

Further Mathematical Methods (Linear Algebra) 2002

Lecture 19: Strong Generalised Inverses

We have seen that WGIs are not unique and that they can be used to give us certain projections. Now, amongst the many WGIs that we could calculate, there is one that is very special. This is called a *strong generalised inverse* and it is special because it is *the* WGI that makes the projections discussed above *orthogonal*. We define it as follows:

Definition 19.1 Let A be an arbitrary $m \times n$ matrix. A *strong generalised inverse* (SGI) of A , denoted by A^G , is any $n \times m$ matrix such that

- $AA^GA = A$.
- $A^GAA^G = A^G$.
- AA^G orthogonally projects \mathbb{R}^m onto $R(A)$.
- A^GA orthogonally projects \mathbb{R}^n parallel to $N(A)$.¹

Bearing in mind our discussion of WGIs, the first thing we have to justify is that the SGI is actually unique:

Theorem 19.2 A matrix A has exactly one SGI.

Proof: See Question 12 on Problem Sheet 10 ♠

Indeed, this ties in nicely with something that we have seen before, namely:

For example: If the matrix A has a left [or right] inverse, then the matrix $(A^tA)^{-1}A^t$ [or $A^t(AA^t)^{-1}$] is *the* SGI of A . (See Question 5 on Problem Sheet 10.) ♣

It turns out that *every* matrix has an SGI, and we can go some way towards establishing this fact by showing how we could calculate the SGI of an $m \times n$ matrix A where $\rho(A) = k \geq 1$.

19.1 A Method For Calculating SGIs

Suppose that A is an $m \times n$ matrix with $\rho(A) = k \geq 1$, that is, A has k linearly independent column vectors, say $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$. Using these, we can construct an $m \times k$ matrix, say

$$B = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \\ | & | & \cdots & | \end{bmatrix},$$

and since every other column vector in A is a linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$, we can find another matrix C such that

$$A = BC.$$

Now, by construction, B is an $m \times k$ matrix of rank k , but what is the rank of C ? Well, clearly, C is going to be a $k \times n$ matrix, and so

- $\rho(C) \leq k$, and
- $k = \rho(A) = \rho(BC) \leq \rho(C)$.

i.e. we must have $\rho(C) = k$ too. Using this, we can construct the matrix

$$C^t(CC^t)^{-1}(B^tB)^{-1}B^t,$$

which is guaranteed to exist as CC^t and B^tB are both $k \times k$ matrices of rank k . We now claim that:

¹That is, A^GA orthogonally projects \mathbb{R}^n onto $N(A)^\perp$.

Theorem 19.3 For any matrix A with rank $\rho(A) = k \geq 1$, the matrix

$$C^t(CC^t)^{-1}(B^tB)^{-1}B^t,$$

where B and C are constructed as above, is the SGI of A .

Proof: See Question 13 on Problem Sheet 10 ♠

Note: The matrices $(B^tB)^{-1}B^t$ and $C^t(CC^t)^{-1}$ used to construct such an SGI are just the left and right inverses which give the SGIs of B and C respectively.

Note: This ‘construction’ will not completely justify the assertion that *every* matrix A has an SGI since it does not deal with the special case where $\rho(A) = 0$.²

For example: Find the strong generalised inverse of the matrix

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

using the method given above. We note that the third column vector of this matrix is linearly dependent on the first two since

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

and so the matrix A is of rank 2 (as the first two column vectors are linearly independent). Thus, taking $k = 2$, we let

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

and given the linear dependence of the column vectors noted above, we have

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

where $A = BC$. So, to find the strong generalised inverse, we note that:

$$B^tB = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \implies (B^tB)^{-1} = \frac{1}{12} \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix},$$

and,

$$CC^t = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \implies (CC^t)^{-1} = \frac{1}{3} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Thus, since

$$(B^tB)^{-1}B^t = \frac{1}{6} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & 1 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 3 & 0 & 3 \\ -1 & 2 & 1 \end{bmatrix},$$

²But, this means that A has no linearly independent column vectors, i.e. we must have $R(A) = \{\mathbf{0}\}$, and so A must be $0_{m,n}$, the $m \times n$ zero matrix. But, this matrix has $A^G = 0_{n,m}$, the $n \times m$ zero matrix as its SGI, since:

- $0_{m,n}0_{n,m}0_{m,n} = 0_{m,n}$ and so $AA^GA = A$.
- $0_{n,m}0_{m,n}0_{n,m} = 0_{n,m}$ and so $A^GAA^G = A^G$.
- The matrix $0_{m,n}0_{n,m}$ clearly orthogonally projects every vector in \mathbb{R}^m onto $R(0_{m,n}) = \{\mathbf{0}\} \subseteq \mathbb{R}^m$ since all such vectors are orthogonal to the null vector $\mathbf{0} \in \mathbb{R}^m$.
- The matrix $0_{n,m}0_{m,n}$ orthogonally projects every vector in \mathbb{R}^n parallel to $N(0_{m,n}) = \mathbb{R}^n$ since for any $\mathbf{x} \in \mathbb{R}^n$, we have $(1_{n,n} - 0_{n,m}0_{m,n})\mathbf{x} = \mathbf{x}$ and all of these vectors are orthogonal to the sole vector in $N(0_{m,n})^\perp = \{\mathbf{0}\}$.

as desired.

and,

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

we have,

$$A^G = C^t(CC^t)^{-1}(B^tB)^{-1}B^t = \frac{1}{18} \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 3 \\ -1 & 2 & 1 \end{bmatrix} = \frac{1}{18} \begin{bmatrix} 5 & 2 & 7 \\ 1 & 4 & 5 \\ 4 & -2 & 2 \end{bmatrix},$$

which is the sought after strong generalised inverse of A . ♣

19.2 Why Are SGIs Useful?

One possible application of SGIs is that they allow us to resolve one of the problems which could arise with our method of least squares fits. You may recall that this method assumed that the inconsistent system of equations formed by using a given relationship and some data had to be such that $\rho(A) = n$ if A was an $m \times n$ matrix. (This was due to the fact that we wanted to calculate the orthogonal projection of the vector \mathbf{b} onto $R(A)$ using the matrix $A(A^tA)^{-1}A^t$ and to do this, we required the inverse of A^tA to exist.) In this case, we discovered that one possible least squares fit was given by

$$\mathbf{x}^* = (A^tA)^{-1}A^t\mathbf{b},$$

as this would minimise the error between the relationship and the data.

Of course, one possible problem with this method is that $\rho(A)$ could be less than n , and so we couldn't use the above result because the inverse that we have to calculate doesn't exist. However, notice that the quantity that we want to calculate above can be written as

$$\mathbf{x}^* = L\mathbf{b},$$

where L is a left inverse, and as it is $(A^tA)^{-1}A^t$, this L is the SGI of A . So, in the cases where our earlier analysis fails, maybe we can try and use SGIs instead? To see why this will work, notice that the SGI of a matrix A^G *always* exists and AA^G orthogonally projects \mathbb{R}^m onto $R(A)$. So, a vector \mathbf{x}^* such that $A\mathbf{x}^* = AA^G\mathbf{b}$ will minimize the least squares error, $\|A\mathbf{x} - \mathbf{b}\|$, as in our previous analysis and clearly,

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

is *one* possible solution of this.

Remark: We are now in a position to discuss how many solutions we will get to a least squares fit analysis of some data modulo a given relationship which we expect them to obey. There are two cases to consider when we have the $m \times n$ matrix A :

- If $\rho(A) = n$, then $\eta(A) = n - \rho(A) = n - n = 0$ and so we have $N(A) = \{\mathbf{0}\}$ as this is the only zero-dimensional vector space.
- If $0 \leq \rho(A) < n$, then $\eta(A) = n - \rho(A) > n - n = 0$, i.e. we have $N(A) \neq \{\mathbf{0}\}$ as $\eta(A) > 0$.

where we have used the rank-nullity theorem in both cases. Now, recalling Theorem 18.3,³ we note that given a matrix equation $A\mathbf{x} = \mathbf{b}$ we can write

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

for any vector $\mathbf{w} \in \mathbb{R}^n$. That is, we have

$$A\mathbf{x} = AA^G\mathbf{b} + A(I - A^GA)\mathbf{w} = AA^G\mathbf{b} + (A - AA^GA)\mathbf{w} = AA^G\mathbf{b} + (A - A)\mathbf{w} = AA^G\mathbf{b},$$

³This applies because all SGIs are WGI's.

where now, since the equations are *not* consistent, it will *not* be the case that $\mathbf{b} = \mathbf{A}\mathbf{A}^G\mathbf{b}$. However, this does mean that any vector of the form

$$\mathbf{x}^* = \mathbf{A}^G\mathbf{b} + \underbrace{(\mathbf{I} - \mathbf{A}^G\mathbf{A})\mathbf{w}}_{\text{in } N(\mathbf{A})},$$

will be a solution to our least squares fit problem (see Question 9 on Problem Sheet 10). In particular, notice that:

- If $\rho(\mathbf{A}) = n$, then as $N(\mathbf{A}) = \{\mathbf{0}\}$, we will get exactly one solution to our least squares fit problem, namely $\mathbf{x}^* = \mathbf{A}^G\mathbf{b}$.
- If $0 \leq \rho(\mathbf{A}) < n$, then as $N(\mathbf{A}) \neq \{\mathbf{0}\}$, we will get an infinite number of solutions to our least squares fit problem, namely

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

for any $\mathbf{w} \in \mathbb{R}^n$. (Of which one will be $\mathbf{x}^* = \mathbf{A}^G\mathbf{b}$.)

The interesting thing is that in the latter case, *the* solution given by $\mathbf{x}^* = \mathbf{A}^G\mathbf{b}$ is the solution that is closest to the origin (see Question 14 on Problem Sheet 10). This is perhaps best seen by considering the illustration given in Figure 1 and noting that using our convention, we have

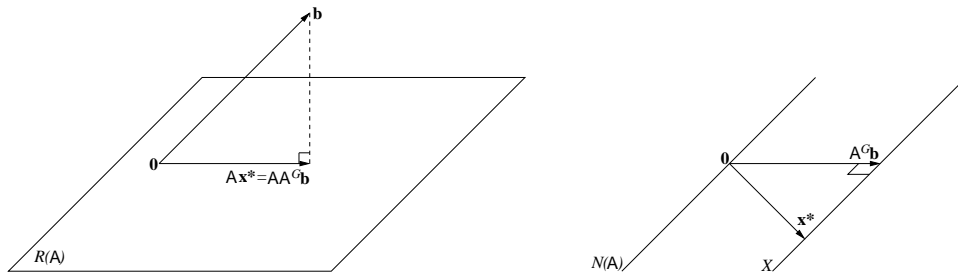


Figure 1: This diagram illustrates the results of a least squares analysis of the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ where $\mathbf{b} \notin R(\mathbf{A})$ and $\rho(\mathbf{A}) < n$. In the diagram on the left we are in \mathbb{R}^m , and here we see the orthogonal projection of the vector \mathbf{b} onto $R(\mathbf{A})$ minimising the least squares error as in our earlier method. However, in the diagram on the right we are in \mathbb{R}^n , and here we see that there are infinitely many solutions to the least squares fit problem and that these all lie in the affine set denoted by X in the diagram. (This is the affine set given by the translate of $N(\mathbf{A})$ by the vector $\mathbf{A}^G\mathbf{b}$.) Also observe that the vector $\mathbf{A}^G\mathbf{b}$ gives the solution to the least squares fit problem that is closest to the origin.

$$\langle (\mathbf{I} - \mathbf{A}^G\mathbf{A})\mathbf{w}, \mathbf{A}^G\mathbf{b} \rangle = \mathbf{w}^t(\mathbf{I} - \mathbf{A}^G\mathbf{A})^t\mathbf{A}^G\mathbf{b},$$

but then, as

$$\begin{aligned} \mathbf{w}^t(\mathbf{I} - \mathbf{A}^G\mathbf{A})^t\mathbf{A}^G\mathbf{b} &= \mathbf{w}^t(\mathbf{I} - \mathbf{A}^G\mathbf{A})\mathbf{A}^G\mathbf{b} && \text{:As } \mathbf{I} - \mathbf{A}^G\mathbf{A} \text{ is an orthogonal projection.} \\ &= \mathbf{w}^t(\mathbf{A}^G - \mathbf{A}^G\mathbf{A}\mathbf{A}^G)\mathbf{b} \\ &= \mathbf{w}^t(\mathbf{A}^G - \mathbf{A}^G)\mathbf{b} && \text{:As } \mathbf{A}^G\mathbf{A}\mathbf{A}^G = \mathbf{A}^G. \end{aligned}$$

$$\therefore \mathbf{w}^t(\mathbf{I} - \mathbf{A}^G\mathbf{A})^t\mathbf{A}^G\mathbf{b} = \mathbf{0},$$

we can see that these two vectors are orthogonal.

For example: Find all of the possible solutions to the least squares fit problem given by the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ where \mathbf{A} is the matrix given in the earlier example and $\mathbf{b} = [-1, 0, 1]^t$. We first note that the range of the matrix \mathbf{A} is the subspace of \mathbb{R}^3 represented by the plane through the origin whose Cartesian equation is given by

$$\begin{vmatrix} x & y & z \\ 1 & 0 & 1 \\ -1 & 2 & 1 \end{vmatrix} = 0 \implies x + y - z = 0,$$

and so we can see that as the vector \mathbf{b} has components such that $-1 + 0 - 1 = -2 \neq 0$, $\mathbf{b} \notin R(\mathbf{A})$ and so this matrix equation is inconsistent. Thus, using the analysis given above, the least squares solutions to this matrix equation will be given by

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

where,

$$\mathbf{A}^G \mathbf{b} = \frac{1}{18} \begin{bmatrix} 5 & 2 & 7 \\ 1 & 4 & 5 \\ 4 & -2 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{18} \begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix},$$

and finding,

$$\mathbf{A}^G \mathbf{A} = \frac{1}{18} \begin{bmatrix} 5 & 2 & 7 \\ 1 & 4 & 5 \\ 4 & -2 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -2 \\ 1 & 1 & 0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix},$$

we have,

$$\mathbf{I} - \mathbf{A}^G \mathbf{A} = \frac{1}{3} \left\{ \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} - \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix} \right\} = \frac{1}{3} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix},$$

that is,

$$(\mathbf{I} - \mathbf{A}^G \mathbf{A}) \mathbf{w} = \frac{1}{3} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{3} \begin{bmatrix} x - y - z \\ -x + y + z \\ -x + y + z \end{bmatrix},$$

where $\mathbf{w} = [x, y, z]^t$ is any vector in \mathbb{R}^3 . Thus, all possible solutions to the least squares fits problem given above are given by

$$\mathbf{x}^* = \frac{1}{9} \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \right\}.$$

where $\lambda = 3(x - y - z) \in \mathbb{R}$

Notice that, as you should expect, the vector $\frac{1}{9}[1, 2, -1]^t$ gives the solution that is closest to the origin since the vector $[1, 2, -1]^t$ is orthogonal to the vector $[1, -1, -1]^t$.⁴ Another way of seeing this is to use the Generalised Theorem of Pythagoras on these two orthogonal vectors, i.e. as

$$\|\mathbf{x}^*\|^2 = \frac{1}{81} \left\| \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \right\|^2 = \frac{1}{81} \left\{ \left\| \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \right\|^2 + \lambda^2 \left\| \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \right\|^2 \right\},$$

we can see that \mathbf{x}^* is minimised when $\|[1, -1, -1]^t\|^2$ term is zero. Further, we can see that $\mathbf{A}\mathbf{x}^* = \frac{1}{3}[-1, 2, 1]^t \in R(\mathbf{A})$ as you should expect. ♣

⁴That is, the length of this vector gives the perpendicular distance from the origin to the line representing all of the possible values of \mathbf{x}^* .