Sociology 761 John Fox

Lecture Notes

Introduction to Matrices

1. The Cast of Characters

- A *matrix* is a rectangular array (i.e., a table) of numbers.
  - For example,

$$\underset{(4\times 3)}{\mathbf{X}} = \begin{bmatrix} 1 & -2 & 3 \\ 4 & -5 & -6 \\ 7 & 8 & 9 \\ 0 & 0 & 10 \end{bmatrix}$$

    – This matrix, with 4 rows and 3 columns, is of *order* 4 by 3. For clarity, I will often show the order in parentheses below a matrix.

    – Note that a matrix is represented by a bold-face upper-case letter.

- A more general example:

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

- This matrix is of order $m \times n$.
- a_{ij} is the *entry* or *element* in the ith row and jth column of the matrix.
- An individual number (e.g., -2 or a_{12}, such as a matrix element, is called a *scalar*.
- Two matrices are *equal* if they are of the same order and all corresponding entries are equal.

▶ A *column vector* is a one-column matrix:

$$\underset{(4\times 1)}{\mathbf{y}} = \begin{bmatrix} -5 \\ 3 \\ 1 \\ 4 \end{bmatrix}, \qquad \underset{(m\times 1)}{\mathbf{a}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

▶ Likewise, a *row vector* is a one-row matrix:

$$\underset{(1\times n)}{\mathbf{b}'} = [b_1, b_2, \ldots, b_n]$$

- I adopt the convention that row vectors are always written with a prime (i.e., $'$).
- Note that bold-face lower-case letters are used to represent vectors.

- More generally, the *transpose* of a matrix $\mathbf{X}$, written as $\mathbf{X}'$ or $\mathbf{X}^T$, interchanges rows and columns; for example,

$$\underset{(4\times 3)}{\mathbf{X}} = \begin{bmatrix} 1 & -2 & 3 \\ 4 & -5 & -6 \\ 7 & 8 & 9 \\ 0 & 0 & 10 \end{bmatrix}, \qquad \underset{(3\times 4)}{\mathbf{X}'} = \begin{bmatrix} 1 & 4 & 7 & 0 \\ -2 & -5 & 8 & 0 \\ 3 & -6 & 9 & 10 \end{bmatrix}$$

- A *square matrix* of order n has n rows and n columns; for example,

$$\underset{(3\times 3)}{\mathbf{B}} = \begin{bmatrix} -5 & 1 & 3 \\ 2 & 2 & 6 \\ 7 & 3 & -4 \end{bmatrix}$$

  - The *main diagonal* of a square matrix $\underset{(n\times n)}{\mathbf{B}}$ consists of the entries b_{ii}, $i = 1, 2, \ldots, n$.
    - In the example, the main diagonal consists of the entries $-5, 2$, and -4.

- The *trace* of a square matrix is the sum of its diagonal elements:

$$\text{trace}(\mathbf{B}) = \sum_{i=1}^{n} b_{ii} = -5 + 2 - 4 = -7$$

- A square matrix is *symmetric* if it is equal to its transpose.
  - That is, $\underset{(n\times n)}{\mathbf{A}}$ is symmetric if $a_{ij} = a_{ji}$ for all i and j.
  - Thus

$$\mathbf{B} = \begin{bmatrix} -5 & 1 & 3 \\ 2 & 2 & 6 \\ 7 & 3 & -4 \end{bmatrix}$$

    is not symmetric, while

$$\mathbf{C} = \begin{bmatrix} -5 & 1 & 3 \\ 1 & 2 & 6 \\ 3 & 6 & -4 \end{bmatrix}$$

    is symmetric.

- A square matrix is *lower-triangular* if all entries above its main diagonal are 0; for example,

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

- Similarly, an *upper-triangular* matrix has zeroes below the main diagonal; for example,

$$\mathbf{U} = \begin{bmatrix} 5 & 5 & 2 \\ 0 & 2 & -4 \\ 0 & 0 & 3 \end{bmatrix}$$

- A *diagonal matrix* is a square matrix with all off-diagonal entries equal to 0; for example,

$$\mathbf{D} = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix} = \text{diag}(6, -2, 0, 7)$$

- A *scalar matrix* is a diagonal matrix with equal diagonal entries; for example,

$$\mathbf{S} = \begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{bmatrix}$$

- An *identity matrix* is a scalar matrix with ones on the diagonal; for example,

$$\mathbf{I}_3 = \underset{(3\times 3)}{\mathbf{I}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

▶ A *zero matrix* has all of its elements equal to 0; for example,

$$\underset{(4\times 3)}{\mathbf{0}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

▶ A *unit vector* has all of its entries equal to 1; for example,

$$\underset{(4\times 1)}{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

2. Basic Matrix Arithmetic

2.1 Addition, Subtraction, Negation, Product of a Matrix and a Scalar

- *Addition* and *subtraction* are defined for matrices of the same order and are elementwise operations; for example, for

$$\underset{(2\times3)}{\mathbf{A}} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad \underset{(2\times3)}{\mathbf{B}} = \begin{bmatrix} -5 & 1 & 2 \\ 3 & 0 & -4 \end{bmatrix}$$

$$\underset{(2\times3)}{\mathbf{C}} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} -4 & 3 & 5 \\ 7 & 5 & 2 \end{bmatrix}$$

$$\underset{(2\times3)}{\mathbf{D}} = \mathbf{A} - \mathbf{B} = \begin{bmatrix} 6 & 1 & 1 \\ 1 & 5 & 10 \end{bmatrix}$$

- *Matrix negation* and the *product of a matrix and a scalar* are also elementwise operations; examples:

$$\underset{(2\times3)}{\mathbf{A}} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad \underset{(2\times3)}{\mathbf{B}} = \begin{bmatrix} -5 & 1 & 2 \\ 3 & 0 & -4 \end{bmatrix}$$

$$\underset{(2\times3)}{\mathbf{E}} = -\mathbf{A} = \begin{bmatrix} -1 & -2 & -3 \\ -4 & -5 & -6 \end{bmatrix}$$

$$\underset{(2\times3)}{\mathbf{F}} = 3 \times \mathbf{B} = \mathbf{B} \times 3 = \begin{bmatrix} -15 & 3 & 6 \\ 9 & 0 & -12 \end{bmatrix}$$

- Consequently, matrix addition, subtraction, and negation, and the product of a matrix by a scalar obey familiar rules for scalar arithmetic:

$$
\begin{aligned}
\mathbf{A}+\mathbf{B} &= \mathbf{B}+\mathbf{A} \text{ (\textit{commutivity})}\\
\mathbf{A}+(\mathbf{B}+\mathbf{C}) &= (\mathbf{A}+\mathbf{B})+\mathbf{C} \text{ (\textit{associativity})}\\
\mathbf{A}-\mathbf{B} &= \mathbf{A}+(-\mathbf{B}) = -(\mathbf{B}-\mathbf{A})\\
\mathbf{A}-\mathbf{A} &= \mathbf{0} \text{ (}-\mathbf{A}\text{ is the \textit{additive inverse} of }\mathbf{A}\text{)}\\
\mathbf{A}+\mathbf{0} &= \mathbf{A} \text{ (}\mathbf{0}\text{ is the \textit{additive identity})}\\
-(-\mathbf{A}) &= \mathbf{A}\\
c\mathbf{A} &= \mathbf{A}c \text{ (\textit{commutivity})}\\
c(\mathbf{A}+\mathbf{B}) &= c\mathbf{A}+c\mathbf{B} \text{ (\textit{distributive laws})}\\
\mathbf{A}(b+c) &= \mathbf{A}b+\mathbf{A}c\\
0\mathbf{A} &= \mathbf{0} \text{ (\textit{zero})}\\
1\mathbf{A} &= \mathbf{A} \text{ (\textit{unit})}\\
(-1)\mathbf{A} &= -\mathbf{A} \text{ (\textit{negation})}
\end{aligned}
$$

where $\mathbf{A}$,$\mathbf{B}$,$\mathbf{C}$, and $\mathbf{0}$ are matrices of the same order, and $b, c, 0,$ and 1 are scalars.

- Another useful rule is that matrix transposition distributes over addition:

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

2.2 Inner Product and Matrix Product

▶ The *inner product* (or *dot product*) of two vectors with equal numbers of elements is the sum of the product of corresponding elements; that is

$$\underset{(1\times n)}{\mathbf{a}'} \cdot \underset{(n\times 1)}{\mathbf{b}} = \sum_{i=1}^{n} a_i b_i$$

- Note: the inner product is defined similarly between two row vectors or two column vectors.

▶ For example,

$$[2, 0, 1, 3] \cdot \begin{bmatrix} -1 \\ 6 \\ 0 \\ 9 \end{bmatrix} = 2(-1) + 0(6) + 1(0) + 3(9) = 25$$

▶ The matrix product $\mathbf{AB}$ is defined if the number of columns of $\mathbf{A}$ equals the number of rows of $\mathbf{B}$.

- In this case, $\mathbf{A}$ and $\mathbf{B}$ are said to be *conformable for multiplication.*
- Some examples:

$$\underset{(2\times 3)}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}} \underset{(3\times 3)}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \quad \text{conformable}$$

$$\underset{(3\times 3)}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \underset{(2\times 3)}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}} \quad \text{not conformable}$$

$$\underset{(1\times 4)}{[\beta_0, \beta_1, \beta_2, \beta_3]} \underset{(4\times 1)}{\begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}} \quad \text{conformable}$$

$$\underset{(2\times2)}{\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}}\underset{(2\times2)}{\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}} \quad \text{conformable}$$

$$\underset{(2\times2)}{\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}}\underset{(2\times2)}{\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}} \quad \text{conformable}$$

- In general, if $\mathbf{A}$ is $(m \times n)$ then, for the product $\mathbf{AB}$ to be defined, $\mathbf{B}$ must be $(n \times p)$, where m and p are unconstrained.
- Note that $\mathbf{BA}$ may not be defined even if $\mathbf{AB}$ is (i.e., unless $m = p$).

▶ The *matrix product* is defined in terms of the inner product.

- Let $\mathbf{a}_i'$ represent the ith row of the $(m \times n)$ matrix $\mathbf{A}$, and $\mathbf{b}_j$ represent the jth column of the $(n \times p)$ matrix $\mathbf{B}$.
- Then $\mathbf{C} = \mathbf{AB}$ is an $(m \times p)$ matrix with

$$c_{ij} = \mathbf{a}_i' \cdot \mathbf{b}_j = \sum_{k=1}^{n} a_{ik} b_{kj}$$

- That is, the matrix product is formed by successively multiplying each row of the right-hand factor into each column of the left-hand factor.

- Some examples:

$$\underset{(2\times3)}{\begin{bmatrix} & \Longrightarrow & \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}} \underset{(3\times3)}{\begin{bmatrix} & 1 & 0 & 0 \\ \Downarrow & 0 & 1 & 0 \\ & 0 & 0 & 1 \end{bmatrix}}$$

$$= \underset{(2\times3)}{\begin{bmatrix} 1(1)+2(0)+3(0), & 1(0)+2(1)+3(0), & 1(0)+2(0)+3(1) \\ 4(1)+5(0)+6(0), & 4(0)+5(1)+6(0), & 4(0)+5(0)+6(1) \end{bmatrix}}$$

$$= \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$\underset{(1\times4)}{[\beta_0, \beta_1, \beta_2, \beta_3]} \underset{(4\times1)}{\begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}} = \underset{(1\times1)}{[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3]}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 5 \\ 8 & 13 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 9 & 12 \\ 5 & 8 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- In summary:

$\mathbf{A}$	$\mathbf{B}$	$\mathbf{AB}$
general matrix $(m \times n)$	matrix $(n \times p)$	matrix $(m \times p)$
square matrix $(n \times n)$	square matrix $(n \times n)$	square matrix $(n \times n)$
row vector $(1 \times n)$	matrix $(n \times p)$	row vector $(1 \times p)$
matrix $(m \times n)$	column vector $(n \times 1)$	column vector $(m \times 1)$
row vector $(1 \times n)$	column vector $(n \times 1)$	"scalar" (1×1)
column vector $(m \times 1)$	row vector $(1 \times n)$	matrix $(m \times n)$

- Some properties of matrix multiplication:

$$\begin{aligned}
\mathbf{A}(\mathbf{BC}) &= (\mathbf{AB})\mathbf{C} \text{ (\textit{associativity})} \\
(\mathbf{A}+\mathbf{B})\mathbf{C} &= \mathbf{AC}+\mathbf{BC} \text{ (\textit{distributive laws})} \\
\mathbf{A}(\mathbf{B}+\mathbf{C}) &= \mathbf{AB}+\mathbf{AC} \\
\underset{(m\times n)}{\mathbf{A}}\mathbf{I}_n &= \mathbf{I}_m\mathbf{A} = \mathbf{A} \text{ (\textit{unit})} \\
\underset{(m\times n)}{\mathbf{A}}\underset{(n\times p)}{\mathbf{0}} &= \underset{(m\times p)}{\mathbf{0}} \text{ (\textit{zero})} \\
\underset{(q\times m)}{\mathbf{0}}\underset{(m\times n)}{\mathbf{A}} &= \underset{(q\times n)}{\mathbf{0}} \\
(\underset{(m\times n)}{\mathbf{A}}\underset{(n\times p)}{\mathbf{B}})' &= \underset{(p\times n)}{\mathbf{B}'}\underset{(n\times m)}{\mathbf{A}'} \text{ (\textit{transpose of a product})} \\
(\mathbf{AB}\cdots\mathbf{F})' &= \mathbf{F}'\cdots\mathbf{B}'\mathbf{A}'
\end{aligned}$$

- But matrix multiplication is **not** in general *commutative*:
  - For $\underset{(m\times n)}{\mathbf{A}}\underset{(n\times p)}{\mathbf{B}}$, the product $\mathbf{BA}$ is not defined unless $m = p$.
  - For $\underset{(m\times n)}{\mathbf{A}}$ and $\underset{(n\times m)}{\mathbf{B}}$, the product $\mathbf{AB}$ is $(m \times m)$ and $\mathbf{BA}$ is $(n \times n)$, so they cannot be equal unless $m = n$.
  - Even when $\mathbf{A}$ and $\mathbf{B}$ are both $(n \times n)$, the products $\mathbf{AB}$ and $\mathbf{BA}$ need not be equal.
  - When $\mathbf{AB}$ and $\mathbf{BA}$ are equal, the matrices **A** and **B** are said to *commute*.

2.3 The Sense Behind Matrix Multiplication

- The definition of matrix multiplication makes it simple to formulate systems of scalar equations as a single matrix equation, often providing a useful level of abstraction.
- For example, consider the following system of two linear equations in two unknowns, x_1 and x_2:

$$\begin{aligned} 2x_1 + 5x_2 &= 4 \\ x_1 + 3x_2 &= 5 \end{aligned}$$

- Writing these equations as a matrix equation,

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$\underset{(2\times 2)}{\mathbf{A}}\underset{(2\times 1)}{\mathbf{x}} = \underset{(2\times 1)}{\mathbf{b}}$$

▶ More generally, for m linear equations in n unknowns,

$$\underset{(m\times n)}{\mathbf{A}}\underset{(n\times 1)}{\mathbf{x}} = \underset{(m\times 1)}{\mathbf{b}}$$

3. Matrix Inversion

▶ In scalar algebra, division is essential to the solution of simple equations. For example,

$$\begin{aligned} 6x &= 12 \\ x &= \frac{12}{6} = 2 \end{aligned}$$

or, equivalently,

$$\frac{1}{6} \cdot 6x = x = \frac{1}{6} \cdot 12 = 2$$

where $\frac{1}{6} = 6^{-1}$ is the scalar inverse of 6.

▶ In matrix algebra, there is no direct analog of division, but most square matrices have a *matrix inverse*.

- The matrix inverse of the $(n \times n)$ matrix $\mathbf{A}$ is an $(n \times n)$ matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$$

- Square matrices that have an inverse are termed *nonsingular*.
- Squares matrices with no inverse are *singular* (i.e., strange, unusual).
- If an inverse of a square matrix exists, it is unique
- Moreover, if, for square matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{AB} = \mathbf{I}$, then necessarily $\mathbf{BA} = \mathbf{I}$, and $\mathbf{B} = \mathbf{A}^{-1}$.

▶ In scalar algebra, only the number 0 has no inverse (i.e., $0^{-1} = \frac{1}{0}$ is undefined).

▶ In matrix algebra, $\underset{(n \times n)}{\mathbf{0}}$ is singular, but there are also nonzero singular matrices.

- For example, take the matrix
$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$
and suppose for the sake of argument that
$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$
is the inverse of $\mathbf{A}$.

- But
$$\mathbf{AB} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ 0 & 0 \end{bmatrix} \neq \mathbf{I}_2$$
contradicting the claim that $\mathbf{B}$ is the inverse of $\mathbf{A}$, and so $\mathbf{A}$ has no inverse.

- A useful fact to remember is that if the matrix $\mathbf{A}$ is singular then at least one column of $\mathbf{A}$ can be written as a multiple of another column or as a weighted sum (*linear combination*) of the other columns; and, equivalently, at least one row can be written as a multiple of another row or as a weighted sum of the other rows.
  - In the example above, the second row is 0 times the first row, and the second column is 0 times the first column.
  - Indeed, any square matrix with a zero row or column is therefore singular.
  - We'll explore this idea in greater depth presently.

- A convenient property of matrix inverses: If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular matrices of the same order, then their product $\mathbf{AB}$ is also nonsingular, with inverse $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
  - The proof of this property is simple:
  $$(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}$$
  - This result extends directly to the product of any number of nonsingular matrices of the same order:
  $$(\mathbf{AB}\cdots\mathbf{F})^{-1} = \mathbf{F}^{-1}\cdots\mathbf{B}^{-1}\mathbf{A}^{-1}$$

- Example: The inverse of the nonsingular matrix
$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$$
is the matrix
$$\begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}$$
  – Check:
$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \checkmark$$
$$\begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \checkmark$$

▶ Matrix inverses are useful for solving systems of linear simultaneous equations where there is an equal number of equations and unknowns.

- The general form of such a problem, with n equations and n unknowns, is
$$\underset{(n\times n)}{\mathbf{A}}\underset{(n\times 1)}{\mathbf{x}} = \underset{(n\times 1)}{\mathbf{b}}$$
- Here, as explained previously, $\mathbf{A}$ is a matrix of known coefficients; $\mathbf{x}$ is the vector of unknowns; and $\mathbf{b}$ is a vector containing the (known) right-hand sides of the equations.
- The solution is obtained by multiplying both sides of the equation on the left by the inverse of the coefficient matrix:
$$\begin{aligned} \mathbf{A}^{-1}\mathbf{A}\mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \\ \mathbf{I}\mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \\ \mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \end{aligned}$$

- For example, consider the following system of 2 equations in 2 unknowns:

$$\begin{aligned} 2x_1 + 5x_2 &= 4 \\ x_1 + 3x_2 &= 5 \end{aligned}$$

  – In matrix form,

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

  – and so

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 5 \end{bmatrix} \\ &= \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \end{bmatrix} \\ &= \begin{bmatrix} -13 \\ 6 \end{bmatrix} \end{aligned}$$

  – That is, the solution is $x_1 = -13, x_2 = 6$.

  – Check:

$$\begin{aligned} 2(-13) + 5(6) &= 4\checkmark \\ -13 + 3(6) &= 5\checkmark \end{aligned}$$

3.1 Matrix Inversion by Gaussian Elimination

▶ There are many methods for finding inverses of nonsingular matrices.

- *Gaussian elimination* (after Carl Friedrich Gauss, the great German mathematician) is one such method, and has other uses as well.
- In practice, finding matrix inverses is tedious, and is a job best left to a computer.

▶ Gaussian elimination proceeds as follows: Suppose that we want to invert the following matrix (or to determine that it is singular):

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

- Adjoin to this matrix an order-3 identify matrix:

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

- Attempt to reduce the original matrix to the identify matrix by a series of *elementary row operations* (*ERO*s) of three types, bringing the adjoined identify matrix along for the ride:

E_I: Multiply each entry in a row of the matrix by a nonzero scalar constant
E_{II}: Add a scalar multiple of one row to another, replacing the other row.
E_{III}: Interchange two rows of the matrix.

- Starting with the first row, and then proceeding to each row in turn:
  (i) Insure that there is a nonzero entry in the diagonal position, called the *pivot* (e.g., the 1,1 position when we are working on the first row), exchanging the current row for a lower row, if necessary.
    · If there is a 0 in the pivot position and there is no lower row with a nonzero entry in the current column, then the original matrix is singular.
    · Numerical accuracy can be increased by *always* exchanging for a lower row if by doing so we can increase the magnitude of the pivot.
  (ii) Divide the current row by the pivot to produce 1 in the pivot position.
  (iii) Add multiples of the current row to other rows to produce 0's in the current column.

- Applying this algorithm to the example:
  – Divide row 1 by 2
$$\left[\begin{array}{rrr|rrr} 1 & -1 & 0 & \frac{1}{2} & 0 & 0 \\ 1 & -1 & 1 & 0 & 1 & 0 \\ 4 & 4 & -4 & 0 & 0 & 1 \end{array}\right]$$
  – Subtract row 1 from row 2
$$\left[\begin{array}{rrr|rrr} 1 & -1 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & -\frac{1}{2} & 1 & 0 \\ 4 & 4 & -4 & 0 & 0 & 1 \end{array}\right]$$
  – Subtract 4 $\times$ row 1 from row 3
$$\left[\begin{array}{rrr|rrr} 1 & -1 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & -\frac{1}{2} & 1 & 0 \\ 0 & 8 & -4 & -2 & 0 & 1 \end{array}\right]$$
  – Interchange rows 2 and 3
$$\left[\begin{array}{rrr|rrr} 1 & -1 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 8 & -4 & -2 & 0 & 1 \\ 0 & 0 & 1 & -\frac{1}{2} & 1 & 0 \end{array}\right]$$

– Divide row 2 by 8

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

– Add row 2 to row 1

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

– Add $\frac{1}{2}\times$ row 3 to each of rows 1 and 2

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

- Thus, the inverse of $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{8} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{8} \\ -\frac{1}{2} & 1 & 0 \end{bmatrix}$$

which can be verified by multiplication.

- Why Gaussian elimination works:
  - Each elementary row operation can be represented as multiplication on the left by an ERO matrix.
    - For example, to interchange rows 2 and 3 in the preceding example, we multiply by

$$\mathbf{E}_{III} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

      (Satisfy yourself that this works as advertised!)
    - The entire sequence of EROs can be represented as

$$\begin{aligned} \mathbf{E}_p \cdots \mathbf{E}_2 \mathbf{E}_1 [\mathbf{A}, \mathbf{I}_n] &= [\mathbf{I}_n, \mathbf{B}] \\ \mathbf{E} [\mathbf{A}, \mathbf{I}_n] &= [\mathbf{I}_n, \mathbf{B}] \end{aligned}$$

      where $\mathbf{E} = \mathbf{E}_p \cdots \mathbf{E}_2 \mathbf{E}_1$.
    - Therefore, $\mathbf{E}\mathbf{A} = \mathbf{I}_n$, implying that $\mathbf{E}$ is $\mathbf{A}^{-1}$; and $\mathbf{E}\mathbf{I}_n = \mathbf{B}$, implying that $\mathbf{B} = \mathbf{E} = \mathbf{A}^{-1}$.

4. Determinants

- Each square matrix $\mathbf{A}$ is associated with a scalar called its determinant, and written $|\mathbf{A}|$ or $\det \mathbf{A}$.
- The determinant is uniquely defined by the following axioms (rules):
  (a) Multiplying a row of $\mathbf{A}$ by a scalar constant multiples the determinant by the same constant.

  (b) Adding a multiple of one row of $\mathbf{A}$ to another does not change the determinant.

  (c) Interchanging two rows of $\mathbf{A}$ changes the sign of the determinant.

  (d) $\det \mathbf{I} = 1$.
- These rules suggest that Gaussian elimination can be used to find the determinant of a square matrix:
  - The determinant is the product of the pivots, reversing the sign of this product if an odd number of row interchanges was performed.

- For the example:

$$\det \mathbf{A} = -(2)(8)(1) = -16$$

▶ As well, since we encounter a 0 pivot when we try to invert a singular matrix, the determinant of a singular matrix is 0 (and a nonsingular matrix has a nonzero determinant).

5. Matrix Rank and the Solution of Linear Simultaneous Equations

▶ As explained, the matrix inverse suffices for the solution of linear simultaneous equations when there are equal numbers of equations and unknowns, and when the coefficient matrix is nonsingular.
- This case covers most, but not all, statistical applications.

▶ In the more general case, we have m linear equations in n unknowns:

$$\underset{(m\times n)}{\mathbf{A}}\underset{(n\times 1)}{\mathbf{x}} = \underset{(m\times 1)}{\mathbf{b}}$$

▶ We can solve such a system of equations by Gaussian elimination, placing the coefficient matrix $\mathbf{A}$ in *reduced row-echelon form* (*RREF*) by a sequence of EROs, and bringing the right-hand-side vector $\mathbf{b}$ along for the ride.
- Once the coefficient matrix is in RREF, the solution(s) of the equation system will be apparent.

- A matrix is in RREF form when
  (i) All of its zero rows (if any) follow its nonzero rows (if any).
  (ii) The *leading entry* in each nonzero row — i.e., the first nonzero entry, proceeding from left to right — is 1.
  (iii) The leading entry in each nonzero row after the first is to the right of the leading entry in the previous row.
  (iv) All other entries in a column containing a leading entry are 0.

▶ The proof that this procedure works proceeds from the observation that none of the elementary row operations changes the solutions of the set of equations, and is similar to the proof that Gaussian elimination suffices to find the inverse of a nonsingular square matrix.

▶ Consider the following system of 3 equations in 4 unknowns:

$$\begin{bmatrix} -2 & 0 & -1 & 2 \\ 4 & 0 & 1 & 0 \\ 6 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

- Adjoin the RHS vector to the coefficient matrix

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

- Reduce the coefficient matrix to row-echelon form:
  – Divide row 1 by -2

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

– Subtract 4 $\times$ row 1 from row 2, and subtract 6 $\times$ row 1 from row 3

$$\left[\begin{array}{cccc|c} 1 & 0 & \frac{1}{2} & -1 & -\frac{1}{2} \\ 0 & 0 & -1 & 4 & 4 \\ 0 & 0 & -2 & 8 & 8 \end{array}\right]$$

– Multiply row 2 by -1

$$\left[\begin{array}{cccc|c} 1 & 0 & \frac{1}{2} & -1 & -\frac{1}{2} \\ 0 & 0 & 1 & -4 & -4 \\ 0 & 0 & -2 & 8 & 8 \end{array}\right]$$

– Subtract $\frac{1}{2}\times$ row 2 from row 1, and add 2 $\times$ row 2 to row 3

$$\left[\begin{array}{cccc|c} {}^*1 & 0 & 0 & 1 & \frac{3}{2} \\ 0 & 0 & {}^*1 & -4 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

(with the leading entries marked by asterisks).

- Writing the result as a scalar system of equations, we get

$$\begin{aligned} x_1 + x_4 &= \frac{3}{2} \\ x_3 - 4x_4 &= -4 \\ 0 &= 0 \end{aligned}$$

– The third equation is uninformative, but it does indicate that the original system of equations is *consistent*.

– The first two equations imply that the unknowns x_2 and x_4 can be given arbitrary values (say x_2^* and x_4^*), and the values of the x_1 and x_3 (corresponding to the leading entries) will then follow:

$$\begin{aligned} x_1 &= \frac{3}{2} - x_4^* \\ x_3 &= -4 + 4x_4^* \end{aligned}$$

and thus any vector

$$\mathbf{x} = [x_1, x_2, x_{3,} x_4]' = \left[\frac{3}{2} - x_4^*, x_2^*, -4 + 4x_4^*, x_4^*\right]'$$

is a solution of the system of equations.

– A system for which there is more than one solution (in this case, an infinity of solutions) is called *underdetermined*.

▶ Now consider the system of equations

$$\begin{bmatrix} -2 & 0 & -1 & 2 \\ 4 & 0 & 1 & 0 \\ 6 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

- Attaching $\mathbf{b}$ to $\mathbf{A}$ and reducing the coefficient matrix to RREF yields

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

- The last equation,

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = 2$$

is contradictory, implying that the original system of equations has no solution.

– Such an *inconsistent* system of equations is termed *overdetermined*.

▶ The third possibility is that the system of equations is consistent and has a *unique solution*. This occurs, for example, when there are equal numbers of equations and unknowns and when the (square) coefficient matrix is nonsingular.

▶ The *rank* of a matrix is its maximum number of *linearly independent* rows or columns.

- A set of rows (or columns) is linearly independent when no row (or column) in the set can be expressed as a linear combination (weighted sum) of the others.
- It turns out that the maximum number of linearly independent rows in a matrix is the same as the maximum number of linearly independent columns.
- Because Gaussian elimination reduces *linearly dependent* rows to $\mathbf{0}$, the rank of a matrix is the number of nonzero rows in its RREF.
  – Thus, in the examples, the matrix $\mathbf{A}$ has rank 2.

- To restate our previous results:
  - When the rank of the coefficient matrix is equal to the number of unknowns, and the system of equations is consistent, there is a unique solution.
  - When the rank of the coefficient matrix is less than the number of unknowns, the system of equations is underdetermined (if consistent) or overdetermined (if inconsistent).

5.1 Homogeneous Systems of Linear Equations

- ▶ When the RHS vector in an equation system is a zero vector, the system of equations is said to be *homogeneous*:

$$\underset{(m\times n)}{\mathbf{A}}\underset{(n\times 1)}{\mathbf{x}} = \underset{(m\times 1)}{\mathbf{0}}$$

- ▶ Homogeneous equations cannot be overdetermined, because the *trivial solution* $\mathbf{x} = \mathbf{0}$ always satisfies the system.
- ▶ When the rank of the coefficient matrix $\mathbf{A}$ is less than the number of unknowns n, a homogeneous system of equations is underdetermined, and consequently has *nontrivial solutions* as well.
- ▶ Consider, for example, the homogeneous system

$$\begin{bmatrix} -2 & 0 & -1 & 2 \\ 4 & 0 & 1 & 0 \\ 6 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- Reducing the coefficient matrix to RREF, we have

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- Thus, the solutions (trivial and nontrivial) may be written in the form

$$\begin{aligned} x_1 &= -x_4^* \\ x_2 &= x_2^* \\ x_3 &= 4x_4^* \\ x_4 &= x_4^* \end{aligned}$$

- That is, x_2 and x_4 can be given arbitrary values, and the values of the other two unknowns follow (from the value assigned to x_4).

6. Eigenvalues and Eigenvectors

▶ Suppose that $\mathbf{A}$ is an order-n square matrix. Then the homogeneous system of equations

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

will have nontrivial solutions only for certain values of the scalar L.

- There will be nontrivial solutions only when the matrix $(\mathbf{A} - L\mathbf{I}_n)$ is singular, that is when

$$\det(\mathbf{A} - L\mathbf{I}_n) = 0$$

- This determinantal equation is called the *characteristic equation* of the matrix $\mathbf{A}$.
- Values of L for which the characteristic equation holds are called *eigenvalues*, *characteristic roots*, or *latent roots* of $\mathbf{A}$. (The German word *“eigen”* means “own” — i.e., the matrix’s own values.)

- Suppose that L_1 is a particular eigenvalue of $\mathbf{A}$. Then a vector $\mathbf{x} = \mathbf{x}_1$ satisfying the homogeneous system of equations

$$(\mathbf{A} - L_1\mathbf{I}_n)\mathbf{x} = \mathbf{0}$$

  is called an eigenvector of $\mathbf{A}$ associated with the eigenvalue L_1.
  – Eigenvectors associated with a particular eigenvalue are never unique, because if $\mathbf{x}_1$ is an eigenvector of $\mathbf{A}$, then so is $c\mathbf{x}_1$ (where c is any nonzero scalar constant).

▶ Because of its simplicity, let's examine the 2×2 case.
- The characteristic equation is

$$\begin{vmatrix} a_{11} - L & a_{12} \\ a_{21} & a_{22} - L \end{vmatrix} = 0$$

- I'll make use of the simple result that the determinant of a 2×2 matrix is the product of the diagonal entries minus the product of the off-diagonal entries, so the characteristic equation is

$$\begin{aligned} (a_{11} - L)(a_{22} - L) - a_{12}a_{21} &= 0 \\ L^2 - (a_{11} + a_{22})L + a_{11}a_{22} - a_{12}a_{21} &= 0 \end{aligned}$$

- This is a quadratic equation, and therefore it can be solved by the quadratic formula, producing the two roots

$$\begin{aligned} L_1 &= \frac{1}{2}\left[a_{11} + a_{22} + \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})}\right] \\ L_2 &= \frac{1}{2}\left[a_{11} + a_{22} - \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})}\right] \end{aligned}$$

  – The roots are real if $(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})$ is nonnegative.
  – Notice that

$$\begin{aligned} L_1 + L_2 &= a_{11} + a_{22} = \text{trace}(\mathbf{A}) \\ L_1 L_2 &= a_{11}a_{22} - a_{12}a_{21} = \det(\mathbf{A}) \end{aligned}$$

  – As well, if $\mathbf{A}$ is singular, and thus $\det(\mathbf{A}) = 0$, then at least one of the eigenvalues must be 0. (Both are 0 only if $\mathbf{A}$ is a $\mathbf{0}$ matrix.)

- Things become simpler when the matrix $\mathbf{A}$ is symmetric (as in most statistical applications of eigenvalues)
  – Then $a_{12} = a_{21}$, and

$$\begin{aligned} L_1 &= \frac{1}{2}\left[a_{11} + a_{22} + \sqrt{(a_{11} - a_{22})^2 + 4a_{12}^2}\right] \\ L_2 &= \frac{1}{2}\left[a_{11} + a_{22} - \sqrt{(a_{11} - a_{22})^2 + 4a_{12}^2}\right] \end{aligned}$$

  – In this case, the eigenvalues are necessarily real numbers, since the quantity within the square-root is a sum of squares.

- Example:

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

$$L_1 = \frac{1}{2}\left[1 + 1 + \sqrt{(1-1)^2 + 4(0.5^2)}\right] = 1.5$$

$$L_2 = \frac{1}{2}\left[1 + 1 - \sqrt{(1-1)^2 + 4(0.5^2)}\right] = 0.5$$

– To find the eigenvectors associated with $L_1 = 1.5$, solve

$$\begin{bmatrix} 1-1.5 & 0.5 \\ 0.5 & 1-1.5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which produces

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = \begin{bmatrix} x_{21}^* \\ x_{21}^* \end{bmatrix}$$

where x_{21}^* can be given an arbitrary nonzero value.

– Likewise, for $L_2 = 0.5$,

$$\begin{bmatrix} 1-0.5 & 0.5 \\ 0.5 & 1-0.5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which produces

$$\mathbf{x}_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} = \begin{bmatrix} -x_{22}^* \\ x_{22}^* \end{bmatrix}$$

where x_{22}^* is arbitrary.

– The two eigenvectors of $\mathbf{A}$ have an inner product of 0:

$$\begin{aligned} \mathbf{x}_1 \cdot \mathbf{x}_2 &= \begin{bmatrix} x_{21}^* \\ x_{21}^* \end{bmatrix} \cdot \begin{bmatrix} -x_{22}^* \\ x_{22}^* \end{bmatrix} \\ &= x_{21}^*(-x_{22}^*) + x_{21}^* x_{22}^* \\ &= 0 \end{aligned}$$

Vectors whose inner product is 0 are termed *orthogonal*.

- ▶ The properties of the 2×2 case generalize as follows:
  - The characteristic equation $\det(\mathbf{A} - L\mathbf{I}_n) = 0$ of an order-n square matrix $\mathbf{A}$ is an nth-order polynomial equation in L.
    - Consequently, there are in general n eigenvalues, though not all are necessarily distinct.
    - There are better ways to find eigenvalues and eigenvectors in the general case than to solve the characteristic equation.
  - The sum of the eigenvalues is the trace,

$$\sum_{i=1}^{n} L_i = \text{trace}(\mathbf{A})$$

  - The product of the eigenvalues is the determinant,

$$\prod_{i=1}^{n} L_i = \det(\mathbf{A})$$

  - The number of nonzero eigenvalues is the rank of $\mathbf{A}$.
  - If $\mathbf{A}$ is symmetric, then all of the eigenvalues are real numbers.

  - If all of the eigenvalues are distinct, then each eigenvector is determined up to a constant factor.
  - Eigenvectors associated with different eigenvalues are linearly independent and, in a symmetric matrix, are orthogonal.