EC9A0: Pre-sessional Advanced Mathematics Course

Slides 1: Matrix Algebra

Peter J. Hammond & Pablo F. Beker
Department of Economics
University of Warwick

Autumn 2013

Slides Outline

System of Equations

- We are interested in solving system of equations like:

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

  where each a_{ij} is called the *coefficient* of the *unknown* x_j in the i-th equation.

Definition

A solution to the system of equations is an $n-$tuple of real numbers $[x_1 \; x_2 \; \ldots \; x_n]$ which satisfies each of the equations.

- For a system like the one above, we are interested in:
  1. Does a solution exists?
  2. How many solutions are there?
- There are essentially three ways of solving such systems:
  1. substitution

Substitution: Two Equations in Two Unknowns

- Consider the system of equations:

$$x + y = 10$$
$$x - y = 6$$

- Use the second equation to write y in terms of x,

$$y = x - 6. \tag{1}$$

- Substitute this expression for y into the 1st equation:

$$x + (x - 6) = 10 \implies x = 8$$

- Use (1) to obtain,

$$y = 2$$

Gaussian Elimination: Two Equations in Two Unknowns

More generally, one can:

1. add the first equation to the second, to eliminate y;
2. obtain y from the first equation.

This leads to the following transformation

$$\left.\begin{array}{r} x + y = 10 \\ x - y = 6 \end{array}\right\} \implies \left\{\begin{array}{r} x + y = 10 \\ 2x = 10 + 6 \end{array}\right. \implies \left\{\begin{array}{r} x + y = 10 \\ x = \dfrac{1}{2}10 + \dfrac{1}{2}6 \end{array}\right.$$

Obviously the solution is

$$x = \tfrac{1}{2}(10 + 6) = 8, \; y = 10 - \tfrac{1}{2}(10 + 6) = 2$$

Using Matrix Notation, I

Matrix notation allows the two equations

$$
\begin{aligned}
1x + 1y &= 10 \\
1x - 1y &= 6
\end{aligned}
$$

to be expressed as

$$
\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}
$$

or as $\mathbf{Az} = \mathbf{b}$, where

$$
\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}.
$$

1. $\mathbf{A}$ is the coefficient matrix;
2. $\mathbf{z}$ is the vector of unknowns;
3. $\mathbf{b}$ is the vector of right-hand sides.

Using Matrix Notation, II

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}$$
$$\Updownarrow$$
$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 10 \\ 6 \end{pmatrix}$$
$$\Updownarrow$$
$$\begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 10 \\ -4 \end{pmatrix}$$
$$\Updownarrow$$
$$\begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 10 \\ -4 \end{pmatrix}$$
$$\Updownarrow$$
$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \end{pmatrix}$$

Using Matrix Notation, III

Also, the solution $x = \frac{1}{2}(10 + 6)$, $y = \frac{1}{2}(10 - 6)$
can be expressed as

$$\begin{aligned} x &= \tfrac{1}{2}10 + \tfrac{1}{2}6 \\ y &= \tfrac{1}{2}10 - \tfrac{1}{2}6 \end{aligned}$$

or as

$$\mathbf{z} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 10 \\ 6 \end{pmatrix} = \mathbf{Cb}, \quad \text{where} \quad \mathbf{C} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

Gaussian Elimination: Two General Equations

Consider the general system

$$\begin{array}{rcccl} ax + by & = & u & = & 1u + 0v \\ cx + dy & = & v & = & 0u + 1v \end{array}$$

of two equations in two unknowns, filled in with 1s and 0s.

In matrix form, these equations can be written as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

In case $a \neq 0$, we can eliminate x from the second equation by adding $-c/a$ times the first row to the second.

After defining the scalar $D := a[d + (-c/a)b] = ad - bc$,

$$\begin{pmatrix} a & b \\ 0 & D/a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c/a & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Gaussian Elimination: Two General Equations-Subcase 1A

In **Subcase 1A** when $D := ad - bc \neq 0$,
multiply the second row by a to obtain

$$\begin{pmatrix} a & b \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Adding $-b/D$ times the second row to the first yields

$$\begin{pmatrix} a & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 + (bc/D) & -ab/D \\ -c & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Recognizing that $1 + (bc/D) = (D + bc)/D = ad/D$,
then dividing the two rows/equations by a and D respectively,
we obtain

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{D} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

which implies the unique solution

$$x = (1/D)(du - bv) \quad \text{and} \quad y = (1/D)(av - cu)$$

Gaussian Elimination: Two General Equations-Subcase 1B

In **Subcase 1B** when $D := ad - bc = 0$,
the multiplier $-ab/D$ is undefined and the system

$$\begin{pmatrix} a & b \\ 0 & D/a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c/a & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

collapses to

$$\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v - c/a \end{pmatrix}.$$

This leaves us with two "subsubcases":

if $c \neq av$, then the left-hand side of the second equation is 0,
but the right-hand side is non-zero,
so there is no solution;

if $c = av$, then the second equation reduces to $0 = 0$,
and there is a continuum of solutions
satisfying the one remaining equation $ax + by = u$,
or $x = (u - by)/a$ where y is any real number.

Gaussian Elimination: Two General Equations-Case 2

In the final case when $a = 0$, simply interchanging the two equations

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} .$$

gives

$$\begin{pmatrix} c & d \\ 0 & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v \\ u \end{pmatrix} .$$

Provided that $b \neq 0$, one has $y = u/b$ and,
assuming that $c \neq 0$, also $x = (v - dy)/c = (bv - du)/bc$.

On the other hand, if $b = 0$,
we are back with two possibilities like those of Subcase 1B.

The Transpose of a Matrix

The transpose of the $m \times n$ matrix $\mathbf{A} = (a_{ij})_{m \times n}$
is the $n \times m$ matrix

$$\mathbf{A}^\top = (a_{ij}^\top)_{n \times m} = (a_{ji})_{n \times m} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}$$

which results from transforming each column m-vector $\mathbf{a}_j = (a_{ij})_{i=1}^m$ $(j = 1, 2, \dots, n)$ of $\mathbf{A}$ into the corresponding row m-vector $\mathbf{a}_j^\top = (a_{ji}^\top)_{i=1}^m$ of $\mathbf{A}^\top$.

Equivalently, for each $i = 1, 2, \dots, m$, the ith row n-vector $\mathbf{a}_i^\top = (a_{ij})_{j=1}^n$ of $\mathbf{A}$ is transformed into the ith column n-vector $\mathbf{a}_i = (a_{ji})_{j=1}^n$ of $\mathbf{A}^\top$.

Either way, one has $a_{ij}^\top = a_{ji}$ for all relevant pairs i, j.

Matrix Multiplication I

A scalar, usually denoted by a Greek letter, is a real number $\alpha \in \mathbb{R}$.

The product of any $m \times n$ matrix $\mathbf{A} = (a_{ij})^{m \times n}$ and any scalar $\alpha \in \mathbb{R}$ is the new $m \times n$ matrix denoted by $\alpha \mathbf{A} = (\alpha a_{ij})^{m \times n}$, each of whose elements αa_{ij} results from multiplying the corresponding element a_{ij} of $\mathbf{A}$ by α.

Matrix Multiplication II

The matrix product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined (whenever possible) as the matrix $\mathbf{C} = \mathbf{AB} = (c_{ij})_{m \times n}$ whose element c_{ij} in row i and column j is the inner product $c_{ij} = \mathbf{a}_i^\top \mathbf{b}_j$ of:

- the ith row vector $\mathbf{a}_i^\top$ of the first matrix $\mathbf{A}$;
- the jth column vector $\mathbf{b}_j$ of the second matrix $\mathbf{B}$.

Note that the resulting matrix product $\mathbf{C}$ must have:

- as many rows as the first matrix $\mathbf{A}$;
- as many columns as the second matrix $\mathbf{B}$.

Laws of Matrix Multiplication

The following laws of matrix multiplication hold whenever the matrices are compatible for multiplication.

associative: $\mathbf{A}(\mathbf{B}\mathbf{C}) = (\mathbf{A}\mathbf{B})\mathbf{C}$;

distributive: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C}$ and $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{A}\mathbf{C} + \mathbf{B}\mathbf{C}$;

transpose: $(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top\mathbf{A}^\top$.

shifting scalars: $\alpha(\mathbf{A}\mathbf{B}) = (\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B})$ for all $\alpha \in \mathbb{R}$.

Exercise

Let $\mathbf{X}$ *be any* $m \times n$ *matrix, and* $\mathbf{z}$ *any column* n*-vector.*

1. *Show that the matrix product* $\mathbf{z}^\top\mathbf{X}^\top\mathbf{X}\mathbf{z}$ *is well-defined, and that its value is a scalar.*

Matrix Multiplication Does Not Commute

The two matrices **A** and **B** commute just in case $\mathbf{AB} = \mathbf{BA}$.

Note that typical pairs of matrices DO NOT commute, meaning that $\mathbf{AB} \neq \mathbf{BA}$ — i.e., the order of multiplication matters.

Indeed, suppose that **A** is $\ell \times m$ and **B** is $m \times n$, as is needed for **AB** to be defined.

Then the reverse product **BA** is undefined except in the special case when $n = \ell$.

Hence, for both **AB** and **BA** to be defined, where **B** is $m \times n$, the matrix **A** must be $n \times m$.

But then **AB** is $n \times n$, whereas **BA** is $m \times m$.

Evidently $\mathbf{AB} \neq \mathbf{BA}$ unless $m = n$.

Thus all four matrices **A**, **B**, **AB** and **BA** are $m \times m = n \times n$.

We must be in the special case when all four are square matrices of the same dimension.

Matrix Multiplication Does Not Commute, II

Even if both **A** and **B** are $n \times n$ matrices,
implying that both **AB** and **BA** are also $n \times n$,
one can still have $\mathbf{AB} \neq \mathbf{BA}$.

Here is a 2×2 example:

Example

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

More Warnings Regarding Matrix Multiplication

Exercise

Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *denote three matrices. Give examples showing that*

1. *The matrix* $\mathbf{AB}$ *might be defined, even if* $\mathbf{B}A$ *is not.*
2. *One can have* $\mathbf{AB} = \mathbf{0}$ *even though* $\mathbf{A} \neq \mathbf{0}$ *and* $\mathbf{B} \neq \mathbf{0}$*.*
3. *If* $\mathbf{AB} = \mathbf{AC}$ *and* $\mathbf{A} \neq \mathbf{0}$*, it does not follow that* $\mathbf{B} = \mathbf{C}$*.*

Square Matrices

A square matrix has an equal number of rows and columns, this number being called its dimension.

The (principal) diagonal of a square matrix $\mathbf{A} = (a_{ij})_{n \times n}$ of dimension n is the list $(a_{ii})_{i=1}^{n} = (a_{11}, a_{22}, \ldots, a_{nn})$ of its diagonal elements a_{ii}.

The other elements a_{ij} with $i \neq j$ are the off-diagonal elements.

A square matrix is often expressed in the form

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

with some extra dots along the diagonal.

Symmetric Matrices

A square matrix $\mathbf{A}$ is symmetric if it is equal to its transpose — i.e., if $\mathbf{A}^\top = \mathbf{A}$.

Example

The product of two symmetric matrices need not be symmetric.

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{but} \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Two Exercises with Symmetric Matrices

Exercise

Let $\mathbf{x}$ *be a column n-vector.*

1. *Find the dimensions of* $\mathbf{x}^\top \mathbf{x}$ *and of* $\mathbf{x}\mathbf{x}^\top$*.*
2. *Show that one is a non-negative number which is positive unless* $\mathbf{x} = \mathbf{0}$*, and that the other is a symmetric matrix.*

Exercise

Let $\mathbf{A}$ *be an* $m \times n$*-matrix.*

1. *Find the dimensions of* $\mathbf{A}^\top \mathbf{A}$ *and of* $\mathbf{A}\mathbf{A}^\top$*.*
2. *Show that both* $\mathbf{A}^\top \mathbf{A}$ *and of* $\mathbf{A}\mathbf{A}^\top$ *are symmetric matrices.*
3. *What is a necessary condition for* $\mathbf{A}^\top \mathbf{A} = \mathbf{A}\mathbf{A}^\top$*?*
4. *What is a sufficient condition for* $\mathbf{A}^\top \mathbf{A} = \mathbf{A}\mathbf{A}^\top$*?*

Diagonal Matrices

A square matrix $\mathbf{A} = (a_{ij})^{n \times n}$ is diagonal just in case all of its off diagonal elements a_{ij} with $i \neq j$ are 0.

A diagonal matrix of dimension n can be written in the form

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_n \end{pmatrix} = \mathbf{diag}(d_1, d_2, d_3, \dots, d_n) = \mathbf{diag}\,\mathbf{d}$$

where the n-vector $\mathbf{d} = (d_1, d_2, d_3, \dots, d_n) = (d_i)_{i=1}^n$ consists of the diagonal elements of $\mathbf{D}$.

Obviously, any diagonal matrix is symmetric.

Multiplying by Diagonal Matrices

Example

Let $\mathbf{D}$ be a diagonal matrix of dimension n.

Suppose that $\mathbf{A}$ and $\mathbf{B}$ are $m \times n$ and $n \times m$ matrices, respectively.

Then $\mathbf{E} := \mathbf{AD}$ and $\mathbf{F} := \mathbf{DB}$ are well defined $m \times n$ and $n \times m$ matrices, respectively.

By the law of matrix multiplication, their elements are

$$e_{ij} = \sum_{k=1}^{n} a_{ik} d_{kj} = a_{ij} d_{jj} \text{ and } f_{ij} = \sum_{k=1}^{n} d_{ik} b_{kj} = d_{ii} b_{ij}$$

Thus, the jth column $\mathbf{e}_j$ of $\mathbf{AD}$ is the product $d_{jj}\mathbf{a}_j$ of the jth element of $\mathbf{D}$ and the jth column $\mathbf{a}_j$ of $\mathbf{A}$.

Similarly, the ith row $\mathbf{f}_i^\top$ of $\mathbf{DB}$ is the product $d_{ii}\mathbf{b}_i^\top$ of the ith element of $\mathbf{D}$ and the ith row $\mathbf{b}_i^\top$ of $\mathbf{B}$.

The Identity Matrix: Definition

The identity matrix of dimension n is the diagonal matrix

$$\mathbf{I}_n = \mathbf{diag}(1, 1, \ldots, 1)$$

whose n diagonal elements are all equal to 1.

Equivalently, it is the $n \times n$-matrix $\mathbf{A} = (a_{ij})^{n \times n}$ whose elements are all given by $a_{ij} = \delta_{ij}$ for the Kronecker delta function $(i, j) \mapsto \delta_{ij}$ defined on $\{1, 2, \ldots, n\}^2$.

Exercise

Given any $m \times n$ matrix $\mathbf{A}$, verify that $\mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}$.

Matrices Representing Elementary Row Operations

$$\mathbf{E}_i(\alpha) = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & \dots & \alpha & \dots & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{pmatrix}, \mathbf{E}_{ij}(\alpha) = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & \alpha & \dots & 1 & \dots & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{pmatrix}$$

1. $\mathbf{E}_{ij}$ is the matrix obtained by interchanging the ith and jth rows of the identity matrix.
2. $\mathbf{E}_i(\alpha)$ is the matrix obtained by multiplying the ith row of the identity matrix by $\alpha \in \mathbb{R}$,
3. $\mathbf{E}_{ij}(\alpha)$ is the matrix obtained by adding $\alpha \in \mathbb{R}$ times row i to row j in the identity matrix.

Theorem

Let E be an elementary $n \times n$ matrix obtained by performing a particular row operation on the $n \times n$ identity. For any $n \times m$ matrix A, EA is the matrix obtained by performing the same row operation on A.

Orthogonal Matrices

An n-dimensional square matrix $\mathbf{Q}$ is said to be orthogonal just in case its columns form an orthonormal set — i.e., they must be pairwise orthogonal unit vectors.

Theorem

A square matrix $\mathbf{Q}$ is orthogonal if and only if it satisfies $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}$.

Proof.

The elements of the matrix product $\mathbf{Q}\mathbf{Q}^\top$ satisfy

$$(\mathbf{Q}\mathbf{Q}^\top)_{ij} = \sum_{k=1}^{n} q_{ik} q_{jk} = \mathbf{q}_i \cdot \mathbf{q}_j$$

where $\mathbf{q}_i$ (resp. $\mathbf{q}_j$) denotes the ith (resp. jth) column vector of $\mathbf{Q}$.

But the columns of $\mathbf{Q}$ are orthonormal iff $\mathbf{q}_i \cdot \mathbf{q}_j = \delta_{ij}$ for all $i, j = 1, 2, \ldots, n$, and so iff $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}$. □

Triangular Matrices: Definition

Definition

A square matrix is upper (resp. lower) triangular if all its non-zero off diagonal elements are above and to the right (resp. below and to the left) of the diagonal — i.e., in the upper (resp. lower) triangle bounded by the principal diagonal.

The elements of an upper (resp. lower) triangular matrix satisfy

$$(\mathbf{U})_{ij} = 0 \text{ whenever } i > j, \text{ and } (\mathbf{L})_{ij} = 0 \text{ whenever } i < j$$

Triangular Matrices: Exercises

Exercise

Prove that the transpose:

1. $\mathbf{U}^\top$ *of any upper triangular matrix* $\mathbf{U}$ *is lower triangular;*
2. $\mathbf{L}^\top$ *of any lower triangular matrix* $\mathbf{U}$ *is upper triangular.*

Exercise

Consider the matrix $\mathbf{E}_{ij}(\alpha)$ *that represents the elementary row operation of adding a multiple of* α *times row* j *to row* i*. Under what conditions is*
$\mathbf{E}_{ij}(\alpha)$
(i) upper triangular?
(ii) lower triangular?

Upper Triangular Matrices: Multiplication

Theorem

The product $\mathbf{W} = \mathbf{UV}$ *of any two upper triangular matrices* $\mathbf{U}, \mathbf{V}$ *is upper triangular, with diagonal elements* $w_{ii} = u_{ii} v_{ii}$ $(i = 1, \ldots, n)$ *equal to the product of the corresponding diagonal elements of* $\mathbf{U}, \mathbf{V}$*.*

Proof.

Given any two upper triangular $n \times n$ matrices $\mathbf{U}$ and $\mathbf{V}$, the elements $(w_{ij})^{n \times n}$ of their product $\mathbf{W} = \mathbf{UV}$ satisfy

$$w_{ij} = \begin{cases} \sum_{k=i}^{j} u_{ik} v_{kj} & \text{if } i \leq j \\ 0 & \text{if } i > j \end{cases}$$

because $u_{ik} v_{kj} = 0$ unless both $i \leq k$ and $k \leq j$.

So $\mathbf{W} = \mathbf{UV}$ is upper triangular.

Finally, putting $j = i$ implies that $w_{ii} = u_{ii} v_{ii}$ for $i = 1, \ldots, n$.

Lower Triangular Matrices: Multiplication

Theorem

The product of any two lower triangular matrices is lower triangular.

Proof.

Given any two lower triangular matrices $\mathbf{L}, \mathbf{M}$, taking transposes shows that $(\mathbf{LM})^\top = \mathbf{M}^\top \mathbf{L}^\top = \mathbf{U}$, where the product $\mathbf{U}$ is upper triangular, as the product of upper triangular matrices. Hence $\mathbf{LM} = \mathbf{U}^\top$ is lower triangular, as the transpose of an upper triangular matrix. □

Row Echelon Matrices: Definition

Definition

For each $i \in \{1, \ldots, m\}$, the leading non-zero element in any non-zero row i of the $m \times n$ matrix $\mathbf{A}$ is a_{i,j_i}, where
$j_i := \arg\min\{j \in \{1, \ldots, n\} \mid a_{i,j} \neq 0\}$.

The $m \times n$ matrix $\mathbf{R} = (a_{ij})^{m \times n}$ is in row echelon form if;

1. each zero row comes after each row that is not zero;
2. if the leading non-zero element in each row $i \in \{1, \ldots, p\}$, where p is the number of non-zero rows, is denoted by a_{i,j_i}, then $1 \leq j_1 < j_2 < \ldots j_p$.

Reduced Row Echelon Matrices

A matrix is in reduced row echelon form (also called row canonical form) if it satisfies the additional condition:

- Every leading non-zero coefficient a_{i,j_i} is 1 and it is the only nonzero entry in column j_i.

Here is an example:

$$\begin{pmatrix} 1 & 0 & \frac{1}{2} & 0 & b_1 \\ 0 & 1 & -\frac{1}{3} & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{pmatrix}$$

Theorem

For any $k \times n$ matrix A there exists elementary matrices $E_1, E_2, ..., E_m$ such that the matrix product $E_m \, E_{m-1} ... E_1 \, A = U$ where U is in (reduced) row echelon form.

Partitioned Matrices: Definition

A partitioned matrix is a rectangular array of different matrices.

Example

Consider the $(m + \ell) \times (n + k)$ matrix

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$$

where the four submatrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$
are of dimension $m \times n$, $m \times k$, $\ell \times n$ and $\ell \times k$ respectively.

For any scalar $\alpha \in \mathbb{R}$,
the scalar multiple of a partitioned matrix is

$$\alpha \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} \alpha\mathbf{A} & \alpha\mathbf{B} \\ \alpha\mathbf{C} & \alpha\mathbf{D} \end{pmatrix}$$

Partitioned Matrices: Addition

Suppose the two partitioned matrices

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}$$

have the property that the following four pairs of corresponding matrices have equal dimensions: (i) $\mathbf{A}$ and $\mathbf{E}$; (ii) $\mathbf{B}$ and $\mathbf{F}$; (iii) $\mathbf{C}$ and $\mathbf{G}$; (iv) $\mathbf{D}$ and $\mathbf{H}$.

Then the sum of the two matrices is

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} + \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix} = \begin{pmatrix} \mathbf{A} + \mathbf{E} & \mathbf{B} + \mathbf{F} \\ \mathbf{C} + \mathbf{G} & \mathbf{D} + \mathbf{H} \end{pmatrix}$$

Partitioned Matrices: Multiplication

Provided that the two partitioned matrices

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}$$

along with their sub-matrices are all compatible for multiplication, the product is defined as

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix} = \begin{pmatrix} \mathbf{AE} + \mathbf{BG} & \mathbf{AF} + \mathbf{BH} \\ \mathbf{CE} + \mathbf{DG} & \mathbf{CF} + \mathbf{DH} \end{pmatrix}$$

This adheres to the usual rule for multiplying rows by columns.

Partitioned Matrices: Transposes

The rule for transposing a partitioned matrix is

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^\top = \begin{pmatrix} \mathbf{A}^\top & \mathbf{C}^\top \\ \mathbf{B}^\top & \mathbf{D}^\top \end{pmatrix}$$

So the original matrix is symmetric iff $\mathbf{A} = \mathbf{A}^\top$, $\mathbf{D} = \mathbf{D}^\top$, $\mathbf{B} = \mathbf{C}^\top$, and $\mathbf{C} = \mathbf{B}^\top$.

It is diagonal iff $\mathbf{A}, \mathbf{D}$ are both diagonal, while $\mathbf{B} = \mathbf{0}$ and $\mathbf{C} = \mathbf{0}$.

The identity matrix is diagonal with $\mathbf{A} = \mathbf{I}$, $\mathbf{D} = \mathbf{I}$, possibly identity matrices of different dimensions.

Linear Functions: Definition

Definition

A linear combination of vectors is the weighted sum $\sum_{h=1}^{k} \lambda_h \mathbf{x}^h$, where $\mathbf{x}^h \in V$ and $\lambda_h \in \mathbb{F}$ for $h = 1, 2, \ldots, k$.

Definition

A function $V : \mathbf{u} \mapsto f(\mathbf{u}) \in \mathbb{F}$ is linear provided that

$$f(\lambda \mathbf{u} + \mu \mathbf{v}) = \lambda f(\mathbf{u}) + \mu f(\mathbf{v})$$

whenever $\mathbf{u}, \mathbf{v} \in V$ and $\lambda, \mu \in V$.

Key Properties of Linear Functions

Exercise

By induction on k, show that if the function $f : V \to \mathbb{F}$ is linear, then

$$f\left(\sum_{h=1}^{k} \lambda_h \mathbf{x}^h\right) = \sum_{h=1}^{k} \lambda_h f(\mathbf{x}^h)$$

for all linear combinations $\sum_{h=1}^{k} \lambda_h \mathbf{x}^h$ in V
— i.e., f preserves linear combinations.

Exercise

In case $V = \mathbb{R}^n$ and $\mathbb{F} = \mathbb{R}$, show that any linear function is homogeneous of degree 1, meaning that $f(\lambda \mathbf{v}) = \lambda f(\mathbf{v})$ for all $\lambda > 0$ and all $\mathbf{v} \in \mathbb{R}^n$.

Affine Functions

Definition

A function $g : V \to \mathbb{F}$ is said to be affine if there is a scalar additive constant $\alpha \in \mathbb{F}$ and a linear function $f : V \to \mathbb{F}$ such that $g(\mathbf{v}) \equiv \alpha + f(\mathbf{v})$.

Quadratic Forms

Definition

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

$$Q(x_1, \ldots, x_n) \equiv \sum_{i \leq j} a_{ij} x_i x_j = x^T A \, x$$

Example

$Q(x_1, x_2) = a_{11} x_1^2 + a_{12} x_1 x_2 + a_{22} x_2^2$ can be written as

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2} a_{12} \\ \frac{1}{2} a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Negative and Semi-Negative Definite Matrices

Definition

Let A be an $n \times n$ symmetric matrix, then A is:

1. Positive definite if $x^T Ax > 0$ for all $x \neq 0$ in $\mathbb{R}^K$,
2. positive semidefinite if $x^T Ax \geq 0$ for all $x \neq 0$ in $\mathbb{R}^K$,
3. negative definite if $x^T Ax < 0$ for all $x \neq 0$ in $\mathbb{R}^K$,
4. negative semidefinite if $x^T Ax \leq 0$ for all $x \neq 0$ in $\mathbb{R}^K$,
5. indefinite if $x^T Ax > 0$ for some $x \in \mathbb{R}^K$ and $x^T Ax < 0$ for some other $x \in \mathbb{R}^K$.

Determinants of Order 2: Definition

Consider again the pair of linear equations

$$\begin{array}{rcrcl} a_{11}x_1 & + & a_{12}x_2 & = & b_1 \\ a_{21}x_1 & + & a_{12}x_2 & = & b_2 \end{array}$$

with its associated coefficient matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Let us define $D := a_{11}a_{22} - a_{21}a_{12}$.

Provided that $D \neq 0$, there is a unique solution given by

$$x_1 = \frac{1}{D}(b_1 a_{22} - b_2 a_{12}), \quad x_2 = \frac{1}{D}(b_2 a_{11} - b_1 a_{21})$$

The number D is called the determinant of the matrix $\mathbf{A}$, and denoted by either $\det(\mathbf{A})$ or more concisely, $|\mathbf{A}|$.

Determinants of Order 2: Simple Rule

Thus, for any 2×2 matrix $\mathbf{A}$, its determinant D is

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

For this special case of order 2 determinants, a simple rule is:

1. multiply the diagonal elements together;
2. multiply the off-diagonal elements together;
3. subtract the product of the off-diagonal elements from the product of the diagonal elements.

Note that

$$|\mathbf{A}| = a_{11}a_{22} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + a_{21}a_{12} \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$$

Cramer's Rule in the 2×2 Case

Using determinant notation, the solution to the equations

$$\begin{array}{ccccc} a_{11}x_1 & + & a_{12}x_2 & = & b_1 \\ a_{21}x_1 & + & a_{12}x_2 & = & b_2 \end{array}$$

can be written in the alternative form

$$x_1 = \frac{1}{D}\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}, \qquad x_2 = \frac{1}{D}\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}$$

This accords with Cramer's rule for the solution to $\mathbf{Ax} = \mathbf{b}$, which is the vector $\mathbf{x} = (x_i)_{i=1}^n$ each of whose components x_i is the fraction with:

1. denominator equal to the determinant D of the coefficient matrix $\mathbf{A}$ (provided, of course, that $D \neq 0$);
2. numerator equal to the determinant of the matrix $(\mathbf{A}_{-i}, \mathbf{b})$ formed from $\mathbf{A}$ by replacing its ith column with the $\mathbf{b}$ vector of right-hand side elements.

Determinants of Order 3: Definition

Determinants of order 3 can be calculated
from those of order 2 according to the formula

$$\begin{aligned} |\mathbf{A}| &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= \sum_{j=1}^{3} (-1)^{1+j} a_{1j} |\mathbf{A}_{1j}| \end{aligned}$$

where, for $j = 1, 2, 3$, the 2×2 matrix $\mathbf{A}_{1j}$ is the sub-matrix obtained by removing both row 1 and column j from $\mathbf{A}$. The scalar $C_{ij} = (-1)^{i+j} |A_{ij}|$ is called the(i, j)th-cofactor of $\mathbf{A}$.

The result is the following sum

$$\begin{aligned} |\mathbf{A}| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} \\ - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \end{aligned}$$

of $3! = 6$ terms, each the product of 3 elements chosen
so that each row and each column is represented just once.

Determinants of Order 3: Cofactor Expansion

The determinant expansion

$$|\mathbf{A}| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

is very symmetric, suggesting (correctly) that the cofactor expansion along the first row (a_{11}, a_{12}, a_{13})

$$|\mathbf{A}| = \sum_{j=1}^{3} (-1)^{1+j} a_{1j} |\mathbf{A}_{1j}|$$

gives the same answer as the two cofactor expansions

$$|\mathbf{A}| = \sum_{j=1}^{3} (-1)^{r+j} a_{rj} |\mathbf{A}_{rj}| = \sum_{i=1}^{3} (-1)^{i+s} a_{is} |\mathbf{A}_{is}|$$

along, respectively:

- the rth row (a_{r1}, a_{r2}, a_{r3})
- the sth column (a_{1s}, a_{2s}, a_{3s})

Eight Basic Rules (Rules A–H of EMEA, Section 16.4)

Let $|\mathbf{A}|$ denote the determinant of any $n \times n$ matrix $\mathbf{A}$.

1. $|\mathbf{A}| = 0$ if all the elements in a row (or column) of $\mathbf{A}$ are 0.
2. $|\mathbf{A}^\top| = |\mathbf{A}|$, where $\mathbf{A}^\top$ is the transpose of $\mathbf{A}$.
3. If all the elements in a single row (or column) of $\mathbf{A}$ are multiplied by a scalar α, so is its determinant.
4. If two rows (or two columns) of $\mathbf{A}$ are interchanged, the determinant changes sign, but not its absolute value.
5. If two of the rows (or columns) of $\mathbf{A}$ are proportional, then $|\mathbf{A}| = 0$.
6. The value of the determinant of $\mathbf{A}$ is unchanged if any multiple of one row (or one column) is added to a different row (or column) of $\mathbf{A}$.
7. The determinant of the product $|\mathbf{AB}|$ of two $n \times n$ matrices equals the product $|\mathbf{A}| \cdot |\mathbf{B}|$ of their determinants.
8. If α is any scalar, then $|\alpha\mathbf{A}| = \alpha^n|\mathbf{A}|$.

The Adjugate Matrix

Definition

The adjugate (or "(classical) adjoint") $\mathbf{adj}\,\mathbf{A}$ of an order n square matrix $\mathbf{A}$ has elements given by $(\mathbf{adj}\,\mathbf{A})_{ij} = \mathbf{C}_{ji}$.

It is therefore the transpose of the cofactor matrix $\mathbf{C}$ whose elements are the respective cofactors of $\mathbf{A}$.

See Example 9.3 in page 195 of Simon and Blume for a detailed example.

Definition of Inverse Matrix

Definition

The $n \times n$ matrix $\mathbf{X}$ is the inverse of the invertible $n \times n$ matrix $\mathbf{A}$ provided that $\mathbf{AX} = \mathbf{XA} = \mathbf{I}_n$.

In this case we write $\mathbf{X} = \mathbf{A}^{-1}$, so $\mathbf{A}^{-1}$ denotes the (unique) inverse.

Big question: does the inverse exist? is it unique?

Existence Conditions

Theorem

An $n \times n$ matrix $\mathbf{A}$ has an inverse if and only if $|\mathbf{A}| \neq 0$, which holds if and only if at least one of the equations $\mathbf{AX} = \mathbf{I}_n$ and $\mathbf{XA} = \mathbf{I}_n$ has a solution.

Proof.

Provided $|\mathbf{A}| \neq 0$, the identity $(\mathbf{adj}\,\mathbf{A})\mathbf{A} = \mathbf{A}(\mathbf{adj}\,\mathbf{A}) = |\mathbf{A}|\mathbf{I}_n$ shows that the matrix $\mathbf{X} := (1/|\mathbf{A}|)\,\mathbf{adj}\,\mathbf{A}$ is well defined and satisfies $\mathbf{XA} = \mathbf{AX} = \mathbf{I}_n$, so $\mathbf{X}$ is the inverse $\mathbf{A}^{-1}$.

Conversely, if either $\mathbf{XA} = \mathbf{I}_n$ or $\mathbf{AX} = \mathbf{I}_n$ has a solution, then the product rule for determinants implies
that $1 = |\mathbf{I}_n| = |\mathbf{AX}| = |\mathbf{XA}| = |\mathbf{A}||\mathbf{X}|$, and so $|\mathbf{A}| \neq 0$. The rest follows from the paragraph above. □

Singularity

So $\mathbf{A}^{-1}$ exists if and only if $|\mathbf{A}| \neq 0$.

Definition

1. In case $|\mathbf{A}| = 0$, the matrix $\mathbf{A}$ is said to be singular;
2. In case $|\mathbf{A}| \neq 0$, the matrix $\mathbf{A}$ is said to be non-singular or invertible.

Example and Application to Simultaneous Equations

Exercise

Verify that

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \Longrightarrow \mathbf{A}^{-1} = \mathbf{C} := \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

by using direct multiplication to show that $\mathbf{AC} = \mathbf{CA} = \mathbf{I}_2$.

Example

Suppose that a system of n simultaneous equations in n unknowns is expressed in matrix notation as $\mathbf{Ax} = \mathbf{b}$.

Of course, $\mathbf{A}$ must be an $n \times n$ matrix.

Suppose $\mathbf{A}$ has an inverse $\mathbf{A}^{-1}$.

Premultiplying both sides of the equation $\mathbf{Ax} = \mathbf{b}$ by this inverse gives $\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b}$, which simplifies to $\mathbf{Ix} = \mathbf{A}^{-1}\mathbf{b}$.

Hence the unique solution of the equation is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Cramer's Rule: Statement

Notation

Given any $m \times n$ matrix $\mathbf{A}$, let $[\mathbf{A}_{-j}, \mathbf{b}]$ denote the new $m \times n$ matrix in which column j has been replaced by the column vector $\mathbf{b}$.

Evidently $[\mathbf{A}_{-j}, \mathbf{a}_j] = \mathbf{A}$.

Theorem

Provided that the $n \times n$ matrix $\mathbf{A}$ is invertible, the simultaneous equation system $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ whose ith component is given by the ratio $x_i = |[\mathbf{A}_{-i}, \mathbf{b}]|/|\mathbf{A}|$.

This result is known as Cramer's rule.

Rule for Inverting Products

Theorem

Suppose that $\mathbf{A}$ *and* $\mathbf{B}$ *are two invertible* $n \times n$ *matrices.*

Then the inverse of the matrix product $\mathbf{AB}$ *exists,*
and is the reverse product $\mathbf{B}^{-1}\mathbf{A}^{-1}$ *of the inverses.*

Proof.

Using the associative law for matrix multiplication repeatedly gives:

$$(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}(\mathbf{I})\mathbf{B} = \mathbf{B}^{-1}(\mathbf{IB}) = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$$

and

$$(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{A}(\mathbf{I})\mathbf{A}^{-1} = (\mathbf{AI})\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}.$$

These equations confirm that $\mathbf{X} := \mathbf{B}^{-1}\mathbf{A}^{-1}$ is the unique matrix satisfying the double equality $(\mathbf{AB})\mathbf{X} = \mathbf{X}(\mathbf{AB}) = \mathbf{I}$. □