Linear Algebra Review Session Day 2

Aaron Zoll

Department of Applied Math and Statistics
Johns Hopkins University

August 21st 2024

Outline

- Vector Space Maps and Linearity
- Null Space and Range
- Rank-Nullity Theorem
- Matrices
- Inverses
- Change of Basis/Similarity
- Gaussian Elimination/Determinants
- Duality (if time)

Maps of Vector Spaces

"Functions describe the Universe" - some guy

Definition

Given two vectors spaces and a function $f : V \to W$, we call this a **map** of vector spaces with input $v \in V$ and output $f(v) =: w \in W$

- $f(x) = x^2$ $\qquad f : \mathbb{R} \to \mathbb{R}$
- $f(x, y) = \frac{x+y}{x^2+y^2}$ $\qquad f : \mathbb{R}^2 \to \mathbb{R}$
- $f(x, y, z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$ $\qquad f : \mathbb{R}^3 \to \mathbb{R}^2$
- $f(a, b) = a + bi$ $\qquad f : \mathbb{R} \to \mathbb{C}$
- $f(z, w) = (2z + w, -3z + 2w, z - 4w)$ $\qquad f : \mathbb{C}^2 \to \mathbb{C}^3$
- $f(z) = \frac{1}{2}\|z\|_2^2$ $\qquad f : \mathbb{F}^n \to \mathbb{F}$
- $f(a + bi) = a - bi$ $\qquad f : \mathbb{C} \to \mathbb{C}$

Linear Maps

In general, functions are simply too complex for us to work with.

Definition

A map $T : V \to W$ is called **linear** if for all $v, w \in V$ and $\lambda \in \mathbb{F}$, we have:

1. Additivity:
$$T(u+v) = T(u) + T(v)$$
2. Homogeneity:
$$T(\lambda v) = \lambda T(v)$$

Remark

Note that these maps "look" like vector spaces. They satisfy the two main requirements. Furthermore, from the requirement of vector spaces being closed under addition and scalar multiplication, these functions are well defined!

The Vector Space of Linear Maps

Definition

The space of linear maps from V to W, we define as

$$\mathcal{L}(V, W) := \{T : V \to W : \text{T is linear}\}$$

Not hard to check that for $S, T \in \mathcal{L}(V, W)$, we have that

$$(S + T)(v) := Sv + Tv \qquad (\lambda S)(v) := \lambda(Sv)$$

are both linear. That is $S + T, \lambda S \in \mathcal{L}(V, W)$ Therefore, $\mathcal{L}(V, W)$ is itself a vector space!

Examples of Linear Maps

- The zero functions $O \in \mathcal{L}(V, W)$ where $v \mapsto 0 \ \forall v \in V$
  - $O(v + w) = 0 = 0 + 0 = O(v) + O(w)$
  - $O(\lambda v) = 0 = \lambda 0 = \lambda O(v)$
  - Only constant linear function (why?)
- The identity map $I \in \mathcal{L}(V, V)$ where $v \mapsto v$
- $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ sending $f \mapsto f'$
- $R \in \mathcal{L}(\mathbb{F}^\infty, \mathbb{F}^\infty)$ sending $(x_1, x_2, \ldots) \mapsto (0, x_1, x_2, \ldots)$
- $M \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ sending $f(x) \mapsto x^2 f(x)$
- Anything with only "linear components" that is:
  - $T(x, y, z) = (2x - y, 7x + 3z, x + y - z)$
  - $T(x, y) = (3x, 2y - x, 4y)$
  - $T(x, y, z) = (3x + a, 4y + b)$ is linear iff $a = b = 0$
  - $T(x, y) = (2x + a, by^2 + 7x, c\sin(x) + d\sqrt{y} + kx)$ is linear iff $a = b = c = d = 0$. k is free to be anything!

"Multiplying" Linear Maps

$$\begin{tikzcd} U \arrow[r, "S"'] \arrow[rr, bend left, dashed, "ST"] & V \arrow[r, "T"'] & W \end{tikzcd}$$

Definition

For linear maps $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$ we define the **left multiplication** of T by S as $ST \in \mathcal{L}(U, W)$:

$$(ST)(u) := S(T(u))$$

Remark

This is exactly just functional composition, but the **linearity** *is nice and preserved:*

$$ST(u_1 + \lambda u_2) = S(T(u_1 + \lambda u_2)) = \underbrace{S(Tu_1 + \lambda Tu_2)}_{\text{linearity of } T} = \underbrace{S(Tu_1) + \lambda S(Tu_2)}_{\text{linearity of } S} = ST(u_1) + \lambda ST(u_2)$$

Null Space (motivation)

Suppose we have the following linear system we are aiming to "solve"

$$T(x) = b \qquad T \in \mathcal{L}(V, W),\ x \in V,\ b \in W$$

Pictorially this can look like:

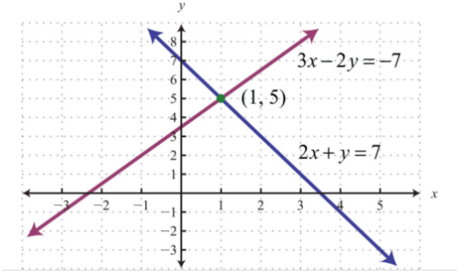
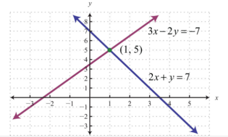


Figure: https://mathpqjq.com/solving-a-linear-system-by-graphing/

Null Space (motivation)



Here we have

$$T(x, y) = (3x - 2y, 2x + y) \text{ and } b = (-7, 7)$$

(in the standard basis) The solution to this problem is

$$(x, y) = (1, 5)$$

which is easily verifiable We then ask, is this the only solution? When are there infinite solutions?

Null Space

Definition

We definite the **null space** of a linear map $T : V \to W$ as

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

Remark

Often this is called the **kernel** *as well and is used in many areas of mathematics*

- $0 \in null(T)$ *always*
  - $\cancel{T(0)} = T(0+0) = T(0) + \cancel{T(0)}$
- *Yields all additional solutions! So if $\hat{x}$ solves $T(x) = b$, then $\hat{x} + v$ for any $v \in null(T)$ does as well*
  - $T(x+v) = T(x) + T(v) = T(x) + 0 = b$

Example

Suppose

$$\varphi(z_1, z_2, z_3) = z_1 + 2z_2 + 3z_3, \quad \varphi \in \mathcal{L}(\mathbb{C}^3, \mathbb{C})$$

$$\begin{aligned}
\text{null}(\varphi) &= \{(z_1, z_2, z_3) \in \mathbb{C}^3 : z_1 + 2z_2 + 3z_3 = 0\} \\
\Rightarrow &= \{(z_1, z_2, z_3) \in \mathbb{C}^3 : z_1 = -2z_2 - 3z_3\} \\
\Rightarrow &= \{(-2z_2 - 3z_3, z_2, z_3) \in \mathbb{C}^3\} \\
\Rightarrow &= \{z_2(-2, 1, 0) + z_3(-3, 0, 1) : z_2, z_3 \in \mathbb{C}\} \\
\Rightarrow &= \text{span}(\{(-2, 1, 0), (-3, 0, 1)\})
\end{aligned}$$

In this example, the null span reduces to the span of just two vectors! These are linearly independent, and so the null space is a two dimensional subspace of $V = \mathbb{C}^3$ with the above basis as a valid one.

Null Space is a Subspace

This is always the case:

Theorem

null(T) is always a subspace of V

Proof.

Suppose $v, w \in \text{null}(T)$

$$T(v + w) = Tv + Tw = 0 + 0 = 0$$

$$T(\lambda v) = \lambda T(v) = \lambda 0 = 0$$

Therefore, $v + w, \lambda v \in \text{null}(T)$ and so it is closed under linearity conditions, so it is a subspace. □

Injective

Definition

A map $T : V \to W$ (not necessarily linear) is **injective** or **one-to-one** if

$$f(x) = f(y) \implies x = y$$

Equivalently,

$$x \neq y \implies f(x) \neq f(y)$$

This condition in general can be difficult to verify because we would need to check this for all $x, y \in V$. Linear maps are far nicer in that $T \in \mathcal{L}(V, W)$ is injective if and only if $\text{null}(T) = \{0\}$

Proof.

($\implies$) If T is injective, then if $Tv = 0 = T(0)$, then $v = 0$, so $\text{null}(T) = \{0\}$
($\impliedby$) If $\text{null}(T) = \{0\}$, then if $Tv = Tw$, by linearity, we have $T(v - w) = 0$, and so $v - w \in \text{null}(T) = \{0\}$
Therefore, $v = w$ and T is injective. $\square$

Previous Graphical Example

Recall when we had

$$T(x, y) = (3x - 2y, 2x + y)$$

To find null(T) we have to simultaneously have

$$\begin{aligned} 3x - 2y &= 0 \\ 2x + y &= 0 \end{aligned}$$

Adding the first equation to twice the second, we get

$$7x = 0$$

so $x = 0$. And subbing that in we get $y = 0$. Therefore, the only vector in the null space is the zero vector, so T is injective, and so the entire set of solutions is

$$\hat{x} + \text{null}(T) = \hat{x} + \{0\} = \{\hat{x}\}$$

There is only one unique solution!

Range

We now have described one fundamental subspace of the input space: $\text{null}(T) \subseteq V$ When working with maps $T : V \to W$, all of V is acted on and sent somewhere in W.

Definition

We define the **range** of a map to be the set of all outputs

$$\text{range}(T) := \{Tv : v \in V\} = \{w \in W : w = Tv \text{ for some } v \in V\}$$

In the first example, we had

$$T(z_1, z_2, z_3) = z_1 + 2z_2 + 3z_3$$

Is it easy to see that $\text{range}(T) = \mathbb{C}$, for if we pick any $z \in \mathbb{C}$, setting

$$(z_1, z_2, z_3) = (z, 0, 0)$$

would work. Similarly, setting

$$(z_1, z_2, z_3) = (0, z/2, 0)$$

or even

$$(z_1, z_2, z_3) = (z/6, z/6, z/6)$$

works. Thus T was not injective (but we know that because it had a nontrivial null space!

Example

Suppose $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ is defined by (note the change to the "basis representation")

$$T(x, y) = (2x, 6y, x + y) = x \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} + y \begin{bmatrix} 0 \\ 5 \\ 1 \end{bmatrix}$$

for any $x, y \in \mathbb{R}$. Therefore, the range the already in "span form"
$\text{range}(T) = \text{span}(\{(2, 0, 1), (0, 5, 1)\})$

Example

Suppose $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ is defined by

$$T(x,y,z) = (2x+4y-3z, x+2y+z) = x\begin{bmatrix} 2 \\ 1 \end{bmatrix} + y\begin{bmatrix} 4 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

for any $x,y,z \in \mathbb{R}$. Still, the range the already in "span form"

$$\text{range}(T) = \text{span}(\{(2,1),(4,2),(-3,1)\})$$

but this is no longer a basis because the list is too long. We can remove one to get a basis of the range.

Which of the following is correct?

1. $\text{range}(T) = \text{span}(\{(4,2),(-3,1)\})$
2. $\text{range}(T) = \text{span}(\{(2,1),(-3,1)\})$
3. $\text{range}(T) = \text{span}(\{(2,1),(4,2)\})$

Range is a Subspace

We saw that the null space describes a subspace of the input. Does the range do the same for the output?

Theorem

For map $T \in \mathcal{L}(V, W)$, range(T) is a subspace of W

Proof.

Suppose $w_1, w_2 \in \text{range}(T)$. Then there exists $v_1, v_2 \in V$ for which

$$w_1 = Tv_1 \qquad w_2 = Tv_2$$

Therefore,

$$w_1 + w_2 = Tv_1 + Tv_2 = T(v_1 + v_2) \qquad \lambda w_1 = \lambda T v_1 = T(\lambda v_1)$$

Since V is a vector space, both $v_1 + v_2$ and λv_1 are in V, so $w_1 + w_2, \lambda w_1 \in \text{range}(T)$

$\square$

Surjective

Definition

A map $T : V \to W$ (not necessarily linear) is **surjective** or **onto** if

$$\text{range}(T) = W$$

Equivalently, for any $w \in W$ there exists a $v \in V$ with $Tv = w$

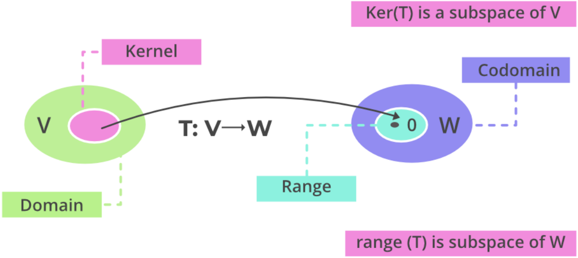


Figure: https://www.geeksforgeeks.org/linear-mapping/

Exact Sequence (quick aside)

Note: Maps "shrink space".

All these tools let us get an understanding of the structure...

$$\cdots \xrightarrow{f_{-1}} G_{-1} \xrightarrow{f_0} G_0 \xrightarrow{f_1} G_1 \xrightarrow{f_2} \cdots$$

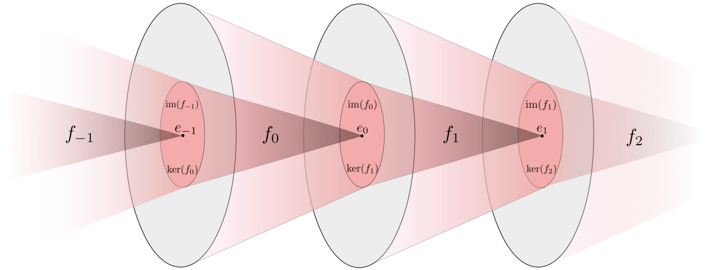


Figure: Wikipedia

Rank-Nullity Theorem (first "Big" theorem)

also known as "The Fundamental Theorem of Linear Maps"

Suppose V is finite dimensional and $T \in \mathcal{L}(V, W)$. Then range(T) is finite dimensional and

$$\boxed{\dim V = \underbrace{\dim \text{null}(T)}_{nullity} + \underbrace{\dim \text{range}(T)}_{rank}}$$

Proof.

Let $\{u_1, ..., u_k\}$ be a basis of null(T) and extend to a basis of all of V : $\{u_1, ..., u_k, v_1, ..., v_r\}$
Therefore, $\dim \text{null}(T) = k$ and $\dim V = k + r$, so all we need to show is that $\dim \text{range}(T) = r$

Suppose $v \in V$, then since the above is a basis, we can write
$v = a_1 u_1 + ... + a_k u_k + b_1 v_1 + ... + b_r v_r$ and since all the $u_i{}'s$ are in the null space:
$T(v) = b_1 Tv_1 + ... b_r Tv_r$. Since this holds for all $v \in V$, we can represent all the $w \in$ range(T) as a linear combination of the $Tv_i's$. Therefore $\{Tv_1, ..., Tv_r\}$ spans the range. Finally, if

$c_1 Tv_1 + ... + c_r Tv_r = 0$, then by linearity $cv_1 + ... + cv_r \in \text{null}(T)$ That is, some linear combination of the $v_i's$ is a linear combination of the $u_i's$. But they are linearly **independent** from each other! The two spans only intersect at the zero vector. So this sum must be 0 as well, and thus each c_i is 0.
i.e. $\{Tv_1, ..., Tv_r\}$ is also linearly independent, and so it forms a basis with dimension r $\square$

Remark on Rank/Nullity

Corollary

The rank is bounded above by the dimension of the input and output space:

$$\dim range(T) \leq \min\{\dim V, \dim W\}$$

Corollary

If $\dim V > \dim W$ *then no* $T \in \mathcal{L}(V, W)$ *is* **injective**

$$\begin{aligned} \dim null(T) &= \dim V - \dim range(T) \\ &\geq \dim V - \dim W > 0 \end{aligned}$$

Corollary

If $\dim V < \dim W$ *then no* $T \in \mathcal{L}(V, W)$ *is* **surjective**

$$\begin{aligned} \dim range(T) &\leq \dim V \\ &< \dim W \end{aligned}$$

Remark on Solving Linear Systems

If we are solving $T(x) = b$ with $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, that is we are solving

$$\begin{aligned} T_{1,1}x_1 + T_{1,2}x_2 + \ldots + T_{1,n}x_n &= b_1 \\ T_{2,1}x_1 + T_{2,2}x_2 + \ldots + T_{2,n}x_n &= b_2 \\ &\vdots \\ T_{m,1}x_1 + T_{m,2}x_2 + \ldots + T_{m,n}x_n &= b_m \end{aligned}$$

with each $T_{i,j} \in \mathbb{R}$. Then if

- $n > m$, so we have more variables than equations: more unknowns that constraints, then T is **not injective** and we have infinite solutions, if ones exist (we call this being consistent)
- $n < m$, so we have more equations than variables: more constraints than unknowns, then T is **not surjective** and we have vectors $b \in \mathbb{R}^m$ without any solutions!

Matrices (finally)

Definition

A **matrix** is an array of numbers

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & A_{2,3} & \dots & A_{2,n} \\ A_{3,1} & A_{3,2} & A_{3,3} & \dots & A_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & A_{m,3} & \dots & A_{m,n} \end{bmatrix}$$

Matrices

The *matrix of a (finite dimensional) linear operator* **with respect to certain bases** is an array of numbers.

Definition

Given a linear map $T : V \to W$ and bases $\mathcal{B}_V = \{v_1, ...v_n\}$ and $\mathcal{B}_W = \{w_1, ..., w_m\}$ then the matrix representation

$$\mathcal{M}(T, \mathcal{B}_V, \mathcal{B}_W) = A = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & A_{2,3} & \dots & A_{2,n} \\ A_{3,1} & A_{3,2} & A_{3,3} & \dots & A_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & A_{m,3} & \dots & A_{m,n} \end{bmatrix}$$

such that for each $x \in V$, we have

$$Tx = y \iff A[x]_{\mathcal{B}_V} = [y]_{\mathcal{B}_W}$$

Matrices

Let's decompose what the previous statement means:

- Recall that given a basis of a vector space: $\mathcal{B}_V$, then for any $x \in V$ we can write $x = a_1 v_1 + ... + a_n v_n$ uniquely.
- Therefore $Tx = T(a_1 v_1 + ... + a_n v_n) = a_1 T(v_1) + ... + a_n T(v_n)$
- Linearity gives us that we *only* need to know where $v_1, ..., v_n$ is sent.
- However, $Tv_i \in W$. So when it is mapped there, it can also be represented uniquely as a sum of finitely many values.
- Thus, $Tv_i = b_1 w_1 + ... b_m w_m$ from the basis vectors in $\mathcal{B}_W$
- However, we need this to hold for each basis vector v_i, so we need two indices. We arrive at

$$Tv_j = A_{1,j} w_1 + ... + A_{m,j} w_m = \sum_{i=1}^{m} A_{i,j} w_i$$

- In other words, to compute/find a matrix of a linear operator, we input each basis vector, see where it goes, and represent it in the output basis

Example

This is a lot to unpack... let's see an example:

$$T(x, y) = (2x + 3y, -x + y, 4x - 7y)$$

If we start with the standard bases

$$\mathcal{B}_V = \{e_1, e_2\} \qquad \mathcal{B}_W = \{f_1, f_2, f_3\}$$

Then we get that

$$\begin{aligned} T(e_1) &= (2, -1, 4) = 2f_1 + (-1)f_2 + 4f_3 \\ T(e_2) &= (3, 1, -7) = 3f_1 + 1f_2 + (-7)f_3 \end{aligned}$$

If we "pattern match" the corresponding coefficients on the output bases to the "array representation" above we get

$$A = \begin{bmatrix} 2 & 3 \\ -1 & 1 \\ 4 & -7 \end{bmatrix}$$

which is exactly what we would get it we get wrote $T(x, y)$ is a standard "vector" form

Example

Again consider

$$T(x, y) = (2x + 3y, -x + y, 4x - 7y)$$

but now let's have the bases be different:

$$\mathcal{B}'_V = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\} \qquad \mathcal{B}'_W = \left\{ \begin{bmatrix} 8 \\ 1 \\ -10 \end{bmatrix}, \begin{bmatrix} 1/3 \\ -1 \\ 5 \end{bmatrix}, \begin{bmatrix} -9 \\ 2 \\ -7 \end{bmatrix} \right\}$$

Doing the same trick we get:

$$T(v_1) = (8, 1, -10) = 1w_1 + 0w_2 + 0w_3$$
$$T(v_w) = (1, -3, 15) = 0w_1 + 3w_2 + 0w_3$$

So this matrix representation is

$$\mathcal{M}(T, \mathcal{B}'_V, \mathcal{B}'_W) = \begin{bmatrix} 1 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix}$$

Example

Okay, this all seems a little contrived and pointless, when would this be needed? Well consider "non-standard" vector spaces. Nothing too crazy, but let's consider $V = \mathcal{P}_3(\mathbb{R})$. Let $D : V \to V$ be differentiation. So $T(p) = p'$ If we just consider the standard basis $\mathcal{B} = \{1, x, x^2, x^3\}$, then

$$\begin{aligned} D(1) &= 0 = 0v_1 + 0v_2 + 0v_3 + 0v_4 \\ D(x) &= 1 = 1v_1 + 0v_2 + 0v_3 + 0v_4 \\ D(x^2) &= 2x = 0v_1 + 2v_2 + 0v_3 + 0v_4 \\ D(x^3) &= 3x^2 = 0v_1 + 0v_2 + 3v_3 + 0v_4 \end{aligned}$$

and so we get a matrix representation for an operation like differentiation. This will be useful in the future: $\mathcal{M}(D) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Recap

In general, recall we have a basis for the input space

$$\mathcal{B}_V = \{v_1, \ldots, v_n\}$$

and a basis for the output space

$$\mathcal{B}_W = \{w_1, \ldots, w_m\}$$

We know that for each input basis vector, we can map it through T, and write it as a linear combination of the output basis vectors:

$$Tv_j = A_{1,j}w_1 + \ldots + A_{m,j}w_m$$

This is precisely then the j^{th} column of the matrix:

$$\mathcal{M}(T) = \begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & \cdots & v_k & \cdots & v_n \end{array} \\ \begin{pmatrix} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{m,k} & & \end{pmatrix} \end{array}.$$

Figure: *Linear Algebra Done Right*

Recap

In other words

$$\mathcal{M}(T, \mathcal{B}_V, \mathcal{B}_W) = \begin{bmatrix} | & | & & | \\ [Tv_1]_{\mathcal{B}_W} & [Tv_2]_{\mathcal{B}_W} & \dots & [Tv_n]_{\mathcal{B}_W} \\ | & | & & | \end{bmatrix}$$

$$\mathcal{M}(T, \mathcal{B}_V, \mathcal{B}_W) = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & A_{2,3} & \dots & A_{2,n} \\ A_{3,1} & A_{3,2} & A_{3,3} & \dots & A_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & A_{m,3} & \dots & A_{m,n} \end{bmatrix}$$

Why do we care?

Setting up matrices in this way let's us

1. Utilize linearity to evaluate a function with much smaller amount of information (just mn numbers).
2. Allows freedom of choice of basis which will be extremely useful later (foreshadowing change of basis/spectral decomposition/Schur decomposition/SVD/etc.)
3. This explains why matrix-vector multiplication works the way it does (and matrix-matrix multiplication too)!

Matrix Vector Multiplication

Suppose $x = c_1 v_1 + ... + c_n v_n$ then

$$T(x) = T(\sum_{i=1}^{n} c_i v_i) = \sum_{i=1}^{n} c_i Tv_i = y \implies [y]_{\mathcal{B}_W} = \sum_{i=1}^{n} \underbrace{c_i [Tv_i]_{\mathcal{B}_W}}_{\text{scaled columns}}$$

Then since we have a matrix representation

$$A = \begin{bmatrix} | & | & & | \\ [Tv_1]_{\mathcal{B}_W} & [Tv_2]_{\mathcal{B}_W} & \dots & [Tv_n]_{\mathcal{B}_W} \\ | & | & & | \end{bmatrix}$$

we have (recall that matrix vector multiplication exactly yields a linear combination of the columns)

$$A[x]_{\mathcal{B}_V} = \begin{bmatrix} | & | & & | \\ [Tv_1]_{\mathcal{B}_W} & [Tv_2]_{\mathcal{B}_W} & \dots & [Tv_n]_{\mathcal{B}_W} \\ | & | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \sum_{i=1}^{n} c_i [Tv_i]_{\mathcal{B}_W}$$

Matrix Multiplication

It'll be useful to define the set of matrices from $\mathbb{F}^n \to \mathbb{F}^m$ for given field $\mathbb{F}$

Definition

The set of matrices from $\mathbb{F}^n \to \mathbb{F}^m$ (so m by n matrices) will be denoted as $M_{m,n}(\mathbb{F})$. If $m = n$ we can just use $M_n(\mathbb{F})$

Consider the follow matrices $A \in M_{m,n}$ and $C \in M_{n,p}$

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & A_{2,3} & \dots & A_{2,n} \\ A_{3,1} & A_{3,2} & A_{3,3} & \dots & A_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & A_{m,3} & \dots & A_{m,n} \end{bmatrix} \qquad C = \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & \dots & C_{1,n} \\ C_{2,1} & C_{2,2} & C_{2,3} & \dots & C_{2,n} \\ C_{3,1} & C_{3,2} & C_{3,3} & \dots & C_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{m,1} & C_{m,2} & C_{m,3} & \dots & C_{m,n} \end{bmatrix}$$

We can view the product of $AC \in M_{m,p}$ in three ways:

Naive Approach

If you had the same first experience with matrices I did, you were taught this "row by column" thing:

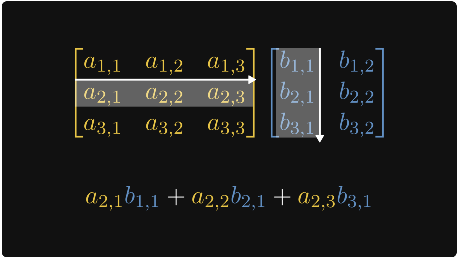


Figure: ThePalindrome.org

It is simple enough to memorize, and is drilled enough to seem natural. In algebraic terms we get an inner product of the j^{th} row of A and the k^{th} column of C

$$\boxed{[AC]_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k} = \langle a_j, C_k \rangle} \qquad 1 \leq j \leq m, \ \ 1 \leq k \leq p$$

Matrix Multiplication (as composition of linear maps)

Consider $T : U \to V$ and $S : V \to W$ with dimensions p, n, m respectively. Thus, we can have bases

$$\mathcal{B}_U = \{u_1, ..., u_p\} \qquad \mathcal{B}_V = \{v_1, ..., v_n\} \qquad \mathcal{B}_W = \{w_1, ..., w_m\}$$

We can then define the matrices

$$A := \mathcal{M}(S, \mathcal{B}_V, \mathcal{B}_W) \qquad C := \mathcal{M}(T, \mathcal{B}_U, \mathcal{B}_V)$$

We already have seen that $ST : U \to W$ is linear, and so we aim to find the matrix $AC := \mathcal{M}(ST, \mathcal{B}_U, \mathcal{B}_W) \in M_{m,p}$. Recall, to find the matrix, we plug in the input basis vectors, and see how they are written in the output. So for any $1 \leq k \leq p$:

$$\begin{aligned} (ST)u_k &= S\left(\sum_{r=1}^{n} C_{r,k} v_r\right) \\ &= \sum_{r=1}^{n} C_{r,k}(Sv_r) \\ &= \sum_{r=1}^{n} C_{r,k} \sum_{j=1}^{m} A_{j,r} w_j \\ &= \sum_{j=1}^{m} \boxed{\left(\sum_{r=1}^{n} A_{j,r} C_{r,k}\right)} w_j \end{aligned}$$

Matrix as a list of vectors

Finally, we can view the right matrix as a list of p vectors in $\mathbb{F}^n$:

$$C = \begin{bmatrix} | & | & & | \\ C_1 & C_2 & \dots & C_p \\ | & | & & | \end{bmatrix}$$

The matrix A then *acts on each vector* simultaneously and we get

$$AC = \begin{bmatrix} | & | & & | \\ AC_1 & AC_2 & \dots & AC_p \\ | & | & & | \end{bmatrix}$$

A good exercise would be to verify from the matrix-vector multiplication (a linear combination of the columns) that this is equivalent to the previous.

Inverses

Definition

For $T \in \mathcal{L}(V, W)$ is **invertible** if there exists a $S \in \mathcal{L}(W, V)$ such that

$$ST = I_V \in \mathcal{L}(V, V) \quad \textit{and} \quad TS = I_W \in \mathcal{L}(W, W)$$

That is, $ST(v) = v$ and $TS(w) = w$ for all $v \in V$ and $w \in W$

Remark (uniqueness of inverses)

If S_1 and S_2 are inverses of T then

$$S_1 = S_1 I_W = S_1(TS_2) = (S_1 T)S_2 = I_V S_2 = S_2$$

So $S_1 = S_2$

Remark

Furthermore $ST = I$ implies that $TS = I$ (in the respective vector spaces). Thus we call S the **inverse of T** *and denote it as*

$$S = T^{-1}$$

Conditions for Invertibility

Theorem

$T \in \mathcal{L}(V, W)$ is invertible if and only if T is injective and surjective

Proof.

- $(\Longrightarrow)$ Suppose T is invertible.
  - (Injectivity) Suppose $T(u) = T(v)$. Then since T^{-1} exists, we multiply on the left on both sides to get $u = T^{-1}Tu = T^{-1}Tv = v$. So $u = v$ and T is injective.
  - (Surjectivity) Now let $w \in W$ be arbitrary and since $T^{-1} \in \mathcal{L}(W, V)$ we have that $T^{-1}w \in V$. So then $T(T^{-1}w) = w$, so there exists something in V that is mapped to any $w \in W$
- $(\Longleftarrow)$ Since T is both injective and surjective, each $w \in W$ has a unique $v \in V$ such that $Tv = w$. Defining $S : W \to V$ element wise, so that $Sw = v$, the unique input of T
  - $TS = I$ by definition
  - $ST = I$ because $T(ST(v)) = (TS)(Tv) = I \circ Tv$, and since T is injective, we get that $(ST)(v) = v$
  - Finally $T(S(w_1) + \lambda S(w_2)) = TS(w_1) + \lambda TS(w_2) = w_1 + \lambda w_2$. Since S is defined by $T(S(w)) = w$, then $T(S(w_1 + \lambda w_2)) = w_1 + \lambda w_2$. Injectivity of T concludes the linearity condition: $S(w_1 + \lambda w_2) = S(w_1) + \lambda S(w_2)$

□

Conditions for Invertibility

When the map is **injective** (one-to-one) and **surjective** (onto), then we get a nice *equivalence* between the input and output space. We can then actively construct the inverse.

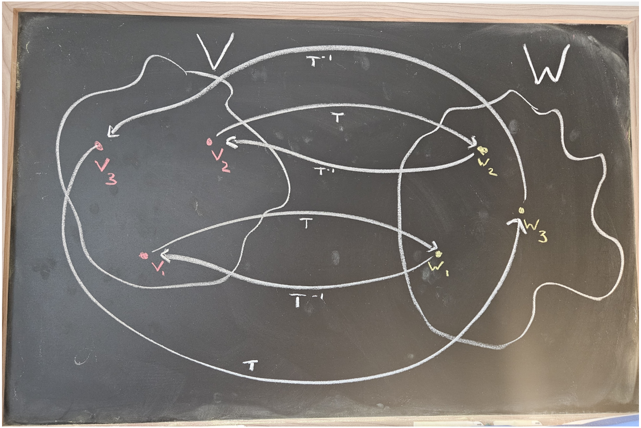


Figure: Both injective and surjective

Conditions for Invertibility

If we don't have injectivity, then we lose the ability to define a unique output of the inverse map. Recall, *any* function must have a unique output (vertical line test).

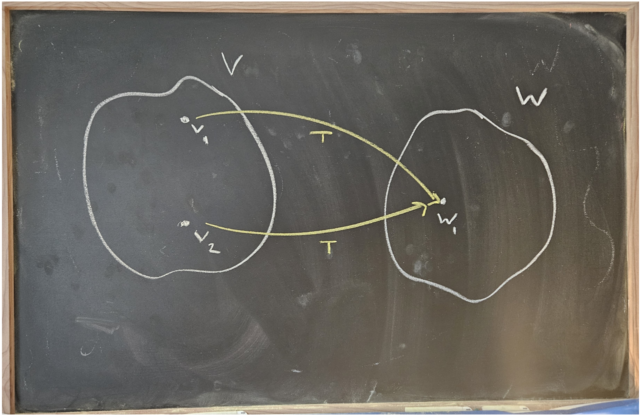


Figure: Not injective

Conditions for Invertibility

If we don't have surjectivity, then we lose the ability to define the map for some of the desired inputs (elements of W).

Note: if the map is injective, then we can always define a map from T^{-1} : range$(T) \to V$. Hence why we want surjectivity to get a map from all of $W \to V$.

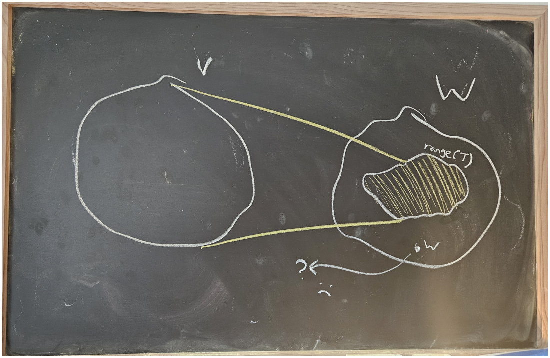


Figure: Not injective

Conditions for Invertibility

Final note of invertibility. In finite dimensions we really only need one of these conditions:

Remark

$$T \text{ invertible} \iff T \text{ injective} \iff T \text{ surjective}$$

with the proof coming directly from the Rank-Nullity Theorem

Change of Basis

Recall that *both* vectors and matrices require a **basis** to represent them. Typically, we just deal with the "standard basis" as this is the easiest to understand and utilize, but there are many reasons to perform a **change of basis**.

Definition

Suppose we have a vector $v \in V$ and a basis $\mathcal{B} = \{v_1, ... v_n\}$. We then can write v in a vector form as

$$[v]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ a_2 \\ \ddots \\ a_n \end{bmatrix} = a_1 v_1 + ... a_n v_n$$

Given a new basis $\mathcal{B}' = \{v_1', ..., v_n'\}$ we can perform a **change of basis** using the **change of basis matrix**

$$S_{\mathcal{B} \to \mathcal{B}'} \in M_n(\mathbb{F})$$

such that $S_{\mathcal{B} \to \mathcal{B}'}[v_j]_{\mathcal{B}} = [v_j]_{\mathcal{B}'}$ for each j and therefore

$$S_{\mathcal{B} \to \mathcal{B}'}[v]_{\mathcal{B}} = [v]_{\mathcal{B}'}$$

Essentially, all we are saying is that we can transform $[v]_{\mathcal{B}} \mapsto [v]_{\mathcal{B}'}$ with a linear transformation simply defined from writing each new vector in the coordinates of the old

Change of Basis

Remark

Note: Change of Basis is, I think, the hardest concept to fully understand in linear algebra. However, it is essential to understanding all the matrix decompositions.

We start with constructing $S_{\mathcal{B}\to\mathcal{B}'}$ be first constructing its inverse. We will explain in a bit why this exists.
Since $\mathcal{B}$ is our original basis, we can write each new basis vector as a linear combination of them

$$\begin{aligned} v_1' &= A_{1,1}v_1 + A_{2,1}v_2 ... + A_{n,1}v_n \\ v_2' &= A_{1,2}v_1 + A_{2,2}v_2 ... + A_{n,2}v_n \\ &\vdots \\ v_n' &= A_{1,n}v_1 + A_{2,n}v_2 ... + A_{n,n}v_n \end{aligned}$$

Letting

$$T = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & \dots & A_{2,n} \\ \vdots & \vdots & \ddots & \ddots \\ A_{n,1} & A_{n,2} & \dots & A_{n,n} \end{bmatrix}$$

we can verify that $T[v_j]_{\mathcal{B}} = T[v_j']_{\mathcal{B}'} = [v_j']_{\mathcal{B}}$ (why?)

Change of Basis

Is is typically easiest to think about change of basis when the first basis is just the standard basis. Then the new basis (in vector form, with respect to the standard basis) looks more as expected*, and moreover:

$$T_{\mathcal{B}'} := S_{\mathcal{B}' \mapsto \mathcal{I}} = \begin{bmatrix} | & | & & | \\ v_1' & v_2' & \dots & v_n' \\ | & | & & | \end{bmatrix}$$

That is, we don't need to worry how to write the new basis vectors in the old basis, and the matrix T just has columns that are precisely the new basis vectors.

Remark

So, plugging in the vector with a one in the j^{th} component and zeros elsewhere extracts the vector v_j'. In other words, plugging in $[v_j']_{\mathcal{B}'}$ yields $[v_j']_{\mathcal{B}}$ (recall here $\mathcal{B}$ is the standard basis.
Thus, $T_{\mathcal{B}'}$ is a map from $\mathcal{B}' \mapsto \mathcal{B}$, this is the wrong direct (and very unintuitive in my opinion).

Change of Basis

However, since $T_{\mathcal{B}'}$ has linearly independent columns (as they are just the basis vectors), any $T_{\mathcal{B}'}x$ just looks like a linear combination of those columns. So $T_{\mathcal{B}'}x = 0 \iff x = 0$, so $\text{null}(T_{\mathcal{B}'}) = \{0\}$, this is trivial and $T_{\mathcal{B}'}$ is invertible. Finally we can define

Definition

The **change of basis matrix** from the standard basis $\mathcal{I} = \{e_1, ..., e_n\}$ to a new basis $\mathcal{B}' = \{v'_1, ..., v'_n\}$ is

$$S_{\mathcal{I}\mapsto\mathcal{B}'} = T_{\mathcal{B}'}^{-1}$$

for $T_{\mathcal{B}'} = \begin{bmatrix} | & | & & | \\ v'_1 & v'_2 & \dots & v'_n \\ | & | & & | \end{bmatrix}$

Remark

To construct a change of basis matrix between any two bases, we can simply compose these linear maps (it is an exercise to show that the composition of invertible maps is still invertible):

$$S_{\mathcal{B}\mapsto\mathcal{B}'} = S_{\mathcal{I}\mapsto\mathcal{B}'}S_{\mathcal{B}\mapsto\mathcal{I}} = S_{\mathcal{B}'\mapsto\mathcal{I}}{}^{-1}S_{\mathcal{B}\mapsto\mathcal{I}} = T_{\mathcal{B}'}{}^{-1}T_{\mathcal{B}}$$

Example

Consider the vector

$$v = (1, 2, 3, 4)$$

and bases

$$\mathcal{I} = \{e_1, e_2, e_3, e_4\} \qquad \mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

We can easily generate the $T_{\mathcal{B}}$ by lining up the columns:

$$T_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}$$

Now we just have to invert this and multiply by

$$[v]_{\mathcal{I}} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \qquad T_{\mathcal{B}}{}^{-1}[v]_{\mathcal{I}} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 1 & 0.5 & -1 & -0.5 \\ 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ -3 \\ 3 \\ 3 \end{bmatrix}$$

Note, we can verify that we must multiply by the inverse, $T_{\mathcal{B}}{}^{-1}$ by checking a basis vector.

Change of Basis (linear maps)

Ok well this all seemed like a round about way to do something we already knew... If we want to find a linear combination of some vectors that equal something desired, it is exactly just solving a linear system:

$$Ax = b \iff \begin{bmatrix}\text{some column vectors}\end{bmatrix}\begin{bmatrix}\text{some unknowns}\end{bmatrix} = \begin{bmatrix}\text{desired vector}\end{bmatrix}$$

In this particular case, all we are doing is solving

$$T_{\mathcal{B}}x = [v]_{\mathcal{I}}$$

However, change of basis is especially useful when considering **linear maps**!

Change of Basis (linear maps)

Theorem

Given a matrix $A = \mathcal{M}(T, \mathcal{B})$ representing linear map $T : V \to V$ with basis $\mathcal{B}$. Then the matrix $D = \mathcal{M}(T, \mathcal{B}')$ for a new basis $\mathcal{B}'$ looks like

$$D = S_{\mathcal{B} \to \mathcal{B}'} A S_{\mathcal{B}' \to \mathcal{B}}$$

If $\mathcal{B} = \mathcal{I}$, the standard basis, then S is the matrix whose **columns are the basis vectors** *and*

$$D = S^{-1}AS \iff A = SDS^{-1}$$

Definition

If two matrices can be related by an invertible (nonsingular) matrix in the above way, we say two (square) matrices are **similar**:

$$A, B \text{ are similar if and only if } A = SBS^{-1} \text{ for some invertible matrix S}$$

Similarity

Remark

As we will see, similar matrices share many, many properties (trace, determinant, eigenvalues, multiplicity, etc.).

Moreover, the big result is that all similar matrices **are the same linear map, just under a different basis**, *so of course they should have similar structure!*

Remark

Furthermore, similarity forms an **equivalence relation**. *That is:*

1. *A is* **similar** *to A for any $A \in M_n$*
2. *If A is* **similar** *to B, then B is* **similar** *to A*
3. *If A is* **similar** *to B, and B is* **similar** *to C, then A is* **similar** *to C*

Therefore, we can choose one matrix to "represent" the equivalence class of all its similar matrices. In other words, there exists **just one matrix**, *let's call it J, that gives all the information about the map.Then any other matrix similar to J, will have similar features, just perhaps in a different basis (e.g. different eigenvectors)*

Gaussian Elimination

Note: In many cases, we set up a system that has an invertible matrix, but finding the inverse is difficult, even for a computer!

Remark

When solving

$$Ax = b$$

or recall the matrix-matrix multiplication can be viewed as repeated matrix-vector multiplication, so this generalizes, we often do not calculate

$$x = A^{-1}b$$

Instead, we utilize something much easier, faster, cheaper, and more importantly, informational.