \in \mathbb{R}^{n \times n}$. If $\det(A) \neq 0$, then A is both injective and surjective, and thus bijective. In this case, there is a unique solution for any given vector $b \in \mathbb{R}^n$.

  If $\det(A) = 0$, A is neither surjective nor injective. In this case, for a given b, there is either no solution, or infinitely many solutions.

  1. If $b \notin \text{Im}(L)$, no solution.
  2. If $b \in \text{Im}(L)$, infinitely many solutions.

- **Rank with rank-nullity theorem**: For $A \in \mathbb{R}^{m \times n}$, using the rank-nullity theorem, we have

$$\text{rank}(A) = n - \dim(\ker(A)).$$

So, we see that the rank of a matrix measures how many dimensions survive under the linear map.

Linear dependence, span, intro to fundamental spaces

- **Linear dependence**: Linear dependence tells us that we will lose information because it implies that some vectors (or columns/rows of a matrix) do not add any new, unique directions to the space. These dependent vectors can be expressed as combinations of other vectors, meaning they don't span new dimensions, and as a result, the transformation collapses part of the input space into a lower-dimensional output space.

- **Linear independence**: No redundancy, all directions are unique

  A set of vectors $\{v_1, v_2, \ldots, v_k\}$ in a vector space is linearly independent if no vector in the set can be written as a linear combination of the others.

  Equivalently, the only solution to the equation

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

  is $c_1 = c_2 = \cdots = c_k = 0$, where c_i are scalars.

- **Span**: The span of a set of vectors is the collection of all possible linear combinations of those vectors

  if a matrix has full rank, the span of its columns (or rows) is the entire codomain (or the entire vector space that the matrix maps to)

  The vectors in a linearly independent set span a space whose dimension equals the number of vectors in the set.

  If a set contains more vectors than the dimension of the vector space, the set cannot be linearly independent (e.g., in $\mathbb{R}^n$, any set with more than n vectors is dependent).

  If a set contains fewer vectors than the dimension of the vector space, it cannot span the entire space.

- **Prove that a set of vectors is linearly dependent**: To show that a set of vectors is linearly independent, you need to prove that the only solution to the linear combination equating the set to the zero vector is the trivial solution

  For a set of vectors $\{v_1, v_2, \ldots, v_k\}$, consider the equation:

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

  where $c_1, c_2, \ldots, c_k$ are scalars (unknowns).

  Place the vectors as columns in a matrix $A = [v_1 \ v_2 \ \ldots \ v_k]$.

  The equation then becomes
  $$Ac = 0$$
  where $c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$ is the column vector of scalars.

Solve the homogeneous system
$$Ac = 0.$$
If the only solution is
$$c_1 = c_2 = \cdots = c_k = 0$$
(the trivial solution), then the vectors are linearly independent.

If the set of vectors forms a square matrix (i.e., the number of vectors equals their dimension), compute the determinant of the matrix.

If $\det(A) \neq 0$, the vectors are linearly independent. If $\det(A) = 0$, they are dependent.

- **Geometric Interpretation**: In $\mathbb{R}^2$, two linearly independent vectors are not collinear (do not lie on the same line through the origin).

  In $\mathbb{R}^3$, three linearly independent vectors are not coplanar (do not lie in the same plane through the origin).

- **Intro to row space, column space, null space (kernel)**:

  - **Column space** $(C(A))$: The set of all linear combinations of the columns of a matrix A. It represents the range of the matrix and consists of all possible outputs of the matrix transformation. It's a subspace of $\mathbb{R}^m$ (for an $m \times n$ matrix)
  - **Row space** $(R(A))$: The set of all linear combinations of the rows of a matrix A. It is the span of the rows of the matrix and forms a subspace of $\mathbb{R}^n$
  - **Kernel/Null space** $(\ker(A)/N(A))$: The set of all vectors $\mathbf{x}$ such that $A\mathbf{x} = 0$, where A is a matrix. It represents the solutions to the homogeneous system and is a subspace of $\mathbb{R}^n$

  Each of these spaces relates to the structure and solvability of linear systems and the transformation properties of the matrix.

- **Orthogonality**: If vectors are orthogonal (mutually perpendicular), they are linearly independent.

  However, linearly independent vectors do not need to be orthogonal.

- **Linear combination of vectors**: A linear combination of vectors is a sum of those vectors, each multiplied by a scalar. So, for a matrix A, if $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ are its column vectors, any vector in the column space can be written as:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n.$$

Where $c_1, c_2, ...c_n$ are scalars

As you vary the scalars in the linear combination of the matrix's columns, you generate all possible vectors in the column space (also known as the range) of the matrix.

Note: The same applies for the row space

- **Row space vs column space vs null space**:

    - The column space tells you what the matrix outputs.
    - The row space tells you about the constraints or conditions that the solutions to the matrix system must satisfy.
    - The null space tells you what the matrix "loses". If a vector $\mathbf{x}$ is in the null space, it gets mapped to the zero vector, meaning it is "annihilated" by the matrix. Vectors in the null space represent dependencies between the columns of the matrix. If the matrix has non-trivial solutions to $A\mathbf{x} = 0$, it indicates that the columns are linearly dependent.

      Geometrically, the null space represents all the directions in which the matrix compresses space to a lower dimension. For example, in $\mathbb{R}^3$, if the null space is a line, the matrix compresses all points along that line to the origin.

      If the null space is non-trivial, it indicates the matrix transformation has lost some dimensions.

      Note: A non-trivial null space refers to a null space that contains vectors other than just the zero vector.

- **More on the row space**: In a system of linear equations, the row space reveals various types of constraints, depending on the number of independent equations and how the rows of the matrix relate to one another.

    - **Unique Solution (Full Rank, Independent Rows)**: If the rows of the matrix are linearly independent and span the entire row space, the system has a unique solution. This means the constraints from the equations are sufficient to pin down exactly one solution.
    - **No Solution (Inconsistent System):**
    - **Infinite Solutions (Dependent Rows, Underdetermined System):** If some rows are linearly dependent, the system will have fewer constraints than unknowns, leading to infinitely many solutions. In this case, the system is underdetermined, meaning there aren't enough independent constraints to specify a unique solution, allowing multiple solutions (often forming a plane, line, or higher-dimensional space).
    - **Zero Solutions (Trivial System):** In a homogeneous system , if the rows are independent but fewer than the number of variables, there is only the trivial solution. This means that the row space spans a subspace of dimension less than the total space, so the only solution is the zero vector.

- **Linear dependence in rows vs columns**: The rows of a matrix are linearly dependent if at least one row can be written as a linear combination of the other rows.

  This means there is redundancy in the information that the rows provide.

  If the rows of a matrix are linearly dependent, the row space has a lower dimension than the number of rows, meaning the system has fewer independent constraints than it might appear.

Consider the matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

Here, the second row is $2\times$ the first row, and the third row is $3\times$ the first row. Therefore, the rows are linearly dependent because each row is a scalar multiple of the first row.

This means there's only one independent constraint in the row space. The row space is spanned by a single vector (the first row), even though the matrix has three rows. Geometrically, the row space collapses to a lower dimension (a line in 3D space).

If the rows are linearly dependent, the system of equations might be underdetermined, leading to infinite solutions or no solutions.

The columns of a matrix are linearly dependent if at least one column can be written as a linear combination of the other columns.

This implies that some columns do not contribute "new" information, and there is a loss of dimensionality in the column space.

If the columns are linearly dependent, the column space has a lower dimension than the number of columns, meaning the matrix cannot map onto all of $\mathbb{R}^m$ (the output space).

Consider the matrix

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

Here, the second and third columns are linearly dependent on the first column (they are multiples of the same vector). In fact, each column is identical in this case, so all the columns are linearly dependent.

This means the column space of this matrix is spanned by a single vector, even though there are three columns.

Geometrically, the matrix can only map vectors to a line in $\mathbb{R}^3$, rather than a full plane or space.

Summary:

– **Row dependence** affects the row space, which corresponds to the constraints on the system of equations. Linearly dependent rows mean some equations are redundant, reducing the number of independent constraints.

– **Column dependence** affects the column space, which corresponds to the set of possible outputs of the matrix. Linearly dependent columns mean the matrix has reduced rank, implying the system might not span the full space, and it could have a non-trivial null space (leading to infinite or no solutions).

Recall: To check if a linear map has full rank, it is sufficient to check whether all the columns of the matrix representing the linear map are linearly independent

Important: The dimension of both the row space and the column space of a matrix is equal to the number of linearly independent rows and linearly independent columns, respectively. This common dimension is called the rank of the matrix.

- **Rank of the null space**: The null space of a matrix has dimension $n-$rank, where n is the number of columns in the matrix. This is a consequence of the Rank-Nullity Theorem (Not yet stated).

- **Nullity**: The nullity of a matrix is the dimension of the null space (i.e., the number of independent vectors that get mapped to the zero vector by the matrix).

- **Zero nullity**: If the nullity is zero, this means that the null space has dimension 0. This implies that the only vector in the null space is the zero vector itself, **0**

- **non-zero nullity**: We know that there are infinitely many vectors in the kernel (null space) of a matrix when the nullity (dimension of the null space) is greater than zero.

Row space, Column space, Null space, and the Rank Nullity Theorem

- **Null space (Kernel)**: Suppose $L : V \to W$ is a linear map between vector spaces V and W. Then the kernel of L is defined as

$$\mathrm{Ker}(L) = \{v \in V : L(v) = 0\} \subseteq V$$

Note that the kernel is a vector space.

- **Image**: Suppose $L : V \to W$ is a linear map between vector spaces V and W. Then, the image of L, also called the range of L is the subspace

$$\mathrm{Im}(L) = \{w \in W : w = L(v), \ v \in V\} \subseteq W$$

In other words, all vectors $w \in W$ such that w is the image of a vector in v. If L is onto, then every vector in the codomain W is the image of a vector in v, thus R_L will be all vectors in W. If W has dimension n and W is onto, then $\dim(R_L) = \dim(W)$. If L is not onto, then $\dim(R_L) \leqslant \dim(w)$

- **Column space**: If $L : V \to W$ is a linear map between vector spaces V and W, and L is represented by a matrix $A \in \mathbb{R}^{m \times n}$, then the **column space** of A, denoted $\mathrm{Col}(A)$ is precisely the Image of L. That is,

$$\mathrm{Col}(A) = \mathrm{Im}(L)$$

The column space of A is the set of all linear combinations of the columns of A. If A has columns $c_1, c_2, ..., c_n$, then $\mathrm{Col}(A)$ is

$$\mathrm{Col}(A) = \{w \in W : w = s_1 c_1 + s_2 c_2 + ... + s_n c_n\}$$

i.e it is the span of the columns of A

- **Row space**: Let $L : V \to W$ be a linear map between vector spaces V and W. If L is represented by a matrix $A \in \mathbb{R}^{m \times n}$.

The row space of A is the span of the rows of A. The row space of A is defined as

$$\mathrm{Row}(A) = \mathrm{Col}(A^\top) = \mathrm{Im}(L^\top)$$

Where $L^\top$ is the transpose map $L^\top : W^* \to V^*$

- **The orthogonal complement of the kernel is the row space**: For a linear map $L : V \to W$ between vector spaces V and W, we have that

$$\mathrm{Ker}(L)^\perp = \mathrm{Im}(L^\top) = \mathrm{Row}(A)$$

Where A is the matrix that represents L.

- **Rank nullity theorem**: The **Rank-Nullity Theorem** states that for a linear map $L : V \to W$ between two vector spaces V and W, the dimension of the domain V is the sum of the rank and the nullity of L. Mathematically, it is expressed as:

$$\dim(V) = \dim(\ker(L)) + \dim(\mathrm{Im}(L))$$

Where:

  - $\dim(V)$ is the dimension of the domain vector space V,

- dim(ker(L)) is the **nullity** of L, i.e., the dimension of the **null space** (the set of vectors in V that map to the zero vector in W),

- dim(im(L)) is the **rank** of L, i.e., the dimension of the **image** (the set of all vectors in W that are the image of some vector in V).

In simpler terms, the dimension of the vector space V is the sum of the number of independent vectors that are mapped to zero and the number of independent vectors that are mapped to non-zero vectors.

Thus, the dimensions of the kernel and image of a linear map must sum to the dimension of the domain.

- **Basis for the column space**: We can find a description of the image by row reducing the matrix associated with a linear map. The number of leading zeros in the rows can give us not only the dimension of the image, but also a basis by using the columns from the original matrix where the leading ones appear in the row reduced matrix. Consider an example.

Suppose we have the matrix

$$A = \begin{pmatrix} 1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 1 & 3 & 1 \\ 3 & 6 & 2 & 5 & 1 \end{pmatrix}.$$

Which represents a linear map $L : \mathbb{R}^5 \to \mathbb{R}^4$ with respect to the standard basis. Row reducing this matrix yields

$$\begin{pmatrix} 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

As we can see, we have three leading ones, in columns 1,3, and 5. Thus, the dimension of the image is three. Furthermore, we can use these columns from the original matrix as a basis for the image. Thus,

$$\begin{pmatrix} 1 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Is a basis for the image. The other columns are linearly dependent on these three vectors.

Note: If the matrix is surjective, the standard basis forms a basis for the column space.

If the matrix is injective, the kernel has no basis.

Inverses

- **Matrix inverse**: A matrix $B \in \mathbb{R}^{n \times n}$ is the inverse of a matrix $A \in \mathbb{R}^{n \times n}$ provided

$$BA = AB = I$$

- **Left and right inverse**: **Left Inverse:** A matrix B is a left inverse of A if $BA = I$, where I is the identity matrix. This means B "undoes" A when multiplied from the left.

  Right Inverse: A matrix C is a right inverse of A if $AC = I$. This means C "undoes" A when multiplied from the right.

- **(Finite dimensions) One sided inverse theorem for matrices (inveribility criterion)**: It states that for square matrices in finite dimensions, the existence of either a left inverse or a right inverse implies the existence of the other, and hence the matrix is invertible. Specifically, if a square matrix A has a left inverse B (such that $BA = I$) or a right inverse C (such that $AC = I$), then both inverses must be equal, making A invertible with the unique inverse $A^{-1} = B = C$.

- **(Finite dimensions) Inverse Uniqueness Theorem for matrices**: If A is an invertible $n \times n$ matrix, then there is exactly one matrix B such that $AB = BA = I$. This unique matrix is called the inverse of A and is denoted by A^{-1}.

- **Solving $Ax = b$ for $A \in \mathbb{R}^{n \times n}$**: Assume $A \in \mathbb{R}^{n \times}$ is nonsingular, then

$$Ax = b \implies x = A^{-1}b$$

- **Matrix products**: If $A, B \in \mathbb{R}^{n \times n}$, and AB nonsingular, then A and B are nonsingular.

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

Determinant

- **Determinant of the identity matrix**: The determinant of the identity matrix is one

$$\det(I) = 1.$$

- **Product of two determinants**: The product of two determinants $\det(A)\det(B) = \det(AB)$. This is particularly useful when we want to find the determinant of a product of two matrices, we can assert, for two matrices A, and B

$$\det(AB) = \det(A)\det(B).$$

- **Determinant of the inverse matrix**: If we have some matrix A, which has an inverse A^{-1}, then

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

- **Determinant of triangular matrices**: The determinant of a triangular matrix is the product of the main diagonal.

- **Determinant of the transpose**: The determinant of the transpose of a matrix is the same as the original matrix.

$$\det(A^\top) = \det(A)$$

- **Determinant after changing basis**: Suppose we have some matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If we introduce a new basis matrix B, then the map becomes $B^{-1}AB$. And the deterimanant is

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

Thus, the determinant remains the same. We say that the determinant is invariant under a change of basis.

Other matrix properties

- **The trace of a matrix**: The trace of a square matrix is the sum of its diagonal elements. For an $n \times n$ matrix A, the trace is defined as:

$$\text{Tr}(A) = \sum_{i=1}^{n} A_{ii}$$

where A_{ii} are the diagonal entries of A. The trace is only defined for square matrices and has several useful properties, such as being invariant under a change of basis.

- **Cyclic property of the trace**: The trace is invariant under cyclic permutations. Observe

$$\text{Tr}(AB) = \text{Tr}(BA).$$

Proof:

$$\text{Tr}(AB) = \sum_{i=1}^{n} (ab)_{ii}$$
$$= \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} b_{ki}.$$

Remark. For finite sums, the order of summation can be swapped because summation is commutative and associative when dealing with real or complex numbers. This means that the sum of a collection of terms does not depend on the order in which the terms are added. So, rearranging the summation over i and k doesn't change the value of the overall sum. Thus

$$\text{Tr}(AB) = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} b_{ki}$$
$$= \sum_{k=1}^{n} \sum_{i=1}^{n} a_{ik} b_{ki}$$
$$= \sum_{k=1}^{n} \sum_{i=1}^{n} b_{ki} a_{ik}$$
$$= \sum_{k=1}^{n} (ba)_{kk}$$
$$= \text{Tr}(BA).$$

Eigenvalues and Eigenvectors

- **Eigenvectors, Eigenvalues**: An **eigenvector** of a square matrix A is a non-zero vector $\mathbf{v}$ such that when A acts on $\mathbf{v}$, the result is a scalar multiple of $\mathbf{v}$. Mathematically, this is written as:

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

  where λ is a scalar known as the eigenvalue corresponding to the eigenvector $\mathbf{v}$.

  An **eigenvalue** λ is the scalar that represents the factor by which the eigenvector is scaled during the transformation. The eigenvalue corresponds to each eigenvector and provides information about the nature of the transformation (scaling, rotation, etc.).

  Since the left side is matrix multiplication and the right side is vector multiplication by a scalar, we can rewrite the equation above as

$$A\mathbf{v} = (\lambda I)\mathbf{v}$$
$$\implies A\mathbf{v} - \lambda I\mathbf{v} = \mathbf{0}$$
$$\implies \mathbf{v}(A - \lambda I) = \mathbf{0}.$$

  This is a homogeneous system. If the transformation map $(A - \lambda I)$ is one-to-one and thus invertible, the only solution would be the trivial solution $(\mathbf{v} = \mathbf{0})$. In order to have non-zero solutions for $\mathbf{v}$ (eigenvectors), the system above would need to not be one-to-one (multiple solutions to the solution vector $\mathbf{0}$), and thus

$$\det(A - \lambda I) = 0.$$

- **Characteristic equation, characteristic polynomial**: The characteristic equation of a square matrix A is the equation obtained by setting the determinant of $A - \lambda I$ equal to zero:

$$\det(A - \lambda I) = 0,$$

  where λ represents the eigenvalues of A and I is the identity matrix. Solving this equation gives the eigenvalues of A. The characteristic polynomial is the polynomial in λ obtained from the determinant $\det(A - \lambda I)$. It is typically expressed as:

$$p(\lambda) = \det(A - \lambda I),$$

  and its roots are the eigenvalues of the matrix A.

- **Finding eigenvectors and eigenvalues**: To find the eigenvalues λ, we need to solve the characteristic equation:

$$\det(A - \lambda I) = 0.$$

  This equation determines the values of λ for which the matrix $A - \lambda I$ is singular (non-invertible), which leads to non-zero solutions for $\mathbf{v}$.

  Once the eigenvalues λ are found, substitute each λ into the equation $(A - \lambda I)\mathbf{v} = 0$ and solve for $\mathbf{v}$. These are the eigenvectors corresponding to each eigenvalue.

- **Repeated eigenvalues (multiplicity)**: A repeated eigenvalue is an eigenvalue that appears more than once when you solve

$$\det(A - \lambda I) = 0.$$

If $(\lambda - \lambda_i)^k$ is a factor of the characteristic polynomial, then λ_i is a repeated eigenvalue with algebraic multiplicity k.

The notation for algebraic multiplicity is

$$\text{mult}_a(\lambda) = m_a(\lambda).$$

And the notation for geometric multiplicity is

$$\text{mult}_g(\lambda) = m_g(\lambda).$$

The geometric multiplicity is the dimension of the eigenspace for a particular eigenvalue:

$$m_g(\lambda_i) = \dim(\ker(A - \lambda_i I)).$$

For an eigenvalue λ of A, the eigenspace for λ is

$$E_\lambda = \{x \neq 0 : Ax = \lambda x\}.$$

Since a member of the space satisfies

$$Ax = \lambda x.$$

We can rearrange to get

$$Ax - \lambda x = (A - \lambda I)x = 0.$$

Thus, the members of the space E_λ are precisely the vectors in the kernel of $A - \lambda I$. This is why the dimension of the space is the dimension of the kernel of $A - \lambda I$.

If the characteristic polynomial is

$$(\lambda - 2)^3(\lambda + 1),$$

then

$$m_a(2) = 3, \ m_a(-1) = 1.$$

If the dimension of the eigenspace for $\lambda = 2$ is 2, then

$$m_g(2) = 2.$$

- **More on geometric multiplicity**: Geometric multiplicity of an eigenvalue quantifies the dimension of the eigenspace associated with that eigenvalue. It measures how many linearly independent eigenvectors correspond to the same eigenvalue.

  Suppose that $A \in \mathbb{F}^{n \times n}$ and let λ be an eigenvalue of A. The geometric multiplicity of λ is

  $$\dim(E_\lambda) = \dim(\ker(A - \lambda I)).$$

  The eigenspace E_λ is a linear subspace of the ambient space.

    - **Geometric multiplicity $= 1$:** the eigenspace is a line through the origin.
    - **Geometric multiplicity $=2$:** the eigenspace is a plane through the origin.
    - Higher values correspond to higher-dimensional invariant subspaces.

  For any eigenvalue λ,

  $$\lambda \leqslant m_g(\lambda) \leqslant m_a(\lambda).$$

  An eigenvalue always has at least one eigenvector. If geometric multiplicity is strictly smaller than algebraic multiplicity, the matrix is defective.

- **The set of eigenvalues (The spectrum)**: The set of eigenvalues of a matrix A is denoted by

$$\text{spec}(A) = \sigma(A) = \{\lambda \in \mathbb{C} : \det(A - \lambda I) = 0\}.$$

This is called the **Spectrum** of A. Since this set does not show repeated eigenvalues, we can list the eigenvalues

$$\lambda_1, \lambda_2, \ldots, \lambda_n.$$

- **The spectrum of the transpose**: Consider a matrix A, the spectrum of A is

$$\sigma(A) = \{\lambda \in \mathbb{C} : \det(A - \lambda I) = 0\}.$$

Now, consider A^T,

$$\sigma(A^T) = \{\lambda \in \mathbb{C} : \det(A^T - \lambda I) = 0\}.$$

But,

$$\det(A^T - \lambda I) = \det((A - \lambda I)^T).$$

Since the determinant of the transpose of a matrix is identical to the determinant of the matrix,

$$\det(A^T - \lambda I) = \det((A - \lambda I)^T) = \det(A - \lambda I).$$

Thus,

$$\sigma(A^T) = \sigma(A).$$

- **The spectrum of the inverse**: Consider a invertible matrix A, with spectrum

$$\sigma(A) = \{\lambda \in \mathbb{C} : \det(A - \lambda I) = 0\}.$$

Then, the spectrum of A^{-1} is the set

$$\sigma(A^{-1}) = \{\frac{1}{\lambda} : \lambda \in \sigma(A)\}.$$

Proof. Let λ be an eigenvalue for A. Then, for some $v \neq 0$

$$Av = \lambda v.$$

Multiply both sides by the inverse of A

$$A^{-1}Av = A^{-1}\lambda v$$
$$\implies v = A^{-1}\lambda v.$$

Now, divide both sides by λ,

$$A^{-1}v = \frac{1}{\lambda}v.$$

Thus, the eigenvalues for A^{-1} are the reciprocals of the eigenvalues of A, call the ith eigenvalue for A^{-1} μ_i. That is,

$$\mu_i = \frac{1}{\lambda_i(A)}.$$

Note: If the eigenvalues of A are listed in ascending order, $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$, then the ascending order for the eigenvalues of A^{-1} is

$$\frac{1}{\lambda_n} \geqslant \cdots \geqslant \frac{1}{\lambda_2} \geqslant \frac{1}{\lambda_1}.$$

- **Number of eigenvalues for a matrix**: Over the complex numbers, every $n \times n$ matrix has exactly n eigenvalues counted with multiplicity.

  This is guaranteed by the Fundamental Theorem of Algebra applied to the characteristic polynomial:

  $$\det(A - \lambda I).$$

  Over the real numbers, a real matrix may have fewer real eigenvalues because some eigenvalues may be complex.

- **Symmetric matrices**: If $A = A^T$ (or $A = A^*$ over $\mathbb{C}$), then

  1. All eigenvalues are real.
  2. There are exactly n real eigenvalues counting multiplicity.
  3. The matrix is diagonalizable by an orthogonal (or unitary) matrix.

- **Triangular matrices with real diagonal**: If a matrix is upper or lower triangular, then the eigenvalues are the entries on the diagonal . So if the diagonal entries are real, you get n real eigenvalues.

- **Note about eigenvalues**: Only square matrices have eigenvalues.

- **Spectral mapping**: Consider the spectrum of A,

  $$\sigma(A) = \{\lambda \in \mathbb{C} : \ \det(A - \lambda I) = 0\}.$$

  The spectrum of A^2, $\sigma(A^2)$ is the set

  $$\sigma(A^2) = \{\lambda^2 \in \mathbb{C} : \ \lambda \in \sigma(A)\}.$$

  Proof. Let λ be in the spectrum of A. Then for any nonzero vector $v \in \mathbb{R}^n$,

  $$Av = \lambda v.$$

  For A^2,

  $$A^2 v = AAv = A\lambda v = \lambda Av = \lambda\lambda v = \lambda^2 v.$$

  In fact, one can use this argument to show that

  $$\sigma(A^k) = \{\lambda^k \in \mathbb{C} : \ \lambda \in \sigma(A)\}.$$

- **Rank of a matrix and eigenvalues**: Let $A \in \mathbb{R}^{m \times n}$, and r be the rank of A, $r = \text{rank}(A)$. If zero is an eigenvalue, $\lambda = 0$, then

  $$Ax = \lambda x = 0.$$

  So, $\lambda = 0$ corresponds exactly to vectors $x \neq 0$ such that

  $$Ax = 0.$$

Thus,

$$\ker(A) \neq \{0\} \iff 0 \text{ is an eigenvalue.}$$

If zero is an eigenvalue, then

$$E_0 = \{x \in \mathbb{R}^n : Ax = 0\}.$$

Thus,

$$E_0 = \ker(A),$$

and

$$\dim(E_0) = \dim(\ker(A)).$$

Therefore, the rank of A is given by

$$\text{rank}(A) = n - \dim(\ker(A)) = n - \dim(E_0),$$

where $\dim(E_0)$ is the geometric multiplicity of $\lambda = 0$. So, we see that

$$m_g(0) = n - \text{rank}(A),$$

where $m_g(0)$ is the geometric multiplicity of $\lambda = 0$.

- **Number of eigenvalues revisited**: Let $A \in \mathbb{C}^{n \times n}$ with rank r. The number of nonzero eigenvalues is given by the rank r, while the remaining $n - r$ eigenvalues are zero.

Basis

- **Basis**: A basis of a vector space is a set of linearly independent vectors that span the entire space. This means any vector in the space can be uniquely expressed as a linear combination of the basis vectors. A basis provides a reference framework for representing vectors in that space.

  For an n-dimensional vector space, a basis will consist of exactly n vectors. The coordinates of a vector relative to a basis are the coefficients used in this linear combination.

- **Basis property**: Any set of n linearly independent vectors in an n-dimensional vector space forms a basis for that space. A basis is a minimal spanning set of vectors.

- **Standard basis (Implied basis)**: From vector calculus, we know that $\hat{i}$, $\hat{j}$ are the unit vectors the describe the 2 dimensional cartesian plane. Where

$$\hat{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\hat{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

  Thus, the standard implied basis when working in $\mathbb{R}^2$ is the matrix

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

  And a vector $\vec{\mathbf{v}} = \begin{bmatrix} x \\ y \end{bmatrix}$ can be repesented as scaling these basis vectors and then adding them. Ie

$$\begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

  Thus, the standard basis for $\mathbb{R}^n$ is the $n \times n$ identity matrix I

- **Basis notation**: When with vectors, the choice of basis determines how we interpret where the vector sits. For the standard basis, the componenets of the vector is precisely where it will be. Thus, for the standard basis, we write

$$[\vec{\mathbf{v}}].$$

  If the basis were non-standard, we would need to specify it. We write

$$[\vec{\mathbf{v}}]_B.$$

  Where B is then defined as the matrix representing the basis.

- **Change of basis explained**: We have

  - $B^{-1}v = v_B$
  - $Bv_B = v$

Suppose we have a vector $v = \begin{pmatrix} a \\ b \end{pmatrix}$ in the standard basis. Define a new basis $b_1 = \begin{pmatrix} u \\ v \end{pmatrix}, b_2 = \begin{pmatrix} x \\ y \end{pmatrix}$, then $B = \begin{pmatrix} u & x \\ v & y \end{pmatrix}$. Where v under this basis becomes $v_B = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. Then

$$\begin{pmatrix} a \\ b \end{pmatrix} = \alpha \begin{pmatrix} u \\ v \end{pmatrix} + \beta \begin{pmatrix} x \\ y \end{pmatrix}$$
$$= \begin{pmatrix} \alpha u + \beta x \\ \alpha v + \beta y \end{pmatrix}$$
$$= \begin{pmatrix} u & x \\ v & y \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$
$$\therefore v = Bv_B.$$

From this,

$$v = Bv_B$$
$$B^{-1}v = B^{-1}Bv_B$$
$$B^{-1}v = Iv_B$$
$$\therefore B^{-1}v = v_B.$$

Note: The new basis must span the same space as the old basis for the change of basis to work correctly

- **Basis in maps**: Suppose we have some linear map

$$L : V \to W$$

where V and W are vector spaces. Suppose V has a basis $\beta = \{b_1, \ldots, b_n\}$, and W has a basis $\gamma = \{c_1, \ldots, c_n\}$. Then, the matrix representation of L with respect to these bases is denoted as

$$[L]_\beta^\gamma$$

Now, suppose we define new bases β' for V and γ' for W. We want to find the matrix representation of the linear map L in the new bases.

We have the following transformations:

$$V_{\beta'} \to W_{\gamma'} \quad \text{and} \quad V_\beta \to W_\gamma$$

To find the matrix of the map $L : V_{\beta'} \to W_{\gamma'}$, we need to relate the new basis vectors to the old basis vectors. Specifically, we perform the following steps:

- To change from the new basis β' to the old basis β, we multiply by the change-of-basis matrix B, so we have $B[v]_{\beta'} = [v]_\beta$.
- To change from the old basis γ to the new basis γ', we multiply by the inverse of the change-of-basis matrix C^{-1}, so we have $C^{-1}[w]_\gamma = [w]_{\gamma'}$.

Thus, the matrix representation of L in the new bases β' and γ' is given by:

$$[L]_{\beta'}^{\gamma'} = C^{-1}[L]_\beta^\gamma B$$

- **Diagonalization and eigenbases**: We want to find some new basis B such that the linear map becomes diagonal. That is

$$B^{-1}LB = L_D.$$

We can achieve this via eigenvectors. By changing the basis to one formed by the eigenvectors, we simplify the linear map so that it acts independently on each direction (each eigenvector). In this new basis, the map scales each eigenvector by its corresponding eigenvalue, without mixing different directions. This independence is what makes the matrix diagonal, making the transformation much easier to understand and work with.

Consider a matrix A and a basis formed from its eigenvectors, say $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. For simplicity, let's assume A has n linearly independent eigenvectors (which guarantees diagonalization).

We can express any vector $\mathbf{x}$ in terms of this new basis as a linear combination of the eigenvectors

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

When we apply the matrix A to $\mathbf{x}$, because each eigenvector $\mathbf{v}_i$ satisfies $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$, we get:

$$A\mathbf{x} = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_n\lambda_n\mathbf{v}_n$$

This shows that A acts on each eigenvector individually by multiplying it by its corresponding eigenvalue. The action of A is now "separated" along each eigenvector direction.

To represent A in the new basis (the eigenvector basis), we express the transformation in matrix form with respect to this new basis. Let's call the matrix P, where the columns of P are the eigenvectors of A

The matrix A in the original basis acts in a complicated way, but in the eigenvector basis, the transformation is simplified. Specifically, in the eigenvector basis, applying A scales each eigenvector by its corresponding eigenvalue. This means that, with respect to this basis, the matrix representing A becomes diagonal

$$P^{-1}AP = D$$
$$\implies A = PDP^{-1}.$$

where D is a diagonal matrix with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ on the diagonal:

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Consider a matrix $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. When this map acts on a vector $\begin{pmatrix} x \\ y \end{pmatrix}$,

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1x \\ 2y \end{pmatrix}.$$

Ie the vector is scaled by 1 in the x direction, and 2 in the y direction

- **Note about diagonalization and eigenbases**: Eigenbases (or eigenspaces) are typically defined only for linear maps of the form $L : V \to V$, where the map L acts on a vector space V and maps vectors within that same space.

If a linear map L is of the form $L : V \rightarrow W$, where W is a different vector space (or even a subspace of V), the concept of eigenvectors and eigenvalues doesn't apply in the usual sense. The reason is that the output of L is not necessarily a scalar multiple of the input vector v, and it may not even belong to the same vector space.

For eigenvectors and eigenvalues to be meaningful, you need the transformation to act within a single vector space, ensuring that the transformed vector remains in the same space, allowing us to compare it directly to the original vector.

- **Diagonalization of a symmetric matrix**: If a matrix $A \in \mathbb{R}^{n \times n}$ is symmetric with real eigenvalues, then the eigenvectors of A are orthogonal.

Thus, we can find an orthonormal set of basis vectors easily.

Vector spaces and subspaces

- **Vector Space**: A vector space is a set of vectors that satisfy

  1. Space has a zero vector
  2. Closed under addition
  3. Closed under multiplication by a scalar

  We also need the space to have the algebric axioms.

  - **Commutativity of addition**: $u + v = v + u$
  - **Associativity of addition**: $(u + v) + w = u + (v + w)$
  - **Additive identity**: There exists a vector $0 \in V$ such that $v + 0 = v$ for all $v \in V$.
  - **Additive inverse**: For every $v \in V$, there exists a vector $-v \in V$ such that $v + (-v) = 0$.
  - **Associativity of scalar multiplication**: $c(dv) = (cd)v$
  - **Distributivity of scalar multiplication over vector addition**: $c(u + v) = cu + cv$
  - **Distributivity of scalar multiplication over scalar addition**: $(c + d)v = cv + dv$
  - **Scalar identity**: $1v = v$, where 1 is the multiplicative identity in the field.

  If we have these properties and algebraic axioms, we have a valid vector space.

- **Abstract vector space**: An abstract vector space is a generalization of this concept, where the elements (vectors) may not have a concrete geometric form, such as functions, polynomials, or matrices, but still follow the same axioms

  For example, A set of matrices can be defined as an abstract vector space

- **Basis of a vector space, span of the basis**: The basis of a vector space is a choice of n vectors $b_1, ... b_n$ such that

$$\mathbf{v} = s_1 b_1 + ... + s_n b_n.$$

  If we are able to generate all vectors in the space by simple scaling the vectors by some constant and adding them, the basis $b_1, ..., b_n$ are said to **span** the vector space.

- **The number of linearly independent vectors per space**: In $\mathbb{R}^n$ the maximum number of linearly independent vectors we can have is n. For example, in $\mathbb{R}^2$, the maximum number of linearly independent vectors we can have is 2. This is why we need exactly n vectors to form a basis in $\mathbb{R}^n$, and having more than n will also result in the case of allowing us to find and throw out the linearly dependent ones.

  In other words, There are only n linearly independent vectors in $\mathbb{R}^n$ because the dimension of $\mathbb{R}^n$ n, which means that the space has exactly n independent directions, or degrees of freedom

- **Definition of dimension**: The dimension of a vector space is the number of vectors in a basis for that space. A basis is a set of linearly independent vectors that spans the entire space. In $\mathbb{R}^n$, any valid basis must have exactly n vectors, because it takes n vectors to fully describe the space.

A set of vectors is linearly independent if no vector in the set can be written as a linear combination of the others. In $\mathbb{R}^n$, if you have more than n vectors, at least one of those vectors can always be written as a linear combination of the others, meaning they will be linearly dependent. This is because there are only n independent directions in $\mathbb{R}^n$

In $\mathbb{R}^2$, the dimension of the space is 2, meaning that any valid set of linearly independent vectors can have at most two vectors. This is because two vectors are sufficient to fully describe the space—they form a basis. Any other vector in $\mathbb{R}^2$ can be expressed as a linear combination of these two vectors.

Once you have two linearly independent vectors, adding any third vector will result in linear dependence, because that third vector will lie in the span of the first two vectors.

- **Discern valid basis**: To give a valid basis for a vector space, we must list a collection of vectors that satisfy

  1. Basis should span the whole space
  2. No redundant basis vectors
  3. Linearly independent set

- **Isomorphic vector spaces**: Two vector spaces are isomorphic if there is a one-to-one correspondence (a bijection) between them that preserves the structure of vector addition and scalar multiplication. This means that if vector spaces V and W are isomorphic, there exists a map (called an isomorphism) $\phi : V \to W$ such that:

  1. ϕ is bijective: Every element in W has a unique preimage in V, and every element in V is mapped to a unique element in W.

  $$\forall w \in W, \ \exists \ v \in V \text{ such that } \phi(v) = w$$

  2. ϕ preserves addition: For any two vectors $u, v \in V$,

  $$\phi(u + v) = \phi(u) + \phi(v)$$

  3. ϕ preserves scalar multiplication: For any scalar $c \in \mathbb{F}$ and any vector $v \in V$,

  $$\phi(cv) = c\phi(v)$$

Example: $M_{2\times 2}$ is isomorphic to $\mathbb{R}^4$

A matrix in $m_{2\times 2}$ can be written as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Where $a, b, c, d \in \mathbb{R}$. This matrix can be uniquely represented as a 4-dimensional vector:

$$(a, b, c, d) \in \mathbb{R}^4.$$

The correspondence between the matrix and the 4-dimensional vector is a linear bijection that preserves both vector addition and scalar multiplication. Thus, there is a one-to-one correspondence between $M_{2\times 2}$ and $\mathbb{R}^4$, and the two spaces are isomorphic.

Moreover, The isomorphism between $M_{2\times 2}$ (the space of 2×2 matrices) and $\mathbb{R}^4$ (the 4-dimensional real vector space) can be described by a linear map that transforms a matrix into a 4-dimensional vector by simply mapping the matrix entries to the components of the vector.

Let's define the linear map $\phi : M_{2\times 2} \to \mathbb{R}^4$.

For any matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2\times 2},$$

the corresponding vector in $\mathbb{R}^4$ under the map ϕ would be:

$$\phi(A) = (a, b, c, d) \in \mathbb{R}^4.$$

This map simply "flattens" the matrix into a 4-tuple of real numbers, with the components arranged in a consistent order (for example, row by row or column by column). In this case, we are mapping the entries of the matrix row by row.

Conversely, given any vector $(a, b, c, d) \in \mathbb{R}^4$, the corresponding matrix in $M_{2\times 2}$ under the inverse map ϕ^{-1} would be:

$$\phi^{-1}(a, b, c, d) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Properties of the Map:

– **Bijectivity:** Every matrix corresponds to a unique vector, and every vector corresponds to a unique matrix.

– **Additivity:** If $A_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}$ and $A_2 = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$, then

$$\phi(A_1 + A_2) = \phi\left(\begin{pmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{pmatrix}\right) = (a_1 + a_2, b_1 + b_2, c_1 + c_2, d_1 + d_2),$$

which is the same as $\phi(A_1) + \phi(A_2)$.

– **Scalar Multiplication:** For any scalar $\lambda \in \mathbb{R}$,

$$\phi(\lambda A) = \phi\left(\begin{pmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{pmatrix}\right) = (\lambda a, \lambda b, \lambda c, \lambda d),$$

which is the same as $\lambda \phi(A)$.

Thus, ϕ is a linear isomorphism between $M_{2\times 2}$ and $\mathbb{R}^4$.

Note: We only have isomorphism if the dimensions are the same.

• **Eigenspaces**: The eigenspace corresponding to an eigenvalue λ of a linear operator T is the set of all eigenvectors associated with λ, along with the zero vector:

$$E_\lambda = \{v \in V \mid T(v) = \lambda v\}.$$

Each eigenvalue λ has its own eigenspace.

- **More on eigenspaces**: Let $A \in \mathbb{R}^{n \times n}$, if λ is an eigenvalue for A, then the eigenspace for λ is

$$E_\lambda = \{x \in \mathbb{R}^n : \ Ax = \lambda x\}.$$

So, $x \in E_\lambda$ implies that

$$Ax = \lambda x \iff Ax - \lambda x = 0 \iff (A - \lambda I)x = 0.$$

Thus,

$$E_\lambda = \ker(A - \lambda I).$$

- **More on function spaces**: Suppose we define a vector space P_4, which contains all polynomials of degree four or less. It has a basis

$$1, x, x^2, x^3, x^4.$$

Thus, is dim 5. The verification of this being a legitimate vector space is not shown. Suppose we define two subspaces of P_4, E and O. Where E is the vector subspace of P_4 that contains all even functions, and O is the vector subspace of P_4 that contains all the odd functions. Recall that a function is even iff $f(-x) = f(x)$, and a function is odd iff $f(-x) = -f(x)$. Both subspaces can be shown that they are indeed vector subspaces. In both subspaces, we use the zero function from P_4, because the zero function is defined here to be both even and odd.

E has basis

$$1, x^2, x^4.$$

Which is dim 3. Note that all constant functions are even, because they satisify the property of even functions $f(-x) = f(x)$.

O has basis

$$x, x^3.$$

Which is dim 2. Let's define a mapping

$$L : \ P_4 \to P_4$$
$$L(p(x)) = p'(x).$$

Which can easily be shown to be linear. Note that this map is not surjective or injective, and thus not bijective. The codomain can not be filled, because no functinos of degree four can be reached by differentiating polynomials of degree four or less. Not injective because all constant functions yield the same derivative (0).

Now, we apply the differentiation operator $L(p(x)) = p'(x)$ to each basis element:

- $L(1) = 0$ because the derivative of a constant is 0.
- $L(x) = 1$.
- $L(x^2) = 2x$.
- $L(x^3) = 3x^2$.
- $L(x^4) = 4x^3$.

Next, express each of these derivatives as a linear combination of the basis elements $\{1, x, x^2, x^3, x^4\}$:

- $L(1) = 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 + 0 \cdot x^4.$
- $L(x) = 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 + 0 \cdot x^4.$
- $L(x^2) = 2x = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 + 0 \cdot x^4.$
- $L(x^3) = 3x^2 = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2 + 0 \cdot x^3 + 0 \cdot x^4.$
- $L(x^4) = 4x^3 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + 4 \cdot x^3 + 0 \cdot x^4.$

We now construct the matrix of the linear map L with respect to the basis $\mathcal{B}$. Each column of the matrix corresponds to the image of one of the basis elements, written as a linear combination of the basis elements

$$[L]_\mathcal{B} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

- The first column corresponds to $L(1) = 0$.

- The second column corresponds to $L(x) = 1$.

- The third column corresponds to $L(x^2) = 2x$.

- The fourth column corresponds to $L(x^3) = 3x^2$.

- The fifth column corresponds to $L(x^4) = 4x^3$.

This matrix represents the differentiation operator on the vector space P_4 in the basis $\{1, x, x^2, x^3, x^4\}$.

Any polynomial $p(x) \in P_4$ can be written as a linear combination of the basis elements $\{1, x, x^2, x^3, x^4\}$, so we can express the polynomial as a vector of coefficients. For example, if:

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4,$$

then the polynomial corresponds to the vector:

$$\mathbf{p} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}.$$

To apply the linear map L, you multiply the matrix representing L by the vector of coefficients for $p(x)$.

If $[L]_\mathcal{B}$ is the matrix of the linear map L, and $\mathbf{p}$ is the vector of coefficients for $p(x)$, then the result is:

$$L(p(x)) = [L]_\mathcal{B} \cdot \mathbf{p}.$$

A linear map from a vector space to itself can be represented as a square matrix whose size corresponds to the dimension of the space. In this case, since L maps P_4 to itself, and P_4 has dimension 5, the matrix representing L must be 5×5.

Now we define a few more maps, first $L : E \to O$, which has a matrix

$$\begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

Then $L : P_4 \to P_3$, which has matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

Then, $L : O \to E$

$$\begin{pmatrix} 1 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix}.$$

Then, $K : P_3 \to P_4$, where $K(p(x)) = \int_0^x p(u) \, du$, defining the integral in this way leads to no constant of integration. Having a constant of integration would lead to a map that is no longer linear, because $K(0) \neq 0$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}.$$

Now, we take a look at $L \circ K : P_3 \to P_3$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

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

Thus, we say that L is a left inverse of K

What about $K \circ L : P_4 \to P_4$?

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}$$

$$= 0 \neq I.$$

Thus, L is not a right inverse of K

- **A quick remark**: We remark that to express a map as a matrix it must satisfy

  1. Linear
  2. Known basis

- **More on the dot product of function spaces**: For two functions f and g in a function space, the dot product (also referred to as the inner product) is typically defined as an integral of their product over a specific interval. Specifically, for real-valued functions f and g on an interval $[a, b]$, the inner product is defined as:

$$\langle p, q \rangle = \int_a^b p(x)q(x)\,dx.$$

  If we define the interval of our function space $x \in [0, 1]$, for example $P_2, \; x \in [0, 1]$. Then the inner product is

$$\langle p, p \rangle = \int_0^1 p(x)q(x)\,dx.$$

- **Properties of the inner product**:

  - $\langle p, 0 \rangle = \int_a^b 0 p(x)\,dx = 0 \int_a^b p(x)q(x)\,dx = 0$
  - $\langle q, p \rangle = \int_a^b q(x)p(x)\,dx = \int_a^b p(x)q(x)\,dx = \langle p, q \rangle$. Thus, we say the inner product is symmetric
  - $\langle p, p \rangle = \int_a^b p^2(x)\,dx \geqslant 0$ and only equal to zero if $p = 0 \; \forall \; x \in [a, b]$
  - $\langle sp, q \rangle = \langle p, sq \rangle = s\langle p, q \rangle$. Note that s cannot depend on x
  - $\langle p + q, r \rangle = \langle p, r \rangle + \langle q, r \rangle$

- **Norm of the inner product**: Recall in vector calculus

$$\mathbf{u} \cdot \mathbf{u} = |\mathbf{u}|^2$$
$$\implies \sqrt{\mathbf{u} \cdot \mathbf{u}} = |\mathbf{u}|.$$

  Given item three above, $\langle p, p \rangle = \int_a^b p^2(x)\,dx$. Thus, $|\langle p, p \rangle| = \sqrt{\int_a^b p^2(x)\,dx}$

- **Angle between function space vectors**: Recall from vector calculus

$$v \cdot w = |v|\,|w| \cos(\theta)$$
$$\implies \cos(\theta) = \frac{v \cdot w}{|v|\,|w|}.$$

- **General Inner product**: If V is a vector space, then the inner product is an operation

$$\langle v_1, v_2 \rangle : \; V \times V \to \mathbb{R}.$$

  With norm $|v| = \sqrt{\langle v, v \rangle}$. And with properties described above. Note that only the zero vector has norm zero.

- **Cauchy Schwarz inequality**: The Cauchy-Schwarz inequality states that for any two vectors u and v in an inner product space, the absolute value of their inner product is less than or equal to the product of their norms:

$$|\langle v, w \rangle| \leqslant |v|\,|w|.$$

Equality holds if and only if v and w are linearly dependent.

- **Inner product space**: An inner product space, also called an *inner space*, is a vector space equipped with an additional structure called an inner product. The inner product is a way to define a notion of "angle" and "length" in the vector space, generalizing the dot product in Euclidean space.

- **Vector plane**: A vector plane (or simply a plane) refers to a two-dimensional flat surface that extends infinitely in all directions within a higher-dimensional space, such as three-dimensional space ($\mathbb{R}^3$). It can be thought of as a set of all possible linear combinations of two linearly independent vectors. Formally, a vector plane can be described as the span of two non-parallel vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in a vector space. If these vectors belong to $\mathbb{R}^3$, their span is the set of all vectors of the form:

$$\mathbf{r} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$$

where c_1 and c_2 are scalars (real numbers). This defines a plane that passes through the origin.

We can find the spanning vectors of a plane given by

$$ax + by + cz = 0.$$

By first finding its normal vector $\vec{\mathbf{n}} = (a, b, c)$ which is normal to all points on the plane. we need to find two linearly independent vectors that lie on the plane and are perpendicular to $\vec{\mathbf{n}}$

To find vectors that satisfy the plane equation, we can make some simple choices by assigning values to two of the coordinates and solving for the third.

For the plane $2x + 3y + z = 0$, let's find two vectors that lie on this plane.

- **First vector**: Choose $x = 1$ and $y = 0$. Substituting these into the plane equation:
$$2(1) + 3(0) + z = 0 \quad \Rightarrow \quad 2 + z = 0 \quad \Rightarrow \quad z = -2$$
So one vector on the plane is:
$$\mathbf{v}_1 = (1, 0, -2)$$

- **Second vector**: Now choose $x = 0$ and $y = 1$. Substituting these into the plane equation:
$$2(0) + 3(1) + z = 0 \quad \Rightarrow \quad 3 + z = 0 \quad \Rightarrow \quad z = -3$$
So another vector on the plane is:
$$\mathbf{v}_2 = (0, 1, -3)$$

These two vectors, $\mathbf{v}_1 = (1, 0, -2)$ and $\mathbf{v}_2 = (0, 1, -3)$, are linearly independent and lie on the plane, so they span the plane.

Note: A vector plane is a vector space

The Adjoint (transpose map)

- **Transpose in inner spaces**: Suppose we have a linear map $L : V \to W$, where V, W are inner spaces. We know $L : V \to W$, and $L^T : W \to V$. Over in W, we have the inner product $\langle L(v), w \rangle_W$, and in V we have $\langle v, L^T(w) \rangle$. We assert

$$\langle L(v), w \rangle_W = \langle v, L^T(w) \rangle_V .$$

- **Definition of the transpose**: The **adjoint** (or transpose) of a linear map $L : V \to W$ between two inner product spaces V and W, denoted $L^T : W \to V$, is the unique linear map that satisfies the following condition for all $v \in V$ and $w \in W$:

$$\langle L(v), w \rangle_W = \langle v, L^T(w) \rangle_V$$

In other words, the adjoint L^T transfers the action of L across the inner product while preserving the result.

- **Linear Functionals**: A **functional** is a special kind of linear map:

$$f : V \to \mathbb{F},$$

where

  - V is a vector space over a field $\mathbb{F}$ (such as $\mathbb{R}$ or $\mathbb{C}$), and
  - f is linear, meaning
  $$f(\alpha v + \beta w) = \alpha f(v) + \beta f(w),$$
  for all $v, w \in V$ and scalars $\alpha, \beta \in \mathbb{F}$.

Thus, a functional is a linear map from a vector space into its field of scalars.

- **Dual space**: Let V be a vector space over a field $\mathbb{F}$ (often $\mathbb{R}$ or $\mathbb{C}$).

The **dual space** of V, denoted V^*, is the set of all linear functionals on V:

$$V^* = \{f : V \to \mathbb{F} \mid f \text{ is linear}\}.$$

That is, each element of V^* is a linear map that takes a vector in V and produces a scalar in $\mathbb{F}$.

- **The adjoint**: Consider the map

$$L : V \to W.$$

We define

$$L^* : W^* \to V^*.$$

Which we call the adjoint

Furthermore, we define

$$L^*(\mu) : V \to \mathbb{R}.$$

For $\mu \in W^*$. Since $\mu \in W^*$, it is a linear map that acts on a vector in W, and takes us to a scalar. We define

$$L^*(\mu)v = \mu(Lv) \in \mathbb{R}.$$

45

In real inner spaces with orthonormal basis,

$$L^* = L^T.$$

We have seen this fact in real spaces, recall for a linear operation $L : V \to V$, for $v, w \in V$

$$\langle L(v), w \rangle = \langle v, L^\top(w) \rangle.$$

If the basis given is not orthonormal, we must convert to an orthonormal basis before using this fact.

In complex inner spaces with orthonormal basis,

$$L^* = \overline{L^T}.$$

- **Self-adjoint and normal maps**

  - A map L is **self-adjoint** if $L^* = L$. In the real setting, we call this map *symmetric*.
  - A map L is said to be **normal** if $L^*L = LL^*$

  All symmetric maps are normal, but not all normal maps are symmetric

- **Real spectral theorem**: Consider a linear operation $L : V \to V$. The following are equivalent (TFAE)

  1. L is symmetric
  2. V has an orthonormal basis consisting of eigenvectors of L
  3. L is represented by a diagonal matrix with respect to some orthonormal basis.

- **Complex spectral theorem**: The only change we make is to number one.

  1. L is normal

Orthogonality

- **Gram-schmidt orthogonalization**: The Gram-Schmidt process is a method used to take a set of linearly independent vectors and convert them into an orthogonal (or orthonormal) set of vectors. It's useful in linear algebra for generating an orthogonal basis from a given set of vectors in an inner product space. Here's how the process works step by step:

  Start with the first vector:
  $$\mathbf{u}_1 = \mathbf{v}_1$$
  This becomes the first orthogonal vector.

  For each subsequent vector $\mathbf{v}_k$,

  subtract the projection of $\mathbf{v}_k$ onto the previously found orthogonal vectors:

  $$\mathbf{u}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \text{proj}_{\mathbf{u}_j}(\mathbf{v}_k)$$

  where $\text{proj}_{\mathbf{u}_j}(\mathbf{v}_k)$ is the projection of $\mathbf{v}_k$ onto $\mathbf{u}_j$:

  $$\text{proj}_{\mathbf{u}_j}(\mathbf{v}_k) = \frac{\langle \mathbf{v}_k, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle} \mathbf{u}_j$$

  If an orthonormal set is desired, normalize each orthogonal vector:

  $$\mathbf{e}_k = \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$$

- **Perpendicular inner spaces (Orthogonal complement)**: Suppose we have an ambient space V, then we take a subset of V, namely $U \subset V$, $U \neq \varnothing$. We define

  $$U^\perp = \{v \in V : \langle v, u \rangle = 0, \ \forall \ u \in U\}.$$

- **Orthogonality and Direct Sum:** One important property of orthogonal complements is that the entire space $\mathbb{R}^n$ can be written as the direct sum of the subspace V and its orthogonal complement $V^\perp$. This means:

  $$\mathbb{R}^n = V \oplus V^\perp.$$

  In other words, every vector in $\mathbb{R}^n$ can be uniquely written as the sum of one vector from V and one vector from $V^\perp$.

  Since these two subspaces are orthogonal and span the entire space, their dimensions must add up to the dimension of the ambient space $\mathbb{R}^n$.

  $$\dim(V) + \dim(V^\perp) = n.$$

- **Orthogonal complement of the kernel**: Recall that the kernel of a linear map acting on vector spaces, $L: V \to W$ is the set of vectors $v \in V$ such that $L(v) = 0$. Formally, for a linear map $L: V \to W$, $\ker(L) = \{v \in V : L(v) = 0\}$. It represents the set of vectors in the domain space that get annihilated by the map. If a map is injective (one-to-one), then the only member of the kernel is the zero vector. Note that the kernel of a linear map is also well defined for linear operations of the form $L: V \to V$. Ie a mapping of a vector space onto itself.

We can find the orthogonal complement of the null space,

$$N^\perp(L) = \ker^\perp(L) = \{v \in V : \langle v, n \rangle = 0 \ \forall \ n \in N(L)\}.$$

If the kernel only contains the zero vector (the map is injective), then the orthogonal complement would be the whole space V, because every vector is perpendicular to the zero vector, even the zero vector itself.

- **Orthonormal matrix**: An orthonormal matrix is a square matrix whose columns (and rows) form an orthonormal set of vectors.

    - **Orthogonality:** Each column is orthogonal to every other column (their dot product is zero).
    - **Normalization:** Each column is a unit vector (its length is 1).

An orthonormal matrix has the following properties

    - $A^T A = A A^T = I$
    - $A^{-1} = A^T$
    - The determinant is ± 1

Special matrices

- **Symmetric matrix**: A matrix $A \in \mathbb{R}^{n \times n}$ is symmetric if $A^\top = A$

- **Skew symmetric matrix**: A matrix $A \in \mathbb{R}^{n \times n}$ is skew symmetric if $A^\top = -A$

- **Properties of symmetric matrices**:

  - **Real Eigenvalues:** A symmetric matrix has real eigenvalues. This follows from the fact that symmetric matrices are self-adjoint in real vector spaces.

  - **Orthogonal Eigenvectors:** The eigenvectors corresponding to distinct eigenvalues of a symmetric matrix are orthogonal.

  - **Diagonalizability**: Every symmetric matrix is diagonalizable, meaning it can be written as

$$A = Q \Lambda Q^\top.$$

    Where Q is an orthogonal matrix (with $Q^\top Q = I$) whose columns are the eigenvectors of A, and Λ is a diagonal matrix whose entries are the eigenvalues of A

  - **Closed under Addition and Scalar Multiplication:** If A and B are symmetric matrices of the same size, then $A + B$ and cA (where c is a scalar) are also symmetric.

  - **Closed under Multiplication (under specific conditions):** If A and B are symmetric and commute ($AB = BA$), then AB is symmetric.

  - **Closed under Powers:** If A is symmetric, then A^n (where n is a positive integer) is also symmetric.

  - **Inverse of symmetric matrices**: If A is symmetric and invertible, then A^{-1} is also symmetric.