THE PENROSE-MOORE PSEUDO INVERSE WITH
DIVERSE STATISTICAL APPLICATIONS
Part I: The General Theory
and Computational Methods

BY
ARTHUR ALBERT

TECHNICAL REPORT NO. 173
APRIL 1, 1971

THIS RESEARCH WAS SPONSORED BY THE ARMY RESEARCH OFFICE
OFFICE OF NAVAL RESEARCH, AND AIR FORCE OFFICE OF
SCIENTIFIC RESEARCH BY CONTRACT NO.
N00014-67-A-0112-0053 (NR-042-267)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE PENROSE-MOORE PSEUDO INVERSE WITH
DIVERSE STATISTICAL APPLICATIONS

Part I: The General Theory
and Computational Methods

by

ARThUR ALBERT

TECHNICAL REPORT NO. 173
April 1, 1971

PREPARED UNDER CONTRACT N00014-67-A-0112-0053
NR-042-267
OFFICE OF NAVAL RESEARCH

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.  Introduction</td>
<td>1</td>
</tr>
<tr>
<td>II. General Background Material</td>
<td>3</td>
</tr>
<tr>
<td>III. Geometric and Analytic Properties of the Penrose-Moore Pseudo Inverse</td>
<td>15</td>
</tr>
<tr>
<td>IV. Pseudo Inverses of Partitioned Matrices and Sums and Products of Matrices</td>
<td>53</td>
</tr>
<tr>
<td>V.  Computational Methods</td>
<td>71</td>
</tr>
<tr>
<td>Bibliography</td>
<td>B1</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

In 1920, Moore [1], introduced the notion of a generalized inverse for matrices. The idea apparently lay dormant for about thirty years, whereupon a renaissance of interest occurred. As may be appreciated after a brief look through the bibliography, a large and varied literature has appeared since the early 1950's.

At the present time, the theory is elegant, the applications are diverse (e.g., least squares, linear equations, projections, statistical regression analysis, filtering and linear programming) and most important, a deeper understanding of these topics is achieved when they are studied in the generalized inverse-context. The results which we present here are, for the most part, scattered throughout the periodical literature or to be found in out of print technical reports. This is the major impetus behind the writing of this monograph.

The level of the material presupposes a familiarity with the notion of "limit" and some of the fundamental properties of finite dimensional Euclidean Spaces. That much will suffice for the first half of the book. The second half is devoted to statistical applications and for these, it would be helpful to have had a first course in probability (and/or statistics).

Many exercises are included. Most are accompanied by solution outlines at the end of the book. Some of the exercises are supplementary to the material in the monograph while others are lemmas to subsequent theorems. The reader is urged to do the exercises. They are the key to a thorough understanding of the material.

Sections and equations are numbered according to a decimal system.
For example, equation (6.3.1) comes after section (6.2) and before definition (6.3.1.9). The last comes before equation (6.4). Every effort has been made to maintain the following typographical conventions:

Sets - upper case script
Matrices - upper case Latin
Vectors - lower case Latin
Scalars - lower case Greek
Vector random variables - lower case Latin bold face
Real random variables - lower case Greek bold face.

Bibliographic references are enclosed in square brackets; equation and section numbers are written in parentheses at the left. Sometimes a section or an equation number appears in the middle of a line or at the right side. In this case, the assertion preceding this equation number is a direct consequence of that (previously established) result.
II. GENERAL BACKGROUND MATERIAL

In this chapter we will review the key results and definitions from the theory of real linear spaces which are relevant to what follows. This survey is informal and presumes previous exposure of the reader to these notions. (For a more formal review, see Appendix A of Karlin [1]. For a rigorous development, see Halmos [1], or the first half of Bellman [1].)

We begin by reminding the reader that linear transformations from one Euclidean Space to another can be represented by matrices, once the coordinate systems for the two spaces have been decided upon. Furthermore, vectors can be represented as long skinny matrices (having one column). If \( A \) is any matrix, we denote the transpose of \( A \) by \( A^t \). In our discussions, all matrices, vectors (and scalars) have real entries. Matrix transposition has the following important properties:

\[
(AB)^t = B^t A^t, \quad (A^t)^t = A, \quad (A+B)^t = A^t + B^t.
\]

If \( x \) and \( y \) are two vectors having the same number of components, the inner product (or scalar product) of \( x \) with \( y \) is the scalar \( x^t y \) (which is the same as \( y^t x \)). The norm of \( x \) is \( \|x\| = (x^t x)^{1/2} \). Two vectors are said to be orthogonal if their inner product vanishes. ("\( x \) is orthogonal to \( y \)" is abbreviated "\( x \perp y \).")

A linear manifold, \( \mathcal{L} \), is a non-empty subset of a Euclidean Space, which is closed under addition and scalar multiplication (if \( x \) and \( y \) are elements of \( \mathcal{L} \) then for any scalars \( \alpha \) and \( \beta \), \( \alpha x + \beta y \) are members of \( \mathcal{L} \).)
A vector, $x$, is orthogonal to the linear manifold, $\mathcal{L}$, if $x$ is orthogonal to every vector in $\mathcal{L}$. (Abbreviated "$x \perp \mathcal{L}$".)

The symbol "\(\epsilon\)" will be reserved for set membership ("$x \in \mathcal{L}$" means that $x$ is a member of the set $\mathcal{L}$.) The symbol "\(\subseteq\)" denotes set inclusion and "\(\subset\)" denotes proper inclusion.

The following Theorem is of fundamental importance in all that follows. We state it without proof:

(2.1) **Theorem:** Let $x$ be a vector in a finite dimensional Euclidean space and let $\mathcal{L}$ be a linear manifold in that space. Then, there is a unique vector, $\hat{x} \in \mathcal{L}$, having the property that $x - \hat{x} \perp \mathcal{L}$.

(2.1.1) **Comment:** An equivalent statement of the theorem is that there is a unique decomposition of $x$:

$$x = \hat{x} + \tilde{x}$$

where $\hat{x} \in \mathcal{L}$ and $\tilde{x} \perp \mathcal{L}$.

The vector, $\hat{x}$, is called the **projection** of $x$ on $\mathcal{L}$. It is the vector in $\mathcal{L}$ which is "nearest" to $x$, as we will now demonstrate:

(2.2) **Theorem:** Let $x$ be a vector and $\mathcal{L}$ a linear manifold. If $x = \hat{x} + \tilde{x}$ where $\hat{x} \in \mathcal{L}$ and $\tilde{x} \perp \mathcal{L}$, then

$$\|x - y\| > \|x - \hat{x}\|$$

if $y \in \mathcal{L}$ and $y \neq \hat{x}$.

**Proof:** If $y \in \mathcal{L}$ then

$$\|x - y\|^2 = \|\hat{x} + \tilde{x} - y\|^2 = \|(\hat{x} - y) + \tilde{x}\|^2$$

$$= \|\hat{x} - y\|^2 + \|\tilde{x}\|^2$$
since $\tilde{x} \perp \mathcal{L}$ and $\tilde{x} - y \in \mathcal{L}$. Therefore

$$\|x - y\|^2 \geq \|\tilde{x}\|^2$$

with strict inequality holding unless $\|\tilde{x} - y\|^2 = 0$. \hfill \Box

(2.1) and (2.2) are "existence theorems". As a consequence of the next theorem we can show how to reduce the computation of $\tilde{x}$ to the solution of simultaneous linear equations. First though, we remind the reader that a linear manifold, $\mathcal{L}$, is spanned by $(y_1, y_2, \ldots, y_n)$ if every vector in $\mathcal{L}$ can be expressed as a linear combination of the $y_j$'s.

(2.3) **Theorem:** a) If $x$ is a vector and $\mathcal{L}$ is a linear manifold, then $\tilde{x}$, the projection of $x$ on $\mathcal{L}$, is the unique vector in $\mathcal{L}$ satisfying the equations

$$(2.3.1) \quad \tilde{x}^t y = x^t y \quad \text{for all} \quad y \in \mathcal{L}. $$

b) If $\mathcal{L}$ is spanned by $y_1, y_2, \ldots, y_n$, $\tilde{x}$ is the unique vector in $\mathcal{L}$ satisfying

$$(2.3.2) \quad \tilde{x}^t y_j = x^t y_j \quad j = 1, 2, \ldots, n. $$

**Proof:** Part (a) follows directly from (2.1). That $\tilde{x}$ satisfies (2.3.2) is a consequence of (a). If $x^*$ is some other vector in $\mathcal{L}$ satisfying

$$x^*^t y_j = x^t y_j \quad j = 1, 2, \ldots, n,$$

then

$$(2.3.3) \quad (x^* - \tilde{x})^t y_j = 0 \quad j = 1, 2, \ldots, n.$$
Since the $y_j$'s span $\mathcal{L}$, any vector in $\mathcal{L}$ is a linear combination of the $y_j$'s, so that $x^* - \hat{x}$ is orthogonal to every vector in $\mathcal{L}$. Since $x^* - \hat{x} \in \mathcal{L}$, it follows that $(x^* - \hat{x})^t (x^* - \hat{x}) = \|x^* - \hat{x}\|^2 = 0$. Therefore, $x^*$ must coincide with $\hat{x}$ if it lies in $\mathcal{L}$ and satisfies (2.3.2).

(2.4) **Exercise:** If $\mathcal{L}$ is spanned by $y_1, y_2, \ldots, y_n$, then $\hat{x}$ is the unique vector of the form

$$\hat{x} = \sum_{j=1}^{n} \alpha_j y_j$$

where the $\alpha_j$'s are any scalars satisfying the simultaneous set of linear equations

(2.4.1) \hspace{1cm} \sum_{j=1}^{n} \alpha_j (y_j^t y_j) = y_j^t x \hspace{1cm} i = 1, 2, \ldots, n.

(2.5.1) **Exercise:** If $\mathcal{L}$ is spanned by $y_1, y_2, \ldots, y_n$ then $x \perp \mathcal{L}$ if $x$ is orthogonal to each of the $y_j$'s.

(2.5.2) **Exercise:** If $x$ and $y$ are vectors in the same Euclidean Space and $\mathcal{L}$ is a linear manifold in that space, then the projection of $\alpha x + \beta y$ on $\mathcal{L}$ is $\hat{\alpha} x + \hat{\beta} y$ where $\hat{x}$ and $\hat{y}$ are the projections of $x$ and $y$ on $\mathcal{L}$.

If $y_1, y_2, \ldots, y_n$ is a set of vectors in the same Euclidean Space, we denote the linear manifold spanned by the $y_j$'s by $\mathcal{L}(y_1, y_2, \ldots, y_n)$. This manifold is the smallest manifold containing all the $y_j$'s and it consists, exactly, of all those vectors which are expressible as a linear combination of the $y_j$'s.

(2.5.3) **Exercise:** The projection of $x$ on $\mathcal{L}(y)$ is $(x^t y) y / \|y\|^2$ if $y \neq 0$. 6
If \( x, y_1, y_2, \ldots, y_n \) is an arbitrary set of vectors in the same Euclidean Space, a particularly simple relationship exists between \( \hat{x}_n \), the projection of \( x \) on \( \mathcal{L}(y_1, \ldots, y_n) \), and \( \hat{x}_{n-1} \), the projection of \( x \) on \( \mathcal{L}(y_1, \ldots, y_{n-1}) \), provided that \( y_n \) is orthogonal to the previous \( y_j \)'s:

(2.6) Theorem: If \( x, y_1, \ldots, y_n \) are vectors in the same Euclidean Space and if \( y_n \perp \mathcal{L}(y_1, \ldots, y_{n-1}) \) then

\[
\hat{x}_n = \hat{x}_{n-1} + \begin{cases} 
0 & \text{if } y_n = 0 \\
\frac{x^t y_n}{\|y_n\|^2} & \text{otherwise.}
\end{cases}
\]

Proof: Since \( \hat{x}_{n-1} \in \mathcal{L}(y_1, \ldots, y_{n-1}) \), \( \hat{x}_n \) is clearly a member of \( \mathcal{L}(y_1, \ldots, y_n) \). It is readily verified that the right side of (2.6.1) satisfies (2.3.2), provided \( y_n \) is orthogonal to \( \mathcal{L}(y_1, \ldots, y_{n-1}) \) (and in particular, orthogonal to \( \hat{x}_{n-1} \)). The conclusion follows from (2.3b). \( \Box \)

As an immediate consequence we can derive the so-called Fourier Expansion Theorem:

(2.7) Theorem: If \( u_1, u_2, \ldots, u_n \) are mutually orthogonal vectors of unit length and \( x \) is an arbitrary vector in the same Euclidean Space then \( \hat{x} \), the projection of \( x \) on \( \mathcal{L}(u_1, \ldots, u_n) \) is given by

\[
(2.7.1) \quad \hat{x} = \left( \sum_{j=1}^{n} u_j^t u_j \right)x
\]

\[
(2.7.2) \quad = \sum_{j=1}^{n} (u_j^t x) u_j .
\]
Comment: If the \( u_j \)'s are \( k \) dimensional vectors, the expression
\[
\sum_{j=1}^{n} u_j u_j^t
\]
is a \( k \times k \) matrix, so that (2.7.1) is a representation of the
matrix which projects \( x \) onto \( \mathcal{L}(u_1, \ldots, u_n) \). (2.7.2) on the other
hand, is an explicit representation of \( \tilde{x} \) as a linear combination of
the \( u_j \)'s.

If \( \mathcal{L}_1 \) and \( \mathcal{L}_2 \) are linear manifolds and \( \mathcal{L}_1 \subseteq \mathcal{L}_2 \), define
\( \mathcal{L}_2^\perp \mathcal{L}_1 \) as the set of vectors in \( \mathcal{L}_2 \) which are orthogonal to \( \mathcal{L}_1 \).
(2.7.3) **Exercise:** If \( \mathcal{L}_1 \subseteq \mathcal{L}_2 \), then \( \mathcal{L}_2^\perp \mathcal{L}_1 \) is a linear
manifold.

(2.7.4) **Exercise:** Let \( x \) be a vector and suppose \( \mathcal{L}_1 \subseteq \mathcal{L}_2 \). Define
\( \tilde{x}_2 \) to be the projection of \( x \) on \( \mathcal{L}_2 \), \( \tilde{x}_{21} \) to be the projection of
\( \tilde{x}_2 \) on \( \mathcal{L}_1 \), \( \tilde{x}_1 \) to be the projection of \( x \) on \( \mathcal{L}_1 \) and \( \tilde{x}_{21} \) to be
the projection of \( x \) on \( \mathcal{L}_2^\perp \mathcal{L}_1 \).

Then

a) \( \tilde{x}_{21} = \tilde{x}_1 \) (the projection of \( x \) on \( \mathcal{L}_1 \) is obtainable by
projecting \( x \) on \( \mathcal{L}_2 \) and then projecting that
vector on \( \mathcal{L}_1 \).

b) \( \tilde{x}_2 = \tilde{x}_1 + \tilde{x}_{21} \).

c) \( \|x - \tilde{x}_1\| \geq \|x - \tilde{x}_2\| \) with strict inequality holding if \( \mathcal{L}_1 \) is a
proper subset of \( \mathcal{L}_2 \) unless \( x \in \mathcal{L}_1 \).

(2.8) **The Gram-Schmidt Orthogonalization Procedure**

This procedure takes an arbitrary collection of vectors, \( h_1, h_2, \ldots, h_n \)
and generates a set of mutually orthogonal vectors, \( u_1, u_2, \ldots, u_n \),
having the properites that

(2.8.1) \( \mathcal{L}(u_1, u_2, \ldots, u_j) = \mathcal{L}(h_1, h_2, \ldots, h_j) \) for \( j = 1, \ldots, n \)
and

\[(2.8.2) \quad \|u_j\| = 1 \text{ if } u_j \neq 0 \quad j=1,2,\ldots,n.\]

The procedure:

\[(2.8.3) \quad u_1 = \begin{cases} \frac{h_1}{\|h_1\|} & \text{if } h_1 \neq 0 \\ 0 & \text{if } h_1 = 0. \end{cases}\]

For \(j=1,2,\ldots,n-1\), define

\[(2.8.4) \quad \hat{h}_{j+1} = \frac{1}{\|h_j\|} \sum_{k=1}^{j} (h_{j+1}^t u_k) u_k \]

and

\[(2.8.5) \quad u_{j+1} = \begin{cases} \frac{(h_{j+1} - \hat{h}_{j+1})/\|h_{j+1} - \hat{h}_{j+1}\|}{h_{j+1} - \hat{h}_{j+1}} & \text{if } h_{j+1} - \hat{h}_{j+1} \neq 0 \\ 0 & \text{otherwise}. \end{cases}\]

The properties (2.8.1) and (2.8.2) are established by induction:

The induction hypothesis is that

\[\mathcal{L}(u_1, u_2, \ldots, u_j) = \mathcal{L}(h_1, \ldots, h_j)\]

and

\[u_{j+1} \perp \mathcal{L}(u_1, \ldots, u_j).\]

By definition of \(u_1\), \(\mathcal{L}(u_1) = \mathcal{L}(h_1)\) and by (2.5.3), \(\hat{h}_2\) is the projection of \(h_2\) on \(\mathcal{L}(u_1)\). By (2.1), \(h_2 - \hat{h}_2 \perp \mathcal{L}(u_1)\) so that \(u_2\) is orthogonal to \(\mathcal{L}(u_1)\). This establishes the induction hypothesis for \(j=1\). If it is assumed that the hypothesis is true for all values of
j up through k, then since \( u_{k+1} \) is a linear combination of \( h_{k+1} \) and \( \hat{h}_{k+1} \) (which lies in \( L(u_1, \ldots, u_k) = L(h_1, \ldots, h_k) \)), we see that any vector which is a linear combination of \( u_1, \ldots, u_{k+1} \) is also a linear combination of \( h_1, \ldots, h_{k+1} \). This means that \( L(u_1, \ldots, u_{k+1}) \subseteq L(h_1, \ldots, h_{k+1}) \).

On the other hand,

\[
h_{k+1} = \|h_{k+1} - \hat{h}_{k+1}\| u_{k+1} + \hat{h}_{k+1},
\]

the right side being a linear combination of the non-zero members of \( \{u_1, u_2, \ldots, u_{k+1}\} \). Since \( L(h_1, \ldots, h_k) = L(u_1, \ldots, u_k) \) under the induction hypothesis, any vector which is expressible as a linear combination of \( h_1, \ldots, h_{k+1} \), is also expressible as a linear combination of \( u_1, \ldots, u_{k+1} \). Therefore, \( L(h_1, \ldots, h_{k+1}) \subseteq L(u_1, \ldots, u_{k+1}) \). This establishes the first half of the induction hypothesis for \( j = k+1 \).

The second half follows since \( \hat{h}_{k+2} \) is the projection of \( h_{k+2} \) on \( L(u_1, \ldots, u_{k+1}) \), (2.7.2). By (2.1), \( h_{k+2} - \hat{h}_{k+2} \perp L(u_1, \ldots, u_{k+1}) \) and therefore, so is \( u_{k+2} \).

(2.8.6) Exercise: \( u_j = 0 \) if and only if \( h_j \) is a linear combination of \( (h_1, \ldots, h_{j-1}) \).

In what follows, two special linear manifolds will be of particular interest: If \( H \) is any matrix, the null space of \( H \), denoted by \( \nabla(H) \), is the set of vectors which \( H \) maps into zero:

\[
\nabla(H) = \{ x : Hx = 0 \}.
\]

(\( \nabla(H) \) always has at least one element, namely the null vector, \( 0 \).)

The range of \( H \), denoted by \( \Omega(H) \), is the set of vectors which are after images of vectors in the Euclidean Space which serves as the
domain of $H$:

$$\mathcal{R}(H) = \{z: z = Hx \text{ for some } x\}$$

It is easy to see that $\mathcal{N}(H)$ and $\mathcal{R}(H)$ are linear manifolds.

(2.9.1) Exercise: Let the column vectors of $H$ be denoted by $h_1, h_2, \ldots, h_n$. Show that $\mathcal{R}(H) = \mathcal{L}(h_1, h_2, \ldots, h_n)$.

(2.9.2) Exercise: Show that $H^t$ is the adjoint of $H$. That is to say, if $H$ is an $n \times m$ matrix, then for any $m$ dimensional vector $x$ and any $n$ dimensional vector $y$, the inner product of $x$ with $Hy$ is the same as the inner product of $y$ with $H^tx$.

If $\mathcal{L}$ is a linear manifold in a Euclidean space $\mathcal{E}$, the orthogonal complement of $\mathcal{L}$ (denoted by $\mathcal{L}^\perp$) is defined to be the set of vectors in $\mathcal{E}$ which are (each) orthogonal to $\mathcal{L}$.

It is easy to see that $\mathcal{L}^\perp$ is itself a linear manifold.

(2.9.3) Exercise:

$$(\mathcal{L}^\perp)^\perp = \mathcal{L}$$

(2.9.4) Exercise: If $x$ is a vector in $\mathcal{E}$ and $x^ty = 0$ for all $y \in \mathcal{E}$, then $x = 0$.

The null space of a matrix is related to the range space of its transpose. In fact, the next theorem shows that the null space of $H$ consists of those vectors which are orthogonal to the column vectors of $H^t$ (i.e., the rows of $H$) which is just another way of saying

(2.10) Theorem: For any matrix $H$, $\mathcal{N}(H) = \mathcal{R}^+(H^t)$.

Proof: $x \in \mathcal{N}(H)$ if and only if $Hx = 0$. Therefore, $x \in \mathcal{N}(H)$ if and only if $y^tHx = 0$ for all $y$ (having the correct number of components, of course). (Use (2.9.4)). Since $y^tHx = (H^ty)^tx$, we see that $Hx = 0$ if and only if $x$ is orthogonal to all vectors of the form $H^ty$. These
vectors, collectively, make up \( \mathcal{R}(H^t) \), thus proving the assertion.

By applying the projection theorem (2.1), we deduce as an immediate consequence of (2.10), that every vector \( z \) (having the correct number of components) has a unique decomposition as the sum of two terms, one lying in \( \mathcal{R}(H) \) and one lying in \( \mathcal{N}(H^t) \):

(2.11) **Theorem:** If \( H \) is an \( n \times m \) matrix and \( z \) is an \( n \)-dimensional vector, we can uniquely decompose \( z \):

\[
z = \hat{z} + \tilde{z}
\]

where \( \hat{z} \in \mathcal{R}(H) \) and \( \tilde{z} \in \mathcal{N}(H^t) \).

(2.11.1) **Exercise:** In (2.11), \( \hat{z} \) is the projection of \( z \) on \( \mathcal{R}(H) \) and \( \tilde{z} \) is the projection of \( z \) on \( \mathcal{N}(H^t) \). Consequently \( H^t \hat{z} = H^t z \).

A matrix is said to be **symmetric** if it is equal to its transpose. Obviously, symmetric matrices are square.

(2.11.2) **Exercise:** Matrices of the form \( H^tH \) and \( HH^t \) are always symmetric.

By virtue of (2.10),

(2.11.3) \[
\begin{align*}
\mathcal{N}(A) &= \mathcal{R}^\perp(A) \\
\mathcal{R}(A) &= \mathcal{N}^\perp(A)
\end{align*}
\]

if \( A \) is symmetric.

Moreover, if \( H \) is any matrix, then

(2.12) **Theorem:** \( \mathcal{R}(H) = \mathcal{R}(HH^t) \), \( \mathcal{R}(H^t) = \mathcal{R}(H^tH) \), \( \mathcal{N}(H) = \mathcal{N}(H^tH) \), \( \mathcal{N}(H^t) = \mathcal{N}(HH^t) \).

Proof: It suffices to prove that \( \mathcal{N}(H^t) = \mathcal{N}(HH^t) \) and \( \mathcal{N}(H) = \mathcal{N}(H^tH) \). Then apply (2.10) and (2.9). To prove that \( \mathcal{N}(H^t) = \mathcal{N}(HH^t) \), we note
that $HH^tx=0$ if $H^tx=0$. On the other hand, if $HH^tx=0$, then $x^tHH^tx=0$ so that $\|H^tx\|^2=0$ which implies that $H^tx=0$. Thus $H^tx=0$ if and only if $HH^tx=0$. The same proof applies to show that $\nabla(H) = \nabla(H^tH)$. 

A square matrix is nonsingular if its null space consists only of the zero vector. If a square matrix is not nonsingular, it is called singular.

(2.12.1) Exercise: If the row vectors of $H$ are linearly independent, then the null space of $H^t$ consists of the zero vector.

(2.12.2) Exercise: Let $h_1, h_2, \ldots, h_n$ be a linearly independent set of vectors. Let $G$ be the $n \times n$ matrix whose $(i,j)$-th entry is $h_i^t h_j$ ($G$ is known as a Grammian.) Show that $G$ is nonsingular. (Hint: $G=HH^t$ where $H$ is the matrix whose rows are $h_1^t, h_2^t, \ldots, h_n^t$. Now apply (2.12.1) and (2.12.2))

If $A$ is a nonsingular matrix, there is a unique matrix, $A^{-1}$, which is the left and right inverse of $A$:

$$A(A^{-1}) = (A^{-1})A = I$$

where $I$ is the identity matrix.

(2.13) Theorem: If $H$ is any matrix and $\delta$ is nonzero, then $H^tH+\delta^2I$ is nonsingular.

Proof: If $(H^tH+\delta^2I)x=0$, then $0=x^t(H^tH+\delta^2I)x = \|Hx\|^2 + \delta^2 \|x\|^2$

which can only occur if $x=0$. 

We close this chapter with a statement of the celebrated diagonalization theorem for symmetric matrices. The proof may be found in Bellman [1].

A (possibly complex) number $\lambda$ is called an eigenvalue of the (square) matrix, $A$, if $A-\lambda I$ is singular.
(2.13.1) Exercise: If A is real and symmetric, its eigenvalues are real. (Hint: If $(A-\lambda I)x=0$ then $(A-\bar{\lambda} I)\bar{x}=0$ where $\bar{\lambda}$ and $\bar{x}$ are the complex conjugates of $\lambda$ and $x$.)

(2.14) Theorem: If A is real and symmetric with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then there is a matrix T such that $T^t = T^{-1}$ and

$$T^tAT = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

(The term "diag($\lambda_1, \ldots, \lambda_n$)" refers to a diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$. If $T^t = T^{-1}$, T is said to be an orthogonal matrix.)

(2.14.1) Exercise: If T is an orthogonal matrix, the rows of T are mutually orthogonal and have unit length. So too, are the columns.
III. GEOMETRIC AND ANALYTIC PROPERTIES OF THE PENROSE-MOORE PSEUDO INVERSE

We begin our treatment of the pseudo inverse by characterizing the minimum norm solution to the classical least squares problem:

(3.1) Theorem: Let \( z \) be an \( n \) dimensional vector and let \( H \) be an \( n \times m \) matrix.

a) There is always a vector, in fact a unique vector, \( \hat{x} \), of minimum norm, which minimizes

\[
\| z - H \hat{x} \|^2
\]

b) \( \hat{x} \) is the unique vector in \( \mathcal{R}(H^t) \) which satisfies the equation

\[
H \hat{x} = \hat{z}
\]

where \( \hat{z} \) is the projection of \( z \) on \( \mathcal{R}(H) \).

Proof: By (2.11) we can write

\[
z = \hat{z} + \tilde{z}
\]

where \( \tilde{z} \) is the projection of \( z \) on \( \mathcal{N}(H^t) \). Since \( \text{Hxe} \mathcal{R}(H) \) for every \( x \), it follows that

\( 2 - \text{Hxe} \mathcal{R}(H) \) and since \( \tilde{z} \in \mathcal{R}(H)^\perp \), \( \tilde{z} \perp 2 - \text{Hx} \).

Therefore

\[
\| z - \text{Hx} \|^2 = \| 2 - \text{Hx} + \tilde{z} \|^2 = \| 2 - \text{Hx} \|^2 + \| \tilde{z} \|^2 \geq \| \tilde{z} \|^2.
\]

This lower bound is attainable since \( \hat{z} \), being in the range of \( H \), is the after-image of some \( x_o \):

\[
\hat{z} = Hx_o.
\]
Thus, for this $x_0$, the bound is attained:

$$\|z-Hx_0\|^2 = \|z-\hat{\alpha}\|^2 = \|\tilde{z}\|^2.$$ 

On the other hand, we just showed that

$$\|z-Hx\|^2 = \|\hat{\alpha}-Hx\|^2 + \|\tilde{\alpha}\|^2$$

so that the lower bound is attained at $x^*$ only if $x^*$ is such that $Hx^* = \hat{\alpha}$. For any such $x^*$, we can decompose it via (2.11) into two orthogonal vectors:

$$x^* = \hat{x}^* + \tilde{x}^*,$$

where

$$\hat{x}^* \in \mathcal{R}(H^t) \quad \text{and} \quad \tilde{x}^* \in \mathcal{N}(H).$$

Thus

$$Hx^* = H\hat{x}^* \quad \text{so that} \quad \|z-Hx^*\|^2 = \|z-H\hat{x}^*\|^2$$

and

$$\|x^*\|^2 = \|\hat{x}^*\|^2 + \|\tilde{x}^*\|^2 \geq \|\hat{x}^*\|^2$$

with strict inequality unless $x^* = \hat{x}^*$ (i.e., unless $x^*$ coincides with its projection on $\mathcal{R}(H^t)$--which is to say, unless $x^* \in \mathcal{R}(H^t)$ to begin with.)

So far we have shown that $x_0$ minimizes $\|z-Hx\|^2$ if and only if $Hx_0 = \hat{\alpha}$, and that, among those vectors which minimize $\|z-Hx\|^2$, any vector of minimum norm must lie in the range of $H^t$. To demonstrate the uniqueness of this vector, suppose that $\hat{\alpha}$ and $x^*$ both are in $\mathcal{R}(H^t)$ and that

$$H\hat{x} = Hx^* = \hat{\alpha}.$$
Then

\[ x^* - \hat{x} \in Q(H^t) . \]

But

\[ H(x^* - \hat{x}) = 0 \]

so that

\[ x^* - \hat{x} \notin \nabla(H) \]

\[ = Q \perp (H^t) \text{ as well, (2.10).} \]

Thus \( x^* - \hat{x} \) is orthogonal to itself, which means that \( \|x^* - \hat{x}\|^2 = 0. \)

(i.e., \( x^* = \hat{x} \).)

(3.1.1) Comment: An alternate statement of (3.1) which is equivalent but perhaps more illuminating is this:

There is always an \( n \) dimensional vector \( y \) such that

\[ \|z - HH^t y\|^2 = \inf_{x} \|z - Hx\|^2 . \]

If

\[ \|z - Hx_0\|^2 = \inf_{x} \|z - Hx\|^2 \]

then

\[ \|x_0\| \geq \|H^t y\|, \text{ with strict inequality holding unless } x_0 = H^t y. \]

\( y \) satisfies the equation

\[ HH^t y = \hat{z} , \]

where \( \hat{z} \) is the projection of \( z \) on \( Q(H) \).

(3.1.2) Exercise: \( \|z - Hx\|^2 \) is minimized by \( x_0 \) if and only if \( Hx_0 = \hat{z} \), where \( \hat{z} \) is the projection of \( z \) on \( Q(H) \).

The minimal least squares solution alluded to in (3.1) can be characterized as a solution to the so-called normal equations:
(3.2) **Theorem:** Among those vectors, \( x \), which minimize \( \|z - Hx\|^2 \), \( \hat{x} \), the one having minimum norm, is the unique vector of the form

\[
(3.2.1) \quad \hat{x} = H^t y
\]

which satisfies

\[
(3.2.2) \quad H^t H \hat{x} = H^t z.
\]

**Comment:** The theorem says that \( \hat{x} \) can be obtained by finding any vector, \( y_0 \), which satisfies the equation

\[
H^t H^t y = H^t z
\]

and then taking

\[
\hat{x} = H^t y_0.
\]

**Proof:** By (2.12), \( \mathcal{R}(H^t) = \mathcal{R}(H^t H) \). Since \( H^t z \) is in the range of \( H^t \), it must therefore be in the range of \( H^t H \) and so, must be the after image of some \( x \) under the transformation \( H^t H \). In other words, (3.2.2) always has at least one solution in \( x \). If \( x \) is a solution to (3.2.2), so then is \( \hat{x} \), the projection of \( x \) on \( \mathcal{R}(H^t) \), since \( Hx = H\hat{x} \), (2.11.1). Since \( \hat{x} \in \mathcal{R}(H^t) \), it is the after image of some vector \( y \) under \( H^t \):

\[
\hat{x} = H^t y.
\]

So far we have shown that there is at least one solution to (3.2.2) of the form (3.2.1). To show uniqueness, suppose

\[
\hat{x}_1 = H^t y_1 \quad \text{and} \quad \hat{x}_2 = H^t y_2
\]

both satisfy (3.2.2). Then
\[ H^t H(H^t y_1 - H^t y_2) = 0 \quad \text{so that} \]
\[ H^t (y_1 - y_2) \in \mathcal{N}(H^t H) = \mathcal{N}(H), \quad (2.12) \]

which implies that
\[ HH^t (y_1 - y_2) = 0. \]

Thus
\[ (y_1 - y_2) \in \mathcal{N}(HH^t) = \mathcal{N}(H^t). \quad (2.12) \]

and so
\[ \hat{x}_1 = H^t y_1 = H^t y_2 = \hat{x}_2. \]

Thus, there is exactly one solution to (3.2.2) of the form (3.2.1). If we can show that this solution also satisfies the equation
\[ Hx = \hat{z} \]

where \( \hat{z} \) is the projection of \( \hat{z} \) on \( \mathcal{R}(H) \) then, by virtue of (3.1b) we will be done.

But, in (2.11.1) we showed that
\[ (3.2.3) \quad H^t z = H^t \hat{z}. \]

In (3.1) we showed that there is a unique solution in \( \mathcal{R}(H^t) \) to the equation
\[ (3.2.4) \quad Hx = \hat{z}. \]

This (unique) solution therefore satisfies the equation
\[ H^t Hx = H^t \hat{z} \]

as well. Since \( H^t z = H^t \hat{z} \), (3.2.3), we see that the (unique) solution
to (3.2.4) which lies in $\mathcal{Q}(H^t)$ must coincide with $\hat{x}$, the unique solution to (3.2.2) which lies in $\mathcal{Q}(H^t)$. In summary, the vector $\hat{x}$ alluded to in the statement of (3.2) coincides exactly with the vector $\hat{x}$ alluded to in (3.1). ■

We are now in a position to exhibit an explicit representation for the minimal norm solution to a least squares problem. A preliminary lemma is needed which, for the sake of continuity, is stated here and proved later:

(3.3) **Lemma:** For any real symmetric matrix, $A$,

$$P_A = \lim_{\delta \to 0} (A + \delta I)^{-1}A = \lim_{\delta \to 0} A(A + \delta I)^{-1}$$

always exists. For any vector $z$,

$$\hat{z} = P_A z$$

is the projection of $z$ on $\mathcal{Q}(A)$.

(3.4) **Theorem:** For any $n \times m$ matrix, $H$,

(3.4.1) $$H^t = \lim_{\delta \to 0} (H^t H + \delta^2 I)^{-1} H^t$$

(3.4.2) $$= \lim_{\delta \to 0} H^t (HH^t + \delta^2 I)^{-1}$$

always exists. For any $n$-vector $z$,

$$\hat{x} = H^t z$$

is the vector of minimum norm among those which minimize

$$\|z - Hx\|^2.$$
Comment: Here and hereafter, we use the symbol "I" to represent the identity matrix, whose dimensionality is to be reckoned from the context. For example, in the expression $H^tH+I$, we refer to the $m \times m$ identity, whereas in $HH^t+I$, we refer to the $n \times n$ identity.

Proof: Since

$$(H^tHH^t+\delta^2I^t) = H^t(HH^t+\delta^2I^t) = (H^tH+\delta^2I)H^t$$

and since $(HH^t+\delta^2I)$ and $(H^tH+\delta^2I)$ have inverses when $\delta^2 > 0$, (2.13), it is clear that the right sides of (3.4.1) and (3.4.2) are equal if either exists.

Let $z$ be a given $n$ dimensional vector, and decompose $z$ into its projections on $\mathcal{Q}(H)$ and $\mathcal{N}(H^t)$ according to (2.11);

$$z = \hat{z} + \tilde{z}.$$ 

Since

$$H^t\tilde{z} = H^t\tilde{z},$$

(2.11.1),

and since $\hat{z} \in \mathcal{Q}(H)$ must be the after image of some vector $x_0$ under $H$, we see that

$$(H^tH+\delta^2I)^{-1}H^t\tilde{z} = (H^tH+\delta^2I)^{-1}H^t\tilde{z}$$

(3.4.3)

$$= (H^tH+\delta^2I)^{-1}H^tHx_0.$$ 

The limit of the last expression always exists and coincides with $x_0$, the projection of $x_0$ on $\mathcal{Q}(H^tH)$, by virtue of (3.3). Since

$\mathcal{Q}(H^t) = \mathcal{Q}(H^tH)$, (2.12), and since

$$\hat{z} = Hx_0$$

$$= Hx_0,$$ 

(2.11.1),

21
we conclude that

$$\hat{z}_0 = \lim_{\delta \to 0} (H^+H + \delta^2 I)^{-1} H^+ z,$$

always exists, is an element of $\mathcal{R}(H^+)$ and satisfies the relation

$$H\hat{z}_0 = \hat{z}$$

where $\hat{z}$ is the projection of $z$ on $\mathcal{R}(H)$. The desired conclusion follows directly from (3.1). $lacksquare$

(3.5) Corollary: For any vector $z$, $H^+ z$ is the projection of $z$ on $\mathcal{R}(H)$ and $(I - HH^+) z$ is the projection of $z$ on $\mathcal{N}(H^+)$. For any vector $x$, $H^+ H x$ is the projection of $x$ on $\mathcal{R}(H^+)$ and $(I - H^+ H) x$ is the projection of $x$ on $\mathcal{N}(H)$.

Proof: By (3.4.2), $H^+ = \lim_{\delta \to 0} (HH^+ + \delta^2 I)^{-1}$ and by (3.4.1),

$$H^+ H = \lim_{\delta \to 0} (H^+ H + \delta^2 I)^{-1} H^+ H.$$  (3.3) tells us that $HH^+ z$ is therefore the projection of $z$ on $\mathcal{R}(HH^+)$, which coincides with the projection of $z$ on $\mathcal{R}(H)$, (2.12). Similarly, (3.3) and (2.12) imply that $H^+ H x$ is the projection of $x$ on $\mathcal{R}(H^+) = \mathcal{R}(H^+)$.

Since $z - \hat{z}$ is the projection of $z$ on $\mathcal{N}(H^+)$ if $\hat{z}$ is the projection of $z$ on $\mathcal{R}(H)$, (2.11), it follows that

$$z - HH^+ z$$

is the projection of $z$ on $\mathcal{N}(H^+)$. By the same token,

$$(I - H^+ H) x$$

is the projection of $x$ on $\mathcal{N}(H)$, $lacksquare$

The matrix, $H^+$, which we explicitly define in (3.4), is the so-called "Penrose-Moore-Generalized Inverse for $H". We will usually refer to it more familiarly as "the pseudo inverse of $H". 

22
Corollary (3.5) is tremendously important and should be noted carefully. It expresses the four most fundamental projection operators in terms of pseudo inverses. The results of (3.4) are attributed (via a slightly different method of proof) to den Broeder and Charnes [1].

(3.5.1) **Exercise:** $H^* = H^{-1}$ if $H$ is square and nonsingular.

(3.5.2) **Exercise:** $H^* = H^t (H H^t)^{-1}$ if the rows of $H$ are linearly independent. (Hint: Apply (2.12.2) to show $(H H^t)$ has an inverse. Then use (3.4.2)).

(3.5.3) **Exercise:** $H^* = (H^t H)^{-1} H^t$ if the columns of $H$ are linearly independent.

Before proceeding to the light task of milking these results for all they are worth, we pause to furnish the proof for lemma (3.3), as promised:

**Proof of (3.3):** If $A$ is any symmetric matrix and $\delta_0$ is a nonzero scalar whose magnitude is less than the magnitude of $A$'s smallest nonzero eigenvalue, then for any $\delta$ with

$$0 < |\delta| < |\delta_0|,$$

$(A + \delta I)$ is nonsingular and hence, for all such $\delta$'s, $(A + \delta I)^{-1}$ exists.

If $z$ is any vector, we write

$$z = \hat{z} + \tilde{z}$$

where

$$\hat{z} \in \mathcal{Q}(A), \tilde{z} \in \mathcal{H}(A),$$

and

$$Az = A\hat{z},$$

(2.11.1)
Since $\hat{z} \in \mathcal{R}(A)$, we can write $\hat{z} = Ax_0$ for some $x_0$ and so

$$(A + 8I)^{-1}Az = (A + 8I)^{-1}A\hat{z}$$

$$= (A + 8I)^{-1}A(Ax_0)$$.

From (2.14), the diagonalization theorem, we conclude that

$$A = TDT^t$$

where

$$D = \text{diag} (\lambda_1, \lambda_2, \ldots, \lambda_n)$$

is the diagonal matrix of $A$'s eigenvalues and $T$ is an orthogonal matrix:

$$T^t = T^{-1}$$.

Thus

$$(A + 8I)^{-1}Az = (A + 8I)^{-1}A^2 x_0 = T(D + 8I)^{-1}D^2 T^t x_0$$.

Element-by-element, it is plain to see that

$$\lim_{\varepsilon \rightarrow 0} (D + 8I)^{-1}D^2 = D^\varepsilon = D^\varepsilon$$

so that

$$\lim_{\varepsilon \rightarrow 0} (A + 8I)^{-1}Az = TDT^t x_0 = Ax_0 = \hat{z}$$,
the projection of $z$ on $\mathcal{R}(A)$.

The same argument works for $\lim_{\varepsilon \to 0} A(A+\varepsilon I)^{-1}z$. $\blacksquare$

In (3.5.1)-(3.5.3), formulas for $H^+$ in terms of inverses were given for the case where the rows and/or the columns of $H$ are linearly independent. There are cases though where these conditions need not obtain:

$$H = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$ 

In such cases, $H^+$ does not have a simple formula in terms of inverses. However, a better understanding of $H^+$ can be had by treating, in turn, the following cases: $H$ a $1 \times 1$ matrix, $H$ diagonal, $H$ symmetric, $H$ rectangular:

If $H$ is a $1 \times 1$ (scalar) matrix, then

$$H^+ = \lim_{\varepsilon^2 \to 0} (H^2 + \varepsilon^2 I)H = \begin{cases} 0 & \text{if } H = 0 \\ 1/H & \text{if } H \neq 0 \end{cases}.$$ 

If $H$ is diagonal:

$$H = \text{diag} (\lambda_1^+, \lambda_2^+, \ldots, \lambda_m^+)$$

then

$$H^+ = \text{diag} (\lambda_1^+, \lambda_2^+, \ldots, \lambda_m^+)$$

where

$$\lambda_j^+ = \begin{cases} 0 & \lambda_j = 0 \\ 1/\lambda_j & \text{if } \lambda_j \neq 0 \end{cases}.$$ 

In terms of least squares, this makes sense, for if
\[
\begin{align*}
    z &= \left( \begin{array}{c} 
    \xi_1 \\
    \vdots \\
    \xi_n 
    \end{array} \right), \\
    H &= \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)
\end{align*}
\]

and

\[
\begin{align*}
    x &= \left( \begin{array}{c} 
    \xi_1 \\
    \vdots \\
    \xi_n 
    \end{array} \right), \\
    \text{is to be chosen to minimize}
\end{align*}
\]

\[
\|z - Hx\|^2 = \sum_{j=1}^{n} (\xi_j - \lambda_j \xi_j)^2,
\]

it is clear that the choice

\[
\xi_j^* = \begin{cases} 
    \frac{\xi_j}{\lambda_j} & \text{if } \lambda_j \neq 0 \\
    \text{arbitrary} & \text{if } \lambda_j = 0
\end{cases}
\]

will minimize the sum of squares, and that

\[
\|x^*\|^2 = \sum_{j=1}^{n} \xi_j^{*2}
\]

is made smallest when

\[
\xi_j^* = 0 \text{ if } \lambda_j = 0.
\]

Thus, the minimum norm solution for the case of a diagonal \( H \) is

\[
\hat{x} = \left( \begin{array}{c} 
    \hat{\xi}_1 \\
    \vdots \\
    \hat{\xi}_n 
    \end{array} \right),
\]

where

\[
\hat{\xi}_j = \lambda_j^+ \xi_j; \quad \text{i.e., } \hat{x} = H^+ z.
\]

26
(3.6) The special case of symmetric matrices.

If $H$ is a symmetric $m \times m$ matrix, the diagonalization theorem allows us to write

$$H = TDT^t,$$

where $T$ is an orthogonal matrix and $D$ is diagonal.

By (3.4),

$$H^+ = \lim_{\delta \to 0} T (D^2 + \delta^2 I)^{-1} D T^t = T \lim_{\delta \to 0} (D^2 + \delta^2 I)^{-1} D T^t = T D^+ T^t.$$

Thus, the pseudo inverse for a symmetric matrix is obtained by pseudo inverting the diagonal matrix of its eigenvalues. Since $H$ is nonsingular if and only if its eigenvalues are non-zero (in which case $D^+ = D^{-1}$) we see that

$$H^+ = T D^{-1} T^t$$

if $H$ is symmetric and nonsingular. Since $T T^t = T^t T = I$, it is easy to see that $H H^+ = H^+ H = I$ in this case, so that $H^+ = H^{-1}$.

The last result can be expressed in a different notation, and gives rise to various so-called spectral representation theorems (c.f. (3.15)): If the column vectors of $T$ are denoted by $t_i$ $(i=1, \ldots, m)$, so that we can write $T$ as a partitioned matrix,

$$T = (t_1 : t_2 : \ldots : t_m),$$

the diagonalization theorem states that

(3.6.1) \hspace{1cm} H = T D T^t = \sum_{j=1}^{m} \lambda_j t_j t_j^t.$
Furthermore, since
\[ T^tT = I, \] this tells us that
\[ t^*_i t_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise} \end{cases} \]
so that the columns of \( T \) are \textbf{orthonormal} (mutually orthogonal with unit length). Furthermore,
\[ HT = TDT^tT = TD \]
which can be read column by column as
\[ Ht_j = \lambda^*_j t_j \quad j=1,2,\ldots,m. \]
Thus, each \( t_j \) is an \textit{eigenvector} of \( H \) associated with the eigenvalue \( \lambda_j \). If the \( \lambda_j \)'s are not all distinct, the \( t_j \)'s are mutually orthogonal nonetheless.

The fact that
\[ H^+ = TD^{-1}T^t \]
can be expressed as
\[ (3.6.2) \quad H^+ = \sum_{j=1}^{m} \lambda^*_j t^*_j t_j. \]

\[ (3.7.1) \textbf{Exercise:} \] Let \( H \) be an \( m \times m \) symmetric matrix and suppose the non-zero eigenvalues of \( H \) are \( \lambda_1, \lambda_2, \ldots, \lambda_k \) (\( k \leq m \)).

\( a) \) Show that there is a representation for \( H \) of the form
\[ H = TDT^t \]
where \( D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k, 0, \ldots, 0) \) and \( T \) is an orthogonal matrix.

(Hint: If \( D \) is not arranged properly, use a permutation matrix, which
is always orthogonal, to do so.)

b) If the columns of \( T \) are denoted by \( t_1, t_2, \ldots, t_m \), show that

\[
\mathcal{R}(\mathbf{H}) = \mathcal{L}(t_1, \ldots, t_k) \quad \text{and} \quad \mathcal{N}(\mathbf{H}) = \mathcal{L}(t_{k+1}, \ldots, t_m).
\]

(3.7.2) Exercise: Without appealing to the diagonalization theorem, show directly that if \( H \) is symmetric and if \( Hx = \lambda_1 x \) and \( Hy = \lambda_2 y \) then \( x \Perp y \) if \( \lambda_1 \neq \lambda_2 \).

The representation (3.6.2) is particularly interesting since it shows clearly the radically discontinuous nature of pseudo-inversion. Two matrices may be very close to each other element by element. But if their ranks differ (e.g., if one is singular while the other is non-singular) their pseudo inverses usually differ greatly. For example, the diagonal matrices

\[
D_1 = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad D_2 = \begin{pmatrix} 4 & 0 \\ 0 & 10^{-10} \end{pmatrix}
\]

are close to each other, but

\[
D_1^+ = \begin{pmatrix} 1/4 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad D_2^+ = \begin{pmatrix} 1/4 & 0 \\ 0 & 10^{10} \end{pmatrix}
\]

differ greatly. In terms of (3.6.2) it is easy to understand why, since the transformation

\[
\lambda^+ = \begin{cases} 
1/\lambda & \text{if } \lambda \neq 0 \\
0 & \text{if } \lambda = 0
\end{cases}
\]

exhibits an infinite discontinuity at \( \lambda = 0 \). This characteristic induces serious computational difficulties which we will discuss at the appropriate time.
In (3.8) we will see that the pseudo inverse of an arbitrary rectangular matrix is expressible in terms of the pseudo inverse of symmetric matrices:

\[ H^+ = (H^tH)^+H^t = H^t(HH^t)^+ \]

so that one can, in theory, diagonalize symmetric matrices (for which well known algorithms are available) and proceed directly to pseudo inversion from there. However, there are other, less tedious methods for computing \( H^+ \) (Chapter V). The problem of roundoff is quite serious as the reader may have already come to appreciate.

(3.7.3) **Exercise:** Let \( H \) be an \( n \times m \) matrix and let \( x_1, x_2, \ldots, x_r \) \( r \leq m \) be an orthonormal set of \( n \) dimensional vectors such that

\[ \mathcal{R}(H) = \mathcal{L}(x_1, x_2, \ldots, x_r) \]

Then

\[ HH^+ = \sum_{j=1}^{r} x_jx_j^t. \]

(Hint: Use (3.5) and (2.7).)

A symmetric matrix, \( P \), is called a *projection matrix* if it is idempotent (i.e., \( P^2 = P \)).

(3.7.4) **Exercise:** The eigenvalues of a projection matrix are either zero or unity.

(3.7.5) **Exercise:** If \( P \) is a projection matrix, \( P^+ = P \) and \( Px \) is the projection of \( x \) on \( \mathcal{R}(P) \). If \( x \in \mathcal{R}(P) \) then \( Px = x \).

(3.7.6) **Exercise:** \( HH^+, H^+H, I - HH^+ \) and \( I - H^+H \) are projection matrices.

(3.7.7) **Exercise:** If \( P_1, P_2, \ldots, P_n \) are projection matrices having the property that \( P_iP_j = 0 \) if \( i \neq j \) and if \( P \) is another projection such that
\[ \mathcal{R}(P) = \mathcal{R}(\sum_{j=1}^{n} P_j) \] then

\[ P = \sum_{j=1}^{n} P_j. \]

(Hint: First show that \( Q = \sum_{j=1}^{n} P_j \) is a projection matrix. Then \( Q^+ = Q \) and so \( Q^+ x = Q x \) which is the projection of \( x \) on \( \mathcal{R}(Q) \). On the other hand \( PP^+ = P \) is the projection of \( x \) on the range of \( P \). Since \( \mathcal{R}(P) = \mathcal{R}(Q) \), we have \( P x = Q x \) for all \( x \).)

(3.7.8) **Exercise:** Let \( h_1, h_2, \ldots, h_n \) be a set of vectors and let \( H \) be the matrix whose column vectors are \( h_1, h_2, \ldots, h_n \). Then for any \( x \), \( HH^+ x \) is the projection of \( x \) on \( \mathcal{L}(h_1, h_2, \ldots, h_n) \).

(3.7.9) **Exercise:** If \( H \) is any matrix, then \( H x = 0 \) if and only if \( x = (I - H^+ H) y \) for some \( y \).

(3.7.10) **Exercise:** If \( H \) is any matrix, then \( z \in \mathcal{R}(H) \) if and only if \( z = H H^+ u \) for some \( u \).

We now turn our attention to an exploration of the most important properties of \( H^+ \).

(3.8) **Theorem:** For any matrix \( H \),

(3.8.1) \[ H^+ = (H^+ H)^+ H^t \]

(3.8.2) \[ (H^t)^+ = (H^+)^t \]

(3.8.3) \[ H^+ = H^t (HH^t)^+ \].

**Proof:**

\[ (H^t H)^+ H^t = \lim_{\delta \to 0} \frac{1}{\delta^2 I + \delta^2 H^t H} H^t \]

and

\[ H^+ = \lim_{\delta \to 0} \left( H^t H + \delta^2 I \right)^{-1} H^t, \]  

(3.4)
By (2.11), any vector, \( z \), can be written as

\[
z = Hx_0 + z^* \quad \text{for some } x_0
\]

where

\[
H^*z = 0.
\]

Thus

\[
(H^H)^+H^*z = \lim_{\epsilon \to 0} [(H^H)^2+\epsilon^2I]\cdot^{-1}(H^H)^2x_0^* = 0,
\]

and

\[
H^*z = \lim_{\epsilon \to 0} [(H^H)^2+\epsilon^2I]\cdot^{-1}H^*x_0.
\]

Using the diagonalization theorem (2.14), we write

\[
H^H = TDT^t
\]

where \( D \) is diagonal and \( T \) is orthogonal. (3.8.4) takes the form

\[
(H^H)^+H^*z = T(\lim_{\epsilon \to 0} [D^2+\epsilon^2I]\cdot^{-1}D^2)T^tx_0
\]

while (3.8.5) takes the form

\[
H^*z = T(\lim_{\epsilon \to 0} [D+\epsilon^2I]\cdot^{-1}D)T^tx_0.
\]

For diagonal \( D \)'s, the matrices in curly brackets are the same, which proves (3.8.1).

To prove (3.8.2), notice that \((HH^t+\epsilon^2I)^{-1}\) is symmetric so by (3.4.1),

\[
(H^t)^+ = \lim_{\epsilon \to 0} (HH^t+\epsilon^2I)^{-1}H
\]

\[
= \lim_{\epsilon \to 0} [H^t(HH^t+\epsilon^2I)^{-1}]^t
\]

\[
= (H^t)^t,
\]

(3.4.2).

To prove (3.8.3), we use (3.8.2) to establish that

\[
([H^t]^t)^t = H^t
\]

(3.8.6)
and (3.8.1) to establish that

\[(3.8.7) \quad (H^t)^+ = (HH^t)^+H.\]

Since \((HH^t)^+\) is symmetric, (3.8.3) follows from (3.8.6) after taking transposes in (3.8.7).

In his paper of 1955, which was probably responsible for the
rebirth of interest in the topic of generalized inverses, Penrose [1],
characterized the pseudo inverse as the (unique) solution to a set of
matrix equations. The pseudo inverse, \(H^+\), that we have been investigat-
ing satisfies the Penrose conditions:

(3.9) **Theorem:** For any matrix \(H\), \(B = H^+\) if and only if

(3.9.1) \(HB\) and \(BH\) are symmetric

(3.9.2) \(HHB = H\)

(3.9.3) \(BBH = B\).

Proof of necessity:

\[HH^+ = \lim_{\delta \to 0} HH^t(HH^t+\delta^2I)^{-1}\]

and

\[H^+H = \lim_{\delta \to 0} (H^tH+\delta^2I)^{-1}H^tH,\quad (3.4).\]

Both are symmetric. This shows that \(H^+\) satisfies (3.9.1). By (3.5),
\(HH^+\) is the projection on \(\mathcal{R}(H)\). Since \(Hx \in \mathcal{R}(H)\) for all \(x\), (3.7.5) assures that \((HH^+)(Hx) = Hx\). This shows that \(H^+\) satisfies (3.9.2).

By (3.8.1),

(3.9.4) \(H^+H = (H^tH)^+(H^tH)\).

By (3.8.1) and (3.9.2)

33
\[ H^+ = (H^+H)^+H^+ = (H^+H)^+(H(H^+H))^+ \]

\[ = (H^+H)^+H^+(HH^+)^+ . \]

Since \( HH^+ \) is symmetric, (3.7.6),

\[ H^+ = (H^+H)^+H^+(HH^+) = (H^+H)^+(H^+H)H^+ \]

and by (3.9.4), the last is equal to \( (H^+H)H^+ \).

This establishes (3.9.3).

Proof of sufficiency: Suppose \( B \) satisfies (3.9.1)-(3.9.3):

Since \( EH = (BH)^t, \) and \( H = H^tB^t \),

\[ H = H^tB^t = HH^tB^t . \]

Since \( HH^+H = H, \)

\[ H^+H = H^+(HH^tB^t) = [H(H^+H)]^tB^t \]

and so

(3.9.5) \[ H^+H = H^tB^t = EH. \]

Since \( B = H^tB^t \) and since \( HB \) is symmetric,

(3.9.6) \[ B^t = H^tB^t . \]

Pre-multiplying (3.9.6) by \( HH^+ \), we find that

\[ HH^+B^t = HH^+H^tB^t = HBB^t , \]

(3.9.2),

and by (3.9.6) the last is equal to \( B^t \). Thus

(3.9.7) \[ B^t = (HH^+)B^t . \]

Taking transposes in (3.9.7) we find that

\[ 34 \]
\[ B = B(HH^+)^t = (BH)H^+ \]

and by (3.9.5) we finally conclude that

\[ B = H^+HH^+ \]

Since \( H^+HH^+ = H^+ \), we see that

\[ B = H^+. \]

The Penrose characterization for pseudo inverses is extremely useful as a method for proving identities. For instance, if one thinks that a certain expression coincides with the pseudo inverse of a certain matrix, \( H \), a handy way of deciding is to run the expression through conditions (3.9.1)-(3.9.3) and observe whether or not they hold.

(3.10) Exercise: If \( A \) and \( B \) are nonsingular, it is well known that \((AB)^{-1} = B^{-1}A^{-1}\). Use (3.9) to show that it is not generally true that \((AB)^+ = B^+A^+\). Where do the conditions break down? Exhibit a counter example. (See (4.10)-(4.16) for a detailed study of this problem.)

(3.11) Exercise: Prove the following:

(3.11.1) \( (H^+)^+ = H \).

(3.11.2) \( (H^+H)^+ = H^+(H^t)^+ \) and \( (HH^t)^+ = (H^t)^+H^+ \).

(3.11.3) If \( A \) is symmetric and \( \alpha > 0 \) then \( (A^\alpha)^+ = (A^+)\alpha \) and \( A^\alpha(A^\alpha)^+ = (A^+)\alpha A^\alpha = AA^+ \).

(3.11.4) \( (H^tH)^+ = H^+(HH^t)^+H = H^t(HH^t)^+(H^t)^+ \).

(3.11.5) \( \mathcal{Q}(H^+) = \mathcal{Q}(H^tH) = \mathcal{Q}(H^t) ; \mathcal{Y}(H) = \mathcal{Y}(H^tH) = \mathcal{Y}[(H^tH)^+] \).

(3.11.6) If \( A \) is symmetric, \( AA^+ = A^+A \).

(3.11.7) \( HH^+ = (HH^t)(HH^t)^+ = (HH^t)^+(HH^t) \) and \( H^+H = (H^+H)(H^+H)^+ \)

\[ = (H^tH)^+(H^tH). \]
(3.11.8) If $A$ is symmetric and $\alpha > 0$, $A^+ \alpha = \hat{A}^+$. 

(3.11.9) If $H$ is an $n \times 1$ matrix (a column vector) $H^+ = H^+H^TH$ and $HH^+ = HH^+/\|H\|^2$.

The properties of pseudo inverses and projections which we have developed can be readily applied to the theory of least squares subject to constraints. But first, we summarize the general results for unconstrained least squares and the related theory of linear equations, in the language of pseudo inverses:

(3.12) **Theorem:**

(a) $x_o$ minimizes

(3.12.1) $\|z - Hx\|^2$

if and only if $x_o$ is of the form

(3.12.2) $x_o = H^+z + (I - H^+H)y$

for some $y$.

(b) The value of $x$ which minimizes (3.12.1) is unique if and only if $H^+H = I$. The last is true if and only if zero is the only null vector of $H$.

(c) The equation

(3.12.3) $Hx = z$

has a solution if and only if

(3.12.4) $HH^+z = z$.

The last is true if and only if $z \in \mathbb{Q}(H)$. $x_o$ is a solution to (3.12.3) if and only if it is of the form (3.12.2). (3.12.3) has a unique solution ($= H^+z$) if and only if $HH^+z = z$, and $H^+H = I$.

**Proof:**

(a) $\|z - Hx\|^2$ is minimized by $x_o$ if and only if $Hx_o = z$ where
\( \hat{z} \) is the projection of \( z \) on \( \mathcal{Q}(H) \), (3.12). By (3.4), \( H^+z \) minimizes \( \|z-Hx\|^2 \) so that \( Hx_o = H(H^+z) \). This means that \( x_o H^+z \) is a null vector of \( H \) if \( x_o \) minimizes (3.12.1). The last is true if and only if
\[
x_o H^+z = (I-H^+H)y \quad \text{for some } y, \quad (3.7.9).
\]

Conversely, if \( x_o \) has the form (3.12.2), then \( Hx_o = H(H^+z) \) since \( H(I-H^+H) = 0 \), (3.9.2). This proves part a).

b) The value of \( x \) which minimizes (3.12.1) is unique if and only if \( (I-H^+H)y \) vanishes for all \( y \). This can only happen if the projection of all vectors on \( \mathcal{N}(H) \) is zero, which means that \( \mathcal{N}(H) \) consists only of the zero vector. \( (I-H^+H)y = 0 \) for all \( y \), by the way, if and only if \( H^+H = I \). This proves b).

c) (3.12.3) has a solution if and only if \( z \) is the after-image of some \( x \) under \( H \). This is, by definition, the same as saying "\( z \in \mathcal{Q}(H) \)". By virtue of (3.7.10), the last holds true if and only if
\[
z = HH^+u
\]
for some \( u \). Since \( HH^+ \) is a projection, (3.7.6), it follows that
\[
HH^+z = (HH^+)^2u = HH^+u = z.
\]

When (3.12.3) has a solution, \( x_o \), this solution must minimize \( \|z-Hx\|^2 \) (the minimal value in this case being zero) and so \( x_o \) must be of the form (3.12.2). The solution is unique if and only if \( H^+H=I \) (part b) and these conclusions collectively serve to establish c).

(3.12.4) **Corollary**: Let \( G \) be a rectangular matrix and suppose \( u \) is a vector in \( \mathcal{Q}(G) \). Then

a) \[
S = \{ x : Gx = u \}
\]
is nonempty and \( x_0 \) minimizes \( \|z-Hx\|^2 \) over \( \mathcal{S} \) if and only if
\[
x_0 = G^+ u + \bar{H}^+ \bar{z} + (I-G^+G)(I-H^+H)y
\]
for some \( y \), where
\[
\bar{z} = z - HG^+u
\]
and
\[
\bar{H} = H(I-G^+G).
\]

b) The vector of minimum norm among those which minimize \( \|z-Hx\|^2 \) over \( \mathcal{S} \) is
\[
G^+ u + \bar{H}^+ \bar{z}.
\]

Proof: a) If \( u \in \mathcal{R}(G) \) then by (3.12), \( \mathcal{S} \) is nonempty and
\[
\mathcal{S} = \{ x : x = G^+u + (I-G^+G)v \text{ for some } v \}.
\]
Therefore
\[
\min_{x \in \mathcal{S}} \|z-Hx\| = \min_{v} \|\bar{z}-\bar{H}v\|,
\]
The latter minimum occurs at \( v_0 \) if and only if
\[
v_0 = \bar{H}^+ \bar{z} + (I-H^+H)y \quad \text{for some } y, \quad (3.12a),
\]
so that \( x_0 \) minimizes \( \|z-Hx\| \) over \( \mathcal{S} \) if and only if
\[
x_0 = G^+u + (I-G^+G)[\bar{H}^+ \bar{z} + (I-H^+H)y]
\]
for some \( y \). Since
\[
\bar{H}^+ = \bar{H}^+(\bar{H}H^t)^+,
\]
and since \( (I-G^+G)^2 = (I-G^+G) = (I-G^+G)^t \), (3.7.6), it follows that
\[
(3.12.4.1) \quad (I-G^+G)\bar{H}^+ = (I-G^+G)^2\bar{H}^t(\bar{H}H^t)^+ = \bar{H}^+
\]
38
and so it is clear that any value of $x$ which minimizes $\|z-Hx\|$ over $S$ is of the form

$$x_0 = G^+u+\bar{H}^Tz+(I-G^+G)(I-\bar{H}^+H)y$$

for some $y$.

b) 

$$[(I-G^+G)(I-\bar{H}^+H)y]^T G^+u = y^T(I-\bar{H}^+H)(I-G^+G)G^+u$$

since $(I-G^+G)$ and $I-\bar{H}^+H$ are symmetric. The last is zero since

$(I-G^+G)G^+ = G^+-G^+GG^+ = 0$. Thus

$$(3.12.4.2) \quad (I-G^+G)(I-\bar{H}^+H)y \perp G^+u.$$ 

On the other hand,

$$[(I-G^+G)(I-\bar{H}^+H)y]^T \bar{H}^+z = y^T(I-\bar{H}^+H)(I-G^+G)\bar{H}^+z$$

$$= y^T(I-\bar{H}^+H)\bar{H}^+z,$$ \hspace{1cm} (3.12.4.1).

Since $\bar{H}^+H-\bar{H}^+H\bar{H}^+ = 0$, we see that $(I-G^+G)(I-\bar{H}^+H)y \perp \bar{H}^+z$ as well, so that if $x_0$ minimizes $\|z-Hx\|^2$ over $S$, then

$$\|x_0\|^2 = \|G^+u+\bar{H}^+z\|^2+\|(I-G^+G)(I-\bar{H}^+H)y\|^2$$

$$\geq \|G^+u+\bar{H}^+z\|^2,$$

with strict inequality holding unless

$$x_0 = G^+u+\bar{H}^+z.$$ ~

(3.12.5) Exercise: a) The equation $Hx=z$ has a solution for all $z$ if the rows of $H$ are linearly independent.

b) If the equation $Hx=z$ has a solution, the solution is unique if and only if the columns of $H$ are linearly independent.

(3.12.6) Exercise: Let $H$ be an $n \times m$ matrix with linearly independent
columns. For any $k \times m$ matrix, $G$, let $\bar{H} = H(I-G^+G)$. Show that 
$$(I-G^+G)(I-\bar{H}^+\bar{H}) = 0.$$ (Hint: If $w = (I-G^+G)(I-\bar{H}^+\bar{H})v$, then $Hw=0$. 
Apply (2.12.1).)

Comment: If the columns of $H$ are linearly independent and $u$ is in the range of $G$, then (3.12.6) and (3.12.4) together, imply that there is a unique vector which minimizes $\|z-Hx\|^2$ over $S$. In general, though, the minimizing vector is not unique. However, the vector, 
$$\hat{x}_o = G^+u + \bar{H}^+z,$$ has minimum norm among those which minimize $\|z-Hx\|^2$ over $S$. The vector $\bar{x} = \bar{H}^+z$ is the vector of minimum norm among those which minimize $\|z-Hx\|^2$ subject to no constraints. $\bar{x}$ and $\hat{x}_o$ differ by $G^+u$, the minimum norm vector in $S$.

(3.12.7) Exercise: Refer to (3.12.4): If $S$ is empty, then $x^* = G^+u + \bar{H}^+z$ is the vector of minimum norm which minimizes $\|z-Hx\|$ over $S^*$ where $S^* = \{x : \|Gx-u\|^2$ is minimized$. (Hint: 
$S^* = \{x : x = G^+u - (I-G^+G)v$ for some $v\}$. The proof of (3.12.4) carries over word for word.)

(3.12.8) Application to Linear Programming.

(Ben Israel and Charnes [2], also Ben Israel, 'Charnes' and Roberts [1].)

Let $a$, $b$ and $c$ be specified $n$-vectors and let $A$ be a given $m \times n$ matrix. Consider the problem of minimizing $c^+x$ with respect to $x$, subject to the constraints 

$$a \leq Ax \leq b$$

(where the vector inequality is to be interpreted componentwise.)

Following the conventional terminology, the problem is said to be feasible if the constraint set is nonempty, and bounded if the minimal value of $c^+x$ on the constraint set is finite.
Assuming that the problem is feasible, it follows readily that the problem is bounded if and only if \( A^+Ac = c \). To see this, suppose first that the problem is bounded. If \( \nabla(A) = \{0\} \) then \( I-A^+A = 0 \), (3.5), and so \( A^+Ac = c \). Otherwise, if \( 0 \not\in \nabla(A) \), and if \( x_0 \) is in the constraint set, so is \( x_0 + \alpha y \) for every scalar \( \alpha \). Hence

\[
\min_{a \leq Ax \leq b} c^t x \leq c^t(x_0 + \alpha y) = c^t x_0 + \alpha c^t y
\]

which can be made arbitrarily small and negative unless \( c \) is orthogonal to every null vector of \( A \). Thus, the problem is bounded if and only if

\[
ce \in \nabla(A) = \mathcal{R}(A^t), \tag{2.10}
\]

Since

\[
A^+Ac \text{ is the projection of } c \text{ on } \mathcal{R}(A^t), \tag{3.5}
\]

we see that boundedness implies \( A^+Ac = c \).

Conversely, if \( A^+Ac = c \), then \( c \in \mathcal{R}(A^t) \) so that \( c = A^t z \) for some \( z \). Thus \( c^t x = z^t Ax \). Each component of \( Ax \) is bounded below on the constraint set so that \( z^t Ax \) is bounded below as \( x \) ranges over the constraint set. We assume henceforth, that \( c \in \mathcal{R}(A^t) \).

We now make the additional assumption that the rows of \( A \) are linearly independent. In this case, the equation \( Ax = z \) has a solution for every \( m \)-dimensional vector \( z \), (3.12.5), and so

\[
\min_{a \leq Ax \leq b} c^t x = \min_{a \leq z \leq b} \min_{z = Ax} c^t x.
\]

The set of \( x \)'s for which \( z = Ax \) is of the form

\[
A^+z + (I-A^+A)y
\]

where \( y \) is free to vary unrestricted over \( n \) space, (3.12). Since
\( cA(A^+)^\perp = \gamma^\perp (A) \), we see that \( c \) is orthogonal to all null vectors of \( A \) (of which \((I-A^+A)y\) is one) so that

\[
c^t x = c^t A^+ z \text{ if } Ax = z.
\]

Consequently,

\[
\min_{a \leq Ax \leq b} c^t x = \min_{a \leq z \leq b} c^t A^+ z
\]

\[
- \min_{a \leq z \leq b} (A^+c)^t z
\]

and if \( \tilde{z} \) minimizes the right side, any \( x \) of the form

\[
\tilde{x} = A^+ \tilde{z} + (I-A^+A)y
\]

will minimize the left side over the constraint set.

The minimization of the right side is trivial: Denote the components of \( a, b, A^+c \) and \( z \) by \( \alpha_i, \beta_i, \gamma_i \) and \( \xi_i \) respectively. Then

\[
\min_{a \leq z \leq b} (A^+c)^t z = \min_{\alpha_i \leq \xi_i \leq \beta_i} \sum_{i=1}^{n} \gamma_i \xi_i
\]

and the components of the solution vector are clearly

\[
\xi_i = \begin{cases} 
\beta_i & \text{if } \gamma_i < 0 \\
\alpha_i & \text{if } \gamma_i > 0 \\
\text{anything between} & \\
\alpha_i \text{ and } \beta_i \text{ (inclusive) if } \gamma_i = 0.
\end{cases}
\]

The present solution depends crucially upon the assumption that the rows of \( A \) are linearly independent. If this assumption is violated, a modified procedure (more complicated) must be resorted to in order to
solve the problem, (Ben Israel and Robers, [1]).

(3.13) Exercises:

(3.13.1) If A and B are matrices with the same number of rows then

$$\mathcal{R}(A) \subseteq \mathcal{R}(B) \text{ if and only if } BB^tA = A.$$ 

(3.13.2) If A and B are matrices with the same number of rows, the matrix equation

$$BX = A$$

has a solution if and only if $$\mathcal{R}(A) \subseteq \mathcal{R}(B)$$. In this case, any matrix of the form

$$X = B^tA + (I - B^tB)Y$$

(where Y has the same number of rows and columns as A) is a solution.

The solution is unique if and only if $$B^tB = I$$ (i.e., $$\mathcal{N}(B) = \{0\}$$).

(3.13.3) If M is nonsingular and A is rectangular with the same number of rows then

$$(MA)^t(MA) = A^tA.$$ 

If N is nonsingular and A is rectangular with the same number of columns,

$$(AN)(AN)^t = AA^t.$$ 

(3.13.4) The matrix equation

$$AXB = C$$

has a solution in X if and only if

$$AA^tCB^tB = C.$$ 

In this case, the general solution is

$$X = A^tCB^t + M - A^tAMBB^t$$
where \( M \) is arbitrary (same number of rows as \( A \), same number of columns as \( B \)).

(3.13.5) For any matrix \( A \),

\[
A^+ = WAX
\]

where \( W \) and \( Y \) are, respectively, any solutions to

\[
WAA^t = A^t
\]

and

\[
A^tAY = A^t
\]

(Decell, [1]).

(3.13.6) A matrix is said to be normal if it commutes with its transpose. In this case, show that \( A^+A = AA^+ \).

(3.13.7) If \( T \) is orthogonal, \( (AT)^t = T^tA^t \).

(3.13.8) If \( \mathcal{O}(B) \subseteq \mathcal{O}(A) \), then among those matrices, \( X \), which satisfy \( AX = B \),

the one which minimizes the trace of \( X^tX \) is \( \hat{X} = A^+B \).

If \( AZ = B \), then

\[
\text{trace}(Z^tZ) > \text{trace}(\hat{X}^t\hat{X}) \quad \text{if} \quad Z \neq \hat{X}.
\]

(3.13.9) The following conditions are equivalent:

a) \( XY^t = 0 \),

b) \( XY^t = 0 \),

c) \( XH^tH = 0 \).

(3.13.10) If \( P \) is a projection, and if \( H \) is any rectangular matrix (having the right number of columns) then

\[
(H^tH)^+ = P(H^tH)^+ = (H^tH)^+P, \quad \overline{H} = H^t,
\]

44
and

$$H^+ = (H^2H)^+H^t$$

where

$$H = HP.$$  

(3.14.1) **Exercise:** Let \( h_1, h_2, \ldots, h_n, \ldots \) be any collection of vectors and define

$$A_0 = I$$

$$A_n = \begin{cases} 
A_{n-1} & \text{if } h_n \text{ is a linear combination of } h_1, h_2, \ldots, h_{n-1} \\
A_{n-1} - \frac{(A_{n-1}h_n)(A_{n-1}h_n)^t}{h_n^tA_{n-1}h_n} & \text{otherwise.}
\end{cases}$$

Show that for each \( n \)

a) \( A_n \) is the projection on \( 
\mathcal{L}^\perp(h_1, \ldots, h_n) \).

b) \( A_nh_{n+1} \) is the (unnormalized) \( (n+1) \)-st vector in the Gramm-Schmidt orthogonalization of \( h_1, h_2, \ldots, h_{n+1} \).

(3.14.2) **Exercise:** Stationary probability for a Markov chain.

A matrix, \( P \), is called a stochastic matrix if its entries are non-negative and its row sums are unity. Such matrices can be used to represent the one step transition probabilities for homogeneous Markov chains. If the process is ergodic (a condition which can be deduced by visual inspection of the transition matrix, Feller [1], Chapter XV) then

$$\lim_{n \to \infty} (p^t)^n = (x; x; \ldots; x)$$
where \( x \) is the unique probability vector (a vector having nonnegative components which sum to unity) satisfying
\[
P^t x = x.
\]

This probability vector is the steady state probability distribution for the Markov chain. The \( i \)-th component of \( x \) represents the (steady state) probability that the process is in state \( i \) at any given instant of time.

Using these facts as a starting point, show that

a) \( y = P^t y \) only if \( y \) is a multiple of \( x \).

b) The row vectors of \( I - P^t \) are linearly dependent.

denoting the columns of \( I - P \) by \( q_1, q_2, \ldots, q_N \), show that

c) \( q_1, q_2, \ldots, q_{N-1} \) are linearly independent.

d) If \( A_0 = I \),
\[
A_n = A_{n-1} - (A_{n-1} q_n)(A_{n-1} q_n)^t / q_n^t A_{n-1} q_n
\]
\( n = 1, 2, \ldots, N-1 \)

and \( u \) is the \( N \) dimensional vector whose components are all ones, then
\[
x = A_{N-1} u / u^t A_{N-1} u.
\]

(Hint: \( A_{N-1} \) is the projection on \( \mathbb{R}^1(q_1, q_2, \ldots, q_{N-1}) \); c.f. Decell and Odell, [1].)

In (3.6.1) and (3.6.2) we showed how a symmetric matrix and its pseudo inverse can be represented in terms of its eigenvalues and eigenvectors. Using the theory which we have developed thus far, we can deduce an analogous result for an arbitrary rectangular matrix, \( A \), in terms of the eigenvalues and eigenvectors of \( A^t A \) and \( AA^t \). (See Good, [1].)

We begin by reminding the reader that any matrix of the form \( A^t A \)
has a unique symmetric square root, which has the explicit representation

\[(A^tA)^{1/2} = TD^{1/2}T^t\]

where \(T\) is the orthogonal matrix which reduces \(A^tA\) to the diagonal matrix \(D\),

\[A^tA = TDT^t\]

and \(D^{1/2}\) is obtained by taking the (positive) square root of \(D\)'s (nonnegative) diagonal entries. (D's entries are clearly nonnegative since

\[D = T^tA^tAT\]

so that for any \(x\)

\[x^tDx = \|ATx\|^2 \geq 0.\]

In particular, if \(D = \text{diag}(d_1, \ldots, d_n)\) and \(x\) has all zero components except for a'one in the \(j\)-th place, \(x^tDx = d_j \geq 0.\)

(3.15) Theorem: Let \(A\) be an \(n \times m\) matrix and let \(L\) be the \(r \times r\) diagonal matrix of \((AA^t)\)'s nonzero eigenvalues arranged in arbitrary order.

Then there is an \(n \times r\) matrix, \(P\), and an \(r \times m\) matrix, \(Q\), such that the following conditions hold:

(3.15.1) \(A = PL^{1/2}Q,\)
(3.15.2) \(AA^t = PLP^t,\)
(3.15.3) \(A^tA = PP^t,\)
(3.15.4) \(P^tP = I,\)
(3.15.5) \(A^tA = Q^tLQ,\)
(3.15.6) \(A^tA = Q^tQ,\)
(3.15.7) \(QQ^t = I.\)

Comment: By (3.15.4) and (3.15.7), the columns of \(P\) are orthonormal as are
the rows of \( Q \). By (3.15.2) and (3.15.4), \( AA^t P = PL \) which shows that the columns of \( P \) are eigenvectors of \( AA^t \), the \( j \)-th column being an eigenvector associated with the \( j \)-th diagonal entry, \( \lambda_j \), of \( L \). By (3.15.5) and (3.15.7), the same goes for the \( j \)-th column of \( Q^t \) (which is the \( j \)-th row of \( Q \)). This result is often referred to as the Singular Decomposition Theorem.

Proof: By (3.11.3) and (3.11.7)

\[
AA^+ = (AA^t)^{1/2}[(AA^t)^{1/2}]^+
\]

so by (3.9.2)

\[
(3.15.8) \quad A = (AA^t)^{1/2}[(AA^t)^{1/2}]^+A.
\]

By virtue of (3.7.1)

\[
(3.15.9) \quad AA^t = TDT^t
\]

where

\[
D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r, 0, 0, \ldots, 0),
\]

\( \lambda_j > 0 \) and \( T \) is an orthogonal matrix. Let us express \( T \) and \( D \) as partitioned matrices:

\[
nT = n[ P : P_0 ]
\]

\[
D = \begin{bmatrix} r & n-r \\ L : 0 \\ \vdots : \vdots \\ n-r & 0 : 0 \end{bmatrix}
\]

where

\[
L = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r).
\]

Then (3.15.9) can be rewritten as

\[
(3.15.10) \quad AA^t = PFP^t \quad \text{which proves (3.15.2)}.
\]

Define
\[(3.15.11)\]
\[Q = P^*[(A^t)^{1/2}]^*A.\]

Since
\[(A^t)^{1/2} = TD^1/2_{T^t},\]
and since
\[D^{1/2} = \begin{bmatrix}
L^{1/2} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0
\end{bmatrix},\]
we can write
\[(3.15.12)\]
\[(A^t)^{1/2} = PL^{1/2}P^*.\]

Therefore, by \((3.15.11)\)
\[PL^{1/2}Q = (PL^{1/2}P^*)[(A^t)^{1/2}]^*A\]
\[= A, \quad (3.15.12) \text{and} (3.15.8).\]

This proves \((3.15.1).\)
\[A^t = (A^t)(A^t)^t, \quad (3.11.7),\]
\[= (TD^t)(TD^+T^t) = TDD^+T^t\]
\[= PP^t \quad \text{which proves} \ (3.15.3).\]

Since \(T\) is an orthogonal matrix, its columns are orthonormal, so
\[P^tP = I \quad \text{which establishes} \ (3.15.4).\]
\[Q^tLQ = A^t[(A^t)^{1/2}]^*PLP^*[(A^t)^{1/2}]^*A, \quad (3.15.11),\]
\[= A^t[(A^t)^{1/2}]^*A, \quad (3.15.12),\]
\[= A^t(A^t)^*A, \quad (3.11.3),(3.11.7),\]
\[= A^tA\]

which proves \((3.15.5).\)
\[ Q^tQ = A^t[(AA^t)^{1/2}]^+P^t[(AA^t)^{1/2}]^+A, \quad (3.15.11) \]

By virtue of (3.15.9)

\[ (AA^t)^+ = TD^+_t \]

\[ = PL^{-1}P^t \]

and so by (3.11.3)

\[ [(AA^t)^{1/2}]^+ = [(AA^t)^+]^{1/2} = PL^{-1/2}P^t. \]

Thus

\[ Q^tQ = A^t(PL^{-1/2}P^t)(PP^t)(PL^{-1/2}P^t)A \]

\[ = A^tPL^{-1}P^tA \]

\[ = A^t(AA^t)^+A \]

\[ = A^+A, \quad (3.8.3). \]

This proves (3.15.6).

Finally, by (3.15.11)

\[ QQ^t = P^t[(AA^t)^{1/2}]^+AA^t[(AA^t)^{1/2}]^+P \]

\[ = P^t(AA^+)P, \quad (3.11.3),(3.11.7), \]

\[ = P^t(PF^t)P, \quad (3.15.3), \]

\[ = I, \quad (3.15.4). \]

(3.15.13) Exercise: The non-zero eigenvalues of \( A^+A \) and \( AA^t \) are identical.

(Hint: Starting with \( A \), we showed in (3.8) that every non-zero eigenvalue of \( AA^t \) is an eigenvalue of \( A^+A \). Use the same method of proof, starting with \( A^t \), to show that every non-zero eigenvalue of \( A^+A \) is an eigenvalue of \( AA^t \).)

(3.15.14) Exercise: \( A^+ = Q^tL^{-1/2}t^t, \quad Q^t = Q, \quad P^t = P^t. \)
(3.16) Exercise: There exists a representation for $A$ of the form

$A = \sum_{j} \lambda_{j}^{1/2} p_{j}^{t} q_{j}$

where the $\lambda_{j}$'s are the non-zero eigenvalues of $A^{t}A$ (or $AA^{t}$) repeated according to their multiplicity,

$AA^{t}p_{j} = \lambda_{j}p_{j}$; $A^{t}Aq_{j} = \lambda_{j}q_{j}$

and

$P_{j}^{t} P_{j} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} = q_{j}^{t} q_{j}$.

The representation is not necessarily unique, but not all such representations (with $p_{j}$ and $q_{j}$ satisfying (3.16.2)(3.16.3)) are valid.

(3.16.4) Exercise: $A^{+} = \sum_{j} \lambda_{j}^{-1/2} q_{j}^{t} p_{j}$ if $A$ satisfies (3.16.1).

(3.17) Exercise: If $A$ is normal and if $\alpha$ is a positive rational number,

$(A^{\alpha})^{+} = (A^{+})^{\alpha}$.

(3.18) Exercise: Any matrix, $A$, can be uniquely represented as a linear combination of partial isometries:

$A = \sum_{j} \lambda_{j}^{1/2} U(\lambda_{j})$

where the $\lambda_{j}$'s are the distinct non-zero eigenvalues of $A^{t}A$ and

$U(\lambda) = \lambda^{-1/2} A(I - (A^{t}A - \lambda I)^{+}(A^{t}A - \lambda I))$

satisfy

$U(\lambda_{j}) U^{t}(\lambda_{k}) = 0$ if $k \neq j$. 

51
\[ U^t(\lambda_j)U(\lambda_k) = 0 \quad \text{if} \quad k \neq j \]
\[ A^+ = \sum_j \lambda_j^{-1/2} U(\lambda_j) . \]

(Hint: \( I - (A^t A - \lambda I)^+ (A^t A - \lambda I) \) is the projection on the null space of \( A^t A - \lambda I \), which is spanned by the eigenvectors (if any) of \( A^t A \) associated with \( \lambda \). Then use (3.16.1).) C.f. Ben-Israel and Charnes [1], Penrose [1] and Golub and Kahan [1] for related work.

(3.19) **Exercise:** Iterative method for finding the dominant eigenvalue and associated eigenvector for \( AA^t \) and \( A^t A \).

Let
\[ y_0 = A x_0 / \| A x_0 \| \]
\[ x_{n+1} = A^t y_n / \| A^t y_n \| \]
\[ y_{n+1} = A x_{n+1} / \| A x_{n+1} \| \quad \text{n=0,1,...} \]

Then
\[ \lim_{n \to \infty} x_n = q \quad \text{and} \quad \lim_{n \to \infty} y_n = p \quad \text{exist} \]

provided that \( x_0 \) is not orthogonal to \( N(A^t A - \lambda_1 I) \) where \( \lambda_1 \) is the largest eigenvalue of \( A^t A \).

Furthermore,
\[ AA^t p = \lambda_1 p \quad \text{and} \quad A^t A q = \lambda_1 q \]
so that
\[ \lambda_1 = \lim_{n \to \infty} \| A^t y_n \|^2 \]
\[ = \lim_{n \to \infty} \| A x_n \|^2 . \]
IV. PSEUDO INVERSES OF PARTITIONED MATRICES, AND SUMS AND PRODUCTS OF MATRICES

If \( c_1, c_2, \ldots, c_m \) are a collection of vectors in an \( n \)-dimensional space, we can write the \( n \times m \) matrix

\[
C_m = (c_1 : c_2 : \ldots : c_m)
\]

whose \( j \)-th column is \( c_j \), in terms of \( C_{m-1} \) and \( c_m \):

\[
(4.1) \quad C_m = (C_{m-1} : c_m), \quad (m=2,3,\ldots).
\]

The pseudo inverse of \( C_1 \) is easy to compute:

\[
(4.2) \quad C_1^+ = c_1^t/c_1^t c_1, \quad (3.11.9),
\]

and so, if we can develop a convenient relationship between \( C_m^+ \) and \( C_{m-1}^+ \), we will have a nice computational procedure for pseudo inverting a rectangular matrix "a column at a time":

(4.3) Theorem: (Greville [2]).

(4.3.1) If

\[
C_{m+1} = (C_m : c_{m+1})
\]

then

\[
(4.3.2) \quad C_{m+1}^+ = \begin{pmatrix}
C_m^+[I-C_m c_{m+1} k_{m+1}^t] \\
\begin{array}{c}
\cdots \\
\cdots \\
k_{m+1}^t
\end{array}
\end{pmatrix}
\]

where

\[
(4.3.3) \quad k_{m+1} = \begin{cases} 
\frac{(I-C_m C_m^+) c_{m+1}}{\| (I-C_m C_m^+) c_{m+1} \|^2} & \text{if } (I-C_m C_m^+) c_{m+1} \neq 0 \\
\frac{C_m^t C_m c_{m+1}}{1 + \| C_m^t C_m c_{m+1} \|^2} & \text{otherwise .}
\end{cases}
\]

53
Comment: \((I-C_mC_m^+)^{c_{m+1}}\) is zero if and only if \(C_mC_m^+c_{m+1} = c_{m+1}\) (i.e., if and only if \(c_{m+1}\) is in the range of \(C_m\), (3.7.5).) The range of \(C_m\) is spanned by \(c_1, \ldots, c_m\) (why?) and so \(k_{m+1}\) is defined by the first part of (4.3.3) if and only if \(c_{m+1}\) is not a linear combination of \(c_1, \ldots, c_m\).

Proof: The proof is a straightforward, though tedious verification that the right side of (4.3.2) satisfies the conditions of (3.9). We leave the details to the reader. (Greville's original proof was constructive. The interested reader will find it instructive, as well.)

(4.3.4) Exercise: \(Q(C_m) = L(c_1, c_2, \ldots, c_m)\).

(4.4) Application to Stepwise Regression:

It is very common for an experimentalist to observe samples of a function of an independent variable (e.g., time) whose functional form is not known a priori and for him to wish to model the behavior of this function (e.g., for the purpose of prediction). Typically, the experimenter has in mind a family of functions, \(\varphi_1(\tau), \varphi_2(\tau), \ldots, \varphi_m(\tau)\) and the data is modeled by choosing an appropriate set of weights and representing (approximating) the observed data as a linear combination of the \(\varphi\)'s.

To be more explicit, if the observations, denoted by \(\xi_1, \xi_2, \ldots, \xi_n\), are made at times \(\tau_1, \tau_2, \ldots, \tau_n\), and if the points \((\tau_j, \xi_j)\) are plotted on cartesian coordinates, the problem boils down to choosing a set of scalars, \(\xi_1, \xi_2, \ldots, \xi_m\) so that the graph of the function

\[ \sum_{j=1}^{m} \xi_j \varphi_j(\tau), \]

54
plotted as a function of \( \tau \), comes as close as possible to the given data points.

The most popular method for choosing the weights is (for good cause) the method of least squares, in which the \( \xi_j \)'s are chosen to minimize the sums of the squares of the (vertical) distances from the data points to the curve. In mathematical terms, the \( \xi_j \)'s are chosen to minimize

\[
(4.4.1) \quad \sum_{k=1}^{n} (\xi_k - \sum_{j=1}^{m} \xi_j \Phi_j(\tau_k))^2.
\]

If a vector notation is used, so that

\[
\begin{align*}
z &= \left( \begin{array}{c} \xi_1 \\ \vdots \\ \xi_n \end{array} \right), \quad x &= \left( \begin{array}{c} \xi_1 \\ \vdots \\ \xi_m \end{array} \right), \quad c_j &= \left( \begin{array}{c} \Phi_j(\tau_1) \\ \vdots \\ \Phi_j(\tau_n) \end{array} \right)
\end{align*}
\]

and

\[
c_m = (c_1, c_2, \ldots, c_m) \quad \text{(an n\times m matrix)}
\]

then the problem of choosing \( \xi_1, \ldots, \xi_m \) to minimize (4.4.1) is the same as choosing \( x \) to minimize

\[
(4.4.2) \quad \|z - c_m x\|^2.
\]

In (3.4) we showed that

\[
(4.4.3) \quad \hat{x}^{(m)} = c_m^+ z
\]

always minimizes (4.4.2) and is in fact the vector of minimum norm which does the job. Suppose \( \hat{x}^{(m)} \) is computed and that the residual sum of squares

\[
\|z - c_m \hat{x}^{(m)}\|^2
\]
(which measures the degree of fit between the data and the model) is unacceptably large.

The standard response to such a situation is to augment the family of functions \( \varphi_1(\cdot), \varphi_2(\cdot), \ldots, \varphi_m(\cdot) \) by adding an \((m+1)\)st function \( \varphi_{m+1}(\cdot) \) and then choosing the \( m+1 \) weights \( \xi_1, \xi_2, \ldots, \xi_{m+1} \) to minimize the new residual sum of squares

\[
\sum_{k=1}^{n} \left( \xi_k - \sum_{j=1}^{m+1} \xi_j \varphi_j(\tau_k) \right)^2
\]

which can be expressed in vector-matrix notation as

\[
\|z - C_{m+1}x\|^2
\]

where \( z \) is as defined before,

\[
C_{m+1} = \begin{pmatrix} C_m & c_{m+1} \end{pmatrix}
\]

\[
c_{m+1} = \begin{pmatrix} \varphi_{m+1}(\tau_1) \\ \vdots \\ \varphi_{m+1}(\tau_n) \end{pmatrix}
\]

and \( x \) is now an \( m+1 \) dimensional vector,

\[
x = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_{m+1} \end{pmatrix}
\]

The minimum norm solution is

\[
(4,4,4) \quad \hat{x}^{(m+1)} = C_{m+1}^+ z
\]

and the results of (4.3) show how \( \hat{x}^{(m+1)} \) is related to \( \hat{x}^{(m)} \):
\[ \hat{x}^{(m+1)} = C_{m+1}^+ z = \left( \frac{C_m^+ - C_m^+ \frac{t}{k_m+1} \hat{x}^{(m+1)} z}{k_m+1 z} \right) \]

\[ = \left( \frac{\hat{x}^{(m)} - (k_{m+1}^t) C_m^+ c_{m+1}}{k_{m+1}^t z} \right). \]

\[(4.5) \text{ Exercise:} \]

\[ \| z - C_{m+1} \hat{x}^{(m+1)} \|^2 \leq \| z - C_m \hat{x}^{(m)} \|^2 \]

with strict inequality holding unless the right side is zero or unless

\[ c_{m+1} \]

is a linear combination of \[ c_1, \ldots, c_m \]. (This means that a better fit between the model and the data always exists when a new "regressor" (i.e., a new \( \varphi_j(\cdot) \), is added to the model, unless it is a linear combination of the previous ones.)

A word of caution is in order with regard to the interpretation of \( (4.5) \). The objective of modeling data is generally predictive in nature. Therefore, the fact that the residual sum of squares can be reduced by the addition of another regressor is not sufficient cause to add regressors ad infinitum. A fine line must be tread between parsimonious modeling (using few regressors) and getting a good fit. In many statistical contexts, under-specifying the number of regressors results in biased estimates. Over specification (too many regressors) results in loss of accuracy. The question "how many regressors is enough?" is partially answered by the theory of the analysis of variance and covariance, under certain assumptions. The question "how should I choose my family of regressors to
begin with?" is much more difficult and not easily answered in quantitative terms.

The results of (4.3) lead directly to an important representation for pseudo inverses of certain types of matrix sums:

(4.6) Theorem: Let $c_1, c_2, \ldots$ be a collection of $n$ dimensional vectors and let

$$ S_m = \sum_{j=1}^{m} c_j c_j^t $$

$$ = c_m c_m^t, \quad (m=1, 2, \ldots), $$

where $c_m$ is as defined in (4.1). Then

$$ S_{m+1}^+ = \begin{cases} 
S_m^+ + \frac{1+c_{m+1}^t S_m c_{m+1}}{1+c_{m+1}^t A_m c_{m+1}} (A_m c_{m+1})(A_m c_{m+1})^t \\
- \frac{(S_m^+ c_{m+1}) (A_m c_{m+1})^t (S_m^+ c_{m+1})^t}{c_{m+1}^t A_m c_{m+1}} & \text{if } c_{m+1} \text{ is not a linear combination of } c_1, \ldots, c_m, \\
S_m^+ - \frac{(S_m^+ c_{m+1}) (S_m^+ c_{m+1})^t}{1+c_{m+1}^t S_m c_{m+1}} & \text{otherwise,}
\end{cases} $$

where

$$ A_m = I - S_m S_m^+. $$

Comment: $A_m$ is the projection on $\mathcal{N}(s_m^t) = \mathcal{R}^\perp(s_m)$, (3.5). Since $S_m = c_m c_m^t$ and

$$ \mathcal{Q}(c_m c_m^t) = \mathcal{Q}(c_m), \quad (2.12), $$

$$ = \mathcal{L}(c_1, c_2, \ldots, c_m), \quad (4.3.4), $$

58
we see that $A_m$ is the projection on $\mathcal{L}^\perp(c_1, \ldots, c_m)$, and so $c_{m+1} \in \mathcal{L}(c_1, \ldots, c_m)$ if and only if $A_m c_{m+1} = 0$. Thus, $c_{m+1}$ is not a linear combination of $c_1, \ldots, c_m$ if and only if $A_m c_{m+1} \neq 0$.

Furthermore, by virtue of (3.14.1), $A_m$ satisfies a recursion of its own (replace $h_m$ by $c_m$ in (3.14.1)).

Proof of (4.6):

\[
S_{m+1} = (C_{m+1} C_{m+1}^t)^+ \quad (4.6.3)
\]

\[
= (C_{m+1}^+)t (C_{m+1}^+) \quad (3.11.2, 3.8.2).
\]

(4.6.1) results when (4.3.2) and (4.3.3) are combined in (4.6.3) and note is taken of the fact that

\[
I - C_m C_m^+ = I - S_m S_m^+ = A_m.
\]

(4.6.4) **Exercise:** (Extension of (3.14.1)). Let

\[
C_m = (c_1 : c_2 : \ldots : c_m)
\]

and

\[
D_m = D_o - D_o C_m (C_m t C_m)^+ C_m t D_o
\]

where $D_o$ is an arbitrary matrix of the form $D_o = R^t R$.

Then,

a) $D_m = R^t Q_m R$

where $Q_m$ is the projection on $\mathcal{L}^\perp(Rc_1, \ldots, Rc_m) = \mathcal{L}^\perp(RC_m)$.

b) $D_{m+1} = \begin{cases} 
(D_m c_{m+1})(D_m c_{m+1})^t & \text{if } Rc_{m+1} \text{ is not a l.c. of } \hspace{1cm} \\
(D_m c_{m+1})^t C_m C_m c_{m+1} & \text{otherwise } \hspace{1cm} \\
D_m & \text{otherwise } \hspace{1cm} 
\end{cases}$

59
c) If $R^t_R$ is nonsingular,

$R_{m+1}$ is a linear combination of $R_1, \ldots, R_m$ if and only if $D_m c_{m+1} = 0$,
so that $D_m c_{m+1}$ is defined by the first half of the recursion if $D_m c_{m+1} \neq 0$, otherwise by the second half.

**Comment:** If $A_m = I - C_m^t C_m$ then $A_0 = I$ and

$$A_{m+1} = \begin{cases} 
A_m - \frac{(A_m c_{m+1})(A_m c_{m+1})^t}{c_{m+1}^t A_m c_{m+1}} & \text{if } A_m c_{m+1} \neq 0 \\
A_m & \text{otherwise},
\end{cases} \quad (3.14.1)$$

If $R^t_R$ is nonsingular, $D_m$ and $A_m$ satisfy the same recursion.

$A_m$ and $D_m$ differ because of the different initial conditions.

(4.6.5) **Exercise:** If $A$ is symmetric, nonsingular and $h^t Ah \neq -1$, then

$$(A + hh^t)^{-1} = A^{-1} - \frac{(A^{-1} h)(A^{-1} h)^t}{1 + h^t A^{-1} h}.$$  

The results of (4.3) and (4.6) can be extended to the case of higher order partitioning. The proofs are computational in nature but are basically the same as (4.3) and (4.6). We state the results without proofs and refer the interested reader to the papers by Cline [2], [3], for details.

(4.7) **Theorem:**

(4.7.1) $$\left(\frac{U}{V}\right)^t = \left(\frac{U^t - U^t V J}{J - V^t U}ight)$$

where

(4.7.2) $$J = C^t (I - C^t C) KY^t U^t U^+ (I - V C^t),$$

(4.7.3) $$G = (I - U U^t)^V$$

and

(4.7.4) $$K = \left[I + [U^t V (I - C^t C)]^t [U^t V (I - C^t C)]\right]^{-1}.$$  

60
Comment: Any matrix of the form $I + D^t D$ is nonsingular, (2.13), so $K$ always exists. The dimension of $K$ is the same as the dimension of $C^t C$ which is the same as the dimension of $V^t V$. If $U^+$ is known, $(U; V)^+$ can be computed at the expense of inverting a square matrix the size of $V^t V$, and finding $I - C^t C$, the projection on $\cap \setminus (C)$.

The extension of (4.6) to higher order partitions is

(4.8) Theorem:

\[(4.8.1) \quad (UU^t + VV^t)^+ = (CC^t)^+ + \]
\[\left[ I - (VC^t)^t \right] [(UU^t)^+ - (UU^t)^+ V(I - C^t C)KV^t(UU^t)^+][I - VC^t] \]

where $C$ and $K$ are as defined in (4.7).

Comment: We can also write

(4.8.2) \[ C = [I - (UU^t)(UU^t)^+]_V \]

and

(4.8.3) \[ K = [I + [(I - C^t C)V^t(UU^t)^+ V(I - C^t C)]]^{-1}. \]

$U$ and $V$ can be interchanged throughout the right side of (4.8.1) without altering the validity of the statement, owing to the symmetry of the left side.

(4.8.4) Exercise: \[ (U; 0)^+ = \begin{pmatrix} U^t \\ 0 \end{pmatrix}. \]

(4.8.5) Exercise: In the special case where $U = C_m$ and $V = C_{m+1}$, show that (4.7) and (4.8) yield formulas which coincide with (4.3) and (4.6).

Theorem (4.8) is extremely important if for no other reason than that it allows the development of an explicit perturbation theory for the pseudo inverse of matrices of the form $A^t A$:  

61
(4.9) Theorem:

(4.9.1) \( (H^tH + \frac{1}{\lambda^2} G^tG)^+ = \)

\( (H^tH)^+ + \lambda^2(I-H^tH)(G^tG)^+(I-H^tH)^t \)

\(-\lambda^4(I-H^tH)[H(G^tG)^+]^t QM(\lambda)Q[H(G^tG)^+](I-H^tH)^t, \)

where

(4.9.2) \( \Pi = H(I-G^tG) = H[I-(G^tG)^+(G^tG)] \)

(4.9.3) \( Q = I-HH^+ \)

and

(4.9.4) \( M(\lambda) = [I+\lambda^2QH(G^tG)^+H^tQ]^{-1}. \)

Proof: Let \( U = G^t/\lambda, \ V = H^t \) and apply (4.8):

\( (H^tH + \frac{1}{\lambda^2} G^tG)^+ = (CC^t)^+ + \)

\( \lambda^2[I-(H^tC^t)^t]^t [(G^tG)^+ - \lambda^2(G^tG)^+H^t(I-C^tC)KH(G^tG)^+](I-H^tC^t) \)

where

\( C = (I-G^tG)H^t = H^t, \ I-C^tC = Q \)

and

\( K = [I+\lambda^2QH(G^tG)^+H^tQ]^{-1} = M(\lambda) . \)

But,

\( (I-H^tC^t)^t = (I-H^tH^+)^t = I-H^tH \)

and

\( G^tH^t = (HG^t)^t. \)

Finally, if \( I+A \) is nonsingular, then

\( (I+A)^{-1} = I-(I+A)^{-1}A \)

so that


\[(I-C^+C)K = QM(\lambda) = QM(\lambda)Q\]

and (4.9.1) follows directly. 

A scalar function \(\varphi(\cdot)\), of a real variable \(\lambda\), is said to be \(O(\lambda^N)\) as \(\lambda \to 0\) if \(\varphi(\lambda)/\lambda^N\) is bounded as \(\lambda \to 0\). A matrix valued function is \(O(\lambda^N)\) if each entry of the matrix is \(O(\lambda^N)\).

\((4.9.5)\) Corollary: a) \(M(\lambda) = I + O(\lambda^2)\) as \(\lambda \to 0\).

b) \[
(H^tH + \frac{1}{\lambda^2} G^tG)^+ = (H^tH)^+ + \lambda^2 (I - H^tH)(G^tG)^+(I - H^tH)^t + O(\lambda^4) \quad \text{as} \quad \lambda \to 0.
\]

Proof: a) \(M(\lambda)\) is obviously \(O(1)\), since \(\lim_{\lambda \to 0} M(\lambda) = I\). Since \(\lim_{\lambda \to 0} (I+A)^{-1} = I - A(I+A)^{-1}\), we see that \(M(\lambda) - I = -\lambda^2CH(G^tG)^+H^tQM(\lambda)\) so that \(\frac{M(\lambda) - I}{\lambda^2}\) is a constant matrix \(\times M(\lambda)\), which is bounded. Thus, \(M(\lambda) - I = O(\lambda^2)\), as asserted.

b) Follows directly.

\((4.9.6)\) Exercise: Show that \(\lim_{\lambda \to \infty} M(\lambda)\) exists, and calculate the limit.

(Hint: Let \(\epsilon = 1/\lambda\) and apply (4.9.5b) to \((I + \frac{1}{\epsilon^2} A^tA)^{-1}\), where \(A = G^tH^tQ\).)

\((4.9.7)\) Exercise: Under what conditions will \(\lim_{\epsilon \to 0} (H^tH + \epsilon^2 G^tG)^+\) exist?

Comment: (4.9.7) again illustrates the extreme discontinuous nature of pseudo inverses. (c.f. the discussion preceding (3.7.3). Also Stewart [1] and Ben Israel [1]).

\((4.9.8)\) Exercise: a) If \(U^tU\) is nonsingular and \(A(\lambda) = O(1)\) as \(\lambda \to 0\) then \(U^tU + \lambda A(\lambda)\) is nonsingular for all suitably small \(\lambda\), and \(b)\) \([U^tU + \lambda A(\lambda)]^{-1} = (U^tU)^{-1} + O(\lambda)\) as \(\lambda \to 0\).

\((4.9.9)\) Exercise: Let \(C\) be an arbitrary symmetric matrix. Then
(4.10) The concept of rank

If \( \mathcal{L} \) is a linear manifold in a Euclidean n-space, the dimension of \( \mathcal{L} \) (abbreviated \( \text{dim} (\mathcal{L}) \)) is defined to be the maximum number of vectors in \( \mathcal{L} \) which can be chosen linearly independent of one another. A fundamental fact that we shall take for granted is that any basis for \( \mathcal{L} \) has exactly \( r \) linearly independent vectors, where \( r = \text{dim} (\mathcal{L}) \).

(4.10.1) Exercise: If \( \mathcal{L}_1 \subseteq \mathcal{L}_2 \) then \( \text{dim} (\mathcal{L}_2 - \mathcal{L}_1) = \text{dim} (\mathcal{L}_2) - \text{dim} (\mathcal{L}_1) \).

If \( A \) is any matrix, the rank of \( A \) (abbreviated \( \text{rk}(A) \)) is defined to be the dimension of \( A \)'s range:

\[ \text{rk}(A) = \text{dim} (\mathcal{R}(A)). \]

Several properties of rank play an important role in the theory of pseudo inverses for products and in the statistical applications of Chapter VI. They are straightforward consequences of already established results and are left as exercises:
(4.10.2) Exercise:

a) For any matrix $A$
\[ \text{rk}(A) = \text{rk}(A^tA) = \text{rk}(A^t) = \text{rk}(AA^t). \]

b) For any matrices $A$ and $B$ (of the right size)
\[ \text{rk}(AB) = \text{rk}(A^tAB) = \text{rk}(AB^t). \]

c) $\text{rk}(AB) \leq \min \{\text{rk}(A), \text{rk}(B)\}$.

d) If $\text{rk}(A) = \text{rk}(AB)$ then $\mathcal{R}(A) = \mathcal{R}(B)$ and the equation $ABX = A$
has a solution in $X$.

(4.10.3) Exercise: a) If $P$ is a projection, $\text{rk}(P) = \text{trace}(P)$.

b) If $P_0, P_1$ and $P_2$ are projections with $\mathcal{R}(P_1) \subseteq \mathcal{R}(P_0)$,
$\mathcal{R}(P_2) \subseteq \mathcal{R}(P_0)$ and $P_1 P_2 = 0$ then
\[ P_0 = P_1 + P_2 \]

if and only if $\text{rk}(P_0) = \text{rk}(P_1) + \text{rk}(P_2)$ (Cochran's Theorem, Scheffé [1]).

We now turn our attention to the question of pseudo inverses for
products of matrixes. If $A$ and $B$ are nonsingular, $(AB)^{-1} = B^{-1}A^{-1}$,
but it is not generally true that $(AB)^+ = B^+A^+$ as evidenced by the example

\[ A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad (AB)^+ = 1, \quad B^+A^+ = \frac{1}{2} \].

Greville [3], has found necessary and sufficient conditions for
$(AB)^+ = B^+A^+$. These we state as the following.

(4.11) Theorem: $(AB)^+ = B^+A^+$ if and only if

(4.11.1) $\mathcal{R}(BB^+A^+) \subseteq \mathcal{R}(A^t)$

and

(4.11.2) $\mathcal{R}(A^tAB) \subseteq \mathcal{R}(B)$.
Proof: Since \( A^+A = A^+A^t, \) (4.11.1) holds if and only if the equation

\[
(4.11.3) \quad A^+ABB^tA^t = BB^tA^t, \quad (3.13.1).
\]

By the same token, (4.11.2) holds if and only if

\[
(4.11.4) \quad BB^tA^tAB = A^tAB.
\]

We will show that (4.11.3) and (4.11.4) are necessary and sufficient for \((AB)^+ = B^+A^+;\)

Suppose (4.11.3) and (4.11.4) hold. Multiply (4.11.3) on the left by \( B^+ \) and on the right by \((AB)^t^+;\)

\[
B^+[A^+A(EE^tA^t)](E^tA^t)^+ =
\]

\[
B^+[A^+(AB)][(AB)^+(AB)]^t =
\]

\[
B^+[AB]
\]

while

\[
B^+[B^tA^t](AB)^t^+ =
\]

\[
B^+[B^tA^t](AB)^t^+ =
\]

\[
(AB)^t(AB)^t^+
\]

\[
= (AB)^+(AB)
\]

so that if (4.11.3) holds

\[
(4.11.5) \quad B^+A^+(AB) = (AB)^+(AB).
\]

By the same token, if both sides of (4.11.4) are premultiplied by \((AB)^t^+\) and postmultiplied by \( A^+\), we find that

\[
(4.11.6) \quad (AB)B^+A^+ = (AB)(AB)^+.
\]

The right hand sides of (4.11.5) and (4.11.6) are symmetric and so
B^+A^+ satisfies (3.9.1). We are done if we can show that B^+A^+ satisfies (3.9.2) and (3.9.3) as well:

If (4.11.5) is premultiplied on both sides by AB, we see that B^+A^+ does indeed satisfy (3.9.2). (3.9.3) is a bit more subtle:

Since

\[ B^+A^+ = (B^+BB^+)(A^+AA^+) = (B^+B^{+t})(B^+A^t)(A^+tA^+), \]

(4.10.2) tells us that

\[ rk(B^+A^+) \leq rk(B^+A^+) = rk(AB). \]

(4.11.7)

On the other hand, (4.10.2b) asserts that

\[ rk(AB) = rk[(AB)^+(AB)] = rk[(B^+A^+) (AB)], \]

(4.11.5),

\[ \leq rk(B^+A^+), \]

(4.10.2c).

If (4.11.7) and (4.11.8) are combined, we find that

\[ rk(B^+A^+) = rk(B^+A^+AB) \]

(4.11.9)

so that the equation

\[ B^+A^+ABX = B^+A^+ \]

(4.11.10)

has a solution in X,

(4.10.2d).

Premultiply (4.11.10) by AB and apply (4.11.6) to deduce

\[ (AB)(B^+A^+)(AB)X = \]

(4.11.11)

\[ (AB)(AB)^+(AB)X = \]

\[ (AB)X = (AB)(B^+A^+). \]
Substituting the last expression for $ABX$ in (4.11.10), we see that
\[(B^+A^+)(AB)(B^+A^+) = B^+A^+\]
which establishes (3.9.3) and shows that $B^+A^+ = (AB)^+$ if (4.11.3) and (4.11.4) hold.

To prove the converse, suppose that $(AB)^+ = B^+A^+$:

Then
\[(AB)^t = [(AB)(AB)^+(AB)]^t\]
\[= (AB)^+(AB)(AB)^t\]
\[= B^+A^+(AB)(AB)^t.\]

If the left and right sides are premultiplied by $ABB^tB$ and use is made of the identity
\[B^t = (BB^tB)^t\]
\[= B^tBB^t\]
we find that
\[ABB^tB(AB)^t = ABB^tBB^tA^+(AB)(AB)^t\]
or equivalently
\[(4.11.12) \quad ABB^t(I-A^tA)BB^tA^t = 0.\]

Since $H^tH = 0$ implies $H = 0$, (4.11.2) tells us that
\[\quad (I-A^tA)BB^tA^t = 0 \quad (\text{take } H=(I-A^tA)BB^tA^t),\]
which is the same as (4.11.3). (4.11.4) is proved the same way, interchanging $A^t$ and $B$ throughout in the proof above.

A general representation for $(AB)^+$ was derived by Cline, [1].

(4.12) **Theorem:** $(AB)^+ = B^+A^+$
where

\[(4.12.1) \quad B_1 = A^+ A B \]

and

\[(4.12.2) \quad A_1 = A B_1 B_1^+. \]

**Proof:** Clearly

\[(4.12.3) \quad A B = A_1 B_1. \]

Furthermore,

\[(4.12.4) \quad B_1 B_1^+ = A^+ A_1 \]

since

\[A^+ A_1 = A^+ (A B_1 B_1^+) = (A^+ A)(A^+ A)BB_1^+ = A^+ ABB_1^+. \]

Similarly

\[(4.12.5) \quad A_1^+ A_1 = A^+ A_1 \]

because

\[(A_1^+ A_1)(B_1 B_1^+) = A_1^+ (A B_1 B_1^+)B_1 B_1^+ = A_1^+ (A B_1 B_1^+) = A_1^+ A_1 \]

from which follows \((4.12.5)\) after transposes are taken and note is made of \((4.12.4)\).

It is now easy to show that \(A_1\) and \(B_1\) satisfy \((4.11.3)\) and \((4.11.4)\):

\[A_1^+ A_1 B_1 (A_1 B_1^t)^t = (A^+ A_1)B_1 (A_1 B_1^t)^t, \quad (4.12.5), \]

\[= A^+ (A B_1) (A_1 B_1^t)^t, \quad (4.12.3), \]

\[= B_1 (A_1 B_1^t)^t, \quad (4.12.1), \]

so \((4.11.3)\) holds.

Similarly,

\[B_1 B_1^t A_1^+ (A_1 B_1^t) = (A_1^+ A_1)(A_1 B_1^t), \quad (4.12.4), \]

69
\[(A_1^+A_1)(A_1^+)(A_1^+B_1), \quad (4.12.5),\]
\[= A_1^+A_1B_1\]
which proves (4.11.4), so that
\[(A_1^+B_1)^+ = B_1^+A_1^+\]
The desired conclusion follows from (4.12.3). \(\Box\)

(4.13) **Exercise:** If A is \(n\times r\) and B is \(r\times m\) then
\[(AB)^+ = B^+A^+\]
if \(rk(A) = rk(B) = r\).

(4.14) **Exercise:** \((AB)^+ = B^+A^+\) if,

a) \(A^+A = I\)

or

b) \(BB^+ = I\)

or

c) \(B = A^+\)

or

d) \(B = A^+.\)

(4.15) **Exercise:** If H is rectangular and S is symmetric and nonsingular,
\[(SH)^+ = H^+S^{-1}[I-(QS^{-1})^+(QS^{-1})]\]
where
\[Q = (I-HH^+).\]

70
V. COMPUTATIONAL METHODS

In recent years, a sizeable literature relating to the computation of pseudo inverses has accompanied the rebirth of interest in the theory. In this chapter we will describe four distinct approaches to the problem. The first method is based upon the Gramm-Schmidt Orthogonalization (hereafter abbreviated by GSO), the second is a modification of the "old faithful", Gauss-Jordan Elimination, the third is based upon the ideas of gradient projection and the last is an exotic procedure derived from the Cayley-Hamilton Theorem.

(5.1) **Method I (GSO, Rust, Burrus and Schneeburger [1]).**

Let $A$ be an $n \times m$ matrix of rank $k \leq \min(n,m)$. It is always possible to rearrange the columns of $A$ so that the first $k$ columns are linearly independent while the remaining columns are linear combinations of the first $k$.

This is the same as saying that for some permutation matrix, $P$ (a square matrix of zeros and ones with exactly one nonzero entry in each row and column)

\[(5.1.1) \quad AP = (R \mid S)\]

where $R$ is $n \times k$ and has rank $k$ and the columns of $S$ are linear combinations of the columns of $R$:

\[(5.1.2) \quad S = RU \quad \text{for some } U, \quad (3.13.2) .\]

$P$ is an orthogonal matrix so that

\[A = (R \mid RU)P^t\]

and

\[A^t = P[R(I \mid U)]^t, \quad (4.14) .\]

The rank of $(I \mid U)$ is the same as the rank of $(I;U)(I;U)^t = I + UU^t, \quad (4.10.2a)$,
which is \( k \). Therefore the rows of \((I^1 U)\) are linearly independent so that

\[
[R(I^1 U)]^+ = (I^1 U)^+R^+,
\]

(4.13),

\[
= (I^1 U)^t (I+UU)^{-1} R^+,
\]

(3.5.2),

hence

\[
A^+ = P(I^1 U)^t (I+UU)^{-1} R^+.
\]

(5.1.3)

The last equation is the starting point for the computational

procedure based on GSO; GSO is used to evaluate \( P, R^+, U \) and \((I+UU)^{-1}\):

a) **Evaluation of \( P \):** Perform a GSO on the columns of \( A \), but do not

normalize: i.e., denote the columns of \( A \) by \( a_1, a_2, \ldots, a_m \), let

\[
c_1^* = a_1
\]

\[
c_j^* = a_j - \sum_{i \in S_j} \frac{a_j c_i^*}{\| c_i^* \|^2} c_i^*
\]

where

\[
S_j = \{i: 1 \leq j-1 \text{ and } c_i^* \neq 0\}.
\]

The vectors \( c_j^* \) are mutually orthogonal and

\[
\langle c_1^*, c_2^*, \ldots, c_i^* \rangle = \langle a_1, a_2, \ldots, a_i \rangle \text{ for each } i, (2.8.1).
\]

If the vectors \( c_1^*, c_2^*, \ldots, c_m^* \) are permuted so that the nonzero

vectors (of which there will be \( k \)) come first, the same permutation

matrix applied to the vectors \( a_1, a_2, \ldots, a_m \) will rearrange them so that

the first \( k \) are linearly independent, while the last \( m-k \) are linear

combinations of the first \( k \), since \( c_j^* = 0 \) if and only if \( a_j \) is a

linear combination of the preceding \( a \)'s. So, if \( P \) is any matrix for

which
\[(5.1.4) \quad \begin{pmatrix} c_1^* & c_2^* & \cdots & c_m^* \end{pmatrix} P = \begin{pmatrix} c_1 & c_2 & \cdots & c_m \end{pmatrix}
\]

where
\[
\|c_j\| \begin{cases} 
> 0 & j=1,2,\ldots,k \\
= 0 & j=k+1,\ldots,m
\end{cases}
\]

then
\[(5.1.5) \quad AP = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \end{pmatrix} P = \begin{pmatrix} R & S \end{pmatrix}
\]

where \(R\) is \(n \times k\) of rank \(k\) and the columns of \(S\) are linear combinations of the columns of \(R\).

b) Computation of \(R^+\):

The (nonzero) vectors, \(c_1, c_2, \ldots, c_k\) defined above, represent a GSO of the columns of \(R\). If we let
\[(5.1.6) \quad Q = \begin{pmatrix} c_1 & c_2 & \cdots & c_k \\
\|c_1\| & \|c_2\| & \cdots & \|c_k\| \end{pmatrix}
\]

then
\[
Q(Q) = Q(R)
\]

so by (3.13.2) there is a \(k \times k\) matrix, \(B\), such that
\[(5.1.7) \quad RB = Q.
\]

Indeed since \(R\) has rank \(k\), \(B = (R^+R)^{-1}Q\). We will derive an algorithm for \(B\) in (5.1.16). Since the columns of \(Q\) are orthonormal, \(Q^tQ = I\) so that \(B\) is nonsingular \((B^{-1} = Q^t(R^+R))\) and hence
\[(5.1.8) \quad R = QB^{-1}.
\]

(4.14) applies again and we find that
\[(5.1.9) \quad R^+ = BQ^+ = B(Q^tQ)^+Q^t = EQ^t.
\]
It remains to evaluate \( B, U \) and \((I+UU^t)^{-1}\).

c) Computation of \( B \) and \( U \)

Denote the columns of \( R \) by \( r_1, r_2, \ldots, r_k \) and the columns of \( S \) by \( s_1, s_2, \ldots, s_{m-k} \). The vectors \( (c_1, c_2, \ldots, c_k, c_{k+1}, \ldots, c_m) \) defined in (5.1.4) represent a nonnormalized GSO of \((r_1, r_2, \ldots, r_k, s_1, \ldots, s_{m-k})\).

Indeed

\[
\begin{align*}
  c_j &= r_j - \sum_{i=1}^{j-1} \frac{r_j^t c_i}{\|c_i\|^2} c_i & j=2, \ldots, k \\
  c_1 &= r_1 \tag{5.1.10}
\end{align*}
\]

and

\[
0 = c_{k+j} = s_j - \sum_{i=1}^{k} \frac{s_j^t c_i}{\|c_i\|^2} c_i & j=1, \ldots, m-k. \tag{5.1.11}
\]

From (5.1.10) it is easy to deduce (by induction on \( j \)) that

\[
\begin{align*}
  c_j &= \sum_{i=1}^{j} \gamma_{ij} r_i & j=1, 2, \ldots, k \\
\end{align*} \tag{5.1.12}
\]

where

\[
\gamma_{ij} = \begin{cases} 
  0 & i > j \\
  1 & i = j \\
  \frac{1}{\sum_{\alpha=1}^{j-1} \frac{(r_j^t c_\alpha)}{\|c_\alpha\|^2} \gamma_{i\alpha}} & i < j
\end{cases} \tag{5.1.13}
\]

On the other hand, (5.1.11) shows that

\[
\begin{align*}
  s_j &= \sum_{i=1}^{k} \omega_{ij} r_i \\
\end{align*} \tag{5.1.14}
\]

where \( \omega_{ij} \) is obtained by substituting (5.1.12) into (5.1.11):
\[ s_j = \sum_{\alpha=1}^{k} \frac{s_j^t c_\alpha}{\|c_\alpha\|^2} (\sum_{i=1}^{\alpha} \gamma_{i\alpha}^r r_i) \]

\[ = \sum_{i=1}^{k} \left( \sum_{\alpha=i}^{k} \frac{s_j^t c_\alpha}{\|c_\alpha\|^2} \gamma_{i\alpha}^r \right) r_i : \]

\[(5.1.15) \quad \omega_{ij} = \sum_{\alpha=i}^{k} \frac{(s_j^t c_\alpha)}{\|c_\alpha\|^2} \gamma_{i\alpha}^r ; \quad \begin{cases} i=1,2,\ldots,k ; \\
j=1,\ldots,m-k . \end{cases} \]

From (5.1.12) and (5.1.6), we see that

\[ Q = \left( \begin{array}{c|c} c_1^t & \cdots & c_k^t \\ \|c_1\| & \cdots & \|c_k\| \end{array} \right) = RB \]

where \( B \) is the \( k \times k \) matrix whose \( (i-j) \)th entry is

\[(5.1.16) \quad \beta_{ij} = \gamma_{ij}/\|c_j\| , \]

while from (5.1.14) we conclude that

\[(5.1.17) \quad S = RU \]

where \( U \) is the \( k \times m-k \) matrix whose \( (i,j) \)th entry is \( \omega_{ij} \).

Notice that (5.1.13) defines \( \gamma_{ij} \) in terms of \( \gamma_{i-1,j-2}, \ldots, \gamma_{i1} \).

Computations are conveniently carried out in the following order:

\[ \gamma_{11} ; \gamma_{22}, \gamma_{12} ; \gamma_{33}, \gamma_{23}, \gamma_{13} ; \gamma_{44}, \gamma_{34}, \gamma_{24}, \gamma_{14} ; \text{ etc.} \]

d) Evaluation of \( (I+UU^t)^{-1} \):

This inversion is achieved via one more GSO:

\[(5.1.18) \text{Theorem: If } U \text{ is a } k \times r \text{ matrix and a GSO is performed on the columns of }\]

\[ k \left( \begin{array}{c} U \\ r \left( \frac{1}{r} \right) \end{array} \right) , \]

75
the resulting matrix of orthonormal vectors is

\[ V = \begin{pmatrix} r \\ k \end{pmatrix} \begin{pmatrix} \frac{V_1}{r} \\ \frac{V_2}{r} \end{pmatrix} \]

where

\[ V_2V_2^t = (I + U^tU)^{-1} \]

and

\[ I - V_1V_1^t = (I + UU^t)^{-1}. \]

**Comment:** The rank of \( \begin{pmatrix} U \\ I \end{pmatrix} \) is \( r \) since \( \begin{pmatrix} U \\ I \end{pmatrix} \) has the same rank as

\( \begin{pmatrix} U^t \\ I \end{pmatrix} = I + U^tU \) (\( r \times r \)) which is nonsingular, (2.13), and hence has rank \( r \). Therefore, the GSO, when performed on the columns of \( \begin{pmatrix} U \\ I \end{pmatrix} \) will generate only nonzero vectors.

**Proof of Theorem:**

Let \( H = \begin{pmatrix} U \\ I \end{pmatrix} \). \( H \) has rank \( r \) (so that \( H^*H = I \), (3.5.3),) and \( \mathcal{R}(V) = \mathcal{R}(H) \), so the equation

\[ HZ = V \]

has the unique solution

\[ Z = H^*V, \quad \text{(3.5.2)} \]

Since the columns of \( V \) are orthonormal, \( V^tV = I \), hence \( V^* = (V^tV)^{1/2}V = V \).

Since \( \mathcal{R}(V) = \mathcal{R}(H) \),

\[ HH^* = VV^* = VV^t. \]

Therefore

\[ ZZ^t = H^*(VV^t)H^* = (H^*H)^+ \]

\[ = (H^*H)^{-1} \quad \text{since } H \text{ has rank } r, (4.10.2a, 3.5). \]

Since

\[ HZ = \begin{pmatrix} U \\ I \end{pmatrix}Z = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \]

we have

\[ UZ = V_1 \]

76
and

$$Z = V_2.$$  

Therefore

$$V_2V_2^t = ZZ^t = (H^tH)^{-1} = (I+U^tU)^{-1}.$$  

The second part of the theorem follows from the identity

$$(I+UU^t)^{-1} = I-U(U^tU+I)^{-1}U^t$$

and the fact that

$$V_1 = UZ = UV_2.$$  

Therefore

$$(I+UU^t)^{-1} = I-U(V_2V_2^t)U^t = I-V_1V_1^t.$$  

Summary of Method I

To find the pseudo inverse of the $n \times m$ matrix, $A$:

1. Perform a GSO on the columns of $A$. Do not normalize. Call this set of vectors $(c_1^*, \ldots, c_m^*)$.

2. Permute the $c_j$'s so that $(c_1 | c_2 | \ldots | c_m) = (c_1^* | c_2^* | \ldots | c_m^*)^P$ where $P$ is a permutation matrix chosen so that

$$c_j \neq 0 \quad j=1,2,\ldots,k$$

$$c_j = 0 \quad j=k+1,\ldots,m.$$  

3. Compute $\gamma_{ij}$ for $j=1+1,\ldots,k; i=1,\ldots,k$, according to (5.1.13) $(\gamma_{ii}=1, \gamma_{ij}=0$ if $i<j$).

4. $B$ is the $k \times k$ matrix whose $(i,j)$th entry is $\gamma_{ij}/\|c_j\|$, $U$ is the $k \times n-k$ matrix whose $(i,j)$th entry is $u_{ij}$ (given by (5.1.15)), $(I+UU^t)^{-1}$ is obtained by performing a (normalized) GSO on the columns of

$$(U)$$

and $Q = (\frac{c_1}{\|c_1\|} | \ldots | \frac{c_k}{\|c_k\|}).$
5. \[ A^+ = P(I + U)^t (I+UU^t)^{-1} B Q^t. \]

Comment: A Fortran listing for this procedure is given in Rust, Burrus and Schneeburger [1].

(5.1.19) Example:

\[
A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}
\]

\[
c_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad c_2 = \begin{pmatrix} -1/2 \\ 1/2 \end{pmatrix}, \quad c_3 = c_4 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\]

\[ k = 2, \quad P = I, \quad R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & 1 \end{pmatrix}.\]

\[ \gamma_1 = 1,\quad \gamma_2 = -1/2,\quad \alpha_1 = 1,\quad \alpha_2 = 1,\quad \alpha_3 = -1,\quad \alpha_4 = 0.\]

\[ \beta_{11} = 1/\sqrt{2},\quad \beta_{22} = 2/\sqrt{6},\quad \beta_{12} = -1/\sqrt{6}.\]

\[ B = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{6} \\ 2/\sqrt{6} & -1/\sqrt{6} \end{pmatrix},\quad Q = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{6} \end{pmatrix}.\]

\[ (U \left| T \right. ) = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{\text{GSQ}} \quad V = \begin{pmatrix} \sqrt{3} & -2/\sqrt{15} \\ 1/\sqrt{3} & -1/\sqrt{15} \\ -3/\sqrt{15} & 1/\sqrt{15} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.\]

\[ I - V_1 V_1^t = \begin{pmatrix} 1/5 & i^2 & 1/3 \end{pmatrix} = (I+UU^t)^{-1}.\]

\[ (I+UU^t)^{-1} B Q^t = \begin{pmatrix} 1/15 & 3 & 0 & 3 \\ -1 & 5 & 4 & 3 \\ -5 & -1 & 4 & 3 \\ 3 & 0 & 3 \end{pmatrix}.\]

\[ A^+ = (U \left| T \right. ) (I+UU^t)^{-1} B Q^t = \begin{pmatrix} 3 & 0 & 3 \\ -1 & 5 & 4 \\ -5 & -1 & 4 \\ 3 & 0 & 3 \end{pmatrix}.\]

78
(5.1.20) Exercise: Alternate method for inverting $(I+UU^t)$:

Let the columns of $U$ be denoted by $u_1, u_2, \ldots, u_r$. Let

$$W_0 = I \quad \quad W_n = W_{n-1} - \frac{(W_{n-1}u_n)(W_{n-1}u_n)^t}{1 + u_n^tW_{n-1}u_n}$$

$n = 1, \ldots, r$.

Then for each $n$,

$$W_n = (I + \sum_{j=1}^{n} u_ju_j^t)^{-1}$$

hence

$$W_r = (I+UU^t)^{-1}.$$  

(Hint: Use (4.6.5) or prove directly by induction on $n$.)

(5.1.21) Exercise: Alternate method for the pseudo inversion of $(I | U)$ (Tewarson, [2]):

a) If $P$ is square and $P^t(I | U) = S^t$ where $S^tS = I$, then $(I | U)^+ = SP^t$.

b) If the GSO is applied to the columns of $\begin{bmatrix} I \\ U \end{bmatrix}$ and the resulting set of orthonormal vectors are denoted by $s_1, s_2, \ldots, s_k$, then there is a $k \times k$ matrix, $P$, such that

$$\begin{bmatrix} I \\ U \end{bmatrix}P = (s_1 | s_2 | \ldots | s_k),$$

c) Let $S = (s_1 | s_2 | \ldots | s_k)$ be partitioned:

$$S = \begin{bmatrix} I \\ U \end{bmatrix}$$

$k$ \quad $r$

Then

$$P = S_1$$

and

$$(I | U)^+ = SS^t.$$  

(5.2) Method II (Based on Gauss-Jordan Elimination, Ben Israel and Wersan [1], Noble [1].)

If $A$ is an $n \times n$ matrix of rank $k$, there always exists a non-singular matrix $E$ and an orthogonal matrix $P$ such that
(5.2.1) \[ EA^tAP = \begin{pmatrix} k & m-k \\ m-k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} I \\ D \\ 0 \end{pmatrix}. \]

Indeed, there are many such E and P. For instance, if P diagonalizes \( A^tA \) in such a way that the nonzero eigenvalues of \( A^tA \) appear in the upper left hand corner:

\[ P^tA^tAP = \begin{pmatrix} k & m-k \\ m-k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D \\ 0 \\ 0 \end{pmatrix}, \]

then

\[ E = \begin{pmatrix} D^{-1} \\ 0 \\ 0 \end{pmatrix} P^t \]

will do the trick:

\[ EA^tAP = \begin{pmatrix} D^{-1} \\ 0 \\ 0 \end{pmatrix} P^t AA^tP = \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix} \]

which is a stronger version of (5.2.1). However, the method to be described here only requires that E and P be chosen to satisfy (5.2.1).

The present method is based upon the following identity:

(5.2.2) Theorem: If E is nonsingular and P is orthogonal, chosen to satisfy (5.2.1), then

\[ A^t = P(EA^tAP)^tEA^t. \]

Proof: The equation

\[ A^tAX = A^t \]

always has a solution in X, (3.13.1, 3.13.2), hence the equation

(5.2.3) \[ A^tAPY = A^t \]

always has a solution in Y since P is nonsingular, hence

(5.2.4) \[ EA^tAPY = EA^t \]

always has a solution in Y.

Among all solutions to (5.2.3), the unique matrix which minimizes
\[ \text{tr}(Y^tY) \text{ is} \]
\[ Y = (EA^t A^t)^t EA^t, \]  
\[ (3.13.8) \]

The set of matrices, \( Y \), satisfying (5.2.4) is the same as the set of matrices satisfying (5.2.3) since \( E \) is nonsingular, so \( \hat{Y} \) also minimizes
\[ \text{tr}(Y^tY) = \text{tr}[(PY)^t(PY)], \quad (P \text{ is orthogonal}), \]
over that class. By (3.13.8),
\[ \hat{PY} = (A^t A^t)^t = A^t \]
and the theorem follows when (5.2.6) and (5.2.5) are combined.

**5.2.7 Corollary:** If \( E \) is nonsingular and \( P \) is orthogonal and
\[ EA^tAP = \begin{pmatrix} H & 0 \\ 0 & 0 \end{pmatrix} \]
then
\[ A^t = P(H^t \mid 0)(EA^t). \]

It remains to show how to compute \( EA^t, H^t \) and \( P \):

a) **Computation of \( H, EA^t \) and \( P \):**

Write down the augmented matrix, \( m(A^t A^t \mid A^t) \), and perform row operations on this matrix until the left block is reduced to the point where a permutation of its columns will result in the row echelon form:

\[
\begin{align*}
\begin{bmatrix} m & n \\ m(A^t A^t \mid A^t) \end{bmatrix} & \xrightarrow{\text{row operations}} \begin{bmatrix} m & n \\ \text{only (step 1)} \end{bmatrix} \\
\begin{bmatrix} m(A^t A^t \mid A^t) \end{bmatrix} & \xrightarrow{\text{permute first}} \begin{bmatrix} m & n \\ \text{m columns} \end{bmatrix} = \begin{bmatrix} k & m-k \\ \text{(step 2)} \end{bmatrix} \begin{bmatrix} I \mid L \\ 0 \mid 0 \end{bmatrix} \begin{bmatrix} EA^t \end{bmatrix}.
\end{align*}
\]

The elementary row operations are reversible so \( E \) is nonsingular, as required. The permutation matrix, \( P \), is orthogonal and is recorded as the columns of \( EA^t \) are permuted to achieve row echelon form. The
matrix $EA^t$ occupies the right hand block of the augmented matrix just after step 1. The matrix, $H$, is $k\times m$ and is given by $(I\mid L)$ after step 2.

b) Computation of $H^+$:

$$H^+ = (I\mid L)^+ = \begin{pmatrix} I_L \mid L^t \end{pmatrix} (I+LL^t)^{-1}.$$ 

Evaluate $(I+LL^t)^{-1}$ either by (5.1.18) or (5.1.20), or evaluate $(I\mid L)^+$ by (5.2.8)

Example:

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

$$(A^tA\mid A) = \begin{bmatrix} 4 & -2 & -2 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ -2 & 4 & -2 & -8 & 0 & 1 & -1 & 1 & -1 & 0 \\ -2 & -2 & 4 & 10 & 1 & 0 & 1 & -1 & 0 & -1 \\ -2 & -8 & 10 & 28 & 2 & -1 & 3 & -3 & 1 & -2 \end{bmatrix}.$$

Pivot on the 4-th element of the first row (which means "reduce the 4-th element of all the other rows to zero by subtracting an appropriate multiple of the first row from each"):

Nextly, pivot on the third element of the second row:

$$\begin{bmatrix} 4 & -2 & -2 & -2 & -2 & -1 & -1 & 0 & 0 & 1 & 1 \\ -18 & 12 & 0 & 4 & 5 & -1 & 1 & -5 & -4 \\ 18 & -12 & -6 & 0 & -4 & -5 & 1 & -1 & 5 & 4 \\ 54 & -36 & -18 & 0 & -12 & -15 & 3 & -3 & 15 & 12 \end{bmatrix}.$$
Next, divide the first row by $-2$, the second by $6$:

\[
\begin{bmatrix}
1 & -1 & 0 & 1 & -1/6 & -2/6 & 1/6 & -1/6 & 2/6 & 1/6 \\
-3 & 2 & 1 & 0 & 4/6 & 5/6 & -1/6 & 1/6 & -5/6 & -4/6 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

$= (E A^t A' + E A^t)$.

This completes step $1$; $E A^t$ is the right hand $4 \times 6$ block. The left hand block is reduced to row echelon form by postmultiplication by

\[
P = \begin{bmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0
\end{bmatrix}
\]

which puts column 4 into column 1, column 3 into column 2, column 1 into column 3, and column 2 into column 4. The result:

\[
E A^t A P = \begin{bmatrix} H \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}
\]

The matrix $L$ is the upper right hand block:

$L = \begin{bmatrix} 1 & -1 \\ -3 & 2 \end{bmatrix}$, \quad $I + LL^t = \begin{bmatrix} 3 & -5 \\ -5 & 14 \end{bmatrix}$.

$(I + LL^t)^{-1} = \frac{1}{17} \begin{bmatrix} 14 & 5 \\ 5 & 3 \end{bmatrix}$.

\[
H^+ = \begin{bmatrix} (I + LL^t)^{-1} \\ L^t(I + LL^t)^{-1} \end{bmatrix} = \frac{1}{17} \begin{bmatrix} 14 & 5 \\ 5 & 3 \\ -1 & -4 \end{bmatrix}
\]

$(H^+ | 0) E A^t =$
\[
\begin{pmatrix}
\frac{1}{102} & (6 & -3 & 9 & -9 & 3 & -6) \\
& (-15 & 5 & 2 & -2 & -5 & -7) \\
& (8 & 13 & 3 & -3 & 18 & 15) \\
& (8 & 13 & -5 & 5 & -13 & -8) \\
& (-6 & 3 & 9 & -9 & 3 & -6)
\end{pmatrix}
\]

\[
P(H^+10E)A^+ = \begin{pmatrix}
-15 & -18 & 3 & -3 & 18 & 15 \\
8 & 13 & -5 & 5 & -13 & -8 \\
7 & 5 & 2 & -2 & -5 & -7 \\
6 & 3 & 9 & -9 & 3 & -6
\end{pmatrix} = A^+.
\]

(5.2.9) Exercise: Let \(E, A\) and \(P\) be as defined in (5.2.1). Show that the last \(m-k\) rows of \(EA^t\) are always zero. (Hint: \(XA^tA = 0\) if and only if \(XA^t = 0\). Specialize to matrices of the form
\[
X = n(0; Y_1).
\]

Comment: A GATE 20 program for this procedure is given in Ben Israel and Ijiri [1]. Additional refinements are contained in Tewarson [1].

(5.3) Method III (Based on Gradient Projection Method, Pyle [1].)

If \(A\) is an \(n \times m\) matrix and \(b \in \mathbb{R}(A)\), the equation
\[
(5.3.1) \quad Ax = b
\]
has at least one solution and
\[
(5.3.2) \quad \hat{x} = A^+ b
\]
is the only solution which lies in \(\mathbb{R}(A^t)\), (3.1b).

Let the columns of \(A^t\) be denoted by \(a_1, a_2, \ldots, a_n\) and denote by \(A_k\) (resp. \(b_k\)) the \(k \times m\) matrix (resp. \(k\)-vector) obtained by deleting the last \(n-k\) rows (resp. components) from \(A\) (resp. \(b\)). Since (5.3.1) has a solution, so does
\[
(5.3.3) \quad A_k x_k = b_k \quad k = 1, 2, \ldots, n,
\]
since the set of simultaneous equations which (5.3.3) represents is a subset of those represented by (5.3.1).
Furthermore,
\[
\hat{x}_k = A^+_k b_k
\]
is the only solution to (5.3.3) which lies in \( \mathcal{Q}(A_k^+) = \mathcal{L}(a_1, \ldots, a_k) \), (5.4.3.1b). The present procedure develops a simple recursion which relates \( \hat{x}_{k+1} \) to \( \hat{x}_k \). Since \( \hat{x} = A^+ b \) is the same as \( \hat{x}_n \), the recursion can be carried out \( n \) times and the desired solution to (5.3.1) results, provided \( b \) is in \( \mathcal{Q}(A) \). Extensions to the general case follow readily as we will show:

By definition, \( \hat{x}_1 \) is the unique vector in \( \mathcal{L}(a_1) \) satisfying
\[
a_1^+ x_1 = \beta_1
\]
where \( \beta_j \) is the \( j \)-th component of \( b \) \((j = 1, \ldots, n)\).

Obviously
\[
\hat{x}_1 = (a_1^+)^+ \beta_1 = \beta_1 a_1 \left( \| a_1 \|^{2} \right)^+.
\]

If \( \hat{x}_{k-1} \) is known, we construct \( \hat{x}_k \) as follows:

Let \( (h_1, h_2, \ldots, h_n) \) be the non-normalized set of vectors which result from a Gramm-Schmidt Orthogonalization of \( (a_1, \ldots, a_n) \):

\[
h_1 = a_1
\]
\[
h_k = a_k - \sum_{j \in S_k} (a_k^+ h_j) h_j / \| h_j \|^2
\]
where
\[
S_k = \{ j : j \leq k-1 \text{ and } \| h_j \| \neq 0 \} \quad k = 1, \ldots, n.
\]

We have shown that \( h_k \perp \mathcal{L}(a_1, \ldots, a_{k-1}) \), (2.8), and since \( \hat{x}_{k-1} \) satisfies
\[
(5.3.7) \quad A_{k-1} \hat{x}_{k-1} = b_{k-1},
\]

85
it follows that

\begin{equation}
\hat{a}_j^t(\hat{x}_{k-1} + \hat{\alpha}_k h_k) = \beta_j \quad j=1,2,\ldots,k-1
\end{equation}

for any scalar \( \hat{\alpha}_k \). (Just write (5.3.7) in componentwise form.)

In particular, if

\begin{equation}
\hat{\alpha}_k = \begin{cases} 
0 & \text{if } h_k = 0 \\
(\beta_k - a_k^t x_{k-1})/(h_k a_k) & \text{otherwise}
\end{cases}
\end{equation}

then

\begin{equation}
\hat{y}_k = \hat{x}_{k-1} + \hat{\alpha}_k h_k
\end{equation}

satisfies

\begin{equation}
\hat{a}_j^t \hat{y} = \beta_j \quad j \in S_{k+1}
\end{equation}

and

\begin{equation}
\hat{y}_k \in \mathcal{L}(h_1,\ldots,h_k).
\end{equation}

By construction,

\[ \mathcal{L}(h_j; j \in S_{k+1}) = \mathcal{L}(h_1,\ldots,h_k) = \mathcal{L}(a_1,\ldots,a_k) = \mathcal{L}(a_j; j \in S_{k+1}) \]

since \( h_j = 0 \) if and only if \( a_j \) is a linear combination of its predecessors, (2.8.6). Therefore, \( \hat{y}_k \) is the unique vector in \( \mathcal{L}(a_j; j \in S_{k+1}) \) satisfying (5.3.11). On the other hand,

\[ \hat{x}_k = A_k^t h_k \in \mathcal{L}(a_1,\ldots,a_k) = \mathcal{L}(a_j; j \in S_{k+1}) \]

satisfies

\begin{equation}
\hat{a}_j^t x = \beta_j \quad j=1,\ldots,k
\end{equation}

and hence it also satisfies (5.3.11). Since \( \hat{y}_k \) is the unique vector in

\[ \text{86} \]
\( \mathcal{L}(x_j; j \in S_{k+1}) \) satisfying (5.3.11), \( \hat{\gamma}_k = \hat{x}_k \).

In summary:

\( (5.3.14) \textbf{Theorem:} \) If \( A^t = (a_1 | a_2 | \ldots | a_n) \) and \( b = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \in \mathcal{Q}(A) \), then

\[
A^+b = \hat{x}_n
\]

where

\[
\hat{x}_k = \begin{cases} 
0 & \text{if } h_k = 0 \\
\hat{x}_{k-1} + \hat{\alpha}_k & \text{otherwise} 
\end{cases}
\]

\( \hat{\alpha}_k = \begin{cases} 
0 & \text{if } h_k = 0 \\
(\beta_k - a_k^t \hat{x}_{k-1})/(h_k a_k) & \text{otherwise} 
\end{cases}
\]

and \( (h_1, h_2, \ldots, h_n) \) are the nonnormalized vectors resulting from a GSO of \( (a_1, a_2, \ldots, a_n) \).

In general, if \( b \notin \mathcal{Q}(A) \) (or if you are not able to decide whether \( b \in \mathcal{Q}(A) \) or not, conveniently) resort to the following trick:

Let

\[
d = A^t b.
\]

and

\[
c = A^+ A.
\]

Then \( d \in \mathcal{Q}(A^t) = \mathcal{Q}(C) \) so that (5.3.14) can be used to compute \( C^+ d \).

Since \( C^+ d = (A^t A)^+ A^t b \), this additional step generalizes the procedure to the case of arbitrary \( b \).

If, instead of \( A^+ b \), it is desired to compute \( A^+ \), proceed in the following manner:

First, perform a GSO on the columns of \( A \). Denote the normalized set of nonzero (orthonormal) vectors by \( d_1, d_2, \ldots, d_r \). These vectors

87
span the same linear manifold that the columns of \( A \) span, namely \( \mathcal{R}(A) \). By (3.7.3), it therefore follows that

\[
AA^+ = \sum_{j=1}^{r} d_j d_j^t .
\]

Denote the columns of \( AA^+ \) by \( b_1, b_2, \ldots, b_n \). Since \( \mathcal{R}(A) = \mathcal{R}(AA^+) = \mathcal{L}(b_1, b_2, \ldots, b_n) \), it must be that \( b_j \in \mathcal{R}(A) \) for each \( j \) so that (5.3.14) can be used \( n \) times to compute \( A^+ b_j \) \((j=1, \ldots, n)\).

Since

\[
(A^+ b_1 \mid A^+ b_2 \mid \ldots \mid A^+ b_n) = A^+(b_1 \mid b_2 \mid \ldots \mid b_n) = A^+(AA^+) = A^+,
\]

this procedure generates the desired result.

**Summary of Method III:**

1) If it is known that \( b \in \mathcal{R}(A) \) then (5.3.14) can be used directly to construct \( A^+ b \).

2) If it is not known whether or not \( b \in \mathcal{R}(A) \), compute \( d = A^+ b \),

\( C = A^+ A \) and use (5.3.14) to compute \( C^+ d \) which coincides with \( A^+ b \).

3) To compute \( A^+ \), first perform a GSO on the columns of \( A \). Denote the resulting orthonormal vectors by \( d_1, d_2, \ldots, d_r \). Construct the matrix \( \sum_{j=1}^{r} d_j d_j^t \) and denote the columns of this matrix by \( b_1, b_2, \ldots, b_n \).

Then for each \( j \), \( b_j \in \mathcal{R}(A) \) and (5.3.14) can be used to construct \( A^+ b_j \) \((j=1, \ldots, n)\).

\[
A^+ = (A^+ b_1 \mid A^+ b_2 \mid \ldots \mid A^+ b_n).
\]

**Exercise:** In (5.3.14) we defined

\[
\hat{\alpha}_k = \begin{cases} 
0 & \text{if } b_k = 0 \\
\frac{(b_k - a_k^+ x_k x_{k-1})}{(b_k^+ a_k)} & \text{otherwise.}
\end{cases}
\]
Show that $h_k^t a_k \neq 0$ if $h_k \neq 0$.

(5.3.16) Example: Compute $A^+$ where $A$ is given by (5.1.19):

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}.$$  

The columns of $A$ are

$$c_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad c_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad c_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad c_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

The GSO generates the orthonormal set:

$$d_1 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}, \quad d_2 = \begin{pmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}.$$  

$$AA^+ = d_1^t d_1 + d_2^t d_2 = \frac{1}{3} \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$  

The columns of $AA^+$ are

$$b_1 = \begin{pmatrix} -2/3 \\ 1/3 \\ 1/3 \end{pmatrix}, \quad b_2 = \begin{pmatrix} -1/3 \\ 2/3 \\ 1/3 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}.$$  

To construct $A^+ b_j$, we perform a GSO on the columns of $A^t$: The columns of $A^t$ are

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$  

The nonnormalized GSO produces

$$h_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad h_2 = \begin{pmatrix} 1/3 \\ 1 \\ -2/3 \\ 1/3 \end{pmatrix}, \quad h_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$
To compute $A^*_b$, use (5.3.14):

$$\hat{\alpha}_1 = 2\sqrt{3}/9 \quad \hat{x}_1 = \frac{2}{9} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\hat{\alpha}_2 = -\sqrt{15}/45 \quad \hat{x}_2 = \frac{1}{15} \begin{pmatrix} 3 \\ -1 \\ -4 \\ 3 \end{pmatrix} = A^*_b$$

To compute $A^*_b^2$:

$$\hat{\alpha}_1 = -\sqrt{3}/9 \quad \hat{x}_1 = \frac{1}{9} \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}$$

$$\hat{\alpha}_2 = \sqrt{15}/9 \quad \hat{x}_2 = \frac{1}{3} \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} = A^*_b^2$$

To compute $A^*_b^3$:

$$\hat{\alpha}_1 = \sqrt{3}/9 \quad \hat{x}_1 = \frac{1}{9} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\hat{\alpha}_2 = 4\sqrt{15}/45 \quad \hat{x}_2 = \frac{1}{15} \begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix} = A^*_b^3$$

Thus $A^* = (A^*_b, A^*_b^2, A^*_b^3)$

$$= \frac{1}{15} \begin{pmatrix} 3 & 0 & 3 \\ -1 & 5 & 4 \\ 5 & 0 & 3 \end{pmatrix}$$

which agrees with (5.1.19).

(5.4) Method IV. (Based on the Cayley-Hamilton Theorem, Decell [2], Ben Israel and Charnes [1].)

The present method is based upon two theorems. The first uses the
classical Cayley-Hamilton relation (which says that any square matrix satisfies its own characteristic equation) to deduce an expression for the generalized inverse of a matrix in terms of its characteristic polynomial. The second theorem is a result of Fadeev and Fadeeva [1], which generates the coefficients of said characteristic polynomial in an efficient manner.

(5.4.1) **Theorem:** a) Let \( A \) be an \( n \times n \) symmetric matrix and let \( \pi(\lambda) = \det(A - \lambda I) \), which can always be factored as follows:

\[
\pi(\lambda) = \alpha \lambda^k (1 - \lambda \varphi(\lambda))
\]

where \( n - k \) is the rank of \( A \).

Then if \( A \) is nonsingular,

\[
A^{-1} = \varphi(A)
\]

and generally

\[
A^+ = \varphi(A) + \varphi(0)[A\varphi(A) - I].
\]

b) If \( H \) is an arbitrary \( m \times n \) matrix and if the characteristic polynomial of \( H^t H \) is given by \( \alpha \lambda^k (1 - \lambda \varphi(\lambda)) \) then

\[
H^+ = \varphi(H^t H)H^t.
\]

**Comment:** The matrix \( \varphi(A) \) is obtained by raising \( A \) to the appropriate powers and summing these powers using the coefficients of \( \varphi(\cdot) \) as weights.

**Proof:** a) The Cayley-Hamilton Theorem, (c.f., Bellman [1]), states that \( A \) satisfies the matrix equation

\[
\pi(A) = 0.
\]

Therefore

\[
\alpha A^k (I - A\varphi(A)) = 0.
\]
If $A$ is nonsingular, $k=0$ and hence

$$A\varphi(A) = I$$

so that $\varphi(A) = A^{-1}$.

Generally,

$$A^k = A^{k+1}\varphi(A).$$

Since

$$(A^+)^{k+1}A = A^+(A^+A)^k$$

$$= A^+(A^+A), \quad (3.11.3),$$

$$= A^+(AA^+), \quad (3.11.6),$$

$$= A^+, \quad (3.9.3),$$

it follows that

$$A^+ = (A^+)^{k+1}A^{k+1}\varphi(A).$$

Again, using (3.11.3) and (3.11.6), we see that

(5.4.1.1) $$A^+ = AA^+\varphi(A).$$

Since $\varphi(A)-\varphi(0)$ involves only positive powers of $A$ and since

$$AA^+A^\alpha = A^\alpha \text{ if } \alpha > 0, \quad (3.11.3),$$

it follows from (5.4.1.1) that

(5.4.1.2) $$A^+ = AA^+[\varphi(A)-\varphi(0)] + AA^+\varphi(0) = \varphi(A)-\varphi(0) + AA^+\varphi(0)$$

and so

$$AA^+ = A\varphi(A) + A(I-AA^+)\varphi(0).$$

Since $AA^+ = A^+A$ for symmetric $A$, (3.11.6), the second term vanishes.
and

(5.4.1.3) \[ AA^+ = A\varphi(A) \, . \]

Since \( \varphi(0) \) is a multiple of the identity, it commutes with \( A\varphi(A) \) and so, if (5.4.1.3) is inserted into the right side of (5.4.1.2) we find

\[ A^+ = \varphi(A) + (A\varphi(A) - I)\varphi(0) \]
\[ = \varphi(A) + \varphi(0)(A\varphi(A) - I). \]

b) From a),

\[ (H^tH)^+ = \varphi(H^tH) + \varphi(0)[H^tH\varphi(H^tH) - I] \, . \]

From (5.4.1.3) and (3.11.7)

\[ H^tH\varphi(H^tH) = H^tH \]

so that

\[ H^+ = (H^tH)^+H^t = \varphi(H^tH)H^t + \varphi(0)[H^tH\varphi(H^tH) - I]H^t \]
\[ = \varphi(H^tH)H^t + \varphi(0)(H^tH - I)H^t \]
\[ = \varphi(H^tH)H^t \, , \text{ (since } H^tHH^t = H^t). \]

Superficially, the last result appears to be of theoretical interest only, owing to the fact that \( \varphi(A) \) is defined in terms of the characteristic polynomial of \( A \) which, in turn, is defined in terms of a determinant which requires a prohibitively large amount of computation for large matrices. The following result of Fadeev and Fadeeva, [1], reduces that problem to manageable proportions:

(5.4.2) Theorem: Let \( A \) be an \( n \times n \) symmetric matrix and define
\[
\begin{aligned}
\begin{cases}
A_1 = A \\
\gamma_k = \text{tr}A_k / k \\
B_k = A_k - \gamma_k I \\
A_{k+1} = AB_k
\end{cases}
\end{aligned}
\]

\[k=1,2,...,n-1.\]

(\text{tr } A_k \text{ is the sum of } A_k\text{'s diagonal elements.})

Let

\[M = \begin{cases} 
\text{the first value of } k \leq n \text{ for which } AB_k = 0 & \text{if such exists,} \\
n & \text{otherwise,}
\end{cases}\]

and let

\[r \text{ be the largest } k \leq M \text{ for which } \gamma_k \neq 0.\]

Then,

\[\pi(\lambda) = \det(A-\lambda I) = (-1)^n \gamma_r \lambda^{n-r} (1-\varphi(\lambda))\]

where

\[\varphi(\lambda) = \frac{1}{\gamma_r} [\lambda^{r-1} - \sum_{j=1}^{r-1} \gamma_j \lambda^{r-1-j}] .\]

Proof: The proof consists of two parts: First we show that \(\gamma_k\) is the coefficient of

\[(-1)^{n+1} \lambda^{n-k}\]

in the expansion of \(\det(A-\lambda I)\) for \(k=1,...,n\). Then we use that result to show that \(\gamma_k = 0\) for \(k \geq r+1\). The desired conclusion follows easily thereafter.

Proof that

\[\det(A-\lambda I) = (-1)^n [\lambda^n - \sum_{k=1}^{n} \gamma_k \lambda^{n-k}] .\]
\[ \gamma_1 = \text{tr} A \] which is the coefficient of \((-1)^{n+1} \lambda^{n-1}\) in the expansion of \(\det(A - \lambda I)\). We proceed by induction, supposing that \(\gamma_1, \gamma_2, \ldots, \gamma_{k-1}\) are, respectively, the coefficients of \((-1)^{n+1} \lambda^{n-1}, (-1)^{n+1} \lambda^{n-2}, \ldots, (-1)^{n+1} \lambda^{n-k+1}\) in the expansion of \(\det(A - \lambda I)\).

Define
\[ s_k = \text{tr}(A^k). \]

From (5.4.2.1),
\[ A_{k+1} = AA_k - \gamma_k A, \]
which can be iterated backwards to obtain
\[ A_k = A^k - \sum_{j=1}^{k-1} \gamma_j A^{k-j} \quad k=1,2,\ldots,n \]
and hence from (5.4.2.7)
\[ \text{tr}(A_k) = s_k - \sum_{j=1}^{k-1} \gamma_j s_{k-j}. \]

It is well known that for any \(A\), \(\text{tr}(A)\) is the sum of \(A\)'s eigenvalues, and if \(\lambda_1, \lambda_2, \ldots, \lambda_n\) are the eigenvalues of the symmetric matrix \(A\), then \(\lambda_1^k, \ldots, \lambda_n^k\) are the eigenvalues of \(A^k\), so that
\[ s_k = \sum_{j=1}^{n} \lambda_j^k. \]

Furthermore, Newton's formula states that for any polynomial, \(\lambda^n = \sum_{j=1}^{n} \beta_j \lambda^{n-j}\), having roots at \(\lambda_1, \lambda_2, \ldots, \lambda_n\), the relations
\[ k \beta_k = s_k - \sum_{j=1}^{k-1} \beta_j s_{k-j} \]

95
hold for \( k=1,2,\ldots,n \), where
\[
s_k = \sum_{j=1}^{k} \lambda_j^k,
\]
(Bocher, [1]).

Writing
\[
\det(A-\lambda I) = (-1)^n[\lambda^n - \sum_{j=1}^{n} \beta_j \lambda^{n-j}]
\]
we see that under the induction hypothesis, \( \beta_j = \gamma_j \) \((j=1,2,\ldots,k-1)\)
so by (5.4.2.10),
\[
(5.4.2.11) \quad k \beta_k = a_k - \sum_{j=1}^{k-1} \gamma_j a_{k-j}.
\]

By (5.4.2.9) and (5.4.2.1)
\[
k \beta_k = \text{tr}(A_k) = k \gamma_k
\]
which shows that \( \gamma_k = \beta_k \) and hence establishes the induction hypothesis
for all \( k \).

The recursion (5.4.2.1) guarantees that \( A_k = 0, B_k = 0 \) and \( \gamma_k = 0 \)
for all \( k > M \) if \( AB_k = 0 \) for some \( M < n \). The definition of \( r \) therefore implies that \( \gamma_k = r \) for \( k > r \). From this and (5.4.2.6),
\[
\det(A-\lambda I) = (-1)^n[\lambda^n - \sum_{j=1}^{r} \gamma_j \lambda^{n-j}]
\]
\[
= (-1)^n \gamma_r \lambda^{n-r}[1 - \frac{\lambda}{\gamma_r} (\lambda^{r-1} - \sum_{j=1}^{r-1} \gamma_j \lambda^{r-1-j})].
\]

This proves (5.4.2.4) and (5.4.2.5). \( \blacksquare \)

The application of (5.4.1) and (5.4.2) to the pseudo inversion of
an \( n \times m \) matrix is particularly charming:

(5.4.3) **Theorem:** Let \( A = H^t H \) (where \( H \) is an arbitrary \( m \times n \) matrix),
and define \( A_k, B_k, \gamma_k, M \) and \( r \) as in (5.4.2). Then
(5.4.3.1) \[ H^+ = \frac{1}{\gamma_r} B_{r-1} H^+ \]

and the rank of \( H \) is \( r \).

**Proof:** The rank of \( H \) is the same as the rank of \( A \). The rank of \( A \) is

\[ n - (\text{the multiplicity of the characteristic root at } \lambda = 0), \]

since \( A \) is symmetric. By (5.4.2) this multiplicity is \( n-r \), which proves that the rank of \( H \) and \( A \) is \( r \).

By (5.4.2.5),

\[ \varphi(A) = \frac{1}{\gamma_r} [A^{r-1} - \sum_{j=1}^{r-1} \gamma_j A^{r-1-j}] \]

whereas

\[ A_{r-1} - \gamma_{r-1} I = A^{r-1} - \sum_{j=1}^{r-1} \gamma_j A^{r-1-j} \quad \text{by (5.4.2.8)}. \]

Thus

\[ \varphi(A) = \frac{1}{\gamma_r} (A_{r-1} - \gamma_{r-1} I) \]

\[ = \frac{1}{\gamma_r} B_{r-1}. \]

The conclusion follows from (5.4.1b). ■

**Summary of Method IV**

To find \( H^+ \) where \( H \) is an \( m \times n \) matrix:

1) Let \( A_1 = H^T H \),

2) \( \gamma_k = \text{tr}(A_k/k), B_k = A_k - \gamma_k I, A_{k+1} = AB_k \),

\[ M = \begin{cases} \text{first value of } k \text{ for which } AB_k = 0 \text{ if such exists and } & \\
\text{is less than } n & \\
n \text{otherwise}, & \end{cases} \]

97
\[ r = \text{largest } k \leq M \text{ for which } \gamma_k \neq 0, \]
\[ H^+ = \frac{1}{\gamma_r} B_{r-1} H^+ . \]

**Comment:** Although it is true that \( A^{-1} = \varphi(A) \) for nonsingular symmetric \( A \), it is not generally true that \( A^+ = \varphi(A) \) as evidenced by the example

\[ A = \text{diag}(0,0,1/2,1/3). \]

For this example

\[ \varphi(\lambda) = 5-6\lambda \]

so that

\[ \varphi(A) = \text{diag}(5,5,2,3) \]

whereas

\[ A^+ = (0,0,2,3) = \varphi(A) + \varphi(0)[A \varphi(A) - I]. \]

(5.4.4) **Example:**

\[
H = \begin{pmatrix}
1 & 1 & 1 \\
1 & 0 & -1 \\
2 & 1 & 0 \\
3 & 1 & -1
\end{pmatrix}
\quad A_1 = A = H^+ H = \begin{pmatrix}
15 & 6 & -3 \\
6 & 3 & 0 \\
-3 & 0 & 3
\end{pmatrix}
\]

\[ \gamma_1 = 21 \]

\[ B_1 = A_1 - 21 I = \begin{pmatrix}
-6 & 6 & -3 \\
6 & -18 & 0 \\
-3 & 0 & -18
\end{pmatrix} \]

\[ A_2 = A_2 = 9 \quad AB_1 = \begin{pmatrix}
-5 & -2 & 1 \\
-2 & -2 & -2 \\
1 & -2 & -5
\end{pmatrix} \]

\[ \gamma_2 = -54 \]

\[ B_2 = A_2 + 54 I = 9 \quad A_2 = \begin{pmatrix}
1 & -2 & 1 \\
-2 & 4 & -2 \\
1 & -2 & 1
\end{pmatrix} \]

\[ A_3 = AB_2 = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix} \]

98
$M=r=2$

$$H^+ = \frac{1}{\gamma_2} R H^t = - \frac{1}{18} \begin{pmatrix}
-1 & -1 & -2 & -3 \\
-4 & 2 & -4 & 0 \\
-7 & 5 & -2 & 3
\end{pmatrix}.$$ 

(5.4.5) Exercise: Prove the Cayley Hamilton theorem for symmetric matrices. (Hint: Use the diagonalization theorem.)

(5.4.6) Exercise: Establish the result analogous to (5.4.3), using the characteristic polynomial for $HH^t$ instead of $H^tH$.

(5.4.7) General Comments:

In Methods I-IV, the option always exists between working with $A$ or $A^t$. For example in Method I, $A^+$ can be found by performing a GSO on the columns of $A$ or one can compute $(A^t)^+$ by performing a GSO on the columns of $A^t$. Similarly, $(I + UU^t)^{-1}$ can be evaluated by performing a GSO on the columns of $U^tI$ or on the columns of $U^tI$. The proper choice is governed mainly by the size of the matrices involved.

All four methods depend crucially upon the computational determination of the rank of the matrix being pseudo inverted. In some of these cases, this is done during the course of a GSO, at that point where one decides whether a vector which is numerically close to zero, is actually zero. In method II, a decision must be made as to whether the Gauss-Jordan elimination has reduced a row to zero or not. If an error (usually due to roundoff) is made at that stage, like as not, the error will be propagated through the remainder of the procedure in a discontinuous manner and cause large errors in the final result. Several authors have addressed themselves to this problem from the practical point of view. The interested reader is referred to Peryra and Rosen [1], Golub and Kahan [1, Golub][1], [2].
The recursions of (4.3) and (4.6) can be interpreted as computational algorithms for $C^+_m$ and $(C_m C_m^\dagger)^+$ but the comments above, apply quite strongly in these cases. Extreme care must be exercised in making the decision whether or not a vector $c_{m+1}$ is a linear combination of its predecessors. We will return to these recursions in Chapters VIII and IX.
BIBLIOGRAPHY

Albert, Arthur

Albert, Arthur, and Sittler, Robert

American Statistical Association
*Proceedings of the 105th regional meeting*, Florida State University (1965).

Anderson, T. W.

Battin, Richard H.

Battin, Richard H. and Levine, Gerald
[1] Application of Kalman filtering techniques to the Apollo program.

Bellman, Richard

Ben-Israel, A.

Ben-Israel, A. and Charnes, A.

Ben-Israel, A. and Charnes, A
Ben-Israel, A., Charnes, A., and Robers, P. D.

Ben-Israel, A. and Ijiri, Y.

Ben-Israel, A. and Robers, P. D.

Ben-Israel, A. and Wersan, S. J.

Bocher, M.

Boullion, Thomas and Odell, Partrick

de Bröder, G. G. and Charnes, A.

Butler, T. and Martin, A. V.

Cline, R. E.

Cline, R. E.

Cline, R. E.
Decell, Henry P.

Decell, Henry P.

Decell, Henry P. and Odell, P. L.

Fadeev, D. K. and Fadeeva, V. N.

Feller, William

Goldman, A. J. and Zelen, J.

Golub, G.

Golub, G.

Golub, G., and Kahan, W.

Good, I. J.

Greville, T. N. E.
Greville, T. N. E.


Greville, T. N. E.


Halmos, Paul R.


Kalman, R. E.


Karlin, Samuel


Kruskal, W.


Marsaglia, G.


Mitra, Soko and Rao, C. R.


Moore, Eliakim Hastings.


Moore, Eliakim Hastings


Noble, Ben

Penrose, R.

Penrose, R.

Pereyra, V. and Rosen, J. B.

Price, C. M.

Pyle, L. D.

Rao, C. R.

Rust, B., Burrus, W. R., and Schneeberger, C.

Scheffe, Henry

Stewart, G. W.

Tewarson, R. P.

Tewarson, R. P.
Watson, Geoffrey S.

Zyskind, George
[1] On canonical forms, nonnegative covariance matrices and best
and simple least squares linear estimators in linear models.

Zyskind, George and Martin, Frank B.
in linear models with arbitrary nonnegative covariance structure.
1. ORIGINATING ACTIVITY (Corporate author)
   Dept. of Statistics
   Stanford University
   Stanford, California

2. REPORT TITLE
   THE FENROSE-MOORE PSEUDO INVERSE WITH DIVERSE STATISTICAL APPLICATIONS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
   TECHNICAL REPORT

5. AUTHOR(S) (Last name, first name, initial(s))
   ALBERT, ARTHUR

6. REPORT DATE
   April 1, 1971

8a. CONTRACT OR GRANT NO.
   N00014-67-A-0112-0053

8b. PROJECT NO.
   NR-042-267

9a. ORIGINATOR'S REPORT NUMBER(s)
   No. 173

9b. OTHER REPORT NO(s) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES
    Unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
   Statistics & Probability Program Code 436
   Office of Naval Research
   Arlington, Va. 22217

13. ABSTRACT.

   The Penrose-Moore Pseudo Inverse extends the notion of "inverse" for square nonsingular matrices to the class of all rectangular matrices. In these reports we develop the essential properties with applications to the theory of equations, constrained and unconstrained least squares, nonnegative definiteness, perturbation theory and the singular decomposition theorem.

   Various computational algorithms are developed and additional results are derived which apply to various statistical topics, such as the General Linear Hypothesis (BLUE's, Orthogonal Designs, tests and confidence sets) Conditional Expectations for vector normal variables and Kalman Filtering.
Penrose-Moore Pseudo Inverse
Projection
Least Squares
Computational Methods
Constrained Least Squares
Recursive Estimation

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual Enter the group number. Also, when applicable, show that original markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capital points in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive date when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year or month, day, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b. 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, effort number, system number, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

   (1) "Qualified requesters may obtain copies of this report from DDC." 
   (2) "Foreign announcement and dissemination of this report by DDC is not authorized." 
   (3) "U.S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through\" 
   (4) "U.S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through\" 
   (5) "All distribution of this report is controlled. Qualified DDC users shall request through\"

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.