Further Mathematical Methods (Linear Algebra) 2002

Solutions for Problem Sheet 7

These are the solutions to Problem Sheet 7. I think that they are probably a bit too detailed, but it is nice to see how different methods are related in some of the questions.

1. There are [at least] three ways of finding the orthogonal complement of S, the subspace of $\mathbb{R}^3$ spanned by the vectors $[0, 0, -1]^t$ and $[1, 2, 3]^t$. Each of these methods, to a greater or lesser extent, relies on the fact that this subspace, spanned by these two linearly independent vectors, is represented by a plane (through the origin) in $\mathbb{R}^3$. This, in turn, means that the orthogonal complement, $S^\perp$ (which is itself a vector space), will be represented by the line through the origin that is perpendicular to this plane, i.e. it is a line in the direction of the *normal* to the plane. The moral is, find the normal and you find a basis for $S^\perp$! For completeness, I will consider all three methods as each of them is informative in some way:

Method 1: As you should all know, the quickest way to find the normal to a plane is to take the vector product of two [linearly independent] vectors in the plane.[1] So, as $[0, 0, -1]^t$ and $[1, 2, 3]$ are two such vectors we find

$$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 0 & 0 & -1 \\ 1 & 2 & 3 \end{vmatrix} = 2\mathbf{e}_1 - \mathbf{e}_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix},$$

where we have expanded along the second row of the determinant (as it contains the most zeros!) and the vectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ represent the standard basis of $\mathbb{R}^3$. Thus, the required normal is the vector $[2, -1, 0]^t$ and so $S^\perp = \text{Lin}\{[2, -1, 0]^t\}$.

Method 2: Of course, another way to find the normal is to use the inner product to find a vector that is perpendicular to two [linearly independent] vectors that lie in the plane. So, as $[0, 0, -1]^t$ and $[1, 2, 3]$ are two such vectors, we require a vector $\mathbf{x} = [x_1, x_2, x_3]^t$ where

$$\left\langle \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \right\rangle = -x_3 = 0 \quad \text{and} \quad \left\langle \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\rangle = x_1 + 2x_2 + 3x_3 = 0.$$

Now, this set of simultaneous equations has a general solution given by $x_3 = 0$ and $x_1 = -2x_2$, which means that the normal is a vector of the form $x_2[-2, 1, 0]^t$ for $x_2 \in \mathbb{R}$. Thus, we can write $S^\perp = \text{Lin}\{[2, -1, 0]^t\}$, as before.

Method 3: Alternatively, you can use the Gram-Schmidt procedure to find the normal. But, unfortunately, this method requires that we find an orthonormal basis for S before we can go on and find $S^\perp$. So, doing this, the first step tells us that

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \Longrightarrow \mathbf{e}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix},$$

where we have used the fact that $\|\mathbf{v}_1\|^2 = \langle \mathbf{v}_1, \mathbf{v}_1 \rangle = (-1)(-1) = 1$. Then, the second step tells us to calculate the vector $\mathbf{u}_2$ where

$$\mathbf{u}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{e}_1 \rangle \mathbf{e}_1 \Longrightarrow \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \left\langle \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \right\rangle \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}.$$

[1] As the vector [or cross] product, $\mathbf{u} \times \mathbf{v}$, of two vectors $\mathbf{u}$ and $\mathbf{v}$ gives another vector that is perpendicular to both $\mathbf{u}$ and $\mathbf{v}$.

Calculating the inner product, we find that

$$\mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - (-3) \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix},$$

and so,

$$\mathbf{e}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix},$$

where we have used the fact that $\|\mathbf{u}_2\|^2 = \langle \mathbf{u}_2, \mathbf{u}_2 \rangle = 1(1) + 2(2) = 5$. Thus, the set of vectors $\{\mathbf{e}_1, \mathbf{e}_2\}$ is an orthonormal basis for the subspace under consideration.

Consequently, to find a normal vector, we take any vector that does not lie in the plane,[2] for instance, the vector $[1, 0, 0]^t$, and then apply the Gram-Schmidt procedure to the expanded set of vectors $\{[0, 0, -1]^t, [1, 2, 3]^t, [1, 0, 0]^t\}$. This means that we have to do a third step, i.e. we have to calculate a vector $\mathbf{u}_3$ such that $\mathbf{u}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{e}_1 \rangle \mathbf{e}_1 - \langle \mathbf{v}_3, \mathbf{e}_2 \rangle \mathbf{e}_2$, which gives

$$\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \right\rangle \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} - \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right\rangle \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

Calculating the inner products, we find that

$$\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - 0 \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} - \frac{1}{\sqrt{5}} \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \frac{2}{5} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix},$$

and so,

$$\mathbf{e}_3 = \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix},$$

where we have used the fact that $\|\mathbf{u}_3\|^2 = \langle \mathbf{u}_3, \mathbf{u}_3 \rangle = 2(2) + 1(1) = 5$. Thus, our normal vector is $[2, -1, 0]^t$ and so we can write $S^\perp = \mathrm{Lin}\{[2, -1, 0]^t\}$, as before.

Thus, we have found that $S^\perp = \mathrm{Lin}\{[2, -1, 0]^t\}$. Geometrically, we have established that S is a plane through the origin in $\mathbb{R}^3$.[3] Further, we have seen that $S^\perp$ is a line through the origin in the direction $[2, -1, 0]^t$.[4] Obviously, this line is perpendicular to the plane.[5]

If you are a fanatic, and you have been paying attention in lectures, there is a fourth method that you may wish to consider. This method doesn't really depend on the geometry of the situation and instead uses the fact that $N(\mathsf{A}^t) = R(\mathsf{A})^\perp$. It runs as follows:

[2]That is, a vector that is not linearly dependent on $\mathbf{e}_1$ and $\mathbf{e}_2$.

[3]Incidentally, the normal that we have calculated gives us the coefficients for the x, y and z terms in the Cartesian equation of this plane. We also know that it is a plane through the origin as S is a subspace of $\mathbb{R}^3$. Thus, the Cartesian equation of the plane representing S is $2x - y = 0$.

[4]Clearly, this means that the vector equation of this line is $\mathbf{r} = \lambda[2, -1, 0]^t$ where $\lambda \in \mathbb{R}$. Thus, eliminating λ from the three corresponding 'component equations', i.e. $x = 2\lambda$, $y = -\lambda$ and $z = 0$, we find that the Cartesian equation of this line is [the intersection of the two planes] given by $x = -2y$ and $z = 0$.

[5]At this point, a word of warning is probably in order. Although Methods 1 and 2 are quicker than Method 3 in this case, within a general vector space Methods 1 and 2 may not be tenable. For instance, we have not defined the vector product in $\mathbb{C}^n$ (see Question 8) and it is not defined in $\mathbb{F}^{\mathbb{R}}$, thus ruling out Method 1 when we work in those vector spaces. Further, Method 2 is not viable when working in $\mathbb{F}^{\mathbb{R}}$, and great care must be taken when using it in $\mathbb{C}^n$ (again, see Question 8). However, Method 3 will work in all of the vector spaces that we will consider and so it is worth knowing how to use it!

Method 4: As the range of a matrix is just its column space, we can form a matrix such that $S = R(\mathsf{A}) = CS(\mathsf{A})$, i.e.

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

Now, we want to calculate $S^\perp$ which, by construction, is $R(\mathsf{A})^\perp$ and we know that $R(\mathsf{A})^\perp = N(\mathsf{A}^t)$. So, all we need to do is find the null-space of A^t, i.e. the set of vectors which satisfy the matrix equation $\mathsf{A}^t\mathbf{x} = \mathbf{0}$, i.e.

$$\begin{bmatrix} 0 & 0 & -1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Longrightarrow \left.\begin{array}{r} z = 0 \\ x + 2y + 3z = 0 \end{array}\right\} \Longrightarrow \left.\begin{array}{r} z = 0 \\ x = -2y \end{array}\right\}$$

Thus, the required vectors are of the form $y[-2, 1, 0]^t$ for $y \in \mathbb{R}$, and so $N(\mathsf{A}^t) = S^\perp = \text{Lin}\{[2, -1, 0]^t\}$, as before.[6]

2. Given the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & -2 \\ -3 & 6 \end{bmatrix} \text{ and its transpose, } \mathsf{A}^t = \begin{bmatrix} 1 & -3 \\ -2 & 6 \end{bmatrix},$$

we can see that as $R(\mathsf{A}) = CS(\mathsf{A})$, and the vectors that form the columns of A and A^t are linearly dependent, the required ranges are:

- $R(\mathsf{A}) = \text{Lin}\{[1, -3]^t\}$, and
- $R(\mathsf{A}^t) = \text{Lin}\{[1, -2]^t\}$,

whereas, the required null-spaces are given by vectors $\mathbf{x}$ such that $\mathsf{A}\mathbf{x} = \mathbf{0}$, i.e.

- $x - 2y = 0$ for $\mathsf{A} \Longrightarrow \mathbf{x} = s[2, 1]^t \in N(\mathsf{A})$, i.e. $N(\mathsf{A}) = \text{Lin}\{[2, 1]^t\}$, and
- $x - 3y = 0$ for $\mathsf{A}^t \Longrightarrow \mathbf{x} = s[3, 1]^t \in N(\mathsf{A}^t)$, i.e. $N(\mathsf{A}^t) = \text{Lin}\{[3, 1]^t\}$.

Further, $R(\mathsf{A}^t) = N(\mathsf{A})^\perp$ as

$$\left\langle \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\rangle = 0,$$

tells us that all of the vectors in $R(\mathsf{A}^t)$ are orthogonal to all of the vectors in $N(\mathsf{A})$. Similarly, $R(\mathsf{A})^\perp = N(\mathsf{A}^t)$ as

$$\left\langle \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\rangle = 0,$$

tells us that all of the vectors in $N(\mathsf{A}^t)$ are orthogonal to all of the vectors in $R(\mathsf{A})$. Geometrically, all of these subsets of $\mathbb{R}^2$ represent lines through the origin (and hence, they are all subspaces of $\mathbb{R}^2$, as one would expect). Indeed, the line $2x + y = 0$ that represents $R(\mathsf{A}^t)$ is perpendicular to the line $x - 2y = 0$ that represents $N(\mathsf{A})$, and similarly, the line $x - 3y = 0$ that represents $N(\mathsf{A}^t)$ is perpendicular to the line $3x + y = 0$ that represents $R(\mathsf{A})$.[7]

3. We are given a finite set of vectors, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, in an inner product space V. The subspace of V spanned by these vectors is $S = \text{Lin}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. We are asked to show that $\mathbf{x} \in S^\perp$ iff $\mathbf{x}$ is orthogonal to every vector in the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. To do this, we have to establish the result in both 'directions', i.e.

[6]Notice the similarity between this method and Method 2!

[7]Incidentally, recall that two lines are perpendicular if the product of their gradients (i.e. the m in $y = mx + c$) is -1 and that this is true of both of the pairs of lines considered above as $(-2) \times (1/2) = -1$ and $(-3) \times (1/3) = -1$. (Further [and trivially] the $y-$intercept, c is zero in all four cases as the lines pass through the origin!)

- **LTR:** Suppose that $\mathbf{x} \in S^{\perp}$. Clearly, as $\mathbf{x}_i \in S$ for $1 \leq i \leq n$, by the definition of $S^{\perp}$, $\mathbf{x}$ must be orthogonal to every vector in the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ (as required).
- **RTL:** Suppose that $\mathbf{x}$ is orthogonal to every vector in the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, that is, $\langle \mathbf{x}, \mathbf{x}_i \rangle = 0$ for $1 \leq i \leq n$. Now consider a general vector $\mathbf{s} \in S$, say

  $$\mathbf{s} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i,$$

  for some scalars α_i and, taking the inner product of $\mathbf{x}$ and $\mathbf{s}$ we see that

  $$\langle \mathbf{x}, \mathbf{s} \rangle = \left\langle \mathbf{x}, \sum_{i=1}^{n} \alpha_i \mathbf{x}_i \right\rangle = \sum_{i=1}^{n} \alpha_i \langle \mathbf{x}, \mathbf{x}_i \rangle = 0.$$

  Thus, $\mathbf{x}$ is orthogonal to a general vector in S (and hence every vector in S) which means that $\mathbf{x} \in S^{\perp}$ (as required).

and so we have established the result.

4. We are asked to prove the following theorems. Firstly,

> **Theorem:** If V and W are subsets of a vector space such that $V \subseteq W$, then $W^{\perp} \subseteq V^{\perp}$.
>
> **Proof:** Suppose that V and W are subsets of a vector space such that $V \subseteq W$.[8] To show that $W^{\perp} \subseteq V^{\perp}$, we need to establish that all of the vectors in $W^{\perp}$ are also in $V^{\perp}$. We do this by considering an arbitrary vector $\mathbf{x}$ in the set $W^{\perp}$ and showing that it is also in the set $V^{\perp}$, i.e.
>
> $$\begin{aligned} \mathbf{x} \in W^{\perp} &\implies \langle \mathbf{y}, \mathbf{x} \rangle = 0 \quad \forall \mathbf{y} \in W && \text{:by the definition of } W^{\perp}. \\ &\implies \langle \mathbf{y}, \mathbf{x} \rangle = 0 \quad \forall \mathbf{y} \in V && \text{:as } V \subseteq W. \\ &\implies \mathbf{x} \in V^{\perp} && \text{:by the definition of } V^{\perp}. \end{aligned}$$
>
> Thus, $W^{\perp} \subseteq V^{\perp}$ as required.

Secondly,

> **Theorem:** If A is any $m \times n$ matrix and B is any $n \times m$ matrix where $n < m$, then AB is singular.
>
> **Proof:** Let us suppose that A is any $m \times n$ matrix and that B is any $n \times m$ matrix. We want to show that if $n < m$, then the matrix given by AB is singular. To do this, we note that AB will be an $m \times m$ matrix and so it is singular (i.e. not invertible) if it does *not* have *full* rank (i.e. if its rank is less than m). So, if we can establish that $\rho(\mathsf{AB}) < m$ when $n < m$ this is equivalent to showing that AB is singular. To do this, note that
>
> $$\begin{aligned} \rho(\mathsf{AB}) &\leq \rho(\mathsf{A}) && \text{:from the lectures.} \\ &\leq \min\{m, n\} && \text{:as } \mathsf{A} \text{ is } m \times n. \\ &= n < m && \text{:given presupposition.} \end{aligned}$$
>
> Thus, $\rho(\mathsf{AB}) < m$ and so the matrix AB is singular, as required.

[8] A common mistake in this question is to assume that V and W are subspaces of the vector space in question. (For example, by introducing bases *et cetera*.) But, this is not general enough! Obviously, not all sub*sets* of a vector space are sub*spaces*.

Thirdly,

> **Theorem:** If B is an invertible square matrix and the matrix product AB is defined, then the rank of AB equals the rank of A.
>
> **Proof:** As B is a square matrix, we can assume that it is $n \times n$ and as we require the matrix product AB to be defined, we can assume that A is an $m \times n$ matrix. Further, the matrix B is invertible and so it has full rank, i.e. $\rho(\mathsf{B}) = n$. We now have to establish that the matrices AB and A have the same rank, i.e. $\rho(\mathsf{AB}) = \rho(\mathsf{A})$. To do this, we note [from the lectures] that
>
> $$\rho(\mathsf{A}) + \rho(\mathsf{B}) - n \leq \rho(\mathsf{AB}) \leq \rho(\mathsf{A}).$$
>
> However, $\rho(\mathsf{B}) = n$, and so this inequality is just
>
> $$\rho(\mathsf{A}) \leq \rho(\mathsf{AB}) \leq \rho(\mathsf{A}),$$
>
> and this can only hold if $\rho(\mathsf{AB}) = \rho(\mathsf{A})$, as required.

5. We are given the subspaces Y and Z of $\mathbb{R}^4$ where

$$Y = \text{Lin}\left\{[1,0,1,0]^t, [0,0,0,1]^t\right\} \quad \text{and} \quad Z = \text{Lin}\left\{[0,1,0,0]^t, [1,0,1,-1]^t\right\} .$$

Recall that a sum is direct if every vector in $Y + Z$ can be written as $\mathbf{y} + \mathbf{z}$ *uniquely* in terms of vectors $\mathbf{y} \in Y$ and $\mathbf{z} \in Z$. So, we can see that the sum $Y + Z$ is not direct as we can write the vector $[1,0,1,-1]^t \in Y + Z$ in *two* ways, i.e.

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\text{in } Y} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix}}_{\text{in } Z} \quad \text{and} \quad \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{\text{in } Y} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\text{in } Z} .$$

Alternatively, using the fact that

$$X = Y \oplus Z \quad \text{iff} \quad X = Y + Z \quad \text{and} \quad Y \cap Z = \{\mathbf{0}\},$$

we can see that the sum is not direct because the intersection of the subspaces Y and Z is not $\{\mathbf{0}\}$. (Indeed, we can see from the analysis above that the vector $[1,0,1,-1]^t$ lies in $Y \cap Z$ too.)

Lastly, it should be clear that

$$Y + Z = \text{Lin}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right\},$$

and so, $Y + Z$ is spanned by the set of vectors

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

Thus, removing the linear dependence from this set of vectors (by removing the vector $[1,0,1,-1]^t$, say) it should be clear that

$$\left\{[1,0,1,0]^t, [0,0,0,1]^t, [0,1,0,0]^t\right\},$$

is a basis for the subspace $Y + Z$ of $\mathbb{R}^4$.

6. We are asked to prove the following theorems. Firstly,

Theorem: If Y and Z are subspaces of a vector space V, then $Y + Z$ is also a subspace of V. Further, $Y + Z$ is the smallest subspace of V containing $Y \cup Z$ (in the sense that every other subspace of V that contains $Y \cup Z$ must contain $Y + Z$).

Proof: As Y and Z are subspaces of the vector space V, they will both be closed under vector addition and scalar multiplication. Indeed, to show that $Y + Z$ is a subspace of V, we need to establish that it is also closed under these operations. To do this, we use the definition of $Y + Z$, i.e.

$$Y + Z = \{\mathbf{y} + \mathbf{z} \,|\, \mathbf{y} \in Y \text{ and } \mathbf{z} \in Z\},$$

and consider two general vectors in $Y + Z$, say

$$\mathbf{x}_1 = \mathbf{y}_1 + \mathbf{z}_1 \quad \text{where} \quad \mathbf{y}_1 \in Y \quad \text{and} \quad \mathbf{z}_1 \in Z,$$

$$\mathbf{x}_2 = \mathbf{y}_2 + \mathbf{z}_2 \quad \text{where} \quad \mathbf{y}_1 \in Y \quad \text{and} \quad \mathbf{z}_2 \in Z,$$

Now, we note that:

- $Y + Z$ is closed under vector addition since:

$$\mathbf{x}_1 + \mathbf{x}_2 = (\mathbf{y}_1 + \mathbf{z}_1) + (\mathbf{y}_2 + \mathbf{z}_2) = (\mathbf{y}_1 + \mathbf{y}_2) + (\mathbf{z}_1 + \mathbf{z}_2),$$

  which is a vector in $Y + Z$ since $\mathbf{y}_1 + \mathbf{y}_2 \in Y$ and $\mathbf{z}_1 + \mathbf{z}_2 \in Z$ (as Y and Z are closed under vector addition).
- $Y + Z$ is closed under scalar multiplication since:

$$\alpha\mathbf{x}_1 = \alpha(\mathbf{y}_1 + \mathbf{z}_1) = \alpha\mathbf{y}_1 + \alpha\mathbf{z}_1,$$

  which is a vector in $Y + Z$ since $\alpha\mathbf{y}_1 \in Y$ and $\alpha\mathbf{z}_1 \in Z$ (as Y and Z are closed under scalar multiplication).

Consequently, as $Y + Z$ is closed under vector addition and scalar multiplication it is a subspace of V (as required).

Further, we are asked to show that $Y + Z$ is the *smallest* subspace of V that contains $Y \cup Z$.[9] We do this in two parts:

- Firstly, we show that $Y \cup Z \subseteq Y + Z$. This is the case since, by definition, $Y + Z$ contains all of the vectors in Y[10] and all of the vectors in Z,[11] and as such it must contain all of the vectors in $Y \cup Z$.
- Secondly, we show that $Y + Z$ is the *smallest* subspace of V for which this holds. To do this, we consider *any other* subspace, $W \subseteq V$ such that $Y \cup Z \subseteq W$, and we need to show that this implies that $Y + Z \subseteq W$ as well. So, as $Y \cup Z \subseteq W$, we have $Y \subseteq W$ and $Z \subseteq W$, i.e.

$$\forall \mathbf{y} \in Y, \ \mathbf{y} \in W \quad \text{and} \quad \forall \mathbf{z} \in Z, \ \mathbf{z} \in W.$$

  Thus, as W is a subspace of V, it is closed under vector addition, and so $\mathbf{y} + \mathbf{z} \in W$ for all $\mathbf{y} \in Y$ and $\mathbf{z} \in Z$. That is, $Y + Z \subseteq W$ too.

Consequently, $Y + Z$ is the smallest subspace of V that contains $Y \cup Z$ (as required).

Secondly,

[9] Recall that in Problem Sheet 1 we discovered that the union of two subspaces is generally not a subspace itself.
[10] Since, for any vector $\mathbf{y} \in Y$, we have $\mathbf{y} = \mathbf{y} + \mathbf{0} \in Y + Z$ as $\mathbf{0} \in Z$.
[11] Since, for any vector $\mathbf{z} \in Z$, we have $\mathbf{z} = \mathbf{0} + \mathbf{z} \in Y + Z$ as $\mathbf{0} \in Y$.

Theorem: If the set of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}$ is a basis of the vector space V, then

$$V = \text{Lin}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \oplus \text{Lin}\{\mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}.$$

Proof: Consider the subspaces of V given by

$$Y = \text{Lin}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \text{ and } Z = \text{Lin}\{\mathbf{x}_{k+1}, \dots, \mathbf{x}_n\},$$

we need to show that $V = Y \oplus Z$. To do this, we use the theorem that was proved in the lectures, namely:

$$V = Y \oplus Z \text{ iff } V = Y + Z \text{ and } Y \cap Z = \{\mathbf{0}\},$$

and so we have to establish that $V = Y + Z$ and $Y \cap Z = \{\mathbf{0}\}$. We shall do each of these in turn:

- Firstly, it should be clear that as $\{\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}$ is a basis of V,

$$\begin{aligned} V &= \text{Lin}\{\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{x}_{k+1}, \dots, \mathbf{x}_n\} \\ &= \{\underbrace{\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k}_{\text{in } Y} + \underbrace{\alpha_{k+1}\mathbf{x}_{k+1} + \cdots + \alpha_n\mathbf{x}_n}_{\text{in } Z} \,|\, \alpha_1, \dots, \alpha_n \text{ are scalars}\} \\ &= \{\mathbf{y} + \mathbf{z} \,|\, \mathbf{y} \in \text{Lin}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \text{ and } \mathbf{z} \in \text{Lin}\{\mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}\} \\ &= \{\mathbf{y} + \mathbf{z} \,|\, \mathbf{y} \in Y \text{ and } \mathbf{z} \in Z\} \\ \therefore \quad V &= Y + Z. \end{aligned}$$

  (as required).

- Secondly, suppose that $\mathbf{u}$ is any vector in $Y \cap Z$ (that is, $\mathbf{u} \in Y$ and $\mathbf{u} \in Z$) and so we can write $\mathbf{u}$ as

$$\mathbf{u} = \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \in Y \text{ and } \mathbf{u} = \alpha_{k+1}\mathbf{x}_{k+1} + \cdots + \alpha_n\mathbf{x}_n \in Z,$$

  which on equating and rearranging gives

$$\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k - \alpha_{k+1}\mathbf{x}_{k+1} - \cdots - \alpha_n\mathbf{x}_n = \mathbf{0}.$$

  But, the vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}$ form a basis of V and so, due to their linear independence, we get

$$\alpha_1 = \cdots = \alpha_k = \alpha_{k+1} = \cdots = \alpha_n = 0,$$

  as the only solution, i.e. $\mathbf{u}$ must be the null vector (i.e. $\mathbf{0}$). Thus, $Y \cap Z = \{\mathbf{0}\}$. Consequently, as $V = Y + Z$ and $Y \cap Z = \{\mathbf{0}\}$ we can conclude that $V = Y \oplus Z$ (as required).

Thus, we can see that in this case,

$$V = \text{Lin}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \oplus \text{Lin}\{\mathbf{x}_{k+1}, \dots, \mathbf{x}_n\},$$

as required.

Thirdly,

Theorem: If Y and Z are subspaces of a vector space V such that $V = Y \oplus Z$, then $\dim(V) = \dim(Y \oplus Z) = \dim(Y) + \dim(Z)$.[12]

---

[12] Notice that the previous part of this question is saying that

If $\{\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}$ is a basis of the vector space V, then $V = \text{Lin}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \oplus \text{Lin}\{\mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}$,

whereas this part is *effectively* saying:

If $V = \text{Lin}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \oplus \text{Lin}\{\mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}$, then $\{\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}$ is a basis of the vector space V,

provided that $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ and $\{\mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}$ are bases of Y and Z respectively. (See the proof!) Consequently, we cannot use the previous part here as it is the *converse* of what we need!

Proof: Y and Z are subspaces of the vector space V, and so we can find a basis $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$ of Y and a basis $\{\mathbf{z}_1, \ldots, \mathbf{z}_s\}$ of Z, i.e. every vector in Y (or Z) can be written uniquely as a linear combination of the basis vectors of Y (or Z).[13] Now, as $V = Y \oplus Z$, we can write every vector $\mathbf{v} \in V$ uniquely as the sum of a vector in Y and a vector in Z. Thus, every vector $\mathbf{v} \in V$ can be written as a unique linear combination of the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_r$ and $\mathbf{z}_1, \ldots, \mathbf{z}_s$, i.e. the set $\{\mathbf{y}_1, \ldots, \mathbf{y}_r, \mathbf{z}_1, \ldots, \mathbf{z}_s\}$ is a basis of V. Consequently, we can see that

$$\dim(V) = r + s = \dim(Y) + \dim(Z),$$

where as $V = Y \oplus Z$, $\dim(V) = \dim(Y \oplus Z)$ too (as required).

Other Problems.

As always, the solutions to these questions will be more detailed as they were not covered in class. You should all look at Question 7 as it illustrates the fact that the null vector is orthogonal to all vectors (even though it is perpendicular to none[14]).

7. Given the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and its transpose, } \mathsf{A}^t = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix},$$

we can see that as $R(\mathsf{A}) = CS(\mathsf{A})$, and the vectors that form the columns of A and A^t are linearly independent, the required ranges are:

- $R(\mathsf{A}) = \text{Lin}\{[1,3]^t, [2,4]^t\}$, and
- $R(\mathsf{A}^t) = \text{Lin}\{[1,2]^t, [3,4]^t\}$,

whereas, the required null-spaces are given by vectors $\mathbf{x}$ such that $\mathsf{A}\mathbf{x} = \mathbf{0}$, i.e.

- $x + 2y = 0$ and $3x + 4y = 0$ for $\mathsf{A} \implies$ a unique trivial solution, i.e. $N(\mathsf{A}) = \text{Lin}\{\mathbf{0}\}$, and
- $x + 3y = 0$ and $2x + 4y = 0$ for $\mathsf{A}^t \implies$ a unique trivial solution, i.e. $N(\mathsf{A}^t) = \text{Lin}\{\mathbf{0}\}$.

Further, $R(\mathsf{A}^t) = N(\mathsf{A})^\perp$ as

$$\left\langle \alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \beta \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\rangle = 0,$$

tells us that all of the vectors in $R(\mathsf{A}^t)$ are orthogonal to the vector in $N(\mathsf{A})$. Similarly, $R(\mathsf{A})^\perp = N(\mathsf{A}^t)$ as

$$\left\langle \alpha \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \beta \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\rangle = 0,$$

tells us that the vector in $N(\mathsf{A}^t)$ is orthogonal to all of the vectors in $R(\mathsf{A})$. Geometrically, the ranges $R(\mathsf{A})$ and $R(\mathsf{A}^t)$ represent $\mathbb{R}^2$ (and hence they are trivially subspaces of $\mathbb{R}^2$, as one would expect). Whereas, the vector in the null-spaces $N(\mathsf{A})$ and $N(\mathsf{A}^t)$ represents the origin in $\mathbb{R}^2$ (and hence they are both the trivial subspace of $\mathbb{R}^2$). Notice that in this case, the null-spaces are orthogonal to the ranges, even though they are not perpendicular to one another (as it is meaningless to talk about the angle between a point and a line).

8. To find an orthonormal basis for the subspace of $\mathbb{C}^3$ spanned by the vectors $[i, 0, 1]^t$ and $[1, 1, 0]^t$ we use the Gram-Schmidt procedure. But, before we start, we recall that the inner product in $\mathbb{C}^3$ is given by

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1^* + x_2 y_2^* + x_3 y_3^*,$$

[13] The connection between bases and unique linear combinations is established in Theorem 2.12 from the handout for Lecture 2.

[14] Obviously there is *no* angle between a point and a line!

where we pay particular attention to the fact that the components of the second vector in the inner product (i.e. $\mathbf{y}$ in this case) are complex conjugated in the calculation of the inner product. Now, the first step of the Gram-Schmidt procedure tells us that

$$\mathbf{v}_1 = \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} \Longrightarrow \mathbf{e}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix},$$

where we have used the fact that $\|\mathbf{v}_1\|^2 = \langle \mathbf{v}_1, \mathbf{v}_1 \rangle = i(-i) + 1(1) = 2$. Then, the second step tells us to calculate the vector $\mathbf{u}_2$ where

$$\mathbf{u}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{e}_1 \rangle \mathbf{e}_1 \Longrightarrow \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \left\langle \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} \right\rangle \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix}.$$

Calculating the inner product, we find that

$$\mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{2}} \frac{-i}{\sqrt{2}} \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -1 \\ 0 \\ i \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 2 \\ i \end{bmatrix},$$

and so,

$$\mathbf{e}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ i \end{bmatrix},$$

where we have used the fact that $\|\mathbf{u}_2\|^2 = \langle \mathbf{u}_2, \mathbf{u}_2 \rangle = 1(1) + 2(2) + i(-i) = 6$. Thus, the set of vectors $\{\mathbf{e}_1, \mathbf{e}_2\}$ is an orthonormal basis for the subspace under consideration.

To find the orthogonal complement of this subspace we can use the analogue of Method 3 from Question 1. That is, we take any vector that does not lie in the linear span of $\mathbf{e}_1$ and $\mathbf{e}_2$,[15] for instance, the vector $[1, 0, 0]^t$, and then apply the Gram-Schmidt procedure to the expanded set of vectors $\{[i, 0, 1]^t, [1, 1, 0]^t, [1, 0, 0]^t\}$. Thus, we have to do a third step, i.e. we have to calculate a vector $\mathbf{u}_3$ such that $\mathbf{u}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{e}_1 \rangle \mathbf{e}_1 - \langle \mathbf{v}_3, \mathbf{e}_2 \rangle \mathbf{e}_2$, which gives

$$\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} \right\rangle \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} - \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ i \end{bmatrix} \right\rangle \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ i \end{bmatrix}.$$

Calculating the inner products, we find that

$$\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{2}} \frac{-i}{\sqrt{2}} \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} - \frac{1}{\sqrt{6}} \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ i \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -1 \\ 0 \\ i \end{bmatrix} - \frac{1}{6} \begin{bmatrix} 1 \\ 2 \\ i \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ -1 \\ i \end{bmatrix},$$

and so,

$$\mathbf{e}_3 = \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ i \end{bmatrix},$$

where we have used the fact that $\|\mathbf{u}_3\|^2 = \langle \mathbf{u}_3, \mathbf{u}_3 \rangle = 1(1) + (-1)(-1) + i(-i) = 3$. Thus, the vector $[1, -1, i]^t$ is orthogonal to the vectors in S, and so we conclude that $S^\perp = \text{Lin}\{[1, -1, i]^t\}$ is the orthogonal complement of S.

Aside: As promised in Footnote 5, we now examine some of the alternatives to the above method for calculating the orthogonal complement of S when we are working in $\mathbb{C}^3$. We start by considering

[15]That is, a vector that is not linearly dependent on $\mathbf{e}_1$ and $\mathbf{e}_2$.

the analogue of Method 2 from Question 1 where we seek a vector that is orthogonal to two [linearly independent] vectors in S, i.e. we seek a vector $\mathbf{x} = [x_1, x_2, x_3]^t$ such that

$$\left\langle \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} \right\rangle = -ix_1 + x_3 = 0 \quad \text{and} \quad \left\langle \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\rangle = x_1 + x_2 = 0.$$

Now, this set of simultaneous equations has a general solution given by $x_2 = -x_1$ and $x_3 = ix_1$, which means that vectors in the orthogonal complement are of the form $x_1[1, -1, i]^t$. Thus, we can write $S^\perp = \text{Lin}\{[1, -1, i]^t\}$ as above.

There is also an analogue of Method 1, which does not *prima facie* involve vector products,[16] although it does involve determinants. To see this, consider two [linearly independent] vectors in S. Clearly, all vectors $\mathbf{x} = [x_1, x_2, x_3]^t$ that are linearly dependent on these will satisfy the determinant equation

$$\begin{vmatrix} x_1 & x_2 & x_3 \\ i & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 0 \implies -x_1 + x_2 + ix_3 = 0 \implies x_1 - x_2 - ix_3 = 0,$$

and so all vectors $[x_1, x_2, x_3]^t$ that satisfy this equation will lie in S. But, this looks like an inner product, for consider

$$\left\langle \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \right\rangle = 0 \implies x_1 n_1^* + x_2 n_2^* + x_3 n_3^* = 0,$$

where the vector $[n_1, n_2, n_3]^t$ is orthogonal to the vector $[x_1, x_2, x_3]^t$. Thus, vectors of the form $[1, -1, i]^t$ (note the complex conjugation!) will be orthogonal to all of the vectors in S, i.e. $S^\perp = \text{Lin}\{[1, -1, i]^t\}$, as above.

Notice that these methods are based (to a greater or lesser extent) on the intuition that we gained when thinking about planes in Question 1. However, we have not taken our geometric notions (such as 'plane,' 'normal,' 'angle' and 'perpendicular') with us into $\mathbb{C}^3$. Consequently, you should be wary of using these methods unless you can provide some sort of justification for their use. Further, notice that the use of Method 4 is also invalid here as the central result, namely $R(\mathsf{A})^\perp = N(\mathsf{A}^t)$, was only established for *real* matrices![17] I suppose that the moral of this aside is two-fold: firstly, if in doubt when finding orthogonal complements use the Gram-Schmidt procedure, and secondly always check your answer!

9. We are asked to prove that if S is any sub*set* of $\mathbb{R}^n$, then $S \subseteq S^{\perp\perp}$. To do this, take any vector $\mathbf{x} \in S$, and observe that the definition of $S^\perp$, i.e.

$$S^\perp = \{\mathbf{y} \,|\, \langle \mathbf{y}, \mathbf{x} \rangle = 0 \;\; \forall \mathbf{x} \in S\},$$

entails that $\langle \mathbf{y}, \mathbf{x} \rangle = 0$ for all $\mathbf{y} \in S^\perp$. But, by definition, $S^{\perp\perp}$ is given by

$$S^{\perp\perp} = (S^\perp)^\perp = \{\mathbf{z} \,|\, \langle \mathbf{z}, \mathbf{y} \rangle = 0 \;\; \forall \mathbf{y} \in S^\perp\},$$

and so, clearly, $\mathbf{x} \in S^{\perp\perp}$.

Further, we have to prove that if S is a sub*space* of $\mathbb{R}^n$, then $S = S^{\perp\perp}$. As S is a sub*space* of $\mathbb{R}^n$, we can find a set of vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k\}$ that form an orthonormal basis for S. Then, using the Gram-Schmidt procedure (say) we could extend this to an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k, \mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$ of $\mathbb{R}^n$. Now, to proceed, we need to prove that

$$S^\perp = \text{Lin}\{\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_n\},$$

and to do this, as usual, we have to do two things:

[16] I am reluctant to introduce vector products in $\mathbb{C}^3$ here as they are not part of the course. However, what we do now could be interpreted in terms of such vector products in $\mathbb{C}^3$.

[17] However, I am sure that an analogue of this result for complex matrices can be found...

- Firstly, we need to show that $\text{Lin}\{\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_n\} \subseteq S^\perp$, i.e. that a general vector

$$\mathbf{x} = \sum_{i=k+1}^{n} \alpha_i \mathbf{e}_i \in \text{Lin}\{\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_n\},$$

  is also in $S^\perp$, and so we have to show that $\mathbf{x}$ is orthogonal to every vector in S. To do this, we take a general vector in S, that is

$$\mathbf{y} = \sum_{j=1}^{k} \beta_j \mathbf{e}_j \in \text{Lin}\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k\} = S,$$

  and show that $\langle \mathbf{y}, \mathbf{x} \rangle = 0$. But, this is clearly the case since

$$\left\langle \sum_{j=1}^{k} \beta_j \mathbf{e}_j, \sum_{i=k+1}^{n} \alpha_i \mathbf{e}_i \right\rangle = \sum_{j=1}^{k} \sum_{i=k+1}^{n} \beta_j \alpha_i \langle \mathbf{e}_j, \mathbf{e}_i \rangle = 0,$$

  as $\langle \mathbf{e}_j, \mathbf{e}_i \rangle = 0$ for all available values of i and j since the $\mathbf{e}$'s form an orthonormal set. Thus, $\mathbf{x} \in S^\perp$, and consequently, $\text{Lin}\{\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_n\} \subseteq S^\perp$, as required.

- Secondly, we need to show that $S^\perp \subseteq \text{Lin}\{\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_n\}$, i.e. that a general vector $\mathbf{x} \in S^\perp$ is also in $\text{Lin}\{\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_n\}$. To do this, let us suppose that such a general vector $\mathbf{x} \in S^\perp$ is

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i,$$

  and as such, it must be the case that $\langle \mathbf{x}, \mathbf{s} \rangle = 0$ for all $\mathbf{s} \in S$. In particular, for each $\mathbf{e}_j \in S$ it must be the case that

$$\left\langle \sum_{i=1}^{n} \alpha_i \mathbf{e}_i, \mathbf{e}_j \right\rangle = 0 \implies \sum_{i=1}^{n} \alpha_i \langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0 \implies \alpha_j \langle \mathbf{e}_j, \mathbf{e}_j \rangle \implies \alpha_j = 0,$$

  for $1 \leq j \leq k$, as the $\mathbf{e}$'s form an orthonormal set. Thus, we can see that $\mathbf{x} \in S^\perp$ implies that

$$\mathbf{x} = \alpha_{k+1} \mathbf{e}_{k+1} + \cdots + \alpha_n \mathbf{e}_n,$$

  and so $\mathbf{x} \in \text{Lin}\{\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_n\}$. Consequently, $S^\perp \subseteq \text{Lin}\{\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_n\}$, as required.

So, we have established that $S^\perp = \text{Lin}\{\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$.[18] From this, as $S^\perp$ is itself a subspace of $\mathbb{R}^n$, we can see that the orthogonal complement of $S^\perp$, namely $S^{\perp\perp}$, is given by

$$S^{\perp\perp} = (S^\perp)^\perp = \text{Lin}\{\mathbf{e}_1, \ldots, \mathbf{e}_k\},$$

that is, S. Thus, we have established that $S^{\perp\perp} = S$, as required.

Consequently, to show that if S is a subspace of $\mathbb{R}^n$, then $\dim(S) + \dim(S^\perp) = n$, we note that using the above notation

$$\dim(S) = k \text{ and } \dim(S^\perp) = n - k \implies \dim(S) + \dim(S^\perp) = k + (n - k) = n,$$

as required.

10. Suppose that $T : V \to V$ is a linear transformation and that X and Y are subspaces of V such that $T(X) \subseteq X$ and $T(Y) \subseteq Y$. We are asked to show that if $V = X \oplus Y$, then $T(V) = T(X) \oplus T(Y)$. We start the proof by taking the opportunity to remark on the obvious notational convention being

[18] This result tells us that if we have an orthonormal basis for a subspace of $\mathbb{R}^n$, and extend this to an orthonormal basis for $\mathbb{R}^n$, then the orthogonal complement of this subspace (which is itself a subspace) has an orthonormal basis given by the vectors required for the extension.

used in this question, namely that for a linear transformation $T : V \to V$ and a [not necessarily proper] subspace W of V,

$$T(W) = \{T(\mathbf{w}) \mid \mathbf{w} \in W \subseteq V\},$$

i.e. $T(W)$ is the set of all vectors in W after they have been mapped by T. Indeed, it should be clear that $T(W)$ is not just a sub*set* of V, as it is also a sub*space* of V. To see this, note that $T(W)$ is just the range of T when the domain of the transformation is restricted to vectors in W, i.e. taking $T : V \to V$ to be a linear transformation and W to be a subspace of V we can write the range of T when just considering vectors in $W \subseteq V$ as

$$R_W(T) = \{T(\mathbf{w}) \mid \mathbf{w} \in W \subseteq V\} = T(W),$$

and we know that the range of a linear transformation for *any* vector space (and hence *any* subspace of a vector space) will itself be a subspace (in this case, of V) by Theorem 3.7 from the handout for Lecture 3.

So, for the proof proper, we start by assuming that $V = X \oplus Y$, that is, from the lectures, we know that

$$V = X + Y \quad \text{and} \quad X \cap Y = \{\mathbf{0}\}.$$

Thus, to establish that $T(V) = T(X) \oplus T(Y)$, by using the same result, we only need to establish that

$$T(V) = T(X) + T(Y) \quad \text{and} \quad T(X) \cap T(Y) = \{\mathbf{0}\}.$$

We shall show each of these in turn:

- Firstly, it should be clear that $T(V) = T(X) + T(Y)$ since

$$\begin{aligned} T(V) &= \{T(\mathbf{v}) \mid \mathbf{v} \in V\} \\ &= \{T(\mathbf{x} + \mathbf{y}) \mid \mathbf{x} \in X \text{ and } \mathbf{y} \in Y\} && \text{:as } V = X + Y. \\ &= \{T(\mathbf{x}) + T(\mathbf{y}) \mid \mathbf{x} \in X \text{ and } \mathbf{y} \in Y\} && \text{:as } T \text{ is a linear transformation.} \\ &= \{T(\mathbf{x}) \mid \mathbf{x} \in X\} + \{T(\mathbf{y}) \mid \mathbf{y} \in Y\} && \text{:definition of sum.} \\ \therefore \quad T(V) &= T(X) + T(Y). \end{aligned}$$

- Secondly, we can see that $T(X) \cap T(Y) = \{\mathbf{0}\}$ since
  - For any vector $\mathbf{z} \in T(X) \cap T(Y)$, we have $\mathbf{z} \in T(X)$ and $\mathbf{z} \in T(Y)$. So, as $T(X) \subseteq X$ and $T(Y) \subseteq Y$, it must be the case that $\mathbf{z} \in X$ and $\mathbf{z} \in Y$, i.e. $\mathbf{z} \in X \cap Y$. But, we know that $X \cap Y = \{\mathbf{0}\}$, and so $\mathbf{z} \in \{\mathbf{0}\}$. Thus, $T(X) \cap T(Y) \subseteq \{\mathbf{0}\}$.
  - As X and Y are both subspaces of V they both contain $\mathbf{0}$, and by Theorem 3.1(i) from the handout for Lecture 3, we know that $T(\mathbf{0}) = \mathbf{0}$. Thus, $\mathbf{0} \in T(X)$ and $\mathbf{0} \in T(Y)$, which implies that $\mathbf{0} \in T(X) \cap T(Y)$, i.e. $\{\mathbf{0}\} \subseteq T(X) \cap T(Y)$.[19]

Consequently, we can conclude that $T(V) = T(X) \oplus T(Y)$, as required.

11. Let us suppose that A is any real $m \times n$ matrix and that $\mathbf{b}$ is an $n \times 1$ column vector. We are asked to show that *precisely one* of the following systems has solutions:

a. $\mathsf{A}\mathbf{x} = \mathbf{b}$.

b. $\mathsf{A}^t\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^t\mathbf{b} \neq \mathbf{0}$.

where $\mathbf{0}$ is the null vector in $\mathbb{R}^n$. This can be done in two steps:

- Suppose that $\mathbf{b} \in R(\mathsf{A})$, i.e. system 1 has solutions. Now, consider *any* $\mathbf{y}$ satisfying $\mathsf{A}^t\mathbf{y} = \mathbf{0}$, clearly such a $\mathbf{y}$ is in $N(\mathsf{A}^t)$ and so as $R(\mathsf{A}) = N(\mathsf{A}^t)^\perp$, it must be the case that $\mathbf{y} \perp \mathbf{b}$, i.e. $\mathbf{y}^t\mathbf{b} = \mathbf{0}$. Thus, system 2 can have no solutions.

[19] This should be obvious! (As $T(X)$ and $T(Y)$ are subspaces of $T(V)$, they must both contain the null vector, i.e. $\mathbf{0} \in T(X) \cap T(Y)$.)

- Suppose that $\mathbf{b} \notin R(\mathsf{A})$, i.e. system 1 has no solutions. Clearly, as $R(\mathsf{A}) = N(\mathsf{A}^t)^\perp$, this implies that $\mathbf{b} \notin N(\mathsf{A}^t)^\perp$. Now, consider a non-zero vector $\mathbf{y} \in N(\mathsf{A}^t)$, that is, a $\mathbf{y} \neq \mathbf{0}$ such that $\mathsf{A}^t\mathbf{y} = \mathbf{0}$.[20] Further, $\mathbf{y}$ and $\mathbf{b}$ cannot be orthogonal because the former lies in $N(\mathsf{A}^t)$ whereas the latter does not lie in $N(\mathsf{A}^t)^\perp$, i.e. $\mathbf{y}^t\mathbf{b} \neq \mathbf{0}$. Consequently, system 2 has solutions.

Thus, *precisely one* of the two systems has solutions (as required).

Harder Problem.

The harder problem for this week is an old exam question and so, theoretically, you should be able to do it in under half an hour. However, the solution presented here will be a bit more detailed as I want to explain what is going on too.

12. Let us suppose that A is an $n \times n$ real matrix, we are asked to prove that

$$\mathbb{R}^n \supseteq R(\mathsf{A}) \supseteq R(\mathsf{A}^2) \supseteq R(\mathsf{A}^3) \supseteq \cdots$$

To do this, we start by showing that $R(\mathsf{A}) \subseteq \mathbb{R}^n$, and this is clearly the case since if we take a vector $\mathbf{x} \in R(\mathsf{A})$, then there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{x} = \mathsf{A}\mathbf{y}$. But, A is an $n \times n$ matrix and so $\mathbf{x}$ must be in $\mathbb{R}^n$ too. Thus, $R(\mathsf{A}) \subseteq \mathbb{R}^n$. To establish the rest of this result, we need to show that $R(\mathsf{A}^{k+1}) \subseteq R(\mathsf{A}^k)$ for all $k \geq 1$. So, consider any vector $\mathbf{x} \in R(\mathsf{A}^{k+1})$, that is, there is a vector $\mathbf{y}$ such that $\mathbf{x} = \mathsf{A}^{k+1}\mathbf{y}$. However, this can be written as $\mathbf{x} = \mathsf{A}^k(\mathsf{A}\mathbf{y})$, and so there is a vector, namely $\mathbf{z} = \mathsf{A}\mathbf{y}$, such that $\mathbf{x} = \mathsf{A}^k\mathbf{z}$. Thus, $\mathbf{x} \in R(\mathsf{A}^k)$ and $R(\mathsf{A}^{k+1}) \subseteq R(\mathsf{A}^k)$ for all $k \geq 1$. Consequently, taking the first result and combining it with the second for $k = 1, 2, 3, \ldots$ we get

$$\cdots \subseteq R(\mathsf{A}^3) \subseteq R(\mathsf{A}^2) \subseteq R(\mathsf{A}) \subseteq \mathbb{R}^n,$$

as required.

Further, we are asked to prove that if $R(\mathsf{A}^s) = R(\mathsf{A}^{s+1})$, then $R(\mathsf{A}^s) = R(\mathsf{A}^q)$ for all $q \geq s$. To do this, let us suppose that for some [positive integer] s, $R(\mathsf{A}^s) = R(\mathsf{A}^{s+1})$. Now, to establish that $R(\mathsf{A}^s) = R(\mathsf{A}^q)$ for all $q \geq s$, we need to show that $R(\mathsf{A}^s) \subseteq R(\mathsf{A}^q)$ (as we already know that $R(\mathsf{A}^q) \subseteq R(\mathsf{A}^s)$ from above). We shall do this by induction, where for simplicity we write $q = s + k$ (with $k \geq 0$), i.e.

- **Induction Hypothesis:** For some $k \geq 0$, if $R(\mathsf{A}^s) = R(\mathsf{A}^{s+1})$, then $R(\mathsf{A}^s) \subseteq R(\mathsf{A}^{s+k})$.

- **Basis:** For k=0, we must show that $R(\mathsf{A}^s) \subseteq R(\mathsf{A}^s)$. But, this is trivially true.

- **Induction Step:** For a general value of k, we must show that given $R(\mathsf{A}^s) = R(\mathsf{A}^{s+1})$, $R(\mathsf{A}^s) \subseteq R(\mathsf{A}^{s+k+1})$. To do this, we take a general vector $\mathbf{x} \in R(\mathsf{A}^s)$, and so by the Induction Hypothesis, $\mathbf{x} \in R(\mathsf{A}^{s+k})$ too. Thus, there exists a vector $\mathbf{y}_1$ such that $\mathbf{x} = \mathsf{A}^{s+k}\mathbf{y}_1$, i.e.

  $$\mathbf{x} = \mathsf{A}^{s+k}\mathbf{y}_1 = \mathsf{A}^{k+s}\mathbf{y}_1 = \mathsf{A}^k(\mathsf{A}^s\mathbf{y}_1) = \mathsf{A}^k\mathbf{y}_2,$$

  where $\mathbf{y}_2 = \mathsf{A}^s\mathbf{y}_1$. So, $\mathbf{y}_2 \in R(\mathsf{A}^s)$, which as $R(\mathsf{A}^s) = R(\mathsf{A}^{s+1})$, entails that $\mathbf{y}_2 \in R(\mathsf{A}^{s+1})$ too. Thus, there exists a vector $\mathbf{y}_3$ such that $\mathbf{y}_2 = \mathsf{A}^{s+1}\mathbf{y}_3$. Hence,

  $$\mathbf{x} = \mathsf{A}^k(\mathsf{A}^{s+1}\mathbf{y}_3) = \mathsf{A}^{s+k+1}\mathbf{y}_3,$$

  and so $\mathbf{x} \in R(\mathsf{A}^{s+k+1})$, which means that $R(\mathsf{A}^s) \subseteq R(\mathsf{A}^{s+k+1})$.

[20] We should maybe take a moment to look at our stipulation that $\mathbf{y} \in N(\mathsf{A}^t)$ should be non-zero. Firstly, notice that taking $\mathbf{y} = \mathbf{0}$ will mean that $\mathbf{y}^t\mathbf{b} = 0$, indicating that $\mathbf{y}$ and $\mathbf{b}$ are orthogonal. But, as $\mathbf{y} \in N(\mathsf{A}^t)$ this contradicts our assumption that $\mathbf{b} \notin N(\mathsf{A}^t)^\perp$. Indeed, if $\mathbf{y} = \mathbf{0}$, then we are in a situation where system 1 has solutions (as $\mathbf{b} \in N(\mathsf{A}^t)^\perp = R(\mathsf{A})$ will hold if $\mathbf{y} = \mathbf{0}$). Secondly, we should justify the assertion that there *are* non-zero $\mathbf{y} \in N(\mathsf{A}^t)$. To do this, we note that as $\mathbf{b} \notin N(\mathsf{A}^t)^\perp$ it must have a 'component' that lies in $N(\mathsf{A}^t)$ (we shall see what this means when we look at projections).

Thus, by the Principle of Induction, if $R(\mathsf{A}^s) = R(\mathsf{A}^{s+1})$, then $R(\mathsf{A}^s) \subseteq R(\mathsf{A}^{s+k})$ for all $k \geq 0$. Consequently, we have shown that if $R(\mathsf{A}^s) = R(\mathsf{A}^{s+1})$, then $R(\mathsf{A}^s) = R(\mathsf{A}^q)$ for all $q \geq s$, as required.

We are also asked to prove that if $R(\mathsf{A}^s) = R(\mathsf{A}^{s+1})$, then $N(\mathsf{A}^s) = N(\mathsf{A}^q)$ for all $q \geq s$. To do this, we begin by showing that for any $k \geq 0$, $N(\mathsf{A}^s) \subseteq N(\mathsf{A}^{s+k})$ where $q = s+k$. So, consider a vector $\mathbf{x} \in N(\mathsf{A}^s)$, that is, $\mathbf{x}$ is such that $\mathsf{A}^s\mathbf{x} = \mathbf{0}$. But then, for any $k \geq 0$,

$$\mathsf{A}^k\mathsf{A}^s\mathbf{x} = \mathsf{A}^k\mathbf{0} \implies \mathsf{A}^{k+s}\mathbf{x} = \mathbf{0} \implies \mathbf{x} \in N(\mathsf{A}^{k+s}),$$

and so $N(\mathsf{A}^s) \subseteq N(\mathsf{A}^q)$. Also, the Dimension Theorem tells us that

$$\rho(\mathsf{A}^s) + \eta(\mathsf{A}^s) = n \text{ and } \rho(\mathsf{A}^q) + \eta(\mathsf{A}^q) = n \implies \eta(\mathsf{A}^s) = \eta(\mathsf{A}^q),$$

as $\rho(\mathsf{A}^s) = \rho(\mathsf{A}^q)$ from the previous part. Thus, using the theorem from Question 12 of Problem Sheet 1, i.e.

> **Theorem:** If V is a subspace of a finite dimensional vector space W, then $\dim(V) \leq \dim(W)$. In particular, $V = W$ iff $\dim(V) = \dim(W)$.

we can conclude that $N(\mathsf{A}^s) = N(\mathsf{A}^q)$, as required.

Hence, we are asked to show that if $\rho(\mathsf{A}) < n$, then

$$\mathbb{R}^n \supset R(\mathsf{A}) \supset R(\mathsf{A}^2) \supset \cdots \supset R(\mathsf{A}^p) = R(\mathsf{A}^{p+1}) = R(\mathsf{A}^{p+2}) = \cdots$$

for some $p \geq 1$ (where $C \supset D$ means that $C \supseteq D$ and $C \neq D$). To start with, it should be clear from the **Theorem** and our earlier result that if $\rho(\mathsf{A}) < n$, then $R(\mathsf{A}) \neq \mathbb{R}^n$ and so, we can conclude that $R(\mathsf{A}) \subset \mathbb{R}^n$. Now, to establish the rest of this result, we have to show that there will be a series of strict containments, followed by equalities. To do this, we bear in mind that $\rho(\mathsf{AB}) \leq \rho(\mathsf{A})$, and so for any $k \geq 1$, $\rho(\mathsf{A}^{k+1}) \leq \rho(\mathsf{A}^k)$. The strict containments occur as long as $\rho(\mathsf{A}^{k+1}) < \rho(\mathsf{A}^k)$, which by the **Theorem**, will entail that $R(\mathsf{A}^{k+1}) \subset R(\mathsf{A}^k)$. But, notice that at each strict containment the rank must drop by at least one (as $\rho(\mathsf{A}^{k+1}) < \rho(\mathsf{A}^k)$ in such cases). However, there can be *at most* n such strict containments as the rank must be a non-negative integer. Thus, there will be some integer p where the rank ceases to decrease and $\rho(\mathsf{A}^{p+1}) = \rho(\mathsf{A}^p)$. At this point, the **Theorem** dictates that $R(\mathsf{A}^p) = R(\mathsf{A}^{p+1})$, and we have seen that once we have this equality, we have $R(\mathsf{A}^p) = R(\mathsf{A}^{p+k})$ for all $k \geq 0$. Consequently, for some p (where, incidentally, $1 \leq p \leq n$), we can see that

$$\mathbb{R}^n \supset R(\mathsf{A}) \supset R(\mathsf{A}^2) \supset \cdots \supset R(\mathsf{A}^p) = R(\mathsf{A}^{p+1}) = R(\mathsf{A}^{p+2}) = \cdots$$

as required.

Further, we are asked to prove that $\mathbb{R}^n = N(\mathsf{A}^p) \oplus R(\mathsf{A}^p)$. To do this we follow the hint and use:

$$V = Y \oplus Z \quad \text{iff} \quad Y \cap Z = \{\mathbf{0}\} \quad \text{and} \quad \dim(V) = \dim(Y) + \dim(Z).$$

Thus, we have to show that $N(\mathsf{A}^p) \cap R(\mathsf{A}^p) = \{\mathbf{0}\}$ and $\dim(\mathbb{R}^n) = \dim(N(\mathsf{A}^p)) + \dim(R(\mathsf{A}^p))$. So, taking each of these in turn, we have:

- Firstly, taking any vector $\mathbf{x} \in N(\mathsf{A}^p) \cap R(\mathsf{A}^p)$, we can see that:
  - As $\mathbf{x} \in N(\mathsf{A}^p)$, it is the case that $\mathsf{A}^p\mathbf{x} = \mathbf{0}$, and
  - As $\mathbf{x} \in R(\mathsf{A}^p)$, there exists a vector $\mathbf{y}$ such that $\mathbf{x} = \mathsf{A}^p\mathbf{y}$.

  So, substituting the second of these into the first we get

  $$\mathsf{A}^p(\mathsf{A}^p\mathbf{y}) = \mathbf{0} \implies \mathsf{A}^{2p}\mathbf{y} = \mathbf{0} \implies \mathbf{y} \in N(\mathsf{A}^{2p}).$$

  But we have shown that if $R(\mathsf{A}^p) = R(\mathsf{A}^{p+1})$, then $N(\mathsf{A}^p) = N(\mathsf{A}^q)$ for all $q \geq p$. (Notice that it is essential that the 'p' in this part is the same as the 'p' in the previous part!) Thus, [as $q = 2p > p$,]

  $$\mathbf{y} \in N(\mathsf{A}^p) \implies \mathsf{A}^p\mathbf{y} = \mathbf{0} \implies \mathbf{x} = \mathbf{0},$$

  as $\mathbf{x} = \mathsf{A}^p\mathbf{y}$. That is, we have shown that $N(\mathsf{A}^p) \cap R(\mathsf{A}^p) = \{\mathbf{0}\}$.

- Secondly, to establish that $\dim(\mathbb{R}^n) = \dim(N(\mathsf{A}^p)) + \dim(R(\mathsf{A}^p))$, we note that this is just the Dimension Theorem, i.e. $n = \eta(\mathsf{A}^p) + \rho(\mathsf{A}^p)$.

Consequently, $\mathbb{R}^n = N(\mathsf{A}^p) \oplus R(\mathsf{A}^p)$, as required.

And, if you were really keen, you may have tried to prove that

$$V = Y \oplus Z \quad \text{iff} \quad Y \cap Z = \{\mathbf{0}\} \quad \text{and} \quad \dim(V) = \dim(Y) + \dim(Z).$$

As this is an 'iff' statement, we must prove it both ways:

LTR: Let us suppose that $V = Y \oplus Z$. Here, we have to show that $Y \cap Z = \{\mathbf{0}\}$ and $\dim(V) = \dim(Y) + \dim(Z)$. But, we have shown in the lectures that if $V = Y \oplus Z$, then $Y \cap Z = \{\mathbf{0}\}$ [and $V = Y + Z$]. Whilst the third part of Question 6 tells us that if $V = Y \oplus Z$, then $\dim(V) = \dim(Y) + \dim(Z)$. Consequently, it is clear that if $V = Y \oplus Z$, then $Y \cap Z = \{\mathbf{0}\}$ and $\dim(V) = \dim(Y) + \dim(Z)$, as required.

RTL: Let us suppose that $Y \cap Z = \{\mathbf{0}\}$ and $\dim(V) = \dim(Y) + \dim(Z)$ where here, we have to show that $V = Y \oplus Z$. So, taking $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ and $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\}$ to be bases for the subspaces Y and Z of V respectively, i.e. $Y = \text{Lin}\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ and $Z = \text{Lin}\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\}$, we want to establish that the union of these two bases, i.e. $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n, \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\}$, is itself a basis. This can be done by showing that the vectors in this set are linearly independent, and we do this by using a Proof by Contradiction:

> Assume, for contradiction, that the set $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n, \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\}$ is linearly dependent. That is, there is a [non-zero] vector $\mathbf{y}_i$ in the basis for Y that can be expressed as a linear combination of the vectors in Z (notice that this $\mathbf{y}_i$ can not be a linear combination of the other $\mathbf{y}$'s as they form a basis for Y), i.e. $\mathbf{y}_i \in Y$ and $\mathbf{y}_i \in Z$. Thus, $\mathbf{y}_i \in Y \cap Z$ and so $Y \cap Z \neq \{\mathbf{0}\}$ contrary to our assumption that $Y \cap Z = \{\mathbf{0}\}$. Thus, the set $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n, \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\}$ is linearly independent.

This means that we have a set of $n+m$ linearly independent vectors in a $\dim(V) = \dim(Y) + \dim(Z) = n + m$ dimensional vector space V and so, by Theorem 2.17 from the handout for Lecture 2, this set is a basis for V. Consequently, from the second part of Question 6 above, $V = \text{Lin}\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\} \oplus \text{Lin}\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\} = Y \oplus Z$, as required.