Understand linear transform (1)

Yinan Huang

Sep 26, 2020

Contents

Abstractness is the price of generality.

John Doe

1 Preface

Linear algebra is a fundamental course at undergraduate level. But for me, it was really difficult to understand what linear algebra is when I first learned it. All I learned from my first linear algebra course were simply technics and tricks of calculation of matrices. I did not feel the beauty of it until I took my quantum mechanics course, for which we used abstract linear algebra a lot. Dirac notations have unbelievable magic and it plays a important role of describing quantum theory. This fall I am taking “vector space methods”, which is an advanced linear algebra course. I think It is a good time to have a review of linear algebra a little bit, especially on topics of linear tranform and their representation. We will begin with abstract linear space definition and head to linear transforms, and try to talk about matrix representation.

Video series “Essence of linear algebra”” is highly recommended by the author, which gives vivid examples and insights to understand linear algebra intuitively. And this note tries to associate these intuitive ideas with abstract linear algebra, and I hope it will be helpful to build a good understanding of what linear algebra is.

2 Abstract vector space

We would like to start with the abstract description of vector space. We do not care about what “vector” really is. Actually it can be anything, so long as the vector space obeies certain rules.

Definition 2.1. *A set V is called a* ***vector space*** *on field F, if V meets the vector space axioms. On a vector space, we have vector addition $v + w$ and scalar multiplication cv. Equivalently, V is a vector space if for any scalar $s \in F$ and any vector $v, w \in V$, vector $sv + w \in V$.*

We are not going to list the axioms here, which can be found in any linear algebra book. Conceptutally, vector can be anything as long as they meet the axioms.

2.1 Linear independence, basis, dimension

Since we define vector addition and scalar multiplication on a vector space, we naturally desire to study properties of these operations on a vector space.

Definition 2.2. *A list of vector $\{v_i\}$ is said to be* ***linear independent****, if and only if linear combination $\sum_i s_i v_i = 0$ implies $s_i = 0$ for all i. If vectors are not linear independent, they are called* ***linear dependent****.*

In other words, linear independence of $\{v_i\}$ implies we cannot find $s_i \neq 0$ such that $\sum s_i v_i = 0$. But why do we care about linear independence? We will see in a second.

Definition 2.3. *A list of vector $\{v_i\}$* ***spans*** *vector space V, or say V is the span of $\{v_i\}$, if for any $v \in V$, $v = \sum_i c_i v_i$.*

Definition 2.4. *A list of vectors $\{v_i\}$ is called the* ***basis*** *of vector space V, if $\{v_i\}$ spans V and $\{v_i\}$ is linear independent.*

Theorem 2.1. *If $\{v_i\}$ is a basis of V and $v \in V$ is any vector. Then linear combination $v = \sum_i s_i v_i$ is **unique**.*

Proof 2.1. *Suppose we have two linear combination $v = \sum_i s_i v_i$ and $v = \sum_i u_i v_i$, then $0 = \sum_i (s_i - u_i) v_i$. Since $\{v_i\}$ is the basis and thus linear independent, so by definition $s_i = u_i$. So the linear combination is unique.*

We see that linear independence of basis leads to uniqueness, which is one of the most vital properties of basis. Because of linear independence, we thus can construct basis along with its wonderful properties.

Theorem 2.2. *If $\{v_1, v_2, ..., v_n\}$ is the basis of V and $\{w_1, w_2, ..., w_m\}$ is linear independent, then $m \leq n$.*

Proof 2.2. *Since $\{v_i\}$ is the basis, so any vector including w_i can be expressed by $\{v_i\}$. We suppose $m > n$, then $\{w_1, v_2, ..., v_n\}$ is also the basis by definition. Repeating replacing v_i by w_i, we finally conclude that $\{w_1, w_2, ..., w_n\}$ is the basis, so $\{w_{n+1}, w_{m+2}, ..., w_m\}$ shoule be the linear combination of $\{w_n\}$, which implies $\{w_1, w_2, ..., w_m\}$ is not linear independence and is a contradiction. So we must have $m \leq n$.*

The theorem shows that a basis has the maximum length of any linear independent vector list. It turns out the length of any basis of a vector space V is fixed, and it allows us to define dimension.

Theorem 2.3. *Let $\{v_i\}$ and $\{w_i\}$ are two basis of V. Then length of $\{v_i\}$ and length of $\{w_i\}$ is the same.*

Proof 2.3. *We can apply the theorem we just proved based on a basis must be linear independent. Since $\{w_i\}$ is linear independent and $\{v_i\}$ is basis of V, so len$\{v_i\} \geq$ len$\{w_i\}$. Since $\{v_i\}$ is linear independent and $\{w_i\}$ is basis of V, so len$\{w_i\} \geq$ len$\{v_i\}$. To sum up, len$\{v_i\} =$ len$\{w_i\}$.*

Definition 2.5. *The length of a basis $\{v_i\}$ of V is called the **dimension** of V.*

2.2 Change of basis

Basis is powerful since it can express any vector in space. Suppose we have basis $\{\hat{e}_i\}$. Then any vector $v \in V$ can be expressed in terms of their linear combination:

$$v = \sum_i v_i \hat{e}_i, \tag{1}$$

where v_i is called the **components** of v in basis $\{\hat{e}_i\}$. What if we want to use another basis? Suppose another basis $\{\hat{e}'_i\}$ is defined by

$$\hat{e}'_j = \sum_i A_{i,j} \hat{e}_i. \tag{2}$$

$A_{i,j}$ is the component of new basis in old basis. Before moving on, a question is posted: what properties should A has such that $\hat{e}'_j$ is a basis indeed. To be a basis, $\hat{e}'_j$ must be linear independent, so consider a linear combination of $\hat{e}'_j$ such that

$$0 = \sum_j c_j \hat{e}'_j = \sum_i \left(\sum_j A_{i,j} c_j \right) \hat{e}_i. \tag{3}$$

Since $\{\hat{e}_i\}$ is a basis. Thus we conclude that

$$\sum_j A_{i,j} c_j = 0 \Rightarrow \forall j, c_j = 0. \tag{4}$$

Then we want to know the new component v'_i of v in new basis:

$$v = \sum_i v'_i \hat{e}'_i = \sum_i v'_i \sum_j A_{j,i} \hat{e}_j = \sum_j \left(\sum_i A_{j,i} v'_i \right) \hat{e}_j. \tag{5}$$

Remember the linear combination is unique, so we must have

$$v_j = \sum_i A_{j,i} v'_i. \tag{6}$$

What we have done to calculate the different component in two basis is called “change of basis”. It is a frequently used technics in problem solving. Sometimes we call it “passive perspective of transform””, since we simply choose another basis and leave the vectors and space unchanged.

2.3 Linear transform

Now we turn to linear transform. Linear transform is just a fancy word for function in linear space.

Definition 2.6. *Let V and W be vector spaces. Function $T : V \to W$ is called a linear transform, if for any vector $v_1, v_2, v \in V$ and scalar $\alpha \in F$,*

- $T(v_1 + v_2) = T(v_1) + T(v_2)$.
- $T(\alpha v) = \alpha T(v)$.

A linear transform is determined by its transform of basis.

Theorem 2.4. *Let $\{v_i\}$ be a basis of V. Given vectors $\{w_i\} \subseteq W$, we can charaterize a unique transform T defined by $T(v_i) = w_i$.*

Proof 2.4. *Since for any vector v, $T(v) = T(\sum_i c_i v_i) = \sum_i c_i T(v_i) = \sum_i c_i w_i$ is characterized by $\{w_i\}$, so the transform is well-defined ans thus unique.*

The theorem above shows that **An transform is uniquely determined once we give its transform on basis.**

2.3.1 Range space and null space

Definition 2.7. *For a linear transform $T : V \to W$, $R(T) = \{w \in W | \exists v \in V, T(v) = w\}$ is called the **range space** of T, and $N(T) = \{v \in V | T(v) = 0\}$ is called the **null space** or kernel of T.*

Theorem 2.5. *Let $\{v_i\}$ be basis of V. $\dim(N(T)) \neq 0$ if and only if vector list $\{w_i | w_i = T(v_i)\}$ is linear dependent.*

Proof 2.5. *Suppose $\dim(N(T)) \neq 0$, then it implies that there exist at least vector $v \neq 0$ such that $T(v) = 0$. Using the basis $\{v_i\}$, we expand v as $v = \sum_i c_i v_i$ and plug into $T(v) = 0$, which turns out to be $T(v) = \sum_i c_i T(v_i) = \sum_i c_i w_i = 0$ with not all $c_i = 0$. By definition, $\{w_i\}$ is linear dependent.*

Theorem 2.6. *Let $R(T)$ and $N(T)$ be range space and null space of T. Then* $\dim(R(T)) + \dim(N(T)) = \dim(V)$.

Proof 2.6. *We shall see that* $\dim(N(T))$ *is actually decided by whether* $\{w_i\}$ *is linear independent. If they are, then* $\dim(N(T)) = 0$. *If there are m unnecessary w_i, then* $\dim(N(T)) = m$. *If* $\dim V = n$*, then* $R(T) = span\{w_i\} = n - m$. *Therefore* $\dim(R(T)) + \dim(N(T)) = n - m + m = n = \dim(V)$.

2.3.2 Eigenvalues and eigenvectors

Definition 2.8. *Let $T : V \to V$ be a linear transform. Equation $T(v) = \lambda v$ is called **eigen equation** of T, and its solution v and corresponding λ are called **eigenvectors** and **eigenvalues**.*

It shows that eigenvectors are special vectors for which transform T acting on them does not change them expect for a scalar scaling. One of the reasons why we are interested in eigenvectors is that they depict a linear transform and it is convenient to work on the basis of eigenvector when dealing with certain transform. Particularly, let us consider transform T has $n = \dim(V)$ linear independent eigenvectors $\{t_i\}$ corresponding to eigenvalue $\{\lambda_i\}$, we can first do a change of basis

$$t_j = \sum_i A_{i,j}\hat{e}_i, \quad \hat{e}_j = \sum_i A^{-1}_{i,j} t_i, \tag{7}$$

where A^{-1} is defined as $\sum_k A^{-1}_{i,k} A_{k,j} = \delta_{i,j}$. and $\sum_k A_{i,k} A^{-1}_{k,j} = \delta_{i,j}$. Then for any vector $v = \sum_i v_i \hat{e}_i$, we find $T(v)$ is

$$T(v) = \sum_i v_i T(\hat{e}_i) = \sum_i v_i T(\sum_j A^{-1}_{j,i} t_j) = \sum_{i,j} v_i A^{-1}_{j,i} \lambda_j t_j = \sum_{i,k} v_i \left(\sum_j A^{-1}_{j,i} A_{k,j} \lambda_j \right) \hat{e}_k. \tag{8}$$

3 Vector space in representation

So our review of abstract vector space is finished and let us turn to the concrete representation. Conceptually, representation is a method to express abstract concepts in concrete objects. To be specific, in vector space, we use **matrix** for representation.

Definition 3.1. *A matrix A is a n by m table defined on field F.*

3.1 Vectors in matrix representation

Vectors can be represented by a column matrix or a row matrix. For this reason, we usually call a column (row) matrix as column (row) vector. As a convention, we will use column vector. To do so, we need a basis first. Suppose we have V and a basis $\{\hat{e}_1, \hat{e}_2, ..., \hat{e}_n\}$, then for any vector $v = \sum_i v_i \hat{e}_i$, we can define its matrix representation as

$$v \to \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ \dots \\ v_n \end{pmatrix} \tag{9}$$

This is reasonable because we can define vector addition and scalar multiplication to the matrix as

$$v + w \to \begin{pmatrix} v_1 \\ v_2 \\ ... \\ ... \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ ... \\ ... \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ ... \\ ... \\ v_n + w_n \end{pmatrix}, \tag{10}$$

and

$$cv \to c \begin{pmatrix} v_1 \\ v_2 \\ ... \\ ... \\ v_n \end{pmatrix} = \begin{pmatrix} cv_1 \\ cv_2 \\ ... \\ ... \\ cv_n \end{pmatrix}. \tag{11}$$

So all the operation to vectors is the same if we do it in representation. This is why it is a legal representation.

One thing to be careful: representation relies on the basis. When we deal with representation, we actually omit the basis information, that is

$$\hat{e}_1 \to \begin{pmatrix} 1 \\ 0 \\ ... \\ ... \\ 0 \end{pmatrix}, \quad \hat{e}_2 \to \begin{pmatrix} 0 \\ 1 \\ ... \\ ... \\ 0 \end{pmatrix}, \quad \hat{e}_n \to \begin{pmatrix} 0 \\ 0 \\ ... \\ ... \\ 1 \end{pmatrix}. \tag{12}$$

So an explicit, rigorous way of representation should be

$$v \to \begin{pmatrix} v_1 \\ v_2 \\ ... \\ ... \\ v_n \end{pmatrix}_{\{\hat{e}_i\}}. \tag{13}$$

3.2 Change of basis in representation

The "change of basis" problem is to change from old basis $\{\hat{e}_i\}$ to new basis $\{\hat{e}'_i\}$ with the relation

$$\hat{e}'_j = \sum_i A_{i,j} \hat{e}_i. \tag{14}$$

This implies if we choose old basis as representation, and new basis are

$$\hat{e}'_i \to \left(\hat{e}'_i\right)_{\{\hat{e}_i\}} = \begin{pmatrix} A_{1,i} \\ A_{2,i} \\ ... \\ ... \\ A_{n,i} \end{pmatrix}_{\{\hat{e}_i\}}. \tag{15}$$

From the previous sections, we know that component of a vector v in different basis is

$$v_j = \sum_i A_{j,i} v'_i. \tag{16}$$

We can list these new basis represented in old basis together to get a basis transform matrix

$$A = \left(\begin{pmatrix} \hat{e}'_1 \end{pmatrix}_{\{\hat{e}_i\}} \quad \begin{pmatrix} \hat{e}'_2 \end{pmatrix}_{\{\hat{e}_i\}} \quad \dots \quad \begin{pmatrix} \hat{e}'_n \end{pmatrix}_{\{\hat{e}_i\}} \right) = \begin{pmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & \dots & A_{2,n} \\ \dots & \dots & \dots & \dots \\ A_{n,1} & A_{n,2} & \dots & A_{n,n} \end{pmatrix}. \tag{17}$$

By carefully defining the matrix multiplication, we find

$$\begin{pmatrix} v'_1 \\ v'_2 \\ \dots \\ v_n \end{pmatrix}_{\{\hat{e}'_i\}} = \left(\begin{pmatrix} \hat{e}'_1 \end{pmatrix}_{\{\hat{e}_i\}} \quad \begin{pmatrix} \hat{e}'_2 \end{pmatrix}_{\{\hat{e}_i\}} \quad \dots \quad \begin{pmatrix} \hat{e}'_n \end{pmatrix}_{\{\hat{e}_i\}} \right) \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix}_{\{\hat{e}'_i\}}. \tag{18}$$

To sum up, to write a vector in a new basis from a old basis, we follow the matrix multiplication

$$\begin{pmatrix} v \end{pmatrix}_{\text{old}} = \left(\begin{pmatrix} \hat{e}'_1 \end{pmatrix}_{\text{old}} \quad \begin{pmatrix} \hat{e}'_2 \end{pmatrix}_{\text{old}} \quad \dots \quad \begin{pmatrix} \hat{e}'_n \end{pmatrix}_{\text{old}} \right) \begin{pmatrix} v \end{pmatrix}_{\text{new}}. \tag{19}$$

Or its inverse version

$$\begin{pmatrix} v \end{pmatrix}_{\text{new}} = \left(\begin{pmatrix} \hat{e}'_1 \end{pmatrix}_{\text{old}} \quad \begin{pmatrix} \hat{e}'_2 \end{pmatrix}_{\text{old}} \quad \dots \quad \begin{pmatrix} \hat{e}'_n \end{pmatrix}_{\text{old}} \right)^{-1} \begin{pmatrix} v \end{pmatrix}_{\text{old}}. \tag{20}$$

Concretely, we first need to know how to represent a new basis in our old basis, and we put the representation of each these new basis in each column of a matrix, and the inverse of the matrix times the vector in old basis gives us the vector in new basis. One thing to be clear, we do not change any vector: what we do is just change a basis, namely using different languages to describe the same vector. We will see that change of basis has very similar formula with linear operation. But linear operation do change the vector but leave the basis unchanged, which is kind of opposite to change of basis.

By the way, to let A a legal basis transform, we require the linear combination of new basis, which in representation, is

$$\sum_i c_i \hat{e}'_i \to \sum_i c_i \begin{pmatrix} \hat{e}'_i \end{pmatrix}_{\text{old}} \tag{21}$$

must be linear independent. In a matrix language, we said the column vectors are linear independent.

3.3 Linear operation in representation

In the previous discussions, we point out that how linear operation change the basis characterize the operation itself. Note that the word "change" here means differently from what we say "change of basis" previously. The subtlety here is that in change of basis, we want to know how to use a brand new basis to represent vector; when we use transform to change vectors, we still use the same basis, but vector is manually changed to another vector.

Consider transform $T : V \to W$. $\{\hat{e}_i\}$ and $\{\hat{\beta}_j\}$ are the basis of T and W respectively.

$$a_j = T(\hat{e}_j) = \sum_i A_{i,j} \hat{\beta}_i. \tag{22}$$

Matrix A describe how T change the old basis in V, namely turn $\hat{e}_j \in V$ to $a_j \in W$. Note that $\{a_j\}$ is not necessarily a basis at all. Appling transform to any vector

$$w = \sum_j w_j \hat{\beta}_j = T(v) = \sum_i v_i T(\hat{e}_i) = \sum_{i,j} v_i A_{j,i} \hat{\beta}_j, \tag{23}$$

for which the representation in basis $\{\hat{e}_i\}$ is

$$\begin{pmatrix} w \end{pmatrix}_W = \left(\begin{pmatrix} T(\hat{e}_1) \end{pmatrix}_W \quad \begin{pmatrix} T(\hat{e}_2) \end{pmatrix}_W \quad \dots \quad \begin{pmatrix} T(\hat{e}_n) \end{pmatrix}_W \right) \begin{pmatrix} v \end{pmatrix}_{V=\{\hat{e}_i\}}. \tag{24}$$

So as we can see, we put $T(\hat{e}_i)$ (called arrival vectors) in the column of the matrix, and this matrix times the old vector gives us a new vector after operation.

Note that this formula is kind of similar to change of basis, especially in the case that $V = W$ and $\{a_j\}$ can form a basis. It seems these two processes are inverse to each other, which is called duality. Note that rotate the basis clockwise is equivalent to rotate vectors counter clockwise in terms of the component of vectors. Though it turns out we get different vectors, but the component of vector in new basis and the component of new vector in old basis is the same. Though it may be mouthful, but just remember two key things in linear algebra when you write matrices: basis and vectors.

It is convenient to use transform's representation A for discussion about range space and null space. The column space is just the range space in representation (depends on how you choose basis in W) and the null space is just the solution x of $Ax = 0$, for which vector x is the representation of vectors in null space (depends on how you choose basis in V).

3.3.1 Eigenvectors and similarity transform

We restrict us to transform $T : V \to V$. So transform's representation is a square matrix. By choosing a basis $\{\hat{e}_i\}$, the eigenequation in representation is

$$T(t_i) = \lambda_i t_i \to \left(\begin{pmatrix} T(\hat{e}_1) \end{pmatrix}_{\{\hat{e}_i\}} \quad \begin{pmatrix} T(\hat{e}_2) \end{pmatrix}_{\{\hat{e}_i\}} \quad \dots \quad \begin{pmatrix} T(\hat{e}_n) \end{pmatrix}_{\{\hat{e}_i\}} \right) \begin{pmatrix} t_i \end{pmatrix}_{\{\hat{e}_i\}} = \lambda_i \begin{pmatrix} t_i \end{pmatrix}_{\{\hat{e}_i\}}. \tag{25}$$

Suppose T has n linear independent eigenvectors, which means $\{t_i\}$ forms a basis. So what if we change our old basis to $\{t_i\}$? From our previous knowledge, we first construct a basis transform matrix by defining

$$U \equiv \left(\begin{pmatrix} t_1 \end{pmatrix}_{\{\hat{e}_i\}} \quad \begin{pmatrix} t_2 \end{pmatrix}_{\{\hat{e}_i\}} \quad \dots \quad \begin{pmatrix} t_n \end{pmatrix}_{\{\hat{e}_i\}} \right). \tag{26}$$

One way to find T in new representation is to use the fact that $T(t) = \lambda t$ does not depend on any representation (a tensor equation). So

$$T_{\{\hat{e}_i\}} t_{\{\hat{e}\}} = \lambda t_{\{\hat{e}\}} \Leftrightarrow T_{\{t_i\}} t_{\{t_i\}} = \lambda t_{\{t_i\}}. \tag{27}$$

Note that $t_{\{\hat{e}_i\}} = U t_{\{t_i\}}$, and thus

$$T_{\{t_i\}} = U^{-1} T_{\{\hat{e}_i\}} U. \tag{28}$$

This is called a **similarity transform**, which transform the matrix representation of a transform from one basis to another basis. In this special case, $T_{\{t_i\}}$ is a diagonal matrix by definition, so

$$\left(\begin{pmatrix} T(\hat{e}_1) \end{pmatrix}_{\{\hat{e}_i\}} \quad \begin{pmatrix} T(\hat{e}_2) \end{pmatrix}_{\{\hat{e}_i\}} \quad \dots \quad \begin{pmatrix} T(\hat{e}_n) \end{pmatrix}_{\{\hat{e}_i\}}\right) = U \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \dots & \\ & & & \lambda_n \end{pmatrix} U^{-1}. \quad (29)$$

There is a intuitive way to explain this equation. U^{-1} is a change of basis, which change representation from old basis to new basis, and then we apply the transform represented in new basis, and then we change the basis to the old one using U. This process is exactly what a transform likes in old basis.

3.3.2 Invertibility

Finally we want to talk a little bit about when inverse of a transform (and matrix) exists.

Theorem 3.1. *Linear transform $T : V \to V$ is inversible if and only if for some basis $\{\hat{e}_i\} \subseteq V$, vector list $\{T(\hat{e}_i)\}$ is linear independent.*

Proof 3.1. *From what we know about functions, a function is inversible if and only if it is surjective and injective. In order to be surjective, the arrival vectors $\{T(\hat{e}_i)\}$ must span V; to be injective, for any $v \neq v'$, we must require $T(v) - T(v') = \sum_i (v - v')T(\hat{e}_i) \neq 0$, which is equivalent to say $if v_i - v'_i \neq 0$, then linear combination $\sum_i (v_i - v'^i)T(\hat{e}_i) \neq 0$. This means $\{T(\hat{e}_i)\}$ is linear independent. In this case, linear independent implies span, so linear independence of $\{T(\hat{e}_i)\}$ makes sure that T is reversible. Note that this is equivalent to say $R(T) = V$ and $N(T) = \{0\}$. In this case we call the transform **full rank** (rank is just the dimension of range space). Does the selection of basis $\{e_i\}$ matter? Intuitively, invertibility is the property of a transform, which does nothing with the representation. But a careful argument can be made by looking at the matrix representation: a change of basis is just a similar transform, which obviouly does not have impact on invertibility.*

Theorem 3.2. *A n by n matrix A is inversible if and only if it is full rank.*

Proof 3.2. *A matrix is just the representation of an transform in some basis. So the theorem is essentially the same as the thorem about inversible transforms above.*

4 Summary

When we want to arrive some abstract and general conclusion, it is always easy to work with the abstract vector space. When dealing with the concrete calculation, we always choose some basis and do calculation based on these basis. The beauty of linear algebra is that it gives us the ability to switch between the abstract perspective and concrete view. We finish by an example, a matrix equation

$$Ax = b, \quad (30)$$

where A is a n by m matrix and x, b are two vectors. Using the abstract perspective, we know that A is a representation of an transform $T : V \to W$. We see each column of A is an arrival vector transformed from basis in V and expressed in basis in W. If $\dim(V) = \dim(W)$, we can further talk about the invertibility of A by looking at its range space (column space). If A is inversible, then it implies there is no solution of $Ax = 0$ except $x = 0$, since A is full rank and its null space is of zero dimension.