Inverse of a matrix

The **inverse** of a matrix $\mathbf{A}$ is the matrix $\mathbf{C}$ that when either left or right multiplied by $\mathbf{A}$, yields the identity matrix. That is, if for a matrix $\mathbf{C}$ it holds that

$$\mathbf{AC} = \mathbf{CA} = \mathbf{I}$$

then $\mathbf{C}$ is the inverse of $\mathbf{A}$ (Definition 1). The inverse of a matrix $\mathbf{A}$ is denoted as $\mathbf{A}^{-1}$.

The inverse of a matrix is not guaranteed to exist. If it does exist, the matrix is called an **invertible matrix**. Otherwise, it is called a **singular matrix**.

Intuition

There are a number of perspectives for which one can view an invertible matrix:

1. An invertible matrix characterizes an invertible linear transformation
2. An invertible matrix preserves information
3. An invertible matrix preserves the dimensionality of a transformed vector whereas a singular matrix collapses vectors into a lower-dimensional subspace
4. An invertible matrix computes a change of coordinates for a vector space

These perspectives are discussed in the following sections.

1. An invertible matrix characterizes an invertible linear transformation

Any matrix that meets the criteria of Definition 1 characterizes an invertible linear transformation. That is given an invertible matrix $\mathbf{A}$, the linear transformation

$$T(\mathbf{x}) := \mathbf{Ax}$$

has an inverse linear transformation $T^{-1}(\mathbf{x})$. Recall, for a function to be invertible it must be both onto and one-to-one. Theorem 3 proves that $T(\mathbf{x})$ is onto. To prove that it is one-to-one, we first must prove that the null space of $T(\mathbf{x})$ consists of only the zero vector (Theorem 1). Then, this fact is used to prove that a linear transformation with this characteristic must be one-to-one (Theorem 2).

Said differently, the inverse of a matrix $\mathbf{A}$ is the matrix that "reverts" vectors transformed by $\mathbf{A}$ back to their original vectors (Figure 4). Thus, since matrix multiplication encodes a composition of the matrices' linear transformations, it follows that a matrix multiplied by its inverse yields the identity matrix $\mathbf{I}$, which characterizes the linear transformation that maps vectors back to themselves. This observation allows us to rigorously define the inverse of a matrix $\mathbf{A}$ as the matrix that when multiplied by $\mathbf{A}^{-1}$ yields the identity matrix as is done by Definition 1.

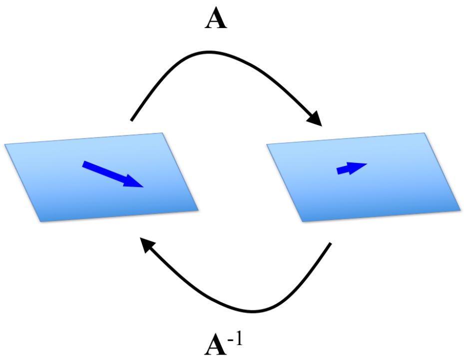

Figure 1: The inverse of matrix $\mathbf{A}$ denoted $\mathbf{A}^{-1}$ maps the vector $\mathbf{Ax}$ back to $\mathbf{x}$.

2. An invertible matrix preserves information

The transformation carried out by an invertible matrix $\mathbf{A}$ can be "reverted." That is, let $\mathbf{b}$ be the vector that results from transforming $\mathbf{x}$ with $\mathbf{A}$. We can recover the original $\mathbf{x}$ by multiplying $\mathbf{b}$ by $\mathbf{A}^{-1}$:

$$\begin{aligned} \mathbf{b} &:= \mathbf{Ax} \\ \implies \mathbf{A}^{-1}\mathbf{b} &= \mathbf{A}^{-1}\mathbf{Ax} \\ \implies \mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \end{aligned}$$

Inherently, $\mathbf{A}$ preserves all of the information of $\mathbf{x}$ in $\mathbf{b}$ as evidenced by the fact that we can recover $\mathbf{x}$ from $\mathbf{b}$ via $\mathbf{A}^{-1}$. If, on the other hand, $\mathbf{A}$ is singular, then we cannot recover $\mathbf{x}$ from $\mathbf{b}$. Intuitively, information about $\mathbf{x}$ is lost in the transformation into $\mathbf{b}$.

3. A singular matrix collapses vectors into a lower-dimensional subspace

The loss of information described in the previous section can be viewed geometrically by the fact that a singular matrix "collapses" or "compresses" vectors into an intrinsic lower dimensional space. That is, a singular matrix reduces the intrinsic dimensionality of the vectors. The loss of these dimensions constitutes the "loss of information" discussed in the previous section.

As shown in Theorem 5, a matrix is invertible if and only if its columns are linearly independent. Recall a set of n linearly independent vectors

$$S := \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$$

spans a space with an intrinsic dimensionality of n because in order to specify any vector $\mathbf{b}$ in the vector space, one must specify the coefficients $c_1, \dots, c_n$ such that

$$\mathbf{b} = c_1\mathbf{x}_1 + \dots + c_n\mathbf{x}_n$$

However, if S is not linearly independent, then we can throw away "redundant" vectors in S that can be constructed from the remaining vectors. Thus, the intrinsic dimensionality of a linearly dependent set S is the maximum sized subset of S that is linearly independent.

When a matrix $\mathbf{A}$ is singular, its columns are linearly dependent and thus, the vectors that constitute the column space of the matrix is inherently of lower dimension that the number of columns. Thus, when $\mathbf{A}$ multiplies a vector $\mathbf{x}$, it transforms $\mathbf{x}$ into this lower dimensional space. Once transformed, there is no way to transform it back to its original vector because certain dimensions of the vector were "lost" in this transformation.

To make this more concrete, an example of this phenomenon can be seen in Figure 4. In Figure 4, a singular matrix $\mathbf{A} \in \mathbb{R}^{3 \times 3}$ maps vectors in $\mathbb{R}^3$ to vectors that lie on a plane in $\mathbb{R}^3$. All vectors on a plane in $\mathbb{R}^3$ are of intrinsic dimensionality of two rather than three because we only need to specify coefficients for two of the column vectors in $\mathbb{A}$ to specify a point on the plane. We can throw away the third. Thus, we see that this singular matrix collapses points from the full 3-dimensional space $\mathbb{R}^3$ to the 2-dimensional space on the plane spanned by the columns of $\mathbb{A}$.

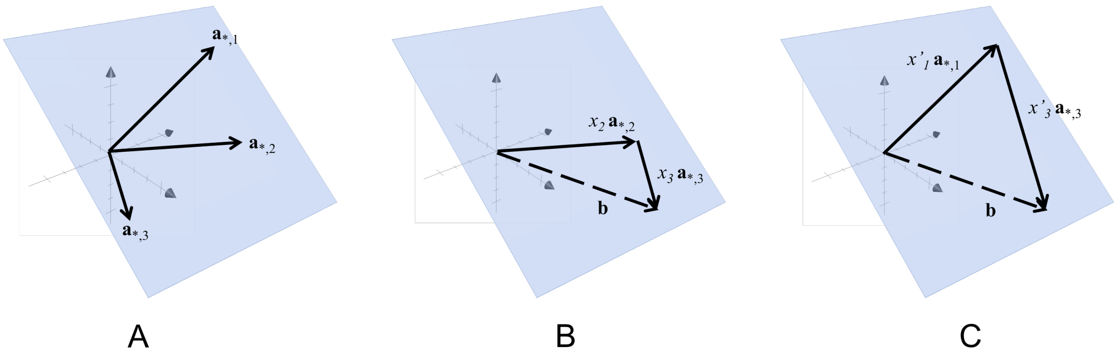

Figure 2: (A) The column vectors of a matrix $\mathbf{A} \in \mathbb{R}^{3 \times 3}$. (B) One solution to the equation $\mathbf{Ax} = \mathbf{b}$. (C) Another solution to $\mathbf{Ax} = \mathbf{b}$. That is, there are multiple $\mathbf{x} \in \mathbb{R}^3$ that map to $\mathbf{b}$. Thus, there does not exist an inverse mapping and therefore no inverse matrix to $\mathbf{A}$. These multiple constructions of mappings from $\mathbb{R}^3$ to $\mathbf{b}$ arise directly from the fact that the columns of $\mathbf{A}$ are linearly dependent.

An invertible matrix computes a change of coordinates for a vector space

A vector $\mathbf{x} \in \mathbb{R}^n$ can be viewed as the coordinates for a point in a coordinate system. That is, for each dimension i, the vector $\mathbf{x}$ provides a value along each dimension (e.g. x_i is the value along dimension i). Of course, the coordinate system we use can be arbitrary. In Figure 3, we can specify locations in $\mathbf{R}^2$ using either the grey coordinate system or the blue coordinate system. Furthermore, there is a one-to-one and onto mapping between coordinates in each of these two alternative coordinate systems. The point $\mathbf{x}$ located at $[-4, -2]$ in the grey coordinate system can be described as $[-1, 1]$ according to the blue coordinate system.

All coordinate systems are, in some sense, equivalent in that each is able to provide an unambiguous location for points in the space. Nonetheless, it often helps to have some coordinate system that acts as a reference to every other coordinate system. This reference coordinate system is the defined by the standard basis vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. All other coordinate systems can then be constructed from this reference coordinate system. In Figure 3, the reference coordinate system is depicted by the grey grid and is constructed by the orthonormal basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. An alternative coordinate system is depicted by the blue grid and is constructed from the basis vectors $\mathbf{a}_1$ and $\mathbf{a}_2$.

An invertible matrix $\mathbf{A} := [\mathbf{a}_1, \dots, \mathbf{a}_n]$ can be viewed as an operator that converts vectors described in terms of the basis vectors $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ back to a description in terms of the standard basis vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. That is, if we have some vector $\mathbf{x}' \in \mathbb{R}^n$, then $\mathbf{A}\mathbf{x}'$ can be understood to be the vector in the standard basis *if* $\mathbf{x}'$ was described according to the basis formed by the columns of $\mathbf{A}$. Another way to think about this is that if we have some vector $\mathbf{x} \in \mathbb{R}^n$ described according to the standard basis, then we can describe $\mathbf{x}$ in terms of an alternative basis $\mathbf{a}_1, \dots, \mathbf{a}_n$ by multiplying $\mathbf{x}$ by the inverse of the matrix $\mathbf{A} := [\mathbf{a}_1, \dots, \mathbf{a}_n]$. That is

$$\mathbf{x}_\mathbf{A} := \mathbf{A}^{-1}\mathbf{x}$$

is the representation of $\mathbf{x}$ in terms of the basis columns of $\mathbf{A}$.

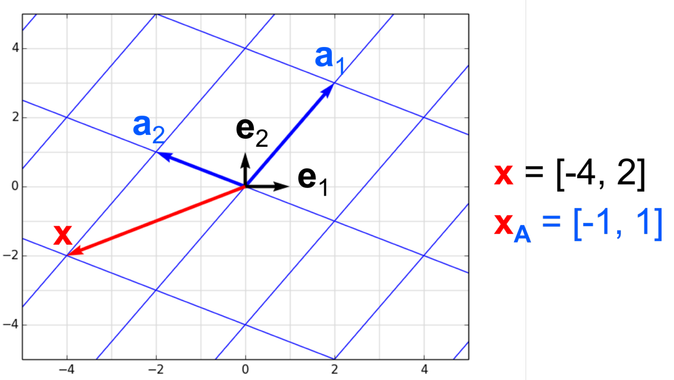

Figure 3: Here we have a vector $\mathbf{x} \in \mathbb{R}^2$. Using the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2$, the vector $\mathbf{x}$ is given by $[-4, -2]$. Using the column vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ of matrix $\mathbf{A}$ as a basis, this vector $\mathbf{x}_\mathbf{A}$ is given by $[-1, 1]$. The matrix $\mathbf{A}$ maps $\mathbf{x}_\mathbf{A}$ to $\mathbf{x}$.

Properties

Below we discuss several properties of invertible matrices that provide more intuition into how they behave and also provide algebraic rules that can be used in derivations.

1. **The columns of an invertible matrix are linearly independent** (Theorem 5).

2. **Taking the inverse of an inverse matrix gives you back the original matrix.** (Theorem 7). Given an invertible matrix $\mathbf{A}$ with inverse $\mathbf{A}^{-1}$, it follows from Definition 1, that $\mathbf{A}^{-1}$ is also invertible with inverse $\mathbf{A}$. That is,

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

   This also follows from the fact that the inverse of an inverse function f^{-1} is simply the original function f.

3. **The result of multiplying invertible matrices is invertible** (Theorem 8). Given two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$, the matrix that results from their multiplication is invertible. That is, $\mathbf{AB}$ is invertible and its inverse is given by

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

   Recall the result of matrix multiplication results in a matrix that characterizes the composition of the linear transformations characterized by the factor matrices. That is, $\mathbf{ABx}$ first transforms $\mathbf{x}$ with $\mathbf{B}$ and then transforms the result with $\mathbf{A}$. It follows that in order to invert this composition of transformations, one must first pass the vector through $\mathbf{B}^{-1}$ and then through $\mathbf{A}^{-1}$.

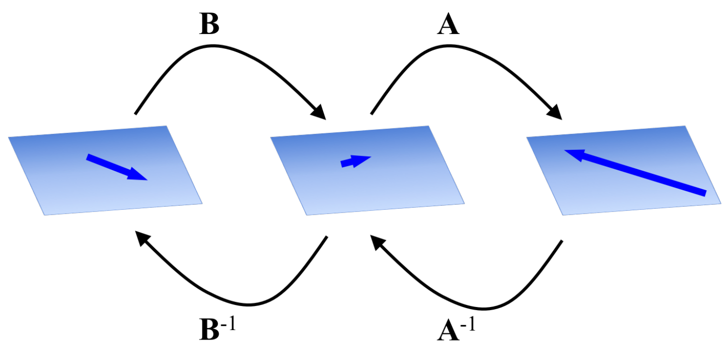

Figure 4: Demonstrating schematically why $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. The matrix $\mathbf{AB}$ first applies the mapping by $\mathbf{B}$ and then applies the mapping by $\mathbf{A}$. The inverse of this function would thus entails applying the inverse of $\mathbf{A}$ followed by the inverse of $\mathbf{B}$.

Definition 1 *A square $n \times n$ matrix $\mathbf{A}$ is said to be* ***invertible*** *if there exists an $n \times n$ matrix $\mathbf{C}$ such that*

$$\mathbf{CA} = \mathbf{I} \text{ and } \mathbf{AC} = \mathbf{I}$$

If $\mathbf{C}$ exists, then it is called the ***inverse*** *of $\mathbf{A}$ and is denoted as $\mathbf{A}^{-1}$. The matrix $\mathbf{A}$ is called* ***invertible****.*

Theorem 1 *The null space of an invertible matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ consists of only the zero vector $\mathbf{0}$.*

Proof:

We must prove that

$$\mathbf{Ax} = \mathbf{0}$$

has only the trivial solution $\mathbf{x} := \mathbf{0}$.

$$\begin{aligned} & \mathbf{Ax} = \mathbf{0} \\ \implies & \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{0} \\ \implies & \mathbf{x} = \mathbf{0} \end{aligned}$$

□

Theorem 2 *A an invertible matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ characterizes a one-to-one linear transformation.*

Proof:

For the sake of contradiction assume that there exists two vectors $\mathbf{x}$ and $\mathbf{x}'$ such that $\mathbf{x} \neq \mathbf{x}'$ and that

$$\mathbf{Ax} = \mathbf{b}$$

and

$$\mathbf{Ax}' = \mathbf{b}$$

where $b \neq \mathbf{0}$. Then,

$$\begin{aligned} & \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x}' = \mathbf{0} \\ \implies & \mathbf{A}(\mathbf{x} - \mathbf{x}') = 0 \end{aligned}$$

By Theorem 1, it must hold that

$$\mathbf{x} - \mathbf{x}' = \mathbf{0}$$

which implies that $\mathbf{x} = \mathbf{x}'$. This contradicts our original assumption. Therefore, it must hold that there does not exist two vectors $\mathbf{x}$ and $\mathbf{x}'$ that map to the same vector via the invertible matrix $\mathbf{A}$. Therefore, $\mathbf{A}$ encodes a one-to-one function.

□

Theorem 3 *An invertible matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ characterizes an onto linear transformation.*

Proof:

Let $\mathbf{x}$ and $b \in \mathbb{R}^n$. Then, there exists a vector, $\mathbf{x} \in \mathbb{R}^n$ such that

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

This solution is precisely

$$\mathbf{x} := \mathbf{A}^{-1}\mathbf{b}$$

$$\begin{aligned} & \mathbf{A}(\mathbf{A}^{-1}\mathbf{b}) = \mathbf{b} & \\ \implies & (\mathbf{A}\mathbf{A}^{-1})\mathbf{b} = \mathbf{b} & \quad \text{associative law} \\ \implies & \mathbf{I}\mathbf{b} = \mathbf{b} & \quad \text{definition of inverse matrix} \\ \implies & \mathbf{b} = \mathbf{b} & \end{aligned}$$

□

Theorem 4 *Let $\mathbf{A}$ be an invertible matrix with inverse $\mathbf{A}^{-1}$. Then,*

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

Proof: Since $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$, it follows that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. This satisfies the definition for $\mathbf{A}$ to be the inverse of $\mathbf{A}^{-1}$.

□

Theorem 5 *Given a matrix* $\mathbf{A} \in \mathbb{R}^{n \times n}$,

$$\mathbf{A} \text{ is invertible} \iff \{\mathbf{a}_{*,1}, \dots, \mathbf{a}_{*,n}\} \text{ are linearly independent}$$

Proof:

We first prove the $\implies$ direction: we assume that $\mathbf{A}$ is invertible and show that under this assumption, the only solution to

$$\mathbf{a}_{*,1}x_1 + \dots + \mathbf{a}_{*,n}x_n = \mathbf{0}$$

is $\mathbf{x} := \mathbf{0}$, which is the condition for linear independence.

$$\begin{aligned}\mathbf{a}_{*,1}x_1 + \dots + \mathbf{a}_{*,n}x_n &= \mathbf{0} \\ \implies \mathbf{A}\mathbf{x} &= \mathbf{0} \\ \implies \mathbf{A}^{-1}\mathbf{A}\mathbf{x} &= \mathbf{A}^{-1}\mathbf{0} \\ \implies \mathbf{x} &= \mathbf{0}\end{aligned}$$

We now prove the $\impliedby$ direction: we assume the columns of $\mathbf{A}$ are linearly independent and show that under this assumption there exists a matrix $\mathbf{C}$ such that

$$\mathbf{C}\mathbf{A} = \mathbf{A}\mathbf{C} = \mathbf{I}$$

Since the columns of $\mathbf{A}$ are linearly independent, then the reduced row echelon form of $\mathbf{A}$ has a pivot in every column. This means that there exists a sequence of elementary row matrices $\mathbf{E}_1, \dots, \mathbf{E}_k$ such that when multiplied by $\mathbf{A}$, they produce the identity matrix. That is,

$$(\mathbf{E}_1 \dots \mathbf{E}_k)\mathbf{A} = \mathbf{I}$$

Though not proven formally, it can be seen that elementary row matrices are invertible. That is, you can always "undo" the transformation imposed by an elementary row matrix (e.g. for an elementary row matrix that swaps rows, you can always swap them back). Furhtermore, since the product of invertible matrices is also invertible, $(\mathbf{E}_1 \dots \mathbf{E}_k)$ is invertible. Thus,

$$\begin{aligned}
&(\mathbf{E}_1 \dots \mathbf{E}_k)\mathbf{A} = \mathbf{I} \\
\implies &(\mathbf{E}_1 \dots \mathbf{E}_k)^{-1}(\mathbf{E}_1 \dots \mathbf{E}_k)\mathbf{A} = (\mathbf{E}_1 \dots \mathbf{E}_k)^{-1}\mathbf{I} \\
\implies &\mathbf{A} = (\mathbf{E}_1 \dots \mathbf{E}_k)^{-1}\mathbf{I} \\
\implies &\mathbf{A} = \mathbf{I}(\mathbf{E}_1 \dots \mathbf{E}_k)^{-1} \\
\implies &\mathbf{A}(\mathbf{E}_1 \dots \mathbf{E}_k)^{-1}(\mathbf{E}_1 \dots \mathbf{E}_k) = \mathbf{I}(\mathbf{E}_1 \dots \mathbf{E}_k)^{-1}(\mathbf{E}_1 \dots \mathbf{E}_k)^{-1}(\mathbf{E}_1 \dots \mathbf{E}_k) \\
\implies &\mathbf{A}(\mathbf{E}_1 \dots \mathbf{E}_k) = \mathbf{I}
\end{aligned}$$

Hence, $\mathbf{C} := (\mathbf{E}_1 \dots \mathbf{E}_k)$ is the matrix for which $\mathbf{AC} = \mathbf{CA} = \mathbf{I}$ and is thus $\mathbf{A}$'s inverse.

Theorem 6 *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Let* $\mathbf{x}$ *and* b *be n-length column vectors. Then,*

$$\mathbf{Ax} = \mathbf{b} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

where $\mathbf{A}^{-1}\mathbf{b}$ *is the unique solution of* $\mathbf{Ax} = \mathbf{b}$.

Proof:

First we show that $\mathbf{A}^{-1}\mathbf{b}$ is a solution: If we let $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, then

$$\begin{aligned}
&\mathbf{A}(\mathbf{A}^{-1}\mathbf{b}) = \mathbf{b} && \\
\implies &(\mathbf{AA}^{-1})\mathbf{b} = \mathbf{b} && \text{associative law} \\
\implies &\mathbf{Ib} = \mathbf{b} && \text{definition of inverse matrix} \\
\implies &\mathbf{b} = \mathbf{b} &&
\end{aligned}$$

Now we show that it is a unique solution. Assume for the sake of contradiction that $\mathbf{u}$ is a solution to $\mathbf{Ax} = \mathbf{b}$ and that $\mathbf{u} \neq \mathbf{A}^{-1}\mathbf{b}$. Then if we multiply both sides by $\mathbf{A}^{-1}$, we get

$$\begin{aligned}
&\mathbf{A}^{-1}\mathbf{Au} = \mathbf{A}^{-1}\mathbf{b} \\
\implies &\mathbf{Iu} = \mathbf{A}^{-1}\mathbf{b} \\
\implies &\mathbf{u} = \mathbf{A}^{-1}\mathbf{b}
\end{aligned}$$

This is a contradiction because by our assumption, $\mathbf{u} \neq \mathbf{A}^{-1}\mathbf{b}$. Thus, are assumption is false.

□

Theorem 7

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

Proof: Since $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$, it follows that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. This satisfies the definition for $\mathbf{A}$ to be the inverse of $\mathbf{A}^{-1}$.

□

Theorem 8 *Given two invertible matrices* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^n$*, the inverse of their product* $\mathbf{AB}$ *is given by* $\mathbf{B}^{-1}\mathbf{A}^{-1}$.

Proof: We seek the inverse matrix $\mathbf{X}$ such that

$$(\mathbf{AB})\mathbf{X} = \mathbf{I}$$

Solving for $\mathbf{X}$:

$$\begin{aligned}
&\mathbf{ABX} = \mathbf{I} \\
\implies &\mathbf{A}^{-1}\mathbf{ABX} = \mathbf{A}^{-1}\mathbf{I} \\
\implies &\mathbf{B}^{-1}\mathbf{BX} = \mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{I} \\
\implies &\mathbf{X} = \mathbf{B}^{-1}\mathbf{A}^{-1}
\end{aligned}$$

□