Lecture Notes 1: Matrix Algebra
Part C: Pivoting and Reduced Row Echelon Form

Peter J. Hammond

revised 2020 September 16th

Lecture Outline

Outline

Three Simultaneous Equations

Consider the following system
of three simultaneous equations in three unknowns,
which depends upon two "exogenous" constants a and b:

$$\begin{array}{rcrcrcl} x & + & y & - & z & = & 1 \\ x & - & y & + & 2z & = & 2 \\ x & + & 2y & + & az & = & b \end{array}$$

It can be expressed, using an augmented 3×4 matrix, as :

$$\left(\begin{array}{rrr|r} 1 & 1 & -1 & 1 \\ 1 & -1 & 2 & 2 \\ 1 & 2 & a & b \end{array}\right)$$

Perhaps even more useful is the doubly augmented 3×7 matrix:

$$\left(\begin{array}{rrr|r|rrr} 1 & 1 & -1 & 1 & 1 & 0 & 0 \\ 1 & -1 & 2 & 2 & 0 & 1 & 0 \\ 1 & 2 & a & b & 0 & 0 & 1 \end{array}\right)$$

whose last 3 columns are those of the 3×3 identity matrix $\mathbf{I}_3$.

The First Pivot Step

Start with the doubly augmented 3×7 matrix:

$$\begin{array}{ccc|c|ccc} 1 & 1 & -1 & 1 & 1 & 0 & 0 \\ 1 & -1 & 2 & 2 & 0 & 1 & 0 \\ 1 & 2 & a & b & 0 & 0 & 1 \end{array}$$

First, we pivot about the element in row 1 and column 1 to eliminate or "zeroize" the other elements of column 1.

This elementary row operation requires us to subtract row 1 from both rows 2 and 3. It is equivalent to multiplying

by the lower triangular matrix $\mathbf{E}_1 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$.

Note: this is the result of applying the same row operations to $\mathbf{I}_3$.

The resulting 3×7 matrix is:

$$\begin{array}{ccc|c|ccc} 1 & 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & -2 & 3 & 1 & -1 & 1 & 0 \\ 0 & 1 & a+1 & b-1 & -1 & 0 & 1 \end{array}$$

The Second Pivot Step

After augmenting again by the identity matrix, we have:

$$\begin{array}{ccc|c|ccc|ccc} 1 & 1 & -1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 3 & 1 & -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & a+1 & b-1 & -1 & 0 & 1 & 0 & 0 & 1 \end{array}$$

Next, we pivot about the element in row 2 and column 2.
Specifically, multiply the second row by $-\frac{1}{2}$,
then subtract the new second row from the third to obtain:

$$\begin{array}{ccc|c|ccc|ccc} 1 & 1 & -1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -\frac{3}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & a+\frac{5}{2} & b-\frac{1}{2} & -\frac{3}{2} & \frac{1}{2} & 1 & 0 & \frac{1}{2} & 1 \end{array}$$

Again, the pivot operation is equivalent to multiplying

by the lower triangular matrix $\mathbf{E}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \end{pmatrix}$,

which is the result of applying the same row operation to $\mathbf{I}_3$.

Case 1: Dependent Equations

In case 1, when $a + \frac{5}{2} = 0$, the equation system reduces to:

$$\begin{array}{rcrcrcr} x & + & y & - & z & = & 1 \\ & & y & - & \frac{3}{2}z & = & -\frac{1}{2} \\ & & & & 0 & = & b - \frac{1}{2} \end{array}$$

In case 1A, when $b \neq \frac{1}{2}$, neither the last equation, nor the system as a whole, has any solution.

In case 1B, when $b = \frac{1}{2}$, the third equation is redundant.

In this case, the first two equations have a general solution with $y = \frac{3}{2}z - \frac{1}{2}$ and $x = z + 1 - y = z + 1 - \frac{3}{2}z + \frac{1}{2} = \frac{3}{2} - \frac{1}{2}z$, where z is an arbitrary scalar.

In particular, there is a one-dimensional set of solutions along the unique straight line in $\mathbb{R}^3$ that passes through both: (i) $(\frac{3}{2}, -\frac{1}{2}, 0)$, when $z = 0$; (ii) $(1, 1, 1)$, when $z = 1$.

Case 2: Three Independent Equations

$$
\left.\begin{array}{ccc} 1 & 1 & -1 \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & a+\frac{5}{2} \end{array}\right| \left.\begin{array}{c} 1 \\ -\frac{1}{2} \\ b-\frac{1}{2} \end{array}\right| \left.\begin{array}{ccc} 1 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{3}{2} & \frac{1}{2} & 1 \end{array}\right| \begin{array}{ccc} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 1 \end{array}
$$

Case 2 occurs when $a + \frac{5}{2} \neq 0$,
and so the reciprocal $c := 1/(a + \frac{5}{2})$ is well defined.
Now divide the last row by $a + \frac{5}{2}$, or multiply by c, to obtain:

$$
\left.\begin{array}{ccc} 1 & 1 & -1 \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 1 \end{array}\right| \left.\begin{array}{c} 1 \\ -\frac{1}{2} \\ (b-\frac{1}{2})c \end{array}\right| \left.\begin{array}{ccc} 1 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{3}{2}c & \frac{3}{2}c & \frac{1}{2}c \end{array}\right|
$$

The system has been reduced to row echelon form
in which the leading zeroes of each successive row form
the steps (in French, *échelons*, meaning rungs)
of a ladder (or *échelle* in French)
which descends steadily as one goes from left to right.

Case 2: Three Independent Equations, Third Pivot

$$\begin{array}{rrr|r|rrr|} 1 & 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -\frac{3}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & (b-\frac{1}{2})c & -\frac{3}{2}c & \frac{3}{2}c & \frac{1}{2}c \end{array}$$

Next, we zeroize the elements in the third column above row 3.

To do so, pivot about the element in row 3 and column 3.

This requires adding 1 times the last row to the first,
and $\frac{3}{2}$ times the last row to the second.

In effect, one multiplies

by the upper triangular matrix $\mathbf{E}_3 := \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{3}{2} \\ 0 & 0 & 1 \end{pmatrix}$

The first three columns of the result are $\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}$

Case 2: Three Independent Equations, Final Pivot

As already remarked, the first three columns of the matrix
we are left with are

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

The final pivoting operation involves subtracting the second row from the first, so the first three columns become the identity matrix

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

This is a matrix in reduced row echelon form because,
given the leading non-zero element of any row (if there is one),
all elements above this element are zero.

Final Exercise

Exercise

1. *Find the last 4 columns of each 3×7 matrix produced by these last two pivoting steps.*
2. *Check that the fourth column solves the original system of 3 simultaneous equations.*
3. *Check that the last 3 columns form the inverse of the original coefficient matrix.*

Outline

Definition of Row Echelon Form

Definition

An $m \times n$ matrix $\mathbf{A}$ is in row echelon form just in case:

1. The first $r \leq m$ rows $i \in \mathbb{N}_r$
   each have a non-zero leading entry a_{i,ℓ_i} in column ℓ_i
   such that $a_{ij} = 0$ for all $j < \ell_i$.
2. Each successive leading entry is in a column to the right of the leading entry in the previous row.

   That is, given the leading element $a_{i,\ell_i} \neq 0$ of row i,
   one has $a_{hj} = 0$ for all $h > i$ and all $j \leq \ell_i$.
3. If $r < m$, then any row $i \in \{r+1, \ldots, m\} = \mathbb{N}_m \setminus \mathbb{N}_r$
   has no leading entry, because all its elements are zero.

   This row without a leading entry
   must be below any row with a leading entry. □

Examples

Assuming that $\alpha, \beta, \gamma \in \mathbb{R} \setminus \{0\}$,
here are three examples of matrices in row echelon form:

$$\mathbf{A} = \begin{pmatrix} \alpha & 2 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \gamma \end{pmatrix};\ \mathbf{B} = \begin{pmatrix} \alpha & 2 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \gamma \\ 0 & 0 & 0 & 0 \end{pmatrix};\ \mathbf{C} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \\ 0 & 0 \end{pmatrix}$$

Here are three examples of matrices
that are not in row echelon form

$$\mathbf{D} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix};\ \mathbf{E} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 1 \end{pmatrix};\ \mathbf{F} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Pivoting to Reach a Generalized Row Echelon Form

Any $m \times n$ matrix $\mathbf{A}$ can be transformed into row echelon form by applying a series of determinant preserving row operations involving non-zero pivot elements.

1. Look for the first or leading non-zero column ℓ_1 in the matrix.
2. Find within column ℓ_1 an element $a_{i_1 \ell_1} \neq 0$ with a large absolute value $|a_{i_1 \ell_1}|$; this will be the first pivot.
3. Interchange rows 1 and i_1, moving the pivot to the top row.
4. To preserve the determinant, adjust the sign of either row 1 or row i_1 (not both) by multiplying that entire row by -1.
5. Subtract $a_{i \ell_1}/a_{1 \ell_1}$ times the new row 1 from each new row $i > 1$.

   This first pivot operation will eliminate all the elements of the pivot column ℓ_1 that lie below the new row 1.

The Intermediate Matrices and Pivot Steps

After $k - 1$ pivoting operations have been completed,
and column ℓ_{k-1} (with $\ell_{k-1} \geq k - 1$) was the last to be used:

1. The first or "top" $k - 1$ rows of the $m \times n$ matrix form a $(k - 1) \times n$ submatrix in row echelon form.
2. The last or "bottom" $m - k + 1$ rows of the $m \times n$ matrix form an $(m - k + 1) \times n$ submatrix whose first ℓ_{k-1} columns are all zero.
3. Find the first column ℓ_k that has at least one non-zero element below row $k - 1$.
4. Choose as the kth pivot element the $a_{i_k \ell_k}$ with $i_k \geq k$ which has the large absolute value $|a_{i_k \ell_k}|$.
5. Interchange rows k and i_k, moving the pivot up to row k, and change the sign of just one of these rows.
6. Subtract $a_{i\ell_k}/a_{k\ell_k}$ times the new row k from each new row $i > k$.

   This kth pivot operation will eliminate all the elements of the pivot column ℓ_k that lie below the new row k.

Ending the Pivoting Process

1. Continue pivoting about successive pivot elements $a_{i_k \ell_k} \neq 0$, moving row $i_k \geq k$ up to row k at each stage k, while leaving all rows above k unchanged.
2. Stop after r steps when either $r = m$, or else all elements in the remaining $m - r$ rows are zero, so no further pivoting is possible.

Outline

Definition of Reduced Row Echelon Form

Definition

An $m \times n$ matrix $\mathbf{A}$ is in reduced row echelon form just in case it is in row echelon form, and the leading entry $a_{i,\ell_i} \neq 0$ in each row i is the only non-zero entry in its column.

That is, $a_{ij} = 0$ for all $j \neq \ell_i$. □

When $m = n$, it is obvious that any diagonal matrix in which all diagonal elements are non-zero is in reduced row echelon form.

Assuming that $\alpha, \beta, \gamma \in \mathbb{R} \setminus \{0\}$, here are three more examples of matrices in reduced row echelon form:

$$\mathbf{A} = \begin{pmatrix} \alpha & 2 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \gamma \end{pmatrix}; \ \mathbf{B} = \begin{pmatrix} \alpha & 2 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \gamma \\ 0 & 0 & 0 & 0 \end{pmatrix}; \ \mathbf{C} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \\ 0 & 0 \end{pmatrix}$$

Reaching a Reduced Row Echelon Form

Consider an $m \times n$ matrix $\mathbf{C}$ that is already in row echelon form.

Suppose it has r leading non-zero elements c_{k,ℓ_k}
in rows $k = 1, 2, \ldots, r$, where ℓ_k is increasing in k.

Starting at the pivot element $c_{r,\ell_r} \neq 0$ in the last pivot row r,
zeroize all the elements in column ℓ_r above this element
by subtracting from each row k above r
the multiple $c_{k,\ell_r}/c_{r,\ell_r}$ of row r of the matrix $\mathbf{C}$,
while leaving row r itself unchanged.

Repeat this pivoting operation for each of the pivot elements c_{k,ℓ_k},
working from $c_{r-1,\ell_{r-1}}$ all the way back and up to c_{1,ℓ_1}.

A Sign Adjusted Permutation of the Columns, I

We have shown how to transform a general $m \times n$ matrix $\mathbf{A}$ into a matrix $\mathbf{C} = \mathbf{RA}$ in reduced row echelon form by applying the row operation $\mathbf{R}$ that equals the product of several determinant preserving row operations.

Denote the leading non-zero elements in the first r rows of $\mathbf{C}$ by $c_{k\ell_k}$, where ℓ_k is increasing in k for $k = 1, 2, \ldots, r$.

For $k = 1, 2, \ldots, r$, post multiply successively by the sign adjusted transposition matrices $\hat{\mathbf{T}}_{k\ell_k}$ in order to interchange columns ℓ_k and k, with an adjusted sign.

Evidently $\hat{\mathbf{T}}_{k\ell_k} = \mathbf{I}$ unless $k < \ell_k$.

Define the product $\hat{\mathbf{P}} = \prod_{k=1}^{r} \hat{\mathbf{T}}_{k\ell_k}$, which is a sign adjusted permutation matrix.

The combined effect of these sign adjusted column interchanges is to form the matrix $\mathbf{C}\hat{\mathbf{P}}$.

A Sign Adjusted Permutation of the Columns, II

Postmultiplying $\mathbf{C}$ by $\hat{\mathbf{P}}$ ensures
that each leading non-zero element $(\mathbf{C})_{k\ell_k}$ in row k
becomes a non-zero element $(\mathbf{C}\hat{\mathbf{P}})_{kk}$ on the diagonal.

Forming $\mathbf{C}\hat{\mathbf{P}}$ also partitions the matrix columns into two sets:

1. first, a complete set of r columns containing all the r pivots, with one pivot in each row and one in each column;
2. then second, the remaining $n - r$ columns without any pivots.

So the resulting matrix $\mathbf{C}\hat{\mathbf{P}} = \mathbf{R}\mathbf{A}\hat{\mathbf{P}}$
has a diagonal sub-matrix $\mathbf{D}_{r \times r}$ in its top left-hand corner.

Moreover, the diagonal elements of $\mathbf{D}_{r \times r}$ are the pivots,
all of which must be non-zero, by construction.

A Maximally Diagonalized Matrix

Our constructions have led to the equality

$$\mathbf{RA\hat{P}} = \begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix}$$

The right-hand side is a partitioned $m \times n$ matrix, whose four sub-matrices have the indicated dimensions.

Because the diagonal matarix $\mathbf{D}_{r \times r}$ is as large as possible, we may call it a maximally diagonalized matrix.

Because the diagonal of $\mathbf{D}_{r \times r} = \mathbf{diag}(d_1, d_2, \ldots, d_r)$ consists of all the non-zero pivots, the inverse $\mathbf{D}_{r \times r}^{-1} = \mathbf{diag}(1/d_1, 1/d_2, \ldots, 1/d_r)$ exists.

Provided that the non-negative integer $r \leq m$ is unique, independent of what pivots are chosen, we may want to call r the pivot rank of the matrix $\mathbf{A}$.

Three Special Cases

So far we have been writing out full partitioned matrices,
as is required when the number of pivots satisfies $r < \min\{m, n\}$.

There are three other special cases when $r = \min\{m, n\}$.

In these three cases, the maximally diagonalized $m \times n$ matrix

$$\mathbf{R}\mathbf{A}\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix}$$

reduces to:

1. $\mathbf{D}_{n \times n}$ in case $r = m = n$, so $m - r = n - r = 0$;
2. $\begin{pmatrix} \mathbf{D}_{m \times m} & \mathbf{B}_{m \times (n-m)} \end{pmatrix}$ in case $r = m < n$, so $m - r = 0$;
3. $\begin{pmatrix} \mathbf{D}_{n \times n} \\ \mathbf{0}_{(m-n) \times n} \end{pmatrix}$ in case $r = n < m$, so $n - r = 0$.

Outline

Finding the Determinant of a Square Matrix

In the case of an $n \times n$ matrix $\mathbf{A}$, our earlier equality becomes

$$\mathbf{R}\mathbf{A}\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{D}_{r\times r} & \mathbf{B}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(n-r)} \end{pmatrix}$$

The determinant of this upper triangular matrix is clearly 0 except in the special case when $r = n$.

When $r = n$, there is a complete set of n pivots.

There are no missing columns, so no need to permute the columns by applying the sign adjusted permutation matrix $\hat{\mathbf{P}}$.

Instead, we have the complete diagonalization $\mathbf{RA} = \mathbf{D}$.

Because $\mathbf{R}$ is determinant preserving,
one has $|\mathbf{RA}| = |\mathbf{A}| = |\mathbf{D}| = \prod_{i=1}^{n} d_i$.

So, to calculate the determinant when $r = n$, it is enough:

1. to pivot to reduce $\mathbf{A}$ to row echelon form or diagonal form;
2. then multiply the diagonal elements.

Inverting a Square Matrix: Necessary Condition

Suppose that $\mathbf{A}$ is $n \times n$, with maximally diagonal form $\mathbf{RA\hat{P}}$.

Consider the equation system $\mathbf{AX} = \mathbf{I}_n$.

Premultiplying this system by $\mathbf{R}$ gives

$$\mathbf{RAX} = \mathbf{RA\hat{P}\hat{P}}^{-1}\mathbf{X} = \begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(n-r) \times r} & \mathbf{0}_{(n-r) \times (n-r)} \end{pmatrix} \mathbf{\hat{P}}^{-1}\mathbf{X} = \mathbf{R}$$

This has a solution only if the last $n - r$ rows of $\mathbf{R}$ are all zero.

But $\mathbf{R}$ is determinant preserving, so $|\mathbf{R}| = 1$, implying that $r = n$.

That is, a necessary condition for $\mathbf{A}$ to be invertible
is that $r = n$, implying that $\mathbf{A}$ has a full set of n pivots.

Inverting a Square Matrix: Sufficient Condition

Conversely, if $r = n$, then there is a complete set of pivots,
so one can take $\hat{\mathbf{P}} = \mathbf{I}$.

Then the maximally diagonalized matrix $\mathbf{R}\mathbf{A}\hat{\mathbf{P}}$ is fully diagonalized,
so $\mathbf{A}\mathbf{X} = \mathbf{I}$ is equivalent to $\mathbf{R}\mathbf{A}\mathbf{X} = \mathbf{D}\mathbf{X} = \mathbf{R}$.

The unique solution is $\mathbf{X} = \mathbf{A}^{-1} = \mathbf{D}^{-1}\mathbf{R}$.

In this case pivoting does virtually all the work of matrix inversion.

This is because all that is left to do is:

1. invert the resulting diagonal matrix $\mathbf{D}$;
2. postmultiply $\mathbf{D}^{-1}$ by the matrix $\mathbf{R}$,
   which represents the product of all the pivoting operations.

Outline

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 $|\alpha \mathbf{A}| = \alpha^n |\mathbf{A}|$.

Verifying the Transpose Rule 2

The transpose rule 2 is very useful: it implies that for any statement $\mathcal{S}$ about how $|\mathbf{A}|$ depends on the rows of $\mathbf{A}$, there is an equivalent "transpose" statement $\mathcal{S}^\top$ about how $|\mathbf{A}|$ depends on the columns of $\mathbf{A}$.

Exercise

Verify Rule 2 directly for 2×2 *and then for* 3×3 *matrices.*

Proof of Rule 2 The expansion formula implies that

$$|\mathbf{A}| = \sum_{\pi \in \Pi} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i\pi(i)} = \sum_{\pi \in \Pi} \operatorname{sgn}(\pi) \prod_{j=1}^{n} a_{\pi^{-1}(j)j}$$

But we proved earlier that $\operatorname{sgn}(\pi^{-1}) = \operatorname{sgn}(\pi)$.

Also $a_{\pi^{-1}(j)j} = a^\top_{j\pi^{-1}(j)}$ by definition of transpose.

Hence, because $\pi \leftrightarrow \pi^{-1}$ is a bijection on the set Π, the expansion formula with π replaced by π^{-1} implies that $|\mathbf{A}| = \sum_{\pi^{-1} \in \Pi} \operatorname{sgn}(\pi^{-1}) \prod_{j=1}^{n} a^\top_{j\pi^{-1}(j)} = |\mathbf{A}^\top|$. $\square$

Verifying the Alternation Rule 4

Recall the notation $\tau_{r,s}$ for the transposition of $r, s \in \mathbb{N}_n$, and $\mathbf{T}_{rs}$ for the associated transposition matrix.

Let $\mathbf{A}_{r\leftrightarrow s}$ denote the matrix that results from applying $\tau_{r,s}$ to the rows of the matrix $\mathbf{A}$ — i.e., interchanging rows r and s.

Theorem

Given any $n \times n$ matrix $\mathbf{A}$ and any transposition $\tau_{r,s}$, one has $\det \mathbf{A}_{r\leftrightarrow s} = \det \mathbf{T}_{rs}\mathbf{A} = -\det \mathbf{A}$.

Proof.

Write τ for $\tau_{r,s}$. Then, because $\pi \leftrightarrow \tau^{-1} \circ \pi$ is a bijection on Π_n and $\operatorname{sgn}(\tau^{-1} \circ \pi) = -\operatorname{sgn}(\pi)$ for all $\pi \in \Pi_n$, we have

$$\begin{aligned}
\det \mathbf{A}_{r\leftrightarrow s} &= \sum_{\pi\in\Pi_n} \operatorname{sgn}(\pi) \prod_{i=1}^n a_{\tau(i),\pi(i)} \\
&= \sum_{\pi\in\Pi_n} \operatorname{sgn}(\pi) \prod_{i=1}^n a_{i,(\tau^{-1}\circ\pi)(i)} \\
&= -\sum_{\pi\in\Pi_n} \operatorname{sgn}(\tau^{-1}\circ\pi) \prod_{i=1}^n a_{i,(\tau^{-1}\circ\pi)(i)} \\
&= -\sum_{\pi\in\Pi_n} \operatorname{sgn}(\pi) \prod_{i=1}^n a_{i,\pi(i)} \;=\; -\det \mathbf{A} \qquad \square
\end{aligned}$$

The Duplication Rule, and Rule 8

The following duplication rule is a special case of Rule 5.

Proposition

If two different rows r and s of $\mathbf{A}$ are equal, then $|\mathbf{A}| = 0$.

Proof.

Suppose that rows r and s of $\mathbf{A}$ are equal.

Then $\mathbf{A}_{r \leftrightarrow s} = \mathbf{A}$, and so $|\mathbf{A}_{r \leftrightarrow s}| = |\mathbf{A}|$.

Yet the alternation Rule 4 implies that $|\mathbf{A}_{r \leftrightarrow s}| = -|\mathbf{A}|$.

Hence $|\mathbf{A}| = -|\mathbf{A}|$, implying that $|\mathbf{A}| = 0$. □

Rule 8: $|\alpha \mathbf{A}| = \alpha^n |\mathbf{A}|$ for any $\alpha \in \mathbb{R}$.

Proof.

The expansion formula implies that

$$\begin{aligned} |\alpha \mathbf{A}| &= \sum_{\pi \in \Pi} \operatorname{sgn}(\pi) \prod_{i=1}^n (\alpha a_{i\pi(i)}) \\ &= \alpha^n \sum_{\pi \in \Pi} \operatorname{sgn}(\pi) \prod_{i=1}^n a_{i\pi(i)}) = \alpha^n |\mathbf{A}| \end{aligned}$$ □

First Implications of Multilinearity: Rules 1 and 3

Recall the notation $\mathbf{A}_{-r}/\mathbf{b}_r^\top$ for the matrix that results after the rth row $\mathbf{a}_r^\top$ of $\mathbf{A}$ has been replaced by $\mathbf{b}_r^\top$.

With this notation, the matrix $\mathbf{A}_{-r}/\alpha\mathbf{a}_r^\top$ is the result of replacing the rth row $\mathbf{a}_r^\top$ of $\mathbf{A}$ by $\alpha\mathbf{a}_r^\top$.

That is, it is the result of multiplying the rth row $\mathbf{a}_r^\top$ of $\mathbf{A}$ by the scalar α.

Rule 3: If all the elements in a single row of $\mathbf{A}$ are multiplied by a scalar α, so is its determinant.

Proof.

By multilinearity one has $|\mathbf{A}_{-r}/\alpha\mathbf{a}_r^\top| = \alpha|\mathbf{A}_{-r}/\mathbf{a}_r^\top| = \alpha|\mathbf{A}|$. □

Rule 1: $|\mathbf{A}| = 0$ if all the elements in a row of $\mathbf{A}$ are 0.

Proof.

This follows from putting $\alpha = 0$ in Rule 3. □

More Implications of Multilinearity: Rules 5 and 6

Rule 5: If two rows of $\mathbf{A}$ are proportional, then $|\mathbf{A}| = 0$.

Proof.

Suppose that $\mathbf{a}_r^\top = \alpha \mathbf{a}_s^\top$ where $r \neq s$.

Then $|\mathbf{A}| = |\mathbf{A}/(\alpha \mathbf{a}_s^\top)_r| = \alpha |\mathbf{A}/(\mathbf{a}_s^\top)_r| = 0$ by duplication. □

Rule 6: $|\mathbf{A}|$ is unchanged if any multiple of one row is added to a different row of $\mathbf{A}$.

Proof.

For the matrix $\mathbf{A}/(\mathbf{a}_r^\top + \alpha \mathbf{a}_s^\top)_r$, where α times row s of $\mathbf{A}$ has been added to row r, row multilinearity implies that

$$|\mathbf{A}/(\mathbf{a}_r^\top + \alpha \mathbf{a}_s^\top)_r| = |\mathbf{A}/(\mathbf{a}_r^\top)_r| + \alpha |\mathbf{A}/(\mathbf{a}_s^\top)_r|$$

But $\mathbf{A}/(\mathbf{a}_r^\top)_r = \mathbf{A}$ and $\mathbf{A}/(\mathbf{a}_s^\top)_r$ has a copy of row s in row r.

By the duplication rule, it follows that

$$|\mathbf{A}/(\mathbf{a}_r^\top + \alpha \mathbf{a}_s^\top)_r| = |\mathbf{A}/(\mathbf{a}_r^\top)_r| + \alpha |\mathbf{A}/(\mathbf{a}_s^\top)_r| = |\mathbf{A}| + 0 = |\mathbf{A}| \quad \square$$

Outline

Verification of the Product Rule 7: Diagonal Case

Recall that Rule 7 is the product rule stating that $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|$.

First we consider the special case
when $\mathbf{A}$ is the $n \times n$ diagonal matrix $\mathbf{D} = \mathbf{diag}(d_1, d_2, \ldots, d_n)$.

Proposition

For any $n \times n$ matrix $\mathbf{B}$, one has $|\mathbf{DB}| = |\mathbf{D}| \cdot |\mathbf{B}| = (\prod_{k=1}^n d_k) |\mathbf{B}|$.

Proof.

First, note that $(\mathbf{DB})_{i,j} = \sum_{k=1}^n d_i \delta_{ik} b_{kj} = d_i b_{ij}$ for all $(i,j) \in \mathbb{N}_n^2$.

Then applying the expansion formula thrice implies that

$$|\mathbf{D}| = \sum_{\pi \in \Pi} \operatorname{sgn}(\pi) \prod_{i=1}^n d_i \delta_{i,\pi(i)} = \prod_{i=1}^n d_i \delta_{ii} = \prod_{i=1}^n d_i$$

because the only non-zero term comes when $\pi = \iota$, and also

$$\begin{aligned} |\mathbf{DB}| &= \sum_{\pi \in \Pi} \operatorname{sgn}(\pi) \prod_{i=1}^n d_i b_{i,\pi(i)} \\ &= (\prod_{k=1}^n d_k) \sum_{\pi \in \Pi} \operatorname{sgn}(\pi) \prod_{i=1}^n b_{i,\pi(i)} = |\mathbf{D}| \cdot |\mathbf{B}| \quad \square \end{aligned}$$

Determinant Preserving Row Operations: Definition

Definition

Let $\mathcal{M}_{m \times n}$ denote the family of all $m \times n$ matrices.

Then any $m \times m$ matrix $\mathbf{R}$ induces, for every $n \in \mathbb{N}$, a row operation $\mathcal{M}_{m \times n} \ni \mathbf{A} \mapsto \mathbf{RA} \in \mathcal{M}_{m \times n}$.

The row operation represented by the $m \times m$ matrix $\mathbf{R}$ is determinant preserving just in case, given any $m \times m$ matrix $\mathbf{X}$, one has $|\mathbf{RX}| = |\mathbf{X}|$. □

Lemma

If the $m \times m$ matrix $\mathbf{R}$ is determinant preserving, then $|\mathbf{R}| = 1$.

Proof.

Putting $\mathbf{X} = \mathbf{I}$ in the definition gives $|\mathbf{R}| = |\mathbf{RI}| = |\mathbf{I}| = 1$. □

Two Basic Determinant Preserving Row Operations

Let $\mathbf{X}$ denote an arbitrary $n \times n$ matrix.

Recall the notation $\mathbf{E}_{r+\alpha q}$ and $\mathbf{E}_{r+\alpha q}\mathbf{X}$ for the matrices which result from applying to $\mathbf{I}$ and $\mathbf{X}$ respectively the elementary row operation of adding α times row q to row r.

Recall too that $\hat{\mathbf{T}}_{rs}$ denotes an elementary row operation of:
(i) first interchanging rows r and s;
(ii) then adjusting the sign of one of these two rows.

We know that $|\hat{\mathbf{T}}_{rs}| = 1$.

Evidently, for any $m \times m$ matrix $\mathbf{X}$ we have $|\mathbf{E}_{r+\alpha q}\mathbf{X}| = |\hat{\mathbf{T}}_{rs}\mathbf{X}| = |\mathbf{X}|$.

So the row operations $\mathbf{E}_{r+\alpha q}$ and $\hat{\mathbf{T}}_{rs}$ are all determinant preserving.

This implies that their inverses all exist, with $\mathbf{E}_{r+\alpha q}^{-1} = \mathbf{E}_{r-\alpha q}$ and $(\mathbf{T}^*_{r\to s\to r^-})^{-1} = \mathbf{T}^*_{s\to r\to s^-}$.

The Subgroup of Determinant Preserving Row Operations

The set of all non-singular $m \times m$ matrices forms a group $\mathcal{G}_m$ under matrix multiplication, with identity $\mathbf{I}$ and matrix inversion.

The set $\mathcal{R}_m$ of all determinant preserving row operations on $m \times n$ matrices is a subgroup of $\mathcal{G}_m$ because:

1. if the $m \times m$ matrix $\mathbf{R}$ is determinant preserving, then it is non-singular because $|\mathbf{R}| = |\mathbf{RI}| = |\mathbf{I}| = 1$;
2. if the two $m \times m$ matrices $\mathbf{R}$ and $\mathbf{S}$ are both determinant preserving, then for every $m \times m$ matrix $\mathbf{X}$ one has

$$|(\mathbf{RS})\mathbf{X}| = |\mathbf{R}(\mathbf{SX})| = |\mathbf{SX}| = |\mathbf{X}|$$

implying that $\mathbf{RS}$ is also determinant preserving.

Verification of the Product Rule 7: Diagonal Case

Proposition

For any two $n \times n$ matrices $\mathbf{D}$ and $\mathbf{B}$ where $\mathbf{D}$ is diagonal, one has $|\mathbf{DB}| = |\mathbf{D}| \cdot |\mathbf{B}|$.

Proof.

Let $\mathbf{D} = \mathbf{diag}(d_1, d_2, \ldots, d_n)$.

Note that $\mathbf{DB}$ is the matrix that results from simultaneously multiplying each row $r = 1, 2, \ldots, n$ of $\mathbf{B}$ by the corresponding diagonal element d_r.

By Rule 3 applied n times, the result of all these simultaneous multiplications is that the determinant is multiplied by $\prod_{r=1}^{n} d_r$.

So $|\mathbf{DB}| = \prod_{r=1}^{n} d_r \cdot |\mathbf{B}|$.

But $|\mathbf{D}| = \prod_{r=1}^{n} d_r$, so $|\mathbf{DB}| = |\mathbf{D}| \cdot |\mathbf{B}|$. □

Verification of the Product Rule 7: Non-Singular Case

Proposition

For any two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ where $|\mathbf{A}| \neq 0$, one has $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|$.

Proof.

Because $|\mathbf{A}| \neq 0$, there exist a non-singular diagonal matrix $\mathbf{D}$ and a sequence of determinant preserving row operations $\langle \mathbf{R}_k \rangle_{k=1}^m$ such that $\mathbf{RA} = \mathbf{D}$ where $\mathbf{R} = \prod_{k=1}^m \mathbf{R}_k$.

Because the family of all determinant preserving row operations is a subgroup, and so closed under matrix multiplication, the matrix $\mathbf{R}$, as well as its inverse $\mathbf{R}^{-1}$, are also determinant preserving row operations.

Hence $|\mathbf{A}| = |\mathbf{R}^{-1}\mathbf{D}| = |\mathbf{D}|$ and also $|\mathbf{AB}| = |\mathbf{R}^{-1}\mathbf{DB}| = |\mathbf{DB}|$.

Because $\mathbf{D}$ is diagonal, it follows that $|\mathbf{DB}| = |\mathbf{D}| \cdot |\mathbf{B}|$ and so

$$|\mathbf{AB}| = |\mathbf{DB}| = |\mathbf{D}| \cdot |\mathbf{B}| = |\mathbf{A}| \cdot |\mathbf{B}| \qquad \square$$

Verification of the Product Rule 7: Singular Case

In case the $n \times n$ matrix $\mathbf{A}$ satisfies $|\mathbf{A}| = 0$, there exists $r < n$ such that the maximally diagonalized matrix takes the form

$$\mathbf{R}\mathbf{A}\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{C}_{r \times (n-r)} \\ \mathbf{0}_{(n-r) \times r} & \mathbf{0}_{(n-r) \times (n-r)} \end{pmatrix}$$

where $n - r \geq 1$, while $\hat{\mathbf{P}}$ is a $n \times n$ permutation matrix, and the $n \times n$ matrix $\mathbf{R}$ is determinant preserving.

So there exist matrices $\mathbf{S}, \mathbf{T}, \mathbf{U}, \mathbf{V}$ of suitable dimension such that $\mathbf{RAB} = (\mathbf{R}\mathbf{A}\hat{\mathbf{P}})\mathbf{P}^{-1}\mathbf{B}$ takes the form

$$\begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{C}_{r \times (n-r)} \\ \mathbf{0}_{(n-r) \times r} & \mathbf{0}_{(n-r) \times (n-r)} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{r \times r} & \mathbf{T}_{r \times (n-r)} \\ \mathbf{U}_{(n-r) \times r} & \mathbf{V}_{(n-r) \times (n-r)} \end{pmatrix}$$

Hence $|\mathbf{AB}| = |\mathbf{RAB}| = \begin{vmatrix} \mathbf{DS} + \mathbf{CU} & \mathbf{DT} + \mathbf{CV} \\ \mathbf{0}_{(n-r) \times r} & \mathbf{0}_{(n-r) \times (n-r)} \end{vmatrix} = 0 = |\mathbf{A}| \cdot |\mathbf{B}|$ also in this case when $|\mathbf{A}| = 0$.

Verification of the Product Rule 7: Summary

Finally, therefore, in view of the previous proposition when $|\mathbf{A}| \neq 0$, we have proved:

Theorem

For any $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, one has $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|$.

Outline

Cofactor Expansion: Theorem

Definition

Given any element a_{rs} of the matrix $n \times n$ matrix $\mathbf{A}$,
the associated (r, s)-cofactor $|\mathbf{C}_{rs}|$
is the determinant of the $(n-1) \times (n-1)$ matrix $\mathbf{C}_{rs}$
obtained by omitting row r and column s from $\mathbf{A}$.

The cofactor expansions of $|\mathbf{A}|$ along any row r or column s
are respectively $\sum_{j=1}^{n} (-1)^{r+j} a_{rj} |\mathbf{C}_{rj}|$ and $\sum_{i=1}^{n} (-1)^{i+s} a_{is} |\mathbf{C}_{is}|$.

Theorem

For every row r and column s of any $n \times n$ matrix $\mathbf{A}$,
these cofactor expansions are valid — i.e., one has

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

The proof of this theorem will occupy the next 6 slides.

Cofactor Expansion: Proof, Part 1

Later we will prove the row expansion formula.

If it is valid, then applying it to the transpose matrix $\mathbf{A}^\top$ gives

$$|\mathbf{A}^\top| = \sum\nolimits_{j=1}^{n} (-1)^{r+j} a_{rj}^\top |\mathbf{C}_{rj}^\top|$$

Taking transposes throughout gives

$$|\mathbf{A}| = \sum\nolimits_{j=1}^{n} (-1)^{r+j} a_{jr} |\mathbf{C}_{jr}|$$

Replacing j by i and r by s, then $s + i$ by $i + s$, one obtains

$$|\mathbf{A}| = \sum\nolimits_{i=1}^{n} (-1)^{i+s} a_{is} |\mathbf{C}_{is}|$$

This is the column expansion formula.

So we have proved that the column expansion formula is implied by the row expansion formula, leaving us to prove the latter.

Cofactor Expansion: Proof, Part 2

To verify the row expansion formula,
first note that the rth row vector satisfies $\mathbf{a}_r^\top = \sum_{j=1}^n a_{rj}\, \mathbf{e}_j^\top$,
where $\mathbf{e}_j^\top$ is defined as the jth unit row vector in $\mathbb{R}^n$,
equal to the jth row of the $n \times n$ identity matrix $\mathbf{I}_n$.

Because the determinant is multilinear, it follows that

$$|\mathbf{A}| = \sum\nolimits_{j=1}^n a_{rj}\, |\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r|$$

which is a linear combination of the n determinants $|\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r|$
in which row $\mathbf{a}_r^\top$ of $\mathbf{A}$ gets successively replaced
by each corresponding jth unit row vector $\mathbf{e}_j^\top$.

Therefore, to verify the row expansion formula

$$|\mathbf{A}| = \sum\nolimits_{j=1}^n (-1)^{r+j} a_{rj} |\mathbf{C}_{rj}|$$

we show that $|\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r| = (-1)^{r+j}|\mathbf{C}_{rj}|$ for each $j \in \mathbb{N}_n$.

Cofactor Expansion: Proof, Part 3

Consider the bordered $n \times n$ matrix $\hat{\mathbf{C}}_{rj} = \begin{pmatrix} \mathbf{C}_{rj} & (\mathbf{a}_j)_{-r} \\ \mathbf{0}^\top & 1 \end{pmatrix}$ whose:

1. top left hand corner
   is the $(n-1) \times (n-1)$ cofactor matrix $\hat{\mathbf{C}}_{rj}$;
2. top right hand border is the column vector $(\mathbf{a}_j)_{-r} \in \mathbb{R}^{n-1}$
   that is constructed by dropping the rth component
   from the jth column $\mathbf{a}_j$ of the original matrix $\mathbf{A}$;
3. bottom left hand border
   is the $n-1$-dimensional row vector $\mathbf{0}^\top$ of zeros;
4. bottom right hand corner is the number 1.

Three lemmas will be used to show that, for each $j \in \mathbb{N}_n$:
(i) the permutations $\pi^{r\nearrow n}$ and $\pi^{j\nearrow n}$
with their associated permutation matrices $\mathbf{P}^{r\nearrow n}$ and $\mathbf{P}^{j\nearrow n}$
together satisfy $\hat{\mathbf{C}}_{rj} = \mathbf{P}^{r\nearrow n}\mathbf{A}\mathbf{P}^{j\nearrow n}$;
(ii) $|\mathbf{A}_{-r}/\mathbf{e}_j^\top| = (-1)^{r+j}|\hat{\mathbf{C}}_{rj}|$; and (iii) $|\hat{\mathbf{C}}_{rj}| = |\mathbf{C}_{rj}|$.

This will complete the proof that $|\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r| = (-1)^{r+j}|\mathbf{C}_{rj}|$.

Cofactor Expansion: Proof, Part 4

Given $k \leq \ell \leq n$, recall that $\pi^{k \nearrow \ell} \in \Pi_n$ moves k to ℓ, and then moves each $q \in \{k+1, \ldots, \ell\}$ to $q - 1$.

Let $\mathbf{P}^{k \nearrow \ell}$ denote the corresponding permutation $\mathbf{P}^{\pi^{k \nearrow \ell}}$.

Lemma

For each $r, j \in \mathbb{N}_n$, one has

$$\hat{\mathbf{C}}_{rj} = \begin{pmatrix} \mathbf{C}_{rj} & (\mathbf{a}_j)_{-r} \\ [1ex]\mathbf{0}^\top & 1 \end{pmatrix} = \mathbf{P}^{r \nearrow n} \, [\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r] \, \mathbf{P}^{j \nearrow n}$$

Proof.

Premultiplying by $\mathbf{P}^{r \nearrow n}$ applies $\pi^{r \nearrow n}$ to the rows, whereas postmultiplying by $\mathbf{P}^{j \nearrow n}$ applies $\pi^{j \nearrow n}$ the columns.

Now the result follows immediately from the definitions of:
(i) the matrix $\hat{\mathbf{C}}_{rj}$; (ii) the permutations $\pi^{r \nearrow n}$ and $\pi^{j \nearrow n}$;
(iii) the associated permutation matrices $\mathbf{P}^{r \nearrow n}$ and $\mathbf{P}^{j \nearrow n}$. □

Cofactor Expansion: Proof, Part 5

Lemma

For each $r, j \in \mathbb{N}_n$ *one has* $|\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r| = (-1)^{r+j}|\hat{\mathbf{C}}_{rj}|$.

Proof.

The previous Lemma implies $|\hat{\mathbf{C}}_{rj}| = |\mathbf{P}^{r\nearrow n}\,[\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r]\,\mathbf{P}^{j\nearrow n}|$.

In earlier results we showed that $|\mathbf{P}^\pi \mathbf{A}| = |\mathbf{A}\mathbf{P}^\pi| = \mathrm{sgn}(\pi)|\mathbf{A}|$
and also that $\mathrm{sgn}(\pi^{k\nearrow \ell}) = (-1)^{\ell-k}$.

Hence we have $|\mathbf{P}^{r\nearrow n}\,[\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r]| = \mathrm{sgn}(\pi^{r\nearrow n})|\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r|$
and so $|\mathbf{P}^{r\nearrow n}\,[\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r]\,\mathbf{P}^{j\nearrow n}| = (-1)^{n-r}(-1)^{n-j}|\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r|$.

Because $(-1)^{2n} = 1$ and $(-1)^k = (-1)^{-k}$ for all $k \in \mathbb{N}$, one has

$$\begin{aligned} |\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r| &= (-1)^{r+j-2n}|\mathbf{P}^{r\nearrow n}\,[\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r]\,\mathbf{P}^{j\nearrow n}| \\ &= (-1)^{r+j}|\hat{\mathbf{C}}_{rj}| \end{aligned}$$

□

Cofactor Expansion: Proof, Part 6

Lemma

For each $j \in \mathbb{N}_n$ *one has* $|\hat{\mathbf{C}}_{rj}| = \begin{vmatrix} \mathbf{C}_{rj} & (\mathbf{a}_j)_{-r} \\ \mathbf{0}^\top & 1 \end{vmatrix} = |\mathbf{C}_{rj}|$.

Proof.

Note that $(\hat{\mathbf{C}}_{rj})_{n,\pi(n)} = \delta_{n,\pi(n)}$, so the expansion formula yields

$$|\hat{\mathbf{C}}_{rj}| = \sum\nolimits_{\pi \in \Pi_n} \prod\nolimits_{i=1}^{n} (\hat{\mathbf{C}}_{rj})_{i,\pi(i)} = \sum\nolimits_{\pi \in \Pi_{n-1}} \prod\nolimits_{i=1}^{n-1} (\hat{\mathbf{C}}_{rj})_{i,\pi(i)}$$

because all other terms are equal to zero.

But then the definition of the bordered matrix $\hat{\mathbf{C}}_{rj}$ implies that

$$|\hat{\mathbf{C}}_{rj}| = \sum\nolimits_{\pi \in \Pi_{n-1}} \prod\nolimits_{i=1}^{n-1} (\mathbf{C}_{rj})_{i,\pi(i)} \; = \; |\mathbf{C}_{rj}| \qquad \square$$

This completes all the parts of the proof
that the row r cofactor expansion of $|\mathbf{A}|$ is valid. $\square$

Outline

Expansion by Alien Cofactors

Expanding along either row r or column s gives

$$|\mathbf{A}| = \sum_{j=1}^{n} a_{rj}|\mathbf{C}_{rj}| = \sum_{i=1}^{n} a_{is}|\mathbf{C}_{is}|$$

when one uses matching cofactors.

Expanding by alien cofactors, however,
from either the wrong row $i \neq r$
or the wrong column $j \neq s$, gives

$$0 = \sum_{j=1}^{n} a_{rj}|\mathbf{C}_{ij}| = \sum_{i=1}^{n} a_{is}|\mathbf{C}_{ij}|$$

This is because the answer will be the determinant
of an alternative matrix in which:

- either row i has been duplicated and put in row r;
- or column j has been duplicated and put in column s.

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 $(\mathbf{C}^+)^\top$ of the cofactor matrix $\mathbf{C}^+$
whose elements $(\mathbf{C}^+)_{ij} = |\mathbf{C}_{ij}|$ are the respective cofactors of $\mathbf{A}$.

Main Property of the Adjugate Matrix

Theorem

For every $n \times n$ square matrix $\mathbf{A}$ one has

$$(\mathbf{adj}\,\mathbf{A})\mathbf{A} = \mathbf{A}(\mathbf{adj}\,\mathbf{A}) = |\mathbf{A}|\mathbf{I}_n$$

Proof.

The (i,j) elements of the two product matrices are respectively

$$[(\mathbf{adj}\,\mathbf{A})\,\mathbf{A}]_{ij} = \sum\nolimits_{k=1}^{n} |\mathbf{C}_{ki}|a_{kj} \text{ and } [\mathbf{A}(\mathbf{adj}\ \mathbf{A})]_{ij} = \sum\nolimits_{k=1}^{n} a_{ik}|\mathbf{C}_{jk}|$$

These are both cofactor expansions, which are expansions by:

- alien cofactors in case $i \neq j$, implying that both equal 0;
- matching cofactors in case $i = j$, implying that both equal $|\mathbf{A}|$.

Hence for each pair (i,j) one has

$$[(\mathbf{adj}\,\mathbf{A})\,\mathbf{A}]_{ij} = [\mathbf{A}\,(\mathbf{adj}\,\mathbf{A})]_{ij} = |\mathbf{A}|\delta_{ij} = |\mathbf{A}|(\mathbf{I}_n)_{ij} \qquad \square$$

Outline

Existence of the Inverse Matrix

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 two matrix equations $\mathbf{AX} = \mathbf{I}_n$ and $\mathbf{XA} = \mathbf{I}_n$ has a solution.

Proof.

Provided that $|\mathbf{A}| \neq 0$, 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 $\mathbf{XA} = \mathbf{I}_n$ has a solution,
then the product rule for determinants implies
that $1 = |\mathbf{I}_n| = |\mathbf{XA}| = |\mathbf{X}||\mathbf{A}|$.

Similarly if $\mathbf{AX} = \mathbf{I}_n$ has a solution.

In either case one has $|\mathbf{A}| \neq 0$.

The rest follows from the paragraph above. □

Singularity versus Invertibility

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 $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Inverting Triangular Matrices

Theorem

If the inverse $\mathbf{U}^{-1}$ *of an upper triangular matrix* $\mathbf{U}$ *exists, then it is upper triangular.*

Taking transposes leads immediately to:

Corollary

If the inverse $\mathbf{L}^{-1}$ *of a lower triangular matrix* $\mathbf{L}$ *exists, then it is lower triangular.*

Inverting Triangular Matrices: Proofs

Recall the $(n-1) \times (n-1)$ cofactor matrix $\mathbf{C}_{rs}$
that results from omitting row r and column s of $\mathbf{U} = (u_{ij})$.

When it exists, $\mathbf{U}^{-1} = (1/|\mathbf{U}|)\, \mathbf{adj}\, \mathbf{U}$, so it is enough to prove
that the $n \times n$ matrix $(|\mathbf{C}_{rs}|)$ of cofactor determinants,
whose transpose $(|\mathbf{C}_{rs}|)^\top$ is the adjugate, is lower triangular.

In case $r < s$, every element below the diagonal of the matrix $\mathbf{C}_{rs}$
is also below the diagonal of $\mathbf{U}$, so must equal 0.

Hence $\mathbf{C}_{rs}$ is upper triangular,
with determinant equal to the product of its diagonal elements.

Yet $s - r$ of these diagonal elements are $u_{i+1,i}$ for $i = r, \ldots, s-1$.
These elements are from below the diagonal of $\mathbf{U}$, so equal zero.

Hence $r < s$ implies that $|\mathbf{C}_{rs}| = 0$, so the $n \times n$ matrix $(|\mathbf{C}_{rs}|)$
of cofactor determinants is indeed lower triangular, as required. $\square$

Cramer's Rule: Statement

Notation

Given any $m \times n$ *matrix* $\mathbf{A}$,
recall that $[\mathbf{A}_{-j}/\mathbf{b}]$ *denotes 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* i*th component is given by the ratio of determinants* $x_i = |[\mathbf{A}_{-i}/\mathbf{b}]|/|\mathbf{A}|$.

This result is known as Cramer's rule.

Cramer's Rule: Proof

Proof.

Given the equation $\mathbf{Ax} = \mathbf{b}$, each cofactor $|\mathbf{C}_{ij}|$ of the coefficient matrix $\mathbf{A}$ is formed by dropping row i and column j of $\mathbf{A}$.

It therefore equals the (i, j) cofactor of the matrix $|[\mathbf{A}_{-j}/\mathbf{b}]|$.

Expanding the determinant by cofactors along column j therefore gives

$$|[\mathbf{A}_{-j}/\mathbf{b}]| = \sum_{i=1}^{n} b_i |\mathbf{C}_{ij}| = \sum_{i=1}^{n} (\mathbf{adj}\,\mathbf{A})_{ji} b_i$$

by definition of the adjugate matrix.

Hence the unique solution to the equation system has components

$$x_i = (\mathbf{A}^{-1}\mathbf{b})_i = \frac{1}{|\mathbf{A}|} \sum_{i=1}^{n} (\mathbf{adj}\,\mathbf{A})_{ji} b_i = \frac{1}{|\mathbf{A}|} |[\mathbf{A}_{-i}/\mathbf{b}]|$$

for $i = 1, 2, \ldots, n$. □

Outline

Definition of Dimension

The dimension of a linear space is the number of elements in the largest linearly independent subset.

Theorem

The dimension of $\mathbb{R}^m$ *is* m*.*

To prove this, we first construct a linearly independent set of m vectors.

Indeed, consider the list $(\mathbf{e}_j)_{j=1}^m$ of m unit column vectors in $\mathbb{R}^m$ with each $\mathbf{e}_j$ equal to jth column of the $m \times m$ identity matrix $\mathbf{I}_m$.

Obviously $\mathbf{0} = \mathbf{I}_m \mathbf{x}$ implies that $\mathbf{x} = \mathbf{0}$, so this list does form a linearly independent set.

Linear Dependence with Too Many Vectors

To complete the proof that $\mathbb{R}^m$ has dimension m, consider any list $(\mathbf{y}_j)_{j=1}^n$ of $n > m$ vectors in $\mathbb{R}^m$.

These n vectors form the columns of an $m \times n$ matrix $\mathbf{Y}$.

After applying enough suitable pivoting operations, the matrix equation $\mathbf{Yx} = \mathbf{0}$ reduces to

$$\mathbf{RY\hat{P}z} = \begin{pmatrix} \mathbf{D}_{r\times r} & \mathbf{B}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix} \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} = \mathbf{0}$$

where: (i) $\mathbf{\hat{P}}$ is a sign adjustified $n \times n$ permutation matrix; (ii) $r \leq m < n$; (iii) $\mathbf{D}_{r\times r}$ is diagonal and non-singlular; (iv) $\mathbf{z} = \mathbf{\hat{P}}^{-1}\mathbf{x}$.

This equation system has many non-trivial solutions of the form $\mathbf{z}_1 = -\mathbf{D}^{-1}\mathbf{Bz}_2$ where $\mathbf{z}_2 \in \mathbb{R}^{n-r} \setminus \{\mathbf{0}\}$ is arbitrary.

Then $\mathbf{Y}\begin{pmatrix} -\mathbf{D}^{-1}\mathbf{Bz}_2 \\ \mathbf{z}_2 \end{pmatrix} = \mathbf{0}$ for all $\mathbf{z}_2 \in \mathbb{R}^{n-r}$.

It follows that the list $(\mathbf{y}_j)_{j=1}^n$ of n vectors is linearly dependent. □

Linear Independence of Matrix Columns

The n column vectors of the $m \times n$ matrix $\mathbf{A}$
are linearly independent just in case
the vector equation $\mathbf{0}_m = \sum_{j=1}^n \xi_j \mathbf{a}_j$ in $\mathbb{R}^m$
implies that $\xi_j = 0$ for each $j = 1, 2, \ldots, n$.

Or equivalently, just in case the only solution of $\mathbf{0}_m = \mathbf{A}\mathbf{x}$
is the trivial solution $\mathbf{x} = \mathbf{0}_n$.

Spanning

Definition

Given any finite set $S = \{\mathbf{x}^j \in \mathbb{R}^n \mid j \in \mathbb{N}_m\}$ of m vectors in $\mathbb{R}^n$, the set of vectors spanned by S, or the span of S, is the set

$$\operatorname{sp} S := \{\mathbf{z} \in \mathbb{R}^n \mid \forall j \in \mathbb{N}_m;\ \exists y_j \in \mathbb{R} : z = \sum_{j=1}^{m} y_j \, \mathbf{x}^j\}$$

Note that any vector $\mathbf{z} \in \operatorname{sp} S$ is a linear combination of the vectors in S.

Exercise

Verify that $\operatorname{sp} \mathbf{A}$ *is a linear subspace of* $\mathbb{R}^n$ *— i.e., it satisfies the vector space axioms.*

The Column and Row Spaces

In case the set $S = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$
consists of the n columns of the $m \times n$ matrix $\mathbf{A}$, one has

$$\text{sp}(\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}) = \{\mathbf{y} \in \mathbb{R}^m \mid \exists \mathbf{x} \in \mathbb{R}^n \mid \mathbf{y} = \mathbf{A}\mathbf{x}\}$$

This is the column space of $\mathbf{A}$; the row space spanned by its rows, which equals the column space of $\mathbf{A}^\top$, is given by

$$\text{sp}(\{\mathbf{a}_1^\top, \ldots, \mathbf{a}_m^\top\}) = \{\mathbf{w}^\top \in \mathbb{R}^n \mid \exists \mathbf{z}^\top \in \mathbb{R}^m \mid \mathbf{w}^\top = z^\top \mathbf{A}\}$$

Column and Row Rank

Definition

The column rank of the $m \times n$ matrix $\mathbf{A}$
is the dimension $r_C \leq m$ of its column space,
which is the maximum number of linearly independent columns.

The row rank of the $m \times n$ matrix $\mathbf{A}$
is the dimension $r_R \leq n$ of its row space,
which is the maximum number of linearly independent rows. □

Obviously, the row rank of $\mathbf{A}$
equals the column rank of the transpose $\mathbf{A}^\top$.

The Column Rank of a Maximally Diagonalized Matrix

Theorem

The maximally diagonalized $m \times n$ matrix

$$\begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix}$$

where $\mathbf{D}_{r \times r}$ is invertible, has column rank r.

Proof.

Given an arbitrary $\mathbf{z} \in \mathbb{R}^r$ and $\mathbf{w} \in \mathbb{R}^{m-r}$, the vector equation

$$\begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{z} \\ \mathbf{w} \end{pmatrix}$$

has a solution given by $\mathbf{x} = \mathbf{D}^{-1}(\mathbf{z} - \mathbf{B}\mathbf{y}) \in \mathbb{R}_r$ iff $\mathbf{w} = \mathbf{0}_{m-r}$.

Hence the column space is $\mathbb{R}^r \times \{\mathbf{0}_{m-r}\}$.

It is isomorphic to $\mathbb{R}^r$, whose dimension is r,
the number of pivots. □

The Row Rank of a Maximally Diagonalized Matrix

Theorem

The maximally diagonalized $m \times n$ matrix

$$\begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix}$$

where $\mathbf{D}_{r \times r}$ is invertible has row rank r.

Proof.

Given an arbitrary row vector $(\mathbf{z}^\top, \mathbf{w}^\top) \in \mathbb{R}^r \times \mathbb{R}^{m-r}$, the equation

$$(\mathbf{x}^\top, \mathbf{y}^\top) \begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix} = (\mathbf{z}^\top, \mathbf{w}^\top)$$

has a solution given by $(\mathbf{x}^\top, \mathbf{y}^\top) = (\mathbf{z}^\top \mathbf{D}^{-1}, \mathbf{0}^\top_{m-r})$
if and only if $\mathbf{w}^\top = \mathbf{z}^\top \mathbf{D}^{-1} \mathbf{B}$.

Hence the row space is $\{(\mathbf{z}^\top, \mathbf{w}^\top) \in \mathbb{R}^r \times \mathbb{R}^{m-r} \mid \mathbf{w}^\top = \mathbf{z}^\top \mathbf{D}^{-1} \mathbf{B}\}$.

It is isomorphic to $\mathbb{R}^r$, whose dimension is r. □

Invariance of Row Space

Theorem

Let $\mathbf{A}$ *be any* $m \times n$ *matrix*
and $\mathbf{R}$ *any determinant preserving row operation.*

Then $\mathbf{A}$ *and* $\mathbf{RA}$ *have the same row space.*

Proof.

Suppose that $\mathbf{w}^\top \in \mathbb{R}^n$ is in the row space of $\mathbf{A}$,
with $\mathbf{w}^\top = \mathbf{z}^\top \mathbf{A}$ where $\mathbf{z}^\top \in \mathbb{R}^m$.

Then $\mathbf{w}^\top = (\mathbf{z}^\top \mathbf{R}^{-1})\mathbf{RA}$,
so $\mathbf{w}^\top \in \mathbb{R}^n$ is in the row space of $\mathbf{RA}$.

Conversely, suppose $\mathbf{w}^\top \in \mathbb{R}^n$ is in the row space of $\mathbf{RA}$,
with $\mathbf{w}^\top = \mathbf{z}^\top \mathbf{RA}$ where $\mathbf{z}^\top \in \mathbb{R}^m$.

Then $\mathbf{w}^\top = (\mathbf{z}^\top \mathbf{R})\mathbf{A}$,
so $\mathbf{w}^\top \in \mathbb{R}^n$ is in the row space of $\mathbf{A}$. □

Isomorphism of Column Spaces

Theorem

Let $\mathbf{A}$ *be any* $m \times n$ *matrix*
and $\mathbf{R}$ *any determinant preserving row operation.*

Then $\mathbf{A}$ *and* $\mathbf{RA}$ *have isomorphic column spaces.*

Proof.

Suppose that $\mathbf{y} \in \mathbb{R}^m$ is in the column space of $\mathbf{A}$,
with $\mathbf{y} = \mathbf{Ax}$ where $\mathbf{x} \in \mathbb{R}^n$.

Then $\mathbf{Ry} = (\mathbf{RA})\mathbf{x}$, so $\mathbf{Ry}$ is in the column space of $\mathbf{RA}$.

Conversely, suppose $\mathbf{Ry}$ is in the column space of $\mathbf{RA}$,
with $\mathbf{Ry} = (\mathbf{RA})\mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^n$.

Because $\mathbf{R}$ is determinant preserving, it is invertible.

Then $\mathbf{y} = \mathbf{R}^{-1}(\mathbf{RA})\mathbf{x} = \mathbf{Ax}$, so $\mathbf{y}$ is in the column space of $\mathbf{A}$.

It follows that $\mathbf{y} \leftrightarrow \mathbf{Ry}$ is a linear bijection
between the column spaces of $\mathbf{A}$ and $\mathbf{RA}$. □

Column Rank Equals Row Rank

Theorem

Suppose the $m \times n$ *matrix* $\mathbf{A}$ *can be maximally diagonalized as* $\mathbf{R}\mathbf{A}\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix}$ *where* $\mathbf{D}_{r \times r}^{-1}$ *exists, while the* $m \times m$ *matrix* $\mathbf{R}$ *is determinant preserving, and the* $n \times n$ *matrix* $\hat{\mathbf{P}}$ *is a sign adjusted permutation.*

Then both the column and row rank of $\mathbf{A}$ *are equal to* r.

Proof.

Because permuting the columns of a matrix makes no difference to its row or column rank, the row and column ranks of $\mathbf{RA}$ are equal to those of $\mathbf{R}\mathbf{A}\hat{\mathbf{P}}$, both of which equal r.

By the previous theorems, the two matrices $\mathbf{A}$ and $\mathbf{RA}$ have isomorphic row and column spaces, with equal dimensions.

So the row and column ranks of $\mathbf{A}$ are equal to the row and column ranks of $\mathbf{RA}$, both of which are r. $\square$

Outline

Two Equations in Two Unknowns Revisited

Consider once again the matrix equation $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix}$ with a, b, c, d all non-zero

In case $D = ad - bc \neq 0$, the coefficient matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and the augmented matrix $\begin{pmatrix} a & b & e \\ c & d & f \end{pmatrix}$ both have rank 2.

Then the two lines $ax + by = e$ and $cx + dy = f$ intersect,

There is a unique solution.

In case $D = 0$, the coefficient matrix has rank 1.

If the augmented matrix has rank 2,
the two lines are parallel and distinct, so there is no solution.

But if the augmented matrix has rank 1,
then the parallel lines coincide, so there are many solutions.

Rank Condition for Existence of a Solution, I

Theorem

Let $\mathbf{A}$ *be an* $m \times n$ *matrix, and* $\mathbf{b}$ *a column m-vector.*

Then the equation $\mathbf{Ax} = \mathbf{b}$ *has a solution* $\mathbf{x} \in \mathbb{R}^n$ *if and only the rank of the* $m \times (n+1)$ *augmented matrix* $(\mathbf{A}, \mathbf{b})$ *equals the rank of* $\mathbf{A}$*.*

Proof.

Necessity: Suppose that $\mathbf{Ax} = \mathbf{b}$ has a solution $\mathbf{x} = (x_j)_{j=1}^n$.

Now apply to $(\mathbf{A}, \mathbf{b})$ the compound column operation of successively subtracting from its last column the multiple x_j of each column j.

This converts $(\mathbf{A}, \mathbf{b})$ to $(\mathbf{A}, \mathbf{0})$ while preserving the column rank.

Hence the ranks of $(\mathbf{A}, \mathbf{b})$ and $(\mathbf{A}, \mathbf{0})$ are equal, with both equal to the rank of $\mathbf{A}$. □

Rank Condition for Existence of a Solution, II

Proof.

Sufficiency: Suppose the common rank of $\mathbf{A}$ and $(\mathbf{A}, \mathbf{b})$ is r.

Then there is an $r \times n$ submatrix $\tilde{\mathbf{A}}$
consisting of r linearly independent columns of $\mathbf{A}$.

Because the rank of $(\mathbf{A}, \mathbf{b})$ equals r, and not $r + 1$,
the $r + 1$ columns of $(\tilde{\mathbf{A}}, \mathbf{b})$ must be linearly independent.

This can only be true because there exists an r-vector $\tilde{\mathbf{x}}$
such that $\mathbf{b} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}$.

By augmenting $\tilde{\mathbf{x}}$ with $n - k$ appropriately placed zero elements,
one can construct $\mathbf{x} \in \mathbb{R}^n$ to satisfy $\mathbf{A}\mathbf{x} = \mathbf{b}$. □

Exercise

Let $\mathbf{A}$ *and* $\mathbf{B}$ *be* $m \times n$ *and* $m \times k$ *matrices.*

Prove that the matrix equation $\mathbf{A}\mathbf{X} = \mathbf{B}$
has one or more solutions for the $n \times k$ *matrix* $\mathbf{X}$ *if and only if*
both $\mathbf{A}$ *and the augmented matrix* $(\mathbf{A}, \mathbf{B})$ *have the same rank.*

Superfluous Equations and Degrees of Freedom, I

Theorem

Let $\mathbf{A}$ be an $m \times n$ matrix, and $\mathbf{b}$ a column m-vector.

Suppose $\mathbf{A}$ and the augmented matrix $(\mathbf{A}, \mathbf{b})$ have both rank r.

1. *If $r < m$, then $\mathbf{Ax} = \mathbf{b}$ has $m - r$ superfluous equations.*
2. *If $r < n$, then there are $n - r$ degrees of freedom in the solution to $\mathbf{Ax} = \mathbf{b}$.*

In the following proof, we assume that the $m \times n$ matrix $\mathbf{A}$ can be maximally diagonalized as

$$\mathbf{RA\hat{P}} = \begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix}$$

where $\mathbf{D}_{r \times r}^{-1}$ exists, while $\mathbf{R}$ is determinant preserving, and $\hat{\mathbf{P}}$ is a sign adjusted permutation.

Superfluous Equations and Degrees of Freedom, II

Proof.

Under the previous assumption, the equation system $\mathbf{Ax} = \mathbf{b}$ is equivalent to $\mathbf{RA\hat{P}z} = \mathbf{w}$ where $\mathbf{z} = \hat{\mathbf{P}}^{-1}\mathbf{x}$ and $\mathbf{w} = \mathbf{Rb}$.

This system can be written as

$$\begin{pmatrix} \mathbf{D}_{r\times r} & \mathbf{B}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix} \begin{pmatrix} \mathbf{z}_r^1 \\ \mathbf{z}_{n-r}^2 \end{pmatrix} = \begin{pmatrix} \mathbf{w}_r^1 \\ \mathbf{w}_{m-r}^2 \end{pmatrix}$$

Here the $m \times (n+1)$ matrix $(\mathbf{RA\hat{P}}, \mathbf{w})$ has rank r if and only if $\mathbf{w}_{m-r}^2 = \mathbf{0}_{m-r}$, in which case the last $m - r$ equations are superfluous.

Then, for each $\mathbf{z}_{n-r}^2 \in \mathbb{R}^{n-r}$ there is a unique solution given by $\mathbf{z}_r^1 = \mathbf{D}_{r\times r}^{-1}(\mathbf{w}_r^1 - \mathbf{B}_{r\times(n-r)}\mathbf{z}_{n-r}^2)$.

Hence there are $n - r$ degrees of freedom. □

Equation Systems: Existence of a Solution

Consider again the matrix equation $\mathbf{AX} = \mathbf{Y}$ in its equivalent form

$$\mathbf{RAX} = \mathbf{RAPP}^{-1}\mathbf{X} = \begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix} \mathbf{P}^{-1}\mathbf{X} = \mathbf{RY}$$

Introduce the partitioned matrix $\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}$ as notation for $\mathbf{Z} = \mathbf{P}^{-1}\mathbf{X}$,

where the $r \times p$ matrix $\mathbf{Z}_1$ consists of the first r rows of $\mathbf{Z}$,

and the $(n-r) \times p$ matrix $\mathbf{Z}_2$ consists of the other $n-r$ rows.

The equation system takes the form

$$\begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix} \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} = \mathbf{RY} = \begin{pmatrix} \mathbf{V}_{r \times p} \\ \mathbf{W}_{(m-r) \times p} \end{pmatrix}$$

Because the matrix $\mathbf{D}_{r \times r}$ of pivots is invertible,

and the last $m-r$ rows of the left-hand side matrix are all zero,

a solution exists if and only if $\mathbf{W}_{(m-r) \times p} = \mathbf{0}_{(m-r) \times p}$.

Equation Systems: The Solution Space

The necessary and sufficient condition for solutions to exist is $\mathbf{W}_{(m-r)\times p} = \mathbf{0}_{(m-r)\times p}$.

In case this is met, the system reduces to $\mathbf{D}\mathbf{Z}_1 + \mathbf{B}\mathbf{Z}_2 = \mathbf{R}\mathbf{Y}_1$.

The general solution is $\mathbf{Z}_1 = \mathbf{D}^{-1}(\mathbf{R}\mathbf{Y}_1 - \mathbf{B}\mathbf{Z}_2)$.

Because the $(n-r) \times p$ matrix $\mathbf{Z}_2$ can be chosen arbitrarily, there are $n-r$ degrees of freedom in each equation system.

The first r rows of the matrix $\mathbf{P}^{-1}\mathbf{X}$ with permuted columns have been expressed as a linear function of $\mathbf{Y}$ and of these last arbitrary $n-r$ rows of $\mathbf{P}^{-1}\mathbf{X}$.

The remaining $m-r$ equations are redundant.

Outline

Minors and Determinantal Rank

Definition

Given any $m \times n$ matrix $\mathbf{A}$, a minor (determinant) of order k is the determinant $|\mathbf{A}_{i_1 i_2 \ldots i_k,\ j_1 j_2 \ldots j_k}|$ of a $k \times k$ submatrix (a_{ij}), whose row numbers satisfy $1 \leq i_1 < i_2 < \ldots < i_k \leq m$ and whose column numbers satisfy $1 \leq j_1 < j_2 < \ldots < j_k \leq n$. □

The matrix $\mathbf{A}_{i_1 i_2 \ldots i_k,\ j_1 j_2 \ldots j_k}$, also denoted by $\mathbf{A}_{I \times J}$, is formed by selecting in the right order all the elements that lie in both:

- one of the k chosen rows in the ordered set $I := \langle i_r \rangle_{r=1}^k$;
- one of the k chosen columns in the ordered set $J := \langle j_s \rangle_{s=1}^k$.

Definition

The determinantal or minor rank of a matrix is the dimension of its largest non-zero minor determinant. □

Minors: Some Examples

Example

1. In case $\mathbf{A}$ is an $n \times n$ matrix:
   - the whole determinant $|\mathbf{A}|$ is the only minor of order n;
   - each of the n^2 cofactors $\mathbf{C}_{ij}$ is a minor of order $n - 1$.
2. In case $\mathbf{A}$ is an $m \times n$ matrix:
   - each element of the mn elements of the matrix is a minor of order 1;
   - the number of minors of order k is

$$\binom{m}{k} \cdot \binom{n}{k} = \frac{m!}{k!(m-k)!} \frac{n!}{k!(n-k)!}$$

Exercise

Verify that the set of elements that make up the minor $|\mathbf{A}_{i_1 i_2 \ldots i_k,\ j_1 j_2 \ldots j_k}|$ of order k is completely determined by its k diagonal elements a_{i_h, j_h} $(h = 1, 2, \ldots, k)$. (These need not be diagonal elements of $\mathbf{A}$).

Principal and Leading Principal Minors

Definition

If $\mathbf{A}$ is an $n \times n$ matrix,
the minor $|\mathbf{A}_{i_1 i_2 \ldots i_k,\ j_1 j_2 \ldots j_k}|$ of order k is:

- a principal minor if $i_h = j_h$ for $h = 1, 2, \ldots, k$,
  implying that its diagonal elements $a_{i_h j_h}$
  are all on the (principal) diagonal of $\mathbf{A}$;
- a leading principal minor if its diagonal elements
  are the leading elements
  of the (principal) diagonal of a_{hh} $(h = 1, 2, \ldots, k)$.

Exercise

Explain why an $n \times n$ determinant has:

1. $2^n - 1$ *principal minors;*
2. n *leading principal minors.*

A First Lemma

Lemma

Given the $m \times n$ *matrix* $\mathbf{A}$*,*
suppose that $|\mathbf{A}_{I \times J}|$ *is any non-zero minor of order* k*.*

Then both the set $\{\mathbf{a}_i^\top \mid i \in I\}$ *of rows of* $\mathbf{A}$
and the set $\{\mathbf{a}_j \mid j \in J\}$ *of columns of* $\mathbf{A}$ *are linearly independent.*

Corollary

Let r *denote the row rank of the* $m \times n$ *matrix* $\mathbf{A}$*,*
which equals its column rank.

Let d *denote the determinantal rank of the* $m \times n$ *matrix* $\mathbf{A}$*.*

Then $r \geq d$*.*

Proof.

There is a non-zero minor $|\mathbf{A}_{I \times J}|$ of order d, so $\# I = d$.

But then the same d rows of $\mathbf{A}$ are linearly dependent,
so the row rank $r \geq d$. □

Proof of First Lemma

Proof.

Suppose that the linear combination $\sum_{i \in I} \xi_i \, \mathbf{a}_i^\top$ of the set of rows $\{\mathbf{a}_i^\top \mid i \in I\}$ equals $\mathbf{0}_n$.

Then $\sum_{i \in I} \xi_i \, \mathbf{a}_{ij} = \mathbf{0}$ for every column $j \in \mathbb{N}_n$.

In particular, $\sum_{i \in I} \xi_i \, \mathbf{a}_{ij} = \mathbf{0}$ for every column $j \in J$.

So the linear combination $\sum_{i \in I} \xi_i \, \tilde{\mathbf{a}}_i^\top$ of the rows of the $k \times k$ matrix $\tilde{\mathbf{A}} = \mathbf{A}_{I \times J}$ is zero.

Since $|\mathbf{A}_{I \times J}| \neq 0$, these rows are linearly independent.

Hence $\sum_{i \in I} \xi_i \, \mathbf{a}_i^\top = \mathbf{0}_n$ implies that $\xi_i = 0$ for all $i \in I$.

Finally, this implies that the set $\{\mathbf{a}_i^\top \mid i \in I\}$ of rows of $\mathbf{A}$ is linearly independent.

To prove the corresponding result for columns, consider the transpose of each matrix. □

A Second Lemma

Lemma

Suppose that the $m \times n$ matrix $\mathbf{A}$ has row rank r.

Then there exist subsets $I \subseteq \mathbb{N}_m$ consisting of r rows and $J \subseteq \mathbb{N}_n$ consisting of r columns such that the minor $|\mathbf{A}_{I \times J}|$ of order r is non-zero.

Proof.

If $\mathbf{A}$ has row rank r,
then there exists a set $I \subseteq \mathbb{N}_m$ of r linearly independent rows.

These form an $r \times n$ submatrix $\mathbf{A}_{I \times \mathbb{N}_n}$ whose row rank is r.

Because row and column rank are equal,
it follows that $\mathbf{A}_{I \times \mathbb{N}_n}$ has column rank r, where $r \leq n$.

So $\mathbf{A}_{I \times \mathbb{N}_n}$ has a subset $J \subseteq \mathbb{N}_n$ of r linearly independent columns.

These form an $r \times r$ submatrix $\mathbf{A}_{I \times J}$ whose rank is r.

So $|\mathbf{A}_{I \times J}|$ is a non-zero minor of order r. □

Determinantal Rank: Theorem and Proof

Theorem

The determinantal rank d of any $m \times n$ matrix $\mathbf{A}$ equals both its row and column rank r.

Proof.

By the corollary to the first lemma, one has $r \geq d$.

But the second lemma implies that $d \geq r$.

Hence $d = r$. □