UNIVERSITY OF PUGET SOUND

MATH 420: ADVANCED TOPICS IN LINEAR ALGEBRA

The Moore-Penrose Inverse and Least Squares

ROSS MACAUSLAND

April 16, 2014

1 Introduction

The inverse of a matrix A can only exist if A is nonsingular. This is an important theorem in linear algebra, one learned in an introductory course. In recent years, needs have been felt in numerous areas of applied mathematics for some kind of inverse like matrix of a matrix that is singular or even rectangular. To fulfill this need, mathematicians discovered that even if a matrix was not invertible, there is still either a left or right sided inverse of that matrix. A matrix $A \in \mathbb{C}^{m\times n}$ is left invertible (right invertible) so that there is a matrix $L(R) \in \mathbb{C}^{n\times m}$ so that

$$LA = I_n \qquad\qquad (AR = I_m)\,.$$

This property, where every matrix has some inverse-like matrix, is what gave way to the defining of the generalized inverse.

The generalized inverse has uses in areas such as inconsistent systems of least squares, properties dealing with eigenvalues and eigenvectors, and even statistics. Though the generalized inverse is generally not used, as it is supplanted through various restrictions to create various different generalized inverses for specific purposes, it is the foundation for any pseudoinverse. Arguably the most important generalized inverses is the Moore-Penrose inverse, or pseudoinverse, founded by two mathematicians, E.H. Moore in 1920 and Roger Penrose in 1955. Just as the generalized inverse the pseudoinverse allows mathematicians to construct an inverse like matrix for any matrix, but the pseudoinverse also yields a unique matrix. The pseudoinverse is what is so important, for example, when solving for inconsistent least square systems as it is constructed in a way that gives the minimum norm and therefore the closest solution.

2 Generalized Inverse

If A is any matrix, there is a generalized inverse, A^- such that,

$$AA^-A = A.$$

As mentioned before, this equation is extrapolated from the conjecture that any matrix has at least a one sided inverse. If we assume that A^- is equal to either L or R we see that

$$ALA = A(LA) = AI = A \qquad\qquad ARA = (AR)A = IA = A$$

If A is a $n\times m$ matrix though, A^- is then a $m\times n$ matrix, and the resultant identity matrix either has its rank equal to the columns or rows of A.It is obvious to point out as well that when $m = n$ and when $rank(A) = n$ then $A^- = A^{-1}$. There are other properties, some trivial, some interesting, but the most important part of the generalized inverse though, is that A^- is not unique.

3 Moore-Penrose Inverse

Definition 1. *If $A \in \mathbb{M}_{n,m}$, then there exists a unique $A^+ \in \mathbb{M}_{m,n}$ that satisfies the four Penrose conditions:*

1. $AA^+A = A$
2. $A^+AA^+ = A^+$
3. $A^+A = (A^+A)^*$ *Hermitian*
4. $AA^+ = (AA^+)^*$ *Hermitian*

Where M^* is the conjugate transpose of matrix M.

If A is nonsingular, it is clear that $A^+ = A^{-1}$ trivially satises the four equations. Since the pseudoinverse is known to be unique, which we prove shortly, it follows that the pseudoinverse of a nonsingular matrix is the same as the ordinary inverse.

Theorem 3.1. *For any $A \in \mathbb{C}_{n,m}$ there exists a $A^+ \in \mathbb{C}_{m,n}$ that satisfies the Penrose conditions.*

Proof. The proof of this existence theorem is lengthy and is not included here, but can be taken as conjecture. A version of the proof can be found in *Generalized Inverses: Theory and Applications* $\square$

Theorem 3.2. *For a matrix $A \in \mathbb{M}_{n,m}$, then there exists a unique $A^+ \in \mathbb{M}_{m,n}$*

Proof. Suppose that there are two matrices, B and C that satisfy the four penrose conditions (1,2,3,4) so that

$$\begin{aligned}
B &= BAB && (2)\\
&= (A^*B^*)B && (4)\\
&= (A^*C^*A^*)B^*B && (1)\\
&= (CA)(A^*B^*B) && (4)\\
&= CAB && (2)
\end{aligned}$$

and

$$\begin{aligned}
C &= CAC && (2)\\
&= C(C^*A^*) && (3)\\
&= CC^*(A^*B^*A^*) && (1)\\
&= (CA)(B) && (3)\\
&= CAB && (2).
\end{aligned}$$

Therefore $B = C$. $\square$

The pseudoinverse is now defined and shown to be a exist and have uniqueness, it also has other properties that are interesting.

Lemma 3.1. *For any* $A \in \mathbb{C}^{n\times m}$*, the* $R(A) \oplus N(A^*) = \mathbb{C}^n$*, and similarly* $R(A^*) \oplus N(A) = \mathbb{C}^m$*.*

Proof. This is the result of orthogonal complements, and two fully fleshed proofs can be found in two theorems in *A Second Course in Linear Algebra, 1.16 & 1.17.* □

Lemma 3.2. *If* $A \in \mathbb{C}^{n\times m}$ *with rank* r*, then then* $A = FR$ *where* $F \in \mathbb{C}^{n\times r}$ *and* $R \in \mathbb{C}^{r\times m}$ *where* $r(F) = r(R) = r$*.*

Proof. Since A has rank r, then A has r linearly independent columns., and assume the first r columns of A are linearly independent. If we let the first r columns of F be the first r columns of A, then the remaining $n-r$ columns are a linear combinations of the first r. We can show that

$$A = [F, FX]$$

where $X \in \mathbb{C}^{r\times(n-r)}$. Now let

$$R = [I_r, X],$$

so that when

$$A = F[I_r, X]$$

we can rewrite it as

$$A = FR$$

and $r(F) = r(R) = r$. □

Theorem 3.3. *for any* $C^{n\times m}$ *with rank* r*,* $A^+ = R^*(RR^*)^{-1}(F * F)^{-1}F^*$*.*

Proof. As proved in the previous lemma, $A = RF$ is a full rank factorization, which can be rewritten as

$$F^*AR^* = F^*FRR^*.$$

Because F^*F and RR^* are now $r\times r$ matrices with full rank, F^*FRR^* is also a $r\times r$ matrices with full rank. Now $(F^*AR^*)^{-1}$ also exists, which allows us to form

$$\begin{aligned} R^*\left(F^*AR^*\right)^{-1}F^* &= R^*(F^*(FR)R^*)F^* \\ &= R^*(F^*F)(RR^*)F^* \\ &= X \end{aligned}$$

X is a $m\times n$ matrix, and satisfies all Penrose conditions, so $X = A^+$ □

Corollary 1. *For any matrix* A*,*

$$N(A^*) = N(A^+) \qquad R(A^*) = R(A^+)$$

so that substituting for lemma 3.1 we get

$$R(A) \oplus N(A^+) = \mathbb{C}^n \qquad R(A^+) \oplus N(A) = \mathbb{C}^m.$$

The next two corollaries from the previous theorem are the two ways the pseudoinverse is generally seen.

Corollary 2. *Full Row Rank Decomposition*
In our previous theorem if A has full row rank,$r = n$, then by lemma 3.2, $F = I_n$, and R is $m \times n$. Then

$$A = I_n C$$

and the equation for A^+ reduces to

$$A^+ = A^*(AA^*)^{-1}$$

Corollary 3. *Full Columns Rank Decomposition*
In our previous theorem if A has full column rank,$r = m$, then by lemma 3.2, $R = I_m$, and F is $m \times n$. Then

$$A = B * I_m$$

and the equation for A^+ reduces to

$$A^+ = (A^*A)^{-1}A^*$$

.

Lemma 3.3. *Let $A \in \mathbb{C}^{n\times m}$ and let $\{\vec{x_1}, ..., \vec{x_r}\}, r \leq m$ be an orthonormal set of $\vec{x} \in \mathbb{C}^n$ such that*

$$R(A) = \{\vec{x_1}, ..., \vec{x_r}\}\rangle\,.$$

Then

$$HH^+ = \sum_{i=1}^{r} \vec{x_i}\vec{x_i}^*$$

shows that HH^+ is an orthogonal projection onto $R(A)$, and using similar lemma, H^+H can be proven to be a orthogonal projection onto $N(A)$.

4 Creating A Pseudoinverse

4.1 QR Decomposition

QR decomposition is a matrix decomposition used in linear algebra. It takes a matrix A and breaks it up into two parts, Q and R, where Q is unitary and R is upper triangular. If A is a square matrix of full rank, then it is invertible and not worth decomposing to find its inverse. There are two instances of the QR decomposition that are useful for finding the pseudoinverse. First is when the matrix $A \in \mathbb{C}^{n\times m}$ where $n > m$, then the R matrix comes out as

$$R = \begin{bmatrix} R_1 \\ \mathcal{O} \end{bmatrix}$$

where R_1 is an $m \times m$ upper triangular matrix, and the zero matrix,$\mathcal{O}$, is $(n-m) \times m$. The pseudoinverse can be solved using QR decomposition where

$$A = QR$$

then,

$$A^+ = \begin{bmatrix} R_1^{-1} & \mathcal{O}^* \end{bmatrix} Q^*.$$

Example 1.

$$Let A = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix}$$

Then the matrix decomposition is Q and R where

$$Q = \begin{bmatrix} 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \end{bmatrix} \qquad R = \begin{bmatrix} 2 & 3 & 2 \\ 0 & 5 & -2 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

Then R_1 is defined by the nonzero rows of R, and R_1^{-1} is

$$R_1 = \begin{bmatrix} 2 & 3 & 2 \\ 0 & 5 & -2 \\ 0 & 0 & 4 \end{bmatrix} \qquad R_1^{-1} = \begin{bmatrix} 1/2 & -3/10 & -2/5 \\ 0 & 1/5 & 1/10 \\ 0 & 0 & 1/4 \end{bmatrix}$$

so that now,

$$\begin{aligned} A^+ &= \begin{bmatrix} R_1^{-1} & \mathcal{O}^* \end{bmatrix} Q^* \\ A^+ &= \begin{bmatrix} 1/2 & -3/10 & -2/5 & 0 \\ 0 & 1/5 & 1/10 & 0 \\ 0 & 0 & 1/4 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -1/2 & 1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \end{bmatrix} \\ &= \begin{bmatrix} 1/5 & 3/10 & -1/10 & 3/5 \\ -1/20 & 1/20 & 3/20 & -3/20 \\ 1/8 & -1/8 & 1/8 & -1/8 \end{bmatrix} \\ A^+A &= \begin{bmatrix} 1/5 & 3/10 & -1/10 & 3/5 \\ -1/20 & 1/20 & 3/20 & -3/20 \\ 1/8 & -1/8 & 1/8 & -1/8 \end{bmatrix} \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

This A^+ satisfies the Penrose conditions and is a pseudoinverse.

The second instance of using QR to find the pseudoinverse is when A is square but not of full rank. In this case

$$A = QRU^*,$$

where $Q \in \mathbb{C}^{n\times n}$ is again unitary, and U^* is an orthogonal $n \times m$ matrix. For this case

$$R = \begin{bmatrix} R_1 & 0 \\ 0 & 0 \end{bmatrix},$$

now R_1 is a $r \times r$ matrix, and rows $\{(n-r)...n\}$, and columns $\{(m-r)...m\}$, of R are full of zeros. Then the pseudoinverse can be found by

$$A^+ = U \begin{bmatrix} R_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q^*,$$

which gives us a $m \times n$ matrix that satisfies the Penrose conditions.

4.2 SVD

Using the singular value decomposition in general is great for visualizing what actions are effecting the matrix and the same is true for using the SVD to find the pseudoinverse.

Definition 2. *For the matrix $A \in \mathbb{C}^{n\times m}$ with rank r, the SVD is*

$$A = UDV^*$$

where $U \in \mathbb{C}^{n\times n}$ and $V \in \mathbb{C}^{m\times m}$ are unitary matrices, and $D \in \mathbb{C}^{n\times m}$ is a diagonal matrix of the singular values, s, of A. The singular values are the square roots of the eigenvalues of the square matrices A^*A or AA^*, which are the same values, and the number of singular values is equal to the rank of A.

$$D = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}$$

S is the $r \times r$ of singular matrices. Rows $\{(n-r)...n\}$, and columns $\{(m-r)...m\}$, of D are full of zeros.

Corollary 4. *The SVD can be rewritten as a summation,*

$$A = \sum_{i=1}^{r} s_i \vec{x_i}\vec{y_i}^*$$

where the columns of U are $\{x_1, ..., x_n\}$ and V are $\{y_1, ..., y_m\}$ and orthonormal.

From this definition, and corollary, we can find the pseudoinverse of any matrix A.

Theorem 4.1. *Let $A \in \mathbb{C}^{n\times m}$ and have rank r, and $A = UDV^*$ as defined above.*

$$X = VD^+U^* \qquad X \in \mathbb{R}^{m\times n}$$

$$D^+ = \begin{bmatrix} S^{-1} & 0 \\ 0 & 0 \end{bmatrix}^*$$

where S^{-1} is still a $r \times r$ matrix, but with $s_i^{-1} = \frac{1}{s_i}$ along the diagonals. Then $X = A^+$.

Proof. Just as in *corollary 4*, we can write X in terms of summation.

$$X = \sum_{i=1}^{r} s_i^{-1} \vec{y_i} \vec{x_i}^*$$

To prove that $X = A^+$ we will check the Penrose conditions.

$$\begin{aligned} AXA &= \sum_{a=1}^{r} s_a \vec{y_a} \vec{x_a}^* \sum_{b=1}^{r} s_b^{-1} \vec{x_b} \vec{y_b}^* \sum_{c=1}^{r} s_c \vec{y_c} \vec{x_c}^* \\ &= \sum_{a=1}^{r} \sum_{b=1}^{r} \sum_{c=1}^{r} \frac{s_a s_c}{s_b} \vec{x_a} (\vec{y_a}^* \vec{y_b})(\vec{x_b}^* \vec{x_c}) \vec{y_c}^* \\ &= \sum_{i=1}^{r} s_i \vec{x_i} \vec{y_i}^* \\ &= A \end{aligned}$$

The reverse, $XAX = X$ is trivial after this step. Next we check that XA is Hermitian.

$$\begin{aligned} AX &= \sum_{a=1}^{r} s_a \vec{y_a} \vec{x_a}^* \sum_{b=1}^{r} s_b^{-1} \vec{x_b} \vec{y_b}^* \\ &= \sum_{a=1}^{r} \sum_{b=1}^{r} \frac{s_a}{s_b} \vec{x_a} (\vec{y_a}^* \vec{y_b}) \vec{x_b}^* \\ &= \sum_{i=1}^{r} \vec{x_i} \vec{x_i}^* \end{aligned}$$

Again, proving XA is Hermitian is now trivial. X has shown to satify the Penrose conditions and can now be shown as $X = A^+$. ☐

Example 2.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \sqrt{3} & 0 \end{bmatrix}$$

By using computing software, we can find that

$$\begin{aligned} U &= \begin{bmatrix} -1/2 & 0 & -0.866025403784 \\ 0 & 1 & 0 \\ -0.866025403784 & 0 & 1/2 \end{bmatrix} \\ V &= \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \\ D &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

Where the

$$(D^{+} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Which we then plug into

$$A^{+} = VD^{+}U = \begin{bmatrix} .25 & 00.433012701892 & \\ 0 & 1 & 0 \end{bmatrix}$$

so that A^{+} *satisfies the Penrose conditions.*

5 Pseudoinverse In Least Squares

The pseudoinverse is most often used to solve least squares systems using the equation $A\vec{x} = \vec{b}$. When $\vec{b}$ is in the range of A, there is at least one or more solutions to the system. If $\vec{b}$ is not in the range of A, then there are no solutions to the system, but it is still desirable to to find a $\vec{x_0}$ that is closest to a solution. The residual vector is a key component to solve these systems, and is given as $\vec{r} = A\vec{x} - \vec{b}$.

Definition 3. *The norm of a vector is written as* $||\vec{a}||$ *such that* $||\vec{a}|| = \sqrt{\vec{a^2}}$.

Definition 4. *A least squares solution to a system is a vector such that*

$$||\vec{r_0}|| = ||A\vec{x_0} - \vec{b}|| \leq ||A\vec{x} - \vec{b}||$$

The unique least squares solution is given when the $\vec{x_0}$ creates a minimum in the norm of the residual vector.

Theorem 5.1. $\vec{x_0} = A^{+}\vec{b}$ *is the best approximate solution of* $A\vec{x} = \vec{b}$.

Proof. For any $x \in \mathbb{C}^m$,

$$A\vec{x} - \vec{b} = A(\vec{x} - A^{+}\vec{b}) + (I - AA^{+})(-\vec{b})$$

where $I - AA^{+}$ is an orthogonal projector onto $N(A^{*})$, which by *corollary 1* of *theorem 3.3* we know is also a projector onto $N(A^{+})$, then the summation on the right hand side is of orthogonal vectors. Using Pythagorean theorem with the norm, we can deduce that

$$\begin{aligned} ||A\vec{x} - \vec{b}||^2 &= ||A(\vec{x} - A^{+}\vec{b})||^2 + ||(I - AA^{+})(-\vec{b})||^2 \\ &= ||A(\vec{x} - \vec{x_0})||^2 + ||A\vec{x_0} - \vec{b}||^2 \\ &\geq ||A\vec{x_0} - \vec{b}||^2. \end{aligned}$$

Now we can say that the norm of the residual vector is at its minimum when $\vec{x} = \vec{x_0}$. □

This theorem allows us to affirm that $A^{+}\vec{b}$ is either the unique least squares solution or is the least squares solution of minimum norm.

Example 3.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad \vec{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Using the corollary 2 of theorem 3.3, $\vec{x_0} = A^+\vec{b} = (A^*A)^{-1}A^*\vec{b}$.

$$\vec{x_0} = A^+\vec{b} \qquad A^+ = \begin{bmatrix} 1/9 & 4/9 & -1/9 \\ 2/9 & -1/9 & 5/18 \end{bmatrix}$$

$$\vec{x_0} = \begin{bmatrix} 1/9 & 4/9 & -1/9 \\ 2/9 & -1/9 & 5/18 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

$$= \begin{pmatrix} 2/3 \\ 5/6 \end{pmatrix}$$

$$A\vec{x_0} = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{pmatrix} 2/3 \\ 5/6 \end{pmatrix}$$

$$= \begin{pmatrix} 7/3 \\ 4/3 \\ 5/3 \end{pmatrix} \qquad ||\vec{r}|| = \begin{pmatrix} 7/3 & 4/3 & 5/3 \end{pmatrix} \begin{pmatrix} 7/3 \\ 4/3 \\ 5/3 \end{pmatrix} = \begin{pmatrix} 10 \end{pmatrix}$$

6 Conclusion

We have effectively shown the basics of the pseudoinverse. From where it is derived from, the generalized inverse, to how to calculate it and its use in applications the pseudoinverse is an interesting tool in linear algebra. Please refer to the bibliography for further readings into the pseudoinverse and higher math applications of it, specifically *Regression And The Moore-Penrose Pseudoinverse*by Albert, and *Generalized Inverses: Theory and Applications*by Ben-Israel and Greville.

References

[1] *Advanced Mathematical Methods* Adam Ostaszewski; Cambridge University Press; 1990.

[2] *Circulant Matrices* Philip J. Davis; John Wiley & Sons; 1979.

[3] *Generalized Inverses: Theory and Applications* Adi Ben-Israel, Thomas N.E. Greville; 2001

[4] *Handbook of Linear Algebra (Discrete Mathematics and Its Applications)4th edition* Leslie Hogben, Richard Brualdi, Anne Greenbaum, Roy Mathias, Chapman Hall.

[5] *Matrix Algebra: Theory, Computations, and Applications in Statistic* James E. Gentle; Addison Wesley, Massachusetts, New York, NY: Springer, 2007.

[6] *Matrix Computations (Johns Hopkins Studies in the Mathematical Sciences)* Gene H. Golub, Charles F. Van Loan; John Hopkins University Press,2013.

[7] *The Moore-Penrose Pseudoinverse. A Tutorial Review of the Theory* J. C. A. Barata, M. S. Hussein; John Hopkins University Press,2013.

[8] *Regression and the Moore-Penrose Inverse* Arthur Albert; Academic Press, 1972.

[9] *A Second Course in Linear Algebra* A Second Course in Linear Algebra Robert A. Beezer;2008.

[10] *The Theory of Matrices* Peter Lancaster, Miron Tismenetsky; Academic Press, 1985.