Essential Mathematics for the Political and Social Research

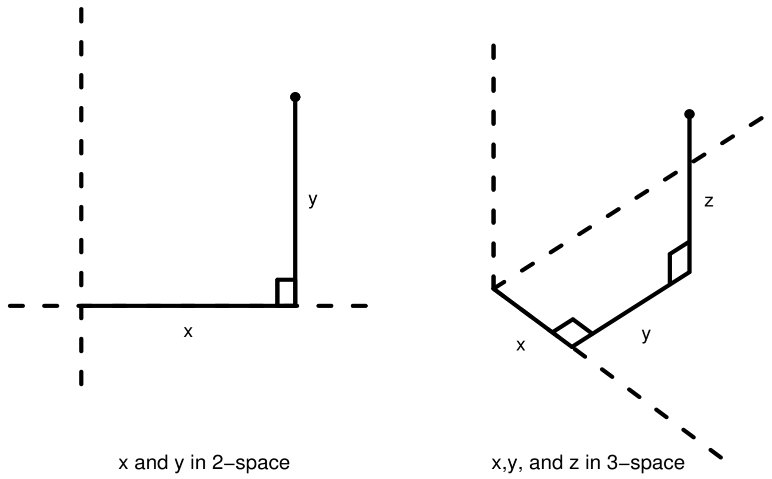
JEFF GILL
Cambridge University Press

Lecture Slides, Chapter 4:
Linear Algebra Continued: Matrix Structure

Space and Time

- The figures shows that vectors are lines that the ordered pair or ordered triple defining a "path" in the associated space that uniquely arrives at a single point.
- This is the intuition behind the basic vector norm we already discussed.

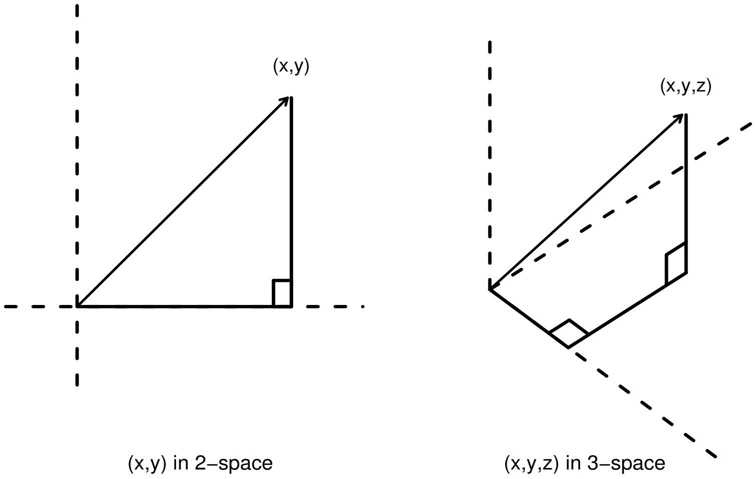
Figure 1: Visualizing Space



Space and Time

- To get the length of the vectors, calculate $\sqrt{x^2 + y^2}$ in the first panel and $\sqrt{x^2 + y^2 + z^2}$ in the second panel, according to the Pythagorean Theorem.

Figure 2: Visualizing Vectors in Spaces



Projections

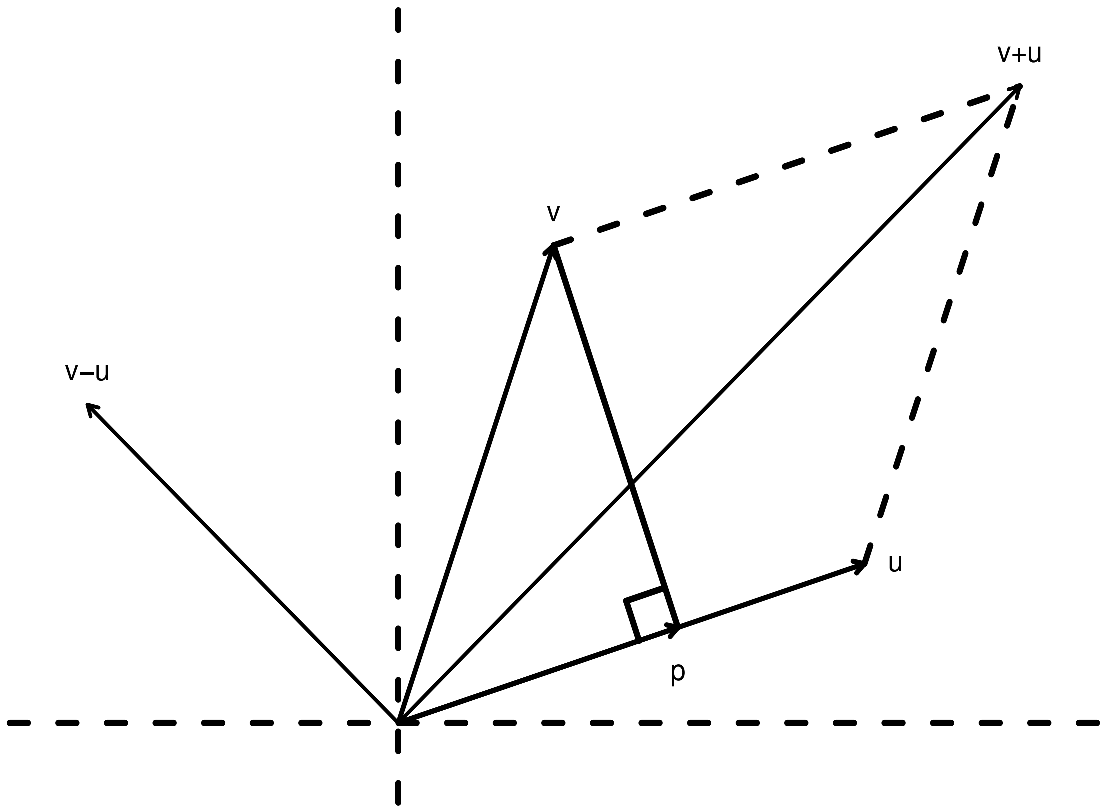
- The two vectors in the previous two figures take up an amount of "space" in the sense that they define a triangular planar region bounded by the vector itself and its two (left panel) or three (right panel) *projections* against the axes
- The angle on the axis from this projection is necessarily a right angle (these are sometimes called *orthogonal projections*).
- Projections define how far along that axis the vector travels in total.
- A projection does not have be just along the axes: project a vector $\mathbf{v}$ against another vector $\mathbf{u}$:

$$p = \text{ projection of } \mathbf{v} \text{ on to } \mathbf{u} \; = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|}\right)\left(\frac{\mathbf{u}}{\|\mathbf{u}\|}\right).$$

- We can think of the second fraction on the right-hand side above as the unit vector in the direction of $\mathbf{u}$.
- The first fraction is a scalar multiplier giving length.
- Since the right angle is preserved, we can rotate this arrangement until $\mathbf{v}$ is lying on the x-axis, as before.

Orthogonal Vectors Produce Zero-Length Projections

Figure 3: Vector Projection, Addition, and Subtraction



Continued...

- Also shown in the figure are the vectors that result from $\mathbf{v} + \mathbf{u}$ and $\mathbf{v} - \mathbf{u}$ with angle θ between them.
- The area of the parallelogram defined by the vector $\mathbf{v} + \mathbf{u}$ shown in the figure is equal to the absolute value of the length of the orthogonal vector that results from the cross product: $\mathbf{u} \times \mathbf{v}$.
- This is related to the projection in the following manner:
  - Call h the length of the line defining the projection in the figure (going from the point p to the point v).
  - The parallelogram has size that is height times length: $h\|\mathbf{u}\|$ from basic geometry.
  - Because the triangle created by the projection is a right triangle, from basic the trigonometry rules $h = \|\mathbf{v}\| \sin\theta$, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.
  - Substituting we get $\mathbf{u} \times \mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\| \sin\theta$.
  - Therefore the size of the parallelogram is $|\mathbf{v} + \mathbf{u}|$ since the order of the cross product could make this negative.

Collinearity

- Sometimes one vector is simply a multiple of another, say $(2, 4)$ and $(4, 8)$.
- The corresponding Cartesian lines are then called *collinear* and the idea of a projection does not make sense.
- The plot of these vectors would be along the exact same line originating at zero, and we are thus adding no new geometric information.
- Therefore the vectors still consume the same "space" even though one is longer.

Higher Dimensions

- Consider matrices as collections of vectors rather than as purely rectangular structures.
- The *column space* of an $i \times j$ matrix $\mathbf{X}$ consists of every possible linear combination of the j columns in $\mathbf{X}$.
- The *row space* of the same matrix consists of every possible linear combination of the i rows in $\mathbf{X}$.
- More formally for the $i \times j$ matrix $\mathbf{X}$:

  Column Space:

  all column vectors $\mathbf{x}_{.1}, \mathbf{x}_{.2}, \ldots, \mathbf{x}_{.j}$,
  and scalars $s_1, s_2, \ldots, s_j$
  producing vectors $s_1\mathbf{x}_{.1} + s_2\mathbf{x}_{.2} + \cdots + s_j\mathbf{x}_{.j}$

- And: Row Space:

  all row vectors $\mathbf{x}_{1.}, \mathbf{x}_{2.}, \ldots, \mathbf{x}_{i.}$,
  and scalars $s_1, s_2, \ldots, s_i$
  producing vectors $s_1\mathbf{x}_{1.} + s_2\mathbf{x}_{2.} + \cdots + s_i\mathbf{x}_{i.}$,

- Here $\mathbf{x}_{.k}$ denotes the kth column vector of $\mathbf{x}$ and $\mathbf{x}_{k.}$ denotes the kth row vector of $\mathbf{x}$.

Higher Dimensions

- The column space here consists of i-dimensional vectors and the row space consists of j-dimensional vectors.
- Note that this is strictly a *linear function* as discussed before (eg. why this is "linear algebra").
- The column space of the matrix $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$ includes (but is not limited to) the following resulting vectors:

$$3\begin{bmatrix} 1 \\ 3 \end{bmatrix} + 1\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 13 \end{bmatrix}, \qquad 5\begin{bmatrix} 1 \\ 3 \end{bmatrix} + 0\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 15 \end{bmatrix}.$$

Example: Linear Transformation of Voter Assessments

- In survey data analysis respondents often answer ordered questions based on their own interpretation of the scale.
- This means that an answer of "strongly agree" may have different meanings across a survey because individuals anchor against different response points, or they interpret the spacing between categories differently.
- Aldrich and McKelvey (1977) apply a linear transformation to data on the placement of presidents on a spatial issue dimension.
- The key to their thinking was that while respondent i places candidate j at X_{ij} on an ordinal scale from the survey instrument, such as a 7-point "dove" to "hawk" measure, their real view was Y_{ij} along some smoother underlying metric with finer distinctions.

Example: Linear Transformation of Voter Assessments

- Hypothethically:

Placement of Candidate Position on the Vietnam War, 1968

	Dove	1	2	3	4	5	6	7	Hawk
Voter 1		H,J,N			W			V	
Voter 2		H	J		N,V			W	
Voter 2		V		H	J,N			W	
Y		$\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|\|$							

H=Humphrey, J=Johnson, N=Nixon, W=Wallace, V=Voter

- The graphic for Y is done to suggest a noncategorical measure such as along $\mathfrak{R}$.

Example: Linear Transformation of Voter Assessments

- To obtain a picture of this latent variable, Aldrich and McKelvey suggested a linear transformation for each voter to relate observed categorical scale to this underlying metric: $c_i + \omega_i X_{ij}$.
- The perceived candidate positions for voter i are given by

$$Y_i = \begin{bmatrix} c_i + \omega_i X_{i1} \\ c_i + \omega_i X_{i2} \\ \vdots \\ c_i + \omega_i X_{iJ} \end{bmatrix},$$

- This gives a better vector of estimates for the placement of all J candidates by respondent i because it accounts for individual-level "anchoring" by each respondent, c_i.
- Aldrich and McKelvey then estimated each of the values of c and ω. The value of this linear transformation is that it allows the researchers to see beyond the limitations of the categorical survey data.

Span and Basis

- Let $\mathbf{x}_{.1}, \mathbf{x}_{.2}, \ldots, \mathbf{x}_{.j}$ be a set of column vectors in $\Re^i$ (i.e., they are all length i).
- The collection of possible linear combinations of these vectors is the *span* of that set.
- Therefore this collection is the span of the matrix if you assemble these as column vectors to form the matrix..
- Any additional vector in $\Re^i$ is spanned by these vectors if and only if it can be expressed as a linear combination of $\mathbf{x}_{.1}, \mathbf{x}_{.2}, \ldots, \mathbf{x}_{.j}$, $j \geq i$.
- For $j = i$ for any set of linearly independent vectors that you can construct spanning $\mathbf{X}$ is called a *basis* of $\mathbf{X}$.

Linear Space

- A *linear space*, $\mathfrak{X}$, is the nonempty set of matrices such that remain *closed* under linear transformation:
  - If $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are in $\mathfrak{X}$,
  - and $s_1, s_2, \ldots, s_n$ are any scalars,
  - then $\mathbf{X}_{n+1} = s_1\mathbf{X}_1 + s_2\mathbf{X}_2 + \cdots + s_n\mathbf{X}_n$ is in $\mathfrak{X}$.
- Meaning: linear combinations of matrices in the linear space have to remain in this linear space.
- Every linear space contains an infinite number of matrices (except for null spaces).
- Also *linear subspaces* represent some enclosed region of the full space.

Span for Matrices

- The span of a finite set of matrices is the set of all matrices that can be achieved by a linear combination of the original matrices.
- This is confusing because a span is also a linear space.
- It is useful is in determining a minimal set of matrices that span a given linear space.
- The finite set of *linearly independent* matrices in a given linear space that span the linear space is called a basis for this linear space
- Note the word "a" here since it is not unique: it cannot be made a smaller set because it would lose the ability to produce parts of the linear space, and it cannot be made a larger set because it would then no longer be linearly independent.

Span for Matrices, Example

- A 3×3 identity matrix is clearly a basis for $\Re^3$ (the three-dimensional space of real numbers) because any three-dimensional coordinate, $[r_1, r_2, r_3]$ can be produced by multiplication of $\mathbf{I}$ by three chosen scalars.
- Yet, the matrices defined by

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

do not qualify as a basis (although the second still *spans* $\Re^3$).

Trace of a Matrix

- The trace of a square matrix is the sum of the diagonal values $\text{tr}(\mathbf{X}) = \sum_{i=1}^{k} x_{ii}$ and is denoted $\text{tr}(\mathbf{X})$ for the trace of square matrix $\mathbf{X}$.
- An $i \times j$ matrix $\mathbf{X}$ is a zero matrix iff $\text{tr}(A'A) = 0)$.
- For example:

$$\text{tr}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 1 + 4 = 5 \qquad \text{tr}\begin{pmatrix} 12 & \frac{1}{2} \\ 9 & \frac{1}{3} \end{pmatrix} = 12 + \frac{1}{3} = \frac{37}{3}.$$

- One property of the trace has implications in statistics: $\text{tr}(\mathbf{X}'\mathbf{X})$ is the sum of the square of every value in the matrix $\mathbf{X}$.
- For example:

$$\text{tr}\left(\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}' \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}\right) = \text{tr}\begin{pmatrix} 2 & 5 \\ 5 & 13 \end{pmatrix} = 15 = 1 + 1 + 4 + 9.$$

Properities of the Trace of a Matrix

Properties of (Conformable) Matrix Trace Operations

$\rightarrowtail$ Identity Matrix	$\text{tr}(\mathbf{I}_n) = n$
$\rightarrowtail$ Zero Matrix	$\text{tr}(\mathbf{0}) = 0$
$\rightarrowtail$ Square **J** Matrix	$\text{tr}(\mathbf{J}_n) = n$
$\rightarrowtail$ Scalar Multiplication	$\text{tr}(s\mathbf{X}) = s\text{tr}(\mathbf{X})$
$\rightarrowtail$ Matrix Addition	$\text{tr}(\mathbf{X} + \mathbf{Y}) = \text{tr}(\mathbf{X}) + \text{tr}(\mathbf{Y})$
$\rightarrowtail$ Matrix Multiplication	$\text{tr}(\mathbf{XY}) = \text{tr}(\mathbf{YX})$
$\rightarrowtail$ Transposition	$\text{tr}(\mathbf{X}') = \text{tr}(\mathbf{X})$

Determinant of a Matrix

- The *determinant* uses all of the values of a square matrix to provide a summary of structure, not just the diagonal like the trace.
- For 2×2 matrices, this is the difference in diagonal products:

$$\det(\mathbf{X}) = |\mathbf{X}| = \begin{vmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{vmatrix} = x_{11}x_{22} - x_{12}x_{21}.$$

- The notation for a determinant is expressed as $\det(\mathbf{X})$ or $|\mathbf{X}|$.
- For example:

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = (1)(4) - (2)(3) = -2$$

$$\begin{vmatrix} 10 & \frac{1}{2} \\ 4 & 1 \end{vmatrix} = (10)(1) - \left(\frac{1}{2}\right)(4) = 8$$

$$\begin{vmatrix} 2 & 3 \\ 6 & 9 \end{vmatrix} = (2)(9) - (3)(6) = 0.$$

Calculating Determinants

- Calculating determinants gets much more involved with square matrices larger than 2×2.
- First define a *submatrix* by deleting rows and/or columns of a matrix, leaving the remaining elements in their respective places.
- For the matrix $\mathbf{X}$, notice the following submatrices whose deleted rows and columns are denoted by subscripting:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{bmatrix},$$

$$\mathbf{X}_{[11]} = \begin{bmatrix} x_{22} & x_{23} & x_{24} \\ x_{32} & x_{33} & x_{34} \\ x_{42} & x_{43} & x_{44} \end{bmatrix}, \quad \mathbf{X}_{[24]} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \end{bmatrix}.$$

Calculating Determinants

- For an $n \times n$ matrix $\mathbf{X}$ define the following:
  - The ijth *minor* of $\mathbf{X}$ for x_{ij}, $|\mathbf{X}_{[ij]}|$ is the determinant of the $(n-1) \times (n-1)$ submatrix that results from taking the ith row and jth column out.
  - The *cofactor* of $\mathbf{X}$ for x_{ij} is the minor signed in this way: $(-1)^{i+j}|\mathbf{X}_{[ij]}|$.
  - To exhaust the entire matrix we cycle recursively through the columns and take sums with a formula that multiplies the cofactor by the determining value:

$$\det(\mathbf{X}) = \sum_{j=1}^{n} (-1)^{i+j} x_{ij} |\mathbf{X}_{[ij]}|$$

for some constant i.

Calculating Determinants

- Here *recursive* means that the algorithm is applied iteratively through progressively smaller submatrices $\mathbf{X}_{[ij]}$.
- This means that we lop off the top row and multiply the values across the resultant submatrices without the associated column.
- We can pick any row or column to perform this operation, because the results will be equivalent.
- For a 3×3 matrix:

$$\begin{vmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{vmatrix}$$

$$= (+1)x_{11} \begin{vmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{vmatrix} + (-1)x_{12} \begin{vmatrix} x_{21} & x_{23} \\ x_{31} & x_{33} \end{vmatrix} + (+1)x_{13} \begin{vmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{vmatrix}.$$

- The problem is easy because the subsequent three determinant calculations are on 2×2 matrices.

Calculating Determinants

- Using this more general process means that one has to be more careful about the alternating signs in the sum since picking the row or column to "pivot" on determines the order.
- For instance, here are the signs for a 7×7 matrix produced from the sign on the cofactor:

$$\begin{bmatrix} + & - & + & - & + & - & + \\ - & + & - & + & - & + & - \\ + & - & + & - & + & - & + \\ - & + & - & + & - & + & - \\ + & - & + & - & + & - & + \\ - & + & - & + & - & + & - \\ + & - & + & - & + & - & + \end{bmatrix}.$$

Example: Structural Shortcuts for Calculating Determinants

- Ishizawa (1991), in looking at the return to scale of public inputs and its effect on the transformation curve of an economy, needed to solve a system of equations by taking the determinant of the matrix

$$\begin{bmatrix} \ell^1 & k^1 & 0 & 0 \\ \ell^2 & k^2 & 0 & 0 \\ L_w^D & L_r^D & \ell^1 & \ell^2 \\ K_w^D & K_r^D & k^1 & k^2 \end{bmatrix}, \qquad \text{where these are all abbreviations for longer vectors or complex terms.}$$

- We can start by being very mechanical about this:

$$\det = \ell^1 \begin{bmatrix} k^2 & 0 & 0 \\ L_r^D & \ell^1 & \ell^2 \\ K_r^D & k^1 & k^2 \end{bmatrix} - k^1 \begin{bmatrix} \ell^2 & 0 & 0 \\ L_w^D & \ell^1 & \ell^2 \\ K_w^D & k^1 & k^2 \end{bmatrix}.$$

- The big help here was the two zeros on the top row that meant that we could stop our 4×4 calculations after two steps, and this trick works again because we have the same structure remaining in the 3×3 case.

- Define the 2×2 lower right matrix as $\mathbf{D} = \left[\begin{smallmatrix} \ell^1 & \ell^2 \\ k^1 & k^2 \end{smallmatrix}\right]$, so that we get the simplification:

$$\det = \ell^1 k^2 |\mathbf{D}| - k^1 \ell^2 |\mathbf{D}| = (\ell^1 k^2 - k^1 \ell^2)|\mathbf{D}| = |\mathbf{D}|^2.$$

Connections

- Kronecker products on square matrices have interesting properties.
- The trace:
$$\mathrm{tr}(\mathbf{X} \otimes \mathbf{Y}) = \mathrm{tr}(\mathbf{X})\mathrm{tr}(\mathbf{Y}).$$
- The determinant:
$$|\mathbf{X} \otimes \mathbf{Y}| = |\mathbf{X}|^{\ell}|\mathbf{Y}|^{j}$$
for the $j \times j$ matrix $\mathbf{X}$ and the $\ell \times \ell$ matrix $\mathbf{Y}$,
- Note the switching of exponents in the RHS of the expression for the determinant.

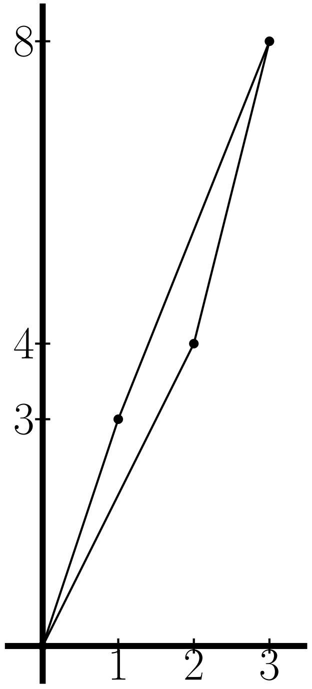
Properties of $(n \times n)$ Matrix Determinants

$\rightarrowtail$ Diagonal Matrix	$\vert\mathbf{D}\vert = \prod_{i=1}^{n} \mathbf{D}_{ii}$
$\rightarrowtail$ (Therefore) Identity Matrix	$\vert\mathbf{I}\vert = 1$
$\rightarrowtail$ Triangular Matrix (upper or lower)	$\vert\boldsymbol{\theta}\vert = \prod_{i=1}^{n} \boldsymbol{\theta}_{ii}$
$\rightarrowtail$ Scalar Times Diagonal	$\vert s\mathbf{D}\vert = s^n\vert\mathbf{D}\vert$
$\rightarrowtail$ Transpose Property	$\vert\mathbf{X}\vert = \vert\mathbf{X}'\vert$
$\rightarrowtail$ $\mathbf{J}$ Matrix	$\vert\mathbf{J}\vert = 0$

Visualization of the Determinant

- If the columns of an $n \times n$ matrix $\mathbf{X}$ are treated as vectors, then the area of the parallelogram created by an n-dimensional space of these vectors is the absolute value of the determinant of $\mathbf{X}$, where the vectors originate at zero and the opposite point of the parallelogram is determined by the product of the columns (a cross product of these vectors).
- View the determinant of the 2×2 matrix $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$.
- This figure indicates that the determinant is somehow a description of the size of a matrix in the geometric sense.

Spatial Representation of a Determinant



Matrix Rank

- An important characteristic of any matrix is its *rank*.
- Rank tells us the "space" in terms of columns or rows that a particular matrix occupies: how much unique information is held in the rows or columns of a matrix.
- For example, a matrix that has three rows but only two rows of unique information is given by

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

- This is also true for the matrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix},$$

because the third row is just two times the second row and therefore has no new relational information to offer.

Matrix Rank

- When any one column of a matrix can be produced by nonzero scalar multiples of other columns added, then we say that the matrix is not *full rank* (sometimes called *short rank*).
- In this case at least one column is *linearly dependent*.
- This means that we can produce the relative relationships defined by this column from the other columns and it thus adds nothing to our understanding of the relationships defined by the matrix.

Linearly Independence

- The collection of vectors determined by the columns is said to be *linearly independent* columns if the only set of scalars, $s_1, s_2, \ldots, s_j$, that satisfies

$$s_1\mathbf{x}_{.1} + s_2\mathbf{x}_{.2} + \cdots + s_j\mathbf{x}_{.j} = \mathbf{0}$$

  is a set of all zero values, $s_1 = s_2 = \ldots = s_j = 0$.

- This is another way of looking at the same idea since such a condition means that we *cannot* reproduce one column vector from a linear combination of the others.

Rows Versus Columns

- This emphasis on columns is somewhat unwarranted because the rank of a matrix is equal to the rank of its transpose.
- Therefore, everything just said about columns can also be said about rows.
- So *the row rank of any matrix is also its column rank.*
- What makes this somewhat confusing is additional terminology.
- An $(i \times j)$ matrix is *full column rank* if its rank equals the number of columns, and it is *full row rank* if its rank equals its number of rows.
- Thus, if $i > j$, then the matrix can be full column rank but never full row rank.
- This does not necessarily mean that it *has* to be full column rank just because there are fewer columns than rows.

Square Matrix Considerations

- A square matrix is full rank if and only if it has a nonzero determinant.
- This is the same thing as saying that a matrix is full rank if it is nonsingular or invertible
- Therefore the determinant is a handy way to calculate whether a matrix is full rank because the linear dependency within can be subtle,

Example: Structural Equation Models

- In their text Hanushek and Jackson (1977, Chapter 9) provided a technical overview of structural equation models where systems of equations are assumed to simultaneously affect each other to reflect endogenous social phenomena.

- Often these models are described in matrix terms, such as their example (p. 265)

$$\mathbf{A} = \begin{bmatrix} \gamma_{24} & 1 & \gamma_{26} & 0 & -1 \\ 0 & -1 & \gamma_{56} & 0 & 0 \\ 0 & \gamma_{65} & -1 & 0 & 0 \\ \beta_{34} & 0 & \beta_{36} & 0 & \beta_{32} \\ \beta_{44} & 0 & \beta_{46} & 0 & \beta_{42} \end{bmatrix}.$$

- Without doing any calculations we can see that this matrix is of rank less than 5 because there is a column of all zeros.

Example: Structural Equation Models

- Matrix determinants are not changed by multiplying the matrix by an identity in advance, multiplying by a permutation matrix in advance, or by taking transformations.
- Therefore we can get a matrix

$$\mathbf{A}^* = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{24} & 1 & \gamma_{26} & 0 & -1 \\ 0 & -1 & \gamma_{56} & 0 & 0 \\ 0 & \gamma_{65} & -1 & 0 & 0 \\ \beta_{34} & 0 & \beta_{36} & 0 & \beta_{32} \\ \beta_{44} & 0 & \beta_{46} & 0 & \beta_{42} \end{bmatrix}' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & \gamma_{65} & 0 & 0 \\ \gamma_{26} & \gamma_{56} & -1 & \beta_{36} & \beta_{46} \\ \gamma_{24} & 0 & 0 & \beta_{34} & \beta_{44} \\ -1 & 0 & 0 & \beta_{32} & \beta_{42} \end{bmatrix}$$

- This is immediately identifiable as having a zero determinant by the general determinant form because each ith minor (the matrix that remains when the ith row and column are removed) is multiplied by the ith value on the first row.

Rank and Idempotent Matrices

- Some rank properties are more specialized. An idempotent matrix has the property that

$$\text{rank}(\mathbf{X}) = \text{tr}(\mathbf{X}),$$

- For any square matrix with the property that $A^2 = sA$, for some scalar s

$$s\text{rank}(\mathbf{X}) = \text{tr}(\mathbf{X}).$$

Properties of Matrix Rank

Properties of Matrix Rank

$\rightarrowtail$ Transpose	$\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}')$
$\rightarrowtail$ Scalar Multiplication (nonzero scalars)	$\text{rank}(s\mathbf{X}) = \text{rank}(\mathbf{X})$
$\rightarrowtail$ Matrix Addition	$\text{rank}(\mathbf{X}+\mathbf{Y}) \leq \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$
$\rightarrowtail$ Consecutive Blocks	$\text{rank}[\mathbf{X}\mathbf{Y}] \leq \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$ $\text{rank}\left[\begin{smallmatrix}\mathbf{X}\\ \mathbf{Y}\end{smallmatrix}\right] \leq \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$
$\rightarrowtail$ Diagonal Blocks	$\text{rank}\left[\begin{smallmatrix}\mathbf{X} & 0\\ 0 & \mathbf{Y}\end{smallmatrix}\right] = \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$
$\rightarrowtail$ Kronecker Product	$\text{rank}(\mathbf{X}\otimes\mathbf{Y}) = \text{rank}(\mathbf{X})\text{rank}(\mathbf{Y})$

Matrix Norms

- Recall that the vectors norm is a measure of length:

$$\|\mathbf{v}\| = (v_1^2 + v_2^2 + \cdots + v_n^2)^{\frac{1}{2}} = (\mathbf{v}'\mathbf{v})^{\frac{1}{2}}.$$

- Matrix norms are a little bit more involved than the vector norms we saw before.

- Euclidean norm or the Frobenius norm:

$$\|\mathbf{X}\|_F = \left[\sum_i \sum_j |x_{ij}|^2\right]^{\frac{1}{2}}$$

  (the square root of the sum of each element squared).

- p-norm:

$$\|\mathbf{X}\|_p = \max_{\|\mathbf{v}\|_p} \|\mathbf{X}\mathbf{v}\|_p,$$

  which is defined with regard to the unit vector $\mathbf{v}$ whose length is equal to the number of columns in $\mathbf{X}$.

Matrix Norms

- For $p = 1$ and an $I \times J$ matrix, this reduces to summing absolute values down columns and taking the maximum:

$$\|\mathbf{X}\|_1 = \max_J \sum_{i=1}^{I} |x_{ij}|.$$

- The infinity version of the matrix p-norm sums across rows before taking the maximum:

$$\|\mathbf{X}\|_\infty = \max_I \sum_{j=1}^{J} |x_{ij}|.$$

- Like the infinity form of the vector norm, this is somewhat unintuitive because there is no apparent use of a limit.

Properties of Matrix Norms

Properties of Matrix Norms, Size $(i \times j)$

$\rightarrowtail$ Constant Multiplication	$\|k\mathbf{X}\| = \|k\|\|\mathbf{X}\|$
$\rightarrowtail$ Addition	$\|\mathbf{X}+\mathbf{Y}\| \leq \|\mathbf{X}\| + \|\mathbf{Y}\|$
$\rightarrowtail$ Vector Multiplication	$\|\mathbf{X}\mathbf{v}\|_p \leq \|\mathbf{X}\|_p \|\mathbf{v}\|_p$
$\rightarrowtail$ Norm Relation	$\|\mathbf{X}\|_2 \leq \|\mathbf{X}\|_F \leq \sqrt{j}\|\mathbf{X}\|_2$
$\rightarrowtail$ Unit Vector Relation	$\mathbf{X}'\mathbf{X}\mathbf{v} = (\|\mathbf{X}\|_2)^2\mathbf{v}$
$\rightarrowtail$ P-norm Relation	$\|\mathbf{X}\|_2 \leq \sqrt{\|\mathbf{X}\|_1 \, \|\mathbf{X}\|_\infty}$
$\rightarrowtail$ Schwarz Inequality	$\|\mathbf{X}\cdot\mathbf{Y}\| \leq \|\mathbf{X}\| \, \|\mathbf{Y}\|$, where $\|\mathbf{X}\cdot\mathbf{Y}\| = \mathrm{tr}(\mathbf{X}'\mathbf{Y})$

Matrix Norm Sum Inequality

- Given matrices

$$\mathbf{X} = \begin{bmatrix} 3 & 2 \\ 5 & 1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{Y} = \begin{bmatrix} -1 & -2 \\ 3 & 4 \end{bmatrix},$$

observe that

$$\begin{array}{l|l} \|\mathbf{X} + \mathbf{Y}\|_\infty & \|\mathbf{X}\|_\infty + \|\mathbf{Y}\|_\infty \\ \left\| \begin{bmatrix} 2 & 0 \\ 8 & 5 \end{bmatrix} \right\|_\infty & \max(5, 6) + \max(3, 7) \\ \max(2, 13) & 13, \end{array}$$

showing the second property above.

Schwarz Inequality for Matrices

- Using the same **X** and **Y** matrices and the $p = 1$ norm, observe that

$$\begin{array}{c|l} |\mathbf{X} \cdot \mathbf{Y}| & \|\mathbf{X}\|_1 \, \|\mathbf{Y}\|_1 \\ (12) + (0) & \max(8, 3) \cdot \max(4, 6) \end{array}$$

showing that the inequality holds: $12 < 48$. This is a neat property because it shows a relationship between the trace and matrix norm.

Matrix Inversion

- Just like scalars have inverses, some *square* matrices have a *matrix inverse*.
- The inverse of a matrix $\mathbf{X}$ is denoted $\mathbf{X}^{-1}$ and defined by the property

$$\mathbf{X}\mathbf{X}^{-1} = \mathbf{X}^{-1}\mathbf{X} = \mathbf{I}.$$

- When a matrix is pre-multiplied or post-multiplied by its inverse the result is an identity matrix of the same size.
- Consider the following matrix and its inverse:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2.0 & 1.0 \\ 1.5 & -0.5 \end{bmatrix} = \begin{bmatrix} -2.0 & 1.0 \\ 1.5 & -0.5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Matrix Inversion

- A singular matrix cannot be inverted, and often "singular" and "noninvertible" are used as synonyms.

- Usually matrix inverses are calculated by computer software because it is quite time-consuming with reasonably large matrices.

- There is a very nice trick for immediately inverting 2×2 matrices, which is given by

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

$$\mathbf{X}^{-1} = \det(\mathbf{X})^{-1} \begin{bmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{bmatrix}.$$

- A matrix inverse is unique: there is only one matrix that meets the multiplicative condition above for a nonsingular square matrix.

Gauss-Jordan elimination

- *Gauss-Jordan elimination* makes use of linear programming to invert the matrix.
- Start with the matrix of interest partitioned next to the identity matrix and allow the following operations:
    - Any row may be multiplied or divided by a scalar.
    - Any two rows may be switched.
    - Any row may be multiplied or divided by a scalar and then added to another row. Note: This operation does not change the original row; its multiple is used but not saved.
- We want to iteratively apply these steps until the identity matrix on the right-hand side is on the left-hand side.

Gauss-Jordan elimination

- Let's perform this process on a 3×3 matrix:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 1 & 8 & 9 & 0 & 0 & 1 \end{array}\right].$$

- Now multiply the first row by -4, adding it to the second row, and multiply the first row by -1, adding it to the third row:

$$\left[\begin{array}{rrr|rrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & 6 & 6 & -1 & 0 & 1 \end{array}\right].$$

Gauss-Jordan elimination

- Multiply the second row by $\frac{1}{2}$, adding it to the first row, and simply add this same row to the third row:

$$\left[\begin{array}{ccc|ccc} 1 & \frac{1}{2} & 0 & -1 & \frac{1}{2} & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & 3 & 0 & -5 & 1 & 1 \end{array}\right].$$

- Multiply the third row by $-\frac{1}{6}$, adding it to the first row, and add the third row (un)multiplied to the second row:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} \\ 0 & 0 & -6 & -9 & 2 & 1 \\ 0 & 3 & 0 & -5 & 1 & 1 \end{array}\right].$$

- Finally, just divide the second row by -6 and the third row by -3, and then switch their places:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} \\ 0 & 1 & 0 & -\frac{5}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{3} & -\frac{1}{6} \end{array}\right],$$

thus completing the operation.

A Singular Case

- We already know that singular matrices cannot be inverted, but consider the described inversion process applied to an obvious case:

$$\left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right].$$

- It is easy to see that there is nothing that can be done to put a nonzero value in the second column of the matrix to the left of the partition.

Properties of Matrix Inversion

Properties of $n \times n$ Nonsingular Matrix Inverse

$\rightarrowtail$ Diagonal Matrix	$\mathbf{D}^{-1}$ has diagonal values $1/d_{ii}$ and zeros elsewhere.
$\rightarrowtail$ (Therefore) Identity Matrix	$\mathbf{I}^{-1} = \mathbf{I}$
$\rightarrowtail$ (Non-zero) Scalar Multiplication	$(s\mathbf{X})^{-1} = \frac{1}{s}\mathbf{X}^{-1}$
$\rightarrowtail$ Iterative Inverse	$(\mathbf{X}^{-1})^{-1} = \mathbf{X}$
$\rightarrowtail$ Exponents	$\mathbf{X}^{-n} = (\mathbf{X}^{n})^{-1}$
$\rightarrowtail$ Multiplicative Property	$(\mathbf{X}\mathbf{Y})^{-1} = \mathbf{Y}^{-1}\mathbf{X}^{-1}$
$\rightarrowtail$ Transpose Property	$(\mathbf{X}')^{-1} = (\mathbf{X}^{-1})'$
$\rightarrowtail$ Orthogonal Property	If $\mathbf{X}$ is orthogonal, then $\mathbf{X}^{-1} = \mathbf{X}'$
$\rightarrowtail$ Determinant	$\lvert\mathbf{X}^{-1}\rvert = 1/\lvert\mathbf{X}\rvert$

Calculating Regression Parameters

- The classic "ordinary least squares" method for obtaining regression parameters proceeds as follows.
- Suppose that $\mathbf{y}$ is the outcome variable of interest and $\mathbf{X}$ is a matrix of explanatory variables with a leading column of 1's.
- What we would like is the vector $\hat{\boldsymbol{\beta}}$ that contains the intercept and the regression slope, which is calculated by the equation $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, which might have seemed hard before this point in the chapter.
- Governments often worry about the economic condition of senior citizens for political and social reasons.
- Typically in a large industrialized society, a substantial portion of these people obtain the bulk of their income from government pensions.
- One important question is whether there is enough support through these payments to provide subsistence above the poverty rate.

Calculating Regression Parameters

- To see if this is a concern, the European Union (EU) looked at this question in 1998 for the (then) 15 member countries with two variables: (1) the median (EU standardized) income of individuals age 65 and older as a percentage of the population age 0–64, and (2) the percentage with income below 60% of the median (EU standardized) income of the national population.

- The data from the European Household Community Panel Survey are

Nation	Relative Income	Poverty Rate
Netherlands	93.00	7.00
Luxembourg	99.00	8.00
Sweden	83.00	8.00
Germany	97.00	11.00
Italy	96.00	14.00
Spain	91.00	16.00
Finland	78.00	17.00
France	90.00	19.00
United.Kingdom	78.00	21.00
Belgium	76.00	22.00
Austria	84.00	24.00
Denmark	68.00	31.00
Portugal	76.00	33.00
Greece	74.00	33.00
Ireland	69.00	34.00

Calculating Regression Parameters

- So the $\mathbf{y}$ vector is the second column of the table and the $\mathbf{X}$ matrix is the first column along with the leading column of 1's added to account for the intercept (also called the constant, which explains the 1's).

- The first quantity that we want to calculate is

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 15.00 & 1252 \\ 1252 & 105982 \end{bmatrix},$$

which has the inverse

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 4.76838 & -0.05633 \\ -0.05633 & 0.00067 \end{bmatrix}.$$

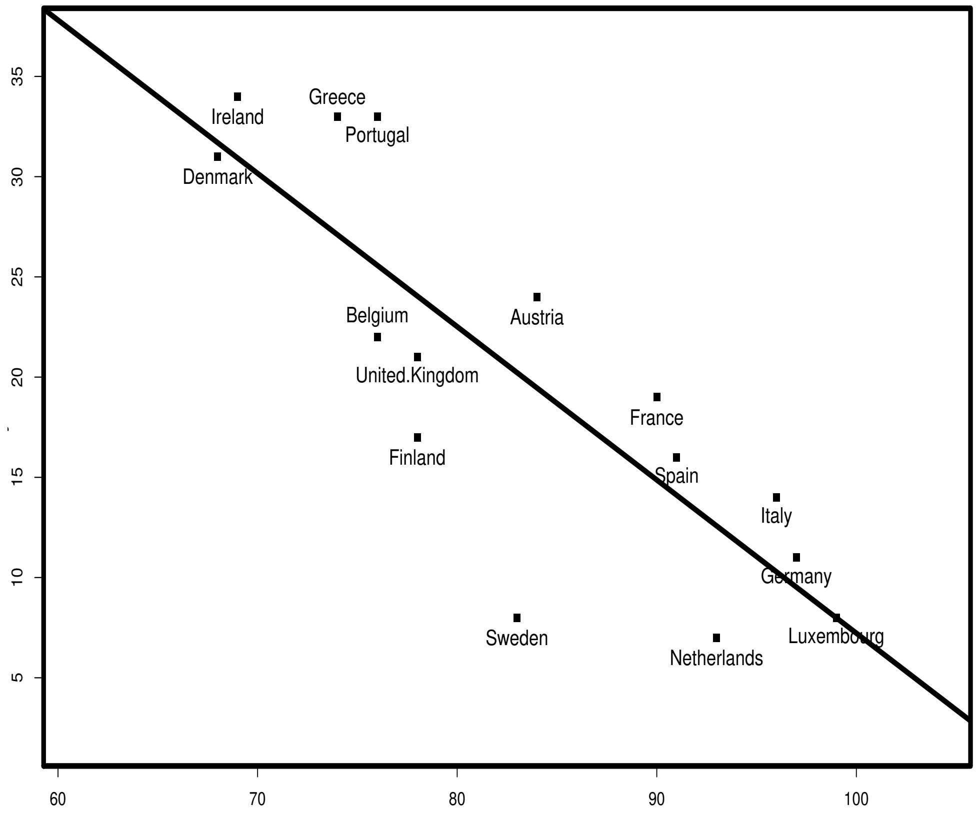
Calculating Regression Parameters

- The final calculation is

$$\begin{bmatrix} 4.76838 & -0.05633 \\ -0.05633 & 0.00067 \end{bmatrix} \begin{bmatrix} 1 & 93 \\ 1 & 99 \\ 1 & 83 \\ 1 & 97 \\ 1 & 96 \\ 1 & 91 \\ 1 & 78 \\ 1 & 90 \\ 1 & 78 \\ 1 & 76 \\ 1 & 84 \\ 1 & 68 \\ 1 & 76 \\ 1 & 74 \\ 1 & 69 \end{bmatrix}' \begin{bmatrix} 7 \\ 8 \\ 8 \\ 11 \\ 14 \\ 16 \\ 17 \\ 19 \\ 21 \\ 22 \\ 24 \\ 31 \\ 33 \\ 33 \\ 34 \end{bmatrix} = \begin{bmatrix} 83.69279 \\ -0.76469 \end{bmatrix}$$

Calculating Regression Parameters

Figure 4: Relative Income and Senior Poverty, EU Countries



Linear Systems of Equations

- A basic and common problem in applied mathematics is the search for a solution, $\mathbf{x}$, to the system of simultaneous linear equations defined by

$$\mathbf{A}\mathbf{x} = \mathbf{y},$$

where $\mathbf{A} \in \Re^{p \times q}$, $\mathbf{x} \in \Re^{q}$, and $\mathbf{y} \in \Re^{p}$.

- If the matrix $\mathbf{A}$ is invertible, then there exists a unique, easy-to-find, solution vector $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ satisfying $\mathbf{A}\mathbf{x} = \mathbf{y}$.
- However, if the system of linear equations in $\mathbf{A}\mathbf{x} = \mathbf{y}$ is not *consistent*, then there exists no solution.
- Consistency simply means that if a linear relationship exists in the rows of $\mathbf{A}$, it must also exist in the corresponding rows of $\mathbf{y}$.

Linear Systems of Equations

- For example, the following simple system of linear equations is consistent:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} \quad \text{because the second row is two times the first across } (\mathbf{x}|\mathbf{y}).$$

- This implies that $\mathbf{y}$ is contained in the linear span of the columns (range) of $\mathbf{A}$, denoted as $\mathbf{y} \in R(\mathbf{A})$.

- Recall that a set of linearly independent vectors (i.e., the columns here) that span a vector subspace is called a basis of that subspace.

- Conversely, the following system of linear equations is not consistent:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \quad \text{because there is no solution for } \mathbf{x} \text{ that satisfies both rows.}$$

- If $\mathbf{A}^{-1}$ exists, then $\mathbf{Ax} = \mathbf{y}$ is always consistent because there exist no linear relationships in the rows of $\mathbf{A}$ that must be satisfied in $\mathbf{y}$.

- The inconsistent case is the more common *statistically* in that a solution that minimizes the squared sum of the inconsistencies is typically applied (ordinary least squares).

Linear Systems of Equations, Singularity

- In addition to the possibilities of the general system of equations $\mathbf{Ax} = \mathbf{y}$ having a unique solution and no solution, this arbitrary system of equations can also have an infinite number of solutions.
- The matrix $\left[\begin{smallmatrix} 1 & 2 \\ 2 & 4 \end{smallmatrix}\right]$ above with $y = [3, 6]'$ is such a case.
- For example, we could solve to obtain $\mathbf{x} = (1, 1)'$, $\mathbf{x} = (-1, 2)'$, $\mathbf{x} = (5, -1)'$, and so on.
- This occurs when the $\mathbf{A}$ matrix is singular: $\text{rank}(\mathbf{A}) = \text{dimension}(R(\mathbf{A})) < q$.
- When the $\mathbf{A}$ matrix is singular at least one column vector is a linear combination of the others, and the matrix therefore contains redundant information.
- In other words, there are $q' < q$ independent column vectors in $\mathbf{A}$.

Solving Systems of Equations by Inversion

- Consider the system of equations

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

where $\mathbf{x} = [x_1, x_2]$, $\mathbf{y} = [4, 3]'$, and $\mathbf{A} = \left[\begin{smallmatrix} 2 & -3 \\ 5 & 5 \end{smallmatrix}\right]$. First invert $\mathbf{A}$:

$$\mathbf{A}^{-1} = \begin{bmatrix} 2 & -3 \\ 5 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 0.2 & 0.12 \\ -0.2 & 0.08 \end{bmatrix}.$$

- Then, to solve for $\mathbf{x}$ we simply need to multiply this inverse by $\mathbf{y}$:

$$\mathbf{A}^{-1}\mathbf{y} = \begin{bmatrix} 0.2 & 0.12 \\ -0.2 & 0.08 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 1.16 \\ -0.56 \end{bmatrix},$$

meaning that $\mathbf{x}_1 = 1.16$ and $\mathbf{x}_2 = -0.56$.

Motivating Example

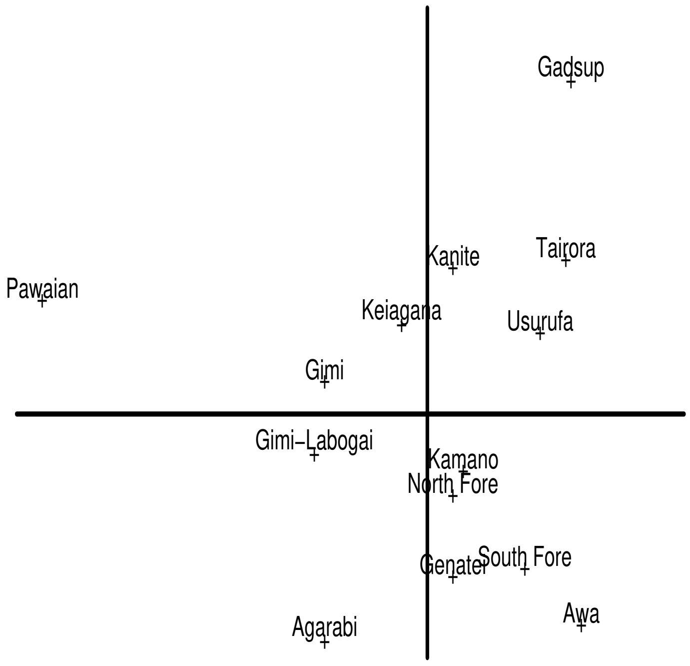
- A single original population undergoes genetic differentiation once it is dispersed into new geographic regions.
- It is interesting anthropologically to compare the rate of this genetic change with changes in nongenetic traits such as language, culture, and use of technology.
- Sorenson and Kenmore (1974) explored the genetic drift of proto-agricultural people in the Eastern Highlands of New Guinea with the idea that changes in horticulture and mountainous geography both determined patterns of dispersion.
- This study uses biological evidence (nine alternative forms of a gene) to make claims about the relatedness of groups that are geographically distinct but similar ethnohistorically and linguistically.
- The raw genetic information can be summarized in a large matrix, but the information in this form is not really the primary interest.

Motivating Example

- To see differences and similarities Sorenson and Kenmore transformed these variables into just two individual factors (new composite variables) that appear to explain the bulk of the genetic variation.

Once that is done it is easy to graph the groups in a single plot and then look at similarities geometrically. This useful result is shown in the figure at right, where we see the placement of these linguistic groups according to the similarity in blood-group genetics. The tool they used for turning the large multidimensional matrix of unwieldy data into an intuitive two-dimensional structure was eigenanalysis.

Linguistic Groups Genetically



Eigen-Analysis of Matrices

- A very useful and theoretically important feature of a given square matrix is the set of *eigenvalues* associated with this matrix.
- Every $p \times p$ matrix $\mathbf{X}$ has p scalar values, $\lambda_i, i = 1, \ldots, p$, such that

  $$\mathbf{X}\mathbf{h}_i = \lambda_i \mathbf{h}_i$$

  for some corresponding vector $\mathbf{h}_i$.
- In this *decomposition*, λ_i is called an eigenvalue of $\mathbf{X}$ and h_i is called an *eigenvector* of $\mathbf{X}$.
- These are also called the *characteristic roots* and *characteristic vectors* of $\mathbf{X}$, and the process is also called *spectral decomposition*.

Eigen-Analysis of Matrices

- The expression above can also be rewritten to produce the *characteristic equation*.
- Start with the algebraic rearrangement

$$(\mathbf{X} - \lambda_i \mathbf{I})\mathbf{h}_i = \mathbf{0}.$$

- If the $p \times p$ matrix in the parentheses has a zero determinant, then there exist eigenvalues that are solutions to the equation:

$$|\mathbf{X} - \lambda_i \mathbf{I}| = \mathbf{0}.$$

Basic Eigenanalysis

- A symmetric matrix $\mathbf{X}$ is given by

$$\mathbf{X} = \begin{bmatrix} 1.000 & 0.880 & 0.619 \\ 0.880 & 1.000 & 0.716 \\ 0.619 & 0.716 & 1.000 \end{bmatrix}.$$

- The eigenvalues and eigenvectors are found by solving the characteristic equation $|\mathbf{X} - \lambda\mathbf{I}| = 0$.

- This produces the matrix

$$\lambda\mathbf{I} = \begin{bmatrix} 2.482 & 0.00 & 0.000 \\ 0.000 & 0.41 & 0.000 \\ 0.000 & 0.00 & 0.108 \end{bmatrix}$$

  from which we take the eigenvalues from the diagonal.

- Note the descending order.

Basic Eigenanalysis

- To see the mechanics of this process more clearly, consider finding the eigenvalues of

$$\mathbf{Y} = \begin{bmatrix} 3 & -1 \\ 2 & 0 \end{bmatrix}.$$

- To do this we expand and solve the determinant of the characteristic equation:

$$|\mathbf{Y} - \lambda \mathbf{I}| = (3 - \lambda)(0 - \lambda) - (-2) = \lambda^2 - 3\lambda + 2$$

  and the only solutions to this quadratic expression are $\lambda_1 = 1, \lambda_2 = 2$.

- In fact, for a $p \times p$ matrix, the resulting characteristic equation will be a polynomial of order p. This is why we had a quadratic expression here.

Basic Eigenanalysis Caveats

- Unfortunately, the eigenvalues that result from the characteristic equation can be zero, repeated (nonunique) values, or even complex numbers.
- All symmetric matrices like the 3×3 example above are guaranteed to have real-valued eigenvalues.

Eigenanalysis Uniqueness

- Eigenvalues and eigenvectors are associated.
- For each eigenvector of a given matrix $\mathbf{X}$ there is exactly one corresponding eigenvalue such that
$$\lambda = \frac{\mathbf{h}'\mathbf{X}\mathbf{h}}{\mathbf{h}'\mathbf{h}}.$$
- For each eigenvalue of the matrix there is an infinite number of eigenvectors, all determined by scalar multiplication.
- If $\mathbf{h}$ is an eigenvector corresponding to the eigenvalue λ, then $s\mathbf{h}$ is also an eigenvector corresponding to this same eigenvalue where s is any nonzero scalar.

Eigenanalysis General Properties

- The number of nonzero eigenvalues is the rank of the $\mathbf{X}$.
- The sum of the eigenvalues is the trace of $\mathbf{X}$.
- The product of the eigenvalues is the determinant of $\mathbf{X}$.
- A matrix is singular if and only if it has a zero eigenvalue, and the rank of the matrix is the number of nonzero eigenvalues.
- If there are no zero-value eigenvalues, then the eigenvectors determine a basis for the space determined by the size of the matrix ($\mathfrak{R}^2$, $\mathfrak{R}^3$, etc.).
- Symmetric nonsingular matrices have eigenvectors that are perpendicular (orthogonal).

Eigenanalysis Properties

Properties of Eigenvalues for a Nonsingular $(n \times n)$ Matrix

$\rightarrowtail$ Inverse Property	If λ_i is an eigenvalue of $\mathbf{X}$, then $\frac{1}{\lambda_i}$ is an eigenvalue of $\mathbf{X}^{-1}$
$\rightarrowtail$ Transpose Property	$\mathbf{X}$ and $\mathbf{X}'$ have the same eigenvalues
$\rightarrowtail$ Identity Matrix	For $\mathbf{I}$, $\sum \lambda_i = n$
$\rightarrowtail$ Exponentiation	If λ_i is an eigenvalue of $\mathbf{X}$, thenλ_i^k is an eigenvalue of $\mathbf{X}^k$ and k a positive integer

Matrix Conditioning

- For a symmetric definite matrix, the ratio of the largest eigenvalue to the smallest eigenvalue is the *condition number*.
- If this number is large, then we say that the matrix is "ill-conditioned," and it usually has poor properties.
- If the matrix is nearly singular (but not quite), then the smallest eigenvalue will be close to zero and the ratio will be large for any reasonable value of the largest eigenvalue.
- As an example of this problem, in the use of matrix inversion to solve systems of linear equations, an ill-conditioned $\mathbf{A}$ matrix means that small changes in $\mathbf{A}$ will produce large changes in $\mathbf{A}^{-1}$ and therefore the calculation of $\mathbf{x}$ will differ dramatically.

Analyzing Social Mobility with Eigens

- Duncan (1966) analyzed social mobility between categories of employment (from the 1962 Current Population Survey) to produce probabilities for blacks and whites.

- This well-known finding is summarized in two *transition* matrices, indicating probabilities for changing between *higher white collar*, *lower white collar*, *higher manual*, *lower manual*, and *farm*:

$$B = \begin{bmatrix} 0.112 & 0.105 & 0.210 & 0.573 & 0.000 \\ 0.156 & 0.098 & 0.000 & 0.745 & 0.000 \\ 0.094 & 0.073 & 0.120 & 0.684 & 0.030 \\ 0.087 & 0.077 & 0.126 & 0.691 & 0.020 \\ 0.035 & 0.034 & 0.072 & 0.676 & 0.183 \end{bmatrix} \qquad W = \begin{bmatrix} 0.576 & 0.162 & 0.122 & 0.126 & 0.014 \\ 0.485 & 0.197 & 0.145 & 0.157 & 0.016 \\ 0.303 & 0.127 & 0.301 & 0.259 & 0.011 \\ 0.229 & 0.124 & 0.242 & 0.387 & 0.018 \\ 0.178 & 0.076 & 0.214 & 0.311 & 0.221 \end{bmatrix},$$

  where the rows and columns are in the order of employment categories given.

- So, for instance, 0.576 in the first row and first column of the W matrix means that we expect 57.6% of the children of white higher white collar workers will themselves become higher white collar workers.

Analyzing Social Mobility with Eigens

- A lot can be learned by staring at these matrices for some time, but what tools will let us understand long-run trends built into the data?
- Since these are transition probabilities, we could multiply one of these matrices to itself a large number of times as a simulation of future events.
- It turns out that the eigenvector produced from $\mathbf{X}\mathbf{h}_i = \lambda_i \mathbf{h}_i$ is the *right eigenvector* because it sits on the right-hand side of $\mathbf{X}$ here.
- This is the default, so when an eigenvector is referenced without any qualifiers, the form derived above is the appropriate one.
- There is also the less-commonly used *left eigenvector* produced from $\mathbf{h}_i\mathbf{X} = \lambda_i \mathbf{h}_i$.
- If $\mathbf{X}$ is a symmetric matrix, then the two vectors are identical (the eigenvalues are the same in either case).
- If $\mathbf{X}$ is not symmetrical, they differ, but the left eigenvector can be produced from using the transpose: $\mathbf{X}'\mathbf{h}_i = \lambda_i \mathbf{h}_i$.

Analyzing Social Mobility with Eigens

- The *spectral component* corresponding to the ith eigenvalue is the square matrix produced from the cross product of the right and left eigenvectors over the dot product of the right and left eigenvectors:

$$S_i = h_{i,\text{right}} \times h_{i,\text{left}} / h_{i,\text{right}} \cdot h_{i,\text{left}}.$$

- This spectral decomposition into constituent components by eigenvalues is especially revealing for probability matrices like the two above, where the rows necessarily sum to 1.

- Because of the probability structure of these matrices, the first eigenvalue is always 1.

- The associated spectral components are

$$B = \begin{bmatrix} 0.09448605 & 0.07980742 & 0.1218223 & 0.6819610 & 0.02114880 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.4293069 & 0.1509444 & 0.1862090 & 0.2148500 & 0.01840510. \end{bmatrix},$$

where only a single row of this 5×5 matrix is given here because all rows are identical (a result of $\lambda_1 = 1$).

Analyzing Social Mobility with Eigens

- The spectral values corresponding to the first eigenvalue give the long-run (stable) proportions implied by the matrix probabilities.
- If conditions do not change, these will be the eventual population proportions.
- If the mobility trends persevere, eventually a little over two-thirds of the black population will be in lower manual occupations, and less than 10% will be in each of the white collar occupational categories.
- For whites, about 40% will be in the higher white collar category with 15–20% in each of the other nonfarm occupational groups.
- Subsequent spectral components from declining eigenvalues give weighted propensities for movement between individual matrix categories.
- The second eigenvalue produces the most important indicator, followed by the third, and so on.

Analyzing Social Mobility with Eigens

- The second spectral components corresponding to the second eigenvalues $\lambda_{2,\text{black}} = 0.177676$, $\lambda_{2,\text{white}} = 0.348045$ are

$$B = \begin{bmatrix} 0.063066 & 0.043929 & 0.034644 & -0.019449 & -0.122154 \\ 0.103881 & 0.072359 & 0.057065 & -0.032037 & -0.201211 \\ -0.026499 & -0.018458 & -0.014557 & 0.008172 & 0.051327 \\ -0.002096 & -0.001460 & -0.001151 & 0.000646 & 0.004059 \\ -0.453545 & -0.315919 & -0.249145 & 0.139871 & 0.878486 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.409172 & 0.055125 & -0.187845 & -0.273221 & -0.002943 \\ 0.244645 & 0.032960 & -0.112313 & -0.163360 & -0.001759 \\ -0.3195779 & -0.043055 & 0.146714 & 0.213396 & 0.002298 \\ -0.6018242 & -0.081080 & 0.276289 & 0.401864 & 0.004328 \\ -1.2919141 & -0.174052 & 0.593099 & 0.862666 & 0.009292 \end{bmatrix}$$

Analyzing Social Mobility with Eigens

- Notice that the full matrix is given here because the rows now differ.
- McFarland noticed the structure highlighted here with the boxes containing positive values.
- For blacks there is a tendency for white collar status and higher manual to be self-reinforcing:
- Once landed in the upper left 2×3 submatrix, there is a tendency to remain and negative influences on leaving.
- The same phenomenon applies for blacks to manual/farm labor: Once there it is more difficult to leave. For whites the phenomenon is the same, except this barrier effect puts higher manual in the less desirable block.
- This suggests a racial differentiation with regard to higher manual occupations.

Quadratic Forms and Descriptions

- An important question is what properties does an $n \times n$ matrix $\mathbf{X}$ possess when pre- and post-multiplied by a conformable nonzero vector $\mathbf{y} \in \mathfrak{R}^n$.
- The quadratic form of the matrix $\mathbf{X}$ is given by

  $$\mathbf{y}'\mathbf{X}\mathbf{y} = s,$$

  where the result is some scalar, s.
- If $s = 0$ for every possible vector $\mathbf{y}$, then $\mathbf{X}$ can only be the null matrix.

Quadratic Forms and Descriptions, Properties

Properties of the Quadratic, y Non-Null

Non-Negative Definite:

$\rightarrowtail$ positive definite $\mathbf{y}'\mathbf{X}\mathbf{y} > 0$

$\rightarrowtail$ positive semidefinite $\mathbf{y}'\mathbf{X}\mathbf{y} \geq 0$

Non-Positive Definite:

$\rightarrowtail$ negative definite $\mathbf{y}'\mathbf{X}\mathbf{y} < 0$

$\rightarrowtail$ negative semidefinite $\mathbf{y}'\mathbf{X}\mathbf{y} \leq 0$

Indefiniteness

- $\mathbf{X}$ is *indefinite* if it is neither nonnegative definite nor nonpositive definite.
- A positive definite matrix is always nonsingular.
- A positive definite matrix is therefore invertible and the resulting inverse will also be positive definite.
- Positive semidefinite matrices are sometimes singular and sometimes not.
- If such a matrix is nonsingular, then its inverse is also nonsingular.
- Every diagonal element of a positive definite matrix is positive, and every diagonal element of a negative definite matrix is negative.
- Every diagonal element of a positive semidefinite matrix is nonnegative, and every diagonal element of a negative semidefinite matrix is nonpositive.

LDU Decomposition

- *Any* $p \times q$ matrix can be decomposed as follows:

$$\underset{(p\times q)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{L}}\underset{(p\times q)}{\mathbb{D}}\underset{(q\times q)}{\mathbf{U}}, \qquad \text{where} \qquad \mathbb{D} = \begin{bmatrix} \mathbf{D}_{r\times r} & 0 \\ 0 & 0 \end{bmatrix},$$

  where $\mathbf{L}$ (lower triangular) and $\mathbf{U}$ (upper triangular) are nonsingular (even given a singular matrix $\mathbf{A}$).

- The diagonal matrix $\mathbf{D}_{r\times r}$ is unique and has dimension and rank r that corresponds to the rank of $\mathbf{A}$.

- If $\mathbf{A}$ is positive definite, and symmetric, then $\mathbf{D}_{r\times r} = \mathbb{D}$ (i.e., $r = q$) and $\mathbf{A} = \mathbf{L}\mathbb{D}\mathbf{L}'$ with unique $\mathbf{L}$.

LDU Decomposition

- For example, consider the LDU decomposition of the 3×3 unsymmetric, positive definite matrix A:

$$\mathbf{A} = \begin{bmatrix} 140 & 160 & 200 \\ 280 & 860 & 1060 \\ 420 & 1155 & 2145 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{bmatrix} \begin{bmatrix} 20 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 10 \end{bmatrix} \begin{bmatrix} 7 & 8 & 10 \\ 0 & 9 & 11 \\ 0 & 0 & 12 \end{bmatrix}.$$

- Now look at the symmetric, positive definite matrix and its LDL' decomposition:

$$\mathbf{B} = \begin{bmatrix} 5 & 5 & 5 \\ 5 & 21 & 21 \\ 5 & 21 & 30 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{bmatrix}',$$

which shows the symmetric principle above.