

Review: Matrix calculus

Example 1. Matrix multiplication is not commutative!

- $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 3 & 10 \end{bmatrix}$

Multiplication (on the right) with that “almost identity matrix” is performing the column operation $C_2 + 2C_1 \Rightarrow C_2$ (i.e. 2 times the first column is added to the second column).

- $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 7 & 10 \\ 3 & 4 \end{bmatrix}$

Multiplication (on the left) with the same matrix is performing the row operation $R_1 + 2R_2 \Rightarrow R_1$.

First comment. This indicates a second interpretation of matrix multiplication: instead of taking linear combinations of columns of the first matrix, we can also take linear combinations of rows of the second matrix.

Second comment. The row operations we are doing during Gaussian elimination can be realized by multiplying (on the left) with “almost identity matrices”.

Example 2. $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = [14]$ whereas $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$.

If you know about the dot product, do you see a connection with the first case?

Example 3. Suppose A is $m \times n$ and B is $p \times q$. When does AB make sense? In that case, what are the dimensions of AB ?

AB makes sense if $n = p$. In that case, AB is a $m \times q$ matrix.

Example 4. $\begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

On the RHS we have the **identity matrix**, usually denoted I or I_2 (since it's the 2×2 identity matrix here).

Hence, the two matrices on the left are inverses of each other: $\begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$, $\begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$.

Example 5. The following formula immediately gives us the inverse of a 2×2 matrix (if it exists). It is worth remembering!

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad \text{provided that } ad-bc \neq 0$$

Let's check that! $\frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} ad-bc & 0 \\ 0 & -cb+ad \end{bmatrix} = I_2$

In particular, a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible $\iff ad-bc \neq 0$.

Recall that this is the **determinant**: $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad-bc$.

In particular:

$$\det(A) = 0 \iff A \text{ is not invertible}$$

Similarly, for $n \times n$ matrices A :

A is invertible	(i.e. there is a matrix A^{-1} such that $AA^{-1} = I$)
$\iff \det(A) \neq 0$	
$\iff Ax = b$ has a unique solution	(namely, $x = A^{-1}b$)

Comment. Why is it not common to write $\frac{1}{A}$ instead of A^{-1} ?

The notation $\frac{1}{A}$ easily leads to ambiguities: for instance, should $\frac{B}{A}$ mean BA^{-1} or should it mean $A^{-1}B$?

[Of course, one could try to avoid this by notations like B/A which would more clearly mean BA^{-1} . It's just not common and doesn't have any real advantages.]

Example 6.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -7 & 2 & 3 \\ -16 & 5 & 6 \\ -25 & 8 & 9 \end{bmatrix}$$

Multiplication (on the right) with that "almost identity matrix" is performing the column operation $C_1 - 4C_2 \Rightarrow C_1$ (i.e. -4 times the second column is added to the first column).

$$\begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{bmatrix}$$

Multiplication (on the left) with the same matrix is performing the row operation $R_2 - 4R_1 \Rightarrow R_2$.

Comment (again). The row operations we are doing during Gaussian elimination can all be realized by multiplying (on the left) with "almost identity matrices".

These matrices are called **elementary matrices** (they are obtained by performing a single elementary row operation on an identity matrix).

Elementary matrices are **invertible** because elementary row operations are reversible:

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & \\ & \frac{1}{2} & \\ & & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Example 7. Let us do Gaussian elimination on $A = \begin{bmatrix} 2 & 1 \\ 4 & -6 \end{bmatrix}$ until we have an echelon form:

$$A = \begin{bmatrix} 2 & 1 \\ 4 & -6 \end{bmatrix} \xrightarrow{R_2 - 2R_1 \Rightarrow R_2} \begin{bmatrix} 2 & 1 \\ 0 & -8 \end{bmatrix}$$

As last class, the row operation can be encoded by multiplication with an “almost identity matrix” E :

$$\underbrace{\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}}_E \underbrace{\begin{bmatrix} 2 & 1 \\ 4 & -6 \end{bmatrix}}_A = \underbrace{\begin{bmatrix} 2 & 1 \\ 0 & -8 \end{bmatrix}}_U$$

Since $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$ (no calculation needed!), this means that

$$A = E^{-1}U = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & -8 \end{bmatrix}.$$

We factored A as the product of a lower and an upper triangular matrix!

$A = LU$ is known as the **LU decomposition** of A .
 L is lower triangular, U is upper triangular.

If A is $m \times n$, then L is an invertible lower triangular $m \times m$ matrix, and U is a usual echelon form of A . Every matrix A has a LU decomposition (after possibly swapping some rows of A first).

- The matrix U is just the echelon form of A produced during Gaussian elimination.
- The matrix L can be constructed, entry-by-entry, by simply recording the row operations used during Gaussian elimination. (No extra work needed!)

Example 8. Determine the LU decomposition of $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Solution. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \xrightarrow{R_2 - 3R_1 \Rightarrow R_2} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}$ translates into $\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}$.

Since $\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$ (no calculation needed!), we therefore have $A = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}$.

Example 9. Determine the LU decomposition of $A = \begin{bmatrix} 1 & 1 & 2 & 1 \\ 3 & 2 & 7 & 2 \\ -2 & 6 & -3 & 1 \end{bmatrix}$.

Solution. We perform Gaussian elimination until we arrive an echelon form:

$$\begin{bmatrix} 1 & 1 & 2 & 1 \\ 3 & 2 & 7 & 2 \\ -2 & 6 & -3 & 1 \end{bmatrix} \xrightarrow{\substack{R_2 - 3R_1 \Rightarrow R_2 \\ R_3 + 2R_1 \Rightarrow R_3}} \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & -1 & 1 & -1 \\ 0 & 8 & 1 & 3 \end{bmatrix} \xrightarrow{R_3 + 8R_2 \Rightarrow R_3} \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & 9 & -5 \end{bmatrix}$$

The corresponding LU decomposition of A is

$$A = LU = \begin{bmatrix} 1 & & & \\ 3 & 1 & & \\ -2 & -8 & 1 & \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 & 1 \\ & -1 & 1 & -1 \\ & & 9 & -5 \end{bmatrix}.$$

Note how the entries in the matrix L are precisely the (negative) coefficients in the row operations. No further computation is required to obtain L .

More details. To see where L is coming from observe that the above row operations are equivalent to the matrix form

$$\begin{bmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & 8 & 1 & \end{bmatrix} \begin{bmatrix} 1 & & & \\ -3 & 1 & & \\ 2 & 0 & 1 & \end{bmatrix} A = \begin{bmatrix} 1 & 1 & 2 & 1 \\ & -1 & 1 & -1 \\ & & 9 & -5 \end{bmatrix}.$$

We therefore have

$$\begin{aligned} A &= \begin{bmatrix} 1 & & & \\ -3 & 1 & & \\ 2 & 0 & 1 & \end{bmatrix}^{-1} \begin{bmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & 8 & 1 & \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 2 & 1 \\ & -1 & 1 & -1 \\ & & 9 & -5 \end{bmatrix} \\ &= \begin{bmatrix} 1 & & & \\ 3 & 1 & & \\ -2 & 0 & 1 & \end{bmatrix} \begin{bmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & -8 & 1 & \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 & 1 \\ & -1 & 1 & -1 \\ & & 9 & -5 \end{bmatrix} \\ &= \begin{bmatrix} 1 & & & \\ 3 & 1 & & \\ -2 & -8 & 1 & \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 & 1 \\ & -1 & 1 & -1 \\ & & 9 & -5 \end{bmatrix}. \end{aligned}$$

Make sure that, in the final step, you can see the fact that, because of their special form, the product of the two lower triangular matrices is just “putting together” the entries.

Review. The RREF (row-reduced echelon form) of A is obtained from an echelon form by

- scaling the pivots to 1, and then
- eliminating the entries above the pivots.

A typical RREF has the shape

[* represents an entry that could be anything]

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

Example 10. Let’s compute the RREF of the 3×4 matrix from Example 9.

Solution.

$$\begin{aligned} &\begin{bmatrix} 1 & 1 & 2 & 1 \\ 3 & 2 & 7 & 2 \\ -2 & 6 & -3 & 1 \end{bmatrix} \xrightarrow[\sim]{\substack{R_2 - 3R_1 \Rightarrow R_2 \\ R_3 + 2R_1 \Rightarrow R_3}} \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & -1 & 1 & -1 \\ 0 & 8 & 1 & 3 \end{bmatrix} \xrightarrow[\sim]{R_3 + 8R_2 \Rightarrow R_3} \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & 9 & -5 \end{bmatrix} \\ &\xrightarrow[\sim]{\substack{-R_2 \Rightarrow R_2 \\ \frac{1}{9}R_3 \Rightarrow R_3}} \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -\frac{5}{9} \end{bmatrix} \xrightarrow[\sim]{\substack{R_1 - 2R_3 \Rightarrow R_1 \\ R_2 + R_3 \Rightarrow R_2}} \begin{bmatrix} 1 & 1 & 0 & \frac{19}{9} \\ 0 & 1 & 0 & \frac{4}{9} \\ 0 & 0 & 1 & -\frac{5}{9} \end{bmatrix} \xrightarrow[\sim]{R_1 - R_2 \Rightarrow R_1} \begin{bmatrix} 1 & 0 & 0 & \frac{5}{3} \\ 0 & 1 & 0 & \frac{4}{9} \\ 0 & 0 & 1 & -\frac{5}{9} \end{bmatrix} \end{aligned}$$

Example 11. The RREF of $A = \begin{bmatrix} 2 & 1 \\ 4 & -6 \end{bmatrix}$ from earlier is the 2×2 identity matrix.

Comment. That’s not surprising: A square matrix is invertible if and only if its RREF is the identity matrix. If that isn’t obvious to you, think about how you invert a matrix using Gaussian elimination (reviewed next).

Review. Recall the Gauss–Jordan method of computing A^{-1} . Starting with the augmented matrix $[A \mid I]$, we do Gaussian elimination until we obtain the RREF, which will be of the form $[I \mid A^{-1}]$ so that we can read off A^{-1} .

Why does that work? By our discussion, the steps of Gaussian elimination can be expressed by multiplication (on the left) with a matrix B . Only looking at the first part of the augmented matrix, and since the RREF of an invertible matrix is I , we have $BA = I$, which means that we must have $B = A^{-1}$. The other part of the augmented matrix (which is I initially) gets multiplied with $B = A^{-1}$ as well, so that, in the end, it is $BI = A^{-1}$. That's why we can read off A^{-1} !

For instance. To invert $\begin{bmatrix} 2 & 1 \\ 4 & -6 \end{bmatrix}$ using the Gauss–Jordan method, we would proceed as follows

$$\left[\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 4 & -6 & 0 & 1 \end{array} \right] \xrightarrow{R_2 - 2R_1 \Rightarrow R_2} \left[\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 0 & -8 & -2 & 1 \end{array} \right] \xrightarrow{\begin{array}{l} \frac{1}{2}R_1 \Rightarrow R_1 \\ -\frac{1}{8}R_2 \Rightarrow R_2 \end{array}} \left[\begin{array}{cc|cc} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & \frac{1}{4} & -\frac{1}{8} \end{array} \right] \xrightarrow{R_1 - \frac{1}{2}R_2 \Rightarrow R_1} \left[\begin{array}{cc|cc} 1 & 0 & \frac{3}{8} & \frac{1}{16} \\ 0 & 1 & \frac{1}{4} & -\frac{1}{8} \end{array} \right]$$

and conclude that $\begin{bmatrix} 2 & 1 \\ 4 & -6 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{3}{8} & \frac{1}{16} \\ \frac{1}{4} & -\frac{1}{8} \end{bmatrix}$.

Of course, for 2×2 matrices it is much simpler to use the formula $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Review: Vector spaces, bases, dimension, null spaces

Review.

- Vectors are things that can be **added** and **scaled**.
- Hence, given vectors v_1, \dots, v_n , the most general we can do is form the **linear combination** $\lambda_1 v_1 + \dots + \lambda_n v_n$. The set of all these linear combinations is the **span** of v_1, \dots, v_n , denoted by $\text{span}\{v_1, \dots, v_n\}$.

- Vector **spaces** are spans.

Equivalently. Vector spaces are sets of vectors so that the result of adding and scaling remains within that set.

Homework. Of course, the latter is a very informal statement. Revisit the formal definition, probably consisting of a list of axioms, and observe how that matches with the above (for instance, several of the axioms are concerned with addition and scaling satisfying the “expected” rules).

- Recall that vectors from a vector space V form a **basis** of V if and only if
 - the vectors span V , and
 - the vectors are (linearly) independent.

Equivalently. v_1, \dots, v_n from V form a basis of V if and only if every vector in V can be expressed as a unique linear combination of v_1, \dots, v_n .

Just checking. Make sure that you can define precisely what it means for vectors v_1, \dots, v_n to be independent.

- The **dimension** of a vector space V is the number of vectors in a basis for V .

No matter what basis one chooses for V , it always has the same number of vectors.

Example 12. \mathbb{R}^3 is the vector space of all vectors with 3 real entries.

\mathbb{R} itself refers to the set of real numbers. We will later also discuss \mathbb{C} , the set of complex numbers.

The **standard basis** of \mathbb{R}^3 is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. The dimension of \mathbb{R}^3 is 3.

Review. The **null space** $\text{null}(A)$ of a matrix A consists of those vectors \mathbf{x} such that $A\mathbf{x} = \mathbf{0}$.

Make sure that you see why $\text{null}(A)$ is a vector space. [For instance, if you pick two vectors in $\text{null}(A)$ why is it that the sum of them is in $\text{null}(A)$ again?]

Example 13. What is $\text{null}(A)$ if the matrix A is invertible?

Solution. If A is invertible, then $A\mathbf{x} = \mathbf{0}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$.

Hence, $\text{null}(A) = \{\mathbf{0}\}$ which is the trivial vector space (consisting of only the null vector) and has dimension 0.

Example 14. Compute a basis for $\text{null}(A)$ where $A = \begin{bmatrix} -1 & 0 & 2 \\ 2 & -3 & 2 \\ 1 & 0 & -2 \end{bmatrix}$.

Solution. We perform row operations and obtain

$$\text{null}\left(\begin{bmatrix} -1 & 0 & 2 \\ 2 & -3 & 2 \\ 1 & 0 & -2 \end{bmatrix}\right) \stackrel{\substack{R_2+2R_1 \Rightarrow R_2 \\ R_3+R_1 \Rightarrow R_3}}{=} \text{null}\left(\begin{bmatrix} -1 & 0 & 2 \\ 0 & -3 & 6 \\ 0 & 0 & 0 \end{bmatrix}\right) \stackrel{\substack{-R_1 \Rightarrow R_1 \\ -\frac{1}{3}R_2 \Rightarrow R_2}}{=} \text{null}\left(\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}\right).$$

From the RREF, we can now read off the general solution to $A\mathbf{x} = \mathbf{0}$:

- x_1 and x_2 are pivot variables. [For each we have an equation expressing it in terms of the other variables; for instance, $x_1 - 2x_3 = 0$ tells us that $x_1 = 2x_3$.]
- x_3 is a free variable. [There is no equation forcing a value on x_3 .]
- Hence, without computation, we see that the general solution is $\begin{bmatrix} 2x_3 \\ 2x_3 \\ x_3 \end{bmatrix}$.

In other words, a basis is $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$.

Comment. We are starting with the three equations $-x_1 + 2x_3 = 0$, $2x_1 - 3x_2 + 2x_3 = 0$, $x_1 - 2x_3 = 0$. Performing row operations on the matrix is the same as combining these equations (with the objective to form simpler equations by eliminating variables).

Example 15. Compute a basis for $\text{null}\left(\begin{bmatrix} 2 & 0 & 2 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix}\right)$.

Solution.

$$\text{null}\left(\begin{bmatrix} 2 & 0 & 2 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix}\right) \stackrel{\substack{R_2-R_1 \Rightarrow R_2 \\ R_3-\frac{1}{2}R_1 \Rightarrow R_3}}{=} \text{null}\left(\begin{bmatrix} 2 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right) \stackrel{\frac{1}{2}R_1 \Rightarrow R_1}{=} \text{null}\left(\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right)$$

This time, x_2 and x_3 are free variables. The general solution is $\begin{bmatrix} -x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

Hence, a basis is $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.