LINEAR VECTOR SPACE DERIVATION OF NEW EXPRESSIONS FOR THE PSEUDO INVERSE OF RECTANGULAR MATRICES

Marco Antonio Murray Lasso

Facultad de Ingeniería, Universidad Nacional Autónoma de México
mamurray@servidor.unam.mx

ABSTRACT

In this paper, a family of simple formulas for the calculation of the pseudo inverse of a rectangular matrix of less than maximum rank is derived using linear vector space methods. The principal result is that the pseudo inverse $\mathbf{A}^+$ of a matrix $\mathbf{A}$ can be calculated as $\mathbf{A}^+ = \mathbf{Q}(\mathbf{P}^T\mathbf{AQ})^{-1}\mathbf{P}^T$, where $\mathbf{P}$ and $\mathbf{Q}$ are rectangular matrices whose r columns are vectors that form a basis for the spaces spanned by the columns and rows, respectively, of matrix $\mathbf{A}$. This leaves the user the liberty to choose the basis to take into consideration other questions such as amount of work needed and ill-conditioning of the matrix that has to be inverted. The formulas are particularized for rectangular matrices that have maximum rank and for the trivial case in which the original matrix is non-singular. Illustrative numerical examples are worked out for several choices of basis vectors and the results are compared with those provided by the program *Mathematica* through its function PseudoInverse[A].

KEY WORDS: *pseudo inverse, generalized inverse, least square solutions, linear vector spaces, orthogonal projections.*

RESUMEN

En este artículo se deduce una familia de fórmulas simples para el cálculo de la seudo inversa de una matriz rectangular de rango menor que el máximo del número de filas y el número de columnas utilizando métodos de espacios vectoriales lineales. El resultado principal es que la seudo inversa $\mathbf{A}^+$ de una matriz **A** se puede calcular con la fórmula
$\mathbf{A}^+ = \mathbf{Q}(\mathbf{P}^T\mathbf{AQ})^{-1}\mathbf{P}^T$, donde **P** y **Q** son matrices rectangulares cuyas r columnas (r es el rango de la matriz **A**) son vectores que forman una base para los espacios generados por las columnas y filas respectivamente de la matriz **A.** Esto le deja al usuario la libertad de escoger la base y tomar en consideración otras cuestiones como la cantidad de trabajo requerida y el mal condicionamiento de la matriz **A**. Las fórmulas se particularizan para matrices rectangulares que tienen máximo rango y para el caso trivial en que la matriz original es no-singular. Se presentan las soluciones de ejemplos numéricos ilustrativos para varias selecciones de vectores base, y los resultados se comparan con los que proporciona el programa *Mathematica* a través de su función `PseudoInverse[A]`.

PALABRAS CLAVE: seudo inversa, inversa generalizada, solución de mínimos cuadrados, espacios vectoriales lineales, proyecciones ortogonales.

1. INTRODUCTION

The pseudo inverse of a matrix is a generalization of the inverse of a matrix which is used in a way similar to the latter to solve (in a least square sense explained below) systems of equations that do not have a solution or whose solution is not unique. All matrices, whether square or not, have a pseudo inverse. When the inverse matrix exists, it coincides with the pseudo inverse. When a rectangular matrix

A has maximum rank, that is, its rank is equal to the minimum of (m, n), where m is the number of rows and n the number of columns, there are simple formulas for the pseudo inverse $\mathbf{A}^+$ involving expressions such as $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ or $\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$. When the rank of **A** is smaller than both m and n, the situation becomes more complicated. Some possibilities are to obtain the *singular value decomposition* of **A** in the form $\mathbf{A} = \mathbf{U}\begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^T$, where **V** and **U** are orthogonal matrices and Λ is a diagonal matrix with r positive entries $\lambda_1, \lambda_2, \ldots, \lambda_r$. The pseudo inverse matrix $\mathbf{A}^+$ is given by $\mathbf{A}^+ = \mathbf{V}\begin{bmatrix} \Gamma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{U}^T$, where Γ is a diagonal matrix with r positive entries $\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_r^{-1}$. (See Dahlquist y Björk [1], pp. 143 – 146.) Although the method appears straightforward, it has the considerable difficulty of having to solve for the eigenvectors and eigenvalues of symmetric matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ and orthogonalize the vectors corresponding to repeated eigenvalues including the zero eigenvalue. There are short cuts that eliminate the zero eigenvalue and consider only the positive ones using rectangular matrices $\mathbf{U}_p$ and $\mathbf{V}_p$, which are portions of **U** and **V**. Golub & Reinsch [2] give alternative methods together with listings of computer programs for obtaining the singular value decomposition and, hence, the pseudo inverse. A second popular method to handle the case $r < m, n$ consists in factoring matrix **A,** $m\times n$ of rank r into two factors **B,** $m\times k$ and **C,** $k\times n$ both of rank r, and apply the formula $\mathbf{A}^+ = \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$. In Noble [3] pp. 142 – 146, the author gives a method for doing the factorization when **A** satisfies certain conditions and explains in a problem how to generalize the method for any **A**.

The generalized inverse of a linear operator was introduced for integral operators by Fredholm in 1903. Subsequently, Moore [4] and Penrose [5] generalized the concept of the inverse of a matrix independently as a way to “solve” systems of linear equations even in the cases in which the matrix of coefficients is singular and is not even square. This *generalized inverse matrix* is also called the *natural inverse* (Lanczos [6], pp.124 – 138) or *pseudo inverse* (Zadeh & Desoer [7].) In the literature, it is known as the *Moore-Penrose pseudoinverse.* The principal characteristic of the pseudo inverse $\mathbf{A}^+$ of a matrix **A** is that in the case of a system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{1}$$

which in the case of a square non-singular matrix **A** can be solved by premultiplying both sides of the equation by the inverse of **A** giving

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{I}\mathbf{x} = \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}, \tag{2}$$

in the case that **A** is singular, its inverse does not exist, and to solve the equation one must examine the situation further to see if the equation does indeed have a solution and, in case it does, if the solution is unique and, in the case it is not unique, how to find all the possible solutions.

With the introduction of the pseudo inverse, independently of the singularity of matrix **A,** a “unique solution” can be found in a very similar way through pseudoinverse $\mathbf{A}^+$ as follows:

$$\mathbf{x}_{ms} = \mathbf{A}^+\mathbf{b} \tag{3}$$

We have placed quotation marks around the words “unique solution” because $\mathbf{x}_{ms}$ may not be a solution at all of equation (1), in case it does not have a solution, or because, in case it does have a solution, we are singling out only one of a possible infinity of solutions. In fact, if $\mathbf{x}_m$ represents all the vectors in the domain of **A** that satisfy

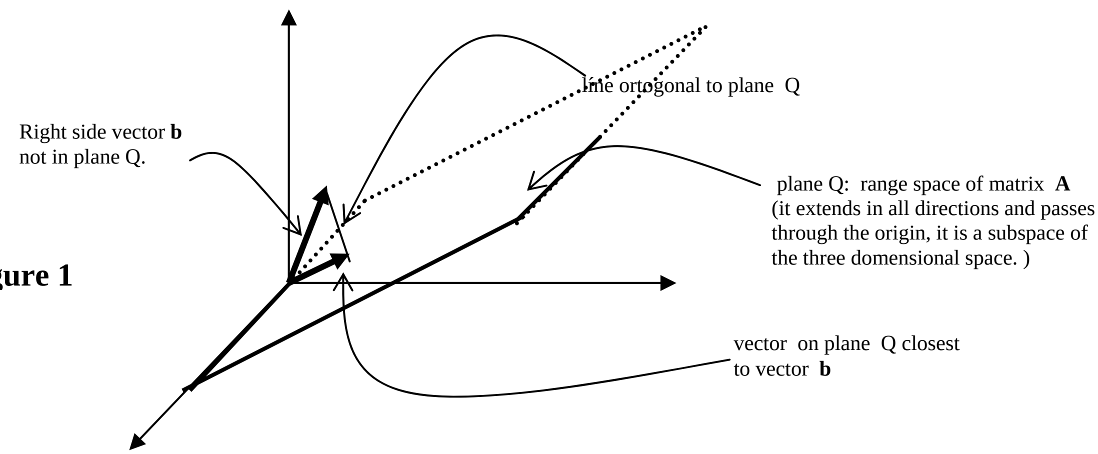
$$\mathbf{x}_m = \min_x \| \mathbf{A}\mathbf{x} - \mathbf{b} \|, \tag{4}$$

where "|| ||" stands for the Euclidean norm or length of a vector which is the square root of the sum of the squares of the components of the vector with respect to the orthonormal natural basis [1, 0, 0, ... 0], [0, 1, 0, ... , 0], ... [0, 0, 0, ... , 1], we will call $\mathbf{x}_{ms}$ the shortest of the vectors satisfying equation (4).

2. GEOMETRICAL INTERPRETATION

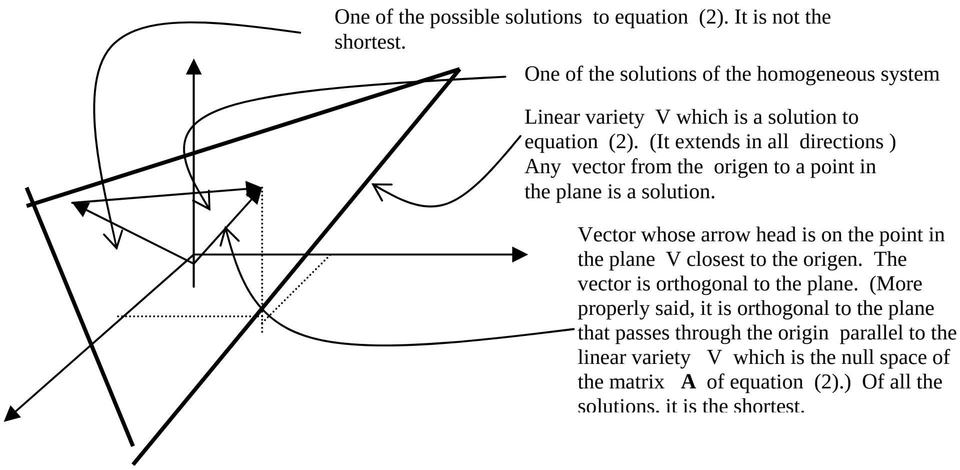
If a linear system of equations has *n* unknowns and the matrix of coeficients has rank *n,* the system has a unique solution which, at the same time, is the solution with minimum length. If this is not the case, the system may not have a solution because the right hand side vector is not in the range space of the matrix of coefficients. In such a case, although there is no solution, a solution whose right side vector is closest to the given right hand side vector can be obtained by orthogonally projecting the given right hand vector onto the range space of the matrix of coefficients. This is shown for a three dimensional case in Fig. 1 that can also be used to illustrate the case in which the right side vector is already in the range space of the coefficient matrix. In such a case, it is not neccessary to perform the projection but any vector in the range space will be a minimum residual solution (with the residual being zero.), so the solution is not unique.

Figure 1



Let us consider the case in which there are many solutions to the system, whether we have projected the right side vector or not. All the solutions have the same residue vector and we are looking for the shortest of these solutions. This is illustrated in Fig. 2.

Figure 2



Linear Vector Space Derivation of an Expression for the Pseudo Inverse.

The first thing we have to consider is that equation (1) may not have a true solution. Let us assume that matrix **A** is $m \times n$ of rank $r < m, n.$ Vector **x,** then, has n components while **b** has m components. For equation (1) to have a solution, the necessary and sufficient condition is that vector **b** lies in the range R(**A**), that is the space spanned by the columns of **A**. If **b** does not lie in the space spanned by the columns of **A,** the best solution, in the sense of a residue vector of minimal Euclidean length, is obtained by replacing vector **b** by a vector that is the orthogonal projection of **b** on the range R(**A**). This can be done by constructing a vector whose components are the inner products of vector **b** with a set of normalized vectors forming a basis for the range space R(**A**). This basis can be formed by finding r linearly independent vectors among the columns of **A** and normalizing them so they have unit length. Let us assume we have done that and that we form a matrix **P,** $m \times r$ whose columns are the normalized linearly independent vectors mentioned. By premultiplying both sides of equation (1) by $\mathbf{P}^T$, the transpose of **P**, we obtain

$$\mathbf{P}^T\mathbf{A}\mathbf{x}_m = \mathbf{P}^T\mathbf{b} \qquad (5)$$

where we have added a subscript to **x** since, unless **b** is in R(**A**), it is not the solution of the original equation. The right side of (5) is the orthogonal projection of **b** on R(**A**). Equation (5) is guaranteed to have a solution since the range of **A** is the same as the range of $\mathbf{P}^T\mathbf{A}$ because the effect of $\mathbf{P}^T$ is to project each column of **A** unto R(**A**). Any solution to equation (5) has the property specified by equation (4).

Because we have assumed that $r < m, n,$ equation (5) will have an infinity of solutions, each of which is formed as the sum of *any* particular solution plus *any* vector in the null space N($\mathbf{P}^T\mathbf{A}$), which is the solution space of the homogeneous system $\mathbf{P}^T\mathbf{A}\mathbf{x} = \mathbf{0}$, which is the same as the null space N(**A**), of all the solutions of equation (5), we are looking for the shortest one. We can form this solution by looking for the shortest possible particular solution and taking the zero vector (which does not increase the length of the particular solution) as the part contributed by the null space. The solution space of a linear system of equations, when it exists, is the subspace N(**A**) of dimension $n - r,$ where n is the number of components of the unknown vector and r is the rank of the matrix of coefficients, displaced parallel to itself from the origin by a particular solution. (See Fig. 1 for a three dimensional illustration.) The shortest possible particular solution must be orthogonal to N(**A**), that is, it must lie in the orthogonal complement of N(**A**), which is the range space of $\mathbf{A}^T$ (Zadeh & Desoer [7], p. C.17.) This shortest vector can be obtained by forcing the solution vector **x** to lie in R($\mathbf{A}^T$). This, in turn, can be accomplished by expressing **x** as a linear combination of base vectors of R($\mathbf{A}^T$). A set of r independent columns of $\mathbf{A}^T$ can serve as the mentioned base, or what is the same, a set of r independent rows of **A** since the columns of $\mathbf{A}^T$ are the rows of **A**.

If we form a matrix **Q,** $n \times r$ whose columns are r linearly independent vectors with n components that span the space generated by rows of matrix **A,** we can introduce a new variable vector ξ obeying

$$\mathbf{x}_{ms} = \mathbf{Q}\xi \qquad (6)$$

into equation (5) obtaining

$$\mathbf{P}^T\mathbf{A}\mathbf{Q}\xi = \mathbf{P}^T\mathbf{b} \qquad (7)$$

We are using the double subindex *ms* because we are restricting vector $\mathbf{x}_m$ to the range of $\mathbf{A}^T$. Since vectors ξ and $\mathbf{P}^T\mathbf{b}$, on both sides of equation (7), are forced to lie in the range of **A** and the range of $\mathbf{A}^T$, respectively, both of dimension r (See Zadeh & Desoer [7], p. C.15.) the $r \times r$ matrix $\mathbf{P}^T\mathbf{A}\mathbf{Q}$ is a one-to-one mapping of the range of $\mathbf{A}^T$ onto the range of **A**; hence, its standard inverse exists, and we can solve equation (7) to obtain

$$\xi = (\mathbf{P}^{\mathrm{T}}\mathbf{AQ})^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{b} \qquad (8)$$

Multiplying both sides of equation (8) by **Q** and taking equation (6) into consideration, we obtain finally

$$\mathbf{x}_{\mathrm{ms}} = \mathbf{Q}\xi = \mathbf{Q}(\mathbf{P}^{\mathrm{T}}\mathbf{AQ})^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{b} \qquad (9)$$

Comparing equation (9) with equation (3), we deduce an expression for pseudo inverse $\mathbf{A}^+$ of **A**

$$\mathbf{A}^+ = \mathbf{Q}(\mathbf{P}^{\mathrm{T}}\mathbf{AQ})^{-1}\mathbf{P}^{\mathrm{T}} \qquad (10)$$

We recall that the r columns of matrix **P** are any set of normalized elements of a basis for the space spanned by the columns of **A**. The r columns of **Q** are any set of elements of a basis for the space spanned by the rows of **A**. We now show that the columns of **P** need not be normalized. Let **P'**= **PD** be the matrix whose columns are the non-nomalized vectors of a basis for the space spanned by the columns of **A**, where **D** is a diagonal matrix whose elements are the lengths of the columns of **P**. The diagonal matrix **D** is non-singular since none of the columns of **P** is the zero vector. We have

$$\mathbf{P} = \mathbf{P}'\mathbf{D}^{-1}$$

which, when the right hand side replaces every instance of **P** in equation (10), becomes

$$\mathbf{A}^+ = \mathbf{Q}\{(\mathbf{P}'\mathbf{D}^{-1})^{\mathrm{T}}\mathbf{AQ}\}^{-1}(\mathbf{P}'\mathbf{D}^{-1})^{\mathrm{T}} = \mathbf{Q}\{\mathbf{P}'^{\mathrm{T}}\mathbf{AQ}\}^{-1}\mathbf{P}'^{\mathrm{T}}$$

where we have replaced the transpose of a product by the product of the transposes of the factors in opposite order, the inverse of a product of non-singular matrices by the product of the inverses of the factors in reverse order and the product of a matrix and its inverse by the unit matrix, which disappears from the expression. Since the form of the expression with or without apostrophes for the **P** is the same, (we can repeat the argument for matrix **Q**) we can interpret equation (10) as a formula for the pseudo inverse with matrices **P** and **Q** having columns which are any complete set of basis vectors of the spaces spanned by the columns and rows of **A**, respectively.

Consider the matrix of coefficients **A** (taken from Noble [3] p. 145.)

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

Matrix **A** has rank 2 because the first two rows are independent (one has a zero where the other one does not) while, by inspection, we see that the third row is equal to the first row minus the second one; the fourth row is equal to the negative of the third row; the fifth row is equal to the negative of the second row, and the sixth row is equal to the negative of the first row. The number of independent rows of any matrix is equal to the number of independent columns of the same matrix. Therefore, we expect two independent columns for **A**. Since the first two columns of **A** are independent (one has a zero where the other does not) we do not have to look further, we can take them as the two columns of **P** and, therefore, its transpose $\mathbf{P}^{\mathrm{T}}$ is

$$\mathbf{P}^T = \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 1 & -1 & 0 \end{bmatrix}$$

To construct matrix **Q,** we take as columns the first two rows of **A;** thus,

$$\mathbf{Q} = \begin{bmatrix} -1 & -1 \\ 0 & 1 \\ 1 & 0 \\ 2 & -1 \end{bmatrix}$$

Using equation (10), we calculate

$$(\mathbf{P}^T\mathbf{AQ})^{-1} = \begin{bmatrix} -10 & -4 \\ -16 & 14 \end{bmatrix}^{-1} = \begin{bmatrix} -\frac{7}{102} & -\frac{1}{51} \\ -\frac{4}{51} & \frac{5}{102} \end{bmatrix}$$

And, finally, we calculate the pseudo inverse from equation (10)

$$\mathbf{A}^+ = \mathbf{Q}(\mathbf{P}^T\mathbf{AQ})^{-1}\mathbf{P}^T =$$

$$\begin{bmatrix} -\frac{5}{34} & -\frac{3}{17} & \frac{1}{34} & -\frac{1}{34} & \frac{3}{17} & \frac{5}{34} \\ \frac{4}{51} & \frac{13}{102} & -\frac{5}{102} & \frac{5}{102} & -\frac{13}{102} & -\frac{4}{51} \\ \frac{7}{102} & \frac{5}{102} & \frac{1}{51} & -\frac{1}{51} & -\frac{5}{102} & -\frac{7}{102} \\ \frac{1}{17} & -\frac{1}{34} & \frac{3}{34} & -\frac{3}{34} & \frac{1}{34} & -\frac{1}{17} \end{bmatrix} = \frac{1}{102}\begin{bmatrix} -15 & -18 & 3 & -3 & 18 & 15 \\ 8 & 13 & -5 & 5 & -13 & -8 \\ 7 & 5 & 2 & -2 & -5 & -7 \\ 6 & -3 & 9 & -9 & 3 & -6 \end{bmatrix}.$$

The results coincide with those given by the program *Mathematica* through its function PseudoInverse[A]. (Wolfram [8], p. 850.) By the way, the book by Noble [3] gives a wrong answer differing only from ours in the sign of the matrix.

There are several alternatives to calculate the pseudo inverse of a matrix, given that we can select any basis for the column space and the row space of **A.** Instead of the **P** and **Q** matrices used, we could have used other independent columns and rows of **A**. For example, if we use the first and fourth columns of **A** to form **P**, and we keep the same **Q** we have

$$\mathbf{P}^T = \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \\ 2 & -1 & 3 & -3 & 1 & -2 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 2 & -1 \end{bmatrix}$$

while for $\mathbf{A}^+$ applying equation (10) we get exactly the same result as before.

In this particular case, we saw by inspection that matrix **A** was of rank 2 and found, also by inspection, two independent columns and rows of **A**. In more complicated cases, we may have to reduce **A** or $\mathbf{A}^T$ to row-reduced echelon form or some other form that will give us the information of the rank *r* of **A** and also provide *r* linearly independent vectors. Those vectors can be used to form matrices **P** and **Q.** For example, if we reduce $\mathbf{A}^T$ and **A** to row-reduced echelon form, we obtain

$$\mathbf{A}^T \Rightarrow \begin{bmatrix} 1 & 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{A} \Rightarrow \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & -1 & -3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

With the non-zero rows of the reduced matrices, we construct matrices $\mathbf{P}^T$ and **Q**

$$\mathbf{P}^T = \begin{bmatrix} 1 & 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 1 & -1 & 0 \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ -2 & -3 \end{bmatrix}$$

and if we use these matrices in equation (10), we obtain identical results for $\mathbf{A}^+$ as before.

Another way of finding the rank of a matrix and obtaining a complete set of basis vectors for the range of **A** and $\mathbf{A}^T$ is to apply to the columns and rows of **A** the Gram-Schimdt orthogonalization process. (Noble [3] p. 314.) By applying this process, we obtain a set of orthonormal vectors that span the space generated by the vectors to which it is applied. It also discovers which of the vectors are linearly dependent on the previously processed vectors because it produces the zero vector when that happens. Instead of normalizing the vectors, which involves taking square roots, the process can be applied without normalization, in which case we end with a set of orthogonal (not *orthonormal*) vectors which span the same space as the vectors processed, and which can be used as basis vectors to construct matrices $\mathbf{P}^T$ and **Q**. If such a non-normalized process is applied to **A** and $\mathbf{A}^T$, we get

$$\mathbf{P}^T = \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \\ -\frac{1}{2} & \frac{1}{2} & -1 & 1 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} -1 & -\frac{7}{6} \\ 0 & 1 \\ 1 & \frac{1}{6} \\ 2 & -\frac{2}{3} \end{bmatrix}$$

When equation (10) is applied with these matrices in place, we obtain results identical to those obtained before.

4. SPECIAL CASES FOR MATRIX A

In the derivation of an expression for the peudo inverse of an $m \times n$ rectangular matrix **A,** we assumed the most complicated case when the rank of the matrix is less than both *m* and *n.* Often the rank of **A** is equal to the smallest of *n*, *m* (it cannot be larger than the smallest of them.) In such a case, the expression for the calculation of the pseudo inverse becomes simpler because one of the matrices **P** or **Q** can be taken to be equal to a unit matrix and the other matrix can be taken to be equal to the transpose of **A**.

Let us first take the case $m > n = r$. This corresponds to the case in which matrix **A** is tall and narrow, there are more equations than variables (the system is overdetermined.) The most common case is that there is no solution to the system of equations, unless the right side vector **b** happens to lie in the space generated by the few columns of **A**. The typical application is the fitting of parameters of a model to experimental points in which, in order to reduce the probability of obtaining poor parameter values, the number m of experimental points is larger than the number n of unknown parameters. The first part of the process is as before, we have to obtain the orthogonal projection of vector **b** into the space spanned by the columns of **A;** hence, matrix **P** is obtained as before without any change. Once we have the right side vector in the range of **A** of dimension $r = n$, we are sure there is a solution to the modified system of equations. Now, since the rank of the new matrix $\mathbf{P}^T\mathbf{A}$ is n, and there are n unknowns in a system where the right hand side has n components, there is nothing else to do; the problem can be solved and it has a unique solution. In this case, the expression for the pseudo inverse $\mathbf{A}^+$ is $\mathbf{A}^+ = (\mathbf{P}^T\mathbf{A})^{-1}\mathbf{P}^T$. Since **P** will have as columns $r = n$ independent vectors that span the same space as the $n = r$ columns of **A**, the columns of **P** can be taken to be the same as the columns of **A**. Therefore, the expression for $\mathbf{A}^+$ could be written

$$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \tag{11}$$

We now consider the case in which $n > m = r$. This is a case of an undetermined system: few equations and many unknowns. Matrix **A** is short and wide. The system will have an $n - r$ fold infinity of solutions. The problem is to obtain the shortest one of them. We do not need to project vector **b** into the range of **A** because it is already in it. What we need to do is to force the solution to lie in the range of $\mathbf{A}^T$. We proceed as in the second part of the general case and find a set of independent vectors spanning the same space as the columns of $\mathbf{A}^T$. This space is r-dimensional and, since $\mathbf{A}^T$ has $r = m$ independent columns (as both **A** and $\mathbf{A}^T$ have the same rank), all its columns are independent and they can be taken as a basis for the m-dimensional space; thus, we might choose $\mathbf{Q} = \mathbf{A}^T$. In this case, the expression for $\mathbf{A}^+$ could be written

$$\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \tag{12}$$

Finally, we mention the case in which $m = n = r$. In this case, matrix **A** is square and non-singular. There is no need to project **b** into the range of **A** since the range of A is the whole space and there is a unique solution for any **b**. Since the solution is unique, there is no need to force the solution to lie in the range of $\mathbf{A}^T$. Thus, for this case, the expression for the pseudo inverse $\mathbf{A}^+$ is

$$\mathbf{A}^+ = \mathbf{A}^{-1} \quad , \tag{13}$$

that is, for this case, the pseudo inverse coincides with the standard inverse.

Although equation (11) is very simple and easy to remember, for numerical reasons the choice of **P** that led to it may result in ill-conditioning of the matrix that is to be inverted. In such a case, orthogonalizing the columns of **A** accurately may improve the accuracy of the solution of the so-called *normal equations* considerably. In Dahlquist & Bjorck [1], pp. 200 – 204, the authors give a simple example to illustrate this situation. This is one of a possible number of situations in which the freedom to choose matrices **P** and **Q** as matrices whose columns are basis for certain subspaces has clear practical advantages.

5. CONCLUSION

We have derived new general expressions for calculating the pseudo inverse of a rectangular matrix (includes square matrices as a particular case) . The matrix may have any rank that does not exceed either the number of rows or columns, as it should. Since the matrices appearing in the resulting formulas have columns or rows which span the column or row space of the matrix, there is considerable leeway which permits the user to choose bases which may have better properties than others, either because of ease of calculation or because of better numerical performance.

6. BIBLIOGRAPHICAL REFERENCES

[1] Dahlquist, G. & A. Bjork (1974): *Numerical Methods.* Prentice – Hall, Inc., Englewood Cliffs, NJ.

[2] Golub, G. H. & C. Reinsch (1971): “Singular Value Decomposition and Least Squares Solutions.” Contribution 1/10 in Wilkinson, J. H. & C. Reinsch: *Handbook for Automatic Computation. Volume II. Linear Algebra.* Springer – Verlag, New York. pp. 134 – 151.

[3] Noble, B. (1969): *Applied Linear Algebra.* Prentice – Hall, Inc., Englewood Cliffs, NJ.

[4] Moore, E. H. (1920): “On the Reciprocal of the General Algebraic Matrix.” *Bulletin of the American Mathematical Society, Vol. 26,* pp. 394 – 395.

[5] Penrose, R. (1955): “A Generalized Inverse for Matrices.” *Proceedings of the Cambridge Philosophical Society, Vol. 51*, pp. 406 – 413.

[6] Lanczos, C. (1961): *Linear Differential Operators.* D. Van Nostrans Company Limited, Londres.

[7] Zadeh, L. & C. Desoer (1963): *Linear System Theory: The State Space Approach.* Mc Graw – Hill Book Company, Inc., New York

[8] Wolfram, S. (1991): *Mathematica: A System for Doing Mathematics by Computer. Second Edition.* Addison – Wesley Publishing Company, Inc., Redwood City, CA.

Authors Biography

Marco Antonio Murray Lasso

He is a mechanical – electrical engineer from Universidad Nacional Autónoma de México (UNAM) (1960) and has a MSc in EE (1962) and ScD in Automatic Control (1965), both from MIT. He has taught at the Newark College of Engineering and at Case Western Reserve University and, for 45 years, at UNAM. He was the Founding President of the Mexican Society for Computers in Education and of the Mexican Chapter of the Society of Systems, Man and Cybernetics of the IEEE, Secretary of the Mexican Academy of Science and Founding President of the National Academy of Engineering of Mexico and President of the Council of Academies of Engineering and Technological Sciences (CAETS). He has been a national researcher and has chaired several faculty evaluation committees at UNAM. He has authored or co – authored about 300 papers in conference proceedings and journals and 10 books. He has been a consultant to several departments in the Mexican government, the Mexican Senate, the International Bureau of Informatics and UNESCO. He is a member of several academies including the Mexican Academy of Sciences, Arts, Technology and Humanities at which he was the first President of the Honor Council. For 27 years, he has been a member of the Educational Council at MIT. He has been included in *Who is Who in Science and Engineering* and *Who is Who in the World.*