Math 4242 Fall 2016 (Darij Grinberg): midterm 2 pratice problems

Exercise 1. Consider the vector space $\mathbb{R}^3$.
  (a) The list $\mathbf{a} = \left((1,2,-1)^T, (1,1,0)^T, (0,1,-1)^T, (1,1,1)^T \right)$ spans $\mathbb{R}^3$. Shrink this list to a basis of $\mathbb{R}^3$ by removing some redundant elements.
  (b) The list $\mathbf{b} = \left((-1,0,1)^T, (2,3,4)^T \right)$ is linearly independent. Extend this list to a basis of $\mathbb{R}^3$ by appending to it some elements from the list $\mathbf{a}$.

Solution. (I am going to be very detailed here. You don't need to write half as much when solving this kind of problem!)

(a) We proceed using the standard algorithm[1]: We scan the list $\mathbf{a}$ from left to right. Each time we read an entry of $\mathbf{a}$, we check if this entry is a linear combination of the entries before it. If it is, then we remove this entry from $\mathbf{a}$ and start from scratch with the new (shorter) $\mathbf{a}$. If it is not, then we proceed to the next entry. If we have arrived at the end of the list, then our list has no redundant entries, and thus is a basis of $\mathbb{R}^3$.
  Let us execute this algorithm step by step:

- We scan the list $\mathbf{a}$ from left to right. Thus, we begin at its first entry, which is $(1,2,-1)^T$.

- Is this first entry $(1,2,-1)^T$ a linear combination of the entries before it? There are no entries before it, and thus the only linear combination of the entries before it is $\overrightarrow{0}$ (because the only linear combination of no vectors is $\overrightarrow{0}$). But our entry $(1,2,-1)^T$ is not $\overrightarrow{0}$; thus, $(1,2,-1)^T$ is not a linear combination of the entries before it. Hence, we proceed to the second entry.

- Is this second entry $(1,1,0)^T$ a linear combination of the entries before it? There is only one entry before it, namely $(1,2,-1)^T$. Hence, we are asking whether $(1,1,0)^T$ is a linear combination of the vector $(1,2,-1)^T$. In other words, we are asking whether $(1,1,0)^T = \lambda_1 (1,2,-1)^T$ for some $\lambda_1 \in \mathbb{R}$. Equivalently, we want to know whether $\begin{cases} 1 = 1\lambda_1; \\ 1 = 2\lambda_1; \\ 0 = -1\lambda_1 \end{cases}$ for some $\lambda_1 \in \mathbb{R}$ (because the equation $(1,1,0)^T = \lambda_1 (1,2,-1)^T$ is equivalent to the system of equations $\begin{cases} 1 = 1\lambda_1; \\ 1 = 2\lambda_1; \\ 0 = -1\lambda_1 \end{cases}$). In other words, we want to know whether the system $\begin{cases} 1 = 1\lambda_1; \\ 1 = 2\lambda_1; \\ 0 = -1\lambda_1 \end{cases}$ of linear equations (in the unknown λ_1) has a solution.

---

[1]In this algorithm, we treat $\mathbf{a}$ as a mutable variable.

But this question is easy to answer (e.g., by Gaussian elimination[2]), and the answer is "no". Thus, our entry $(1, 1, 0)^T$ is not a linear combination of the entries before it. Hence, we proceed to the third entry.

- Is this third entry $(0, 1, -1)^T$ a linear combination of the entries before it? The entries before it are $(1, 2, -1)^T$ and $(1, 1, 0)^T$. Hence, we are asking whether $(0, 1, -1)^T$ is a linear combination of the vectors $(1, 2, -1)^T$ and $(1, 1, 0)^T$. In other words, we are asking whether $(0, 1, -1)^T = \lambda_1 (1, 2, -1)^T + \lambda_2 (1, 1, 0)^T$

  for some $\lambda_1 \in \mathbb{R}$. Equivalently, we want to know whether $\begin{cases} 0 = 1\lambda_1 + 1\lambda_2; \\ 1 = 2\lambda_1 + 1\lambda_2; \\ -1 = -1\lambda_1 + 0\lambda_2 \end{cases}$

  for some $\lambda_1, \lambda_2 \in \mathbb{R}$ (because the equation $(0, 1, -1)^T = \lambda_1 (1, 2, -1)^T + \lambda_2 (1, 1, 0)^T$ is equivalent to the system of equations $\begin{cases} 0 = 1\lambda_1 + 1\lambda_2; \\ 1 = 2\lambda_1 + 1\lambda_2; \\ -1 = -1\lambda_1 + 0\lambda_2 \end{cases}$).

  In other words, we want to know whether the system $\begin{cases} 0 = 1\lambda_1 + 1\lambda_2; \\ 1 = 2\lambda_1 + 1\lambda_2; \\ -1 = -1\lambda_1 + 0\lambda_2 \end{cases}$

  of linear equations (in the unknowns λ_1, λ_2) has a solution. But this question is easy to answer (e.g., by Gaussian elimination[3]), and the answer is

---

[2]Of course, for this particular system, it is clear by inspection. But Gaussian elimination is a method that works in all situations.

[3]Let me give details on how this is done. (I will not be using the matrix form of Gaussian elimination, but work with plain equations instead, since I want to avoid the overhead of translating between matrices and their entries.)

Our system $\begin{cases} 0 = 1\lambda_1 + 1\lambda_2; \\ 1 = 2\lambda_1 + 1\lambda_2; \\ -1 = -1\lambda_1 + 0\lambda_2 \end{cases}$ rewrites as $\begin{cases} 1\lambda_1 + 1\lambda_2 = 0; \\ 2\lambda_1 + 1\lambda_2 = 1; \\ -1\lambda_1 + 0\lambda_2 = -1 \end{cases}$ (just to bring the unknowns onto the left-hand side). Subtracting twice the first equation from the second equation, we transform it into $\begin{cases} 1\lambda_1 + 1\lambda_2 = 0; \\ -1\lambda_2 = 1; \\ -1\lambda_1 + 0\lambda_2 = -1 \end{cases}$. Adding the first equation to the third equation, we transform this further into $\begin{cases} 1\lambda_1 + 1\lambda_2 = 0; \\ -1\lambda_2 = 1; \\ 1\lambda_2 = -1 \end{cases}$. Adding the second equation to the third equation, we transform this further into $\begin{cases} 1\lambda_1 + 1\lambda_2 = 0; \\ -1\lambda_2 = 1; \\ 0 = 0 \end{cases}$. The system we now have can be solved by back-substitution: The third equation ($0 = 0$) says nothing and thus can be discarded; the second equation ($-1\lambda_2 = 1$) lets us compute λ_2 (namely, $\lambda_2 = -1$); finally, the first equation ($1\lambda_1 + 1\lambda_2 = 0$) lets us compute λ_1 (namely, $\lambda_1 = 1$). Thus, we see that the only solution is $(\lambda_1, \lambda_2) = (1, -1)$.

Remark: In general, solving a system of linear equations can lead to free variables (when there is more than one solution). However, this won't happen in this particular kind of situation, because the first redundant element in a list can always be written as a linear combination of the elements before it in a **unique** way (this is not hard to prove, if you are so inclined), and thus the system of equations that determines this combination will have a unique solution.

"yes". Thus, our entry $(0,1,-1)^T$ is a linear combination of the entries before it[4]. Thus, we remove the entry from **a**, and start from scratch with the new (shorter) **a**.

- We scan the new list $\mathbf{a} = \left((1,2,-1)^T, (1,1,0)^T, (1,1,1)^T \right)$ (the result of removing $(0,1,-1)^T$ from the old list **a**) from left to right. Thus, we begin at its first entry, which is $(1,2,-1)^T$.

- Is this first entry $(1,2,-1)^T$ a linear combination of the entries before it? We have already answered this question during our previous scan of the list (since the segment of our list **a** up to its first entry has not changed when we removed $(0,1,-1)^T$), and thus we already know that the answer is "no". Hence, we proceed to the second entry.

- Is this second entry $(1,1,0)^T$ a linear combination of the entries before it? Again, this is a question we have already answered during our previous scan of the list (since the segment of our list **a** up to its second entry has not changed when we removed $(0,1,-1)^T$), and thus we already know that the answer is "no". Hence, we proceed to the third entry.

- Is this third entry $(1,1,1)^T$ a linear combination of entries before it? The entries before it are $(1,2,-1)^T$ and $(1,1,0)^T$. Hence, we are asking whether $(1,1,1)^T$ is a linear combination of the vectors $(1,2,-1)^T$ and $(1,1,0)^T$. In other words, we are asking whether $(1,1,1)^T = \lambda_1 (1,2,-1)^T + \lambda_2 (1,1,0)^T$ for some $\lambda_1 \in \mathbb{R}$. Equivalently, we want to know whether $\begin{cases} 1 = 1\lambda_1 + 1\lambda_2; \\ 1 = 2\lambda_1 + 1\lambda_2; \\ 1 = -1\lambda_1 + 0\lambda_2 \end{cases}$ for some $\lambda_1, \lambda_2 \in \mathbb{R}$ (because the equation $(1,1,1)^T = \lambda_1 (1,2,-1)^T + \lambda_2 (1,1,0)^T$ is equivalent to the system of equations $\begin{cases} 1 = 1\lambda_1 + 1\lambda_2; \\ 1 = 2\lambda_1 + 1\lambda_2; \\ 1 = -1\lambda_1 + 0\lambda_2 \end{cases}$). In other words, we want to know whether the system $\begin{cases} 1 = 1\lambda_1 + 1\lambda_2; \\ 1 = 2\lambda_1 + 1\lambda_2; \\ 1 = -1\lambda_1 + 0\lambda_2 \end{cases}$ of linear equations (in the unknowns λ_1, λ_2) has a solution. But this question is easy to answer (e.g., by Gaussian elimination), and the answer is "no". Thus, our entry $(1,1,1)^T$ is not a linear combination of the entries before it. Thus, we have arrived at the end of the list.

We have thus ended up with the list $\left((1,2,-1)^T, (1,1,0)^T, (1,1,1)^T \right)$. This list is therefore a basis of $\mathbb{R}^3$ obtained by shrinking our (old) list **a**.

---

[4]Namely, $(0,1,-1)^T = 1(1,2,-1)^T + (-1)(1,1,0)^T$. But we don't need to know these specifics.

(b) There are various ways to do this. One particularly simple way is the following[5]: We scan the list **a** from left to right. Each time we read an entry of **a**, we check if this entry is a linear combination of the (current) entries of **b**. If it isn't, then we append this entry to **b**. In either case, we proceed to the next entry. By the time we have scanned all entries of **a**, the list **b** has become a basis of $\mathbb{R}^3$. (This is easy to prove[6].)

In order to simplify our life, we use not the original list

$$\mathbf{a} = \left((1, 2, -1)^T, (1, 1, 0)^T, (0, 1, -1)^T, (1, 1, 1)^T \right),$$

but the shorter list

$$\mathbf{a} = \left((1, 2, -1)^T, (1, 1, 0)^T, (1, 1, 1)^T \right)$$

obtained at the end of the shrinking process in part **(a)** of the problem. Indeed, this shorter list works just as well (it is a basis of $\mathbb{R}^3$ and thus spans $\mathbb{R}^3$), and clearly its elements are elements of the original list **a** as well.

Let us now execute our algorithm step by step:

- We scan the list **a** from left to right. Thus, we begin at its first entry, which is $(1, 2, -1)^T$.

- Is this first entry $(1, 2, -1)^T$ a linear combination of the entries of **b**? The entries of **b** are $(-1, 0, 1)^T$ and $(2, 3, 4)^T$. Hence, we are asking whether $(1, 2, -1)^T$ is a linear combination of the vectors $(-1, 0, 1)^T$ and $(2, 3, 4)^T$. In other words, we are asking whether $(1, 2, -1)^T = \lambda_1 (-1, 0, 1)^T + \lambda_2 (2, 3, 4)^T$ for some $\lambda_1, \lambda_2 \in \mathbb{R}$. Equivalently, we want to know whether $\begin{cases} 1 = (-1)\lambda_1 + 2\lambda_2; \\ 2 = 0\lambda_1 + 3\lambda_2; \\ -1 = 1\lambda_1 + 4\lambda_2 \end{cases}$ for some $\lambda_1, \lambda_2 \in \mathbb{R}$ (because the equation $(1, 2, -1)^T = \lambda_1 (-1, 0, 1)^T +$

---

[5]In this algorithm, we treat **b** as a mutable variable.

[6]*Proof.* Consider the following:

- Every time we append an entry of **a** to the list **b**, the list **b** remains linearly independent (because we append an entry to **b** only if this entry is not a linear combination of the existing entries of **b**; but this guarantees that the linear independence of the list **b** is preserved).

- By the time we have scanned all entries of **a**, the list **b** has the property that each entry of **a** is a linear combination of the entries of **b** (because when we scanned this entry, we have ensured that it became such a linear combination by appending it to **b**, if it wasn't already one). In other words, every entry of **a** belongs to span (**b**). Thus, span (**a**) $\subseteq$ span (**b**). But since **a** spans $\mathbb{R}^3$, we have span (**a**) $= \mathbb{R}^3$, so that $\mathbb{R}^3 =$ span (**a**) $\subseteq$ span (**b**) and thus span (**b**) $= \mathbb{R}^3$.

Hence, by the time we have scanned all entries of **a**, the list **b** is linearly independent and satisfies span (**b**) $= \mathbb{R}^3$. In other words, this list **b** has become a basis of $\mathbb{R}^3$.

$\lambda_2 (2,3,4)^T$ is equivalent to the system of equations $\begin{cases} 1 = (-1)\lambda_1 + 2\lambda_2; \\ 2 = 0\lambda_1 + 3\lambda_2; \\ -1 = 1\lambda_1 + 4\lambda_2 \end{cases}$).

In other words, we want to know whether the system $\begin{cases} 1 = (-1)\lambda_1 + 2\lambda_2; \\ 2 = 0\lambda_1 + 3\lambda_2; \\ -1 = 1\lambda_1 + 4\lambda_2 \end{cases}$

of linear equations (in the unknowns λ_1, λ_2) has a solution. But this question is easy to answer (e.g., by Gaussian elimination), and the answer is "no". Thus, our entry $(1, 2, -1)^T$ is not a linear combination of the entries of **b**. Thus, we append this entry to **b**, so that **b** becomes $\left((-1, 0, 1)^T, (2, 3, 4)^T, (1, 2, -1)^T \right)$. We now proceed to the second entry of **a**.

- Is this second entry $(1, 1, 0)^T$ a linear combination of the entries of **b**? The entries of **b** are $(-1, 0, 1)^T$, $(2, 3, 4)^T$ and $(1, 2, -1)^T$ (keep in mind that **b** has changed in the previous step!). Hence, we are asking whether $(1, 1, 0)^T$ is a linear combination of the vectors $(-1, 0, 1)^T$, $(2, 3, 4)^T$ and $(1, 2, -1)^T$. By now, we have seen often enough how to answer such questions (of course, we now have to solve a system of equations in three unknowns $\lambda_1, \lambda_2, \lambda_3$). The answer is "yes". Thus, our entry $(1, 1, 0)^T$ is a linear combination of the entries of **b**. Hence, we proceed to the third entry of **a** (without adding anything to **b**). Recall that we have used the shorter list $\mathbf{a} = \left((1, 2, -1)^T, (1, 1, 0)^T, (1, 1, 1)^T \right)$ as our **a**, so this third entry is $(1, 1, 1)^T$.

- Is this third entry $(1, 1, 1)^T$ a linear combination of the entries of **b**? The answer is "yes" (found in the same way as many times before). Hence, we arrive at the end of **a** (without adding anything to **b**).

We have thus ended up with the list $\mathbf{b} = \left((-1, 0, 1)^T, (2, 3, 4)^T, (1, 2, -1)^T \right)$. This list is therefore a basis of $\mathbb{R}^3$ obtained by appending some elements from **a** to the (old) list **b**.

[*Remark:* We could have made our life much easier. In fact, we could have stopped our algorithm immediately after adding $(1, 2, -1)^T$ to the list **b**, because the list **b** had become a basis of $\mathbb{R}^3$ at that moment (being a linearly independent list of 3 vectors in $\mathbb{R}^3$).

There are other ways to solve this exercise, and some of them lead to different results. For example, $\left((-1, 0, 1)^T, (2, 3, 4)^T, (0, 1, -1)^T \right)$ is an equally valid answer to part **(b)**.] $\qquad \square$

Exercise 2. (a) Find bases of the four subspaces of the 3×4-matrix $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \end{pmatrix}$.

> **(b)** [Too tricky for a midterm, but worth thinking about!] More generally: Let
> $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A_{n \times m}$ be the $n \times m$-matrix $\left(\min \{i, j\} \right)_{1 \le i \le n, \ 1 \le j \le m}$. (This
> is the $n \times m$-matrix whose (i, j)-th entry is $\min \{i, j\}$. For example, $A_{3 \times 4}$ is the
> matrix A from part **(a)** of this exercise.)
> Find bases of the four subspaces of $A_{n \times m}$.

Solution. **(a)** Let me first give the answers, then remind you how they can be obtained by safe-and-stupid algorithms, and finally how they can be obtained quickly using certain shortcuts.

The answers:

Column space: A basis of Col A is $\left((1, 1, 1)^T, (1, 2, 2)^T, (1, 2, 3)^T \right)$. (There are other bases, of course. For example, (e_1, e_2, e_3) is a basis of Col A, since Col A is simply $\mathbb{R}^3$.)

Row space: A basis of Row A is $((1, 1, 1, 1), (1, 2, 2, 2), (1, 2, 3, 3))$. (Again, there are other bases. For example, $((1, 1, 1, 1), (0, 1, 1, 1), (0, 0, 1, 1))$ is another basis of Row A.)

Kernel: A basis of Ker A is $\left((0, 0, 1, -1)^T \right)$.

Left kernel: A basis of $\left(\mathrm{Ker} \left(A^T \right) \right)^T$ is $()$ (that is, the empty list).

The algorithms:

Column space: How to find a basis of Col A ? Recall that Col A is the span of the columns of A. Thus, the columns of A span Col A. Now, it remains to shrink the list of te columns of A to a basis of Col A by removing redundant elements. We know how to do such shrinking (see Exercise 1 **(a)** for an example). Once we have done it, we are left with a basis of Col A.

(There are faster algorithms for finding a basis of Col A around; for example, Olver and Shakiban show two such algorithms on pp. 117–118 of their book. Feel free to use them.)

Row space: Finding a basis of Row A is similar to finding a basis of Col A, except that we are now dealing with rows instead of columns.

(Again, there are faster algorithms for finding a basis of Row A. Probably the fastest one is to transform A into row echelon form, and take the nonzero rows of the resulting matrix. These form a basis of Row A because they are linearly independent[7] and span Row A.)

Kernel: See Examples 4.34, 4.35 and 4.36 in my notes for how to find a basis of Ker A.

Left kernel: In order to find a basis of $\left(\mathrm{Ker} \left(A^T \right) \right)^T$, we just have to find a basis of Ker $\left(A^T \right)$ and then transpose each vector. Again, we know how to do this.

The shortcuts:

Column space: The matrix formed by the first three columns of A can be trans-

---

[7]This is not hard to see, using the fact that they have their pivot entries in different positions.

formed by row operations as follows:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix} \xrightarrow{A_{2,1}^{-1}} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \xrightarrow{A_{3,1}^{-1}} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

$$\xrightarrow{A_{3,2}^{-1}} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

The result is a row echelon matrix with three pivot entries and thus rank 3. Since the rank of a matrix does not change when you apply row operations (check this!), we thus conclude that the original matrix formed by the first three columns of A has rank 3. Hence, the first three columns of A are linearly independent. Hence, $\dim(\operatorname{Col} A) \geq 3$, so that $\operatorname{Col} A = \mathbb{R}^3$ (since A is a subspace of $\mathbb{R}^3$). Hence, to find a basis of $\operatorname{Col} A$ is to find a basis of $\mathbb{R}^3$. We can either take the first three columns of A, or take any basis of $\mathbb{R}^3$ that we know (for example, (e_1, e_2, e_3)).

Row space: We have $\dim(\operatorname{Row} A) = \operatorname{rank} A = \dim(\operatorname{Col} A) \geq 3$ (as we have just seen). Thus, the three rows of A are linearly independent (since otherwise, at least one of them would be redundant, and thus $\operatorname{Row} A$ would have a basis of size < 3, which contradicts $\dim(\operatorname{Row} A) \geq 3$). Hence, the three rows of A form a basis of $\operatorname{Row} A$.

Kernel: We have just shown that the three rows of A form a basis of $\operatorname{Row} A$. Hence, $\dim(\operatorname{Row} A) = 3$, so that $\operatorname{rank} A = \dim(\operatorname{Row} A) = 3$.

The rank-nullity theorem yields $\operatorname{rank} A + \dim(\operatorname{Ker} A) = 4$, so that $\dim(\operatorname{Ker} A) = 4 - \underbrace{\operatorname{rank} A}_{=3} = 4 - 3 = 1$. Thus, in order to find a basis of $\operatorname{Ker} A$, it suffices to find one linearly independent vector in $\operatorname{Ker} A$.

It is easy to observe that $(0,0,1,-1)^T \in \operatorname{Ker} A$ (because the product $A(0,0,1,-1)^T$ equals the third column of A minus the fourth column of A, but this is clearly $\overrightarrow{0}$ because the third and the fourth columns of A are equal). Thus, $\left((0,0,1,-1)^T\right)$ is a list of 1 linearly independent element of $\operatorname{Ker} A$, and therefore is a basis of $\operatorname{Ker} A$ (since $\dim(\operatorname{Ker} A) = 1$).

Left kernel: The rank-nullity theorem yields $\operatorname{rank}(A^T) + \dim(\operatorname{Ker}(A^T)) = 3$. Thus, $\dim(\operatorname{Ker}(A^T)) = 3 - \underbrace{\operatorname{rank}(A^T)}_{=\operatorname{rank} A = 3} = 3 - 3 = 0$, so that $\operatorname{Ker}(A^T) = \left\{\overrightarrow{0}\right\}$ and therefore $\left(\operatorname{Ker}(A^T)\right)^T = \left\{\overrightarrow{0}\right\}$. Hence, a basis of $\left(\operatorname{Ker}(A^T)\right)^T$ is the empty list.

(b) I am going to just state the answers and very briefly hint at the proofs.
The answers:
I shall abbreviate the $n \times m$-matrix $A_{n \times m}$ by A.
Column space: The first $\min\{n, m\}$ columns of A form a basis of $\operatorname{Col} A$.
Row space: The first $\min\{n, m\}$ rows of A form a basis of $\operatorname{Row} A$.
Kernel: A basis of $\operatorname{Ker} A$ consists of all vectors $e_j - e_{j+1}$ with $j \in \{n, n+1, \ldots, m-1\}$.

(When $n \geq m$, the set $\{n, n+1, \ldots, m-1\}$ is empty, whence the basis of Ker A is an empty list in this case.)

Left kernel: A basis of $\left(\text{Ker}\left(A^T\right)\right)^T$ consists of all vectors $(e_i - e_{i+1})^T$ with $j \in \{m, m+1, \ldots, n-1\}$. (When $m \geq n$, the set $\{m, m+1, \ldots, n-1\}$ is empty, whence the basis of $\left(\text{Ker}\left(A^T\right)\right)^T$ is an empty list in this case.)

Proof ideas:

First of all, recall that if we know the four subspaces of some matrix B, then we get the four subspaces of B^T for free (since Row $\left(B^T\right) = \left(\text{Col } B\right)^T$, etc.). Observe that $\left(A_{n \times m}\right)^T = A_{m \times n}$. Thus, it suffices to solve the case when $n \leq m$ (because in the case $n \geq m$, we can just switch n and m). Hence, WLOG assume that $n \leq m$. Let $A = A_{n \times m}$.

The first n columns of A are linearly independent (indeed, Gaussian elimination transforms them into the upper-triangular $n \times n$-matrix $\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$, which is a row-echelon matrix with n pivots and thus has rank n). Hence, rank $A \geq n$.

But rank $A \leq n$ as well (since A is an $n \times m$-matrix), and thus rank $A = n$. Therefore, the first n columns of A form a basis of Col A.

Also, the n rows of A are linearly independent (since $\dim (\text{Row } A) = \text{rank } A = n$), and thus form a basis of Row A.

The rank-nullity theorem yields rank $A + \dim (\text{Ker } A) = m$, so that $\dim (\text{Ker } A) = m - \underbrace{\text{rank } A}_{=n} = m - n$. The vectors $e_j - e_{j+1}$ with $j \in \{n, n+1, \ldots, m-1\}$ belong to Ker A (because $A\left(e_j - e_{j+1}\right)$ is the difference between the j-th and the $(j+1)$-th columns of A, but these two columns are equal, and thus the difference is $\vec{0}$), and are linearly independent (this is not hard to check). Hence, they form a basis of Ker A (since they are $m - n$ linearly independent elements of the $(m-n)$-dimensional vector space Ker A).

The rank-nullity theorem yields rank $\left(A^T\right) + \dim \left(\text{Ker}\left(A^T\right)\right) = n$, so that $\dim \left(\text{Ker}\left(A^T\right)\right) = n - \underbrace{\text{rank}\left(A^T\right)}_{=\text{rank } A = n} = n - n = 0$. Hence, $\text{Ker}\left(A^T\right) = \left\{\vec{0}\right\}$, so that $\left(\text{Ker}\left(A^T\right)\right)^T = \left\{\vec{0}\right\}$. Thus, the empty list () is a basis of $\left(\text{Ker}\left(A^T\right)\right)^T$. $\square$

If A is an $n \times k$-matrix whose columns are linearly independent, then a *QR decomposition* of A means a way to write A in the form $A = QR$, where:

- Q is an $n \times k$-matrix with orthonormal columns (this is equivalent to saying that Q is an $n \times k$-matrix satisfying $Q^T Q = I_k$);

- R is an upper-triangular $k \times k$-matrix with nonzero diagonal entries.

For example, a QR decomposition of $\begin{pmatrix} 2 & 17 \\ 4 & 13 \\ 8 & 5 \end{pmatrix}$ is

$$\begin{pmatrix} 2 & 17 \\ 4 & 13 \\ 8 & 5 \end{pmatrix} = \underbrace{\begin{pmatrix} \dfrac{1}{\sqrt{21}} & \dfrac{2}{\sqrt{6}} \\ \dfrac{2}{\sqrt{21}} & \dfrac{1}{\sqrt{6}} \\ \dfrac{4}{\sqrt{21}} & \dfrac{-1}{\sqrt{6}} \end{pmatrix}}_{\text{this is the } Q} \underbrace{\begin{pmatrix} 2\sqrt{21} & 3\sqrt{21} \\ 0 & 7\sqrt{6} \end{pmatrix}}_{\text{this is the } R}.$$

Exercise 3. (a) Find a QR decomposition of the matrix $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$.

(b) Find a QR decomposition of the matrix $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$.

(c) Find a QR decomposition of the matrix $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

[**Hint:** Two of the three parts are easy and can be done with no computations whatsoever!]

Solution. **(a)** Let A be our matrix $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. Let w_1, w_2, w_3 be the three columns of A; thus,

$$w_1 = (0, 1, 1)^T, \qquad w_2 = (1, 0, 1)^T, \qquad w_3 = (1, 1, 0)^T.$$

Now, we apply the Gram-Schmidt process to w_1, w_2, w_3:

1. At the first step, we set $u_1 = w_1$. Thus,

$$u_1 = w_1 = (0, 1, 1)^T.$$

2. At the second step, we set $u_2 = w_2 - \lambda_{2,1} u_1$, where $\lambda_{2,1} = \dfrac{\langle w_2, u_1 \rangle}{\langle u_1, u_1 \rangle}$. We compute these explicitly:

$$\lambda_{2,1} = \frac{\langle w_2, u_1 \rangle}{\langle u_1, u_1 \rangle} = \frac{\left\langle (1, 0, 1)^T, (0, 1, 1)^T \right\rangle}{\left\langle (0, 1, 1)^T, (0, 1, 1)^T \right\rangle} = \frac{1}{2}$$

and thus

$$u_2 = w_2 - \lambda_{2,1} u_1 = (1, 0, 1)^T - \frac{1}{2}(0, 1, 1)^T = \left(1, -\frac{1}{2}, \frac{1}{2}\right)^T.$$

3. At the third step, we set $u_3 = w_3 - \lambda_{3,1} u_1 - \lambda_{3,2} u_2$, where $\lambda_{3,1} = \dfrac{\langle w_3, u_1 \rangle}{\langle u_1, u_1 \rangle}$ and

$\lambda_{3,2} = \dfrac{\langle w_3, u_2 \rangle}{\langle u_2, u_2 \rangle}$. We compute these explicitly:

$$\lambda_{3,1} = \frac{\langle w_3, u_1 \rangle}{\langle u_1, u_1 \rangle} = \frac{\left\langle (1,1,0)^T, (0,1,1)^T \right\rangle}{\left\langle (0,1,1)^T, (0,1,1)^T \right\rangle} = \frac{1}{2}$$

and

$$\lambda_{3,2} = \frac{\langle w_3, u_2 \rangle}{\langle u_2, u_2 \rangle} = \frac{\left\langle (1,1,0)^T, \left(1, -\frac{1}{2}, \frac{1}{2}\right)^T \right\rangle}{\left\langle \left(1, -\frac{1}{2}, \frac{1}{2}\right)^T, \left(1, -\frac{1}{2}, \frac{1}{2}\right)^T \right\rangle} = \frac{\left(\frac{1}{2}\right)}{\left(\frac{3}{2}\right)} = \frac{1}{3}$$

and thus

$$u_3 = w_3 - \lambda_{3,1} u_1 - \lambda_{3,2} u_2$$
$$= (1,1,0)^T - \frac{1}{2}(0,1,1)^T - \frac{1}{3}\left(1, -\frac{1}{2}, \frac{1}{2}\right)^T$$
$$= \left(\frac{2}{3}, \frac{2}{3}, \frac{-2}{3}\right)^T .$$

Next, we normalize the vectors u_1, u_2, u_3 – that is, we divide them by their lengths so they become orthonormal and not just orthogonal. The resulting vectors will be called q_1, q_2, q_3. Explicitly:

$$q_1 = \frac{1}{||u_1||} u_1 = \frac{1}{\sqrt{2}}(0,1,1)^T \qquad \left(\text{since } ||u_1|| = \sqrt{\langle u_1, u_1 \rangle} = \sqrt{2} \right)$$
$$= \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T,$$

$$q_2 = \frac{1}{||u_2||} u_2 = \frac{1}{\sqrt{\frac{3}{2}}}\left(1, -\frac{1}{2}, \frac{1}{2}\right)^T \qquad \left(\text{since } ||u_2|| = \sqrt{\langle u_2, u_2 \rangle} = \sqrt{\frac{3}{2}} \right)$$
$$= \left(\frac{1}{\sqrt{\frac{3}{2}}}, \frac{\left(-\frac{1}{2}\right)}{\sqrt{\frac{3}{2}}}, \frac{\left(\frac{1}{2}\right)}{\sqrt{\frac{3}{2}}}\right)^T = \left(\frac{\sqrt{2}}{\sqrt{3}}, \frac{-1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)^T,$$

$$q_3 = \frac{1}{||u_3||}u_3 = \frac{1}{\sqrt{\frac{4}{3}}}\left(\frac{2}{3}, \frac{2}{3}, \frac{-2}{3}\right)^T \qquad \left(\text{since } ||u_3|| = \sqrt{\langle u_3, u_3 \rangle} = \sqrt{\frac{4}{3}}\right)$$

$$= \left(\frac{\left(\frac{2}{3}\right)}{\sqrt{\frac{4}{3}}}, \frac{\left(\frac{2}{3}\right)}{\sqrt{\frac{4}{3}}}, \frac{\left(\frac{-2}{3}\right)}{\sqrt{\frac{4}{3}}}\right)^T = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{-1}{\sqrt{3}}\right)^T.$$

Now, the Q and R in the QR decomposition $A = QR$ of A can be determined as follows:

- The matrix Q will be the 3×3-matrix with columns q_1, q_2, q_3. Plugging in the values of q_1, q_2, q_3 already computed, we thus find

$$Q = \begin{pmatrix} 0 & \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{3}} \end{pmatrix}.$$

- The matrix R will be the 3×3-matrix whose (i, j)-th entry (for all i and j) is

$$R_{i,j} = \begin{cases} \lambda_{j,i} ||u_i||, & \text{if } i < j; \\ ||u_j||, & \text{if } i = j; \\ 0, & \text{if } i > j \end{cases}.$$

In other words,

$$R = \begin{pmatrix} ||u_1|| & \lambda_{2,1} ||u_1|| & \lambda_{3,1} ||u_1|| \\ 0 & ||u_2|| & \lambda_{3,2} ||u_2|| \\ 0 & 0 & ||u_3|| \end{pmatrix}.$$

Plugging in the values of $||u_i||$ and $\lambda_{j,i}$ (which have already been computed), we obtain

$$R = \begin{pmatrix} \sqrt{2} & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ 0 & \sqrt{\frac{3}{2}} & \frac{1}{3}\sqrt{\frac{3}{2}} \\ 0 & 0 & \sqrt{\frac{4}{3}} \end{pmatrix}.$$

Thus, Q and R have both been found.

(b) We can use the same algorithm as in **(a)**. But we can also save ourselves the hassle and read off the answer from the problem: Namely, set $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$.
Then, the matrix A itself is upper-triangular. Hence, setting $Q = I_3$ and $R = A$ yields a QR decomposition $A = QR$ of A.

(c) Once again, the answer can be read off from the problem: Namely, set $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$. Then, the matrix A has orthonormal columns (in fact, its columns are distinct standard basis vectors). Hence, setting $Q = A$ and $R = I_3$ yields a QR decomposition $A = QR$ of A. $\qquad\qquad\square$

Exercise 4. (a) Apply the Gram-Schmidt process to the two vectors

$$w_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \qquad w_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

in $\mathbb{R}^3$.

(b) Let U be the subspace of $\mathbb{R}^3$ spanned by w_1, w_2. Find the projection u of the vector $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ on the subspace U.

Solution. **(a)** We apply the Gram-Schmidt process to w_1, w_2:

1. At the first step, we set $u_1 = w_1$. Thus,

$$u_1 = w_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

2. At the second step, we set $u_2 = w_2 - \lambda_{2,1} u_1$, where $\lambda_{2,1} = \dfrac{\langle w_2, u_1 \rangle}{\langle u_1, u_1 \rangle}$. We compute these explicitly:

$$\lambda_{2,1} = \frac{\langle w_2, u_1 \rangle}{\langle u_1, u_1 \rangle} = \frac{\left\langle \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \right\rangle}{\left\langle \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \right\rangle} = \frac{1}{5}$$

and thus

$$u_2 = w_2 - \lambda_{2,1} u_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} - \frac{1}{5} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\dfrac{2}{5} \\ \dfrac{4}{5} \\ 2 \end{pmatrix}.$$

(b) We have just obtained an orthogonal basis (u_1, u_2) of U. Recall the general formula that says that if b is a vector in $\mathbb{R}^n$, and if $(u_1, u_2, \ldots, u_k)$ is an orthogonal basis of a subspace U of $\mathbb{R}^n$, then the projection of b on U is

$$\frac{\langle b, u_1 \rangle}{\|u_1\|^2} u_1 + \frac{\langle b, u_2 \rangle}{\|u_2\|^2} u_2 + \cdots + \frac{\langle b, u_k \rangle}{\|u_k\|^2} u_k.$$

Applying this to our situation (in which $n = 3$ and $k = 2$), we conclude that the projection of b on U is

$$\frac{\langle b, u_1 \rangle}{\|u_1\|^2} u_1 + \frac{\langle b, u_2 \rangle}{\|u_2\|^2} u_2 = \frac{\left\langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \right\rangle}{\left\| \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \right\|^2} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \frac{\left\langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -\frac{2}{5} \\ \frac{4}{5} \\ 2 \end{pmatrix} \right\rangle}{\left\| \begin{pmatrix} -\frac{2}{5} \\ \frac{4}{5} \\ 2 \end{pmatrix} \right\|^2} \begin{pmatrix} -\frac{2}{5} \\ \frac{4}{5} \\ 2 \end{pmatrix}$$

$$= \frac{3}{5} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \frac{\left(\frac{12}{5} \right)}{\left(\frac{24}{5} \right)} \begin{pmatrix} -\frac{2}{5} \\ \frac{4}{5} \\ 2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

[In hindsight, this is obvious! The projection of b on U is b itself, since b belongs to U to begin with. If you notice this early on, you save yourself the whole messy computation.] □

Exercise 5. Find the least-squares solution to the equation $Ax = b$, where $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$.

Solution. We follow the usual method: We set $K = A^T A$ and $f = A^T b$, and then $x = K^{-1} f$. This works because the columns of A are linearly independent.

Here are the computations:

$$K = A^T A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}^T \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix}$$

and

$$f = A^T b = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}^T \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 9 \\ 8 \\ 7 \end{pmatrix}.$$

Thus,

$$x = K^{-1} f = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 9 \\ 8 \\ 7 \end{pmatrix} = \begin{pmatrix} \dfrac{15}{7} \\ \dfrac{8}{7} \\ \dfrac{1}{7} \end{pmatrix}.$$

$\square$

Exercise 6. Let $p \in \mathbb{N}$. Find the least-squares solution $x \in \mathbb{R}^2$ to the equation $Ax = b$, where $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ -1 \\ 0 \end{pmatrix}$. (The matrix A has $p + 2$ rows and 2 columns, and the column vector b has size $p + 2$. All entries of A are 1's except for the last two entries of the second column. All entries of b are 1 except for the last two entries.)

(For example, if $p = 3$, then $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ 0 \end{pmatrix}$, and the least-squares solution is $\begin{pmatrix} \dfrac{9}{10} \\ -\dfrac{1}{2} \end{pmatrix}$.)

[Feel free to check your result visually: This exercise is a data-fitting problem,

where you are trying to fit a line $y = \alpha t + \beta$ through the $p + 2$ points

$$\underbrace{(1,1),\ (1,1),\ \ldots,\ (1,1)}_{p \text{ times}},\ (2,-1),\ (0,0).$$

Thus, the least-squares solution $x = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ should lead to a line $y = \alpha + \beta t$ that comes relatively close to all these points, but gets pulled closer and closer to $(1,1)$ when p grows (because with growing p, the point $(1,1)$ gets repeated more often and thus "pulls more weight").]

Solution. Again, we follow the standard method: We set $K = A^T A$ and $f = A^T b$, and then $x = K^{-1} f$. This works because the columns of A are linearly independent.
Here are the computations:

$$K = A^T A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{pmatrix}^T \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p+2 & p+2 \\ p+2 & p+4 \end{pmatrix}$$

and

$$f = A^T b = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{pmatrix}^T \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} p-1 \\ p-2 \end{pmatrix}.$$

Thus,

$$x = K^{-1} f = \begin{pmatrix} p+2 & p+2 \\ p+2 & p+4 \end{pmatrix}^{-1} \begin{pmatrix} p-1 \\ p-2 \end{pmatrix} = \begin{pmatrix} \dfrac{3p}{2(p+2)} \\ -\dfrac{1}{2} \end{pmatrix}.$$

$\square$