Further Mathematical Methods (Linear Algebra) 2002

Solutions For Problem Sheet 10

In this Problem Sheet, we calculated some left and right inverses and verified the theorems about them given in the lectures. We also calculated an SGI and proved one of the theorems that apply to such matrices.

1. To find the equations that u, v, w, x, y and z must satisfy so that

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} u & v \\ w & x \\ y & z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

we expand out this matrix equation to get

$$\begin{aligned} u - w + y &= 1 \\ u + w + 2y &= 0 \end{aligned} \quad \text{and} \quad \begin{aligned} v - x + z &= 0 \\ v + x + 2z &= 1 \end{aligned}$$

and notice that we have two independent sets of equations. Hence, to find all of the right inverses of the matrix,

$$\mathsf{A} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 2 \end{bmatrix},$$

i.e. all matrices R such that $\mathsf{AR} = \mathsf{I}$, we have to solve these sets of simultaneous equations. Now, as each of these sets consist of two equations in three variables, we select one of the variables in each set, say y and z, to be free parameters.[1] Thus, 'mucking about' with the equations, we find

$$\begin{aligned} 2u &= 1 - 3y \\ 2w &= -1 - y \end{aligned} \quad \text{and} \quad \begin{aligned} 2v &= 1 - 3z \\ 2x &= 1 - z \end{aligned}$$

and so, all of the right inverses of the matrix A are given by

$$\mathsf{R} = \frac{1}{2} \begin{bmatrix} 1 - 3y & 1 - 3z \\ -1 - y & 1 - z \\ 2y & 2z \end{bmatrix},$$

where $y, z \in \mathbb{R}$. (You might like to verify this by checking that $\mathsf{AR} = \mathsf{I}$, the 2×2 identity matrix.)

We are also asked to verify that the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ has a solution for every vector $\mathbf{b} = [b_1, b_2]^t \in \mathbb{R}^2$. To do this, we note that for any such vector $\mathbf{b} \in \mathbb{R}^2$, the vector $\mathbf{x} = \mathsf{R}\mathbf{b}$ is a solution to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ since

$$\mathsf{A}(\mathsf{R}\mathbf{b}) = (\mathsf{AR})\mathbf{b} = \mathsf{I}\mathbf{b} = \mathbf{b}.$$

Consequently, for any one of the right inverses calculated above (i.e. any choice of $y, z \in \mathbb{R}$) and any vector $\mathbf{b} = [b_1, b_2]^t \in \mathbb{R}^2$, the vector

$$\mathbf{x} = \mathsf{R}\mathbf{b} = \frac{1}{2} \begin{bmatrix} 1 - 3y & 1 - 3z \\ -1 - y & 1 - z \\ 2y & 2z \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} b_1 + b_2 \\ -b_1 + b_2 \\ 0 \end{bmatrix} - \frac{yb_1 + zb_2}{2} \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix},$$

is a solution to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ as

$$\mathsf{A}\mathbf{x} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 2 \end{bmatrix} \left\{ \frac{1}{2} \begin{bmatrix} b_1 + b_2 \\ -b_1 + b_2 \\ 0 \end{bmatrix} - \frac{yb_1 + zb_2}{2} \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix} \right\} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \mathbf{0} = \mathbf{b}.$$

[1] If you choose different variables to be the free parameters, you will get a general form for the right inverse that looks different to the one given here. However, they will be equivalent.

Thus, we can see that the infinite number of different right inverses available to us give rise to an infinite number of different solutions to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$.

Indeed, the choice of right inverse [via the parameters $y, z \in \mathbb{R}$] only affects the solutions through vectors in $\text{Lin}\{[3, 1, -2]^t\}$ which is the null space of the matrix A. As such, the infinite number of different solutions only differ due to the relationship between the choice of right inverse and the corresponding vector in the null space of A.

2. To find the equations that u, v, w, x, y and z must satisfy so that

$$\begin{bmatrix} u & v & w \\ x & y & z \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

we expand out this matrix equation to get

$$\begin{aligned} u - v + w &= 1 \\ u + v + 2w &= 0 \end{aligned} \quad \text{and} \quad \begin{aligned} x - y + z &= 0 \\ x + y + 2z &= 1 \end{aligned}$$

and notice that we have two independent sets of equations. Hence, to find all of the left inverses of the matrix,

$$\mathsf{A} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 2 \end{bmatrix},$$

i.e. all matrices L such that $\mathsf{LA} = \mathsf{I}$, we have to solve these sets of simultaneous equations. Now, as each of these sets consist of two equations in three variables, we select one of the variables in each set, say w and z, to be free parameters.[2] Thus, 'mucking about' with the equations, we find

$$\begin{aligned} 2u &= 1 - 3w \\ 2v &= -1 - w \end{aligned} \quad \text{and} \quad \begin{aligned} 2x &= 1 - 3z \\ 2y &= 1 - z \end{aligned}$$

and so, all of the left inverses of the matrix A are given by

$$\mathsf{L} = \frac{1}{2} \begin{bmatrix} 1 - 3w & -1 - w & 2w \\ 1 - 3z & 1 - z & 2z \end{bmatrix},$$

where $w, z \in \mathbb{R}$.[3] (You might like to verify this by checking that $\mathsf{LA} = \mathsf{I}$, the 2×2 identity matrix.)

Using this, we can find the solutions of the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ for an arbitrary vector $\mathbf{b} = [b_1, b_2, b_3]^t \in \mathbb{R}^3$ since

$$\mathsf{A}\mathbf{x} = \mathbf{b} \implies \mathsf{LA}\mathbf{x} = \mathsf{L}\mathbf{b} \implies \mathsf{I}\mathbf{x} = \mathsf{L}\mathbf{b} \implies \mathbf{x} = \mathsf{L}\mathbf{b}.$$

Therefore, the desired solutions are given by:

$$\mathbf{x} = \mathsf{L}\mathbf{b} = \frac{1}{2} \begin{bmatrix} 1 - 3w & -1 - w & 2w \\ 1 - 3z & 1 - z & 2z \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} b_1 - b_2 \\ b_1 + b_2 \end{bmatrix} - \frac{3b_1 + b_2 - 2b_3}{2} \begin{bmatrix} w \\ z \end{bmatrix},$$

and this appears to give us an infinite number of solutions (via $w, z \in \mathbb{R}$) contrary to the second part of Question 10. But, we know that the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ is consistent (i.e. has solutions) if

[2] If you choose different variables to be the free parameters, you will get a general form for the left inverse that looks different to the one given here. However, they will be equivalent.

[3] Having done Question 1, this should look familiar since if A is an $m \times n$ matrix of rank m, then it has a right inverse — i.e. there is an $n \times m$ matrix R such that $\mathsf{AR} = \mathsf{I}$ (see Question 11). Now, taking the transpose of this we have $\mathsf{R}^t\mathsf{A}^t = \mathsf{I}$ — i.e. the $m \times n$ matrix R^t is a left inverse of the $n \times m$ rank m matrix A^t (see Question 10). Thus, it should be clear that the desired left inverses are given by R^t where R was found in Question 1.

and only if $\mathbf{b} \in R(\mathsf{A})$. Hence, since $\rho(\mathsf{A}) = 2$, we know that the range of A is a plane through the origin given by

$$\mathbf{b} = [b_1, b_2, b_3]^t \in R(\mathsf{A}) \iff \begin{vmatrix} b_1 & b_2 & b_3 \\ 1 & -1 & 1 \\ 1 & 1 & 2 \end{vmatrix} = 0 \iff 3b_1 + b_2 - 2b_3 = 0.$$

That is, *if* the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ has a solution, then $3b_1 + b_2 - 2b_3 = 0$ and, as such, this solution is given by

$$\mathbf{x} = \frac{1}{2}\begin{bmatrix} b_1 - b_2 \\ b_1 + b_2 \end{bmatrix},$$

which is unique (i.e. independent of the choice of left inverse via $w, z \in \mathbb{R}$).

Alternatively, you may have noted that given the above solutions to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$, we can multiply through by A to get

$$\mathsf{A}\mathbf{x} = \frac{1}{2}\begin{bmatrix} 2b_1 \\ 2b_2 \\ 3b_1 + b_2 \end{bmatrix} - \frac{3b_1 + b_2 - 2b_3}{2}\begin{bmatrix} w + z \\ -w + z \\ w + 2z \end{bmatrix}.$$

Thus, it should be clear that the vector $\mathbf{x}$ will only be a solution to the system of equations if

$$3b_1 + b_2 - 2b_3 = 0,$$

as this alone makes $\mathsf{A}\mathbf{x} = \mathbf{b}$. Consequently, when $\mathbf{x}$ represents a solution, it is given by

$$\mathbf{x} = \frac{1}{2}\begin{bmatrix} b_1 - b_2 \\ b_1 + b_2 \end{bmatrix},$$

which is unique.

3. To show that the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

has no left inverse, we have two methods available to us:

> **Method 1:** A la Question 2, a left inverse exists if there are u, v, w, x, y and z that satisfy the matrix equation
>
> $$\begin{bmatrix} u & v \\ w & x \\ y & z \end{bmatrix}\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$
>
> Now, expanding out the top row and looking at the first and third equations, we require u and v to be such that $u + v = 1$ and $u + v = 0$. But, these equations are inconsistent, and so there are no solutions for u and v, and hence, no left inverse.
>
> **Method 2:** From the lectures (or Question 10), we know that an $m \times n$ matrix has a left inverse if its rank is n. Now, this is a 2×3 matrix and so for a left inverse to exist, we require that it has a rank of three. But, its rank can be *no more* than $\min\{2, 3\} = 2$, and so this matrix cannot have a left inverse.

Similarly, to show that the matrix has a right inverse, we use the analogue of Method 2.[4] So, from the lectures (or Question 11), we know that an $m \times n$ matrix has a right inverse if and only if its rank

[4] Or, we *could* use the analogue of Method 1, but it seems pointless here as we are just trying to show that the right inverse exists. (Note that using this method we would have to show that the six equations were consistent. But, to do this, we would probably end up solving them and finding the general right inverse in the process — which is exactly what the question says we need not do!)

is m. Now, this is a 2×3 matrix and so for a right inverse to exist, we require that it has a rank of two. In this case, the matrix has two linearly independent rows[5] and so it has the required rank, and hence a right inverse. Further, to find *one* of these right inverses, we use the formula $\mathsf{R} = \mathsf{A}^t(\mathsf{A}\mathsf{A}^t)^{-1}$ from the lectures (or Question 11). Thus, as

$$\mathsf{A}\mathsf{A}^t = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \implies (\mathsf{A}\mathsf{A}^t)^{-1} = \frac{1}{8} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix},$$

we get

$$\mathsf{R} = \mathsf{A}^t(\mathsf{A}\mathsf{A}^t)^{-1} = \frac{1}{8} \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 2 & 2 \\ -4 & 4 \\ 2 & 2 \end{bmatrix}.$$

So, tidying up, we get

$$\mathsf{R} = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ -2 & 2 \\ 1 & 1 \end{bmatrix}.$$

(Again, you can check that this is a right inverse of A by verifying that $\mathsf{A}\mathsf{R} = \mathsf{I}$, the 2×2 identity matrix.)

4. We are asked to find the strong generalised inverse of the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 4 & 5 & 3 \\ 2 & 3 & 5 & 1 \\ 3 & 2 & 5 & -1 \\ 4 & 1 & 5 & -3 \end{bmatrix}.$$

To do this, we note that if the column vectors of A are denoted by $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$, then

$$\mathbf{x}_3 = \mathbf{x}_1 + \mathbf{x}_2 \quad \text{and} \quad \mathbf{x}_4 = -\mathbf{x}_1 + \mathbf{x}_2,$$

where the first two column vectors of the matrix, i.e. $\mathbf{x}_1$ and $\mathbf{x}_2$, are linearly independent. Consequently, $\rho(\mathsf{A}) = 2$ and so we use these two linearly independent column vectors to form a 4×2 matrix B of rank two, i.e.

$$\mathsf{B} = \begin{bmatrix} 1 & 4 \\ 2 & 3 \\ 3 & 2 \\ 4 & 1 \end{bmatrix}.$$

We then seek a 2×4 matrix C such that $\mathsf{A} = \mathsf{B}\mathsf{C}$ and $\rho(\mathsf{C}) = 2$, i.e.

$$\mathsf{C} = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

(Notice that this matrix reflects how the column vectors of A are linearly dependent on the column vectors of B.) The strong generalised inverse of A, namely A^G, is then given by

$$\mathsf{A}^G = \mathsf{C}^t(\mathsf{C}\mathsf{C}^t)^{-1}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t,$$

as we saw in the lectures (or Question 13). So, we find that

$$\mathsf{C}\mathsf{C}^t = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \implies (\mathsf{C}\mathsf{C}^t)^{-1} = \frac{1}{3} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

[5] Notice that the two row vectors are not scalar multiples of each other.

and so,

$$\mathsf{C}^t(\mathsf{CC}^t)^{-1} = \frac{1}{3}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix},$$

whereas,

$$\mathsf{B}^t\mathsf{B} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 2 & 3 \\ 3 & 2 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 20 \\ 20 & 30 \end{bmatrix} \implies (\mathsf{B}^t\mathsf{B})^{-1} = \frac{1}{50}\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix},$$

and so,

$$(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t = \frac{1}{50}\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix} = \frac{1}{50}\begin{bmatrix} -5 & 0 & 5 & 10 \\ 10 & 5 & 0 & -5 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} -1 & 0 & 1 & 2 \\ 2 & 1 & 0 & -1 \end{bmatrix}.$$

Thus,

$$\mathsf{A}^G = \mathsf{C}^t(\mathsf{CC}^t)^{-1}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t = \frac{1}{30}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} -1 & 0 & 1 & 2 \\ 2 & 1 & 0 & -1 \end{bmatrix} = \frac{1}{30}\begin{bmatrix} -1 & 0 & 1 & 2 \\ 2 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 \\ 3 & 1 & -1 & -3 \end{bmatrix},$$

is the required strong generalised inverse.[6] (You may like to check this by verifying that $\mathsf{A}^G\mathsf{A}\mathsf{A}^G = \mathsf{A}^G$, or indeed that $\mathsf{A}\mathsf{A}^G\mathsf{A} = \mathsf{A}$. But, then again, you may have better things to do.)

5. We are asked to show that: If A has a right inverse, then the matrix $\mathsf{A}^t(\mathsf{AA}^t)^{-1}$ is the strong generalised inverse of A. To do this, we start by noting that, by Question 11, as A has a right inverse this matrix exists. So, to establish the result, we need to show that this matrix satisfies the four conditions that need to hold for a matrix A^G to be a strong generalised inverse of the matrix A and these are given in Definition 19.1, i.e.

1. $\mathsf{A}\mathsf{A}^G\mathsf{A} = \mathsf{A}$.
2. $\mathsf{A}^G\mathsf{A}\mathsf{A}^G = \mathsf{A}^G$.
3. $\mathsf{A}\mathsf{A}^G$ is an orthogonal projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$.
4. $\mathsf{A}^G\mathsf{A}$ is an orthogonal projection of $\mathbb{R}^n$ parallel to $N(\mathsf{A})$.

Now, one way to proceed would be to show that each of these conditions are satisfied by the matrix $\mathsf{A}^t(\mathsf{AA}^t)^{-1}$.[7] But, to save time, we shall adopt a different strategy here and utilise our knowledge of weak generalised inverses. That is, noting that all right inverses are weak generalised inverses,[8] we know that they must have the following properties:

- $\mathsf{ARA} = \mathsf{A}$.
- AR is a projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$.
- RA is a projection of $\mathbb{R}^n$ parallel to $N(\mathsf{A})$.

[6] In Questions 1 (and 2), we found that a matrix can have many right (and left) inverses. However, the strong generalised inverse of a matrix is unique. (See Question 12.)

[7] This will establish that the matrix $\mathsf{A}^t(\mathsf{AA}^t)^{-1}$ is *a* strong generalised inverse of A. The fact that it is *the* strong generalised inverse of A then follows by the result proved in Question 12.

[8] See the handout for Lecture 18.

and it is easy to see that RAR = R as well since

$$\mathsf{RAR} = \mathsf{R}(\mathsf{AR}) = \mathsf{RI} = \mathsf{R}.$$

Thus, we only need to establish that the projections given above are *orthogonal*, i.e. we only need to show that the matrices AR and RA are *symmetric*. But, as we are only concerned with the right inverse given by $\mathsf{A}^t(\mathsf{AA}^t)^{-1}$, this just involves noting that:

- Since $\mathsf{AR} = \mathsf{I}$, we have $(\mathsf{AR})^t = \mathsf{I}^t = \mathsf{I}$ too, and so $\mathsf{AR} = (\mathsf{AR})^t$.
- $[\mathsf{RA}]^t = [\mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathsf{A}]^t = \mathsf{A}^t[\mathsf{A}^t(\mathsf{AA}^t)^{-1}]^t = \mathsf{A}^t[(\mathsf{AA}^t)^{-1}]^t[\mathsf{A}^t]^t = \mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathsf{A} = \mathsf{RA}$.[9]

Consequently, the right inverse of A given by the matrix $\mathsf{A}^t(\mathsf{AA}^t)^{-1}$ is the strong generalised inverse of A (as required).

Further, we are asked to show that $\mathbf{x}^* = \mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathbf{b}$ is the solution of $\mathsf{A}\mathbf{x} = \mathbf{b}$ nearest to the origin. To do this, we recall that if an $m \times n$ matrix A has a right inverse, then the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ has a solution for all vectors $\mathbf{b} \in \mathbb{R}^m$.[10] Indeed, following the analysis of Question 1, it should be clear that these solutions can be written as

$$\mathbf{x} = \mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathbf{b} + \mathbf{u},$$

where $\mathbf{u} \in N(\mathsf{A})$.[11] Now, the vector $\mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathbf{b}$ is in the range of the matrix A^t and so, since $R(\mathsf{A}^t) = N(\mathsf{A})^\perp$, the vectors $\mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathbf{b}$ and $\mathbf{u}$ are orthogonal. Thus, using the Generalised Theorem of Pythagoras, we have

$$\|\mathbf{x}\|^2 = \|\mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathbf{b} + \mathbf{u}\|^2 = \|\mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathbf{b}\|^2 + \|\mathbf{u}\|^2 \implies \|\mathbf{x}\|^2 \geq \|\mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathbf{b}\|^2,$$

as $\|\mathbf{u}\|^2 \geq 0$. That is, $\mathbf{x}^* = \mathsf{A}^t(\mathsf{AA}^t)^{-1}\mathbf{b}$ is the solution of $\mathsf{A}\mathbf{x} = \mathbf{b}$ nearest to the origin (as required).[12]

Other Problems.

In the Other Problems, we looked at a singular values decomposition and investigated how these can be used to calculate strong generalised inverses. We also look at some other results concerning such matrices.

6. To find the singular values decomposition of the matrix

$$\mathsf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix},$$

we note that the matrix given by

$$\mathsf{A}^\dagger\mathsf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

has eigenvalues given by

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = 0 \implies (2-\lambda)^2 - 1 = 0 \implies \lambda = 2 \pm 1,$$

and the corresponding eigenvectors are given by:

[9]Notice that this is the only part of the proof that requires the use of the 'special' right inverse given by the matrix $\mathsf{A}^t(\mathsf{AA}^t)^{-1}$!

[10]See Questions 1 and 11.

[11]Notice that multiplying this expression by A we get

$$\mathsf{A}\mathbf{x} = \mathsf{AA}^t(\mathsf{AA}^t)^{-1}\mathbf{b} + \mathsf{A}\mathbf{u} = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

which confirms that these vectors are solutions.

[12]I trust that it is obvious that the quantity $\|\mathbf{x}\|$ measures the distance of the solution $\mathbf{x}$ from the origin.

- For $\lambda = 1$: The matrix equation $(\mathsf{A}^\dagger\mathsf{A} - \lambda\mathsf{I})\mathbf{x} = \mathbf{0}$ gives

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{0},$$

  and so we have $x + y = 0$. Thus, taking x to be the free parameter, we can see that the eigenvectors corresponding to this eigenvalue are of the form $x[1, -1]^t$.

- For $\lambda = 3$: The matrix equation $(\mathsf{A}^\dagger\mathsf{A} - \lambda\mathsf{I})\mathbf{x} = \mathbf{0}$ gives

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{0},$$

  and so we have $x - y = 0$. Thus, taking x to be the free parameter, we can see that the eigenvectors corresponding to this eigenvalue are of the form $x[1, 1]^t$.

So, an orthonormal[13] set of eigenvectors corresponding to the positive eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 3$ of the matrix $\mathsf{A}^\dagger\mathsf{A}$ is $\{\mathbf{y}_1, \mathbf{y}_2\}$ where

$$\mathbf{y}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ and } \mathbf{y}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Thus, by Theorem 19.4, the positive eigenvalues of the matrix $\mathsf{A}\mathsf{A}^\dagger$ are $\lambda_1 = 1$ and $\lambda_2 = 3$ too[14] and the orthonormal set of eigenvectors corresponding to these eigenvalues is $\{\mathbf{x}_1, \mathbf{x}_2\}$ where[15]

$$\mathbf{x}_1 = \frac{1}{\sqrt{1}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},$$

and

$$\mathbf{x}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}.$$

Consequently, having found two appropriate sets of orthonormal eigenvectors corresponding to the positive eigenvalues of the matrices $\mathsf{A}^\dagger\mathsf{A}$ and $\mathsf{A}\mathsf{A}^\dagger$, we can find the singular values decomposition of the matrix A as defined in Theorem 19.5, i.e.

$$\begin{aligned} \mathsf{A} &= \sqrt{\lambda_1}\,\mathbf{x}_1\mathbf{y}_1^\dagger + \sqrt{\lambda_2}\,\mathbf{x}_2\mathbf{y}_2^\dagger \\ &= \sqrt{1}\,\frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \end{bmatrix} + \sqrt{3}\,\frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \therefore \quad \mathsf{A} &= \frac{1}{2} \begin{bmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix}. \end{aligned}$$

(Note: you can verify that this is correct by adding these two matrices together and checking that you do indeed get A.)

---

[13] Notice that the eigenvectors that we have found are already orthogonal and so we just have to 'normalise' them. (You should have expected this since the matrix $\mathsf{A}^\dagger\mathsf{A}$ is Hermitian with two distinct eigenvalues.)

[14] Note that since $\mathsf{A}\mathsf{A}^\dagger$ is a 3×3 matrix it has three eigenvalues and, in this case, the third eigenvalue must be zero.

[15] Using the fact that

$$\mathbf{x}_i = \frac{1}{\sqrt{\lambda_i}} \mathsf{A}\mathbf{y}_i,$$

as given by Theorem 19.4.

Hence, to calculate the strong generalised inverse of A, we use the result given in Theorem 19.6 to get

$$\begin{aligned}
\mathsf{A}^G &= \frac{1}{\sqrt{\lambda_1}}\mathbf{y}_1\mathbf{x}_1^\dagger + \frac{1}{\sqrt{\lambda_2}}\mathbf{y}_2\mathbf{x}_2^\dagger \\
&= \frac{1}{\sqrt{1}}\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}-1 & 1 & 0\end{bmatrix} + \frac{1}{\sqrt{3}}\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}\frac{1}{\sqrt{6}}\begin{bmatrix}1 & 1 & 2\end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix}-1 & 1 & 0\\ 1 & -1 & 0\end{bmatrix} + \frac{1}{6}\begin{bmatrix}1 & 1 & 2\\ 1 & 1 & 2\end{bmatrix} \\
\therefore\ \mathsf{A}^G &= \frac{1}{3}\begin{bmatrix}-1 & 2 & 1\\ 2 & -1 & 1\end{bmatrix}.
\end{aligned}$$

Thus, using Theorem 19.2, the orthogonal projection of $\mathbb{R}^3$ onto $R(\mathsf{A})$ is given by

$$\mathsf{AA}^G = \begin{bmatrix}0 & 1\\ 1 & 0\\ 1 & 1\end{bmatrix}\frac{1}{3}\begin{bmatrix}-1 & 2 & 1\\ 2 & -1 & 1\end{bmatrix} = \frac{1}{3}\begin{bmatrix}2 & -1 & 1\\ -1 & 2 & 1\\ 1 & 1 & 2\end{bmatrix},$$

and the orthogonal projection of $\mathbb{R}^2$ parallel to $N(\mathsf{A})$ is given by

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

(Notice that both of these matrices are symmetric and idempotent.)

7. Given that the singular values decomposition of an $m \times n$ matrix A is

$$\mathsf{A} = \sum_{i=1}^{k}\sqrt{\lambda_i}\mathbf{x}_i\mathbf{y}_i^\dagger,$$

we are asked to prove that the matrix given by

$$\sum_{i=1}^{k}\frac{1}{\sqrt{\lambda_i}}\mathbf{y}_i\mathbf{x}_i^\dagger,$$

is the strong generalised inverse of A. To do this, we have to show that this expression gives a matrix which satisfies the four conditions that 'define' a strong generalised inverse and these are given in Definition 19.1, i.e.

1. $\mathsf{AA}^G\mathsf{A} = \mathsf{A}$.

2. $\mathsf{A}^G\mathsf{AA}^G = \mathsf{A}^G$.

3. AA^G is an orthogonal projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$.

4. $\mathsf{A}^G\mathsf{A}$ is an orthogonal projection of $\mathbb{R}^n$ parallel to $N(\mathsf{A})$.

We shall proceed by showing that each of these conditions are satisfied by the matrix defined above.[16]

[16]This will establish that this matrix is *a* strong generalised inverse of A. The fact that it is *the* strong generalised inverse of A then follows by the result proved in Question 12.

1: It is fairly easy to see that the first condition is satisfied since

$$\begin{aligned}
\mathsf{AA}^G\mathsf{A} &= \left[\sum_{p=1}^{k} \sqrt{\lambda_p}\mathbf{x}_p\mathbf{y}_p^\dagger\right]\left[\sum_{q=1}^{k} \frac{1}{\sqrt{\lambda_q}}\mathbf{y}_q\mathbf{x}_q^\dagger\right]\left[\sum_{r=1}^{k} \sqrt{\lambda_r}\mathbf{x}_r\mathbf{y}_r^\dagger\right] \\
&= \sum_{p=1}^{k}\sum_{q=1}^{k}\sum_{r=1}^{k} \sqrt{\frac{\lambda_p\lambda_r}{\lambda_s}}\mathbf{x}_p(\mathbf{y}_p^\dagger\mathbf{y}_q)(\mathbf{x}_q^\dagger\mathbf{x}_r)\mathbf{y}_r^\dagger \\
&= \sum_{p=1}^{k} \sqrt{\lambda_p}\mathbf{x}_p\mathbf{y}_p^\dagger \qquad \text{(As the } \mathbf{x}_i \text{ and } \mathbf{y}_i \text{ form orthonormal sets.)} \\
\therefore\quad \mathsf{AA}^G\mathsf{A} &= \mathsf{A}
\end{aligned}$$

as required.

2: It is also fairly easy to see that the second condition is satisfied since

$$\begin{aligned}
\mathsf{A}^G\mathsf{AA}^G &= \left[\sum_{p=1}^{k} \frac{1}{\sqrt{\lambda_p}}\mathbf{y}_p\mathbf{x}_p^\dagger\right]\left[\sum_{q=1}^{k} \sqrt{\lambda_q}\mathbf{x}_q\mathbf{y}_q^\dagger\right]\left[\sum_{r=1}^{k} \frac{1}{\sqrt{\lambda_r}}\mathbf{y}_r\mathbf{x}_r^\dagger\right] \\
&= \sum_{p=1}^{k}\sum_{q=1}^{k}\sum_{r=1}^{k} \sqrt{\frac{\lambda_q}{\lambda_p\lambda_r}}\mathbf{y}_p(\mathbf{x}_p^\dagger\mathbf{x}_q)(\mathbf{y}_q^\dagger\mathbf{y}_r)\mathbf{x}_r^\dagger \\
&= \sum_{p=1}^{k} \frac{1}{\sqrt{\lambda_p}}\mathbf{y}_p\mathbf{x}_p^\dagger \qquad \text{(As the } \mathbf{x}_i \text{ and } \mathbf{y}_i \text{ form orthonormal sets.)} \\
\therefore\quad \mathsf{A}^G\mathsf{AA}^G &= \mathsf{A}^G
\end{aligned}$$

as required.

Remark: Throughout this course, when we have been asked to show that a matrix represents an *orthogonal* projection, we have shown that it is [idempotent and] symmetric. But, we have only developed the theory of orthogonal projections for *real* matrices and, in this case, our criterion suffices. However, we could ask when a *complex* matrix represents an orthogonal projection, and we may *suspect* that such matrices will be Hermitian. However, to avoid basing the rest of this proof on mere suspicions, we shall now restrict our attention to the case where the eigenvectors $\mathbf{x}_i$ and $\mathbf{y}_i$ (for $i = 1, 2, \ldots, k$) of the matrices $\mathsf{AA}^\dagger$ and $\mathsf{A}^\dagger\mathsf{A}$ are in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively.[17] Under this assumption, we can take the matrices $\mathbf{x}_i\mathbf{x}_i^\dagger$ and $\mathbf{y}_i\mathbf{y}_i^\dagger$ (for $i = 1, 2, \ldots, k$) to be $\mathbf{x}_i\mathbf{x}_i^t$ and $\mathbf{y}_i\mathbf{y}_i^t$ respectively in the proofs of conditions **3** and **4**.[18]

3: To show that AA^G is an orthogonal projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$, we note that AA^G can be written as

$$\mathsf{AA}^G = \left[\sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{x}_i\mathbf{y}_i^\dagger\right]\left[\sum_{j=1}^{k} \frac{1}{\sqrt{\lambda_j}}\mathbf{y}_j\mathbf{x}_j^\dagger\right] = \sum_{i=1}^{k}\sum_{j=1}^{k} \sqrt{\frac{\lambda_i}{\lambda_j}}\mathbf{x}_i(\mathbf{y}_i^\dagger\mathbf{y}_j)\mathbf{x}_j^\dagger = \sum_{i=1}^{k} \mathbf{x}_i\mathbf{x}_i^\dagger,$$

since the $\mathbf{y}_j$ form an orthonormal set.[19] Thus, AA^G is idempotent as

$$(\mathsf{AA}^G)^2 = \left[\sum_{i=1}^{k} \mathbf{x}_i\mathbf{x}_i^\dagger\right]\left[\sum_{j=1}^{k} \mathbf{x}_j\mathbf{x}_j^\dagger\right] = \sum_{i=1}^{k}\sum_{j=1}^{k} \mathbf{x}_i(\mathbf{x}_i^\dagger\mathbf{x}_j)\mathbf{x}_j^\dagger = \sum_{i=1}^{k} \mathbf{x}_i\mathbf{x}_i^\dagger = \mathsf{AA}^G$$

since the $\mathbf{x}_j$ form an orthonormal set and it is symmetric as

$$(\mathsf{AA}^G)^t = \left[\sum_{i=1}^{k} \mathbf{x}_i\mathbf{x}_i^t\right]^t = \sum_{i=1}^{k} (\mathbf{x}_i\mathbf{x}_i^t)^t = \sum_{i=1}^{k} \mathbf{x}_i\mathbf{x}_i^t = \mathsf{AA}^G$$

[17] Recall that, since the matrices $\mathsf{AA}^\dagger$ and $\mathsf{A}^\dagger\mathsf{A}$ are Hermitian, we already know that their eigenvalues are real.

[18] Recall that we also adopted this position in Question 2 of Problem Sheet 8.

[19] Notice that this gives us a particularly simple expression for an orthogonal projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$.

assuming that the $\mathbf{x}_i$ are in $\mathbb{R}^m$. Also, we can see that $R(\mathsf{A}) = R(\mathsf{AA}^G)$ as

- If $\mathbf{u} \in R(\mathsf{A})$, then there is a vector $\mathbf{v}$ such that $\mathbf{u} = \mathsf{A}\mathbf{v}$. Thus, we can write

$$\mathbf{u} = \sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{x}_i\mathbf{y}_i^\dagger \mathbf{v},$$

and multiplying both sides of this expression by AA^G we get

$$\mathsf{AA}^G\mathbf{u} = \left[\sum_{j=1}^{k} \mathbf{x}_j\mathbf{x}_j^\dagger\right]\left[\sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{x}_i\mathbf{y}_i^\dagger \mathbf{v}\right] = \sum_{j=1}^{k}\sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{x}_j(\mathbf{x}_j^\dagger\mathbf{x}_i)\mathbf{y}_i^\dagger \mathbf{v} = \sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{x}_i\mathbf{y}_i^\dagger \mathbf{v} = \mathbf{u},$$

i.e. $\mathbf{u} = \mathsf{AA}^G\mathbf{u}$.[20] Thus, $\mathbf{u} \in R(\mathsf{AA}^G)$ and so, $R(\mathsf{A}) \subseteq R(\mathsf{AA}^G)$.

- If $\mathbf{u} \in R(\mathsf{AA}^G)$, then there is a vector $\mathbf{v}$ such that $\mathbf{u} = \mathsf{AA}^G\mathbf{v}$. Thus, as we can write $\mathbf{u} = \mathsf{A}(\mathsf{A}^G\mathbf{v})$, we can see that $\mathbf{u} \in R(\mathsf{A})$ and so, $R(\mathsf{AA}^G) \subseteq R(\mathsf{A})$.

Consequently, AA^G is an orthogonal projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$ (as required).

4: To show that $\mathsf{A}^G\mathsf{A}$ is an orthogonal projection of $\mathbb{R}^n$ parallel to $N(\mathsf{A})$, we note that $\mathsf{A}^G\mathsf{A}$ can be written as

$$\mathsf{A}^G\mathsf{A} = \left[\sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}}\mathbf{y}_i\mathbf{x}_i^\dagger\right]\left[\sum_{j=1}^{k} \sqrt{\lambda_j}\mathbf{x}_j\mathbf{y}_j^\dagger\right] = \sum_{i=1}^{k}\sum_{j=1}^{k} \sqrt{\frac{\lambda_j}{\lambda_i}}\mathbf{y}_i(\mathbf{x}_i^\dagger\mathbf{x}_j)\mathbf{y}_j^\dagger = \sum_{i=1}^{k} \mathbf{y}_i\mathbf{y}_i^\dagger,$$

since the $\mathbf{x}_j$ form an orthonormal set.[21] Thus, $\mathsf{A}^G\mathsf{A}$ is idempotent as

$$(\mathsf{A}^G\mathsf{A})^2 = \left[\sum_{i=1}^{k} \mathbf{y}_i\mathbf{y}_i^\dagger\right]\left[\sum_{j=1}^{k} \mathbf{y}_j\mathbf{y}_j^\dagger\right] = \sum_{i=1}^{k}\sum_{j=1}^{k} \mathbf{y}_i(\mathbf{y}_i^\dagger\mathbf{y}_j)\mathbf{y}_j^\dagger = \sum_{i=1}^{k} \mathbf{y}_i\mathbf{y}_i^\dagger = \mathsf{A}^G\mathsf{A},$$

since the $\mathbf{y}_j$ form an orthonormal set and it is symmetric as

$$(\mathsf{A}^G\mathsf{A})^t = \left[\sum_{i=1}^{k} \mathbf{y}_i\mathbf{y}_i^t\right]^t = \sum_{i=1}^{k} (\mathbf{y}_i\mathbf{y}_i^t)^t = \sum_{i=1}^{k} \mathbf{y}_i\mathbf{y}_i^t = \mathsf{A}^G\mathsf{A},$$

assuming the $\mathbf{y}_i$ are in $\mathbb{R}^n$. Also, we can see that $N(\mathsf{A}) = N(\mathsf{A}^G\mathsf{A})$ as

- If $\mathbf{u} \in N(\mathsf{A})$, then $\mathsf{A}\mathbf{u} = \mathbf{0}$ and so clearly, $\mathsf{A}^G\mathsf{A}\mathbf{u} = \mathbf{0}$ too. Thus, $\mathbf{u} \in N(\mathsf{A}^G\mathsf{A})$ and so, $N(\mathsf{A}) \subseteq N(\mathsf{A}^G\mathsf{A})$.

- If $\mathbf{u} \in N(\mathsf{A}^G\mathsf{A})$, then $\mathsf{A}^G\mathsf{A}\mathbf{u} = \mathbf{0}$. As such, we have $\mathsf{AA}^G\mathsf{A}\mathbf{u} = \mathbf{0}$ and so, by **1**, we have $\mathsf{A}\mathbf{u} = \mathbf{0}$. Thus, $N(\mathsf{A}^G\mathsf{A}) \subseteq N(\mathsf{A})$.

Consequently, AA^G is an orthogonal projection of $\mathbb{R}^n$ parallel to $N(\mathsf{A})$ (as required).

8. Suppose that the real matrices A and B are such that $\mathsf{AB}^t\mathsf{B} = 0$. We are asked to prove that $\mathsf{AB}^t = 0$. To do this, we note that for any real matrix C with column vectors $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m$, the diagonal elements of the matrix product $\mathsf{C}^t\mathsf{C}$ are given by $\mathbf{c}_1^t\mathbf{c}_1, \mathbf{c}_2^t\mathbf{c}_2, \ldots, \mathbf{c}_m^t\mathbf{c}_m$. So, if $\mathsf{C}^t\mathsf{C} = 0$, all of the elements of the matrix $\mathsf{C}^t\mathsf{C}$ are zero and, in particular, this means that each of the diagonal elements is zero. But then, for $1 \leq i \leq m$, we have

$$\mathbf{c}_i^t\mathbf{c}_i = \|\mathbf{c}_i\|^2 = 0 \implies \mathbf{c}_i = \mathbf{0},$$

[20] Alternatively, as $\mathbf{u} = \mathsf{A}\mathbf{v}$, we have $\mathsf{AA}^G\mathbf{u} = \mathsf{AA}^G\mathsf{A}\mathbf{v}$ and so, by **1**, this gives $\mathsf{AA}^G\mathbf{u} = \mathsf{A}\mathbf{v}$. Thus, again, $\mathsf{AA}^G\mathbf{u} = \mathbf{u}$, as desired.

[21] Notice that this gives us a particularly simple expression for an orthogonal projection of $\mathbb{R}^n$ parallel to $N(\mathsf{A})$.

and so $\mathsf{C}^t\mathsf{C} = 0$ implies that $\mathsf{C} = 0$ as all of the column vectors must be null. Thus, noting that

$$\mathsf{AB}^t\mathsf{B} = 0 \implies \mathsf{AB}^t\mathsf{BA}^t = 0 \implies (\mathsf{BA}^t)^t\mathsf{BA}^t = 0,$$

we can deduce that $\mathsf{BA}^t = 0$. Consequently, taking the transpose of both sides of this expression we have $\mathsf{AB}^t = 0$ (as required).

9. Let A be an $m \times n$ matrix. We are asked to show that the general least squares solution of the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ is given by

$$\mathbf{x} = \mathsf{A}^G\mathbf{b} + (\mathsf{I} - \mathsf{A}^G\mathsf{A})\mathbf{z},$$

where $\mathbf{z}$ is any vector in $\mathbb{R}^n$. To do this, we note that a least squares solution to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ is a vector $\mathbf{x}$ which minimises the quantity $\|\mathsf{A}\mathbf{x} - \mathbf{b}\|$. Thus, as we have seen before, the vector $\mathsf{A}\mathbf{x}^*$ obtained from an orthogonal projection of $\mathbf{b}$ onto $R(\mathsf{A})$ will give such a least squares solution. So, as AA^G is an orthogonal projection onto $R(\mathsf{A})$, we have

$$\mathsf{A}\mathbf{x}^* = \mathsf{AA}^G\mathbf{b} \implies \mathbf{x}^* = \mathsf{A}^G\mathbf{b},$$

is a particular least squares solution. Consequently, the general form of a least squares solution to $\mathsf{A}\mathbf{x} = \mathbf{b}$ will be given by

$$\mathbf{x}^* = \mathsf{A}^G\mathbf{b} + (\mathsf{I} - \mathsf{A}^G\mathsf{A})\mathbf{z},$$

for any vector $\mathbf{z} \in \mathbb{R}^n$ (as required).[22]

Harder Problems.

Here are some results from the lectures on generalised inverses that you might have tried to prove.

Note: In the next two questions we are asked to show that three statements are equivalent. Now, saying that two statements p and q are equivalent is the same as saying $p \Leftrightarrow q$ (i.e. p iff q). Thus, if we have three statements p_1, p_2 and p_3, and we are asked to show that they are equivalent, then we need to establish that $p_1 \Leftrightarrow p_2$, $p_2 \Leftrightarrow p_3$ and $p_1 \Leftrightarrow p_3$. Or, alternatively, we can show that $p_1 \Rightarrow p_2 \Rightarrow p_3 \Rightarrow p_1$. (Can you see why?)

10. We are asked to prove that the following statements about an $m \times n$ matrix A are equivalent:

1. A has a left inverse, i.e. there is a matrix L such that $\mathsf{LA} = \mathsf{I}$. (For example $(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t$.)
2. $\mathsf{A}\mathbf{x} = \mathbf{b}$ has a *unique* solution when it has a solution.
3. A has rank n.

Proof: We prove this theorem using the method described in the note above.

1 ⇒ 2: Suppose that the matrix A has a left inverse, L and that the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ has a solution. By definition, L is such that $\mathsf{LA} = \mathsf{I}$, and so multiplying both sides of our matrix equation by L we get

$$\mathsf{LA}\mathbf{x} = \mathsf{L}\mathbf{b} \implies \mathsf{I}\mathbf{x} = \mathsf{L}\mathbf{b} \implies \mathbf{x} = \mathsf{L}\mathbf{b}.$$

Thus, a solution to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ is $\mathsf{L}\mathbf{b}$ when it has a solution.

[22] Notice that this *is* the general form of a least squares solution to $\mathsf{A}\mathbf{x} = \mathbf{b}$ since it satisfies the equation $\mathsf{A}\mathbf{x}^* = \mathsf{AA}^G\mathbf{b}$ given above. Not convinced? We can see that this is the case since, using the fact that $\mathsf{AA}^G\mathsf{A} = \mathsf{A}$, we have

$$\mathsf{A}\mathbf{x}^* = \mathsf{A}\big[\mathsf{A}^G\mathbf{b} + (\mathsf{I} - \mathsf{A}^G\mathsf{A})\mathbf{z}\big] = \mathsf{AA}^G\mathbf{b} + (\mathsf{A} - \mathsf{AA}^G\mathsf{A})\mathbf{z} = \mathsf{AA}^G\mathbf{b} + (\mathsf{A} - \mathsf{A})\mathbf{z} = \mathsf{AA}^G\mathbf{b} + \mathbf{0} = \mathsf{AA}^G\mathbf{b},$$

as expected.

However, a matrix A can have many left inverses and so, to establish the uniqueness of the solutions, consider two [possibly distinct] left inverses L and L'. Using these, the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ will have two [possibly distinct] solutions given by

$$\mathbf{x} = \mathsf{L}\mathbf{b} \quad \text{and} \quad \mathbf{x}' = \mathsf{L}'\mathbf{b},$$

and so, to establish uniqueness, we must show that their difference, i.e. $\mathbf{x} - \mathbf{x}' = (\mathsf{L} - \mathsf{L}')\mathbf{b}$, is $\mathbf{0}$. To do this, we start by noting that

$$\mathsf{A}\mathbf{x} - \mathsf{A}\mathbf{x}' = \mathsf{A}(\mathsf{L} - \mathsf{L}')\mathbf{b} \implies \mathbf{b} - \mathbf{b} = \mathsf{A}(\mathsf{L} - \mathsf{L}')\mathbf{b} \implies \mathbf{0} = \mathsf{A}(\mathsf{L} - \mathsf{L}')\mathbf{b},$$

and so, the vector $(\mathsf{L} - \mathsf{L}')\mathbf{b} \in N(\mathsf{A})$, i.e. the vectors $\mathbf{x}$ and $\mathbf{x}'$ can only differ by a vector in $N(\mathsf{A})$. Then, we can see that:

- For any vector $\mathbf{u} \in N(\mathsf{A})$, we have $\mathsf{A}\mathbf{u} = \mathbf{0}$ and so $\mathsf{L}\mathsf{A}\mathbf{u} = \mathbf{0}$ too. However, as $\mathsf{L}\mathsf{A} = \mathsf{I}$, this means that $\mathbf{u} = \mathbf{0}$, $N(\mathsf{A}) \subseteq \{\mathbf{0}\}$.

- As $N(\mathsf{A})$ is a vector space, $\mathbf{0} \in N(\mathsf{A})$, which means that $\{\mathbf{0}\} \subseteq N(\mathsf{A})$.

and so $N(\mathsf{A}) = \{\mathbf{0}\}$.[23] Thus, the vectors $\mathbf{x}$ and $\mathbf{x}'$ can only differ by the vector $\mathbf{0}$, i.e. we have $\mathbf{x} - \mathbf{x}' = \mathbf{0}$. Consequently, $\mathsf{L}\mathbf{b}$ is the unique solution to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$when it has a solution (as required).

2 ⇒ 3: The matrix equation $\mathsf{A}\mathbf{x} = \mathbf{0}$ has a solution, namely $\mathbf{x} = \mathbf{0}$, and so by **2**, this solution is unique. Thus, the null space of A is $\{\mathbf{0}\}$ and so $\eta(\mathsf{A}) = 0$. Hence, by the Dimension Theorem, the rank of A is given by

$$\rho(\mathsf{A}) = n - \eta(\mathsf{A}) \implies \rho(\mathsf{A}) = n - 0 = n,$$

as required.

3 ⇒ 1: Suppose that the rank of the $m \times n$ matrix A is n. That is, the matrix $\mathsf{A}^t\mathsf{A}$ is invertible (as $\rho(\mathsf{A}^t\mathsf{A}) = \rho(\mathsf{A})$ and $\mathsf{A}^t\mathsf{A}$ is $n \times n$) and so the matrix given by $(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t$ exists. Further, this matrix is a left inverse since

$$\left[(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\right]\mathsf{A} = (\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathsf{A} = (\mathsf{A}^t\mathsf{A})^{-1}\left[\mathsf{A}^t\mathsf{A}\right] = \mathsf{I},$$

as required.

Remark: An immediate consequence of this theorem is that if A is an $m \times n$ matrix with $m < n$, then as $\rho(\mathsf{A}) = \min\{m, n\} = m < n$, A can have no left inverse. (Compare this with what we found in Question 3.) Further, when $m = n$, A is a square matrix and so if it has full rank, it will have a left inverse which is identical to the inverse A^{-1} that we normally consider.[24]

11. We are asked to prove that the following statements about an $m \times n$ matrix A are equivalent:

1. A has a right inverse, i.e. there is a matrix R such that $\mathsf{A}\mathsf{R} = \mathsf{I}$. (For example $\mathsf{A}^t(\mathsf{A}\mathsf{A}^t)^{-1}$.)

2. $\mathsf{A}\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$.

3. A has rank m.

[23] We should have expected this since, looking ahead to **3**, we have $\rho(\mathsf{A}) = n$ if A is an $m \times n$ matrix. Thus, by the rank-nullity theorem, we have

$$\rho(\mathsf{A}) + \eta(\mathsf{A}) = n \implies n + \eta(\mathsf{A}) = n \implies \eta(\mathsf{A}) = 0,$$

and so, the null space of A must be zero-dimensional, i.e. $N(\mathsf{A}) = \{\mathbf{0}\}$.

[24] Since if A is a square matrix with full rank, then it has a left inverse L such that $\mathsf{L}\mathsf{A} = \mathsf{I}$ and an inverse A^{-1} such that $\mathsf{A}^{-1}\mathsf{A} = \mathsf{I}$. Equating these two expressions we get $\mathsf{L}\mathsf{A} = \mathsf{A}^{-1}\mathsf{A}$, and so multiplying both sides by A^{-1} (say), we can see that $\mathsf{L} = \mathsf{A}^{-1}$ (as expected).

Proof: We prove this theorem using the method described in the note above.

1 ⇒ 2: Suppose that the matrix A has a right inverse, R. Then, for *any* $\mathbf{b} \in \mathbb{R}^m$ we have

$$\mathbf{b} = \mathsf{I}\mathbf{b} = (\mathsf{AR})\mathbf{b} = \mathsf{A}(\mathsf{R}\mathbf{b}),$$

as $\mathsf{AR} = \mathsf{I}$. Thus, $\mathbf{x} = \mathsf{R}\mathbf{b}$ is a solution to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ for every $\mathbf{b}$, as required.

2 ⇒ 3: Suppose that the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b} \in \mathbb{R}^m$. Recalling the fact that the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ has a solution iff $\mathbf{b} \in R(\mathsf{A})$, this implies that $R(\mathsf{A}) = \mathbb{R}^m$. Thus, the rank of A is m, as required.

3 ⇒ 1: Suppose that the rank of the $m \times n$ matrix A is m. That is, the matrix AA^t is invertible (as $\rho(\mathsf{AA}^t) = \rho(\mathsf{A})$ and AA^t is $m \times m$) and so the matrix given by $\mathsf{A}^t(\mathsf{AA}^t)^{-1}$ exists. Further, this matrix is a right inverse since

$$\mathsf{A}\big[\mathsf{A}^t(\mathsf{AA}^t)^{-1}\big] = \mathsf{AA}^t(\mathsf{AA}^t)^{-1} = \big[\mathsf{AA}^t\big](\mathsf{AA}^t)^{-1} = \mathsf{I},$$

as required.

Remark: An immediate consequence of this theorem is that if A is an $m \times n$ matrix with $n < m$, then as $\rho(\mathsf{A}) = \min\{m, n\} = n < m$, A can have no right inverse. (Compare this with what we found in Question 3.) Further, when $m = n$, A is a square matrix and so if it has full rank, it will have a right inverse which is identical to the inverse A^{-1} that we normally consider.[25]

12. We are asked to prove that a matrix A has exactly one strong generalised inverse. To do this, let us suppose that we have two matrices, G and H, which both have the properties attributed to a strong generalised inverse of A. (There are four such properties, and they are given in Definition 19.1.) Thus, working on the first two terms of $\mathsf{G} = \mathsf{GAG}$, we can see that

$$\begin{aligned}
\mathsf{G} &= \mathsf{GAG} && \text{:Prop 2.}\\
&= (\mathsf{GA})^t\mathsf{G} && \text{:Prop 4. (As orthogonal projections are symmetric.)}\\
&= \mathsf{A}^t\mathsf{G}^t\mathsf{G} &&\\
&= (\mathsf{AHA})^t\mathsf{G}^t\mathsf{G} && \text{:Prop 1. (Recall that H is an SGI too.)}\\
&= \mathsf{A}^t\mathsf{H}^t\mathsf{A}^t\mathsf{G}^t\mathsf{G} &&\\
&= (\mathsf{HA})^t(\mathsf{GA})^t\mathsf{G} &&\\
&= (\mathsf{HA})(\mathsf{GA})\mathsf{G} && \text{:Prop 4. (Again, as orthogonal projections are symmetric.)}\\
&= \mathsf{H}(\mathsf{AGA})\mathsf{G} &&\\
\therefore\ \mathsf{G} &= \mathsf{HAG} && \text{:Prop 1.}
\end{aligned}$$

whereas, working on the last two terms of $\mathsf{H} = \mathsf{HAH}$, we have

$$\begin{aligned}
\mathsf{H} &= \mathsf{HAH} && \text{:Prop 2.}\\
&= \mathsf{H}(\mathsf{AH})^t && \text{:Prop 3. (As orthogonal projections are symmetric.)}\\
&= \mathsf{HH}^t\mathsf{A}^t &&\\
&= \mathsf{HH}^t(\mathsf{AGA})^t && \text{:Prop 1. (Recall that G is an SGI too.)}\\
&= \mathsf{HH}^t\mathsf{A}^t\mathsf{G}^t\mathsf{A}^t &&\\
&= \mathsf{H}(\mathsf{AH})^t(\mathsf{AG})^t &&\\
&= \mathsf{H}(\mathsf{AH})(\mathsf{AG}) && \text{:Prop 3. (Again, as orthogonal projections are symmetric.)}\\
&= \mathsf{H}(\mathsf{AHA})\mathsf{G} &&\\
\therefore\ \mathsf{H} &= \mathsf{HAG} && \text{:Prop 1.}
\end{aligned}$$

[25] Since if A is such a square matrix with full rank, then it has a right inverse R such that $\mathsf{AR} = \mathsf{I}$ and an inverse A^{-1} such that $\mathsf{AA}^{-1} = \mathsf{I}$. Equating these two expressions we get $\mathsf{AR} = \mathsf{AA}^{-1}$, and so multiplying both sides by A^{-1} (say), we can see that $\mathsf{R} = \mathsf{A}^{-1}$ (as expected).

and we have made prodigious use of the fact that [for any two matrices X and Y,] $(\mathsf{XY})^t = \mathsf{Y}^t\mathsf{X}^t$. Consequently, $\mathsf{H} = \mathsf{HAG} = \mathsf{G}$, and so the strong generalised inverse of a matrix is unique (as required).

13. We are asked to consider an $m \times n$ matrix A that has been decomposed into the product of an $m \times k$ matrix B and an $k \times n$ matrix C such that the ranks of B and C are both k. Under these circumstances, we need to show that the matrix

$$\mathsf{A}^G = \mathsf{C}^t(\mathsf{CC}^t)^{-1}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t,$$

is a strong generalised inverse of A. To do this, we have to establish that this matrix satisfies the four conditions that 'define' a strong generalised inverse A^G of A and these are given in Definition 19.1, i.e.

1. $\mathsf{AA}^G\mathsf{A} = \mathsf{A}$.
2. $\mathsf{A}^G\mathsf{AA}^G = \mathsf{A}^G$.
3. AA^G is an orthogonal projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$.
4. $\mathsf{A}^G\mathsf{A}$ is an orthogonal projection of $\mathbb{R}^n$ parallel to $N(\mathsf{A})$.

We shall proceed by showing that each of these properties are satisfied by the matrix defined above.

1: It is easy to see that the first property is satisfied since

$$\begin{aligned} \mathsf{AA}^G\mathsf{A} &= \mathsf{BC}\big[\mathsf{C}^t(\mathsf{CC}^t)^{-1}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\big]\mathsf{BC} \\ &= \mathsf{B}\big[\mathsf{CC}^t(\mathsf{CC}^t)^{-1}\big]\big[(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\mathsf{B}\big]\mathsf{C} \\ \therefore \quad \mathsf{AA}^G\mathsf{A} &= \mathsf{BC} = \mathsf{A}, \end{aligned}$$

as required.

2: It is also easy to see that the second property is satisfied since

$$\begin{aligned} \mathsf{A}^G\mathsf{AA}^G &= \big[\mathsf{C}^t(\mathsf{CC}^t)^{-1}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\big]\mathsf{BC}\big[\mathsf{C}^t(\mathsf{CC}^t)^{-1}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\big] \\ &= \mathsf{C}^t(\mathsf{CC}^t)^{-1}\big[(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\mathsf{B}\big]\big[\mathsf{CC}^t(\mathsf{CC}^t)^{-1}\big](\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t \\ \therefore \quad \mathsf{A}^G\mathsf{AA}^G &= \mathsf{C}^t(\mathsf{CC}^t)^{-1}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t = \mathsf{A}^G, \end{aligned}$$

as required.

For the last two parts, we recall that an $m \times n$ matrix P is an orthogonal projection of $\mathbb{R}^n$ onto $R(\mathsf{A})$ [parallel to $N(\mathsf{A})$] if

a. P is idempotent (i.e. $\mathsf{P}^2 = \mathsf{P}$).
b. P is symmetric (i.e. $\mathsf{P}^t = \mathsf{P}$).
c. $R(\mathsf{P}) = R(\mathsf{A})$.

So, using this, we have

3: To show that AA^G is an orthogonal projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$, we note that AA^G can be written as

$$\mathsf{AA}^G = \big[\mathsf{BC}\big]\big[\mathsf{C}^t(\mathsf{CC}^t)^{-1}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\big] = \mathsf{B}\big[\mathsf{CC}^t(\mathsf{CC}^t)^{-1}\big](\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t = \mathsf{B}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t.$$

Thus, AA^G is idempotent as

$$\begin{aligned} \big(\mathsf{AA}^G\big)^2 &= \big[\mathsf{B}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\big]^2 \\ &= \big[\mathsf{B}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\big]\big[\mathsf{B}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\big] \\ &= \mathsf{B}(\mathsf{B}^t\mathsf{B})^{-1}\big[\mathsf{B}^t\mathsf{B}(\mathsf{B}^t\mathsf{B})^{-1}\big]\mathsf{B}^t \\ &= \mathsf{B}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t \\ \therefore \quad \big(\mathsf{AA}^G\big)^2 &= \mathsf{AA}^G, \end{aligned}$$

and it is symmetric as

$$
\begin{aligned}
\left(\mathsf{AA}^G\right)^t &= \left[\mathsf{B}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\right]^t \\
&= \mathsf{B}\left[(\mathsf{B}^t\mathsf{B})^{-1}\right]^t\mathsf{B}^t \\
&= \mathsf{B}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t \\
\therefore \quad \left(\mathsf{AA}^G\right)^t &= \mathsf{AA}^G,
\end{aligned}
$$

where we have used the fact that [for any matrices X and Y,] $(\mathsf{XY})^t = \mathsf{Y}^t\mathsf{X}^t$ and [for any invertible square matrix X,] $(\mathsf{X}^{-1})^t = (\mathsf{X}^t)^{-1}$. Also, we can see that $R(\mathsf{AA}^G) = R(\mathsf{A})$ as

- If $\mathbf{x} \in R(\mathsf{A})$, then there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{x} = \mathsf{A}\mathbf{y}$. But, from the first property, we know that $\mathsf{A} = \mathsf{AA}^G\mathsf{A}$, and so $\mathbf{x} = \mathsf{AA}^G\mathsf{A}\mathbf{y} = \mathsf{AA}^G(\mathsf{A}\mathbf{y})$. Thus, there exists a vector, namely $\mathbf{z} = \mathsf{A}\mathbf{y}$, such that $\mathbf{x} = \mathsf{AA}^G\mathbf{z}$ and so $\mathbf{x} \in R(\mathsf{AA}^G)$. Consequently, $R(\mathsf{A}) \subseteq R(\mathsf{AA}^G)$.

- If $\mathbf{x} \in R(\mathsf{AA}^G)$, then there is a vector $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{x} = \mathsf{AA}^G\mathbf{y} = \mathsf{A}(\mathsf{A}^G\mathbf{y})$. Thus, there exists a vector, namely $\mathbf{z} = \mathsf{A}^G\mathbf{y}$, such that $\mathbf{x} = \mathsf{A}\mathbf{z}$ and so $\mathbf{x} \in R(\mathsf{A})$. Consequently, $R(\mathsf{AA}^G) \subseteq R(\mathsf{A})$.

Thus, AA^G is an orthogonal projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$, as required.

4: To show that $\mathsf{A}^G\mathsf{A}$ is an orthogonal projection of $\mathbb{R}^n$ parallel to $N(\mathsf{A})$, we note that $\mathsf{A}^G\mathsf{A}$ can be written as

$$
\mathsf{A}^G\mathsf{A} = \left[\mathsf{C}^t(\mathsf{CC}^t)^{-1}(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\right]\left[\mathsf{BC}\right] = \mathsf{C}^t(\mathsf{CC}^t)^{-1}\left[(\mathsf{B}^t\mathsf{B})^{-1}\mathsf{B}^t\mathsf{B}\right]\mathsf{C} = \mathsf{C}^t(\mathsf{CC}^t)^{-1}\mathsf{C}.
$$

Thus, $\mathsf{A}^G\mathsf{A}$ is idempotent as

$$
\begin{aligned}
\left(\mathsf{A}^G\mathsf{A}\right)^2 &= \left[\mathsf{C}^t(\mathsf{CC}^t)^{-1}\mathsf{C}\right]^2 \\
&= \left[\mathsf{C}^t(\mathsf{CC}^t)^{-1}\mathsf{C}\right]\left[\mathsf{C}^t(\mathsf{CC}^t)^{-1}\mathsf{C}\right] \\
&= \mathsf{C}^t(\mathsf{CC}^t)^{-1}\left[\mathsf{CC}^t(\mathsf{CC}^t)^{-1}\right]\mathsf{C} \\
&= \mathsf{C}^t(\mathsf{CC}^t)^{-1}\mathsf{C} \\
\therefore \quad \left(\mathsf{A}^G\mathsf{A}\right)^2 &= \mathsf{A}^G\mathsf{A},
\end{aligned}
$$

and it is symmetric as

$$
\begin{aligned}
\left(\mathsf{A}^G\mathsf{A}\right)^t &= \left[\mathsf{C}^t(\mathsf{CC}^t)^{-1}\mathsf{C}\right]^t \\
&= \mathsf{C}^t\left[(\mathsf{CC}^t)^{-1}\right]^t\mathsf{C} \\
&= \mathsf{C}^t(\mathsf{CC}^t)^{-1}\mathsf{C} \\
\therefore \quad \left(\mathsf{A}^G\mathsf{A}\right)^t &= \mathsf{A}^G\mathsf{A},
\end{aligned}
$$

where we have used the fact that [for any matrices X and Y,] $(\mathsf{XY})^t = \mathsf{Y}^t\mathsf{X}^t$ and [for any invertible square matrix X,] $(\mathsf{X}^{-1})^t = (\mathsf{X}^t)^{-1}$. Also, we can see that $R(\mathsf{A}^G\mathsf{A}) = R(\mathsf{A}^t)$ as

- If $\mathbf{x} \in R(\mathsf{A}^t)$, then there is a vector $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{x} = \mathsf{A}^t\mathbf{y}$. But, from the first property, we know that $\mathsf{A} = \mathsf{AA}^G\mathsf{A}$, and so $\mathbf{x} = (\mathsf{AA}^G\mathsf{A})^t\mathbf{y} = (\mathsf{A}^G\mathsf{A})^t\mathsf{A}^t\mathbf{y} = \mathsf{A}^G\mathsf{AA}^t\mathbf{y} = \mathsf{A}^G\mathsf{A}(\mathsf{A}^t\mathbf{y})$ (recall that $(\mathsf{A}^G\mathsf{A})^t = \mathsf{A}^G\mathsf{A}$ by the symmetry property established above). Thus, there exists a vector, namely $\mathbf{z} = \mathsf{A}^t\mathbf{y}$, such that $\mathbf{x} = \mathsf{A}^G\mathsf{A}\mathbf{z}$ and so $\mathbf{x} \in R(\mathsf{A}^G\mathsf{A})$. Consequently, $R(\mathsf{A}^t) \subseteq R(\mathsf{A}^G\mathsf{A})$.

- If $\mathbf{x} \in R(\mathsf{A}^G\mathsf{A})$, then there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{x} = \mathsf{A}^G\mathsf{A}\mathbf{y} = (\mathsf{A}^G\mathsf{A})^t\mathbf{y} = \mathsf{A}^t(\mathsf{A}^G)^t\mathbf{y} = \mathsf{A}^t[(\mathsf{A}^G)^t\mathbf{y}]$ (notice that we have used the fact that $(\mathsf{A}^G\mathsf{A})^t = \mathsf{A}^G\mathsf{A}$ again). Thus, there exists a vector, namely $\mathbf{z} = (\mathsf{A}^G)^t\mathbf{y}$, such that $\mathbf{x} = \mathsf{A}^t\mathbf{z}$ and so $\mathbf{x} \in R(\mathsf{A}^t)$. Consequently, $R(\mathsf{A}^G\mathsf{A}) \subseteq R(\mathsf{A}^t)$.

Thus, $\mathsf{A}^G\mathsf{A}$ is an orthogonal projection of $\mathbb{R}^n$ onto $R(\mathsf{A}^t)$ parallel to $R(\mathsf{A}^t)^\perp$. However, recall that $R(\mathsf{A}^t)^\perp = N(\mathsf{A})$, and so we have established that $\mathsf{A}^G\mathsf{A}$ is an orthogonal projection of $\mathbb{R}^n$ parallel to $N(\mathsf{A})$, as required.

14. Let us suppose that $\mathsf{A}\mathbf{x} = \mathbf{b}$ is an inconsistent set of equations. We are asked to show that $\mathbf{x} = \mathsf{A}^G\mathbf{b}$ is the least squares solution that is closest to the origin. To do this, we note that a least squares solution to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ is given by

$$\mathbf{x}^* = \mathsf{A}^G\mathbf{b} + (\mathsf{I} - \mathsf{A}^G\mathsf{A})\mathbf{z},$$

for any vector $\mathbf{z}$ (see Question 9) and we need to establish that, of all these solutions, the one given by $\mathbf{x}^* = \mathsf{A}^G\mathbf{b}$ is closest to the origin. To do this, we use the fact that $\mathsf{A}^G = \mathsf{A}^G\mathsf{A}\mathsf{A}^G$, to see that

$$\mathsf{A}^G\mathbf{b} = \mathsf{A}^G\mathsf{A}\mathsf{A}^G\mathbf{b} = \mathsf{A}^G\mathsf{A}(\mathsf{A}^G\mathbf{b}),$$

and so $\mathsf{A}^G\mathbf{b} \in N(\mathsf{A})^\perp$.[26] Then, as $\mathsf{A}^G\mathsf{A}$ is an orthogonal projection parallel to $N(\mathsf{A})$, the matrix $\mathsf{I} - \mathsf{A}^G\mathsf{A}$ is an orthogonal projection onto $N(\mathsf{A})$, and so we can see that $(\mathsf{I} - \mathsf{A}^G\mathsf{A})\mathbf{z} \in N(\mathsf{A})$ for all vectors $\mathbf{z}$. Thus, the vectors given by $\mathsf{A}^G\mathbf{b}$ and $(\mathsf{I} - \mathsf{A}^G\mathsf{A})\mathbf{z}$ are orthogonal, and so applying the Generalised Theorem of Pythagoras, we get

$$\|\mathbf{x}^*\|^2 = \|\mathsf{A}^G\mathbf{b}\|^2 + \|(\mathsf{I} - \mathsf{A}^G\mathsf{A})\mathbf{z}\|^2 \implies \|\mathbf{x}^*\|^2 \geq \|\mathsf{A}^G\mathbf{b}\|^2,$$

as $\|(\mathsf{I} - \mathsf{A}^G\mathsf{A})\mathbf{z}\|^2 \geq 0$. Consequently, the least squares solution given by $\mathbf{x}^* = \mathsf{A}^G\mathbf{b}$ is the one which is closest to the origin.

Is it necessary that the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ is inconsistent? Well, clearly not! To see why, we note that if the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ is consistent, i.e. $\mathbf{b} \in R(\mathsf{A})$, it will have solutions given by

$$\mathbf{x} = \mathsf{A}^G\mathbf{b} + \mathbf{u},$$

where $\mathbf{u} \in N(\mathsf{A})$.[27] Then, since the vector $\mathsf{A}^G\mathbf{b}$ can be written as

$$\mathsf{A}^G\mathbf{b} = \mathsf{A}^G\mathsf{A}\mathsf{A}^G\mathbf{b} = \mathsf{A}^G\mathsf{A}(\mathsf{A}^G\mathbf{b}),$$

we have $\mathsf{A}^G\mathbf{b} \in N(\mathsf{A})^\perp$ (as seen above), and so the vectors given by $\mathsf{A}^G\mathbf{b}$ and $\mathbf{u}$ are orthogonal. Thus, applying the Generalised Theorem of Pythagoras, we get

$$\|\mathbf{x}\|^2 = \|\mathsf{A}^G\mathbf{b} + \mathbf{u}\|^2 = \|\mathsf{A}^G\mathbf{b}\|^2 + \|\mathbf{u}\|^2 \implies \|\mathbf{x}\|^2 \geq \|\mathsf{A}^G\mathbf{b}\|^2,$$

as $\|\mathbf{u}\|^2 \geq 0$. So, once again, the solution $\mathbf{x} = \mathsf{A}^G\mathbf{b}$ is the one which is closest to the origin.

[26] Too quick? Spelling this out, we see that there is a vector, namely $\mathbf{y} = \mathsf{A}^G\mathbf{b}$, such that $\mathsf{A}^G\mathsf{A}\mathbf{y} = \mathsf{A}^G\mathsf{A}\mathsf{A}^G\mathbf{b} = \mathsf{A}^G\mathbf{b}$ and so $\mathsf{A}^G\mathbf{b} \in R(\mathsf{A}^G\mathsf{A})$. But, we know that $\mathsf{A}^G\mathsf{A}$ is an orthogonal projection parallel to $N(\mathsf{A})$ and so $R(\mathsf{A}^G\mathsf{A}) = N(\mathsf{A})^\perp$. Thus, $\mathsf{A}^G\mathbf{b} \in N(\mathsf{A})^\perp$, as claimed.

[27] We can see that these are the solutions since multiplying both sides of this expression by the matrix A we get

$$\mathsf{A}\mathbf{x} = \mathsf{A}\mathsf{A}^G\mathbf{b} + \mathsf{A}\mathbf{u} = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

where $\mathsf{A}\mathbf{u} = \mathbf{0}$ as $\mathbf{u} \in N(\mathsf{A})$ and $\mathsf{A}\mathsf{A}^G\mathbf{b} = \mathbf{b}$ as $\mathbf{b} \in R(\mathsf{A})$ and $\mathsf{A}\mathsf{A}^G$ is an orthogonal projection onto $R(\mathsf{A})$.