Matrix Algebra for Econometrics and Statistics

GARTH TARR

2011

THE UNIVERSITY OF SYDNEY

Matrix fundamentals

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

- A matrix is a rectangular array of numbers.
- Size: (rows)×(columns). E.g. the size of $\mathbf{A}$ is 2×3.
- The size of a matrix is also known as the dimension.
- The element in the ith row and jth column of $\mathbf{A}$ is referred to as a_{ij}.
- The matrix $\mathbf{A}$ can also be written as $\mathbf{A} = (a_{ij})$.

Matrix addition and subtraction

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

Definition (Matrix Addition and Subtraction)

- Dimensions must match:

$$(r \times c) \pm (r \times c) \implies (r \times c)$$

- $\mathbf{A}$ and $\mathbf{B}$ are both 2×3 matrices, so

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} \end{bmatrix}$$

- More generally we write:

$$\mathbf{A} \pm \mathbf{B} = (a_{ij}) \pm (b_{ij}).$$

Matrix multiplication

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}; \quad \mathbf{D} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \\ d_{31} & d_{32} \end{bmatrix}$$

Definition (Matrix Multiplication)

- Inner dimensions need to match:

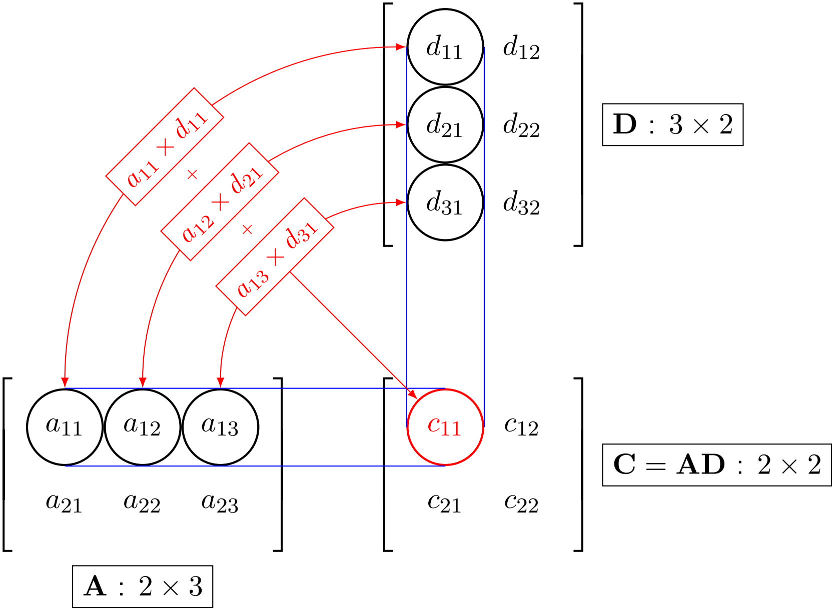
$$(r \times c) \times (c \times p) \implies (r \times p)$$

- $\mathbf{A}$ is a 2×3 and $\mathbf{D}$ is a 3×2 matrix, so the inner dimensions match and we have: $\mathbf{C} = \mathbf{A} \times \mathbf{D} =$

$$\begin{bmatrix} a_{11}d_{11} + a_{12}d_{21} + a_{13}d_{31} & a_{11}d_{12} + a_{12}d_{22} + a_{13}d_{32} \\ a_{21}d_{11} + a_{22}d_{21} + a_{23}d_{31} & a_{21}d_{12} + a_{22}d_{22} + a_{23}d_{32} \end{bmatrix}$$

- Look at the pattern in the terms above.

Matrix multiplication



Determinant

Definition (General Formula)

- Let $\mathbf{C} = (c_{ij})$ be an $n \times n$ square matrix.
- Define a cofactor matrix, C_{ij}, be the determinant of the square matrix of order $(n-1)$ obtained from $\mathbf{C}$ by removing row i and column j multiplied by $(-1)^{i+j}$.
- For fixed i, i.e. focusing on one row: $\det(\mathbf{C}) = \sum_{j=1}^{n} c_{ij} C_{ij}$.
- For fixed j, i.e. focusing on one column: $\det(\mathbf{C}) = \sum_{j=1}^{n} c_{ij} C_{ij}$.
- Note that this is a recursive formula.

More

- The trick is to pick a row (or column) with a lot of zeros (or better yet, use a computer)!

2×2 determinant

Apply the general formula to a 2×2 matrix: $\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$.

- Keep the first row fixed, i.e. set $i = 1$.
- General formula when $i = 1$ and $n = 2$: $\det(\mathbf{C}) = \sum_{j=1}^{2} c_{1j} C_{1j}$
- When $j = 1$, C_{11} is one cofactor matrix of $\mathbf{C}$, i.e. the determinant after removing the first row and first column of $\mathbf{C}$ multiplied by $(-1)^{i+j} = (-1)^2$. So

$$C_{11} = (-1)^2 \det(c_{22}) = c_{22}$$

  as c_{22} is a scalar and the determinant of a scalar is itself.
- $C_{12} = (-1)^3 \det(c_{21}) = -c_{21}$ as c_{21} is a scalar and the determinant of a scalar is itself.
- Put it all together and you get the familiar result:

$$\det(\mathbf{C}) = c_{11}C_{11} + c_{12}C_{12} = c_{11}c_{22} - c_{12}c_{21}$$

3×3 determinant

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

- Keep the first row fixed, i.e. set $i = 1$. General formula when $i = 1$ and $n = 3$:

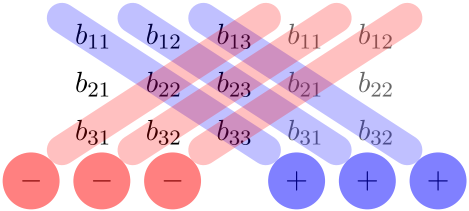
$$\det(\mathbf{B}) = \sum_{j=1}^{3} b_{1j} B_{1j} = b_{11} B_{11} + b_{12} B_{12} + b_{13} B_{13}$$

- For example, B_{12} is the determinant of the matrix you get after removing the first row and second column of $\mathbf{B}$ multiplied by $(-1)^{i+j} = (-1)^{1+2} = -1$: $B_{12} = - \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix}$.
- $\det(\mathbf{B}) = b_{11} \begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix} - b_{12} \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} + b_{13} \begin{vmatrix} b_{21} & b_{22} \\ b_{31} & b_{32} \end{vmatrix}$

Sarrus' scheme for the determinant of a 3×3

- French mathematician: Pierre Frédéric Sarrus (1798-1861)

$$\begin{aligned}
\det(\mathbf{B}) &= \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \\
&= b_{11} \begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix} - b_{12} \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} + b_{13} \begin{vmatrix} b_{21} & b_{22} \\ b_{31} & b_{32} \end{vmatrix} \\
&= \left({\color{blue} b_{11}b_{22}b_{33} + b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32}} \right) \\
&\quad - \left({\color{red} b_{13}b_{22}b_{31} + b_{11}b_{23}b_{32} + b_{12}b_{21}b_{33}} \right)
\end{aligned}$$

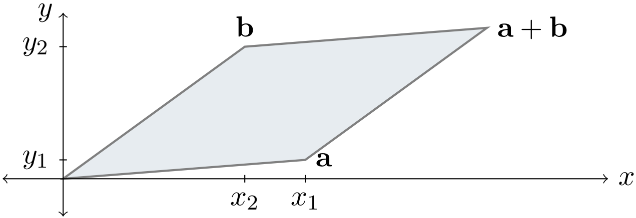


Write the first two columns of the matrix again to the right of the original matrix. Multiply the diagonals together and then add or subtract.

Determinant as an area

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$$

- For a 2×2 matrix, $\det(\mathbf{A})$ is the oriented area[1] of the parallelogram with vertices at $\mathbf{0} = (0,0)$, $\mathbf{a} = (x_1, y_1)$, $\mathbf{a} + \mathbf{b} = (x_1 + x_2, y_1 + y_2)$, and $\mathbf{b} = (x_2, y_2)$.



- In a sense, the determinant "summarises" the information in the matrix.

[1] The oriented area is the same as the usual area, except that it is negative when the vertices are listed in clockwise order.

Identity matrix

Definition (Identity matrix)

- A square matrix, $\mathbf{I}$, with ones on the main diagonal and zeros everywhere else:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

- Sometimes you see $\mathbf{I}_r$ which indicates that it is an $r \times r$ identity matrix.
- If the size of $\mathbf{I}$ is not specified, it is assumed to be "conformable", i.e. as big as necessary.

Identity matrix

- An identity matrix is the matrix analogue of the number 1.
- If you multiply any matrix (or vector) with a conformable identity matrix the result will be the same matrix (or vector).

Example (2×2)

$$
\begin{aligned}
\mathbf{AI} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} a_{11} \times 1 + a_{12} \times 0 & a_{11} \times 0 + a_{12} \times 1 \\ a_{21} \times 1 + a_{22} \times 0 & a_{21} \times 0 + a_{22} \times 1 \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \mathbf{A}.
\end{aligned}
$$

Inverse

Definition (Inverse)

- Requires a square matrix i.e. dimensions: $r \times r$
- For a 2×2 matrix, $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$,

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

- More generally, a square matrix A is invertible or nonsingular if there exists another matrix $\mathbf{B}$ such that

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}.$$

- If this occurs then $\mathbf{B}$ is uniquely determined by $\mathbf{A}$ and is denoted $\mathbf{A}^{-1}$, i.e. $\mathbf{AA}^{-1} = \mathbf{I}$.

Vectors

Vectors are matrices with only one row or column. For example, the column vector:

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

Definition (Transpose Operator)

Turns columns into rows (and vice versa):

$$\mathbf{x}' = \mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

Example (Sum of Squares)

$$\mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2$$

Transpose

Say we have some $m \times n$ matrix:

$$\mathbf{A} = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Definition (Transpose Operator)

- Flips the rows and columns of a matrix:

$$\mathbf{A}' = (a_{ji})$$

- The subscripts gets swapped.
- $\mathbf{A}'$ is a $n \times m$ matrix: the columns in $\mathbf{A}$ are the rows in $\mathbf{A}'$.

Symmetry

Definition (Square Matrix)

A matrix, $\mathbf{P}$ is square if it has the same number of rows as columns. I.e.

$$\dim(\mathbf{P}) = n \times n$$

for some $n \geq 1$.

Definition (Symmetric Matrix)

A square matrix, $\mathbf{P}$ is symmetric if it is equal to its transpose:

$$\mathbf{P} = \mathbf{P}'$$

Idempotent

Definition (Idempotent)

A square matrix, $\mathbf{P}$ is idempotent if when multiplied by itself, yields itself. I.e.

$$\mathbf{PP} = \mathbf{P}.$$

1. When an idempotent matrix is subtracted from the identity matrix, the result is also idempotent, i.e. $\mathbf{M} = \mathbf{I} - \mathbf{P}$ is idempotent.
2. The trace of an idempotent matrix is equal to the rank.
3. $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is an idempotent matrix.

Order of operations

- Matrix multiplication is non-commutative, i.e. the order of multiplication is important: $\mathbf{AB} \neq \mathbf{BA}$. ▸ Commutativity
- Matrix multiplication is associative, i.e. as long as the order stays the same, $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$. ▸ Associativity
- $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
- $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$

Example

Let $\mathbf{A}$ be a $k \times k$ matrix and $\mathbf{x}$ and $\mathbf{c}$ be $k \times 1$ vectors:

$$\begin{aligned} \mathbf{Ax} &= \mathbf{c} \\ \mathbf{A}^{-1}\mathbf{Ax} &= \mathbf{A}^{-1}\mathbf{c} \qquad \text{(PRE-multiply both sides by } \mathbf{A}^{-1}\text{)} \\ \mathbf{Ix} &= \mathbf{A}^{-1}\mathbf{c} \\ \mathbf{x} &= \mathbf{A}^{-1}\mathbf{c} \end{aligned}$$

Note: $\mathbf{A}^{-1}\mathbf{c} \neq \mathbf{c}\mathbf{A}^{-1}$

Matrix Differentiation

If $\boldsymbol{\beta}$ and $\mathbf{a}$ are both $k \times 1$ vectors then, $\dfrac{\partial \boldsymbol{\beta}' \mathbf{a}}{\partial \boldsymbol{\beta}} = \mathbf{a}$.

Proof.

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} \left(\boldsymbol{\beta}' \mathbf{a} \right) &= \frac{\partial}{\partial \boldsymbol{\beta}} \left(\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_k a_k \right) \\
&= \begin{bmatrix} \frac{\partial}{\partial \beta_1} \left(\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_k a_k \right) \\ \frac{\partial}{\partial \beta_2} \left(\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_k a_k \right) \\ \vdots \\ \frac{\partial}{\partial \beta_k} \left(\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_k a_k \right) \end{bmatrix} \\
&= \mathbf{a}
\end{aligned}
$$

$\square$

▸ Matrix Calculus

Matrix Differentiation

Let $\boldsymbol{\beta}$ be a $k \times 1$ vector and $\mathbf{A}$ be a $k \times k$ symmetric matrix then

$$\frac{\partial \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = 2\mathbf{A}\boldsymbol{\beta}.$$

Proof.

By means of proof, say $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$, then

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} \left(\boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} \right) &= \frac{\partial}{\partial \boldsymbol{\beta}} \left(\beta_1^2 a_{11} + 2a_{12}\beta_1\beta_2 + \beta_2^2 a_{22} \right) \\
&= \begin{bmatrix} \frac{\partial}{\partial \beta_1} \left(\beta_1^2 a_{11} + 2a_{12}\beta_1\beta_2 + \beta_2^2 a_{22} \right) \\ \frac{\partial}{\partial \beta_2} \left(\beta_1^2 a_{11} + 2a_{12}\beta_1\beta_2 + \beta_2^2 a_{22} \right) \end{bmatrix} \\
&= \begin{bmatrix} 2\beta_1 a_{11} + 2a_{12}\beta_2 \\ 2\beta_1 a_{12} + 2a_{22}\beta_2 \end{bmatrix} \\
&= 2\mathbf{A}\boldsymbol{\beta}
\end{aligned}$$

□

Matrix Differentiation

Let $\boldsymbol{\beta}$ be a $k \times 1$ vector and $\mathbf{A}$ be a $n \times k$ matrix then $\dfrac{\partial \mathbf{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}'} = \mathbf{A}$.

Proof.

By means of proof, say $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, then

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}'}\left(\mathbf{A}\boldsymbol{\beta}\right) &= \frac{\partial}{\partial \boldsymbol{\beta}'}\begin{bmatrix} a_{11}\beta_1 + a_{12}\beta_2 \\ a_{21}\beta_1 + a_{22}\beta_2 \end{bmatrix} \\
&= \begin{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \beta_1} & \frac{\partial}{\partial \beta_2} \end{bmatrix} (a_{11}\beta_1 + a_{12}\beta_2) \\ \begin{bmatrix} \frac{\partial}{\partial \beta_1} & \frac{\partial}{\partial \beta_2} \end{bmatrix} (a_{21}\beta_1 + a_{22}\beta_2) \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial}{\partial \beta_1}(a_{11}\beta_1 + a_{12}\beta_2) & \frac{\partial}{\partial \beta_2}(a_{11}\beta_1 + a_{12}\beta_2) \\ \frac{\partial}{\partial \beta_1}(a_{21}\beta_1 + a_{22}\beta_2) & \frac{\partial}{\partial \beta_2}(a_{21}\beta_1 + a_{22}\beta_2) \end{bmatrix} \\
&= \mathbf{A}.
\end{aligned}
$$

$\square$

Rank

- The rank of a matrix $\mathbf{A}$ is the maximal number of linearly independent rows or columns of $\mathbf{A}$.
- A family of vectors is linearly independent if none of them can be written as a linear combination of finitely many other vectors in the collection.

Example (Dummy variable trap)

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{bmatrix} = \begin{bmatrix} \overbrace{1 \quad 1 \quad 0}^{\text{independent}} & 0 \\ 1 \quad 1 \quad 0 & 0 \\ 1 \quad 0 \quad 1 & 0 \\ \underbrace{1 \quad 0 \quad 0 \quad 1}_{\text{dependent}} \end{bmatrix}$$

$\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ are independent but $\mathbf{v}_4 = \mathbf{v}_1 - \mathbf{v}_2 - \mathbf{v}_3$.

Rank

- The maximum rank of an $m \times n$ matrix is $\min(m, n)$.
- A full rank matrix is one that has the largest possible rank, i.e. the rank is equal to either the number of rows or columns (whichever is smaller).
- In the case of an $n \times n$ square matrix $\mathbf{A}$, then $\mathbf{A}$ is invertible if and only if $\mathbf{A}$ has rank n (that is, $\mathbf{A}$ has full rank).
- For some $n \times k$ matrix, $\mathbf{X}$, $\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}'\mathbf{X})$
- This is why the dummy variable trap exists, you need to drop one of the dummy categories otherwise $\mathbf{X}$ is not of full rank and therefore you cannot find the inverse of $\mathbf{X}'\mathbf{X}$.

Trace

Definition

The trace of an $n \times n$ matrix $\mathbf{A}$ is the sum of the elements on the main diagonal: $\mathrm{tr}(\mathbf{A}) = a_{11} + a_{22} + \ldots + a_{nn} = \sum_{i=1}^{n} a_{ii}$.

Properties

- $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$
- $\mathrm{tr}(c\mathbf{A}) = c\mathrm{tr}(\mathbf{A})$
- If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times m$ matrix then
$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$$
- More generally, for conformable matrices:
$$\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB}) = \mathrm{tr}(\mathbf{BCA})$$
BUT: $\mathrm{tr}(\mathbf{ABC}) \neq \mathrm{tr}(\mathbf{ACB})$. You can only move from the front to the back (or back to the front)!

Eigenvalues

- An eigenvalue λ and an eigenvector $\mathbf{x} \neq \mathbf{0}$ of a square matrix $\mathbf{A}$ is defined as

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}.$$

- Since the eigenvector $\mathbf{x}$ is different from the zero vector (i.e. $\mathbf{x} \neq \mathbf{0}$) the following is valid:

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

- We know $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ because:
  - if $(\mathbf{A} - \lambda\mathbf{I})^{-1}$ existed, we could just pre multiply both sides by $(\mathbf{A} - \lambda\mathbf{I})^{-1}$ and get the solution $\mathbf{x} = \mathbf{0}$.
  - but we have assumed $\mathbf{x} \neq \mathbf{0}$ so we require that $(\mathbf{A} - \lambda\mathbf{I})$ is NOT invertible which implies[2] that $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.
- To find the eigenvalues, we can solve $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

[2] A matrix is invertible if and only if the determinant is non-zero

Eigenvalues

Example (Finding eigenvalues)

Say $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. We can find the eigenvaules of $\mathbf{A}$ by solving

$$
\begin{aligned}
\det(\mathbf{A} - \lambda \mathbf{I}) &= 0 \\
\det\left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) &= 0 \\
\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} &= 0 \\
(2-\lambda)(2-\lambda) - 1 \times 1 &= 0 \\
\lambda^2 - 4\lambda + 3 &= 0 \\
(\lambda - 1)(\lambda - 3) &= 0
\end{aligned}
$$

The eigenvalues are the roots of this quadratic: $\lambda = 1$ and $\lambda = 3$.

Why do we care about eigenvalues?

- An $n \times n$ matrix $\mathbf{A}$ is positive definite if all eigenvalues of $\mathbf{A}$, $\lambda_1, \lambda_2, \ldots, \lambda_n$ are positive. Definiteness
- A matrix is negative-definite, negative-semidefinite, or positive-semidefinite if and only if all of its eigenvalues are negative, non-positive, or non-negative, respectively.
- The eigenvectors corresponding to different eigenvalues are linearly independent. So if a $n \times n$ matrix has n nonzero eigenvalues, it is of full rank. Rank
- The trace of a matrix is the sum of the eigenvectors: $\mathrm{tr}(\mathbf{A}) = \lambda_1 + \lambda_2 + \ldots + \lambda_n$. Trace
- The determinant of a matrix is the product of the eigenvectors: $\det(\mathbf{A}) = \lambda_1 \lambda_2 \cdots \lambda_n$. Determinant
- The eigenvectors and eigenvalues of the covariance matrix of a data set data are also used in principal component analysis (similar to factor analysis). Factor Analysis

Useful rules

- $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$
- $\det(\mathbf{A}) = \det(\mathbf{A}')$
- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
- $\det(\mathbf{A}^{-1}) = \dfrac{1}{\det(\mathbf{A})}$
- $\mathbf{AI} = \mathbf{A}$ and $\mathbf{xI} = \mathbf{x}$
- If $\boldsymbol{\beta}$ and $\mathbf{a}$ are both $k \times 1$ vectors, $\dfrac{\partial \boldsymbol{\beta}'\mathbf{a}}{\partial \boldsymbol{\beta}} = \mathbf{a}$
- If $\mathbf{A}$ is a $n \times k$ matrix, $\dfrac{\partial \mathbf{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}'} = \mathbf{A}$
- If $\mathbf{A}$ is a $k \times k$ symmetric matrix, $\dfrac{\partial \boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = 2\mathbf{A}\boldsymbol{\beta}$
- If $\mathbf{A}$ is a $k \times k$ (not necessarily symmetric) matrix, $\dfrac{\partial \boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}$

Quadratic forms

- A quadratic form on $\mathbb{R}^n$ is a real-valued function of the form

$$Q(x_1, \ldots, x_n) = \sum_{i \le j} a_{ij} x_i x_j.$$

- E.g. in $\mathbb{R}^2$ we have $Q(x_1, x_2) = a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2$.
- Quadratic forms can be represented by a *symmetric* matrix $\mathbf{A}$ such that:

$$Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$$

- E.g. if $\mathbf{x} = (x_1, x_2)'$ then

$$\begin{aligned} Q(\mathbf{x}) &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} \\ \frac{1}{2}a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= a_{11}x_1^2 + \frac{1}{2}(a_{12} + a_{21})x_1x_2 + a_{22}x_2^2 \end{aligned}$$

  but $\mathbf{A}$ is symmetric, i.e. $a_{12} = a_{21}$, so we can write,

$$= a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2.$$

Quadratic forms

If $\mathbf{x} \in \mathbb{R}^3$, i.e. $\mathbf{x} = (x_1, x_2, x_3)'$ then the general three dimensional quadratic form is:

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}'\mathbf{A}\mathbf{x} \\ &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} & \frac{1}{2}a_{13} \\ \frac{1}{2}a_{12} & a_{22} & \frac{1}{2}a_{23} \\ \frac{1}{2}a_{13} & \frac{1}{2}a_{23} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3. \end{aligned}$$

Quadratic Forms and Sum of Squares

Recall sums of squares can be written as $\mathbf{x}'\mathbf{x}$ and quadratic forms are $\mathbf{x}'\mathbf{A}\mathbf{x}$. Quadratic forms are like generalised and weighted sum of squares. Note that if $\mathbf{A} = \mathbf{I}$ then we recover the sums of squares exactly.

Definiteness of quadratic forms

- A quadratic form always takes on the value zero at the point $\mathbf{x} = \mathbf{0}$. This is not an interesting result!
- For example, if $\mathbf{x} \in \mathbb{R}$, i.e. $\mathbf{x} = x_1$ then the general quadratic form is ax_1^2 which equals zero when $x_1 = 0$.
- Its distinguishing characteristic is the set of values it takes when $\mathbf{x} \neq \mathbf{0}$.
- We want to know if $\mathbf{x} = \mathbf{0}$ is a max, min or neither.
- Example: when $\mathbf{x} \in \mathbb{R}$, i.e. the quadratic form is ax_1^2,

  $a > 0$ means $ax^2 \geq 0$ and equals 0 only when $x = 0$. Such a form is called positive definite; $x = 0$ is a global minimiser.

  $a < 0$ means $ax^2 \leq 0$ and equals 0 only when $x = 0$. Such a form is called negative definite; $x = 0$ is a global maximiser.

Positive definite

If $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ then $Q_1(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 + x_2^2$.

- Q_1 is greater than zero at $\mathbf{x} \neq \mathbf{0}$ i.e. $(x_1, x_2) \neq (0, 0)$.
- The point $\mathbf{x} = \mathbf{0}$ is a global minimum.
- Q_1 is called positive definite.

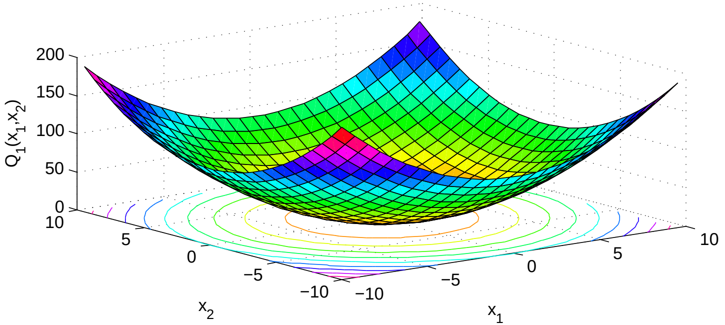


Figure 1: $Q_1(x_1, x_2) = x_1^2 + x_2^2$

Code

Negative definite

If $\mathbf{A} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ then $Q_2(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = -x_1^2 - x_2^2$.

- Q_2 is less than zero at $\mathbf{x} \neq \mathbf{0}$ i.e. $(x_1, x_2) \neq (0, 0)$.
- The point $\mathbf{x} = \mathbf{0}$ is a global maximum.
- Q_2 is called negative definite.

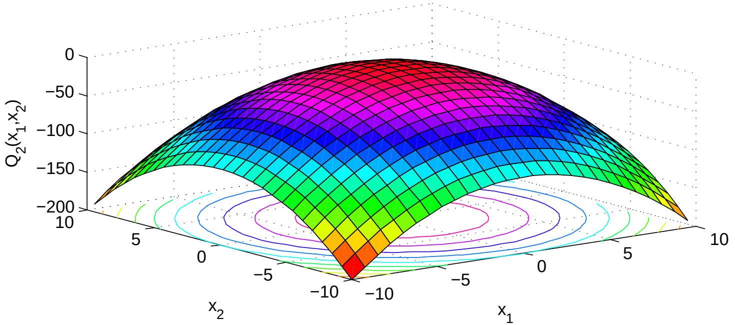


Figure 2: $Q_2(x_1, x_2) = -x_1^2 - x_2^2$

Code

Indefinite

If $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ then $Q_3(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 - x_2^2$.

- Q_3 can be take both positive and negative values.
- E.g. $Q_3(1,0) = +1$ and $Q_3(0,1) = -1$.
- Q_3 is called indefinite.

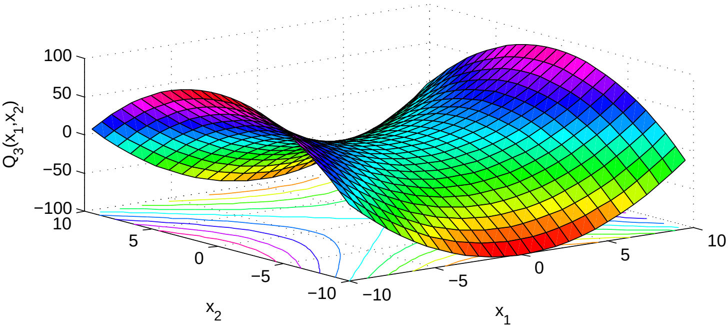


Figure 3: $Q_3(x_1, x_2) = x_1^2 - x_2^2$

Code

Positive semidefinite

If $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ then $Q_4(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 + 2x_1x_2 + x_2^2$.

- Q_4 is always ≥ 0 but does equal zero at some $\mathbf{x} \neq \mathbf{0}$.
- E.g. $Q_4(10, -10) = 0$.
- Q_4 is called positive semidefinite.

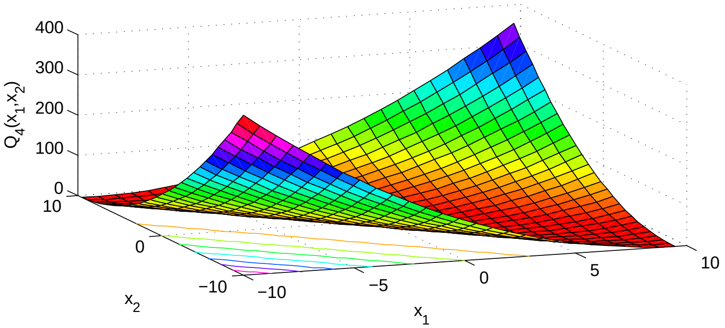


Figure 4: $Q_4(x_1, x_2) = x_1^2 + 2x_1x_2 + x_2^2$

Code

Negative semidefinite

If $\mathbf{A} = \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix}$ then $Q_5(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = -(x_1 + x_2)^2$.

- Q_4 is always ≤ 0 but does equal zero at some $\mathbf{x} \neq \mathbf{0}$
- E.g. $Q_5(10, -10) = 0$
- Q_5 is called negative semidefinite.

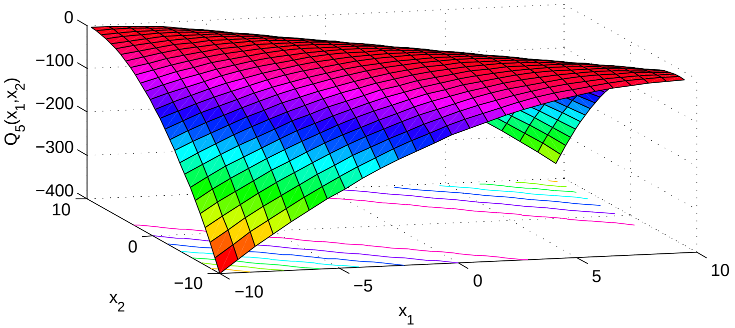


Figure 5: $Q_5(x_1, x_2) = -(x_1 + x_2)^2$

Code

Definite symmetric matrices

A symmetric matrix, $\mathbf{A}$, is called positive definite, positive semidefinite, negative definite, etc. according to the definiteness of the corresponding quadratic form $Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$.

Definition

Let $\mathbf{A}$ be a $n \times n$ symmetric matrix, then $\mathbf{A}$ is

1. positive definite if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$
2. positive semidefinite if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$
3. negative definite if $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$
4. negative semidefinite if $\mathbf{x}'\mathbf{A}\mathbf{x} \leq 0$ for all $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$
5. indefinite if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for some $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$ and < 0 for some other $\mathbf{x}$ in $\mathbb{R}^n$

- We can check the definiteness of a matrix by show that one of these definitions holds as in the example ◂ Example
- You can find the eigenvalues to check definiteness ◂ Eigenvalues

How else to check for definiteness?

You can check the sign of the sequence of determinants of the leading principal minors:

Positive Definite

An $n \times n$ matrix $\mathbf{M}$ is positive definite if all the following matrices have a positive determinant:

- the top left 1×1 corner of $\mathbf{M}$ (1st order principal minor)
- the top left 2×2 corner of $\mathbf{M}$ (2nd order principal minor)

$\vdots$

- $\mathbf{M}$ itself.

In other words, all of the leading principal minors are positive.

Negative Definite

A matrix is negative definite if all kth order leading principal minors are negative when k is odd and positive when k is even.

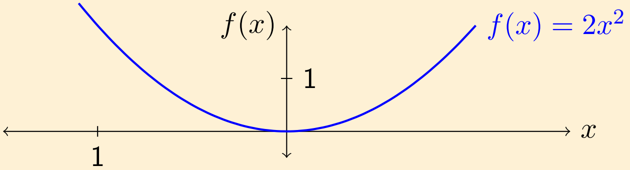
Why do we care about definiteness?

Useful for establishing if a (multivariate) function has a maximum, minimum or neither at a critical point.

- If we have a function, $f(x)$, we can show that a minimum exists at a critical point, i.e. when $f'(x) = 0$, if $f''(x) > 0$.

Example ($f(x) = 2x^2$)

- $f'(x) = 4x$
- $f'(x) = 0 \implies x = 0$
- $f''(x) = 4 > 0 \implies$ minimum at $x = 0$.



Why do we care about definiteness?

- In the special case of a univariate function $f''(x)$ is a 1×1 Hessian matrix and showing that $f''(x) > 0$ is equivalent to showing that the Hessian is positive definite.
- If we have a bivariate function $f(x, y)$ we find critical points when the first order partial derivatives are equal to zero:
  1. Find the first order derivatives and set them equal to zero
  2. Solve simultaneously to find critical points
- We can check if max or min or neither using the Hessian matrix, $\mathbf{H}$, the matrix of second order partial derivatives:

$$\mathbf{H} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

  1. (If necessary) evaluate the Hessian at a critical point
  2. Check if $\mathbf{H}$ is positive or negative definite: ◂ Check definiteness
     - $|\mathbf{H}| > 0$ and $f_{xx} > 0 \implies$ positive definite $\implies$ minimum
     - $|\mathbf{H}| > 0$ and $f_{xx} < 0 \implies$ negative definite $\implies$ maximum
  3. Repeat for all critical points

Why do we care about definiteness?

- If we find the second order conditions and show that it is a positive definite matrix then we have shown that we have a minimum.
- Positive definite matrices are non-singular, i.e. we can invert them. So if we can show $\mathbf{X}'\mathbf{X}$ is positive definiteness, we can find $[\mathbf{X}'\mathbf{X}]^{-1}$.
- Application: showing that the Ordinary Least Squares (OLS) minimises the sum of squared residuals.

◂ Application

Matrices as systems of equations

- A system of equations:

$$
\begin{aligned}
y_1 &= x_{11}b_1 + x_{12}b_2 + \ldots + x_{1k}b_k \\
y_2 &= x_{21}b_1 + x_{22}b_2 + \ldots + x_{2k}b_k \\
&\vdots \\
y_n &= x_{n1}b_1 + x_{n2}b_2 + \ldots + x_{nk}b_k
\end{aligned}
$$

- The matrix form:

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}.
$$

Matrices as systems of equations

- More succinctly: $\mathbf{y} = \mathbf{X}\mathbf{b}$ where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}; \quad \mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix}$$

for $i = 1, 2, \ldots, n$ and

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_n \end{bmatrix}.$$

- $\mathbf{x}_i$ is the "covariate vector" for the ith observation.

DM §1.4

Matrices as systems of equations

- We can write $\mathbf{y} = \mathbf{X}\mathbf{b}$ as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix} \mathbf{b}.$$

- Returning to the original system, we can write each individual equation using vectors:

$$\begin{aligned} y_1 &= \mathbf{x}_1'\mathbf{b} \\ y_2 &= \mathbf{x}_2'\mathbf{b} \\ &\vdots \\ y_n &= \mathbf{x}_n'\mathbf{b} \end{aligned}$$

Mixing matrices, vectors and summation notation

Often we want to find $\mathbf{X}'\mathbf{u}$ or $\mathbf{X}'\mathbf{X}$. A convenient way to write this is as a sum of vectors. Say we have a 3×2 matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \mathbf{x}_3' \end{bmatrix}; \quad \mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix}; \text{ and } \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

We can write,

$$\begin{aligned} \mathbf{X}'\mathbf{u} &= \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \\ &= \begin{bmatrix} x_{11}u_1 + x_{21}u_2 + x_{31}u_3 \\ x_{12}u_1 + x_{22}u_2 + x_{32}u_3 \end{bmatrix} \\ &= \mathbf{x}_1 u_1 + \mathbf{x}_2 u_2 + \mathbf{x}_3 u_3 \\ &= \sum_{i=1}^{3} \mathbf{x}_i u_i \end{aligned}$$

Mixing matrices, vectors and summation notation

In a similar fashion, you can also show that $\mathbf{X}'\mathbf{X} = \sum_{i=1}^{3} \mathbf{x}_i \mathbf{x}_i'$.

$$
\begin{aligned}
\mathbf{X}'\mathbf{X} &= \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \mathbf{x}_3' \end{bmatrix} \\
&= \mathbf{x}_1\mathbf{x}_1' + \mathbf{x}_2\mathbf{x}_2' + \mathbf{x}_3\mathbf{x}_3' \\
&= \sum_{i=1}^{3} \mathbf{x}_i \mathbf{x}_i'
\end{aligned}
$$

Application: variance-covariance matrix

- For the univariate case, $\mathrm{var}(Y) = \mathbb{E}\left([Y-\mu]^2\right)$.
- In the multivariate case $\mathbf{Y}$ is a vector of n random variables.
- Without loss of generality, assume $\mathbf{Y}$ has mean zero, i.e. $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu} = \mathbf{0}$. Then,

$$
\begin{aligned}
\mathrm{cov}(\mathbf{Y},\mathbf{Y}) = \mathrm{var}(\mathbf{Y}) &= \mathbb{E}\left([\mathbf{Y}-\boldsymbol{\mu}][\mathbf{Y}-\boldsymbol{\mu}]'\right) \\
&= \mathbb{E}\left(\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}\begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix}\right) \\
&= \mathbb{E}\begin{bmatrix} Y_1^2 & Y_1Y_2 & \cdots & Y_1Y_n \\ Y_2Y_1 & Y_2^2 & \cdots & Y_2Y_n \\ \vdots & \vdots & & \vdots \\ Y_nY_1 & Y_nY_2 & \cdots & Y_n^2 \end{bmatrix}
\end{aligned}
$$

Application: variance-covariance matrix

- Hence, we have a variance-covariance matrix:

$$\text{var}(\mathbf{Y}) = \begin{bmatrix} \text{var}(Y_1) & \text{cov}(Y_1, Y_2) & \cdots & \text{cov}(Y_1, Y_n) \\ \text{cov}(Y_2, Y_1) & \text{var}(Y_2) & \cdots & \text{cov}(Y_2, Y_n) \\ \vdots & \vdots & & \vdots \\ \text{cov}(Y_n, Y_1) & \text{cov}(Y_n, Y_2) & \cdots & \text{var}(Y_n) \end{bmatrix}.$$

- What if we weight the random variables with a vector of constants, $\mathbf{a}$?

$$\begin{aligned} \text{var}(\mathbf{a}'\mathbf{Y}) &= \mathbb{E}\left([\mathbf{a}'\mathbf{Y} - \mathbf{a}'\boldsymbol{\mu}][\mathbf{a}'\mathbf{Y} - \mathbf{a}'\boldsymbol{\mu}]'\right) \\ &= \mathbb{E}\left(\mathbf{a}'[\mathbf{Y} - \boldsymbol{\mu}](\mathbf{a}'[\mathbf{Y} - \boldsymbol{\mu}])'\right) \\ &= \mathbb{E}\left(\mathbf{a}'[\mathbf{Y} - \boldsymbol{\mu}][\mathbf{Y} - \boldsymbol{\mu}]'\mathbf{a}\right) \\ &= \mathbf{a}'\mathbb{E}\left([\mathbf{Y} - \boldsymbol{\mu}][\mathbf{Y} - \boldsymbol{\mu}]'\right)\mathbf{a} \\ &= \mathbf{a}'\text{var}(\mathbf{Y})\mathbf{a} \end{aligned}$$

Application: variance of sums of random variables

Let $\mathbf{Y} = (Y_1, Y_2)'$ be a vector of random variables and $\mathbf{a} = (a_1, a_2)'$ be some constants,

$$\mathbf{a}'\mathbf{Y} = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = a_1 Y_1 + a_2 Y_2$$

Now, $\mathrm{var}(a_1 Y_1 + a_2 Y_2) = \mathrm{var}(\mathbf{a}'\mathbf{Y}) = \mathbf{a}'\mathrm{var}(\mathbf{Y})\mathbf{a}$ where

$$\mathrm{var}(\mathbf{Y}) = \begin{bmatrix} \mathrm{var}(Y_1) & \mathrm{cov}(Y_1, Y_2) \\ \mathrm{cov}(Y_1, Y_2) & \mathrm{var}(Y_2) \end{bmatrix},$$

is the (symmetric) variance-covariance matrix.

$$\begin{aligned} \mathrm{var}(\mathbf{a}'\mathbf{Y}) &= \mathbf{a}'\mathrm{var}(\mathbf{Y})\mathbf{a} \\ &= \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} \mathrm{var}(Y_1) & \mathrm{cov}(Y_1, Y_2) \\ \mathrm{cov}(Y_1, Y_2) & \mathrm{var}(Y_2) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \\ &= a_1^2 \mathrm{var}(Y_1) + a_2^2 \mathrm{var}(Y_2) + 2 a_1 a_2 \mathrm{cov}(Y_1, Y_2) \end{aligned}$$

Application: Given a linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ derive the OLS estimator $\hat{\boldsymbol{\beta}}$. Show that $\hat{\boldsymbol{\beta}}$ achieves a minimum.

- The OLS estimator $\boldsymbol{\beta}$ minimises the sum of squared residuals, $\mathbf{u}'\mathbf{u} = \sum_{i=1}^{n} u_i^2$ where $\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ or $u_i = y_i - \mathbf{x}_i'\boldsymbol{\beta}$.

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$
$$= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

- Take the first derivative of $S(\boldsymbol{\beta})$ and set it equal to zero:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0 \implies \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}.$$

- Assuming $\mathbf{X}$ (and therefore $\mathbf{X}'\mathbf{X}$) is of full rank (so is $\mathbf{X}'\mathbf{X}$ invertible) we get,

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

Application: Given a linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ derive the OLS estimator $\hat{\boldsymbol{\beta}}$. Show that $\hat{\boldsymbol{\beta}}$ achieves a minimum.

- For a minimum we need to use the second order conditions:
$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = 2\mathbf{X}'\mathbf{X}.$$
- The solution will be a minimum if $\mathbf{X}'\mathbf{X}$ is a positive definite matrix. Let $q = \mathbf{c}'\mathbf{X}'\mathbf{X}\mathbf{c}$ for some $\mathbf{c} \neq \mathbf{0}$. Then
$$q = \mathbf{v}'\mathbf{v} = \sum_{i=1}^{n} v_i^2, \quad \text{where } \mathbf{v} = \mathbf{X}\mathbf{c}.$$
- Unless $\mathbf{v} = \mathbf{0}$, q is positive. But, if $\mathbf{v} = \mathbf{0}$ then $\mathbf{v}$ or $\mathbf{c}$ would be a linear combination of the columns of $\mathbf{X}$ that equals $\mathbf{0}$ which contradicts the assumption that $\mathbf{X}$ has full rank.
- Since $\mathbf{c}$ is arbitrary, q is positive for every $\mathbf{c} \neq \mathbf{0}$ which establishes that $\mathbf{X}'\mathbf{X}$ is positive definite.

◂ Definiteness

- Therefore, if $\mathbf{X}$ has full rank, then the least squares solution $\hat{\boldsymbol{\beta}}$ is unique and minimises the sum of squared residuals.

Matrix Operations

Operation	R	Matlab
$\mathbf{A} = \begin{bmatrix} 5 & 7 \\ 10 & 2 \end{bmatrix}$	`A=matrix(c(5,7,10,2),` `ncol=2,byrow=T)`	`A = [5,7;10,2]`
$\det(\mathbf{A})$	`det(A)`	`det(A)`
$\mathbf{A}^{-1}$	`solve(A)`	`inv(A)`
$\mathbf{A}+\mathbf{B}$	`A + B`	`A + B`
$\mathbf{AB}$	`A %*% B`	`A * B`
$\mathbf{A}'$	`t(A)`	`A'`

Matrix Operations

Operation	R	Matlab
eigenvalues & eigenvectors	`eigen(A)`	`[V,E] = eig(A)`
covariance matrix of $\mathbf{X}$	`var(X)` or `cov(X)`	`cov(X)`
estimate of $\mathrm{rank}(\mathbf{A})$	`qr(A)$rank`	`rank(A)`
$r \times r$ identity matrix, $\mathbf{I}_r$	`diag(1,r)`	`eye(r)`

Matlab Code

Figure 1

◂ Figure 1

```
[x,y] = meshgrid(-10:0.75:10,-10:0.75:10);
surfc(x,y,x.^2 + y.^2)
ylabel('x_2')
xlabel('x_1')
zlabel('Q_1(x_1,x_2)')
```

Figure 2

◂ Figure 2

```
[x,y] = meshgrid(-10:0.75:10,-10:0.75:10);
surfc(x,y,-x.^2 - y.^2)
ylabel('x_2')
xlabel('x_1')
zlabel('Q_2(x_1,x_2)')
```

Matlab Code

Figure 3

Figure 3

```
[x,y] = meshgrid(-10:0.75:10,-10:0.75:10);
surfc(x,y,x.^2 - y.^2)
ylabel('x_2')
xlabel('x_1')
zlabel('Q_3(x_1,x_2)')
```

Figure 4

Figure 4

```
[x,y] = meshgrid(-10:0.75:10,-10:0.75:10);
surfc(x,y,x.^2 + 2.*x.*y + y.^2)
ylabel('x_2')
xlabel('x_1')
zlabel('Q_4(x_1,x_2)')
```

Matlab Code

Figure 5

Figure 5

```
[x,y] = meshgrid(-10:0.75:10,-10:0.75:10);
surfc(x,y,-(x+y).^2)
ylabel('x_2')
xlabel('x_1')
zlabel('Q_5(x_1,x_2)')
```