matrix basics

- An m-by-n *matrix* A is a rectangular array of real numbers, with m rows and n columns.
- We use a capital unbolded letter for a matrix.
- Each *entry* (also called *element*) in a matrix is addressed using two indices. Such an entry is denoted as a_{ij}.
  - The first index determines the row and the second index determines the column. The upper left entry is a_{11}.
- We say that the *dimension* of such an A is m-by-n.

example

- here is a 2-by-3 matrix

$$A := \begin{bmatrix} 6 & 12 & -9 \\ 2 & 0 & 77 \end{bmatrix}$$

- And $a_{13} = -9$.

row of vectors form

- We may also write an m-by-n matrix in the form a row of n vectors from $\mathbb{R}^m$, as in:

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$$

  where, for example,

$$\mathbf{a}_1 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$$

more words

- If a matrix has the same number of rows as columns, it is called *square*. A non-square matrix may be tall and skinny, or short and wide.
- The *zero* matrix, called 0 has all zero entries.
- The entries with indices of the form a_{ii} are called the *diagonal* entries.
- A square matrix, with all non-diagonal entries set to zero is called a *diagonal* matrix.
- A n-by-n diagonal matrix, will all of its diagonal entries equal to 1 is called an *identity* matrix, and denoted by I_n or just I.

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

vetor-like operations

- We can add two m-by-n matrices A and B (same m and n) by adding the corresponding entries. This is denoted as $A + B$.
- We can multiply a matrix by a real scalar value r, by multiplying each of its entries by r. This is denoted as rA.
- With these operations, the set of matrices of a fixed dimension, satisfy all of the vector-space properties, so we can do regular calculations on matrices.

Matrix-Vector Multiplication

- The following expression describes the vector $\mathbf{b}$ in $\mathbb{R}^m$ as a linear combination of n vectors $\mathbf{a}_i$ in $\mathbb{R}^m$, with n real coefficients c_i.

$$\mathbf{b} := c_1\mathbf{a}_1 + ...c_n\mathbf{a}_n = \sum_i c_i\mathbf{a}_i$$

- picture form and two-finger rule

$$\mathbf{b} := \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \dots \\ c_n \end{bmatrix} \tag{1}$$

  – our vertical c_i form a single vector $\mathbf{c}$ in $\mathbb{R}^n$, and that our horizontal row of n-vectors in $\mathbb{R}^m$ forms an m-by-n matrix A.

- Thus we can also write this as

$$\mathbf{b} := A\mathbf{c}$$

- the matrix must go on the left, and the vector must go on the right.
- the size of the multiplied vector must match the number of columns of the matrix. The size of the vector output will match the number of rows of the matrix.
- this is called *matrix-vector multiplication*.

entry format

- 2-by-3 example

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \tag{2}$$

- The output vector $\mathbf{b}$ has its entries defined as

$$b_i = \sum_j a_{ij} c_j$$

  – This means that the ith entry $\mathbf{b}$ is obtained by applying the two-finger rule to the ith row of A and the (column) vector $\mathbf{c}$.

forms

- Equation 1 is useful when thinking of $\mathbf{b}$ as a linear combination of the $\mathbf{a}_i$.
- Equation 2 is useful for calculations
- from this first point of view we see that following must be true

  Lemma. *Suppose that $\mathbf{b} = A\mathbf{c}$. Then $\mathbf{b}$ must be in the span of the vectors $\mathbf{a}_i$ that comprise the columns of A.*

- Q: what is $A\mathbf{e}_1$? $A\mathbf{e}_i$?

easy thms

Theorem. *For all vectors $\mathbf{v} \in \mathbb{R}^n$, we have $\mathbf{v} = I_n\mathbf{v}$.*

Theorem. *For all m-by-n matrices A, vectors $\mathbf{v}$ and $\mathbf{u}$ in $\mathbb{R}^n$ and scalars c, we have*

- $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$
- $A(c\mathbf{u}) = c(A\mathbf{u})$
- Taken together, we refer to the two properties of this theorem as *linearity* of matrix-vector multiplication.

Linear Maps

Definition. A *map* or *transform* $\mathcal{A}$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ is an assignment to each (input) $\mathbf{v} \in \mathbb{R}^n$ some (output) vector $\mathbf{w}$ in $\mathbb{R}^m$, which we denote as $\mathcal{A}(\mathbf{v})$. ■

Definition. We say that a map $\mathcal{A}$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ is *linear* if it satisfies the following two properties for all $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ and scalar c.

- $\mathcal{A}(\mathbf{u} + \mathbf{v}) = \mathcal{A}(\mathbf{u}) + \mathcal{A}(\mathbf{v})$

- $\mathcal{A}(c\mathbf{u}) = c(\mathcal{A}(\mathbf{u}))$

■

- A linear map behaves in a somewhat controlled way.
  - For example a linear map must take a subspace V of $\mathbb{R}^n$ to some subspace of $\mathbb{R}^m$ (not necessarily of the same dimension) It must map an affine set of vectors to an affine set.
  - If $\mathbf{u} = 1/2\mathbf{v} + 1/2\mathbf{w}$ is the average, say of two vectors, then the same property must persist after the map.
  - The zero vector must map to the zero vector.
- Examples in 2 or 3 dimensions of linear maps include things like "rotations" and "reflections" about the origin. and "scales" and "shears" It also includes projections (which can lower the dimension of an input subspace).
  - Later on we will learn about things like lengths of vectors and angles between pairs of vectors. These properties need not be preserved under a linear map.

matrices and maps 1

Definition. Let A be an m-by-n matrix. Then A *determines* a transformation $\mathcal{A}$ that maps a vector $\mathbf{u} \in \mathbb{R}^n$ to the vector $A\mathbf{u} \in \mathbb{R}^m$. ■

- Under this transformation, we have

$$\mathcal{A}(\mathbf{e}_i) = A\mathbf{e}_i = \mathbf{a}_i$$

- the following is obvious from above:

  Theorem (mat=map 1)**.** *Let A be an m-by-n matrix. Then the map determined by A is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$.*

matrices and maps 2

- in fact the other way works as well

  Theorem (mat=map 2)**.** *Let $\mathcal{A}$ be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, then there is a unique m-by-n matrix A that determines this map.*

- summary: every matrix gives me a linear map, and every linear map can be described using a matrix.
- The columns $\mathbf{a}_i$ of A must be $\mathcal{A}(\mathbf{e}_i)$
  - so all i need to do is measure how the map acts on the indicator vectors, and then I can write down the map's matrix!

basis map

- I can determine a linear map by measruing how it acts on the indicator basis for $\mathbb{R}^n$.
- "basis map thm": I can determine a unique linear map (and a matrix) by measuring how it acts on any chosen basis of $\mathbb{R}^n$
  - not as easy to compute the matrix.

Rank

- now for lots of terminology:

  Definition. Given an m-by-n matrix A, let us denote its ith column as $\mathbf{a}_i$. This is a vector in $\mathbb{R}^m$. We denote by $Col(A)$ the span of the set $\{\mathbf{a}_i\}$, and call this the *column span* of A. We denote by $Rank(A)$ the dimension of $Col(A)$. ■

- BQ: Prove that $Rank(A) \leq n$.

- BQ: Prove that $Rank(A) \leq m$.

- BQ: How could you easily compute the rank of a diagonal matrix?

surjective characterization

Definition. We say that a linear map determined by an m-by-n matrix A is *surjective* if for all $\mathbf{b} \in \mathbb{R}^m$, there exists an $\mathbf{x} \in \mathbb{R}^n$ so that $\mathbf{b} = A\mathbf{x}$. This is the same as $Col(A) = \mathbb{R}^m$. This is the same as $Rank(A) = m$. ■

- Q: Why does Surjectivity require $m \leq n$.
- Surjectivity means that we can hit any target vector using an appropriate input to the linear map.

null space

Lemma. *Given an m-by-n matrix A. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two vectors in $\mathbb{R}^n$ such that $A\mathbf{x}_1 = A\mathbf{x}_2 = \mathbf{0}$. Then $A(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{0}$.*
Let c be any scalar, and let $\mathbf{x}$ be a vector in $\mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$. Then $A(c\mathbf{x}) = \mathbf{0}$.
The set of vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$ form a subspace.

Proof. The two main claims follow from the linearity Theorem. This gives closedness under addition and scalar multiplication, giving rise to a subspace. □

Definition. Given an m-by-n matrix A. We will denote by $Null(A)$ the subspace of vectors $\mathbf{x} \in \mathbb{R}^n$ so that $A\mathbf{x} = \mathbf{0}$. This is called the *null space* or *kernel* of A. We define $Nullity(A) = Dim(Null(A))$. ■

injective

Definition. We say that a linear map determined by an m-by-n matrix A is *injective* if for all $\mathbf{x}_1 \neq \mathbf{x}_2$ both in $\mathbb{R}^n$, then we have $A\mathbf{x}_1 \neq A\mathbf{x}_2$. ■

- This means that there are no "collisions" under the map.
- We also say that the map is a one-to-one map.
- Another way to say this is that if there is some $\mathbf{x}$ that hits $\mathbf{b}$ under the map, then this $\mathbf{x}$ is the only such vector.

injectivity and nullity

Lemma. *The linear map determined by an m-by-n matrix A is injective iff $Nullity(A) = 0$.*

- this is an 'iff'. our proof below will actually prove the two contrapositives.

  Proof. Lets start with proving that zero nullity implies injectivity, using contrapositive. Suppose the map is not injective. Then we have two vectors, $\mathbf{x}_1 \neq \mathbf{x}_2$ with $A\mathbf{x}_1 = A\mathbf{x}_2$. Then from linearity (Theorem), we have

$$\begin{aligned} \mathbf{0} &= A\mathbf{x}_1 - A\mathbf{x}_2 \\ &= A(\mathbf{x}_1 - \mathbf{x}_2) \end{aligned}$$

  putting $\mathbf{y} := \mathbf{x}_1 - \mathbf{x}_2 \neq \mathbf{0}$ in $Null(A)$. Then $Nullity(A) > 0$.

  For the converse lets prove that injectivity implies zero nullity, using contrapositive. Conversely, if $Nullity(A) > 0$, then there must exist some non-zero vector with $A\mathbf{y} = \mathbf{0}$. But by we also must have $A\mathbf{0} = \mathbf{0}$ giving us a collision and violating injectivity. □

injectivity and linInd

Lemma (linInd)**.** *The linear map determined by an m-by-n matrix A is injective iff iff the columns of A are linearly independent.*

- "iff" means if and only if. this means that we have an implication as well as its converse.

  Proof. A null vector $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{0}$ which means $\sum_i x_i \mathbf{a}_i = 0$. By definition, linear independence means that the only satisfying x_i weights must all be zeros. Since $\mathbf{0}$ is the only null vector, this is injectivity from the prev lemma. □

- BQ: Why does $Nullity(A) = 0$ require $m \geq n$?

bijectivity

- The nicest thing we could have is both surjectivity and injectivity. Clearly this requires $n = m$.

Definition. If the linear map determined by an n-by-n matrix A is both surjective and injective, then we say it is *bijective*. This means that $Rank(A) = n$ and $Nullity(A) = 0$. ■

- Such a map has the following property: for any $\mathbf{b} \in \mathbb{R}^n$ there exists a unique $\mathbf{x} \in \mathbb{R}^n$ with $A\mathbf{x} = \mathbf{b}$.

collapse

- Indeed we can now show that in the square case, lots of things collapse

  Theorem (Collapse)**.** *Let A be a square n-by-n matrix. Then its map is surjective iff it is injective.*

- you only need to check for one property and you get the other for free.

  Proof. Injective iff its columns are lin. ind.

  n vectors in $\mathbb{R}^n$, an n dimensional space are lin ind iff they span (Basis theorem).

  Span iff $Rank(A) = n$ iff surjective. □

Linear Systems

- Consider the following system of $m = 2$ equations with $n = 3$ real unknowns x_1, x_2, x_3.

$$\begin{aligned} 1x_1 - 2x_2 + 1x_3 &= 0 \\ 0x_1 + 2x_2 - 8x_3 &= 8 \end{aligned}$$

- A novel but equivalent way we could write this is

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -8 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}$$

- in this new view of a linear system, we are trying to find x_i so that express $[0; 8]$ as a linear combination of the columns of A.

...

- This can also be expressed as

$$A\mathbf{x} = \mathbf{b}$$

- where A is the 2-by-3 matrix

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \end{bmatrix},$$

  $\mathbf{b} \in \mathbb{R}^2$ is

$$\begin{bmatrix} 0 \\ 8 \end{bmatrix},$$

- and $\mathbf{x}$ is a variable vector in $\mathbb{R}^3$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Solvability

- When is a linear system *solvable*?

  Lemma. *A linear system $A\mathbf{x} = \mathbf{b}$ has a solution iff $\mathbf{b} \in Col(A)$.*

  Proof. From the definition of matrix-vector multiply, a solution to $A\mathbf{x} = \mathbf{b}$ is equivalent to the existance of a solution to the equation $\sum_i x_i \mathbf{a}_i = \mathbf{b}$, where the $\mathbf{a}_i$ are the columns of A. This is equivalent to saying that that $\mathbf{b}$ is a linear combination of the $\mathbf{a}_i$. This is equivalent to saying that $\mathbf{b}$ in $Span(\{\mathbf{a}_i\}) = Col(A)$. □

- proof notes: since each step is an equivalence, we have proven an iff statement.

always solvable

- Some matrices A have the nice property that a linear system $A\mathbf{x} = \mathbf{b}$ are solvable for any chosen $\mathbf{b}$.

  Definition. We say that an m-by-n matrix A is *always solvable* if for all $\mathbf{b} \in \mathbb{R}^m$, the system $A\mathbf{x} = \mathbf{b}$ has a solution. ■

alwasy solve and surjective

Theorem. *Let A be an m-by-n matrix. A, is always solvable iff the map determined by A is surjective.*

- just follows from the defintions.

homogeneous

Definition. A linear system of the form $A\mathbf{x} = \mathbf{0}$ is called *homogeneous*. By definition, the solutions $\mathbf{x}$ to such a system must form the null space of A. ■

- So we see that the solution set to a homogeneous equation is a subspace of $\mathbb{R}^n$! The dimension of this space is $Nullity(A)$.
- A homogeneous system can be always solved using by setting $\mathbf{x} := \mathbf{0}$, which is always a null vector.

inhomogeneous

Definition. A linear system of the form $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b} \neq \mathbf{0}$ is called *inhomogeneous*. ■

- no guarantee that an inhomogeneous linear system is solvable
- , but suppose that we do assume that it is solvable. So we are assuming the existence of a vector $\mathbf{x}_1 \in \mathbb{R}^n$ with $A\mathbf{x}_1 = \mathbf{b}$.

solutions

Theorem. *Suppose that $\mathbf{x}_1$ and $\mathbf{x}_2$ are both solutions to the (homogeneous or inhomogeneous) system $A\mathbf{x} = \mathbf{b}$. Then $\mathbf{k} := \mathbf{x}_1 - \mathbf{x}_2$ is in $Null(A)$, and thus a solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$. In different words $\mathbf{x}_2 = \mathbf{x}_1 - \mathbf{k}$, with $\mathbf{k} \in Null(A)$.*
Conversely, suppose $\mathbf{x}_1$ is a solution to the (homogeneous or inhomogeneous) system $A\mathbf{x} = \mathbf{b}$, and $\mathbf{k}$ is a solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then $\mathbf{x}_2 := \mathbf{x}_1 + \mathbf{k}$ is also a solution to $A\mathbf{x} = \mathbf{b}$.

proof

Proof. For the first statement: By assumption, $A\mathbf{x}_1 = \mathbf{b} = A\mathbf{x}_2$. So

$$\begin{aligned} \mathbf{0} &= A\mathbf{x}_1 - A\mathbf{x}_2 \\ &= A(\mathbf{x}_1 - \mathbf{x}_2) \\ &=: A\mathbf{k} \end{aligned}$$

Here we used linearity
For the second statement:

$$\begin{aligned} A(\mathbf{x}_1 + \mathbf{k}) &= A\mathbf{x}_1 + A\mathbf{k} \\ &= \mathbf{b} + \mathbf{0} \\ &= \mathbf{b} \end{aligned}$$

again, using linearity. This makes $\mathbf{x}_2 := \mathbf{x}_1 + \mathbf{k}$ a solution to $A\mathbf{x} = \mathbf{b}$. □

affinity

- This theorem tells us the shape of the solution set to a solvable linear system:

  Corollary (solution dimension)**.** *Suppose the (homogeneous or inhomogeneous) system $A\mathbf{x} = \mathbf{b}$ is solvable. Then its set of solutions form an affine set of dimension $Nullity(A)$.*

  Proof. The theorem tells us that the solution set is exactly the vectors of the form $\{\mathbf{x}_1 + \mathbf{k} | \mathbf{k} \in Null(A)\}$, where $\mathbf{x}_1$ is any solution. This is an affine set of dimension $Nullity(A)$. □

- This corollary tells us that once we have fixed A, as we vary the rhs $\mathbf{b}$, either we get no solutions, or we get an affine solution set of some unchanging dimension!

Uniqueness

Definition. Suppose that a linear system $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}_1$. We say that the system's solution is *unique* if it has no other solution.
We say that the matrix A is *never ambiguous* if for any $\mathbf{b}$, the system $A\mathbf{x} = \mathbf{b}$ has no solution, or has a unique solution. ■

uniqueness and injectivity

Theorem. *Let A be an m-by-n matrix. A is never ambiguous iff the map determined by A is injective.*

Proof. From the corollary, never ambiguity is the same as zero-nullity. From Lemma, this is the same as injectivity. □

summary

Theorem. *Let A be an m-by-n matrix. The following properties are equivalent:*

1. *The map determined by A is injective.*
2. *$Nullity(A) = 0$.*
3. *The columns, $\mathbf{a}_i$, form a linearly independent set.*
4. *A is never ambiguous.*

Additionally, the following properties are equivalent:

1. *The map determined by A is surjective.*
2. *$Rank(A) = m$.*
3. *The columns, $\mathbf{a}_i$, span $\mathbb{R}^m$.*
4. *A is always solvable.*

When A is square, and so $m = n$, all 8 of these properties are equivlalent.

Rank-Nullity

- Q: given a set of n linearly independent vectors in $\mathbb{R}^m$ what can we say about the dimension of their span?
- We know that for an m-by-n matrix A: $Nullity(A) = 0 \Leftrightarrow Rank(A) = n$.
- With some work, this idea can be beefed up significantly in to the following non-obvious theorem.

  Theorem (Rank-nullity)**.** *Let A be an m-by-n matrix. Then*

  $$Rank(A) = n - Nullity(A)$$

RN intuition

- let us just consider matrices of the following form

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- what is the column space? what is the null space?
  - note that the column space is in $\mathbb{R}^m$, while the null space is in $\mathbb{R}^n$.
- what is the rank? what is the nullity?

- even thought this is a very special form, every linear map is somewhat like this with some extra pre/post- linear "deformation" on the domain and range that does not effect these counts.

RN intuition 2

- we saw from linear systems that fixing $\mathbf{x}_1$, then for all $\mathbf{k} \in Null(A)$ we have $A\mathbf{x}_1 = A(\mathbf{x}_1 + \mathbf{k})$.
- so in our current context, we start in $\mathbb{R}^n$ with n-dimensions of input to A.
- during the map under A, we collide together $nullity(A)$ dimensions of stuff together.
- this leaves $n - nullity(A)$ dimensions of outputs. Meanwhile, $rank(A)$ represents the dimensions of possible outputs
- see deck of cards gadget.

RN intuition 3

- in the end, some number of the columnts of A can be used to build up the column span
- the rest $n - Rank$ of them are redundant.
- that redundancy quantity is captured by the nullity.

hyperplane

Corollary. *Let A be a non-zero matrix with exactly one row. Then for any $\mathbf{b} \in \mathbb{R}^1$ the solution set to the equation $A\mathbf{x} = \mathbf{b}$ is a hyperplane in $\mathbb{R}^n$.*

Proof. As long as A is non-zero, its columns must span $\mathbb{R}^1$, making making A always solvable. From Rank-Nullity, we have $Null(A) = n - Rank(A) = n - 1$. This gives us the solution set dimension. $\square$

MATLAB

- Let us look at how these ideas play out in MATLAB.

```
b = [1; -5; 4; 1; 7];
A = [2, -7, 2, -5, 8; 0, 1, -3, -3, -1; 0, 0, 0, 1,- 1];
C=colspace(A);
r=size(C,2);
N=null(A);
z=size(N,2);

q=M*b;
```

- The command `colspace(A)` returns a matrix, whose columns form a basis for $Col(A)$.
- The command `size(C,2)` returns the number of columns in `C` which here is the same as $Rank(A)$.
- The command `null(A)` returns a matrix, whose columns form a basis for $Null(A)$.
- The command `size(N,2)` returns the number of columns in `N` which here is the same as $Nullity(A)$.
- The command `A*b` implements matrix-vector multiplication.