

Linear Algebra Lecture Notes¹

Stanford GSE Math Camp 2019

Do Not Distribute Outside GSE

1 Matrices

- A **matrix** is a rectangular set of elements ordered in rows and columns:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1C} \\ a_{21} & a_{22} & \dots & a_{2C} \\ \dots & \dots & \dots & \dots \\ a_{R1} & a_{R2} & \dots & a_{RC} \end{pmatrix}$$

In this matrix there are R rows and C columns.

- If $R = C$ this matrix is a **square** matrix, otherwise it is **rectangular**.
 - The upper left to lower right **diagonal** of a square matrix is called the **main** or **principal** diagonal if it is square. Sometimes this is indicated as $\text{diag}(\mathbf{A})$
 - The lower left to upper right diagonal of a square matrix is called the **opposite** diagonal.
- The generic element of the matrix is a_{ij} . The first subscript i indexes the row where this element is located. The second subscript indexes the column j where this element is located. So a_{25} is the element in the 2nd row and the 5th column of the matrix \mathbf{A} .
- A matrix is said to be **symmetric** if $a_{ij} = a_{ji}$.
- The dimension of a matrix is defined by the number of rows and columns it has. So, a matrix with R rows and C columns has dimension $(R \times C)$. This will sometimes be written under the \mathbf{A} , but is often omitted. The dimension of a matrix \mathbf{A} is indicated as $\text{dim}(\mathbf{A})$.
- A **vector** is a particular type of matrix with either one row (in which case it is called a row vector) or one column (called a column vector). A vector is conventionally assumed to be a column vector if it is not otherwise indicated.

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1C} \end{pmatrix}$$

$$\mathbf{D} = \begin{pmatrix} d_{11} \\ d_{21} \\ \dots \\ d_{R1} \end{pmatrix}$$

Above, \mathbf{B} is a row vector and \mathbf{D} is a column vector.

- A **scalar** is a particular type of matrix with just one row and one column, i.e. a singleton.

¹This lecture is a slightly modified version of one written by Luigi Pistaferri. If you find errors, please let us know so that we may correct them. Thanks!

2 Solving Linear Systems of Equations with Matrices

- **Quick Review of Elementary Equation Operations** The following operations are used to transform the equations of a linear system, while maintaining an *equivalent* linear system:

1. Interchange two equations
2. Multiply two sides of an equation by a constant
3. Add equations to each other

By *equivalence*, we mean that the same values of x_j solve both the original and the transformed systems.

- **Quick Review of Gaussian Elimination:** Gaussian elimination is a method by which we start with some linear system of m equations in n unknowns and use the elementary equation operations to eliminate variables, until we arrive at an equivalent system of the form:

$$\begin{array}{ccccccccc}
 \boxed{a'_{11}x_1} & + & a'_{12}x_2 & + & a'_{13}x_3 & + & \cdots & + & a'_{1n}x_n & = & b'_1 \\
 & & \boxed{a'_{22}x_2} & + & a'_{23}x_3 & + & \cdots & + & a'_{2n}x_n & = & b'_2 \\
 & & & & \boxed{a'_{33}x_3} & + & \cdots & + & a'_{3n}x_n & = & b'_3 \\
 & & & & & & \vdots & & \vdots & & \\
 & & & & & & & & \boxed{a'_{mn}x_n} & = & b'_m
 \end{array}$$

where a'_{ij} denotes the coefficient of the j th unknown in the i th equation after performing elementary equation operations. At each stage of the elimination process, we want to change some coefficient of our system to 0 by adding a multiple of an earlier equation to the given equation. The coefficients $\boxed{a'_{11}}$, $\boxed{a'_{22}}$, etc. in the boxes are referred to as **pivots**², since they are the terms used to eliminate the variables in the rows below them in their respective columns. Once the linear system is in the above reduced form, we then use back substitution to find the values of the x_j s.

- **Quick review of Gauss-Jordan Elimination:** Gauss-Jordan Elimination takes over where Gaussian Elimination left off. Once the linear system is in the *reduced form*, as shown in the Gaussian Elimination section above, we use elementary equation operations and Gaussian elimination to

1. Change the coefficient of the pivot term in each equation to 1
2. Eliminate all terms above each pivot in its column This process results in a reduced, equivalent system. For example, a system of m equations in m unknowns would look something like this:

$$\begin{array}{ccccccc}
 \boxed{x_1} & & & & & & = & b_1^* \\
 & \boxed{x_2} & & & & & = & b_2^* \\
 & & \boxed{x_3} & & & & = & b_3^* \\
 & & & \ddots & & & \vdots & \\
 & & & & \boxed{x_n} & & = & b_m^*
 \end{array}$$

Notice that we have fully solved for the x_j s.

²Pivots do not need to be on the ij diagonal, where $i = j$. Also, there are times when we will eliminate variables in the rows above a pivot

- **Back to Matrices...** Matrices provide an efficient way to represent linear systems such as

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

The above system of equations can be written compactly as:

$$\mathbf{Ax} = \mathbf{b}$$

where

1. The $m \times n$ **coefficient matrix** \mathbf{A} is an array of mn real numbers arranged in m rows by n columns:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

2. The unknown quantities are represented by the vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

3. The right hand side of the linear system is represented by the vector

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Putting this all together, we have:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

- We can solve a system of linear equations using an **Augmented Matrix**. When we append \mathbf{b}

to the coefficient matrix \mathbf{A} , we get the augmented matrix $\hat{\mathbf{A}} = [\mathbf{A}|\mathbf{b}]$

$$\hat{\mathbf{A}} = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_m \end{array} \right)$$

Notice that the \mathbf{x} vector is not explicitly represented in the augmented matrix.

- We can conduct **elementary row operations** to transform some augmented matrix representation of a linear system into another augmented matrix that represents an *equivalent* linear system. Specifically, we can do the following:

1. Interchange two rows
2. Multiply a row by a constant
3. Add two rows to each other

Notice how similar these operations are when compared with elementary equation operations. Let's illustrate each of the row operations below.

- **Interchanging Rows:** Suppose we have the augmented matrix

$$\hat{\mathbf{A}} = \left(\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{array} \right)$$

If we interchange the two rows, we get the following augmented matrix:

$$\left(\begin{array}{cc|c} a_{21} & a_{22} & b_2 \\ a_{11} & a_{12} & b_1 \end{array} \right)$$

which represents a linear system *equivalent* to the matrix $\hat{\mathbf{A}}$.

- **Multiplying by a Constant:** If we multiply the second row of matrix $\hat{\mathbf{A}}$ by a constant c , we obtain the following augmented matrix

$$\left(\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ ca_{21} & ca_{22} & cb_2 \end{array} \right)$$

which represents a linear system *equivalent* to the matrix $\hat{\mathbf{A}}$.

- **Adding Rows:** If we add the first row of matrix $\hat{\mathbf{A}}$ to the second row, we obtain the following augmented matrix:

$$\left(\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{11} + a_{21} & a_{12} + a_{22} & b_1 + b_2 \end{array} \right)$$

which represents a linear system *equivalent* to the matrix $\hat{\mathbf{A}}$.

- **Row Echelon Form:** We use row operations to change coefficients in the augmented matrix to 0 and to put it in a matrix form representing the final linear system of Gaussian elimination. An

augmented matrix of the form

$$\left(\begin{array}{ccccc|c} \boxed{a'_{11}} & a'_{12} & a'_{13} & \cdots & a'_{1n} & b'_1 \\ 0 & \boxed{a'_{22}} & a'_{23} & \cdots & a'_{2n} & b'_2 \\ 0 & 0 & \boxed{a'_{33}} & \cdots & a'_{3n} & b'_3 \\ 0 & 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \boxed{a'_{mn}} & b'_m \end{array} \right)$$

is said to be in *row echelon form*. Each row has more leading zeros than the row preceding it.

- **Reduced Row Echelon Form:** Reduced row echelon form is the matrix representation of a linear system after Gauss-Jordan elimination. For a system of m equations in n unknowns, with no all-zero rows, the reduced row echelon form would be:

$$\left(\begin{array}{ccccc|c} \boxed{1} & 0 & 0 & \cdots & 0 & b_1^* \\ 0 & \boxed{1} & 0 & \cdots & 0 & b_2^* \\ 0 & 0 & \boxed{1} & \cdots & 0 & b_3^* \\ 0 & 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \boxed{1} & b_m^* \end{array} \right)$$

3 Properties of Matrices

1. Suppose you have two matrices \mathbf{A} and \mathbf{B} with generic elements a_{ij} and b_{ij} . Then $\mathbf{A} = \mathbf{B}$ if and only if (also denoted iff) $a_{ij} = b_{ij}$ for all (also denoted \forall) i, j .
2. Suppose you have two matrices \mathbf{A} and \mathbf{B} with generic elements a_{ij} and b_{ij} . You can add or subtract ($\mathbf{A} \pm \mathbf{B}$) them iff they have the same dimension. The resulting matrix \mathbf{C} has a generic element: $c_{ij} = a_{ij} \pm b_{ij}$.

- Example: Let \mathbf{A} and \mathbf{B} be defined:

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

$$\mathbf{B} = \begin{pmatrix} -7 & 0 \\ 4 & 1 \end{pmatrix}$$

Then

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} -6 & 5 \\ 2 & 4 \end{pmatrix}$$

And

$$\mathbf{A} - \mathbf{B} = \begin{pmatrix} 8 & 5 \\ -6 & 2 \end{pmatrix}$$

3. Multiplying \mathbf{A} by a scalar, λ is always a feasible operation. The resulting matrix $\mathbf{B} = \lambda\mathbf{A}$ will have generic elements $b_{ij} = \lambda a_{ij}$.
4. A matrix can be **transposed**. This means exchanging the rows with the columns, i.e., creating a new matrix whose rows are the columns of the original matrix (and vice versa). If this original

matrix is \mathbf{A} its transpose is denoted \mathbf{A}' (you may also see this at \mathbf{A}^T). If this dimension of \mathbf{A} is $(R \times C)$, then the dimension of \mathbf{A}' is $(C \times R)$. If the general element of \mathbf{A} is a_{ij} then the generic element of \mathbf{A}' will be a_{ji} . Note that if \mathbf{A} is symmetric, $\mathbf{A} = \mathbf{A}'$.

- Example:

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

Then,

$$\mathbf{A}' = \begin{pmatrix} 1 & -2 \\ 5 & 3 \end{pmatrix}$$

5. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
6. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
7. $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$
8. $(\lambda + \mu)\mathbf{A} = \lambda\mathbf{A} + \mu\mathbf{A}$

4 Important Matrices

1. The **identity matrix**. This is a square matrix whose elements are 1 along the main diagonal and 0 otherwise. It is denoted as \mathbf{I} or \mathbf{I}_n where the subscript n tells us the dimension, $(n \times n)$ of the matrix.

- Example:

$$\mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

2. The **scalar matrix**:

$$\mathbf{A} = a\mathbf{I}_n = \begin{pmatrix} a & 0 & \dots & 0 \\ 0 & a & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a \end{pmatrix}$$

3. The **diagonal matrix**:

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

Note that $a_{ii} = a$ for all i gives a scalar matrix, and that $a = 1$ gives an identity matrix.

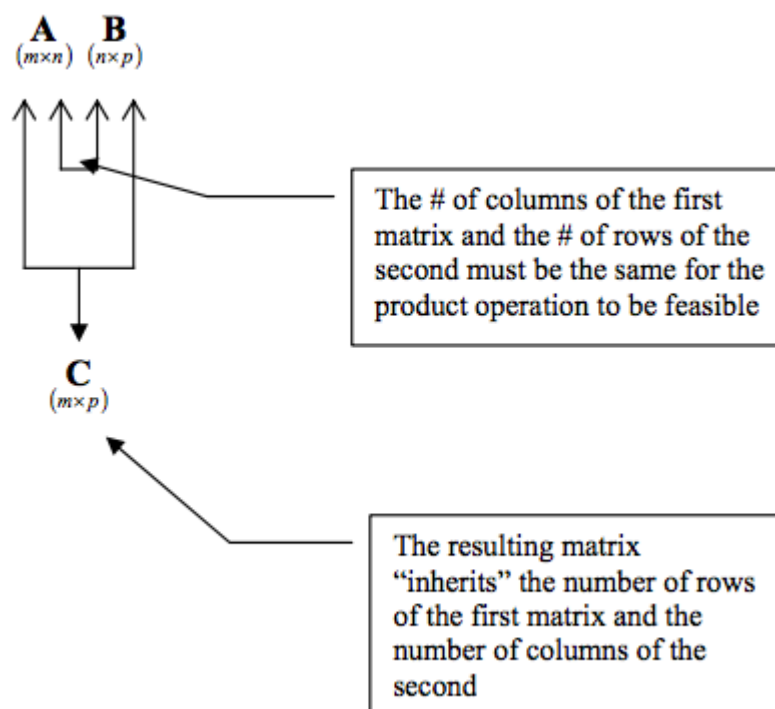
4. An **upper triangular matrix**:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

5. An **idempotent matrix** is such that it is equal to its second power, i.e. $\mathbf{M} = \mathbf{M}^2$. Moreover, we will assume from now on that an idempotent matrix is also symmetric, $\mathbf{M} = \mathbf{M}'$.

5 Products of Matrices

- Suppose you have two matrices \mathbf{A} and \mathbf{B} with generic elements a_{ij} and b_{ij} . You can multiply them, \mathbf{AB} , iff the number of columns of \mathbf{A} equals the number of rows of \mathbf{B} . The order is important (i.e., the commutative property isn't true for matrices!). Thus in general $\mathbf{AB} \neq \mathbf{BA}$. In fact, you may have cases in which \mathbf{AB} is feasible, but \mathbf{BA} is not. The product \mathbf{AB} reads: \mathbf{A} pre-multiplies \mathbf{B} or equivalently \mathbf{B} post-multiplies \mathbf{A} .
- Suppose \mathbf{A} has dimension $(m \times n)$ and \mathbf{B} has dimension $(n \times p)$. Thus \mathbf{AB} is feasible, and the resulting matrix, $\mathbf{C} = \mathbf{AB}$ will have dimension $(m \times p)$. This means that:



The generic element of \mathbf{C} is $c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$. In other words, to obtain c_{ij} , take the i -th row of \mathbf{A} and the j -th column of \mathbf{B} .

- Example 1:

$$\mathbf{A} = \begin{pmatrix} 1 & -3 & 2 & 11 \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} -3 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

$$\mathbf{C} = \mathbf{AB} = (1 \times -3) + (-3 \times 0) + (2 \times 1) + (11 \times 1) = 10$$

- Example 2:

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 4 & 0 \\ -1 & 0 & 5 & 4 \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 3 \\ -4 & 1 & -1 \\ 0 & 1 & 0 \\ 9 & 6 & 0 \end{pmatrix}$$

$$\mathbf{C} = \mathbf{AB} = \begin{pmatrix} \begin{pmatrix} 1 & -2 & 4 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -4 \\ 0 \\ 9 \end{pmatrix} & \begin{pmatrix} 1 & -2 & 4 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 1 \\ 6 \end{pmatrix} & \begin{pmatrix} 1 & -2 & 4 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \\ 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} -1 & 0 & 5 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ -4 \\ 0 \\ 9 \end{pmatrix} & \begin{pmatrix} -1 & 0 & 5 & 4 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 1 \\ 6 \end{pmatrix} & \begin{pmatrix} -1 & 0 & 5 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \\ 0 \\ 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 9 & 1 & 5 \\ 35 & 30 & -3 \end{pmatrix}$$

6 More Properties

1. $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$
2. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
3. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
4. $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$
5. $\mathbf{A}'\mathbf{A}$ and \mathbf{AA}' are both symmetric matrices
6. If

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$$

then (note how important the order is):

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \sum_{i=1}^n x_i^2$$

returns a scalar (specifically, a sum of squares). In contrast:

$$\mathbf{X}\mathbf{X}' = \begin{pmatrix} x_1^2 & x_1x_2 & \dots & x_1x_n \\ x_2x_1 & x_2^2 & \dots & x_2x_n \\ \dots & \dots & \dots & \dots \\ x_1x_n & x_2x_n & \dots & x_n^2 \end{pmatrix}$$

is an $(n \times n)$ matrix.

7. If we have a diagonal matrix:

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

then $\mathbf{X}'\mathbf{A}\mathbf{X} = \sum_{i=1}^n a_{ii}x_i^2$, a weighted sum of squares.

8. The identity vector, \mathbf{i} is a column vector of 1s.

$$\mathbf{i} = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}$$

It can be used to write sums compactly, i.e.,

$$\mathbf{X}'\mathbf{i} = \mathbf{i}'\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} = \sum_{i=1}^n x_i.$$

Note also that $\mathbf{i}'\mathbf{i} = n$. So, the expression for the average of x is $(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\mathbf{X} = \frac{\sum_{i=1}^n x_i}{n}$.

9. When \mathbf{a} and \mathbf{b} are vectors, $\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a}$. This is called the **inner product** or **dot product**.

10. $(\mathbf{A}')' = \mathbf{A}$

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

12. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

13. $\mathbf{I}' = \mathbf{I}$

7 Rank of a Matrix

- The column rank of a matrix \mathbf{A} is the maximum number of linearly independent columns of the matrix \mathbf{A} . A column is linearly independent from the other columns if we cannot write it as a linear combination of the other columns.

– Example:

$$\mathbf{a} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} 3 \\ 4 \\ 8 \end{pmatrix}$$

Then, \mathbf{c} is not linearly independent of \mathbf{a} and \mathbf{b} because $\mathbf{c} = 2\mathbf{a} + \mathbf{b}$.

- The row rank of a matrix \mathbf{A} is the maximum number of linearly independent rows of the matrix \mathbf{A} . The column rank of a matrix \mathbf{A} is the maximum number of linearly independent columns of the matrix \mathbf{A} . The column rank and the row rank are always equal, so we talk about the general rank of a matrix denoted as $rank(\mathbf{A})$.
- In general, finding the rank of a matrix may not be straightforward and we often rely on computers to do the calculation.
- For a rectangular matrix \mathbf{A} with $dim(\mathbf{A}) = (R \times C)$, then $rank(\mathbf{A}) \leq \min(R, C)$.
 - When $rank(\mathbf{A}) = \min(R, C)$ exactly, we say that \mathbf{A} has full rank.
 - Otherwise, if $rank(\mathbf{A}) < \min(R, C)$ we say that \mathbf{A} is rank-deficient.
- Theorems:
 1. $rank(\mathbf{A}) = rank(\mathbf{A}'\mathbf{A}) = rank(\mathbf{A}\mathbf{A}')$
 2. $rank(\mathbf{AB}) \leq \min(rank(\mathbf{A}), rank(\mathbf{B}))$

8 Trace of a Matrix

- The trace is the sum of the elements on the main diagonal and is ONLY defined for a square matrix:

$$tr(\mathbf{A}) = \sum_{i=1}^n a_{ii}$$

- Properties:
 1. $tr(\mathbf{A} \pm \mathbf{B}) = tr(\mathbf{A}) \pm tr(\mathbf{B})$
 2. $tr(\mathbf{I}_k) = k$
 3. $tr(\mathbf{ABC}) = tr(\mathbf{BCA}) = tr(\mathbf{CAB})$
 4. $tr(\lambda) = \lambda$ if λ is a scalar
 5. $tr(\lambda\mathbf{A}) = \lambda tr(\mathbf{A})$ if λ is a scalar
 6. $tr(\mathbf{AA}') = tr(\mathbf{A}'\mathbf{A})$
 7. $tr(\mathbf{A}) = rank(\mathbf{A})$ if \mathbf{A} is idempotent

9 Matrix Inversion

- This is the most complicated operation we do with matrices, and we will often rely on computers to do it.
- Suppose you have a matrix \mathbf{A} with generic elements a_{ij} . Its **inverse**, if it exists, is denoted \mathbf{A}^{-1} . Only square matrices can be inverted, and $\dim(\mathbf{A}) = \dim(\mathbf{A}^{-1})$ if \mathbf{A}^{-1} exists.
- Note that in general the generic element of \mathbf{A}^{-1} will not be $a_{ij}^{-1} = \frac{1}{a_{ij}}$. However, this is true for diagonal matrices (including the identity matrix and the scalar matrix).
- Motivation: Consider three scalars x, y , and z . Suppose we have the following relationship:

$$xy = z$$

And we can solve for x to get

$$x = zy^{-1}$$

(unless $y = 0$, in which case it is not invertible). Note also that $yy^{-1} = 1$.

- The operation of matrix inversion has some analogies with the operation of inverting the scalar y above.
 - In particular, the operation of matrix inversion may sometimes not be feasible. It is certainly not feasible for non-square matrices. However, it may also be infeasible for square matrices that do not satisfy an invertibility condition (detailed below).
 - Moreover, the matrix \mathbf{A} and its inverse \mathbf{A}^{-1} (if it exists) satisfy the identity $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$.

This suggests a first way to find the inverse of a matrix \mathbf{A} (given that \mathbf{A} and \mathbf{I} are known).

- Suppose:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$$

Then its inverse (with general element b_{ij}) must satisfy the identity:

$$\mathbf{A}\mathbf{A}^{-1} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and b_{ij} could be found by solving the system of 9 equations with 9 unknowns:

$$\begin{aligned} b_{11} + 2b_{21} + b_{31} &= 1 \\ b_{12} + 2b_{22} + b_{32} &= 0 \\ &\dots \\ b_{13} + 4b_{33} &= 1 \end{aligned} \tag{1}$$

which is very time consuming.

- Another way to find the inverse of \mathbf{A} is to follow an algorithm.

- To see how the algorithm works, we need to introduce a few more concepts. Assume we have a $(C \times C)$ matrix \mathbf{A} with generic element a_{ij} .

1. The **minor** \mathbf{A}_{ij} is obtained by sweeping out the i -th row and j -th column of the matrix

\mathbf{A} . So if $\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$, then $\mathbf{A}_{23} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{A}_{13} = \begin{pmatrix} -1 & 3 \\ 1 & 0 \end{pmatrix}$.

2. The **co-factor**, which is indicated as $c_{ij} = (-1)^{i+j}|\mathbf{A}_{ij}|$. Note here that $|\mathbf{A}|$ is a symbol used to define the determinant of a matrix.

3. The **determinant** of a matrix \mathbf{A} is defined as $|\mathbf{A}| = \sum_{j=1}^C a_{ij}c_{ij} = \sum_{j=1}^C a_{ij}(-1)^{i+j}|\mathbf{A}_{ij}|$ for any row i or column j of \mathbf{A} . Note that $(-1)^{i+j} = 1$ if $(i+j)$ is even and $(-1)^{i+j} = -1$ if $(i+j)$ is odd.

- The determinant of a scalar is the scalar itself.
- If \mathbf{A} is a (2×2) matrix, the determinant is the product of the elements in the main diagonal minus the product of the elements on the opposite diagonal.

4. Example:

$$\mathbf{B} = \begin{pmatrix} 5 & 1 \\ 4 & -3 \end{pmatrix}$$

The minors are then $\mathbf{B}_{11} = -3$, $\mathbf{B}_{12} = 4$, $\mathbf{B}_{21} = 1$, and $\mathbf{B}_{22} = 5$. And, the determinant is:

$$|\mathbf{B}| = (5 \times (-3)) - (1 \times 4) = -19$$

Alternatively, we could have used the full formula above for any row or column.

Using the first row:

$$|\mathbf{B}| = \sum_{j=1}^C b_{1j}c_{1j} = \sum_{j=1}^C b_{1j}(-1)^{1+j}|\mathbf{B}_{1j}| = (5(-1)^{1+1}(-3)) + (1(-1)^{1+2}4) = -19$$

Using the second row:

$$|\mathbf{B}| = \sum_{j=1}^C b_{2j}c_{2j} = \sum_{j=1}^C b_{2j}(-1)^{2+j}|\mathbf{B}_{2j}| = (4(-1)^{2+1}1) + (-3(-1)^{2+2}5) = -19$$

Using the first column:

$$|\mathbf{B}| = \sum_{j=1}^C b_{ij}c_{ij} = \sum_{j=1}^C b_{ij}(-1)^{i+j}|\mathbf{B}_{ij}| = (5(-1)^{1+1}(-3)) + (4(-1)^{2+1}1) = -19$$

Using the second column:

$$|\mathbf{B}| = \sum_{j=1}^C b_{ij}c_{ij} = \sum_{j=1}^C b_{ij}(-1)^{i+j}|\mathbf{B}_{ij}| = (1(-1)^{1+2}(-3)) + (-3(-1)^{2+2}5) = -19$$

This shows that using any row or column yields the same result. Note that the determinant is always a scalar.

Steps to finding the inverse:

1. Calculate the minors \mathbf{A}_{ij} of the matrix and the minors' determinants $|\mathbf{A}_{ij}|$.
 2. Calculate the co-factors c_{ij} associated to each minor using $c_{ij} = (-1)^{i+j}|\mathbf{A}_{ij}|$.
 3. Create a matrix of co-factors \mathbf{C} with generic elements c_{ij} .
 4. Transpose \mathbf{C} . This is so-called **adjoint** matrix of \mathbf{A} , i.e. $\mathbf{C}' = \text{adj}(\mathbf{A})$
 5. Calculate the determinant $|\mathbf{A}|$ using any row or column of \mathbf{A} .
 6. The define $\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\text{adj}(\mathbf{A})$
- Example:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$$

Step 1:

$$\begin{aligned} \mathbf{A}_{11} &= \begin{pmatrix} 3 & 2 \\ 0 & 4 \end{pmatrix}, \mathbf{A}_{12} = \begin{pmatrix} -1 & 2 \\ 1 & 4 \end{pmatrix}, \mathbf{A}_{13} = \begin{pmatrix} -1 & 3 \\ 1 & 0 \end{pmatrix} \\ \mathbf{A}_{21} &= \begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix}, \mathbf{A}_{22} = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}, \mathbf{A}_{23} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} \\ \mathbf{A}_{31} &= \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}, \mathbf{A}_{32} = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}, \mathbf{A}_{33} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} \end{aligned}$$

Step 2:

$$\begin{aligned} |\mathbf{A}_{11}| &= 12, |\mathbf{A}_{12}| = -6, |\mathbf{A}_{13}| = -3, \\ |\mathbf{A}_{21}| &= 8, |\mathbf{A}_{22}| = 3, |\mathbf{A}_{23}| = -2, \\ |\mathbf{A}_{31}| &= 1, |\mathbf{A}_{32}| = 3, |\mathbf{A}_{33}| = 5 \end{aligned}$$

and so:

$$\begin{aligned} c_{11} &= 12, c_{12} = 6, c_{13} = -3 \\ c_{21} &= -8, c_{22} = 3, c_{23} = 2 \\ c_{31} &= 1, c_{32} = -3, c_{33} = 5 \end{aligned}$$

Step 3:

$$\mathbf{C} = \begin{pmatrix} 12 & 6 & -3 \\ -8 & 3 & 2 \\ 1 & -3 & 5 \end{pmatrix}$$

Step 4:

$$\text{adj}(\mathbf{A}) = \mathbf{C}' = \begin{pmatrix} 12 & -8 & 1 \\ 6 & 3 & -3 \\ -3 & 2 & 5 \end{pmatrix}$$

Step 5:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$$

$$|\mathbf{A}| = \sum_{j=1}^C a_{ij} c_{ij} = \sum_{j=1}^C a_{ij} (-1)^{i+j} |\mathbf{A}_{ij}|$$

$$|\mathbf{A}| = (1 \times 12) + (2 \times 6) + (1 \times (-3)) = 21$$

Step 6:

$$\mathbf{A}^{-1} = \frac{1}{21} \begin{pmatrix} 12 & -8 & 1 \\ 6 & 3 & -3 \\ -3 & 2 & 5 \end{pmatrix} = \begin{pmatrix} \frac{4}{7} & \frac{-8}{21} & \frac{1}{21} \\ \frac{2}{7} & \frac{1}{7} & \frac{-1}{7} \\ \frac{-1}{7} & \frac{2}{21} & \frac{5}{21} \end{pmatrix}$$

Step 7: Check that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$

(Do this on your own)

- A matrix that is not invertible is said to be **singular**. A non-invertible matrix is one with a determinant equal to zero. Equivalently, it does not have full rank.
- Consider a matrix that is not full rank, like:

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

in which the second column can be obtained by multiplying the first column by 3. Here, the determinant is $1 \times 6 - 2 \times 3 = 0$ and hence the matrix is not invertible.

- Properties for Inverses:

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
2. $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$
3. $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

10 Eigenvalues and Eigenvectors

- Consider the following problem. We want to find the values of the scalar λ for which there exist vectors $\mathbf{x} \neq \mathbf{0}$ satisfying the equation:

$$\mathbf{Ax} = \lambda \mathbf{x}$$

- The values of λ that solve this equation are called the characteristics roots (or eigenvalues) of a matrix \mathbf{A} . If \mathbf{A} is a $(n \times n)$ matrix, there are n eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$

$$\mathbf{Ax} = \lambda \mathbf{Ix}$$

$$\mathbf{Ax} - \lambda \mathbf{Ix} = \mathbf{0}$$

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

$$|\mathbf{A} - \lambda \mathbf{I}| = 0$$

Why? Assume that instead $\mathbf{B} = (\mathbf{A} - \lambda \mathbf{I})$ is invertible. Then we can write $\mathbf{Bx} = \mathbf{0}$, $\mathbf{B}^{-1}\mathbf{Bx} = \mathbf{0}$ and $\mathbf{x} = \mathbf{0}$, which is a contradiction (we assumed that $\mathbf{x} \neq \mathbf{0}$).

- The vectors associated to the eigenvalues, λ_i 's, are called eigenvectors (typically, denoted $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$).
- Consider the following example:

$$\mathbf{A} = \begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}$$

Let's find $|\mathbf{A} - \lambda \mathbf{I}|$. We have:

$$\left| \begin{pmatrix} 2-\lambda & 2 \\ 2 & -1-\lambda \end{pmatrix} \right| = -(2-\lambda)(1+\lambda) - 4 = 0$$

$$\lambda^2 - \lambda - 6 = 0$$

$$\lambda_1 = 3, \lambda_2 = -2$$

Now, let's find the eigenvectors. For $\lambda = 3$ we have:

$$\begin{pmatrix} -x_1 + 2x_2 \\ 2x_1 - 4x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

i.e., $x_1 = 2x_2$. For $\lambda = -2$ we have $x_1 = \frac{-1}{2}x_2$. We need to impose a normalization constraint (otherwise we cannot pin down x_1 and x_2 separately). The typical normalization constraint is: $\mathbf{x}'\mathbf{x} = 1$, and hence we get solutions.

When $\lambda = 3$: $x_1^2 + x_2^2 = 1$, or $x_1 = \frac{2}{\sqrt{5}}$, $x_2 = \frac{1}{\sqrt{5}}$

When $\lambda = -2$: $x_1^2 + x_2^2 = 1$, or $x_1 = \frac{-1}{\sqrt{5}}$, $x_2 = \frac{2}{\sqrt{5}}$

Hence:

$$\mathbf{v}_1 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} \frac{-1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$$

For a symmetric matrix, characteristic vectors corresponding to distinct characteristics roots are

pairwise orthogonal. That is $\mathbf{v}_i \mathbf{v}'_i = 1$ and $\mathbf{v}_i \mathbf{v}'_j = 0$. This is easy to verify in our example above. Try doing it on your own.

The $(n \times n)$ matrix of eigenvectors is:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{pmatrix}$$

Now, notice that

$$\mathbf{Q}' = \begin{pmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \\ \dots \\ \mathbf{v}_n' \end{pmatrix}$$

And, therefore

$$\mathbf{Q}\mathbf{Q}' = \begin{pmatrix} \mathbf{v}_1' \mathbf{v}_1 & \mathbf{v}_1' \mathbf{v}_2 & \dots & \mathbf{v}_1' \mathbf{v}_n \\ & \mathbf{v}_2' \mathbf{v}_2 & \dots & \mathbf{v}_2' \mathbf{v}_n \\ & & \dots & \mathbf{v}_3' \mathbf{v}_n \\ & & & \mathbf{v}_n' \mathbf{v}_n \end{pmatrix}$$

And so also $\mathbf{Q}' = \mathbf{Q}^{-1}$. Therefore \mathbf{Q} is said to be an orthogonal matrix.

Now write $\mathbf{A}\mathbf{Q}$:

$$\mathbf{A}\mathbf{Q} = \begin{pmatrix} \mathbf{A}\mathbf{v}_1 & \mathbf{A}\mathbf{v}_2 & \dots & \mathbf{A}\mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \lambda \mathbf{v}_1 & \lambda \mathbf{v}_2 & \dots & \lambda \mathbf{v}_n \end{pmatrix} = \mathbf{Q}\mathbf{\Lambda}$$

Where $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues on the main diagonal. We conclude that:

$$\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{Q}'\mathbf{Q}\mathbf{\Lambda} = \mathbf{\Lambda}$$

The operation is called the diagonalization of a matrix \mathbf{A} . We have found a matrix \mathbf{Q} such that the transformation $\mathbf{Q}'\mathbf{A}\mathbf{Q}$ produces a diagonal matrix with \mathbf{A} 's characteristic roots along the main diagonal.

- Useful things to know:
 1. For a square matrix \mathbf{A} , the product of the eigenvalues is equal to the determinant of the matrix.
 2. For a square matrix \mathbf{A} , the rank of \mathbf{A} is equal to the number of nonzero eigenvalues.
- Consider now an idempotent matrix \mathbf{M} . For example:

$$\mathbf{M} = \begin{pmatrix} \frac{1}{2} & \frac{-1}{2} \\ \frac{11}{2} & \frac{1}{2} \end{pmatrix}$$

It's easy to show that the only eigenvalues that solve the characteristics equation $\mathbf{M}\mathbf{x} = \lambda\mathbf{x}$ are $\lambda_1 = 0$ and $\lambda_2 = 1$. In general, one can show that an idempotent $(n \times n)$ matrix of rank $k \leq n$ has k eigenvalues equal to 1 and $(n - k)$ eigenvalues equal to 0. What about eigenvectors of \mathbf{M} ? Let's find them for our example:

$$\mathbf{v}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix}$$

Now write \mathbf{MQ} :

$$\mathbf{MQ} = \begin{pmatrix} \mathbf{M}\mathbf{v}_1 & \mathbf{M}\mathbf{v}_2 & \dots & \mathbf{M}\mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \dots & \lambda_n\mathbf{v}_n \end{pmatrix} = \mathbf{Q}\mathbf{\Lambda}$$

Where in our case $\mathbf{\Lambda}$ is a diagonal matrix that has k 1's and $(n - k)$ 0's in the main diagonal. And also

$$\mathbf{Q}'\mathbf{MQ} = \mathbf{\Lambda}$$

is the diagonalization of \mathbf{M} .

11 Matrix Calculus

- Suppose you have the following:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix}, \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_k \end{pmatrix}, \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix}$$

- And then you have the mapping $y = f(\beta)$. The objective is to find the derivative of y with respect to β . This can be defined as the vector:

$$\frac{\partial y}{\partial \beta} = \frac{\partial f(\beta)}{\partial \beta} = \begin{pmatrix} \frac{\partial f(\beta)}{\partial \beta_1} \\ \frac{\partial f(\beta)}{\partial \beta_2} \\ \dots \\ \frac{\partial f(\beta)}{\partial \beta_k} \end{pmatrix}$$

This is the so-called vector of gradients.

- We will consider the following mappings:

$$y = f(\beta) = \mathbf{c}'\beta = \beta'\mathbf{c} = \sum_{i=1}^k \beta_i c_i$$

$$y = f(\beta) = \beta'\mathbf{A}\beta$$

- The corresponding gradient vectors are:

$$\frac{\partial y}{\partial \beta} = \frac{\partial \beta'\mathbf{c}}{\partial \beta} = \frac{\partial \mathbf{c}'\beta}{\partial \beta} = \begin{pmatrix} \frac{\partial(\beta_1 c_1 + \beta_2 c_2 + \dots + \beta_k c_k)}{\partial \beta_1} \\ \frac{\partial(\beta_1 c_1 + \beta_2 c_2 + \dots + \beta_k c_k)}{\partial \beta_2} \\ \dots \\ \frac{\partial(\beta_1 c_1 + \beta_2 c_2 + \dots + \beta_k c_k)}{\partial \beta_k} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_k \end{pmatrix} = \mathbf{c}$$

$$\frac{\partial y}{\partial \beta} = \frac{\partial \beta'\mathbf{A}\beta}{\partial \beta} = \mathbf{A}\beta + \mathbf{A}'\beta = 2\mathbf{A}\beta$$

This last equation is true only if \mathbf{A} is a symmetric matrix.

12 The OLS Problem

- Consider a model with k regressors for N individuals ($N > k$). Write the regression model for each individual:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} \\ &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} \end{aligned}$$

- Note that in x_{ij} , the i subscript indexes the individual or unit, and the j subscript the variable. This can be put in matrix format as:

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{pmatrix}$$

Or more compactly as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

- The OLS problem is to solve:

$$\min_{\boldsymbol{\beta}} f(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^n \hat{\mathbf{u}}_i^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{y}' - \hat{\boldsymbol{\beta}}' \mathbf{X}')(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}' \mathbf{y} - \mathbf{y}' \mathbf{X} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} + \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}$$

- To find the minimum, take the derivative of this objective and set equal to zero (i.e. $\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = 0$):

$$\begin{aligned} 0 &= \frac{\partial(\mathbf{y}' \mathbf{y} - \mathbf{y}' \mathbf{X} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} + \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \\ 0 &= \frac{\partial(\mathbf{y}' \mathbf{X} \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} - \frac{\partial(\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y})}{\partial \hat{\boldsymbol{\beta}}} + \frac{\partial(\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \\ 0 &= -\mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{y} + 2\mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} = -2\mathbf{X}' \mathbf{y} + 2\mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} = -\mathbf{X}' \mathbf{y} + \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} \end{aligned}$$

Or:

$$\mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}' \mathbf{y}$$

- If $\mathbf{X}' \mathbf{X}$ is invertible, then we can pre-multiply both sides by $(\mathbf{X}' \mathbf{X})^{-1}$ to yield the OLS solution:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$$

- When $\mathbf{X}' \mathbf{X}$ is not invertible, does it prevent solving the OLS problem? Only when it does not have full rank. Recall the theorem:

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}' \mathbf{A}) = \text{rank}(\mathbf{A} \mathbf{A}')$$

- This means that the rank of $\mathbf{X}' \mathbf{X}$ is equal to the rank of the matrix of the covariates \mathbf{X} . Thus,

if we have columns of \mathbf{X} that are not linearly independent, the matrix $\mathbf{X}'\mathbf{X}$ cannot be inverted. This is what we call multicollinearity.