Linear Algebra | A Crash Course

Parin Chaipunya

Foreword

This lecture note is designed to recap the background knowledge for new students in the Master of Science Program in Applied Mathematics at KMUTT. This note together with another Topology note should include the necessary credentials to persue the Functional Analysis and Application course.

– Welcome to KMUTT.

Contents

§ 1. Introduction

Linear algebra focuses on the vector space structure and linear transformations between two vector spaces. The theory was originally developed from the study of a linear system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1.1}$$

of m equations and n variables. Letting $A = [a_{ij}]$ be the coefficient matirx, $x = (x_1, \cdots, x_n)^\top$ be the unknown vector, and $b = (b_1, \cdots, b_m)^\top$ be the target vector, we can reduce (1.1) into the following simple matrix equation

$$Ax = b. \tag{1.2}$$

The first question concerning this equation is about its solution. Here, the concepts of determinant and rank were introduced in order to characterize the solution of (1.2). As we dive into such concepts, we found that there are even deeper connections between them and would involve many more concepts and structures like Euclidean space, subspaces, basis and dimension.

Let us now take a new look at the equation (1.2) from another perspective. We may view the matrix A as a rule of *transforming* any (appropriate) vector x into Ax. That is, it inherits a mapping (or transformation, or operator) T such that $T(x) = Ax$. This mapping satisfies the property $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ for any possible inputs x, y and any scalars α, β. This property is called *linearity*. Hence, the equation (1.2) can be rephrased equivalently to "find a vector x that transformed (under the rule of A) in to the target vector b."

It turns out that the fascinating albegraic structure (called the vector space structure, or the linear structure) of a Euclidean space can also be found in several other sets. This allows also for a linear operator to be defined on any pair of sets with linear structure. In this lecture, we shall be exploring linear algebra of vector spaces and linear operators which, together with another intensive topology class, will help lay a good background for our Functional Analysis course.

§ 2. Euclidean spaces and matrix theory

Let us begin with the definition of a Euclidean space.

Definition 2.1. Let $n \in \mathbb{N}$. The set $\mathbb{R}^n := \{(x_1, \cdots, x_n)^\top \mid x_i \in \mathbb{R} \ (\forall i = 1, 2, \cdots, n)\}$ is called an n-dimensional Euclidean space, whose elements are called n-dimensional (column) vectors. The prefix "n-dimensional" is usually tacit if no confusion should arise. Moreover, the notation 0 shall refer to a zero vector $0 \equiv (0, \cdots, 0)^\top$ of appropriate dimension as well as the number 0.

2.1 *Matrix equations and ranks*

If A is an $m \times n$ matrix and $b \in \mathbb{R}^m$, then one may seek an unknown vector $x \in \mathbb{R}^n$ for which the matrix equation (1.2) holds, i.e. $Ax = b$. The need to characterize the solution of this equation leads to several interesting concepts about a matrix. Since a matrix equation can be thought of as a linear system, one may perform row operations to obtain an equivalent linear system (in the sense that there is no loss or gain of information about the unknowns). An equivalent row-reduced echelon form of a matrix A can be used to reflect the solution propertis of a matrix equation. Recall that we write a_{i*} and a_{*j} to denote the i^{th} row (as a row vector) and the j^{th} column (as a column vector) of A.

Definition 2.2. An $m \times n$ matrix U is said to be a *row echelon matrix* if

a. all rows that consisting entirely of zero are stacked below nonzero rows, and

b. in each nonzero row, the first nonzero entry (called the leading entry) is to the left of the leading entries below it.

If a row echelon matrix U satisfies the following additional condition

c. all leading entries are 1 and the leading entries are the only nonzero entry in their rows,

then we say that U is a *row reduced echelon matrix.*

By the above definition, one may notice for a row echelon matrix that the first nonzero entries along the rows must lie in the staircase shape and that all the zero rows (if any) are stacked in the buttom of the matrix. This may be illustrated in the following:

$$U = \begin{bmatrix} (*) & * & * & * & * & * & * & * \\ 0 & (*) & * & * & * & * & * & * \\ 0 & 0 & 0 & (*) & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & (*) & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The leading entries of each row (entries in the parentheses) are called the *pivots* of U. Obviously, the number of pivots is equal to the number of nonzero rows.

Definition 2.3. If the matrix A can be reduced (by row operations) to a row (resp., reduced) echelon matrix U, then we say that U is *a row echelon form (or REF)* (resp., *row reduced echelon form (or RREF)*) of A.

Since there are many ways to generally perform row operations, one may reach at different RREF matrix of A. However, there is a nice consistent property across all the RREFs of A as follows:

Proposition 2.4. *All the RREFs of A has the same number of pivots (and hence the same number of nonzero rows).*

This leads us to define the *rank* of a matrix.

Definition 2.5. The *rank* of a matrix A, denoted by $\text{rank}(A)$, is the number of pivots in any RREF of A.

The rank of a matrix is a great tool to classify the solvability of a given matrix equation. The equation $Ax = b$ is said to be *consistent* if it has at least a solution, otherwise it is called *inconsistent*.

Theorem 2.6. *Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then the following solution characterizations hold:*

1. *If* $\text{rank}(A) < \text{rank}([A\ b])$, *then $Ax = b$ is inconsistent.*
2. *If* $\text{rank}(A) = \text{rank}([A\ b]) = n$, *then $Ax = b$ is consistent and it has a unique solution in $\mathbb{R}^n$.*
3. *If* $\text{rank}(A) = \text{rank}([A\ b]) < n$, *then $Ax = b$ is consistent but it has infinitely many solutions.*

It is important to add that in the third case (infinitely many solutions), the solutions is a translation of a subspace in $\mathbb{R}^n$. We shall state a more precise result in the Rank-Nullity Theorem below.

2.2 Linear independence, basis, dimension and subspaces

The description of an RREF of a given matrix is not so convenient to work with in greater details. Hence we may need to look for a friendlier approach. Let us recall some necessary concepts first.

Definition 2.7. A family $\{v_1, \cdots, v_k\}$ of vectors in $\mathbb{R}^n$ is said to be *linearly independent* if the equation $a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$ holds only if $a_1 = a_2 = \cdots = a_k = 0$.

Put $V = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{bmatrix}$ and $A = (a_1, \cdots, a_k)^\top$. If the only solution to $VA = 0$ is $A = 0$, then this would imply that $\{v_1, \cdots, v_k\}$ is linearly independent. Conversely, if $VA = 0$ has a nonzero solution, then the family is not linearly independent.

A linear combination of vectors $v_1, \cdots, v_k$ with coefficients $\lambda_1, \cdots, \lambda_k \in \mathbb{R}$ is the summation $\sum_{i=1}^k \lambda_i v_i = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$. Then if the family $\{v_1, \cdots, v_k\}$ is not linearly independent, that would mean some vector v_i is either zero or it can be written as a *nontrivial* linear combination of the remaining vectors. An infinite set of vectors is said to be *linearly independent* if all of its finite subset is linearly independent.

In $\mathbb{R}^n$, the maximal number of linearly independent vectors is always n. We therefore say that the dimension of $\mathbb{R}^n$ is n, or symbolically that $\dim(\mathbb{R}^n) = n$. The maximal linearly independent set of vectors in R^n is called the *basis* of R^n. The most natural choice of basis is the *standard basis*, which is the set $\{e_1, e_2, \cdots, e_n\}$ where $e_1 = (1, 0, \cdots, 0)^\top, e_2 = (0, 1, 0, \cdots, 0)^\top, \cdots, e_n = (0, \cdots, 0, 1)^\top$. Notice that the standard vector representation $x = (x_1, \cdots, x_n)^\top \in \mathbb{R}^n$ corresponds to the expansion

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n.$$

Take any subset $V \subseteq \mathbb{R}^n$. Then the *span* (or *linear span*) of X, denoted by $\operatorname{span}(V)$, is the set of all linear combinations of vectors in V, i.e.

$$\operatorname{span}(X) := \{\lambda_1 v_1 + \cdots + \lambda_k v_k \,|\, \lambda_i \in \mathbb{R}\,(\forall i = 1, \cdots, k)\}.$$

The set $\operatorname{span}(V)$ is closed under addition and scalar multiplication, which means we have $\alpha x + \beta y \in \operatorname{span}(V)$ whenever $x, y \in \operatorname{span}(V)$ and $\alpha, \beta \in \mathbb{R}$. The set $\operatorname{span}(V$ is called the *subspace of* $\mathbb{R}^n$ *spanned by* V. If $\{v_1, \cdots, v_m\}$ is a maximal linearly independent subset in $\operatorname{span}(V)$, we say that $\operatorname{span}(V)$ has dimension m and write $\dim(\operatorname{span}(V)) = m$. On the other hand, if a subset $W \subseteq \mathbb{R}^n$ is closed under addition and scalar multiplication, then it is a subspace of $\mathbb{R}^n$. In this case, W is necessarily equal to $\operatorname{span}(\{w_1, \cdots, w_m\})$ for some vectors $w_1, \cdots, w_m \in \mathbb{R}^n$.

The following examples of subspaces are obtained by regarding a matrix A as a mapping.

Example 2.8. If A is an $m \times n$ matrix, then the set $\ker(A) := \{x \in \mathbb{R}^n \,|\, Ax = 0\}$ is a subspace of $\mathbb{R}^n$. The subspace $\ker(A)$ is called the *kernel* or the *null space* of A. The dimension of $\ker(A)$ is usually referred to as the *nullity* of A, or symbolically as $\operatorname{nullity}(A)$.

Example 2.9. If A is an $m \times n$ matrix, then the set $\text{ran}(A) := \{y \in \mathbb{R}^m \,|\, y = Ax,\ \exists x \in \mathbb{R}^n\}$ is a subspace of $\mathbb{R}^m$. The subspace $\text{ran}(A)$ is called the *range* or the *image* of A.

The rank of a matrix A can be stated conveniently in terms of linear independence as well.

Proposition 2.10. *The rank of a matrix A is equal to the maximal number of linearly independent rows. In particular,* $\dim(\text{ran}(A)) = \text{rank}(A)$.

Next, we shall look at the relationship between the rank (dimension of $\text{ran}(A)$) and the nullity (dimension of $\ker(A)$) of a matrix A.

Theorem 2.11 (Rank-Nullity Theorem)**.** *If A is an $m \times n$ matrix, then* $\text{rank}(A) + \text{nullity}(A) = n$.

Lastly, we shall mention that the solution of the nonhomogeneous matrix equation $Ax = b$ can be related to that of the homogeneous one $Ax = 0$ (which is $\ker(A)$). If $Ax = b$ has a solution x_0, then the set of all solutions of $Ax = b$ is $x_0 + \ker(A) = \{x_0 + v \,|\, v \in \ker(A)\}$. If $\ker(A)$ is trivial, then this corresponds to the case where $\text{rank}(A) = n$ and $x_0 + \ker(A)$ reduces to a single point x_0.

2.3 Determinant

In this section, we give a special treatment to a square matrix by looking at its determinant. Most of us know how to calculate the determinant of any 2×2 or 3×3 matrices. For larger matrices, it is usually frustrating to calculate the determinant. Even when we know the determinant of a matrix, it is still not so clear what this value tells us about a matrix.

The *determinant* of a square matrix A can be equivalently and formally defined in many ways. The most used definition is the one that is involved with the column permutation.

Recall that a permutation on the set $\{1, \cdots, n\}$ is any bijection $\sigma : \{1, \cdots, n\} \to \{1, \cdots, n\}$. The shorthand notation $(\sigma(1), \sigma(2), \cdots, \sigma(n))$ is often used to denote the permutation σ. The sign of a permutation σ, $\text{sgn}(\sigma)$, is the quantity $(-1)^{N(\sigma)}$ where $N(\sigma)$ is the number of inversions in σ. The set of all permutations on $\{1, \cdots, n\}$ will be denoted with $\text{Perm}(n)$.

Definition 2.12. Let A be an $n \times n$ matrix. Then its determinant $\det(A)$ is defined by

$$\det(A) := \sum_{\sigma \in \mathrm{Perm}(n)} \left(\mathrm{sgn}(A) \prod_{i=1}^{n} a_{i\sigma(i)} \right).$$

The above definition of determinant is not so easy to calculate in practice. However it is simple enough for deducing important properties.

Proposition 2.13. *Let A be an $n \times n$ matrix. Then the following properties hold:*

1. $\det(A^\top) = \det(A)$.
2. $\det(AB) = \det(A)\det(B)$.
3. $\det(\lambda A) = \lambda^n A$.

The meaning of the determinant can be explained geometrically by viewing the matrix A as a mapping. Suppose that f_i denotes the column vector of A for each $i = 1, \cdots, n$. Then we get $f_i = Ae_i$ which means the n-dimensional unit box is transformed into an n-dimensional parallelogram with sides f_i (see Figure 2.1). A careful computation reveals that the volume of this parallelogram equals to $|\det(A)|$. If we consider a solid $S \subseteq \mathbb{R}^n$, then the determinant $|\det(A)|$ gives a scaling constant of the volume S after being transformed by A to that of S before the transformation. In other words

$$|\det(A)| = \frac{\text{volume of } AS}{\text{volume of } S}.$$

Now, the sign of $\det(A)$ reveals the change of orientation sign of the transformed basis. For example, if A swaps e_1 and e_2 of $\mathbb{R}^2$, then $\det(A) = -1$ which means the output changes the right-hand orientation into the left-hand orientation.

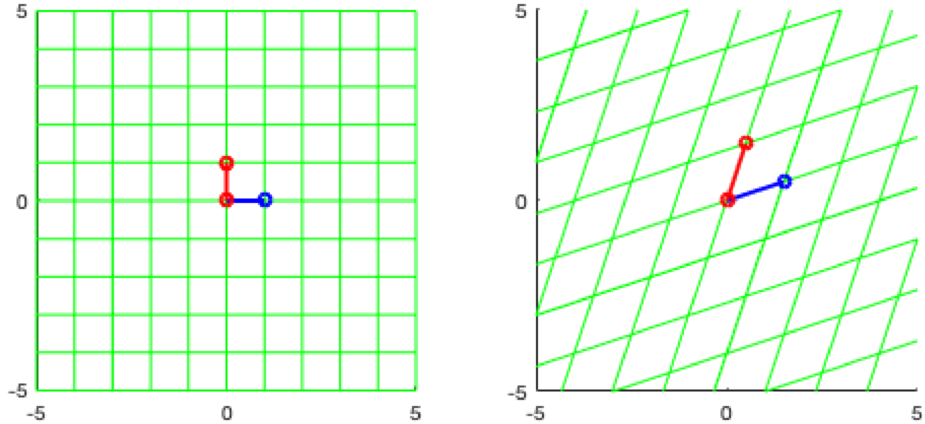

Figure 2.1: The transformation of the standard basis.

If $\det(A) = 0$, then some two basis elements collapse unto a common line. This collapse implies that the output dimension $\dim(\text{ran}(A))$ must be less than n. In fact, we have the following conclusion.

Proposition 2.14. *Let A be an $n \times n$ matrix. Then the following statements are equivalent:*

1. $\det(A) = 0$.
2. $\text{rank}(A) < n$.
3. $\ker(A)$ *is nontrivial.*

This proposition has a very crucial consequence: $\det(A) \neq 0 \iff \text{rank}(A) = n$. The latter condition implies $\text{rank}([A\ b]) = n$ for all vectors $b \in \mathbb{R}^n$. So the matrix equation $Ax = b$ has a unique solution for any $b \in \mathbb{R}^n$. Now, suppose that v_i is the unique solution to $Ax = e_i$ for each $i = 1, \cdots, n$ and let V be the matrix whose i^{th} column is v_i, then

$$AV = I.$$

Now that we know $\det(A^\top) = \det(A) = n$, we may similarly get a unique solution u_i to the equation $A^\top x = e_i$ for each $i = 1, \cdots, n$. If U is the matrix whose i^{th} column is u_i, then

$$A^\top U = I = U^\top A.$$

Therefore we get

$$I = U^\top A = U^\top (AV)A = (U^\top A)VA = VA.$$

Since $AV = VA = I$, we conclude that A is invertible and $A^{-1} = V$. This procedure proves the following well-known fact:

Proposition 2.15. *A square matrix A is invertible if and only if* $\det(A) \neq 0$.

2.4 Linear transformtation

Let us formally introduce a linear transformation between Euclidean spaces once again.

Definition 2.16. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* if

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$$

for any $x, y \in \mathbb{R}^n$ and any $\alpha, \beta \in \mathbb{R}$.

The defining feature of a linear transformation lets understand the big picture by only knowing the important ones. For instance, if $\{e_1, \cdots, e_n\}$ is the standard basis for $\mathbb{R}^n$, then $T(e_1), \cdots, T(e_n)$ are sufficient to describe the whole T. This is because any vector $x \in \mathbb{R}^n$ can be represented as a linear combination of the basis elements $x = x_1 e_1 + \cdots + x_n e_n$. Therefore $T(x) = T(x_1 e_1 + \cdots + x_n e_n) = x_1 T(e_1) + \cdots + x_n T(e_n)$. If we let A_T to be a matrix whose i^{th} column is $T(e_1)$, then A_T is an $m \times n$ matrix and we have

$$T(x) = A_T x$$

for all $x \in \mathbb{R}^n$. The matrix A_T obtained this way is called the *matrix representation* of T. Together with the mapping perspective of a matrix, we conclude the following result which means we do not really distinguish a matrix and a linear transformation when we are in Euclidean settings.

Proposition 2.17. *The linear maps from $\mathbb{R}^n$ into $\mathbb{R}^m$ is an equivalent object with an $m \times n$ matrix in the following sense:*

1. *Every $m \times n$ matrix A induces a unique linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ such that $T_A(x) = Ax$ for all $x \in \mathbb{R}^n$.*

2. *Every linear transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ *induces a unique* $m \times n$ *matrix* A_T *such that* $T(x) = A_T x$ *for all* $x \in \mathbb{R}^n$.

It is possible to define the *kernel*, *range*, *rank*, *determinant*, and other related notions for a linear transformation as well. In this case, we consider its matrix representation and calculate the correspondings: $\ker(T) = \ker(A_T)$, $\operatorname{ran}(T) = \operatorname{ran}(A_T)$, $\operatorname{rank}(T) = \operatorname{rank}(A_T)$, $\det(T) = \det(A_T)$, etc.

2.5 Eigenvalues, eigenvectors and diagonalization

Eigenvalues and eigenvectors are important concepts with applications in operator theory, differential equations and also in digital image processing. In this section, those applications are not in our scope to discuss.

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformtaion with the matrix representation A. We are interested in finding a nonzero vector $v \in \mathbb{R}^n$ that does not change the direction under the transformation T. In terms of its matrix representation, we wish to find a nonzero vector $v \in \mathbb{R}^n$ such that

$$Av = \lambda v \tag{2.1}$$

for some $\lambda \in \mathbb{R}$. The unknowns in this equation are the vector v and the scalar λ. Due to the product of λ and v, this equation is nonlinear. If the pair (λ, v) (with $v \neq 0$) solves (2.1), we say that v is an *eigenvector* of A associated with the eigenvalue λ.

Rearranging (2.1) yields $(A - \lambda I)v = 0$. Observe that if $A - \lambda I$ is invertible, then the only solution to (2.1) is $v = 0$ which is not qualified as an eigenvector. Conversely, if $v = 0$ is the only solution to (2.1), then $A - \lambda I$ must be invertible (by Theorem 2.6). Therefore, an eigenvector v corresponds to (2.1) only when $A - \lambda I$ is not invertible, i.e. $\det(A - \lambda I) = 0$. This equation $\det(A - \lambda I) = 0$ is referred to as the *characteristic polynomial* since it resolves into a polonomial of degree n. Since the characteristic polynomial involves only the scalar unknown λ, it becomes a standard procedure of determining the eigenvalues of A (or of T). To find eigenvalues and eigenvectors of A, one then follow the following steps:

1. Solve the characteristic polynomial $\det(A - \lambda I) = 0$ for eigenvalues λ.

2. For each eigenvalue λ, solve the linear system $(A - \lambda I)v = 0$ for the eigenvectors v.

Note that there can be the situation where (2.1) has no real eigenvalues (in which case A does not preserve any directions). However, the characteristic polynomial is guaranteed to have n repeatable complex roots. If $\lambda \in \mathbb{C}$ solves $\det(A - \lambda I) = 0$, we still would call λ an *eigenvalue* of A and similar principle applies to the eigenvectors. If λ is complex, then there is no doubt v can be complex. However, one can also find a complex eigenvector as well for a real eigenvalue. This will not be considered here. Whether A has real or complex eigenvalues is an important property in the study of differential equations. The following propositions can be useful in confirming the real eigenvalues.

Proposition 2.18. *If A is a triangular matrix, then eigenvalues of A are its diagonal entries.*

Proposition 2.19. *If A is a symmetric matrix, then all eigenvectors of A are real.*

We also need to address that $(A - \lambda I)v = 0$ is consistent with infinitely many solutions. The set of all eigenvectors associated with λ is therefore $\ker(A - \lambda I)$, which is a subspace. This subspace is called the *eigenspace* of A associated with λ and is denoted as $E_A(\lambda)$. The following result is helpful in many situations.

Proposition 2.20. *If v_1 and v_2 are eigenvectors of A associated respectively to different eigenvalues λ_1 and λ_2, then v_1 and v_2 are linearly independent.*

Suppose that an $n \times n$ matrix A has n linearly independent eigenvectors $v_1, \cdots, v_n$ which are associated to (possible repeated) eigenvalues $\lambda_1, \cdots, \lambda_n$, respectively. Let S be the matrix whose i^{th} column is v_i and Λ be the diagonal matrix whose diagonal entries are $\lambda_1, \cdots, \lambda_n$, respectively. Then we get $A = S\Lambda S^{-1}$. Note that S^{-1} exists due to the fact that $\text{rank}(S) = n$. Conversely, if $A = PDP^{-1}$ for some invertible matrix P and diagonal matrix D (where we would say that A is *diagonalizable* in the future), then one have $AP = PD$ which implies that the columns of P are eigenvectors of A and the diagonal entries of D are eigenvalues whose columns of P are associated to. In conclusion, we have the following result.

Theorem 2.21. *A square matrix A is diagonalizable if and only if A has n independent eigenvectors. Moreover, if A is diagonalizable, then $A = S\Lambda S^{-1}$ where S and Λ are defined as above.*

We may obtain the diagonalizability for certain matrices.

Proposition 2.22. *If A has n different eigenvalues, then it is diagonalizable.*

Proposition 2.23. *If A is symmetric, then it is diagonalizable.*

One advantage for a diagonalizable matrix A is that its matrix power is very efficiently computed. That is, if $A = S\Lambda S^{-1}$, then

$$A^k = \underbrace{AA\cdots A}_{k \text{ times}} = \underbrace{(S\Lambda S^{-1})(S\Lambda S^{-1})\cdots(S\Lambda S^{-1})}_{k \text{ times}} = S\Lambda^k S^{-1},$$

where Λ^k can be calculated by $\Lambda^k = \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix}$.

§ 3. General vector spaces

There are sets which possess the same structures as the Euclidean spaces. Hence, in this chapter, we introduce the concept of a general *vector space* which is defined under the influence of the algebraic structures that the Euclidean spaces have. That being said, on $\mathbb{R}^n$ we have the vector addition $p+q$ and the scalar multiplication $\alpha \cdot p$, for any vectors $p, q \in \mathbb{R}^n$ and any scalar $\alpha \in \mathbb{R}$.

We are not framed only to the real territory. In fact, we will consider both the real field $\mathbb{R}$ and the complex field $\mathbb{C}$. The concept of a vector space to be introduced, will be over either the real or complex fields.

Definition 3.1. A set V, together with the addition "+" and scalar multiplication "·", is said to be a *vector space* (or a *linear space*) over the field $\mathbb{F}$ ($\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$) if V it is closed under $+$ and $\cdot$ and all the following properties hold for all $p, q, r \in V$, and all scalars $\alpha, \beta \in \mathbb{R}$:

1. $p + q = q + p$ (commutativity of addition),
2. $(p + q) + r = p + (q + r)$ (associativity of addition),
3. There is a vector $0 \in V$ such that $p+0 = p$ for every p (existence of zero vector),
4. For every vector p there is an associated vector $-p \in V$ such that $p + (-p) = 0$ (existence of additive inverse),
5. $1 \cdot p = p$ (rule of multiplication by 1),
6. $\alpha(\beta \cdot p) = (\alpha\beta)\cdot$ (associativity of multiplication by scalars),
7. $(\alpha + \beta)\cdot = \alpha \cdot p + \beta \cdot p$ (first distributive law),
8. $\alpha \cdot (p + q) = \alpha \cdot p + \alpha \cdot q$ (second distributive law).

We sometimes write $(V, +, \cdot)$ to indicate all the algebraic operations of V. In most cases, we omit the symbol "·" of the scalar multiplication, e.g. $\alpha \cdot x$ instead of αx.

Remark. From the above definition, it is not automatic that $-p = (-1) \cdot p$. However, it follows from the existence of an additive inverse and the first distributive law that

$p + (-p) = 0 = (1-1)p = p + (-1) \cdot p$. Therefore, we incidentally get $-p = (-1) \cdot p$ coincides. We always write $q - p = q + (-p)$ in this case.

In many cases, we do not need to specify the field $\mathbb{F}$ that V is defined over. In some other cases, it might be necessary to indicate. If V is a vector space over $\mathbb{R}$, we say that V is a *real* vector space. In the other case, we say that V is a *complex* vector space. We also skip the indication that a vector space is real or complex if it is clear from the context.

Of course, every Euclidean space $\mathbb{R}^n$ is the first and obvious example of a vector space. For other examples, we have to be careful about the definition of the addition "+" and the scalar multiplication "·". Let us consider some other important vector spaces apart from the Euclidean spaces.

Example 3.2.

1. Any subspace of $\mathbb{R}^n$ is a real vector space.
2. $\mathbb{C}^n$ over the field $\mathbb{C}$ is a complex vector space.
3. $\mathbb{C}^n$ over the field $\mathbb{R}$ is a real vector space.
4. The set $\mathcal{M}_{m,n}$ (also denoted with $\mathbb{R}^{m\times n}$) of all $m \times n$ real matrices is a vector space with the usual matrix addition and scalar multiplication.
5. If V and W are vector spaces, then the set $\mathscr{F}(V, W)$ of all mappings from V into W is a vector space.
6. The set $\mathscr{S}(\mathbb{R})$ of all real sequences is a real vector space.
7. The set $\mathscr{S}(\mathbb{C})$ of all real sequences is a vector space (real or complex, depending on the choice of field).

Let us now give the general definition of a (vector) subspace.

Definition 3.3. If W is a vector space which is a subset of another vector space V with the inherited algebraic operations, then we call W a *vector subspace* or simply a *subspace* of V.

We have a relatively simple way to verify a vector subspace.

Theorem 3.4. *If a set W is a subset of a known vector space V, then it is a vector subspace if $\alpha p + q \in W$ for any $p, q \in W$ and $\alpha \in \mathbb{R}$.*

Note that if W is not a vector (sub)space, it is often easier to use Definition 3.1 directly. Let us consider some interesting subspaces in the next example.

Example 3.5.

1. The set $\mathcal{S}_n$ of all $n \times n$ symmetric matrices is a vector space.

2. The set $\mathcal{T}_n^{\mathcal{U}}$ of all $n \times n$ diagonal matrices is a vector space.

3. Fix a nonzero matrix $X \in \mathcal{M}_{n,n}$. Then the set $\mathcal{C}(X) = \{A \in \mathcal{M}_{n,n}, \mid AX = XA\}$ is a vector space.

4. The set $C(\mathbb{R}, \mathbb{R})$ of continuous functions from $\mathbb{R}$ into itself is a vector space.

5. The set $C(\mathbb{C}, \mathbb{C})$ of continuous functions from $\mathbb{C}$ into itself is a vector space.

6. The set $C^1(\mathbb{R}, \mathbb{R})$ of all continuously differentiable functions is a vector space with usual addition and scalar multiplication.

7. The set $\mathcal{P}(\mathbb{R}, \mathbb{R})$ of all polynomial functions is a vector space.

The following are examples of a non-subspace.

Example 3.6.

1. The set $\mathcal{I}_n$ of all $n \times n$ invertible matrices is *not* a vector space.

2. The set $\mathcal{P}^n(\mathbb{R}, \mathbb{R})$ of all polynomial functions of degree n is a *not* vector space.

Let us demonstrate the gererality of a vector space so that the attached albegraic operations does not need to be the usual one that we commonly use.

Example 3.7 (Vector space with weird algebra)**.** Let us consider the set $\mathbb{R}^+$. If $\mathbb{R}^+$ is equipped with the usual additiona and (scalar) multiplication, then it is not a vector space. However, if we equip a new (weird) addition $\oplus$ and multiplication $\odot$ defined by $x \oplus y = xy$ and $\alpha \odot x = x^\alpha$ for $x, y, \alpha \in \mathbb{R}$, then $(\mathbb{R}^+, \oplus, \odot)$ is a (real) vector space.

3.1 Basis and dimension

The notion of linear independence, span, basis and dimension can be directly extended to a general vector space but they can behave much differently.

Let V be a vector space over the field $\mathbb{F}$ and $B \subseteq V$. Then span(B) is the set of all finite linear combinations of vectors in B, i.e.

$$\operatorname{span}(B) := \left\{ \sum_{i=1}^{k} \alpha_i v_i \,|\, k \in \mathbb{N}\; v_i \in B,\ \alpha_i \in \mathbb{F},\ i = 1, \cdots, k \right\}.$$

Notice that B can be an infinite set, but span(B) consists only of finite linear combinations. We may immediately see that span(B) is a vector subspace of V. We say that B is a linearly independent set if for any finite subset $\{v_1, \cdots, v_k\} \subseteq B$, the equation $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$ implies $\alpha_i = 0$ for all $i = 1, \cdots, k$.

Definition 3.8. Let V be a vector space. A subset $B \subseteq V$ is a *basis* of V if it is linearly independent and span$(B) = V$.

From the definitions of a basis and a span, it is required that each element of V must be representable with a finite linear combination of elements from B. Let us see some examples in simple cases.

Example 3.9.

1. The $\{e_1, \cdots, e_n\}$ is a basis for $\mathbb{R}^n$.
2. The set $\{1\}$ is a basis for $\mathbb{C}$.
3. The set $\{e_1, \cdots, e_n\}$ is a basis for $\mathbb{C}^n$.
4. The set $\{E_{11}, E_{12}, \cdots, E_{ij}, \cdots, E_{mn}\}$, where E_{ij} is the matrix whose ij-entry is 1 and 0 elsewhere, is a basis for $\mathcal{M}_{mn}$.
5. The set $\{E_{11}, E_{22}, \cdots, E_{nn}\} \cup \{E_{ij} + E_{ji} \,|\, i = 1, \cdots, n,\ j > i\}$ is a basis for $\mathcal{S}_n$.

Now let us look at the vector spaces like $\mathscr{F}(\mathbb{R}, \mathbb{R})$ or $\mathscr{S}(\mathbb{R})$. It is not obvious to choose a basis for such vector spaces, nor to see that a basis exists at all. For this, we cling to the following elegant theorem.

Theorem 3.10. *Every vector space has a basis.*

Though the theorem gives an affirmative answer to the existence of a basis, it does not give any information about its construction due to the fact that it is based on the Zorn's lemma.

Regardless of the choice of basis, we have the following result which guarantees the consistence of number of elements in a basis.

Proposition 3.11. *If a vector space V has a basis B consisting of n vectors. Then any basis of V consists of n vectors.*

This leads to the definition of a vector space dimension.

Definition 3.12. A vector space V is said to have a *finite dimension* with $\dim(V) = n$ if its basis consists of exactly n vectors. If V does not have a finite dimension, then we say that it has an *infinite dimension* and write $\dim(V) = \infty$.

It is important to note that the same vector space can has different dimensionality depending on the field it is defined over.

Example 3.13.

1. The vector space $V = \mathbb{C}$ over the field $\mathbb{C}$ has a basis $\{1\}$. So $\dim(\mathbb{C}) = 1$ when $\mathbb{C}$ is defined over $\mathbb{C}$.

2. The vector space $V = \mathbb{C}$ over the field $\mathbb{R}$ has a basis $\{1, i\}$. So $\dim(\mathbb{C}) = 2$ when $\mathbb{C}$ is defined over $\mathbb{R}$.

One might be able to guess that the vector spaces $\mathscr{F}(\mathbb{R}, \mathbb{R})$ or $\mathscr{S}(\mathbb{R})$ that we cannot find their bases are infinite dimensional. However, not every function space has complicate basis or even being infinite dimensional. Let us see the next examples.

Example 3.14.

1. The set $\{x^k \mid k \in \mathbb{N}\}$ is a basis for $\mathcal{P}^{\leq n}$. Therefore, $\dim(\mathcal{P}^{\leq n}) = \infty$.

2. If $\mathcal{P}^{\leq n}(\mathbb{R}, \mathbb{R})$ is a vector space of all polynomials of degrees $\leq n$. Then the set $\{x^k \mid k \leq n\}$ is a basis for $\mathcal{P}^{\leq n}(\mathbb{R}, \mathbb{R})$ and hence $\dim(\mathcal{P}^{\leq n}(\mathbb{R}, \mathbb{R})) = n$.

We shall end this section by remarking the existence of a better concept of basis, called the Schauder basis, which allows an infinite linear combination to be used. However, it requires the notion of convergence which is not included in this note. Additionally, not every infinite dimensional vector space admits a Schauder basis.

3.2 Linear operators

It is natural to also extend the notion of linearity of mappings to any vector spaces. The term "operator" is often preferred over "transformation" in general vector space setting. However, any terms operator, transformation, mapping, map, etc. are all correct.

Definition 3.15. Let V and W be two vector spaces over the same field $\mathbb{F}$. A mapping $T : V \to W$ is called a *linear operator* if

$$T(\alpha p + \beta q) = \alpha T(p) + \beta T(q) \tag{3.1}$$

for all $p, q \in V$ and $\alpha, \beta \in \mathbb{R}$.

Let us see some examples of linear operator outside the Euclidean settings.

Example 3.16.

1. Take any matrices $X_{c\times m}$ and $Y_{n\times d}$. Then $T : \mathcal{M}_{m,n} \to \mathcal{M}_{c,d}$ defined by

$$T(A) = XAY \qquad (A \in \mathcal{M}_{m,n})$$

is a linear operator.

2. The mapping $T : \mathcal{M}_{m,n} \to \mathcal{M}_{n,m}$ defined by

$$T(A) = A^\top \qquad (A \in \mathcal{M}_{m,n})$$

is a linear operator.

3. The mapping $T : C^1(\mathbb{R}, \mathbb{R}) \to C(\mathbb{R}, \mathbb{R})$ defined by

$$T(f) = \frac{df}{dx} \qquad (f \in \mathcal{D}(\mathbb{R}, \mathbb{R}))$$

is a linear operator.

4. The shift operator $T : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R})$ defined by

$$T((x_1, x_2, \cdots)) = (0, x_1, x_2, \cdots) \qquad ((x_1, x_2, \cdots) \in \mathscr{S}(\mathbb{R}))$$

is a linear operator.

Let us, on the contrary, see a non-linear operator example.

Example 3.17. Consider the linear mapping $T : \mathcal{F}(\mathbb{R}, \mathbb{R}) \to \mathcal{F}(\mathbb{R}, \mathbb{R})$ defined by

$$T(f) = f^2 \qquad (f \in \mathcal{D}(\mathbb{R}, \mathbb{R}))$$

is *not* a linear transformation.

In some occasions, the vector spaces V and W can be of different fields. In particular, the problem can be raised only when the domain V is defined over the complex field $\mathbb{C}$. In this case we need to specify over which field the scalars α, β are taken from. If α and β are taken from $\mathbb{R}$ and (3.1) holds, we say that T is a linear operator over $\mathbb{R}$ or that it is $\mathbb{R}$-linear. Let us look at an example.

Example 3.18. The complex conjugation is $\mathbb{R}$-linear but not $\mathbb{C}$-linear.

3.3 *Matrix representation*

If V and W are finite dimensional vector spaces over the same field $\mathbb{F}$, then we can represent any linear operator between them by a matrix.

Suppose that $\dim(V) = n$ and $B = \{b_1, \cdots, b_n\}$ is the chosen basis for V. Then every element v of V can be represented uniquely with a linear combination of $b_1, \cdots, b_n$ as

$$v = v_1 b_1 + \cdots + v_n b_n$$

with $v_1, \cdots, v_n \in \mathbb{F}$. In this case, we adopt the vectorial representation $v = (v_1, \cdots, v_n)_B^\top$. Similarly for W, let $H = \{h_1, \cdots, h_m\}$ be the chosen basis for W and we have the vectorial representation of any element $w \in W$ with respect to the basis H written as $w = (w_1, \cdots, w_m)_H^\top$ if $w = w_1 h_1 + \cdots + w_m h_m$ with $w_1, \cdots, w_m \in \mathbb{F}$. The linear operator $T : V \to W$ then has a matrix representation with respect to the bases B and H by the column vectors $T(b_1), \cdots, T(b_n)$, each written in the vectorial form with respect to H. That is, if A_T is the matrix given by

$$A_T := \begin{bmatrix} | & | & & | \\ T(b_1) & T(b_2) & \cdots & T(b_n) \\ | & | & & | \end{bmatrix}.$$

Therefore, we get

$$T(v) = A_T v, \quad \text{or more precisely} \quad [T(v)]_H = [A_T]_B^H [v]_B$$

where the input and output vectors v and $T(v) = Av$ are written in the bases B and H, respectively. The matrix A_T, or more precisely $[A_T]_B^H$, is called the *matrix representation* of T with respect to the bases B and H.

Definition 3.19. A linear operator $T : V \to W$ is *invertible* if there exists a unique linear operator $T^{-1} : W \to V$ such that $T^{-1} \circ T(v) = v$ and $T \circ T^{-1}(w) = w$ for all $v \in V$ and $w \in W$.

If V and W are both finite dimensional, then T has a matrix representation A_T (with respect to some chosen bases). In this case, we have T is invertible if and only if A_T is an invertible matrix. In fact, all the theory of matrices in the earlier chapter are all applicable to this the matrix representation. If one of V or W is infinite dimensional, then the situation can be much more complicated and we will not discuss here.

3.4 Vector space of linear operators and duality

In this final section, we would like to expose the vector space $\mathscr{L}(V, W)$. Suppose that V and W are vector spaces over the same field $\mathbb{F}$. Then the set $\mathscr{L}(V, W)$, consisting of all linear operators from V into W, is a vector space.

There is a simplification in finite dimensional setting. Assume that $\dim(V) = n$ and $\dim(W) = m$. Let us fix the bases B and H for V and W, respectively. Then the each $T \in \mathscr{L}(V, W)$ has a matrix representation A_T in $\mathcal{M}_{mn}$. This is coincidence is so nice and in fact we have the following property.

Proposition 3.20. *There is a vector space isomorphism $T \mapsto A_T$ between $\mathscr{L}(V, W)$ and $\mathcal{M}_{mn}$.*

The isomorphism property says that the two vector spaces are essentially the same. However the isomorphism is not canonical in the sense that it depends on the choice of bases on V and W.

There is a case which is extremely useful in functional analysis, namely the (algebraic) dual space. Note that this is not the same as the topological dual which more commonly adopted.

Definition 3.21. Let V be a vector space over the field $\mathbb{F}$. The algebraic dual of V is the vector space $V^* = \mathscr{L}(V, \mathbb{F})$.

Of course, if V is finite dimensional with $\dim(V) = n$, then V^* can just be identified with the set of row vectors $\mathbb{F}^{1\times n}$. More importantly, if $B = \{b_1, \cdots, b_n\}$ is a basis for V, then we can create a basis B' on V^* in correspondence to B by letting $B' = \{b_1^*, \cdots, b_n^*\}$ with each b_i^* defined by

$$b_i^*(c_1 b_1 + c_2 b_2 + \cdots + c_n b_n) := c_i.$$

This basis B' is called the *dual basis* of V'.

We end the lecture with a remark for infinite dimensional setting. If $\dim(V) = \infty$ and B is a basis for V, then B' constructed in the same way is linearly independent but not necessarily spans V'. Hence B' may not be a basis in this case. In particular, we may view the construction of B' as an injection of V into V' which mnay not be surjective. We may roughly view this situation that V' has a greater infinite dimension compared to V.

Concluding remarks

In this note, we have discussed the main idea of linear algebra developed upon the study of linear systems and the corresponding matrix equation. The structure we found in Euclidean spaces is then extended to an abstract formulation which applies largely to various settings and turns into a building block of functional analysis. For more detailed discussion, the reader may consult the textbooks listed in the References section.

References

[1] S. Axler. *Linear algebra done right*. Cham: Springer, 2015. ISBN 978-3-319-11079-0; 978-3-319-11080-6. doi: 10.1007/978-3-319-11080-6.

[2] N. Johnston. *Advanced linear and matrix algebra*. Cham: Springer, 2021. ISBN 978-3-030-52814-0; 978-3-030-52815-7. doi: 10.1007/978-3-030-52815-7.

[3] N. Johnston. *Introduction to linear and matrix algebra*. Cham: Springer, 2021. ISBN 978-3-030-52810-2; 978-3-030-52811-9. doi: 10.1007/978-3-030-52811-9.

[4] S. Lang. *Linear algebra. 3rd ed.* Springer, Cham, 1987. ISBN 0-387-96412-6.