Day two: Linear Algebra

Rebecca Johnson

September 7th, 2016

Outline

- **Vectors**
  - Basic notation
  - Multiplying by a scalar
  - Addition and subtraction (sidenote on conformability)
  - Two forms of vector multiplication:
    1. Dot/inner product
    2. Cross product
  - Norms
- **From Vectors to matrices**
  - Typology of matrix types
  - Basic matrix algebra: addition/subtraction, scalar multiplication
  - More complex matrix algebra: matrix multiplication (and transpose to help with multiplication)
  - Summary measures: matrix rank, inverse, determinant
  - Interspersed with the matrix material are preview of two applications:
    1. Transforming rectangular matrices of individual-level measures to square matrices of dyad-level distance measures
    2. Matrix representation of linear regression

Data source for vectors and matrix: senators' co-sponsorship of bills during 2004 session

Source for data: James H. Fowler: Connecting the Congress: A Study of Cosponsorship Networks , *Political Analysis* 14 (4): 456-487 (Fall 2006) and Legislative Cosponsorship Networks in the U.S. House and Senate, *Social Networks* 28 (4): 454-465 (October 2006). Cleaned senate network data provided as part of Skyler Cranmer, ICPSR 2016 Network Analysis workshop.

Motivation for vectors

- Compact way of storing information, instead of:

Cosponsors: S.1775 — 110th Congress (2007-2008)

All Bill Information (Except Text)

Sponsor: Sen. Burr, Richard [R-NC] (includes 1 original)

Party

Check all

Republican [3]

Cosponsors by U.S. State or Territory

Georgia [1]

New Hampshire [1]

Tennessee [1]

* = Original cosponsor

Sort by First to Last

Cosponsor	Date Cosponsored
Sen. Gregg, Judd [R-NH]*	07/12/2007
Sen. Alexander, Lamar [R-TN]	11/06/2007
Sen. Isakson, Johnny [R-GA]	11/06/2007

- Use:

	Gregg	Alexander	Isaakson	Lieberman
Burr	1	1	1	0

Vector notation

- **Example**: two vectors: John McCain's versus Russ Feingold's cosponsorship of bills with Hillary Clinton, Lincoln Chaffee, Joseph Lieberman, and Strom Thurmond, stored in that order in the vector
- Let:

$$John\ McCain's\ cosponsorship = \mathbf{u} = \begin{bmatrix} HRC & LC & JL & ST \end{bmatrix} = \begin{bmatrix} 1 & 1 & 10 & 5 \end{bmatrix}$$

$$Russ\ Feingold's\ cosponsorship = \mathbf{v} = \begin{bmatrix} HRC & LC & JL & ST \end{bmatrix} = \begin{bmatrix} 2 & 2 & 8 & 1 \end{bmatrix}$$

- **Bold** = entire vector; non-bold = element of vector. For instance:
  - $\mathbf{u}$: all of John McCain's sponsorship information
  - u_3: John McCain's sponsorship with Joe Lieberman (ten bills)
  - v_4: Russ Feingold's sponsorship with Strom Thurmond (one bill)

Vector notation: continued

Can arrange data in multiple ways depending on purpose:

- Row vector:
$$\mathbf{u} = \begin{bmatrix} 1 & 1 & 10 & 5 \end{bmatrix}$$
- Column vector:
$$\mathbf{u}^T = \begin{bmatrix} 1 \\ 1 \\ 10 \\ 5 \end{bmatrix}$$

Basic operations with vectors and properties of those operations

Can group operations into two types:

1. Vector *operation* scalar
2. Vector *operation* other vector (can include the vector itself)

Operations with a scalar: motivation

- **Example**: right now, the co-sponsorship is coded as a continuous variable ranging from 0 to 10 in the case of these particular vectors. What if we want to rescale so that each element is between 0 and 1 to more easily compare John McCain and Russ Feingold's sponsorship patterns?
- Two potential ways to rescale:
    1. Multiply each cosponsorship vector by the maximum cosponsorship value *across all* senators: $\frac{1}{max(\mathbf{u}, \mathbf{v})}$
    2. Multiply each cosponsorship vector by the maximum cosponsorship value *within each* senator: $\frac{1}{max(\mathbf{u})}, \frac{1}{max(\mathbf{v})}$

Operations with a scalar: mechanics

1. Rescaling one (maximum cosponsorship across all senators). Let $s = \frac{1}{max(\mathbf{u},\mathbf{v})}$:

$$\mathbf{u}_{scaledallmax} = s\mathbf{u} = \frac{1}{10}\begin{bmatrix}1 & 1 & 10 & 5\end{bmatrix} = \begin{bmatrix}\frac{1}{10}\times 1 & \frac{1}{10}\times 1 & \frac{1}{10}\times 10 & \frac{1}{10}\times 5\end{bmatrix}$$
$$= \begin{bmatrix}0.1 & 0.1 & 1 & 0.5\end{bmatrix}$$

$$\mathbf{v}_{scaledallmax} = s\mathbf{v} = \begin{bmatrix}0.2 & 0.2 & 0.8 & 0.1\end{bmatrix}$$

2. Rescaling two (maximum cosponsorship within each senator). Let $s_1 = \frac{1}{max(\mathbf{u})} \implies s_1 = \frac{1}{10}$, $s_2 = \frac{1}{max(\mathbf{v})} \implies s_2 = \frac{1}{8}$:

$$\mathbf{u}_{scaledwithinmax} = s_1\mathbf{u} = \begin{bmatrix}0.1 & 0.1 & 1 & 0.5\end{bmatrix}$$
$$\mathbf{v}_{scaledallmax} = s_2\mathbf{v} = \begin{bmatrix}0.25 & 0.25 & 1 & 0.125\end{bmatrix}$$

Operations with a scalar: substantive interpretation

Table 1: John McCain's cosponsorship

	Hillary	Lincoln	Joe	Strom
Original	1	1	10	5
Max across all	0.10	0.10	1.00	0.50
Max within senator	0.10	0.10	1.00	0.50

Table 2: Russ Feingold's cosponsorship

	Hillary	Lincoln	Joe	Strom
Original	2	2	8	1
Max across all	0.20	0.20	0.80	0.10
Max within senator	0.25	0.25	1.00	0.12

Some differences with whether you do or don't adjust for different senator-specific levels of cosponsorship activity for the pressing question that has been keeping you up at night: with whom is Joe Lieberman closer, John McCain or Russ Feingold?

Operations with other vectors: motivation

- We used scalar multiplication to compare two vectors, but what if we want to compare more directly?
- For instance, which pair of senators exhibits a higher degree of similarity in terms of their cosponsorship patterns?:
  1. John McCain and Russ Feingold
  2. John McCain and Rick Santorum
- Answering this question requires performing operations on pairs of vectors

The overarching concern with vector by vector calculations: are the vectors conformable?

- When we multiplied the McCain vector by the $\frac{1}{max}$ scalar, we were doing something similar to the following:

$$\underset{1\times 1}{\begin{bmatrix}\frac{1}{10}\end{bmatrix}}\underset{1\times 4}{\begin{bmatrix}1 & 1 & 10 & 5\end{bmatrix}} = \underset{1\times 4}{\begin{bmatrix}0.1 & 0.1 & 1 & 0.5\end{bmatrix}}$$

- But unlike a scalar (always 1×1) and always distributes to all elements of a vector or matrix, when performing operations with vectors, we need to make sure the dimensions of the vectors are compatible
- **Conformable**: special term for when the dimensions of a vector (which is equivalent to a one-row or one-column matrix) or matrix allow us to perform some operation; will discuss more when we move to matrices but important to emphasize that conformability is always in the context of *some operation* (e.g., addition versus multiplication)

Vector addition and subtraction

- **Conformable in this case**: vectors must have same number of elements
- **Motivating example**: finding the residual, where y = vector of observed values for the outcome variable and $\hat{y}$ = vector of fitted values for the outcome variable ($\hat{e}$ is sometimes denoted $\hat{u}$):

$$\hat{e} = y - \hat{y}$$

- **Example from cosponsorship data**: fit a linear regression that regresses a senator's total number of cosponsorships against a measure of how liberal versus conservative they are on economic issues (DW-NOMINATE score, dim2). Residuals are:

$$\hat{e} = \text{observed cosponsorship count - fitted cosponsorship count}$$

Vector addition and subtraction

- **Conformable**: When we could properly subtract the fitted cosponsorship count from the observed count to find each senator's residual– same dimensions for fitted vector as observe vector:

$$\underset{5\times 1}{\overset{\textit{Senators}}{\begin{bmatrix} \textit{John McCain} \\ \textit{Joe Lieberman} \\ \textit{Rick Santorum} \\ \textit{Joe Biden} \\ \textit{John Edwards} \end{bmatrix}}} : \underset{5\times 1}{\overset{\textit{Observed cosponsorships}}{\begin{bmatrix} 176 \\ 351 \\ 158 \\ 247 \\ 203 \end{bmatrix}}} - \underset{5\times 1}{\overset{\textit{Fitted cosponsorships}}{\begin{bmatrix} 215.21 \\ 300.92 \\ 211.23 \\ 320.14 \\ 304.06 \end{bmatrix}}} = \underset{5\times 1}{\overset{\textit{Residuals}}{\begin{bmatrix} -39.21 \\ 50.08 \\ -53.23 \\ -73.14 \\ -101.06 \end{bmatrix}}}$$

- **Not conformable**: When we could not (cosponsorship data from before Obama joined senate but pretend we have Obama's ideology score to get a fitted cosponsorship value), residuals vector is undefined:

$$\underset{6\times 1}{\overset{\textit{Senators}}{\begin{bmatrix} \textit{John McCain} \\ \textit{Joe Lieberman} \\ \textit{Rick Santorum} \\ \textit{Joe Biden} \\ \textit{John Edwards} \\ \textit{Barack Obama} \end{bmatrix}}} : \underset{5\times 1}{\overset{\textit{Observed cosponsorships}}{\begin{bmatrix} 176 \\ 351 \\ 158 \\ 247 \\ 203 \end{bmatrix}}} - \underset{6\times 1}{\overset{\textit{Fitted cosponsorships}}{\begin{bmatrix} 215.21 \\ 300.92 \\ 211.23 \\ 320.14 \\ 304.06 \\ 332.74 \end{bmatrix}}}$$

Two forms of vector multiplication: motivation

- **(Stylized!) example**: want to calculate likelihood of two senators cosponsoring a bill in the future using their pattern of collaboration with the three senators with highest cosponsorship counts: Mary Landrieu (conservative dem), Tim Johnson (centrist dem), and Jon Corzine
- Will use this example to review two forms of vector multiplication (there are others out there):
    1. Dot or inner product: $\mathbf{u} \bullet \mathbf{v}$
    2. Cross product: $\mathbf{u} \times \mathbf{v}$

Dot product: motivation

- Senators of interest:

$$
\begin{aligned}
&= \begin{bmatrix} \textit{Mary L.} & \textit{Tim J.} & \textit{Jon C.} \end{bmatrix} \\
\textit{Paul Wellstone} = \mathbf{u} &= \begin{bmatrix} 0 & 8 & 2 \end{bmatrix} \\
\textit{Joe Lieberman} = \mathbf{v} &= \begin{bmatrix} 4 & 2 & 6 \end{bmatrix} \\
\textit{Dianne Feinstein} = \mathbf{z} &= \begin{bmatrix} 3 & 1 & 1 \end{bmatrix}
\end{aligned}
$$

- Intuition behind **dot or inner product**:
  - Each additional non-zero cosponsorship with one of those three senators provides an extra "boost" towards the pair's own likelihood of collaborating on a bill together but that boost is lost if *either* of the senators has sponsored no bills with a given senator
  - For instance, Joe Lieberman's high degree of collaboration with Mary Landrieu does not provide a boost towards collaboration with Paul Wellstone, since Paul Wellstone has zero collaboration with Mary
  - **Result**: scalar (single value) for each pair of senators that provides a summary measure of what we think matters for their likelihood of collaboration

Dot product: mechanics

- **Conformable in this case**: vectors have same dimension
- Let n = number of elements in the vector:

$$\mathbf{u} \bullet \mathbf{v} = u_1 v_1 + u_2 v_2 \ldots u_n v_n = \sum_{i=1}^{n} u_i * v_i$$

- Example with Paul Wellstone and Joe Lieberman:

$$\begin{aligned} \textit{Paul Wellstone} = \mathbf{u} &= \begin{bmatrix} 0 & 8 & 2 \end{bmatrix} \\ \textit{Joe Lieberman} = \mathbf{v} &= \begin{bmatrix} 4 & 2 & 6 \end{bmatrix} \end{aligned}$$

$$\mathbf{u} \bullet \mathbf{v} = 0 * 4 + 8 * 2 + 2 * 6 = 28$$

Dot product: find all combinations and interpret

$$
\begin{aligned}
&= \begin{bmatrix} \textit{Mary L.} & \textit{Tim J.} & \textit{Jon C.} \end{bmatrix} \\
\textit{Paul Wellstone} = \mathbf{u} &= \begin{bmatrix} 0 & 8 & 2 \end{bmatrix} \\
\textit{Joe Lieberman} = \mathbf{v} &= \begin{bmatrix} 4 & 2 & 6 \end{bmatrix} \\
\textit{Dianne Feinstein} = \mathbf{z} &= \begin{bmatrix} 3 & 1 & 1 \end{bmatrix}
\end{aligned}
$$

On your worksheet, find the dot product by hand for the other two combinations: Wellstone and Feinstein, Lieberman and Feinstein. Then, use R with the cleaned data we provide to confirm your solution by writing a function or for loop that either 1) you feed two vectors and apply separately each time; 2) you feed a matrix containing all three vectors and the function produces the dot product for each unique combination

Hint: if you pursue strategy two, R's *combn* function might be helpful. It takes the form: combn(source for combinations, number of elements to choose)

Dot product: find all combinations and interpret

$$
\begin{array}{rl}
 & = \begin{bmatrix} \textit{Mary L.} & \textit{Tim J.} & \textit{Jon C.} \end{bmatrix} \\
\textit{Paul Wellstone} = \mathbf{u} = & \begin{bmatrix} 0 & 8 & 2 \end{bmatrix} \\
\textit{Joe Lieberman} = \mathbf{v} = & \begin{bmatrix} 4 & 2 & 6 \end{bmatrix} \\
\textit{Dianne Feinstein} = \mathbf{z} = & \begin{bmatrix} 3 & 1 & 1 \end{bmatrix}
\end{array}
$$

- **Our R solution pursuing strategy two**:

```
rowcombos <- t(combn(1:nrow(senfordot), 2))
dotresults <- c()
for(i in 1:nrow(rowcombos)){
    dotresults[i] <- sum(senfordot[rowcombos[i, 1], ] *
                         senfordot[rowcombos[i, 2], ])
}
dotresults
```

Dot product: find all combinations and interpret

$$
\begin{array}{rl}
 & = \begin{bmatrix} \textit{Mary L.} & \textit{Tim J.} & \textit{Jon C.} \end{bmatrix} \\
\textit{Paul Wellstone} = \mathbf{u} = & \begin{bmatrix} 0 & 8 & 2 \end{bmatrix} \\
\textit{Joe Lieberman} = \mathbf{v} = & \begin{bmatrix} 4 & 2 & 6 \end{bmatrix} \\
\textit{Dianne Feinstein} = \mathbf{z} = & \begin{bmatrix} 3 & 1 & 1 \end{bmatrix}
\end{array}
$$

- **Results**:
  1. Paul Wellstone and Joe Lieberman (28)
  2. Joe Lieberman and Dianne Feinstein (20)
  3. Paul Wellstone and Dianne Feinstein (10)
- **Interpretation**: the fact that Dianne Feinstein's highest collaborator is Mary L. hurts her potential-for-collaboration score with Paul Wellstone, since his zero collaborations with Mary L. makes her high score count for nothing towards their potential

Dot product: alternate notation

See Gill page 88 for formal properties (commutative, associative, etc.) but for now, worth highlighting the following:

- Alternate way to write the inner product: $\mathbf{u} \bullet \mathbf{v} = \mathbf{u'v}$
- Can see with the Wellstone and Lieberman vectors that:

$$\mathbf{u'} = \begin{bmatrix} 0 \\ 8 \\ 2 \end{bmatrix}$$

$$\mathbf{v} = \begin{bmatrix} 4 & 2 & 6 \end{bmatrix}$$

$$\mathbf{u'v} = 0 * 4 + 8 * 2 + 6 * 2 = 28 = \mathbf{u} \bullet \mathbf{v}$$

Dot product: other applications

- We've used the dot product as one measure of the similarity of a given pair of senators' co-sponsorship patterns, but more broadly, we can use the dot product to assess the orientation of two vectors
- In what case does the dot product equal zero and what does it imply about the vectors? Imagine the following senator who we're trying to find a collaboration score with Wellstone ($\mathbf{u}$):

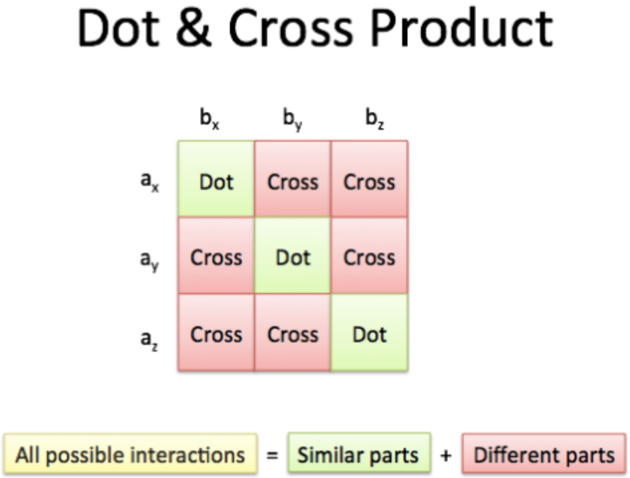
$$\mathbf{u} = \begin{bmatrix} 0 & 8 & 2 \end{bmatrix}, \ \mathbf{w} = \begin{bmatrix} 4 & 0 & 0 \end{bmatrix}$$
$$\mathbf{u} \bullet \mathbf{w} = 0$$

- When the dot product of two vectors is zero, these vectors are **orthogonal/perpendicular**
- This has important applications for linear regression (will discuss more in 500/504):
  - **Assumption of linear regression**: the vector of residuals is orthogonal to every column of X (the covariate matrix). More formally: $X^T e = 0$. This is more colloquially discussed as the independence assumption.

Cross product: motivation

- With the **dot product**, a senator pair's "potential for collaboration" score increases when they share a *common cosponsor*; the resulting score is a single value (a scalar)
  - If Paul Wellstone sponsors a bill with Jon Corzine and Dianne Feinstein sponsors a bill with Tim Johnson, the product of those cosponsors *does not increase* Paul and Dianne's potential for collaboration score
- With the **cross product**, we can form a different "potential for collaboration" measure that increases not if the senators share a *common cosponsor*, but instead if the senators share a *dissimilar cosponsor*(e.g., want to accumulate diverse cosponsors to help bridge disparate cliques in the senate); the resulting score is a vector composed of each interaction
  - If Paul Wellstone sponsors a bill with Jon Corzine and Dianne Feinstein sponsors a bill with Tim Johnson, the product of those cosponsors *does* appear in their potential for collaboration vector
- *Note*: the dot product is defined for vectors of any dimension (as long as for $\mathbf{u} \bullet \mathbf{v}$, $\mathbf{u}$ has same n as $\mathbf{v}$); the cross product is defined only for vectors in $\mathbb{R}^3$

Cross product: intuition behind mechanics



Source: Betterexplained.com

Cross product: mechanics

1. Stack the vectors on top of each into a matrix (*note*: $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$, so stacking in opposite order will produce same magnitude but differently-signed result vector):

$$\mathbf{A} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \textit{Joe Lieberman} \\ \textit{Dianne Feinstein} \end{bmatrix} = \begin{bmatrix} 4 & 2 & 6 \\ 3 & 1 & 1 \end{bmatrix}$$

2. Extract all 2×2 sub-matrices from that matrix in the following order:

$$\mathbf{A}[1,2;\ 2,3] = \begin{bmatrix} 2 & 6 \\ 1 & 1 \end{bmatrix} \ \mathbf{A}[1,2;3,1] = \begin{bmatrix} 6 & 4 \\ 1 & 3 \end{bmatrix} \ \mathbf{A}[1,2;1,2] = \begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix}$$

3. Find the determinant of each sub-matrix (when 2×2, and $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $det(A) = ad - bc$) and arrange into a vector:

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} -4 & 14 & -2 \end{bmatrix}$$

Cross product: find for remaining vector pairs

On your worksheet, find the cross product the remaining two pairs of vectors by hand. Then, use the vector cross product function (*xprod*) we provide in the R file (▸ source) to find the cross product

Answers (signs will depend on order you specified)

```
> jldfcrossprod <- xprod(a = senfordot["jl", ], b = senfordot["df", ])
> jlpwcrossprod <- xprod(a = senfordot["jl", ], b = senfordot["pw", ])
> pwdfcrossprod <- xprod(a = senfordot["pw", ], b = senfordot["df", ])
> pwdfcrossprod
[1]   6   6 -24
> jlpwcrossprod
[1] -44  -8  32
> jldfcrossprod
[1] -4 14 -2
```

Cross product: interpretation

Geometric interpretation of the *magnitude* is as the size of the area between the two vectors if we ▸ plot them in the (x, y, z) plane, More stylized example:

1. Senators with very different co-sponsorship patterns: large-magnitude cross product

$$\mathbf{A} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} Sen1 \\ Sen2 \end{bmatrix} = \begin{array}{c} [\textit{Mary L.} \quad \textit{Tim J.} \quad \textit{Jon C.}] \\ \begin{bmatrix} 8 & 0 & 2 \\ 1 & 9 & 7 \end{bmatrix} \end{array}$$

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} -18 & -54 & 72 \end{bmatrix}$$

2. Senators with very similar co-sponsorship patterns (vectors are almost overlapping, can see that the zero element stems from the leftmost sub-matrix where determinant $= 0$):

$$\mathbf{A} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} Sen1 \\ Sen2 \end{bmatrix} = \begin{array}{c} [\textit{Mary L.} \quad \textit{Tim J.} \quad \textit{Jon C.}] \\ \begin{bmatrix} 8 & 8 & 2 \\ 7 & 7 & 3 \end{bmatrix} \end{array}$$

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} 10 & -10 & 0 \end{bmatrix}$$

What is the cross product in the following case?

$$\mathbf{A} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} Sen1 \\ Sen2 \end{bmatrix} = \begin{array}{c} \begin{bmatrix} \textit{Mary L.} & \textit{Tim J.} & \textit{Jon C.} \end{bmatrix} \\ \begin{bmatrix} 2 & 4 & 1 \\ 6 & 12 & 3 \end{bmatrix} \end{array}$$

What is the cross product in the following case?

$$\mathbf{A} = \begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}Sen1\\ Sen2\end{bmatrix} = \begin{array}{c} \begin{bmatrix} \textit{Mary L.} & \textit{Tim J.} & \textit{Jon C.} \end{bmatrix} \\ \begin{bmatrix} 2 & 4 & 1 \\ 6 & 12 & 3 \end{bmatrix} \end{array}$$

What is the cross product in the following case?

$$\mathbf{A} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} Sen1 \\ Sen2 \end{bmatrix} = \begin{array}{c} [\textit{Mary L.} \quad \textit{Tim J.} \quad \textit{Jon C.}] \\ \begin{bmatrix} 2 & 4 & 1 \\ 6 & 12 & 3 \end{bmatrix} \end{array}$$

- Cross product is 0, which results from the fact that Senator 1's cosponsorship can be written as a linear combination of senator 2's cosponsorship vector ($Sen\,2 = 3 * Sen\,1$), more formally: $\mathbf{v} = 3\mathbf{u}$)
- Less formally, the vectors fully overlap; more formally, they are *linearly dependent*

Implications of dot versus cross product equaling zero

Different implications, and remember that we can find the dot product of any two vectors of the same dimension, but cross product (and potential for cross product to equal zero) is only defined in $\mathbb{R}^3$:

- $\mathbf{u} \bullet \mathbf{v} = 0$: vectors are perpendicular/orthogonal
- $\mathbf{u} \times \mathbf{v} = 0$: vectors are linearly dependent, which in geometric terms, means the vectors are *parallel*

Vector length and distance between: motivation

- Thus far, we've been constructing our own "potential for collaboration" score as either a scalar measuring the extent to which two senators' cosponsorship vectors overlap (dot product) or as a vector measuring the extent to which two senators' help "bridge" disparate parts of the cosponsorship space (a large area between them, the cross product)
- But more general/common uses of the dot and cross product are to calculate:
  - The length of a vector (more formally, the vector's *norm*)
  - The distance between two vectors

Vector length and distance between: motivation

- Returning to our previous example:

$$\begin{array}{rl} & = \begin{bmatrix} \textit{Mary L.} & \textit{Tim J.} & \textit{Jon C.} \end{bmatrix} \\ \textit{Paul Wellstone} = \mathbf{u} & = \begin{bmatrix} 0 & 8 & 2 \end{bmatrix} \\ \textit{Joe Lieberman} = \mathbf{v} & = \begin{bmatrix} 4 & 2 & 6 \end{bmatrix} \\ \textit{Dianne Feinstein} = \mathbf{w} & = \begin{bmatrix} 3 & 1 & 1 \end{bmatrix} \end{array}$$

- For this example, it is easy to see which senator has the highest magnitude of cosponsorship (Joe Lieberman) because all elements of the vector have positive values
- But what if we were dealing with, for instance, how far away the senator's *observed* cosponsorship count with another senator was from his/her *predicted* cosponsorship count based on some model. And we want undershooting to count for the same as overshooting:

$$\begin{array}{rl} & = \begin{bmatrix} \textit{Mary L.} & \textit{Tim J.} & \textit{Jon C.} \end{bmatrix} \\ \textit{Paul Wellstone} = \mathbf{u} & = \begin{bmatrix} 2 & -3 & 5 \end{bmatrix} \\ \textit{Joe Lieberman} = \mathbf{v} & = \begin{bmatrix} -2 & -3 & -4 \end{bmatrix} \end{array}$$

Vector length and distance between: motivation

- Squaring to the rescue! To help both negative and positive prediction errors (and elements more generally) contribute the same magnitude to the vector's overall size
- We could depict that squaring in a lengthy way, e.g. if we had Wellstone's vector **u**, we could begin to find its length using some function of:

$$u_1^2 + u_2^2....u_n^2$$

- But what's a more compact way to write this...?

Vector length and distance between: mechanics

- **Euclidean norm**: (one option for measuring vector length; others out there). It is also denoted as $\|\mathbf{u}\|_2$ to indicate that you're squaring $\mathbf{u}$ and then raising it to the $\frac{1}{2}$ ($\sqrt{\ }$):

$$\|\mathbf{u}\|^2 = \mathbf{u} \bullet \mathbf{u} = \mathbf{u}^T\mathbf{u}$$
$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \bullet \mathbf{u}}$$

- **Difference norm**: one measure of the distance between two vectors:
  1. Start with:
  $$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u} - \mathbf{v}\|\|\mathbf{u} - \mathbf{v}\|$$
  2. Expand/distribute:
  $$\|\mathbf{u}\|^2 - 2(\mathbf{u} \bullet \mathbf{v}) + \|\mathbf{v}\|^2$$
  3. Can also rewrite using dot product or vector-transpose notation:
  $$= \mathbf{u} \bullet \mathbf{u} - 2(\mathbf{u} \bullet \mathbf{v}) + \mathbf{v} \bullet \mathbf{v}$$
  $$= \mathbf{u}^T\mathbf{u} - 2(\mathbf{u}^T\mathbf{v}) + \mathbf{v}^T\mathbf{v}$$

Applying with the (fake) example of three vectors of prediction errors for cosponsorship counts

$$
\begin{array}{rcccc}
 & & \textit{Mary L.} & \textit{Tim J.} & \textit{Jon C.} \\
\textit{Paul Wellstone} = \mathbf{u} = & [& 2 & -3 & 5 \;] \\
\textit{Joe Lieberman} = \mathbf{v} = & [& -2 & -3 & -4 \;] \\
\textit{Dianne Feinstein} = \mathbf{w} = & [& 0 & 5 & -1 \;]
\end{array}
$$

Using the vectors in your .rmd and R's built-in calculator, calculate and interpret following:

1. $\|u\|$, $\|v\|$, $\|w\|$ For whom did the model do the best job of prediction (smallest magnitude prediction error)? For whom did the model do the worst job?.
2. $\|u - v\|$, $\|u - w\|$, $\|w - v\|$ (remember that we gave you the expression for $\|u - v\|^2$ rather than $\|u - v\|$, so adapt appropriately). Which senator pair had the smallest difference in how well the model predicted their cosponsorship? Which senator pair had the largest difference?

Answers

1. Who had the smallest prediction error (smallest norm)?: Feinstein ($\mathbf{w}$; largest prediction error: Wellstone ($\mathbf{u}$)
2. Smallest difference in prediction errors? $\|\mathbf{v} - \mathbf{w}\|$ (Lieberman and Feinstein); largest difference in prediction errors $\|\mathbf{u} - \mathbf{w}\|$ (Wellstone and Feinstein) (makes sense given those had the smallest and largest norms)
3. Reviewed code to generate in class and will post on Blackboard

Vector norms: properties

Less obvious than properties of vector addition/subtraction so will briefly review the more important ones (discussed on Gill page 97)
Some product properties useful for proofs of vector norm properties (more on Gill page 88):

- Commutative: $\mathbf{u} \bullet \mathbf{v} = \mathbf{u} \bullet \mathbf{v}$
- Distributive: $\mathbf{w} \bullet (\mathbf{u} \bullet \mathbf{v}) = \mathbf{w} \bullet \mathbf{u} + \mathbf{w} \bullet \mathbf{v}$
- Cauchy-schwarz inequality: $|\mathbf{u} \bullet \mathbf{v}| \leq \|\mathbf{u}\| \|\mathbf{v}\|$, can also write as: $|\mathbf{u} \bullet \mathbf{v}|^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$
  - Application you might run into in probability/statistics: can use to show that $|cov(X, Y)| \leq \sqrt{E(X - \mu_x)^2} \sqrt{E(Y - \mu_y)^2}$
- Triangle inequality: $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
  - On your worksheet, to help you practice dot product properties, write an informal proof of the triangle inequality. We provide an outline of key steps on the worksheet

Answers

1. Square each side of the triangle inequality:
$$\|\mathbf{u}+\mathbf{v}\|\|\mathbf{u}+\mathbf{v}\| \leq (\|\mathbf{u}\|+\|\mathbf{v}\|)(\|\mathbf{u}\|+\|\mathbf{v}\|)$$
2. Focusing on the LHS, we can expand the $\|\mathbf{u}+\mathbf{v}\|$:
$$\|\mathbf{u}+\mathbf{v}\| = (\mathbf{u} \bullet \mathbf{u} + 2(\mathbf{u} \bullet \mathbf{v}) + \mathbf{v} \bullet \mathbf{v})^{\frac{1}{2}}$$
3. Since $x^{\frac{1}{2}} \times x^{\frac{1}{2}} = x$, the LHS becomes:
$$\mathbf{u} \bullet \mathbf{u} + 2(\mathbf{u} \bullet \mathbf{v}) + \mathbf{v} \bullet \mathbf{v}$$
4. Focusing on the RHS, we can distribute the multiplication, first writing in terms of norms:
$$\|\mathbf{u}\|^2 + 2(\|\mathbf{u}\|\|\mathbf{v}\|) + \|\mathbf{v}\|^2$$
5. Then, use $\|\mathbf{u}\|^2 = \mathbf{u} \bullet \mathbf{u}$ to rewrite in terms of dot products to be similar to the left hand side:
$$\mathbf{u} \bullet \mathbf{u} + 2(\|\mathbf{u}\|\|\mathbf{v}\|) + \mathbf{v} \bullet \mathbf{v}$$
6. Now, let's plug those into the original expression, highlighting in red what differs:
$$\mathbf{u} \bullet \mathbf{u} + 2\color{red}{(\mathbf{u} \bullet \mathbf{v})} + \mathbf{v} \bullet \mathbf{v} \overset{?}{\leq} \mathbf{u} \bullet \mathbf{u} + 2\color{red}{(\|\mathbf{u}\|\|\mathbf{v}\|)} + \mathbf{v} \bullet \mathbf{v}$$
7. Cauchy-schwartz tells us that $|\mathbf{u} \bullet \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|$, and since $\mathbf{u} \bullet \mathbf{v} \leq |\mathbf{u} \bullet \mathbf{v}|$, $\mathbf{u} \bullet \mathbf{v} \leq \|\mathbf{u}\|\|\mathbf{v}\|$, and since that's the only part that differs, the LHS $\leq$ RHS

We've been assessing the relationship between *pairs* of vectors; how can we generalize to broader patterns?

$$
\begin{array}{r c}
 & \begin{matrix} \text{\textit{Mary L.}} & \text{\textit{Tim J.}} & \text{\textit{Jon C.}} \end{matrix} \\
\textit{Paul Wellstone} = \mathbf{u} = & \begin{bmatrix} 0 & 8 & 2 \end{bmatrix} \\
\textit{Joe Lieberman} = \mathbf{v} = & \begin{bmatrix} 4 & 2 & 6 \end{bmatrix} \\
\textit{Dianne Feinstein} = \mathbf{w} = & \begin{bmatrix} 3 & 1 & 1 \end{bmatrix}
\end{array}
$$

- Thus far, to evaluate the structure of these co-sponsorship patterns, we've been pulling out each pair of vectors
- We'll keep doing that, but we can also treat the stacked vectors as *matrices* and discern new information about cosponsorship patterns

And with that somewhat forced transition, from vectors to matrices...

Where we're going:

- Types of matrices
- Matrix algebra:
  - Basic: addition, subtraction, scalar multiplication
  - More complex: matrix multiplication and transposition
- Ways to use matrices
  - Preview of role matrices play in linear regression
  - (Briefly) Solving systems of equations, draws upon:
    - What is the rank of the matrix?
    - Is the matrix invertible? To assess that, we need to ask:
    - What is the determinant of the matrix?

Types of matrices: motivation

- We've been drawing out vectors from the senate cosponsorship data, but now let's view the structure of the entire matrix (or more precisely, a matrix with more obscure senators cruelly culled to help it fit on the slide):

$$\mathbf{A} = \begin{array}{c|ccccc} & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} & \textit{John McCain} & \textit{Joe Biden} \\ \textit{Hillary Clinton} & 0 & 3 & 9 & 7 & 12 \\ \textit{Rick Santorum} & 0 & 0 & 6 & 0 & 4 \\ \textit{Joe Lieberman} & 9 & 3 & 0 & 20 & 9 \\ \textit{John McCain} & 1 & 1 & 10 & 0 & 3 \\ \textit{Joe Biden} & 8 & 0 & 2 & 9 & 0 \end{array}$$

- You reviewed over the summer that this is a 5×5 matrix (the names are just labels) and, for instance, we'd depict John McCain's cosponsorship of Hillary Clinton's bills a_{41}, and Hillary's cosponsorship of John McCain's bills as a_{14}
- Now we'll review some other special features of this particular matrix

First special feature: the matrix is square

- If we denote matrix dimensions as $m \times n$:
  - Square: $m = n$, which for whatever reason, we often describe as a $k \times k$ matrix. Another way to describe is that it is *order* k. What order is the matrix from the previous slide?
  - Rectangular: $m \neq n$
- Typical square matrices in social science: matrices to summarize pairwise measures (e.g., our co-sponsorship data; a correlation matrix summarizing correlations between any two variables in a dataset; etc.)
- Typical rectangular matrices in social science: rows = observations, columns = predictor variables (unless your number of observations happens to equal number of covariates)

Application: how to go from rectangular to square using our measure of vector distance

- **Example**: in addition to the cosponsorship data, we have a rectangular 101×2 matrix summarizing each senator's ideology along two dimensions (based on their vote record)
- Viewing this rectangular matrix for our five senators of interest, we have:

	ideology1	ideology2
Biden Joseph R. Jr.	-0.335	0.016
Clinton Hillary Rodham	-0.344	0.017
Lieberman Joseph I.	-0.219	-0.130
McCain John	0.298	-0.445
Santorum Rick	0.322	-0.263

- We can transform into a 5×5 square matrix of ideological distance for each senator pair by going pair by using our measure of euclidean distance on each pair of senators:

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{\mathbf{u} \bullet \mathbf{u} - 2(\mathbf{u} \bullet \mathbf{v}) + \mathbf{v} \bullet \mathbf{v}}$$

Application: how to go from rectangular to square using our measure of vector distance

- On your worksheet, using the below values (which you can find by subsetting the ideologydf matrix), find the ideological distance (difference norm/euclidean, defined on previous slide) between Joe Biden and John McCain

	ideology1	ideology2
Biden Joseph R. Jr.	-0.335	0.016
Clinton Hillary Rodham	-0.344	0.017
Lieberman Joseph I.	-0.219	-0.130
McCain John	0.298	-0.445
Santorum Rick	0.322	-0.263

- Use the following R command to calculate the ideological distance for all pairs of senators and confirm $A_{Biden,McCain}$ is equivalent to your hand calculation, where *ideomat* is the ideology matrix subsetted to only include those 5 senators:

$$\text{as.matrix(dist(ideomat, method = "euclidean"))}$$

Answer: hand calculation of distance between McCain and Biden

1. Let $McCain = \mathbf{u} = \begin{bmatrix} 0.298 & -0.445 \end{bmatrix}$, and $Biden = \mathbf{v} = \begin{bmatrix} -0.335 & 0.016 \end{bmatrix}$
2. Separate $\|\mathbf{u} - \mathbf{v}\| = \sqrt{\mathbf{u} \bullet \mathbf{u} - 2(\mathbf{u} \bullet \mathbf{v}) + \mathbf{v} \bullet \mathbf{v}}$ into sub-components and calculate each:

$$\begin{aligned}
\mathbf{u} \bullet \mathbf{u} &= 0.298^2 + (-0.445)^2 = 0.286829 \\
\mathbf{u} \bullet \mathbf{v} &= 0.298 * -0.335 + (-0.455 * 0.016) = -0.10695 \\
\mathbf{v} \bullet \mathbf{v} &= -0.335^2 + 0.016^2 = 0.112481
\end{aligned}$$

3. Combine and take the square root:

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{0.286829 + 2 * -0.10695 + 0.112481} = 0.783$$

Entire square ideological distance matrix, confirms our hand calculation is correct!

	Hillary Clinton	*Rick Santorum*	*Joe Lieberman*	*John McCain*	*Joe Biden*
Hillary Clinton	0.000	0.722	0.193	0.791	0.009
Rick Santorum	0.722	0.000	0.557	0.184	0.714
Joe Lieberman	0.193	0.557	0.000	0.605	0.186
John McCain	0.791	0.184	0.605	0.000	0.783
Joe Biden	0.009	0.714	0.186	0.783	0.000

Now we have two square matrices, what other special properties do we notice about each of these matrices?

Note: these properties are dependent upon the matrix being square. In other words, being square is necessary but not sufficient for...

$\mathbf{A} =$

	Hillary Clinton	*Rick Santorum*	*Joe Lieberman*	*John McCain*	*Joe Biden*
Hillary Clinton	0	3	9	7	12
Rick Santorum	0	0	6	0	4
Joe Lieberman	9	3	0	20	9
John McCain	1	1	10	0	3
Joe Biden	8	0	2	9	0

$\mathbf{X} =$

	Hillary Clinton	*Rick Santorum*	*Joe Lieberman*	*John McCain*	*Joe Biden*
Hillary Clinton	0.000	0.722	0.193	0.791	0.009
Rick Santorum	0.722	0.000	0.557	0.184	0.714
Joe Lieberman	0.193	0.557	0.000	0.605	0.186
John McCain	0.791	0.184	0.605	0.000	0.783
Joe Biden	0.009	0.714	0.186	0.783	0.000

Symmetric matrix

- Focusing on the ideology matrix, the way we defined distance meant that $distance_{McCain,Biden} = distance_{Biden,McCain}$
- Since we defined distance similarly for every pair, the matrix is symmetric ($a_{ij} = a_{ji}$ for all i, j), which informally means that if you split the matrix in two along the diagonal (bolded below), the two halves are mirror images (which when we discuss transposes, means, $\mathbf{X}^T = \mathbf{X}$):

$$
\mathbf{X} = \begin{array}{r c c c c c}
 & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} & \textit{John McCain} & \textit{Joe Biden} \\
\textit{Hillary Clinton} & \mathbf{0.000} & 0.722 & 0.193 & 0.791 & 0.009 \\
\textit{Rick Santorum} & 0.722 & \mathbf{0.000} & 0.557 & 0.184 & 0.714 \\
\textit{Joe Lieberman} & 0.193 & 0.557 & \mathbf{0.000} & 0.605 & 0.186 \\
\textit{John McCain} & 0.791 & 0.184 & 0.605 & \mathbf{0.000} & 0.783 \\
\textit{Joe Biden} & 0.009 & 0.714 & 0.186 & 0.783 & \mathbf{0.000}
\end{array}
$$

Symmetrix matrix

To discuss: is the cosponsorship matrix a symmetric matrix? If not, what's one senator pair who are making the matrix non-symmetric?

$$
\mathbf{A} = \begin{array}{r|ccccc}
 & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} & \textit{John McCain} & \textit{Joe Biden} \\
\textit{Hillary Clinton} & 0 & 3 & 9 & 7 & 12 \\
\textit{Rick Santorum} & 0 & 0 & 6 & 0 & 4 \\
\textit{Joe Lieberman} & 9 & 3 & 0 & 20 & 9 \\
\textit{John McCain} & 1 & 1 & 10 & 0 & 3 \\
\textit{Joe Biden} & 8 & 0 & 2 & 9 & 0
\end{array}
$$

Diagonal matrix

- If a matrix is symmetric *and* square, it may also be a diagonal matrix, where $a_{ij} = 0$ for all $i \neq j$
- For instance, if we were to form a square matrix that simply represented each senator's age (so a non pairwise measure), it might look like (ages made up):

$$\mathbf{A} = \begin{array}{c|ccccc} & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} & \textit{John McCain} & \textit{Joe Biden} \\ \textit{Hillary Clinton} & 68 & 0 & 0 & 0 & 0 \\ \textit{Rick Santorum} & 0 & 50 & 0 & 0 & 0 \\ \textit{Joe Lieberman} & 0 & 0 & 65 & 0 & 0 \\ \textit{John McCain} & 0 & 0 & 0 & 67 & 0 \\ \textit{Joe Biden} & 0 & 0 & 0 & 0 & 63 \end{array}$$

Special diagonal matrix: identity matrix

- If a matrix is symmetric *and* square *and* diagonal, it may also be an identity matrix, where $a_{ij} = 0$ for all $i \neq j$ and $a_{ij} = 1$ for all $i = j$
- More formally, the notation for the identity matrix is $\mathbf{I}_n$
- In this context, $\mathbf{I}_5$:

$$\mathbf{A} = \begin{array}{r|ccccc} & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} & \textit{John McCain} & \textit{Joe Biden} \\ \hline \textit{Hillary Clinton} & 1 & 0 & 0 & 0 & 0 \\ \textit{Rick Santorum} & 0 & 1 & 0 & 0 & 0 \\ \textit{Joe Lieberman} & 0 & 0 & 1 & 0 & 0 \\ \textit{John McCain} & 0 & 0 & 0 & 1 & 0 \\ \textit{Joe Biden} & 0 & 0 & 0 & 0 & 1 \end{array}$$

Summing up thus far and where we're going next

Note: *italicized* nodes on the tree are where the nodes represent special cases of the matrix in question rather than an exhaustive set of cases. So for instance, all square matrices are either symmetric or non-symmetric, but there are symmetric matrices that are *not* diagonal matrices

- Matrix
  - Square
    - Symmetric
      - *Diagonal*
        - *Identity matrix*
    - Non-symmetric
  - Rectangular

Focusing on our case of a symmetric matrix (ideological distance matrix), we could delete all redundant elements and rewrite as:

$$\mathbf{X} = \begin{array}{r|ccccc} & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} & \textit{John McCain} & \textit{Joe Biden} \\ \hline \textit{Hillary Clinton} & \mathbf{0.000} & 0.722 & 0.193 & 0.791 & 0.009 \\ \textit{Rick Santorum} & 0.722 & \mathbf{0.000} & 0.557 & 0.184 & 0.714 \\ \textit{Joe Lieberman} & 0.193 & 0.557 & \mathbf{0.000} & 0.605 & 0.186 \\ \textit{John McCain} & 0.791 & 0.184 & 0.605 & \mathbf{0.000} & 0.783 \\ \textit{Joe Biden} & 0.009 & 0.714 & 0.186 & 0.783 & \mathbf{0.000} \end{array}$$

Becomes:

$$\mathbf{X}_2 = \begin{array}{r|ccccc} \textit{Hillary Clinton} & \mathbf{0.000} & 0 & 0 & 0 & 0 \\ \textit{Rick Santorum} & 0.722 & \mathbf{0.000} & 0 & 0 & 0 \\ \textit{Joe Lieberman} & 0.193 & 0.557 & \mathbf{0.000} & 0 & 0 \\ \textit{John McCain} & 0.791 & 0.184 & 0.605 & \mathbf{0.000} & 0 \\ \textit{Joe Biden} & 0.009 & 0.714 & 0.186 & 0.783 & \mathbf{0.000} \end{array}$$

Triangular matrix

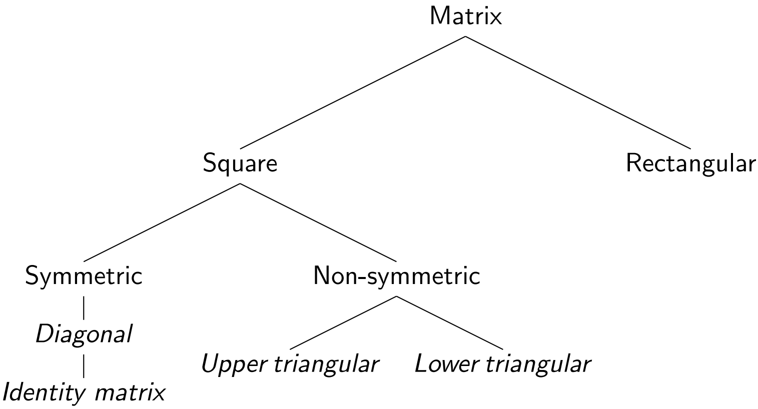
- The matrix is no longer symmetric, but it's what's called a *lower triangular* matrix where all the entries above the diagonal are zero (formally: $a_{ij} = 0$ for $j > i$):

$$\mathbf{X}_2 = \begin{array}{r|ccccc} \textit{Hillary Clinton} & \mathbf{0.000} & 0 & 0 & 0 & 0 \\ \textit{Rick Santorum} & 0.722 & \mathbf{0.000} & 0 & 0 & 0 \\ \textit{Joe Lieberman} & 0.193 & 0.557 & \mathbf{0.000} & 0 & 0 \\ \textit{John McCain} & 0.791 & 0.184 & 0.605 & \mathbf{0.000} & 0 \\ \textit{Joe Biden} & 0.009 & 0.714 & 0.186 & 0.783 & \mathbf{0.000} \end{array}$$

- An *upper triangular* matrix is where all entries below the diagonal are zero (formally: $a_{ij} = 0$ for $j < i$):

$$\mathbf{X}_3 = \begin{array}{ccccc} \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} & \textit{John McCain} & \textit{Joe Biden} \\ \hline \mathbf{0.000} & 0.722 & 0.193 & 0.791 & 0.009 \\ 0 & \mathbf{0.000} & 0.557 & 0.184 & 0.714 \\ 0 & 0 & \mathbf{0.000} & 0.605 & 0.186 \\ 0 & 0 & 0 & \mathbf{0.000} & 0.783 \\ 0 & 0 & 0 & 0 & \mathbf{0.000} \end{array}$$

Updated tree



Summing up thus far and where we're going next

- We're reviewed a taxonomy of special matrix types: is a matrix square? If so, is it symmetric? If so, is it a diagonal matrix? If so, is it an identity matrix?
- These might seem abstract for now, but they come up, for instance, in matrix-based approach to linear regression and in matrix decomposition (how to represent matrix $\mathbf{A}$ as the product of other matrices)
  - For instance, if a matrix is square, we can represent it as the product of a upper and lower triangular matrix
- In addition, the matrix algebra and summary operations we review next are mechanically easier for certain special matrices
  - For instance, the determinant of a diagonal matrix is just the product of the diagonal elements

Overview

- Basic operations: addition/subtraction, scalar multiplication
- More complex operations: multiplication, transpose
- For each, we'll cover:
  - Motivation/preview of applications
  - Mechanics: what counts as conformable matrices for the purposes of the operation and how to perform
  - Practice with our cosponsorship or ideology matrices

Addition/subtraction: motivation

- **Example**: you want to conduct a time-series analysis of senate cosponsorship patterns that investigates factors that explain a senator's *deviation* from his or her average cosponsorship patterns (e.g., do his patterns change after a close election relative to his own average across all years?)
- Represent as ($\bar{A}$ made up), with a smaller subset: $\mathbf{A}_{closeelect} - \bar{\mathbf{A}} = \tilde{\mathbf{A}}$

$$\mathbf{A}_{closeelect} = \begin{array}{r|ccc} & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} \\ \hline \textit{Hillary Clinton} & 0 & 3 & 9 \\ \textit{Rick Santorum} & 0 & 0 & 6 \\ \textit{Joe Lieberman} & 9 & 3 & 0 \end{array}$$

$$\bar{\mathbf{A}} = \begin{array}{r|ccc} & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} \\ \hline \textit{Hillary Clinton} & 0 & 2 & 12 \\ \textit{Rick Santorum} & 0 & 0 & 3 \\ \textit{Joe Lieberman} & 7 & 5 & 0 \end{array}$$

Addition/subtraction: mechanics and application

- **Conformable in this case**: matrices have exact same dimensions (so in this case, both 3×3)
- How to do: element by element addition/subtraction

$$\tilde{\mathbf{A}} = \mathbf{A}_{closeelect} - \bar{\mathbf{A}} = \begin{array}{cccc} & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} \\ \textit{Hillary Clinton} & 0 & 3-2 & 9-12 \\ \textit{Rick Santorum} & 0 & 0 & 6-3 \\ \textit{Joe Lieberman} & 9-7 & 3-5 & 0 \end{array}$$

Multiplying by a scalar: motivation, mechanics, application

- Similar motivation as in vector case: can simultaneously rescale all the elements by some constant
- **Conformable in this case**: since scalar is applied to every element of the matrix, works for a matrix of any dimension
- Example: rescale cosponsorship by maximum cosponsorship value (multiply by $\frac{1}{max(\mathbf{A})} = \frac{1}{20}$) to constrain to be between 0 and 1, show example mechanics for first line:

	Hillary Clinton	*Rick Santorum*	*Joe Lieberman*	*John McCain*	*Joe Biden*
Hillary Clinton	0.00	$\frac{1}{20} * 3 = 0.15$	$\frac{1}{20} * 9 = 0.45$	$\frac{1}{20} * 7 = 0.35$	$\frac{1}{20} * 12 = 0.60$
Rick Santorum	0.00	0.00	0.30	0.00	0.20
Joe Lieberman	0.45	0.15	0.00	1.00	0.45
John McCain	0.05	0.05	0.50	0.00	0.15
Joe Biden	0.40	0.00	0.10	0.45	0.00

Matrix multiplication: motivation

- Many applications- important one is linear regression (learn much more in Soc 500), where we can begin with typical way of writing the regression equation, and rewrite using matrices and vectors:

$$Y = \beta_0 + \beta_1 X_1 \ldots \beta_n X_n$$

- How can we begin to write the following using vectors, matrices, and matrix multiplication? (we'll pretend there are four observations)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Brief side note: example of linear regression you might run with our cosponsorship data

- Outcome variable (Y): number of bills a senator cosponsors
- We have intuitions about what might explain variation between senators in Y:

$$Y = \beta_0 + ideology * \beta_1 + tenure * \beta_2 + donations * \beta_3$$

- Linear regression: which set of weights gets us closest to the observed cosponsorship count? Where the weights can be any set of linear combinations in $\mathbb{R}^n$, so they might be:

Option one:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 0.5 \\ 0.8 \\ 0.9 \end{bmatrix}$$

Option two:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0.1 \\ 0.001 \\ 5 \end{bmatrix}$$

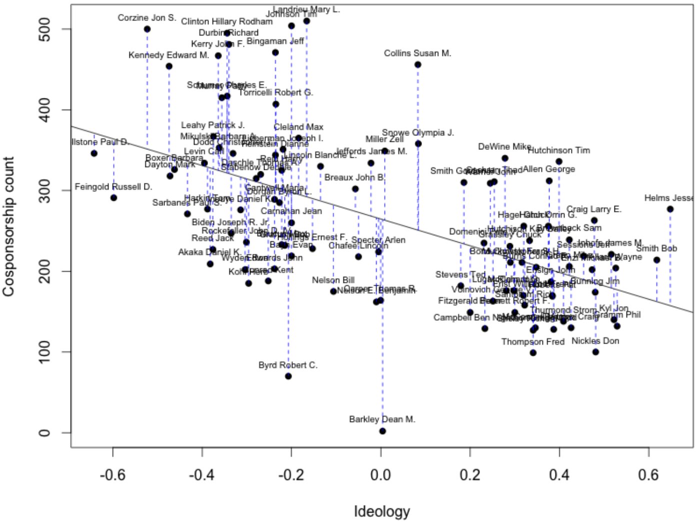
Option three:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0.4 \\ 2.3 \\ 4.7 \end{bmatrix}$$

The game: single variable case

We might be familiar, and if not, you'll learn this year, with how to adjudicate between these options (or a two-element version of these options) in the univariate case (so where we only need to find: $\beta_0 =$ weight on intercept and $\beta_1 =$ weight on predictor) by using a "best fit" line:

The game: single variable case



Once we move into more than one predictor variable, the geometric representation gets more complicated but highlights important properties of matrix and vector algebra

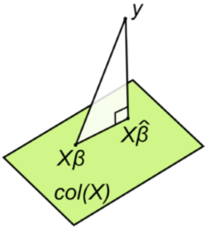
- Let's start by representing our variables of interest, along with an intercept term, in vector form, pretending we only have four observations:

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \textbf{ideology} = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ x_{41} \end{bmatrix}, \ \textbf{tenure} = \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \\ x_{42} \end{bmatrix}, \ \textbf{donations} = \begin{bmatrix} x_{13} \\ x_{23} \\ x_{33} \\ x_{43} \end{bmatrix}$$

- Again, we're interested in which weight we should multiply each vector by (four vectors...four weights) to get closest to the observed cosponsorship count

Visualization of the new game in the multivariate case

- How do we find a set of $\hat{\beta}$ (remember: options 1, 2, 3, etc... for weights on the ideology, tenure, and donations variables) that get us closest to the observed cosponsorship count (y)?
- The requirements of the game:
  - We can't leave what's called the *column space* of $\mathbf{X}$, which is the green space on the graphic formed by:
    1. Our collection of predictor vectors: intercept, ideology, tenure, and donations
    2. All combinations of weights we might place on those predictors vectors
  - What gets us closest to y (the observed sponsorship count) without leaving that column space?



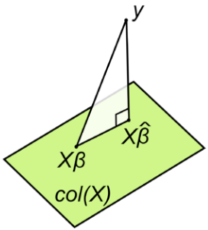
Source: stackexchange

Visualization of the new game in the multivariable case

- The way we can get as close as possible while still remaining on the column space composed of our predictors and combinations of different weights is by finding the vector that is perpendicular/orthogonal to the column space, so the vector where:

$$X^T(Y - X\hat{\beta}) = 0$$

- If we tried any vector other than the one perpendicular to our observed cosponsorship count of y (so moved elsewhere along the column space), we would be further from that observed cosponsorship count, which we want to avoid



Source: stackexchange

Now that we've visualized what we need to do to get as close as possible to the cosponsorship values, what tools involving vectors and matrices do we need in order to solve for the set of weights ($\hat{\beta}$)?

$$X^T(Y - X\hat{\beta}) = 0$$

1. Matrix multiplication, and tools that assist us with multiplication (how to check conformability, how to transpose):

$$X^T(Y - X\hat{\beta}) = 0$$

2. How to invert a matrix (and check whether it can be inverted in the first place): to help us get $\hat{\beta}$ on one side of the equation, we're going to need to do some division which in the matrix case, means multiplying by the inverse
3. Assorted other tools (e.g., whether given our matrix of predictors, we can even solve for $\hat{\beta}$ in the first place, which gets into matrix rank)

Matrix multiplication: we can start by writing out the vectors we need...

$$Y = \beta_0 + \beta_1 ideology + \beta_2 tenure + \beta_3 donations$$

1. LHS: $\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$
2. RHS:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mathbf{ideology} = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ x_{41} \end{bmatrix} \mathbf{tenure} = \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \\ x_{42} \end{bmatrix} \mathbf{donations} = \begin{bmatrix} x_{13} \\ x_{23} \\ x_{33} \\ x_{43} \end{bmatrix}$$

Then, we can combine the four covariate vectors into a matrix by placing them side by side

- Four vectors:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \textbf{ideology} = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ x_{41} \end{bmatrix} \quad \textbf{tenure} = \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \\ x_{42} \end{bmatrix} \quad \textbf{donations} = \begin{bmatrix} x_{13} \\ x_{23} \\ x_{33} \\ x_{43} \end{bmatrix}$$

- Become a 4×4 matrix:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ 1 & x_{41} & x_{42} & x_{43} \end{bmatrix}$$

- Rewriting the original expression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Becomes:

$$Y = X\beta$$

Now, focusing on the RHS, we have the following:

- What we have, with dimensions indicated:

$$\underset{4\times 1}{\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}, \quad \mathbf{X} = \underset{4\times 4}{\begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ 1 & x_{41} & x_{42} & x_{43} \end{bmatrix}}$$

- Which order of multiplication is permissible with a vector and matrix with these dimensions?

Matrix multiplication mechanics: checking conformability

Conformable for multiplication: number of columns in first matrix must equal number of rows in second matrix, so:

- Not conformable:

$$\beta = \underset{4\times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}, \quad \mathbf{X} = \underset{4\times 3}{\begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ 1 & x_{41} & x_{42} & x_{43} \end{bmatrix}}$$

- Conformable (here, switched order; soon, we'll discuss transposing a matrix):

$$\mathbf{X}\beta = \underset{4\times 3}{\begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{bmatrix}} \underset{3\times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}$$

- Result: 4×1 matrix (dimensions are drawn from rows in the first matrix, columns in second)

Now that we have a conformable matrix and vector, how do we multiply?

- Example with two columns, result will be a 4×2 matrix:

$$\mathbf{X}\beta = \underset{4\times 3}{\begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{bmatrix}} \underset{3\times 2}{\begin{bmatrix} \beta_0 & \gamma_0 \\ \beta_1 & \gamma_1 \\ \beta_2 & \gamma_2 \end{bmatrix}}$$

- Formally, we take the dot product of each row vector from the first matrix and column vector from the second matrix (red = changes from row to row of results):

$$\mathbf{X}\beta = \underset{4\times 2}{\begin{bmatrix} (1 \quad x_{11} \quad x_{12}) \bullet (\beta_0 \quad \beta_1 \quad \beta_2) & (1 \quad x_{11} \quad x_{12}) \bullet (\gamma_0 \quad \gamma_1 \quad \gamma_2) \\ (1 \quad x_{21} \quad x_{22}) \bullet (\beta_0 \quad \beta_1 \quad \beta_2) & (1 \quad x_{21} \quad x_{22}) \bullet (\gamma_0 \quad \gamma_1 \quad \gamma_2) \\ (1 \quad x_{31} \quad x_{32}) \bullet (\beta_0 \quad \beta_1 \quad \beta_2) & (1 \quad x_{31} \quad x_{32}) \bullet (\gamma_0 \quad \gamma_1 \quad \gamma_2) \\ (1 \quad x_{41} \quad x_{42}) \bullet (\beta_0 \quad \beta_1 \quad \beta_2) & (1 \quad x_{41} \quad x_{42}) \bullet (\gamma_0 \quad \gamma_1 \quad \gamma_2) \end{bmatrix}}$$

- Informally, I 1) draw out a shell matrix with correct dimensions for results; 2) circle rows in first, columns in second and proceed (illustrated on board)

Matrix multiplication: some properties

Note: these properties assume the matrices are conformable so that should always be your first step! Full list is on Gill page 112.

- Associative: $(XY)Z = X(YZ)$
- Distributive for addition: $(X + Y)Z = XZ + YZ$
- What's missing from above?
  - **NOT COMMUTATIVE FOR MULTIPLICATION**: except in certain cases, $XY \neq YX$
  - One of those cases where multiplication is commutative: identity matrix $\times$ any other (conformable) matrix. Example:

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 2 & 3 \\ 4 & 2 \end{bmatrix}$$

  - On your worksheet, use matrix multiplication to confirm that $I_2X = XI_2$. What do the results highlight about multiplication by an identity matrix?

Matrix multiplication: more practice

1. For the following matrices:

$$\mathbf{Y} = \begin{bmatrix} 3 & 1 & -2 \\ 6 & 3 & 4 \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} 4 & 2 \\ 3 & 0 \\ 1 & 2 \end{bmatrix}$$

   1.1 Write out dimensions of each
   1.2 Arrange multiplication in a way that makes matrices conformable to multiply
   1.3 Multiply by hand

2. Confirm your results using the following command in R, with mat1 and mat2 referring to first and second matrices. You can use the matrix command to create these matrices:

```
mat1 %*% mat2
```

Answers

$$\mathbf{Y} = \begin{bmatrix} 3 & 1 & -2 \\ 6 & 3 & 4 \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} 4 & 2 \\ 3 & 0 \\ 1 & 2 \end{bmatrix}$$

1. Write out dimensions of each: $\mathbf{X} = 3 \times 2$; $\mathbf{Y} = 2 \times 3$
2. Arrange multiplication in a way that makes matrices conformable to multiply: order that makes conformable, and will result in a 3 x 3 matrix:

$$\mathbf{X} = \begin{bmatrix} 4 & 2 \\ 3 & 0 \\ 1 & 2 \end{bmatrix} \mathbf{Y} = \begin{bmatrix} 3 & 1 & -2 \\ 6 & 3 & 4 \end{bmatrix},$$

3. Multiply by hand:

$$\begin{bmatrix} 12+12 & 4+6 & -8+8 \\ 9 & 3 & -6 \\ 3+12 & 1+6 & -2+8 \end{bmatrix}$$

Where we're going next

- We've emphasized the importance of checking to make sure matrices are conformable before matrix multiplication
- One way we made conformable: switching around order of multiplication
- But what about the following case, which incidentally is the total count of bills a senator cosponsored ($\mathbf{Y}$) and the two measures of senator ideology along with an intercept term ($\mathbf{X}$):

$$\mathbf{Y} = \begin{bmatrix} HRC \\ RS \\ JL \\ JM \\ JB \end{bmatrix} = \underset{5\times 1}{\begin{bmatrix} 31 \\ 10 \\ 41 \\ 15 \\ 19 \end{bmatrix}}, \quad \mathbf{X} = \underset{5\times 3}{\begin{bmatrix} 1 & -0.34 & 0.02 \\ 1 & -0.34 & 0.02 \\ 1 & -0.22 & -0.13 \\ 1 & 0.30 & -0.45 \\ 1 & 0.32 & -0.26 \end{bmatrix}}$$

- $Y_{5\times 1}X_{5\times 3}$ is not conformable
- $X_{5\times 3}Y_{5\times 1}$ is also not conformable

Transpose: motivation

- **Transpose**: we've already reviewed with vectors, but with matrices, written as $\mathbf{A}^T$ or $\mathbf{A}'$ and just means switching the rows and columns
- We've run into the situation where we want to multiply two matrices:
  1. $\mathbf{X}$: 5×3
  2. $\mathbf{Y}$: 5×1
- Can use transpose to 1) get the inner dimensions to match in a way that makes the matrices conformable; 2) produces a results matrix with dimensions that are appropriate for the question at hand
- Dimensions after transposing:
  1. $\mathbf{X}^T : 3 \times 5$
  2. $\mathbf{Y} : 5 \times 1$
  3. Check conformability– we're good!: $(3 \times 5)(5 \times 1)$
  4. Dimensions after multiplying: $\mathbf{X}^T\mathbf{Y} : 3 \times 1$

Transpose: mechanics

Just flip: Visual depiction of $\mathbf{X}^T$. To do in R: t(matrix).

$$\underset{5\times 3}{\mathbf{X}} = \begin{bmatrix} 1 & -0.34 & 0.02 \\ 1 & -0.34 & 0.02 \\ 1 & -0.22 & -0.13 \\ 1 & 0.30 & -0.45 \\ 1 & 0.32 & -0.26 \end{bmatrix}$$

$$\underset{3\times 5}{\mathbf{X}^T} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -0.34 & -0.34 & -0.22 & 0.30 & 0.32 \\ 0.02 & 0.02 & -0.13 & -0.45 & -0.26 \end{bmatrix}$$

Important properties of transpose

These are invaluable for matrix-based proofs of regression properties (properties listed on Gill page 116):

- Invertibility (or: if you transpose a transposed matrix, you get back the original): $(\mathbf{X}^T)^T = \mathbf{X}$
- Additive (or: can distribute a transpose to addition without switching order since matrix addition is commutative): $(\mathbf{X}+\mathbf{Y})^T = \mathbf{X}^T + \mathbf{Y}^T = \mathbf{Y}^T + \mathbf{X}^T$
- Multiplicative (or: can distribute a transpose to matrix multiplication but need to switch the order): $(\mathbf{XYZ})^T = \mathbf{Z}^T\mathbf{Y}^T\mathbf{X}^T$
- For a symmetric matrix (e.g., our ideological distance one): $\mathbf{X}^T = \mathbf{X}$

Practice with transpose properties

On your worksheet, for the matrices with the dimensions below, write

1. The dimensions of the $Y - X\beta$. Hint: what are the dimensions of $X\beta$ and then what are the dimensions of Y minus that result?
2. Given those dimensions, how would you would use transpose to make the following multiplication 1) conformable, 2) produce a 1×1 result?: $(Y - X\beta)(Y - X\beta)$
3. After step two, if it involves transposing one or both of the $Y - X\beta$, how would those transposes be distributed using the properties on the previous slide (we can flip back),

Matrices:

$$\mathbf{X} = \underset{5\times 3}{\begin{bmatrix} x_{11} & \dots & x_{31} \\ \vdots & & \vdots \\ x_{51} & \dots & x_{53} \end{bmatrix}}, \quad \mathbf{Y} = \underset{5\times 1}{\begin{bmatrix} y_{11} \\ \vdots \\ y_{51} \end{bmatrix}}, \quad \beta = \underset{3\times 1}{\begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{31} \end{bmatrix}}$$

Answers

1. Since $X\beta = (5 \times 3)(3 \times 1) = 5 \times 1$ (outer dimensions), we can do $Y - X\beta$, and since subtraction doesn't change dimensions, result is 5×1
2. We want to be able to multiply: $(5 \times 1)(5 \times 1)$. If we transpose the first matrix, we multiply: $(1 \times 5)(5 \times 1) = (1 \times 1)$. If we transpose the second matrix, we multiply: $(5 \times 1)(1 \times 5) = (5 \times 5)$. Therefore, we want to transpose the first matrix to get the 1×1 results we want.
3. $(Y - X\beta)^T$. Can break into steps:
   3.1 Distribute the transpose across subtraction (no reordering needed):
   $$Y^T - (X\beta)^T$$
   3.2 Distribute the transpose across multiplication (need to reorder):
   $$Y^T - \beta^T X^T$$
   3.3 You might notice that the result is not conformable for that expression alone, but it would be if we then multiply by $Y - X\beta$ and distribute (we can try as a group if enough time)

Summing up thus far and where we're going next

- We've reviewed types of matrices and basic and more complex matrix operations, some of which are simplified if the matrices are of special types (e.g., $\mathbf{X}^T = \mathbf{X}$ for symmetric matrix)
- We've also, sometimes less explicitly, touched upon various applications:

Summing up thus far and where we're going next

- *Residuals using vector subtraction*: one way to represent residuals after we fit the regression (observed - fitted) relies on vector subtraction and the vectors being of the same dimension (below, $n =$ number of observations in data):

$$\underset{n \times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} - \underset{n \times 1}{\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}}$$

- *Dyadic measures of distance*: to turn a single variable (e.g., a set of responses on a scale) into a dyadic variable (e.g., the distance between two respondents' responses), can calculate Euclidean distance as: $\|\mathbf{u} - \mathbf{v}\| = \sqrt{\mathbf{u} \bullet \mathbf{u} - 2(\mathbf{u} \bullet \mathbf{v}) + \mathbf{v} \bullet \mathbf{v}}$
- *Beginning tools for proof of linear regression!*: will learn more in SOC 500, but we know colloquially that the purpose of linear regression is to minimize the sum of squared errors $(Y - X\beta)^T(Y - X\beta)$. On previous slide, we reviewed steps for checking conformability and distributing the transpose

Where we're going next: motivation

- We learned about how to do various *operations* with matrices
- But there are also important summary-level measures that help succinctly describe the structure of a matrix
- Oftentimes, you'll learn these summary measures in the context of manually solving systems of linear equations
- Since this application is a bit outdated, we'll show other reasons why these summary measures are important before moving on to mechanics of how to find. These summary measures are:
- What is the matrix's rank?
  - *How is this useful?* Helps us assess the number of unique solutions to a system of linear equations
- Can we invert the matrix? (also known as: is the matrix nonsingular)? If so, what is the inverse?
  - *How is this useful?* when solving an equation involving matrices, we can't just divide a matrix by another matrix in the normal sense (e.g., $\frac{\mathbf{X}}{\mathbf{Y}}$). Instead, we multiply by the inverse: $\mathbf{X}Y^{-1}$. Likewise, to divide a matrix by itself to get the identity matrix, we multiply a matrix by its own inverse: $\mathbf{X}\mathbf{X}^{-1} = \mathbf{I}$

Matrix rank: motivation

- Basic idea: how much unique information a matrix holds
- Imagine two versions of the senator cosponsorship matrix, the original:

$$\mathbf{A} = \begin{array}{r|ccccc} & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} & \textit{John McCain} & \textit{Joe Biden} \\ \hline \textit{Hillary Clinton} & 0 & 3 & 9 & 7 & 12 \\ \textit{Rick Santorum} & 0 & 0 & 6 & 0 & 4 \\ \textit{Joe Lieberman} & 9 & 3 & 0 & 20 & 9 \\ \textit{John McCain} & 1 & 1 & 10 & 0 & 3 \\ \textit{Joe Biden} & 8 & 0 & 2 & 9 & 0 \end{array}$$

- And a modified version, where Joe Biden, hoping to curry favor with Hillary Clinton, decides to take any senator who she cosponsors a bill with and cosponsor with them twice as many times (ignore the inconsistency in terms of Biden cosponsoring with himself)

$$\mathbf{A}_{shortrank} = \begin{array}{r|ccccc} & \textit{Hillary Clinton} & \textit{Rick Santorum} & \textit{Joe Lieberman} & \textit{John McCain} & \textit{Joe Biden} \\ \hline \textit{Hillary Clinton} & 0 & 3 & 9 & 7 & 12 \\ \textit{Rick Santorum} & 0 & 0 & 6 & 0 & 4 \\ \textit{Joe Lieberman} & 9 & 3 & 0 & 20 & 9 \\ \textit{John McCain} & 1 & 1 & 10 & 0 & 3 \\ \textit{Joe Biden} & 0 & 6 & 18 & 14 & 24 \end{array}$$

Matrix rank, motivation

- The original matrix is what we call **full rank**: none of its rows or columns are linearly dependent upon another column (To discuss, how could we use vector multiplication to assess linear dependence?)
- The modified matrix is what we call not full rank or short rank, since we can take the Hillary row, multiply it by a scalar (in this case 2) and end up with the Biden row of the matrix (aka, the two rows are linearly dependent)
- If *either* any rows *or* any columns are linearly dependent, the matrix is not full rank- more broadly, we get excited because based on its dimensions in this case, we think the cosponsorship matrix will provide five rows worth of unique information, when in reality, it's only providing four rows of unique information

Matrix rank, mechanics

- By hand (happy to discuss more details if interest):
  1. Form what's called an augmented matrix where we place our matrix of interest side by side with an identity matrix:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 7 & 6 & 4 & 0 & 1 & 0 \\ 2 & 3 & 4 & 0 & 0 & 1 \end{array}\right]$$

  2. Perform elementary row operations (multiply any row by a scalar, switch any two rows, combine those first two) to try to get the left hand side into identity matrix form
  3. If those row operations result in any rows in the LHS matrix composed of all zero's, the matrix is not full rank and the number of non-zero rows = the rank of the matrix
- In R: 1) use the rank command, 2) can write a function to iterate through the vectors, take the cross product, and check if any of the vectors cross products equal zero (indicating linear dependence)

Matrix inverse: motivation

- Know colloquially that in linear regression, we want to minimize the sum of squared residuals (so minimize: $(Y - x\beta)^T(Y - X\beta)$
- You'll learn how to find $\hat{\beta}$ (the value that minimizes this expression) later, but for now, we can recognize that after we find $\hat{\beta}$, we also have a new quantity, $\hat{Y}$, distinct from the observed Y
- Getting from our observed covariate matrix X and our observed outcome vector Y to $\hat{Y}$ involves what is called projection, we're projecting the Y vector into a space defined by the columns of our covariate matrix (source: statexchange.com):

- The expression for that projection involves an inverse, so important to know:

$$\hat{Y} = X(X^TX)^{-1}X^TY$$

- Which we also sometimes write as, with P_x called the projection matrix, as:

$$\hat{Y} = P_xY$$

Matrix inverse: mechanics

Two separate questions:

1. Can we invert a matrix? (invertible/nonsingular)
   - *Minimal condition*: a matrix must be square
   - If a matrix is square, it still may not be invertible
2. If yes to 1, what is the matrix's inverse?
   - For a 2×2 matrix $\mathbf{X}$, $\mathbf{X}^{-1} = \frac{1}{det\mathbf{X}}$ (reviewed earlier: when 2×2, and $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $det(A) = ad - bc$)
   - For larger, perform similar process as finding the rank with an augmented matrix (the identity matrix stacked to the right) and elementary row operations- if the left becomes the identity matrix, the right becomes $\mathbf{X}^{-1}$
   - In R, use command: solve(matrix)

Summing up

- **Vectors**
  - Basic notation
  - Multiplying by a scalar
  - Addition and subtraction (sidenote on conformability)
  - Two forms of vector multiplication:
    1. Dot/inner product
    2. Cross product
  - Norms
- **From Vectors to matrices**
  - Typology of matrix types
  - Basic matrix algebra: addition/subtraction, scalar multiplication
  - More complex matrix algebra: matrix multiplication (and transpose to help with multiplication)
  - Summary measures: matrix rank, inverse, determinant
  - Interspersed with the matrix material are preview of two applications:
    1. Transforming rectangular matrices of individual-level measures to square matrices of dyad-level distance measures
    2. Matrix representation of linear regression