

BIOS 7345: Advanced Regression Analysis I

Andrew J. Spieker, Ph.D.

Assistant Professor of Biostatistics
Vanderbilt University Medical Center

Set 1: A quick review of linear algebra

Version: 08/31/2023

Supplementary notes:

- Note: Linear algebra is *shockingly* self-referential. These distilled notes are not a substitute for a good first course in linear algebra; rather, they are intended to serve as a refresher on key concepts, ideas, and results.
- We will not go through everything in these notes, but you should review everything independently.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces
- 4 Matrix inverses
- 5 Very special matrices
- 6 Determinants
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices
- 9 Very important matrix decompositions
- 10 Generalized inverses
- 11 Matrix calculus

NOTATION AND KEY DEFINITIONS

Vectors:

- Typically given by bold, lower-case letters (e.g, \mathbf{x}).
- Almost exclusively *column* vectors:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$$

- Absolute (L_1) norm, denoted by $\|\mathbf{x}\|_1$, given by $\sum_{i=1}^N |x_i|$.
- Euclidean (L_2) norm, denoted by $\|\mathbf{x}\|_2$, given by $\sqrt{\sum_{i=1}^N x_i^2}$.
- The zero-vector is denoted by $\mathbf{0}$.
- Similarly, $\mathbf{1}$ denotes a vector of ones.

NOTATION AND KEY DEFINITIONS

Matrices:

- Typically given by bold, upper-case letters (e.g, **A**).

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{bmatrix} = [\mathbf{a}_{\cdot 1} \quad \mathbf{a}_{\cdot 2} \quad \cdots \quad \mathbf{a}_{\cdot K}] = \begin{bmatrix} \mathbf{a}_{1\cdot} \\ \mathbf{a}_{2\cdot} \\ \vdots \\ \mathbf{a}_{N\cdot} \end{bmatrix}.$$

- a_{ij} denotes the entry in the i^{th} row and j^{th} column of **A**.
- $\mathbf{a}_{i\cdot}$ denotes the i^{th} row vector of **A**.
- $\mathbf{a}_{\cdot j}$ denotes the j^{th} column vector of **A**.
- We say **A** is an $N \times K$ matrix (read “ N by K ,” meaning it has N rows and K columns).

Vectors and matrices: Definitions

- A **matrix** is a rectangular array of numbers.
- A **vector** is a matrix that consists of only one column.
- A **scalar** is a vector that consists of only one element (i.e., a matrix with one row and one column).
- Note: matrix/vector elements will generally be assumed to be finite and real-valued.

Square matrices:

- **A** is said to be *square* if $N = K$.
 - ▶ That is, if it has the same number of rows and columns.
- For instance:

$$\mathbf{A} = [1], \mathbf{B} = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}, \text{ and } \mathbf{C} = \begin{bmatrix} 1 & 2 & -3 \\ -1 & 0 & 0 \\ 0 & -3 & 1 \end{bmatrix}$$

are all examples of square matrices.

NOTATION AND KEY DEFINITIONS

Diagonal matrices:

- A square matrix, \mathbf{A} , is said to be *diagonal* if $a_{ij} = 0$ when $i \neq j$.
 - ▶ That is, if its non-diagonal entries are zero.
- For instance:

$$\mathbf{A} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

is an example of a diagonal matrix.

- Importantly, it is not required for the entries on the diagonals to be non-zero for the matrix to qualify as a diagonal matrix.
- Notation (coding in R, and in this course):
 - ▶ $\text{diag}(\mathbf{A})$: a vector of diagonal entries of matrix \mathbf{A} .
 - ▶ $\text{diag}(\mathbf{x})$: a diagonal matrix with entries of \mathbf{x} along the diagonal.
 - ▶ $\text{diag}(K)$: a $K \times K$ diagonal matrix with ones along the diagonal.

Identity matrices:

- \mathbf{A} is an *identity* matrix if $a_{ij} = 1$ if $i = j$ and 0 if $i \neq j$.
 - ▶ It is a diagonal matrix with ones along the diagonal entries.
- For instance:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is a 3×3 identity matrix.

- Note: We use notation \mathbf{I}_K to denote a $K \times K$ identity matrix.
- Note: If context makes the value of K evident, we simplify the notation by dropping the subscript.

Matrices of ones:

- We use the notation \mathbf{J}_N to denote an $N \times N$ matrix of all ones.
- For instance:

$$\mathbf{J}_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Upper-triangular matrices:

- \mathbf{A} is an *upper-triangular* if $a_{ij} = 0$ for $i > j$.
 - ▶ The entries “below” the diagonal are zero.
- For instance:

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & 3 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

is an example of an upper triangular matrix.

- Importantly, it is not required for the entries in the upper triangle to be non-zero for the matrix to qualify as an upper-triangular matrix.

Lower-triangular matrices:

- \mathbf{A} is a *lower-triangular* matrix if $a_{ij} = 0$ for $i < j$.
 - ▶ The entries “above” the diagonal are zero.
- For instance:

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 2 & -2 & 0 \\ -1 & 3 & 2 \end{bmatrix}$$

is an example of a lower-triangular matrix.

- Importantly, it is not required for the entries in the lower triangle to be non-zero for the matrix to qualify as an lower-triangular matrix.

Matrix addition:

- If **A** and **B** are of the same dimension, they may be added.
 - ▶ To add two matrices, you add element-wise.
- For instance:

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 1 \\ 3 & 0 & -1 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 5 & -4 & 0 \\ -2 & 6 & -2 \end{bmatrix}.$$

Therefore,

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 + 5 & 4 + (-4) & 1 + 0 \\ 3 + (-2) & 0 + 6 & -1 + (-2) \end{bmatrix} = \begin{bmatrix} 6 & 0 & 1 \\ 1 & 6 & -3 \end{bmatrix}.$$

Matrix addition: Properties

- Matrix addition is associative:

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}).$$

- Matrix addition is commutative:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}.$$

Scalar multiplication:

- If \mathbf{A} is any matrix, it can be multiplied by a scalar.
 - ▶ To multiply a matrix by a scalar, multiply element-wise.
- For instance:

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 1 \\ 3 & 0 & -1 \end{bmatrix} \text{ and } c = 2.$$

Therefore,

$$c\mathbf{A} = \begin{bmatrix} 2 \times 1 & 2 \times 4 & 2 \times 1 \\ 2 \times 3 & 2 \times 0 & 2 \times (-1) \end{bmatrix} = \begin{bmatrix} 2 & 8 & 2 \\ 6 & 0 & -2 \end{bmatrix}.$$

The dot product:

- If \mathbf{x} and \mathbf{y} are of length K , then $\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^K x_k y_k$.
 - ▶ The dot product is the sum of the element-wise products (well defined if vectors are of the same length).
- For instance:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix}.$$

Therefore,

$$\mathbf{x} \cdot \mathbf{y} = (1 \times 5) + (4 \times -4) + (1 \times 0) = -11.$$

- Note: $\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta$, where θ marks the angle between \mathbf{x} and \mathbf{y} .
- If \mathbf{x} and \mathbf{y} are orthogonal vectors, then $\mathbf{x} \cdot \mathbf{y} = 0$.

Matrix transposition:

- If \mathbf{A} is an $N \times K$ matrix, its *transpose*, \mathbf{A}^T is a $K \times N$ matrix that reverses the role of the rows and columns of \mathbf{A} .
 - ▶ The (i, j) entry of \mathbf{A} is the (j, i) entry of \mathbf{A}^T .
- For instance:

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 1 \\ 3 & 0 & -1 \end{bmatrix} \implies \mathbf{A}^T = \begin{bmatrix} 1 & 3 \\ 4 & 0 \\ 1 & -1 \end{bmatrix}.$$

Matrix transposition: Properties

- $(\mathbf{A}^T)^T = \mathbf{A}$.
- $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.
- $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$.
- $(c\mathbf{A})^T = c\mathbf{A}^T$.
- $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}$.
- $\mathbf{x} \cdot \mathbf{x} = \mathbf{x}^T \mathbf{x} = \sum_{k=1}^K x_i^2 = \|\mathbf{x}\|_2^2$.

A matrix times a vector:

- If \mathbf{A} is an $N \times K$ and \mathbf{x} is a vector of length K , then the product \mathbf{Ax} is well defined.
 - ▶ The product \mathbf{Ax} produces a vector of length N .
 - ▶ Elements of \mathbf{x} inform you how to combine columns of \mathbf{A} .
- For instance:

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 1 \\ 3 & 0 & -1 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix}.$$

Therefore,

$$\mathbf{Ax} = \left(5 \times \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right) + \left(-4 \times \begin{bmatrix} 4 \\ 0 \end{bmatrix} \right) + \left(0 \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} -11 \\ 15 \end{bmatrix}.$$

MATRIX OPERATIONS

A matrix times a vector:

- In practice, we combine steps: the i^{th} element of \mathbf{Ax} is given by the dot product of the i^{th} row of \mathbf{A} with \mathbf{x} .
 - ▶ In other words, $[\mathbf{Ax}]_i = \mathbf{a}_i^T \cdot \mathbf{x} = \mathbf{a}_i \cdot \mathbf{x}$.
- Mental image associated with first step in this operation:

$$\begin{bmatrix} \boxed{1} & 4 & 1 \\ 3 & 0 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix} = \begin{bmatrix} \boxed{-11} \\ ? \end{bmatrix}.$$

- Mental image associated with second step in this operation:

$$\begin{bmatrix} 1 & 4 & 1 \\ \boxed{3} & 0 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix} = \begin{bmatrix} -11 \\ \boxed{15} \end{bmatrix}.$$

MATRIX OPERATIONS

A matrix times a matrix:

- Matrix product \mathbf{AB} well defined if number of columns of \mathbf{A} matches number of rows of \mathbf{B} .
 - ▶ Each column obtained by multiplying \mathbf{A} by columns of \mathbf{B} .
- For instance:

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 1 \\ 3 & 0 & -1 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 5 & -2 \\ -4 & 6 \\ 0 & -2 \end{bmatrix}.$$

Therefore,

$$[\mathbf{AB}]_{.1} = 5 \times \begin{bmatrix} 1 \\ 3 \end{bmatrix} + (-4) \times \begin{bmatrix} 4 \\ 0 \end{bmatrix} + (0) \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -11 \\ 15 \end{bmatrix}, \text{ and}$$

$$[\mathbf{AB}]_{.2} = (-2) \times \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 6 \times \begin{bmatrix} 4 \\ 0 \end{bmatrix} + (-2) \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 20 \\ -4 \end{bmatrix}.$$

MATRIX OPERATIONS

A matrix times a matrix:

- We use the visual tool to obtain elements of **AB**.
- Mental image associated with first step in this operation:

$$\begin{bmatrix} \boxed{1} & 4 & 1 \\ 3 & 0 & -1 \end{bmatrix} \begin{bmatrix} \boxed{5} & -2 \\ -4 & 6 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} \boxed{-11} & ? \\ ? & ? \end{bmatrix}.$$

- ... and so on ...
- Mental image associated with fourth step in this operation:

$$\begin{bmatrix} 1 & 4 & 1 \\ \boxed{3} & 0 & -1 \end{bmatrix} \begin{bmatrix} 5 & \boxed{-2} \\ -4 & 6 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} -11 & 20 \\ 15 & \boxed{-4} \end{bmatrix}.$$

A matrix times a matrix: Properties

- Matrix multiplication is associative:

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}).$$

- Matrix multiplication is distributive over matrix addition:

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \text{ and } (\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}.$$

- Matrix multiplication is non-commutative:

$$\mathbf{AB} \neq \mathbf{BA} \text{ (generally).}$$

A matrix's trace:

- For a square matrix, **A**, its trace is given by $\text{tr}(\mathbf{A}) = \sum_{n=1}^N A_{nn}$.
 - ▶ That is, the trace is the sum of the entries on the main diagonal.
- For instance:

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

Therefore,

$$\text{tr}(\mathbf{A}) = 2 + (-1) = 1.$$

A matrix's trace: Properties

- $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^T)$.
- $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$.
- $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$.
- $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) = \text{tr}(\mathbf{CAB})$.
- $\text{tr}(\mathbf{xx}^T) = \mathbf{x}^T \mathbf{x}$.
 - ▶ $\mathbf{x}^T \mathbf{x} = \langle \mathbf{x}, \mathbf{x} \rangle$ is referred to as an “inner” product, producing a scalar.
 - ▶ $\mathbf{xx}^T = \mathbf{x} \otimes \mathbf{x}$ is referred to as an “outer” product, producing a matrix.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices**
- 3 Understanding the key linear subspaces
- 4 Matrix inverses
- 5 Very special matrices
- 6 Determinants
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices
- 9 Very important matrix decompositions
- 10 Generalized inverses
- 11 Matrix calculus

Gaussian elimination:

- Suppose we seek to solve $\mathbf{Ax} = \mathbf{c}$ for \mathbf{x} (a very convenient matrix notation to represent a system of N equations with K unknowns).
 - ▶ The proper way to think about this from the standpoint of linear algebra is that \mathbf{x} is telling you the correct linear combination (if one exists) of the *columns* of \mathbf{A} that produce the vector \mathbf{c} .
- The typical way we learn how to do this is through a process called Gaussian elimination, sequentially scaling, switching, and adding multiples of rows from the augmented matrix $[\mathbf{A} \quad \mathbf{c}]$ until \mathbf{A} has been transformed into an upper-triangular matrix.
- We then back-substitute to learn the values of \mathbf{x} .

ELIMINATION WITH MATRICES

Gaussian elimination: Example

- Equations:

$$x + 2y + z = 2$$

$$3x + 8y + z = 12$$

$$4y + z = 2$$

- From “good” to “better”: keep variables aligned and refer to them as x_1 , x_2 , and x_3 :

$$x_1 + 2x_2 + x_3 = 2$$

$$3x_1 + 8x_2 + x_3 = 12$$

$$0x_1 + 4x_2 + x_3 = 2$$

ELIMINATION WITH MATRICES

Gaussian elimination: Example

- From “good” to “better”: keep variables aligned.

$$x_1 + 2x_2 + x_3 = 2$$

$$3x_1 + 8x_2 + x_3 = 12$$

$$0x_1 + 4x_2 + x_3 = 2$$

- From “better” to “best”: matrix notation ($\mathbf{Ax} = \mathbf{c}$).

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 12 \\ 2 \end{bmatrix}$$

ELIMINATION WITH MATRICES

Gaussian elimination: Example

- Augmented matrix $[\mathbf{A} \ \mathbf{c}]$ includes a vertical slash for clarity:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 3 & 8 & 1 & 12 \\ 0 & 4 & 1 & 2 \end{array} \right]$$

- Next step: want a 0 in the (2,1) position.
- Natural choice: subtract three of row 1 from row 2.
- Proper way to think about this from a linear algebra standpoint: multiply both sides by 3×3 matrix \mathbf{E}_{21} that leaves rows 1 and 3 unchanged but subtracts three of row 1 from row 2:

$$\mathbf{E}_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

ELIMINATION WITH MATRICES

Gaussian elimination: Example

- Applying the desired operation $\mathbf{E}_{21} [\mathbf{A} \ \mathbf{c}]$:

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

- Next step: want 0 in the (3,2) position.
- Natural choice: subtract two of row 2 from row 3.
- Proper way to think about this from a linear algebra standpoint: multiply both sides by 3×3 matrix \mathbf{E}_{32} that leaves rows 1 and 2 unchanged but subtracts two of row 2 from row 3:

$$\mathbf{E}_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}.$$

ELIMINATION WITH MATRICES

Gaussian elimination: Example

- Applying the desired operation $\mathbf{E}_{32}\mathbf{E}_{21}$ $[\mathbf{A} \ \mathbf{c}]$:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 0 & 5 & -10 \end{array} \right]$$

- The algorithm is complete and we use back-substitution to find:

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$$

- If the rows (equations) aren't in the "right" order, may need to switch rows (equivalent to multiplying by permutation matrix) to get to the right place.

ELIMINATION WITH MATRICES

Gaussian elimination: Properties in this nice case

- In this special case, the matrix was placed in row echelon form.
- \mathbf{A} had three “pivots” (and is thus characterized as *rank three*).
- Can be read as the first non-zero column of each row in echelon form.
- Though we haven't yet discussed the determinant, it is equal to the product of the pivots ($\det(\mathbf{A}) = 1 \times 2 \times 5 = 10$).
- Note that the echelon form is upper-triangular.
- Though we haven't yet discussed inverses, note that $(\mathbf{E}_{32}\mathbf{E}_{21})^{-1}$ is lower-triangular:

$$(\mathbf{E}_{32}\mathbf{E}_{21})^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 6 & 2 & 1 \end{bmatrix}$$

- This is the idea behind the famous decomposition $\mathbf{A} = \mathbf{LU}$.

ELIMINATION WITH MATRICES

More about the rank:

- $\text{rank}(\mathbf{A}) = \#$ of pivots in row echelon form (zeros can be in a pivot position, but a pivot cannot be zero).
- A square ($N \times N$) matrix is said to be of full rank if it has N pivots.
- For any matrix \mathbf{A} ,
 - ▶ $\text{rank}(\mathbf{A}) \leq N$ (number of rows).
 - ▶ $\text{rank}(\mathbf{A}) \leq K$ (number of columns).
 - ▶ $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T) = \text{rank}(\mathbf{A}\mathbf{A}^T) = \text{rank}(\mathbf{A}^T\mathbf{A})$.
- If the matrix product \mathbf{AB} is well defined, then the following are true:
 - ▶ $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$.
 - ▶ $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{B})$.
 - ▶ $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A})$ if \mathbf{B} is full-rank.
- If the matrix sum $\mathbf{A} + \mathbf{B}$ is well defined, then the following is true:
 - ▶ $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$.

More about the rank:

- To put some of these ideas more concretely/elegantly, define the augmented matrix $\mathbf{A}_{\text{aug}} = [\mathbf{A} \ \mathbf{c}]$.
- Then,
 - ① If $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}_{\text{aug}}) = K$, there is a unique solution to $\mathbf{Ax} = \mathbf{c}$.
 - ② If $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}_{\text{aug}}) < K$, there are many solutions to $\mathbf{Ax} = \mathbf{c}$.
 - ③ If $\text{rank}(\mathbf{A}) < \text{rank}(\mathbf{A}_{\text{aug}})$, there are no solutions to $\mathbf{Ax} = \mathbf{c}$.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces**
- 4 Matrix inverses
- 5 Very special matrices
- 6 Determinants
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices
- 9 Very important matrix decompositions
- 10 Generalized inverses
- 11 Matrix calculus

UNDERSTANDING THE KEY LINEAR SUBSPACES

Key ideas: Let $\mathbf{v}_1, \dots, \mathbf{v}_K \in \mathbb{R}^N$

- $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_K) = \sum_{k=1}^K a_k \mathbf{v}_k$ (all possible linear combinations).
- Vectors said to be linearly dependent if there exist c_1, \dots, c_K (not all zero) such that

$$\sum_{k=1}^K c_k \mathbf{v}_k = \mathbf{0},$$

and are otherwise said to be linearly independent.

- If $\mathbf{v}_1, \dots, \mathbf{v}_K$ are linearly independent, each $\mathbf{y} \in \mathbb{R}^N$ can be written in the form for some values a_1, \dots, a_K :

$$\mathbf{y} = \sum_{k=1}^K a_k \mathbf{v}_k,$$

- Linearly independent vectors that *span* a vector space in this way are said to form a *basis* for that vector space.

Motivation:

- Square matrices represent the setting in which the number of equations and number of unknown variables match, which is what gives us the hope of finding a unique solution to $\mathbf{Ax} = \mathbf{c}$ for any \mathbf{c} .
 - ▶ $N > K$: number of equations exceeds number of unknowns.
 - ▶ $N < K$: number of unknowns exceeds number of equations.
- Of course, sometimes the equations in a system can represent either redundant or inconsistent information.

UNDERSTANDING THE KEY LINEAR SUBSPACES

Example: Redundancy even in the case of $N = K$

- Despite the fact that there are three equations and three unknowns below, only two of the equations contribute unique information:

$$2x_1 + 3x_2 + 5x_3 = 0$$

$$2x_1 + 2x_2 + 4x_3 = 0$$

$$2x_1 + 4x_2 + 6x_3 = 0$$

- The “redundancy” in this example is clearer by looking at the columns of the associated matrix:

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

UNDERSTANDING THE KEY LINEAR SUBSPACES

Example: Redundancy even in the case of $N = K$

- Matrix form:

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

- The third column is the sum of the other two.
- Very importantly: $\text{rank}(\mathbf{A})$ marks *both* the number of linearly independent rows *and* the number of linearly independent columns.
- Therefore, I know that any row can be written as a linear combination of the other two without having to do the calculation.

UNDERSTANDING THE KEY LINEAR SUBSPACES

Example: Redundancy even in the case of $N = K$

- The linear dependence of the columns (or equivalently, rows) will make itself apparent when reducing to row echelon form:

$$\begin{bmatrix} 2 & 3 & 5 \\ 2 & 2 & 4 \\ 2 & 4 & 6 \end{bmatrix} \xrightarrow{\mathbf{E}_{21}} \begin{bmatrix} 2 & 3 & 5 \\ 0 & -1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{\mathbf{E}_{32}} \begin{bmatrix} 2 & 3 & 5 \\ 0 & -1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

- We learn ALL of the following:
 - ▶ The pivots of \mathbf{A} are 2 and -1 ($\text{rank}(\mathbf{A}) = 2$).
 - ▶ There are two linearly independent columns.
 - ▶ There are two linearly independent rows.
 - ▶ There are infinitely many solutions to the equation $\mathbf{Ax} = \mathbf{0}$.
 - ▶ The values \mathbf{c} for which $\mathbf{Ax} = \mathbf{c}$ can be expressed as a linear combination of any two linearly independent columns of \mathbf{A} (that is, any two linearly independent columns of \mathbf{A} form a *basis* for its column space).

The row space:

- The *row space* of \mathbf{A} , denoted by $\mathcal{R}(\mathbf{A})$, is defined as:

$$\mathcal{R}(\mathbf{A}) = \left\{ \mathbf{y} \in \mathbb{R}^K : \exists \mathbf{x} \in \mathbb{R}^N \text{ with } \mathbf{y} = \mathbf{A}^T \mathbf{x} \right\}.$$

- Characterizations:

- ▶ The space *spanned* by the rows of \mathbf{A} .
- ▶ Note that in my notation above, I'm avoiding row vectors by noting that the row space of \mathbf{A} is simply the “column space” of \mathbf{A}^T .

- Properties:

- ▶ $\mathbf{0}_K \in \mathcal{R}(\mathbf{A})$.
- ▶ If $\mathbf{v}_1 \in \mathcal{R}(\mathbf{A})$ and $\mathbf{v}_2 \in \mathcal{R}(\mathbf{A})$, then $c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 \in \mathcal{R}(\mathbf{A})$.
- ▶ $\dim(\mathcal{R}(\mathbf{A})) = \text{rank}(\mathbf{A}) = \# \text{ pivots in echelon form}$.

The null space:

- The *null space* of \mathbf{A} , denoted by $\mathcal{N}(\mathbf{A})$, is defined as:

$$\mathcal{N}(\mathbf{A}) = \left\{ \mathbf{x} \in \mathbb{R}^K : \mathbf{A}\mathbf{x} = \mathbf{0}_N \right\}.$$

- Characterizations:

- ▶ The set of vectors that vanish when \mathbf{A} is “applied” to them.

- Properties:

- ▶ $\mathbf{0}_K \in \mathcal{N}(\mathbf{A})$.
- ▶ If $\mathbf{v}_1 \in \mathcal{N}(\mathbf{A})$ and $\mathbf{v}_2 \in \mathcal{N}(\mathbf{A})$, then $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \in \mathcal{N}(\mathbf{A})$.
- ▶ If $\mathcal{N}(\mathbf{A})$ *only* includes $\mathbf{0}_K$, then \mathbf{A} is invertible (meaning that $\mathbf{A}\mathbf{x} = \mathbf{c}$ has a solution for every \mathbf{c}).
- ▶ Every vector $\mathbf{v}_1 \in \mathcal{N}(\mathbf{A})$ is orthogonal to every vector $\mathbf{v}_2 \in \mathcal{R}(\mathbf{A})$. For this reason, we say that the null space and row space are *orthogonal complements* ($\mathcal{N}(\mathbf{A}) = [\mathcal{R}(\mathbf{A})]^\perp$).

The left null space:

- The *left null space* of \mathbf{A} , denoted by $\mathcal{N}(\mathbf{A}^T)$, is defined as:

$$\mathcal{N}(\mathbf{A}^T) = \left\{ \mathbf{x} \in \mathbb{R}^N : \mathbf{x}^T \mathbf{A} = \mathbf{0}_K \right\}.$$

- Characterizations:

- ▶ The set of vectors that vanish when “applied to” \mathbf{A} .
- ▶ The term “left null space” comes from the fact that the \mathbf{x} is on the left rather than on the right.
- ▶ The notation $\mathcal{N}(\mathbf{A}^T)$ come from the idea that the left null space of \mathbf{A} is simply the null space of \mathbf{A}^T .

- Properties:

- ▶ $\mathbf{0}_N \in \mathcal{N}(\mathbf{A}^T)$.
- ▶ If $\mathbf{v}_1 \in \mathcal{N}(\mathbf{A}^T)$ and $\mathbf{v}_2 \in \mathcal{N}(\mathbf{A}^T)$, then $c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 \in \mathcal{N}(\mathbf{A}^T)$.

UNDERSTANDING THE KEY LINEAR SUBSPACES

The column space:

- The *column space* of \mathbf{A} , denoted by $\mathcal{C}(\mathbf{A})$, is defined as:

$$\mathcal{C}(\mathbf{A}) = \left\{ \mathbf{y} \in \mathbb{R}^N : \exists \mathbf{x} \in \mathbb{R}^K \text{ with } \mathbf{y} = \mathbf{A}\mathbf{x} \right\}.$$

- Characterizations:

- ▶ The space *spanned* by the columns of \mathbf{A} .
- ▶ Characterizes all vectors \mathbf{c} for which $\mathbf{A}\mathbf{x} = \mathbf{c}$ has a solution.

- Properties:

- ▶ $\mathbf{0}_N \in \mathcal{C}(\mathbf{A})$
- ▶ If $\mathbf{v}_1 \in \mathcal{C}(\mathbf{A})$ and $\mathbf{v}_2 \in \mathcal{C}(\mathbf{A})$, then $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \in \mathcal{C}(\mathbf{A})$.
- ▶ $\dim(\mathcal{C}(\mathbf{A})) = \text{rank}(\mathbf{A}) = \# \text{ pivots in echelon form}$.
- ▶ $\dim(\mathcal{N}(\mathbf{A})) + \dim(\mathcal{C}(\mathbf{A})) = K = \# \text{ columns in } \mathbf{A}$.
- ▶ Every vector $\mathbf{v}_1 \in \mathcal{C}(\mathbf{A})$ is orthogonal to every vector $\mathbf{v}_2 \in \mathcal{N}(\mathbf{A}^T)$. For this reason, we say that the left null space and column space are *orthogonal complements* ($\mathcal{N}(\mathbf{A}^T) = [\mathcal{C}(\mathbf{A})]^\perp$).

UNDERSTANDING THE KEY LINEAR SUBSPACES

Return to prior example:

- Equations:

$$\begin{bmatrix} 2 & 3 & 5 \\ 2 & 2 & 4 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{0}.$$

- We found that $\text{rank}(\mathbf{A}) = 2$. From the prior properties, we know:
 - ▶ $\mathcal{C}(\mathbf{A})$ is a plane in \mathbb{R}^3 through the origin.
 - ★ Specifically, the plane is characterized by the span of any two of the columns of \mathbf{A} and has equation $4x_1 - 2x_2 - 2x_3 = 0$, which I learned by taking the cross product of the first two linearly independent column vectors.
 - ▶ $\mathcal{N}(\mathbf{A})$ is a line in \mathbb{R}^3 through the origin.
 - ★ Specifically, the line is spanned by the vector $(1, 1, -1)^T$, which I learned from the rank-nullity theorem and by recognizing this as a vector in the null space.

Example: Linear models

- We often find ourselves in the case where $N > K$, with more equations than unknowns:

$$x_1 + 2x_2 = 4$$

$$x_1 + 3x_2 = 3$$

$$x_1 + 4x_2 = 6$$

$$x_1 + 3x_2 = 5$$

$$x_1 + 1x_2 = 2$$

- In this case, the equations are inconsistent and the most we can hope for is a “best” solution for x_1 and x_2 (which are playing the role of the intercept and a slope in a simple linear regression model).
- It is extraordinarily important to unify these ideas with properties regarding matrix-related subspaces.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces
- 4 Matrix inverses**
- 5 Very special matrices
- 6 Determinants
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices
- 9 Very important matrix decompositions
- 10 Generalized inverses
- 11 Matrix calculus

A matrix's inverse:

- A square matrix **A** is invertible if there exists a matrix **A**⁻¹ such that **A**⁻¹**A** = **AA**⁻¹ = **I**.
- An invertible matrix is also referred to as a non-singular matrix.
- For instance, if

$$\mathbf{A} = \begin{bmatrix} 2 & -2 \\ 1 & 4 \end{bmatrix}.$$

you can verify that

$$\mathbf{A}^{-1} = \begin{bmatrix} 0.4 & 0.2 \\ -0.1 & 0.2 \end{bmatrix}.$$

Algorithm:

- You can attempt to find a matrix's inverse via the Gauss-Jordan algorithm, which extends the Gaussian elimination algorithm.
- Idea: Take augmented matrix $[\mathbf{A} \quad \mathbf{I}]$ and apply the steps of elimination—but keep going until you can get \mathbf{A} into the form \mathbf{I} .
- If this is not possible, \mathbf{A} is not invertible. But if it is, then the matrix to which the identity transforms is nothing other than \mathbf{A}^{-1} .

Gauss-Jordan: Example

- Let's take one of our prior examples:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 3 & 8 & 1 & 0 & 1 & 0 \\ 0 & 4 & 1 & 0 & 0 & 1 \end{array} \right]$$

- Step 1 from elimination: subtract three of row 1 from row 2.
- Step 2 from elimination: subtract two of row 2 from row 3.
- Result:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 2 & -2 & -3 & 1 & 0 \\ 0 & 0 & 5 & 6 & -2 & 1 \end{array} \right]$$

- Gauss says stop, but Jordan says “keep going.”

Gauss-Jordan: Example

- New starting point:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 2 & -2 & -3 & 1 & 0 \\ 0 & 0 & 5 & 6 & -2 & 1 \end{array} \right]$$

- Step 3: Take one of row 2 away from row 1.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 4 & -1 & 0 \\ 0 & 2 & -2 & -3 & 1 & 0 \\ 0 & 0 & 5 & 6 & -2 & 1 \end{array} \right]$$

Gauss-Jordan: Example

- Not done yet; keep going!

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 4 & -1 & 0 \\ 0 & 2 & -2 & -3 & 1 & 0 \\ 0 & 0 & 5 & 6 & -2 & 1 \end{array} \right]$$

- Step 4: Take $3/5$ of row 3 away from row 1.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 - 18/5 & -1 + 6/5 & -3/5 \\ 0 & 2 & -2 & -3 & 1 & 0 \\ 0 & 0 & 5 & 6 & -2 & 1 \end{array} \right]$$

Gauss-Jordan: Example

- Not done yet; keep going!

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 - 18/5 & -1 + 6/5 & -3/5 \\ 0 & 2 & -2 & -3 & 1 & 0 \\ 0 & 0 & 5 & 6 & -2 & 1 \end{array} \right]$$

- Step 5: Add $2/5$ of row 3 to row 2.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 - 18/5 & -1 + 6/5 & -3/5 \\ 0 & 2 & 0 & -3 + 12/5 & 1 - 4/5 & 2/5 \\ 0 & 0 & 5 & 6 & -2 & 1 \end{array} \right]$$

MATRIX INVERSES

Gauss-Jordan: Example

- Not done yet; keep going!

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 - 18/5 & -1 + 6/5 & -3/5 \\ 0 & 2 & 0 & -3 + 12/5 & 1 - 4/5 & 2/5 \\ 0 & 0 & 5 & 6 & -2 & 1 \end{array} \right]$$

- Steps 6 and 7: Divide row 2 by 2 and row 5 by 5

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 - 18/5 & -1 + 6/5 & -3/5 \\ 0 & 1 & 0 & -3/2 + 12/10 & 1/2 - 4/10 & 2/10 \\ 0 & 0 & 1 & 6/5 & -2/5 & 1/5 \end{array} \right]$$

- Cleaning up:

$$\mathbf{A}^{-1} = \begin{bmatrix} 2/5 & 1/5 & -3/5 \\ -3/10 & 1/10 & 1/5 \\ 6/5 & -2/5 & 1/5 \end{bmatrix}.$$

Gauss-Jordan: Example

- We can use this to verify that our prior solution to $\mathbf{Ax} = \mathbf{c}$ was correct when $\mathbf{c} = (2, 12, 2)$:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{c} = \begin{bmatrix} 2/5 & 1/5 & -3/5 \\ -3/10 & 1/10 & 1/5 \\ 6/5 & -2/5 & 1/5 \end{bmatrix} \begin{bmatrix} 2 \\ 12 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}.$$

- We can also use this to find solutions involving other values of \mathbf{c} without having to cycle through the elimination steps again.

A matrix's inverse: Meaning and properties

- Meaning: If \mathbf{A} ($N \times N$) is non-singular (invertible), then...
 - ▶ The equation $\mathbf{Ax} = \mathbf{c}$ can be solved for \mathbf{x} for each $\mathbf{c} \in \mathbb{R}^N$ and those solutions are unique ($\mathbf{x} = \mathbf{A}^{-1}\mathbf{c}$).
 - ▶ $\mathcal{C}(\mathbf{A}) = \mathbb{R}^N$.
 - ▶ $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$.
 - ▶ \mathbf{A} has N pivots.
 - ▶ \mathbf{A} is of full rank (that is, $\text{rank}(\mathbf{A}) = N$).
- Properties:
 - ▶ If \mathbf{A} and \mathbf{B} are invertible $N \times N$ matrices, then $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
 - ▶ $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.
 - ▶ $(c\mathbf{A})^{-1} = c^{-1}(\mathbf{A}^{-1})$.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces
- 4 Matrix inverses
- 5 Very special matrices**
- 6 Determinants
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices
- 9 Very important matrix decompositions
- 10 Generalized inverses
- 11 Matrix calculus

Symmetric matrices:

- A square matrix, \mathbf{A} is said to be *symmetric* if $a_{ij} = a_{ji}$.
 - ▶ Or, to put it another way, $\mathbf{A}^T = \mathbf{A}$, meaning that reversing the roles of the rows and columns does not alter a symmetric matrix.
- For instance:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & -1 \\ 3 & -1 & 0 \end{bmatrix}$$

is an example of a symmetric matrix.

- Example: $\mathbf{A}^T \mathbf{A}$ is a symmetric matrix.
- Example: For $\mathbf{v} \in \mathbb{R}^N$, $\mathbf{v}\mathbf{v}^T$ is a symmetric $N \times N$ matrix (referred to as the outer product, just as $\mathbf{v}^T \mathbf{v}$ is referred to as the inner product).

Symmetric matrices:

- Matrices of the form $\mathbf{A}^T \mathbf{A}$ (or $\mathbf{A} \mathbf{A}^T$) are generally very special in linear algebra and you'll see this form all the time in this course.

Lemma 1.1: Property of $\mathbf{A}^T \mathbf{A}$

$\mathbf{A}^T \mathbf{A} = \mathbf{0}$ if and only if $\mathbf{A} = \mathbf{0}$.

Lemma 1.1: Proof

- If $\mathbf{A}^T \mathbf{A} = \mathbf{0}$, then $\text{trace}(\mathbf{A}^T \mathbf{A}) = 0$, which can only happen if $\mathbf{A} = \mathbf{0}$.
- The proof in the other direction is trivial.

Lemma 1.2: Another property involving $\mathbf{A}^T \mathbf{A}$

If $\mathbf{B}\mathbf{A}^T \mathbf{A} = \mathbf{C}\mathbf{A}^T \mathbf{A}$, then $\mathbf{B}\mathbf{A}^T = \mathbf{C}\mathbf{A}^T$.

Lemma 1.2: Proof

- This can be shown using Lemma 1.1:

$$\begin{aligned}(\mathbf{BA}^T - \mathbf{CA}^T)(\mathbf{BA}^T - \mathbf{CA}^T)^T &= (\mathbf{BA}^T - \mathbf{CA}^T)\mathbf{A}(\mathbf{B} - \mathbf{C})^T \\ &= (\mathbf{BA}^T\mathbf{A} - \mathbf{CA}^T\mathbf{A})(\mathbf{B} - \mathbf{C})^T \\ &= \mathbf{0}(\mathbf{B} - \mathbf{C})^T \\ &= \mathbf{0}.\end{aligned}$$

- Therefore, by Lemma 1.1, $\mathbf{BA}^T - \mathbf{CA}^T = \mathbf{0}$.

Idempotent matrices:

- An *idempotent* matrix is one for which $\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$.
 - ▶ Multiplying the matrix by itself gives back the original matrix.
- For instance:

$$\mathbf{A} = \begin{bmatrix} 3 & -6 \\ 1 & -2 \end{bmatrix}$$

is an example of an idempotent matrix.

- Note: Idempotent matrices must be square.
- Note: the rank of an idempotent matrix equals its trace.
- Note: an idempotent matrix can only have eigenvalues of zero and one (though we haven't yet discussed eigenvalues).

Projection matrices:

- An orthogonal projection matrix, \mathbf{P} , is symmetric and idempotent.
 - ▶ Non-symmetric idempotent matrices are *oblique* projection matrices.
 - ▶ Unless otherwise specified, any reference to projection matrices refers to orthogonal projection matrices in this course.
- If \mathbf{P} is a projection matrix, then $\text{rank}(\mathbf{P}) = \text{trace}(\mathbf{P})$ by a property of idempotent matrices.
- If \mathbf{P} is a projection matrix, $\mathbf{P}^T \mathbf{P} = \mathbf{0}$ if and only if $\mathbf{P} = \mathbf{0}$ by a property of symmetric matrices.
- Matrices of the form $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ are very important examples of a projection matrices in which $\mathbf{v} = \mathbf{P}\mathbf{x}$ marks the projection of vector \mathbf{x} onto the space spanned by the columns of \mathbf{A} (presumes \mathbf{A} is of full rank, in which case $\mathbf{A}^T \mathbf{A}$ is invertible).
- So important, that I think we should play with some examples of projection matrices having this nice form.

Projection matrices:

- For the time being, let me use \mathbf{X} to denote my “matrix of interest” to follow familiar notation from regression modeling, and let $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ denote the projection matrix.
- Let's verify that \mathbf{P} is a projection matrix. To do this, we need to check whether \mathbf{P} is symmetric:

$$\begin{aligned}\mathbf{P}^T &= (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\ &= ((\mathbf{X}^T)^T(\mathbf{X}^T\mathbf{X})^{-T}(\mathbf{X})^T) \\ &= (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) = \mathbf{P}.\end{aligned}$$

- I also need to check whether \mathbf{P} is idempotent:

$$\begin{aligned}\mathbf{P}^2 &= \mathbf{P}\mathbf{P} = (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) \\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{P}.\end{aligned}$$

Projection matrices:

- Let's further show that $\mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is a projection matrix by verifying that it too is symmetric and idempotent:

$$\begin{aligned}(\mathbf{I} - \mathbf{P})^T &= \mathbf{I}^T - \mathbf{P}^T \\ &= \mathbf{I} - \mathbf{P}.\end{aligned}$$

- Further,

$$\begin{aligned}(\mathbf{I} - \mathbf{P})^2 &= (\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P}) \\ &= \mathbf{I}^2 - \mathbf{IP} - \mathbf{PI} + \mathbf{PP} \\ &= \mathbf{I} - 2\mathbf{P} + \mathbf{P} \\ &= \mathbf{I} - \mathbf{P}.\end{aligned}$$

Projection matrices:

- Continuing our example, you should be able to show that $\mathbf{PX} = \mathbf{X}$ easily. What about $(\mathbf{I} - \mathbf{P})\mathbf{X}$?

$$\begin{aligned}(\mathbf{I} - \mathbf{P})\mathbf{X} &= \mathbf{X} - \mathbf{PX} \\ &= \mathbf{X} - \mathbf{X} = \mathbf{0}\end{aligned}$$

- How does this square with our intuition about what \mathbf{P} is doing?
 - ▶ \mathbf{P} projects vectors onto the space spanned by the columns of \mathbf{X} , while $\mathbf{I} - \mathbf{P}$ projects vectors onto the orthogonal complement of \mathbf{X} .
 - ▶ Projecting \mathbf{X} onto its own column space should leave \mathbf{X} unchanged. On the other hand, projecting the columns of \mathbf{X} onto a space that is orthogonal to the columns of \mathbf{X} should indeed give the zero vector.

Projection matrices:

- As an example, let $\mathbf{P} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, a projection matrix.
- $\mathbf{P}\mathbf{x} = \frac{1}{2} \begin{bmatrix} x_1 + x_2 \\ x_1 + x_2 \end{bmatrix}$. \mathbf{P} projects vectors onto the line spanned by $(1, 1)^T$.
- On the other hand,

$$\mathbf{I} - \mathbf{P} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

- $(\mathbf{I} - \mathbf{P})\mathbf{x} = \frac{1}{2} \begin{bmatrix} x_1 - x_2 \\ x_2 - x_1 \end{bmatrix}$. $\mathbf{I} - \mathbf{P}$ projects vectors onto the line spanned by $(1, -1)^T$.

Projection matrices:

- More generally, \mathbf{J}_N/N , an $N \times N$ matrix with all entries given by $1/N$, is a projection matrix.
- Clearly, \mathbf{J}_N is symmetric.
- To show idempotence, it is helpful to write $\mathbf{J} = \mathbf{1}_N \mathbf{1}_N^T$, where $\mathbf{1}_N$ is a length- N vector of ones.
- Then,

$$\begin{aligned}\left(\frac{1}{N}\mathbf{J}_N\right)^2 &= \left(\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T\right)\left(\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T\right) \\ &= \frac{1}{N^2}\mathbf{1}_N(\mathbf{1}_N^T\mathbf{1}_N)\mathbf{1}_N^T = \frac{1}{N^2}\mathbf{1}_N(N)\mathbf{1}_N^T \\ &= \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T = \frac{1}{N}\mathbf{J}_N.\end{aligned}$$

Orthogonal matrices:

- A square matrix, \mathbf{A} , is said to be *orthogonal* if its columns are mutually orthogonal and each unit length (orthonormal).
- For instance:

$$\mathbf{A} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

is an example of an orthogonal matrix.

- Note: If \mathbf{A} is orthogonal, then $\mathbf{A}^T \mathbf{A} = \mathbf{I}$, which is another way of saying that $\mathbf{A} = \mathbf{A}^{-1}$.
- Geometrically, these matrices are marked by rotations, reflections, and rotoinversions.
- Example: Permutation matrices (zeros everywhere except an entry of “1” in each column and in each row) are orthogonal.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces
- 4 Matrix inverses
- 5 Very special matrices
- 6 Determinants**
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices
- 9 Very important matrix decompositions
- 10 Generalized inverses
- 11 Matrix calculus

A (square) matrix's determinant:

- The determinant of **A** is typically denoted by $\det(\mathbf{A})$ or $|\mathbf{A}|$.
- It is a quantity that shows up *repeatedly* in linear algebra.
- The value of a determinant is uniquely defined/identified through its possession of the following properties:
 - 1 The identity matrix has a determinant of one.
 - 2 Row exchanges reverse the sign of the determinant.
 - 3 The determinant has a “row-wise linearity,” which is easiest to describe in the case of a 2×2 matrix:

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix} \text{ and } \begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

A matrix's determinant:

- You should be able to verify most of the below properties of the determinant of an $N \times N$ matrix **A**:
 - 1 If **A** has two equal rows, $|\mathbf{A}| = 0$.
 - 2 If **A** has a row of zeros, $|\mathbf{A}| = 0$.
 - 3 Subtracting a multiple of row i from row k does not change $|\mathbf{A}|$.
 - 4 The determinant of an upper-triangular matrix is the product of the diagonal entries.
 - 5 $|\mathbf{A}| = 0$ if and only if **A** is singular.
 - 6 $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$.
 - 7 If **A** is non-singular, $|\mathbf{A}^{-1}| = 1/|\mathbf{A}|$.
 - 8 $|c\mathbf{A}| = c^N|\mathbf{A}|$.
 - 9 $|\mathbf{A}^T| = |\mathbf{A}|$.
 - 10 $|\mathbf{I} + \mathbf{xy}^T| = 1 + \mathbf{x}^T\mathbf{y}$.

A matrix's determinant:

- The absolute value of the determinant can be conceptualized geometrically as $\text{vol}(P)$, where

$$P = \{c_1 \mathbf{A}_{\cdot 1} + \cdots + c_K \mathbf{A}_{\cdot K} \mid 0 \leq c_i \leq 1 \forall i\}.$$

- This represents the volume of the parallelepiped with each of its sides being represented/implied by the columns of \mathbf{A} .

DETERMINANTS

A matrix's determinant:

- There are many formulas and procedures to determine $|\mathbf{A}|$.
- In the 2×2 case, this is given by:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$$

which is absolutely necessary to commit to memory.

- In the general 3×3 case, we often break down:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

- Note that we could have gone “down a column” of \mathbf{A} rather than “across a row” and we would still get the same answer.

Connection to inverses:

- If \mathbf{A} is invertible, then:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \mathbf{C}^T,$$

where \mathbf{C} is the matrix of cofactors, and $\mathbf{C}^T = \text{adj}(\mathbf{A})$ is referred to as the adjugate matrix.

- In the 2×2 case, we have that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Connection to inverses:

- In the 3×3 case, we have that

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} + \begin{vmatrix} e & f \\ h & i \end{vmatrix} & - \begin{vmatrix} d & f \\ g & i \end{vmatrix} & + \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ - \begin{vmatrix} b & c \\ h & i \end{vmatrix} & + \begin{vmatrix} a & c \\ g & i \end{vmatrix} & - \begin{vmatrix} a & b \\ g & h \end{vmatrix} \\ + \begin{vmatrix} b & c \\ e & f \end{vmatrix} & - \begin{vmatrix} a & c \\ d & f \end{vmatrix} & + \begin{vmatrix} a & b \\ d & e \end{vmatrix} \end{bmatrix}$$

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces
- 4 Matrix inverses
- 5 Very special matrices
- 6 Determinants
- 7 Eigenvalues and eigenvectors**
- 8 Positive definite matrices
- 9 Very important matrix decompositions
- 10 Generalized inverses
- 11 Matrix calculus

Motivation:

- Often useful to motivate linear algebra problems geometrically.
- Suppose you want to find vectors \mathbf{x} for which $\mathbf{y} = \mathbf{Ax} \in \mathcal{C}(\mathbf{A})$ is “in the same direction” as \mathbf{x} (or in the reverse direction).
- Put algebraically, can we figure out which vectors \mathbf{x} solve $\mathbf{Ax} = \lambda\mathbf{x}$?
- These vectors are called the *eigenvectors* of \mathbf{A} .
- The values of λ for which the above equation is satisfied are called the eigenvalues.
- Naturally, this concept only makes sense if \mathbf{A} is square.

EIGENVALUES AND EIGENVECTORS

Ideas:

- If in fact $\mathbf{Ax} = \lambda\mathbf{x}$, then $\mathbf{Ax} - \lambda\mathbf{x} = (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$.
- By convention, we don't consider $\mathbf{x} = \mathbf{0}$ to be an eigenvector of \mathbf{A} (although λ can be zero).
- Now, there is a solution $\mathbf{x} \neq \mathbf{0}$ exactly when the columns of $\mathbf{A} - \lambda\mathbf{I}$ are linearly dependent (or, $\mathbf{A} - \lambda\mathbf{I}$ is singular and has determinant zero).
- This gives us a clue on how to find the eigenvalues, which is to solve $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ for λ . This left-hand side is called the characteristic polynomial (with degree given by the number of columns/rows).
- To then find the eigenvectors (i.e., the vectors $\mathbf{x} \in \mathcal{N}(\mathbf{A} - \lambda\mathbf{I})$), we plug in the eigenvalues into $\mathbf{Ax} = \lambda\mathbf{x}$ and solve for \mathbf{x} .

Example:

- Take the matrix $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ as an example.
- We have that $\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} 3 - \lambda & 1 \\ 1 & 3 - \lambda \end{bmatrix}$.
- Now, $\det(\mathbf{A} - \lambda\mathbf{I}) = (3 - \lambda)^2 - 1 = \lambda^2 - 6\lambda + 8 = (\lambda - 4)(\lambda - 2)$.
- This tells us that $\lambda = 2$ and $\lambda = 4$ are the eigenvalues of \mathbf{A} .
- To characterize the eigenvectors, we want to solve $(\mathbf{A} - 2\mathbf{I})\mathbf{x} = \mathbf{0}$ and $(\mathbf{A} - 4\mathbf{I})\mathbf{x} = \mathbf{0}$ for \mathbf{x} .

Example:

- Now, $\mathbf{A} - 2\mathbf{I} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.
- We should be able to identify that this matrix will map any vector of the form $(c, -c)^T$ to zero. We typically take $c = 1/\sqrt{2}$ so that the eigenvector is of length one.
- Next, $\mathbf{A} - 4\mathbf{I} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$.
- We should be able to identify that this matrix will map any vector of the form $(c, c)^T$ to zero. We typically take $c = 1/\sqrt{2}$ so that the eigenvector is of length one.
- A lot of really nice things happened in this example that were of course no accident.

Properties: All of which check out with the prior example

- The sum of the eigenvalues of \mathbf{A} is equal to $\text{trace}(\mathbf{A})$.
- The product of the eigenvalues of \mathbf{A} is equal to $\det(\mathbf{A})$.
- Symmetric matrices in particular have real eigenvalues.
- Symmetric matrices in particular have orthogonal eigenvectors.
- Positive definite matrices have positive eigenvalues (although we have not talked about positive definite matrices yet).
- If λ is an eigenvalue of \mathbf{A} , then $\lambda + c$ is an eigenvalue of $\mathbf{A} + c\mathbf{I}$.

More properties:

- If \mathbf{A} is triangular, then the diagonal elements of \mathbf{A} are the eigenvalues.
- If λ is an eigenvalue of an invertible matrix \mathbf{A} with eigenvector \mathbf{v} , then λ^{-1} is an eigenvalue of \mathbf{A}^{-1} with eigenvector \mathbf{x} .
- A matrix is invertible if and only if it has *only* nonzero eigenvalues.
- The eigenvalues of a matrix *can be complex*, but they come in pairs when they are.
- If an $N \times N$ matrix has N distinct eigenvalues, then there are N linearly independent eigenvectors.
- The eigenvalues of a matrix *can be repeated*. If the N eigenvalues of an $N \times N$ matrix are not distinct, you *may not* be able to find N linearly independent eigenvectors.

Example: Repeated eigenvalues

- Take the matrix $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ as an example.
- We have that $\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} 1 - \lambda & 0 \\ 0 & 1 - \lambda \end{bmatrix}$.
- Now, $\det(\mathbf{A} - \lambda\mathbf{I}) = (1 - \lambda)^2$.
- This tells us that $\lambda = 1$ is the eigenvalue of \mathbf{A} , but it's been repeated.
- To characterize the eigenvectors, we want to solve $(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{0}$.
- Every vector is an eigenvector in this example.

EIGENVALUES AND EIGENVECTORS

Example: Repeated eigenvalues

- Take the matrix $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ as an example.
- We have that $\mathbf{A} - \lambda \mathbf{I} = \begin{bmatrix} 1 - \lambda & 1 \\ 0 & 1 - \lambda \end{bmatrix}$.
- Now, $\det(\mathbf{A} - \lambda \mathbf{I}) = (1 - \lambda)^2$.
- This tells us that $\lambda = 1$ is the eigenvalue of \mathbf{A} , but it's been repeated.
- To characterize the eigenvectors, we want to solve $(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{0}$.
- You should verify that the eigenvectors are all of the form $(c, 0)^T$.
- In this case, the *algebraic* multiplicity of the eigenvalue does not match the *geometric* multiplicity, in which case we call \mathbf{A} a defective matrix (which seems kind of mean).
- \mathbf{A} is an example of a shear matrix, representing the addition of a multiple of one row/column to another. Shear matrices are defective.

Tying into projection matrices:

- If \mathbf{P} is an $N \times N$ projection matrix of rank R , it has R eigenvalues of one and $N - R$ eigenvalues of zero (you will use this fact on a homework assignment when \mathbf{X} is of full column rank).
- Further, there must be an orthogonal matrix \mathbf{Q} such that

$$\mathbf{Q}^T \mathbf{P} \mathbf{Q} = \mathbf{\Lambda},$$

where $\mathbf{\Lambda} = \text{diag}(\mathbf{1}_R, \mathbf{0}_{N-R})$.

- This property has to do with the eigendecomposition, which we will discuss later.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces
- 4 Matrix inverses
- 5 Very special matrices
- 6 Determinants
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices**
- 9 Very important matrix decompositions
- 10 Generalized inverses
- 11 Matrix calculus

Reminder about symmetric matrices:

- Recall that all eigenvalues of a symmetric matrix are real and its eigenvectors are perpendicular.
 - ▶ In the case where every vector is an eigenvector, we can follow the convention of *choosing* the vectors to be perpendicular.
- When \mathbf{A} is symmetric, it also happens that the signs of its pivots match the signs of its eigenvalues, which is computationally useful!

POSITIVE DEFINITE MATRICES

Definition:

- A symmetric matrix \mathbf{A} is said to be positive definite if for every $\mathbf{x} \neq \mathbf{0}$, we have that $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$. Put another way, \mathbf{A} is positive definite if $\mathbf{x}^T \mathbf{A} \mathbf{x} = 0$ only when $\mathbf{x} = \mathbf{0}$.
- The quantity $\mathbf{x}^T \mathbf{A} \mathbf{x}$ is known as a *quadratic form*.
- To see why, consider the 2×2 case:

$$\begin{aligned}\mathbf{x}^T \mathbf{A} \mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} \\ &= a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2.\end{aligned}$$

- When \mathbf{A} is positive definite, the graph of $\mathbf{x}^T \mathbf{A} \mathbf{x}$ would have its minimum at the origin and “face upward” as an elliptical paraboloid.

Properties of positive definite matrices:

- If \mathbf{A} is a positive definite matrix, then the following properties hold:
 - ① \mathbf{A} has all positive eigenvalues (you can use this as a test for positive definiteness).
 - ② \mathbf{A} has all positive pivots.
 - ③ \mathbf{A} has a positive determinant (because the eigenvalues are all positive).

Positive semi-definite matrices: Definition and properties

- A symmetric matrix \mathbf{A} is said to be positive semi-definite if for every \mathbf{x} , we have that $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$.
- This property is sometimes referred to as non-negative definite.
- If \mathbf{A} is positive semi-definite, then \mathbf{A} has all non-negative eigenvalues.
- If \mathbf{A} is positive definite, then \mathbf{A} is positive semi-definite.
- If \mathbf{A} is positive semi-definite but not positive definite, then \mathbf{A} is singular.
- Projection matrices are positive semi-definite because they are symmetric and have non-negative eigenvalues.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces
- 4 Matrix inverses
- 5 Very special matrices
- 6 Determinants
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices
- 9 Very important matrix decompositions**
- 10 Generalized inverses
- 11 Matrix calculus

Decomposition 1: Lower times upper triangular

- We have already learned how to decompose the matrix \mathbf{A} into the form $\mathbf{A} = \mathbf{LU}$ via Gaussian elimination when there were no row exchanges.
- In fact, we already have an example of this from an earlier section. The \mathbf{LU} decomposition will not directly come into play so much in this class, but the next two decompositions we learn certainly will.

Decomposition 2: Eigendecomposition

- An $N \times N$ matrix \mathbf{A} with N linearly independent eigenvectors can be expressed in the form $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$, where:
 - 1 \mathbf{S} is an $N \times N$ matrix of eigenvectors of \mathbf{A} .
 - 2 $\mathbf{\Lambda}$ is an $N \times N$ diagonal matrix with the eigenvalues of \mathbf{A} along the diagonal (in descending order, by convention).
- Interestingly, $\mathbf{A}^r = \mathbf{S}\mathbf{\Lambda}^r\mathbf{S}^{-1}$ (verify this when r is a positive integer).
- When \mathbf{A} is symmetric, the eigenvectors are orthogonal and so, if we scale their lengths to unity, the eigendecomposition can be expressed as $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$.
- Matrices with this factorization are called diagonalizable. Defective matrices are not diagonalizable because they do not have a complete basis of eigenvectors.

Decomposition 2: Eigendecomposition

- The eigendecomposition is useful for quickly computing powers of \mathbf{A} , particularly large ones.

```
set.seed(7345)
K <- 100
A <- matrix(sample(1:(K^2)), nrow = K)
A <- A/rep(rowSums(A), each = K)
```

- An \mathbf{A} generated in this way will almost certainly be diagonalizable.

Decomposition 2: Eigendecomposition

- See below:

```
eA <- eigen(A)
Lambda <- diag(eA$values)
S <- eA$vectors
```

```
Acb.1 <- S %**% Lambda^3 %**% solve(S)
Acb.2 <- A %**% A %**% A
```

```
> Acb.1[1:3,1:3]
      [,1]      [,2]      [,3]
[1,] 0.010315443-0i 0.009203797-0i 0.01187848+0i
[2,] 0.010653311-0i 0.009514142-0i 0.01230300+0i
[3,] 0.009243248-0i 0.008278312-0i 0.01058362+0i
```

```
> Acb.2[1:3,1:3]
      [,1]      [,2]      [,3]
[1,] 0.010315443 0.009203797 0.01187848
[2,] 0.010653311 0.009514142 0.01230300
[3,] 0.009243248 0.008278312 0.01058362
```

```
> sum(abs(Acb.2 - Acb.1))
[1] 1.211255e-12
```

- You can imagine trying to do \mathbf{A}^{100} (for loop?).

Theorem 1.1: Symmetric matrices and projection matrices

A symmetric $N \times N$ matrix, \mathbf{A} , can be realized as a linear combination of N mutually orthogonal projection matrices.

Theorem 1.1: Argument

- Let $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ denote the eigendecomposition of \mathbf{A} (recall that since \mathbf{A} is symmetric, \mathbf{S} is orthogonal).
- Let \mathbf{q}_i denote the i^{th} column of \mathbf{Q} and let λ_i the i^{th} diagonal entry of $\mathbf{\Lambda}$. Then we can rewrite \mathbf{A} as $\lambda_1\mathbf{q}_1\mathbf{q}_1^T + \cdots + \lambda_N\mathbf{q}_N\mathbf{q}_N^T$.
- Now, note that because \mathbf{q}_i is of unit length, $\mathbf{q}_i^T\mathbf{q}_i = 1$, and

$$\mathbf{q}_i\mathbf{q}_i^T = \mathbf{q}_i(\mathbf{q}_i^T\mathbf{q}_i)^{-1}\mathbf{q}_i.$$

- That is to say that $\mathbf{q}_i\mathbf{q}_i^T$ is an $N \times N$ orthogonal projection matrix that projects any vector onto the linear subspace spanned by the eigenvector \mathbf{q}_i (all of which are mutually orthogonal).
- This theorem is known in mathematics by a famous name. What is it?

MATRIX DECOMPOSITIONS

Decomposition 3: Motivation for singular value decomposition (SVD)

- Consider an $N \times K$ matrix \mathbf{A} . Want to find orthonormal vectors in \mathbb{R}^K that, after applying \mathbf{A} , are orthonormal in \mathbb{R}^N :

$$[\mathbf{A}\mathbf{v}_1 \quad \cdots \quad \mathbf{A}\mathbf{v}_K] = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_N] \begin{bmatrix} D_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & D_{22} & \ddots & \vdots & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & D_{PP} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & & & \\ \vdots & \vdots & \ddots & \vdots & & & \\ 0 & 0 & \cdots & 0 & & & \end{bmatrix}$$

- If $N > K$, need the blue zeros to make dimensions match ($P = K$).
- If $N < K$, need the red zeros to make dimensions match ($P = N$).
- If $N = K$, no extra zeros needed ($P = K = N$).

Decomposition 3: Motivation for singular value decomposition (SVD)

- Consider an $N \times K$ matrix \mathbf{A} . Want to find orthonormal vectors in \mathbb{R}^K that, after applying \mathbf{A} , are orthonormal in \mathbb{R}^N :

$$[\mathbf{A}\mathbf{v}_1 \quad \cdots \quad \mathbf{A}\mathbf{v}_K] = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_N] \begin{bmatrix} D_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & D_{22} & \ddots & \vdots & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & D_{PP} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & & & \\ \vdots & \vdots & \ddots & \vdots & & & \\ 0 & 0 & \cdots & 0 & & & \end{bmatrix}$$

- By convention, take $D_{11} > \cdots > D_{PP} \geq 0$ (call them d_1, \dots, d_P , the *singular values*). The first $R = \text{rank}(\mathbf{A})$ are positive.

Decomposition 3: Motivation for singular value decomposition (SVD)

- Consider an $N \times K$ matrix \mathbf{A} . Want to find orthonormal vectors in \mathbb{R}^K that, after applying \mathbf{A} , are orthonormal in \mathbb{R}^N :

$$[\mathbf{A}\mathbf{v}_1 \quad \cdots \quad \mathbf{A}\mathbf{v}_K] = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_N] \begin{bmatrix} D_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & D_{22} & \ddots & \vdots & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & D_{PP} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & & & \\ \vdots & \vdots & \ddots & \vdots & & & \\ 0 & 0 & \cdots & 0 & & & \end{bmatrix}$$

- More compactly, find \mathbf{U} , \mathbf{V} , and \mathbf{D} such that $\mathbf{AV} = \mathbf{UD}$.
- Since \mathbf{V} is orthogonal, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, and we can write $\mathbf{A} = \mathbf{UDV}^T$.

Decomposition 3: Clues to determine \mathbf{U} and \mathbf{D}

- Note first that, assuming such a decomposition is possible, that:

$$\begin{aligned}\mathbf{AA}^T &= \mathbf{UDV}^T(\mathbf{UDV}^T)^T \\ &= \mathbf{UDV}^T\mathbf{VD}^T\mathbf{U}^T \\ &= \mathbf{UDD}^T\mathbf{U}^T\end{aligned}$$

- This is the eigendecomposition of \mathbf{AA}^T , so \mathbf{U} is based on the eigenvectors of \mathbf{AA}^T and \mathbf{D} on the eigenvalues of \mathbf{AA}^T .
- The eigenvalues of \mathbf{AA}^T will be real. Why?
- The eigenvectors of \mathbf{AA}^T will be orthogonal. Why?

Decomposition 3: Clues to determine \mathbf{V}

- Note further that:

$$\begin{aligned}\mathbf{A}^T \mathbf{A} &= (\mathbf{U} \mathbf{D} \mathbf{V}^T)^T \mathbf{U} \mathbf{D} \mathbf{V}^T \\ &= \mathbf{V} \mathbf{D}^T \mathbf{U}^T \mathbf{U} \mathbf{D} \mathbf{V}^T \\ &= \mathbf{V} \mathbf{D}^T \mathbf{D} \mathbf{V}^T\end{aligned}$$

- This is the eigendecomposition of $\mathbf{A}^T \mathbf{A}$, so \mathbf{V} is based on the eigenvectors of $\mathbf{A}^T \mathbf{A}$ and \mathbf{D} on the eigenvalues of $\mathbf{A}^T \mathbf{A}$.
- The eigenvalues of $\mathbf{A}^T \mathbf{A}$ will be real. Why?
- The eigenvectors of $\mathbf{A}^T \mathbf{A}$ will be orthogonal. Why?

Decomposition 3: Singular value decomposition (SVD)

- Although we didn't prove it, it turns out every $N \times K$ matrix \mathbf{A} can be expressed in the form $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, and the previous slides more or less show you how to do it:
 - 1 \mathbf{U} is an $N \times N$ matrix of orthonormal eigenvectors of $\mathbf{A}\mathbf{A}^T$.
 - 2 \mathbf{D} is an $N \times K$ "diagonal" matrix with the positive square-rooted eigenvalues of $\mathbf{A}\mathbf{A}^T$ along the diagonal and are arranged in descending order (called the singular values).
 - 3 \mathbf{V} is a $K \times K$ matrix of orthonormal eigenvectors of $\mathbf{A}^T\mathbf{A}$.
- Must be *careful* when setting up the eigenvectors, because if \mathbf{v} is an eigenvector, so is $-\mathbf{v}$.

Decomposition 3: Singular value decomposition (SVD)

- As an example, suppose I want to find the SVD of the following matrix:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

- We could just cycle through this by brute force using the steps on the prior slide.

Decomposition 3: Singular value decomposition (SVD)

- Step 1: Find the eigenvectors of \mathbf{AA}^T (because \mathbf{A} is symmetric, they should exist and they should be orthogonal).

$$\mathbf{AA}^T = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

- Can identify $\mathbf{u}_1 = (1, 1)^T$ and $\mathbf{u}_2 = (1, -1)^T$ as eigenvectors with corresponding eigenvalues of $\lambda_1 = 3$ and $\lambda_2 = 1$.
 - ▶ Careful!! Choosing eigenvectors this way (and not $(-1, -1)^T$ and/or $(-1, 1)^T$, for instance) affects choice of eigenvectors that form \mathbf{V} .
- Suitably scaling, this suggests that:

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ and } \mathbf{D} = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Decomposition 3: Singular value decomposition (SVD)

- Step 2: Find the eigenvectors of $\mathbf{A}^T \mathbf{A}$ (because \mathbf{A} is symmetric, they should exist and they should be orthogonal).

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

- Can identify eigenvalues $\lambda_1 = 3$, $\lambda_2 = 1$, and $\lambda_3 = 0$ (no accident).
- An eigenvector corresponding to λ_1 could be chosen as, say, $\mathbf{v}_1 = (1, 2, 1)^T$ or $\mathbf{v}_1 = (-1, -2, -1)^T$. Which one is correct?
- Recall that the whole point of this we need $\mathbf{A}\mathbf{v}_1 = \mathbf{u}_1 d_1$ for non-negative d_1 . This only checks out with for the first choice.

Decomposition 3: Singular value decomposition (SVD)

- Step 2: Find the eigenvectors of $\mathbf{A}^T \mathbf{A}$ (because \mathbf{A} is symmetric, they should exist and they should be orthogonal).

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

- Can identify eigenvalues $\lambda_1 = 3$, $\lambda_2 = 1$, and $\lambda_3 = 0$ (no accident).
- An eigenvector corresponding to λ_2 could be chosen as, say, $\mathbf{v}_2 = (1, 0, -1)^T$ or $\mathbf{v}_2 = (-1, 0, 1)^T$. Which one is correct?
- Recall that the whole point of this we need $\mathbf{A}\mathbf{v}_2 = \mathbf{u}_2 d_2$ for non-negative d_2 . This only checks out with for the second choice.

Decomposition 3: Singular value decomposition (SVD)

- Step 2: Find the eigenvectors of $\mathbf{A}^T \mathbf{A}$ (because \mathbf{A} is symmetric, they should exist and they should be orthogonal).

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

- Can identify eigenvalues $\lambda_1 = 3$, $\lambda_2 = 1$, and $\lambda_3 = 0$ (no accident).
- An eigenvector corresponding to λ_3 could be chosen as, say, $\mathbf{v}_3 = (1, -1, 1)^T$ or $\mathbf{v}_3 = (-1, 1, -1)^T$. Which one is correct?
- If I'm not mistaken, it shouldn't make a difference because there is no \mathbf{u}_3 with which I'm trying to synergize the direction of \mathbf{v}_3 . Let's try it both ways and pray?

Decomposition 3: Singular value decomposition (SVD)

- Step 2: Find the eigenvectors of $\mathbf{A}^T \mathbf{A}$ (because \mathbf{A} is symmetric, they should exist and they should be orthogonal).
- Suitably scaling, *one* choice of \mathbf{V} is given by:

$$\mathbf{V} = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}.$$

- The other is given by:

$$\tilde{\mathbf{V}} = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \end{bmatrix}.$$

Decomposition 3: Singular value decomposition (SVD)

- Step 3: Does this check out when we use \mathbf{V} ?

$$\begin{aligned}
 \mathbf{UDV}^T &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \\
 &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{3}/\sqrt{6} & 2\sqrt{3}/\sqrt{6} & \sqrt{3}/\sqrt{6} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1/2 & 1 & 1/2 \\ -1/2 & 0 & 1/2 \end{bmatrix} \\
 &= \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}
 \end{aligned}$$

- And we breathe a huge sigh of relief!

Decomposition 3: Singular value decomposition (SVD)

- Step 3: Does this check out when we use $\tilde{\mathbf{V}}$?

$$\begin{aligned} \mathbf{UD}\tilde{\mathbf{V}}^T &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ -1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{3}/\sqrt{6} & 2\sqrt{3}/\sqrt{6} & \sqrt{3}/\sqrt{6} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \dots \text{umm} \dots \end{aligned}$$

- We can stop here. Do you see what happened?
- Because \mathbf{D} has a right column of zeros, the third *row* of $\tilde{\mathbf{V}}^T$ never got picked up. We could have chosen $\mathbf{v}_3 = (\log(\pi), e^2, -2 \sinh(7.3))^T$ and the math would have worked just fine. But of course \mathbf{V} would then not be orthogonal and wouldn't qualify as a valid SVD.

Decomposition 3: Singular value decomposition (SVD)

- Given the convention of ordering the singular values from largest to smallest, the SVD is unique up to sign changes in the eigenvectors (and, for non-square matrices, re-ordering of the “extraneous” eigenvectors).
- As we just saw, when $N \neq K$, the SVD contains some “extra” information that we can theoretically toss and come up with a decomposition (it’s just not the SVD in particular).

MATRIX DECOMPOSITIONS

Decomposition 3: SVD for wide matrices (*only* the top is an SVD)

$$X = U D V^T$$

$$= U D_{\text{mini}} (V_{\text{mini}})^T$$

Decomposition 3: Singular value decomposition (SVD)

- If \mathbf{A} is positive definite, then the SVD is none other than the eigendecomposition.
 - ▶ Because \mathbf{A} is symmetric, it has real eigenvalues and orthogonal eigenvectors, so it can be written as $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$.
 - ▶ Because \mathbf{A} is positive definite in particular, it has only positive eigenvalues and so $\mathbf{\Lambda}$ is a diagonal matrix of only positive values.

Theorem 1.2: A condition for $\mathbf{A}^T \mathbf{A}$ to be positive definite

If \mathbf{A} is an $N \times K$ matrix of rank K , then $\mathbf{A}^T \mathbf{A}$ is positive definite.

Theorem 1.2: Proof

- We will use the SVD. Let $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$.
- $\mathbf{A}^T\mathbf{A} = (\mathbf{U}\mathbf{D}\mathbf{V}^T)^T(\mathbf{U}\mathbf{D}\mathbf{V}^T) = \mathbf{V}\mathbf{D}^T\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{V}\mathbf{D}^2\mathbf{V}^T$.
- This must be the eigendecomposition of $\mathbf{A}^T\mathbf{A}$ with $\mathbf{\Lambda} = \mathbf{D}^2$; since the rank of \mathbf{A} is K , the diagonal entries of \mathbf{D} are nonzero and so the diagonal entries of $\mathbf{\Lambda} = \mathbf{D}^2$ are positive.
- Therefore, $\mathbf{A}^T\mathbf{A}$ has only positive eigenvalues, which concludes the proof.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces
- 4 Matrix inverses
- 5 Very special matrices
- 6 Determinants
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices
- 9 Very important matrix decompositions
- 10 Generalized inverses**
- 11 Matrix calculus

Motivation:

- As we know, not all matrices are invertible.
- A generalized inverse (or pseudoinverse) of \mathbf{A} , denoted by \mathbf{A}^- , is one that satisfies $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$.
 - ▶ Such a matrix always exists.
 - ▶ If \mathbf{A} is non-singular, then \mathbf{A}^- is unique and $\mathbf{A}^- = \mathbf{A}^{-1}$.
 - ▶ Otherwise, \mathbf{A}^- is not unique.

Example:

- Let $\mathbf{A} = \begin{bmatrix} 2 & 2 & 3 \\ 1 & 0 & 1 \\ 3 & 2 & 4 \end{bmatrix}$, which is singular.
- Both of the following are generalized inverses (g-inverses, henceforth):

$$\mathbf{A}_1^- = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{A}_2^- = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -3/2 & 1/2 \\ 0 & 0 & 0 \end{bmatrix}$$

- You can verify that $\mathbf{AA}_1^- \mathbf{A} = \mathbf{AA}_2^- \mathbf{A} = \mathbf{A}$.

Example:

- Let $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$, which is not even square!

- Both of the following are g-inverses:

$$\mathbf{x}_1^- = [1 \ 0 \ 0 \ 0] \text{ and } \mathbf{x}_2^- = [0 \ 1/2 \ 0 \ 0]$$

- You can verify that $\mathbf{xx}_1^- \mathbf{x} = \mathbf{xx}_2^- \mathbf{x} = \mathbf{x}$.

Properties of g-inverses:

- Let \mathbf{A} have rank R .
- Let \mathbf{A}^- be a g-inverse for \mathbf{A} .
- Let \mathbf{G}_1 and \mathbf{G}_2 be two g-inverses of $\mathbf{A}^T \mathbf{A}$.
- Then, the following hold:
 - ① $\text{rank}(\mathbf{A}^- \mathbf{A}) = \text{rank}(\mathbf{A} \mathbf{A}^-) = \text{rank}(\mathbf{A}) = R$.
 - ② $(\mathbf{A}^-)^T$ is a g-inverse of \mathbf{A}^T .
 - ③ $\mathbf{A} = \mathbf{A} \mathbf{G}_1 \mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{G}_2 \mathbf{A}^T \mathbf{A} = \mathbf{A} (\mathbf{A}^T \mathbf{A})^- \mathbf{A}^T \mathbf{A}$.
 - ④ $\mathbf{G}_1 \mathbf{A}^T$ is a g-inverse of \mathbf{A} ; $\mathbf{A}^- = \mathbf{G}_1 \mathbf{A}^T = (\mathbf{A}^T \mathbf{A})^- \mathbf{A}^T$.
 - ⑤ $\mathbf{A} \mathbf{G}_1 \mathbf{A}^T = \mathbf{A} \mathbf{G}_2 \mathbf{A}^T$ is a projection matrix.
- I'll have you prove some of these results on a homework assignment. These key results will ultimately help us deal with ANOVA problems.

Properties of g-inverse:

- If the system $\mathbf{Ax} = \mathbf{c}$ is consistent, then for any g-inverse \mathbf{A}^- of \mathbf{A} , $\mathbf{x} = \mathbf{A}^- \mathbf{c}$ is a solution (different choices for \mathbf{A}^- yield different solutions to the equation).
- The system $\mathbf{Ax} = \mathbf{c}$ has a solution if and only if $\mathbf{AA}^- \mathbf{c} = \mathbf{c}$ for any g-inverse \mathbf{A}^- of \mathbf{A} .

The Moore-Penrose pseudoinverse:

- This is a g -inverse for \mathbf{A} that satisfies the following criteria:
 - ① $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$.
 - ② $\mathbf{A}^{-}\mathbf{AA}^{-} = \mathbf{A}^{-}$.
 - ③ $\mathbf{A}^{-}\mathbf{A}$ is symmetric.
 - ④ \mathbf{AA}^{-} is symmetric.
- A g -inverse that satisfies the first two properties is *reflexive*.
- The Moore-Penrose pseudoinverse is unique.

The Moore-Penrose pseudoinverse: What is it?

- Let $\mathbf{A} = \mathbf{UDV}^T$ denote the SVD. The Moore-Penrose pseudoinverse is given by $\mathbf{A}^- = \mathbf{V}(\mathbf{D}^+)^T \mathbf{U}^T$, where \mathbf{D}^+ is obtained by replacing each nonzero diagonal element of \mathbf{D} with its reciprocal.
- The properties of the Moore-Penrose pseudoinverse are easy to verify:
 - $\mathbf{AA}^- \mathbf{A} = \mathbf{UDV}^T \mathbf{V}(\mathbf{D}^+)^T \mathbf{U}^T \mathbf{UDV}^T = \mathbf{UDV}^T = \mathbf{A}$.
 - $\mathbf{A}^- \mathbf{AA}^- = \mathbf{V}(\mathbf{D}^+)^T \mathbf{U}^T \mathbf{UDV}^T \mathbf{V}(\mathbf{D}^+)^T \mathbf{U}^T = (\mathbf{VDU}^T)^T = \mathbf{UDV}^T = \mathbf{A}$.
 - $\mathbf{A}^- \mathbf{A} = \mathbf{V}(\mathbf{D}^+)^T \mathbf{U}^T \mathbf{UDV}^T = \mathbf{VV}^T$ (symmetric).
 - $\mathbf{AA}^- = \mathbf{UDV}^T \mathbf{V}(\mathbf{D}^+)^T \mathbf{U}^T = \mathbf{UU}^T$ (symmetric).
- Our first (but NOT our last) demonstration of how useful the SVD is.
- Honestly, the SVD is very often the best possible way to think about what a matrix is and what it does.

The Moore-Penrose pseudoinverse: Example

- Let's determine the Moore-Penrose pseudoinverse for

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

- We did the SVD for this matrix already, with:

$$\mathbf{V} = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}.$$

$$(\mathbf{D}^+)^T = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{U}^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

The Moore-Penrose pseudoinverse: Example

- Leveraging our knowledge of how this will play out, let's convert to "mini" matrices by chopping out the extraneous information.

$$\mathbf{V}_{\text{mini}} = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} \\ 2/\sqrt{6} & 0 \\ 1/\sqrt{6} & 1/\sqrt{2} \end{bmatrix}.$$

$$(\mathbf{D}_{\text{mini}}^+)^T = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{U}^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

The Moore-Penrose pseudoinverse: Example

- Continuing on,

$$\begin{aligned}
 \mathbf{A}^{-} &= \mathbf{V}_{\text{mini}}(\mathbf{D}_{\text{mini}}^{+})^T \mathbf{U}^T \\
 &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} \\ 2/\sqrt{6} & 0 \\ 1/\sqrt{6} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\
 &= \begin{bmatrix} 1/2\sqrt{3} & -1/2 \\ 1/\sqrt{3} & 0 \\ 1/2\sqrt{3} & 1/2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} \\ 1 & -1 \end{bmatrix} \\
 &= \begin{bmatrix} 1/6 - 1/2 & 1/6 + 1/2 \\ 1/3 & 1/3 \\ 1/6 + 1/2 & 1/6 - 1/2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} -1 & 2 \\ 1 & 1 \\ 2 & -1 \end{bmatrix}
 \end{aligned}$$

The Moore-Penrose pseudoinverse: Example

- We should be able to verify at the very least that it is a g-inverse.
- We also should be able to verify that it satisfies the additional properties of the Moore-Penrose pseudoinverse.

TABLE OF CONTENTS

- 1 Notation and operations
- 2 Elimination with matrices
- 3 Understanding the key linear subspaces
- 4 Matrix inverses
- 5 Very special matrices
- 6 Determinants
- 7 Eigenvalues and eigenvectors
- 8 Positive definite matrices
- 9 Very important matrix decompositions
- 10 Generalized inverses
- 11 Matrix calculus**

Basic ideas:

- Let $f(\mathbf{x}) \in \mathbb{R}$ for $\mathbf{x} \in \mathbb{R}^N$, then:

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{bmatrix}.$$

- It is very convenient to be able to write down derivatives of (multivariable) functions in terms of matrices and vectors.
- I am following the “denominator layout” convention for vector/matrix calculus in this course. Since I am treating vectors as column vectors in this course, using denominator layout substantially reduces the number of transposes we’ll need in our notation.

Example:

- If $f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} = \mathbf{x}^T \mathbf{c} = c_1 x_1 + \cdots + c_N x_N$, then:

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix} = \mathbf{c}.$$

- By numerator layout, $\partial f / \partial \mathbf{x} = \mathbf{c}^T$; we are not going to use numerator layout, but don't be confused if when you look it up you see this discrepancy.

Example:

- If $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$, then $\partial f / \partial \mathbf{x} = \mathbf{A}^T$.
- If $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A}$, then $\partial f / \partial \mathbf{x} = \mathbf{A}$.
- This is reminiscent of the rule you already know from single-variable calculus: if $f(x) = ax$, then $f'(x) = a$.
- This example also underscores the point that choosing numerator vs. denominator layout doesn't itself avoid having to keep track of transposes altogether; we just have to choose where we want the transpose.
- Try these on your own!

Example:

- Let \mathbf{A} denote a symmetric $K \times K$ matrix.
- If $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, then $\partial f / \partial \mathbf{x} = 2\mathbf{A} \mathbf{x}$.
- To prove this, write in summation notation:

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^K \sum_{j=1}^K a_{ij} x_i x_j$$

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^K \sum_{j=1}^K \frac{\partial}{\partial x_k} a_{ij} x_i x_j = \sum_{i=1}^K a_{ki} x_i + \sum_{j=1}^K a_{jk} x_j$$

$$= 2 \sum_{i=1}^K a_{ki} x_i = 2\mathbf{a}_k \cdot \mathbf{x}$$

Example:

- If $f(\mathbf{x}) = \|\mathbf{y} - \mathbf{Ax}\|^2$, then $\partial f / \partial \mathbf{x} = -2\mathbf{A}^T(\mathbf{y} - \mathbf{Ax})$.
- To see this, note that:

$$\begin{aligned} \|\mathbf{y} - \mathbf{Ax}\|^2 &= (\mathbf{y} - \mathbf{Ax})^T(\mathbf{y} - \mathbf{Ax}) \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T(\mathbf{Ax}) - (\mathbf{Ax})^T\mathbf{y} + (\mathbf{Ax})^T(\mathbf{Ax}) \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T(\mathbf{Ax}) - (\mathbf{Ax})^T\mathbf{y} + \mathbf{x}^T(\mathbf{A}^T\mathbf{A})\mathbf{x}. \end{aligned}$$

- Apply rules on previous slides and rearrange to arrive at result.
- Notice that this sort of behaves like the power rule and the chain rule.

Example:

- If \mathbf{A} is non-singular with each entry a real-valued function of x , then:

$$\frac{\partial}{\partial x} \mathbf{A}^{-1} = -\mathbf{A}^{-1} \left(\frac{\partial \mathbf{A}}{\partial x} \right) \mathbf{A}^{-1}.$$

- To see this, note that $\mathbf{A}^{-1} \mathbf{A} = \mathbf{I}$ and differentiate both sides:

$$\frac{\partial}{\partial x} \mathbf{A}^{-1} \mathbf{A} = \frac{\partial}{\partial x} \mathbf{I} = \mathbf{0}.$$

$$\Rightarrow \left(\frac{\partial \mathbf{A}^{-1}}{\partial x} \right) \mathbf{A} + \mathbf{A}^{-1} \left(\frac{\partial \mathbf{A}}{\partial x} \right) = \mathbf{0}$$

$$\Rightarrow \left(\frac{\partial \mathbf{A}^{-1}}{\partial x} \right) \mathbf{A} = -\mathbf{A}^{-1} \left(\frac{\partial \mathbf{A}}{\partial x} \right)$$

$$\Rightarrow \frac{\partial \mathbf{A}^{-1}}{\partial x} = -\mathbf{A}^{-1} \left(\frac{\partial \mathbf{A}}{\partial x} \right) \mathbf{A}^{-1}.$$

Basic ideas:

- Let $f(\mathbf{X}) \in \mathbb{R}$ for an $N \times K$ matrix \mathbf{X} , then:

$$\frac{\partial f}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1K}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{N1}} & \cdots & \frac{\partial f}{\partial x_{NK}} \end{bmatrix}.$$

Example:

- For an invertible matrix \mathbf{X} :

$$\frac{\partial}{\partial \mathbf{X}} \log(|\mathbf{X}|) = (\mathbf{X}^{-1})^T.$$

- Try verifying this for a 2×2 or a 3×3 matrix!

Example:

- For a $K \times K$ matrix \mathbf{X} , an $N \times K$ matrix \mathbf{A}_1 , and a $K \times N$ matrix \mathbf{A}_2 :

$$\frac{\partial}{\partial \mathbf{X}} \text{trace}(\mathbf{A}_1 \mathbf{X} \mathbf{A}_2) = \mathbf{A}_1^T \mathbf{A}_2^T.$$

- Try verifying this for a 2×2 or a 3×3 matrix!

Chain rule:

- If $f = f(\mathbf{x}) : \mathbb{R}^K \rightarrow \mathbb{R}$ is a function of \mathbf{x} and $\mathbf{x}(\mathbf{y}) : \mathbb{R}^M \rightarrow \mathbb{R}^K$ is a function of \mathbf{y} , then:

$$\frac{\partial f}{\partial \mathbf{y}} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \frac{\partial f}{\partial \mathbf{x}}.$$

- Notice the “cancellation of $\partial \mathbf{x}$.”
- Notice the reverse-ordering. The dimensions need to match!
 - ▶ $\partial f / \partial \mathbf{y}$ is a length- M vector.
 - ▶ $\partial \mathbf{x} / \partial \mathbf{y}$ is an $M \times K$ matrix.
 - ▶ $\partial f / \partial \mathbf{x}$ is a length- K vector.
- This extends in the way you might expect:

$$\frac{\partial f}{\partial \mathbf{z}} = \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \frac{\partial f}{\partial \mathbf{x}}.$$

Matrix times a vector:

- Suppose \mathbf{A} is an $N \times K$ matrix of constants, $\mathbf{z} = \mathbf{z}(\mathbf{x})$ is a length- K vector, and \mathbf{x} is a length- M vector. Then,

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{z}) = \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \mathbf{A}^T.$$

- Note: $\partial \mathbf{z} / \partial \mathbf{x}$ is an $M \times K$ matrix.
- We'll invoke this implicitly many times.

Matrix times a vector:

- Suppose $\mathbf{A} = \mathbf{A}(\mathbf{x})$ is a diagonal $N \times N$ matrix, $\mathbf{z} = \mathbf{z}(\mathbf{x})$ is a length- N vector, and \mathbf{x} is a length- K vector. Then,

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{z}) = \frac{\partial \mathbf{A}}{\partial \mathbf{x}} \text{diag}(\mathbf{z}) + \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \mathbf{A}.$$

- Note:
 - ▶ $\partial \mathbf{A} / \partial \mathbf{x}$ is a $K \times N$ matrix.
 - ▶ $\text{diag}(\mathbf{z})$ is an $N \times N$ matrix.
 - ▶ $\partial \mathbf{z} / \partial \mathbf{x}$ is a $K \times N$ matrix.
- This formula will come up when we start dealing with observed vs. expected information for generalized linear models.

Matrix times a vector: Example

- Let \mathbf{A} and \mathbf{z} be given by:

$$\mathbf{A} = \begin{bmatrix} x_1 - x_2 & 0 \\ 0 & x_2 + x_1 \end{bmatrix} \quad \text{and} \quad \mathbf{z} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}.$$

- Then,

$$\mathbf{Az} = \begin{bmatrix} x_1^2 - x_1x_2 \\ -x_2^2 - x_1x_2 \end{bmatrix},$$

which has derivative matrix

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{Az}) = \begin{bmatrix} 2x_1 - x_2 & -x_2 \\ -x_1 & -2x_2 - x_1 \end{bmatrix}.$$

- Can we reproduce this using the formula on the previous slide?

Matrix times a vector: Example

- Let \mathbf{A} and \mathbf{z} be given by:

$$\mathbf{A} = \begin{bmatrix} x_1 - x_2 & 0 \\ 0 & x_2 + x_1 \end{bmatrix} \quad \text{and} \quad \mathbf{z} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}.$$

- Then,

$$\frac{\partial \mathbf{A}}{\partial \mathbf{x}} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad \frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

- Therefore,

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

- This verifies the formula.

So far:

- Everything you wanted to know about linear algebra! Maybe more.
- These ideas are central to the methods, concepts, procedures, and algorithms we will implement in this class (sometimes very explicitly and other times more subtly).

Up next:

- Random vectors and matrices.
- The multivariate normal distribution and its nice properties.