Princeton University MAT 202 Spring 2008

Willie W. Wong

Feb 18 - 22

FEB 18, 2008

- Topic of today: Matrix multiplication.
- Remember from last week that, for all intents and purposes, a linear transformation of vectors is the same as multiplication of vectors by a matrix. In other words, we have $T(\boldsymbol{x}) = A\boldsymbol{x}$.
- Now suppose we have this picture

  $$\mathbb{R}^M \xrightarrow{T} \mathbb{R}^N \xrightarrow{S} \mathbb{R}^P$$

  and we want to define the composition of the transformation. E.g., we want to define the transformation $U : \mathbb{R}^M \to \mathbb{R}^P$ such that $U(\boldsymbol{x}) = S(T(\boldsymbol{x}))$. We write $U = S \circ T$, read "U is S composed with T".
- Now, suppose T has a corresponding $N \times M$ matrix A and S has a corresponding $P \times N$ matrix B. The operation

  $$S(T(\boldsymbol{x})) = S(A\boldsymbol{x}) = B(A\boldsymbol{x})$$

  On the other hand, U is obviously a linear transformation (it takes lines to lines and origin to origin). So U has a corresponding $P \times M$ matrix C such that

  $$U(\boldsymbol{x}) = C\boldsymbol{x}$$

  Comparing the two expressions, we want to establish some sort of associativity rule so that

  $$B(A\boldsymbol{x}) = (BA)\boldsymbol{x} = C\boldsymbol{x}$$

  and so we want to define the product of the matrices B and A.
- This leads to our first fact: motivated by our definition of linear transformations. The product matrix BA is only defined if the number of columns (the width) of B is the same as the number of rows (the height) of A. Pictorially

  $$P \times M = P \times \underbrace{Q \cdot N}_{\text{The same!}} \times M$$

The number of rows of BA is the same as the number of rows of B, and the number of columns of BA is the same as the number of columns of A, should BA be defined.

- How do we calculate the matrix $C = BA$? Remember the rules of linearity and rules for multiplication of a vector by a matrix: suppose A can be decomposed a column vectors

$$\begin{aligned} A &= \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_M \end{bmatrix} \\ A\boldsymbol{x} &= x_1\boldsymbol{v}_1 + \cdots + x_M\boldsymbol{v}_M \\ S(T(\boldsymbol{x})) = S(A\boldsymbol{x}) &= x_1 S(\boldsymbol{v}_1) + \cdots x_M S(\boldsymbol{v}_M) \\ U(\boldsymbol{x}) &= x_1 B\boldsymbol{v}_1 + \cdots x_M B\boldsymbol{v}_M \\ C &= \begin{bmatrix} B\boldsymbol{v}_1 & \dots & B\boldsymbol{v}_M \end{bmatrix} \end{aligned}$$

- Now, we use the row-vector decomposition of B

$$B = \begin{bmatrix} \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_P \end{bmatrix}$$

Then the vector

$$B\boldsymbol{v}_i = \begin{bmatrix} \boldsymbol{w}_1 \cdot \boldsymbol{v}_i \\ \vdots \\ \boldsymbol{w}_P \cdot \boldsymbol{v}_i \end{bmatrix}$$

So we finally have the rule that the entry of $C = BA$ in the ith row and jth column is given by

$$c_{ij} = \boldsymbol{w}_i \cdot \boldsymbol{v}_j$$

- In other words, let A B C be as above

$$c_{ij} = b_{i1}a_{1j} + b_{i2}a_{2j} + \cdots + b_{iN}a_{Nj}$$

- Properties of matrix multiplication
  - NONcommutative: remember our example of shear-then-rotate vs. rotate-then-shear. Applying linear transformation in succession is, by definition, composing those linear transformations, and thus corresponds to multiplication, in different order, of the matrices. In general,

$$AB \neq BA$$

where A and B are matrices. (Notice that AB and BA don't even need to have the same size: A can be an $N \times M$ matrix and B can be a $M \times N$ matrix. AB is then $N \times N$ and BA is then $M \times M$.)

– Example:

$$\begin{bmatrix}1 & 2 & 3\\ 2 & -4 & 5\\ 0 & 2 & 1\end{bmatrix}\begin{bmatrix}-3 & 5 & 4\\ 2 & 0 & 1\\ 1 & 3 & 5\end{bmatrix} = \begin{bmatrix}4 & 14 & 21\\ -9 & 25 & 29\\ 5 & 3 & 7\end{bmatrix}$$
$$\begin{bmatrix}-3 & 5 & 4\\ 2 & 0 & 1\\ 1 & 3 & 5\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\ 2 & -4 & 5\\ 0 & 2 & 1\end{bmatrix} = \begin{bmatrix}7 & -18 & 20\\ 2 & 6 & 7\\ 7 & 0 & 23\end{bmatrix}$$

– But: multiplication by scalars is commutative:

$$(kA)B = A(kB) = k(AB)$$

where k is a scalar and A B are matrices.

– Associative:

$$(AB)C = A(BC)$$

if a string of matrices are multiplied together in succession, it doesn't matter which two you multiply first. You can see this by using the definition using linear transformations: no matter how you group the terms, the product

$$[(AB)C]\boldsymbol{x} = A(B(C(\boldsymbol{x}))) = [A(BC)]\boldsymbol{x}$$

– Distributive:

$$A(C + D) = AC + AD$$

and

$$(A + B)C = AC + BC$$

provided C and D are matrices of the same size, and A and B are matrices of the same size.

– Multiplication by identity: let A be $N \times M$, then

$$AI_M = I_N A = A$$

again, obvious using the interpretation of linear transformations where the Identity matrix corresponds to the identity transformation where nothing is changed.

– Multiplication by inverse (square matrices):

$$AA^{-1} = A^{-1}A = I_M$$

for square $M \times M$ matrix A that is invertible. This follows from the definition of the inverse matrix.

- Inverse of a product of matrix:

$$(AB)^{-1} = B^{-1}A^{-1}$$

where A, B are both square invertible matrices. Important that *the order of the product is reversed when you take the inverse*. Check:

$$\begin{aligned}(AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} \\ &= AI_M A^{-1} \\ &= AA^{-1} = I_M \\ (B^{-1}A^{-1})(AB) &= B^{-1}(A^{-1}A)B \\ &= B^{-1}I_M B \\ &= B^{-1}B = I_M\end{aligned}$$

where the first step we used the associativity, the second step multiplication by inverse, third step multiplication by identity.

- A intuitive reason for the reversal of order, again, let $S : \mathbb{R}^M \to \mathbb{R}^M$ be invertible with the matrix B, and $T : \mathbb{R}^M \to \mathbb{R}^M$ be invertible with the matrix A. $S \circ T$ is

$$\mathbb{R}^M \xrightarrow[A]{T} \mathbb{R}^M \xrightarrow[B]{S} \mathbb{R}^M$$

The inverse of this operation goes backwards

$$\mathbb{R}^M \xleftarrow[A^{-1}]{T^{-1}} \mathbb{R}^M \xleftarrow[B^{-1}]{S^{-1}} \mathbb{R}^M$$

So on one way, you hit A first, then B, on the reverse, you hit B^{-1} first, then A^{-1}.

- A important fact about inverses: let A and B be $M \times M$ matrices. Suppose

$$BA = I_M$$

Then we can conclude:

  * A, B both invertible
  * $A^{-1} = B$ and $B^{-1} = A$
  * $AB = I_M$

• Partitioned matrices: we can partition a big matrix into smaller parts for ease of calculation. For example, we can partition the following 5×5 matrix A into 4 parts, A_{11} which is 3×3, A_{22} which is 2×2, A_{12} which is 3×2 and A_{21} which is 2×3

$$A = \begin{bmatrix} 1 & 5 & 2 & \vdots & 3 & 4 \\ 2 & 3 & 1 & \vdots & 5 & 8 \\ 1 & 3 & 4 & \vdots & 2 & 5 \\ \cdots & \cdots & \cdots & & \cdots & \cdots \\ 5 & 7 & 1 & \vdots & 9 & 1 \\ 2 & 2 & 1 & \vdots & 3 & 5 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

with

$$A_{11} = \begin{bmatrix} 1 & 5 & 2 \\ 2 & 3 & 1 \\ 1 & 3 & 4 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 3 & 4 \\ 5 & 8 \\ 2 & 5 \end{bmatrix}, \quad A_{21} = \begin{bmatrix} 5 & 7 & 1 \\ 2 & 2 & 1 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 9 & 1 \\ 3 & 5 \end{bmatrix}$$

- What is the point of partitioning a matrix? Well, the multiplication formulat of matrices can be generalized to the blocks! Let A be an $N \times M$ matrix and B a $P \times N$ matrix. And suppose A is partitioned into n blocks on the row and m blocks on the column, i.e.

$$A = \begin{bmatrix} A_{11} & \dots & A_{1m} \\ \vdots & & \vdots \\ A_{n1} & \dots & A_{nm} \end{bmatrix}$$

and similarly for B into p blocks on the row and n blocks on the column (we are trying to multiply BA, so we need the number of divisions of columns of B be the same as the number of divisions as rows of A). Furthermore, we also require that the number of columns in the jth division (e.g., the width of the B_{1j} submatrix) of B be the same as the number of rows in the jth division (e.g., the height of the A_{j1} submatrix) for every j. (This means that the matrix product $B_{ij}A_{jk}$ can be defined for any $1 \leq i \leq p$, $1 \leq j \leq n$, $1 \leq k \leq m$.) Then let $C = BA$, the $P \times M$ matrix. Then C has the partition

$$C = \begin{bmatrix} C_{11} & \dots & C_{1m} \\ \vdots & & \vdots \\ C_{p1} & \dots & C_{pm} \end{bmatrix}$$

with

$$C_{ik} = B_{i1}A_{1k} + B_{i2}A_{2k} + \cdots + B_{in}A_{nk}$$

- This process seems complicated, but it makes life simpler when your matrix has a lot of zeroes. For example, suppose we want to invert the following matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

supposing that A_{11} is $N \times N$, A_{22} is $M \times M$, A_{12} is $N \times M$, and 0 denote the $M \times N$ matrix with entries all zero. Remember our criterion for inverse: we want to find a square matrix B such that

$$BA = I_{N+M} = \begin{bmatrix} I_N & 0 \\ 0 & I_M \end{bmatrix}$$

Then our multiplication process above shows that we need to decompose B in the same way as A, and write down

$$B_{11}A_{11} = I_N, B_{21}A_{11} = 0, B_{11}A_{12} + B_{12}A_{22} = 0, B_{21}A_{12} + B_{22}A_{22} = I_M$$

Now suppose A_{11}, A_{22} are invertible, then $B_{11} = A_{11}^{-1}$. And $B_{21} = 0$. So $B_{22} = A_{22}^{-1}$, and $B_{12} = -A_{11}^{-1}A_{12}A_{22}^{-1}$

- Typically, the computational difficulty for finding inverses increases very fast as the size of the matrix gets bigger. So being able to break down the problem into chuncks makes calculation a lot easier.

FEB 20, 2008

- Some announcements: 1) Remember that you have a quiz on chapters 1 and 2 (everything up to what I covered on Monday) next Wednesday 2) Your midterms have been scheduled to be Wednesday, March 12th, 7:30pm - 9pm. Please let me know by e-mail if you have conflicts so we can schedule a make up exam.
- Given a transformation $T : X \to Y$, remember that we say $y \in Y$ is an image of $x \in X$ if $T(x) = y$. We can generalize this concept to the entire domain. We write
$$im(T)$$
for the set given by all $y \in Y$ such that $T(x) = y$ for some $x \in X$. (Draw picture)
- Example: let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the projection operation $proj_L$ onto some line L. Then $im(T)$ is the entire line L. Suppose we limit the domain of T to some subset of T (draw a figure), the image will be the "shadow" casted by the subset on the line.
- Example: let $T : \mathbb{R} \to \mathbb{R}^2$ be given by
$$T(x) = \begin{bmatrix} \cos(x) \\ \sin(x) \end{bmatrix}$$
The image of T is the circle. We call T a *parametrization* of the circle. In general, a parametrization of a curve C in $\mathbb{R}^N$ is the transformation $T : \mathbb{R} \to \mathbb{R}^N$ such that $im(T) = C$.
- Example: let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the transformation
$$T(\boldsymbol{x}) = \begin{bmatrix} 1 & 5 & 3 \\ 0 & 4 & 4 \\ 2 & 2 & -2 \end{bmatrix} \boldsymbol{x}$$
We write out the transformation
$$T(\boldsymbol{x}) = x_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 2+3 \\ 0+4 \\ 4+-2 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix} = (x_1 + 2x_2) \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + (x_2 + x_3) \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix}$$
So the image is all the vectors in $\mathbb{R}^3$ of the form $a \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + b \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix}$, which is a 2-dimensional plane in space. Notice that
$$rref(\begin{bmatrix} 1 & 5 & 3 \\ 0 & 4 & 4 \\ 2 & 2 & -2 \end{bmatrix}) = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

and so the $rank(T) = 2$. For a linear transformation, the rank of its corresponding matrix will be the number of dimensions of the image.

- Remember that a linear transformation can be thought of as a linear combination of the column vectors of the corresponding matrix, i.e., suppose

$$T(\boldsymbol{x}) = A\boldsymbol{x} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_M \end{bmatrix} \boldsymbol{x} = x_1\boldsymbol{v}_1 + \cdots x_M\boldsymbol{v}_M$$

  the image of a linear transformation is the set of all vectors that can be written as a linear combination of the column vectors $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_M\}$. We call this set the *span* of the vectors. We write

$$\begin{aligned} span\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_M\} &= \{ \text{ all linear combinations of the vectors } \} \\ &= \{c_1\boldsymbol{v}_1 + \cdots + c_M\boldsymbol{v}_M \text{ where } c_1 \dots c_M \in \mathbb{R}\} \end{aligned}$$

- Properties of the image of a linear transformation
  - The 0 vector is in the image. (Let $c_1 = c_2 = \cdots = c_M = 0$.)
  - $im(T)$ is closed under addition. I.e., if $\boldsymbol{w}_1, \boldsymbol{w}_2 \in im(T)$, then $\boldsymbol{w}_1 + \boldsymbol{w}_2$ is also in $im(T)$.
  - $im(T)$ is closed under multiplication by scalars. I.e., if $\boldsymbol{w} \in im(T)$, then $k\boldsymbol{w} \in im(T)$ for any number k as well.

  The second and third properties follows from the linearity of the transformation.

- A related concept to the image is the *kernel*. Remember that the criterion for whether a linear transformation (given by a square matrix) is invertible is whether there exists a non-zero $\boldsymbol{x}$ such that $T(\boldsymbol{x}) = 0$.

- The *kernel* of a linear transformation $T : X \to Y$ is $ker(T)$ which is defined as the set of all $\boldsymbol{x} \in X$ such that $T(\boldsymbol{x}) = 0$. In other words, the kernel is the solution set of

$$T(\boldsymbol{x}) = 0$$

- The kernel is sometimes called the *null space*.

- For a linear transformation, the image is a subset of the codomain, and the kernel is a subset of the domain.

- Example: find the kernel of the transformation given by

$$A = \begin{bmatrix} 2 & 5 & 1 & 3 & 4 \\ 1 & 2 & 0 & 6 & 2 \\ 0 & 1 & 1 & 4 & 0 \\ 1 & 3 & 1 & 5 & 2 \end{bmatrix}$$

We want to solve $A\boldsymbol{x} = 0$. So we perform G-J elimination to find the solution

$$\left[\begin{array}{ccccc:c} 2 & 5 & 1 & 3 & 4 & 0 \\ 1 & 2 & 0 & 6 & 2 & 0 \\ 0 & 1 & 1 & 4 & 0 & 0 \\ 1 & 3 & 1 & 5 & 2 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{ccccc:c} 1 & 5/2 & 1/2 & 3/2 & 2 & 0 \\ 0 & -1/2 & -1/2 & 9/2 & 0 & 0 \\ 0 & 1 & 1 & 4 & 0 & 0 \\ 0 & 1/2 & 1/2 & 7/2 & 0 & 0 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccccc:c} 1 & 0 & -2 & 24 & 2 & 0 \\ 0 & 1 & 1 & -9 & 0 & 0 \\ 0 & 0 & 0 & 13 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{ccccc:c} 1 & 0 & -2 & 0 & 2 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

So the solution to the problem is

$$\begin{aligned} x_1 &= 2x_3 - 2x_5 \\ x_2 &= -x_3 \\ x_4 &= 0 \end{aligned}$$

In terms of the free parameters $x_3 = s, x_5 = t$, we write

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = s \begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

So the kernel of the transformation is given by a two-parameter family. Using the notation given prior, we say that the kernel is the *span* of the vectors $\begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

- Using the linearity of the transformation, we have the same three properties for the kernel set
  - The 0 vector is in $ker(T)$
  - $ker(T)$ is closed under addition.
  - $ker(T)$ is closed under multiplication by a scalar.
- Fact: a square matrix A is invertible if and only if $ker(A) = \{0\}$, the set containing only the 0 vector and nothing else.

- When, in general, is the $ker(A) = 0$? Notice first that the equation

$$A\boldsymbol{x} = 0$$

has at least 1 solution: the 0 on the RHS means the equation cannot be inconsistent. For the solution to be unique, it cannot have free parameters. Remember from chapter 1 that the requirement is the same as saying $rank(A) = M$ for an $N \times M$ matrix A. In particular, we must have that $N \geq M$ for this to be true…

- Given A an $N \times N$ matrix, and T its linear transformation. The following are equivalent (that is, either all true or all false)
  - A is invertible
  - $T(\boldsymbol{x}) = \boldsymbol{b}$ has unique solution for all $\boldsymbol{b} \in \mathbb{R}^M$
  - rref(A) $= I_N$
  - rank(A) $= N$
  - $im(T) = \mathbb{R}^N$
  - $ker(T) = \{0\}$

- Notice that the image and kernel of a linear transformation satisfies some of the same properties. This motivates us to define a category in which both of them resides.

- A subset W of $\mathbb{R}^N$ is called a *subspace* if it 1) contains the 0 vector 2) is closed under addition and 3) is closed under multiplication by a scalar.

- A subspace is the generalization of the concept of lines: if two points are in the subspace W, the line connecting them must also be in the subspace W. For this reason a subspace is also called a *linear subspace*.

- So letting $T : \mathbb{R}^M \to \mathbb{R}^N$ be a linear transformation, we have that $im(T)$ is a subspace of $\mathbb{R}^N$ and $ker(T)$ is a subspace of $\mathbb{R}^M$.

- What are all the subspaces of $\mathbb{R}^2$? Ans: $\mathbb{R}^2$ itself, the $\{0\}$ set, and all the lines through the origin.

- Given a set of some vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ in $\mathbb{R}^N$, its span is a subspace.

FEB 22, 2008

- Let's go back a revisit an example from Wednesday

$$T(\boldsymbol{x}) = \begin{bmatrix} 1 & 5 & 3 \\ 0 & 4 & 4 \\ 2 & 2 & -2 \end{bmatrix} \boldsymbol{x}$$

On Wednesday we showed that the image of this linear transformation can be spanned by the column vectors

$$\begin{bmatrix}1\\0\\2\end{bmatrix}, \quad \begin{bmatrix}3\\4\\-2\end{bmatrix}$$

(this is from the fact that

$$\begin{bmatrix}5\\4\\2\end{bmatrix} = 2\begin{bmatrix}1\\0\\2\end{bmatrix} + \begin{bmatrix}3\\4\\-2\end{bmatrix}$$

the middle column is a linear combination of the left and right columns.)

- On the other hand, the image of the linear transformation is obviously spanned by all three of the vectors (this is the definition of images of linear transformation after all). So the natural question to ask is, "what is the least number of vectors needed to span a given subspace?"

- Some definitions
  - Given a list of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$. We say that a vector $\boldsymbol{v}_i$ in the list is *redundant* if $\boldsymbol{v}_i$ is a linear combination of the ones preceding it, i.e., there exists numbers $c_1, \ldots, c_{i-1}$ such that

    $$\boldsymbol{v}_i = c_1\boldsymbol{v}_1 + \cdots + c_{i-1}\boldsymbol{v}_{i-1}$$

    The first vector in the list, $\boldsymbol{v}_1$ is only redundant when it is 0. (The term *redundant* is not in the linear algebra canon. Depending on textbook it might be defined differently, or even not used at all. Here I follow Bretscher in the definition.)
  - The list $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ is called *linearly independent* if none of them is redundant. The list is called *linearly dependent* if at least one of them is redundant.
  - We say that the list $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ form a *basis* of a subspace V of $\mathbb{R}^N$ if the following three conditions hold
    * Every vector in the list is an element of V
    * The list spans V
    * The list is linearly independent

- Example: let

  $$\boldsymbol{v}_1 = \begin{bmatrix}1\\0\\2\end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix}5\\4\\2\end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix}3\\4\\-2\end{bmatrix}$$

  we ask whether the list is linearly independent? Well, $\boldsymbol{v}_1$ is not the 0 vector, so it is not redundant. $\boldsymbol{V}_2$ cannot be written as a multiple of $\boldsymbol{v}_1$, so it is not redundant. But since $\boldsymbol{v}_3 = \boldsymbol{v}_2 - 2\boldsymbol{v}_1$, we have that $\boldsymbol{v}_3$ is redundant. So the set is linearly dependent.

- Coming from this example and the definition, we arrive at the following fact: starting from a list of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$, let $V = span\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M\}$. Then if we take the list and remove all the redundant vectors, what we have left will form a basis of V.

- Another example: Consider the column vectors of the matrix

$$\begin{bmatrix} 1 & 5 & 3 & 2 \\ 0 & 4 & 0 & 3 \\ 3 & 2 & 1 & 4 \\ 0 & 0 & 0 & 7 \\ 2 & 1 & 3 & 2 \\ 0 & 0 & 1 & 9 \end{bmatrix}$$

  Are they linearly independent? We start with $\boldsymbol{v}_1$, it is not zero, so it is not redundant. Looking at $\boldsymbol{v}_2$, we see immediately that it is not a multiple of $\boldsymbol{v}_1$. This is from the second entry in the two vectors: the second entry of $\boldsymbol{v}_1$ is 0 and the entry of $\boldsymbol{v}_2$ is 4, so there are no multiples of $\boldsymbol{v}_1$ that can give $\boldsymbol{v}_2$. For $\boldsymbol{v}_3$, we can look at the last entry. Since the entry is 0 for both $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, no combination of the two vectors can get the entry 1 for $\boldsymbol{v}_3$, and so $\boldsymbol{v}_3$ is not redundant. Lastly, look at the fourth entry of the vectors and using the same argument we arrive at the result that $\boldsymbol{v}_4$ is also not redundant. So the 4 column vectors are linearly independent.

- Let's look at the process of deciding whether a vector $\boldsymbol{v}_i$ out of a list is redundant or not. What we are looking for is numbers $c_1, \dots, c_{i-1}$ such that $c_1\boldsymbol{v}_1 + \cdots = \boldsymbol{v}_i$. This is a linear system of equations. Its augmented matrix is

$$\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \dots & \boldsymbol{v}_{i-1} & \vdots & \boldsymbol{v}_i \end{bmatrix}$$

  If this system is inconsistent, then $\boldsymbol{v}_i$ cannot be written as a linear combination of the preceding vectors, and so is not redundant. If this system is consistent (meaning that it possesses at least 1 solution), then $\boldsymbol{v}_i$ is redundant.

- Instead of writing

$$c_1\boldsymbol{v}_1 + \cdots c_{i-1}\boldsymbol{v}_{i-1} = \boldsymbol{v}_i$$

  we like to write instead

$$c_1\boldsymbol{v}_1 + \cdots + c_i\boldsymbol{v}_i = 0$$

  (just a notational issue). A linear equation of the form above, i.e.

$$(\text{ something about some vectors }) = 0$$

  is called a *linear relation* of those vectors. In other words, we want to look at, instead of the augmented matrix

$$\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \dots & \boldsymbol{v}_{i-1} & \vdots & \boldsymbol{v}_i \end{bmatrix}$$

  the matrix

$$\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \dots & \boldsymbol{v}_i & \vdots & 0 \end{bmatrix}$$

  This second formulation always admits a *trivial* solution $c_1 = \cdots = c_i = 0$.

- A list of vectors is linearly independent if and only if the only linear relation among them is the trivial one. Since if there exists a non-trivial linear relation, for the relation, we can find the largest j such that $c_j \neq 0$. Then we can solve for $\boldsymbol{v}_j$ in terms of its preceding vectors, and thus $\boldsymbol{v}_j$ is dependent.

- Now, let's go back to a definition we gave on Wednesday. Remember that the kernel of a linear transformation (given by the matrix A) is the set of $\boldsymbol{x}$ such that

$$A\boldsymbol{x} = 0$$

  By writing A as column vectors, we see that this the equation is the same as giving a linear relation

$$x_1\boldsymbol{v}_1 + \cdots + x_M\boldsymbol{v}_M = 0$$

  among the column vectors. Therefore, we have that the following three are equivalent for an $N \times M$ matrix A:

  - the column vectors of A are linear independent
  - $ker(A) = \{0\}$
  - $rank(A) = M$ (which implies that $M \leq N$).

- The last condition (more precisely, its corollary that $M \leq N$) implies that *there are at most N linear independent vectors in $\mathbb{R}^N$*. (Why?)

- A summary of all the equivalent statements we have seen up to now about linear independence: Given the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ in $\mathbb{R}^N$, the following are equivalent

  - They are linearly independent
  - None of the vectors are redundant (i.e., none of the vectors can be written as a linear combination of the preceding vectors)
  - None of the vectors can be written as a linear combination of the others, i.e. $\boldsymbol{v}_i$ cannot be written as a linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_{i+1}, \ldots, \boldsymbol{v}_M$.
  - The only linear relation among them is the trivial one
  - $ker(\begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_M\end{bmatrix}) = \{0\}$
  - $rank(\begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_M\end{bmatrix}) = M$

- Now, let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ be a basis for the subspace V. Remember that this means $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ spans V, or that every vector in V can be written as a linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$. We ask whether this decomposition of a vector $\boldsymbol{v} \in V$ into $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ parts is unique. The answer is yes. Suppose we can write

$$c_1\boldsymbol{v}_1 + \cdots + c_M\boldsymbol{v}_M = v = d_1\boldsymbol{v}_1 + \cdots + d_M\boldsymbol{v}_M$$

Then we can subtract the two representations from each other:

$$(c_1 - d_1)\boldsymbol{v}_1 + \cdots + (c_M - d_M)\boldsymbol{v}_M = v - v = 0$$

But $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ is a basis! And so are linearly independent, which means that the only linear relation among them is the trivial one. So we have that $c_i = d_i$ for every $1 \leq i \leq M$.

- This relation goes the other way: suppose we have a subspace V and a bunch of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ in it. If we assume that every vector in V can be uniquely decomposed as a linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$, then $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ is a basis. For the definition of a basis, we need to check that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ 1) are in V, 2) span V, 3) are linearly independent. The first condition is given by our assumption. The second is by definition, since we assumed every vector in V is a linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$. So we only need to check that the vectors are linearly independent. By assumption, every vector in V has a uniqe decomposition. Now, the 0 vector is a vector in V, since V is a subspace. So the 0 vector has a unique linear decomposition in terms of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$. On the other hand, we know that the trivial linear relation holds true:
$$0\boldsymbol{v}_1 + \cdots + 0\boldsymbol{v}_M = 0$$
so by assumption, the only linear relation among the vectors is the trivial one, and so the vectors are linearly independent.

- So far: lots of theory. Let's do one example. Give a basis of the image of the linear transformation given by the matrix
$$A = \begin{bmatrix} 1 & 5 & 3 & 5 & 3 & 4 & 2 & 1 \\ 0 & 4 & 5 & 7 & 2 & 3 & 1 & 4 \\ 1 & 0 & 4 & 6 & 3 & 0 & 1 & 1 \\ 0 & 1 & 1 & 4 & 3 & 0 & 1 & 4 \\ 2 & 3 & 3 & 4 & 3 & 2 & 3 & 1 \\ 2 & 3 & 3 & 2 & 1 & 1 & 4 & 5 \\ 1 & 5 & 4 & 5 & 2 & 3 & 3 & 4 \end{bmatrix}$$
  - Remember that the image of the linear transformation is spanned by its column vectors. So all we need to do is to remove all the redundant column vectors from the list.
  - So we try to look at the augmented matrix
$$\left[\begin{array}{cccccccc|c} 1 & 5 & 3 & 5 & 3 & 4 & 2 & 1 & 0 \\ 0 & 4 & 5 & 7 & 2 & 3 & 1 & 4 & 0 \\ 1 & 0 & 4 & 6 & 3 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 4 & 3 & 0 & 1 & 4 & 0 \\ 2 & 3 & 3 & 4 & 3 & 2 & 3 & 1 & 0 \\ 2 & 3 & 3 & 2 & 1 & 1 & 4 & 5 & 0 \\ 1 & 5 & 4 & 5 & 2 & 3 & 3 & 4 & 0 \end{array}\right]$$

Any solution of this system will give a linear relation among the column vectors. The rref of this augmented matrix is

$$\left[\begin{array}{cccccccc:c} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

which gives a 2 paramter family of solutions

$$\begin{aligned} x_1 &= -s - t \\ x_2 &= -t \\ x_3 &= s \\ x_4 &= -s \\ x_5 &= s \\ x_6 &= t \\ x_7 &= t \\ x_8 &= 0 \end{aligned}$$

- The two parameter family of solutions implies that, writing the column vectors of A as $\boldsymbol{v}_i$,
$$-(s+t)\boldsymbol{v}_1 - t\boldsymbol{v}_2 + s\boldsymbol{v}_3 - s\boldsymbol{v}_4 + s\boldsymbol{v}_5 + t\boldsymbol{v}_6 + t\boldsymbol{v}_7 = 0$$
is true for any s and t.
- So, first we pick $s = 0$ and $t = 1$. Then we have
$$-\boldsymbol{v}_1 - \boldsymbol{v}_2 + \boldsymbol{v}_6 + \boldsymbol{v}_7 = 0$$
or
$$\boldsymbol{v}_7 = \boldsymbol{v}_1 + \boldsymbol{v}_2 - \boldsymbol{v}_6$$
So $\boldsymbol{v}_7$ is redundant.
- Similarly, picking $s = 1$ and $t = 0$, we have
$$\boldsymbol{v}_5 = \boldsymbol{v}_1 - \boldsymbol{v}_3 + \boldsymbol{v}_4$$
and so $\boldsymbol{v}_5$ is redundant.

– Since the rank of A is 6, this means that we should have 6 vectors in the basis (this will be shown next class). Once we remove $\boldsymbol{v}_5$ and $\boldsymbol{v}_7$, we have 6 vectors left. And you can easily check that the rref of the matrix formed by the 6 remaining vectors is

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

– Notice that you could just as well have removed either of $\boldsymbol{v}_2$ or $\boldsymbol{v}_6$ instead of $\boldsymbol{v}_7$, and either of $\boldsymbol{v}_3$ or $\boldsymbol{v}_4$ instead of $\boldsymbol{v}_5$. If you try to remove $\boldsymbol{v}_1$ (it is possible), you need to be more careful which other one you will remove. BUT you cannot remove $\boldsymbol{v}_8$.

Homework for this week

2.4: 4, 17, 22, 24, 28, 36, 40, 44

3.1: 5, 15, 20, 24, 31, 34, 44