The Gist of Linear Algebra

Hugh Van Deventer

April 28, 2025

Contents

1 Vector Spaces: The Foundation

Linear algebra is built upon the concept of a vector space. While the formal definition involves a list of axioms, the core idea is simpler.

Definition 1.1 (Vector Space)**.** A *vector space* V over a field $\mathbb{F}$ (typically $\mathbb{R}$ or $\mathbb{C}$) is a set equipped with two operations: vector addition ($+$) and scalar multiplication ($\cdot$), satisfying ten axioms (commutativity, associativity, identity, inverse for addition; distributivity, compatibility, identity for scalar multiplication). See a textbook for the full list.

Remark 1.2 (The Gist of Vector Spaces)*.* Forget the axioms for a moment. Think about familiar "vectors": arrows in 2D or 3D space ($\mathbb{R}^2, \mathbb{R}^3$). What can you do with them?

- **Add them**: Place arrows tip-to-tail. The result is another arrow (vector).
- **Scale them**: Stretch, shrink, or reverse their direction by multiplying by a number (scalar). The result is still an arrow (vector).

A vector space is essentially *any* collection of objects (not just arrows!) where these two operations (addition and scaling) are meaningfully defined and behave in the expected, consistent way dictated by the axioms. This abstraction is powerful because concepts developed for $\mathbb{R}^n$ can apply to spaces of polynomials, functions, matrices, etc., as long as they fit the vector space rules. The zero vector ($\mathbf{0}$) is the crucial "origin" or additive identity element.

Example 1.3 (Common Vector Spaces)**.** Condiser the following:

- $\mathbb{R}^n$: Vectors as n-tuples of real numbers. $(x_1, \dots, x_n)$. Standard addition and scaling.
- $P_n(\mathbb{R})$: Polynomials with real coefficients of degree at most n. E.g., $p(x) = 3x^2 - 2x + 5$ is in $P_2(\mathbb{R})$. Add polynomials term-by-term, scalar multiply coefficients.
- $M_{m\times n}(\mathbb{F})$: $m \times n$ matrices with entries from field $\mathbb{F}$. Add matrices component-wise, scalar multiply each entry.
- $C([a, b])$: Continuous real-valued functions on the interval $[a, b]$. $(f + g)(x) = f(x) + g(x)$, $(\alpha f)(x) = \alpha f(x)$.

To understand the structure *within* a vector space, we need a few more ideas:

Definition 1.4 (Subspace)**.** A subset $U \subseteq V$ of a vector space V is a *subspace* if U is itself a vector space using the same operations as V. **Test**: U is a subspace if and only if:

1. $\mathbf{0} \in U$ (Contains the zero vector)
2. Closed under addition: If $\boldsymbol{u}, \boldsymbol{w} \in U$, then $\boldsymbol{u} + \boldsymbol{w} \in U$.
3. Closed under scalar multiplication: If $\boldsymbol{u} \in U$ and $\alpha \in \mathbb{F}$, then $\alpha\boldsymbol{u} \in U$.

Intuition: A subspace is a "smaller" vector space living inside a larger one, passing through the origin (like a line or plane through the origin in $\mathbb{R}^3$). These ideas are very important in high dimensional machine learning, where a common theme is that stuff you care about tend to live in lower dimensional subspaces.

Definition 1.5 (Linear Combination and Span)**.** A *linear combination* of vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k \in V$ is any vector of the form $\boldsymbol{w} = \alpha_1\boldsymbol{v}_1 + \cdots + \alpha_k\boldsymbol{v}_k$, where $\alpha_i \in \mathbb{F}$ are scalars. The *span* of these vectors, denoted $\text{span}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$, is the set of *all* possible linear combinations.

Intuition: The span is the smallest subspace containing the given vectors. It's everything you can "reach" by just adding and scaling those initial vectors.

Definition 1.6 (Linear Independence/Dependence)**.** A set of vectors $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ is *linearly independent* if the only solution to $\alpha_1\boldsymbol{v}_1 + \cdots + \alpha_k\boldsymbol{v}_k = \mathbf{0}$ is $\alpha_1 = \cdots = \alpha_k = 0$. If there's a non-trivial solution (at least one $\alpha_i \neq 0$), the set is *linearly dependent.*

Intuition: Independent vectors provide unique directional information; none can be expressed as a combination of the others. Dependent means there's redundancy; at least one vector lies within the span of the others.

Definition 1.7 (Basis and Dimension)**.** A *basis* for a vector space V is a set of vectors $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ that is:

1. Linearly independent.

2. Spans V (i.e., $\text{span}(\mathcal{B}) = V$).

The number of vectors in any basis for V is the same, called the *dimension* of V, denoted $\dim(V)$.

Intuition: A basis is the smallest set of "building blocks" needed to construct every vector in the space. The dimension tells you how many independent directions or degrees of freedom the space has. E.g., $\mathbb{R}^3$ has dimension 3, requiring 3 basis vectors (like $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$).

2 Linear Transformations: Mapping Spaces

Vector spaces are interesting, but linear transformations are what let us relate them and perform operations.

Definition 2.1 (Linear Transformation)**.** A function $T : V \to W$ between vector spaces V, W over the same field $\mathbb{F}$ is a *linear transformation* if it preserves the vector space operations:

$$T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v}) \quad \text{(Additivity)} \tag{1}$$

$$T(\alpha \boldsymbol{v}) = \alpha T(\boldsymbol{v}) \quad \text{(Homogeneity)} \tag{2}$$

for all $\boldsymbol{u}, \boldsymbol{v} \in V$ and $\alpha \in \mathbb{F}$. Equivalently, $T(\alpha \boldsymbol{u} + \beta \boldsymbol{v}) = \alpha T(\boldsymbol{u}) + \beta T(\boldsymbol{v})$.

Remark 2.2 (The Gist of Linear Transformations)*.* Think of a linear transformation T as a "well-behaved" mapping between vector spaces. It doesn't warp the space arbitrarily; it respects the underlying structure. Straight lines remain straight lines (or become points), and the origin $\mathbf{0}_V$ always maps to the origin $\mathbf{0}_W$ (since $T(\mathbf{0}_V) = T(0 \cdot \boldsymbol{v}) = 0 \cdot T(\boldsymbol{v}) = \mathbf{0}_W$). Examples include rotations, reflections, projections, and scaling in $\mathbb{R}^n$, but also differentiation and integration for spaces of functions.

Two crucial subspaces are associated with any linear transformation:

Definition 2.3 (Kernel / Null Space)**.** The *kernel* (or *null space*) of $T : V \to W$ is the set of vectors in the domain V that map to the zero vector in the codomain W:

$$\text{Ker}(T) \coloneqq \{\boldsymbol{v} \in V \mid T(\boldsymbol{v}) = \mathbf{0}_W\}$$

The dimension of the kernel is the *nullity* of T: $\text{nullity}(T) \coloneqq \dim(\text{Ker}(T))$.

Intuition: The kernel tells you what vectors get "lost" or "squashed" to zero by the transformation. If $\text{Ker}(T) = \{\mathbf{0}_V\}$, then T is injective (one-to-one), meaning no information is lost. A larger kernel means more vectors collapse to zero. $\text{Ker}(T)$ is always a subspace of the domain V.

Definition 2.4 (Image / Range)**.** The *image* (or *range*) of $T : V \to W$ is the set of all possible output vectors in the codomain W:

$$\text{Im}(T) \coloneqq \{T(\boldsymbol{v}) \mid \boldsymbol{v} \in V\} = \{\boldsymbol{w} \in W \mid \exists \boldsymbol{v} \in V \text{ s.t. } T(\boldsymbol{v}) = \boldsymbol{w}\}$$

The dimension of the image is the *rank* of T: $\text{rank}(T) \coloneqq \dim(\text{Im}(T))$.

Intuition: The image is the actual subspace of W that is "covered" or "reached" by the transformation T. It might not be all of W (if T is not surjective/onto). The rank measures the dimension of this output space. $\text{Im}(T)$ is always a subspace of the codomain W.

Remark 2.5 (Connection to Matrices)*.* For finite-dimensional spaces, linear transformations are intimately linked to matrices. If $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, any $T : V \to W$ can be uniquely represented by multiplication by an $m \times n$ matrix $\mathbf{A}$. That is, $T(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x}$. In this context:

- $\mathrm{Ker}(T)$ is the *null space* of $\mathbf{A}$, $N(\mathbf{A})$ (solutions to $\mathbf{A}\boldsymbol{x} = \mathbf{0}$).
- $\mathrm{Im}(T)$ is the *column space* of $\mathbf{A}$, $C(\mathbf{A})$ (the span of the columns of $\mathbf{A}$).
- $\mathrm{rank}(T)$ is the *rank* of the matrix $\mathbf{A}$ (dimension of column space).
- $\mathrm{nullity}(T)$ is the *nullity* of the matrix $\mathbf{A}$ (dimension of null space).

This makes matrix techniques (like row reduction) essential for computing kernels, images, rank, and nullity.

These concepts culminate in a fundamental theorem:

Theorem 2.6 (Rank-Nullity Theorem)**.** Let V be a finite-dimensional vector space and let $T : V \to W$ be a linear transformation. Then:

$$\dim(V) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{Im}(T))$$

Or, using the common terms:

$$\dim(\mathrm{Domain}) = \mathrm{nullity}(T) + \mathrm{rank}(T)$$

Remark 2.7 (The Gist of Rank-Nullity). This theorem provides a crucial link between the domain, kernel, and image. It essentially says that the dimension of the input space (V) is completely accounted for by the dimension of the part that gets collapsed to zero (the nullity) and the dimension of the part that constitutes the output (the rank). Dimensions aren't created or destroyed arbitrarily by a linear map; the dimensions of the domain are partitioned between the kernel and the image. This is a powerful tool for understanding the behavior of linear transformations and matrices. For instance, if you know the dimension of the domain and the rank, you immediately know the dimension of the kernel.

3 Eigenvalues and Eigenvectors: The Invariant Directions

Definition 3.1 (Eigenvalue and Eigenvector)**.** Let $T : V \to V$ be a linear operator on V. A non-zero vector $\boldsymbol{v} \in V$ is an *eigenvector* of T if $T(\boldsymbol{v})$ is simply a scalar multiple of $\boldsymbol{v}$:

$$T(\boldsymbol{v}) = \lambda \boldsymbol{v} \tag{3}$$

for some scalar $\lambda \in \mathbb{F}$. The scalar λ is the *eigenvalue* corresponding to $\boldsymbol{v}$.

For a square matrix $\mathbf{A} \in M_n(\mathbb{F})$, representing T in some basis, this is:

$$\mathbf{A}\boldsymbol{v} = \lambda \boldsymbol{v} \quad (\boldsymbol{v} \neq \mathbf{0}) \tag{4}$$

Intuition: Eigenvectors are special vectors whose direction isn't changed by the transformation T (or multiplication by matrix $\mathbf{A}$). They are only stretched, shrunk, or flipped (if $\lambda < 0$). The eigenvalue λ is the factor by which they are scaled along their own direction. Finding these invariant directions can greatly simplify the analysis of a linear transformation.

Definition 3.2 (Characteristic Polynomial)**.** To find eigenvalues for a matrix $\mathbf{A}$, we rewrite Equation (4) as $(\mathbf{A} - \lambda\mathbf{I})\boldsymbol{v} = \mathbf{0}$. Since $\boldsymbol{v} \neq \mathbf{0}$, this system must have non-trivial solutions, which means the matrix $(\mathbf{A} - \lambda\mathbf{I})$ must be singular (not invertible). This occurs precisely when its determinant is zero. The *characteristic polynomial* of $\mathbf{A}$ is:

$$p(\lambda) \coloneqq \det(\mathbf{A} - \lambda\mathbf{I}) \tag{5}$$

The eigenvalues λ of $\mathbf{A}$ are exactly the roots of this polynomial ($p(\lambda) = 0$).