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 $\begin{array}{c} R_2 + 3R_1 \Rightarrow R_2 \\ R_3 - 2R_1 - 8R_2 \Rightarrow R_3 \end{array}$.

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 $\begin{array}{c} R_2 - 3R_1 \Rightarrow R_2 \\ R_3 + 2R_1 \Rightarrow R_3 \end{array}$ 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 $\begin{array}{c} R_2 - 3R_1 \Rightarrow R_2 \\ R_3 - 22R_1 + 8R_2 \Rightarrow R_3 \end{array}$ (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 $\begin{array}{c} R_2 + 3R_1 \Rightarrow R_2 \\ R_3 - 2R_1 \Rightarrow R_3 \end{array}$ simply results in $\begin{array}{c} R_2 + 3R_1 \Rightarrow R_2 \\ R_3 - 2R_1 - 8R_2 \Rightarrow R_3 \end{array}$, 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.

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

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

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.

Armin Straub
straub@southalabama.edu

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** $\operatorname{null}(A)$ of a matrix A consists of those vectors $\boldsymbol{x}$ such that $A\boldsymbol{x} = \mathbf{0}$.

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

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

Solution. If A is invertible, then $A\boldsymbol{x} = \mathbf{0}$ has the unique solution $\boldsymbol{x} = A^{-1}\mathbf{0} = \mathbf{0}$.
Hence, $\operatorname{null}(A) = \{\mathbf{0}\}$ which is the trivial vector space (consisting of only the null vector) and has dimension 0.

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

Solution. We perform row operations and obtain

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

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

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

Solution.

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

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

Review: Eigenvalues and eigenvectors

If $A\boldsymbol{x} = \lambda\boldsymbol{x}$ (and $\boldsymbol{x} \neq \mathbf{0}$), 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:

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

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 $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$.

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

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: $AP = PD$

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

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

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

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

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

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

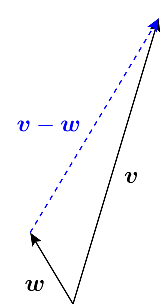
Definition 26.

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

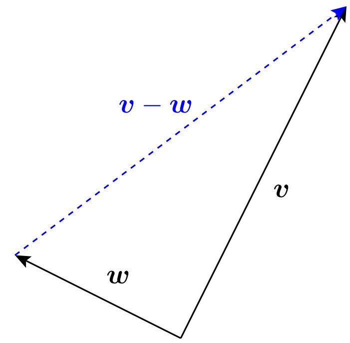
Comment. This is a vector version of $(x - y)^2 = x^2 - 2xy + y^2$.

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, $\boldsymbol{v} \times \boldsymbol{w} = -\boldsymbol{w} \times \boldsymbol{v}$).

Armin Straub
straub@southalabama.edu

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 $V = \operatorname{col}\left(\begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 1 & 2 \end{bmatrix}\right)$ and found that $V^{\perp} = \operatorname{null}\left(\begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \end{bmatrix}\right)$.

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

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

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 $\operatorname{null}(A^T)$. 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, $V = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}\right\}$ and we are looking for the orthogonal complement $V^{\perp}$.

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

Computing a basis for $\operatorname{null}(\begin{bmatrix} 1 & 2 & 1 \end{bmatrix})$ 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 $V^{\perp} = \operatorname{null}(\begin{bmatrix} 1 & 2 & 1 \end{bmatrix})$ 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 $\mathrm{null}(A) \iff A\boldsymbol{x} = \mathbf{0}$.

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

[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, $A\boldsymbol{x} = \mathbf{0}$. But think about what $A\boldsymbol{x} = \mathbf{0}$ 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.)

Armin Straub
straub@southalabama.edu

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:

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

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 $\iff$ $\boldsymbol{b}$ is orthogonal to $\operatorname{null}(A^T)$

Proof. $A\boldsymbol{x} = \boldsymbol{b}$ is consistent $\iff$ $\boldsymbol{b}$ is in $\operatorname{col}(A)$ $\overset{\text{FTLA}}{\iff}$ $\boldsymbol{b}$ 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.

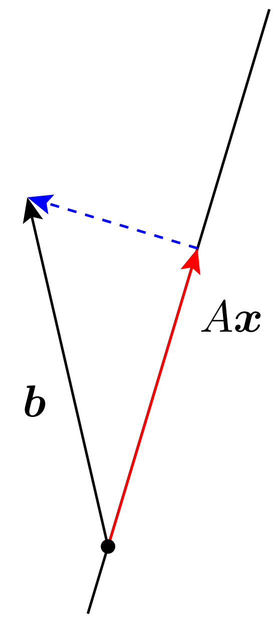
In fact, for many applications, data needs to be fitted and there is no hope for a perfect match.

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 $\mathrm{col}(A)$ 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 $\mathrm{col}(A)$ (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 $\mathrm{col}(A)$

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

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

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

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 $\operatorname{null}(A^T)$, 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()
```

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

Armin Straub
straub@southalabama.edu

Application: least squares lines

Given data points (x_i, y_i), we wish to find optimal parameters a, b such that $y_i \approx a + bx_i$ 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:

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

In matrix form, this is: $\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix} X} \begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } \boldsymbol{y}}$ (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 $y_i - (a + bx_i)$.

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

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).

Example 49. (again) Determine the least squares line for the points $(2,1),(5,2),(7,3),(8,3)$.

Solution. Let's repeat the computation we did in the previous example. This time, we let Sage do the actual work for us:

```
>>> X = matrix([[1,2],[1,5],[1,7],[1,8]]); y = vector([1,2,3,3])
>>> (X.transpose()*X).solve_right(X.transpose()*y)
```

$$\left(\frac{2}{7}, \frac{5}{14}\right)$$

Here are some intermediate steps to help see what's going on (and that it matches our earlier work):

```
>>> X.transpose()*X
```

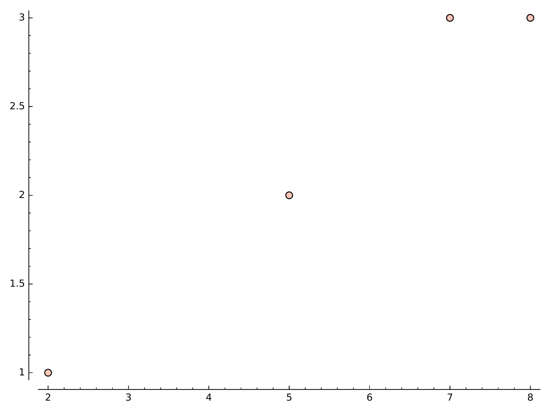
$$\begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix}$$

```
>>> X.transpose()*y
```

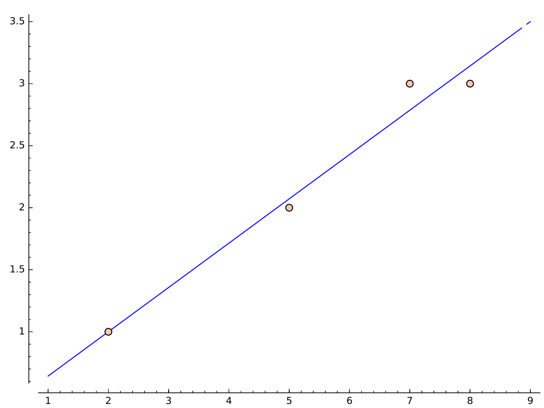
$$(9, 57)$$

Let's plot the least squares line $y = \frac{2}{7} + \frac{5}{14}x$ in Sage to marvel at the good fit!

```
>>> points = [[2,1],[5,2],[7,3],[8,3]]
>>> scatter_plot(points)
```




```
>>> scatter_plot(points) + plot(2/7+5/14*x,1,9)
```



Armin Straub
straub@southalabama.edu

Comment. As mentioned earlier, the least squares line minimizes the (sum of squares of the) vertical offsets:

http://mathworld.wolfram.com/LeastSquaresFitting.html

Comment. We get a (slightly) different "best fit" line if we change the role of x and y! Can you explain that?

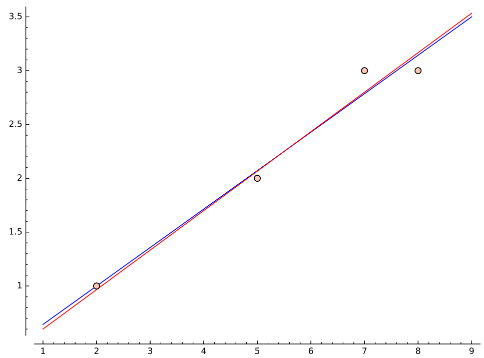
```
>>> X = matrix([[1,1],[1,2],[1,3],[1,3]]); y = vector([2,5,7,8])
```

```
>>> (X.transpose()*X).solve_right(X.transpose()*y)
```

$$\left(-\frac{7}{11}, \frac{30}{11}\right)$$

Note that $x = -\frac{7}{11} + \frac{30}{11}y$ is equivalent to $y = \frac{7}{30} + \frac{11}{30}x$.

```
>>> scatter_plot([[2,1],[5,2],[7,3],[8,3]]) + plot(2/7+5/14*x,1,9) + plot(7/30+11/30*x,1,
    9,color='red')
```



The explanation is that (see pictures at the beginning of this example) we are minimizing vertical offsets in one case and horizontal offsets in the other case.

In linear regression, the relationship between a dependent variable and one or more explanatory variables is modeled. If y is the dependent variable, with x the explanatory variable, then it is natural to minimize the error we make in "predicting y through x" (vertical offsets). See next example.

Example 50. A car rental company wants to predict the annual maintenance cost y (in 100USD/year) of a car using the age x (in years) of that car (as an explanatory variable). Based on the observations $(x, y) = (2, 1), (5, 2), (7, 3), (8, 3)$, predict the cost for a 4.5 year old car (using linear regression).

Solution. Once we compute the regression line $y = a + bx$ (we already did that: $y = \frac{2}{7} + \frac{5}{14}x$), our prediction is $\frac{2}{7} + \frac{5}{14} \cdot 4.5 = \frac{53}{28} \approx 1.89$, that is, 189 USD/year.

Armin Straub
straub@southalabama.edu

Application: multiple linear regression

> *In statistics, **linear regression** is an approach for modeling the relationship between a scalar dependent variable and one or more explanatory variables.*
>
> *The case of one explanatory variable is called simple linear regression.*
>
> *For more than one explanatory variable, the process is called multiple linear regression.*
>
> http://en.wikipedia.org/wiki/Linear_regression

The experimental data might be of the form (x_i, y_i, z_i), where now the dependent variable z_i depends on two explanatory variables x_i, y_i (instead of just x_i).

Example 51. Set up a linear system to find values for the parameters a, b, c such that $z = a + bx + cy$ best fits some given points $(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots$

Solution. The equations $a + bx_i + cy_i = z_i$ translate into the system:

$$\underbrace{\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{design matrix } A} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \underbrace{\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \end{bmatrix}}_{\substack{\text{observation} \\ \text{vector } \boldsymbol{z}}}$$

Of course, this is usually inconsistent. To find the best possible a, b, c we compute a least squares solution by solving $A^T A \begin{bmatrix} a \\ b \\ c \end{bmatrix} = A^T \boldsymbol{z}$.

Application: Fitting data to other curves

We can also fit the experimental data (x_i, y_i) using other curves.

Example 52. Set up a linear system to find values for the parameters a, b, c that result in the quadratic curve $y = a + bx + cx^2$ that best fits some given points $(x_1, y_1), (x_2, y_2), \ldots$

Solution. $y_i \approx a + bx_i + cx_i^2$ with parameters a, b, c.

The equations $y_i = a + bx_i + cx_i^2$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{design matrix } A} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}}_{\substack{\text{observation} \\ \text{vector } \boldsymbol{y}}}$$

Again, we determine values for a, b, c by computing a least squares solution to that system.

That is, we need to solve the system $A^T A \begin{bmatrix} a \\ b \\ c \end{bmatrix} = A^T \boldsymbol{y}$.

Example 53. (homework) Use Sage to find values for a, b, c that result in the quadratic curve $y = a + bx + cx^2$ that best fits the points $(0,1), (1,2), (2,3), (3,-4), (4,-7), (5,-12)$.

Solution. We first input the points:

```
>>> points = [[0,1],[1,2],[2,3],[3,-4],[4,-7],[5,-12]]
```

We set up the system described in the previous example, then determine a least-squares solution.

```
>>> X = matrix([[1,0,0],[1,1,1],[1,2,4],[1,3,9],[1,4,16],[1,5,25]])
```

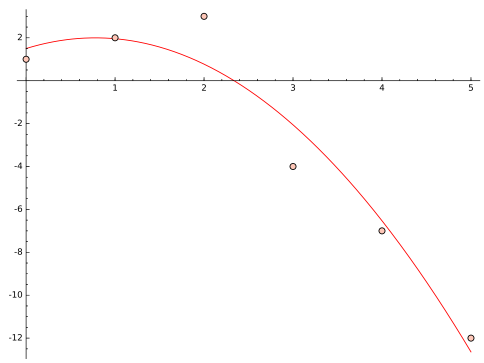
```
>>> y = vector([1,2,3,-4,-7,-12])
```

```
>>> (X.transpose()*X).solve_right(X.transpose()*y)
```

$$\left(\frac{3}{2}, \frac{179}{140}, -\frac{23}{28}\right)$$

Hence, the best fitting quadratic curve is $y = \frac{3}{2} + \frac{179}{140} - \frac{23}{28}x^2$. Here's a plot:

```
>>> scatter_plot(points) + plot(3/2+179/140*x-23/28*x^2,0,5,color='red')
```



Advanced comment. If you are comfortable with Python, you can avoid typing out X and $\boldsymbol{y}$:
[The plot command above now won't work anymore because we are overwriting x with numbers.]

```
>>> X = matrix([[1,x,x^2] for x,y in points])
```

```
>>> y = vector([y for x,y in points])
```

Armin Straub
straub@southalabama.edu

More on orthogonality

Example 54. (review) Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

Solution. First, $A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$ and $A^T\boldsymbol{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$.

Hence, the normal equations $A^TA\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ take the form $\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}\hat{\boldsymbol{x}} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$. Solving, we find $\hat{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Check. The error $A\hat{\boldsymbol{x}} - \boldsymbol{b} = \begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}$ is indeed orthogonal to $\operatorname{col}(A)$. Because $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} = 0$ and $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = 0$.

Orthogonal projections

The **(orthogonal) projection** $\hat{\boldsymbol{b}}$ of a vector $\boldsymbol{b}$ onto a subspace W is the vector in W closest to $\boldsymbol{b}$.

We can compute $\hat{\boldsymbol{b}}$ as follows:

- Write $W = \operatorname{col}(A)$ for some matrix A.
- Then $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ where $\hat{\boldsymbol{x}}$ is a least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$. (i.e. $\hat{\boldsymbol{x}}$ solves $A^TA\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$)

Why? Why is $A\hat{\boldsymbol{x}}$ the projection of $\boldsymbol{b}$ onto $\operatorname{col}(A)$?

Because, if $\hat{\boldsymbol{x}}$ is a least squares solution then $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible (and any element in $\operatorname{col}(A)$ is of the form $A\boldsymbol{x}$ for some $\boldsymbol{x}$).

Note. This is a recipe for computing any orthogonal projection! That's because every subspace W can be written as $\operatorname{col}(A)$ for some choice of the matrix A (take, for instance, A so that its columns are a basis for W).

Assuming A^TA is invertible (which, as discussed in the lemma below, is automatically the case if the columns of A are independent), we have $\hat{\boldsymbol{x}} = (A^TA)^{-1}A^T\boldsymbol{b}$ and hence:

(projection matrix) The projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto $\operatorname{col}(A)$ is (assuming cols of A are independent)

$$\hat{\boldsymbol{b}} = \underbrace{A(A^TA)^{-1}A^T}_{P}\boldsymbol{b}.$$

The matrix $P = A(A^TA)^{-1}A^T$ is the **projection matrix** for projecting onto $\operatorname{col}(A)$.

Example 55.

(a) What is the orthogonal projection of $\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto $W = \operatorname{span}\left\{\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}\right\}$?

(b) What is the matrix P for projecting onto $W = \operatorname{span}\left\{\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}\right\}$?

(c) **(once more)** Using P, what is the orthogonal projection of $\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto W?

(d) Using P, what is the orthogonal projection of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ onto W?

Armin Straub
straub@southalabama.edu

Solution.

(a) In other words, what is the orthogonal projection of $\boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto $\operatorname{col}(A)$ with $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$.

In Example 54, we found that the system $A\boldsymbol{x} = \boldsymbol{b}$ has the least squares solution $\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

The projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto $\operatorname{col}(A)$ thus is $A\hat{\boldsymbol{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$.

Check. The error $\hat{\boldsymbol{b}} - \boldsymbol{b} = \begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}$ needs to be orthogonal to $\operatorname{col}(A)$. Indeed: $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} = 0$ and $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = 0$.

(b) Note that $W = \operatorname{col}(A)$ for $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ and that $A^T A = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$. Thus $(A^T A)^{-1} = \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$.

$P = A(A^T A)^{-1} A^T = \frac{1}{84}\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}\begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 20 & -2 & 4 \\ -2 & 17 & 8 \\ 4 & 8 & 5 \end{bmatrix}$

(c) The orthogonal projection of $\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto W is $P\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 20 & -2 & 4 \\ -2 & 17 & 8 \\ 4 & 8 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 84 \\ 84 \\ 63 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$.

Note. Of course, that agrees with what our computations in the first part. Note that computing P is more work than what we did in in the first part. However, after having computed P once, we can easily project many vectors onto W.

(d) The orthogonal projection of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ onto W is $P\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 20 & -2 & 4 \\ -2 & 17 & 8 \\ 4 & 8 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 20 \\ -2 \\ 4 \end{bmatrix}$.

Check. The error $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{21}\begin{bmatrix} 20 \\ -2 \\ 4 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix}$ is indeed orthogonal to both $\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$.

Example 56. (extra)

(a) What is the orthogonal projection of $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ onto $\operatorname{span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}\right\}$?

(b) What is the orthogonal projection of $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ onto $\operatorname{span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\}$?

Solution. (final answer only) The projections are $\left(\frac{11}{6}, \frac{1}{3}, \frac{7}{6}\right)^T$ and $\left(\frac{3}{2}, 0, \frac{3}{2}\right)^T$.

Armin Straub
straub@southalabama.edu

Example 57. (extra)

(a) What is the matrix P for projecting onto $W = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$?

(b) Using the projection matrix, project $\begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$ onto $W = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$.

Solution.

(a) Choosing $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$, the projection matrix P is $A(A^TA)^{-1}A^T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$
$= \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \frac{1}{8} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 2 & -4 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

Comment. We can choose A in any way such that its columns are a basis for W. The final projection matrix will always be the same.

(b) The projection is $\frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix}$.

Check. The error $\begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$ is indeed orthogonal to W.

Lemma 58. If the columns of a matrix A are independent, then A^TA is invertible.

Proof. Assume A^TA is not invertible, so that $A^TA\boldsymbol{x} = \boldsymbol{0}$ for some $\boldsymbol{x} \neq \boldsymbol{0}$. Multiply both sides with $\boldsymbol{x}^T$ to get

$$\boldsymbol{x}^TA^TA\boldsymbol{x} = (A\boldsymbol{x})^TA\boldsymbol{x} = \|A\boldsymbol{x}\|^2 = 0,$$

which implies that $A\boldsymbol{x} = 0$. Since the columns of A are independent, this shows that $\boldsymbol{x} = \boldsymbol{0}$. A contradiction! □

Example 59. If P is a projection matrix, then what is P^2?

For instance. For P as in Example 57, $P^2 = \frac{1}{4} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}^2 = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} = P$.

Solution. Can you see why it is always true that $P^2 = P$?
[Recall that P projects a vector onto a space W (actually, $W = \operatorname{col}(P)$). Hence P^2 takes a vector $\boldsymbol{b}$, projects it onto W to get $\hat{\boldsymbol{b}}$, and then projects $\hat{\boldsymbol{b}}$ onto W again. But the projection of $\hat{\boldsymbol{b}}$ onto W is just $\hat{\boldsymbol{b}}$ (why?!), so that P^2 always has the exact same effect as P. Therefore, $P^2 = P$.]

Example 60. True or false? If P is the matrix for projecting onto W, then $W = \operatorname{col}(P)$.

Solution. True!

Why? The columns of P are the projections of the standard basis vectors and hence in W. On the other hand, for any vector $\boldsymbol{w}$ in W, we have $P\boldsymbol{w} = \boldsymbol{w}$ so that $\boldsymbol{w}$ is a combination of the columns of P.

[This may take several readings to digest but do read (or ask) until it makes sense!]

In particular. $\operatorname{rank}(P) = \dim W$ (because, for any matrix, $\operatorname{rank}(A) = \dim \operatorname{col}(A)$)

Armin Straub
straub@southalabama.edu

Review. The **projection matrix** for projecting onto $\text{col}(A)$ is $P = A(A^TA)^{-1}A^T$.

Projecting onto 1-dimensional spaces

When we project onto a 1-dimensional space $\text{span}\{\boldsymbol{w}\}$, we usually just say that we are projecting onto $\boldsymbol{w}$.

The (orthogonal) projection of $\boldsymbol{v}$ onto $\boldsymbol{w}$ is $\frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\|\boldsymbol{w}\|^2}\boldsymbol{w}$.

Why? Replace $\boldsymbol{b}$ with $\boldsymbol{v}$ and A with $\boldsymbol{w}$ in our general projection matrix formula to get $\boldsymbol{w}(\boldsymbol{w}^T\boldsymbol{w})^{-1}\boldsymbol{w}^T\boldsymbol{v}$, which equals $\frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\|\boldsymbol{w}\|^2}\boldsymbol{w}$ (note that $\boldsymbol{w}^T\boldsymbol{v} = \boldsymbol{w}\cdot\boldsymbol{v}$ and $\boldsymbol{w}^T\boldsymbol{w} = \|\boldsymbol{w}\|^2$ are scalars).

Comment. If you have taken Calculus 3, you have seen that formula before. Most likely, you were deriving it using angles at that time. Namely, the dot product has the following connection to angles:

$\boldsymbol{v}\cdot\boldsymbol{w} = \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|\cos\theta$ where $\theta \in [0,\pi]$ is the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$

Why? You can derive this by repeating what we did, right after Definition 29 to show that $\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal if and only if $\boldsymbol{v}\cdot\boldsymbol{w} = 0$. Just replace Pythagoras with the law of cosines ($c^2 = a^2 + b^2 - 2ab\cos\theta$ holds in any triangle!).

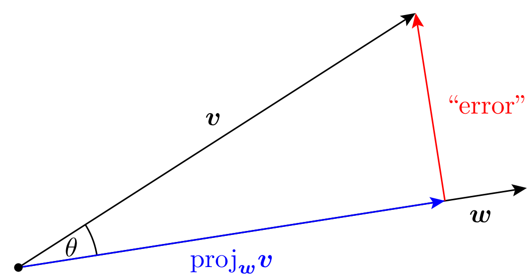
Two obvious cases. Observe that the cases $\theta = 0$ and $\theta = 90°$ are clearly true.

We will not discuss angles much further in this class. Just in case it is helpful, here is the typical argument given in Calculus 3 to determine the projection $\text{proj}_{\boldsymbol{w}}\boldsymbol{v}$ of $\boldsymbol{v}$ onto $\boldsymbol{w}$:

From the sketch, we see that "error" $= \boldsymbol{v} - \text{proj}_{\boldsymbol{w}}\boldsymbol{v}$ and that this error is orthogonal to $\boldsymbol{w}$.

Basic trigonometry tells us that the length of $\text{proj}_{\boldsymbol{w}}\boldsymbol{v}$ is $\|\boldsymbol{v}\|\cos\theta$. Hence:

$$\text{proj}_{\boldsymbol{w}}\boldsymbol{v} = \underbrace{\|\boldsymbol{v}\|\cos\theta}_{\text{length}}\ \underbrace{\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}}_{\text{direction}} = \frac{\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|\cos\theta}{\|\boldsymbol{w}\|}\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} = \left(\frac{\boldsymbol{v}\cdot\boldsymbol{w}}{\|\boldsymbol{w}\|^2}\right)\boldsymbol{w}$$



Orthogonal bases

Review. Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are a **basis** for V.

$\Longleftrightarrow$ $V = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are linearly independent.

$\Longleftrightarrow$ Any vector $\boldsymbol{w}$ in V can be written as $\boldsymbol{w} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n$ in a unique way.

The latter is the practical reason why we care so much about bases!

V could be some abstract vector space (of polynomials or Fourier series), meaning that vectors are abstract objects and not just our usual column vectors. However, as soon as we pick a basis of V, then we can represent every (abstract) vector $\boldsymbol{w}$ by the (usual) column vector $(c_1, c_2, \ldots, c_n)^T$.

This means all of our results can be used, too, when working with these abstract spaces!

Definition 61. A basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of a vector space V is an **orthogonal basis** if the vectors are (pairwise) orthogonal. If, in addition, the basis vectors have length 1, then this is called an **orthonormal basis**.

Armin Straub
straub@southalabama.edu

Example 62. The standard basis $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is an orthonormal basis for $\mathbb{R}^3$.

Example 63. Are the vectors $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ an orthogonal basis for $\mathbb{R}^3$? Is it orthonormal?

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

So, this is an orthogonal basis.

On the other hand, the vectors do not all have length 1, so that this basis is not orthonormal.

Note. Orthogonal vectors are always linearly independent (see next class). Here, this certifies that the three vectors are linearly independent (and hence a basis for $\mathbb{R}^3$).

Normalize the vectors to produce an orthonormal basis.

Solution.

$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ has length $\sqrt{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} = \sqrt{2} \implies$ normalized: $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ has length $\sqrt{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} = \sqrt{2} \implies$ normalized: $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ has length $\sqrt{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}} = 1 \implies$ is already normalized: $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

The resulting orthonormal basis is $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Theorem 64. Suppose that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are nonzero and pairwise orthogonal. Then $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are linearly independent.

Proof. Suppose that $c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n = \boldsymbol{0}$. In order to show that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are independent, we need to show that $c_1 = c_2 = \ldots = c_n = 0$.

Take the dot product of $\boldsymbol{v}_1$ with both sides:

$$\begin{aligned} 0 &= \boldsymbol{v}_1 \cdot (c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n) \\ &= c_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 + c_2 \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 + \ldots + c_n \boldsymbol{v}_1 \cdot \boldsymbol{v}_n \\ &= c_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 = c_1 \|\boldsymbol{v}_1\|^2 \end{aligned}$$

But $\|\boldsymbol{v}_1\| \neq 0$ and hence $c_1 = 0$. Likewise, we find $c_2 = 0$, …, $c_n = 0$. Hence, the vectors are independent. □

Comment. Note that this result is intuitively obvious: if the vectors were linearly dependent, then one of them could be written as a linear combination of the others. However, all these other vectors (and hence any combination of them) are orthogonal to it.

Armin Straub
straub@southalabama.edu

Orthogonal projections if we have an orthogonal basis

Lemma 65. (orthogonal projection if we have an orthogonal basis)
If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthogonal, then the orthogonal projection of $\boldsymbol{w}$ onto $\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is

$$\hat{\boldsymbol{w}} = \underbrace{\frac{\boldsymbol{w}\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\boldsymbol{v}_1}_{\substack{\text{proj of } \boldsymbol{w} \\ \text{onto } \boldsymbol{v}_1}} + \ldots + \underbrace{\frac{\boldsymbol{w}\cdot\boldsymbol{v}_n}{\boldsymbol{v}_n\cdot\boldsymbol{v}_n}\boldsymbol{v}_n}_{\substack{\text{proj of } \boldsymbol{w} \\ \text{onto } \boldsymbol{v}_n}}.$$

Proof. It suffices to show that the error $\boldsymbol{w} - \hat{\boldsymbol{w}}$ is orthogonal to each $\boldsymbol{v}_i$. Indeed:

$$(\boldsymbol{w} - \hat{\boldsymbol{w}})\cdot\boldsymbol{v}_i = \left(\boldsymbol{w} - \frac{\boldsymbol{w}\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\boldsymbol{v}_1 - \ldots - \frac{\boldsymbol{w}\cdot\boldsymbol{v}_n}{\boldsymbol{v}_n\cdot\boldsymbol{v}_n}\boldsymbol{v}_n\right)\cdot\boldsymbol{v}_i = \boldsymbol{w}\cdot\boldsymbol{v}_i - \frac{\boldsymbol{w}\cdot\boldsymbol{v}_i}{\boldsymbol{v}_i\cdot\boldsymbol{v}_i}\boldsymbol{v}_i\cdot\boldsymbol{v}_i = 0.$$

Alternatively, can you deduce the formula (say, in the case of an orthonormal basis) from our earlier formula for the projection matrix? □

Important consequence. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is an orthogonal basis of V, and $\boldsymbol{w}$ is in V, then

$$\boldsymbol{w} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n \qquad \text{with} \quad c_j = \frac{\boldsymbol{w}\cdot\boldsymbol{v}_j}{\boldsymbol{v}_j\cdot\boldsymbol{v}_j}.$$

If the $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are a basis, but not orthogonal, then we have to solve a system of equations to find the c_i. That is a lot more work than simply computing a few dot products.

Note. In other words, $\boldsymbol{w}$ decomposes as the sum of its projections onto each basis vector.

Note. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthonormal, then the denominators are all 1.

Example 66. What is the projection of $\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$ onto $W = \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ with $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $\boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$?

Comment. We know how to do this using least squares. (Do it for practice!)
However, realizing that $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are orthogonal makes things easier.
[Actually, here, it is obvious what the projection is going to be if we realized that W is the x-y-plane.]

Solution. (using orthogonality) Because $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are orthogonal, the projection is

$$\underbrace{\frac{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}\cdot\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}\cdot\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}_{\text{projection onto } \boldsymbol{v}_1} + \underbrace{\frac{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}\cdot\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\cdot\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\text{projection onto } \boldsymbol{v}_2} = \frac{-4}{2}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{10}{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 0 \end{bmatrix}.$$

Important note. Note that, at this point, we can easily extend $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ to an orthogonal basis of $\mathbb{R}^3$:

That is because the error $\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 7 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$ is orthogonal to both of the existing basis vectors.

Therefore $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$ is an orthogonal basis of $\mathbb{R}^3$.

This observation underlies the Gram-Schmidt process, which we will discuss next class.

Example 67. Express $\underbrace{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}}_{\boldsymbol{x}}$ in terms of the basis $\underbrace{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}_{\boldsymbol{v}_1}, \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{v}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{v}_3}$.

Solution. Because $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ is an orthogonal basis of $\mathbb{R}^3$, we get (much as in the previous example):

$$\begin{aligned}
\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} &= c_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
&= \underbrace{\frac{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}_{\text{projection of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_1} + \underbrace{\frac{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\text{projection of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_2} + \underbrace{\frac{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\text{projection of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_3} \\
&= \frac{-4}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{10}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{4}{1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}
\end{aligned}$$

Because we spelled out all the details this looks more involved than it is. We only computed 6 dot products!

Alternative. We could have solved $\begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$ to also find $\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \\ 4 \end{bmatrix}$.

The numbers are particularly easy here but in general, to find this solution, we have to go through the entire process of Gaussian elimination. On the other hand, if we have an orthogonal basis, the former approach requires less work, because it is just computing a few dot products.

Example 68. Express $\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$ in terms of the basis $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

Solution. This is not an orthogonal basis, so we cannot proceed as in the previous example.

To write $\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, we need to solve $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$.

Solving that system (do it!), we find $\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$.

Review. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthogonal, the orthogonal projection of $\boldsymbol{w}$ onto $\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is

$$\hat{\boldsymbol{w}} = \frac{\boldsymbol{w}\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\boldsymbol{v}_1 + \ldots + \frac{\boldsymbol{w}\cdot\boldsymbol{v}_n}{\boldsymbol{v}_n\cdot\boldsymbol{v}_n}\boldsymbol{v}_n.$$

Example 69.

(a) Project $\begin{bmatrix}3\\2\\1\end{bmatrix}$ onto $W = \operatorname{span}\left\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}2\\-1\\0\end{bmatrix}\right\}$.

(b) Express $\begin{bmatrix}3\\2\\1\end{bmatrix}$ in terms of the basis $\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}2\\-1\\0\end{bmatrix}, \begin{bmatrix}1\\2\\-5\end{bmatrix}$.

Solution.

(a) We note that the vectors $\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}2\\-1\\0\end{bmatrix}$ are orthogonal to each other.

Therefore, the projection can be computed as $\dfrac{\begin{bmatrix}3\\2\\1\end{bmatrix}\cdot\begin{bmatrix}1\\2\\1\end{bmatrix}}{\begin{bmatrix}1\\2\\1\end{bmatrix}\cdot\begin{bmatrix}1\\2\\1\end{bmatrix}}\begin{bmatrix}1\\2\\1\end{bmatrix} + \dfrac{\begin{bmatrix}3\\2\\1\end{bmatrix}\cdot\begin{bmatrix}2\\-1\\0\end{bmatrix}}{\begin{bmatrix}2\\-1\\0\end{bmatrix}\cdot\begin{bmatrix}2\\-1\\0\end{bmatrix}}\begin{bmatrix}2\\-1\\0\end{bmatrix} = \frac{8}{6}\begin{bmatrix}1\\2\\1\end{bmatrix} + \frac{4}{5}\begin{bmatrix}2\\-1\\0\end{bmatrix}$.

Comment. If we didn't have an orthogonal basis for $W = \operatorname{col}\left(\begin{bmatrix}1 & 2\\2 & -1\\1 & 0\end{bmatrix}\right)$, then we would have to solve the least squares problem $\begin{bmatrix}1 & 2\\2 & -1\\1 & 0\end{bmatrix}\boldsymbol{x} = \begin{bmatrix}3\\2\\1\end{bmatrix}$ instead to get the same final result (with more work).

(b) Note that this basis is orthogonal! Therefore, we can compute $\begin{bmatrix}3\\2\\1\end{bmatrix} = \frac{8}{6}\begin{bmatrix}1\\2\\1\end{bmatrix} + \frac{4}{5}\begin{bmatrix}2\\-1\\0\end{bmatrix} + \frac{5}{30}\begin{bmatrix}1\\2\\-5\end{bmatrix}$.

(We proceed exactly as in the previous part to compute each coefficient as a quotient of dot products.)

Gram–Schmidt

(Gram–Schmidt orthogonalization)

Given a basis $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots$ for W, we produce an orthogonal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ for W as follows:

- $\boldsymbol{q}_1 = \boldsymbol{w}_1$
- $\boldsymbol{q}_2 = \boldsymbol{w}_2 - \begin{pmatrix}\text{projection of}\\ \boldsymbol{w}_2 \text{ onto } \boldsymbol{q}_1\end{pmatrix}$
- $\boldsymbol{q}_3 = \boldsymbol{w}_3 - \begin{pmatrix}\text{projection of}\\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_1\end{pmatrix} - \begin{pmatrix}\text{projection of}\\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_2\end{pmatrix}$
- $\boldsymbol{q}_4 = \ldots$

Note. Since $\boldsymbol{q}_1, \boldsymbol{q}_2$ are orthogonal, $\begin{pmatrix}\text{projection of}\\ \boldsymbol{w}_3 \text{ onto } \operatorname{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\}\end{pmatrix} = \begin{pmatrix}\text{projection of}\\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_1\end{pmatrix} + \begin{pmatrix}\text{projection of}\\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_2\end{pmatrix}$.

Important comment. When working numerically on a computer it actually saves time to compute an orthonormal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ by the same approach but always normalizing each $\boldsymbol{q}_i$ along the way. The reason this saves time is that now the projections onto $\boldsymbol{q}_i$ only require a single dot product (instead of two). This is called **Gram–Schmidt orthonormalization**. When working by hand, it is usually simpler to wait until the end to normalize (so as to avoid working with square roots).

Note. When normalizing, the orthonormal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ is the unique one (up to $\pm$ signs) with the property that $\operatorname{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_k\} = \operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ for all $k = 1, 2, \ldots$.

Example 70. Using Gram–Schmidt, find an orthogonal basis for $W = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$.

Solution. We already have the basis $\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{w}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ for W. However, that basis is not orthogonal.

We can construct an orthogonal basis $\boldsymbol{q}_1, \boldsymbol{q}_2$ for W as follows:

- $\boldsymbol{q}_1 = \boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
- $\boldsymbol{q}_2 = \boldsymbol{w}_2 - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_2 \text{ onto } \boldsymbol{q}_1 \end{pmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 \\ -4 \\ 2 \end{bmatrix}$

  Note. $\boldsymbol{q}_2$ is the error of the projection of $\boldsymbol{w}_2$ onto $\boldsymbol{q}_1$. This guarantees that it is orthogonal to $\boldsymbol{q}_1$. On the other hand, since $\boldsymbol{q}_2$ is a combination of $\boldsymbol{w}_2$ and $\boldsymbol{q}_1$, we know that $\boldsymbol{q}_2$ actually is in W.

We have thus found the orthogonal basis $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{2}{3}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ for W (if we like, we can, of course, drop that $\frac{2}{3}$).

Important comment. By normalizing, we get an orthonormal basis for W: $\frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$.

Practical comment. When implementing Gram–Schmidt on a computer, it is beneficial (slightly less work) to normalize each $\boldsymbol{q}_i$ during the Gram–Schmidt process. This typically introduces square roots, which is why normalizing at the end is usually preferable when working by hand.

Comment. There are, of course, many orthogonal bases $\boldsymbol{q}_1, \boldsymbol{q}_2$ for W. Up to the length of the vectors, ours is the unique one with the property that $\operatorname{span}\{\boldsymbol{q}_1\} = \operatorname{span}\{\boldsymbol{w}_1\}$ and $\operatorname{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\} = \operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$.

A matrix Q has orthonormal columns $\iff$ $Q^T Q = I$

Why? Let $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ be the columns of Q. By the way matrix multiplication works, the entries of $Q^T Q$ are dot products of these columns:

$$\begin{bmatrix} — & \boldsymbol{q}_1^T & — \\ — & \boldsymbol{q}_2^T & — \\ & \vdots & \end{bmatrix} \begin{bmatrix} | & | & \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \ddots \end{bmatrix}$$

Hence, $Q^T Q = I$ if and only if $\boldsymbol{q}_i^T \boldsymbol{q}_j = 0$ (that is, the columns are orthogonal), for $i \neq j$, and $\boldsymbol{q}_i^T \boldsymbol{q}_i = 1$ (that is, the columns are normalized).

Example 71. $Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}$ obtained from Example 70 satisfies $Q^T Q = I$.

The QR decomposition

Just like the LU decomposition encodes the steps of Gaussian elimination, the QR decomposition encodes the steps of Gram–Schmidt.

(QR decomposition) Every $m \times n$ matrix A of rank n can be decomposed as $A = QR$, where

- Q has orthonormal columns, $(m \times n)$
- R is upper triangular and invertible. $(n \times n)$

How to find Q and R?

- Gram–Schmidt orthonormalization on (columns of) A, to get (columns of) Q
- $R = Q^T A$
  Why? If $A = QR$, then $Q^T A = Q^T Q R$ which simplifies to $R = Q^T A$ (since $Q^T Q = I$).

The decomposition $A = QR$ is unique if we require the diagonal entries of R to be positive (and this is exactly what happens when applying Gram–Schmidt).

Practical comment. Actually, no extra work is needed for computing R. All of its entries have been computed during Gram–Schmidt.

Variations. We can also arrange things so that Q is an $m \times m$ **orthogonal** matrix (this means Q is square and has orthonormal columns) and R a $m \times n$ upper triangular matrix. This is a tiny bit more work (and not required for many applications): we need to complement "our" Q with additional orthonormal columns and add corresponding zero rows to R. For square matrices this makes no difference.

Example 72. Determine the QR decomposition of $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$.

Solution. The first step is Gram–Schmidt orthonormalization on the columns of A. We then use the resulting orthonormal vectors (they need to be normalized!) as the columns of Q.

We already did Gram–Schmidt in Example 70: from that work, we have $Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}$.

Hence, $R = Q^T A = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 1/\sqrt{3} \\ 0 & 4/\sqrt{6} \end{bmatrix}$.

Comment. The entries of R have actually all been computed during Gram–Schmidt, so that, if we pay attention, we could immediately write down R (no extra work required). Looking back at Example 70, can you see this?

Check. Indeed, $QR = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 1/\sqrt{3} \\ 0 & 4/\sqrt{6} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$ equals A.

Example 73. Using Gram–Schmidt, find an orthogonal basis for $W = \operatorname{span}\left\{ \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

Solution. We begin with the (not orthogonal) basis $\boldsymbol{w}_1 = \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}$, $\boldsymbol{w}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $\boldsymbol{w}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

We then construct an orthogonal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$:

- $\boldsymbol{q}_1 = \boldsymbol{w}_1 = \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}$
- $\boldsymbol{q}_2 = \boldsymbol{w}_2 - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_2 \text{ onto } \boldsymbol{q}_1 \end{pmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{3}{9} \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$
- $\boldsymbol{q}_3 = \boldsymbol{w}_3 - \begin{pmatrix} \text{projection of } \boldsymbol{w}_3 \\ \text{onto span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\} \end{pmatrix} = \boldsymbol{w}_3 - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_1 \end{pmatrix} - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_2 \end{pmatrix}$
  $= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{9} \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix} - \frac{2}{4} \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

  Make sure you understand how $\boldsymbol{q}_3$ was designed to be orthogonal to both $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$!

  Also note that breaking up the projection onto $\operatorname{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$ into the projections onto $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ is only possible because $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are orthogonal.

Hence, $\begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ is an orthogonal basis of W.

Important. Normalizing, we obtain an orthonormal basis: $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$.

Example 74. Determine the QR decomposition of $A = \begin{bmatrix} 0 & 2 & 1 \\ 3 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

Solution. The first step is Gram–Schmidt orthonormalization on the columns of A. We then use the resulting orthonormal vectors as the columns of Q.

We already did Gram–Schmidt in Example 73: from that work, we have $Q = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}$.

Hence, $R = Q^T A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 2 & 1 \\ 3 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$.

Comment. As commented earlier, the entries of R have actually all been computed during Gram–Schmidt, so that, if we pay attention, we could immediately write down R (no extra work required). Looking back at Example 73, can you see this?

Armin Straub
straub@southalabama.edu