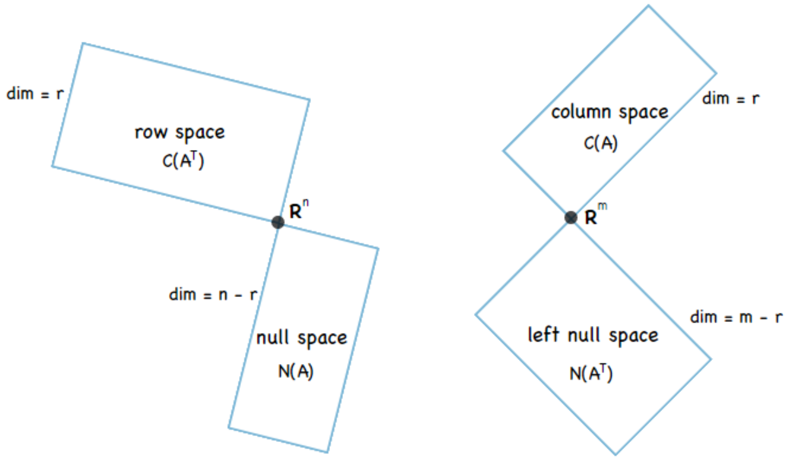
Four fundamental subspaces



Inverse of a matrix $\mathbf{A}$: $\mathbf{A}^{-1}$
Applying the matrix and its inverse in succession restore the original vector, so $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
There is no inverse if det($\mathbf{A}$) = 0

$\mathbf{Av} = \mathbf{x}$
$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{v}$
$\mathbf{x} = \mathbf{A}^{-1}\mathbf{v}$

1. Column space C(A) is all possible linear combinations of column vectors, zero vector is always in the column space. It is the span of the columns of the matrix.

$C(\mathbf{A}) = \text{span}\{\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_n}\} = \{x_1\mathbf{v_1} + x_2\mathbf{v_2} + \ldots + x_n\mathbf{v_n} \mid x_1, x_2, \ldots, x_n \in R, \mathbf{v_i} \in R^n\}$
If $\mathbf{v_1}$ $\mathbf{v_2}$ … $\mathbf{v_n}$ are linearly independent set, they are valid basis of column space

Rank r = Dim(C($\mathbf{A}$)): number of linearly independent column vectors, it equals to the dimensions in the output

2. Null space N(A) = $\{\mathbf{x} \in R^n \mid \mathbf{Ax} = \mathbf{0}\}$, also termed 'Kernel', gives all of the possible solutions for the equation $\mathbf{Ax} = \mathbf{0}$

N($\mathbf{A}$) = N(rref($\mathbf{A}$))

Nullity = Dim(N($\mathbf{A}$)): Number of free variables in rref(A)

Column vectors of A are linearly independent ⇔ N($\mathbf{A}$) = $\{\mathbf{0}\}$
Because $x_1\mathbf{v_1} + x_2\mathbf{v_2} + \ldots + x_n\mathbf{v_n} = \mathbf{0}$, $x_1, x_2, \ldots, x_n = 0$

3. Row space $R(\mathbf{A})$ or $C(\mathbf{A}^T)$ is all possible linear combinations of row vectors, zero vector is always in the row space; equal to column space of $\mathbf{A}^T$

4. Left null space $N(\mathbf{A}^T)$ is the null space of $\mathbf{A}^T$

The big picture of linear algebra

MIT lecture by Prof. Gilbert Strang on the orthogonality and dimensionality of the four fundamental subspaces:
https://www.youtube.com/watch?v=ggWYkes-n6E

Let $\mathbf{A}$ be a matrix with m rows and n columns, then there are 4 fundamental subspaces:

- Row space $R(\mathbf{A}) = C(\mathbf{A}^T)$, it contains all linear combinations of the rows of $\mathbf{A}$, or columns of $\mathbf{A}^T$
- Column space $C(\mathbf{A})$, it contains all linear combinations of the columns of $\mathbf{A}$
- Null space $N(\mathbf{A})$, it contains all solutions to the system $\mathbf{Ax} = \mathbf{0}$
- Left null space $N(\mathbf{A}^T)$, it contains all solutions to the system $\mathbf{A}^T\mathbf{y} = \mathbf{0}$

- The null space $N(\mathbf{A})$, is perpendicular to/orthogonal complement of its row space $R(\mathbf{A})$
- The left null space $N(\mathbf{A}^T)$, is perpendicular to/orthogonal complement of its column space $C(\mathbf{A})$

- $r = \text{rank}(\mathbf{A}) = \dim(C(\mathbf{A})) = \dim(R(\mathbf{A}))$
- $\dim(N(\mathbf{A})) = n - r$
- $\dim(N(\mathbf{A}^T)) = m - r$
- $\text{rank}(\mathbf{A}^T) + \text{nullity}(\mathbf{A}^T) = m$
- $\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n$

Change of basis

https://www.youtube.com/watch?v=P2LTAUO1TdA

Let V be a vector space and let $B = \{ \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \}$ be a set of vectors in V.

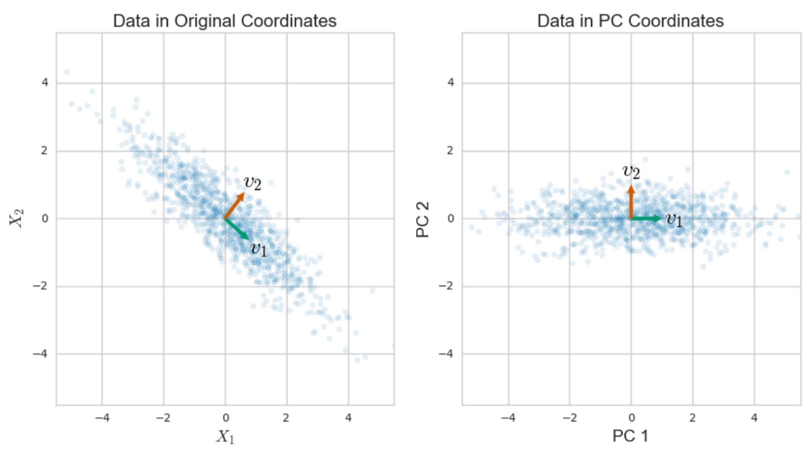
Recall that B forms a basis for V if the following two conditions hold:

1. $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent
2. S spans V

If $S = \{ \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \}$ is a basis for V, then every $\mathbf{v} \in V$ can be expressed uniquely as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$:
$\mathbf{v} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_n\mathbf{v_n}$

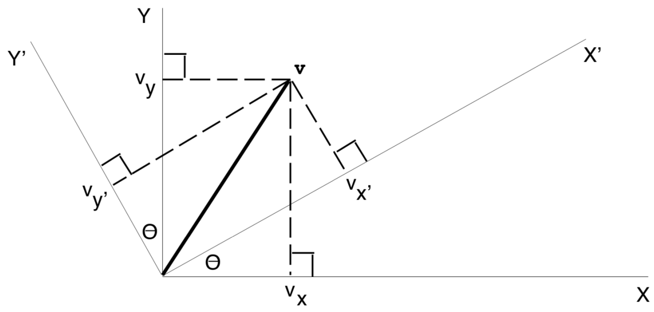
$c_1, c_2, \ldots c_n$ are just the coordinates of $\mathbf{v}$ relative to basis B. If V has a dimension of n, then every set of n linearly independent vectors in V forms a basis for V.

We have a nice coordinate representation of our vector $\mathbf{v}$, which we can construct by projecting our vector onto each basis vector individually. Now, what happens if we don't find the standard basis particularly convenient for our problem, and we would like to look at $\mathbf{v}$ from a different perspective?
We have a choice as to what basis to use!



Here we will focus on vectors in R^2, although all of this generalizes to R^n. The standard basis in R^2 is $\{ \binom{1}{0} \binom{0}{1} \}$

Let us specify other bases with reference to this rectangular coordinate system



Let **B** = {**x**, **y**} and **B'** = {**x'**, **y'**} be two bases for R^2. For a vector **v** ∈ V, given its coordinates $[\mathbf{v}]_B$ in basis B, we would like to be able to express **v** in terms of its coordinates $[\mathbf{v}]_{B'}$ in basis **B'**, and vice versa.

Suppose the basis vectors **x'** and **y'** for **B'** have the following coordinates relative to the basis **B**: $[\mathbf{x}']_B = \begin{bmatrix} a \\ b \end{bmatrix}$ $\quad [\mathbf{y}']_B = \begin{bmatrix} c \\ d \end{bmatrix}$

The change of coordinate matrix from **B'** to B: **C** = $\begin{bmatrix} a & c \\ b & d \end{bmatrix}$, meaning

x' = a**x** + b**y**

y' = c**x** + d**y**

Matrix **C** governs the change of coordinates of **v** ∈ V:

$$[\mathbf{v}]_B = \mathbf{C}[\mathbf{v}]_{B'}$$
$$[\mathbf{v}]_{B'} = \mathbf{C}^{-1}[\mathbf{v}]_B$$

Suppose we know the coordinates $[\mathbf{v}]_{B'}$ in the new basis **B'**, and we can find the coordinate of $[\mathbf{v}]_B$ in the old basis. The underlying mapping remains the same when matrix **M** is used in the original basis and the matrix **M'** is used in the new basis:

$$\mathbf{M}' = \mathbf{C}^{-1}\mathbf{M}\mathbf{C}$$
$$\mathbf{M} = \mathbf{C}\mathbf{M}'\mathbf{C}^{-1}$$

The big take away: there are arbitrary different bases we can use to represent R^n, so we can have different matrices to represent the same linear transformation under different coordinate systems!

Exercise1: We have a vector **d** = $\begin{bmatrix} 8 \\ -6 \\ 2 \end{bmatrix}$, with the change of basis matrix from standard coordinates: $\begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 1 \end{bmatrix}$, find its new coordinates in the new basis.

Hint: equivalent to solving a linear system of equations: $\begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 8 \\ -6 \\ 2 \end{bmatrix}$

a = -3, b = 11

Exercise2: Let us alternate R^2 basis to **B** = $\{\binom{1}{2}\binom{2}{1}\}$, what is the representation of the transformation matrix **M** = $\begin{bmatrix} 3 & -2 \\ 2 & -2 \end{bmatrix}$ in the new basis?

The change of basis matrix for **B** would be **C** = $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$
det(**C**) = -3
$\mathbf{C}^{-1} = -1/3\begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}$

$\mathbf{M'} = \mathbf{C}^{-1}\mathbf{MC} = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}$ It is a diagonal matrix!

When we transfer to a new basis, all of a sudden the transformation matrix is much simpler: we are only taking scaling factors of the corresponding terms during matrix multiplication. This is a neat result! It is super easy to multiply, to invert, to calculate the determinant, etc.

Linear algebra is the art of choosing the right basis!

Later we will learn eigenvectors and eigenvalues, we will see that eigenvectors make for good bases vectors/coordinate systems.

Eigenvectors and Eigenvalues

https://www.youtube.com/watch?v=PFDu9oVAE-g

In general, vectors will change direction as well as length when multiplied by a matrix. However, there will be vectors that might change length, but not direction. In other words, for these vectors, multiplication by a matrix is no different than multiplication by a simple scalar:
$\mathbf{Av} = \lambda\mathbf{v}$

A therefore acts by stretching **v** (by a scalar factor of λ), but not changing its direction. We say that **v** is an eigenvector, and λ is the corresponding eigenvalue which determines how much **v** is shortened or lengthened by a linear transformation.

Example: the shear transformation we saw last time, **A** = $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, λ = 1

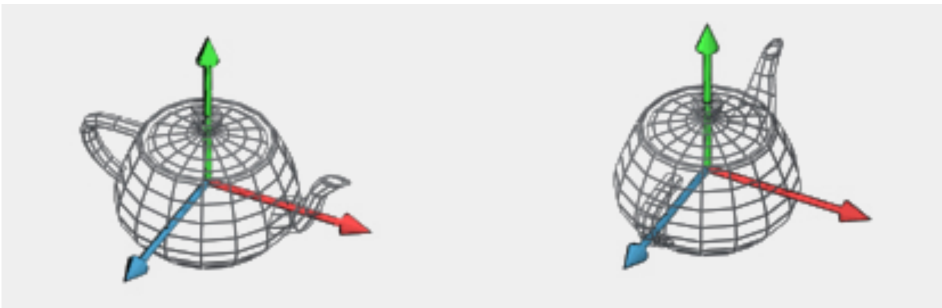
L.T. knocks off vectors from its original span, i.e. rotate them, but eigenvectors remain on their own span after transformation, the corresponding eigenvalue is the factor it gets stretched or squished after L.T.

*Any vector in eigenvector direction is still an eigenvector with the same corresponding eigenvalue, though we usually consider eigenvectors as unit vectors

Q: What does it mean when an eigenvalue is negative?
The vector gets flipped after L.T., but stays on the same line it spans out

Intuitive example: Consider spinning a globe, every location faces a new direction, except the north and south pole.

3D rotation of a tea pot, if we can find an eigenvector who remains on its own span, what does it mean? We find the axis of rotation (green vector)!



An example of calculating eigenvalues and eigenvectors of a 2x2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

Exercise: look for eigenvalues and eigenvectors of matrix: $\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$
1. Left matrix multiplication, right scalar multiplication
2. Rewrite the right hand side by some matrix multiplication:

$\mathbf{Av} = (\lambda\mathbf{I})\mathbf{v}$
$\mathbf{Av} - (\lambda\mathbf{I})\mathbf{v} = \mathbf{0}$
$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$

We want a non-zero eigenvector, a product of a matrix with a non-zero vector equals to zero, the transformation therefore squishes the space into a lower dimension, corresponding to a determinant of zero:

$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
$(a-\lambda)(d-\lambda) - bc = 0$
$(3-\lambda)(1-\lambda) - 8 = 0$

$\lambda_1 = 5$, $\lambda_2 = -1$

$\mathbf{v}_1 = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

$\text{Eigenspace}_5 = \text{span}\{\begin{bmatrix} 0.5 \\ 1 \end{bmatrix}\}$, $\text{Eigenspace}_{-1} = \text{span}\{\begin{bmatrix} -1 \\ 1 \end{bmatrix}\}$

Eigenvectors and eigenvalues pop up in many areas such as:
Image compression: using SVD to find eigenvectors of the covariance matrix, and project data onto the eigenvectors of the largest eigenvalues, thereby throwing away the small eigenvalues.

Dimensionality reduction for machine learning and data analysis: using PCA to find the principal components that correspond to the largest eigenvalues of the covariance matrix. It allows us to understand where most of the variation in a data comes from.

Carsen will lecture on PCA, SVD in details in the following week.

Related to linear regression week:

Linear algebra view of least-squares regression

The goal of regression is to fit a mathematical model, i.e. a linear equation, to a set of observed points. We believe there is an underlying relationship that maps a to b, in the form of b = ca + d. Here, c and d are the regression coefficients we are looking for.

We can write the observed points as a simple linear system such as:
$b_1 = ca_1 + d$
$b_2 = ca_2 + d$
$b_3 = ca_3 + d$

We have a simple linear system and we only need to deal with vectors and matrices!
Here is our linear system in the matrix form: $\mathbf{Ax} = \mathbf{b}$

$$\begin{bmatrix} a1 & 1 \\ a2 & 1 \\ a3 & 1 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} b1 \\ b2 \\ b3 \end{bmatrix}$$

What this is saying is that we hope the vector **b** lies in the column space of **A**, i.e. C(**A**). In other word, we wish to find a linear combination of the columns of **A** that gives us our vector **b**.
But we know that most of the time **b** doesn't fit our model perfectly, meaning that it is outside the column space of **A**. We cannot simply solve the equation **Ax** = **b** for vector **x**.

The linear regression answer is that we intend to swap out for **b** for another vector that is very close to it but fits in our model. Specifically, we want to pick a vector **p** that is in the column space of **A**, but is also as close as possible to **b**.

This is termed least squares approximation in linear algebra. We would like to find **p**, that could minimize the difference between observed **b** and the projected **p**. Here, **p** is the projection of **b** in C(**A**).

Geometry makes it pretty clear what's going on. We started with **b**, which doesn't fit the model, and then switched to **p**, which gives a good approximation and has the virtue of sitting in C(**A**).
$\mathbf{p} = \mathbf{Ax}^*$

Now solving the regression coefficients becomes solving for $\mathbf{x}^*$, which is the estimated regression coefficients c and d.

$\mathbf{x}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$