Linear Algebra

How does it develop and why do we study this?

WANG, Anbang

WANG, Anbang
awangas@connect.ust.hk

Preface

Linear algebra is the fundamental introduction of algebra. However, when I first learned this course, I usually found myself confused. In this course, there are quite a number of concepts that do not seem to be useful when you first learn them, but later when I was going to use them, I've already forgotten the specific content.

Later I found out, if we understand what's the propose of the content we are studying, it itself will be very helpful! That's why I've decided to write this "lecture notes". In fact, we don't have lectures. (Maybe in the future?) But I hope this material will be helpful for the one's who are going to learn this subject, or someone who have already learnt this but usually find it confusing about "What is linear algebra", the latter usually conclude that this course is used for solving linear equations.(laugh)

I gave the small book a subtitle called "How does it develop and why do we study this?". For the first part, I will try my best to show you the idea of developing the concepts. And what I want to mention about the second part is that, the "usefulness" of mathematics usually appears in both "mathematics" propose and "application" propose. They are, in fact, somehow connected. For example, the invention of imaginary number $\mathtt{i}$, was invented for the "existence" of root of a quadratic equation. And this invention is also very useful for the later application, especially in EE or some other industries. This is also true for linear algebra. I will include adequate examples to elaborate this point.

The content of the book mainly comes from lecture notes of two of my professors, Professor Ivan Ip and Professor YAN Min in HKUST. Thanks for their courses for me to learn this course.

Now let's begin!

Content

4 | Vector Space

9 | linear transformation

Vector Space

1.1 Function and linear transformation

If you ask me, what's linear algebra studying for? My answer would be: it studies the property of linear transformation. We begin with the more general version: a function.

Definición 1.1. For given two set V and W, if $\forall \mathbf{v} \in V, !\exists \mathbf{w} \in W$ corresponding to it, we define the correspondence (also called mapping) as a function from V to W.

Example. $f(x) = x + 1, \mathbf{R} \to \mathbf{R}$

Example. $\frac{d}{dx} f(x) = f'(x), \mathbf{C}^\infty \to \mathbf{C}^\infty$, where C^∞ mean the infinitely times differentiable functions.

The second example might seems a little weird, it's a function that takes in a function and returns a function. But, checking the definition, the differentiation function is well-defined, since a differentiable function can have only one derivative.

In case of linear algebra, we usually focus on one specific type of function, or we call transformation. And the linear transformation is defined as following:

Definición 1.2. if T is a function that mapping from V to W, $T: V \to W$, and

$$\forall \mathbf{v}_1, \mathbf{v}_2 \in V, \forall c \in \mathcal{F}, T(\mathbf{v}_1) + cT(\mathbf{v}_2) = T(\mathbf{v}_1 + c\mathbf{v}_2)$$

We say that T is a linear transformation.

Here, the $\mathcal{F}$ is a field set, but we can simply consider it as a "factor", in this notes, the set can be $\mathbf{R}$ or $\mathbf{C}$, the real numbers or complex numbers.

This definition seems good enough, and enough intuitive, but something bad might happen if $\mathbf{v}_1 + c\mathbf{v}_2 \notin V$, then T is not a well defined function. Then we have the definition of vector space, which ensures the inner operation will not "get out of range".

1.2 Vector space

Definición 1.3. A vector space over a field $\mathcal{F}$ is a set that

$$\forall \mathbf{v}_1, \mathbf{v}_2 \in V, \forall c \in \mathcal{F}, \mathbf{v}_1 + c\mathbf{v}_2 \in V$$

One direct consequence is that the unique zero vector must exist in the vector space. We can show that by taking $\mathbf{v}_1 = \mathbf{v}_2, c = -1$.

Following are examples of vector spaces:

Example. The zero vector space $\{\mathbf{0}\}$.

Example. Euclidean space, $\mathbf{R}^n$, which is a n-tuple consisting of n real numbers. With element-wise adding and scaling. (When we say scaling, we mean multiply the original vector by a number).

We usually write the Eucliden vector in this way:

$$\mathbf{v} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \in \mathbf{R}^n$$

Example. The set $\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : x + y + z = 0\} \subset \mathbf{R}^3$

Example. All continuous functions defined on $\mathbf{R}$.

This is a counter example of vector space.

Counter Example. he set $\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : x + y + z = 1\} \subset \mathbf{R}^3$

Now that we have a special set for linear transformation. Next the question is, how do we describe the transformation? Since a non-zero vector space has infinitely many vectors, it seems foolish to express them one by one. Our goal is now try to find a way to use less vectors to represent the transformation, in the best case, using finitely many vectors. To do this, we must dig deeper into some properties of vector space.

1.3 Subspace and Span

Consider a plane in a 3-D space, say the set $\{\begin{pmatrix} x \\ y \\ z \end{pmatrix} : x+y+z=0\} \subset \mathbf{R}^3$ We also notice that this set itself is also a vector space. Then we call the set a subspace.

> **Definición 1.4.** Subspace is a subset of a vector space, which itself is also a vector space.

Subspace is a good thing. It uses several unordered vector and fill the whole area making it "symmetric" and closed. It will later be used for decomposing the vector space, but we can leave it later.

Since subspace is a good thing, we want to find a way to generate a subspace using an arbitrary set of vectors. That's when "spanning set" comes in.

> **Definición 1.5.** Let $\mathcal{S}$ be a subset of a vector space V. 1.A linear combination of $\mathcal{S}$ is any finite sum of the form:
> $$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n \in V$$
> 2.We define the spanning set of $\mathcal{S}$ be the set of all linear combination of $\mathcal{S}$, denoted by $Span(\mathcal{S})$

Then in the beginning example, we take $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, then $Span(\{\mathbf{v}_1, \mathbf{v}_2\}) = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} : x+y+z=0\} \subset \mathbf{R}^3\}$
Directly following from definition, a spanning set MUST be a subspace of the original vector space.

1.4 Linear independence

We still use the above example, but this time we use two different vectors. $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix}$, and we now get a different spanning set:$\{\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}, x+y=0\}$. Now we wonder, what have led to the different?
In second example, $\mathbf{v}_2 = 2\mathbf{v}_1$, then a linear combination of the two vector is $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = (c_1 + 2c_2)v_1$, only one vecor left! In this case, we say $\mathbf{v}_1, \mathbf{v}_2$ are linearly dependent. Formally,

Definición 1.6. A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is linearly dependent if for some $c_i \neq 0$, $\sum_{k=1}^{n} c_k \mathbf{v}_k = 0$

In this case, we can divide c_i at each of the vectors and represent $\mathbf{v}_i$ by other vectors.
For the contrary case,

Definición 1.7. A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is linearly independent if $\sum_{k=1}^{n} c_k \mathbf{v}_k = 0$ implies $c_i = 0, \forall i$

1.5 Basis

We have almost get our wanted results! Remember:

> Our goal is now try to find a way to use less vectors to represent the transformation, in the best case, using finitely many vectors.

Consider this, if all vectors in V is a linear combination of a set of vectors in V, then the given result of linear transformation of the original vector is the linear combination of mappings of vectors in V. Formally,

$$T(\mathbf{v}) = \sum_{k=1}^{n} c_k T(\mathbf{v}_k)$$

. All we need to do now is to find that set of vectors. We can use the following algorithm.

```
set = EmptySet;
while(We can find a vector not in Span(set))
    add the vector to set;
return set;
```

Nota. For some vector spaces, we can never find a finite set of vector that span the whole space, but we may first not discuss about that.

For the vector space that can be spanned by finitely many vectors (in short finitely dimensional space), the set of vectors that spans the whole set must exists. In this case, we want a more ideal set. We hence introduce the concept of basis.

Definición 1.8. A set of vectors $\mathcal{B} \subset V$ is a basis of V if it satisfies

1. $\mathcal{B}$ is a linearly independent set.
2. $\text{Span}(\mathcal{B}) = V$.

We can understand the role of a basis as a smallest spanning set, and also the largest linear independent set. This will give us a quite beautiful result. We call it Unique Representation theorem.

Teorema 1.9. If $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a basis of V, then for all $\mathbf{v} \in V$, there exists unique scalars c_i such that

$$\mathbf{v} = \sum_{k=1}^{n} c_k v_k$$

Proof.

1. Existence is trivial by Definition 1.5 and Definition 1.8.
2. Uniqueness: We assume we have two different way of decomposition, subtracting on both sides, we have

   $$(c_1 - c_1')\mathbf{v}_1 + (c_2 - c_2')\mathbf{v}_2 + \cdots + (c_n - c_n')\mathbf{v}_n$$

   Since $\mathcal{B}$ is linear independent, all factors are 0.

□

If you are someone who care very much about mathematical rigorousness, we have two jobs left, One is to show the existence of such basis. One is to show the number of vectors in a basis is the same. I will assume we have proven this part and if you are really interested, you can look up Spanning Set theorem and Steinitz Exchange lemma, proofs can be found online. Combining the information that all the basis has the same number of vectors. It becomes natural for us to give a name for the "number".

Definición 1.10. The dimension of a vector space is the number of vectors in a basis of the vector space.

We have the criterion for basis:

Teorema 1.11. Let V be a n-dimensional vector space, let $\mathcal{S} \subset V$ be a set with n numbers.Then,

1. If $\mathcal{S}$ is linearly independent, $\mathcal{S}$ is a basis.
2. If $\text{Span}(\mathcal{S}) = V$, $\mathcal{S}$ is a basis.

1.6 Summary

Following is a summary of what we have done in this chapter:

1. We guarantee we don't have to check whether the $T(\mathbf{v}_1 + c\mathbf{v}_2)$ is well-defined because all vector space is closed in adding and scaling.
2. We show all finitely dimensional vector space can be decomposed into an unique sum of basis vectors.

Yeah, That mainly what we did. Next chapter is of more fun. We will finally introduce the concept of matrix!

linear transformation

Here's a short introduction about what this chapter will be about. We have many kinds of vector spaces, and the computation is quite complex. Therefore, we want to come up with a more general way so that we can compute different linear transformations in the same form. This universal language of linear transformation is called matrix. To develop this chapter, we follow the following logic:

1. Show that a matrix stands for a linear transformation in Euclidean space.

2. Converting any linear transformation into a matrix. We need to study some of the properties of linear transformations before that.

2.1 Matrix

We now introduce the concept of matrix. In different context, matrix represents different kinds of things. In computer science, people sometimes regard 2-d array as matrix. But in the context of linear algebra, a matrix is usually a linear transformation on Euclidean space.

We consider the space $\mathbf{R}^n$, it's a n-dimensional vector space, the most natural way of finding a basis is

$$\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\} := \{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \}$$

Therefore, when we want to describe a linear transformation, we only need to describe how each of the elements was mapped to.For a linear transformation T, we put all resulting vectors in a list of columns. For example:

$$T(\mathbf{e}_1) = \mathbf{a}_1, T(\mathbf{e}_2) = \mathbf{a}_2, \cdots, T(\mathbf{e}_2) = \mathbf{a}_n$$

The matrix of linear transformation is

$$\begin{pmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{pmatrix}$$

. Since

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \cdots + c_n\mathbf{e}_n$$

We have

$$T(\mathbf{v}) = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n$$

We usually denote the matrix with each columns as the "sent" vectors from the $\mathbf{e}_1 \cdots$ as the matrix of linear transformation of Euclidean space. For example, if

$$T(\mathbf{e}_1) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, T(\mathbf{e}_2) = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

we can say

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$$

is the matrix of the linear transformation. Also:

$$T(\begin{pmatrix} 1 \\ 2 \end{pmatrix}) = A \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 2 \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

2.2 Matrix multiplication

Intuitively, a combination of linear transformation is still a linear transformation. That means we can still represent combination of linear transformation with another matrix. Now lets figure out the relation between them. Then let's consider:

$$A = \begin{pmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{pmatrix} \quad B = \begin{pmatrix} | & & | \\ \mathbf{b}_1 & \cdots & \mathbf{b}_n \\ | & & | \end{pmatrix}$$

The i-th column of the combination of the two LT (Later I will use LT as linear transformation) is given by $AB\mathbf{e}_i = A\mathbf{b}_i$. Therefore, we express

$$AB = \begin{pmatrix} | & & | \\ A\mathbf{b}_1 & \cdots & A\mathbf{b}_n \\ | & & | \end{pmatrix}$$

That makes a lot of sense! And if we consider an vector as a LT from a real number to vector, the computation of matrix is compatible with matrix multiplication! (Which is good)

2.2.1. Nonlinearity of neural network

There's some fun fact about deep learning. In deep learning, there's a very important structure called MLP (multi-layers preceptor). If we consider the easiest case in deep learning, maybe recognizing numbers? What we actually do is to map an image into it's label, which can be both considered to be Euclidean space.

To further elaborate this point, consider an image with size 20×20. We can in fact consider it as an vector in $\mathbf{R}^{400}$, since it has 400 blocks. And we want to find a mapping from $\mathbf{R}^{400}$ to $\mathbf{R}^{10}$, which is the range of all the numbers (In machine learning, we call these labels). The idealistic situation is, we input an image(vector) and output a vector say $\mathbf{e}_3$, which correspond to 3.

The most natural way people would come up is to use LT to simulate this mapping, because computer would really love to compute LT (easier). They later want to increase the number

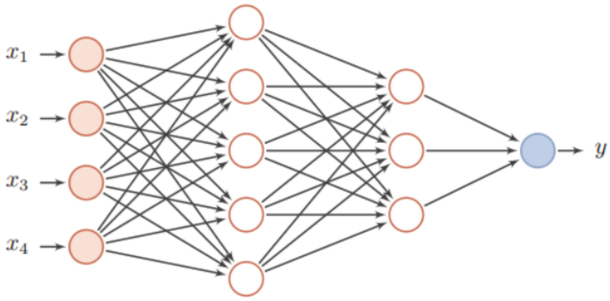

Figura 2.1: This is an example

of parameters (the number of matrices), so that the performance of the network might be better. However, if we just arbitrarily put many matrices together, nothing will work. Since the combination of LT is still LT, which can be represented by a single matrix.
To deal with that, people usually add an element-wise non-linear function. Therefore, the combination of all matrices will not become a simple matrix!
I think I haven't explain this part clear enough, if you are interested, maybe you can check out the video of 3Blue1Brown. He has a wonderful video talking about this concept.

2.3 Injective, Surjective and Bijective

In this part, we will deal with some properties of LT, which is essential when we want to convert an arbitrary LT into a matrix.

Definición 2.1. Consider a LT T from V to W.
A LT is injective (one to one) if $T(\mathbf{v}) = T(\mathbf{w})$ will imply $\mathbf{v} = \mathbf{w}$
A LT is surjective (onto) if $\forall \mathbf{w} \in W\ \exists \mathbf{v} \in V$ with $T(\mathbf{v}) = \mathbf{w}$
A LT is bijective if it's both injective and surjective.

Note that sometimes we call a bijective function an isomophism. These definitions might be a little weird to you, let's explain their meanings. These properties usually serve for the existence of inverse, another function T^{-1} with $TT^{-1} = \mathrm{I}$, where I is the LT map every element to itself. We will get back to it later.

If you have some background information in calculus, you would know that e^x is the inverse of $\log(x)$. we have $e^{\log(x)} = \log(e^x) = x$. In the context LT, we also want something similar. But the first important thing is that we have to make sure the co-domain and domain have one to one relation. If $f(1) = 2$ $f(3) = 2$, what is $f^{-1}(2)$ supposed to be? We cannot expect a funcion will output two values at the same time. Therefore, injective is the first necessary condition of existence of inverse.

Now if a function is injective, can we conclude that it's invertible? No! We also need to pay attention to the value that nothing would map to. They are like the single guys from a party. If you want to find a invertible matrix, you have to find each of them a partner, that is exactly the expression of surjective!

Hence we can sum up.

Teorema 2.2. If LT is bijective, then dim $V = \dim W$. Also, $\exists T^{-1}$, with $TT^{-1} = \mathrm{I}$, $T^{-1}T = \mathrm{I}$.

Note that even TT^{-1} and $T^{-1}T$ will both give an I, but they are actually different. Because they are defined in different spaces. One map W to W, and other map V to V. The proof is omitted there because I'm lazy to type. But I think the above explaination will be a good source for understand.

Nota. We also have two equivalent way to express injective and surjective.
If T is injective, $\mathrm{Ker}T := \{\mathbf{v}\colon T(\mathbf{v}) = \mathbf{0}\} = \{\mathbf{0}\}\}$.(:= means defined as)
If T is surjective, $T(V) := \{\mathbf{w} : \exists\mathbf{v} \text{ with } T(\mathbf{v}) = w\} = W$
If T is injective, we can show that $\dim T(V) = \dim W$, combining this and T is surjective, we can show the dimension of two vector space are the same.

Sometimes if we want to show that two set are equal, we can show that $A \subset B$ and $B \subset A$. Try this to show If T is surjective, $T(V) := \{\mathbf{w} : \exists\mathbf{v} \text{ with } T(\mathbf{v}) = w\} = W$.
We also should mention the inverse of combiniation of invertible LT.

Teorema 2.3. If LT S and T are both invertible, then their combination is also invertible, with $(ST)^{-1} = T^{-1}S^{-1}$

You can simply multiply the inverse with the original LT to show it's indeed the inverse.

2.4 Use Matrix to represent a linear transformation

It's such a long section name, but the following indeed will give us a very beautiful result, or even the most important result of linear algebra. That is, every LT can be represented using a matrix.

We know that every finitely dimensional LT will map from a vector space to another. A vector space has an integer dimension. Then we have the following map.

Teorema 2.4. If we have a vector space V whose dimension is n and a basis $\mathcal{B}\,\{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_n\}$, then we have the following mapping from V to $\mathbf{R}^n$:

$$\psi_{\mathcal{B}}(\mathbf{v}) = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

This is a direct consequence of Theorem 1.9, which is decomposition theorem. By checking the property, we know that this function is bijective. (Leave as an exercise for readers) Therefore, without doubt, it's an invertible function!

Then let's consider this, if we first map $\mathbf{v}$ to another vector in $\mathbf{R}^n$, passing through a matrix, and map it from $\mathbf{R}^m$ back to V (dimension might have changed here), we are then able to express the general LT with matrix. In other word

$$T = \psi_{\mathcal{B}_2}^{-1} A \psi_{\mathcal{B}_1}$$

Very Importantly! We usually consider combination of functions from right to left (even though we might read from right to left).

Then the matrix is given by

$$[T]_{\mathcal{B}_2\mathcal{B}_1} = A = \psi_{\mathcal{B}_2} T \psi_{\mathcal{B}_1}^{-1}$$

$[T]_{\mathcal{B}_2\mathcal{B}_1}$ is considered to be: the matrix representing linear transformation T from the vector space V represented by basis $\mathcal{B}_1$, to vector space W represented by basis $\mathcal{B}_2$, which itself would be a matrix mapping from $\mathbf{R}^n$ to $\mathbf{R}^m$. Let's see an example

2.4.1. Vandermonde Matrix

There's an very interesting property of polynomials, that any polynomial of degree n can be represented by $n+1$ points. On the other hand, if we know the $n+1$ points and the coooresponding coordinate, we can determine the polynomial.

It becomes trivial to consider the following map called evaluation map.

> **Definición 2.5.** Given n+1 distinct points $t_0, t_1, \cdots, t_n$ and a polynomial $p(x)$, we have the following map:
>
> $$T(p(t)) = \begin{pmatrix} p(t_0) \\ p(t_1) \\ \vdots \\ p(t_n) \end{pmatrix}$$

We know that a polynomial can be represented by a list of coefficients. For example, $3x^2+2x+1$ can be represented by $(1, 2, 3)$. Then we have the following coordinate mapping :

$$\psi_{\{1,t,t^2,\cdots,t^n\}}(x_0 + x_1 t + \cdots + x_n t^n) = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Then we want to express the evaluation map with a matrix. We follow the following logic:

$$[T]_{\mathcal{B}_2\mathcal{B}_1}(\mathbf{e}_i) = A = \psi_{\mathcal{B}_2} T \psi_{\mathcal{B}_1}^{-1}(\mathbf{e}_i) = \psi_{\mathcal{B}_2} T(t^i) = \psi_{\mathcal{B}_2} \begin{pmatrix} t_0^i \\ t_1^i \\ \vdots \\ t_n^i \end{pmatrix} = \begin{pmatrix} t_0^i \\ t_1^i \\ \vdots \\ t_n^i \end{pmatrix}$$

Note that the T will map from $P_n[t]$ to $\mathbf{R}^{n+1}$. Here, $\mathcal{B}_1$ is the standard basis of polynomial, and $\mathcal{B}_2$ is just standard basis of Euclidean space. The last step is basicly an identity map. This will indeed give us the Vandermonde matrix.

$$\begin{pmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^n \\ 1 & t_1 & t_1^2 & \cdots & t_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^n \end{pmatrix}$$

Please! Feel free to test whether this will give you the desired results. Also, the matrix is invertible, if we want to solve the inverse matrix, we should consider Lagrange interpolate. Relevant information can be found in wikipedia.

2.5 Change of basis and Similar matrix

In last chapter, when we use a matrix to represent a LT, we have an important move, we select a basis from the original V and build a bijective coordinate map. It becomes natural for us to consider, if we change to another basis, what will actually happen?

This discussion lead to the research on "change of basis". For simplicity, we usually discuss the LT whose V and W are of the same dimension. Then we have

$$[T]_{\mathcal{B}_1\mathcal{B}_1} = \psi_{\mathcal{B}_1} T \psi_{\mathcal{B}_1}^{-1}$$

$$[T]_{\mathcal{B}_2\mathcal{B}_2} = \psi_{\mathcal{B}_2} T \psi_{\mathcal{B}_2}^{-1}$$

Therefore

$$[T]_{\mathcal{B}_2\mathcal{B}_2} = \psi_{\mathcal{B}_2}\psi_{\mathcal{B}_1}^{-1}[T]_{\mathcal{B}_1\mathcal{B}_1}\psi_{\mathcal{B}_1}\psi_{\mathcal{B}_2}^{-1}$$

We can then try to understand the meaning of $\psi_{\mathcal{B}_1}\psi_{\mathcal{B}_2}^{-1}$, this function will receive a vector from $\mathbf{R}^n$ and convert it into V, then coordinate map it to $\mathbf{R}^n$ again. Obviously, it's a LT on Euclidean space. Then it can be represented as a matrix, we denote it as $P_{\mathcal{B}_1\mathcal{B}_2}$.

The i-th column of $P_{\mathcal{B}_1\mathcal{B}_2}$ will be the obtained vector from $\mathbf{e}_i$, then $\psi_{\mathcal{B}_1}\psi_{\mathcal{B}_2}^{-1}(\mathbf{e}_i) = \psi_{\mathcal{B}_1}(\mathbf{b_{2i}}) := [\mathbf{b_{2i}}]_{\mathcal{B}_1}$ Therefore

$$P_{\mathcal{B}_1\mathcal{B}_2} = \begin{pmatrix} | & & | \\ [\mathbf{b_{21}}]_{\mathcal{B}_1} & \cdots & [\mathbf{b_{2n}}]_{\mathcal{B}_1} \\ | & & | \end{pmatrix}$$

The notation here might be a little confusing and different textbook might use different notation. In this case, $\mathbf{b_{2i}}$ would mean the i-th vector in the basis $\mathcal{B}_2$. And $[\mathbf{b_{2i}}]_{\mathcal{B}_1}$ would mean the coordinate vector of $\mathbf{b_{2i}}$ in terms of basis $\mathcal{B}_1$.

We are also aware of the fact that $\psi_{\mathcal{B}_2}\psi_{\mathcal{B}_1}^{-1} = (\psi_{\mathcal{B}_1}\psi_{\mathcal{B}_2}^{-1})^{-1}$. To sum up, we have

$$[T]_{\mathcal{B}_2\mathcal{B}_2} = P_{\mathcal{B}_1\mathcal{B}_2}^{-1}[T]_{\mathcal{B}_1\mathcal{B}_1}P_{\mathcal{B}_1\mathcal{B}_2} = P_{\mathcal{B}_2\mathcal{B}_1}[T]_{\mathcal{B}_1\mathcal{B}_1}P_{\mathcal{B}_2\mathcal{B}_1}^{-1}$$

Basicly, we obtain the matrix in terms of $\mathcal{B}_2$, we first change vector to basis of $\mathcal{B}_1$, operate the obtained vector and convert it into $\mathcal{B}_2$.

Sometimes, it's a little troublesome to find $P_{\mathcal{B}_1\mathcal{B}_2}$. A practical experience is to use standard basis of $\mathbf{R}^n$, because $P_{\varepsilon\mathcal{B}} = \begin{pmatrix} | & & | \\ \mathbf{b}_1 & \cdots & \mathbf{b}_n \\ | & & | \end{pmatrix}$, which is much easier to compute. Therefore

$$P_{\mathcal{B}_1\mathcal{B}_2} = P_{\varepsilon\mathcal{B}_1}^{-1} P_{\varepsilon\mathcal{B}_2} = \begin{pmatrix} | & & | \\ \mathbf{b}_{11} & \cdots & \mathbf{b}_{1n} \\ | & & | \end{pmatrix}^{-1} \begin{pmatrix} | & & | \\ \mathbf{b}_{21} & \cdots & \mathbf{b}_{2n} \\ | & & | \end{pmatrix}$$

Example. Consider the two basis:

$$\mathcal{B}_1 = \{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\}, \mathcal{B}_2 = \{\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\}$$

Then

$$P_{\varepsilon\mathcal{B}_1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, P_{\varepsilon\mathcal{B}_2} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$P_{\mathcal{B}_1\mathcal{B}_2} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

So far I haven't told you about how do we compute the inverse of a matrix. I will leave this kind of work to next chapter.
One thing that I should mention is the concept of similar matrix.

Definición 2.6. Considering two matrix of the same size, if there exists an invertible matrix P, such that

$$B = PAP^{-1}$$

Then we say A and B are similar.

In this case, we can consider the two matrix are representing the same LT, that's why we call it similar. Since they describe the same LT, any quantities defined on the LT are shared by the matrices.

Moreover, sometimes if we choose a "nice" basis, the matrix will look very nice, this is one of the key propose we will be working on.

2.6 Summary

Let's do some summary. In this chapter, we first studied a special type of LT, which is LT in Euclidean space, and we represent that by a beautiful structure, matrix. Then by introducing some properties of LT (will be discussed later), we connect an arbitrary LT with matrix, then discussed different matrix to represent a LT and their relations.

In next chapter, we will solve some problems that we have left in this chapter. Like how do we know the inverse of the matrix? Good luck!