MAT240: Algebra I

Review Notes

Author: Joseph Siu

Email: joseph.siu@mail.utoronto.ca

Date: April 20, 2024

Contents

1   Fields

Commutativity, Associativity, Distributivity, Identity, Inverse.

2   Modular arithmetic

$[a] + [b] = [a + b]$, $[a] \cdot [b] = [a \cdot b]$.

Claim 1

All elements in $\mathbb{Z}_n$ are invetible if and only if n is a prime.

Proof. Assume $n = ab$, then assume for contradiction $ac = 1 \pmod{n}$ for some c, then $ac - ny = 1$ for some y, and $ac - aby = 1$ so $a(c - by) = 1$ so a divides 1, contradiction.

QUOD
ERAT
DEM■

Claim 2

a is invertible in $\mathbb{Z}_n$ if and only if $\gcd(a, n) = 1$.

Proof. Assume a invertible where $ac = 1 \pmod{n}$, and assume $\gcd(a, n) = b > 1$, then $ac - nx = 1$ for some x, however right side not divisible by b so contradiction.
Assume $\gcd(a, n) = 1$, then $ax + ny = 1$ for some x, y, so $ax = 1 \pmod{n}$.

QUOD
ERAT
DEM■

3   Complex number

Modulus is just the length of the line, which by pathagorean theorem is $\sqrt{x^2 + y^2}$.

To find multiplicative inverse, we first multiply the conjugate, then divide by modulus squared so that it is projected to the real line then scaled to 1. That is, $\left(\frac{1}{x^2 + y^2}\right)(x - yi)(x + yi) = 1$.

Polar form is a scaled radius of the unit circle, that is, first rotate the line from 0 to 1 by θ, then scale by the modulus. Here $\tan\theta = \frac{y}{x}$. To calculate we can also consider sin and cos separately, that is, modulus $x = r\cos\theta$ and $y = r\sin\theta$.

4   Polynomial

Definition 1 – Polynomial Division

$f(x) = q(x)g(x) + r(x)$, where $\begin{cases} \deg r < \deg g & \text{if } g \neq 0 \\ r = 0 & \text{if } g = 0 \end{cases}$.

g divides f if $r \equiv 0$.

To find the irreducible polynomials, we can first eliminate the combination of coefficients that guarantee an integer root.

> **Theorem 1 – FTA**
>
> Every non-constant polynomial has a root over $\mathbb{C}$ (every non-constant polynomial can be factored into products of linear factors over $\mathbb{C}$).

5 Matrix

(This section assumed that the reader is already familiar with most of the definitions of MAT240 regarding matrices. This section will reinforce the connections between the ideas of matrices and linear transformations.)

5.1 How matrices are linear transformations

Since every matrix represents a linear transformation and vice versa, to understand matrix, it is equivalent to understand the linear transformation that corresponds to the matrix.

First consider a linear transformation and its matrix. The set that contains the columns of the matrix spans the image of the linear transformation. That is, consider a vector in the domain, then the output of the linear transformation given this vector is the summation of the ith entry of the vector times the ith column of the matrix. What this means is that the matrix stores some vectors (columns) that can be scaled and added (using the given vector) to form any vector in the image.

5.2 Matrix multiplications in general

What we have discussed was the case that a matrix is multipled with a vector at the right, if a matrix is multiplied with another matrix at the right instead, we can treat the right matrix as a set of vectors (columns), then we perform the same operation as before, then put all the results together and form another matrix (set of column vectors). Furthermore, If two matrices are multiplied together, this can be treated as a composition of linear transformations from right to left.

5.3 Check linearly independency of columns

Now, we can see the columns of the matrix span the image of the linear transformation, but how do we know if the columns form a basis[1] of the image (column space)[2]? We need to check if they are linearly independent or not, the easiest way would be change the matrix to RREF, and see if one column can be expressed as a linear combination of the others or not (if it is full rank then the columns are dependent thus form a basis). Here note that when performing row reductions, the column space (image) has changed, however since we only care whether they are independent or not, so row operations won't affect the result.

5.4 From the row perspective

To now, we have discussed matrices from a column perspective as combinations / sets of column vectors. However, when solving solution like $Ax = B$, we can think this as a system of linear equations, where row operations won't affect the solution since all operations are reversible, for example:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \end{array}\right): \quad \left(\begin{array}{ccc|c} 1 & 0 & 0 & a \\ 1 & 1 & 0 & b \to b+a \\ 0 & 0 & 1 & c \end{array}\right) \quad \left(\begin{array}{ccc|c} 1 & 0 & 0 & a \\ 0 & 0 & 1 & b \to c \\ 0 & 1 & 0 & c \to b \end{array}\right) \quad \left(\begin{array}{ccc|c} 1 & 0 & 0 & a \\ 1 & \frac{1}{2} & 0 & b \to \frac{b}{2} \\ 0 & 0 & 1 & c \end{array}\right)$$

are the transformations that add the first row to the second row; swap the last 2 rows; and multiply the second row by $\frac{1}{2}$ respectively. We can see that the vectors that satisfy the later equations must also satisfy

---

[1] An ordered linearly independent set that spans the space.

[2] Space, or vector space, is a set of vectors that are closed under addition and scalar multiplication.

the original equations since we can reverse the operations without changing the equalities:

$$
\left(\begin{array}{ccc|c}
1 & 0 & 0 & a \\
-1 & 1 & 0 & b \to b + a \to b + a - a = b \\
0 & 0 & 1 & c
\end{array}\right)
\quad
\left(\begin{array}{ccc|c}
1 & 0 & 0 & a \\
0 & 0 & 1 & b \to c \to b \\
0 & 1 & 0 & c \to b \to c
\end{array}\right)
\quad
\left(\begin{array}{ccc|c}
1 & 0 & 0 & a \\
1 & 2 & 0 & b \to \frac{b}{2} \to 2\frac{b}{2} = b \\
0 & 0 & 1 & c
\end{array}\right)
$$

From here, we can also see why when finding the null space[3] of the image (column space), our row operations won't affect the result, and performing row operations is very helpful not only because we can decrease the numbers of values to look at, but also it is easier to see *how* the columns are depending on others if there are any.

5.5 Purpose of row reductions

Relating to this discussion, we can see the purpose of row reductions is to transform the given matrix to a diagonalized (or mostly diagonalized) form, so that *some* solutions of the diagonalized matrix remain the same as the original matrix, for example the null space is preserved.

5.6 Other properties of similar matrices

However just knowing the points that do not change after "diagonalization" [4] is probably not that interesting enough, in fact, similar matrices also preserve another important property: the eigenvalues.[5] Geometrically the eigenspace contains vectors that are only being streatched by some factor λ after the linear tranasformation. Before we begin, the idea of determinant needs to be introduced.

5.7 Determinants

As commonly known, the determinant of a transformation is simply the "area / volume / hyper-volume" determined by the column vectors (the vectors that span the image). We first defind the identity matrix (that is, transformation that its image is spanned by the standard basis) to have determinant 1. Then we define swapping rows or columns as negating the original determinant. What this does intuitively is that the order of "axis / ~~dimension~~" we used to calculate the "hyper-volume" of the transformation got changed, which results in flipping the entire "hyper-volume" from an external perspective. Lastly, as we can see intuitively, scaling or adding determinants on one "axis / ~~dimension~~" will scale or add the entire "hyper-volume" by the same factor or value.

To summerize, we define determinant using the following three definitions:

1. $\det I = 1$.

2. $\det A' = -\det A$ if A' is obtained by swapping 2 rows or 2 columns of A.

3. $\begin{vmatrix} ta + x & tb + y \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} x & y \\ c & d \end{vmatrix}$.

Moreover, by Leibniz formula, we can calculate the determinant for any square $n \times n$ matrix A as:

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det\left(\tilde{A}_{ij}\right) \quad \forall i \in \{1, \ldots, n\},$$

where $\tilde{A}_{ij}$ is defined by removing the ith row and j^{th} column of A.

---

[3]Null space contains vectors that are being sent to 0 by the linear transformation / matrix.

[4]The process of changing a matrix into a diagonalized form, but as we all know what some matrices (not full rank) cannot be fully diagonalized, here comes the Jordan Normal Form which will be discussed in MAT247.

[5]Note that the converse is not true. Similar matrices have same determinant, same rank, same eigenvalues, same trace, same characteristic polynomial... However matrices with same eigenvalues / determinants does not imply they are similar.

Using these definitions, we can come up with useful properties such as $\det AB = \det A \det B$ and $\det A^T = \det A$, that is, linear transformation through column vectors has the same determinant as linear transformation through row vectors. Last but not least, we can see that similar matrices have the same determinant as their diagonal matrix

$$\det A = \det Q^{-1} \det A \det Q = \det Q^{-1} AQ = \det B,$$

and

$$\det A = \det P^{-1} \det D \det P = \det P^{-1} DP = \det D,$$

for some similar matrices $B = Q^{-1}AQ$ and diagonal matrix D where $A = P^{-1}DP$. Notice here that the determiant of all corresponding diagonal matrices are the same which the reason will be explained in the next paragraph.

5.8   Eigenvectors, eigenvalues, and eigenspaces

Finally, we can discuss what eigenvalues, eigenvectors, and eigenspaces are. To start, eigenvectors are vectors that are only stretched after applying matrix A (or linear transformation T_A), namely $Av = \lambda v$ for some vector v and some constant λ. By distributivity we may rearrange this equation to $(A - \lambda I)v = 0$, here λ is a constant to be determined. Since v is non-zero, and is being transformed to the 0-vector through matrix $A - \lambda I$ (and thus it's linear transformation $T_{A-\lambda I}$), so v exists if and only if $T_{A-\lambda I}$ is *not* full rank, or in other words, its columns are linearly dependent. Which we also know that if there exists a column that is dependent from the others, then we may perform column operations to make the row to be 0, which by Leibniz formula gives the determinant is equal to 0. Thus, we may conclude that as long as $\det(A - \lambda I) = 0$, there exists a non-zero vector v that satisfies $Av = \lambda v$, and we will also call such vector an eigenvector with eigenvalue λ.

Since the entries of A are constants, and λI is a diagonal matrix where all diagonal entries are λ, to find all possible λ, we will treat it as a variable, and apply the Leibniz formula to get a polynomial from $\det(A - \lambda I) = 0$, which is called the *characteristic* polynomial. Then we can see that the roots of such polynomial gives all possible λ that satisfies the equation $Av = \lambda v$ for some vector v in the domain.

So, the eigenvalues of a linear transformation / matrix are the roots of the characteristic polynomial determined using $\det(A - \lambda I) = 0$, the eigenvectors are the vectors that satisfy $Av = \lambda v$ for some eigenvalue λ, and the eigenspace is the space that contains all eigenvectors that correspond to the *same* eigenvalue.

> **Claim 3**
>
> Similar matrices have the same characteristic polynomial, thus same eigenvalues.   ⚥

$$\textit{Proof. } \det\left(P^{-1}AP - \lambda I\right) = \det\left(P^{-1}AP - \lambda P^{-1}P\right) = \det P^{-1} \det\left(A - \lambda I\right) \det P = \det\left(A - \lambda I\right).$$

<div align="right">Q<small>UOD</small> E<small>RAT</small> D<small>EM</small>■</div>

6   Vector spaces

$0 \in V, ax + y \in V$.

The dimension of a vector space is the size (cardinality) of the basis of the vector space. By replacement theorem, all bases of a vector space have the same size.

7   Subspaces

$\{0\}$ and the space itself are trivial subspaces, others are called non-trivial subspaces. Subspace is just a subset of another vector space that is also a vector space.

Common subspaces include the null space (kernel) and the column space (image) of a matrix[6].

If $V = W_1 + W_2$ and $W_1 \cap W_2 = \{0\}$, then $V = W_1 \oplus W_2$. That is, if V is the sum of two "disjoint" subspaces (only intersect at 0), then V is called the direct sum of W_1 and W_2.

8   System of equations

To solve $Ax = B$, we augment the matrix A to $A \mid B$, then perform row operations to make the left side to be in its RREF form, then set the free (non-leading) variables and express the pivot (leading) variables in terms of the free variables, then we can get a basis of the solutions. In particular, to find $N(A)$, we solve for $Ax = 0$, that is, row reduce A to RREF, the make free variables to be parameters and express pivot variables in terms of the variables, finally isolate the parameters then we can claim that the (null space) solutions are spanned by linear combinations of the vectors we obtained.

We call the solutions that are not the 0-vector the non-trivial solutions, and 0 the trivial solution. And call the system of equations consistent if there exists a solution, and inconsistent otherwise.

We will also call the system of equations homogeneous if the constants are all 0's. We are particular interested in homogeneous equatioins since they must either have only the trivial solution or also have non-trivial solutions, and the set of all solutions form a subspace of the domain. That is, they are always consistent.

9   Independence and spanning sets

We used the word linearly independent throughout this notes, but we haven't formally defined what is linearly independent. We say vectors $v_1, \ldots, v_n$ are linearly independent if the only solution to $c_1 v_1 + \cdots + c_n v_n = 0$ is $c_1 = \cdots = c_n = 0$. Otherwise, we say they are linearly dependent. In other words, if one vector can be expressed by the other ones, then they are linearly dependent.

10   Basis

The span of a set of vectors is the set of all linear combinations of the vectors. We will call this set a basis of the span if its vectors are linearly independent.

That is, a basis of a vector space is a set of vectors that are linearly independent and span the vector space.

As we can see a vector space may have different bases, but all the bases are linearly independent sets and have the same size. So intuitively, we may swap the elements in different bases and the resulting bases are still linearly independent and with the same size, and also still spanning the same space.

Another way of thinking this is that for some linearly independent set of a vector space, we are still able to add more independent vectors as long as the size of the set is less than the dimension / the size of the basis, and once the size reaches the dimension, we can claim that this set becomes a basis, and thus we cannot find more independent vectors to add to the set.

11   Dimension

The dimension of a vector space is the size of its basis.

$\dim T_A = \operatorname{rank} A$, where A is the corresponding matrix of linear transformation T_A.

For finite dimensional vector space V, if W is a subspace, then $\dim W \leq \dim V$.

---

[6]By dimension theorem we have that $\dim V = \dim \operatorname{im} T + \dim N(T)$ for all linear transformation $T : V \to W$.

12   Linear transformations

$T(ax + y) = aT(x) + T(y)$, $T(0) = 0$.

As explained earlier, every linear transformation is associated with a matrix.

Null space contains vectors in domain that are sent to 0 by the linear transformation, the image is just the column space, that is, the space spanned by the column vectors.

Injection, surjective, and isomorphism are equivalent since they all describe the null space to contain only the 0-vector, thus the image is also the entire space. An important idea here is that if a particular vector satisfies a solution, so is and only such particular vector added with any veector in the null space would also give the solution. In other words, the difference between any two solutions is in the null space, namely $D(s - p) = D(s) - D(p) = 0$ for any two solutions s and p, so $s - p \in N(T)$ and is also a solution.

Consider the dimension theorem $\dim V = \dim \text{im} T + \dim N(T)$, this basically tells us the relating between the null space and the image when the size of the basis is fixed. That is, first record the basis vectors in the domain, then if we know what basis remain in the image, we can also know what basis remain in the null space. And here as we have noticed that the solutions are spanned by a particular solution and a null space basis, so the uniqueness of the solutions can be guaranteed if the null space is $\{0\}$, this shows that every vector in the image can still be expressed as a combination of the basis vectors in the domain, that is, the basis after applied linear transformation still spans the image since some basis vectors span the null space and the remaining span the image.

13   Isomorphism and invertibility

Isomorphism is just a fancy term for bijective linear transformations. If there exists a bijection between two vector spaces, then the linearality of the transformation guarantees that the transformed basis is still a basis (of the image), and so the null space must be only $\{0\}$.

14   Change of coordinates from standard bases' perspectives

Throughout the discussion we have been treating the input as a vector. A vector is just a list of numbers, but how do we know about the vector with respect to the whole system, we might implicitly assume a standard basis and "plot" the vector onto the grid created by the standard basis. But as we can see the bases are often not standard, and matrices are not always the identity matrix.

As the amazing 3blue1brown video has mentioned, since the columns of a matrix span the image, input a vector, then the vector consists of "scalars" of the columns, and the output is the linear combination of the columns, that is, another vector. The entries of the output vector have no meaning if we do not assign a coordinate system, or a basis to it. They are only meaningful with respect to some basis, for example, the standard basis.

So how do we know the different vectors in different bases are essentially the same with respect to their coordinate system[7]? The most straightforward way would be just "undo" the transformations by applying the matrices that have columns as their corresponding bases, then the resulting vector can be directly viewed with respect to the standard basis:

$$\underbrace{v}_{\substack{\text{a linear combination of} \\ \text{the standard basis}}} = \underbrace{A}_{\substack{\text{columns are vectors} \\ \text{in another basis}}} \underbrace{v'}_{\substack{\text{a linear combination} \\ \text{of the other basis}}},$$

---

[7]We use the notation $[v]_\beta$ to denote v as a linear combination of the vectors in basis β. Coordinate systems can be treated as bases, when a vector is in a coordinate system we say it is a linear combination of the coresponding basis, that is, from that basis' perspective.

to change the standard basis linear combination to the other basis', we can just multiply the inverse[8]:

$$\underbrace{A^{-1}}_{\substack{\text{inverse of a matrix where} \\ \text{the columns are vectors} \\ \text{in another basis}}} \underbrace{v}_{\substack{\text{a linear combination of} \\ \text{the standard basis}}} = \underbrace{v'}_{\substack{\text{a linear combination} \\ \text{of the other basis}}} .$$

14.1   Combine linear transformations and change of bases

What we have discussed was the scanario when we want to simply convert a vector from a basis to the standard basis, or from the standard basis to another basis (of the same vector space), however what if we want to involve a linear transformation from a vector space to itself with respect to the standard basis? For example, a rotation where the matrix is given in the standard basis, and we want to perform the same rotation with respect to other bases.

First consider $v = Av'$ (where A is the change of basis matrix from given basis to the standard basis), then both sides are represented in the standard basis, now apply a linear transformation M (where T is a linear transformation from a vector space to itself with respect to the standard basis, and M is the corresponding matrix of T) and we have $Mv = MAv'$, both sides are still represented as a linear combination of the standard basis. Finally apply A^{-1} to "undoes" the change of basis and we get $A^{-1}Mv = A^{-1}MAv'$. What we have been doing at the right hand side is that, we first consider a vector that is a linear combination of the second basis, transform it into the standard basis, then apply the linear transformation for the standard basis vectors, and finally convert the result back to the second basis.

Does $A^{-1}MA$ look familiar? Consider $M' = A^{-1}MA$, then we can see that M' and M are similar, that is, the same transformation under different bases are similar. If we let M to be the identity transformation for standard basis, then $A^{-1}A = I$, the identity transformation with respect to any basis is still the identity matrix.

So far, we have discussed the change of basis and linear transformations within a vector space V, we use $[T]_\beta^{\beta'}$ to denote a transformation that takes in a vector in coordinate system β, and ouputs a vector in coordinate system β'. To extend the scanario to any general case, let's consider linear transformations that the codomain and domain are different.

14.2   Extending to different vector spaces

Given a linear transformation T from V to W and its M. What we really is doing here is that we convert the vector v in V that is represented in the standard basis, using the linear transformation T, to a new vector v' that is represented in the standard basis of W.

What if our given vector is not represented in the standard basis' notation, or what if we want the output to be in a different basis respective (in another coordinate system)?

Let T be a linear transformation from V and W, and M be its matrix. Let β, γ be bases for V and W respectively. Then what $[T]_\beta^\gamma$ means is that consider a vector v in V that is a linear combination of β coordinate system, then we first change it to the standard coordinate system, then apply the linear transformation T which accepts vectors in the standard coordinate system and outputs vectors in the standard basis of W, then we convert the output to the γ coordinate system.

That is, let A be the change of coordinate matrix of V from β to standard basis, let B be the change of coordinate matrix of W from standard basis to γ, then $[T]_\beta^\gamma = BMA$.

Since A is from β to standard, then A is just the matrix where its columns are the vectors in β[9], oppositely B^{-1} is the matrix where its columns are the vectors in γ.

---

[8] Basis vectors are linearly independent, combining the fact that A is a square matrix, thus the matrix is always a full rank square matrix so it is invertible.

[9] Here we can see the idea that A is translating a vector in langauge β to the standard language, and A^{-1} is translating a standard language back to β.

Example 1. Consider $\beta = \{1, x, x^2, x^3, x^4\}, \gamma = \{1, x, x^2, x^3\}.T : P_4(\mathbb{R}) \to P_3(\mathbb{R}), T(p(x)) = p'(x)$.

$$
B^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad M = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \quad A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.
$$

Combine them then we get the same answer as the textbook:

$$
\begin{aligned}
[T]_\beta^\gamma = BMA &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}
\end{aligned}
$$

Same logic applies to the mid term problem too.

♋

So, since we can see that $[T]_\beta^\gamma = BMA$ for some invertible matrices A and B, and there exists a bijection between linear transformations (T) and matrices (M), we can see the function that reads β, γ, T and outputs $[T]_\beta^\gamma = BMA$ is also a bijection. Namey the map $\phi_\beta^\gamma : \mathcal{L}(V, W) \to \mathcal{M}_{m \times n}(\mathbb{F})$ given by $\phi_\beta^\gamma(T) = [T]_\beta^\gamma$ is a bijection, thus the vector spaces V and M have an isomorphism.

14.3 Composing linear transformations with change of bases

Let's end this discussion with linear transformation compositions.

To understand why $[S \circ T]_\beta^\gamma = [S]_\alpha^\gamma [T]_\beta^\alpha$ for some linear transformations S, T and some bases α, β, γ, intuitively we can see $[T]_\beta^\alpha$ reads a vector in β coordinate system, and outputs a vector in α coordinate system, then we read such vector into $[S]_\alpha^\gamma$ and outputs our final vector in the langauge of γ. So for the same input, through *same* two transformations, we should expect the output to be the same.

Through our previous equality, we can directly see that $[S]_\alpha^\gamma [T]_\beta^\alpha = GM_S A^{-1} \cdot AM_T B = GM_S M_T B = [S \circ T]_\beta^\gamma$ where A changes standard to α, B changes β to standard, and G changes standard to γ. Similarly $[T^{-1}]_\gamma^\beta = B^{-1} M^{-1} G^{-1}$ and $\left([T]_\beta^\gamma\right)^{-1} = (GMB)^{-1} = B^{-1} M^{-1} G^{-1}$, thus $[T^{-1}]_\gamma^\beta = \left([T]_\beta^\gamma\right)^{-1}$.
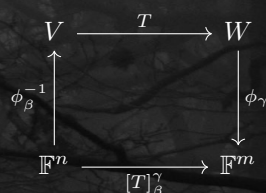
14.4 Understand the fancy graph

$$
\begin{array}{ccc}
V & \xrightarrow{\;\;T\;\;} & W \\
\phi_\beta \downarrow & & \downarrow \phi_\gamma \\
\mathbb{F}^n & \xrightarrow[\;[T]_\beta^\gamma\;]{} & \mathbb{F}^m
\end{array}
$$

First, ϕ_β and ϕ_γ are the linear transformations that convert standard basis vectors to the bases β and γ respectively, because every vector space is bijective thus isomorphic to itself, we know the change of basis function must also be bijective, thus all processes are reversable / invertible.

Moreover, $V \xrightarrow{T} W$ is a linear transformation from V to W with respect to their standard bases. And $[T]_\beta^\gamma$ is the matrix that takes in a vector in β coordinate system, and outputs a vector in γ coordinate system, where we implicitly changed the input vector to the standard coordinate system and changed the resulting vector from standard to γ.

To sum up, we explained $[T]_\beta^\gamma$ through the path

$$
\begin{array}{ccc}
V & \xrightarrow{\quad T \quad} & W \\
\phi_\beta^{-1} \uparrow & & \downarrow \phi_\gamma \\
\mathbb{F}^n & \xrightarrow[{[T]_\beta^\gamma}]{} & \mathbb{F}^m
\end{array}
$$

14.5   Another interesting observation

Since $[T]_{\beta'} = [I_V]_\beta^{\beta'}[T]_\beta[I_V]_{\beta'}^\beta$, and $[I_V]_\beta^{\beta'}$ is the inverse of $[I_V]_{\beta'}^\beta$, this means that $[T]_{\beta'}$ is similar to $[T]_\beta$. That is, let A be the change of basis (coordinate) matrix from β to standard, then $[T]_\beta = A^{-1}TA$, let B be the one for β', then $[T]_{\beta'} = B^{-1}TB$. Combine them and we have $A^{-1}TA = Q^{-1}B^{-1}TBQ$ where Q is an invertible matrix since $[T]_{\beta'}$ is similar to $[T]_\beta$. Hence, by iscolating Q, we get $A = BQ$ thus $Q = B^{-1}A$. Here the equality makes sense: let v be a vector in β, we apply transformation A to get its standard coordinate, then apply B^{-1} which translate standard coordinate to β' coordinate, thus overall we have a series of transformations that translate a vector in β to β'.

15   Rank

Since we define rank to be the dimension of the image (column space) of the matrix, despite we perform row operations to figure out the rank of a matrix, we still count the number of *columns* to determine the rank instead of rows.

Notice that here row operations will not change the linear independency of the columns, but the column space (image) *is* changed, however we are not interested in that since we just need the rank to be unchange; similarly we can see column operations also preserve the linear independency and also the column space thus also the rank.

16   Determinants

Here are the essential properties of determinants:

- $\det A = \sum\limits_{j=1}^{n} (-1)^{i+j} A_{ij} \det\left(\tilde{A}_{ij}\right) \quad \forall i \in \{1, \ldots, n\}$

- $\det A^T = \det A$

- each **row** of the matrix is a linear function of the determinant.

By the first property we may derive the ones like

1. row of zeros implies $\det A = 0$.

2. same rows implies $\det A = 0$.

3. swap rows implies $\det A' = -\det A$.

4. Upper / Lower triangular implies $\det A = \prod_{i=1}^{n} A_{ii}$.

By the second property, we maye perform the first property vertically instead of horizontally for a fixed i.

By the thid property, we can see that add one row to another does not affect the determinant.

16.1 Similar matrices and determinants

From properties 1 and 3 we may also deduce $\det(AB) = \det(A)\det(B)$, and this property is useful for showing the determinants of similar matrices.

First, since $\det(A^{-1}A) = \det(I) = 1$, by the above equality we have $\det(A^{-1})\det(A) = 1$, thus by letting A, B be similar, then $A = P^{-1}BP$, and so $\det A = \det P^{-1}\det B \det P = \det B$. This shows similar matrices have same determinants.

In addition, since $T = A^{-1}TA = [T]_\beta$ for some change of basis matrix A from a basis β to the standard basis, we can see $T \sim [T]_\beta$, thus $\det T = \det[T]_\beta$.

17 Diagonalization

Sometimes computing matrix A^{100} is sophisticated, but if A is a diagonal matrix, then A^{100} is just the diagonal matrix with each diagonal entry to the power of 100. So, if $A = P^{-1}DP$ for some diagonal matrix D, then $A^{100} = \left(P^{-1}DP\right)^{100} = P^{-1}D^{100}P$ where the intermediate $P^{-1}P$ cancels out. The idea of diagonalization is to simplify the matrix into the way that is much easier for us to perform computation, and other properties.

The matrix definition of diagonalizable and the linear transformation definition of diagonalizable are equivalent. As explained, let M be the matrix of the linear transformation T, then $[T]_\beta = A^{-1}TA$ for some change of basis matrix A, if $[T]_\beta = A^{-1}TA = D$ for some diagonal matrix D. then $T = ADA^{-1}$, thus T is diagonalizable. All our steps are reversible, thus the two definitions are equivalent.

Since we proved that similar matrices have the same eigenvalues, and a matrix M is diagonalizable if and only if it is similar to a diagonal matrix, this means that the diagonal matrix and the original matrix *must* have the same eigenvalues.

To see how we will show a matrix is diagonalizable, let's first discuss the eigenbases.

17.1 Eigenbases

Let $E_\lambda = \{v \in V \mid (A - \lambda I)\,v = 0\}$ be the eigenspace of eigenvalue λ, since $Av = \lambda v \neq \lambda'v$ for some v in E_λ and some other eigenvalue λ', this shows the union of the bases of the eigenspaces form a linearly independent set. Moreover, in order for these bases vectors to span the entire vector space, the size of the set must be equal to the dimension of the space, that is, the *sum* of the dimension of *all* E_λ needs to be equal to the dimension of the vector space.[10]

1. If the sum of the dimensions are precisely the dimention of the vector space, then let P be the matrix that the columns are the bases vectors of eigenspaces, since it is linearly independent and the column vectors form a basis, this means P is full rank thus invertible, and also represents a change of coordinate matrix.

---

[10] Geometric multiplicity is the dimension of the null space of an eigenspace; Algebric multiplicity is the power of the factor in the characteristic polynomial, that is, if $C_T(x) = (x - \lambda)^k p(x)$, then the algebric multiplicity of λ is k. Since the geometric multiplicity is always at most the algebric multiplicity, to show the matrix is not diagonalizable, it suffices to find one eigenvalue λ such that its geometric multiplicity is strictly less than its algebric multiplicity, this can be done by comparing the dimension of its null space and the power of the factor in the characteristic polynomial.

> **Claim 4**
>
> If we change the coordinate of our original matrix A to using the eigenbasis, then $P^{-1}AP$ is a diagonal matrix.   ⚥

Proof. Let the eigenbasis be $\{w_1, ..., w_n\}$ where w_i is an eigenvector of λ_i. Consider arbitrary vector v in vector space in eigenbasis coordinate system, let c_i be the ith entry of v, apply P to the entry then it becomes $c_i w_i$, since w_i is an eigenvector, apply transformation matrix A then we have $Ac_i w_i = \lambda_i c_i w_i$, now apply P^{-1} and convert the coordinate back to eigenbasis', and we can see it becomes $\lambda_i c_i$. Since our entry was arbitrary, this means

$$P^{-1}APv = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} v,$$

which is a diagonal matrix, as needed.

$\qquad$ Quod Erat Dem ∎

In this way, we first find the eigenvalues from the characteristic polynomial, then checked the existence of an eigenbasis, if it exists, then change the basis of the matrix to the eigenbasis, then the matrix is diagonal. [11] So now we have found its similar diagonal matrix (and *surprisingly* the diagonals are engenvalues [12]).

2. If the sum of the dimensions are strictly less than the dimension of the vector space, then assume for contradiction it is diagonalizable, then there exists an eigenbasis that changes A to the diagonal matrix D, however there cannot be such eigenbasis since there is no enough independent eigenvectors to span the entire vector space, thus contradiction. In this case the matrix is not diagonalizable.

17.2   Final observation

If $T : V \to V$ is diagonalizable when $\dim V = n$, then we can find n independent eigenvectors such that each eigenvector spans a one-dimensional subspace of V, and is invariant due to the vectors being eigenvectors, so $V = E_1 \oplus \cdots \oplus E_n$ for eigenspaces E_n.

Does the converse also hold? If $V = E_1 \oplus \cdots \oplus E_n$ for invariant subspaces E_n, then we can find n independent eigenvectors that their images span the entire vector space, thus by replacement theorem it has an eigenbasis, which means the matrix is diagonalizable.

---

[11] the order of the vectors in eigenbasis does not matter, so many rearrangement of the eigenvalues work here.

[12] If the diagonal matrix has entry that is not the eigenvalue, then through its characteristic polynomial they have different eigenvalues, thus not similar, which contradicts our previuos observation.