Notes for Lecture 1

Mon, 1/12/2026

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. $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 \end{bmatrix}$ 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!

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

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:

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

Armin Straub
straub@southalabama.edu

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

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

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:

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

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 is the row operation $R_2 + 2R_1 \Rightarrow R_2$ which reverses our above operation), 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. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \overset{R_2 - 3R_1 \Rightarrow R_2}{\rightsquigarrow} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}$ 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 at an echelon form:

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

Observe that we can reverse both of these steps using the row operations $\begin{matrix} R_2 + 3R_1 \Rightarrow R_2 \\ R_3 - 2R_1 - 8R_2 \Rightarrow R_3 \end{matrix}$.

Encoding these in L, 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 that no further computation was required to obtain L. (The entries in the matrix L are precisely the (negative) coefficients in the original row operations.)

Armin Straub
straub@southalabama.edu

Comment. By contrast, combining the operations $\begin{matrix} R_2 - 3R_1 \Rightarrow R_2 \\ R_3 + 2R_1 \Rightarrow R_3 \end{matrix}$ and $R_3 + 8R_2 \Rightarrow R_3$ requires computation. That is because we change R_2 in the first step, and then use the changed R_2 in the second step. Indeed, note that

$$\begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & 8 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ -3 & 1 & \\ 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ -3 & 1 & \\ -22 & 8 & 1 \end{bmatrix},$$

so the combined operations are $\begin{matrix} R_2 - 3R_1 \Rightarrow R_2 \\ R_3 - 22R_1 + 8R_2 \Rightarrow R_3 \end{matrix}$ (you can also see that directly from the operations).

On the other hand, there was no such complication when combining the reversed operations:

Combining $R_3 - 8R_2 \Rightarrow R_3$ and $\begin{matrix} R_2 + 3R_1 \Rightarrow R_2 \\ R_3 - 2R_1 \Rightarrow R_3 \end{matrix}$ simply results in $\begin{matrix} R_2 + 3R_1 \Rightarrow R_2 \\ R_3 - 2R_1 - 8R_2 \Rightarrow R_3 \end{matrix}$, as used above.

The difference is that, here, we change R_3 in the first step but then don't use the changed R_3 in the second step. In terms of matrix multiplication, we have

$$\begin{bmatrix} 1 & & \\ 3 & 1 & \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & -8 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 3 & 1 & \\ -2 & -8 & 1 \end{bmatrix},$$

where, because of their special form, the product of the two lower triangular matrices is just "putting together" the entries (unlike in the non-reversed product).

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.

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

$$\underset{\rightsquigarrow}{\begin{matrix} -R_2 \Rightarrow R_2 \\ \frac{1}{9}R_3 \Rightarrow R_3 \end{matrix}} \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -\frac{5}{9} \end{bmatrix} \underset{\rightsquigarrow}{\begin{matrix} R_1 - 2R_3 \Rightarrow R_1 \\ R_2 + R_3 \Rightarrow R_2 \end{matrix}} \begin{bmatrix} 1 & 1 & 0 & \frac{19}{9} \\ 0 & 1 & 0 & \frac{4}{9} \\ 0 & 0 & 1 & -\frac{5}{9} \end{bmatrix} \underset{\rightsquigarrow}{R_1 - R_2 \Rightarrow R_1} \begin{bmatrix} 1 & 0 & 0 & \frac{5}{3} \\ 0 & 1 & 0 & \frac{4}{9} \\ 0 & 0 & 1 & -\frac{5}{9} \end{bmatrix}$$

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:

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

We 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 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, the most general we can do is form the **linear combination** $\lambda_1 \boldsymbol{v}_1 + \ldots + \lambda_n \boldsymbol{v}_n$. The set of all these linear combinations is the **span** of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, denoted by $\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$.
- 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. $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ from V form a basis of V if and only if every vector in V can be expressed as a unique linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$.

  Just checking. Make sure that you can define precisely what it means for vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 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** $\mathrm{null}(A)$ of a matrix A consists of those vectors $\boldsymbol{x}$ such that $A\boldsymbol{x} = \boldsymbol{0}$.

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

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

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

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

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

Solution. We perform row operations and obtain

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

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

- 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 $\mathrm{null}\left(\begin{bmatrix} 2 & 0 & 2 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix}\right)$.

Solution.

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

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 $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

Armin Straub
straub@southalabama.edu

Review: Eigenvalues and eigenvectors

If $A\boldsymbol{x} = \lambda\boldsymbol{x}$ (and $\boldsymbol{x} \neq \boldsymbol{0}$), then $\boldsymbol{x}$ is an **eigenvector** of A with **eigenvalue** λ (just a number).

Note that for the equation $A\boldsymbol{x} = \lambda\boldsymbol{x}$ to make sense, A needs to be a square matrix (i.e. $n \times n$).

Key observation:

$$\begin{aligned} & A\boldsymbol{x} = \lambda\boldsymbol{x} \\ \Longleftrightarrow\ & A\boldsymbol{x} - \lambda\boldsymbol{x} = \boldsymbol{0} \\ \Longleftrightarrow\ & (A - \lambda I)\boldsymbol{x} = \boldsymbol{0} \end{aligned}$$

This homogeneous system has a nontrivial solution $\boldsymbol{x}$ if and only if $\det(A - \lambda I) = 0$.

To find eigenvectors and eigenvalues of A:

(a) First, find the eigenvalues λ by solving $\det(A - \lambda I) = 0$.

$\det(A - \lambda I)$ is a polynomial in λ, called the **characteristic polynomial** of A.

(b) Then, for each eigenvalue λ, find corresponding eigenvectors by solving $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$.

More precisely, we find a basis of eigenvectors for the λ-**eigenspace** $\operatorname{null}(A - \lambda I)$.

Example 16. $A = \begin{bmatrix} 4 & 0 & 2 \\ 2 & 2 & 2 \\ 1 & 0 & 3 \end{bmatrix}$ has one eigenvector that is "easy" to see. Do you see it?

Solution. Note that $A\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Hence, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ is a 2-eigenvector.

Just for contrast. Note that $A\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} \neq \lambda\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Hence, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is not an eigenvector.

Suppose that A is $n \times n$ and has independent eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.
Then A can be **diagonalized** as $A = PDP^{-1}$, where

- the columns of P are the eigenvectors, and
- the diagonal matrix D has the eigenvalues on the diagonal.

Such a diagonalization is possible if and only if A has enough (independent) eigenvectors.

Comment. If you don't quite recall why these choices result in the diagonalization $A = PDP^{-1}$, note that the diagonalization is equivalent to $AP = PD$.

- Put the eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ as columns into a matrix P.

$$A\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i \quad \Longrightarrow \quad A\begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1\boldsymbol{x}_1 & \cdots & \lambda_n\boldsymbol{x}_n \\ | & & | \end{bmatrix}$$
$$= \begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ | & & | \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- In summary: $AP = PD$

Armin Straub
straub@southalabama.edu

Example 17. Let $A = \begin{bmatrix} 4 & 0 & 2 \\ 2 & 2 & 2 \\ 1 & 0 & 3 \end{bmatrix}$.

(a) Find the eigenvalues and bases for the eigenspaces of A.

(b) Diagonalize A. That is, determine matrices P and D such that $A = PDP^{-1}$.

Solution.

(a) By expanding by the second column, we find that the characteristic polynomial $\det(A - \lambda I)$ is

$$\begin{vmatrix} 4-\lambda & 0 & 2 \\ 2 & 2-\lambda & 2 \\ 1 & 0 & 3-\lambda \end{vmatrix} = (2-\lambda)\begin{vmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{vmatrix} = (2-\lambda)[(4-\lambda)(3-\lambda)-2] = (2-\lambda)^2(5-\lambda).$$

Hence, the eigenvalues are $\lambda = 2$ (with multiplicity 2) and $\lambda = 5$.

Comment. At this point, we know that we will find one eigenvector for $\lambda = 5$ (more precisely, the 5-eigenspace definitely has dimension 1). On the other hand, the 2-eigenspace might have dimension 2 or 1. In order for A to be diagonalizable, the 2-eigenspace must have dimension 2. (Why?!)

- The 5-eigenspace is $\text{null}\left(\begin{bmatrix} -1 & 0 & 2 \\ 2 & -3 & 2 \\ 1 & 0 & -2 \end{bmatrix}\right)$. Proceeding as in Example 14, we obtain

$$\text{null}\left(\begin{bmatrix} -1 & 0 & 2 \\ 2 & -3 & 2 \\ 1 & 0 & -2 \end{bmatrix}\right) \overset{\text{RREF}}{=} \text{null}\left(\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}\right) = \text{span}\left\{\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}\right\}.$$

In other words, the 5-eigenspace has basis $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$.

- The 2-eigenspace is $\text{null}\left(\begin{bmatrix} 2 & 0 & 2 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix}\right)$. Proceeding as in Example 15, we obtain

$$\text{null}\left(\begin{bmatrix} 2 & 0 & 2 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix}\right) \overset{\text{RREF}}{=} \text{null}\left(\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right) = \text{span}\left\{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}$$

In other words, the 2-eigenspace has basis $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

Comment. So, indeed, the 2-eigenspace has dimension 2. In particular, A is diagonalizable.

(b) A possible choice is $P = \begin{bmatrix} 2 & 0 & -1 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

Comment. However, many other choices are possible and correct. For instance, the order of the eigenvalues in D doesn't matter (as long as the same order is used for P). Also, for P, the columns can be chosen to be any other set of eigenvectors.

Armin Straub
straub@southalabama.edu

Example 18. (extra practice) Diagonalize, if possible, the matrices

$$A = \begin{bmatrix} 3 & 4 & 1 \\ 0 & 2 & 0 \\ 1 & 4 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Solution. For instance, $A = PDP^{-1}$ with $P = \begin{bmatrix} 1 & -4 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 4 & & \\ & 2 & \\ & & 2 \end{bmatrix}$. B is not diagonalizable.

For instance, $C = PDP^{-1}$ with $P = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & & \\ & 0 & \\ & & 0 \end{bmatrix}$.

Review: Computing determinants using cofactor expansion

Review. Let A be an $n \times n$ matrix. The **determinant** of A, written as $\det(A)$ or $|A|$, is a number with the property that:

$$\begin{aligned} \det(A) \neq 0 &\iff A \text{ is invertible} \\ &\iff A\boldsymbol{x} = \boldsymbol{b} \text{ has a (unique) solution } \boldsymbol{x} \text{ (for all } \boldsymbol{b}) \\ &\iff A\boldsymbol{x} = 0 \text{ is only solved by } \boldsymbol{x} = 0 \end{aligned}$$

Example 19. $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$

Example 20. Compute $\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix}$ by **cofactor expansion**.

Solution. We expand by the first row:

$$\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix} = 1 \cdot \begin{vmatrix} + & & \\ & -1 & 2 \\ & 0 & 1 \end{vmatrix} - 2 \cdot \begin{vmatrix} & - & \\ 3 & & 2 \\ 2 & & 1 \end{vmatrix} + 0 \cdot \begin{vmatrix} & & + \\ 3 & -1 & \\ 2 & 0 & \end{vmatrix}$$

$$\overset{\text{i.e.}}{=} 1 \cdot \begin{vmatrix} -1 & 2 \\ 0 & 1 \end{vmatrix} - 2 \cdot \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} + 0 \cdot \begin{vmatrix} 3 & -1 \\ 2 & 0 \end{vmatrix} = 1 \cdot (-1) - 2 \cdot (-1) + 0 = 1$$

Each term in the cofactor expansion is ± 1 times an entry times a smaller determinant (row and column of entry deleted).

The ± 1 is assigned to each entry according to $\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$.

Solution. We expand by the second column:

$$\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix} = -2 \cdot \begin{vmatrix} & - & \\ 3 & & 2 \\ 2 & & 1 \end{vmatrix} + (-1) \cdot \begin{vmatrix} 1 & & 0 \\ & + & \\ 2 & & 1 \end{vmatrix} - 0 \cdot \begin{vmatrix} 1 & & 0 \\ 3 & & 2 \\ & - & \end{vmatrix}$$

$$= -2 \cdot (-1) + (-1) \cdot 1 - 0 = 1$$

Armin Straub
straub@southalabama.edu

Example 21. Compute $\begin{vmatrix} 1 & 0 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 2 & 0 & 8 & 5 \end{vmatrix}$.

Solution. We can expand by the second column:

$$\begin{vmatrix} 1 & 0 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 2 & 0 & 8 & 5 \end{vmatrix} = -0\begin{vmatrix} 0 & 1 & 5 \\ 0 & 2 & 1 \\ 2 & 8 & 5 \end{vmatrix} + 2\begin{vmatrix} 1 & 3 & 4 \\ 0 & 2 & 1 \\ 2 & 8 & 5 \end{vmatrix} - 0\begin{vmatrix} 1 & 3 & 4 \\ 0 & 1 & 5 \\ 2 & 8 & 5 \end{vmatrix} + 0\begin{vmatrix} 1 & 3 & 4 \\ 0 & 1 & 5 \\ 0 & 2 & 1 \end{vmatrix}$$

[Of course, you don't have to spell out the 3×3 matrices that get multiplied with 0.]

We can compute the remaining 3×3 matrix in any way we prefer. One option is to expand by the first column:

$$2\begin{vmatrix} 1 & 3 & 4 \\ 0 & 2 & 1 \\ 2 & 8 & 5 \end{vmatrix} = 2\left(+1\begin{vmatrix} 2 & 1 \\ 8 & 5 \end{vmatrix} + 2\begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix}\right) = 2(1\cdot 2 + 2\cdot(-5)) = -16$$

Comment. For cofactor expansion, choosing to expand by the second column is the best choice because this column has more zeros than any other column or row.

The determinant of a triangular matrix is the product of the diagonal entries.

Why? Can you explain this (you can use the next example) using cofactor expansion?

Example 22. Compute $\begin{vmatrix} 1 & 0 & 3 & -1 \\ 0 & 3 & 1 & 5 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 5 \end{vmatrix}$.

Solution. Since the matrix is (upper) triangular, $\begin{vmatrix} 1 & 0 & 3 & -1 \\ & 3 & 1 & 5 \\ & & -2 & 1 \\ & & & 5 \end{vmatrix} = 1\cdot 3\cdot(-2)\cdot 5 = -30$.

Review.

- Effect of row (or column) operations on determinant.
- $\det(AB) = \det(A)\det(B)$
- In particular, the LU decomposition provides us with a way to compute determinants: If $A = LU$, then $\det(A) = \det(L)\det(U)$ and the latter determinants are just products of diagonal entries (because both L and U are triangular).

  Comment. Unless a row swap is required, we can compute the LU decomposition of $A = LU$ using only row operations of the form $R_i + cR_j \Rightarrow R_i$ (those don't change the determinant!).

  In that case, the matrix L will have 1's on the diagonal. In particular, $\det(L) = 1$.

  Consequently, in that case, $\det(A) = \det(U)$.

  Practical comment. For larger matrices, cofactor expansion is a terribly inefficient way of computing determinants. Instead, Gaussian elimination (i.e. LU decomposition) is much more efficient.

  On the other hand, cofactor expansion is a good choice when working by hand with small matrices.

Example 23. (review) If $A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$, then its **transpose** is $A^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

Recall that $(AB)^T = B^T A^T$. This reflects the fact that, in the column-centric versus the row-centric interpretation of matrix multiplication, the order of the matrices is reversed.

Comment. When working with complex numbers, the fundamental role is not played by the transpose but by the **conjugate transpose** instead (we'll see that in our discussion of orthogonality): $A^* = \overline{A^T}$.

For instance, if $A = \begin{bmatrix} 1-3i & 5i \\ 2+i & 3 \end{bmatrix}$, then $A^* = \begin{bmatrix} 1+3i & 2-i \\ -5i & 3 \end{bmatrix}$.

Orthogonality

The inner product and distances

Definition 24. The **inner product** (or **dot product**) of $\boldsymbol{v}$, $\boldsymbol{w}$ in $\mathbb{R}^n$:

$$\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}^T \boldsymbol{w} = v_1 w_1 + \ldots + v_n w_n.$$

Because we can think of this as a special case of the matrix product, it satisfies the basic rules like associativity and distributivity.

In addition: $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{w} \cdot \boldsymbol{v}$.

Example 25. $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} = 2 - 2 + 12 = 12$

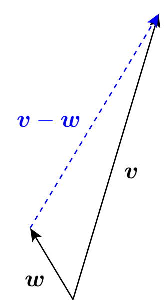
Definition 26.

- The **norm** (or **length**) of a vector $\boldsymbol{v}$ in $\mathbb{R}^n$ is

$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{v_1^2 + \ldots + v_n^2}.$$

- The **distance** between points $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^n$ is

$$\operatorname{dist}(\boldsymbol{v}, \boldsymbol{w}) = \|\boldsymbol{v} - \boldsymbol{w}\|.$$



Example 27. For instance, in $\mathbb{R}^2$, $\operatorname{dist}\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) = \left\|\begin{bmatrix} x_1 - x_2 \\ y_1 - y_2 \end{bmatrix}\right\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

Example 28. Write $\|\boldsymbol{v} - \boldsymbol{w}\|^2$ as a dot product, and multiply it out.

Solution. $\|\boldsymbol{v} - \boldsymbol{w}\|^2 = (\boldsymbol{v} - \boldsymbol{w}) \cdot (\boldsymbol{v} - \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{v} - \boldsymbol{v} \cdot \boldsymbol{w} - \boldsymbol{w} \cdot \boldsymbol{v} + \boldsymbol{w} \cdot \boldsymbol{w} = \|\boldsymbol{v}\|^2 - 2\boldsymbol{v} \cdot \boldsymbol{w} + \|\boldsymbol{w}\|^2$

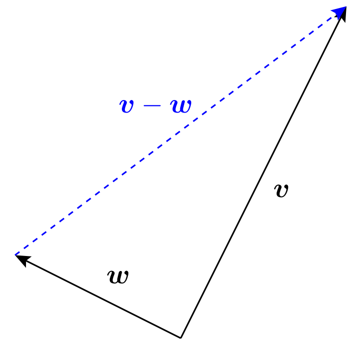
Comment. This is a vector version of $(x-y)^2 = x^2 - 2xy + y^2$.

The reason we were careful and first wrote $-\boldsymbol{v} \cdot \boldsymbol{w} - \boldsymbol{w} \cdot \boldsymbol{v}$ before simplifying it to $-2\boldsymbol{v} \cdot \boldsymbol{w}$ is that we should not take rules such as $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{w} \cdot \boldsymbol{v}$ for granted. For instance, for the cross product $\boldsymbol{v} \times \boldsymbol{w}$, that you may have seen in Calculus, we have $\boldsymbol{v} \times \boldsymbol{w} \neq \boldsymbol{w} \times \boldsymbol{v}$ (instead, $\boldsymbol{v} \times \boldsymbol{w} = -\boldsymbol{w} \times \boldsymbol{v}$).

Orthogonal vectors

Definition 29. $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^n$ are **orthogonal** if

$$\boldsymbol{v}\cdot\boldsymbol{w}=0.$$



Why? How is this related to our understanding of right angles?

Pythagoras!

$\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal

$\iff \|\boldsymbol{v}\|^2+\|\boldsymbol{w}\|^2=\underbrace{\|\boldsymbol{v}-\boldsymbol{w}\|^2}_{\substack{=\|\boldsymbol{v}\|^2-2\boldsymbol{v}\cdot\boldsymbol{w}+\|\boldsymbol{w}\|^2\\ \text{(by previous example)}}}$

$\iff -2\boldsymbol{v}\cdot\boldsymbol{w}=0$

$\iff \boldsymbol{v}\cdot\boldsymbol{w}=0$

Definition 30. We say that two subspaces V and W of $\mathbb{R}^n$ are **orthogonal** if and only if every vector in V is orthogonal to every vector in W.

The **orthogonal complement** of V is the space $V^\perp$ of all vectors that are orthogonal to V.

Exercise. Show that the orthogonal complement is indeed a vector space. Alternatively, this follows from our discussion in the next example which leads to Theorem 32. Namely, every space V can be written as $V=\operatorname{col}(A)$ for a suitable matrix A (for instance, we can choose the columns of A to be basis vectors of V). It then follows that $V^\perp=\operatorname{null}(A^T)$ (which is clearly a space).

Example 31. Determine a basis for the orthogonal complement of $V=\operatorname{span}\left\{\begin{bmatrix}1\\2\\1\end{bmatrix},\begin{bmatrix}3\\1\\2\end{bmatrix}\right\}$.

Solution. The orthogonal complement $V^\perp$ consists of all vectors $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$ that are orthogonal to $\begin{bmatrix}1\\2\\1\end{bmatrix}$ and $\begin{bmatrix}3\\1\\2\end{bmatrix}$.

Using the dot product, this means we must have $\begin{bmatrix}1\\2\\1\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=0$ as well as $\begin{bmatrix}3\\1\\2\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=0$.

Note that this is equivalent to the equations $1x_1+2x_2+1x_3=0$ and $3x_1+1x_2+2x_3=0$.

In matrix-vector form, these two equations combine to $\begin{bmatrix}1&2&1\\3&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$.

This is the same as saying that $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$ has to be in $\operatorname{null}\left(\begin{bmatrix}1&2&1\\3&1&2\end{bmatrix}\right)$. This means that $V^\perp=\operatorname{null}\left(\begin{bmatrix}1&2&1\\3&1&2\end{bmatrix}\right)$.

[Note that we have done no computations up to this point! Instead, we have derived Theorem 32 below.]

We compute (fill in the work!) that $V^\perp=\operatorname{null}\left(\begin{bmatrix}1&2&1\\3&1&2\end{bmatrix}\right)\overset{\text{RREF}}{=}\operatorname{null}\left(\begin{bmatrix}1&0&3/5\\0&1&1/5\end{bmatrix}\right)=\operatorname{span}\left\{\begin{bmatrix}-3/5\\-1/5\\1\end{bmatrix}\right\}$.

Check. $\begin{bmatrix}-3/5\\-1/5\\1\end{bmatrix}$ is indeed orthogonal to both $\begin{bmatrix}1\\2\\1\end{bmatrix}$ and $\begin{bmatrix}3\\1\\2\end{bmatrix}$.

Note. If $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$ is orthogonal to both basis vectors $\begin{bmatrix}1\\2\\1\end{bmatrix}$ and $\begin{bmatrix}3\\1\\2\end{bmatrix}$, then it is orthogonal to every vector in V.

Indeed, vectors in V are of the form $\boldsymbol{v}=a\begin{bmatrix}1\\2\\1\end{bmatrix}+b\begin{bmatrix}3\\1\\2\end{bmatrix}$ and we have $\boldsymbol{v}\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=a\underbrace{\begin{bmatrix}1\\2\\1\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}}_{=0}+b\underbrace{\begin{bmatrix}3\\1\\2\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}}_{=0}=0$.

Just to make sure. Why is it geometrically clear that the orthogonal complement of V is 1-dimensional?

Armin Straub
straub@southalabama.edu

The following theorem follows by the same reasoning that we used in the previous example.

In that example, we started with $V = \mathrm{col}\left(\begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 1 & 2 \end{bmatrix}\right)$ and found that $V^{\perp} = \mathrm{null}\left(\begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \end{bmatrix}\right)$.

Theorem 32. If $V = \mathrm{col}(A)$, then $V^{\perp} = \mathrm{null}(A^T)$.
In particular, if V is a subspace of $\mathbb{R}^n$ with $\dim(V) = r$, then $\dim(V^{\perp}) = n - r$.

For short. $\mathrm{col}(A)^{\perp} = \mathrm{null}(A^T)$

Note that the second part can be written as $\dim(V) + \dim(V^{\perp}) = n$.

To see that this is true, suppose we choose the columns of A to be a basis of V. If V is a subspace of $\mathbb{R}^n$ with $\dim(V) = r$, then A is a $r \times n$ matrix with r pivot columns. Correspondingly, A^T is a $n \times r$ matrix with r pivot rows. Since $n \geqslant r$ there are $n - r$ free variables when computing a basis for $\mathrm{null}(A^T)$. Hence, $\dim(V^{\perp}) = n - r$.

Example 33. Suppose that V is spanned by 3 linearly independent vectors in $\mathbb{R}^5$. Determine the dimension of V and its orthogonal complement $V^{\perp}$.

Solution. This means that $\dim V = 3$. By Theorem 32, we have $\dim V^{\perp} = 5 - 3 = 2$.

Example 34. Determine a basis for the orthogonal complement of (the span of) $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

Solution. Here, $V = \mathrm{span}\left\{\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}\right\}$ and we are looking for the orthogonal complement $V^{\perp}$.

Since $V = \mathrm{col}\left(\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}\right)$, it follows from Theorem 32 that $V^{\perp} = \mathrm{null}(\begin{bmatrix} 1 & 2 & 1 \end{bmatrix})$.

Computing a basis for $\mathrm{null}(\begin{bmatrix} 1 & 2 & 1 \end{bmatrix})$ is easy since $\begin{bmatrix} 1 & 2 & 1 \end{bmatrix}$ is already in RREF.

Note that the general solution to $\begin{bmatrix} 1 & 2 & 1 \end{bmatrix}\boldsymbol{x} = 0$ is $\begin{bmatrix} -2s - t \\ s \\ t \end{bmatrix} = s\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

A basis for $V^{\perp} = \mathrm{null}(\begin{bmatrix} 1 & 2 & 1 \end{bmatrix})$ therefore is $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

Check. We easily check (do it!) that both of these are indeed orthogonal to the original vector $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

The fundamental theorem

Review. The four **fundamental subspaces** associated with a matrix A are

$$\mathrm{col}(A), \quad \mathrm{row}(A), \quad \mathrm{null}(A), \quad \mathrm{null}(A^T).$$

Note that $\mathrm{row}(A) = \mathrm{col}(A^T)$. (In particular, we usually write vectors in $\mathrm{row}(A)$ as column vectors.)

Comment. $\mathrm{null}(A^T)$ is called the **left null space** of A.

Why that name? Recall that, by definition $\boldsymbol{x}$ is in $\mathrm{null}(A) \iff A\boldsymbol{x} = \boldsymbol{0}$.

Likewise, $\boldsymbol{x}$ is in $\mathrm{null}(A^T) \iff A^T\boldsymbol{x} = \boldsymbol{0} \iff \boldsymbol{x}^T A = \boldsymbol{0}$.

[Recall that $(AB)^T = B^T A^T$. In particular, $(A^T\boldsymbol{x})^T = \boldsymbol{x}^T A$, which is what we used in the last equivalence.]

Review. The **rank** of a matrix is the number of pivots in its RREF.

Equivalently, as showcased in the next result, the rank is the dimension of either the column or the row space.

Theorem 35. **(Fundamental Theorem of Linear Algebra, Part I)**
Let A be an $m \times n$ matrix of **rank** r.

- $\dim \mathrm{col}(A) = r$ (subspace of $\mathbb{R}^m$)
- $\dim \mathrm{row}(A) = r$ (subspace of $\mathbb{R}^n$) $\qquad \mathrm{row}(A) = \mathrm{col}(A^T)$
- $\dim \mathrm{null}(A) = n - r$ (subspace of $\mathbb{R}^n$)
- $\dim \mathrm{null}(A^T) = m - r$ (subspace of $\mathbb{R}^m$)

Example 36. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}$. Determine bases for all four fundamental subspaces.

Solution. Make sure that, for such a simple matrix, you can see all of these that at a glance!

$\mathrm{col}(A) = \mathrm{span}\left\{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right\}$, $\mathrm{row}(A) = \mathrm{span}\left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right\}$, $\mathrm{null}(A) = \mathrm{span}\left\{\begin{bmatrix} -2 \\ 1 \end{bmatrix}\right\}$, $\mathrm{null}(A^T) = \mathrm{span}\left\{\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}\right\}$

Important observation. The basis vectors for $\mathrm{row}(A)$ and $\mathrm{null}(A)$ are orthogonal! $\begin{bmatrix} -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 0$

The same is true for the basis vectors for $\mathrm{col}(A)$ and $\mathrm{null}(A^T)$: $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} = 0$ and $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} = 0$

Always. Vectors in $\mathrm{null}(A)$ are orthogonal to vectors in $\mathrm{row}(A)$. In short, $\mathrm{null}(A)$ is orthogonal to $\mathrm{row}(A)$.

Why? Suppose that $\boldsymbol{x}$ is in $\mathrm{null}(A)$. That is, $A\boldsymbol{x} = \boldsymbol{0}$. But think about what $A\boldsymbol{x} = \boldsymbol{0}$ means (row-product rule). It means that the inner product of every row with $\boldsymbol{x}$ is zero. Which implies that $\boldsymbol{x}$ is orthogonal to the row space.

Theorem 37. **(Fundamental Theorem of Linear Algebra, Part II)**

- $\mathrm{null}(A)$ is orthogonal to $\mathrm{row}(A)$. (both subspaces of $\mathbb{R}^n$)

Note that $\dim \mathrm{null}(A) + \dim \mathrm{row}(A) = n$. Hence, the two spaces are orthogonal complements.

- $\mathrm{null}(A^T)$ is orthogonal to $\mathrm{col}(A)$.

Again, the two spaces are orthogonal complements. (This is just the first part with A replaced by A^T.)

Example 38. Let $A = \begin{bmatrix} 1 & 2 & 1 & 4 \\ 2 & 4 & 0 & 2 \\ 3 & 6 & 0 & 3 \end{bmatrix}$. Check that $\operatorname{null}(A)$ and $\operatorname{row}(A)$ are orthogonal complements.

Solution.

$$\begin{bmatrix} 1 & 2 & 1 & 4 \\ 2 & 4 & 0 & 2 \\ 3 & 6 & 0 & 3 \end{bmatrix} \overset{\substack{R_2 - 2R_1 \Rightarrow R_2 \\ R_3 - 3R_1 \Rightarrow R_3}}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & 1 & 4 \\ 0 & 0 & -2 & -6 \\ 0 & 0 & -3 & -9 \end{bmatrix} \overset{R_3 - \frac{3}{2}R_2 \Rightarrow R_3}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & 1 & 4 \\ 0 & 0 & -2 & -6 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\overset{-\frac{1}{2}R_2 \Rightarrow R_2}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & 1 & 4 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \overset{R_1 - R_2 \Rightarrow R_1}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence, $\operatorname{null}(A) = \operatorname{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -3 \\ 1 \end{bmatrix} \right\}$, $\operatorname{row}(A) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 3 \end{bmatrix} \right\}$.

$\operatorname{null}(A)$ and $\operatorname{row}(A)$ are indeed orthogonal, as certified by:

$$\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} = 0, \quad \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ 3 \end{bmatrix} = 0, \quad \begin{bmatrix} -1 \\ 0 \\ -3 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} = 0, \quad \begin{bmatrix} -1 \\ 0 \\ -3 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ 3 \end{bmatrix} = 0.$$

In fact, $\operatorname{null}(A)$ and $\operatorname{row}(A)$ are orthogonal complements because the dimensions add up to $2 + 2 = 4$.

In particular, $\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 3 \end{bmatrix}$ form a basis of all of $\mathbb{R}^4$.

Example 39. (extra) Determine bases for all four fundamental subspaces of

$$A = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 2 & 4 & 0 & 1 \\ 3 & 6 & 0 & 1 \end{bmatrix}.$$

Verify all parts of the Fundamental Theorem, especially that $\operatorname{null}(A)$ and $\operatorname{row}(A)$ (as well as $\operatorname{null}(A^T)$ and $\operatorname{col}(A)$) are orthogonal complements.

Partial solution. One can almost see that $\operatorname{rank}(A) = 3$. Hence, the dimensions of the fundamental subspaces are ...

Consistency of a system of equations

Example 40. (warmup) $\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$

Note that this means that the system of equations $\begin{array}{rcrcl} x_1 & + & 2x_2 & = & 1 \\ 3x_1 & + & x_2 & = & 1 \\ & & 5x_2 & = & 1 \end{array}$ can also be written as $\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

[This was the motivation for introducing matrix-vector multiplication.]

In the same way, any system can be written as $A\boldsymbol{x} = \boldsymbol{b}$, where A is a matrix and $\boldsymbol{b}$ a vector.
In particular, this makes it obvious that:

$$A\boldsymbol{x} = \boldsymbol{b} \text{ is consistent} \quad \Longleftrightarrow \quad \boldsymbol{b} \text{ is in } \operatorname{col}(A)$$

Recall that, by the FTLA, $\operatorname{col}(A)$ and $\operatorname{null}(A^T)$ are orthogonal complements.

Theorem 41. $A\boldsymbol{x} = \boldsymbol{b}$ is consistent $\quad \Longleftrightarrow \quad$ $\boldsymbol{b}$ is orthogonal to $\operatorname{null}(A^T)$

Proof. $A\boldsymbol{x} = \boldsymbol{b}$ is consistent $\Longleftrightarrow$ $\boldsymbol{b}$ is in $\operatorname{col}(A)$ $\overset{\text{FTLA}}{\Longleftrightarrow}$ $\boldsymbol{b}$ is orthogonal to $\operatorname{null}(A^T)$

Note. $\boldsymbol{b}$ is orthogonal to $\operatorname{null}(A^T)$ means that $\boldsymbol{y}^T\boldsymbol{b} = 0$ whenever $\boldsymbol{y}^T A = \boldsymbol{0}$. Why?!

Example 42. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix}$. For which $\boldsymbol{b}$ does $A\boldsymbol{x} = \boldsymbol{b}$ have a solution?

Solution. (old)

$$\left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 3 & 1 & b_2 \\ 0 & 5 & b_3 \end{array}\right] \overset{R_2 - 3R_1 \Rightarrow R_2}{\rightsquigarrow} \left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 0 & -5 & -3b_1 + b_2 \\ 0 & 5 & b_3 \end{array}\right] \overset{R_3 + R_2 \Rightarrow R_3}{\rightsquigarrow} \left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 0 & -5 & -3b_1 + b_2 \\ 0 & 0 & -3b_1 + b_2 + b_3 \end{array}\right]$$

So, $A\boldsymbol{x} = \boldsymbol{b}$ is consistent if and only if $-3b_1 + b_2 + b_3 = 0$.

Solution. (new) We determine a basis for $\operatorname{null}(A^T)$:

$$\begin{bmatrix} 1 & 3 & 0 \\ 2 & 1 & 5 \end{bmatrix} \overset{R_2 - 2R_1 \Rightarrow R_2}{\rightsquigarrow} \begin{bmatrix} 1 & 3 & 0 \\ 0 & -5 & 5 \end{bmatrix} \overset{-\frac{1}{5}R_2 \Rightarrow R_2}{\rightsquigarrow} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & -1 \end{bmatrix} \overset{R_1 - 3R_2 \Rightarrow R_1}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix}$$

We read off from the RREF that $\operatorname{null}(A^T)$ has basis $\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$.

$\boldsymbol{b}$ has to be orthogonal to $\operatorname{null}(A^T)$. That is, $\boldsymbol{b} \cdot \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = 0$. As above!

Armin Straub
straub@southalabama.edu

Comment. Below is how we can use Sage to (try and) solve $A\boldsymbol{x} = \boldsymbol{b}$ for $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

```
>>> A = matrix([[1,2],[3,1],[0,5]])
>>> A.solve_right(vector([1,1,2]))
```

$$\left(\frac{1}{5}, \frac{2}{5}\right)$$

```
>>> A.solve_right(vector([1,1,1]))
```

```
ValueError: matrix equation has no solutions

During handling of the above exception, another exception occurred:

ValueError: matrix equation has no solutions
```

Least squares

Example 43. Not all linear systems have solutions.

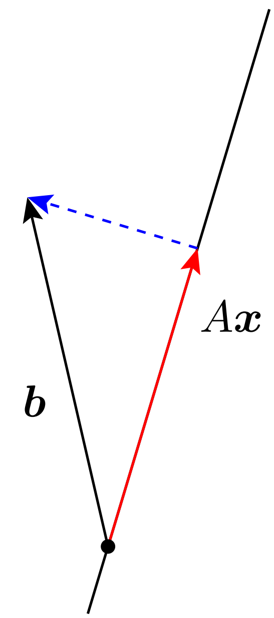
In fact, for many applications, data needs to be fitted and there is no hope for a perfect match.

For instance, $A\boldsymbol{x} = \boldsymbol{b}$ with

$$\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

has no solution:

- $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is not in $\operatorname{col}(A)$ since $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \neq 0$ (see previous example).
- Instead of giving up, we want the $\boldsymbol{x}$ which makes $A\boldsymbol{x}$ and $\boldsymbol{b}$ as close as possible.
- Such $\boldsymbol{x}$ is characterized by the error $A\boldsymbol{x} - \boldsymbol{b}$ being **orthogonal** to $\operatorname{col}(A)$ (i.e. all possible $A\boldsymbol{x}$).



Definition 44. $\hat{\boldsymbol{x}}$ is a **least squares solution** of the system $A\boldsymbol{x} = \boldsymbol{b}$ if $\hat{\boldsymbol{x}}$ is such that $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible (i.e. minimal norm).

- If $A\boldsymbol{x} = \boldsymbol{b}$ is consistent, then $\hat{\boldsymbol{x}}$ is just an ordinary solution. (in that case, $A\hat{\boldsymbol{x}} - \boldsymbol{b} = \boldsymbol{0}$)
- Interesting case: $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent. (in particular, if the system is overdetermined)

The normal equations

The following result provides a straightforward recipe (thanks to the FTLA) to find least squares solutions for all systems $A\boldsymbol{x} = \boldsymbol{b}$.

Theorem 45. $\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$
$\iff A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ (the **normal equations**)

Proof.

$\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$

$\iff A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible

$\iff A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is orthogonal to $\operatorname{col}(A)$

$\overset{\text{FTLA}}{\iff} A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is in $\operatorname{null}(A^T)$

$\iff A^T(A\hat{\boldsymbol{x}} - \boldsymbol{b}) = \boldsymbol{0}$

$\iff A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ □

Armin Straub
straub@southalabama.edu

Example 46. Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

Solution. First, $A^T A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ and $A^T \boldsymbol{b} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

Hence, the normal equations $A^T A \hat{\boldsymbol{x}} = A^T \boldsymbol{b}$ take the form $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \hat{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.
Solving, we immediately find $\hat{\boldsymbol{x}} = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix}$.

Check. Since $A\hat{\boldsymbol{x}} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, the error is $A\hat{\boldsymbol{x}} - \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$. Recall that the error must be orthogonal to $\operatorname{col}(A)$!

This error is indeed orthogonal to $\operatorname{col}(A)$ because $\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = 0$ and $\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0$.

Comment. Why are the normal equations so particularly simple (compare with example below for the typical case) here? Note how each entry of the product $A^T A$ is computed as the dot product of two columns of A (matrix products of a row of A^T times a column of A). That $A^T A$ is a diagonal matrix reflects the fact that the two columns of A are orthogonal to each other.

Example 47. Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Solution. First, $A^T A = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 10 & 5 \\ 5 & 30 \end{bmatrix}$ and $A^T \boldsymbol{b} = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 1 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$.

Hence, the normal equations $A^T A \hat{\boldsymbol{x}} = A^T \boldsymbol{b}$ take the form $\begin{bmatrix} 10 & 5 \\ 5 & 30 \end{bmatrix} \hat{\boldsymbol{x}} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$.

Since $\begin{bmatrix} 10 & 5 \\ 5 & 30 \end{bmatrix}^{-1} = \frac{1}{275} \begin{bmatrix} 30 & -5 \\ -5 & 10 \end{bmatrix} = \frac{1}{55} \begin{bmatrix} 6 & -1 \\ -1 & 2 \end{bmatrix}$, we find $\hat{\boldsymbol{x}} = \frac{1}{55} \begin{bmatrix} 6 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 8 \end{bmatrix} = \frac{1}{55} \begin{bmatrix} 16 \\ 12 \end{bmatrix}$.

Check. Since $A\hat{\boldsymbol{x}} = \frac{1}{55} \begin{bmatrix} 40 \\ 60 \\ 60 \end{bmatrix}$, the error $A\hat{\boldsymbol{x}} - \boldsymbol{b} = \frac{1}{55} \begin{bmatrix} -15 \\ 5 \\ 5 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$ must be orthogonal to $\operatorname{col}(A)$.

The error is indeed orthogonal to $\operatorname{col}(A)$ because $\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \cdot \frac{1}{11} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = 0$ and $\begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} \cdot \frac{1}{11} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = 0$.

Any serious linear algebra problems are done by a machine. Let us see how to use the open-source computer algebra system **Sage** to do basic computations for us.

Sage is freely available at sagemath.org. Instead of installing it locally (it's huge!) we can conveniently use it in the cloud at cocalc.com from any browser. For short computations, like the one below, you can also just use the input field on our course website.

Sage is built as a **Python** library, so any Python code is valid. Here, we will just use it as a fancy calculator.

Let's revisit Example 38 and let Sage do the work for us:

```
>>> A = matrix([[1,2,1,4],[2,4,0,2],[3,6,0,3]])
```

```
>>> A.rref()
```

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Similarly, if we wanted to compute a basis for $\mathrm{null}(A^T)$, we can simply do:

```
>>> A.transpose().rref()
```

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{3}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Here are some other standard things we might be interested in (compare with Example 17):

```
>>> A = matrix([[4,0,2],[2,2,2],[1,0,3]])
```

```
>>> A.eigenvalues()
```

$$[5, 2, 2]$$

```
>>> A.eigenvectors_right()
```

$$\left[\left(5, \left[\left(1,\ 1,\ \frac{1}{2}\right)\right], 1\right), (2, [(1,\ 0,\ -1), (0,\ 1,\ 0)], 2)\right]$$

```
>>> A.eigenmatrix_right()
```

$$\left(\begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ \frac{1}{2} & -1 & 0 \end{bmatrix}\right)$$

```
>>> A.rank()
```

3

```
>>> A.determinant()
```

$$20$$

```
>>> A.inverse()
```

$$\begin{pmatrix} \frac{3}{10} & 0 & -\frac{1}{5} \\ -\frac{1}{5} & \frac{1}{2} & -\frac{1}{5} \\ -\frac{1}{10} & 0 & \frac{2}{5} \end{pmatrix}$$

Application: least squares lines

Given data points (x_i, y_i), we wish to find optimal parameters a, b such that $y_i \approx a + b x_i$ for all i.

Example 48. Determine the line that "best fits" the data points $(2,1), (5,2), (7,3), (8,3)$.

Comment. Can you see that there is no line fitting the data perfectly? (Check out the last two points!)

Solution. We need to determine the values a, b for the best-fitting line $y = a + bx$.
If there was a line that fit the data perfectly, then:

$$\begin{array}{ll} a + 2b = 1 & (2,1) \\ a + 5b = 2 & (5,2) \\ a + 7b = 3 & (7,3) \\ a + 8b = 3 & (8,3) \end{array}$$

In matrix form, this is: $\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix} X} \begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\substack{\text{observation} \\ \text{vector } \boldsymbol{y}}}$ (writing the points as (x_i, y_i))

Using our points, these equations become $\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. [This system is inconsistent (as expected).]

We compute a least squares solution.

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}, \qquad X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

Solving the normal equations $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, we find $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$.

Hence, the least squares line is $y = \frac{2}{7} + \frac{5}{14} x$.

The plot above shows our points together with this line. It does look like a very good fit!

Important comment. In what sense is this the line of "best fit"? By computing a least squares solution the way we do, we are minimizing the error $\boldsymbol{y} - X \begin{bmatrix} a \\ b \end{bmatrix}$. The components of that error are $y_i - (a + b x_i)$.

Hence, we see that we are minimizing the **residual sum of squares** $\mathrm{SS}_{\mathrm{res}} = \sum_i [y_i - (a + b x_i)]^2$.

Also see the discussion after the next example (where we swap the role of x and y) as well as the example at the beginning of next class (where we discuss making predictions and why minimizing $\mathrm{SS}_{\mathrm{res}}$ corresponds to minimizing the error of those predictions).

Armin Straub
straub@southalabama.edu