Section 4: Matrices, Determinants and Inverse Matrices

MATH241 Lecture Notes, notes pages 32–52

Notation. Scalars are plain italic (a, α, z), vectors are bold ($\boldsymbol{u}, \boldsymbol{x}$), and matrices are italic capitals (A, B; zero matrix O, identity I). The handwritten notes underline vectors once and matrices twice.

These notes are built on top of Orhan Arikan's linear algebra course notes.

notes page 32

Matrices and Linear Equations

($\star$ Chapter 10 in the textbook)

A *matrix* is a rectangular array of numbers. It has m rows and n columns, and its entries a_{ij} can be real or complex:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \quad n \text{ columns, } m \text{ rows}, \quad a_{ij} \in \mathbb{R} \text{ or } a_{ij} \in \mathbb{C}.$$

ex/ $A = \begin{bmatrix} 2 & 1 \\ 0 & 4 \\ -5 & 7 \end{bmatrix}$ is a real 3×2 matrix, and $B = \begin{bmatrix} 1+j & 2 \\ 3-j & 4+2j \end{bmatrix}$ is a complex 2×2 matrix.

Def'n Two matrices $A = \{a_{ij}\}_{m\times n}$ and $B = \{b_{ij}\}_{k\times l}$ are *equal* if they have the same size, $m = k$ and $n = l$, and all entries agree: $a_{ij} = b_{ij}$ for $1 \le i \le m$ and $1 \le j \le n$.

Def'n Matrices of the same size are added entry by entry. For $A = \{a_{ij}\}_{m\times n}$ and $B = \{b_{ij}\}_{m\times n}$,

$$C = A + B = \{a_{ij} + b_{ij}\}_{m\times n}.$$

ex/ Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 & -5 \\ -1 & 3 & 7 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Then

$$A + B = \begin{bmatrix} 2 & 4 & -2 \\ 3 & 8 & 13 \end{bmatrix},$$

but $A + C$ and $B + C$ are not defined, because C has a different size.

Def'n Multiplying a matrix by a scalar c multiplies every entry by c: for $A = \{a_{ij}\}_{m\times n}$, $cA = \{ca_{ij}\}_{m\times n}$, where $c \in \mathbb{R}$ (or $c \in \mathbb{C}$ for complex matrices).

ex/ For $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ we get $2A = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}$ and $(-1)A = -A = \begin{bmatrix} -1 & -2 & -3 \\ -4 & -5 & -6 \end{bmatrix}$.

Theorem 10.2.1 Properties of matrix addition and scalar multiplication

Addition and scalar multiplication of matrices obey the same rules as for numbers. If A, B and C are $m \times n$ matrices, O is the $m \times n$ zero matrix, and α, β are scalars, then:

(i) $A + B = B + A$ (commutativity).

(ii) $(A + B) + C = A + (B + C)$ (associativity).

(iii) $A + O = A$ (identity element of matrix addition).

—— notes page 33 ——

The remaining properties are:

(iv) $A + (-A) = O$ (inverse element of matrix addition).

(v) $\alpha(\beta A) = (\alpha\beta)A$ (associativity of scalar multiplication).

(vi) $(\alpha + \beta)A = \alpha A + \beta A$ (distributivity).

(vii) $\alpha(A + B) = \alpha A + \alpha B$ (distributivity).

(viii) $1 \cdot A = A$ (identity element of scalar multiplication).

(ix) $0A = O$.

(x) $\alpha O = O$.

Def'n Matrix multiplication Matrices can be multiplied when the number of columns of the first equals the number of rows of the second. If $A = \{a_{ij}\}_{m\times n}$ and $B = \{b_{ij}\}_{n\times p}$, their product is the $m \times p$ matrix $C = AB = \{c_{ij}\}_{m\times p}$ whose entry c_{ij} is the dot product of row i of A with column j of B:

Matrix multiplication

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}, \qquad 1 \le i \le m,\ 1 \le j \le p.$$

ex/ Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 & 4 \\ -1 & 0 & 5 \\ 1 & 3 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 2 & 11 & 17 \\ 5 & 26 & 47 \end{bmatrix}, \qquad CA = \begin{bmatrix} 13 & 17 & 21 \\ 18 & 24 & 30 \end{bmatrix}$$

but AC, BC, BA and CB are not defined, because the sizes do not match.

Note that even when both AB and BA exist, AB is not necessarily equal to BA. The order matters. If $AB = BA$, we say that A and B *commute* under multiplication.

The definition of matrix multiplication comes from linear systems. A system of m equations in n unknowns,

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= c_m \end{aligned}\right\}$$

is exactly the statement that a matrix times a column of unknowns equals a column of right-hand sides:

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{A\ m\times n} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\boldsymbol{x}\ n\times 1} = \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}}_{\boldsymbol{c}\ m\times 1}.$$

A linear system in one line

$$A\boldsymbol{x} = \boldsymbol{c}.$$

—— notes page 34 ——

Def'n For a square matrix $A = \{a_{ij}\}_{n\times n}$, the *power* A^p is the product of p copies of A:

$$A^p = \underbrace{A \cdot A \cdots A}_{p \text{ factors}}.$$

Def'n A square matrix $A = \{a_{ij}\}_{n\times n}$ is *diagonal* if $a_{ij} = 0$ for $i \neq j$, that is, if all its non-zero entries lie on the main diagonal.

ex/ $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix}$ is not diagonal, because it is not square. $B = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$ is a diagonal matrix.

Def'n The *identity matrix* is the diagonal matrix with $a_{ii} = 1$ on the diagonal.

ex/ $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the 2×2 identity matrix, and $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is the 3×3 identity matrix.

The identity matrix leaves any matrix unchanged: $A_{m\times n}I_n = A_{m\times n}$ and $I_m A_{m\times n} = A_{m\times n}$. This holds because in each entry of the product, only one term of the sum is non-zero, and it is multiplied by 1.

Theorem 10.2.2 Exceptional Properties of matrix multiplication:

Matrix multiplication differs from multiplication of numbers in four important ways, so some familiar rules fail:

(i) $AB \neq BA$ in general. For example, with $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ we get $AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ but $BA = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

(ii) Even if $AB = AC$, the matrices B and C need not be equal. For example, with $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$, both products equal $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

(iii) $AB = O$ does not imply that $A = O$ or $B = O$. For example, $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = O$.

(iv) $A^2 = I$ does not imply $A = I$ or $A = -I$. For example, $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ has $A^2 = I$.

Theorem 10.2.3 Ordinary properties of matrix multiplication

The rules that do hold are the ones we need for algebra:

(i) $(\alpha A)B = \alpha(AB) = A(\alpha B)$ (scalars can be moved around).

(ii) $(AB)C = A(BC)$ (associativity).

(iii) $(A + B)C = AC + BC$ (distributivity).

(iv) $C(A + B) = CA + CB$ (distributivity).

(v) $A(\alpha B + \beta C) = \alpha AB + \beta AC$ (linearity).

notes page 35

Partitioning in Matrix Operations

A large matrix can be cut by horizontal and vertical lines into smaller matrices called *blocks*, and the blocks can then be treated as if they were entries:

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & \cdots & B_{1q} \\ \vdots & & \vdots \\ B_{p1} & \cdots & B_{pq} \end{bmatrix}.$$

If $m = p$, $n = q$ and each block A_{ij} has the same size as the corresponding block B_{ij}, the sum is computed block by block:

$$A + B = \begin{bmatrix} A_{11} + B_{11} & \cdots & A_{1n} + B_{1n} \\ \vdots & & \vdots \\ A_{m1} + B_{m1} & \cdots & A_{mn} + B_{mn} \end{bmatrix}.$$

If $n = p$, the product $C = AB$ has blocks

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj},$$

provided that each product $A_{ik}B_{kj}$ is defined.

ex/ Cut two 3×3 matrices into blocks as shown by the lines:

$$A = \left[\begin{array}{cc|c} 2 & 4 & 1 \\ -1 & 3 & 0 \\ \hline 5 & 4 & 6 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \left[\begin{array}{cc|c} 0 & 1 & 3 \\ 2 & -4 & -1 \\ \hline 5 & 8 & 2 \end{array}\right] = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

The sum and the product are then computed block by block:

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{bmatrix} = \left[\begin{array}{cc|c} 2 & 5 & 4 \\ 1 & -1 & -1 \\ \hline 10 & 12 & 8 \end{array}\right].$$

$$\begin{aligned} AB &= \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} \\ &= \left[\begin{array}{cc|c} 13 & -6 & 4 \\ 6 & -13 & -6 \\ \hline 38 & 37 & 23 \end{array}\right]. \end{aligned}$$

Partitioning is especially helpful for sparse matrices, where most of the entries are zero and many blocks can be skipped.

notes page 36

Def'n The *transpose* of $A = \{a_{ij}\}_{m \times n}$ swaps its rows and columns:

$$A^T = \{a_{ij}^T\}_{n \times m} \quad \text{where} \quad a_{ij}^T = a_{ji}.$$

ex/ For $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ we get $A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$.

Theorem 10.3.1

Transposition behaves well with the other operations, except that it reverses the order of a product:

(i) $(A^T)^T = A$.

(ii) $(A+B)^T = A^T + B^T$.

(iii) $(\alpha A)^T = \alpha A^T$.

(iv) $(AB)^T = B^T A^T$, and so $(ABCD)^T = D^T C^T B^T A^T$.

Proof for (iv) Let $C = AB$, so $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$. The (i,j) entry of C^T is the (j,i) entry of C:

$$c_{ij}^T = c_{ji} = \sum_{k=1}^{n} a_{jk} b_{ki} = \sum_{k=1}^{n} a_{kj}^T b_{ik}^T = \sum_{k=1}^{n} b_{ik}^T a_{kj}^T \quad \Rightarrow C^T = B^T A^T. \quad \blacksquare$$

Note that for column vectors $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$, the matrix product of a row with a column is a dot product:

Dot product as a matrix product

$$\boldsymbol{x}^T \boldsymbol{y} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x} \cdot \boldsymbol{y}.$$

Def'n

(i) If $A^T = A$, the matrix A is called *symmetric*.

(ii) If $A^T = -A$, the matrix A is called *skew-symmetric*.

Every square matrix is the sum of a symmetric part and a skew-symmetric part:

$$A = \tfrac{1}{2}(A + A^T) + \tfrac{1}{2}(A - A^T).$$

The first term is symmetric, because $(A + A^T)^T = A^T + A$, and the second is skew-symmetric, because $(A - A^T)^T = A^T - A = -(A - A^T)$.

— notes page 37 —

Determinants

The determinant is a single number computed from a square matrix. Among other things, it tells us whether the matrix is invertible, as we will see later in this section. For a square matrix $A = \{a_{ij}\}_{n \times n}$ it is defined by *cofactor expansion* along a row:

Determinant by cofactor expansion

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{k=1}^{n} a_{ik} C_{ik} \qquad \text{(any fixed row } i\text{)},$$

where the *cofactor* is $C_{ik} = (-1)^{i+k} M_{ik}$.

The *minor* M_{ik} is the determinant of the $(n-1)\times(n-1)$ matrix that is left when row i and column k of A are removed. The determinant of a 1×1 or 2×2 matrix is the starting point of the recursion.

ex/ Small determinants.

(i) For a 1×1 matrix, $\det[a_{11}] = a_{11}$.

(ii) For a 2×2 matrix, expanding along the first row gives the familiar formula

$$\det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{bmatrix} = a_{11}(-1)^{1+1}|a_{22}| + a_{12}(-1)^{1+2}|a_{21}| = a_{11}a_{22} - a_{12}a_{21}.$$

(iii) For a 3×3 matrix, expanding along the first row reduces the problem to three 2×2 determinants:

$$\begin{aligned}\det\begin{bmatrix} 4 & 7 & -2 \\ 0 & 3 & 2 \\ 1 & 5 & 6\end{bmatrix} &= 4(-1)^{1+1}\begin{vmatrix} 3 & 2 \\ 5 & 6\end{vmatrix} + 7(-1)^{1+2}\begin{vmatrix} 0 & 2 \\ 1 & 6\end{vmatrix} \\ &\quad + (-2)(-1)^{1+3}\begin{vmatrix} 0 & 3 \\ 1 & 5\end{vmatrix} \\ &= 4\cdot 8 - 7\cdot(-2) - 2\cdot(-3) = 32 + 14 + 6 = 52.\end{aligned}$$

Comments

(i) The expansion works along any row, so it is best to choose the row with the most zero entries, since zero entries contribute nothing.

(ii) Cofactor expansion of large matrices is extremely time consuming: the number of operations for an $n\times n$ matrix grows like $N(n)\sim e\cdot n!$, which is enormous even for moderate n.

—— notes page 38 ——

Properties of Determinants

Row operations change a determinant in a simple, predictable way. This gives a much faster way to compute determinants than cofactor expansion:

> **D1.** Adding a multiple of one row to another does not change the determinant.

> **D2.** Interchanging two rows changes the sign of the determinant.

> **D3.** The determinant of a triangular matrix is the product of its diagonal entries.

example To compute the determinant of the matrix below, reduce it to triangular form using D1 and D2, and then multiply the diagonal. We first interchange the first and third rows (which flips the sign), then use $-2R_1 + R_2 \to R_2$ and $-\frac{2}{3}R_2 + R_3 \to R_3$.

$$\begin{aligned}\det\begin{bmatrix} 0 & 2 & -1 \\ 4 & 3 & 5 \\ 2 & 0 & -4\end{bmatrix} &= -\begin{vmatrix} 2 & 0 & -4 \\ 4 & 3 & 5 \\ 0 & 2 & -1\end{vmatrix} = -\begin{vmatrix} 2 & 0 & -4 \\ 0 & 3 & 13 \\ 0 & 2 & -1\end{vmatrix} \\ &= -\begin{vmatrix} 2 & 0 & -4 \\ 0 & 3 & 13 \\ 0 & 0 & -29/3\end{vmatrix} = -2\cdot 3\cdot(-29/3) = 58.\end{aligned}$$

Reducing to triangular form takes about $N_T(n)\sim \frac{2n^3}{3}$ operations, which is far fewer than $e\cdot n!$ when n is large.

Additional Properties of Determinants

> **D4.** If A has a row or a column of zeros, then $\det A = 0$.

> **D5.** If two rows (or two columns) are proportional, then $\det A = 0$.

> **D6.** If a row or column is a linear combination of the other rows or columns, then $\det A = 0$.

> **D7.** Multiplying every entry of one row or column by α multiplies the determinant by α.

> **D8.** $\det(\alpha A) = \alpha^n \det A$ for an $n \times n$ matrix, because all n rows are multiplied by α.

> **D9.** If A, B and C are identical except in one row (or column), and that row of C is the sum of the corresponding rows of A and B, then $\det C = \det A + \det B$.

> **D10.** $\det(A^T) = \det A$.

> **D11.** In general $\det(A + B) \neq \det A + \det B$. D9 is only a special case in which the sum of determinants does work.

> **D12.** $\det(AB) = (\det A)(\det B)$.

notes page 39

Row and Column Spaces

A matrix can be read in two ways: as a stack of rows, or as a row of columns. Write the rows of A as $\boldsymbol{r}_i^T$ (so each $\boldsymbol{r}_i \in \mathbb{R}^n$) and the columns as $\boldsymbol{a}_i$ (so each $\boldsymbol{a}_i \in \mathbb{R}^m$):

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_1^T \\ \boldsymbol{r}_2^T \\ \vdots \\ \boldsymbol{r}_m^T \end{bmatrix}, \quad \boldsymbol{r}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} \in \mathbb{R}^n.$$

$$A = \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{bmatrix} \quad \text{where } \boldsymbol{a}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix} \in \mathbb{R}^m.$$

Def'n The *row space* of A is the span of its rows, and the *column space* is the span of its columns:

$$\text{Row}(A) = \text{span}\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m\}, \qquad \text{Col}(A) = \text{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}.$$

Def'n The *row rank* is $\dim(\text{Row}(A))$ and the *column rank* is $\dim(\text{Col}(A))$.

ex/ Take the echelon matrix below, with rows $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_4$.

$$A = \begin{bmatrix} 1 & -3 & 2 & 5 & 3 \\ 0 & 0 & 1 & -4 & 2 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{r}_1 = \begin{bmatrix} 1 \\ -3 \\ 2 \\ 5 \\ 3 \end{bmatrix}, \ \boldsymbol{r}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -4 \\ 2 \end{bmatrix}, \ \boldsymbol{r}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 7 \end{bmatrix}, \ \boldsymbol{r}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The zero row adds nothing, so the row space is spanned by the three non-zero rows, and the row rank is 3:

$$\text{Row}(A) = \text{span}\{\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{r}_4\} = \text{span}\{\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3\}.$$

The columns that contain the pivots (the leading or pivot columns) are $\boldsymbol{a}_1, \boldsymbol{a}_3, \boldsymbol{a}_4$, and they span the column space, so the column rank is also 3:

$$\mathrm{Col}(A) = \mathrm{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4, \boldsymbol{a}_5\} = \mathrm{span}\{\boldsymbol{a}_1, \boldsymbol{a}_3, \boldsymbol{a}_4\}.$$

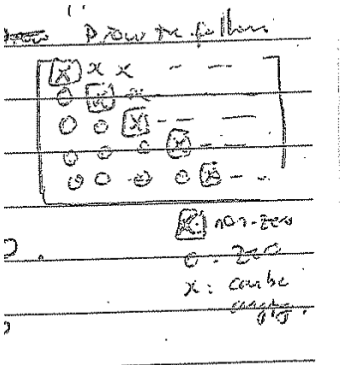
Row-equivalent matrices have the same row space
If two matrices are row equivalent, they have the same row space and the same row rank.

Proof If A is row equivalent to B, then B is obtained from A by a finite sequence of elementary row operations. Every elementary row operation replaces a row by a linear combination of rows, so every row of B is a linear combination of the rows of A. Hence $\mathrm{Row}(B) \subset \mathrm{Row}(A)$. Row operations can be undone by row operations, so A can also be obtained from B, and the same argument gives $\mathrm{Row}(A) \subset \mathrm{Row}(B)$. Therefore $\mathrm{Row}(A) = \mathrm{Row}(B)$. ■

notes page 40

Rows of an echelon matrix
The non-zero row vectors of an echelon matrix are linearly independent, so they form a basis for its row space.

Proof Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_k$ be the non-zero rows of an echelon matrix E, and let the first non-zero entry of $\boldsymbol{e}_i$ be in column p_i. Since E is in echelon form, these positions move to the right as we go down: $p_1 < p_2 < \cdots < p_k$. The picture shows the pattern, with the first non-zero entry of each row marked.



Suppose, for contradiction, that these rows are linearly dependent. Then there are coefficients $\alpha_1, \ldots, \alpha_k$, not all zero, such that

$$\alpha_1 \boldsymbol{e}_1 + \cdots + \alpha_k \boldsymbol{e}_k = \boldsymbol{0}.$$

Look at column p_1 of this equation:

$$\alpha_1 e_{1p_1} + \alpha_2 e_{2p_1} + \cdots + \alpha_k e_{kp_1} = 0.$$

All the rows below the first have a zero in column p_1, that is, $e_{2p_1} = e_{3p_1} = \cdots = e_{kp_1} = 0$, while $e_{1p_1} \neq 0$. Hence $\alpha_1 = 0$.

Now look at column p_2. The first row's coefficient is already zero, and the rows below the second have zeros in this column, so $\alpha_2 e_{2p_2} = 0$ and $\alpha_2 = 0$.

Continuing like this, we find $\alpha_i = 0$ for $1 \leq i \leq k$. This contradicts the assumption that not all the coefficients are zero, so the non-zero rows are linearly independent. ■

So the row space and the row rank of any matrix can be found by reducing it to echelon form with elementary row operations and counting the non-zero rows.

Now we turn to the column space and the column rank.

Column rank is preserved
The column rank of A is the same as the column rank of its row echelon form E.

Proof Consider the solutions of $A\boldsymbol{x} = \mathbf{0}$. Elementary row operations turn this into $E\boldsymbol{x} = \mathbf{0}$ without changing the solution set, so any dependence among the columns of A is also a dependence among the columns of E, and vice versa. The proof continues on the next page.

notes page 41

Writing this out in terms of the columns $\boldsymbol{a}_i$ of A and the columns $\boldsymbol{e}_i$ of E, we get

$$\begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n\end{bmatrix}\begin{bmatrix}\alpha_1\\ \vdots \\ \alpha_n\end{bmatrix} = \mathbf{0} \iff \begin{bmatrix}\boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n\end{bmatrix}\begin{bmatrix}\alpha_1\\ \vdots \\ \alpha_n\end{bmatrix} = \mathbf{0}.$$

So the columns of A and of E satisfy exactly the same dependence relations.

The pivot columns of the echelon matrix E span its column space and are linearly independent. (To see independence, suppose a combination of the pivot columns is zero. In the last pivot row only the last pivot column has a non-zero entry, so its coefficient must be zero; then the same argument applies to the previous pivot row, and so on, which is the same argument as in the theorem about the rows of an echelon matrix.) Denote the pivot columns by $\{\boldsymbol{e}_{p_1}, \ldots, \boldsymbol{e}_{p_r}\}$. Every other column $\boldsymbol{e}_k$ of E is a combination of these, so for every non-pivot column k,

$$\boldsymbol{e}_k = \sum_{i=1}^{r} c_{ki}\boldsymbol{e}_{p_i} \Rightarrow \boldsymbol{e}_k - \sum_{i=1}^{r} c_{ki}\boldsymbol{e}_{p_i} = \mathbf{0}.$$

Let $\boldsymbol{c}_k$ be the vector that has the entry c_{ki} in position p_i and zeros elsewhere. Then the relation above says

$$E\left(\begin{bmatrix}0\\ \vdots \\ 1\\ \vdots \\ 0\end{bmatrix} - \boldsymbol{c}_k\right) = \mathbf{0},$$

where the 1 is in the k^{th} position. By the equivalence of the two solution sets,

$$A\left(\begin{bmatrix}0\\ \vdots \\ 1\\ \vdots \\ 0\end{bmatrix} - \boldsymbol{c}_k\right) = \mathbf{0}.$$

This says that if a column $\boldsymbol{e}_k$ of E is a combination of the pivot columns of E, then the matching column $\boldsymbol{a}_k$ of A is the same combination of the matching columns $\{\boldsymbol{a}_{p_1}, \ldots, \boldsymbol{a}_{p_r}\}$ of A. Therefore

$$\operatorname{span}\{\boldsymbol{a}_{p_1}, \ldots, \boldsymbol{a}_{p_r}\} = \operatorname{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}.$$

These columns are also linearly independent. We show this by contradiction. Suppose some non-zero coefficients solve

$$\begin{bmatrix}\boldsymbol{a}_{p_1} & \boldsymbol{a}_{p_2} & \cdots & \boldsymbol{a}_{p_r}\end{bmatrix}\begin{bmatrix}\beta_{p_1}\\ \vdots \\ \beta_{p_r}\end{bmatrix} = \mathbf{0}.$$

Here p_i is the index of the i^{th} pivot column. This means $A\boldsymbol{z} = \mathbf{0}$ for the vector $\boldsymbol{z}$ with

$$z_i = \begin{cases} 0 & \text{if } i \notin \{p_1, \ldots, p_r\} \\ \beta_i & \text{if } i \in \{p_1, \ldots, p_r\}. \end{cases}$$

Since $A\boldsymbol{x} = \mathbf{0}$ and $E\boldsymbol{x} = \mathbf{0}$ have the same solutions, also $E\boldsymbol{z} = \mathbf{0}$, that is,

$$\begin{bmatrix} \boldsymbol{e}_{p_1} & \cdots & \boldsymbol{e}_{p_r} \end{bmatrix} \begin{bmatrix} \beta_{p_1} \\ \vdots \\ \beta_{p_r} \end{bmatrix} = \mathbf{0},$$

so $\{\boldsymbol{e}_{p_1}, \ldots, \boldsymbol{e}_{p_r}\}$ would be LD. This is a contradiction, because pivot columns are linearly independent. Thus $\{\boldsymbol{a}_{p_1}, \ldots, \boldsymbol{a}_{p_r}\}$ is LI and spans $\text{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$, so it is a basis for $\text{Col}(A)$.

—— notes page 42 ——

Hence $\dim(\text{Col}(A)) = r = \dim(\text{Col}(E))$. ■

We can now compare rows and columns. This is one of the central facts of linear algebra.

Row rank equals column rank
The column rank of a matrix A equals its row rank. In particular, for any matrix A the number of linearly independent rows equals the number of linearly independent columns. (Theorem 10.5.2)

Proof Let E be an echelon form of A. We have shown two things:

(i) $\text{Row}(A) = \text{Row}(E)$, so the row rank of A is the number of pivots of E.

(ii) The column rank of A equals the column rank of E, which is also the number of pivots of E.

Both ranks equal the number of pivots, so they are equal. ■

We call this common number the *rank* of A and write $r(A)$.

ex/ How many linearly independent vectors are there in

$$\left\{ \begin{bmatrix} 2 \\ 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ -2 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 3 \\ -2 \end{bmatrix} \right\} ?$$

Put the vectors in the columns of a matrix A and reduce it to echelon form:

$$\begin{bmatrix} 2 & 1 & -3 & 4 \\ 2 & 4 & -2 & 5 \\ 0 & 3 & 1 & 3 \\ 2 & 1 & -3 & -2 \end{bmatrix} \xrightarrow{R_1/2 \to R_1} \begin{bmatrix} 1 & 1/2 & -3/2 & 2 \\ 2 & 4 & -2 & 5 \\ 0 & 3 & 1 & 3 \\ 2 & 1 & -3 & -2 \end{bmatrix}$$

$$\xrightarrow[R_4 - 2R_1 \to R_4]{R_2 - 2R_1 \to R_2} \begin{bmatrix} 1 & 1/2 & -3/2 & 2 \\ 0 & 3 & 1 & 1 \\ 0 & 3 & 1 & 3 \\ 0 & 0 & 0 & -6 \end{bmatrix}$$

$$\xrightarrow{R_3 - R_2 \to R_3} \begin{bmatrix} 1 & 1/2 & -3/2 & 2 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -6 \end{bmatrix} \xrightarrow{R_4 + 3R_3 \to R_4} \begin{bmatrix} 1 & 1/2 & -3/2 & 2 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

There are three pivots, in columns 1, 2 and 4, so three of the vectors are linearly independent (the first, second and fourth). The third vector is a combination of the first two: it equals $-\frac{5}{3}$ of the first plus $\frac{1}{3}$ of the second.

ex/ Now consider a system of four equations in six unknowns:

$$\begin{bmatrix} 1 & -1 & 1 & 3 & 0 & 2 \\ 1 & 0 & 3 & 3 & -1 & 6 \\ 2 & -1 & 2 & 1 & -1 & 7 \\ 1 & 0 & 5 & 8 & -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 9 \\ 1 \end{bmatrix}.$$

Reduce the augmented matrix to echelon form:

$$\begin{bmatrix} 1 & -1 & 1 & 3 & 0 & 2 & 4 \\ 1 & 0 & 3 & 3 & -1 & 6 & 3 \\ 2 & -1 & 2 & 1 & -1 & 7 & 9 \\ 1 & 0 & 5 & 8 & -1 & 7 & 1 \end{bmatrix} \xrightarrow[\text{Row Operations}]{\text{Elementary}} \begin{bmatrix} 1 & -1 & 1 & 3 & 0 & 2 & 4 \\ 0 & 1 & 2 & 0 & -1 & 4 & -1 \\ 0 & 0 & 2 & 5 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here x_1, x_2, x_3 are the pivot variables and x_4, x_5, x_6 are the free variables.

— notes page 43 —

Observations

(i) The rank of A and the rank of the augmented matrix $[A : \boldsymbol{c}]$ are the same, so the system is consistent.

(ii) There are three free variables, so the solutions form a family with three parameters.

We give the free variables the names $x_6 = \alpha$, $x_5 = \beta$ and $x_4 = \gamma$, and solve for the pivot variables by back substitution:

$$\begin{aligned} x_3 &= -1 - \tfrac{1}{2}\alpha - \tfrac{5}{2}\gamma \\ x_2 &= 1 - 3\alpha + \beta + 5\gamma \\ x_1 &= 6 - \tfrac{9}{2}\alpha + \beta + \tfrac{9}{2}\gamma. \end{aligned}$$

In vector form this is a constant vector plus a combination of three direction vectors:

$$\boldsymbol{x} = \underbrace{\begin{bmatrix} 6 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_0} + \alpha \underbrace{\begin{bmatrix} -9/2 \\ -3 \\ -1/2 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{x}_1} + \beta \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_2} + \gamma \underbrace{\begin{bmatrix} 9/2 \\ 5 \\ -5/2 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_3}.$$

Observe that:

(i) $\boldsymbol{x}_0$ is a particular solution of $A\boldsymbol{x} = \boldsymbol{c}$ (take $\alpha = \beta = \gamma = 0$).

(ii) $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ solve the homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$. Indeed, for every α, β, γ,

$$\begin{aligned} A\boldsymbol{x} &= A(\boldsymbol{x}_0 + \alpha\boldsymbol{x}_1 + \beta\boldsymbol{x}_2 + \gamma\boldsymbol{x}_3) \\ &= \underbrace{A\boldsymbol{x}_0}_{\boldsymbol{c}} + \alpha A\boldsymbol{x}_1 + \beta A\boldsymbol{x}_2 + \gamma A\boldsymbol{x}_3 = \boldsymbol{c}, \end{aligned}$$

so $\alpha A\boldsymbol{x}_1 + \beta A\boldsymbol{x}_2 + \gamma A\boldsymbol{x}_3 = \boldsymbol{0}$ for all α, β, γ, and therefore $A\boldsymbol{x}_i = \boldsymbol{0}$ for $1 \le i \le 3$.

(iii) The matrix of these three vectors has rank 3 (its last three rows contain an identity matrix), so $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are linearly independent:

$$\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 \end{bmatrix} = \begin{bmatrix} -9/2 & 1 & 9/2 \\ -3 & 1 & 5 \\ -1/2 & 0 & -5/2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

This example illustrates a general fact. For $A\boldsymbol{x} = \boldsymbol{c}$ with $A \in \mathbb{R}^{m\times n}$ (or $\mathbb{C}^{m\times n}$), writing $r(A)$ for the rank:

How many solutions?

(i) If $r(A) < r([A : \boldsymbol{c}])$, there are no solutions.

(ii) If $r(A) = r([A : \boldsymbol{c}])$, there is at least one solution. If moreover $n = r(A)$, the solution is unique. If $n > r(A)$, the solution is non-unique, and the solutions form a family with $n - r(A)$ parameters.

notes page 44

(iii) If the system is consistent and $n > r(A)$, every solution can be written as

Particular plus homogeneous solution

$$\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n,$$

where $\boldsymbol{x}_p$ is one particular solution of $A\boldsymbol{x} = \boldsymbol{c}$ and $\boldsymbol{x}_n$ solves the homogeneous system $A\boldsymbol{x}_n = \boldsymbol{0}$.

Def'n The set of all solutions of $A\boldsymbol{x} = \boldsymbol{0}$ is called the *null space* of A.

The number of pivots and the size of the null space always add up to the number of unknowns:

Rank–nullity theorem
$r(A) + \dim(\mathrm{Null}(A)) = n.$

Proof Consider $A\boldsymbol{x} = \boldsymbol{0}$ and reduce A to an echelon matrix E to find the pivots and the free variables. The rank $r(A)$ is the number of pivot variables, and $\dim \mathrm{Null}(A)$ is the number of free variables, since each free variable contributes one independent direction. Every one of the n unknowns is either a pivot variable or a free variable, so

$$\underbrace{\#\text{pivots}}_{r(A)} + \underbrace{\#\text{free variables}}_{\dim(\mathrm{Null}(A))} = n. \quad \blacksquare$$

Inverse Matrix; Cramer's Rule, Factorization

To solve the scalar equation $ax = c$ we divide by a, which is the same as multiplying by a^{-1}, as long as $a \neq 0$:

$$\underbrace{a^{-1}a}_{1} x = a^{-1}c \Rightarrow x = a^{-1}c.$$

We would like to do the same for $A\boldsymbol{x} = \boldsymbol{c}$: multiply both sides by a suitable matrix A^{-1} such that

$$\underbrace{A^{-1}A}_{\text{we want this to be } I} \qquad \boldsymbol{x} = A^{-1}\boldsymbol{c}.$$

But we have to be careful. For example, the matrix below has a left inverse, a matrix that multiplies it to give the identity:

$$\begin{bmatrix} -1 & -2 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and note that } \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} -1 & -2 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

notes page 45

Here is what goes wrong. We might multiply both sides of the system by the matrix $\begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \end{bmatrix}$:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

and conclude that $x_1 = 3$, $x_2 = 4$ is a solution. But it is not. The new system is not equivalent to the original one: the original system has no solution at all, and multiplying by a matrix can destroy that information. A matrix that is not square, or not of full rank, cannot serve as an inverse in this way, as the next theorem shows.

Inverses need full-rank square matrices
Consider $A\boldsymbol{x} = \boldsymbol{c}$. If A is not a full-rank square matrix, there is no matrix A^{-1} with $A^{-1}A = I$ such that $A\boldsymbol{x} = \boldsymbol{c}$ and $(A^{-1}A)\boldsymbol{x} = A^{-1}\boldsymbol{c}$ have the same solutions for every $\boldsymbol{c}$.

Proof Suppose such an A^{-1} existed for a matrix A that is not a full-rank square matrix. We show that this leads to a contradiction. There are two cases to examine.

(i) Suppose $A\boldsymbol{x} = \boldsymbol{c}$ is inconsistent for some $\boldsymbol{c}$, so its solution set is empty. But $(A^{-1}A)\boldsymbol{x} = A^{-1}\boldsymbol{c}$ reads $\boldsymbol{x} = A^{-1}\boldsymbol{c}$ and always has exactly one solution. The two systems would then have different solution sets. So $A\boldsymbol{x} = \boldsymbol{c}$ must be consistent for all $\boldsymbol{c}$, which means that A has full row rank, $r(A_{m\times n}) = m$.

(ii) Suppose $A\boldsymbol{x} = \boldsymbol{c}$ has more than one solution for some $\boldsymbol{c}$. Again $(A^{-1}A)\boldsymbol{x} = A^{-1}\boldsymbol{c}$ has the single solution $\boldsymbol{x} = A^{-1}\boldsymbol{c}$, so the solution sets would differ. Hence $A\boldsymbol{x} = \boldsymbol{c}$ can never have more than one solution, which means that A has full column rank, $r(A_{m\times n}) = n$.

Combining the two cases gives $r(A_{m\times n}) = m = n$, so A is square and of full rank. This contradicts the assumption. ■

notes page 46

The converse is also true: a full-rank square matrix has an inverse.

Existence of the inverse
If A is a square matrix of full rank, there exists a matrix A^{-1} with $A^{-1}A = I$.

Proof If $A^{-1}A = I$, then transposing gives $A^T(A^{-1})^T = I$. We show that this transposed equation has a unique solution. Let $(A^{-1})^T = B$; if we can find B, then $A^{-1} = B^T$. The equation $A^TB = I$ can be written column by column, where $\boldsymbol{e}_i$ is the i^{th} column of the identity (a 1 in position i, zeros elsewhere):

$$A^T \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \end{bmatrix}, \quad \boldsymbol{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This is n separate linear systems:

$$A^T \boldsymbol{b}_i = \boldsymbol{e}_i, \qquad 1 \le i \le n.$$

Since A is square and of full rank, so is A^T. Therefore each system $A^T\boldsymbol{b}_i = \boldsymbol{e}_i$ is consistent and has a unique solution. So there is exactly one B with $A^T B = I$. Then $A^{-1} = B^T$ exists, because $B^T A = (A^T B)^T = I$, and it is unique. ■

Def'n A full-rank square matrix is called *invertible* (or non-singular), and its unique inverse is written A^{-1}.

So far we only know $A^{-1}A = I$. The next theorem shows that the inverse also works from the other side.

Two-sided inverse
If A is invertible, then $A^{-1}A = AA^{-1} = I$.

Proof We have shown that $A^{-1}A = I$, so it remains to show that AA^{-1} is also I. Multiply $A^{-1}A = I$ on the left by A:

$$A^{-1}A = I \Rightarrow AA^{-1}A = A \Rightarrow (AA^{-1})A = A.$$

notes page 47

So $(AA^{-1})A\boldsymbol{x} = A\boldsymbol{x}$ for every $\boldsymbol{x}$. Since A has full rank, the vectors $A\boldsymbol{x}$ fill all of $\mathbb{R}^n$, so

$$(AA^{-1})\boldsymbol{z} = \boldsymbol{z} \quad \text{for every } \boldsymbol{z} \text{ in } \mathbb{R}^n.$$

Hence $(AA^{-1} - I)\boldsymbol{z} = \boldsymbol{0}$ for every $\boldsymbol{z}$. Taking $\boldsymbol{z} = \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ shows that all the entries of $AA^{-1} - I$ are zero. Thus $AA^{-1} = I$. ■

Def'n Negative powers are powers of the inverse: $A^{-n} \equiv (A^{-1})^n = \underbrace{(A^{-1})\cdots(A^{-1})}_{n \text{ times}}$.

Theorem Algebra of inverse matrices:

Inverses combine with the other operations as follows:

Rules for inverses

(a) $(A^{-1})^{-1} = A$.

(b) $(A^n)^{-1} = (A^{-1})^n$.

(c) $(AB)^{-1} = B^{-1}A^{-1}$, if both A and B are invertible (note the reversed order).

(d) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

Proof

(a) $A^{-1}A = I$ and $AA^{-1} = I$, so A is the inverse of A^{-1}.

(b) This follows from (c) and the definition of the power:

$$A^n = \underbrace{AA\cdots A}_{n \text{ times}} \Rightarrow (A^n)^{-1} = [AA\cdots A]^{-1} = \underbrace{(A^{-1})(A^{-1})\cdots(A^{-1})}_{n \text{ times}} \equiv A^{-n} = (A^{-1})^n.$$

(c) We check both products with the claimed inverse $B^{-1}A^{-1}$:

$$AB(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I,$$
$$(B^{-1}A^{-1})AB = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$

(d) $A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$ and $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$.

Theorem Inverses solve linear systems directly.

Solving with the inverse
If A is invertible, the unique solution of $A\boldsymbol{x} = \boldsymbol{c}$ is $\boldsymbol{x} = A^{-1}\boldsymbol{c}$.

Proof $A(A^{-1}\boldsymbol{c}) = (AA^{-1})\boldsymbol{c} = \boldsymbol{c}$, so $A^{-1}\boldsymbol{c}$ is a solution. It is the only one, because A has full column rank (it is invertible), so all n columns are pivot columns. ■

notes page 48

Determinants and Invertibility

The determinant gives a quick test for invertibility.

Determinant test
A is invertible if and only if $\det A \neq 0$.

ex/ For $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ we have $\det A = 4 - 6 = -2 \neq 0$, so A is invertible.

Proof Let E be the echelon form obtained from A by Gauss elimination. Row interchanges and adding multiples of one row to another change the determinant at most by a sign, so $\det A = \pm \det E$. In particular, $\det A \neq 0$ exactly when $\det E \neq 0$. Since E is triangular, $\det E$ is the product of its diagonal entries, the pivots. Now we prove both directions.

(i) If A is invertible, it has full rank, so E has full rank: there are n pivots, all non-zero. Then $\det(E) \neq 0$, and so $\det(A) \neq 0$.

(ii) If $\det A \neq 0$, then $\det(E) \neq 0$, so all the pivots are non-zero and E has full rank. Hence A has full rank, and A is invertible.

Theorem Cramer's Rule

For a square system with $\det A \neq 0$ there is an explicit formula for each unknown, in terms of determinants. Consider the $n \times n$ system $A\boldsymbol{x} = \boldsymbol{b}$ with $A = \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{bmatrix}$ written by its columns.

Cramer's rule
If $\det A \neq 0$, the system is uniquely solvable, and the i^{th} entry of the solution is

$$x_i = \frac{\det \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_{i-1} & \boldsymbol{b} & \boldsymbol{a}_{i+1} & \cdots & \boldsymbol{a}_n \end{bmatrix}}{\det A},$$

that is, the determinant of A with its i^{th} column replaced by $\boldsymbol{b}$, divided by $\det A$.

Proof Since $\det A \neq 0$, the matrix A is invertible, so the system has a unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. We now find the entries of $\boldsymbol{x}$.

notes page 49

Let $\boldsymbol{x}$ be the unique solution. Writing the product $A\boldsymbol{x}$ as a combination of the columns of A shows that this combination equals $\boldsymbol{b}$:

$$A\boldsymbol{x} = \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \underbrace{\boldsymbol{a}_1 x_1 + \cdots + \boldsymbol{a}_n x_n}_{= \boldsymbol{b}} = \sum_{k=1}^{n} \boldsymbol{a}_k x_k.$$

Now replace the i^{th} column of A by $\boldsymbol{b}$, and substitute this combination for $\boldsymbol{b}$. Because the determinant is additive and homogeneous in a single column (D9 and D7), it splits into a sum of n determinants. In every term with $k \neq i$ the column $\boldsymbol{a}_k$ appears twice, so that determinant is zero (D5). Only the term $k = i$ survives, and it is $x_i \det A$:

$$\begin{aligned}
&\det\begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_{i-1} & \boldsymbol{b} & \boldsymbol{a}_{i+1} & \cdots & \boldsymbol{a}_n\end{bmatrix} \\
&= \det\begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_{i-1} & \sum_{k=1}^{n} \boldsymbol{a}_k x_k & \boldsymbol{a}_{i+1} & \cdots & \boldsymbol{a}_n\end{bmatrix} \\
&= \sum_{k=1}^{n} x_k \underbrace{\det\begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_{i-1} & \boldsymbol{a}_k & \boldsymbol{a}_{i+1} & \cdots & \boldsymbol{a}_n\end{bmatrix}}_{0 \text{ if } k\neq i;\ \det A \text{ if } k=i}.
\end{aligned}$$

Dividing by $\det A$ gives the formula:

$$\Rightarrow \det\begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_{i-1} & \boldsymbol{b} & \boldsymbol{a}_{i+1} & \cdots & \boldsymbol{a}_n\end{bmatrix} = x_i \det A \quad \Rightarrow \quad x_i = \frac{\det\begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_{i-1} & \boldsymbol{b} & \boldsymbol{a}_{i+1} & \cdots & \boldsymbol{a}_n\end{bmatrix}}{\det A}. \quad \blacksquare$$

Cramer's rule gives a closed-form expression for $\boldsymbol{x}$, which is useful in analytical investigations.

ex/ Use Cramer's rule to solve

$$\begin{aligned}
x_1 + 4x_2 + 5x_3 &= 2 \\
4x_1 + 2x_2 + 5x_3 &= 3 \\
-3x_1 + 3x_2 - x_3 &= 1.
\end{aligned}$$

First compute $\det A$ by reducing to triangular form:

$$\det A = \begin{vmatrix} 1 & 4 & 5 \\ 4 & 2 & 5 \\ -3 & 3 & -1 \end{vmatrix} = \det\begin{bmatrix} 1 & 4 & 5 \\ 0 & -14 & -15 \\ 0 & 15 & 14 \end{bmatrix} = -14 \cdot 14 + 15 \cdot 15 = 29.$$

Then each unknown is a determinant, with one column replaced by the right-hand side, divided by 29:

$$x_1 = \frac{\det\begin{bmatrix} 2 & 4 & 5 \\ 3 & 2 & 5 \\ 1 & 3 & -1 \end{bmatrix}}{29} = \frac{33}{29}, \qquad x_2 = \frac{\det\begin{bmatrix} 1 & 2 & 5 \\ 4 & 3 & 5 \\ -3 & 1 & -1 \end{bmatrix}}{29} = \frac{35}{29}, \qquad x_3 = \frac{\det\begin{bmatrix} 1 & 4 & 2 \\ 4 & 2 & 3 \\ -3 & 3 & 1 \end{bmatrix}}{29} = -\frac{23}{29}.$$

Inverses and Adjoint Matrix

The inverse of a square, full-rank matrix A can be found by solving $AX = I$. Written column by column, this means solving n linear systems, one for each column of the identity:

$$AX = I \quad \text{or} \quad \begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n\end{bmatrix} = \begin{bmatrix}\boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n\end{bmatrix}, \qquad A\boldsymbol{x}_j = \boldsymbol{e}_j, \quad 1 \leq j \leq n.$$

notes page 50

By Cramer's rule, the i^{th} entry of $\boldsymbol{x}_j$ is a determinant in which column i is replaced by $\boldsymbol{e}_j$. Expanding that determinant along the replaced column leaves a single cofactor, C_{ji}:

$$x_{ij} = \frac{|\boldsymbol{a}_1 \ \cdots \ \boldsymbol{a}_{i-1} \ \boldsymbol{e}_j \ \boldsymbol{a}_{i+1} \ \cdots \ \boldsymbol{a}_n|}{\det A} = \frac{C_{ji}}{\det A},$$

where C_{ji} is the (j, i) cofactor of A. Collecting these entries gives the inverse as the transposed matrix of cofactors, called the *adjoint* (or adjugate) of A, divided by the determinant:

Inverse by the adjoint

$$X = A^{-1} = \frac{[C_{ij}]^T}{\det A} = \frac{\operatorname{Adj} A}{\det A}.$$

ex/ Take again $A = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 5 \\ -3 & 3 & -1 \end{bmatrix}$ with $\det A = 29$. We solve $A\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \end{bmatrix}$, where

$$\boldsymbol{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \end{bmatrix}, \quad \boldsymbol{x}_3 = \begin{bmatrix} x_{13} \\ x_{23} \\ x_{33} \end{bmatrix}.$$

By Cramer's rule, the entries of the first column $\boldsymbol{x}_1$ are

$$x_{11} = \frac{\det \begin{bmatrix} 1 & 4 & 5 \\ 0 & 2 & 5 \\ 0 & 3 & -1 \end{bmatrix}}{\det A} = \frac{C_{11}}{\det A},$$

$$x_{21} = \frac{\det \begin{bmatrix} 1 & 1 & 5 \\ 4 & 0 & 5 \\ -3 & 0 & -1 \end{bmatrix}}{\det A} = \frac{C_{12}}{\det A},$$

$$x_{31} = \frac{\det \begin{bmatrix} 1 & 4 & 1 \\ 4 & 2 & 0 \\ -3 & 3 & 0 \end{bmatrix}}{\det A} = \frac{C_{13}}{\det A},$$

and the other columns follow in the same way. Together, the inverse is the transposed cofactor matrix over $\det A$:

$$X = A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \frac{1}{\det A} \underbrace{\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}^T}_{\operatorname{adj} A}.$$

A more practical technique for manual computation of A^{-1}: the adjoint formula needs many determinants, which is slow by hand. A much faster way is to solve all n systems $AA^{-1} = I$ at once.

notes page 51

Let $A^{-1} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix}$, where the columns $\boldsymbol{x}_i$ are the unknowns. Then

$$AA^{-1} = A\begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix} = I = \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n \end{bmatrix},$$

where $\boldsymbol{e}_i$ are the standard basis vectors. So $A\boldsymbol{x}_i = \boldsymbol{e}_i$ for each i. These could be solved one by one, but they can all be solved simultaneously by putting the identity next to A and row reducing:

Finding A^{-1} by row reduction
Form $\begin{bmatrix} A & \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n \end{bmatrix} \end{bmatrix} = [A\ I]$ and use elementary row operations to reduce it to

$$\begin{bmatrix} I & \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix} \end{bmatrix} = [I\ A^{-1}].$$

ex/ Find the inverse of $A = \begin{bmatrix} 1 & 3 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}$.

$$[A\ I] = \left[\begin{array}{ccc|ccc} 1 & 3 & 0 & 1 & 0 & 0 \\ -2 & 3 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{R_2\to R_2+2R_1}\left[\begin{array}{ccc|ccc}1&3&0&1&0&0\\0&9&1&2&1&0\\0&1&1&0&0&1\end{array}\right]\xrightarrow{R_2\to R_2/9}\left[\begin{array}{ccc|ccc}1&3&0&1&0&0\\0&1&1/9&2/9&1/9&0\\0&1&1&0&0&1\end{array}\right]$$

$$\xrightarrow{R_3\to R_3-R_2}\left[\begin{array}{ccc|ccc}1&3&0&1&0&0\\0&1&1/9&2/9&1/9&0\\0&0&8/9&-2/9&-1/9&1\end{array}\right]\xrightarrow{R_3\to R_3(9/8)}\left[\begin{array}{ccc|ccc}1&3&0&1&0&0\\0&1&1/9&2/9&1/9&0\\0&0&1&-1/4&-1/8&9/8\end{array}\right]$$

$$\xrightarrow{R_2\to R_2-\frac{1}{9}R_3}\left[\begin{array}{ccc|ccc}1&3&0&1&0&0\\0&1&0&1/4&1/8&-1/8\\0&0&1&-1/4&-1/8&9/8\end{array}\right]$$

$$\xrightarrow{R_1\to R_1-3R_2}\left[\begin{array}{ccc|ccc}1&0&0&1/4&-3/8&3/8\\0&1&0&1/4&1/8&-1/8\\0&0&1&-1/4&-1/8&9/8\end{array}\right].$$

The three columns on the right of the bar are A^{-1}.

LU-Factorization

LU factorization is another way to solve $A\boldsymbol{x}=\boldsymbol{c}$ for a square matrix A. The idea is to write A as a product of two simple matrices, one lower triangular and one upper triangular. If Gauss elimination of A needs no row interchanges, then

LU factorization

$$A=\underbrace{L}_{\text{lower triangular}}\underbrace{U}_{\text{upper triangular}}.$$

If row interchanges are needed, one instead factorizes $PA=LU$ with a permutation matrix P.

— notes page 52 —

The factorization is not unique. To see why, count unknowns. A 3×3 matrix A has 9 entries. The two triangular factors have 6 parameters each:

$$A=\underbrace{\begin{bmatrix}l_{11}&0&0\\l_{21}&l_{22}&0\\l_{31}&l_{32}&l_{33}\end{bmatrix}}_{\text{6 parameters}}\underbrace{\begin{bmatrix}u_{11}&u_{12}&u_{13}\\0&u_{22}&u_{23}\\0&0&u_{33}\end{bmatrix}}_{\text{6 parameters}}.$$

That is 12 unknowns for only 9 equations (for an $n\times n$ matrix, n^2+n unknowns). In *Doolittle's method* we fix the freedom by setting $l_{11}=l_{22}=l_{33}=1$ and solving for the remaining 9 parameters. For an $n\times n$ matrix, $l_{11}=\cdots=l_{nn}=1$ and there are n^2 remaining parameters.

ex/ Factorize $A=\begin{bmatrix}2&-3&3\\6&-8&7\\-2&6&-1\end{bmatrix}$ with Doolittle's method. Multiplying out the factors gives

$$\underbrace{\begin{bmatrix}2&-3&3\\6&-8&7\\-2&6&-1\end{bmatrix}}_{A}=\begin{bmatrix}1&0&0\\l_{21}&1&0\\l_{31}&l_{32}&1\end{bmatrix}\begin{bmatrix}u_{11}&u_{12}&u_{13}\\0&u_{22}&u_{23}\\0&0&u_{33}\end{bmatrix}$$

$$=\begin{bmatrix}u_{11}&u_{12}&u_{13}\\l_{21}u_{11}&l_{21}u_{12}+u_{22}&l_{21}u_{13}+u_{23}\\l_{31}u_{11}&l_{31}u_{12}+l_{32}u_{22}&l_{31}u_{13}+l_{32}u_{23}+u_{33}\end{bmatrix}.$$

Comparing entries with A, row by row, each unknown can be found from the ones already known:

$$\begin{aligned}
2 &= u_{11}, \quad -3 = u_{12}, \quad 3 = u_{13},\\
6 &= l_{21}u_{11} \Rightarrow l_{21} = 6/2 = 3; \quad -8 = l_{21}u_{12} + u_{22}\\
&\Rightarrow u_{22} = -8 - 3\cdot(-3) = 1,\\
7 &= l_{21}u_{13} + u_{23} \Rightarrow u_{23} = 7 - 3\cdot 3 = -2,\\
-2 &= l_{31}u_{11} \Rightarrow l_{31} = -1; \quad 6 = l_{31}u_{12} + l_{32}u_{22} \Rightarrow l_{32} = 3,\\
-1 &= l_{31}u_{13} + l_{32}u_{23} + u_{33} \Rightarrow u_{33} = -1 - (-1)\cdot 3 - 3\cdot(-2) = 8.
\end{aligned}$$

$$\Rightarrow A = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & -3 & 3 \\ 0 & 1 & -2 \\ 0 & 0 & 8 \end{bmatrix}.$$

Once $A = LU$ is known, the system $A\boldsymbol{x} = \boldsymbol{c}$ is solved in two easy steps. Name the product $U\boldsymbol{x} = \boldsymbol{y}$. First solve $L\boldsymbol{y} = \boldsymbol{c}$ for $\boldsymbol{y}$ by forward substitution, then solve $U\boldsymbol{x} = \boldsymbol{y}$ for $\boldsymbol{x}$ by backward substitution:

$$L\underbrace{U\boldsymbol{x}}_{\boldsymbol{y}} = \boldsymbol{c} \Rightarrow L\boldsymbol{y} = \boldsymbol{c} \text{ (forward substitution)}, \qquad U\boldsymbol{x} = \boldsymbol{y} \text{ (backward substitution)}.$$

This needs fewer operations than Gauss–Jordan: about $n^3/3$ multiplications for the factorization, versus about $n^3/2$ for Gauss–Jordan, and the factors L and U can be reused for every new right-hand side $\boldsymbol{c}$.