Lecture Notes: Matrix Algebra
Part D: Determinants, Inverses, and Rank

Peter J. Hammond

revised 2022 September 15th
typeset from `matrixAlgD23.tex`

Outline

Definition of Maximal Diagonalization

Definition

A maximal diagonalization of an $m \times n$ matrix $\mathbf{A}$ takes the form

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

where:

1. the integer $r \in \mathbb{Z}$ satisfying $0 \leq r \leq \min\{m, n\}$ is the diagonalization rank;
2. $\mathbf{D}_{r \times r}$ is an $r \times r$ diagonal matrix which is invertible because all its r diagonal elements are non-zero;
3. $\mathbf{R}$ is an invertible $m \times m$ matrix that represents a determinant preserving row operation;
4. $\hat{\mathbf{P}}$ is a sign adjusted invertible $n \times n$ permutation matrix that represents a determinant preserving column operation;
5. $\mathbf{B}_{r \times (n-r)}$ denotes an $r \times (n - r)$ matrix.

Four Special Cases

In case $0 < r < \min\{m, n\}$, the maximal diagonalization

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

of the $m \times n$ matrix $\mathbf{A}$ needs the full expression for the 2×2 partitioned matrix on the right-hand side.

Otherwise, there are four special cases when this partitioned matrix reduces to:

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

Our notation is intended to include all these special cases.

Existence Theorem

Theorem

Any $m \times n$ matrix $\mathbf{A}$ has a maximal diagonalization

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

The name "maximal diagonalization" is used
because the invertible diagonal matrix $\mathbf{D}_{r\times r}$ is as big as possible.

The next several slides offer a constructive proof.

The construction is based on a pivoting algorithm
that is a version of Gaussian elimination.

It is somewhat related to the "Doolittle algorithm"
which does a lot for square matrices!

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 $\mathbf{A}$.

Outline

Definition of Straightforward Pivoting, I

Let $\mathbf{A}$ be any $m \times n$ matrix, with a_{pq} in row p and column q where $1 < p < m$ and that $1 < q < n$.

In partitioned form, we have $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{<p,<q} & \mathbf{a}_{<p,q} & \mathbf{A}_{<p,>q} \\ \mathbf{a}^\top_{p,<q} & a_{pq} & \mathbf{a}^\top_{p,>q} \\ \mathbf{A}_{>p,<q} & \mathbf{a}_{>p,q} & \mathbf{A}_{>p,>q} \end{pmatrix}$.

If $a_{pq} \neq 0$, a determinant preserving row operation can zeroize many of the elements of column q by pivoting about a_{pq}:

- either downwards, to reach $\begin{pmatrix} \mathbf{A}_{<p,<q} & \mathbf{a}_{<p,q} & \mathbf{A}_{<p,>q} \\ \mathbf{a}^\top_{p,<q} & a_{pq} & \mathbf{a}^\top_{p,>q} \\ \mathbf{A}^\downarrow_{>p,<q} & \mathbf{0}_{>p,q} & \mathbf{A}^\downarrow_{>p,>q} \end{pmatrix}$;
- or upwards, to reach $\begin{pmatrix} \mathbf{A}^\uparrow_{<p,<q} & \mathbf{0}_{<p,q} & \mathbf{A}^\uparrow_{<p,>q} \\ \mathbf{a}^\top_{p,<q} & a_{pq} & \mathbf{a}^\top_{p,>q} \\ \mathbf{A}_{>p,<q} & \mathbf{a}_{>p,q} & \mathbf{A}_{>p,>q} \end{pmatrix}$.

Definition of Straightforward Pivoting, II

Starting from $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{<p,<q} & \mathbf{a}_{<p,q} & \mathbf{A}_{<p,>q} \\ \mathbf{a}^\top_{p,<q} & a_{pq} & \mathbf{a}^\top_{p,>q} \\ \mathbf{A}_{>p,<q} & \mathbf{a}_{>p,q} & \mathbf{A}_{>p,>q} \end{pmatrix}$ where $a_{pq} \neq 0$,

one pivots by adding $\alpha_i := -a_{iq}/a_{pq}$ times row p:

1. if pivoting downwards, to each row i with $p < i \leq m$,

   until one reaches $\mathbf{E}^\downarrow_{pq}\mathbf{A} = \begin{pmatrix} \mathbf{A}_{<p,<q} & \mathbf{a}_{<p,q} & \mathbf{A}_{<p,>q} \\ \mathbf{a}^\top_{p,<q} & a_{pq} & \mathbf{a}^\top_{p,>q} \\ \mathbf{A}^\downarrow_{>p,<q} & \mathbf{0}_{>p,q} & \mathbf{A}^\downarrow_{>p,>q} \end{pmatrix}$;

2. if pivoting upwards, to each row i with $1 \leq i < p$,

   until one reaches $\mathbf{E}^\uparrow_{pq}\mathbf{A} = \begin{pmatrix} \mathbf{A}^\uparrow_{<p,<q} & \mathbf{0}_{<p,q} & \mathbf{A}^\uparrow_{<p,>q} \\ \mathbf{a}^\top_{p,<q} & a_{pq} & \mathbf{a}^\top_{p,>q} \\ \mathbf{A}_{>p,<q} & \mathbf{a}_{>p,q} & \mathbf{A}_{>p,>q} \end{pmatrix}$.

Upper and Lower Triangular Pivot Matrices

The pre-multiplying matrices $\mathbf{E}^{\downarrow}$ and $\mathbf{E}^{\uparrow}$ involved
in downward and upward pivoting respectively are the products

$$\mathbf{E}^{\downarrow}_{pq} = \prod\nolimits_{i=p+1}^{m} \mathbf{E}_{i+\alpha_i p} \quad \text{and} \quad \mathbf{E}^{\uparrow}_{pq} = \prod\nolimits_{i=1}^{p-1} \mathbf{E}_{i+\alpha_i p}$$

of the relevant elementary row operations $\mathbf{E}_{i+\alpha_i p}$
which all focus on the pivot row p.

Recall that:

1. the matrix $\mathbf{E}_{i+\alpha_i p}$ is upper or lower triangular
according as the pivot row p is above or below row i;
2. the product of any pair of upper/lower triangular matrices
is upper/lower triangular.

It follows that downward pivoting results in a lower triangular $\mathbf{E}^{\downarrow}_{pq}$,
whereas upward pivoting results in an upper triangular $\mathbf{E}^{\uparrow}_{pq}$.

Straightforward Double Pivoting

In our earlier motivating example of 3 equations in 3 unknowns,
we allowed double pivoting
that combines both downward and upward pivoting.

Starting from $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{<p,<q} & \mathbf{a}_{<p,q} & \mathbf{A}_{<p,>q} \\ \mathbf{a}^\top_{p,<q} & a_{pq} & \mathbf{a}^\top_{p,>q} \\ \mathbf{A}_{>p,<q} & \mathbf{a}_{>p,q} & \mathbf{A}_{>p,>q} \end{pmatrix}$ where $a_{pq} \neq 0$,

double pivoting involves adding $\alpha_i := -a_{iq}/a_{pq}$ times row p
to each row i with $i \neq p$, resulting in the matrix

$$\mathbf{E}^{\updownarrow}_{pq}\mathbf{A} = \begin{pmatrix} \mathbf{A}^{\uparrow}_{<p,<q} & \mathbf{0}_{<p,q} & \mathbf{A}^{\uparrow}_{<p,>q} \\ \mathbf{a}^\top_{p,<q} & a_{pq} & \mathbf{a}^\top_{p,>q} \\ \mathbf{A}^{\downarrow}_{>p,<q} & \mathbf{0}_{>p,q} & \mathbf{A}^{\downarrow}_{>p,>q} \end{pmatrix}$$

Straightforward Double Pivoting, First Step

Start with any $m \times n$ matrix $\mathbf{A}^{(0)} = \mathbf{A}$.

Assuming that $a_{11} \neq 0$, we can pivot about a_{11} in order to go from $\mathbf{A}^{(0)} = \begin{pmatrix} a_{11} & \mathbf{A}_{1,>1} \\ \mathbf{A}_{>1,1} & \mathbf{A}_{>1,>1} \end{pmatrix}$ to $\mathbf{A}^{(1)} = \begin{pmatrix} a_{11} & \mathbf{A}^{(1)}_{1,>1} \\ \mathbf{0}_{>1,1} & \mathbf{A}^{(1)}_{>1,>1} \end{pmatrix}$,
a matrix whose only non-zero element in column 1 is a_{11}.

For each row $i > 1$, we must pre-multiply $\mathbf{A}^{(0)}$ by $\mathbf{E}_{i+\alpha_i 1}$,
the determinant-preserving elementary row operation
which adds $\alpha_{i1} = -a_{i1}/a_{11}$ times row 1 to row i.

The combination of all these $m-1$ row operations
is represented by the lower triangular downward pivot matrix

$$\mathbf{E}^{\downarrow}_{11} := \prod_{i=2}^{m} \mathbf{E}_{i+\alpha_{i1}1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \alpha_2 & 1 & 0 & \dots & 0 \\ \alpha_3 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_m & 0 & 0 & \dots & 1 \end{pmatrix}$$

Straightforward Double Pivoting, Start of Step k

For each $k \in \mathbb{N}$ with $1 < k < r := \min\{m, n\}$,
step k starts with the $m \times n$ matrix $\mathbf{A}^{(k-1)}$ which,
by induction on k, takes the form

$$\mathbf{A}^{(k-1)} = \begin{pmatrix} \mathbf{D}^{(k-1)}_{<k,<k} & \mathbf{a}^{(k-1)}_{<k,k} & \mathbf{A}^{(k-1)}_{<k,>k} \\ \mathbf{0}^\top_{k,<k} & a^{(k-1)}_{kk} & (\mathbf{a}^\top)^{(k-1)}_{k,>k} \\ \mathbf{0}_{>k,<k} & \mathbf{a}^{(k-1)}_{>k,k} & \mathbf{A}^{(k-1)}_{>k,>k} \end{pmatrix}$$

where $\mathbf{D}^{(k-1)}_{<k,<k}$ is a $(k-1) \times (k-1)$ diagonal matrix
whose diagonal elements are all non-zero.

Straightforward Double Pivoting, Step k

Assuming that $a_{kk}^{(k-1)} \neq 0$, we can pre-multiply

$$\mathbf{A}^{(k-1)} = \begin{pmatrix} \mathbf{D}_{<k,<k}^{(k-1)} & \mathbf{a}_{<k,k}^{(k-1)} & \mathbf{A}_{<k,>k}^{(k-1)} \\ \mathbf{0}_{k,<k}^{\top} & a_{kk}^{(k-1)} & (\mathbf{a}^{\top})_{k,>k}^{(k-1)} \\ \mathbf{0}_{>k,<k} & \mathbf{a}_{>k,k}^{(k-1)} & \mathbf{A}_{>k,>k}^{(k-1)} \end{pmatrix}$$

by the double pivot matrix $\mathbf{E}_{kk}^{\updownarrow}$ in order to arrive at

$$\mathbf{A}^{(k)} = \mathbf{E}_{kk}^{\updownarrow} \mathbf{A}^{(k-1)} = \begin{pmatrix} \mathbf{D}_{\leq k,\leq k}^{(k)} & \mathbf{a}_{\leq k,k+1}^{(k)} & \mathbf{A}_{\leq k,>k+1}^{(k)} \\ \mathbf{0}_{k+1,\leq k}^{\top} & a_{k+1,k+1}^{(k)} & (\mathbf{a}^{\top})_{k+1,>k+1}^{(k)} \\ \mathbf{0}_{>k+1,\leq k} & \mathbf{a}_{>k+1,k+1}^{(k)} & \mathbf{A}_{>k+1,>k+1}^{(k)} \end{pmatrix}$$

Here the $k \times k$ diagonal matrix $\mathbf{D}_{\leq k,\leq k}^{(k)}$
is the $(k-1) \times (k-1)$ diagonal matrix $\mathbf{D}_{<k,<k}^{(k-1)}$
with one extra non-zero diagonal element $a_{kk}^{(k-1)}$.

Conclusion of Straightforward Double Pivoting

Let $\ell := \min\{m, n\}$, and let $\mathbf{D}$ denote a diagonal matrix whose diagonal elements are always non-zero.

Provided a_{11} and then $a_{kk}^{(k-1)}$ $(k = 2, \ldots, \ell - 1)$ are all non-zero, pivoting can continue for $\ell - 1$ stages until k reaches ℓ.

1. In case $\ell = m = n$, pivoting ends with the $\ell \times \ell$ diagonal matrix $\mathbf{A}^{(\ell-1)} = \begin{pmatrix} \mathbf{D}^{(\ell-1)}_{<\ell,<\ell} & \mathbf{0}_{<\ell,\ell} \\ \mathbf{0}_{\ell,<\ell} & a^{(\ell-1)}_{\ell,\ell} \end{pmatrix} = \mathbf{D}_{\ell\times\ell}$.
2. In case $\ell = m < n$, pivoting ends with the $m \times n$ matrix
$$\mathbf{A}^{(\ell-1)} = \begin{pmatrix} \mathbf{D}^{(\ell-1)}_{<\ell,<\ell} & \mathbf{0}_{<\ell,\ell} & \mathbf{A}^{(\ell-1)}_{<\ell,>\ell} \\ \mathbf{0}_{\ell,<\ell} & a^{(\ell-1)}_{\ell,\ell} & (\mathbf{a}^\top)^{(\ell-1)}_{\ell,>\ell} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_{m\times m} & \mathbf{B}_{m\times(n-m)} \end{pmatrix}$$
3. In case $\ell = n < m$, pivoting ends with the $m \times n$ matrix
$$\mathbf{A}^{(\ell-1)} = \begin{pmatrix} \mathbf{D}^{(\ell-1)}_{<\ell,<\ell} & \mathbf{0}_{<\ell,\ell} \\ \mathbf{0}_{\ell,<\ell} & a^{(\ell-1)}_{\ell,\ell} \\ \mathbf{0}_{>\ell,<\ell} & \mathbf{0}^{(\ell-1)}_{>\ell,\ell} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_{n\times n} \\ \mathbf{0}_{(m-n)\times n} \end{pmatrix}$$

Outline

Adjusted Double Pivoting: The Matrix Sequence

Straightforward double pivoting works as long as all the successive pivot elements $a_{kk}^{(k-1)}$ ($k \in \mathbb{N}_{\ell-1}$) are non-zero.

Adjusted double pivoting allows for the possibility that any prospective pivot $a_{kk}^{(k-1)} = 0$.

The adjusted double pivoting process that lasts at least r steps will generate, for each $k \in \mathbb{N}_r$, an $m \times n$ matrix $\tilde{\mathbf{A}}^{(k)}$ that takes the partitioned form $\tilde{\mathbf{A}}^{(k)} = \begin{pmatrix} \tilde{\mathbf{D}}^{(k)}_{k \times k} & \tilde{\mathbf{A}}^{(k)}_{\leq k, > k} \\ \mathbf{0}_{(m-k) \times k} & \tilde{\mathbf{A}}^{(k)}_{>k, >k} \end{pmatrix}$.

Here $\tilde{\mathbf{D}}^{(k)}_{k \times k}$ is the diagonal matrix whose non-zero pth diagonal element, for each $p \leq k$, is the adjusted pivot $\tilde{a}_{pp}^{(p-1)}$ that was used at step p.

Adjusted Double Pivoting: When Not to Adjust

The next $(k+1)$th step starts from $\tilde{\mathbf{A}}^{(k)} = \begin{pmatrix} \tilde{\mathbf{D}}^{(k)}_{k\times k} & \tilde{\mathbf{A}}^{(k)}_{\leq k, >k} \\ \mathbf{0}_{(m-k)\times k} & \tilde{\mathbf{A}}^{(k)}_{>k,>k} \end{pmatrix}$.

Case 1: Suppose that the top left element $\tilde{a}^{(k)}_{k+1,k+1}$ of the submatrix $\tilde{\mathbf{A}}^{(k)}_{>k,>k}$ is non-zero.

In this case $\tilde{a}^{(k)}_{k+1,k+1}$ is the obvious possible pivot element for the next step $k+1$.

Pivot adjustment is optional, but not needed.

Adjusted Double Pivoting: The End

The $(k+1)$th step starts from $\tilde{\mathbf{A}}^{(k)} = \begin{pmatrix} \tilde{\mathbf{D}}^{(k)}_{k \times k} & \tilde{\mathbf{A}}^{(k)}_{\leq k, > k} \\ \mathbf{0}_{(m-k) \times k} & \tilde{\mathbf{A}}^{(k)}_{> k, > k} \end{pmatrix}$.

Case 2: If the bottom right submatrix $\tilde{\mathbf{A}}^{(k)}_{> k, > k} = \mathbf{0}_{(m-k) \times (n-k)}$,
the zero matrix of the right dimension,
then the $(k+1)$th pivot step is impossible.

All the r pivoting steps that are possible have been completed.

The final matrix takes the form $\begin{pmatrix} \mathbf{D}_{r \times r} & \mathbf{B}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix}$,
where:

1. $\mathbf{D}_{r \times r}$ is an $r \times r$ diagonal matrix
   whose diagonal elements are all non-zero;
2. $\mathbf{B}_{r \times (n-r)}$ is an arbitrary $r \times (n-r)$ matrix.

Sign Corrected Transposition Matrices

Suppose that $\mathbf{G}$ is any $m \times n$ matrix and that $r, s \in \mathbb{N}_m$.

Recall the notation $\hat{\mathbf{T}}^{rs}_{m \times m}$
for the $m \times m$ sign corrected transposition matrix
having the property that $\hat{\mathbf{T}}^{rs}_{m \times m}\mathbf{G}$ equals $\mathbf{G}$
with rows r and s interchanged,
and one of these rows multiplied by -1.

In case $r = s \leq m$,
the matrix $\hat{\mathbf{T}}^{rs}_{m \times m}$ becomes the identity matrix $\mathbf{I}_m$.

Recall too that the $n \times n$ sign corrected transposition matrix $\hat{\mathbf{T}}^{rs}_{n \times n}$
has the property that $\mathbf{G}\hat{\mathbf{T}}^{rs}_{n \times n}$
equals $\mathbf{G}$ with columns r and s interchanged,
and one of these columns multiplied by -1.

In case $r = s \leq n$,
the matrix $\hat{\mathbf{T}}^{rs}_{n \times n}$ becomes the identity matrix $\mathbf{I}_n$.

Adjusted Double Pivoting: How to Adjust

The $(k+1)$th step starts from $\tilde{\mathbf{A}}^{(k)} = \begin{pmatrix} \tilde{\mathbf{D}}^{(k)}_{k \times k} & \tilde{\mathbf{A}}^{(k)}_{\leq k, >k} \\ \mathbf{0}_{(m-k) \times k} & \tilde{\mathbf{A}}^{(k)}_{>k, >k} \end{pmatrix}$.

Case 3: Suppose that $\tilde{a}^{(k)}_{k+1,k+1} = 0$ but $\tilde{\mathbf{A}}^{(k)}_{>k,>k} \neq \mathbf{0}_{(m-k)\times(n-k)}$.

Then $\tilde{\mathbf{A}}^{(k)}_{>k,>k}$ has at least one non-zero element $\tilde{a}^{(k)}_{pq}$ with $k < p \leq m$ and $k < q \leq n$.

We adjust the pivot by applying two sign corrected transpositions:

1. first, we pre-multiply $\tilde{\mathbf{A}}^{(k)}$ by the $m \times m$ matrix $\hat{\mathbf{T}}^{p,k+1}_{m \times m}$, which swaps rows p and $k+1$;
2. then we post-multiply $\hat{\mathbf{T}}^{p,k+1}_{m \times m} \tilde{\mathbf{A}}^{(k)}$ by the $n \times n$ matrix $\hat{\mathbf{T}}^{q,k+1}_{n \times n}$, which swaps columns q and $k+1$.

Together, these two sign corrected transpositions move up and left the original non-zero element $\tilde{a}^{(k)}_{pq}$ in $\tilde{\mathbf{A}}^{(k)}$ to the $k+1, k+1$ position in the adjusted matrix $\hat{\mathbf{T}}^{p,k+1}_{m \times m} \tilde{\mathbf{A}}^{(k)} \hat{\mathbf{T}}^{q,k+1}_{n \times n}$.

Double Pivoting: The Adjusted Next Step

Recall that the $(k+1)$th step started

from $\tilde{\mathbf{A}}^{(k)} = \begin{pmatrix} \tilde{\mathbf{D}}^{(k)}_{k \times k} & \tilde{\mathbf{A}}^{(k)}_{\leq k, > k} \\ \mathbf{0}_{(m-k) \times k} & \tilde{\mathbf{A}}^{(k)}_{> k, > k} \end{pmatrix}$.

The adjustment that replaces $\tilde{\mathbf{A}}^{(k)}$ by $\hat{\mathbf{T}}^{p,k+1}_{m \times m} \tilde{\mathbf{A}}^{(k)} \hat{\mathbf{T}}^{q,k+1}_{n \times n}$
moves the element $\tilde{a}^{(k)}_{pq} \neq 0$ into position
so that it can replace $\tilde{a}^{(k)}_{k+1,k+1}$ as the next pivot element.

This adjustment allows the standard double pivoting row operation
that is represented by the matrix $\mathbf{E}^{\updownarrow}_{k+1,k+1}$
to be applied to the new version of the matrix $\tilde{\mathbf{A}}^{(k)}$.

The result of this $(k+1)$th pivot step is the next matrix $\tilde{\mathbf{A}}^{(k+1)}$.

The End of Double Pivoting

The double pivoting process can continue
through steps $k = 1, 2, \ldots, r$, until it reaches a terminal matrix.

After double pivoting is over, the following four cases are possible:

1. row exhaustion with $r = m < n$
   and terminal matrix $\tilde{\mathbf{A}}^{(m)} = \begin{pmatrix} \tilde{\mathbf{D}}^{(m)}_{m \times m} & \tilde{\mathbf{A}}^{(m)}_{m \times (n-m)} \end{pmatrix}$;
2. column exhaustion with $r = n < m$
   and terminal matrix $\tilde{\mathbf{A}}^{(n)} = \begin{pmatrix} \tilde{\mathbf{D}}^{(n)}_{n \times n} \\ \mathbf{0}_{(m-n) \times n} \end{pmatrix}$;
3. simultaneous row and column exhaustion with $r = m = n$
   and terminal matrix $\tilde{\mathbf{A}}^{(r)} = \tilde{\mathbf{D}}^{(r)}_{r \times r}$;
4. non-zero pivot exhaustion with $r < \min\{m, n\}$
   and terminal matrix $\tilde{\mathbf{A}}^{(r)} = \begin{pmatrix} \tilde{\mathbf{D}}^{(r)}_{r \times r} & \tilde{\mathbf{A}}^{(r)}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{pmatrix}$.

Together these four possible results of double pivoting
can all be summarized in one theorem, stated on the next slide.

Existence of a Maximal Diagonalization

We have completed a constructive proof of the following:

Theorem

Any $m \times n$ matrix $\mathbf{A}$ has a maximal diagonalization

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

where $r \in \mathbb{Z}$ satisfies $0 \leq r \leq \min\{m, n\}$.

Moreover, for each $k \in \mathbb{N}_r$, there exist two pivot adjustments in the form of sign corrected transposition matrices $\hat{\mathbf{T}}^{p,k}_{m\times m}$ and $\hat{\mathbf{T}}^{q,k}_{n\times n}$ with $k \leq p \leq m$ and $k \leq q \leq n$ such that $\mathbf{R} := \prod_{j=0}^{r-1} \left(\mathbf{E}^{\updownarrow}_{r-j,r-j} \hat{\mathbf{T}}^{p,r-j}_{m\times m} \right)$ and $\hat{\mathbf{P}} := \prod_{k=1}^{r} \hat{\mathbf{T}}^{q,k}_{n\times n}$ are well defined $m \times m$ and $n \times n$ matrix products which represent determinant preserving operations.

Conclusion

This result seems novel, but will not surprise most mathematicians.

It states that any matrix can be reduced to one
that is not too different from a diagonal matrix
whose diagonal elements are all non-zero.

This implies that the diagonal matrix
has an inverse that is trivial to find.

This plays a significant role
in simplifying much of our subsequent discussion.

Outline

Finding the Determinant of a Square Matrix

In the case of an $n \times n$ matrix $\mathbf{A}$, when $m = n$,
the maximal diagonalization takes the form

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

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

In case $r = n$ we have the complete diagonalization $\mathbf{R}\mathbf{A}\hat{\mathbf{P}} = \mathbf{D}$.

Because multiplication by $\mathbf{R}$ or $\hat{\mathbf{P}}$ is determinant preserving,
one has $|\mathbf{R}\mathbf{A}\hat{\mathbf{P}}| = |\mathbf{A}| = |\mathbf{D}| = \prod_{i=1}^{n} d_i$
where $\mathbf{D} = \mathbf{diag}(d_1, d_2, \ldots, d_n)$.

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

1. to pivot to reduce $\mathbf{A}$ to maximal diagonal form;
2. then multiply the resulting diagonal elements.

For $n > 3$, this is far more efficient
than trying to sum the $n!$ products of n elements.

Condition for a Square Matrix to Be Invertible

Theorem

Let $\mathbf{A}$ *be an* $n \times n$ *matrix, with maximally diagonalized form*

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

Then $\mathbf{A}$ *has an inverse matrix* $\mathbf{X}$ *satisfying* $\mathbf{AX} = \mathbf{XA} = \mathbf{I}_n$
if and only if $r = n$.

Proof.

Separate proofs of necessity and sufficiency,
which do not assume the product rule for determinants
(yet to be proved), appear on the next two slides. □

Remark

The condition $r = n$ *holds*
if and only if the maximally diagonalized matrix $\mathbf{R}\mathbf{A}\hat{\mathbf{P}}$
is a fully diagonalized and so invertible $n \times n$ *matrix* $\mathbf{D}$.

Condition for Invertibility: Proof of Necessity

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

Premultiplying the equation system $\mathbf{AX} = \mathbf{I}_n$
by the particular $n \times n$ matrix $\mathbf{R}$ used in pivoting
gives the equation $\mathbf{RAX} = \mathbf{RI}_n = \mathbf{R}$, or

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

Now, the equality $\begin{pmatrix} \mathbf{D}_{r\times r} & \mathbf{B}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(n-r)} \end{pmatrix} \mathbf{\hat{P}}^{-1}\mathbf{X} = \mathbf{Y}$ holds
for the $n \times n$ matrix $\mathbf{Y}$ only if each of its last $n - r$ rows equals $\mathbf{0}_n^\top$.

But $\mathbf{R}$ is determinant preserving, so $|\mathbf{R}| = 1$,
implying that none of its rows is $\mathbf{0}_n^\top$, so $r = n$.

It follows that $\mathbf{AX} = \mathbf{I}_n$, and so $\mathbf{RAX} = \mathbf{R}$, holds only if $r = n$.

That is, a necessary condition for $\mathbf{A}$ to be invertible
is that the pivot rank $r = n$. □

Condition for Invertibility: Proof of Sufficiency

Conversely, suppose that $r = n$, implying that $\mathbf{RA\hat{P}} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal $n \times n$ matrix with no zero diagonal elements, which is therefore invertible.

Recall that the matrix $\mathbf{R}$ that results from all the pivot operations is the combined product of determinant preserving row operations, all of which are invertible.

Furthermore, the sign corrected permutation matrix $\mathbf{\hat{P}}$ is invertible, as the combined product of sign corrected transposition matrices, all of which are invertible.

So the matrices $\mathbf{R}$ and $\mathbf{\hat{P}}$ are both invertible.

It follows that we can:
(i) premultiply $\mathbf{RA\hat{P}} = \mathbf{D}$ by $\mathbf{R}^{-1}$ to obtain $\mathbf{A\hat{P}} = \mathbf{R}^{-1}\mathbf{D}$;
(ii) then postmultiply the result by $\mathbf{\hat{P}}^{-1}$ to obtain $\mathbf{A} = \mathbf{R}^{-1}\mathbf{D}\mathbf{\hat{P}}^{-1}$.

But the three matrices $\mathbf{R}^{-1}$, $\mathbf{D}$, and $\mathbf{\hat{P}}^{-1}$ are all invertible.

It follows that $\mathbf{A}$ is invertible,
with inverse $\mathbf{A}^{-1} = (\mathbf{R}^{-1}\mathbf{D}\mathbf{\hat{P}}^{-1})^{-1} = \mathbf{\hat{P}}\mathbf{D}^{-1}\mathbf{R}$. □

Use Pivoting to Invert a Matrix!

Note that when $\mathbf{A}^{-1}$ exists,
pivoting to construct the three matrices $\hat{\mathbf{P}}, \mathbf{D}, \mathbf{R}$
does virtually all the work of matrix inversion.

Outline

Existence of the Inverse Matrix

We have just proved:

Theorem

An $n \times n$ matrix $\mathbf{A}$ has an inverse if and only if its determinant $|\mathbf{A}| \neq 0$.

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 $\mathbf{Ax} = \mathbf{b}$,
where the matrix $\mathbf{A}$ is $n \times n$.

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

Let $\mathbf{U}$ be any upper triangular square matrix.
Provided that $|\mathbf{U}| \neq 0$, the inverse $\mathbf{U}^{-1}$ exists, and is upper triangular.

Taking transposes leads immediately to:

Corollary

Let $\mathbf{L}$ be any lower triangular square matrix.
Provided that $|\mathbf{L}| \neq 0$, the inverse $\mathbf{L}^{-1}$ exists, and is lower triangular.

Inverting an Upper Triangular Matrix: Proof

Let $\mathbf{U}$ be any upper triangular $n \times n$ matrix
which, because $u_{ii} \neq 0$ for all $i \in \mathbb{N}_n$, satisfies $|\mathbf{U}| \neq 0$.

By the straightforward pivoting lemma,
the $n \times n$ upward pivoting matrix $\mathbf{R} := \mathbf{E}^{\nwarrow}$
is well defined, invertible, and upper triangular,
as well as representing a determinant preserving row operation.

Also, the result of the upward pivoting operation $\mathbf{R} = \mathbf{E}^{\nwarrow}$ on $\mathbf{U}$
is a diagonal matrix $\mathbf{D} = \mathbf{RU}$.

Then, because $\mathbf{R}$ is determinant preserving,
it follows that $|\mathbf{D}| = |\mathbf{U}| \neq 0$, which implies that $\mathbf{D}$ is invertible.

But $\mathbf{R}$ is invertible, so $\mathbf{RU} = \mathbf{D}$ implies that $\mathbf{U} = \mathbf{R}^{-1}\mathbf{D}$.

Because both $\mathbf{R}^{-1}$ and $\mathbf{D}$ are invertible,
it follows that $\mathbf{U}^{-1}$ exists and that $\mathbf{U}^{-1} = (\mathbf{R}^{-1}\mathbf{D})^{-1} = \mathbf{D}^{-1}\mathbf{R}$.

Finally, the inverse matrix $\mathbf{U}^{-1} = \mathbf{D}^{-1}\mathbf{R}$ is upper triangular
as the product of two upper triangular matrices. □

Outline

Verifying the Product Rule 7: Non-Singular Case

Proposition

If $\mathbf{A}$ *and* $\mathbf{B}$ *are* $n \times n$ *matrices with* $|\mathbf{A}| \neq 0$*, then* $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|$*.*

Proof.

Because $|\mathbf{A}| \neq 0$, its maximally diagonalized form is $\mathbf{R}\mathbf{A}\hat{\mathbf{P}} = \mathbf{D}$
where $\mathbf{R}$ is the product of determinant preserving row operations,
which is determinant preserving,
whereas $\hat{\mathbf{P}}$ is a sign adjusted permutation matrix,
and the diagonal matrix $\mathbf{D}$ is non-singular.

But the determinant preserving operations $\mathbf{R}$ and $\hat{\mathbf{P}}$
have inverses $\mathbf{R}^{-1}$ and $\hat{\mathbf{P}}^{-1}$ that are also determinant preserving.

So $\mathbf{A} = \mathbf{R}^{-1}\mathbf{D}\hat{\mathbf{P}}^{-1}$ and also $|\mathbf{A}| = |\mathbf{R}^{-1}(\mathbf{D}\hat{\mathbf{P}}^{-1})| = |\mathbf{D}\hat{\mathbf{P}}^{-1}| = |\mathbf{D}|$.

Then, because the product rule holds for the product $\mathbf{DC}$
when $\mathbf{D}$ is any diagonal matrix, for any $n \times n$ matrix $\mathbf{B}$ one has

$$|\mathbf{AB}| = |\mathbf{R}^{-1}\mathbf{D}\hat{\mathbf{P}}^{-1}\mathbf{B}| = |\mathbf{D}(\hat{\mathbf{P}}^{-1}\mathbf{B})| = |\mathbf{D}||\hat{\mathbf{P}}^{-1}\mathbf{B}| = |\mathbf{A}||\mathbf{B}| \quad \square$$

Verifying 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 an $n \times n$ 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 $\mathbf{RAB} = (\mathbf{RA}\hat{\mathbf{P}})\hat{\mathbf{P}}^{-1}\mathbf{B}$ 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 $|\mathbf{A}| = 0$.

Verifying 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

Definition of Cofactors

Definition

Given any element a_{ij} of the matrix $n \times n$ matrix $\mathbf{A}$,
the associated (i,j)-cofactor C_{ij} is the sign adjusted determinant
of the $(n-1) \times (n-1)$ matrix $\mathbf{A}_{-i,-j}$, as shown below.

$$C_{ij} = (-1)^{i+j} \begin{vmatrix} a_{11} & \dots & a_{1,j-1} & a_{1j} & a_{1,j+1} & \dots & a_{1n} \\ a_{21} & \dots & a_{2,j-1} & a_{2j} & a_{2,j+1} & \dots & a_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i1} & \dots & a_{i,j-1} & \boxed{a_{ij}} & a_{i,j+1} & \dots & a_{in} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{n,j-1} & a_{nj} & a_{n,j+1} & \dots & a_{nn} \end{vmatrix}$$

The box indicates the particular element a_{ij}.

The two blue lines it contains cross out row i and column j,
both of which have to be deleted from the matrix $\mathbf{A}$
before finding and adjusting the resulting determinant $|\mathbf{A}_{-i,-j}|$.

Cofactor Expansion: Theorem

Definition

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

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

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

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 cofactor expansion along any row r of the transposed matrix $\mathbf{A}^\top$ with its transposed cofactors $C_{rj}^\top$ gives $|\mathbf{A}^\top| = \sum_{j=1}^n a_{rj}^\top C_{rj}^\top$.

Taking transposes throughout gives $|\mathbf{A}| = \sum_{j=1}^n a_{jr} C_{jr}$.

Replacing j by i and then r by s, one obtains $|\mathbf{A}| = \sum_{i=1}^n a_{is} C_{is}$.

This is the formula for the cofactor expansion of $\mathbf{A}$ along column s.

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 prove 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 function is multilinear, it follows that

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

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

Therefore, to verify the formula $|\mathbf{A}| = \sum_{j=1}^n a_{rj} C_{rj}$
for the cofactor expansion of $\mathbf{A}$ along row any row r,
we show that, for each $j \in \mathbb{N}_n$, one has $|\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r| = C_{rj}$.

Cofactor Expansion: Proof, Part 3

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

whose:

1. top left hand corner is the $(n-1) \times (n-1)$ matrix $\mathbf{A}_{-r,-j}$ obtained by removing row r and column j from $\mathbf{A}$;
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 row vector $\mathbf{0}_{n-1}^\top$ 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 $\hat{\mathbf{A}}_{rj} = \mathbf{P}^{r \nearrow n} \mathbf{A} \mathbf{P}^{j \nearrow n}$;

(ii) $|\mathbf{A}_{-r}/\mathbf{e}_j^\top| = |\hat{\mathbf{A}}_{rj}|$; and (iii) $|\hat{\mathbf{A}}_{rj}| = C_{rj}$.

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

Cofactor Expansion: Proof, Part 4

Given $k \leq \ell \leq n$, recall that the permutation $\pi^{k \nearrow \ell} \in \Pi_n$ 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 matrix $\mathbf{P}^{\pi^{k \nearrow \ell}}$.

Lemma (A)

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

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

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 $\hat{\mathbf{A}}_{rj}$; (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 (B)

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

Proof.

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

In earlier results we showed that, for every permutation π, one has

$$|\mathbf{P}^\pi \mathbf{A}| = |\mathbf{A}\mathbf{P}^\pi| = \text{sgn}(\pi)|\mathbf{A}| \quad \text{and} \quad \text{sgn}(\pi^{k \nearrow \ell}) = (-1)^{\ell - k}$$

Hence we have $|\mathbf{P}^{r \nearrow n}[\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r]| = \text{sgn}(\pi^{r \nearrow n})|\mathbf{A}_{-r}/(\mathbf{e}_j^\top)_r|$
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

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

□

Cofactor Expansion: Proof, Part 6

Lemma (C)

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

Proof.

Note that the elements of row n satisfy $(\hat{\mathbf{A}}_{rj})_{n,\pi(n)} = \delta_{n,\pi(n)}$.

Using the determinant expansion formula along with the definitions of the bordered matrix $\hat{\mathbf{A}}_{rj}$ and cofactor C_{rj},
this special property of row n implies that

$$|\hat{\mathbf{A}}_{rj}| = \sum_{\pi \in \Pi_n} \prod_{i=1}^{n} (\hat{\mathbf{A}}_{rj})_{i,\pi(i)}$$

$$= \sum_{\pi \in \Pi_{n-1}} \prod_{i=1}^{n-1} (\hat{\mathbf{A}}_{rj})_{i,\pi(i)} = C_{rj}$$

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

Outline

Expansion by Alien Cofactors

Expanding by matching cofactors cofactors
along either row r or column s gives

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

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 (zero) determinant
resulting from the cofactor expansion
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

Given any $n \times n$ square matrix $\mathbf{A}$,
its adjugate (or "(classical) adjoint") matrix $\mathbf{adj}\,\mathbf{A}$
whose (i, j) element $(\mathbf{adj}\,\mathbf{A})_{ij}$,
for all $(i, j) \in \mathbb{N}_n \times \mathbb{N}_n$, is the cofactor C_{ji} of $\mathbf{A}$. □

The adjoint is therefore the transpose $(\mathbf{A}^+)^\top$
of the cofactor matrix $\mathbf{A}^+$
whose (i, j) element $(\mathbf{A}^+)_{ij} = C_{ij}$,
for all $(i, j) \in \mathbb{N}_n \times \mathbb{N}_n$, is the cofactor C_{ij} 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 first two matrices are respectively

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

These are cofactor expansions
along, first column j, and second row i, using:

- 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 $(i, j) \in \mathbb{N}_n \times \mathbb{N}_n$ one has

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

□

Outline

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 ith component is given by the ratio of determinants

$$x_i = \frac{|[\mathbf{A}_{-i}/\mathbf{b}]|}{|\mathbf{A}|}$$

This result is known as Cramer's rule.

Cramer's Rule: Proof

Proof.

Given the equation $\mathbf{Ax} = \mathbf{b}$, each cofactor $|\mathbf{C}_{ij}|$ equals the determinant of the $(n-1) \times (n-1)$ matrix formed by dropping row i and column j from the coefficient matrix $\mathbf{A}$.

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

By definition of the adjugate matrix $\mathbf{adj\,A}$, therefore, expanding the determinant by cofactors along column j gives

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

Hence the unique solution to the equation system has components

$$x_i = (\mathbf{A}^{-1}\mathbf{b})_i = \frac{1}{|\mathbf{A}|} \sum\nolimits_{j=1}^{n} (\mathbf{adj\,A})_{ij} b_j = \frac{1}{|\mathbf{A}|} |[\mathbf{A}_{-i}/\mathbf{b}]|$$

for $i = 1, 2, \ldots, n$. $\square$

Outline

Recall the Definition of a Block Diagonal Matrix

For each $i \in \mathbb{N}_k$, let $\mathbf{A}^{(i)}_{n_i \times n_i}$ be an $n_i \times n_i$ matrix.

The notation $\mathbf{diag}(\mathbf{A}^{(1)}_{n_1 \times n_1}, \ldots, \mathbf{A}^{(k)}_{n_k \times n_k})$ was introduced for the block diagonal matrix that results if the k matrices $\mathbf{A}^{(i)}_{n_i \times n_i}$ are arranged as blocks along the diagonal of a $k \times k$ square array, with all off diagonal blocks equal to the zero matrix.

That is

$$\mathbf{diag}(\mathbf{A}^{(1)}_{n_1 \times n_1}, \ldots, \mathbf{A}^{(k)}_{n_k \times n_k}) = \begin{pmatrix} \mathbf{A}^{(1)}_{n_1 \times n_1} & \mathbf{0}_{n_1 \times n_2} & \cdots & \mathbf{0}_{n_1 \times n_k} \\ \mathbf{0}_{n_2 \times n_1} & \mathbf{A}^{(2)}_{n_2 \times n_2} & \cdots & \mathbf{0}_{n_2 \times n_k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_k \times n_1} & \mathbf{0}_{n_k \times n_2} & \cdots & \mathbf{A}^{(k)}_{n_k \times n_k} \end{pmatrix}$$

The Determinant of a Block Diagonal Matrix

Theorem

The determinant $|\mathbf{diag}(\mathbf{A}^{(1)}_{n_1 \times n_1}, \ldots, \mathbf{A}^{(k)}_{n_k \times n_k})|$
of a block diagonal matrix
equals the product $\prod_{i=1}^{k} |\mathbf{A}^{(i)}_{n_i \times n_i}|$ *of the determinants*
of the matrices that form the k *blocks along the diagonal.*

To prove this theorem, it is enough:

1. first to prove that, if $\mathbf{X}$ and $\mathbf{Y}$ are both square matrices, then $\begin{vmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Y} \end{vmatrix} = |\mathbf{X}|\,|\mathbf{Y}|$;
2. second, to prove the result for $k > 2$ by induction.

Here I will only complete the first step as a lemma
and leave the second as an exercise.

Basic Lemma for a Block Lower Triangular Matrix, I

In fact, we confirm that $\begin{vmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{Z} & \mathbf{Y} \end{vmatrix} = \begin{vmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Y} \end{vmatrix} = |\mathbf{X}|\,|\mathbf{Y}|$
whenever $\mathbf{Z}$ is any matrix with appropriate dimensions.

Lemma

Given the block lower triangular matrix $\begin{pmatrix} \mathbf{X}_{m\times m} & \mathbf{0}_{m\times n} \\ \mathbf{Z}_{n\times m} & \mathbf{Y}_{n\times n} \end{pmatrix}$,
its determinant D satisfies $\begin{vmatrix} \mathbf{X}_{m\times m} & \mathbf{0}_{m\times n} \\ \mathbf{Z}_{n\times m} & \mathbf{Y}_{n\times n} \end{vmatrix} = |\mathbf{X}_{m\times m}|\,|\mathbf{Y}_{n\times n}|$.

The proof will be by induction on m.

First, suppose that $m = 1$, and that $\mathbf{X}_{1\times 1} = x$, whereas $\mathbf{Z}_{n\times 1} = \mathbf{z}_n$.

Then the cofactor expansion of $D = \begin{vmatrix} x & \mathbf{0}_n^\top \\ \mathbf{z}_n & \mathbf{Y}_{n\times n} \end{vmatrix}$
along the first row evidently reduces to $x|\mathbf{Y}_{n\times n}|$, as required.

Basic Lemma for a Block Lower Triangular Matrix, II

As the induction hypothesis, suppose that when $m = p - 1$, one has $D = \begin{vmatrix} \mathbf{X}_{m\times m} & \mathbf{0}_{m\times n} \\ \mathbf{Z}_{n\times m} & \mathbf{Y}_{n\times n} \end{vmatrix} = |\mathbf{X}_{m\times m}|\,|\mathbf{Y}_{n\times n}|$.

When $m = p$, the cofactor expansion of D along the first row reduces to $\sum_{j=1}^{p} x_{1j} C_{1j}$.

Each cofactor, after using the induction hypothesis, becomes

$$C_{1j} = \begin{vmatrix} \mathbf{X}^{-1,-j}_{(p-1)\times(p-1)} & \mathbf{0}_{(p-1)\times n} \\ \mathbf{Z}^{-j}_{n\times(p-1)} & \mathbf{Y}_{n\times n} \end{vmatrix} = |\mathbf{X}^{-1,-j}_{(p-1)\times(p-1)}|\,|\mathbf{Y}_{n\times n}|$$

where $\mathbf{X}^{-1,-j}_{(p-1)\times(p-1)}$ and $\mathbf{Z}^{-j}_{n\times(p-1)}$ denote the matrices $\mathbf{X}_{p\times p}$ and $\mathbf{Z}_{n\times p}$ with one indicated row and/or column omitted.

But $\sum_{j=1}^{p} x_{1j}|\mathbf{X}^{-1,-j}_{(p-1)\times(p-1)}|$ is the cofactor expansion of $|\mathbf{X}_{p\times p}|$, so $D = \sum_{j=1}^{p} x_{1j}|\mathbf{X}^{-1,-j}_{(p-1)\times(p-1)}|\,|\mathbf{Y}_{n\times n}| = |\mathbf{X}_{p\times p}|\,|\mathbf{Y}_{n\times n}|$.

This confirms the induction step. □

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, first consider the canonical basis of $\mathbb{R}^m$.

It is the list $(\mathbf{e}^j)_{j=1}^m$ of the m unit column vectors

$$\mathbf{e}^j = (e_i^j)_{i=1}^m = (\delta_{ij})_{i=1}^m \in \mathbb{R}^m$$

That is, each $\mathbf{e}^j$ equals the jth column of the identity matrix $\mathbf{I}_m$.

Now, if $\sum_{j=1}^m x_j \mathbf{e}^j = \mathbf{0}$,
then $\mathbf{x} = (x_j)_{j=1}^m$ satisfies $\mathbf{I}_m \mathbf{x} = \mathbf{0}$, implying that $\mathbf{x} = \mathbf{0}$.

So this list of m unit column m-vectors is linearly independent.

Linear Independence of Matrix Columns

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

Or equivalently, if and only if the only solution of $\mathbf{Ax} = \mathbf{0}_m$ in $\mathbb{R}^n$ is the trivial solution $\mathbf{x} = \mathbf{0}_n$.

Linear Dependence with Too Many Vectors, I

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

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

Let a maximal diagonalization of $\mathbf{Y}$ be

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

where $r \leq m \leq n$, and the matrices $\mathbf{R}$ and $\hat{\mathbf{P}}$ are both invertible.

Suppose that the n columns of $\mathbf{Y}$ are linearly independent because $\mathbf{Y}\mathbf{x} = \mathbf{0} \Longrightarrow \mathbf{x} = \mathbf{0}$ and so, with $\mathbf{x} = \hat{\mathbf{P}}\mathbf{z}$, one has

$$\mathbf{R}\mathbf{Y}\hat{\mathbf{P}}\mathbf{z} = \mathbf{0} \Longrightarrow \mathbf{R}\mathbf{Y}\mathbf{x} = \mathbf{0} \Longrightarrow \mathbf{Y}\mathbf{x} = \mathbf{R}^{-1}\mathbf{0}$$
$$\Longrightarrow \mathbf{Y}\mathbf{x} = \mathbf{0} \Longrightarrow \mathbf{x} = \mathbf{0} \Longrightarrow \mathbf{z} = \hat{\mathbf{P}}^{-1}\mathbf{x} = \mathbf{0}$$

Thus, the n columns of the $m \times n$ matrix $\mathbf{R}\mathbf{Y}\hat{\mathbf{P}}$ must be linearly independent.

Linear Dependence with Too Many Vectors, II

We have shown that the only solution $\mathbf{z}$ of the equation

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

must satisfy $\mathbf{z}_1 = \mathbf{0}_r$ and $\mathbf{z}_2 = \mathbf{0}_{n-r}$.

But then

$$\begin{aligned} \mathbf{R}\mathbf{Y}\hat{\mathbf{P}}\mathbf{z} = \mathbf{0}_n &\iff \mathbf{D}\mathbf{z}_1 + \mathbf{B}\mathbf{z}_2 = \mathbf{0}_n \\ &\iff \mathbf{z}_2 \in \mathbb{R}^{n-r} \text{ and } \mathbf{z}_1 = -\mathbf{D}^{-1}\mathbf{B}\mathbf{z}_2 \end{aligned}$$

So $\mathbf{R}\mathbf{Y}\hat{\mathbf{P}}\mathbf{z} = \mathbf{0}_n \implies \mathbf{z} = \mathbf{0}_n$ only if $\mathbb{R}^{n-r} = \{\mathbf{0}_{n-r}\}$ because $r = n$.

Yet $r \leq m \leq n$, so $r = n$ implies that $m = n$. □

Outline

Spanning

Definition

Let $S = \{\mathbf{x}^j \in \mathbb{R}^n \mid j \in \mathbb{N}_m\}$ be any finite set of m vectors in $\mathbb{R}^n$.

Then the set of vectors spanned by S, or the span of S, is

$$\operatorname{sp} S := \{\mathbf{z} \in \mathbb{R}^n \mid \exists \mathbf{y} = (\mathbf{y}_j)_{j=1}^m \in \mathbb{R}^m : \mathbf{z} = \sum\nolimits_{j=1}^m y_j \mathbf{x}^j\}$$

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

Exercise

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

The Column Space and Row Space

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

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

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

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

Column Rank and Row Rank

Definition

The column rank of the $m \times n$ matrix $\mathbf{A}$
is the dimension $r_{\text{col}} \leq n$ 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 $r_{\text{row}} \leq m$ 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

Suppose the $m \times n$ matrix $\mathbf{A}$ has the maximally diagonalized form

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

Then the column rank r_{col} of $\mathbf{R}\mathbf{A}\hat{\mathbf{P}}$ equals its pivot 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 evidently isomorphic to $\mathbb{R}^r$, whose dimension r equals the number of non-zero pivots in the diagonal matrix $\mathbf{D}_{r\times r}$. □

The Row Rank of a Maximally Diagonalized Matrix, I

Theorem

Suppose the $m \times n$ matrix $\mathbf{A}$ has the maximally diagonalized form

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

Then the row rank r_{row} of $\mathbf{R}\mathbf{A}\hat{\mathbf{P}}$ equals its pivot rank r.

Note first that, given this maximal diagonalization
and an arbitrary row vector $(\mathbf{z}^\top, \mathbf{w}^\top) \in \mathbb{R}^r \times \mathbb{R}^{m-r}$,
the equation $(\mathbf{x}^\top, \mathbf{y}^\top)\mathbf{R}\mathbf{A}\hat{\mathbf{P}} = (\mathbf{z}^\top, \mathbf{w}^\top)$ reduces to the equivalent
pair of vector equations $\mathbf{x}^\top\mathbf{D} = \mathbf{z}^\top$ and $\mathbf{x}^\top\mathbf{B} = \mathbf{w}^\top$.

The rest of the proof appears on the next slide.

Proof that Row Rank Equals Pivot Rank

Proof.

Given any $\mathbf{z}^\top \in \mathbb{R}^r$, because $\mathbf{D}$ is invertible, the pair of vector equations $\mathbf{x}^\top \mathbf{D} = \mathbf{z}^\top$ and $\mathbf{x}^\top \mathbf{B} = \mathbf{w}^\top$ has a non-empty solution set of pairs $(\mathbf{x}^\top, \mathbf{y}^\top) \in \mathbb{R}^m$ if and only if the unique solution $\mathbf{x}^\top = \mathbf{D}^{-1}\mathbf{z}^\top$ of the first equation satisfies the second equation $\mathbf{x}^\top \mathbf{B} = \mathbf{w}^\top$.

This is evidently true if and only if $\mathbf{w}^\top = \mathbf{z}^\top \mathbf{D}^{-1}\mathbf{B}$.

This proves that the row space is

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

Then the mapping $\mathbb{R}^r \ni \mathbf{z}^\top \leftrightarrow (\mathbf{z}^\top, \mathbf{z}^\top \mathbf{D}^{-1}\mathbf{B})$ is a linear bijection between $\mathbb{R}^r$, whose dimension is r, and the row space.

This establishes that the row space also has dimension r. □

Invariance of Row Space

Theorem

Let $\mathbf{A}$ *be any* $m \times n$ *matrix and* $\mathbf{R}$ *any* $m \times m$ *matrix representing a 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* $m \times m$ *matrix representing a 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 $\mathbf{R}$ preserves determinants, and $\hat{\mathbf{P}}$ is a 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 two previous theorems, the two matrices $\mathbf{A}$ and $\mathbf{RA}$ have identical row spaces and isomorphic column spaces, with equal dimensions.

So the respective row and column ranks of $\mathbf{A}$ are equal to the row and column ranks of $\mathbf{R}\mathbf{A}\hat{\mathbf{P}}$, both of which are r. □

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

Proof.

Necessity: Suppose that $\mathbf{A}\mathbf{x} = \mathbf{b}$ 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 $(\mathbf{A}, \mathbf{0})$ must be 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 dependent.

But the rank of $\tilde{\mathbf{A}}$ is r, so 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 $n - r$ 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

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

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

Superfluous Equations and Degrees of Freedom, II

Proof.

Under the stated supposition, the vector equation $\mathbf{A}\mathbf{x} = \mathbf{b}$
is equivalent to $\mathbf{R}\mathbf{A}\mathbf{x} = (\mathbf{R}\mathbf{A}\hat{\mathbf{P}})\hat{\mathbf{P}}^{-1}\mathbf{x} = \mathbf{R}\mathbf{b}$,
and so to $\mathbf{R}\mathbf{A}\hat{\mathbf{P}}\mathbf{z} = \mathbf{w}$ where $\mathbf{z} = \hat{\mathbf{P}}^{-1}\mathbf{x}$ and $\mathbf{w} = \mathbf{R}\mathbf{b}$.

Because $\mathbf{R}\mathbf{A}\hat{\mathbf{P}}$ is a maximal diagonalization,
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 $m \times (n+1)$ augmented matrix $(\mathbf{R}\mathbf{A}\hat{\mathbf{P}}, \mathbf{w})$ 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.

But 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

Let $\mathbf{A}$, $\mathbf{X}$ and $\mathbf{Y}$ be $m \times n$, $n \times p$, and $m \times p$ respectively.

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.

With this notation, 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

We are considering the reduced equation system

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

We have shown that the necessary and sufficient condition for solutions to exist is that $\mathbf{RY}_2 = \mathbf{W}_{(m-r)\times p} = \mathbf{0}_{(m-r)\times p}$.

In case this is met, the system reduces to $\mathbf{DZ}_1 + \mathbf{BZ}_2 = \mathbf{RY}_1$.

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

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

The first r rows of the matrix $\mathbf{Z} = \mathbf{P}^{-1}\mathbf{X}$,
which is $\mathbf{X}$ with permuted columns, have been expressed as a linear function of $\mathbf{Y}_1$, the first r rows of $\mathbf{Y}$,
and of $\mathbf{Z}_2$, the last arbitrary $n-r$ rows of $\mathbf{P}^{-1}\mathbf{X}$.

The remaining $m-r$ equations are redundant.

Outline

Minors of a Determinant: Definition

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 k 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 $\mathbf{A}_{i_1 i_2 \ldots i_k,\ j_1 j_2 \ldots j_k}$ will also be denoted by $\mathbf{A}_{I \times J}$.

The $k \times k$ matrix $\mathbf{A}_{I \times J}$ 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 **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 **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

In case **A** *is* $n \times n$*, 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 **A**.*)*

Principal Minors of a Square Matrix

Definition

Suppose that $\mathbf{A}$ is an $n \times n$ matrix.

Then 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 k 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 k leading elements $\langle a_{hh} \rangle_{h=1}^k$ of the (principal) diagonal of $\mathbf{A}$.

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 of Corollary.

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 satisfies $r \geq d$. □

Proof of First Lemma

Proof.

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

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

In particular, $\sum_{i \in I} \xi_i \, \mathbf{a}_{ij} = \mathbf{0}_n$ for all the k columns $j \in J$.

So the linear combination $\sum_{i \in I} \xi_i \, \tilde{\mathbf{a}}_i^\top$
of all the k 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 k rows of $\tilde{\mathbf{A}}$ are linearly independent.

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

This implies that the set $\{\mathbf{a}_i^\top \mid i \in I\}$ of k rows
selected from $\mathbf{A}$ 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 $d = r$. □