Linear Algebra

Last modified: 2026-09-28 23:45:45 (PDT)

Vectors

Definition 0.1 (Column vector)**.** A **column vector** of length p is an ordered list of p numbers, written vertically:

$$\tilde{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

Column vectors are the default convention in these notes and in most statistics textbooks. They are also called $p \times 1$ *matrices*.

Definition 0.2 (Transpose)**.** The **transpose** of a column vector $\tilde{x}$ is the row vector with the same sequence of entries, written horizontally:

$$\tilde{x}^\top \equiv \tilde{x}' \equiv [x_1,\ x_2,\ \dots,\ x_p]$$

The transpose operation converts a column vector to a row vector, or more generally, swaps the rows and columns of a matrix (Definition 0.12).

Definition 0.3 (Vector addition)**.** The **sum** of two column vectors $\tilde{x}$ and $\tilde{y}$ of the same length p is the column vector $\tilde{x} + \tilde{y}$ of length p obtained by adding entry by entry:

$$(\tilde{x} + \tilde{y})_i \stackrel{\text{def}}{=} x_i + y_i, \quad i = 1, \dots, p$$

Example 0.1 (Adding two vectors).

$$\begin{bmatrix}1\\2\\3\end{bmatrix} + \begin{bmatrix}4\\5\\6\end{bmatrix} = \begin{bmatrix}1+4\\2+5\\3+6\end{bmatrix} = \begin{bmatrix}5\\7\\9\end{bmatrix}$$

Definition 0.4 (Dot product). For any two real-valued vectors $\tilde{x} = (x_1, \ldots, x_p)$ and $\tilde{y} = (y_1, \ldots, y_p)$ of the same length p, the **dot product** of $\tilde{x}$ and $\tilde{y}$ is:

$$\tilde{x} \cdot \tilde{y} \stackrel{\text{def}}{=} \sum_{i=1}^{p} x_i y_i$$

Note

The dot product $\tilde{x} \cdot \tilde{y}$ is a *linear combination* of the entries of $\tilde{y}$, with coefficients $x_1, \ldots, x_p$. It is also the standard *inner product* on $\mathbb{R}^p$; "inner product" is the general notion, of which the dot product is one example.
See also the definitions in

- Dobson and Barnett (2018), Section 1.3 (equation 1.1, page 7)
- Kaplan (2022), chapter on vectors
- wikipedia

"Linear combination" can also refer to weighted sums of vectors, or in other words matrix-vector multiplication.
The dot-product has a different generalization for two matrices; see wikipedia for more.

Example 0.2 (A dot product). For $\tilde{x} = (1, 2, 3)$ and $\tilde{y} = (4, 5, 6)$:

$$\begin{aligned}
\tilde{x} \cdot \tilde{y} &= 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 && \text{(definition of the dot product)} \\
&= 4 + 10 + 18 && \text{(multiply)} \\
&= 32 && \text{(add)}
\end{aligned}$$

Theorem 0.1 (Dot product is symmetric). *The dot product is symmetric:*

$$\tilde{x} \cdot \tilde{y} = \tilde{y} \cdot \tilde{x}$$

Proof

Proof.

$$\begin{aligned}
\tilde{x} \cdot \tilde{y} &= \sum_{i=1}^{p} x_i y_i && \text{(definition of the dot product)} \\
&= \sum_{i=1}^{p} y_i x_i && \text{(products of numbers are symmetric)} \\
&= \tilde{y} \cdot \tilde{x} && \text{(definition of the dot product)}
\end{aligned}$$

□

Special vectors

Definition 0.5 (Zero vector)**.** The **zero vector** $\tilde{0}$ of length p has all entries equal to zero:

$$\tilde{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The zero vector is the additive identity for vector addition: $\tilde{x} + \tilde{0} = \tilde{x}$ for any vector $\tilde{x}$ of the same length.

Definition 0.6 (Ones vector)**.** The **ones vector** $\tilde{1}$ of length p has all entries equal to one:

$$\tilde{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

The dot product $\tilde{1} \cdot \tilde{x} = \sum_{i=1}^{p} x_i$ is the sum of all entries of $\tilde{x}$.

Definition 0.7 (Indicator vector / standard basis vector)**.** The j-th **indicator vector** (or *standard basis vector*) $\tilde{e}_j$ of length p has a 1 in position j and 0s elsewhere:

$$(\tilde{e}_j)_i = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad \tilde{e}_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \text{position } j$$

They are also called *unit vectors* or *standard basis vectors*.

Theorem 0.2 (Indicator vectors select entries)**.** *For any vector $\tilde{x}$ of length p and any $j \in \{1, \dots, p\}$:*

$$\tilde{e}_j \cdot \tilde{x} = x_j$$

Proof

Proof. Writing the dot product componentwise:

$$\begin{aligned} \tilde{e}_j \cdot \tilde{x} &= \sum_{i=1}^{p} (\tilde{e}_j)_i \, x_i && \text{(definition of the dot product)} \\ &= \sum_{i=1}^{p} \begin{cases} 1 \cdot x_i & \text{if } i = j \\ 0 \cdot x_i & \text{if } i \neq j \end{cases} && \text{(definition of } \tilde{e}_j\text{)} \\ &= x_j && \text{(only the } i = j \text{ term is nonzero)} \end{aligned}$$

□

Orthogonality

Definition 0.8 (Orthogonal vectors)**.** Two vectors $\tilde{x}$ and $\tilde{y}$ of the same length are **orthogonal** (written $\tilde{x} \perp \tilde{y}$) if their dot product (Definition 0.4) is zero:

$$\tilde{x} \perp \tilde{y} \iff \tilde{x} \cdot \tilde{y} = 0$$

Orthogonality generalizes the geometric notion of perpendicularity to arbitrary dimensions.

Definition 0.9 (Euclidean norm)**.** The **Euclidean norm** (or *length*) of a vector $\tilde{x}$ of length p is

$$\|\tilde{x}\| \stackrel{\text{def}}{=} \sqrt{\tilde{x} \cdot \tilde{x}} = \sqrt{\sum_{i=1}^{p} x_i^2}$$

Example 0.3 (The length of a vector)**.** For $\tilde{x} = (3, 4)$:

$$\begin{aligned}
\|\tilde{x}\| &= \sqrt{3^2 + 4^2} && \text{(definition of the norm)} \\
&= \sqrt{25} && \text{(square and add)} \\
&= 5
\end{aligned}$$

The vector $(0.6, 0.8)$ has norm $\sqrt{0.36 + 0.64} = 1$.

Definition 0.10 (Orthonormal vectors)**.** A set of vectors $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\}$ is **orthonormal** if the vectors are mutually orthogonal (Definition 0.8) and each has norm 1 (Definition 0.9), that is, unit length:

$$\tilde{x}_i \cdot \tilde{x}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The indicator vectors $\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_p$ (Definition 0.7) form an orthonormal set.

Video lecture

Hutchinson's Linear Function Basics (24 min) covers dot products and the Euclidean norm, and goes on to hyperplanes and level sets (Hutchinson, n.d.). The login for the video site is posted on Canvas.

Matrices

Definition 0.11 (Matrix)**.** A **matrix** of dimensions $m \times n$ is a rectangular array of $m \cdot n$ numbers, arranged in m rows and n columns:

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

The entry in row i and column j is denoted a_{ij} or $(\mathbf{A})_{ij}$. A column vector of length p is a special case: a $p \times 1$ matrix. A row vector of length p is a $1 \times p$ matrix.

Matrix transpose

Definition 0.12 (Matrix transpose)**.** The **transpose** of an $m \times n$ matrix $\mathbf{A}$ is the $n \times m$ matrix $\mathbf{A}^\top$ obtained by swapping the rows and columns of $\mathbf{A}$:

$$(\mathbf{A}^\top)_{ij} = a_{ji}$$

Matrix addition

Definition 0.13 (Zero matrix)**.** The $m \times n$ **zero matrix** $\mathbf{0}_{m\times n}$ (or $\mathbf{0}$ when dimensions are clear from context) has all entries equal to zero:

$$\mathbf{0}_{m\times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Definition 0.14 (Matrix addition)**.** Two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimensions $m \times n$ can be added element-wise; their **matrix sum** is:

$$(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij}$$

Theorem 0.3 (Matrix addition is commutative)**.** *For $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$:*

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

Theorem 0.4 (Matrix addition is associative)**.** *For $m \times n$ matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$:*

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

Theorem 0.5 (Zero matrix is the additive identity)**.** *For an $m \times n$ matrix $\mathbf{A}$:*

$$\mathbf{A} + \mathbf{0}_{m\times n} = \mathbf{A}$$

Theorem 0.6 (Additive inverse)**.** *For any $m \times n$ matrix $\mathbf{A}$, the $m \times n$ matrix $-\mathbf{A}$ (defined by $(-\mathbf{A})_{ij} = -a_{ij}$) satisfies:*

$$\mathbf{A} + (-\mathbf{A}) = \mathbf{0}_{m\times n}$$

Scalar multiplication

Definition 0.15 (Scalar multiplication)**.** The **scalar multiple** of a matrix $\mathbf{A}$ by a scalar c is:

$$(c\mathbf{A})_{ij} = c \cdot a_{ij}$$

Matrix multiplication

Definition 0.16 (Matrix multiplication)**.** The **product** of an $m \times k$ matrix $\mathbf{A}$ and a $k \times n$ matrix $\mathbf{B}$ is the $m \times n$ matrix $\mathbf{C} = \mathbf{AB}$ with entries:

$$c_{ij} = \sum_{s=1}^{k} a_{is}\, b_{sj}$$

Matrix multiplication is only defined when the number of columns in $\mathbf{A}$ equals the number of rows in $\mathbf{B}$.

Matrix multiplication is **not** commutative in general: $\mathbf{AB} \neq \mathbf{BA}$.

Example 0.4 (Dot product as matrix multiplication)**.** The dot product of two column vectors $\tilde{x}$ and $\tilde{\beta}$ can be written as a matrix product of the row vector $\tilde{x}^\top$ with the column vector $\tilde{\beta}$:

$$\begin{aligned}
\tilde{x} \cdot \tilde{\beta} &= \tilde{x}^\top\, \tilde{\beta} \\
&= [x_1,\ x_2,\ \dots,\ x_p] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \\
&= x_1\beta_1 + x_2\beta_2 + \cdots + x_p\beta_p
\end{aligned}$$

Theorem 0.7 (Matrix multiplication is associative)**.** *For an $m \times k$ matrix $\mathbf{A}$, a $k \times l$ matrix $\mathbf{B}$, and an $l \times n$ matrix $\mathbf{C}$:*

$$\underbrace{(\mathbf{AB})\mathbf{C}}_{m\times n} = \underbrace{\mathbf{A}(\mathbf{BC})}_{m\times n}$$

Proof

Proof. Entry (i, j) of each side, from Definition 0.16:

$$\begin{aligned}
[(\mathbf{AB})\mathbf{C}]_{ij} &= \sum_{t=1}^{l} (\mathbf{AB})_{it}\, c_{tj} && \text{(definition of } (\mathbf{AB})\mathbf{C})\\
&= \sum_{t=1}^{l} \left(\sum_{s=1}^{k} a_{is}\, b_{st} \right) c_{tj} && \text{(definition of } \mathbf{AB})\\
&= \sum_{s=1}^{k} a_{is} \left(\sum_{t=1}^{l} b_{st}\, c_{tj} \right) && \text{(distribute and swap the finite sums)}\\
&= \sum_{s=1}^{k} a_{is}\, (\mathbf{BC})_{sj} && \text{(definition of } \mathbf{BC})\\
&= [\mathbf{A}(\mathbf{BC})]_{ij} && \text{(definition of } \mathbf{A}(\mathbf{BC}))
\end{aligned}$$

□

Theorem 0.8 (Matrix multiplication is distributive over addition)**.** *For an $m \times k$ matrix $\mathbf{A}$ and $k \times n$ matrices $\mathbf{B}$ and $\mathbf{C}$:*

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

For $m \times k$ matrices $\mathbf{A}$ and $\mathbf{B}$ and a $k \times n$ matrix $\mathbf{C}$:

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$$

Matrix-vector multiplication

Definition 0.17 (Matrix-vector multiplication)**.** The **matrix-vector product** of an $m \times p$ matrix $\mathbf{A}$ and a $p \times 1$ column vector $\tilde{x}$ is the $m \times 1$ column vector $\mathbf{A}\tilde{x}$ with entries:

$$(\mathbf{A}\tilde{x})_i = \sum_{j=1}^{p} a_{ij}\, x_j$$

Matrix-vector multiplication is a generalization of the dot product. Each entry of the result is a dot product of a row of $\mathbf{A}$ with the vector $\tilde{x}$.

Transposes of sums and products

Theorem 0.9 (Transpose of a sum)**.** *For $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$:*

$$(\mathbf{A}+\mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$$

In particular, for column vectors $\tilde{x}$ and $\tilde{y}$ of the same length:

$$(\tilde{x}+\tilde{y})^\top = \tilde{x}^\top + \tilde{y}^\top$$

ℹ Proof

Proof. Entry (i,j) of each side:

$$\begin{aligned}
\left[(\mathbf{A}+\mathbf{B})^\top\right]_{ij} &= (\mathbf{A}+\mathbf{B})_{ji} && \text{(definition of the transpose)}\\
&= a_{ji} + b_{ji} && \text{(definition of matrix addition)}\\
&= (\mathbf{A}^\top)_{ij} + (\mathbf{B}^\top)_{ij} && \text{(definition of the transpose)}\\
&= [\mathbf{A}^\top + \mathbf{B}^\top]_{ij} && \text{(definition of matrix addition)}
\end{aligned}$$

□

Theorem 0.10 (Transpose of a product)**.** *For an $m \times k$ matrix $\mathbf{A}$ and a $k \times n$ matrix $\mathbf{B}$:*

$$\underbrace{(\mathbf{AB})^\top}_{n\times m} = \underbrace{\mathbf{B}^\top}_{n\times k}\ \underbrace{\mathbf{A}^\top}_{k\times m}$$

ℹ Proof

Proof. Entry (i,j) of each side:

$$\begin{aligned}
\left[(\mathbf{AB})^\top\right]_{ij} &= (\mathbf{AB})_{ji} && \text{(definition of the transpose)}\\
&= \sum_{s=1}^{k} a_{js}\, b_{si} && \text{(definition of matrix multiplication)}\\
&= \sum_{s=1}^{k} (\mathbf{B}^\top)_{is}\, (\mathbf{A}^\top)_{sj} && \text{(definition of the transpose; reorder each product)}\\
&= [\mathbf{B}^\top\, \mathbf{A}^\top]_{ij} && \text{(definition of matrix multiplication)}
\end{aligned}$$

□

The order of the factors reverses when transposing a product.

Rank

Definition 0.18 (Linearly independent vectors)**.** Vectors $\tilde{v}_1, \dots, \tilde{v}_k$ of the same length p are **linearly independent** if the only numbers $c_1, \dots, c_k$ with

$$c_1\tilde{v}_1 + \cdots + c_k\tilde{v}_k = \tilde{0}$$

are $c_1 = \cdots = c_k = 0$.

Example 0.5 (Independent and dependent pairs of vectors)**.**

- $\tilde{v}_1 = (1,0)$ and $\tilde{v}_2 = (1,1)$ are linearly independent: $c_1\tilde{v}_1 + c_2\tilde{v}_2 = (c_1 + c_2, c_2)$, which is $\tilde{0}$ only if $c_2 = 0$ and then $c_1 = 0$.
- $\tilde{v}_1 = (1,2)$ and $\tilde{v}_2 = (2,4)$ are not: $2\tilde{v}_1 - \tilde{v}_2 = \tilde{0}$.

Definition 0.19 (Rank)**.** The **rank** of a matrix $\mathbf{A}$, written $\operatorname{rank}(\mathbf{A})$, is the largest number of columns of $\mathbf{A}$ that are linearly independent (Definition 0.18).

An $n \times p$ matrix whose rank is p, so that all of its columns are linearly independent, is said to have *full column rank*; that can only happen when $p \le n$, because more than n vectors of length n are never linearly independent.

Example 0.6 (The rank of two 3×2 matrices)**.** The matrix $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ has rank 2: if $c_1(1,1,1) + c_2(1,2,3) = \tilde{0}$, then subtracting the first entry from the second gives $c_2 = 0$, and then $c_1 = 0$.

The matrix $\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$ has rank 1: its second column is twice its first, so the two columns are not linearly independent, but the first column on its own is.

Outer product

Definition 0.20 (Outer product)**.** The **outer product** of a vector $\tilde{u}$ of length m and a vector $\tilde{v}$ of length n is the $m \times n$ matrix product (Definition 0.16) $\tilde{u}\,\tilde{v}^\top$ of the column vector $\tilde{u}$ with the row vector $\tilde{v}^\top$. Its entries are

$$\Big(\underbrace{\tilde{u}}_{m\times 1}\ \underbrace{\tilde{v}^\top}_{1\times n}\Big)_{ij} = u_i v_j \qquad \text{(definition of matrix multiplication; the sum has one term)}$$

The dot product $\tilde{u}^\top \tilde{v}$ (Example 0.4) needs two vectors of the same length and gives a number. The outer product $\tilde{u}\,\tilde{v}^\top$ takes vectors of any two lengths and gives a matrix (Banerjee and Roy 2014, chap. 1, p. 11).

Example 0.7 (An outer product)**.** For $\tilde{u} = (1, 2, 3)$ and $\tilde{v} = (4, 5)$:

$$\begin{aligned} \tilde{u}\,\tilde{v}^\top &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 4 & 5 \end{bmatrix} && \text{(definition of the outer product)} \\ &= \begin{bmatrix} 1\cdot 4 & 1\cdot 5 \\ 2\cdot 4 & 2\cdot 5 \\ 3\cdot 4 & 3\cdot 5 \end{bmatrix} && \text{(entry (i,j) is $u_i v_j$)} \\ &= \begin{bmatrix} 4 & 5 \\ 8 & 10 \\ 12 & 15 \end{bmatrix} && \text{(multiply)} \end{aligned}$$

Theorem 0.11 (An outer product has rank one)**.** *If $\tilde{u}$ is a nonzero vector of length m and $\tilde{v}$ is a nonzero vector of length n, then* $\operatorname{rank}(\tilde{u}\,\tilde{v}^\top) = 1$ *(Definition 0.19).*

Proof

Proof. Column j of $\tilde{u}\,\tilde{v}^\top$ has entries $u_1 v_j, \ldots, u_m v_j$, so it is the vector $v_j \tilde{u}$.
The rank is at least one. Because $\tilde{v} \neq \tilde{0}$, some entry v_j is nonzero. Then column j, $v_j\tilde{u}$, is a nonzero vector, and a single nonzero vector is linearly independent (Definition 0.18): $c\,(v_j\tilde{u}) = \tilde{0}$ forces $c = 0$.
The rank is at most one. Take any two columns $j \neq k$. If $v_j = v_k = 0$, both columns are $\tilde{0}$, and $1\cdot(v_j\tilde{u}) + 1\cdot(v_k\tilde{u}) = \tilde{0}$. Otherwise, use the coefficients $c_j = v_k$ and $c_k = -v_j$, which are not both zero:

$$\begin{aligned} v_k\,(v_j\tilde{u}) - v_j\,(v_k\tilde{u}) &= (v_k v_j - v_j v_k)\,\tilde{u} && \text{(collect the scalar multiples of $\tilde{u}$)} \\ &= 0\cdot\tilde{u} && \text{(multiplication of numbers is commutative)} \\ &= \tilde{0} && \text{(a zero multiple of a vector is $\tilde{0}$)} \end{aligned}$$

Either way, every pair of columns has a combination with coefficients that are not all zero and that equals $\tilde{0}$. Now take any set of two or more columns. It contains a pair

$j \neq k$. Use that pair's coefficients for columns j and k, and the coefficient 0 for every other column in the set. This combination equals $\tilde{0}$, and its coefficients are not all zero, so the set is not linearly independent (Definition 0.18). The largest linearly independent set of columns therefore has one column. □

Example 0.8 (The rank of an outer product)**.** In Example 0.7, the second column $(5, 10, 15)$ of $\tilde{u}\,\tilde{v}^\top$ is $\frac{5}{4}$ times the first column $(4, 8, 12)$, because the columns are $v_1\tilde{u} = 4\tilde{u}$ and $v_2\tilde{u} = 5\tilde{u}$. So the two columns are not linearly independent, and the 3×2 matrix has rank 1.

Special Matrices

See also Definition 0.13 for the zero matrix.

Definition 0.21 (Square matrix)**.** A matrix is **square** if it has the same number of rows as columns.

Definition 0.22 (Order of a square matrix)**.** The **order** of a square matrix (Definition 0.21) is its number of rows, which equals its number of columns.

Definition 0.23 (Matrix power)**.** For a square matrix $\mathbf{A}$ of order p and a positive integer k, the k-th **power** of $\mathbf{A}$ is:

$$\mathbf{A}^k = \underbrace{\mathbf{A}\,\mathbf{A} \cdots \mathbf{A}}_{k \text{ copies}}$$

In particular, $\mathbf{A}^2 = \mathbf{A}\mathbf{A}$.

Definition 0.24 (Identity matrix)**.** The $p \times p$ **identity matrix** $\mathbf{I}_p$ (or $\mathbf{I}$ when the size is clear from context) has ones on the main diagonal and zeros elsewhere:

$$(\mathbf{I}_p)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad \mathbf{I}_p = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Theorem 0.12 (Identity matrix is a multiplicative identity)**.** *For any $m \times p$ matrix $\mathbf{A}$:*

$$\mathbf{A}\,\mathbf{I}_p = \mathbf{A}$$

$$\mathbf{I}_m\,\mathbf{A} = \mathbf{A}$$

Definition 0.25 (Symmetric matrix)**.** A square matrix $\mathbf{A}$ is **symmetric** if $\mathbf{A}^\top = \mathbf{A}$, i.e., $a_{ij} = a_{ji}$ for all i and j.

Covariance matrices and information matrices are symmetric.

Definition 0.26 (Diagonal matrix)**.** A square matrix $\mathbf{D}$ is a **diagonal matrix** if all off-diagonal entries are zero: $d_{ij} = 0$ whenever $i \neq j$:

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}$$

Diagonal matrices are denoted $\mathbf{D} = \operatorname{diag}(d_1, d_2, \ldots, d_p)$, where $d_1, \ldots, d_p$ are the diagonal entries.

Definition 0.27 (Matrix inverse)**.** For a square $p \times p$ matrix $\mathbf{A}$, the **inverse** $\mathbf{A}^{-1}$ (if it exists) is the unique matrix satisfying:

$$\mathbf{A}\,\mathbf{A}^{-1} = \mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}_p$$

Definition 0.28 (Invertible matrix)**.** A square matrix $\mathbf{A}$ is **invertible** (or *non-singular*) if it has an inverse $\mathbf{A}^{-1}$ (Definition 0.27). A square matrix with no inverse is *singular*.

Example 0.9 (An invertible matrix and a singular one)**.** The matrix $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$ is invertible, with $\mathbf{A}^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ 0 & 1 \end{bmatrix}$: multiplying out gives $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_2$. The matrix $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ is singular: for any 2×2 matrix $\mathbf{C}$, the two rows of $\mathbf{BC}$ are equal, so $\mathbf{BC}$ can never be $\mathbf{I}_2$, whose two rows differ.

Theorem 0.13 (Inverse of a product). *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are invertible* $p \times p$ *matrices, then* $\mathbf{AB}$ *is invertible, and*

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

Proof

Proof. Multiply $\mathbf{AB}$ by the candidate inverse on the right:

$$\begin{aligned}
(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) &= \mathbf{A}(\mathbf{B}\mathbf{B}^{-1})\mathbf{A}^{-1} && \text{(matrix multiplication is associative)} \\
&= \mathbf{A}\,\mathbf{I}_p\,\mathbf{A}^{-1} && \text{(definition of } \mathbf{B}^{-1}\text{)} \\
&= \mathbf{A}\mathbf{A}^{-1} && \text{(identity matrix)} \\
&= \mathbf{I}_p && \text{(definition of } \mathbf{A}^{-1}\text{)}
\end{aligned}$$

and on the left:

$$\begin{aligned}
(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) &= \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} && \text{(matrix multiplication is associative)} \\
&= \mathbf{B}^{-1}\,\mathbf{I}_p\,\mathbf{B} && \text{(definition of } \mathbf{A}^{-1}\text{)} \\
&= \mathbf{B}^{-1}\mathbf{B} && \text{(identity matrix)} \\
&= \mathbf{I}_p && \text{(definition of } \mathbf{B}^{-1}\text{)}
\end{aligned}$$

So $\mathbf{B}^{-1}\mathbf{A}^{-1}$ satisfies Definition 0.27 for $\mathbf{AB}$. The steps use Theorem 0.7 and Theorem 0.12. □

Theorem 0.14 (Inverse of a transpose). *If* $\mathbf{A}$ *is an invertible* $p \times p$ *matrix, then* $\mathbf{A}^\top$ *is invertible, and*

$$\left(\mathbf{A}^\top\right)^{-1} = \left(\mathbf{A}^{-1}\right)^\top$$

Proof

Proof. Multiply $\mathbf{A}^\top$ by the candidate inverse on each side:

$$\begin{aligned}
\mathbf{A}^\top\left(\mathbf{A}^{-1}\right)^\top &= \left(\mathbf{A}^{-1}\mathbf{A}\right)^\top && \text{(transpose of a product)} \\
&= {\mathbf{I}_p}^\top && \text{(definition of } \mathbf{A}^{-1}\text{)} \\
&= \mathbf{I}_p && (\mathbf{I}_p \text{ is symmetric})
\end{aligned}$$

$$
\begin{aligned}
\left(\mathbf{A}^{-1}\right)^{\top}\mathbf{A}^{\top} &= \left(\mathbf{A}\mathbf{A}^{-1}\right)^{\top} && \text{(transpose of a product)}\\
&= \mathbf{I}_p{}^{\top} && \text{(definition of } \mathbf{A}^{-1}\text{)}\\
&= \mathbf{I}_p && \text{(}\mathbf{I}_p \text{ is symmetric)}
\end{aligned}
$$

So $\left(\mathbf{A}^{-1}\right)^{\top}$ satisfies Definition 0.27 for $\mathbf{A}^{\top}$. The first step of each display is Theorem 0.10.

□

Example 0.10 (Inverting a transpose)**.** For $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$, Example 0.9 gives $\mathbf{A}^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ 0 & 1 \end{bmatrix}$, so Theorem 0.14 says

$$
\left(\mathbf{A}^{\top}\right)^{-1} = \left(\begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}\right)^{-1} = \left(\mathbf{A}^{-1}\right)^{\top} = \begin{bmatrix} 0.5 & 0 \\ -0.5 & 1 \end{bmatrix},
$$

and multiplying $\begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 0.5 & 0 \\ -0.5 & 1 \end{bmatrix}$ out does give $\mathbf{I}_2$.

Corollary 0.1 (The inverse of a symmetric matrix is symmetric)**.** *If* $\mathbf{A}$ *is symmetric and invertible, then* $\mathbf{A}^{-1}$ *is symmetric.*

Proof

Proof.

$$
\begin{aligned}
\left(\mathbf{A}^{-1}\right)^{\top} &= \left(\mathbf{A}^{\top}\right)^{-1} && \text{(inverse of a transpose)}\\
&= \mathbf{A}^{-1} && \text{(}\mathbf{A} \text{ is symmetric)}
\end{aligned}
$$

The first step is Theorem 0.14. □

Example 0.11 (Inverting a symmetric matrix)**.** $\mathbf{S} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ is symmetric, and $\mathbf{S}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ is symmetric too, as Corollary 0.1 says.

Definition 0.29 (Idempotent matrix)**.** A square matrix $\mathbf{A}$ is **idempotent** if

$$
\mathbf{A}^2 = \mathbf{A}
$$

Definition 0.30 (Orthogonal projection matrix)**.** A square matrix $\mathbf{P}$ is an **orthogonal projection matrix** if it is both symmetric (Definition 0.25) and idempotent (Definition 0.29):

$$\mathbf{P}^\top = \mathbf{P} \qquad \text{and} \qquad \mathbf{P}^2 = \mathbf{P}$$

Any idempotent matrix is called a *projection matrix*; an idempotent matrix that is not symmetric is an *oblique* projection. Regression texts often say "projection matrix" when they mean an orthogonal one. In these notes, "projection matrix" always means an orthogonal projection matrix in the sense of Definition 0.30.

Example 0.12 (An orthogonal projection and an oblique one)**.** $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is symmetric, and $\mathbf{P}^2 = \mathbf{P}$, so $\mathbf{P}$ is an orthogonal projection matrix; it maps (v_1, v_2) to $(v_1, 0)$.
$\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ is idempotent:

$$\mathbf{Q}^2 = \begin{bmatrix} 1\cdot 1 + 1\cdot 0 & 1\cdot 1 + 1\cdot 0 \\ 0\cdot 1 + 0\cdot 0 & 0\cdot 1 + 0\cdot 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \mathbf{Q},$$

but $\mathbf{Q}^\top \neq \mathbf{Q}$, so $\mathbf{Q}$ is an oblique projection, not an orthogonal projection matrix.

Theorem 0.15 (Complement of a projection matrix)**.** *If $\mathbf{P}$ is a $p \times p$ orthogonal projection matrix (Definition 0.30), then $\mathbf{I}_p - \mathbf{P}$ is also an orthogonal projection matrix.*

Proof

Proof. We verify symmetry and idempotency.
Symmetry:

$$(\mathbf{I} - \mathbf{P})^\top = \mathbf{I}^\top - \mathbf{P}^\top = \mathbf{I} - \mathbf{P}$$

Idempotency:

$$\begin{aligned}(\mathbf{I} - \mathbf{P})^2 &= (\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P}) \\ &= \mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P}^2 \\ &= \mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P} \\ &= \mathbf{I} - \mathbf{P}\end{aligned}$$

□

Theorem 0.16 (Projection matrices produce orthogonal decompositions)**.** *If $\mathbf{P}$ is a $p \times p$ orthogonal projection matrix (Definition 0.30) and $\tilde{v}$ is any vector of length p, then the two components of the decomposition*

$$\tilde{v} = \underbrace{\mathbf{P}\tilde{v}}_{projected} + \underbrace{(\mathbf{I}_p - \mathbf{P})\tilde{v}}_{residual}$$

are orthogonal:

$$\mathbf{P}\tilde{v} \perp (\mathbf{I}_p - \mathbf{P})\tilde{v}$$

Proof

Proof.

$$\begin{aligned}
(\mathbf{P}\tilde{v})^\top (\mathbf{I} - \mathbf{P})\tilde{v} &= \tilde{v}^\top \mathbf{P}^\top (\mathbf{I} - \mathbf{P})\tilde{v} \\
&= \tilde{v}^\top \mathbf{P} (\mathbf{I} - \mathbf{P})\tilde{v} \\
&= \tilde{v}^\top (\mathbf{P} - \mathbf{P}^2)\tilde{v} \\
&= \tilde{v}^\top (\mathbf{P} - \mathbf{P})\tilde{v} \\
&= \tilde{v}^\top \mathbf{0}\, \tilde{v} \\
&= 0
\end{aligned}$$

where the second line uses symmetry ($\mathbf{P}^\top = \mathbf{P}$) and the fourth line uses idempotency ($\mathbf{P}^2 = \mathbf{P}$). □

Definition 0.31 (Orthogonal matrix)**.** A $p \times p$ matrix $\mathbf{Q}$ is **orthogonal** if

$$\underbrace{\mathbf{Q}^\top}_{p\times p} \underbrace{\mathbf{Q}}_{p\times p} = \mathbf{I}_p$$

Entry (i, j) of $\mathbf{Q}^\top\mathbf{Q}$ is the dot product of column i and column j of $\mathbf{Q}$, so $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_p$ says that the columns of $\mathbf{Q}$ are orthonormal (Definition 0.10). For a square matrix, $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_p$ also implies $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}_p$, so $\mathbf{Q}^{-1} = \mathbf{Q}^\top$ (Definition 0.27) (Banerjee and Roy 2014, chap. 8, Theorem 8.1 and Definition 8.1, p. 209).

Example 0.13 (A rotation matrix)**.** The matrix

$$\mathbf{Q} = \begin{bmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{bmatrix}$$

rotates each vector in the plane counterclockwise by the angle θ with $\cos\theta = 0.6$ and $\sin\theta = 0.8$ (about 53 degrees) (Banerjee and Roy 2014, chap. 8, Example 8.1, p. 207). It is orthogonal:

$$
\begin{aligned}
\mathbf{Q}^\top \mathbf{Q} &= \begin{bmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{bmatrix} \begin{bmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{bmatrix} && \text{(definition of the transpose)} \\
&= \begin{bmatrix} 0.36 + 0.64 & -0.48 + 0.48 \\ -0.48 + 0.48 & 0.64 + 0.36 \end{bmatrix} && \text{(definition of matrix multiplication)} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} && \text{(add)}
\end{aligned}
$$

Theorem 0.17 (Orthogonal matrices preserve length)**.** *If* $\mathbf{Q}$ *is a* $p \times p$ *orthogonal matrix (Definition 0.31) and* $\tilde{x}$ *is a vector of length* p*, then*

$$\|\mathbf{Q}\tilde{x}\| = \|\tilde{x}\|$$

Proof

Proof.

$$
\begin{aligned}
\|\mathbf{Q}\tilde{x}\|^2 &= (\mathbf{Q}\tilde{x}) \cdot (\mathbf{Q}\tilde{x}) && \text{(definition of the Euclidean norm)} \\
&= (\mathbf{Q}\tilde{x})^\top (\mathbf{Q}\tilde{x}) && \text{(dot product as a matrix product)} \\
&= \tilde{x}^\top \mathbf{Q}^\top (\mathbf{Q}\tilde{x}) && \text{(transpose of a product)} \\
&= \tilde{x}^\top (\mathbf{Q}^\top \mathbf{Q})\, \tilde{x} && \text{(regroup; matrix multiplication is associative)} \\
&= \tilde{x}^\top \mathbf{I}_p\, \tilde{x} && \text{(definition of an orthogonal matrix)} \\
&= \tilde{x}^\top \tilde{x} && \text{(identity matrix)} \\
&= \|\tilde{x}\|^2 && \text{(definition of the Euclidean norm)}
\end{aligned}
$$

Both norms are nonnegative square roots, so equal squares give $\|\mathbf{Q}\tilde{x}\| = \|\tilde{x}\|$. The second step is Example 0.4, the third is Theorem 0.10, and the fourth is Theorem 0.7. □

Example 0.14 (Rotating a vector keeps its length)**.** With $\mathbf{Q}$ from Example 0.13 and $\tilde{x} = (3, 4)$ from Example 0.3:

$$
\begin{aligned}
\mathbf{Q}\tilde{x} &= \begin{bmatrix} 0.6 \cdot 3 - 0.8 \cdot 4 \\ 0.8 \cdot 3 + 0.6 \cdot 4 \end{bmatrix} && \text{(definition of matrix-vector multiplication)} \\
&= \begin{bmatrix} 1.8 - 3.2 \\ 2.4 + 2.4 \end{bmatrix} && \text{(multiply)} \\
&= \begin{bmatrix} -1.4 \\ 4.8 \end{bmatrix} && \text{(add)}
\end{aligned}
$$

$$
\begin{aligned}
\|\mathbf{Q}\tilde{x}\| &= \sqrt{(-1.4)^2 + 4.8^2} && \text{(definition of the norm)}\\
&= \sqrt{1.96 + 23.04} && \text{(square)}\\
&= \sqrt{25} && \text{(add)}\\
&= 5 && \text{(take the square root)}
\end{aligned}
$$

which is $\|\tilde{x}\| = 5$.

Quadratic Forms

Definition 0.32 (Quadratic form)**.** A **quadratic form** is a mathematical expression of the structure

$$\tilde{x}^\top \, \mathbf{S} \, \tilde{x}$$

where $\tilde{x}$ is a $p \times 1$ vector and $\mathbf{S}$ is a $p \times p$ matrix.

Quadratic forms are the matrix generalizations of the scalar expression cx^2. They occur frequently in statistics:

- The residual sum of squares in linear regression (see Vector Calculus) is a quadratic form.
- The variance of a linear combination of estimates (see Inference about Gaussian Linear Regression Models) is a quadratic form: $\text{Var}\left(\tilde{x}^\top \hat{\tilde{\beta}}\right) = \tilde{x}^\top \, \text{Var}\left(\hat{\tilde{\beta}}\right) \tilde{x}$.

Definition 0.33 (Symmetric part of a square matrix)**.** The **symmetric part** of a $p \times p$ matrix $\mathbf{S}$ is

$$\frac{1}{2}(\mathbf{S} + \mathbf{S}^\top)$$

Example 0.15 (The symmetric part of a 2×2 matrix)**.** For $\mathbf{S} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$, the symmetric part is

$$\frac{1}{2}\left(\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} 2 & 2 \\ 2 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix},$$

which is symmetric.

Theorem 0.18 (A quadratic form depends only on the symmetric part)**.** *If* $\mathbf{S}$ *is a* $p \times p$ *matrix and* $\tilde{x}$ *is a vector of length* p*, then*

$$\tilde{x}^\top \mathbf{S} \tilde{x} = \tilde{x}^\top \left(\frac{1}{2}(\mathbf{S} + \mathbf{S}^\top) \right) \tilde{x}.$$

So the value of a quadratic form depends only on the symmetric part (Definition 0.33) of $\mathbf{S}$*.*

Proof

Proof. The quadratic form $\tilde{x}^\top \mathbf{S} \tilde{x}$ is a 1×1 matrix, so it equals its own transpose:

$$\begin{aligned}
\tilde{x}^\top \mathbf{S} \tilde{x} &= (\tilde{x}^\top \mathbf{S} \tilde{x})^\top && \text{(a } 1 \times 1 \text{ matrix is symmetric)} \\
&= \tilde{x}^\top\, \mathbf{S}^\top\, (\tilde{x}^\top)^\top && \text{(transpose of a product, twice)} \\
&= \tilde{x}^\top\, \mathbf{S}^\top\, \tilde{x} && \text{(transposing twice changes nothing)}
\end{aligned}$$

Averaging the two expressions for $\tilde{x}^\top \mathbf{S} \tilde{x}$:

$$\begin{aligned}
\tilde{x}^\top \mathbf{S} \tilde{x} &= \frac{1}{2}(\tilde{x}^\top \mathbf{S} \tilde{x} + \tilde{x}^\top\, \mathbf{S}^\top\, \tilde{x}) && \text{(average of two equal numbers)} \\
&= \frac{1}{2}\, \tilde{x}^\top (\mathbf{S} + \mathbf{S}^\top) \tilde{x} && \text{(distributive law)} \\
&= \tilde{x}^\top \left(\frac{1}{2}(\mathbf{S} + \mathbf{S}^\top) \right) \tilde{x} && \text{(move the scalar } \tfrac{1}{2} \text{ inside)}
\end{aligned}$$

The distributive step is Theorem 0.8, and the transpose step is Theorem 0.10. ☐

Example 0.16 (Replacing a matrix by its symmetric part)**.** With $\mathbf{S} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$ from Example 0.15 and $\tilde{x} = (1, 1)$:

$$\tilde{x}^\top \mathbf{S} \tilde{x} = 1 + 2 + 0 + 3 = 6 \qquad \tilde{x}^\top \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \tilde{x} = 1 + 1 + 1 + 3 = 6$$

Both quadratic forms equal the sum of their matrix's entries, because every entry of $\tilde{x}$ is 1.

Trace and Matrix Inner Product

Definition 0.34 (Trace)**.** The **trace** of a $p \times p$ matrix $\mathbf{M}$ is the sum of its diagonal entries:

$$\operatorname{tr}(\mathbf{M}) \overset{\text{def}}{=} \sum_{i=1}^{p} M_{ii}$$

Example 0.17 (The trace of a 2×2 matrix)**.**

$$\operatorname{tr}\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right) = 1 + 4 = 5$$

Theorem 0.19 (The trace of a product does not depend on the order)**.** *For an $m \times n$ matrix $\mathbf{A}$ and an $n \times m$ matrix $\mathbf{B}$:*

$$\operatorname{tr}\Bigg(\underbrace{\mathbf{AB}}_{m \times m}\Bigg) = \operatorname{tr}\Bigg(\underbrace{\mathbf{BA}}_{n \times n}\Bigg)$$

Proof

Proof.

$$\begin{aligned}
\operatorname{tr}(\mathbf{AB}) &= \sum_{i=1}^{m} (\mathbf{AB})_{ii} && \text{(definition of the trace)} \\
&= \sum_{i=1}^{m} \sum_{s=1}^{n} a_{is}\, b_{si} && \text{(definition of matrix multiplication)} \\
&= \sum_{s=1}^{n} \sum_{i=1}^{m} a_{is}\, b_{si} && \text{(swap the order of two finite sums)} \\
&= \sum_{s=1}^{n} \sum_{i=1}^{m} b_{si}\, a_{is} && \text{(multiplication of numbers is commutative)} \\
&= \sum_{s=1}^{n} (\mathbf{BA})_{ss} && \text{(definition of matrix multiplication)} \\
&= \operatorname{tr}(\mathbf{BA}) && \text{(definition of the trace)}
\end{aligned}$$

□

$\mathbf{AB}$ and $\mathbf{BA}$ need not have the same size, so Theorem 0.19 can equate the traces of an $m \times m$ matrix and an $n \times n$ matrix. Applying it to $\mathbf{A}$ and the product $\mathbf{BC}$ moves the last factor of a triple product to the front: $\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{CAB})$ (Banerjee and Roy 2014, chap. 1, Theorem

1.5 and eq. 1.17, p. 19).

Example 0.18 (Traces of the two products of a 2×3 and a 3×2 matrix)**.** Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 3 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix}$$

The products are

$$\mathbf{AB} = \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix} \qquad \mathbf{BA} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & 3 \\ 0 & 1 & 3 \end{bmatrix}$$

The trace needs only the diagonal entries. For $\mathbf{AB}$, which is 2×2:

$$\begin{aligned} \operatorname{tr}(\mathbf{AB}) &= (\mathbf{AB})_{11} + (\mathbf{AB})_{22} && \text{(definition of the trace)} \\ &= (1 \cdot 1 + 2 \cdot 2 + 0 \cdot 0) + (0 \cdot 0 + 1 \cdot 1 + 3 \cdot 1) && \text{(definition of matrix multiplication)} \\ &= (1 + 4 + 0) + (0 + 1 + 3) && \text{(multiply)} \\ &= 5 + 4 && \text{(add within each entry)} \\ &= 9 && \text{(add)} \end{aligned}$$

For $\mathbf{BA}$, which is 3×3:

$$\begin{aligned} \operatorname{tr}(\mathbf{BA}) &= (\mathbf{BA})_{11} + (\mathbf{BA})_{22} + (\mathbf{BA})_{33} && \text{(definition of the trace)} \\ &= (1 \cdot 1 + 0 \cdot 0) + (2 \cdot 2 + 1 \cdot 1) + (0 \cdot 0 + 1 \cdot 3) && \text{(definition of matrix multiplication)} \\ &= (1 + 0) + (4 + 1) + (0 + 3) && \text{(multiply)} \\ &= 1 + 5 + 3 && \text{(add within each entry)} \\ &= 9 && \text{(add)} \end{aligned}$$

The products have different sizes, but their traces agree.

Definition 0.35 (Matrix inner product)**.** The **inner product** of two $n \times p$ matrices $\mathbf{A}$ and $\mathbf{B}$ is

$$\langle \mathbf{A}, \mathbf{B} \rangle \overset{\text{def}}{=} \operatorname{tr}\left(\underbrace{\mathbf{A}^\top \mathbf{B}}_{p \times p} \right)$$

Banerjee and Roy (2014, chap. 15, eq. 15.7, p. 491) defines the same inner product on $n \times p$ matrices with the trace.

Theorem 0.20 (The matrix inner product multiplies matching entries)**.** *For two $n \times p$ matrices* $\mathbf{A}$ *and* $\mathbf{B}$*:*

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^{n} \sum_{j=1}^{p} a_{ij}\, b_{ij}$$

Proof

Proof.

$$\begin{aligned}
\langle \mathbf{A}, \mathbf{B} \rangle &= \operatorname{tr}(\mathbf{A}^\top \mathbf{B}) && \text{(definition of the inner product)} \\
&= \sum_{j=1}^{p} (\mathbf{A}^\top \mathbf{B})_{jj} && \text{(definition of the trace)} \\
&= \sum_{j=1}^{p} \sum_{i=1}^{n} (\mathbf{A}^\top)_{ji}\, b_{ij} && \text{(definition of matrix multiplication)} \\
&= \sum_{j=1}^{p} \sum_{i=1}^{n} a_{ij}\, b_{ij} && \text{(definition of the transpose)} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{p} a_{ij}\, b_{ij} && \text{(swap the order of two finite sums)}
\end{aligned}$$

□

Theorem 0.20 says that the matrix inner product is the dot product (Definition 0.4) of the two matrices' entries, each listed as one vector of length np.

Example 0.19 (The inner product of two 2×2 matrices)**.** Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}$$

From Definition 0.35:

$$
\begin{aligned}
\langle \mathbf{A}, \mathbf{B} \rangle &= \operatorname{tr}(\mathbf{A}^\top \mathbf{B}) && \text{(definition of the inner product)} \\
&= \operatorname{tr}\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^\top \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix} \right) && \text{(substitute)} \\
&= \operatorname{tr}\left(\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix} \right) && \text{(definition of the transpose)} \\
&= \operatorname{tr}\left(\begin{bmatrix} -3 & 7 \\ -4 & 10 \end{bmatrix} \right) && \text{(multiply)} \\
&= -3 + 10 && \text{(definition of the trace)} \\
&= 7 && \text{(add)}
\end{aligned}
$$

From Theorem 0.20:

$$
\begin{aligned}
\langle \mathbf{A}, \mathbf{B} \rangle &= 1 \cdot 0 + 2 \cdot 1 + 3 \cdot (-1) + 4 \cdot 2 && \text{(multiply matching entries and add)} \\
&= 0 + 2 - 3 + 8 && \text{(multiply)} \\
&= 7 && \text{(add)}
\end{aligned}
$$

Definition 0.36 (Frobenius norm)**.** The **Frobenius norm** of an $n \times p$ matrix $\mathbf{A}$ is

$$
\|\mathbf{A}\|_F \stackrel{\text{def}}{=} \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle}
$$

where $\langle \cdot, \cdot \rangle$ is the matrix inner product (Definition 0.35).

By Theorem 0.20, $\|\mathbf{A}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p a_{ij}^2$, a sum of squares, so the square root is always defined. $\|\mathbf{A}\|_F$ is the Euclidean norm (Definition 0.9) of the entries of $\mathbf{A}$, listed as one vector of length np (Banerjee and Roy 2014, chap. 15, Definition 15.4, p. 492).

Example 0.20 (The Frobenius norm of a 2×2 matrix)**.** For $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$:

$$
\begin{aligned}
\|\mathbf{A}\|_F &= \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} && \text{(definition of the Frobenius norm)} \\
&= \sqrt{1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 + 4 \cdot 4} && \text{(sum of products of matching entries)} \\
&= \sqrt{1 + 4 + 9 + 16} && \text{(multiply)} \\
&= \sqrt{30} && \text{(add)}
\end{aligned}
$$

The second step is Theorem 0.20.

Matrix Decompositions

Definition 0.37 (Eigenvalue and eigenvector)**.** Let $\mathbf{A}$ be a $p \times p$ matrix. A real number λ is an **eigenvalue** of $\mathbf{A}$ if some real vector $\tilde{v} \neq \tilde{0}$ of length p satisfies

$$\underbrace{\mathbf{A}}_{p\times p}\ \underbrace{\tilde{v}}_{p\times 1} = \lambda\ \underbrace{\tilde{v}}_{p\times 1}$$

Any such $\tilde{v}$ is an **eigenvector** of $\mathbf{A}$ for λ.

Multiplying by $\mathbf{A}$ rescales an eigenvector by λ and does not change the line it lies on. Any nonzero multiple $c\,\tilde{v}$ of an eigenvector is also an eigenvector for the same λ, because $\mathbf{A}(c\,\tilde{v}) = c\,\mathbf{A}\tilde{v} = c\,\lambda\tilde{v} = \lambda\,(c\,\tilde{v})$.

These notes take λ and $\tilde{v}$ to be real. Banerjee and Roy (2014, chap. 11, Definition 11.1, pp. 312-313) also allows complex eigenvalues and eigenvectors, because some real matrices, such as a rotation by 90 degrees, have no real eigenvalues (Banerjee and Roy 2014, chap. 11, Example 11.2, p. 312).

Example 0.21 (Eigenvectors of a 2×2 matrix)**.** Let $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

$(1, 1)$ is an eigenvector for the eigenvalue 3:

$$\begin{aligned} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} 2+1 \\ 1+2 \end{bmatrix} && \text{(definition of matrix-vector multiplication)} \\ &= 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} && \text{(factor out 3)} \end{aligned}$$

$(1, -1)$ is an eigenvector for the eigenvalue 1:

$$\begin{aligned} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} &= \begin{bmatrix} 2-1 \\ 1-2 \end{bmatrix} && \text{(definition of matrix-vector multiplication)} \\ &= 1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} && \text{(factor out 1)} \end{aligned}$$

Example 0.22 (A vector that is not an eigenvector)**.** For the same $\mathbf{A}$, $(1, 0)$ is not an eigenvector:

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

and no number λ gives $(2, 1) = \lambda\,(1, 0)$, because the second entries would need $1 = \lambda \cdot 0$.

Theorem 0.21 (Spectral theorem for symmetric matrices)**.** *If* $\mathbf{A}$ *is a* $p \times p$ *symmetric matrix (Definition 0.25) with real entries, then there are a* $p \times p$ *orthogonal matrix* $\mathbf{Q}$ *(Definition 0.31) and a* $p \times p$ *diagonal matrix (Definition 0.26) with real diagonal entries* $\lambda_1, \dots, \lambda_p$ *such that*

$$\underbrace{\mathbf{A}}_{p\times p} = \underbrace{\mathbf{Q}}_{p\times p} \underbrace{}_{p\times p} \underbrace{\mathbf{Q}^\top}_{p\times p}$$

Each λ_i *is an eigenvalue of* $\mathbf{A}$ *(Definition 0.37), and column* i *of* $\mathbf{Q}$ *is an eigenvector of* $\mathbf{A}$ *for* λ_i*.*

The proof, by induction on p, is outside the scope of these notes; see Banerjee and Roy (2014, chap. 11, Theorem 11.27, p. 349), which states the result in the equivalent form $\mathbf{Q}^\top \mathbf{A} \mathbf{Q} = .$ The two forms are equivalent because $\mathbf{Q}^\top \mathbf{Q} = \mathbf{Q}\mathbf{Q}^\top = \mathbf{I}_p$: multiply $\mathbf{A} = \mathbf{Q}\mathbf{Q}^\top$ by $\mathbf{Q}^\top$ on the left and by $\mathbf{Q}$ on the right.

Definition 0.38 (Eigendecomposition)**.** Let $\mathbf{A}$ be a $p \times p$ symmetric matrix with real entries. An **eigendecomposition**, or **spectral decomposition**, of $\mathbf{A}$ is a factorization

$$\mathbf{A} = \mathbf{Q}\mathbf{Q}^\top$$

with $\mathbf{Q}$ a $p \times p$ orthogonal matrix (Definition 0.31) and a $p \times p$ diagonal matrix (Definition 0.26). Theorem 0.21 says that every such $\mathbf{A}$ has one.

An eigendecomposition is not unique. For example, reordering the eigenvalues on the diagonal of , and the columns of $\mathbf{Q}$ with them, gives another one, and so does multiplying a column of $\mathbf{Q}$ by -1.

Example 0.23 (An eigendecomposition of a 2×2 symmetric matrix)**.** For $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, Example 0.21 found the eigenvectors $(1, 1)$ for 3 and $(1, -1)$ for 1. They are orthogonal (Definition 0.8):

$$\begin{aligned}
(1,1) \cdot (1,-1) &= 1 \cdot 1 + 1 \cdot (-1) && \text{(definition of the dot product)} \\
&= 1 - 1 && \text{(multiply)} \\
&= 0 && \text{(add)}
\end{aligned}$$

Each has norm $\sqrt{1^2 + 1^2} = \sqrt{2}$ (Definition 0.9), so dividing each by $\sqrt{2}$ gives two orthonormal eigenvectors (Definition 0.10). Put them in the columns of $\mathbf{Q}$, and the matching eigenvalues on the diagonal of :

$$\mathbf{Q} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

This $\mathbf{Q}$ is symmetric, so $\mathbf{Q}^\top = \mathbf{Q}$. Then

$$\begin{aligned}
\mathbf{Q}\mathbf{Q}^\top &= \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} && \text{(substitute)}\\
&= \frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} && \text{(move the scalars to the front)}\\
&= \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} && \text{(multiply the scalars)}\\
&= \frac{1}{2}\begin{bmatrix}1\cdot 3 + 1\cdot 0 & 1\cdot 0 + 1\cdot 1\\ 1\cdot 3 + (-1)\cdot 0 & 1\cdot 0 + (-1)\cdot 1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} && \text{(definition of matrix multiplication, first two matrices)}\\
&= \frac{1}{2}\begin{bmatrix}3+0 & 0+1\\ 3+0 & 0-1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} && \text{(multiply)}\\
&= \frac{1}{2}\begin{bmatrix}3 & 1\\ 3 & -1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} && \text{(add)}\\
&= \frac{1}{2}\begin{bmatrix}3\cdot 1 + 1\cdot 1 & 3\cdot 1 + 1\cdot(-1)\\ 3\cdot 1 + (-1)\cdot 1 & 3\cdot 1 + (-1)\cdot(-1)\end{bmatrix} && \text{(definition of matrix multiplication)}\\
&= \frac{1}{2}\begin{bmatrix}3+1 & 3-1\\ 3-1 & 3+1\end{bmatrix} && \text{(multiply)}\\
&= \frac{1}{2}\begin{bmatrix}4 & 2\\ 2 & 4\end{bmatrix} && \text{(add)}\\
&= \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} && \text{(multiply by } \tfrac{1}{2}\text{)}
\end{aligned}$$

which is $\mathbf{A}$.

Theorem 0.22 (Singular value decomposition)**.** *Let* $\mathbf{A}$ *be an* $n \times p$ *matrix with real entries and* $\operatorname{rank}(\mathbf{A}) = r$ *(Definition 0.19). Then there are an* $n \times n$ *orthogonal matrix* $\mathbf{U}$*, a* $p \times p$ *orthogonal matrix* $\mathbf{V}$ *(Definition 0.31), and numbers* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ *such that*

$$\underbrace{\mathbf{A}}_{n\times p} = \underbrace{\mathbf{U}}_{n\times n}\ \underbrace{\mathbf{D}}_{n\times p}\ \underbrace{\mathbf{V}^\top}_{p\times p}$$

where $\mathbf{D}$ *has entries* $d_{ii} = \sigma_i$ *for* $i = 1, \ldots, r$ *and every other entry* 0*.*

The proof is outside the scope of these notes; see Banerjee and Roy (2014, chap. 12, Theorem 12.1, p. 373). Unlike the spectral theorem (Theorem 0.21), Theorem 0.22 applies to every real matrix, including one that is not square or not symmetric.

Definition 0.39 (Singular value decomposition and singular values)**.** Let $\mathbf{A}$ be an $n \times p$ matrix with real entries and $\text{rank}(\mathbf{A}) = r$. A **singular value decomposition** (SVD) of $\mathbf{A}$ is a factorization $\mathbf{A} = \mathbf{UDV}^\top$ with $\mathbf{U}$, $\mathbf{D}$ and $\mathbf{V}$ as in Theorem 0.22. The numbers $\sigma_1 \geq \cdots \geq \sigma_r > 0$ on the diagonal of $\mathbf{D}$ are the **singular values** of $\mathbf{A}$.

An SVD is not unique (Banerjee and Roy 2014, chap. 12, Examples 12.2 and 12.3, p. 378). For example, for any $i \leq r$, multiplying column i of both $\mathbf{U}$ and $\mathbf{V}$ by -1 gives another one. The singular values do not depend on which SVD is chosen (Banerjee and Roy 2014, chap. 12, pp. 371 and 378).

Some texts, and R's `svd()`, also count $\min(n, p) - r$ singular values equal to 0, so that every $n \times p$ matrix has $\min(n, p)$ singular values.

Example 0.24 (An SVD of a 3×2 matrix)**.** Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \qquad \mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \qquad \mathbf{V} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$\mathbf{V}$ is the orthogonal matrix $\mathbf{Q}$ from Example 0.23. $\mathbf{U}$ is orthogonal too. Entry (i, j) of $\mathbf{U}^\top\mathbf{U}$ is the dot product of columns i and j of $\mathbf{U}$, and those columns are $\tilde{u}_1 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$, $\tilde{u}_2 = (0, 1, 0)$ and $\tilde{u}_3 = (\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}})$:

$$\begin{aligned} \tilde{u}_1 \cdot \tilde{u}_1 &= \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{\sqrt{2}} + 0 \cdot 0 + \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{\sqrt{2}} && \text{(definition of the dot product)} \\ &= \tfrac{1}{2} + 0 + \tfrac{1}{2} && \text{(multiply)} \\ &= 1 && \text{(add)} \end{aligned}$$

$$\begin{aligned} \tilde{u}_2 \cdot \tilde{u}_2 &= 0 \cdot 0 + 1 \cdot 1 + 0 \cdot 0 && \text{(definition of the dot product)} \\ &= 0 + 1 + 0 && \text{(multiply)} \\ &= 1 && \text{(add)} \end{aligned}$$

$$\begin{aligned} \tilde{u}_3 \cdot \tilde{u}_3 &= \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{\sqrt{2}} + 0 \cdot 0 + \left(-\tfrac{1}{\sqrt{2}}\right) \cdot \left(-\tfrac{1}{\sqrt{2}}\right) && \text{(definition of the dot product)} \\ &= \tfrac{1}{2} + 0 + \tfrac{1}{2} && \text{(multiply)} \\ &= 1 && \text{(add)} \end{aligned}$$

$$\begin{aligned} \tilde{u}_1 \cdot \tilde{u}_2 &= \tfrac{1}{\sqrt{2}} \cdot 0 + 0 \cdot 1 + \tfrac{1}{\sqrt{2}} \cdot 0 && \text{(definition of the dot product)} \\ &= 0 + 0 + 0 && \text{(multiply)} \\ &= 0 && \text{(add)} \end{aligned}$$

$$
\begin{aligned}
\tilde{u}_1 \cdot \tilde{u}_3 &= \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{\sqrt{2}} + 0 \cdot 0 + \tfrac{1}{\sqrt{2}} \cdot \left(-\tfrac{1}{\sqrt{2}}\right) && \text{(definition of the dot product)} \\
&= \tfrac{1}{2} + 0 - \tfrac{1}{2} && \text{(multiply)} \\
&= 0 && \text{(add)}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{u}_2 \cdot \tilde{u}_3 &= 0 \cdot \tfrac{1}{\sqrt{2}} + 1 \cdot 0 + 0 \cdot \left(-\tfrac{1}{\sqrt{2}}\right) && \text{(definition of the dot product)} \\
&= 0 + 0 + 0 && \text{(multiply)} \\
&= 0 && \text{(add)}
\end{aligned}
$$

The dot product is symmetric, so these six values fill in all nine entries, and $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_3$. $\mathbf{V}$ is symmetric, so $\mathbf{V}^\top = \mathbf{V}$. Then

$$
\begin{aligned}
\mathbf{D}\mathbf{V}^\top &= \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(substitute)} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(move the scalar to the front)} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 2 \cdot 1 + 0 \cdot 1 & 2 \cdot 1 + 0 \cdot (-1) \\ 0 \cdot 1 + \sqrt{2} \cdot 1 & 0 \cdot 1 + \sqrt{2} \cdot (-1) \\ 0 \cdot 1 + 0 \cdot 1 & 0 \cdot 1 + 0 \cdot (-1) \end{bmatrix} && \text{(definition of matrix multiplication)} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 2 + 0 & 2 + 0 \\ 0 + \sqrt{2} & 0 - \sqrt{2} \\ 0 + 0 & 0 + 0 \end{bmatrix} && \text{(multiply)} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 2 \\ \sqrt{2} & -\sqrt{2} \\ 0 & 0 \end{bmatrix} && \text{(add)} \\
&= \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 1 & -1 \\ 0 & 0 \end{bmatrix} && \text{(divide by } \sqrt{2}\text{)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{U}\mathbf{D}\mathbf{V}^\top &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 1 & -1 \\ 0 & 0 \end{bmatrix} && \text{(substitute)} \\
&= \begin{bmatrix} \frac{1}{\sqrt{2}} \cdot \sqrt{2} + 0 \cdot 1 + \frac{1}{\sqrt{2}} \cdot 0 & \frac{1}{\sqrt{2}} \cdot \sqrt{2} + 0 \cdot (-1) + \frac{1}{\sqrt{2}} \cdot 0 \\ 0 \cdot \sqrt{2} + 1 \cdot 1 + 0 \cdot 0 & 0 \cdot \sqrt{2} + 1 \cdot (-1) + 0 \cdot 0 \\ \frac{1}{\sqrt{2}} \cdot \sqrt{2} + 0 \cdot 1 - \frac{1}{\sqrt{2}} \cdot 0 & \frac{1}{\sqrt{2}} \cdot \sqrt{2} + 0 \cdot (-1) - \frac{1}{\sqrt{2}} \cdot 0 \end{bmatrix} && \text{(definition of matrix multiplication)} \\
&= \begin{bmatrix} 1+0+0 & 1+0+0 \\ 0+1+0 & 0-1+0 \\ 1+0-0 & 1+0-0 \end{bmatrix} && \text{(multiply)} \\
&= \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} && \text{(add)}
\end{aligned}
$$

which is $\mathbf{A}$. $\mathbf{A}$ has rank 2, because its two columns are linearly independent, and its singular values are $\sigma_1 = 2$ and $\sigma_2 = \sqrt{2}$.

Theorem 0.23 (An SVD gives an eigendecomposition of $\mathbf{A}^\top\mathbf{A}$)**.** *Let $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ be a singular value decomposition (Definition 0.39) of an $n \times p$ matrix $\mathbf{A}$ with singular values $\sigma_1, \ldots, \sigma_r$. Then $\mathbf{A}^\top\mathbf{A}$ is symmetric (Definition 0.25), and*

$$
\underbrace{\mathbf{A}^\top\mathbf{A}}_{p\times p} = \underbrace{\mathbf{V}}_{p\times p} \underbrace{}_{p\times p} \underbrace{\mathbf{V}^\top}_{p\times p}
$$

where $\overset{def}{=} \mathbf{D}^\top\mathbf{D}$ is the $p \times p$ diagonal matrix whose i-th diagonal entry λ_i is σ_i^2 for $i \leq r$ and 0 for $i > r$. So $\mathbf{V}\mathbf{V}^\top$ is an eigendecomposition (Definition 0.38) of $\mathbf{A}^\top\mathbf{A}$: column i of $\mathbf{V}$ is an eigenvector of $\mathbf{A}^\top\mathbf{A}$ for the eigenvalue λ_i.

Proof

Proof. **Symmetry.**

$$
\begin{aligned}
\left(\mathbf{A}^\top\mathbf{A}\right)^\top &= \mathbf{A}^\top \left(\mathbf{A}^\top\right)^\top && \text{(transpose of a product)} \\
&= \mathbf{A}^\top\mathbf{A} && \text{(transposing twice changes nothing)}
\end{aligned}
$$

The first step is Theorem 0.10, and the second follows from Definition 0.12, which swaps rows and columns, so swapping them again restores $\mathbf{A}$.
The factorization.

$$
\begin{aligned}
\mathbf{A}^\top \mathbf{A} &= \left(\mathbf{U}\mathbf{D}\mathbf{V}^\top\right)^\top \mathbf{U}\mathbf{D}\mathbf{V}^\top && \text{(substitute the SVD)} \\
&= \left(\mathbf{D}\mathbf{V}^\top\right)^\top \mathbf{U}^\top \mathbf{U}\mathbf{D}\mathbf{V}^\top && \text{(transpose of the product of } \mathbf{U} \text{ and } \mathbf{D}\mathbf{V}^\top\text{)} \\
&= \left(\mathbf{V}^\top\right)^\top \mathbf{D}^\top \mathbf{U}^\top \mathbf{U}\mathbf{D}\mathbf{V}^\top && \text{(transpose of the product of } \mathbf{D} \text{ and } \mathbf{V}^\top\text{)} \\
&= \mathbf{V}\, \mathbf{D}^\top \mathbf{U}^\top \mathbf{U}\mathbf{D}\mathbf{V}^\top && \text{(transposing twice changes nothing)} \\
&= \mathbf{V}\, \mathbf{D}^\top \left(\mathbf{U}^\top \mathbf{U}\right) \mathbf{D}\mathbf{V}^\top && \text{(regroup; matrix multiplication is associative)} \\
&= \mathbf{V}\, \mathbf{D}^\top \mathbf{I}_n\, \mathbf{D}\mathbf{V}^\top && \text{(}\mathbf{U} \text{ is orthogonal)} \\
&= \mathbf{V}\left(\mathbf{D}^\top \mathbf{D}\right) \mathbf{V}^\top && \text{(identity matrix)} \\
&= \mathbf{V}\mathbf{V}^\top && \text{(definition of)}
\end{aligned}
$$

The entries of . Write λ_{ij} for entry (i, j) of , so $\lambda_i = \lambda_{ii}$. Only the entries d_{kk} with $k \le r$ of $\mathbf{D}$ can be nonzero.

$$
\begin{aligned}
\lambda_{ij} &= \sum_{k=1}^{n} (\mathbf{D}^\top)_{ik}\, d_{kj} && \text{(definition of matrix multiplication)} \\
&= \sum_{k=1}^{n} d_{ki}\, d_{kj} && \text{(definition of the transpose)}
\end{aligned}
$$

The term for k is nonzero only if $k = i \le r$ and $k = j$. So $\lambda_{ij} = 0$ when $i \ne j$. On the diagonal, only the term $k = i$ can be nonzero, and for $i \le r$

$$
\begin{aligned}
\lambda_i &= \lambda_{ii} && \text{(notation)} \\
&= d_{ii}\, d_{ii} && \text{(the only term that can be nonzero)} \\
&= \sigma_i \cdot \sigma_i && \text{(definition of } \mathbf{D}\text{)} \\
&= \sigma_i^2 && \text{(notation for a square)}
\end{aligned}
$$

For $i > r$, every term is 0, so $\lambda_i = 0$.
The eigenvectors. Write $\tilde{v}_i$ for column i of $\mathbf{V}$, and $\tilde{e}_i$ for the indicator vector (Definition 0.7) of length p. Column i of $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_p$ says $\mathbf{V}^\top \tilde{v}_i = \tilde{e}_i$. Then

$$
\begin{aligned}
\mathbf{A}^\top \mathbf{A}\, \tilde{v}_i &= \mathbf{V}\, \mathbf{V}^\top \tilde{v}_i && \text{(the factorization)} \\
&= \mathbf{V}\, \tilde{e}_i && (\mathbf{V}^\top \tilde{v}_i = \tilde{e}_i) \\
&= \mathbf{V}\left(\lambda_i\, \tilde{e}_i\right) && \text{(column } i \text{ of the diagonal matrix)} \\
&= \lambda_i\, \mathbf{V}\tilde{e}_i && \text{(move the scalar } \lambda_i \text{ to the front)} \\
&= \lambda_i\, \tilde{v}_i && (\mathbf{V}\tilde{e}_i \text{ is column } i \text{ of } \mathbf{V})
\end{aligned}
$$

and $\tilde{v}_i \ne \tilde{0}$ because it has norm 1. □

Theorem 0.23 matches Banerjee and Roy (2014, chap. 12, pp. 372-373), which constructs an SVD from a spectral decomposition of $\mathbf{A}^\top \mathbf{A}$ and sets $\sigma_i = \sqrt{\lambda_i}$.

Example 0.25 ($\mathbf{A}^\top\mathbf{A}$ for the matrix of the SVD example)**.** For $\mathbf{A}$ in Example 0.24:

$$\begin{aligned}
\mathbf{A}^\top\mathbf{A} &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} && \text{(definition of the transpose)} \\
&= \begin{bmatrix} 1+1+1 & 1-1+1 \\ 1-1+1 & 1+1+1 \end{bmatrix} && \text{(definition of matrix multiplication)} \\
&= \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} && \text{(add)}
\end{aligned}$$

Theorem 0.23 says its eigenvalues are $\sigma_1^2 = 4$ and $\sigma_2^2 = 2$, with the columns of $\mathbf{V}$ as eigenvectors. Checking the first column, up to its factor $\frac{1}{\sqrt{2}}$:

$$\begin{aligned}
\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} 3 \cdot 1 + 1 \cdot 1 \\ 1 \cdot 1 + 3 \cdot 1 \end{bmatrix} && \text{(definition of matrix-vector multiplication)} \\
&= \begin{bmatrix} 3+1 \\ 1+3 \end{bmatrix} && \text{(multiply)} \\
&= \begin{bmatrix} 4 \\ 4 \end{bmatrix} && \text{(add)} \\
&= 4 \begin{bmatrix} 1 \\ 1 \end{bmatrix} && \text{(factor out 4)}
\end{aligned}$$

and the second:

$$\begin{aligned}
\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} &= \begin{bmatrix} 3 \cdot 1 + 1 \cdot (-1) \\ 1 \cdot 1 + 3 \cdot (-1) \end{bmatrix} && \text{(definition of matrix-vector multiplication)} \\
&= \begin{bmatrix} 3-1 \\ 1-3 \end{bmatrix} && \text{(multiply)} \\
&= \begin{bmatrix} 2 \\ -2 \end{bmatrix} && \text{(add)} \\
&= 2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} && \text{(factor out 2)}
\end{aligned}$$

Definite Matrices

Definition 0.40 (Positive semidefinite matrix)**.** A $p \times p$ matrix $\mathbf{A}$ is **positive semidefinite** if it satisfies both conditions:

- $\mathbf{A}$ is symmetric (Definition 0.25).

- Every quadratic form (Definition 0.32) in $\mathbf{A}$ is non-negative: $\tilde{x}^\top \mathbf{A}\tilde{x} \geq 0$ for every vector $\tilde{x}$ of length p.

Example 0.26 (A positive semidefinite matrix)**.** Let $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, which is symmetric. For any $\tilde{x} = (x_1, x_2)$:

$$\begin{aligned} \tilde{x}^\top \mathbf{B}\tilde{x} &= x_1^2 + x_1x_2 + x_2x_1 + x_2^2 && \text{(multiply out the quadratic form)} \\ &= (x_1 + x_2)^2 && \text{(complete the square)} \\ &\geq 0 && \text{(a square is non-negative)} \end{aligned}$$

So $\mathbf{B}$ is positive semidefinite.

Definition 0.41 (Positive definite matrix)**.** A $p \times p$ matrix $\mathbf{A}$ is **positive definite** if it satisfies both conditions:

- $\mathbf{A}$ is symmetric (Definition 0.25).
- Every quadratic form (Definition 0.32) in $\mathbf{A}$ at a nonzero vector is positive: $\tilde{x}^\top \mathbf{A}\tilde{x} > 0$ for every vector $\tilde{x} \neq \tilde{0}$ of length p.

Remark. A positive definite matrix is positive semidefinite (Definition 0.40), because $\tilde{0}^\top \mathbf{A}\tilde{0} = 0$. Some sources drop the symmetry condition from both definitions. These notes keep it, because without it a matrix can pass the quadratic-form condition and still have no real eigenvalues: $\mathbf{C} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ has $\tilde{x}^\top \mathbf{C}\tilde{x} = x_1^2 + x_2^2$, which is positive for every $\tilde{x} \neq \tilde{0}$, yet no real λ solves $\mathbf{C}\tilde{v} = \lambda\tilde{v}$ with $\tilde{v} \neq \tilde{0}$ (see also https://en.wikipedia.org/wiki/Definite_matrix).

Example 0.27 (Positive definite, semidefinite, and neither)**.**

- The identity matrix $\mathbf{I}_p$ (Definition 0.24) is positive definite: $\tilde{x}^\top \mathbf{I}_p\tilde{x} = \sum_{i=1}^p x_i^2$, which is positive unless every x_i is 0.
- $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ from Example 0.26 is positive semidefinite but not positive definite: at $\tilde{x} = (1, -1) \neq \tilde{0}$, $\tilde{x}^\top \mathbf{B}\tilde{x} = (1-1)^2 = 0$.
- $\mathbf{D} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is symmetric but not positive semidefinite: at $\tilde{x} = (1, -1)$, $\tilde{x}^\top \mathbf{D}\tilde{x} = 1 - 2 - 2 + 1 = -2 < 0$.

Theorem 0.24 (Definiteness and eigenvalues)**.** *Let* $\mathbf{A}$ *be a* $p \times p$ *symmetric matrix with real entries, with eigendecomposition* $\mathbf{A} = \mathbf{Q}\mathbf{Q}^\top$ *(Definition 0.38) and eigenvalues* $\lambda_1, \dots, \lambda_p$ *on the diagonal of . Then:*

- $\mathbf{A}$ *is positive semidefinite (Definition 0.40) if and only if every* $\lambda_i \geq 0$*.*
- $\mathbf{A}$ *is positive definite (Definition 0.41) if and only if every* $\lambda_i > 0$*.*

Proof

Proof. For any vector $\tilde{x}$ of length p, let $\tilde{y} = \mathbf{Q}^\top \tilde{x}$. Then:

$$\begin{aligned}
\tilde{x}^\top \mathbf{A} \tilde{x} &= \tilde{x}^\top \mathbf{Q}\mathbf{Q}^\top \tilde{x} && \text{(substitute the eigendecomposition)} \\
&= (\mathbf{Q}^\top \tilde{x})^\top (\mathbf{Q}^\top \tilde{x}) && \text{(transpose of a product)} \\
&= \tilde{y}^\top \tilde{y} && \text{(definition of } \tilde{y}\text{)} \\
&= \sum_{i=1}^{p} \lambda_i y_i^2 && \text{(is diagonal)}
\end{aligned}$$

The second step is Theorem 0.10. Also, $\tilde{x} = \tilde{0}$ exactly when $\tilde{y} = \tilde{0}$: $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}_p$ for an orthogonal matrix (Definition 0.31), so $\tilde{x} = \mathbf{Q}\tilde{y}$.
If every $\lambda_i \geq 0$, then every term $\lambda_i y_i^2 \geq 0$, so $\tilde{x}^\top \mathbf{A} \tilde{x} \geq 0$. *If every* $\lambda_i > 0$ and $\tilde{x} \neq \tilde{0}$, then some $y_i \neq 0$, so at least one term is positive and the rest are non-negative, and $\tilde{x}^\top \mathbf{A} \tilde{x} > 0$.
Conversely, take $\tilde{x} = \tilde{q}_i$, column i of $\mathbf{Q}$, which is not $\tilde{0}$ because it has norm 1. Then $\tilde{y} = \mathbf{Q}^\top \tilde{q}_i$ is column i of $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_p$, so $y_i = 1$ and every other entry of $\tilde{y}$ is 0, and the display gives $\tilde{q}_i^\top \mathbf{A} \tilde{q}_i = \lambda_i$. So if $\mathbf{A}$ is positive semidefinite, $\lambda_i \geq 0$, and if $\mathbf{A}$ is positive definite, $\lambda_i > 0$. □

Example 0.28 (Reading definiteness off the eigenvalues)**.**

- $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has eigenvalues 3 and 1 (Example 0.21), both positive, so it is positive definite.
- $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ has eigenvalues 2, for the eigenvector $(1, 1)$, and 0, for the eigenvector $(1, -1)$, so it is positive semidefinite but not positive definite, as Example 0.27 found directly.

Theorem 0.25 (A positive definite matrix has a positive definite inverse)**.** *Let* $\mathbf{A}$ *be a* $p \times p$ *positive definite matrix (Definition 0.41) with real entries, with eigendecomposition* $\mathbf{A} = \mathbf{Q}\mathbf{Q}^\top$ *(Definition 0.38) and eigenvalues* $\lambda_1, \dots, \lambda_p$*. Then* $\mathbf{A}$ *is invertible (Definition 0.28),*

$$\mathbf{A}^{-1} = \mathbf{Q}\,^{-1}\,\mathbf{Q}^\top, \qquad {}^{-1} = \begin{bmatrix} 1/\lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\lambda_p \end{bmatrix},$$

and $\mathbf{A}^{-1}$ *is positive definite.*

Proof

Proof. By Theorem 0.24, every $\lambda_i > 0$, so ${}^{-1}$ is well defined, and multiplying the two diagonal matrices entry by entry gives ${}^{-1} = {}^{-1} = \mathbf{I}_p$. Write $\mathbf{B} = \mathbf{Q}^{-1}\mathbf{Q}^\top$. Then:

$$\begin{aligned} \mathbf{AB} &= \mathbf{Q}\mathbf{Q}^\top\mathbf{Q}^{-1}\mathbf{Q}^\top && \text{(substitute)} \\ &= \mathbf{Q}\mathbf{I}_p{}^{-1}\mathbf{Q}^\top && (\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_p) \\ &= \mathbf{Q}^{-1}\mathbf{Q}^\top && \text{(identity matrix)} \\ &= \mathbf{Q}\mathbf{Q}^\top && ({}^{-1} = \mathbf{I}_p) \\ &= \mathbf{I}_p && (\mathbf{Q} \text{ is orthogonal}) \end{aligned}$$

The same steps, with ${}^{-1}$ and swapped, give $\mathbf{BA} = \mathbf{I}_p$, so $\mathbf{B} = \mathbf{A}^{-1}$ (Definition 0.27). The steps with $\mathbf{Q}$ use Definition 0.31, whose remark records that $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}_p$ too, and the identity step uses Theorem 0.12.
$\mathbf{A}^{-1}$ is symmetric by Corollary 0.1. For $\tilde{x} \neq \tilde{0}$, let $\tilde{y} = \mathbf{Q}^\top\tilde{x}$, which is not $\tilde{0}$ because $\tilde{x} = \mathbf{Q}\tilde{y}$. As in the proof of Theorem 0.24:

$$\begin{aligned} \tilde{x}^\top\mathbf{A}^{-1}\tilde{x} &= \tilde{y}^\top{}^{-1}\tilde{y} && (\text{substitute } \mathbf{A}^{-1} = \mathbf{Q}^{-1}\mathbf{Q}^\top) \\ &= \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} && ({}^{-1} \text{ is diagonal}) \\ &> 0 && (\text{each } \lambda_i > 0, \text{ and some } y_i \neq 0) \end{aligned}$$

So $\mathbf{A}^{-1}$ is positive definite. □

Example 0.29 (Inverting a positive definite matrix)**.** For $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, Example 0.23 gives $\mathbf{Q} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and eigenvalues 3 and 1, so:

$$
\begin{aligned}
\mathbf{A}^{-1} &= \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1/3 & 0\\ 0 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} && \text{(substitute; } \mathbf{Q} \text{ is symmetric)}\\
&= \frac{1}{2}\begin{bmatrix}1/3 & 1\\ 1/3 & -1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} && \text{(multiply the first two matrices)}\\
&= \frac{1}{2}\begin{bmatrix}4/3 & -2/3\\ -2/3 & 4/3\end{bmatrix} && \text{(multiply)}\\
&= \frac{1}{3}\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix} && \text{(simplify)}
\end{aligned}
$$

Multiplying out, $\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \cdot \frac{1}{3}\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix} = \frac{1}{3}\begin{bmatrix}3 & 0\\ 0 & 3\end{bmatrix} = \mathbf{I}_2$.

Determinants

Definition 0.42 (Determinant)**.** The **determinant** of a $p \times p$ matrix $\mathbf{A}$ with entries a_{ij}, written $\det(\mathbf{A})$ or $|\mathbf{A}|$, is the number defined recursively in p:

- For $p = 1$, $\det(\mathbf{A}) = a_{11}$.
- For $p \geq 2$,

  $$\det(\mathbf{A}) = \sum_{j=1}^{p} (-1)^{1+j}\, a_{1j} \det\big(\mathbf{A}_{(1j)}\big),$$

  where $\mathbf{A}_{(1j)}$ is the $(p-1) \times (p-1)$ matrix left after deleting row 1 and column j of $\mathbf{A}$.

Remark. This recursive formula is the *cofactor expansion* along the first row. Other sources define the determinant as a sum over all orderings of the columns and derive this expansion from it; the two definitions agree [Banerjee and Roy (2014); see also https://en.wikipedia.org/wiki/Determinant].

Example 0.30 (Determinant of a 2×2 matrix)**.** For $\mathbf{A} = \begin{bmatrix}a & b\\ c & d\end{bmatrix}$, deleting row 1 and column 1 leaves $\begin{bmatrix}d\end{bmatrix}$, and deleting row 1 and column 2 leaves $\begin{bmatrix}c\end{bmatrix}$. So:

$$
\begin{aligned}
\det(\mathbf{A}) &= (-1)^{1+1}\, a \det\big(\begin{bmatrix}d\end{bmatrix}\big) + (-1)^{1+2}\, b \det\big(\begin{bmatrix}c\end{bmatrix}\big) && \text{(definition, } p = 2\text{)}\\
&= ad - bc && \text{(definition, } p = 1\text{)}
\end{aligned}
$$

For example, $\det\left(\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}\right) = 2 \cdot 2 - 1 \cdot 1 = 3$.

Theorem 0.26 (Determinant of a diagonal matrix)**.** *The determinant (Definition 0.42) of a $p \times p$ diagonal matrix (Definition 0.26) is the product of its diagonal entries:*

$$\det\left(\begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_p \end{bmatrix}\right) = \prod_{i=1}^{p} d_i$$

Proof

Proof. By induction on p. For $p = 1$, the matrix is $\left[d_1\right]$, and its determinant is d_1 by definition.
For $p \geq 2$, suppose the result holds for $(p-1) \times (p-1)$ diagonal matrices, and let $\mathbf{D}$ be $p \times p$ and diagonal. Row 1 of $\mathbf{D}$ is $(d_1, 0, \dots, 0)$, so only the $j = 1$ term of the definition can be nonzero, and deleting row 1 and column 1 of $\mathbf{D}$ leaves the diagonal matrix with diagonal entries $d_2, \dots, d_p$:

$$\begin{aligned}
\det(\mathbf{D}) &= (-1)^{1+1}\, d_1 \det\left(\mathbf{D}_{(11)}\right) + \sum_{j=2}^{p} (-1)^{1+j} \cdot 0 \cdot \det\left(\mathbf{D}_{(1j)}\right) && \text{(definition)} \\
&= d_1 \det\left(\mathbf{D}_{(11)}\right) && \text{(drop the zero terms)} \\
&= d_1 \prod_{i=2}^{p} d_i && \text{(induction hypothesis)} \\
&= \prod_{i=1}^{p} d_i && \text{(combine)}
\end{aligned}$$

□

Example 0.31 (Determinant of a scaled identity matrix)**.** For a number c, $c\,\mathbf{I}_p$ is diagonal with every diagonal entry equal to c, so $\det(c\,\mathbf{I}_p) = c^p$. In particular, $\det(\mathbf{I}_p) = 1$.

Theorem 0.27 (Determinant of a product)**.** *For $p \times p$ matrices $\mathbf{A}$ and $\mathbf{B}$,*

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B}).$$

Remark. The proof needs properties of the determinant that these notes don't develop; see Banerjee and Roy (2014) and https://en.wikipedia.org/wiki/Determinant.

Example 0.32 (Checking the product rule)**.** For $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$, Example 0.30 gives $\det(\mathbf{A}) = 2 \cdot 1 - 1 \cdot 0 = 2$ and $\det(\mathbf{B}) = 1 \cdot 1 - 0 \cdot 3 = 1$. The product is

$$\mathbf{AB} = \begin{bmatrix} 2 \cdot 1 + 1 \cdot 3 & 2 \cdot 0 + 1 \cdot 1 \\ 0 \cdot 1 + 1 \cdot 3 & 0 \cdot 0 + 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 3 & 1 \end{bmatrix},$$

with $\det(\mathbf{AB}) = 5 \cdot 1 - 1 \cdot 3 = 2 = \det(\mathbf{A})\det(\mathbf{B})$.

Theorem 0.28 (The determinant of a symmetric matrix is the product of its eigenvalues)**.** *Let* $\mathbf{A}$ *be a* $p \times p$ *symmetric matrix with real entries, with eigendecomposition* $\mathbf{A} = \mathbf{Q}\mathbf{Q}^\top$ *(Definition 0.38) and eigenvalues* $\lambda_1, \dots, \lambda_p$ *on the diagonal of . Then*

$$\det(\mathbf{A}) = \prod_{i=1}^{p} \lambda_i.$$

Proof

Proof. First, $\det(\mathbf{Q}^\top)\det(\mathbf{Q}) = 1$:

$$\begin{aligned} \det(\mathbf{Q}^\top)\det(\mathbf{Q}) &= \det(\mathbf{Q}^\top\mathbf{Q}) && \text{(determinant of a product)} \\ &= \det(\mathbf{I}_p) && \text{($\mathbf{Q}$ is orthogonal)} \\ &= 1 && \text{(determinant of the identity)} \end{aligned}$$

Then:

$$\begin{aligned} \det(\mathbf{A}) &= \det(\mathbf{Q}\mathbf{Q}^\top) && \text{(substitute the eigendecomposition)} \\ &= \det(\mathbf{Q})\det()\det(\mathbf{Q}^\top) && \text{(determinant of a product, twice)} \\ &= \det() \cdot \det(\mathbf{Q}^\top)\det(\mathbf{Q}) && \text{(reorder the three numbers)} \\ &= \det() && \text{(the first display)} \\ &= \prod_{i=1}^{p} \lambda_i && \text{(determinant of a diagonal matrix)} \end{aligned}$$

The product steps are Theorem 0.27, the orthogonality step is Definition 0.31, and the identity and diagonal steps are Theorem 0.26 and Example 0.31. □

Corollary 0.2 (A positive definite matrix has a positive determinant)**.** *If* $\mathbf{A}$ *is positive definite (Definition 0.41), then* $\det(\mathbf{A}) > 0$.

Proof

Proof. By Theorem 0.24, every eigenvalue of $\mathbf{A}$ is positive, so their product, which is $\det(\mathbf{A})$ by Theorem 0.28, is positive. □

Example 0.33 (Two ways to the same determinant). $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has eigenvalues 3 and 1 (Example 0.21), so Theorem 0.28 gives $\det(\mathbf{A}) = 3 \cdot 1 = 3$, matching $2 \cdot 2 - 1 \cdot 1 = 3$ from Example 0.30.

Design Matrix

Definition 0.43 (Design matrix). In a regression model with n observations and p predictors, the **design matrix** (or *model matrix*) $\mathbf{X}$ is the $n \times p$ matrix whose i-th row is the covariate vector $\tilde{x}_i^\top$ for observation i:

$$\mathbf{X} = \begin{bmatrix} \tilde{x}_1^\top \\ \tilde{x}_2^\top \\ \vdots \\ \tilde{x}_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

The product $\mathbf{X}\tilde{\beta}$ collects all the linear predictors $\tilde{x}_i^\top \tilde{\beta}$ into a single $n \times 1$ vector:

$$\mathbf{X}\tilde{\beta} = \begin{bmatrix} \tilde{x}_1^\top \tilde{\beta} \\ \vdots \\ \tilde{x}_n^\top \tilde{\beta} \end{bmatrix}$$

The matrix $\mathbf{X}^\top\mathbf{X}$ is a $p \times p$ symmetric matrix that appears in the OLS estimator $\hat{\tilde{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\tilde{y}$.

Theorem 0.29 ($\mathbf{X}^\top\mathbf{X}$ is invertible when $\mathbf{X}$ has full column rank). *If $\mathbf{X}$ is an $n \times p$ matrix with* $\operatorname{rank}(\mathbf{X}) = p$ *(Definition 0.19), then the $p \times p$ matrix $\mathbf{X}^\top\mathbf{X}$ is invertible.*

Proof

Proof. Let $\tilde{c}$ be a vector of length p with $\mathbf{X}^\top\mathbf{X}\tilde{c} = \tilde{0}$. Then

$$\begin{aligned}
0 &= \tilde{c}^\top \mathbf{X}^\top \mathbf{X}\tilde{c} && \text{(multiply } \mathbf{X}^\top\mathbf{X}\tilde{c} = \tilde{0} \text{ on the left by } \tilde{c}^\top\text{)} \\
&= (\mathbf{X}\tilde{c})^\top (\mathbf{X}\tilde{c}) && \text{(transpose of a product)} \\
&= \|\mathbf{X}\tilde{c}\|^2 && \text{(definition of the Euclidean norm)}
\end{aligned}$$

so $\mathbf{X}\tilde{c} = \tilde{0}$. Because $\mathbf{X}\tilde{c} = c_1(\text{column } 1) + \cdots + c_p(\text{column } p)$ and the columns of $\mathbf{X}$ are linearly independent, $\tilde{c} = \tilde{0}$. So the only solution of $\mathbf{X}^\top\mathbf{X}\tilde{c} = \tilde{0}$ is $\tilde{c} = \tilde{0}$, which says that the columns of the square matrix $\mathbf{X}^\top\mathbf{X}$ are linearly independent, and a square matrix with linearly independent columns is invertible (Banerjee and Roy 2014, Corollary 5.7, p. 143). □

Example 0.34 (Inverting $\mathbf{X}^\top\mathbf{X}$)**.** For the rank-2 matrix $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ from Example 0.6:

$$\mathbf{X}^\top\mathbf{X} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \qquad (\mathbf{X}^\top\mathbf{X})^{-1} = \frac{1}{6}\begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix},$$

and multiplying the two out gives $\mathbf{I}_2$.

Definition 0.44 (Hat matrix)**.** For an $n \times p$ design matrix $\mathbf{X}$ (Definition 0.43) with $\operatorname{rank}(\mathbf{X}) = p$, the **hat matrix** is the $n \times n$ matrix

$$\underbrace{\mathbf{H}}_{n\times n} \overset{\text{def}}{=} \underbrace{\mathbf{X}}_{n\times p} \underbrace{(\mathbf{X}^\top\mathbf{X})^{-1}}_{p\times p} \underbrace{\mathbf{X}^\top}_{p\times n}$$

The inverse exists by Theorem 0.29.

Example 0.35 (The hat matrix of an intercept-only model)**.** With $n = 2$ observations and only an intercept, $\mathbf{X} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (2×1, rank 1), so $\mathbf{X}^\top\mathbf{X} = 2$ and

$$\mathbf{H} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \frac{1}{2} \cdot \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}.$$

Then $\mathbf{H}\tilde{y} = (\bar{y}, \bar{y})$, where $\bar{y} = (y_1 + y_2)/2$: the fitted values of an intercept-only model are the sample mean.

Theorem 0.30 (Hat matrix is a projection matrix)**.** *If $\mathbf{X}$ is an $n \times p$ design matrix with $\operatorname{rank}(\mathbf{X}) = p$, then the hat matrix $\mathbf{H}$ (Definition 0.44) is an orthogonal projection matrix (Definition 0.30).*

ℹ Proof

Proof. We verify symmetry and idempotency. Both use that $\mathbf{X}^\top\mathbf{X}$ is symmetric, $(\mathbf{X}^\top\mathbf{X})^\top = \mathbf{X}^\top(\mathbf{X}^\top)^\top = \mathbf{X}^\top\mathbf{X}$ (Theorem 0.10), so its inverse is symmetric too (Corollary 0.1).

Symmetry:

$$\begin{aligned}
\mathbf{H}^\top &= \left(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\right)^\top && \text{(definition of } \mathbf{H}\text{)}\\
&= \left(\mathbf{X}^\top\right)^\top \cdot \left((\mathbf{X}^\top\mathbf{X})^{-1}\right)^\top \cdot \mathbf{X}^\top && \text{(transpose of a product, twice)}\\
&= \mathbf{X} \cdot \left((\mathbf{X}^\top\mathbf{X})^{-1}\right)^\top \cdot \mathbf{X}^\top && \text{(transposing twice changes nothing)}\\
&= \mathbf{X} \cdot (\mathbf{X}^\top\mathbf{X})^{-1} \cdot \mathbf{X}^\top && \text{(the inverse of a symmetric matrix is symmetric)}\\
&= \mathbf{H} && \text{(definition of } \mathbf{H}\text{)}
\end{aligned}$$

Idempotency:

$$\begin{aligned}
\mathbf{H}^2 &= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \cdot \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top && \text{(definition of } \mathbf{H}\text{)}\\
&= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{X}^\top\mathbf{X})(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top && \text{(regroup; matrix multiplication is associative)}\\
&= \mathbf{X}\,\mathbf{I}_p\,(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top && \text{(definition of the inverse)}\\
&= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top && \text{(identity matrix)}\\
&= \mathbf{H} && \text{(definition of } \mathbf{H}\text{)}
\end{aligned}$$

□

Example 0.36 (The intercept-only hat matrix is a projection)**.** For $\mathbf{H} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$ from Example 0.35, $\mathbf{H}^\top = \mathbf{H}$, and

$$\mathbf{H}^2 = \begin{bmatrix} 0.5\cdot 0.5 + 0.5\cdot 0.5 & 0.5\cdot 0.5 + 0.5\cdot 0.5 \\ 0.5\cdot 0.5 + 0.5\cdot 0.5 & 0.5\cdot 0.5 + 0.5\cdot 0.5 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} = \mathbf{H},$$

as Theorem 0.30 says.

The hat matrix appears in the formula for fitted values in linear regression: $\hat{\tilde{y}} = \mathbf{X}\hat{\tilde{\beta}} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\tilde{y} = \mathbf{H}\tilde{y}$. It "puts a hat" on $\tilde{y}$ — hence the name.

Additional resources

- Fieller (2016)
- Banerjee and Roy (2014)
- Searle and Khuri (2017)

References

Banerjee, Sudipto, and Anindya Roy. 2014. *Linear Algebra and Matrix Analysis for Statistics*. Vol. 181. Crc Press Boca Raton. https://www.routledge.com/Linear-Algebra-and-Matrix-Analysis-for-Statistics/Banerjee-Roy/p/book/9781420095388.

Dobson, Annette J, and Adrian G Barnett. 2018. *An Introduction to Generalized Linear Models*. 4th ed. CRC press. https://doi.org/10.1201/9781315182780.

Fieller, Nick. 2016. *Basics of Matrix Algebra for Statistics with R*. Chapman; Hall/CRC. https://doi.org/10.1201/9781315370200.

Hutchinson, Brian. n.d. *DATA 471/571 (Machine Learning) and CSCI 481/581 (Deep Learning) Video Lectures*. Western Washington University. Accessed September 28, 2026. https://facultyweb.cs.wwu.edu/~hutchib2/video_lectures/data371/.

Kaplan, Daniel. 2022. *MOSAIC Calculus*. Www.mosaic-web.org. www.mosaic-web.org.

Searle, Shayle R, and Andre I Khuri. 2017. *Matrix Algebra Useful for Statistics*. John Wiley & Sons.