A Revisit to Linear Algebra

Abhilash Patel

1 Motivation

Before we begin, it is worth asking why a course on control systems should spend time on linear algebra at all. The answer lies in what mathematics is. Mathematics is a language, constructed to describe the physical universe with a precision that ordinary language cannot offer. Like any language, it has elements: symbols, rules of combination, and a grammar that tells us which combinations are meaningful. A control system is, at its core, a physical process that changes with time and responds to inputs. To describe this process precisely, we need a language capable of representing quantities, collections of quantities, and the rules by which one collection transforms into another. That language is linear algebra.

The objects of this language, scalar, vector, vector space, map, matrix, eigenvalue, are not arbitrary constructions. Each one exists to answer a specific representational need. In this note, we build the language in order, starting from the smallest unit and ending with the tools that let us characterize how a linear transformation behaves. Once this language is in place, describing a dynamical system in state space becomes a matter of using a vocabulary you already possess, rather than learning new rules under time pressure.

2 Scalar

Definition 2.1. *A **scalar** is a single quantity used to scale, that is, to stretch or shrink, another quantity.*

The simplest use of a scalar is multiplication of a real number: if x is a quantity and α is a scalar, then αx is the same quantity, scaled by a factor of α. A scalar by itself carries no direction and no structure beyond magnitude and sign. It is the most basic unit of the language, and every other object we define is built by combining scalars in specific ways.

3 Set

Definition 3.1. *A **set** is a well-defined collection of distinct objects, called its **elements**. We write $x \in S$ to mean x is an element of the set S, and $S \subseteq T$ to mean every element of S is also an element of T.*

A scalar does not exist in isolation. It belongs to a set of scalars, and before we can scale anything, we must specify this set and the rules by which its elements combine. A set by itself carries no operations and no structure: it only specifies membership. The next step is to equip a set with operations, addition and multiplication, and require that these operations behave consistently. A set with this additional structure is called a field.

4 Field

Definition 4.1. *A **field** $\mathbb{F}$ is a set equipped with two operations, addition and multiplication, satisfying the following properties.*

(i) $(\mathbb{F}, +)$ *is closed, associative, commutative, has an identity element* 0*, and every element has an additive inverse.*

(ii) $(\mathbb{F} \setminus \{0\}, \times)$ *is closed, associative, commutative, has an identity element* 1*, and every nonzero element has a multiplicative inverse.*

(iii) Multiplication distributes over addition: $a(b + c) = ab + ac$ *for all* $a, b, c \in \mathbb{F}$*.*

The sets $\mathbb{R}$ and $\mathbb{C}$ are fields, and they are the two fields used throughout this course. Every scalar coefficient we write from this point onward is assumed to be an element of one of these fields. The set of integers is not a field, since most integers do not have a multiplicative inverse within the integers. This distinction matters because a vector space, defined next, requires its scalars to come from a field. Without that guarantee, operations such as scaling a vector by $1/\alpha$ would not be well defined in general.

5 Vector

Definition 5.1. *A* ***vector*** *is a structure formed by arranging scalars from a field according to a fixed rule of combination, together with two operations, addition and scalar multiplication, that respect the field's operations.*

The familiar picture of a vector as an arrow in $\mathbb{R}^2$ or $\mathbb{R}^3$ is one instance of this definition: an ordered tuple of scalars, added component-wise and scaled component-wise. But the definition is broader than the picture. A polynomial of degree at most n is a vector, since it is built from a finite list of scalar coefficients, and two polynomials add coefficient by coefficient. A discrete-time signal of finite length is a vector for the same reason. What makes an object a vector is not its shape or its physical interpretation, but whether it supports addition and scalar multiplication consistent with the underlying field.

6 Vector Space

Definition 6.1. *A* ***vector space*** V *over a field* $\mathbb{F}$ *is a set of vectors, together with addition and scalar multiplication, satisfying:*

(i) $(V, +)$ *is closed, associative, commutative, has a zero vector, and every vector has an additive inverse.*

(ii) Scalar multiplication is compatible with field multiplication: $\alpha(\beta v) = (\alpha\beta)v$ *for* $\alpha, \beta \in \mathbb{F}$*,* $v \in V$*.*

(iii) Distributive laws hold: $\alpha(u + v) = \alpha u + \alpha v$ *and* $(\alpha + \beta)v = \alpha v + \beta v$*.*

(iv) $1 \cdot v = v$ *for all* $v \in V$*.*

Example 6.1. $\mathbb{R}^n$*, the set of* n*-tuples of real numbers, is a vector space over* $\mathbb{R}$*. The set of real polynomials of degree at most* n *is a vector space over* $\mathbb{R}$*. The set of continuous real-valued functions on an interval* $[0, T]$ *is a vector space over* $\mathbb{R}$*.*

Definition 6.2. *A subset* $S \subseteq V$ *is a* ***subspace*** *of* V *if* S *is itself a vector space under the operations inherited from* V*. Equivalently,* S *is nonempty, closed under addition, and closed under scalar multiplication.*

7 Basis, Linear Independence, and Dimension

Given a vector space, we want the smallest set of vectors from which every other vector in the space can be constructed.

Definition 7.1. *Given vectors $v_1, \ldots, v_k \in V$ and scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$, the vector $\alpha_1 v_1 + \cdots + \alpha_k v_k$ is called a **linear combination** of $v_1, \ldots, v_k$. The set of all linear combinations of $v_1, \ldots, v_k$ is called their **span**.*

Definition 7.2. *A set of vectors $\{v_1, \ldots, v_k\}$ is **linearly independent** if the only scalars $\alpha_1, \ldots, \alpha_k$ satisfying*

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0$$

*are $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$. If some other combination also gives zero, the set is **linearly dependent**, meaning at least one vector in the set can be written as a combination of the others.*

Definition 7.3. *A set of vectors $\{e_1, \ldots, e_n\}$ is a **basis** for V if it is linearly independent and its span equals V. The number of vectors in a basis is called the **dimension** of V, written $\dim V$.*

Remark 7.1. *Every basis for a given finite-dimensional vector space has the same number of vectors. The specific vectors chosen as a basis are not unique, but the count is. Dimension is therefore a property of the space itself, while a basis is a choice made by whoever is describing the space. Coordinates of a vector, that is, the scalars $\alpha_1, \ldots, \alpha_n$ used to express it as $\sum_i \alpha_i e_i$, depend on the basis chosen, even though the vector itself does not change.*

8 Linear Maps

A vector space by itself only describes objects. To describe how one collection of objects transforms into another, we need a map between vector spaces.

Definition 8.1. *Let V and W be vector spaces over the same field $\mathbb{F}$. A map $T : V \to W$ is called **linear** if, for all $u, v \in V$ and $\alpha, \beta \in \mathbb{F}$,*

$$T(\alpha u + \beta v) = \alpha\, T(u) + \beta\, T(v).$$

This single condition is strong. It implies that a linear map is entirely determined by its action on a basis of V: once $T(e_1), \ldots, T(e_n)$ are known for a basis $\{e_1, \ldots, e_n\}$ of V, then $T(v)$ is known for every $v \in V$, since $v = \sum_i \alpha_i e_i$ implies $T(v) = \sum_i \alpha_i T(e_i)$.

8.1 Domain, Codomain, Range

Definition 8.2. *For a map $T : V \to W$, the space V is called the **domain** and the space W is called the **codomain**. The **range** (or image) of T is*

$$\text{range}(T) = \{T(v) : v \in V\} \subseteq W.$$

Remark 8.1. *The range is generally a proper subset of the codomain, not the whole of it. The codomain specifies where outputs are permitted to lie; the range specifies where outputs actually lie.*

Definition 8.3. *The **null space** (or kernel) of T is*

$$\text{null}(T) = \{v \in V : T(v) = 0\}.$$

Definition 8.4 (Rank-nullity theorem)**.** *For a linear map $T : V \to W$ with V finite-dimensional,*

$$\dim(\text{range}(T)) + \dim(\text{null}(T)) = \dim(V).$$

The dimension of the range is called the **rank** of T. A map is **injective** if and only if $\text{null}(T) = \{0\}$, and **surjective** if and only if $\text{range}(T) = W$.

9 Matrix

A linear map is an abstract rule. To compute with it, we represent it using numbers, and this representation is the matrix.

Definition 9.1. *Let $T : V \to W$ be linear, with $\dim V = n$ and $\dim W = m$. Fix a basis $\{e_1, \ldots, e_n\}$ for V and a basis $\{f_1, \ldots, f_m\}$ for W. Write each $T(e_j)$ in terms of the basis of W:*

$$T(e_j) = a_{1j} f_1 + a_{2j} f_2 + \cdots + a_{mj} f_m, \qquad j = 1, \ldots, n.$$

*The $m \times n$ array $A = [a_{ij}]$ is called the **matrix** of T with respect to these two bases. The j-th column of A is the coordinate vector of $T(e_j)$ in the basis of W.*

Remark 9.1. *The matrix A depends on the choice of bases for V and W. If either basis changes, the entries of A change, even though the underlying map T does not. This is the reason a fixed physical transformation can be written using different-looking matrices, depending on the coordinate system used to describe it.*

The dictionary between the map and its matrix is direct: the range of T corresponds to the column space of A (the span of its columns), the null space of T corresponds to the null space of A (solutions of $Ax = 0$), and the rank of T equals the rank of A.

10 Eigenvalues and Eigenvectors

We now restrict attention to a linear map from a vector space to itself, $T : V \to V$, represented by a square matrix A.

Definition 10.1. *A nonzero vector $v \in V$ is an **eigenvector** of A if there exists a scalar λ such that*

$$Av = \lambda v.$$

*The scalar λ is called the corresponding **eigenvalue**.*

An eigenvector is a direction along which the action of A reduces to multiplication by a single scalar. If A possesses a full set of n linearly independent eigenvectors for an $n \times n$ matrix, then choosing these eigenvectors as a basis represents A as a diagonal matrix, and repeated application of A reduces to independent scalar multiplications along each eigenvector direction.

Definition 10.2. *The eigenvalues of A are the roots of the **characteristic equation***

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

For an $n \times n$ matrix, this is a degree-n polynomial in λ, so A has exactly n eigenvalues, counted with multiplicity, possibly complex and possibly repeated.

Definition 10.3. *The **algebraic multiplicity** of an eigenvalue is its multiplicity as a root of the characteristic polynomial. The **geometric multiplicity** of an eigenvalue is the dimension of the space of eigenvectors associated with it, that is,* $\dim \operatorname{null}(A - \lambda I)$.

Geometric multiplicity never exceeds algebraic multiplicity. When the two are equal for every eigenvalue, A is **diagonalizable**. When geometric multiplicity falls short for some eigenvalue, A cannot be diagonalized, and a more general structure, the Jordan form, is required to represent it.

Remark 10.1. *Two standard identities relate eigenvalues to the matrix directly:*

$$\operatorname{trace}(A) = \sum_{i=1}^{n} \lambda_i, \qquad \det(A) = \prod_{i=1}^{n} \lambda_i.$$

A matrix is singular, that is, non-invertible, if and only if at least one eigenvalue is zero.

11 Similarity Transformations

Definition 11.1. *Two square matrices A and A' are **similar** if there exists an invertible matrix P such that*

$$A' = P^{-1}AP.$$

If A is the matrix of a linear map $T : V \to V$ in one basis, and P is the change-of-basis matrix to a second basis, then A' is the matrix of the same map T in the second basis. Similar matrices represent the same underlying map in different coordinate systems, and consequently share the same eigenvalues, the same trace, and the same determinant. If A is diagonalizable, choosing P from the eigenvectors of A gives $A' = P^{-1}AP$ as a diagonal matrix of eigenvalues.

12 Singular Values

Eigenvalues require a square matrix and are meaningful only for a map from a space to itself. For a general matrix A of size $m \times n$, we require a different tool to describe how the map stretches vectors.

Definition 12.1 (Singular value decomposition)**.** *Every matrix $A \in \mathbb{R}^{m\times n}$ can be written as*

$$A = U\Sigma V^T,$$

*where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal with non-negative entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ along the leading diagonal. The values σ_i are the **singular values** of A.*

Singular values relate to eigenvalues through

$$\sigma_i(A) = \sqrt{\lambda_i(A^TA)},$$

since A^TA is always square and symmetric, regardless of the shape of A. The largest singular value gives the maximum possible amplification of a unit vector under A:

$$\sigma_{\max}(A) = \max_{\|x\|=1} \|Ax\| = \|A\|_2.$$

13 Condition Number

Definition 13.1. *For a matrix A with $\sigma_{\min}(A) > 0$, the **condition number** is*

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}.$$

A condition number close to 1 indicates that A amplifies vectors by roughly the same factor in every direction. A large condition number indicates that A amplifies some directions far more than others, and consequently that A is close to a singular matrix along its least-amplified direction. In this case, small errors in the input can produce disproportionately large errors in Ax, or in the solution of $Ax = b$.

14 Quadratic Forms

Definition 14.1. *For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, the function*

$$x^TAx, \qquad x \in \mathbb{R}^n,$$

*is called a **quadratic form**.*

Definition 14.2. *A symmetric matrix A is*

- ***positive definite*** *if $x^T Ax > 0$ for every $x \neq 0$,*
- ***positive semidefinite*** *if $x^T Ax \geq 0$ for every x,*
- ***negative definite*** *(respectively **semidefinite**) if $-A$ is positive definite (respectively semidefinite),*
- ***indefinite*** *if $x^T Ax$ takes both positive and negative values depending on x.*

Remark 14.1. *A symmetric matrix A is positive definite if and only if every eigenvalue of A is strictly positive. It is positive semidefinite if and only if every eigenvalue is nonnegative.*

15 Rayleigh Quotient

Definition 15.1. *For a symmetric matrix $A \in \mathbb{R}^{n\times n}$ and a nonzero vector x, the **Rayleigh quotient** is*

$$R(x) = \frac{x^T Ax}{x^T x}.$$

If x is an eigenvector of A, then $R(x)$ equals the corresponding eigenvalue exactly. For any nonzero x, the Rayleigh quotient satisfies the **Rayleigh inequality**:

$$\lambda_{\min}(A) \leq R(x) \leq \lambda_{\max}(A).$$

This gives bounds on the extreme eigenvalues of a symmetric matrix without solving the characteristic equation, and it connects the sign of a quadratic form directly to the sign of the eigenvalues stated in the previous section.

16 Linear Matrix Inequalities

The notion of definiteness extends naturally from a single matrix to conditions involving several matrices, and this extension is used extensively later in this course for stability analysis and controller design.

Definition 16.1. *Given symmetric matrices $F_0, F_1, \ldots, F_p \in \mathbb{R}^{n\times n}$, an expression of the form*

$$F(y) = F_0 + \sum_{i=1}^{p} y_i F_i \succ 0$$

*in the unknowns $y = (y_1, \ldots, y_p)$ is called a **linear matrix inequality (LMI)**, where $F(y) \succ 0$ means $F(y)$ is positive definite. The inequality $F(y) \succeq 0$ denotes positive semidefiniteness.*

Remark 16.1. *An LMI is linear in the unknowns y_i, even though the constraint $F(y) \succ 0$ is a statement about the eigenvalues of $F(y)$, which are generally nonlinear functions of the entries of $F(y)$. The set of y satisfying $F(y) \succeq 0$ is convex, since positive semidefiniteness is preserved under nonnegative combinations of matrices. This convexity is what makes LMIs computationally tractable: a feasible y, or an optimal y under a linear objective, can be found using standard convex optimization methods, without resorting to search over all possible values of y.*

Example 16.1. *A single symmetric matrix inequality $A^T P + PA \prec 0$, with $P = P^T \succ 0$ treated as the unknown, is an LMI in the entries of P. This particular inequality is encountered later in the course in the context of Lyapunov stability.*