Definite Matrices

Last modified: 2026-11-24 09:19:45 (PST)

Definite Matrices

Definition 0.1 (Positive semidefinite matrix)**.** A $p \times p$ matrix $\mathbf{A}$ is **positive semidefinite** [1] if it satisfies both conditions:

- $\mathbf{A}$ is symmetric (Definition 3 in Special Matrices, Quadratic Forms and Trace).
- Every quadratic form (Definition 13 in Special Matrices, Quadratic Forms and Trace) in $\mathbf{A}$ is non-negative: $\underset{\sim}{x}^\top \mathbf{A} \underset{\sim}{x} \geq 0$ for every vector $\underset{\sim}{x}$ of length p.

Example 0.1 (A positive semidefinite matrix)**.** Let $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, which is symmetric. For any $\underset{\sim}{x} = (x_1, x_2)$:

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

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

Theorem 0.1 (Entries and symmetry of a Gram matrix)**.** *Let $\mathbf{A}$ be an $m \times n$ matrix with columns $\underset{\sim}{a}_1, \ldots, \underset{\sim}{a}_n$, and let $\mathbf{G} = \mathbf{A}^\top \mathbf{A}$ (Definition 6 in Column Space and Null Space).*

1. *Entry (i, j) of $\mathbf{G}$ is $\underset{\sim}{a}_i \cdot \underset{\sim}{a}_j$.*
2. *$\mathbf{G}$ is symmetric (Definition 3 in Special Matrices, Quadratic Forms and Trace).*
3. *$\underset{\sim}{x}^\top \mathbf{G} \underset{\sim}{x} = \|\mathbf{A}\underset{\sim}{x}\|^2 \geq 0$ for every $\underset{\sim}{x} \in \mathbb{R}^n$, so $\mathbf{G}$ is positive semidefinite (Definition 0.1).*

Proof

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

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

Positive semidefiniteness. For any $\underset{\sim}{x}$,

$$\begin{aligned} \underset{\sim}{x}^\top \mathbf{G} \underset{\sim}{x} &= \underset{\sim}{x}^\top \mathbf{A}^\top \mathbf{A} \underset{\sim}{x} && \text{(substitute } \mathbf{G}) \\ &= (\mathbf{A}\underset{\sim}{x})^\top (\mathbf{A}\underset{\sim}{x}) && \text{(Theorem 6 in Matrix Multiplication)} \\ &= \|\mathbf{A}\underset{\sim}{x}\|^2 && \text{(Equation 5 in Vectors, squared)} \\ &\geq 0. && \text{(a squared norm is nonnegative)} \end{aligned}$$

□

Example 0.2 (Checking the properties on a small Gram matrix)**.** For $\mathbf{A}$ in Example 8 in Column Space and Null Space, $\mathbf{G} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. It is symmetric: $g_{12} = g_{21} = 1$. Take $\underset{\sim}{x} = (1, -1)$. Then

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

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

Corollary 0.1 (A self outer product is positive semidefinite)**.** *For a vector $\underset{\sim}{x}$, the self outer product $\underset{\sim}{x}\underset{\sim}{x}^\top$ is positive semidefinite.*

Proof

Proof. Let $\mathbf{A} = \underset{\sim}{x}^\top$, the $1 \times n$ matrix whose one row is $\underset{\sim}{x}$. Then

$$
\begin{aligned}
\mathbf{A}^\top \mathbf{A} &= \mathbf{A}^\top \mathbf{A} && \text{(definition of the Gram matrix)} \\
&= (\underset{\sim}{x}^\top)^\top \underset{\sim}{x}^\top && \text{(substitute } \mathbf{A}\text{)} \\
&= \underset{\sim}{x}\underset{\sim}{x}^\top && \text{(the transpose of a transpose is the original matrix)} \\
&= \underset{\sim}{x}\underset{\sim}{x}^\top && \text{(definition of the self outer product)}
\end{aligned}
$$

So $\underset{\sim}{x}\underset{\sim}{x}^\top$ is the Gram matrix of $\mathbf{A}$, and Theorem 0.1 says that a Gram matrix is positive semidefinite. □

Example 0.3 (A quadratic form in a self outer product)**.** For $\underset{\sim}{x} = (1, 2)$, the example of a self outer product found $\underset{\sim}{x}\underset{\sim}{x}^\top = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, which is symmetric. Take $\underset{\sim}{w} = (2, -1)$. The quadratic form in $\underset{\sim}{x}\underset{\sim}{x}^\top$ at $\underset{\sim}{w}$ is

$$
\begin{aligned}
\underset{\sim}{w}^\top \underset{\sim}{x}\underset{\sim}{x}^\top \underset{\sim}{w} &= \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} && \text{(definition of the quadratic form)} \\
&= \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} && (1 \cdot 2 + 2 \cdot (-1) = 0 \text{ and } 2 \cdot 2 + 4 \cdot (-1) = 0) \\
&= 0 && \text{(multiply and add)}
\end{aligned}
$$

The quadratic form is 0 here, which is not negative, as Corollary 0.1 says. It is not positive either, because $\underset{\sim}{w}$ is orthogonal to $\underset{\sim}{x}$.

Definition 0.2 (Positive definite matrix)**.** A $p \times p$ matrix $\mathbf{A}$ is **positive definite** [1] if it satisfies both conditions:

- $\mathbf{A}$ is symmetric (Definition 3 in Special Matrices, Quadratic Forms and Trace).
- Every quadratic form (Definition 13 in Special Matrices, Quadratic Forms and Trace) in $\mathbf{A}$ at a nonzero vector is positive: $\underset{\sim}{x}^\top \mathbf{A} \underset{\sim}{x} > 0$ for every vector $\underset{\sim}{x} \neq \underset{\sim}{0}$ of length p.

Example 0.4 (Positive definite, semidefinite, and neither)**.**

- The identity matrix $\mathbf{I}_p$ (Definition 7 in Matrix Multiplication) is positive definite: $\underset{\sim}{x}^\top \mathbf{I}_p \underset{\sim}{x} = \sum_{i=1}^p x_i^2$, which is positive unless every x_i is 0.
- $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ from Example 0.1 is positive semidefinite but not positive definite: at $\underset{\sim}{x} = (1, -1) \neq \underset{\sim}{0}$,

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

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

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

Remark (Why the definition requires symmetry)

Remark 0.1 (Why the definition requires symmetry). A positive definite matrix is positive semidefinite (Definition 0.1): $\underset{\sim}{x}^\top \mathbf{A} \underset{\sim}{x} > 0$ for every $\underset{\sim}{x} \neq \underset{\sim}{0}$, and $\underset{\sim}{0}^\top \mathbf{A} \underset{\sim}{0} = 0$. For example, $\mathbf{I}_p$ in Example 0.4 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 1 in Eigenvalues and Matrix Decompositions). For example, $\mathbf{C} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ is not symmetric, and for any $\underset{\sim}{x} = (x_1, x_2)$:

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

which is positive for every $\underset{\sim}{x} \neq \underset{\sim}{0}$. But suppose $\mathbf{C}\underset{\sim}{v} = \lambda \underset{\sim}{v}$ for a real number λ. 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 λ has an eigenvector $\underset{\sim}{v} \neq \underset{\sim}{0}$.

See also [1].

Theorem 0.2 (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 3 in Eigenvalues and Matrix Decompositions) and eigenvalues* $\lambda_1, \ldots, \lambda_p$ *on the diagonal of* $\mathbf{\Lambda}$*. Then:*

- $\mathbf{A}$ *is positive semidefinite (Definition 0.1) if and only if every* $\lambda_i \geq 0$.

- $\mathbf{A}$ *is positive definite (Definition 0.2) if and only if every* $\lambda_i > 0$.

Proof

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

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

The second step is Theorem 6 in Matrix Multiplication. Also, $\underset{\sim}{x} = \underset{\sim}{0}$ exactly when $\underset{\sim}{y} = \underset{\sim}{0}$: $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}_p$ for an orthogonal matrix (Definition 10 in Special Matrices, Quadratic Forms and Trace), so $\underset{\sim}{x} = \mathbf{Q}\underset{\sim}{y}$.
If every $\lambda_i \geq 0$, then every term $\lambda_i y_i^2 \geq 0$, so $\underset{\sim}{x}^\top \mathbf{A} \underset{\sim}{x} \geq 0$. *If every* $\lambda_i > 0$ and $\underset{\sim}{x} \neq \underset{\sim}{0}$, then some $y_i \neq 0$, so at least one term is positive and the rest are non-negative, and $\underset{\sim}{x}^\top \mathbf{A} \underset{\sim}{x} > 0$.
Conversely, take $\underset{\sim}{x} = q_i$, column i of $\mathbf{Q}$, which is not $\underset{\sim}{0}$ because it has norm 1. Then $\underset{\sim}{y} = \mathbf{Q}^\top q_i$ is column i of $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_p$, so $y_i = 1$ and every other entry of $\underset{\sim}{y}$ is 0, and the display gives ${q_i}^\top \mathbf{A} q_i = \lambda_i$. So if $\mathbf{A}$ is positive semidefinite, $\lambda_i \geq 0$, and if $\mathbf{A}$ is positive definite, $\lambda_i > 0$. □

Example 0.5 (Reading definiteness off the eigenvalues)**.**

- $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has eigenvalues 3 and 1 (Example 1 in Eigenvalues and Matrix Decompositions), 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 0.4 found directly.

Definition 0.3 (Condition number of a positive definite matrix)**.** Let $\mathbf{S}$ be a $p \times p$ positive definite matrix with eigenvalues $\lambda_{\min} \leq \cdots \leq \lambda_{\max}$, all positive by definiteness and eigenvalues. The **condition number** of $\mathbf{S}$ is

$$\kappa(\mathbf{S}) \stackrel{\text{def}}{=} \frac{\lambda_{\max}}{\lambda_{\min}}.$$

The condition number is at least 1, that is, $\kappa(\mathbf{S}) \geq 1$, because $\lambda_{\max} \geq \lambda_{\min} > 0$, and

$\kappa(\mathbf{S}) = 1$ exactly when all eigenvalues coincide.

Example 0.6 (A condition number of 3)**.** $\mathbf{S} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has eigenvalues 3 and 1 (eigenvalue example), so $\kappa(\mathbf{S}) = 3/1 = 3$. The identity matrix $\mathbf{I}_2$ has eigenvalues 1 and 1, so $\kappa(\mathbf{I}_2) = 1$.

Definition 0.4 (Condition number of a matrix)**.** The **condition number** of an $n \times p$ matrix $\mathbf{A}$ is the ratio of its largest to its smallest singular value, counting all $\min(n, p)$ singular values of $\mathbf{A}$, zeros included (Remark 7 in Eigenvalues and Matrix Decompositions) [2]:

$$\kappa(\mathbf{A}) \stackrel{\text{def}}{=} \frac{\sigma_1}{\sigma_{\min(n,p)}}.$$

This generalizes Definition 0.3, which covers positive definite matrices. For a nonzero matrix, $\kappa(\mathbf{A}) \geq 1$. When the columns of a tall matrix $\mathbf{A}$ ($n \geq p$) are linearly dependent, its smallest singular value is 0, and $\kappa(\mathbf{A}) = \infty$.

Example 0.7 (The condition number of a matrix with nearly parallel columns)**.** Let $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1.01 \end{bmatrix}$. Its columns, $(1, 1)$ and $(1, 1.01)$, are nearly parallel. We find its singular values and their ratio, then repeat after multiplying every entry by 1000:

R

```
entries_mcn <- c(1, 1, 1, 1.01)
a_mcn <- matrix(entries_mcn, nrow = 2)
svd_mcn <- svd(a_mcn)
sv_mcn <- svd_mcn$d
a_mcn_scaled <- 1000 * a_mcn
svd_mcn_scaled <- svd(a_mcn_scaled)
sv_mcn_scaled <- svd_mcn_scaled$d
sv_mcn
#> [1] 2.0050125 0.0049875
sv_mcn[1] / sv_mcn[2]
#> [1] 402.0075
sv_mcn_scaled[1] / sv_mcn_scaled[2]
#> [1] 402.0075
```

Python

```python
import numpy as np

a_mcn = np.array([[1.0, 1.0], [1.0, 1.01]])
sv_mcn = np.linalg.svd(a_mcn, compute_uv=False)
sv_mcn_scaled = np.linalg.svd(1000 * a_mcn, compute_uv=False)
print(sv_mcn)
#> [2.0050125 0.0049875]
print(sv_mcn[0] / sv_mcn[1])
#> 402.0075124842948
print(sv_mcn_scaled[0] / sv_mcn_scaled[1])
#> 402.00751248429236
```

The singular values are 2.005 and 0.005, so $\kappa(\mathbf{A}) = 402$. Multiplying every entry by 1000 multiplies both singular values by 1000, so the condition number is still 402.

i Remark (Other condition numbers)

Remark 0.2 (Other condition numbers). Other norms give other condition numbers. This definition uses the 2-norm, which is the one built from the singular values. For a positive definite matrix, the singular values are the eigenvalues, and this definition agrees with the ratio of its largest to its smallest eigenvalue in Definition 0.3.

Theorem 0.3 (Forming $\mathbf{A}^\top\mathbf{A}$ squares the condition number)**.** *Let* $\mathbf{A}$ *be a nonzero* $n \times p$ *matrix with* $n \geq p$*. Then*

$$\kappa(\mathbf{A}^\top\mathbf{A}) = \kappa(\mathbf{A})^2.$$

i Proof

Proof. By Theorem 4 in Eigenvalues and Matrix Decompositions, the eigenvalues of the symmetric matrix $\mathbf{A}^\top\mathbf{A}$ are $\sigma_1^2 \geq \cdots \geq \sigma_p^2$, where $\sigma_1 \geq \cdots \geq \sigma_p$ are the singular values of $\mathbf{A}$, with zeros counted (Remark 7 in Eigenvalues and Matrix Decompositions). These eigenvalues are non-negative, so for the symmetric matrix $\mathbf{A}^\top\mathbf{A}$ they are also its singular values. Therefore

$$\kappa(\mathbf{A}^\top\mathbf{A}) = \frac{\sigma_1^2}{\sigma_p^2} = \left(\frac{\sigma_1}{\sigma_p}\right)^2 = \kappa(\mathbf{A})^2.$$

If $\sigma_p = 0$, both sides are ∞. □

Example 0.8 (Squaring the condition number)**.** For the matrix $\mathbf{A}$ from Example 0.7, compare $\kappa(\mathbf{A})$ with $\kappa(\mathbf{A}^\top\mathbf{A})$:

R

```
# Uses a_mcn and sv_mcn from the earlier example
gram_cng <- crossprod(a_mcn)
kappa_a_cng <- sv_mcn[1] / sv_mcn[2]
svd_gram_cng <- svd(gram_cng)
sv_gram_cng <- svd_gram_cng$d
sv_gram_cng[1] / sv_gram_cng[2]
#> [1] 161610
kappa_a_cng^2
#> [1] 161610
```

Python

```
# Uses a_mcn and sv_mcn from the earlier example
gram_cng = a_mcn.T @ a_mcn
kappa_a_cng = sv_mcn[0] / sv_mcn[1]
sv_gram_cng = np.linalg.svd(gram_cng, compute_uv=False)
print(sv_gram_cng[0] / sv_gram_cng[1])
#> 161610.0400912735
print(kappa_a_cng ** 2)
#> 161610.04009381044
```

The condition number of $\mathbf{A}^\top\mathbf{A}$ is 161610, and $\kappa(\mathbf{A})^2 = 161610$, as Theorem 0.3 says.

Definition 0.5 (Ill-conditioned matrix)**.** A matrix $\mathbf{A}$ is **ill-conditioned** when its condition number is so large that small relative changes in its entries, such as those from rounding, can change the solution to a linear system by a much larger relative amount. The relative change in the solution can be up to $\kappa(\mathbf{A})$ times the relative change in the entries. A solution computed with $\kappa(\mathbf{A}) \approx 10^k$ therefore loses about k of the digits that the floating-point arithmetic carries [3], Lecture 12.
There is no sharp threshold: how large is too large depends on how many accurate digits the answer needs.

Definition 0.6 (Well-conditioned matrix)**.** A matrix $\mathbf{A}$ is **well-conditioned** when its condition number is close to 1, so it is not ill-conditioned.

Example 0.9 (A small change in the right-hand side moves the solution a long way)**.** Let $\mathbf{A}$ be the matrix from Example 0.7, and let $\underset{\sim}{b} = (2, 2.01)$. The system $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{b}$ has solution $\underset{\sim}{x} = (1, 1)$. Change the second entry of $\underset{\sim}{b}$ from 2.01 to 2.011, and solve again:

R

```
# Uses a_mcn and sv_mcn from the previous example
b_ic <- c(2, 2.01)
b_new_ic <- c(2, 2.011)
x_ic <- solve(a_mcn, b_ic)
x_new_ic <- solve(a_mcn, b_new_ic)
x_ic
#> [1] 1 1
x_new_ic
#> [1] 0.9 1.1

# Relative changes, measured with the Euclidean length
db_ic <- (b_new_ic - b_ic)^2 |> sum() |> sqrt()
nb_ic <- b_ic^2 |> sum() |> sqrt()
dx_ic <- (x_new_ic - x_ic)^2 |> sum() |> sqrt()
nx_ic <- x_ic^2 |> sum() |> sqrt()
rel_b_ic <- db_ic / nb_ic
rel_x_ic <- dx_ic / nx_ic
c(rel_b_ic, rel_x_ic, rel_x_ic / rel_b_ic)
#> [1] 3.526706e-04 1.000000e-01 2.835507e+02
```

Python

```
# Uses a_mcn and sv_mcn from the previous example
b_ic = np.array([2.0, 2.01])
b_new_ic = np.array([2.0, 2.011])
x_ic = np.linalg.solve(a_mcn, b_ic)
x_new_ic = np.linalg.solve(a_mcn, b_new_ic)
print(x_ic)
#> [1. 1.]
print(x_new_ic)
#> [0.9 1.1]

# Relative changes, measured with the Euclidean length
rel_b_ic = np.linalg.norm(b_new_ic - b_ic) / np.linalg.norm(b_ic)
rel_x_ic = np.linalg.norm(x_new_ic - x_ic) / np.linalg.norm(x_ic)
print(rel_b_ic, rel_x_ic, rel_x_ic / rel_b_ic)
#> 0.0003526706147110207 0.1000000000000334 283.55070093371313
```

The right-hand side changes by a relative amount of 0.00035, and the solution changes by a relative amount of 0.100, about 284 times as much. That amplification is what Definition 0.5 describes, and it cannot exceed $\kappa(\mathbf{A}) = 402.0$.

Theorem 0.4 (A positive definite matrix has a positive definite inverse)**.** *Let* $\mathbf{A}$ *be a* $p \times p$ *positive definite matrix (Definition 0.2) with real entries, with eigendecomposition* $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$ *(Definition 3 in Eigenvalues and Matrix Decompositions) and eigenvalues* $\lambda_1, \dots, \lambda_p$*. Then* $\mathbf{A}$ *is invertible (Definition 6 in Special Matrices, Quadratic Forms and Trace),*

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

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

Proof

Proof. By Theorem 0.2, 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:

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

The same steps, with $\mathbf{\Lambda}^{-1}$ and $\mathbf{\Lambda}$ swapped, give $\mathbf{BA} = \mathbf{I}_p$, so $\mathbf{B} = \mathbf{A}^{-1}$ (Definition 5 in Special Matrices, Quadratic Forms and Trace). The steps with $\mathbf{Q}$ use Definition 10 in Special Matrices, Quadratic Forms and Trace, whose remark records that $\mathbf{QQ}^\top = \mathbf{I}_p$ too, and the identity step uses Theorem 4 in Matrix Multiplication.

$\mathbf{A}^{-1}$ is symmetric by Corollary 1 in Special Matrices, Quadratic Forms and Trace. For $\underset{\sim}{x} \neq \underset{\sim}{0}$, let $\underset{\sim}{y} = \mathbf{Q}^\top\underset{\sim}{x}$, which is not $\underset{\sim}{0}$ because $\underset{\sim}{x} = \mathbf{Q}\underset{\sim}{y}$. As in the proof of Theorem 0.2:

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

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

Example 0.10 (Inverting a positive definite matrix)**.** For $\mathbf{A} = \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}$, Example 4 in Eigenvalues and Matrix Decompositions 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 0.7 (Paraboloid)**.** A **paraboloid** [4] in $\mathbb{R}^{p+1}$ is the graph $\{(\underset{\sim}{x}, f(\underset{\sim}{x})) \mid \underset{\sim}{x} \in \mathbb{R}^p\}$ of a function

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

where $\mathbf{S}$ is a nonzero symmetric $p \times p$ matrix (Definition 3 in Special Matrices, Quadratic Forms and Trace), $\underset{\sim}{b} \in \mathbb{R}^p$, and $c \in \mathbb{R}$. The first term is a quadratic form (Definition 13 in Special Matrices, Quadratic Forms and Trace).

Definition 0.8 (Elliptic paraboloid)**.** The paraboloid of Definition 0.7 is **elliptic** if $\mathbf{S}$ is positive definite (Definition 0.2); it is then a bowl that opens upward.

Example 0.11 (A bowl with its lowest point at the origin)**.** Take $p = 2$, $\mathbf{S} = \mathbf{I}_2$, $\underset{\sim}{b} = \underset{\sim}{0}_{2\times 1}$ and $c = 0$. Then $\mathbf{S}$ is positive definite (Example 0.4), so the graph of

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

is an elliptic paraboloid. At $\underset{\sim}{x} = \underset{\sim}{0}$, $f(\underset{\sim}{x}) = 0$; at $\underset{\sim}{x} = (1, 1)$, $f(\underset{\sim}{x}) = 2$; and $f(\underset{\sim}{x}) > 0$ at every $\underset{\sim}{x} \neq \underset{\sim}{0}$, 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 0.9 (Matrix-induced inner product ($\mathbf{A}$-inner product))**.** Let $\mathbf{A}$ be a $p \times p$ symmetric positive definite matrix. The **matrix-induced inner product** (or **A-inner product**) on $\mathbb{R}^p$ is the symmetric bilinear form of $\mathbf{A}$, used as an inner product: for all $\underset{\sim}{x}, \underset{\sim}{y} \in \mathbb{R}^p$,

$$\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle_{\mathbf{A}} \stackrel{\text{def}}{=} B_{\mathbf{A}}(\underset{\sim}{x}, \underset{\sim}{y}).$$

Example 0.12 (A matrix-induced inner product calculation)**.** Consider the symmetric matrix

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

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

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

which is strictly positive because $x_1^2 + x_2^2 > 0$ when $\underset{\sim}{x} \neq \underset{\sim}{0}$. Thus $\mathbf{A}$ is positive definite (Definition 0.2).
For vectors $\underset{\sim}{x} = (1, 2)$ and $\underset{\sim}{y} = (3, -1)$, their inner product induced by $\mathbf{A}$ is:

$$\begin{aligned} \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle_{\mathbf{A}} &= \underset{\sim}{x}^\top \mathbf{A} \underset{\sim}{y} && \text{(definition of } \langle \cdot, \cdot \rangle_{\mathbf{A}}) \\ &= \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} && \text{(substitute vectors and matrix)} \\ &= \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \cdot 3 + 1 \cdot (-1) \\ 1 \cdot 3 + 2 \cdot (-1) \end{bmatrix} && \text{(Definition 8 in Matrix Multiplication)} \\ &= \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \end{bmatrix} && \text{(multiply and add)} \\ &= 1 \cdot 5 + 2 \cdot 1 && \text{(Definition 8 in Matrix Multiplication)} \\ &= 7. && \text{(multiply and add)} \end{aligned}$$

In contrast, the standard dot product yields:

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

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 $\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle_{\mathbf{I}_p} = \underset{\sim}{x} \cdot \underset{\sim}{y}$.

ℹ Remark (A symmetric bilinear form needs a positive definite matrix to be an inner product)

Remark 0.3 (A symmetric bilinear form needs a positive definite matrix to be an inner product). Definition 0.9 uses the symmetric bilinear form of $\mathbf{A}$ as an inner product only when $\mathbf{A}$ is also positive definite. A symmetric matrix $\mathbf{A}$ that is not positive definite has a nonzero vector $\underset{\sim}{x}$ with $B_{\mathbf{A}}(\underset{\sim}{x}, \underset{\sim}{x}) \leq 0$. If $B_{\mathbf{A}}(\underset{\sim}{x}, \underset{\sim}{x}) < 0$, the form breaks positivity; if $B_{\mathbf{A}}(\underset{\sim}{x}, \underset{\sim}{x}) = 0$, it breaks definiteness. Either way, that form is not an inner product. For

example, take $\mathbf{S}$ from Example 18 in Special Matrices, Quadratic Forms and Trace and the vector $\underset{\sim}{z} = (1, -1)^\top$:

$$
\begin{aligned}
B_{\mathbf{S}}(\underset{\sim}{z}, \underset{\sim}{z}) &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} && \text{(Definition 11 in Special Matrices, Quadratic Forms and Trace)} \\
&= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \cdot 1 + 2 \cdot (-1) \\ 2 \cdot 1 + 1 \cdot (-1) \end{bmatrix} && \text{(multiply the matrix by } \underset{\sim}{z}\text{)} \\
&= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} && \text{(multiply and add)} \\
&= 1 \cdot (-1) + (-1) \cdot 1 && \text{(multiply the row by the column)} \\
&= -2. && \text{(multiply and add)}
\end{aligned}
$$

The quadratic form of $\mathbf{S}$ is negative at $\underset{\sim}{z}$, so $\mathbf{S}$ is not positive definite.

Theorem 0.5 (Properties of the matrix-induced inner product)**.** *For every $p \times p$ symmetric positive definite matrix $\mathbf{A}$, the function $\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle_{\mathbf{A}} = \underset{\sim}{x}^\top \mathbf{A} \underset{\sim}{y}$ satisfies the four inner product axioms (Definition 5 in Inner Products and the Cauchy-Schwarz Inequality): 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 6 in Inner Products and the Cauchy-Schwarz Inequality). Taking $\mathbf{A} = \mathbf{I}_p$ recovers the standard dot product (Example 1 in Inner Products and the Cauchy-Schwarz Inequality), while taking $\mathbf{A} = \operatorname{diag}(c_1, \dots, c_p)$ recovers the weighted inner product (Example 3 in Inner Products and the Cauchy-Schwarz Inequality).*
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 $A_{ij} = \langle \underset{\sim}{e}_i, \underset{\sim}{e}_j \rangle$, where $\underset{\sim}{e}_1, \dots, \underset{\sim}{e}_p$ are the standard basis vectors (Definition 21 in Vectors).

Proof

Proof.

- **positivity**: $\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle_{\mathbf{A}} = \underset{\sim}{x}^\top \mathbf{A} \underset{\sim}{x} \geq 0$ for all $\underset{\sim}{x} \in \mathbb{R}^p$, because $\mathbf{A}$ is positive definite (Definition 0.2);
- **definiteness**: $\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle_{\mathbf{A}} = 0$ if and only if $\underset{\sim}{x} = \underset{\sim}{0}_{p \times 1}$, by definition of a positive definite matrix (Definition 0.2);

- **linearity in the first slot**: for all $\underset{\sim}{x}, \underset{\sim}{y}, \underset{\sim}{z} \in \mathbb{R}^p$ and real numbers a, b,

$$\begin{aligned}
\langle a\,\underset{\sim}{x} + b\,\underset{\sim}{y}, \underset{\sim}{z}\rangle_{\mathbf{A}} &= (a\,\underset{\sim}{x} + b\,\underset{\sim}{y})^{\top}\,\mathbf{A}\underset{\sim}{z} && \text{(definition of } \langle\cdot,\cdot\rangle_{\mathbf{A}}\text{)} \\
&= (a\,\underset{\sim}{x}^{\top} + b\,\underset{\sim}{y}^{\top})\,\mathbf{A}\underset{\sim}{z} && \text{(Theorem 5 in Matrix Multiplication, Definition 9 in Matrices, I} \\
&= a\,\underset{\sim}{x}^{\top}\mathbf{A}\underset{\sim}{z} + b\,\underset{\sim}{y}^{\top}\mathbf{A}\underset{\sim}{z} && \text{(Theorem 2 in Matrix Multiplication)} \\
&= a\,\langle \underset{\sim}{x}, \underset{\sim}{z}\rangle_{\mathbf{A}} + b\,\langle \underset{\sim}{y}, \underset{\sim}{z}\rangle_{\mathbf{A}} && \text{(definition of } \langle\cdot,\cdot\rangle_{\mathbf{A}}\text{)}
\end{aligned}$$

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

$$\begin{aligned}
\langle \underset{\sim}{y}, \underset{\sim}{x}\rangle_{\mathbf{A}} &= \underset{\sim}{y}^{\top}\mathbf{A}\underset{\sim}{x} && \text{(definition of } \langle\cdot,\cdot\rangle_{\mathbf{A}}\text{)} \\
&= (\underset{\sim}{y}^{\top}\mathbf{A}\underset{\sim}{x})^{\top} && \text{(a scalar equals its transpose)} \\
&= \underset{\sim}{x}^{\top}\mathbf{A}^{\top}\underset{\sim}{y} && \text{(Theorem 6 in Matrix Multiplication, twice)} \\
&= \underset{\sim}{x}^{\top}\mathbf{A}\underset{\sim}{y} && (\mathbf{A} \text{ is symmetric}) \\
&= \langle \underset{\sim}{x}, \underset{\sim}{y}\rangle_{\mathbf{A}} && \text{(definition of } \langle\cdot,\cdot\rangle_{\mathbf{A}}\text{)}
\end{aligned}$$

For the converse, let $\langle\cdot,\cdot\rangle$ be any inner product on $\mathbb{R}^p$. Any vectors $\underset{\sim}{x}, \underset{\sim}{y} \in \mathbb{R}^p$ can be written in the standard basis as $\underset{\sim}{x} = \sum_{i=1}^{p} x_i\,\underset{\sim}{e}_i$ and $\underset{\sim}{y} = \sum_{j=1}^{p} y_j\,\underset{\sim}{e}_j$ (Definition 1 in Bases and Dimension, Definition 21 in Vectors). By linearity in both slots (Remark 1 in Inner Products and the Cauchy-Schwarz Inequality):

$$\begin{aligned}
\langle \underset{\sim}{x}, \underset{\sim}{y}\rangle &= \left\langle \sum_{i=1}^{p} x_i\,\underset{\sim}{e}_i, \sum_{j=1}^{p} y_j\,\underset{\sim}{e}_j \right\rangle && \text{(expand in the standard basis)} \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} x_i y_j\,\langle \underset{\sim}{e}_i, \underset{\sim}{e}_j\rangle && \text{(linearity in both slots)} \\
&= \sum_{i=1}^{p} x_i \sum_{j=1}^{p} A_{ij} y_j && (\text{define } A_{ij} \overset{\text{def}}{=} \langle \underset{\sim}{e}_i, \underset{\sim}{e}_j\rangle) \\
&= \underset{\sim}{x}^{\top}\mathbf{A}\underset{\sim}{y}. && \text{(Definition 8 in Matrix Multiplication)}
\end{aligned}$$

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

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

so $\mathbf{A}$ is symmetric (Definition 3 in Special Matrices, Quadratic Forms and Trace). For any $\underset{\sim}{x} \neq \underset{\sim}{0}$, $\underset{\sim}{x}^{\top}\mathbf{A}\underset{\sim}{x} = \langle \underset{\sim}{x}, \underset{\sim}{x}\rangle > 0$ by positivity and definiteness of the inner product, so $\mathbf{A}$ is positive definite (Definition 0.2).

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

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

so $\mathbf{B} = \mathbf{A}$. □

Example 0.13 (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 0.12, the vectors $\underset{\sim}{x} = (1, 2)$, $\underset{\sim}{y} = (3, -1)$ and $\underset{\sim}{z} = (0, 4)$, and the numbers $a = 2$ and $b = -3$. The code computes both sides of each axiom in Theorem 0.5, and then the numbers $\langle \underset{\sim}{e}_i, \underset{\sim}{e}_j \rangle_{\mathbf{A}}$ 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 $\underset{\sim}{x}$ 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 $\langle \underset{\sim}{e}_i, \underset{\sim}{e}_j \rangle_{\mathbf{A}}$ is equal to $\mathbf{A}$, as the converse in Theorem 0.5 says.

Definition 0.10 (Matrix-induced norm, **A**-norm)**.** Let $\mathbf{A}$ be a $p \times p$ positive definite matrix. The **matrix-induced norm**, or **A-norm**, on $\mathbb{R}^p$ is the induced norm of the matrix-induced inner product $\langle \cdot, \cdot \rangle_{\mathbf{A}}$: for every $\underset{\sim}{x} \in \mathbb{R}^p$,

$$\|\underset{\sim}{x}\|_{\mathbf{A}} \stackrel{\text{def}}{=} \sqrt{\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle_{\mathbf{A}}}.$$

Example 0.14 (A matrix-induced norm calculation)**.** Take the positive definite matrix $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ of Example 0.12 and the vector $\underset{\sim}{x} = (1, 2)$. Then

$$\begin{aligned}
\|\underset{\sim}{x}\|_{\mathbf{A}} &= \sqrt{\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle_{\mathbf{A}}} && \text{(Definition 0.10)} \\
&= \sqrt{\underset{\sim}{x}^\top \mathbf{A} \underset{\sim}{x}} && \text{(Definition 0.9)} \\
&= \sqrt{\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}} && \text{(substitute the vector and the matrix)} \\
&= \sqrt{\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \cdot 1 + 1 \cdot 2 \\ 1 \cdot 1 + 2 \cdot 2 \end{bmatrix}} && \text{(Definition 8 in Matrix Multiplication)} \\
&= \sqrt{\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \end{bmatrix}} && \text{(multiply and add)} \\
&= \sqrt{1 \cdot 4 + 2 \cdot 5} && \text{(Definition 8 in Matrix Multiplication)} \\
&= \sqrt{14}. && \text{(multiply and add)}
\end{aligned}$$

The Euclidean norm of the same vector is $\|\underset{\sim}{x}\| = \sqrt{1^2 + 2^2} = \sqrt{5}$, so the **A**-norm of this vector is larger than its Euclidean norm.

Example 0.15 (A symmetric matrix that induces no norm)**.** The symmetric matrix $\mathbf{B} = \operatorname{diag}(1, -1)$ is not positive definite, so it has no matrix-induced norm. The square root $\sqrt{\underset{\sim}{x}^\top \mathbf{B} \underset{\sim}{x}}$ is undefined at $\underset{\sim}{x} = (0, 1)$:

$$\begin{aligned}
\underset{\sim}{x}^\top \mathbf{B} \underset{\sim}{x} &= 1 \cdot 0^2 + (-1) \cdot 1^2 && \text{(expand the quadratic form)} \\
&= -1, && \text{(multiply and add)}
\end{aligned}$$

and -1 has no square root in $\mathbb{R}$.

Corollary 0.2 (A positive definite matrix is a transpose product)**.** *Let* $\mathbf{A}$ *be a* $p \times p$ *symmetric positive definite matrix (Definition 0.2). Then* $\mathbf{A} = \mathbf{W}^\top \mathbf{W}$ *for some invertible* $p \times p$ *matrix* $\mathbf{W}$*.*

Proof

Proof. By the spectral theorem (Theorem 1 in Eigenvalues and Matrix Decompositions), $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$ with $\mathbf{Q}$ orthogonal and $\mathbf{\Lambda}$ diagonal with the eigenvalues λ_i of $\mathbf{A}$ on its diagonal. These are positive (Theorem 0.2), 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, $\left(\mathbf{\Lambda}^{1/2}\right)^\top = \mathbf{\Lambda}^{1/2}$, 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 1 in Special Matrices, Quadratic Forms and Trace). □

Theorem 0.6 (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* $\underset{\sim}{x}, \underset{\sim}{y} \in \mathbb{R}^p$*, the* $\mathbf{A}$*-inner product (Definition 0.9) satisfies*

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

Proof

Proof.

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

□

Remark (Inner product spaces in machine learning)

Remark 0.4 (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 0.2), the precision matrix $\mathbf{A} = \mathbf{\Sigma}^{-1}$ (Theorem 0.4) defines the Mahalanobis inner product $\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle_{\mathbf{\Sigma}^{-1}} = \underset{\sim}{x}^\top\mathbf{\Sigma}^{-1}\underset{\sim}{y}$. The induced distance $d_{\mathbf{\Sigma}^{-1}}(\underset{\sim}{x}, \underset{\sim}{y}) =$

$\sqrt{(\underset{\sim}{x}-\underset{\sim}{y})^\top \mathbf{\Sigma}^{-1}(\underset{\sim}{x}-\underset{\sim}{y})}$ 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 0.2), this inner product is the Euclidean dot product of the whitened vectors $\mathbf{W}\underset{\sim}{x}$ and $\mathbf{W}\underset{\sim}{y}$ (Theorem 0.6).

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 $\langle \underset{\sim}{x}_i, \underset{\sim}{x}_j \rangle$. By mapping data into a higher- or infinite-dimensional inner product space via a feature map $D : \mathbb{R}^p \to \mathcal{H}$, the kernel trick replaces the computation of $D(\underset{\sim}{x}_i)$ with an evaluation of a positive definite kernel function:

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

   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 14 in Inner Products and the Cauchy-Schwarz Inequality):

$$\operatorname{cossim}(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle}{\|\underset{\sim}{x}\| \, \|\underset{\sim}{y}\|}.$$

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

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 matrix-induced norm of the residual, with $\mathbf{W} = \mathbf{\Omega}^{-1}$:

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

[1] Wikipedia contributors, "Definite matrix — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Definite_matrix

[2] Wikipedia contributors, "Condition number — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Condition_number

[3] L. N. Trefethen and D. Bau III, *Numerical linear algebra*. SIAM, 1997. doi: 10.1137/1.9780898719574.

[4] Wikipedia contributors, "Paraboloid — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Paraboloid