Matrix Algebra

Hormoz Ramian
Imperial College London
Department of Finance

h.ramian@imperial.ac.uk.

Overview

Outline

- Vectors and Operations
- Vectors Differential Calculus
- Applications to Least Squares

Motivation

- Most economic and econometric models analyze several variables
- Compact notations and useful operations to multivariate problems
- Computational advantages

Vectors

Compact notations (numbers, characteristics or mathematically quantifiable concepts). Suppose there are n commodities in a market, we can show collection of

- trade quantities $(q_1, q_2, \ldots, q_n)$, and
- their prices $(p_1, p_2, \ldots, p_n)$

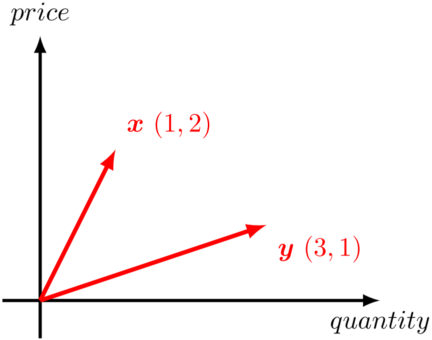
in the following way:

$$\boldsymbol{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} \qquad \boldsymbol{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}$$

1. Boldface notation (lower or upper case)
2. Each entry is referred to as an *element* or a *co-ordinate*
3. Size or Dimension: $(n \times 1)$ or $(1 \times n)$
4. Each element has a domain such as real numbers $\mathbb{R}$
5. Vectors inherit characteristics of elements e.g. $\boldsymbol{q} \in \mathbb{R}^n_+$ and $\boldsymbol{p} \in \mathbb{R}^n_+$ (**vector space**)

Vectors: Geometry

Compact way to generalize scalars



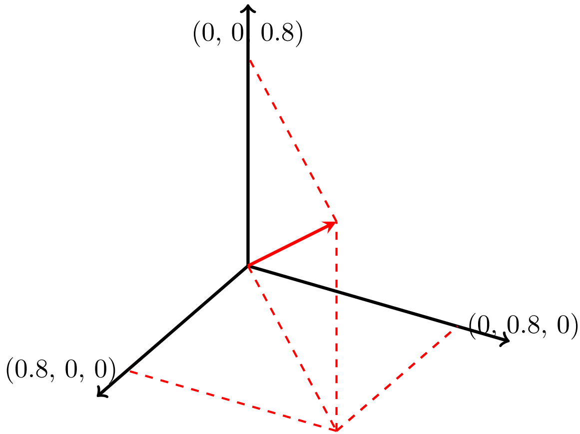
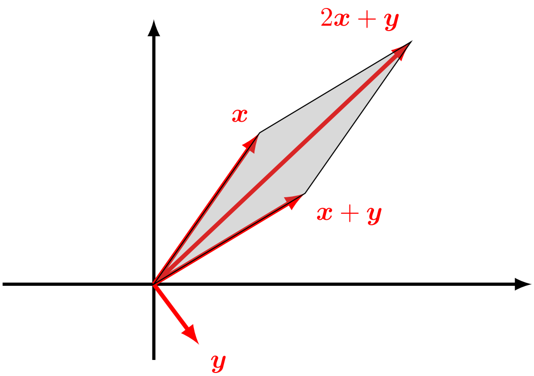
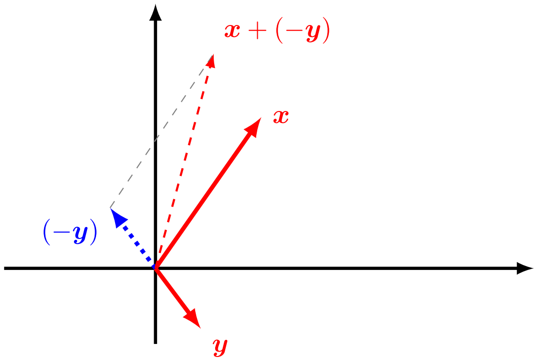


Figure: The left figure illustrates two 2-dimensional vectors. The first element of each vector measures the quantity and the second element measures price on the horizontal and vertical axes, respectively. The figure on the right shows a more general case where an additional characteristic, e.g. currency, is depicted on the third axis (the figure on the right shows only one vector in $\mathbb{R}^3_+$ space.)

Addition and Subtraction of vectors: Element-by-element addition and subtraction

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \boldsymbol{x}+\boldsymbol{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \qquad \boldsymbol{x}-\boldsymbol{y} = \begin{bmatrix} x_1 + (-y_1) \\ x_2 + (-y_2) \\ \vdots \\ x_n + (-y_n) \end{bmatrix}$$





Data

Suppose we observe

- *education* ($\boldsymbol{x}$) in years, *income* ($\boldsymbol{y}$) in log-dollars and *experience* ($\boldsymbol{z}$) in years
- each vector is length n
- each $i \in n$ represents person i

$$\boldsymbol{x} = \begin{bmatrix} 12 \\ 16 \\ \vdots \\ 21 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 4.23 \\ 4.69 \\ \vdots \\ 5.47 \end{bmatrix} \qquad \boldsymbol{z} = \begin{bmatrix} 4 \\ 4 \\ \vdots \\ 4 \end{bmatrix}$$

Reshape to construct

- attribute $j = 1, 2, 3$ shows education, income and experience
- each column vector $\boldsymbol{A}_i$ represents person i's attributes

$$\boldsymbol{A}_1 = \begin{bmatrix} 12 \\ 4.23 \\ 4 \end{bmatrix} \qquad \boldsymbol{A}_2 = \begin{bmatrix} 16 \\ 4.69 \\ 4 \end{bmatrix} \qquad \dots \qquad \boldsymbol{A}_n = \begin{bmatrix} 21 \\ 5.47 \\ 4 \end{bmatrix}$$

Axioms

Let $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ denote arbitrary n-dimensional vectors (r and s real-valued scalars):

1. Commutativity: $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$
2. Associativity of vector addition: $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$
3. Existence of "*additive inverse*": $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}_n$
4. Associativity of scalar multiplication: $r \times (s \times \boldsymbol{x}) = (r \times s) \times \boldsymbol{x}$
5. Distributivity of scalar sums: $(r + s) \times \boldsymbol{x} = r \times \boldsymbol{x} + s \times \boldsymbol{x}$
6. Distributivity of vector sums: $r \times (\boldsymbol{x} + \boldsymbol{y}) = r \times \boldsymbol{x} + r \times \boldsymbol{y}$

Transpose and Product

Transpose transformation requires switching the row (n) and column (1) identifiers:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{x}' = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$$

Product (inner) of two vectors yields a scalar:

$$z \;=\; \boldsymbol{x}'\boldsymbol{y} \;=\; \sum_{i=1}^{n} x_i \times y_i \;=\; \begin{bmatrix} x_1 \times y_1 + x_2 \times y_2 + \ldots + x_n \times y_n \end{bmatrix}$$

For any (real) vector, the norm measures the magnitude or length of the vector:

$$||\boldsymbol{x}|| \;=\; \sqrt{\sum_{i=1}^{n} x_i^2} \;=\; \sqrt{\boldsymbol{x}'\boldsymbol{x}}$$

Equivalent notations: $||\boldsymbol{x}||$ or $||\boldsymbol{x}||_2$ or $||\boldsymbol{x}||^2$.

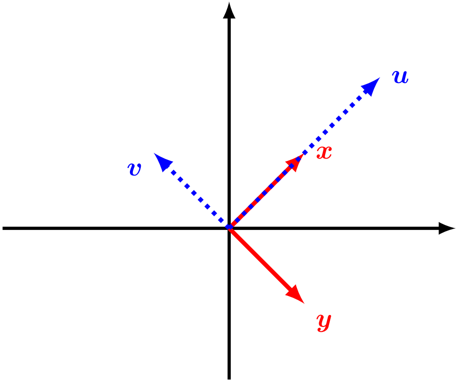
Orthogonality

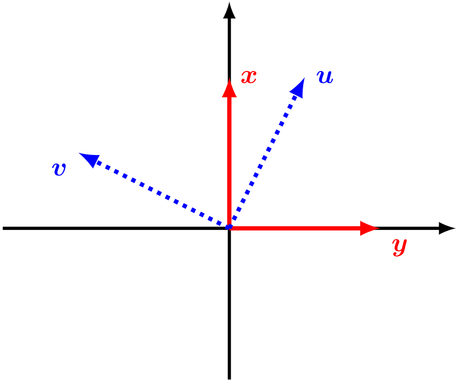
Orthogonality (using inner product):

$$\boldsymbol{x}'\boldsymbol{y} = 0$$

If two vectors are orthogonal, then

- any *scaled transformation* of them are also orthogonal to each other
- suppose λ_1 and λ_2 are two real-valued scalars,
- two transformed vectors $\lambda_1\boldsymbol{x}$ and $\lambda_2\boldsymbol{y}$ are also orthogonal ($\lambda_1\boldsymbol{x} \perp \lambda_2\boldsymbol{y}$) iff $\boldsymbol{x} \perp \boldsymbol{y}$.





Functions

Two generic possibilities

- scalar function $\{f \,\big|\, f : \mathbb{R}^n \to \mathbb{R}\}$
- vector function $\{f \,\big|\, f : \mathbb{R}^n \to \mathbb{R}^n\}$

For example, for any arbitrary real-valued n-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and scalars λ_i:

$$\begin{aligned}
\underset{(1\times 1)}{f(\boldsymbol{x})} &= 1 + \lambda_0.\boldsymbol{x}'\boldsymbol{x} \\
\underset{(n\times 1)}{g(\boldsymbol{x}, \boldsymbol{y})} &= \frac{\lambda_1}{\lambda_1 + \lambda_2}\boldsymbol{x} + \frac{\lambda_2}{\lambda_1 + \lambda_2}\boldsymbol{y} \\
\underset{(n\times 1)}{h(\boldsymbol{x})} &= ||\boldsymbol{x}||_2^{-1}\boldsymbol{x}
\end{aligned}$$

$h(\boldsymbol{x})$ is called normalized vector (each element divided by vector's norm and $||h(\boldsymbol{x})|| = 1$)

Vectors of vectors

Generalization of vectors

- A vector is a collection of elements and each element can also be itself a vector
- Suppose $\boldsymbol{X}$ is an n-dimensional column vector:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \dots & \boldsymbol{x_n} \end{bmatrix}'$$

where each individual element $\boldsymbol{x}_i$ is a k-dimensional row vector:

$$\boldsymbol{x}_i = \begin{bmatrix} x_{i,1} & x_{i,2} & \dots & x_{i,k} \end{bmatrix}$$

Matrices

Generalize vector notations

- a matrix is an $n \times k$ array of scalars, vectors or even smaller matrices.

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

1. boldface notation (lower or upper case)
2. each entry is referred to as an *element* or a *co-ordinate*
3. size or dimension: $(n \times k)$
4. each element has a domain e.g. real numbers $\mathbb{R}$
5. matrices inherit their elements' characteristics $\boldsymbol{X} \in \mathbb{R}^{n \times k}$
6. entries per element required

Variable	Vector Notation	Matrix Notation
Scalar	1×1 vector	1×1 matrix
—	n-dimensional vector	$n \times 1$ or $1 \times n$ matrix

Characterizations

General notation

- matrix $\boldsymbol{X} = \{x_{i,j}\}$
- $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, k$

$$\boldsymbol{X} \quad = \quad \{x_{i,j}\} \quad = \quad \{x_{i,j}\}_{i=1,\ldots,n}^{j=1,\ldots,k} \quad = \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \in \mathbb{R}^{n \times k}$$

Basic characterizations,

1. **Zero** or **Null** matrix ($\mathbf{0}_{n\times k}$ or $\mathbf{O}_{n\times k}$) with zeros everywhere
2. **One** matrix ($\mathbf{1}_{n\times k}$) with ones everywhere
3. **Square** matrix has equal number of rows and columns $n = k$

The set of elements with $i = j$ is called the main diagonal of a (square) matrix.

4. **Symmetric** matrix is *square* and neutral to swapping rows and column indexes $\boldsymbol{X} = \boldsymbol{X}' = (\boldsymbol{X}')'$
5. **Diagonal** matrix is *symmetric* with entries on the main diagonal and zeros elsewhere
6. **Identity** matrix ($\mathbf{I}_n$) is *diagonal*, with ones on the main diagonal and zeros elsewhere
7. **Upper (Lower) Triangular** is *square* with zeros below (above) the main diagonal
8. **Block Diagonal** generalized *diagonal* matrix but each element is a *square* matrix
9. **(Block) Tridiagonal** matrix has entries on/above/below its main diagonal, and zeros elsewhere

Examples:

- $\mathbf{0}_n = \mathbf{0}'_n$ but $\mathbf{0}_{n\times m} \neq \mathbf{0}'_{n\times m}$
- $\mathbf{1}_n = \mathbf{1}'_n$, $\mathbf{I} = \mathbf{I}'$
- An $n \times k$ matrix $\boldsymbol{X}$ can be transposed to a $k \times n$ matrix $\boldsymbol{X}'$ (or $\boldsymbol{X}^\top$ or $\boldsymbol{X}^T$), e.g.

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \text{ then } \boldsymbol{X}' = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- Symmetric matrix (square)

$$\boldsymbol{Y} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ then } \boldsymbol{Y}' = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \boldsymbol{X}\boldsymbol{X}' = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Equality, Addition and Subtraction

1. Commutativity

$$\boldsymbol{X} + \boldsymbol{Y} = \boldsymbol{Y} + \boldsymbol{X} \tag{1}$$

2. Associativity of vector addition

$$(\boldsymbol{X} + \boldsymbol{Y}) + \boldsymbol{Z} = \boldsymbol{X} + (\boldsymbol{Y} + \boldsymbol{Z}) \tag{2}$$

3. Existence of "*additive inverse*"

$$\boldsymbol{X} + (-\boldsymbol{X}) = \mathbf{0}_n \tag{3}$$

4. Associativity of scalar multiplication

$$r \times (s \times \boldsymbol{X}) = (r \times s) \times \boldsymbol{X} \tag{4}$$

5. Distributivity of scalar sums

$$(r + s) \times \boldsymbol{X} = r \times \boldsymbol{X} + s \times \boldsymbol{X} \tag{5}$$

6. Distributivity of vector sums

$$r \times (\boldsymbol{X} + \boldsymbol{Y}) = r \times \boldsymbol{X} + r \times \boldsymbol{Y} \tag{6}$$

Linear Dependence

Definition

- A set of m individual n-dimensional vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ are linearly independent, if and only if none of the vectors in the set can be expressed as a linear combination of the rest.
- If at least one vector $\boldsymbol{x}_i$ $(i = 1, \ldots, m)$ can be expressed as a linear combination of the rest, then the set is linearly dependent:

$$\sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i \quad = \quad \lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 + \ldots \lambda_m \boldsymbol{x}_m \quad = \quad \mathbf{0} \quad \text{for some} \quad \lambda_j \neq 0$$

otherwise the set is linearly independent. Intuitively, linear independence indicates that one vector can be constructed using linear transformation of another vector.

- Suppose $\boldsymbol{x} = [1 \;\; 2]'$ and $\boldsymbol{y} = [3 \;\; 6]'$ then $\boldsymbol{x} = 3\boldsymbol{y}$, thus satisfying the sum $\boldsymbol{x} - 3\boldsymbol{y} = \mathbf{0}'$.
- Linear dependence is the basis for collinearity (multi-linearity) in econometrics
- e.g. temperature measurements in Fahrenheit and Celsius ($C = \frac{F - 32.00}{1.80}$) is only one variable.

Linear Combinations

Let $\boldsymbol{X}$ denote an arbitrary $n \times k$ matrix with $k \leq n$. It is sometimes possible to express one of the rows (columns) of $\boldsymbol{X}$ in terms of one or a linear combination of other rows (columns) of $\boldsymbol{X}$:

$$\boldsymbol{X} = \begin{array}{c c} & \begin{array}{cc} C_1 & C_2 \end{array} \\ \begin{array}{c} R_1 \\ R_2 \end{array} & \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \end{array}$$

- R_1 and R_2 are row labels and C_1 and C_2 are column labels
- express $C_1 = 2C_2$ when $x_{11} = 2x_{12}$ and $x_{21} = 2x_{22}$
- express $R_1 = 2R_2$ when $x_{11} = 2x_{21}$ and $x_{12} = 2x_{22}$
- linear combination can [cannot] be established when rows (columns) are **linearly dependant [independent]**
- A square matrix without any possible linear combination is a full rank matrix

$$\boldsymbol{X} = \begin{array}{c c} & \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} & \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 12 \end{pmatrix} \end{array} \qquad \boldsymbol{I} = \begin{array}{c c} & \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

Elementary Matrix Operations

A row (columns) or a linear combination of rows (columns) of a matrix can be added to (subtracted from) any other rows (columns) with that matrix. Such operations change the original matrix but leave the span of the matrix intact. For example:

$$\boldsymbol{X} = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \end{array} \begin{array}{c} \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 12 \end{pmatrix} \end{array}$$

can also be re-written in the following way:

$$\begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \end{array} \begin{array}{c} \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{pmatrix} 1 & 2+1 & 4 \\ 2 & 4+2 & 8 \\ 3 & 6+3 & 12 \end{pmatrix} \end{array} \rightarrow \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \end{array} \begin{array}{c} \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{pmatrix} 1 & 2+1 & 4 \\ 2+1 & 4+2+2+1 & 8+4 \\ 3 & 6+3 & 12 \end{pmatrix} \end{array}$$

Alternatively, add $-\frac{1}{2}C_2$ to C_1, then:

$$\boldsymbol{X} = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \end{array} \begin{array}{c} \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{pmatrix} 1-\frac{1}{2}\times 2 & 2 & 4 \\ 2-\frac{1}{2}\times 4 & 4 & 8 \\ 3-\frac{1}{2}\times 6 & 6 & 12 \end{pmatrix} \end{array} \rightarrow \begin{pmatrix} 0 & 2 & 4 \\ 0 & 4 & 8 \\ 0 & 6 & 12 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 4 \\ 4 & 8 \\ 6 & 12 \end{pmatrix}$$

Multiplication I

A $k \times n$ dimensional matrix $\boldsymbol{Y}$ can be pre-multiplied by an $n \times k$ dimensional matrix $\boldsymbol{Z}$:

$$\underset{n\times n}{\boldsymbol{X}} = \underset{(n\times k)}{\boldsymbol{Y}}\underset{(k\times n)}{\boldsymbol{Z}}$$

because their inner dimensions are identical. When $\boldsymbol{X}$ is an $n \times n$ square matrix, then:

$$\boldsymbol{X} = \boldsymbol{I}_n\boldsymbol{X} = \boldsymbol{X}\boldsymbol{I}_n$$

where the identity matrix is multiplicatively neutral (pre-multiplying or post-multiplying). However, this is not the case in general for two arbitrary $n \times m$ matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$:

$$\boldsymbol{X}'\boldsymbol{Y} \neq \boldsymbol{X}\boldsymbol{Y}'$$

Multiplication II: Kronecker and Hadamard Products

$$\boldsymbol{X} \otimes \boldsymbol{Y} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \otimes \begin{bmatrix} y_{11} & y_{21} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \vdots & \vdots & & \vdots \\ y_{s1} & y_{s2} & \dots & y_{sm} \end{bmatrix} = \begin{bmatrix} x_{11}\boldsymbol{Y} & x_{21}\boldsymbol{Y} & \dots & x_{1k}\boldsymbol{Y} \\ x_{21}\boldsymbol{Y} & x_{22}\boldsymbol{Y} & \dots & x_{2k}\boldsymbol{Y} \\ \vdots & \vdots & & \vdots \\ x_{n1}\boldsymbol{Y} & x_{n2}\boldsymbol{Y} & \dots & x_{nk}\boldsymbol{Y} \end{bmatrix}$$

where $\otimes$ is the Kronecker product operation. This product has limited applications but it can be a very helpful method when working with certain multivariate methods such Vector Autoregressive Regression (VAR).

$$\boldsymbol{X} \circ \boldsymbol{Y} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \circ \begin{bmatrix} y_{11} & y_{21} & \dots & y_{1k} \\ y_{21} & y_{22} & \dots & y_{2k} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nk} \end{bmatrix} = \begin{bmatrix} x_{11}y_{11} & x_{21}y_{21} & \dots & x_{1k}y_{1k} \\ x_{21}y_{21} & x_{22}y_{22} & \dots & x_{2k}y_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1}y_{n1} & x_{n2}y_{n2} & \dots & x_{nk}y_{nk} \end{bmatrix}$$

Hadamard product is an element-wise product of two matrices. This product requires both the pre- and post-multiplying matrices to be of exact size.

Trace

Trace operator is the sum of main diagonal of a matrix (*square*):

$$tr(\boldsymbol{X}) = \sum_{i=j}^{n} x_{i,j}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ then } tr(\boldsymbol{X}) = 1 + 4$$

Properties:

$$\begin{aligned} tr(\boldsymbol{X} + \boldsymbol{Y}) &= tr(\boldsymbol{X}) + tr(\boldsymbol{Y}) \\ tr(\lambda.\boldsymbol{X}) &= \lambda.tr(\boldsymbol{X}) \\ tr(\boldsymbol{X}) &= tr(\boldsymbol{X}') \\ tr(\mathbf{I}_n) &= n \end{aligned}$$

$$tr(\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3) = tr(\boldsymbol{X}_2\boldsymbol{X}_3\boldsymbol{X}_1) = tr(\boldsymbol{X}_3\boldsymbol{X}_1\boldsymbol{X}_2) = tr(\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3)$$

but,

$$tr(\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3) \neq tr(\boldsymbol{X}_1\boldsymbol{X}_3\boldsymbol{X}_2)$$

Determinant

The **determinant** of a (square) matrix is a measure to distinguish linear independence of rows or columns of a matrix. For a 2×2 matrix $\boldsymbol{X}$:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{2,1} & x_{22} \end{bmatrix}$$

$$|\boldsymbol{X}| = det(\boldsymbol{X}) \equiv \begin{vmatrix} x_{11} & x_{12} \\ x_{2,1} & x_{22} \end{vmatrix} = x_{11} \times x_{22} - x_{12} \times x_{21}$$

The determinant of a 3×3 matrix is defined as:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

$$det(\boldsymbol{X}) = (-1)^{1+1} x_{11} \times \begin{vmatrix} x_{2,2} & x_{2,3} \\ x_{3,2} & x_{3,3} \end{vmatrix} + (-1)^{1+2} x_{12} \times \begin{vmatrix} x_{2,1} & x_{2,3} \\ x_{3,1} & x_{3,3} \end{vmatrix} + (-1)^{i+j} x_{13} \times \begin{vmatrix} x_{2,1} & x_{2,2} \\ x_{3,1} & x_{3,2} \end{vmatrix}$$

Inversion

When a square matrix $\boldsymbol{X}$ is full rank then it is referred to an invertible matrix:

$$\boldsymbol{X}^{-1} = \frac{1}{\det(\boldsymbol{X})}\operatorname{adj}(\boldsymbol{X})$$

where $\operatorname{adj}(\boldsymbol{X})$ is the adjunct (or adjugate) matrix and for a 2×2 matrix $\boldsymbol{X}$:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

is defined as:

$$\operatorname{adj}(\boldsymbol{X}) = \begin{bmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{bmatrix}$$

Inversion

for a 3×3 matrix defined as:

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{12} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \end{bmatrix}$$

$$\mathrm{adj}(\boldsymbol{X}) = \begin{bmatrix} + \begin{vmatrix} x_{2,2} & x_{2,3} \\ x_{3,2} & x_{3,3} \end{vmatrix} & - \begin{vmatrix} x_{2,1} & x_{2,3} \\ x_{3,1} & x_{3,3} \end{vmatrix} & + \begin{vmatrix} x_{2,1} & x_{2,2} \\ x_{3,1} & x_{3,2} \end{vmatrix} \\ - \begin{vmatrix} x_{1,2} & x_{1,3} \\ x_{3,2} & x_{3,3} \end{vmatrix} & + \begin{vmatrix} x_{1,1} & x_{1,3} \\ x_{3,1} & x_{3,3} \end{vmatrix} & - \begin{vmatrix} x_{1,1} & x_{1,2} \\ x_{3,1} & x_{3,2} \end{vmatrix} \\ + \begin{vmatrix} x_{1,2} & x_{1,3} \\ x_{2,2} & x_{2,3} \end{vmatrix} & - \begin{vmatrix} x_{1,1} & x_{1,3} \\ x_{2,1} & x_{2,3} \end{vmatrix} & + \begin{vmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{vmatrix} \end{bmatrix}'$$

- each individual element of the adjunct matrix is itself a determinant of four elements
- sign of each element, in general, is driven by $(-1)^{i+j}$ where i is the row and j is the column identifiers of the element inside the matrix.
- for example, the first element sits at the first row and first column with $i = 1$ and $j = 1$ hence $(-1)^2 = 1$.

Consider the following square, non-symmetric matrix $\boldsymbol{X}$:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 3 & 0 \\ 3 & 5 & 3 \end{bmatrix}$$

First, $|\boldsymbol{X}| = 2$ then,

- full rank (non-singular) hence invertible
- no linear combination between rows (columns) is possible

Second,

$$[adj_{1,2}(\boldsymbol{X})]' = (-1)^{2+1} \begin{vmatrix} . & . & . \\ 2 & . & 0 \\ 3 & . & 3 \end{vmatrix} = -\begin{vmatrix} 2 & 0 \\ 5 & 3 \end{vmatrix} = -6$$

repeating for each $adj_{i,j}(\boldsymbol{X})$:

$$adj(\boldsymbol{X}) = \begin{bmatrix} 9 & 19 & -15 \\ -6 & -12 & 10 \\ 1 & 1 & -1 \end{bmatrix}, \quad \boldsymbol{X}^{-1} = \begin{bmatrix} \frac{9}{2} & \frac{19}{2} & -\frac{15}{2} \\ -3 & -6 & 5 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Inversion

Given an invertible (square, full rank ($det \neq 0$)) matrix $\boldsymbol{X}$ then,

1. $(\boldsymbol{X}^{-1})^{-1} = \boldsymbol{X}$
2. $\boldsymbol{X}\boldsymbol{X}^{-1} = \boldsymbol{I}_n$
3. $(\boldsymbol{X}\boldsymbol{Y})^{-1} = \boldsymbol{Y}^{-1}\boldsymbol{X}^{-1}$
4. $(\boldsymbol{X}^{-1})' = (\boldsymbol{X}')^{-1}$ if non-symmetric
5. $(\boldsymbol{X}^{-1})' = (\boldsymbol{X}')^{-1} = \boldsymbol{X}^{-1}$ if symmetric
6. $\boldsymbol{X}^{-1} = \{x_{ii}^{-1}\}$ if $\boldsymbol{X}$ is diagonal

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & 0 & 0 \\ 0 & x_{22} & 0 \\ 0 & 0 & x_{33} \end{bmatrix} \quad \boldsymbol{X}^{-1} = \begin{bmatrix} x_{11}^{-1} & 0 & 0 \\ 0 & x_{22}^{-1} & 0 \\ 0 & 0 & x_{33}^{-1} \end{bmatrix}$$

Quadratic Forms

Quadratic form is a polynomial with terms all of degree two:

$$\begin{aligned} F(x_1, x_2) &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2 \\ &= \sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}x_ix_j \end{aligned}$$

e.g. $(x_1 + x_2)^2 = x_1^2 + x_1x_2 + x_2x_1 + x_2^2$ when all coefficients equal to one.

In general:

$$\begin{aligned} F(\boldsymbol{x}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j \\ &= a_{11}x_1^2 + a_{12}x_1x_2 + \ldots + a_{1n}x_1x_n + \\ & \quad a_{21}x_2x_1 + a_{22}x_2^2 + \ldots + a_{2nx_2x_n} + \\ & \quad \ldots \\ & \quad a_{n1}x_nx_1 + a_{n2}x_nx_2 + \ldots + a_{nn}x_n^2 \end{aligned}$$

Quadratic Form

Definition:

- Let $c_i = \sum_j a_{ij}x_j$ or $\boldsymbol{c} = \boldsymbol{A}\boldsymbol{x}$
- where $\boldsymbol{A} = \{a_{ij}\}$ is square
- then the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is equal to its transpose.

$$F(\boldsymbol{x}) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \;=\; \sum_{i=1}^{n} x_i \left\{\sum_{j=1}^{n} a_{ij}x_j\right\} \;=\; \sum_{i=1}^{n} x_i c_i \;=\; \boldsymbol{x}'\boldsymbol{c} \;=\; \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$$

This can be seen through applying the transpose operation , $(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})' = \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{x}$ which implies that matrix $\boldsymbol{A}$ is symmetric. We can write $2F(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{x}$, then pre- and post-factorizing gives:

$$2F(\boldsymbol{x}) \;=\; \boldsymbol{x}'(\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x} \;=\; \boldsymbol{x}'(2\boldsymbol{B})\boldsymbol{x}$$

where $\boldsymbol{B} = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}')$ therefore $F(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$.

Example: The following two quadratic forms are identical:

$$\boldsymbol{x}'\begin{bmatrix}4 & 2\\ 4 & 6\end{bmatrix}\boldsymbol{x}, \qquad \boldsymbol{x}'\begin{bmatrix}4 & 3\\ 3 & 6\end{bmatrix}\boldsymbol{x}$$

because

$$\frac{1}{2} \times \left(\begin{bmatrix}4 & 2\\ 4 & 6\end{bmatrix} + \begin{bmatrix}4 & 4\\ 2 & 6\end{bmatrix}\right) = \begin{bmatrix}4 & 3\\ 3 & 6\end{bmatrix}$$

Definiteness

Consider

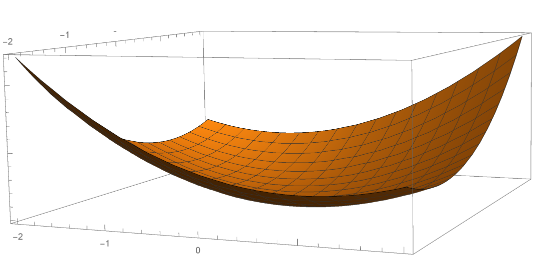
- Quadratic form $F(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$
- Arbitrary real-valued symmetric coefficient matrix $\boldsymbol{A}$

Sign of $F(\boldsymbol{x})$ given any possible choice of value for elements within $\boldsymbol{x}$:

Restrict scope to 2×2 matrices. A quadratic form is,

1. p.d. ($\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$) iff $a_{11} > 0$ and $det(\boldsymbol{A}) > 0$
2. p.s.d. ($\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$) iff $a_{11} \geq 0$ and $det(\boldsymbol{A}) \geq 0$
3. n.d. ($\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} < 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$) iff $a_{11} < 0$ and $det(\boldsymbol{A}) > 0$
4. n.s.d. ($\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$) iff $a_{11} \leq 0$ and $det(\boldsymbol{A}) \geq 0$
5. indefinite iff $det(\boldsymbol{A}) < 0$

Figure on the left shows a pd quadratic form since the surface tends to increase from all edges and hence the entire surface value is above the zero plane.



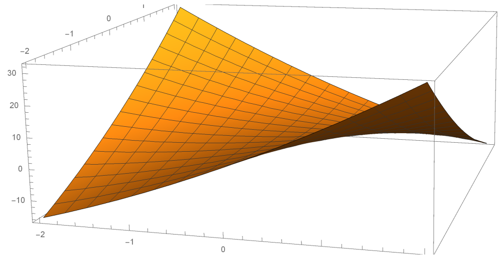


Figure on the right is an indefinite quadratic form since the surface edges tend to increase from two corners but tend to decrease from the other two corners. Hence part of the surface sits above the zero plane but another part fall below.

Expectation

Vector notation in multivariate statistics

- Let an $n \times k$-dimensional matrix $\boldsymbol{X} = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \dots \ \boldsymbol{x}_k]$ collect n-observations of k-variables
- Elements on each column are drawn from different distributions.
- Suppose joint normality:

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}$$

Expectation is a linear operator:

1. $\mathbb{E}[s\boldsymbol{X}] = s.\boldsymbol{\mu}$
2. $\mathbb{E}[\mathbf{1}_n + r\boldsymbol{x}] = \mathbf{1}_n + r.\boldsymbol{\mu}$

$$\mathbb{E}[\boldsymbol{X}] = \begin{bmatrix} \mathbb{E}[\boldsymbol{x}_1] \\ \vdots \\ \mathbb{E}[\boldsymbol{x}_k] \end{bmatrix} = \boldsymbol{\mu}$$

Reduces to $\mu.\mathbf{1}_k$ if all share a common mean.

Variance-Covariance Matrix

Variance is a quadratic operator:

1. Scalar transformation: $\text{var}[\lambda_0 \boldsymbol{X}] = \lambda_0^2 \,\text{var}[\boldsymbol{X}]$
2. Neutral to adding (subtracting) constants: $\text{var}[\lambda_1 \mathbf{1}_{n\times k} + \lambda_2 \boldsymbol{X}] = \lambda_2^2 \,\text{var}[\boldsymbol{X}]$
3. De-meaned transformation: $\text{var}\left\{[\boldsymbol{x}_1 - \mu_1 \quad \boldsymbol{x}_2 - \mu_2 \quad \dots \quad \boldsymbol{x}_k - \mu_k]'\right\} = \text{var}[\boldsymbol{X}]$

$$\text{var}[\boldsymbol{X}] \quad = \quad \mathbb{E}[\boldsymbol{X}'\boldsymbol{X}] - \mathbb{E}[\boldsymbol{X}]'\mathbb{E}[\boldsymbol{X}]$$

Recall the notion of orthogonality $\boldsymbol{x}_i'\boldsymbol{x}_j = 0$:

$$\text{var}[\boldsymbol{X}] \;=\; \begin{bmatrix} \mathbb{E}[\boldsymbol{x_1}'\boldsymbol{x_1}] - \mathbb{E}[\boldsymbol{x_1}]'\mathbb{E}[\boldsymbol{x_1}] & \mathbb{E}[\boldsymbol{x_1}'\boldsymbol{x_2}] - \mathbb{E}[\boldsymbol{x_1}]'\mathbb{E}[\boldsymbol{x_2}] & \dots & \mathbb{E}[\boldsymbol{x_1}'\boldsymbol{x_k}] - \mathbb{E}[\boldsymbol{x_1}]'\mathbb{E}[\boldsymbol{x_k}] \\ \mathbb{E}[\boldsymbol{x_2}'\boldsymbol{x_1}] - \mu_2\mu_1 & \mathbb{E}[\boldsymbol{x_2}'\boldsymbol{x_2}] - \mu_2^2 & \dots & \mathbb{E}[\boldsymbol{x_2}'\boldsymbol{x_k}] - \mu_2\mu_k \\ \vdots & \vdots & & \vdots \\ \mathbb{E}[\boldsymbol{x_k}'\boldsymbol{x_1}] - \mu_k\mu_1 & \mathbb{E}[\boldsymbol{x_k}'\boldsymbol{x_2}] - \mu_k\mu_2 & \dots & \mathbb{E}[\boldsymbol{x_k}'\boldsymbol{x_k}] - \mu_k^2 \end{bmatrix}$$

heteroskedastic variance with zero cross-correlations,

$$\text{var}[\boldsymbol{X}] = \boldsymbol{\Sigma} = \underset{k\times k}{\text{diag}}(\sigma_i^2)$$

homoskedastic if all variances are equal, $\sigma_i^2 = \sigma^2$:

$$\text{var}[\boldsymbol{X}] = \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_k$$

Idempotency and Projection

A square (not necessarily symmetric) matrix is called an **idempotent** matrix when its (general) product with itself results in the original matrix.

Let $\boldsymbol{M}$ be an idempotent matrix, then:

$$\boldsymbol{M}^2 = \boldsymbol{M}$$

When an idempotent matrix $\boldsymbol{M}$ is subtracted from the identity matrix $\boldsymbol{I}_n$, the result is also idempotent:

$$\boldsymbol{I}_n - \boldsymbol{M} = (\boldsymbol{I}_n - \boldsymbol{M})(\boldsymbol{I}_n - \boldsymbol{M})$$

Given an invertible matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ then:

$$\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

is a **projection** matrix which is a special case of an idempotent matrix

- a projection matrix is a square, symmetric and idempotent matrix
- because it inherits properties of idempotent matrices, then $\boldsymbol{P}^2 = \boldsymbol{P} = \boldsymbol{P}'\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}' = \boldsymbol{P}'$
- $tr(\boldsymbol{P}) = rank(\boldsymbol{P})$

Example:

$$\boldsymbol{X} = \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix}$$

Least Squares in Matrix Form

Consider the Ordinary Least Squares (OLS) problem:

$$min_{\beta} \qquad (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

- $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variable
- $\boldsymbol{X}$ is an $n \times k$ matrix of independent variable
- $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients

Minimize matrix product $(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$

1. by choosing the regression coefficients $\boldsymbol{\beta}$
2. solution is $\widehat{\boldsymbol{\beta}}_{\text{OLS}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$
3. search tool is vector differentiation

Least Squares in Matrix Form

Difference between the fitted value $\boldsymbol{X}\widehat{\boldsymbol{\beta}}$ and the dependent variable $\boldsymbol{Y}$:

$$\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$$

Substituting for $\widehat{\boldsymbol{\beta}}$

$$\widehat{\boldsymbol{e}} = (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y}$$

Let $\boldsymbol{M} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

$$\boldsymbol{M}^2 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \ \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

$\boldsymbol{M}$ satisfies properties of an idempotent matrix.

Derivatives

Suppose

- $\boldsymbol{\beta}$ and $\boldsymbol{X}$ are $N \times 1$ vectors
- Y an $N \times N$ non-symmetric matrix
- $f(\boldsymbol{\beta}) = \boldsymbol{X}'\boldsymbol{\beta}$ and $g(\boldsymbol{\beta}) = \boldsymbol{\beta}'\boldsymbol{Y}\boldsymbol{\beta}$

Vector differentials:

1. $\partial f(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \boldsymbol{X}'$
2. $\partial g(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \boldsymbol{\beta}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{Y}' = \boldsymbol{\beta}'(\boldsymbol{Y} + \boldsymbol{Y}')$
3. $\partial^2 g(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = \boldsymbol{Y} + \boldsymbol{Y}'$

$\boldsymbol{Y}$ is a symmetric matrix then:

1. $\partial^2 g(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = 2\boldsymbol{Y}$
2. $\boldsymbol{\beta}'(\boldsymbol{Y} + \boldsymbol{Y}') = 2\boldsymbol{\beta}'\boldsymbol{Y}$
3. $\boldsymbol{Y}' + \boldsymbol{Y} = 2\boldsymbol{Y} = 2\boldsymbol{Y}'$

$$\min_{\beta} \quad (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$
$$\boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$$

- $-\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{\beta} = -\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y}$ are in fact equal
- Differentiating with respect to $\boldsymbol{\beta}$ gives:

$$\frac{\partial(\boldsymbol{Y}' - \boldsymbol{\beta}'\boldsymbol{X}')'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$$

This is the first order condition (FOC) of the OLS problem which is also referred to as system of normal equations (true or false?)

First and Second Order Conditions

Re-arrange:

$$\begin{aligned} \boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{y} \\ \widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \end{aligned}$$

- take second derivative to examine second order conditions (SOC)
- SOC together with results on definiteness determines that solution is minimizer

$$\nabla^2_{\boldsymbol{\beta}}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) = \frac{\partial^2(\boldsymbol{Y}'-\boldsymbol{\beta}'\boldsymbol{X}')'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = \partial\frac{-2\boldsymbol{X}'\boldsymbol{y}+2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}}{\partial\boldsymbol{\beta}'} = 2\boldsymbol{X}'\boldsymbol{I}\boldsymbol{X}$$

- is a quadratic form with identity coefficients
- $det(\boldsymbol{X}'\boldsymbol{X}) \geq 0$ if $det(\boldsymbol{X}'\boldsymbol{X}) \neq 0$
- $\exists\ (\boldsymbol{X}'\boldsymbol{X})^{-1}$ iff $\boldsymbol{X}'\boldsymbol{X}$ is non-singular
- therefore $\widehat{\boldsymbol{\beta}}$ is a minimizer

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\beta} \ (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})$$

Practice I

Let

$$\boldsymbol{X} = \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix}$$

1. Find the rank and determinant of $\boldsymbol{X}$.
2. Is $\boldsymbol{X}$ an idempotent matrix?

Answer:

1. $det(\boldsymbol{X}) = 0$, and $rank(\boldsymbol{X}) = 2$ (a singular matrix or not full rank). The maximum rank possible for this matrix is 3, however, since the determinant is zero, there exists at least one linear combination among rows or columns (in this case $C_3 = C_2 - 2C_1$). After establishing this linear combination, no further linear combination is possible and rank becomes 2, as there are only two linearly independent vectors within $\boldsymbol{X}$, $rank(\boldsymbol{X}) = 2$
2. Yes, because $\boldsymbol{X}\boldsymbol{X} = \boldsymbol{X}$.

Practice II

Let

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \boldsymbol{Y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

1. What is $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.
2. Find $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.
3. Denote $\boldsymbol{M} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and find $\boldsymbol{M}\boldsymbol{Y}$.
4. Find $\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.
5. Find $(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})'(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})$.

Answer:

1. $(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$.
2. $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = (3, -1, -1)'$.
3. $\boldsymbol{M}\boldsymbol{Y} = (3, 2, 1)'$.
4. $\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = (0, 0, 0)'$.
5. $(0, 0, 0)'(0, 0, 0) = 0$ (scalar).

Practice III

Let

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \quad \boldsymbol{Y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

1. What is $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.
2. Find $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.
3. Denote $\boldsymbol{M} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and find $\boldsymbol{M}\boldsymbol{Y}$.
4. Find $\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$
5. Find $(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})'(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})$.
6. Compare scalars found in this exercise and the previous, and comment.

Practice IV

- Let Ω be an $N \times N$ positive definite (PD) matrix that is $b'\Omega b > 0$ (for any non-null $N \times 1$ vector b). Define a $T \times N$ matrix Ψ with $rank(\Psi) = N$. Show that $\Psi'\Psi$ is a PD matrix.
- Let the expectation of a random variable Y conditional on a set of other random variables $X_1, \ldots, X_k$ be the deterministic function $h(X_1, \ldots, X_k)$ of the condition variables. Let Ω be the information set consisting of all deterministic functions of the X_i, $i = 1, \ldots, k$. Show that $\mathbb{E}[Y|\Omega] = h(X_1, \ldots, X_k)$.