

# Special Matrices and Decompositions

Last modified: 2026-10-27 20:53:54 (PDT)

## Special Matrices

Some special matrices appeared earlier:

- the zero matrix ([Definition 4 in Matrices](#)), in the section on matrix addition
- square matrices ([Definition 8 in Matrices](#)), in the section on square and identity matrices
- the identity matrix ([Definition 12 in Matrices](#)), in that same section

**Definition 0.1** (Order of a square matrix). The **order** of a square matrix ([Definition 8 in Matrices](#)) is its number of rows, which equals its number of columns.

**Example 0.1** (Orders of square matrices).  $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$  has order 2, and  $[5]$  has order 1. A  $2 \times 3$  matrix has no order, since it is not square ([Example 8 in Matrices](#)).

**Definition 0.2** (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,  $\mathbf{A}^2 = \mathbf{A}\mathbf{A}$ .

**Example 0.2** (Powers of a  $2 \times 2$  matrix). Let  $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ . Then

$$\begin{aligned}
\mathbf{A}^2 &= \begin{bmatrix} 1 \cdot 1 + 1 \cdot 0 & 1 \cdot 1 + 1 \cdot 1 \\ 0 \cdot 1 + 1 \cdot 0 & 0 \cdot 1 + 1 \cdot 1 \end{bmatrix} && \text{(definition of matrix multiplication)} \\
&= \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, && \text{(multiply and add)} \\
\mathbf{A}^3 = \mathbf{A}^2 \mathbf{A} &= \begin{bmatrix} 1 \cdot 1 + 2 \cdot 0 & 1 \cdot 1 + 2 \cdot 1 \\ 0 \cdot 1 + 1 \cdot 0 & 0 \cdot 1 + 1 \cdot 1 \end{bmatrix} && \text{(definition of matrix multiplication)} \\
&= \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}. && \text{(multiply and add)}
\end{aligned}$$

A  $2 \times 3$  matrix  $\mathbf{B}$  is not square ([Definition 8 in Matrices](#)), so  $\mathbf{B}^2$  is not defined; indeed  $\mathbf{B}\mathbf{B}$  would multiply a matrix with 3 columns by one with 2 rows, which [Definition 7 in Matrices](#) does not allow.

**Definition 0.3** (Symmetric matrix). A square matrix  $\mathbf{A}$  is **symmetric** [1] if  $\mathbf{A}^\top = \mathbf{A}$ , i.e.,  $a_{ij} = a_{ji}$  for all  $i$  and  $j$ .

**i** Remark (Symmetric matrices in statistics)

*Remark 0.1* (Symmetric matrices in statistics). Covariance matrices are symmetric: entry  $(i, j)$  of the covariance matrix of a random vector  $\mathbf{Y}$  is  $\text{Cov}(Y_i, Y_j)$ , and  $\text{Cov}(Y_i, Y_j) = \text{Cov}(Y_j, Y_i)$ . For example, if  $\text{Var}(Y_1) = 4$ ,  $\text{Var}(Y_2) = 9$ , and  $\text{Cov}(Y_1, Y_2) = 1$ , the covariance matrix of  $(Y_1, Y_2)$  is

$$\begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}.$$

**Example 0.3** (A matrix that is not symmetric).  $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$  is not symmetric:

$$\begin{aligned}
a_{12} &= 2 \\
&\neq 3 \\
&= a_{21}.
\end{aligned}$$

**Definition 0.4** (Diagonal matrix). A square matrix  $\mathbf{D}$  is a **diagonal matrix** [2] if all off-diagonal entries are zero:  $d_{ij} = 0$  whenever  $i \neq j$ :

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

**i** Remark (diag notation)

*Remark 0.2* (diag notation). For numbers  $d_1, \dots, d_p$ ,  $\text{diag}(d_1, \dots, d_p)$  is the  $p \times p$  diagonal matrix with diagonal entries  $d_1, \dots, d_p$ . For example,

$$\text{diag}(2, 5) = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix},$$

and the  $p \times p$  identity matrix is  $\mathbf{I}_p = \text{diag}(1, \dots, 1)$ .

**Example 0.4** (A matrix that is not diagonal).  $\mathbf{D} = \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix}$  is not diagonal: its off-diagonal entry  $d_{12} = 1 \neq 0$ .

**Definition 0.5** (Matrix inverse). For a square  $p \times p$  matrix  $\mathbf{A}$ , the **inverse**  $\mathbf{A}^{-1}$  (if it exists) is the unique matrix satisfying:

$$\begin{aligned} \mathbf{A} \mathbf{A}^{-1} &= \mathbf{A}^{-1} \mathbf{A} \\ &= \mathbf{I}_p \end{aligned}$$

**Definition 0.6** (Invertible matrix, non-singular matrix). A  $p \times p$  matrix  $\mathbf{A}$  is **invertible** [3] (or **non-singular**) if some  $p \times p$  matrix  $\mathbf{B}$  satisfies

$$\begin{aligned} \mathbf{AB} &= \mathbf{BA} \\ &= \mathbf{I}_p, \end{aligned}$$

where  $\mathbf{I}_p$  is the identity matrix ([Definition 12 in Matrices](#)). A square matrix that is not invertible is **singular**.

**Example 0.5** (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

$$\begin{aligned}\mathbf{A}\mathbf{A}^{-1} &= \mathbf{A}^{-1}\mathbf{A} \\ &= \mathbf{I}_2.\end{aligned}$$

The matrix  $\mathbf{M} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$  is singular: for any  $2 \times 2$  matrix  $\mathbf{C}$ , the two rows of  $\mathbf{MC}$  are equal, so  $\mathbf{MC}$  can never be  $\mathbf{I}_2$ , whose two rows differ.

**i** Remark (Invertible matrices and inverses)

*Remark 0.3* (Invertible matrices and inverses). Let  $\mathbf{A}$  be a  $p \times p$  matrix, and suppose that two  $p \times p$  matrices  $\mathbf{B}$  and  $\mathbf{C}$  satisfy

$$\begin{aligned}\mathbf{AB} &= \mathbf{BA} \\ &= \mathbf{I}_p\end{aligned}$$

and

$$\begin{aligned}\mathbf{AC} &= \mathbf{CA} \\ &= \mathbf{I}_p.\end{aligned}$$

Then

$$\begin{aligned}\mathbf{B} &= \mathbf{B}\mathbf{I}_p && \text{(identity matrix)} \\ &= \mathbf{B}(\mathbf{AC}) && (\mathbf{AC} = \mathbf{I}_p) \\ &= (\mathbf{BA})\mathbf{C} && \text{(matrix multiplication is associative)} \\ &= \mathbf{I}_p\mathbf{C} && (\mathbf{BA} = \mathbf{I}_p) \\ &= \mathbf{C} && \text{(identity matrix)}\end{aligned}$$

The steps use [Theorem 5 in Matrices](#) and [Theorem 8 in Matrices](#). Since  $\mathbf{B} = \mathbf{C}$ , at most one matrix satisfies these equations, which is the uniqueness that [Definition 0.5](#) asserts. When  $\mathbf{A}$  is invertible, the matrix  $\mathbf{B}$  in [Definition 0.6](#) is therefore the inverse  $\mathbf{A}^{-1}$  of  $\mathbf{A}$ , and  $\mathbf{A}$  is invertible exactly when it has an inverse. For example, for  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$  in

[Example 0.5](#),  $\begin{bmatrix} 0.5 & -0.5 \\ 0 & 1 \end{bmatrix}$  is the only  $2 \times 2$  matrix  $\mathbf{B}$  with

$$\begin{aligned}\mathbf{AB} &= \mathbf{BA} \\ &= \mathbf{I}_2,\end{aligned}$$

so it is  $\mathbf{A}^{-1}$ .

**Theorem 0.1** (Inverse of a product). *If  $\mathbf{A}$  and  $\mathbf{B}$  are invertible  $p \times p$  matrices, then  $\mathbf{AB}$  is invertible, and*

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

### **i** Proof

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

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

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 0.5 for  $\mathbf{AB}$ . The steps use [Theorem 5 in Matrices](#) and [Theorem 8 in Matrices](#). □

**Theorem 0.2** (Inverse of a transpose). *If  $\mathbf{A}$  is an invertible  $p \times p$  matrix, then  $\mathbf{A}^\top$  is invertible, and*

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

### **i** Proof

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

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

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

So  $(\mathbf{A}^{-1})^\top$  satisfies Definition 0.5 for  $\mathbf{A}^\top$ . The first step of each display is [Theorem 17 in Matrices](#). □

**Example 0.6** (Inverting a transpose). For  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$ , Example 0.5 gives  $\mathbf{A}^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ 0 & 1 \end{bmatrix}$ , so Theorem 0.2 says

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

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.*

#### **i** Proof

*Proof.*

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

The first step is Theorem 0.2. □

**Example 0.7** (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.

**Theorem 0.3** (An invertible factor can be cancelled). *Let  $\mathbf{A}$  be an invertible  $p \times p$  matrix (Definition 0.6), and let  $\mathbf{B}$  and  $\mathbf{C}$  be  $p \times n$  matrices. If  $\mathbf{AB} = \mathbf{AC}$ , then  $\mathbf{B} = \mathbf{C}$ .*

**i** Proof

*Proof.*

$$\begin{aligned}
 \mathbf{B} &= \mathbf{I}_p \mathbf{B} && \text{(Theorem 8 in Matrices)} \\
 &= (\mathbf{A}^{-1} \mathbf{A}) \mathbf{B} && \text{(Definition 0.5)} \\
 &= \mathbf{A}^{-1} (\mathbf{AB}) && \text{(Theorem 5 in Matrices)} \\
 &= \mathbf{A}^{-1} (\mathbf{AC}) && (\mathbf{AB} = \mathbf{AC}) \\
 &= (\mathbf{A}^{-1} \mathbf{A}) \mathbf{C} && \text{(Theorem 5 in Matrices)} \\
 &= \mathbf{I}_p \mathbf{C} && \text{(Definition 0.5)} \\
 &= \mathbf{C}. && \text{(Theorem 8 in Matrices)}
 \end{aligned}$$

□

□

**Corollary 0.2** (An invertible factor can be cancelled on the right). *Let  $\mathbf{A}$  be an invertible  $p \times p$  matrix, and let  $\mathbf{B}$  and  $\mathbf{C}$  be  $n \times p$  matrices. If  $\mathbf{BA} = \mathbf{CA}$ , then  $\mathbf{B} = \mathbf{C}$ . The proof of Theorem 0.3 works with every product reversed: multiply both sides on the right by  $\mathbf{A}^{-1}$ .*

**Example 0.8** (Solving for a matrix by cancelling an invertible factor). Let  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$ , which is invertible with  $\mathbf{A}^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ 0 & 1 \end{bmatrix}$  (Example 0.5). If  $\mathbf{AB} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$  for a  $2 \times 1$  matrix  $\mathbf{B}$ , then multiplying both sides on the left by  $\mathbf{A}^{-1}$ , as in the proof of Theorem 0.3,

$$\begin{aligned}
 \mathbf{B} &= \mathbf{A}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix} \\
 &= \begin{bmatrix} 1 \\ 1 \end{bmatrix}.
 \end{aligned}$$

Check:  $\mathbf{A} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2+1 \\ 0+1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$

**Example 0.9** (A product of two nonzero matrices can be the zero matrix). With

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

neither matrix is the zero matrix ([Definition 4 in Matrices](#)), but

$$\begin{aligned}\mathbf{AB} &= \begin{bmatrix} 1 \cdot 0 + 0 \cdot 0 & 1 \cdot 0 + 0 \cdot 1 \\ 0 \cdot 0 + 0 \cdot 0 & 0 \cdot 0 + 0 \cdot 1 \end{bmatrix} \\ &= \mathbf{0}_{2 \times 2}.\end{aligned}$$

For numbers,  $ab = 0$  forces  $a = 0$  or  $b = 0$ . For matrices it does not.

Because  $\mathbf{A}\mathbf{0}_{2 \times 2} = \mathbf{0}_{2 \times 2} = \mathbf{AB}$  while  $\mathbf{B} \neq \mathbf{0}_{2 \times 2}$ , the factor  $\mathbf{A}$  cannot be cancelled from  $\mathbf{AB} = \mathbf{A}\mathbf{0}_{2 \times 2}$ . This does not contradict [Theorem 0.3](#), because  $\mathbf{A}$  is singular: its second row is zero, so  $\mathbf{AC}$  has a zero second row for every  $2 \times 2$  matrix  $\mathbf{C}$ , and  $\mathbf{AC}$  can never be  $\mathbf{I}_2$ .

**i** Remark (Matrix arithmetic compared with scalar arithmetic)

*Remark 0.4* (Matrix arithmetic compared with scalar arithmetic). **Sums and differences.** Matrix addition has every property that addition of numbers has: it is commutative ([Theorem 1 in Matrices](#)), associative ([Theorem 2 in Matrices](#)), has the zero matrix as its identity ([Theorem 3 in Matrices](#)), and gives each matrix a negative ([Theorem 4 in Matrices](#)). The only new condition is that  $\mathbf{A} + \mathbf{B}$  is defined only when  $\mathbf{A}$  and  $\mathbf{B}$  have the same dimensions.

**Products.** Matrix multiplication keeps three properties of multiplication of numbers: it is associative ([Theorem 5 in Matrices](#)), it is distributive over addition ([Theorem 6 in Matrices](#)), and it has an identity,  $\mathbf{I}_p$  ([Theorem 8 in Matrices](#)). It differs from multiplication of numbers in four ways:

- $\mathbf{AB}$  is defined only when the number of columns of  $\mathbf{A}$  equals the number of rows of  $\mathbf{B}$  ([Definition 7 in Matrices](#)).
- It is not commutative:  $\mathbf{AB} \neq \mathbf{BA}$  in general ([Remark 2 in Matrices](#)).
- A nonzero square matrix need not have an inverse ([Example 0.5](#)).
- $\mathbf{AB} = \mathbf{0}$  can hold with  $\mathbf{A} \neq \mathbf{0}$  and  $\mathbf{B} \neq \mathbf{0}$  ([Example 0.9](#)), so a factor cannot be cancelled from  $\mathbf{AB} = \mathbf{AC}$  unless it is invertible ([Theorem 0.3](#)).

**Definition 0.7** (Idempotent matrix). A square matrix  $\mathbf{A}$  is **idempotent** [4] if

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

**Example 0.10** (An idempotent matrix). For  $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ ,

$$\mathbf{P}^2 = \begin{bmatrix} 1 \cdot 1 + 0 \cdot 0 & 1 \cdot 0 + 0 \cdot 0 \\ 0 \cdot 1 + 0 \cdot 0 & 0 \cdot 0 + 0 \cdot 0 \end{bmatrix} \quad (\text{definition of matrix multiplication})$$

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

$$= \mathbf{P}, \quad (\text{multiply and add})$$

so  $\mathbf{P}$  is idempotent.

**Example 0.11** (A matrix that is not idempotent). For  $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ ,

$$\mathbf{A}^2 = \begin{bmatrix} 2 \cdot 2 + 0 \cdot 0 & 2 \cdot 0 + 0 \cdot 0 \\ 0 \cdot 2 + 0 \cdot 0 & 0 \cdot 0 + 0 \cdot 0 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\neq \mathbf{A},$$

so  $\mathbf{A}$  is not idempotent.

**Definition 0.8** (Orthogonal projection matrix). A square matrix  $\mathbf{P}$  is an **orthogonal projection matrix** (also called an **orthogonal projector**) if it is both symmetric (Definition 0.3) and idempotent (Definition 0.7):

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

**Definition 0.9** (Oblique projection). A square matrix is an **oblique projection** if it is idempotent (Definition 0.7) but not symmetric (Definition 0.3).

**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.

**i** Remark (What “projection matrix” means in these notes)

*Remark 0.5* (What “projection matrix” means in these notes). Some texts call any idempotent matrix a projection matrix, so that both  $\mathbf{P}$  and  $\mathbf{Q}$  in Example 0.12 would count as one. 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 0.8, such as  $\mathbf{P}$  in Example 0.12, and never an oblique projection (Definition 0.9) such as  $\mathbf{Q}$ .

**Theorem 0.4** (Complement of a projection matrix). *If  $\mathbf{P}$  is a  $p \times p$  orthogonal projection matrix (Definition 0.8), then  $\mathbf{I}_p - \mathbf{P}$  is also an orthogonal projection matrix.*

**i** Proof

*Proof.* We verify symmetry and idempotency.

**Symmetry:**

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

**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 0.5** (Projection matrices produce orthogonal decompositions). *If  $\mathbf{P}$  is a  $p \times p$  orthogonal projection matrix (Definition 0.8) and  $\tilde{v}$  is any vector of length  $p$ , then the two components of the decomposition*

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

*are orthogonal:*

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

**i** 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 ( $\mathbf{P}^2 = \mathbf{P}$ ).  $\square$

**Definition 0.10** (Orthogonal matrix). A  $p \times p$  matrix  $\mathbf{Q}$  is **orthogonal** [5] if

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

**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  $\theta$  with  $\cos \theta = 0.6$  and  $\sin \theta = 0.8$  (about 53 degrees) [6, Ch. 8], Example 8.1, p. 207. It is orthogonal:

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

**Example 0.14** (Orthogonal columns are not enough).  $\mathbf{Q} = 2\mathbf{I}_2$  has orthogonal columns  $(2, 0)$  and  $(0, 2)$ , but

$$\begin{aligned}
\mathbf{Q}^\top \mathbf{Q} &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \\
&= \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \\
&\neq \mathbf{I}_2,
\end{aligned}$$

so  $\mathbf{Q}$  is not an orthogonal matrix: its columns have norm 2, not 1.

**i** Remark (The columns of an orthogonal matrix are orthonormal)

*Remark 0.6* (The columns of an orthogonal matrix are orthonormal). 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 23 in Vectors](#)). 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 0.5](#)) [6, Ch. 8], Theorem 8.1 and Definition 8.1, p. 209.

For example, the columns of  $\mathbf{Q}$  in [Example 0.13](#) are  $(0.6, 0.8)$  and  $(-0.8, 0.6)$ . The diagonal entries  $0.36 + 0.64 = 1$  of  $\mathbf{Q}^\top \mathbf{Q}$  are their squared norms, and the off-diagonal entries  $-0.48 + 0.48 = 0$  are their dot product. Multiplying in the other order also gives  $\mathbf{I}_2$ :

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

So  $\mathbf{Q}^{-1} = \mathbf{Q}^\top$ .

**Theorem 0.6** (Orthogonal matrices preserve length). *If  $\mathbf{Q}$  is a  $p \times p$  orthogonal matrix ([Definition 0.10](#)) and  $\tilde{x}$  is a vector of length  $p$ , then*

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

## i 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 6 in Matrices](#), the third is [Theorem 17 in Matrices](#), and the fourth is [Theorem 5 in Matrices](#).  $\square$

**Example 0.15** (Rotating a vector keeps its length). With  $\mathbf{Q}$  from [Example 0.13](#) and  $\tilde{x} = (3, 4)$  from [Example 15 in Vectors](#):

$$\begin{aligned}
\mathbf{Q}\tilde{x} &= \begin{bmatrix} 0.6 \cdot 3 - 0.8 \cdot 4 \\ 0.8 \cdot 3 + 0.6 \cdot 4 \end{bmatrix} && \text{(definition of matrix-vector multiplication)} \\
&= \begin{bmatrix} 1.8 - 3.2 \\ 2.4 + 2.4 \end{bmatrix} && \text{(multiply)} \\
&= \begin{bmatrix} -1.4 \\ 4.8 \end{bmatrix} && \text{(add)} \\
\|\mathbf{Q}\tilde{x}\| &= \sqrt{(-1.4)^2 + 4.8^2} && \text{(definition of the norm)} \\
&= \sqrt{1.96 + 23.04} && \text{(square)} \\
&= \sqrt{25} && \text{(add)} \\
&= 5 && \text{(take the square root)}
\end{aligned}$$

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

## Quadratic Forms

**Definition 0.11** (Quadratic form). A **quadratic form** [7] 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.

### **i** Remark (Quadratic forms in statistics)

*Remark 0.7* (Quadratic forms in statistics). A quadratic form extends the scalar expression  $cx^2$  to vectors: with  $p = 1$  and  $\mathbf{S} = [c]$ ,  $\tilde{x}^\top \mathbf{S} \tilde{x} = cx^2$ . With  $p = 2$ ,

$$\tilde{x}^\top \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \tilde{x} = x_1^2 + 2x_1x_2 + 3x_2^2,$$

which is  $1 + 4 + 12 = 17$  at  $\tilde{x} = (1, 2)$ .

Quadratic forms occur often 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:  $\text{Var}(\tilde{x} \cdot \hat{\beta}) = \tilde{x}^\top \text{Var}(\hat{\beta}) \tilde{x}$ .

**Theorem 0.7** (A quadratic form is additive in its matrix). *If  $\mathbf{A}$  and  $\mathbf{B}$  are  $p \times p$  matrices,  $c$  is a number and  $\tilde{x}$  is a vector of length  $p$ , then*

$$\begin{aligned} \tilde{x}^\top (\mathbf{A} + \mathbf{B}) \tilde{x} &= \tilde{x}^\top \mathbf{A} \tilde{x} + \tilde{x}^\top \mathbf{B} \tilde{x}, \\ \tilde{x}^\top (c \mathbf{A}) \tilde{x} &= c \tilde{x}^\top \mathbf{A} \tilde{x}. \end{aligned}$$

*So for a fixed  $\tilde{x}$ , the map  $\mathbf{S} \mapsto \tilde{x}^\top \mathbf{S} \tilde{x}$  is linear in  $\mathbf{S}$ .*

### **i** Proof

*Proof.* By the distributive law ([Theorem 6 in Matrices](#)), applied on the left and then on the right,

$$\begin{aligned} \tilde{x}^\top (\mathbf{A} + \mathbf{B}) \tilde{x} &= (\tilde{x}^\top \mathbf{A} + \tilde{x}^\top \mathbf{B}) \tilde{x} \quad (\text{distribute } \tilde{x}^\top \text{ over the sum}) \\ &= \tilde{x}^\top \mathbf{A} \tilde{x} + \tilde{x}^\top \mathbf{B} \tilde{x} \quad (\text{distribute } \tilde{x} \text{ over the sum}) \end{aligned}$$

For the scalar, a number can be moved out of any product, so  $\tilde{x}^\top(c\mathbf{A})\tilde{x} = c\tilde{x}^\top\mathbf{A}\tilde{x}$ .  $\square$

**Example 0.16** (Combining two quadratic forms into one). Take  $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$ ,  $\mathbf{B} = \mathbf{I}$  and  $\tilde{x} = (1, 2)$ . Then

$$\begin{aligned}\tilde{x}^\top\mathbf{A}\tilde{x} &= 1 + 4 + 12 = 17, \\ \tilde{x}^\top\mathbf{I}\tilde{x} &= 1 + 4 = 5, \\ \tilde{x}^\top(\mathbf{A} + \mathbf{I})\tilde{x} &= \tilde{x}^\top \begin{bmatrix} 2 & 2 \\ 0 & 4 \end{bmatrix} \tilde{x} \\ &= 2 + 4 + 16 \\ &= 22,\end{aligned}$$

which equals  $17 + 5$ . In particular,  $\tilde{x}^\top(\mathbf{A} + c\mathbf{I})\tilde{x} = \tilde{x}^\top\mathbf{A}\tilde{x} + c\|\tilde{x}\|^2$  for every number  $c$ , because  $\tilde{x}^\top\mathbf{I}\tilde{x} = \tilde{x}^\top\tilde{x} = \|\tilde{x}\|^2$ . This is how two quadratic terms in the same vector combine into a single quadratic form.

**Definition 0.12** (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 0.17** (The symmetric part of a  $2 \times 2$  matrix). For  $\mathbf{S} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$ , the symmetric part is

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

which is symmetric.

**Theorem 0.8** (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 0.12) of*

**S.**

**i** Proof

*Proof.* The quadratic form  $\tilde{x}^\top \mathbf{S} \tilde{x}$  is a  $1 \times 1$  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  $\tilde{x}^\top \mathbf{S} \tilde{x}$ :

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

The distributive step is [Theorem 6 in Matrices](#), and the transpose step is [Theorem 17 in Matrices](#).  $\square$

**Example 0.18** (Replacing a matrix by its symmetric part). With  $\mathbf{S} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$  from Example 0.17 and  $\tilde{x} = (1, 1)$ :

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

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 0.13** (Trace). The **trace** [8] of a  $p \times p$  matrix  $\mathbf{M}$  is the sum of its diagonal entries:

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

**Example 0.19** (The trace of a  $2 \times 2$  matrix).

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

**Theorem 0.9** (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  $\mathbf{B}$ :*

$$\text{tr}\left(\underbrace{\mathbf{AB}}_{m \times m}\right) = \text{tr}\left(\underbrace{\mathbf{BA}}_{n \times n}\right)$$

### **i** Proof

*Proof.*

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

□

**Example 0.20** (Traces of the two products of a  $2 \times 3$  and a  $3 \times 2$  matrix). Let

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

The products are

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

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

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

For  $\mathbf{BA}$ , which is  $3 \times 3$ :

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

The products have different sizes, but their traces agree.

### **i** Remark (Moving a factor of a triple product)

*Remark 0.8* (Moving a factor of a triple product). Let  $\mathbf{A}$  be  $m \times n$ ,  $\mathbf{B}$  be  $n \times k$ , and  $\mathbf{C}$  be  $k \times m$ , so that  $\mathbf{AB}$  is  $m \times k$  and  $\mathbf{ABC}$  is square. Applying Theorem 0.9 to the  $m \times k$  matrix  $\mathbf{AB}$  and the  $k \times m$  matrix  $\mathbf{C}$  moves the last factor of the triple product to the front [6, Ch. 1], Theorem 1.5 and eq. 1.17, p. 19:

$$\begin{aligned} \text{tr}(\mathbf{ABC}) &= \text{tr}((\mathbf{AB})\mathbf{C}) && \text{(matrix multiplication is associative)} \\ &= \text{tr}(\mathbf{C}(\mathbf{AB})) && \text{(the theorem, applied to the two factors } \mathbf{AB} \text{ and } \mathbf{C}) \\ &= \text{tr}(\mathbf{CAB}) && \text{(matrix multiplication is associative)} \end{aligned}$$

For example, let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Then

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

so

$$(\mathbf{AB})\mathbf{C} = \begin{bmatrix} 3 \cdot 0 + 2 \cdot 1 & 3 \cdot 1 + 2 \cdot 0 \\ 1 \cdot 0 + 1 \cdot 1 & 1 \cdot 1 + 1 \cdot 0 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} \quad (\mathbf{CA})\mathbf{B} = \begin{bmatrix} 0 \cdot 1 + 1 \cdot 1 & 0 \cdot 0 + 1 \cdot 1 \\ 1 \cdot 1 + 2 \cdot 1 & 1 \cdot 0 + 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix}$$

and both traces are the same:

$$\begin{aligned} \text{tr}(\mathbf{ABC}) &= 2 + 1 \\ &= 3 \end{aligned}$$

and

$$\begin{aligned} \text{tr}(\mathbf{CAB}) &= 1 + 2 \\ &= 3. \end{aligned}$$

**Definition 0.14** (Matrix inner product). The **inner product** of two  $n \times p$  matrices  $\mathbf{A}$  and  $\mathbf{B}$  is

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

Banerjee and Roy [6, Ch. 15], eq. 15.7, p. 491 defines the same inner product on  $n \times p$  matrices with the trace.

**Theorem 0.10** (The matrix inner product multiplies matching entries). *For two  $n \times p$  matrices  $\mathbf{A}$  and  $\mathbf{B}$ :*

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

**i** Proof

*Proof.*

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

□

**Example 0.21** (The inner product of two  $2 \times 2$  matrices). Let

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

From Definition 0.14:

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

From Theorem 0.10:

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

**i** Remark (The matrix inner product as a dot product)

*Remark 0.9* (The matrix inner product as a dot product). Theorem 0.10 says that the matrix inner product is the dot product (Definition 8 in Vectors) of the two matrices' entries, each listed as one vector of length  $np$ , with both matrices' entries listed in the same order.

For example, listing the entries of  $\mathbf{A}$  and  $\mathbf{B}$  in Example 0.21 row by row gives  $(1, 2, 3, 4)$  and  $(0, 1, -1, 2)$ , and

$$\begin{aligned}(1, 2, 3, 4) \cdot (0, 1, -1, 2) &= 7 \\ &= \langle \mathbf{A}, \mathbf{B} \rangle.\end{aligned}$$

**Definition 0.15** (Frobenius norm). The **Frobenius norm** [9] 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 0.14).

**Example 0.22** (The Frobenius norm of a  $2 \times 2$  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 0.10.

**i** Remark (The Frobenius norm is the Euclidean norm of the entries)

*Remark 0.10* (The Frobenius norm is the Euclidean norm of the entries). By Theorem 0.10,  $\|\mathbf{A}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p a_{ij}^2$ , a sum of squares, so the square root is always defined.  $\|\mathbf{A}\|_F$  is the Euclidean norm (Definition 20 in Vectors) of the entries of  $\mathbf{A}$ , listed as one vector of length  $np$  [6, Ch. 15], Definition 15.4, p. 492.

For example, listing the entries of  $\mathbf{A}$  in Example 0.22 row by row gives the vector  $(1, 2, 3, 4)$  of length  $2 \cdot 2 = 4$ , and

$$\begin{aligned}
\|(1, 2, 3, 4)\| &= \sqrt{1 + 4 + 9 + 16} \\
&= \sqrt{30} \\
&= \|\mathbf{A}\|_F.
\end{aligned}$$

**Theorem 0.11** (The squared Frobenius norm adds up the squared column norms). *If  $\mathbf{A}$  is an  $n \times p$  matrix with columns  $\underline{a}_1, \dots, \underline{a}_p$ , then*

$$\|\mathbf{A}\|_F^2 = \sum_{j=1}^p \|\underline{a}_j\|^2.$$

### **i** Proof

*Proof.* Entry  $i$  of  $\underline{a}_j$  is  $a_{ij}$ , so

$$\begin{aligned}
\|\mathbf{A}\|_F^2 &= \sum_{i=1}^n \sum_{j=1}^p a_{ij}^2 \quad (\text{Remark 0.10}) \\
&= \sum_{j=1}^p \sum_{i=1}^n a_{ij}^2 \quad (\text{swap the order of two finite sums}) \\
&= \sum_{j=1}^p \|\underline{a}_j\|^2 \quad (\text{Definition 20 in Vectors, applied to each column})
\end{aligned}$$

□

**Example 0.23** (The Frobenius norm of a  $2 \times 2$  matrix, column by column). The matrix

$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$  in Example 0.22 has columns  $\underline{a}_1 = (1, 3)$  and  $\underline{a}_2 = (2, 4)$ , so

$$\begin{aligned}
\|\mathbf{A}\|_F^2 &= \|\underline{a}_1\|^2 + \|\underline{a}_2\|^2 && (\text{Theorem 0.11}) \\
&= (1^2 + 3^2) + (2^2 + 4^2) && (\text{definition of the norm}) \\
&= 10 + 20 && (\text{square and add}) \\
&= 30 && (\text{add})
\end{aligned}$$

which matches  $\|\mathbf{A}\|_F = \sqrt{30}$  from Example 0.22.

## Matrix Decompositions

**Definition 0.16** (Eigenvalue). Let  $\mathbf{A}$  be a  $p \times p$  matrix. A real number  $\lambda$  is an **eigenvalue** of  $\mathbf{A}$  if some real vector  $\underset{\sim}{v} \neq \underset{\sim}{0}$  of length  $p$  satisfies

$$\underset{p \times p}{\mathbf{A}} \underset{p \times 1}{\underset{\sim}{v}} = \lambda \underset{p \times 1}{\underset{\sim}{v}}$$

**Definition 0.17** (Eigenvector). In Definition 0.16, any such  $\underset{\sim}{v}$  is an **eigenvector** of  $\mathbf{A}$  for  $\lambda$ .

**Example 0.24** (Eigenvectors of a  $2 \times 2$  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:

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

### Remark (Multiples of an eigenvector)

*Remark 0.11* (Multiples of an eigenvector). Multiplying by  $\mathbf{A}$  rescales an eigenvector by  $\lambda$  and does not change the line it lies on. Any nonzero multiple  $c \underset{\sim}{v}$  of an eigenvector is also an eigenvector for the same  $\lambda$ :

$$\begin{aligned} \mathbf{A}(c \underset{\sim}{v}) &= c \mathbf{A} \underset{\sim}{v} && \text{(move the scalar } c \text{ to the front)} \\ &= c \lambda \underset{\sim}{v} && (\underset{\sim}{v} \text{ is an eigenvector for } \lambda) \\ &= \lambda (c \underset{\sim}{v}) && \text{(multiplication of numbers is commutative)} \end{aligned}$$

For example, in Example 0.24,  $2(1, 1) = (2, 2)$  is also an eigenvector for the eigenvalue 3:

$$\begin{aligned}
\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} &= \begin{bmatrix} 2 \cdot 2 + 1 \cdot 2 \\ 1 \cdot 2 + 2 \cdot 2 \end{bmatrix} && \text{(definition of matrix-vector multiplication)} \\
&= \begin{bmatrix} 4 + 2 \\ 2 + 4 \end{bmatrix} && \text{(multiply)} \\
&= \begin{bmatrix} 6 \\ 6 \end{bmatrix} && \text{(add)} \\
&= 3 \begin{bmatrix} 2 \\ 2 \end{bmatrix} && \text{(factor out 3)}
\end{aligned}$$

**Example 0.25** (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  $\lambda$  gives  $(2, 1) = \lambda(1, 0)$ , because the second entries would need  $1 = \lambda \cdot 0$ .

### Remark (Real and complex eigenvalues)

*Remark 0.12* (Real and complex eigenvalues). These notes take  $\lambda$  and  $\tilde{v}$  to be real. Banerjee and Roy [6, Ch. 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 [6, Ch. 11], Example 11.2, p. 312.

For example,  $\mathbf{R} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$  rotates each vector in the plane counterclockwise by 90 degrees:  $\mathbf{R}(v_1, v_2) = (-v_2, v_1)$ . Suppose  $\mathbf{R}\tilde{v} = \lambda\tilde{v}$  for a real number  $\lambda$ . Matching entries gives  $-v_2 = \lambda v_1$  and  $v_1 = \lambda v_2$ . Substituting the second equation into the first gives  $-v_2 = \lambda^2 v_2$ , so  $(1 + \lambda^2)v_2 = 0$ . Since  $1 + \lambda^2 > 0$ ,  $v_2 = 0$ , and then

$$\begin{aligned}
v_1 &= \lambda v_2 \\
&= 0.
\end{aligned}$$

So the only solution is  $\tilde{v} = \mathbf{0}$ , and  $\mathbf{R}$  has no real eigenvalue.

### Video

Chapter 14 of the [Essence of linear algebra](#) series by 3Blue1Brown shows the geometric meaning of eigenvectors and eigenvalues:

- [Eigenvectors and eigenvalues](#) shows how certain vectors maintain their direction

during a linear transformation and are merely stretched or squished by an eigenvalue factor (Definition 0.16).

**Theorem 0.12** (Spectral theorem for symmetric matrices). *If  $\mathbf{A}$  is a  $p \times p$  symmetric matrix (Definition 0.3) with real entries, then there are a  $p \times p$  orthogonal matrix  $\mathbf{Q}$  (Definition 0.10) and a  $p \times p$  diagonal matrix  $\mathbf{\Lambda}$  (Definition 0.4) with real diagonal entries  $\lambda_1, \dots, \lambda_p$  such that*

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

*Each  $\lambda_i$  is an eigenvalue of  $\mathbf{A}$  (Definition 0.16), and column  $i$  of  $\mathbf{Q}$  is an eigenvector of  $\mathbf{A}$  for  $\lambda_i$ .*

**i** Remark (An equivalent form of the spectral theorem)

*Remark 0.13* (An equivalent form of the spectral theorem). The proof, by [induction](#) on  $p$ , is outside the scope of these notes; see Banerjee and Roy [6, Ch. 11], Theorem 11.27, p. 349, which states the result in the equivalent form  $\mathbf{Q}^\top \mathbf{A} \mathbf{Q} = \mathbf{\Lambda}$ . The two forms are equivalent because

$$\begin{aligned} \mathbf{Q}^\top \mathbf{Q} &= \mathbf{Q} \mathbf{Q}^\top \\ &= \mathbf{I}_p \end{aligned}$$

(Remark 0.6). Multiplying  $\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top$  by  $\mathbf{Q}^\top$  on the left and by  $\mathbf{Q}$  on the right gives

$$\begin{aligned} \mathbf{Q}^\top \mathbf{A} \mathbf{Q} &= \mathbf{Q}^\top (\mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top) \mathbf{Q} && \text{(substitute)} \\ &= (\mathbf{Q}^\top \mathbf{Q}) \mathbf{\Lambda} (\mathbf{Q}^\top \mathbf{Q}) && \text{(regroup; matrix multiplication is associative)} \\ &= \mathbf{I}_p \mathbf{\Lambda} \mathbf{I}_p && \text{(\mathbf{Q} is orthogonal)} \\ &= \mathbf{\Lambda} && \text{(multiplying by } \mathbf{I}_p \text{ changes nothing)} \end{aligned}$$

Multiplying  $\mathbf{Q}^\top \mathbf{A} \mathbf{Q} = \mathbf{\Lambda}$  by  $\mathbf{Q}$  on the left and by  $\mathbf{Q}^\top$  on the right gives back the first form:

$$\begin{aligned} \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top &= \mathbf{Q} (\mathbf{Q}^\top \mathbf{A} \mathbf{Q}) \mathbf{Q}^\top && \text{(substitute)} \\ &= (\mathbf{Q} \mathbf{Q}^\top) \mathbf{A} (\mathbf{Q} \mathbf{Q}^\top) && \text{(regroup; matrix multiplication is associative)} \\ &= \mathbf{I}_p \mathbf{A} \mathbf{I}_p && \text{(\mathbf{Q} \mathbf{Q}^\top = \mathbf{I}_p)} \\ &= \mathbf{A} && \text{(multiplying by } \mathbf{I}_p \text{ changes nothing)} \end{aligned}$$

For example, let  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$  and  $\mathbf{Q} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ , which is symmetric, so  $\mathbf{Q}^\top = \mathbf{Q}$ . Then

$$\begin{aligned}
\mathbf{Q}^\top \mathbf{A} \mathbf{Q} &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(substitute)} \\
&= \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(move the scalars to the front)} \\
&= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(multiply the scalars)} \\
&= \frac{1}{2} \begin{bmatrix} 1 \cdot 2 + 1 \cdot 1 & 1 \cdot 1 + 1 \cdot 2 \\ 1 \cdot 2 + (-1) \cdot 1 & 1 \cdot 1 + (-1) \cdot 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(definition of matrix multiplication, first two matrices)} \\
&= \frac{1}{2} \begin{bmatrix} 2+1 & 1+2 \\ 2-1 & 1-2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(multiply)} \\
&= \frac{1}{2} \begin{bmatrix} 3 & 3 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(add)} \\
&= \frac{1}{2} \begin{bmatrix} 3 \cdot 1 + 3 \cdot 1 & 3 \cdot 1 + 3 \cdot (-1) \\ 1 \cdot 1 + (-1) \cdot 1 & 1 \cdot 1 + (-1) \cdot (-1) \end{bmatrix} && \text{(definition of matrix multiplication)} \\
&= \frac{1}{2} \begin{bmatrix} 3+3 & 3-3 \\ 1-1 & 1+1 \end{bmatrix} && \text{(multiply)} \\
&= \frac{1}{2} \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} && \text{(add)} \\
&= \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} && \text{(multiply by } \frac{1}{2} \text{)}
\end{aligned}$$

a diagonal matrix with the eigenvalues 3 and 1 of  $\mathbf{A}$  from Example 0.24 on its diagonal.

**Theorem 0.13** (Eigenvectors of a symmetric matrix for distinct eigenvalues are orthogonal). *Let  $\mathbf{A}$  be a symmetric  $p \times p$  matrix (Definition 0.3), and let  $\tilde{u}$  and  $\tilde{v}$  be eigenvectors of  $\mathbf{A}$  (Definition 0.17) for eigenvalues  $\lambda_1 \neq \lambda_2$ . Then  $\tilde{u} \perp \tilde{v}$  (Definition 19 in Vectors).*

**i** Proof

*Proof.*

$$\begin{aligned}
 \lambda_1 (\underline{u} \cdot \underline{v}) &= (\lambda_1 \underline{u}) \cdot \underline{v} && \text{(Definition 8 in Vectors)} \\
 &= (\mathbf{A}\underline{u}) \cdot \underline{v} && (\underline{u} \text{ is an eigenvector for } \lambda_1) \\
 &= (\mathbf{A}\underline{u})^\top \underline{v} && \text{(Definition 8 in Vectors)} \\
 &= \underline{u}^\top \mathbf{A}^\top \underline{v} && \text{(Theorem 17 in Matrices)} \\
 &= \underline{u}^\top \mathbf{A} \underline{v} && (\mathbf{A} \text{ is symmetric}) \\
 &= \underline{u}^\top (\lambda_2 \underline{v}) && (\underline{v} \text{ is an eigenvector for } \lambda_2) \\
 &= \lambda_2 (\underline{u} \cdot \underline{v}). && \text{(Definition 8 in Vectors)}
 \end{aligned}$$

So  $(\lambda_1 - \lambda_2) (\underline{u} \cdot \underline{v}) = 0$ , and  $\lambda_1 - \lambda_2 \neq 0$ , so  $\underline{u} \cdot \underline{v} = 0$ . □

**Example 0.26** (Eigenvectors of a symmetric  $2 \times 2$  matrix).  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$  is [symmetric](#). Its [eigenvectors](#)  $(1, 1)$  for  $\lambda_1 = 3$  and  $(1, -1)$  for  $\lambda_2 = 1$  (Example [0.24](#)) satisfy

$$\begin{aligned}
 (1, 1) \cdot (1, -1) &= 1 \cdot 1 + 1 \cdot (-1) \\
 &= 0.
 \end{aligned}$$

**Definition 0.18** (Eigendecomposition, spectral decomposition). Let  $\mathbf{A}$  be a  $p \times p$  symmetric matrix with real entries. An **eigendecomposition** [10], or **spectral decomposition**, of  $\mathbf{A}$  is a factorization

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

with  $\mathbf{Q}$  a  $p \times p$  orthogonal matrix (Definition [0.10](#)) and  $\mathbf{\Lambda}$  a  $p \times p$  diagonal matrix (Definition [0.4](#)).

**Example 0.27** (An eigendecomposition of a  $2 \times 2$  symmetric matrix). For  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ , Example [0.24](#) found the eigenvectors  $(1, 1)$  for 3 and  $(1, -1)$  for 1. They are orthogonal ([Definition 19 in Vectors](#)):

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

Each has norm  $\sqrt{1^2 + 1^2} = \sqrt{2}$  ([Definition 20 in Vectors](#)), so dividing each by  $\sqrt{2}$  gives two orthonormal eigenvectors ([Definition 23 in Vectors](#)). Put them in the columns of  $\mathbf{Q}$ ,

and the matching eigenvalues on the diagonal of  $\mathbf{\Lambda}$ :

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

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

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

which is  $\mathbf{A}$ .

**i** Remark (An eigendecomposition is not unique)

*Remark 0.14* (An eigendecomposition is not unique). A symmetric matrix can have more than one eigendecomposition. Reordering the eigenvalues on the diagonal of  $\mathbf{\Lambda}$ , and the columns of  $\mathbf{Q}$  with them, gives another one, and so does multiplying a column of  $\mathbf{Q}$  by  $-1$ .

For example, take  $\mathbf{A}$  from Example 0.27. Swapping the two eigenvalues and the two columns gives

$$\mathbf{Q}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \mathbf{\Lambda}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

and multiplying the second column of  $\mathbf{Q}$  by  $-1$  gives

$$\begin{aligned} \mathbf{Q}_2 &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \\ \mathbf{\Lambda}_2 &= \mathbf{\Lambda} \\ &= \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

The columns of  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  are the columns of  $\mathbf{Q}$ , reordered or multiplied by  $-1$ , so they are still orthonormal, and  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  are orthogonal. Both products give back  $\mathbf{A}$ :

$$\begin{aligned} \mathbf{Q}_1 \mathbf{\Lambda}_1 \mathbf{Q}_1^\top &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \left( \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \right)^\top && \text{(substitute)} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} && \text{(definition of the transpose)} \\ &= \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} && \text{(move the scalars to the front)} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} && \text{(multiply the scalars)} \\ &= \frac{1}{2} \begin{bmatrix} 1 \cdot 1 + 1 \cdot 0 & 1 \cdot 0 + 1 \cdot 3 \\ -1 \cdot 1 + 1 \cdot 0 & -1 \cdot 0 + 1 \cdot 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} && \text{(definition of matrix multiplication, first two matrices)} \\ &= \frac{1}{2} \begin{bmatrix} 1 + 0 & 0 + 3 \\ -1 + 0 & 0 + 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} && \text{(multiply)} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 3 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} && \text{(add)} \\ &= \frac{1}{2} \begin{bmatrix} 1 \cdot 1 + 3 \cdot 1 & 1 \cdot (-1) + 3 \cdot 1 \\ -1 \cdot 1 + 3 \cdot 1 & -1 \cdot (-1) + 3 \cdot 1 \end{bmatrix} && \text{(definition of matrix multiplication)} \\ &= \frac{1}{2} \begin{bmatrix} 1 + 3 & -1 + 3 \\ -1 + 3 & 1 + 3 \end{bmatrix} && \text{(multiply)} \\ &= \frac{1}{2} \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} && \text{(add)} \\ &= \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} && \text{(multiply by } \frac{1}{2} \text{)} \end{aligned}$$

$$\begin{aligned}
\mathbf{Q}_2 \mathbf{\Lambda}_2 \mathbf{Q}_2^\top &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \left( \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \right)^\top && \text{(substitute)} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} && \text{(definition of the transpose)} \\
&= \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} && \text{(move the scalars to the front)} \\
&= \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} && \text{(multiply the scalars)} \\
&= \frac{1}{2} \begin{bmatrix} 1 \cdot 3 + (-1) \cdot 0 & 1 \cdot 0 + (-1) \cdot 1 \\ 1 \cdot 3 + 1 \cdot 0 & 1 \cdot 0 + 1 \cdot 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} && \text{(definition of matrix multiplication, first two matrices)} \\
&= \frac{1}{2} \begin{bmatrix} 3 + 0 & 0 - 1 \\ 3 + 0 & 0 + 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} && \text{(multiply)} \\
&= \frac{1}{2} \begin{bmatrix} 3 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} && \text{(add)} \\
&= \frac{1}{2} \begin{bmatrix} 3 \cdot 1 + (-1) \cdot (-1) & 3 \cdot 1 + (-1) \cdot 1 \\ 3 \cdot 1 + 1 \cdot (-1) & 3 \cdot 1 + 1 \cdot 1 \end{bmatrix} && \text{(definition of matrix multiplication)} \\
&= \frac{1}{2} \begin{bmatrix} 3 + 1 & 3 - 1 \\ 3 - 1 & 3 + 1 \end{bmatrix} && \text{(multiply)} \\
&= \frac{1}{2} \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} && \text{(add)} \\
&= \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} && \text{(multiply by } \frac{1}{2} \text{)}
\end{aligned}$$

**Theorem 0.14** (Singular value decomposition). *Let  $\mathbf{A}$  be an  $n \times p$  matrix with real entries and  $\text{rank}(\mathbf{A}) = r$  ([Definition 2 in Subspaces and Rank](#)). Then there are an  $n \times n$  orthogonal matrix  $\mathbf{U}$ , a  $p \times p$  orthogonal matrix  $\mathbf{V}$  ([Definition 0.10](#)), and numbers  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  such that*

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

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

**i** Remark (The SVD applies to every real matrix)

*Remark 0.15* (The SVD applies to every real matrix). The proof is outside the scope of these notes; see Banerjee and Roy [6, Ch. 12], Theorem 12.1, p. 373. Unlike the spectral theorem (Theorem 0.12), Theorem 0.14 applies to every real matrix, including one that is not square or not symmetric.

For example, the  $1 \times 2$  matrix  $\mathbf{A} = \begin{bmatrix} 1 & 1 \end{bmatrix}$  is not square, so Theorem 0.12 does not apply to it. It has rank 1, and Theorem 0.14 holds with  $\mathbf{U} = \begin{bmatrix} 1 \end{bmatrix}$ ,  $\mathbf{D} = \begin{bmatrix} \sqrt{2} & 0 \end{bmatrix}$ , and  $\mathbf{V}$  the orthogonal matrix  $\mathbf{Q}$  from Example 0.27, which is symmetric, so  $\mathbf{V}^\top = \mathbf{V}$ :

$$\begin{aligned}
 \mathbf{UDV}^\top &= \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(substitute)} \\
 &= \begin{bmatrix} \sqrt{2} & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(multiplying by the } 1 \times 1 \text{ identity changes nothing)} \\
 &= \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} && \text{(move the scalar to the front)} \\
 &= \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} \cdot 1 + 0 \cdot 1 & \sqrt{2} \cdot 1 + 0 \cdot (-1) \end{bmatrix} && \text{(definition of matrix multiplication)} \\
 &= \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} + 0 & \sqrt{2} + 0 \end{bmatrix} && \text{(multiply)} \\
 &= \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} & \sqrt{2} \end{bmatrix} && \text{(add)} \\
 &= \begin{bmatrix} 1 & 1 \end{bmatrix} && \text{(divide by } \sqrt{2})
 \end{aligned}$$

$\mathbf{U}$  is orthogonal because

$$\begin{aligned}
 \mathbf{U}^\top \mathbf{U} &= \begin{bmatrix} 1 \end{bmatrix} \\
 &= \mathbf{I}_1,
 \end{aligned}$$

and the single singular value is  $\sigma_1 = \sqrt{2}$ .

**Definition 0.19** (Singular value decomposition). Let  $\mathbf{A}$  be an  $n \times p$  matrix with real entries and  $\text{rank}(\mathbf{A}) = r$ . A **singular value decomposition** [11] (SVD) of  $\mathbf{A}$  is a factorization  $\mathbf{A} = \mathbf{UDV}^\top$  with  $\mathbf{U}$ ,  $\mathbf{D}$  and  $\mathbf{V}$  as in Theorem 0.14.

**Definition 0.20** (Singular values). The numbers  $\sigma_1 \geq \dots \geq \sigma_r > 0$  on the diagonal of  $\mathbf{D}$  in a singular value decomposition of  $\mathbf{A}$  (Definition 0.19) are the **singular values** of  $\mathbf{A}$ .

**Example 0.28** (An SVD of a  $3 \times 2$  matrix). Let

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

$\mathbf{V}$  is the orthogonal matrix  $\mathbf{Q}$  from Example 0.27.  $\mathbf{U}$  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  $\underline{u}_1 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$ ,  $\underline{u}_2 = (0, 1, 0)$  and  $\underline{u}_3 = (\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}})$ :

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

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

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

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

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

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

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

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

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  $\sigma_2 = \sqrt{2}$ .

**i** Remark (An SVD is not unique)

*Remark 0.16* (An SVD is not unique). An SVD is not unique [6, Ch. 12], Examples 12.2 and 12.3, p. 378. 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 [6, Ch. 12], pp. 371 and 378.

For example, in Example 0.28, multiplying the first columns of  $\mathbf{U}$  and  $\mathbf{V}$  by  $-1$  gives

$$\mathbf{U}_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \quad \mathbf{V}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}$$

Their columns are the columns of  $\mathbf{U}$  and  $\mathbf{V}$ , some multiplied by  $-1$ , so they are still orthonormal. With the same  $\mathbf{D}$ :

$$\begin{aligned} \mathbf{D}\mathbf{V}_1^\top &= \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \left( \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix} \right)^\top && \text{(substitute)} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix} && \text{(definition of the transpose)} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix} && \text{(move the scalar to the front)} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 2 \cdot (-1) + 0 \cdot 1 & 2 \cdot (-1) + 0 \cdot (-1) \\ 0 \cdot (-1) + \sqrt{2} \cdot 1 & 0 \cdot (-1) + \sqrt{2} \cdot (-1) \\ 0 \cdot (-1) + 0 \cdot 1 & 0 \cdot (-1) + 0 \cdot (-1) \end{bmatrix} && \text{(definition of matrix multiplication)} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} -2 + 0 & -2 + 0 \\ 0 + \sqrt{2} & 0 - \sqrt{2} \\ 0 + 0 & 0 + 0 \end{bmatrix} && \text{(multiply)} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} -2 & -2 \\ \sqrt{2} & -\sqrt{2} \\ 0 & 0 \end{bmatrix} && \text{(add)} \\ &= \begin{bmatrix} -\sqrt{2} & -\sqrt{2} \\ 1 & -1 \\ 0 & 0 \end{bmatrix} && \text{(divide by } \sqrt{2}) \end{aligned}$$

and

$$\begin{aligned}
\mathbf{U}_1 \mathbf{D} \mathbf{V}_1^\top &= \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -\sqrt{2} & -\sqrt{2} \\ 1 & -1 \\ 0 & 0 \end{bmatrix} && \text{(substitute)} \\
&= \begin{bmatrix} -\frac{1}{\sqrt{2}} \cdot (-\sqrt{2}) + 0 \cdot 1 + \frac{1}{\sqrt{2}} \cdot 0 & -\frac{1}{\sqrt{2}} \cdot (-\sqrt{2}) + 0 \cdot (-1) + \frac{1}{\sqrt{2}} \cdot 0 \\ 0 \cdot (-\sqrt{2}) + 1 \cdot 1 + 0 \cdot 0 & 0 \cdot (-\sqrt{2}) + 1 \cdot (-1) + 0 \cdot 0 \\ -\frac{1}{\sqrt{2}} \cdot (-\sqrt{2}) + 0 \cdot 1 - \frac{1}{\sqrt{2}} \cdot 0 & -\frac{1}{\sqrt{2}} \cdot (-\sqrt{2}) + 0 \cdot (-1) - \frac{1}{\sqrt{2}} \cdot 0 \end{bmatrix} && \text{(definition of matrix multiplication)} \\
&= \begin{bmatrix} 1 + 0 + 0 & 1 + 0 + 0 \\ 0 + 1 + 0 & 0 - 1 + 0 \\ 1 + 0 - 0 & 1 + 0 - 0 \end{bmatrix} && \text{(multiply)} \\
&= \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} && \text{(add)}
\end{aligned}$$

which is  $\mathbf{A}$  again. Both SVDs have the singular values  $\sigma_1 = 2$  and  $\sigma_2 = \sqrt{2}$ .

**i** Remark (Counting zero singular values)

*Remark 0.17* (Counting zero singular values). 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.

For example,  $\mathbf{B} = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$  has rank  $r = 1$ . Taking

$$\begin{aligned}
\mathbf{U} &= \mathbf{V} \\
&= \mathbf{I}_2
\end{aligned}$$

and  $\mathbf{D} = \mathbf{B}$  gives an SVD  $\mathbf{B} = \mathbf{I}_2 \mathbf{B} \mathbf{I}_2^\top$ , so by Definition 0.20,  $\mathbf{B}$  has one singular value,  $\sigma_1 = 3$ . R's `svd()` reports  $\min(2, 2) = 2$  singular values for  $\mathbf{B}$ : 3 and 0.

**Theorem 0.15** (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 0.19) of an  $n \times p$  matrix  $\mathbf{A}$  with singular values  $\sigma_1, \dots, \sigma_r$ . Then  $\mathbf{A}^\top \mathbf{A}$  is symmetric (Definition 0.3), and*

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

where  $\mathbf{\Lambda} \stackrel{\text{def}}{=} \mathbf{D}^\top \mathbf{D}$  is the  $p \times p$  diagonal matrix whose  $i$ -th diagonal entry  $\lambda_i$  is  $\sigma_i^2$  for  $i \leq r$  and 0 for  $i > r$ . So  $\mathbf{V} \mathbf{\Lambda} \mathbf{V}^\top$  is an eigendecomposition (Definition 0.18) of  $\mathbf{A}^\top \mathbf{A}$ : column  $i$  of  $\mathbf{V}$  is an eigenvector of  $\mathbf{A}^\top \mathbf{A}$  for the eigenvalue  $\lambda_i$ .

## i Proof

*Proof.* **Symmetry.**

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

The first step is [Theorem 17 in Matrices](#), and the second follows from [Definition 3 in Matrices](#), which swaps rows and columns, so swapping them again restores  $\mathbf{A}$ .

**The factorization.**

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

**The entries of  $\mathbf{\Lambda}$ .** Write  $\lambda_{ij}$  for entry  $(i, j)$  of  $\mathbf{\Lambda}$ , so  $\lambda_i = \lambda_{ii}$ . Only the entries  $d_{kk}$  with  $k \leq r$  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  $k = i \leq r$  and  $k = j$ . So  $\lambda_{ij} = 0$  when  $i \neq j$ . On the diagonal, only the term  $k = i$  can be nonzero, and for  $i \leq r$

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

For  $i > r$ , every term is 0, so  $\lambda_i = 0$ .

**The eigenvectors.** Write  $v_i$  for column  $i$  of  $\mathbf{V}$ , and  $e_i$  for the indicator vector ([Definition 18 in Vectors](#)) of length  $p$ . Column  $i$  of  $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_p$  says  $\mathbf{V}^\top v_i = e_i$ . Then

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

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

**Example 0.29** ( $\mathbf{A}^\top \mathbf{A}$  for the matrix of the SVD example). For  $\mathbf{A}$  in Example 0.28:

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

Theorem 0.15 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:

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

**i** Remark (Building an SVD from an eigendecomposition)

*Remark 0.18* (Building an SVD from an eigendecomposition). Theorem 0.15 matches Banerjee and Roy [6, Ch. 12], pp. 372-373, which works in the other direction: it constructs an SVD from a spectral decomposition of  $\mathbf{A}^\top \mathbf{A}$  and sets  $\sigma_i = \sqrt{\lambda_i}$ . For example, in Example 0.29 the eigenvalues of  $\mathbf{A}^\top \mathbf{A}$  are 4 and 2, and  $\sqrt{4} = 2$  and  $\sqrt{2}$  are the singular values  $\sigma_1$  and  $\sigma_2$  of  $\mathbf{A}$  from Example 0.28.

## Definite Matrices

**Definition 0.21** (Positive semidefinite matrix). A  $p \times p$  matrix  $\mathbf{A}$  is **positive semidefinite** [12] if it satisfies both conditions:

- $\mathbf{A}$  is symmetric (Definition 0.3).
- Every quadratic form (Definition 0.11) in  $\mathbf{A}$  is non-negative:  $\tilde{x}^\top \mathbf{A} \tilde{x} \geq 0$  for every vector  $\tilde{x}$  of length  $p$ .

**Example 0.30** (A positive semidefinite matrix). Let  $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ , which is symmetric.

For any  $\tilde{x} = (x_1, x_2)$ :

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

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

**Theorem 0.16** (Entries and symmetry of a Gram matrix). Let  $\mathbf{A}$  be an  $m \times n$  matrix with columns  $\tilde{a}_1, \dots, \tilde{a}_n$ , and let  $\mathbf{G} = \mathbf{A}^\top \mathbf{A}$  (Definition 6 in Column Space, Null Space and Rank-Nullity).

1. Entry  $(i, j)$  of  $\mathbf{G}$  is  $\tilde{a}_i \cdot \tilde{a}_j$ .
2.  $\mathbf{G}$  is symmetric (Definition 0.3).
3.  $\tilde{x}^\top \mathbf{G} \tilde{x} = \|\mathbf{A} \tilde{x}\|^2 \geq 0$  for every  $\tilde{x} \in \mathbb{R}^n$ , so  $\mathbf{G}$  is positive semidefinite (Definition 0.21).

**i** Proof

*Proof. Entries.* Row  $i$  of  $\mathbf{A}^\top$  is  $\tilde{a}_i^\top$ , so entry  $(i, j)$  of  $\mathbf{A}^\top \mathbf{A}$  is the dot product of that row with column  $j$  of  $\mathbf{A}$ , namely  $\tilde{a}_i^\top \tilde{a}_j = \tilde{a}_i \cdot \tilde{a}_j$ .

**Symmetry.** The dot product is commutative, so

$$\begin{aligned} g_{ij} &= \underline{a}_i \cdot \underline{a}_j \\ &= \underline{a}_j \cdot \underline{a}_i \\ &= g_{ji}. \end{aligned}$$

**Positive semidefiniteness.** For any  $\underline{x}$ ,

$$\begin{aligned} \underline{x}^\top \mathbf{G} \underline{x} &= \underline{x}^\top \mathbf{A}^\top \mathbf{A} \underline{x} && \text{(substitute } \mathbf{G}) \\ &= (\mathbf{A} \underline{x})^\top (\mathbf{A} \underline{x}) && \text{(Theorem 17 in Matrices)} \\ &= \|\mathbf{A} \underline{x}\|^2 && \text{(Equation 2 in Vectors, squared)} \\ &\geq 0. && \text{(a squared norm is nonnegative)} \end{aligned}$$

□

**Example 0.31** (Checking the properties on a small Gram matrix). For  $\mathbf{A}$  in [Example 8 in Column Space, Null Space and Rank-Nullity](#),  $\mathbf{G} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ . It is symmetric:  $g_{12} = g_{21} = 1$ . Take  $\underline{x} = (1, -1)$ . Then

$$\begin{aligned} \underline{x}^\top \mathbf{G} \underline{x} &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} && (\mathbf{G} \underline{x} = (1, -1)) \\ &= 2, && \text{(multiply and add)} \end{aligned}$$

and  $\mathbf{A} \underline{x} = (1, 0, -1)$  has  $\|\mathbf{A} \underline{x}\|^2 = 1 + 0 + 1 = 2$ , as the theorem says. The converse (every positive semidefinite matrix is a Gram matrix) is [this theorem](#).

**Definition 0.22** (Positive definite matrix). A  $p \times p$  matrix  $\mathbf{A}$  is **positive definite** [12] if it satisfies both conditions:

- $\mathbf{A}$  is symmetric (Definition 0.3).
- Every quadratic form (Definition 0.11) in  $\mathbf{A}$  at a nonzero vector is positive:  $\underline{x}^\top \mathbf{A} \underline{x} > 0$  for every vector  $\underline{x} \neq \underline{0}$  of length  $p$ .

**Example 0.32** (Positive definite, semidefinite, and neither).

- The identity matrix  $\mathbf{I}_p$  ([Definition 12 in Matrices](#)) is positive definite:  $\underline{x}^\top \mathbf{I}_p \underline{x} = \sum_{i=1}^p x_i^2$ , which is positive unless every  $x_i$  is 0.
- $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$  from [Example 0.30](#) is positive semidefinite but not positive definite: at

$$\tilde{x} = (1, -1) \neq \tilde{0},$$

$$\begin{aligned} \tilde{x}^\top \mathbf{B} \tilde{x} &= (1 - 1)^2 \\ &= 0. \end{aligned}$$

- $\mathbf{D} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$  is symmetric but not positive semidefinite: at  $\tilde{x} = (1, -1)$ ,

$$\begin{aligned} \tilde{x}^\top \mathbf{D} \tilde{x} &= 1 - 2 - 2 + 1 \\ &= -2 \\ &< 0. \end{aligned}$$

### **i** Remark (Why the definition requires symmetry)

*Remark 0.19* (Why the definition requires symmetry). A positive definite matrix is positive semidefinite (Definition 0.21):  $\tilde{x}^\top \mathbf{A} \tilde{x} > 0$  for every  $\tilde{x} \neq \tilde{0}$ , and  $\tilde{0}^\top \mathbf{A} \tilde{0} = 0$ . For example,  $\mathbf{I}_p$  in Example 0.32 is both.

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 (Definition 0.16). For example,  $\mathbf{C} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$  is not symmetric, and for any  $\tilde{x} = (x_1, x_2)$ :

$$\begin{aligned} \tilde{x}^\top \mathbf{C} \tilde{x} &= x_1(x_1 + x_2) + x_2(-x_1 + x_2) && \text{(multiply out the quadratic form)} \\ &= x_1^2 + x_1x_2 - x_2x_1 + x_2^2 && \text{(distribute)} \\ &= x_1^2 + x_2^2 && \text{(the middle terms cancel)} \end{aligned}$$

which is positive for every  $\tilde{x} \neq \tilde{0}$ . But suppose  $\mathbf{C}\tilde{v} = \lambda\tilde{v}$  for a real number  $\lambda$ . Matching entries gives  $v_1 + v_2 = \lambda v_1$  and  $-v_1 + v_2 = \lambda v_2$ , so  $v_2 = (\lambda - 1)v_1$  and  $-v_1 = (\lambda - 1)v_2$ . Substituting the first into the second gives  $-v_1 = (\lambda - 1)^2 v_1$ , so  $(1 + (\lambda - 1)^2)v_1 = 0$ . Since  $1 + (\lambda - 1)^2 > 0$ ,  $v_1 = 0$ , and then

$$\begin{aligned} v_2 &= (\lambda - 1)v_1 \\ &= 0. \end{aligned}$$

So no real  $\lambda$  has an eigenvector  $\tilde{v} \neq \tilde{0}$ .

See also [12].

**Theorem 0.17** (Definiteness and eigenvalues). *Let  $\mathbf{A}$  be a  $p \times p$  symmetric matrix with real entries, with eigendecomposition  $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$  (Definition 0.18) and eigenvalues*

$\lambda_1, \dots, \lambda_p$  on the diagonal of  $\mathbf{A}$ . Then:

- $\mathbf{A}$  is positive semidefinite (Definition 0.21) if and only if every  $\lambda_i \geq 0$ .
- $\mathbf{A}$  is positive definite (Definition 0.22) 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:

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

The second step is [Theorem 17 in Matrices](#). Also,  $\tilde{x} = \tilde{0}$  exactly when  $\tilde{y} = \tilde{0}$ :  $\mathbf{Q} \mathbf{Q}^\top = \mathbf{I}_p$  for an orthogonal matrix (Definition 0.10), 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,  $\lambda_i > 0$ .  $\square$

**Example 0.33** (Reading definiteness off the eigenvalues).

- $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$  has eigenvalues 3 and 1 (Example 0.24), 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 0.32 found directly.

**Theorem 0.18** (A positive definite matrix has a positive definite inverse). *Let  $\mathbf{A}$  be a  $p \times p$  positive definite matrix (Definition 0.22) with real entries, with eigendecomposition  $\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top$  (Definition 0.18) and eigenvalues  $\lambda_1, \dots, \lambda_p$ . Then  $\mathbf{A}$  is invertible (Definition 0.6),*

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

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

### **i** Proof

*Proof.* By Theorem 0.17, every  $\lambda_i > 0$ , so  $\mathbf{\Lambda}^{-1}$  is well defined, and multiplying the two diagonal matrices entry by entry gives

$$\begin{aligned} \mathbf{\Lambda} \mathbf{\Lambda}^{-1} &= \mathbf{\Lambda}^{-1} \mathbf{\Lambda} \\ &= \mathbf{I}_p. \end{aligned}$$

Write  $\mathbf{B} = \mathbf{Q} \mathbf{\Lambda}^{-1} \mathbf{Q}^\top$ . Then:

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

The same steps, with  $\mathbf{\Lambda}^{-1}$  and  $\mathbf{\Lambda}$  swapped, give  $\mathbf{B} \mathbf{A} = \mathbf{I}_p$ , so  $\mathbf{B} = \mathbf{A}^{-1}$  (Definition 0.5). The steps with  $\mathbf{Q}$  use Definition 0.10, whose remark records that  $\mathbf{Q} \mathbf{Q}^\top = \mathbf{I}_p$  too, and the identity step uses Theorem 8 in Matrices.

$\mathbf{A}^{-1}$  is symmetric by Corollary 0.1. For  $x \neq 0$ , let  $y = \mathbf{Q}^\top x$ , which is not 0 because  $x = \mathbf{Q} y$ . As in the proof of Theorem 0.17:

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

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

**Example 0.34** (Inverting a positive definite matrix). For  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ , Example 0.27

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,

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

**Definition 0.23** (Paraboloid). A **paraboloid** [13] in  $\mathbb{R}^{p+1}$  is the graph  $\{(\underline{x}, f(\underline{x})) \mid \underline{x} \in \mathbb{R}^p\}$  of a function

$$f(\underline{x}) = \underline{x}^\top \mathbf{S} \underline{x} + \underline{b} \cdot \underline{x} + c,$$

where  $\mathbf{S}$  is a nonzero symmetric  $p \times p$  matrix (Definition 0.3),  $\underline{b} \in \mathbb{R}^p$ , and  $c \in \mathbb{R}$ . The first term is a quadratic form (Definition 0.11).

**Definition 0.24** (Elliptic paraboloid). The paraboloid of Definition 0.23 is **elliptic** if  $\mathbf{S}$  is positive definite (Definition 0.22); it is then a bowl that opens upward.

**Example 0.35** (A bowl with its lowest point at the origin). Take  $p = 2$ ,  $\mathbf{S} = \mathbf{I}_2$ ,  $\underline{b} = \underline{0}_{2 \times 1}$  and  $c = 0$ . Then  $\mathbf{S}$  is positive definite (Example 0.32), so the graph of

$$\begin{aligned}
f(\underline{x}) &= \underline{x}^\top \mathbf{I}_2 \underline{x} \\
&= x_1^2 + x_2^2
\end{aligned}$$

is an elliptic paraboloid. At  $\underline{x} = \underline{0}$ ,  $f(\underline{x}) = 0$ ; at  $\underline{x} = (1, 1)$ ,  $f(\underline{x}) = 2$ ; and  $f(\underline{x}) > 0$  at every  $\underline{x} \neq \underline{0}$ , so the origin is the only lowest point of the bowl. Figure 1 shows the surface.

*(This 3D surface is interactive; see the HTML or slide version of this page to view it.)*

Figure 1: The surface  $z = x_1^2 + x_2^2$  over  $[-2, 2]^2$ : a bowl that opens upward, with its lowest point above the origin. Horizontal slices of the bowl are circles around the origin.

**Definition 0.25** (Matrix-induced inner product (**A**-inner product)). Let **A** be a  $p \times p$  symmetric positive definite matrix (Definition 0.22) ( $\mathbf{A} = \mathbf{A}^\top$  and  $\tilde{x}^\top \mathbf{A} \tilde{x} > 0$  for all nonzero  $\tilde{x} \in \mathbb{R}^p$ ). The **matrix-induced inner product** (or **A-inner product**) on  $\mathbb{R}^p$  is defined for all  $\tilde{x}, \tilde{y} \in \mathbb{R}^p$  by

$$\langle \tilde{x}, \tilde{y} \rangle_{\mathbf{A}} \stackrel{\text{def}}{=} \tilde{x}^\top \mathbf{A} \tilde{y}.$$

**Example 0.36** (A matrix-induced inner product calculation). Consider the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

For any nonzero vector  $\tilde{x} = (x_1, x_2) \in \mathbb{R}^2$ :

$$\begin{aligned} \tilde{x}^\top \mathbf{A} \tilde{x} &= 2x_1^2 + 2x_1x_2 + 2x_2^2 && \text{(expand the quadratic form)} \\ &= (x_1 + x_2)^2 + x_1^2 + x_2^2 && \text{(complete the square)} \end{aligned}$$

which is strictly positive because  $x_1^2 + x_2^2 > 0$  when  $\tilde{x} \neq \tilde{0}$ . Thus **A** is positive definite (Definition 0.22).

For vectors  $\tilde{x} = (1, 2)$  and  $\tilde{y} = (3, -1)$ , their inner product induced by **A** is:

$$\begin{aligned} \langle \tilde{x}, \tilde{y} \rangle_{\mathbf{A}} &= \tilde{x}^\top \mathbf{A} \tilde{y} && \text{(definition of } \langle \cdot, \cdot \rangle_{\mathbf{A}} \text{)} \\ &= \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} && \text{(substitute vectors and matrix)} \\ &= \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \cdot 3 + 1 \cdot (-1) \\ 1 \cdot 3 + 2 \cdot (-1) \end{bmatrix} && \text{(Definition 13 in Matrices)} \\ &= \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \end{bmatrix} && \text{(multiply and add)} \\ &= 1 \cdot 5 + 2 \cdot 1 && \text{(Definition 13 in Matrices)} \\ &= 7. && \text{(multiply and add)} \end{aligned}$$

In contrast, the standard dot product yields:

$$\begin{aligned} \tilde{x} \cdot \tilde{y} &= 1 \cdot 3 + 2 \cdot (-1) \\ &= 1. \end{aligned}$$

The off-diagonal entries

$$\begin{aligned} A_{12} &= A_{21} \\ &= 1 \end{aligned}$$

couple the coordinate directions, altering lengths and angles from their Euclidean values. When  $\mathbf{A} = \mathbf{I}_p$ , the matrix-induced inner product recovers the standard dot product  $\langle \tilde{x}, \tilde{y} \rangle_{\mathbf{I}_p} = \tilde{x} \cdot \tilde{y}$ .

**Theorem 0.19** (Properties of the matrix-induced inner product). *For every  $p \times p$  symmetric positive definite matrix  $\mathbf{A}$ , the function  $\langle \tilde{x}, \tilde{y} \rangle_{\mathbf{A}} = \tilde{x}^\top \mathbf{A} \tilde{y}$  satisfies the four inner product axioms (Definition 1 in Inner Products and Orthogonality): positivity, definiteness, linearity in the first slot, and symmetry. Thus  $(\mathbb{R}^p, \langle \cdot, \cdot \rangle_{\mathbf{A}})$  is an inner product space (Definition 2 in Inner Products and Orthogonality).*

*Taking  $\mathbf{A} = \mathbf{I}_p$  recovers the standard dot product (Example 1 in Inner Products and Orthogonality), while taking  $\mathbf{A} = \text{diag}(c_1, \dots, c_p)$  recovers the weighted inner product (Example 3 in Inner Products and Orthogonality).*

*Conversely, every inner product on  $\mathbb{R}^p$  has this form for a unique symmetric positive definite matrix  $\mathbf{A}$ , whose  $(i, j)$  entry is  $A_{ij} = \langle \tilde{e}_i, \tilde{e}_j \rangle$ , where  $\tilde{e}_1, \dots, \tilde{e}_p$  are the standard basis vectors (Definition 18 in Vectors).*

### i Proof

*Proof.*

- **positivity:**  $\langle \tilde{x}, \tilde{x} \rangle_{\mathbf{A}} = \tilde{x}^\top \mathbf{A} \tilde{x} \geq 0$  for all  $\tilde{x} \in \mathbb{R}^p$ , because  $\mathbf{A}$  is positive definite (Definition 0.22);
- **definiteness:**  $\langle \tilde{x}, \tilde{x} \rangle_{\mathbf{A}} = 0$  if and only if  $\tilde{x} = \mathbf{0}_{p \times 1}$ , by definition of a positive definite matrix (Definition 0.22);
- **linearity in the first slot:** for all  $\tilde{x}, \tilde{y}, \tilde{z} \in \mathbb{R}^p$  and real numbers  $a, b$ ,

$$\begin{aligned} \langle a\tilde{x} + b\tilde{y}, \tilde{z} \rangle_{\mathbf{A}} &= (a\tilde{x} + b\tilde{y})^\top \mathbf{A} \tilde{z} && \text{(definition of } \langle \cdot, \cdot \rangle_{\mathbf{A}} \text{)} \\ &= (a\tilde{x}^\top + b\tilde{y}^\top) \mathbf{A} \tilde{z} && \text{(Theorem 16 in Matrices, Definition 6 in Matrices, Definition 3)} \\ &= a\tilde{x}^\top \mathbf{A} \tilde{z} + b\tilde{y}^\top \mathbf{A} \tilde{z} && \text{(Theorem 6 in Matrices)} \\ &= a\langle \tilde{x}, \tilde{z} \rangle_{\mathbf{A}} + b\langle \tilde{y}, \tilde{z} \rangle_{\mathbf{A}} && \text{(definition of } \langle \cdot, \cdot \rangle_{\mathbf{A}} \text{)} \end{aligned}$$

- **symmetry:** because  $\tilde{y}^\top \mathbf{A} \tilde{x}$  is a  $1 \times 1$  scalar and  $\mathbf{A}$  is symmetric ( $\mathbf{A}^\top = \mathbf{A}$ ),

$$\begin{aligned} \langle \tilde{y}, \tilde{x} \rangle_{\mathbf{A}} &= \tilde{y}^\top \mathbf{A} \tilde{x} && \text{(definition of } \langle \cdot, \cdot \rangle_{\mathbf{A}} \text{)} \\ &= (\tilde{y}^\top \mathbf{A} \tilde{x})^\top && \text{(a scalar equals its transpose)} \\ &= \tilde{x}^\top \mathbf{A}^\top \tilde{y} && \text{(Theorem 17 in Matrices, twice)} \\ &= \tilde{x}^\top \mathbf{A} \tilde{y} && \text{(\mathbf{A} is symmetric)} \\ &= \langle \tilde{x}, \tilde{y} \rangle_{\mathbf{A}} && \text{(definition of } \langle \cdot, \cdot \rangle_{\mathbf{A}} \text{)} \end{aligned}$$

For the converse, let  $\langle \cdot, \cdot \rangle$  be any inner product on  $\mathbb{R}^p$ . Any vectors  $\tilde{x}, \tilde{y} \in \mathbb{R}^p$  can be written in the standard basis as  $\tilde{x} = \sum_{i=1}^p x_i \tilde{e}_i$  and  $\tilde{y} = \sum_{j=1}^p y_j \tilde{e}_j$  (Definition 6 in

Subspaces and Rank, [Definition 18 in Vectors](#)). By linearity in both slots ([Remark 1 in Inner Products and Orthogonality](#)):

$$\begin{aligned}
\langle \tilde{x}, \tilde{y} \rangle &= \left\langle \sum_{i=1}^p x_i \tilde{e}_i, \sum_{j=1}^p y_j \tilde{e}_j \right\rangle && \text{(expand in the standard basis)} \\
&= \sum_{i=1}^p \sum_{j=1}^p x_i y_j \langle \tilde{e}_i, \tilde{e}_j \rangle && \text{(linearity in both slots)} \\
&= \sum_{i=1}^p x_i \sum_{j=1}^p A_{ij} y_j && \text{(define } A_{ij} \stackrel{\text{def}}{=} \langle \tilde{e}_i, \tilde{e}_j \rangle) \\
&= \tilde{x}^\top \mathbf{A} \tilde{y}. && \text{(Definition 13 in Matrices)}
\end{aligned}$$

Symmetry of  $\langle \cdot, \cdot \rangle$  ensures

$$\begin{aligned}
A_{ji} &= \langle \tilde{e}_j, \tilde{e}_i \rangle \\
&= \langle \tilde{e}_i, \tilde{e}_j \rangle \\
&= A_{ij},
\end{aligned}$$

so  $\mathbf{A}$  is symmetric ([Definition 0.3](#)). For any  $\tilde{x} \neq 0$ ,  $\tilde{x}^\top \mathbf{A} \tilde{x} = \langle \tilde{x}, \tilde{x} \rangle > 0$  by positivity and definiteness of the inner product, so  $\mathbf{A}$  is positive definite ([Definition 0.22](#)).

For uniqueness, if another matrix  $\mathbf{B}$  satisfies  $\langle \tilde{x}, \tilde{y} \rangle = \tilde{x}^\top \mathbf{B} \tilde{y}$  for all  $\tilde{x}, \tilde{y}$ , evaluating at  $\tilde{x} = \tilde{e}_i$  and  $\tilde{y} = \tilde{e}_j$  gives

$$\begin{aligned}
B_{ij} &= \tilde{e}_i^\top \mathbf{B} \tilde{e}_j \\
&= \langle \tilde{e}_i, \tilde{e}_j \rangle \\
&= A_{ij},
\end{aligned}$$

so  $\mathbf{B} = \mathbf{A}$ . □

**Example 0.37** (Checking the inner product axioms, and recovering the matrix from the inner product). Take  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$  from [Example 0.36](#), the vectors  $\tilde{x} = (1, 2)$ ,  $\tilde{y} = (3, -1)$  and  $\tilde{z} = (0, 4)$ , and the numbers  $a = 2$  and  $b = -3$ . The code computes both sides of each axiom in [Theorem 0.19](#), and then the numbers  $\langle \tilde{e}_i, \tilde{e}_j \rangle_{\mathbf{A}}$  for the standard basis vectors.

```

A <- matrix(c(2, 1, 1, 2), nrow = 2)
ip <- function(x, y) drop(t(x) %*% A %*% y)

x <- c(1, 2)
y <- c(3, -1)
z <- c(0, 4)
a <- 2
b <- -3

c(
  positivity = ip(x, x),
  symmetry_xy = ip(x, y),
  symmetry_yx = ip(y, x),
  linearity_left = ip(a * x + b * y, z),
  linearity_right = a * ip(x, z) + b * ip(y, z)
)
#>      positivity      symmetry_xy      symmetry_yx      linearity_left      linearity_right
#>           14              7              7              28              28

e <- diag(2)
recovered <- outer(1:2, 1:2, Vectorize(function(i, j) ip(e[, i], e[, j])))
recovered
#>      [,1] [,2]
#> [1,]    2    1
#> [2,]    1    2

```

The inner product of  $x$  with itself is 14, which is positive. The two orders give the same number, 7 and 7. Both sides of the linearity axiom equal 28. The matrix of numbers  $\langle \tilde{e}_i, \tilde{e}_j \rangle_{\mathbf{A}}$  is equal to  $\mathbf{A}$ , as the converse in Theorem 0.19 says.

**Corollary 0.3** (A positive definite matrix is a transpose product). *Let  $\mathbf{A}$  be a  $p \times p$  symmetric positive definite matrix (Definition 0.22). Then  $\mathbf{A} = \mathbf{W}^\top \mathbf{W}$  for some invertible  $p \times p$  matrix  $\mathbf{W}$ .*

### **i** Proof

*Proof.* By the spectral theorem (Theorem 0.12),  $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$  with  $\mathbf{Q}$  orthogonal and  $\mathbf{\Lambda}$  diagonal with the eigenvalues  $\lambda_i$  of  $\mathbf{A}$  on its diagonal. These are positive (Theorem 0.17), so  $\mathbf{\Lambda}^{1/2}$ , the diagonal matrix with entries  $\sqrt{\lambda_i}$ , is defined. Take  $\mathbf{W} = \mathbf{\Lambda}^{1/2}\mathbf{Q}^\top$ . Since  $\mathbf{\Lambda}^{1/2}$  is diagonal,  $(\mathbf{\Lambda}^{1/2})^\top = \mathbf{\Lambda}^{1/2}$ , and

$$\begin{aligned}
\mathbf{W}^\top \mathbf{W} &= \mathbf{Q} \mathbf{\Lambda}^{1/2} \mathbf{\Lambda}^{1/2} \mathbf{Q}^\top \\
&= \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top \\
&= \mathbf{A}.
\end{aligned}$$

$\mathbf{W}$  is invertible, because  $\mathbf{Q}^\top$  has inverse  $\mathbf{Q}$  and  $\mathbf{\Lambda}^{1/2}$  is diagonal with positive entries, so it has the diagonal inverse with entries  $1/\sqrt{\lambda_i}$ , and a product of invertible matrices is invertible (Theorem 0.1).  $\square$

**Theorem 0.20** (A matrix-induced inner product is a dot product of transformed vectors). *Let  $\mathbf{A}$  be a  $p \times p$  symmetric positive definite matrix, and let  $\mathbf{W}$  be any  $p \times p$  matrix with  $\mathbf{A} = \mathbf{W}^\top \mathbf{W}$ . Then for all  $\tilde{x}, \tilde{y} \in \mathbb{R}^p$ , the  $\mathbf{A}$ -inner product (Definition 0.25) satisfies*

$$\langle \tilde{x}, \tilde{y} \rangle_{\mathbf{A}} = (\mathbf{W}\tilde{x}) \cdot (\mathbf{W}\tilde{y}).$$

#### **i** Proof

*Proof.*

$$\begin{aligned}
\langle \tilde{x}, \tilde{y} \rangle_{\mathbf{A}} &= \tilde{x}^\top \mathbf{W}^\top \mathbf{W} \tilde{y} \\
&= (\mathbf{W}\tilde{x})^\top (\mathbf{W}\tilde{y}) \\
&= (\mathbf{W}\tilde{x}) \cdot (\mathbf{W}\tilde{y}).
\end{aligned}$$

$\square$

#### **i** Remark (Inner product spaces in machine learning)

*Remark 0.20* (Inner product spaces in machine learning). In machine learning and statistics, choosing or learning an inner product space tailors geometric notions of distance, angle, and projection to the structure of the data:

1. **Mahalanobis metric and whitening:** Standard Euclidean distance treats all coordinate axes identically, ignoring feature correlations and variance differences. If data features have a symmetric positive definite covariance matrix  $\mathbf{\Sigma}$  (Definition 0.22), the precision matrix  $\mathbf{A} = \mathbf{\Sigma}^{-1}$  (Theorem 0.18) defines the Mahalanobis inner product  $\langle \tilde{x}, \tilde{y} \rangle_{\mathbf{\Sigma}^{-1}} = \tilde{x}^\top \mathbf{\Sigma}^{-1} \tilde{y}$ . The induced distance  $d_{\mathbf{\Sigma}^{-1}}(\tilde{x}, \tilde{y}) = \sqrt{(\tilde{x} - \tilde{y})^\top \mathbf{\Sigma}^{-1} (\tilde{x} - \tilde{y})}$  measures statistical distance normalized by variance along each principal direction. Factoring  $\mathbf{\Sigma}^{-1} = \mathbf{W}^\top \mathbf{W}$  (via Cholesky or spectral decomposition, Corollary 0.3), this inner product is the Euclidean dot product of the whitened vectors  $\mathbf{W}\tilde{x}$  and  $\mathbf{W}\tilde{y}$  (Theorem 0.20).
2. **The kernel trick and reproducing kernel Hilbert spaces (RKHS):** Algorithms such as support vector machines (SVMs), kernel ridge regression, and

Gaussian processes depend on training vectors solely through pairwise inner products  $\langle \tilde{x}_i, \tilde{x}_j \rangle$ . By mapping data into a higher- or infinite-dimensional inner product space via a feature map  $D : \mathbb{R}^p \rightarrow \mathcal{H}$ , the kernel trick replaces the computation of  $D(\tilde{x}_i)$  with an evaluation of a positive definite kernel function:

$$k(\tilde{x}_i, \tilde{x}_j) \stackrel{\text{def}}{=} \langle D(\tilde{x}_i), D(\tilde{x}_j) \rangle_{\mathcal{H}}.$$

Mercer's theorem guarantees that every symmetric positive definite kernel function implicitly defines a valid inner product space  $\mathcal{H}$  (a reproducing kernel Hilbert space) without requiring explicit coordinate representations.

3. **Embeddings and cosine similarity:** In natural language processing and recommendation systems, data items are embedded as vectors in high-dimensional representations. The angle between embeddings captures semantic similarity through cosine similarity ([Definition 9 in Inner Products and Orthogonality](#)):

$$\text{cossim}(\tilde{x}, \tilde{y}) = \frac{\langle \tilde{x}, \tilde{y} \rangle}{\|\tilde{x}\| \|\tilde{y}\|}.$$

When embedding vectors are normalized to unit Euclidean norm ( $\|\tilde{x}\| = 1$  and  $\|\tilde{y}\| = 1$ ), cosine similarity equals the inner product  $\langle \tilde{x}, \tilde{y} \rangle$  ([Corollary 2 in Inner Products and Orthogonality](#)). Similarly, the self-attention mechanism in Transformers computes query-key alignment scores through scaled inner products  $\tilde{q} \cdot \tilde{k} / \sqrt{d_k}$ .

4. **Weighted and generalized least squares:** In linear regression  $y \approx \mathbf{X}\beta$ , when observation errors have unequal variances or known correlations with positive definite covariance matrix  $\mathbf{\Omega}$ , ordinary least squares is inefficient. Generalized least squares minimizes the residual norm in the inner product space induced by  $\mathbf{W} = \mathbf{\Omega}^{-1}$ :

$$\min_{\tilde{\beta}} \|\tilde{y} - \mathbf{X}\tilde{\beta}\|_{\mathbf{W}}^2 = \min_{\tilde{\beta}} \langle \tilde{y} - \mathbf{X}\tilde{\beta}, \tilde{y} - \mathbf{X}\tilde{\beta} \rangle_{\mathbf{W}}.$$

## Determinants

**Definition 0.26** (Determinant). The **determinant** [14] 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  $p \geq 2$ ,

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

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

**Example 0.38** (Determinant of a  $2 \times 2$  matrix). For  $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ , deleting row 1 and column 1 leaves  $[d]$ , and deleting row 1 and column 2 leaves  $[c]$ . So:

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

For example,

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

### Remark (Cofactor expansion)

*Remark 0.21* (Cofactor expansion). The recursive formula in Definition 0.26 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 [6]. Each ordering picks one entry from each row, from the column that the ordering assigns to that row, and multiplies them. Its sign is  $+1$  if the ordering takes an even number of swaps of two columns to reach from  $(1, 2, \dots, p)$ , and  $-1$  if it takes an odd number. For  $p = 2$ , the columns have two orderings. The ordering  $(1, 2)$  keeps each column in place, picks the entries  $a_{11}$  and  $a_{22}$ , and has sign  $+1$ . The ordering  $(2, 1)$  swaps the two columns, picks the entries  $a_{12}$  and  $a_{21}$ , and has sign  $-1$ . The sum  $a_{11}a_{22} - a_{12}a_{21}$  is the value  $ad - bc$  from Example 0.38.

See also [14].

### Video

Chapter 6 of the *Essence of linear algebra* series by 3Blue1Brown shows the geometric meaning of the determinant:

- [The determinant](#) shows how the determinant measures the factor by which a linear transformation scales areas and volumes, and why a zero determinant corresponds to compressing space into a lower dimension.

**Theorem 0.21** (Determinant of a diagonal matrix). *The determinant (Definition 0.26) of a  $p \times p$  diagonal matrix (Definition 0.4) is the product of its diagonal entries:*

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

### **i** Proof

*Proof.* By [induction](#) on  $p$ . For  $p = 1$ , the matrix is  $[d_1]$ , 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  $d_2, \dots, d_p$ :

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

□

**Example 0.39** (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 0.22** (Determinant of a product). *For  $p \times p$  matrices  $\mathbf{A}$  and  $\mathbf{B}$ ,*

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

**Example 0.40** (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 0.38](#) gives

$$\begin{aligned}\det(\mathbf{A}) &= 2 \cdot 1 - 1 \cdot 0 \\ &= 2\end{aligned}$$

and

$$\begin{aligned}\det(\mathbf{B}) &= 1 \cdot 1 - 0 \cdot 3 \\ &= 1.\end{aligned}$$

The product is

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

with

$$\begin{aligned}\det(\mathbf{AB}) &= 5 \cdot 1 - 1 \cdot 3 \\ &= 2 \\ &= \det(\mathbf{A}) \det(\mathbf{B}).\end{aligned}$$

**i** Remark (An example is not a proof)

*Remark 0.22* (An example is not a proof). Example 0.40 checks Theorem 0.22 for one pair of  $2 \times 2$  matrices, which shows the theorem holds there but does not prove it for every pair. The proof needs properties of the determinant that these notes don't develop; see Banerjee and Roy [6].

See also [14].

**Theorem 0.23** (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  $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$  (Definition 0.18) and eigenvalues  $\lambda_1, \dots, \lambda_p$  on the diagonal of  $\mathbf{\Lambda}$ . Then*

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

**i** Proof

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

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

Then:

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

The product steps are Theorem 0.22, the orthogonality step is Definition 0.10, and the identity and diagonal steps are Theorem 0.21 and Example 0.39.  $\square$

**Corollary 0.4** (A positive definite matrix has a positive determinant). *If  $\mathbf{A}$  is positive definite (Definition 0.22), then  $\det(\mathbf{A}) > 0$ .*

#### **i** Proof

*Proof.* By Theorem 0.17, every eigenvalue of  $\mathbf{A}$  is positive, so their product, which is  $\det(\mathbf{A})$  by Theorem 0.23, is positive.  $\square$

**Example 0.41** (Two ways to the same determinant).  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$  has eigenvalues 3 and 1 (Example 0.24), so Theorem 0.23 gives

$$\begin{aligned}
\det(\mathbf{A}) &= 3 \cdot 1 \\
&= 3,
\end{aligned}$$

matching  $2 \cdot 2 - 1 \cdot 1 = 3$  from Example 0.38.

- [1] Wikipedia contributors, “Symmetric matrix — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Symmetric\\_matrix](https://en.wikipedia.org/wiki/Symmetric_matrix)
- [2] Wikipedia contributors, “Diagonal matrix — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Diagonal\\_matrix](https://en.wikipedia.org/wiki/Diagonal_matrix)
- [3] Wikipedia contributors, “Invertible matrix — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Invertible\\_matrix](https://en.wikipedia.org/wiki/Invertible_matrix)

- [4] Wikipedia contributors, “Idempotent matrix — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Idempotent\\_matrix](https://en.wikipedia.org/wiki/Idempotent_matrix)
- [5] Wikipedia contributors, “Orthogonal matrix — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Orthogonal\\_matrix](https://en.wikipedia.org/wiki/Orthogonal_matrix)
- [6] S. Banerjee and A. Roy, *Linear algebra and matrix analysis for statistics*, vol. 181. Crc Press Boca Raton, 2014. Available: <https://www.routledge.com/Linear-Algebra-and-Matrix-Analysis-for-Statistics/Banerjee-Roy/p/book/9781420095388>
- [7] Wikipedia contributors, “Quadratic form — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Quadratic\\_form](https://en.wikipedia.org/wiki/Quadratic_form)
- [8] Wikipedia contributors, “Trace (linear algebra) — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Trace\\_\(linear\\_algebra\)](https://en.wikipedia.org/wiki/Trace_(linear_algebra))
- [9] Wikipedia contributors, “Frobenius norm — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Frobenius\\_norm](https://en.wikipedia.org/wiki/Frobenius_norm)
- [10] Wikipedia contributors, “Eigendecomposition of a matrix — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Eigendecomposition\\_of\\_a\\_matrix](https://en.wikipedia.org/wiki/Eigendecomposition_of_a_matrix)
- [11] Wikipedia contributors, “Singular value decomposition — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Singular\\_value\\_decomposition](https://en.wikipedia.org/wiki/Singular_value_decomposition)
- [12] Wikipedia contributors, “Definite matrix — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Definite\\_matrix](https://en.wikipedia.org/wiki/Definite_matrix)
- [13] Wikipedia contributors, “Paraboloid — Wikipedia, the free encyclopedia.” 2026. Available: <https://en.wikipedia.org/wiki/Paraboloid>
- [14] Wikipedia contributors, “Determinant — Wikipedia, the free encyclopedia.” 2026. Available: <https://en.wikipedia.org/wiki/Determinant>