NASA TECHNICAL
MEMORANDUM

NASA TM X-58069
September 1971

ON THE p-q GENERALIZED INVERSE

A Dissertation Presented to
the Graduate Faculty of
Texas Technological University
In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

NASA TM X-58069

ON THE p-q GENERALIZED INVERSE

Dennis Addington Johnston
Manned Spacecraft Center
Houston, Texas

CONTENTS

CHAPTER I

INTRODUCTION

1.1 Introductory Remarks

The subject of this dissertation is a calculation algorithm for the p-q solution of the degenerate linear system

$$Y = AX \tag{1.1}$$

where A is an $m{\times}n$ linear transformation matrix with Y and X elements of the real m-dimensional and n-dimensional normed linear spaces V_m and V_n with norms $||\cdot||_m$ and $||\cdot||_n$, respectively [49, p. 83]. After Frame [13], the system is said to be *degenerate* in that $m \neq n$ or there is no exact solution X to (1.1) for a given A and Y. The *p-q solution* X of (1.1) is a special case of a best approximate solution of (1.1) when V_m and V_n are restricted, respectively, to the finite dimensional normed linear spaces $\ell^p(m)$ and $\ell^q(n)$ with norms

$$||Y||_p = (|Y_1|^p + \cdots + |Y_m|^p)^{1/p}$$

and

$$||X||_q = (|X_1|^q + \cdots + |X_n|^q)^{1/q}$$

for $1 < p < \infty$, $1 < q < \infty$ [12], [49, pp. 87-88], where a best approximate solution is defined as follows:

DEFINITION 1.1 [39]: *A best approximate solution of the equation* $f(X) = G$ *is* X_0 *if for all* X , *either*

a) $||f(X) - G|| > ||f(X_0) - G||$

or

b) $||f(X) - G|| = ||f(X_0) - G||$
and $||X|| \geq ||X_0||$

This definition is similar to the fundamental definitions found in references [5, p. 13], [6, p. 3], [29, p. 16], [34, p. 1], [37], [50, p. 79].

If we let S(A,Y) be the set of all best approximate solutions of the equation $Y = AX$, then, for a given operator A , the set-valued operator B_A which maps Y onto S(Y,A) will be called the *norm generalized inverse*. If we specialize V_m and V_n to $\ell^p(m)$ and $\ell^q(n)$, then we will call B_A the *p-q generalized inverse*.

This generalized inverse was suggested by P. L. Odell, introduced by M. Meicler [30, p. 39], and developed by Meicler, Odell, and Newman [33], [36], [37].

Some properties of the norm generalized inverse of A are given in Chapter II as well as defining a norm generalized inverse for the norm generalized inverse B_A of A and examining some of its properties. The definition and convergence theorems for an algorithm to calculate the p-q generalized inverse are developed in Chapter III. A basic definition and some notation needed in the subsequent chapters will be presented next.

1.2 A Basic Definition and Some Notation

DEFINITION 2.1: *For any* $m \times n$ *matrix* A *and* $n \times m$ *matrix* B, *consider the four equations*

1. $ABA = A$ (2.1)
2. $BAB = B$ (2.2)
3. $(BA)^T = BA$ (2.3)
4. $(AB)^T = AB$ (2.4)

where ^{T} *indicates matrix transpose. If* B *satisfies*

a) *equation 1, then* B *is said to be a* ***generalized inverse*** *of* A *and is denoted by* $B = A^g$ *;*

b) *equations* 1 *and* 2, *then* B *is said to be a* <u>*reflexive generalized inverse*</u> *of* A *and is denoted by* $B = A^r$;

c) *equations* 1, 2, *and* 3, *then* B *is said to be a* <u>*left weak generalized inverse*</u> *of* A *and is denoted by* $B = A^n$;

d) *equations* 1, 2, *and* 4, *then* B *is said to be a* <u>*right weak generalized inverse*</u> *of* A *and is denoted by* $B = A^w$;

e) *equations* 1, 2, 3, *and* 4, *then* B *is said to be a* <u>*pseudoinverse*</u> *of* A *and is denoted by* $B = A^+$.

The four equations were introduced by Penrose [38]. His notation is used for the pseudoinverse. The names for the inverses defined in statements a), b), and e) and the notation for a), b), and c) are due to Rohde [47]. The name, weak generalized inverse, originated with Goldman and Zelen [15], but the left and right designations are due to Cline [8]. The notation for d) is from Boullion and Odell [3].

These generalized inverses will be used throughout the paper.

Also used throughout are the letters I and ϕ which are the identity matrix and the zero vector or matrix of zeros. Usage will indicate the order with I_k and ϕ_k denoting the k×k identity and k×1 column

vector of zeros if necessary. Also used is $\emptyset$ for the null or empty set. Boldface $\mathbf{N}$ and $\mathbf{R}$ are used for the null set of the operator Q

$$N(Q) = \left\{X\varepsilon V : Q(X) = \phi\right\}$$

and the range set of the operator Q

$$R(Q) = \left\{Y : Q(X) = Y, X\varepsilon V\right\} .$$

The operator Q is not necessarily linear. The symbol $\oplus$ will denote the direct sum of two subspaces [17, p. 24]. For a matrix A , A^i will denote the *ith* row of A , A_j will denote the *jth* column, and A^i_j the element in the *jth* column, *ith* row. Scalars are real numbers and are denoted by lower case Roman and Greek letters. For typing convenience, the Greek letter epsilon (ε) will be used for the set theory "element of" symbol $\in$ except where some confusion may occur with an epsilon used in limit proofs. In these cases, the symbol $\in$ will be used to denote "element of."

When a theorem or definition is known in the literature, this fact will be noted by a reference after the statement of the theorem or definition. If a known proof is included for completeness, then this fact will

be noted by a reference after the identifier "Proof" or the identifier of a subsection of the proof.

CHAPTER II

THE NORM GENERALIZED INVERSE

2.1 The Metric Projection in a Finite Dimensional Normed Linear Space

DEFINITION 1.1 [35]: *Let* V *be a real normed vector space and* M *be a subset. For* X *in* V, *let* $E_M(X)$ *denote the set of nearest points in* M *to* X, *i.e.,*

$$E_M(X) = \{Y \in M: ||X - Y|| \leq ||X - Z|| \text{ for all } Z \in M\} .$$

The set-valued mapping E_M *is called the metric projection onto* M. *Let* M *denote the set of all metric projections onto subspaces of* V.

The concept of a metric projection has been discussed by several authors, among them Blatter and Morris [2], Brown [4], Cheney and Wulbert [7], Lazar, Morris, and Wulbert [24], and others [25], [35], [40], [48], [53].

The existence of a metric projection where $E_M(X)$ is unique for all elements X is given below.

DEFINITION 1.2 [6, p. 22]: *A normed linear vector*

space is said to be <u>*strictly convex*</u> *if and only if, for all elements of the space* X *and* Y *,*
$||X|| = ||Y|| = ||(X + Y)/2|| = 1$ *implies* $X = Y$.

<u>THEOREM 1.3</u> [6, P. 23]: *In a strictly convex normed linear space* V *a nonempty real finite dimensional subspace* M *contains a unique point closest to any given point of* V .

To prove the theorem it is only necessary that M be convex, closed, complete and contained in a finite dimensional subspace. We can see this by letting $Y \varepsilon V$ and $d = \inf_{X \varepsilon M} ||X - Y||$. If $Y \varepsilon M$, Y itself is the unique point closest to Y . So let $Y \notin M$. Then for $Z \varepsilon M$ define

$$S = \left\{X : ||X-Y|| \leq ||Z-Y|| , X \varepsilon M\right\} .$$

Then

$$\begin{aligned} d &= \inf_{X \varepsilon M} \left\{||X - Y||\right\} \\ &= \inf_{X \varepsilon M} \left\{||X - Y|| : ||X - Y|| \leq ||Z - Y||\right\} \\ &= \inf_{X \varepsilon S} \left\{||X - Y||\right\} \\ &= \inf f(S) \end{aligned}$$

where $f(X) = ||X - Y||$; since ($d \leq f(X)$ for all $X \in M$) is unaffected by the requirement $f(X) \leq f(Z)$ when $d \leq f(Z)$. S is closed in M and thus in V since M is closed. Also, S is bounded by $f(Z)$.

Since $d = \inf f(S)$, there exists a monotone decreasing sequence of real numbers $\{f_i\} \in f(S)$ such that $\lim_{i \to \infty} f_i = d$. Now $f_i \in f(S)$ implies there are points $X_i \in S$ such that $f(X_i) = f_i$. Using the standard distance function ρ , calculate, if $n \geq m$

$$\begin{aligned} \rho(X_n, X_m) &= ||X_n - X_m|| \\ &\leq ||X_n - Y|| + ||X_m - Y|| \\ &= f_m + f_n \\ &\leq 2f_m \end{aligned}$$

showing that $\{X_i\}$ is a bounded sequence. Therefore, there exists a convergent subsequence $\{X_{k_i}\}$ having a limit, say $X_0 \in M$. Thus

$$d = f(X_0) = ||X_0 - Y|| .$$

The uniqueness follows the proof by Cheney [6, p. 23].

From the above remarks, notice that strict convexity was not required to prove existence. Uniqueness may be

lost without strict convexity. For example, let $V = R^2$, $||X|| = \max\{X_1, X_2\}$, $M = \{X : X_2 = 0\}$. Let $Y = (0,1)$. The problem is then to minimize $\max_{X_1 \varepsilon R}\{X_1, 1\}$ which is minimized for all points $-1 \leq X \leq 1$.

THEOREM 1.4 [37]: *The following are properties of the metric projection mapping* $E_M = E$ *on a subspace* M *with norm* $||\cdot||$:

a) $E(\alpha X) = \alpha E(X)$, *for any scalar* α;

b) $E^2 = E$;

c) $E(X) = X$ *if and only if* $X \varepsilon M$;

d) $E(X + Y) = E(X) + Y$ *for* $X \varepsilon V$, $Y \varepsilon M$;

e) $E\big(X + E(Y)\big) = E(X) + E(Y)$ *for all* $X, Y \varepsilon V$;

f) $E\big(X - E(X)\big) = \phi$ *for all* $X \varepsilon V$.

THEOREM 1.5: *In a normed linear space* V, *let* Q *be an operator from* V *into* V *and consider the properties*:

1. $Q^2 = Q$
2. $Q(\phi) = \phi$
3. $Q(Y - X) = Q(Y) - X$ *for* $X, Y \varepsilon V$ *and* $Q(X) = X$
4. $Q(Y \pm X) = Q(Y) \pm X$ *for* $X, Y \varepsilon V$ *and* $Q(X) = X$
5. $Q(Y + \alpha X) = Q(Y) + \alpha X$ *for* $X, Y \varepsilon V$ *and* $Q(X) = X$ *and* α *any real scalar*

It follows that

a) *if properties* 1, 2, *and* 3 *hold, then each* $Y \varepsilon V$ *can be written as* $Y = X + Z$, *where* $Q(X) = X$ *and* $Q(Z) = \phi$;

b) *if properties* 1, 2, *and* 4 *hold, then* a *is true and* X *and* Z *are a unique pair;*

c) *if properties* 1, 2, *and* 5 *hold, then* b *is true and*

$$U = \left\{X \varepsilon V : Q(X) = X\right\}$$

is a subspace;

d) *if properties* 1, 2 *and* 5 *hold and*

$$W = \{Z \varepsilon V : Q(Z) = \phi\}$$

is a subspace, then Q *is a linear operator.*

Proof: Let $Y \varepsilon V$, $X = Q(Y)$, $Z = Y - X$.

a) We need only show that

$$\begin{aligned} X &= Q(Y) \\ &= Q^2(Y) \\ &= Q\big(Q(Y)\big) = Q(X) \end{aligned}$$

and

$$\begin{aligned} Q(Z) &= Q(Y - X) \\ &= Q(Y) - X = \phi \ . \end{aligned}$$

b) We need only show uniqueness so assume that $Y = X_1 + Z_1$ also, with $X_1 \varepsilon U$, $Z_1 \varepsilon W$. Then using 4. we find that

$$X = Q(Y) = Q(X_1 + Z_1) = Q(X_1) + Q(Z_1) = X_1$$

By subtraction, $Z_1 = Z$.

c) Using property 5 twice, we find that $Q(\alpha X_1 + \beta X_2) = Q(\alpha X_1 + \phi) + \beta X_2 = \alpha X_1 + \beta X_2$ for any $X_1, X_2 \varepsilon U$ and α, β scalars.

d) Let $Y_1, Y_2 \varepsilon V$, $X_1 = Q(Y_1)$, $X_2 = Q(Y_2)$, so that $Y_1 - X_1, Y_2 - X_2 \varepsilon W$ making $\alpha(Y_1 - X_1) + \beta(Y_2 - X_2) \varepsilon W$ and

$$\begin{aligned} \phi &= Q\left[\alpha(Y_1 - X_1) + \beta(Y_2 - X_2)\right] \\ &= Q(\alpha Y_1 + \beta Y_2 - \alpha X_1 - \beta X_2) \\ &= Q(\alpha Y_1 + \beta Y_2) - \alpha Q(Y_1) - \beta Q(Y_2) \end{aligned}$$

showing Q is a linear operator.

<u>DEFINITION 1.6:</u> a) *If* Q *is an operator from the normed linear space* V *into itself,* Q *is called a* <u>*projection operator*</u> *if* Q *has the properties* $Q^2 = Q$

and $Q(\phi) = \phi$. *Let* P *denote the set of projections.*

b) *If* Q *is a projection and* $Q(Y - X) = Q(Y) - X$ *for all* $X, Y \varepsilon V$ *such that* $Q(X) = X$, *then* Q *is called a* *true projection operator.* *Let* T *denote the set of true projections.*

c) *If* Q *is a projection and* $Q(Y \pm X) = Q(Y) \pm X$ *when* $Q(X) = X$, *then* Q *is called a* *unique projection operator.* *Let* U *denote the set of unique projections.*

d) *If* Q *is a projection operator and* $Q(Y + \alpha X) = Q(Y) + \alpha X$ *when* $Q(X) = X$ *and* α *is a scalar, then* Q *is called a* *spatial projection operator.* *Let* S *denote the set of spatial projections.*

Corollary 1.7 establishes some relationships between the types of projection operators and the properties of the sets they generate.

COROLLARY 1.7: *Let* V *be a normed linear space,* Q *a projection on* V , $Y \varepsilon V$, $X = Q(Y)$, $Z = Y - X$, $U = \left\{X : Q(X) = X\right\}$, *and* $W = \left\{Z : Q(Z) = \phi\right\}$. *Then*

a) *if* $Q \varepsilon T$, $Z \varepsilon W$;

b) *if* $Q \varepsilon U$ $Z \varepsilon W$ *and the pair of vectors* X,Z *is unique for any* $Y \varepsilon V$;

c) *if* $Q \varepsilon S$, $Z \varepsilon W$ *and* U *is a subspace* ;

d) Q *is a linear projection if and only if* Q, $I - Q \in S$ *making* $V = U \oplus W$;

e) $M \subset S$;

f) *letting* L *denote the set of linear projections,* $L \subset M \subset S \subset U \subset T \subset P$.

Proof: Parts a through e are reformulations of Theorem 1.5 using the notation of Definition 1.6.

f) $M \subset S \subset U \subset T \subset P$ and $L \subset S \subset U$ follow immediately from a through e. To show $L \subset M$, let $K \in L$. A norm must be found on V so that for any $Y \in V$, $||Y - KY|| \leq ||Y - KX||$ for any $X \in V$. Consider the weighted square norm

$$||Z|| = \left\{Z^T\left[K^TK + (I - K)^T(I - K)\right]Z\right\}^{1/2} .$$

This is a norm since K^TK is positive semidefinite, and if $Z \neq \phi$, then either $KZ \neq \phi$ or $(I - K)Z \neq \phi$ making either $Z^TK^TKZ > 0$ or $Z^T(I - K)^T(I - K)Z > 0$. Consequently we need only show that KY minimizes the norm for $M = \{X:KX = X\} = \{X:KY = X, Y \in V\}$. This is accomplished utilizing the following steps for $Y, X \in V$:

1. $$\begin{aligned}(Y - KY)^TK^TK(Y - KY) &= Y^T(I - K)^TK^TK(I - K)Y \\ &= Y^T(I - K)^TK^T(K - K^2)Y \\ &= 0\end{aligned}$$

2. $(Y - KY)^T(I - K)^T(I - K)(Y - KY)$

$$= Y^T(I - K)^T(I - K)^T(I - K)(I - K)Y$$

$$= Y^T(I - K)^T(I - K)Y$$

3. $(Y - KX)^T K^T K(Y - KX) \geq 0 \;=\; (Y - KY)^T K^T K(Y - KY)$

4. $(Y - KX)^T(I - K)^T(I - K)(Y - KX)$

$$= Y^T(I - K)^T(I - K)Y$$

$$- X^T K^T(I - K)^T(I - K)Y$$

$$- Y^T(I - K)^T(I - K)KX$$

$$+ X^T K^T(I - K)^T(I - K)KX$$

$$= Y^T(I - K)^T(I - K)Y$$

$$= (Y - KY)^T(I - K)^T(I - K)(Y - KY)$$

$$5.\quad ||Y - KY||^2 = (Y - KY)^T\Big(K^TK + (I - K)^T(I - K)\Big) \cdot (Y - KY)$$

$$= (Y - KY)^TK^TK(Y - KY) + (Y - KY)^T(I - K)^T(I - K)(Y - KY)$$

$$\leq (Y - KX)^TK^TK(Y - KX) + (Y - KX)^T(I - K)^T(I - K)(Y - KX)$$

$$= (Y - KX)^T\Big(K^TK + (I - K)^T(I - K)\Big) \cdot (Y - KX)$$

$$= ||Y - KX||^2$$

One set of interesting linear metric projections is defined on the range space $R(A)$ and null space $N(A)$ of a linear operator A. These projections are defined in terms of the generalized inverses of Definition 2.1 in Chapter I. Early versions of this theorem were proved by Desoer and Whalen [10] and by Ben-Israel and Charnes [1].

THEOREM 1.8 [3, p. 15]: *Let* A *and* B *be* $m \times n$ *and* $n \times m$ *matrices, respectively, with* A *mapping* V_n *into* V_m *and* B *mapping* V_m *into* V_n. *Then*

a) *if* B *is a generalized inverse of* A, *there are unique subspaces* U *and* W *such that*

$V_m = R(A) \oplus U$ *and* $V_n = N(A) \oplus W$;

b) *If* B *is a reflexive generalized inverse, then* $V_m = R(A) \oplus N(B)$, $V_n = R(A) \oplus N(B)$;

c) *if* B *is a left weak generalized inverse, then* $N(A)$ *and* $R(B)$ *are orthogonal;*

d) *if* B *is a right weak generalized inverse, then* $R(A)$ *and* $N(B)$ *are orthogonal;*

e) *if* B *is the pseudoinverse of* A , $R(A)$ *and* $N(B)$ *as well as* $N(A)$ *and* $R(B)$ *are orthogonal.*

The next theorem obtains the metric projections onto $R(A)$ and $N(A)$ when $V_n = \ell^2(n)$ and $V_m = \ell^2(m)$ in terms of the pseudoinverse A^+ of A .

THEOREM 1.9 [37]: *Let* A *be an* $m \times n$ *matrix. Then*

a) *if* $V_m = \ell^2(m)$, *the metric projection onto* $R(A)$ *is* AA^+:

b) *if* $V_n = \ell^2(n)$, *the metric projection onto* $N(A)$ *is* $(I - A^+A)$.

Theorem 1.9 shows that there are some metric projections which are also linear projections. However, all metric projections are not linear as can be seen from the example which was quoted by Newman and Odell [37]

and attributed to Charles Anderson of Southern Methodist University:

Let $V = \ell^p(3)$ with $1 < p < \infty$ and $M = \{X : X = \alpha(1,1,1)$, α a real scalar$\}$. Suppose E , the metric projection on M , is linear; then

$$E(1,0,0) = \min_{\alpha}\left(|1 - \alpha|^p + |\alpha|^p + |\alpha|^p\right)^{1/p}$$
$$= E(0,1,0) = E(0,0,1)$$

which implies

$$E(3,0,0) = 3E(1,0,0)$$
$$= E(1,0,0) + E(0,1,0)$$
$$+ E(0,0,1)$$
$$= E(1,1,1) = (1,1,1)$$

by Theorem 1.4 and the linearity of E . Therefore, the function $f(\alpha) = ||(\alpha - 3),\alpha,\alpha||^p$ is minimized uniquely for $\alpha = 1$ since the $\ell^p(3)$ norm is strictly convex for $1 < p < \infty$. Since $f(\alpha)$ is differentiable for $1 < p < \infty$, it must be true that $0 = P'(1) = -2p2^{p-2} + 2p$ or that $2^{p-2} = 1$ which is true if and only if $p = 2$. Therefore, E is linear if and only if $p = 2$. This result suggests a lemma and a theorem.

LEMMA 1.10 [37]: *Let* M *be a hyperplane contained in the normed linear vector space* V *of dimension* n, *then the metric projection* E *of* V *on* M *is a linear transformation.*

THEOREM 1.11 [37]: *Consider the spaces* $\ell^p(n)$, $1 < p < \infty$. *For every non-null subspace* M, *the metric projection* E *is linear if and only if* $n \leq 2$ *or* $p = 2$.

LEMMA 1.12 [4]: *For any metric projection* E *and any sequence* $\{X_n\}$ *such that* $\lim_{n\to\infty} X_n = Y$, *if* $z = \lim_{n\to\infty} E(X_n) \in R(E)$, *then* $Z \in E(Y)$.

Proof similar to that of Brown [4]:

$$\begin{aligned}
||Y - Z|| &= ||Y - X_n + X_n - Z_n + Z_n - Z|| \\
&\leq ||Y - X_n|| + ||X_n - Z_n|| + ||Z_n - Z|| \\
&= ||Y - X_n|| + ||X_n - E(X_n)|| \\
&\quad + ||Z_n - Z|| .
\end{aligned}$$

Now since $E(Y) \in R(E)$,

$$\begin{aligned}
||X_n - E(X_n)|| &\leq ||X_n - E(Y)|| \\
&\leq ||X_n - Y|| + ||Y - E(Y)||
\end{aligned}$$

so that

$$||Y - Z|| \leq 2||X_n - Y|| + ||Y - E(Y)|| + ||Z_n - Z||$$

and since $||X_n - Y|| > 0$, $||Z_n - Z|| > 0$ and $\lim_{n\to\infty} X_n = Y$, $\lim_{n\to\infty} Z_n = Z$, then given ε_1 , $\varepsilon_2 > 0$ there is an N such that if $n > N$,

$$||X_n - Y|| < \varepsilon_1$$

$$||Z_n - Y|| < \varepsilon_2$$

and thus

$$||Y - Z|| < 2\varepsilon_1 + ||Y - E(Y)|| + \varepsilon_2 .$$

Now since the above is true for any $\varepsilon_1,\varepsilon_2 > 0$ then

$$||Y - Z|| \leq ||Y - E(Y)||$$

which implies $Z = E(Y)$ since $Z \in R(E)$.

<u>COROLLARY 1.13</u> [4]: *For any metric projection* E *on a finite dimensional subspace* M *in a strictly convex space* V , $E(X)$ *is a continuous function of* X *on* V .

2.2 Properties of the Norm Generalized Inverse

Existence, uniqueness, and properties of the norm generalized inverse of a m×n matrix A will be developed in terms of the metric projections on the corresponding spaces V_m and V_n. The norm generalized inverse of the norm generalized inverse of A will be defined and some properties established.

THEOREM 2.1 [37]: *For each* $Y \varepsilon V_m$ *and every pair of strictly convex norms, there exists a unique best approximate solution* $X_0 \varepsilon V_n$ *of the system* $AX = Y$. *If* E *and* F *are the metric projections onto* $R(A)$ *and* $N(A)$, *respectively,* B *is the norm generalized inverse of* A, *and* A^g *is any generalized inverse of* A, *then the solution can be written symbolically as*

$$X_0 = B(Y) = (I - F)A^gE(Y)$$

Proof: We will show that X_0 satisfies Definition 1.1 of Chapter I. Let $Y \varepsilon V_m$. Now since A is a linear operator, $R(A)$ is a subspace of the strictly convex space V_m which implies there exists a metric projection E onto $R(A)$ such that $E(Y)$ is unique by applying Theorem 1.3. Let $Y_0 = E(Y) \varepsilon R(A)$. Let A^g be a generalized inverse of A and consider $X_1 = A^gY_0 \varepsilon V_n$. Observe that $AX_1 = AA^gY_0 = Y_0$, using Theorem 1.8 a),

so that X_1 is a solution to $AX = Y_0$. If we choose another generalized inverse A_2^g and let $X_2 = A_2^g Y_0$, then $AX_2 = Y_0$. Now if X_2 is such that $AX_2 = Y_0$, then $A(X_1 - X_2) = \phi_m$. Consequently the difference between any two solutions and any two generalized inverses is an element of N(A), and the set S of X's satisfying $AX = Y_0$ is characterized as

$$S = \left\{X : X = X_1 - Z, Z \varepsilon N(A)\right\} .$$

This is then the set of all points in V_n which can be best approximate solutions to $AX = Y$. We must now find that subset of S, say S', of minimum norm in V_n, or in other words, those $X_0 \varepsilon S'$ such that $||X_0||_n \leq ||X||_n$ for all $X \varepsilon S$. Using the N(A) characterization, we must find those $Z_0 \varepsilon N(A)$ such that $||X_1 - Z_0||_n \leq ||X_1 - Z||_n$ for all $Z \varepsilon N(A)$. But this is simply finding the metric projection $F(X_1)$ on N(A), which yields a unique Z_0 since V_n is strictly convex and N(A) is a subspace by Theorem 1.3. Therefore, the best approximate solution is

$$
\begin{aligned}
X_0 &= X_1 - Z_0 \\
&= X_1 - F(X_1) \\
&= (I - F)X_1 \\
&= (I - F)A^g Y_0 \\
&= (I - F)A^g E(Y) \\
&= B(Y) \quad .
\end{aligned}
$$

COROLLARY 2.2 [37]: *In the notation of Theorem* 2.1 *the following properties hold:*

a) $AF = \underline{0}$, *the zero or null operator*

b) $EA = A$

c) $BE = B$

d) $AA^gE = E$

e) $AB = E$

f) $BA = (I - F)A^gA$

g) $ABA = A$

h) $BAB = B$

<u>COROLLARY 2.3</u> [37]:

a) *If* $V_n = \ell^2(n)$, *then* $B = A^+E$;

b) *If* $V_m = \ell^2(m)$, *then* $B = (I - F)A^+$;

c) *If* $V_n = \ell^2(n)$ *and* $V_m = \ell^2(m)$, *then* $B = A^+$;

d) *If the rank of* A *is* $n \leq m$, *then* $F = \underline{0}$ *and* $B = A^gE$.

The result of part c was first shown by Penrose [39]. The dependence of the linearity of the norm generalized inverse B upon the linearity of the metric projections E and F will be shown next.

<u>THEOREM 2.4</u>: *The norm generalized inverse* B *of an* m×n *matrix* A *is linear if and only if the metric projections* F *and* E *are linear over* V_n *and* V_m , *respectively* .

Proof: If F and E are linear, then B is linear since $B = (I - F)A^gE$ and A^g is a matrix and therefore linear.

If B is linear, then for α,β scalars and $Y_1, Y_2 \in V_m$,

$$B(\alpha Y_1 + \beta Y_2) = \alpha Y_1 + \beta Y_2$$

which becomes

$$A^g\left[E(\alpha Y_1 + \beta Y_2) - \alpha E(Y_1) - \beta E(Y_2)\right]$$
$$= FA^gE(\alpha Y_1 + \beta Y_2) - \alpha FA^gE(Y_1)$$
$$- \beta FA^gE(Y_2) \qquad (2.1)$$

after substitution and some algebraic manipulation. Observing that $E(\alpha Y_1 + \beta Y_2) - \alpha E(Y_1) - \beta E(Y_2) \varepsilon R(A)$ and that $AF = \underline{0}$ by Corollary 2.2 a), we find that multiplication of (2.1) by A produces

$$E(\alpha Y_1 + \beta Y_2) - \alpha E(Y_1) - \beta E(Y_2) = \phi_m$$

showing that E is linear. Including this result in equation (2.1), we obtain

$$\phi_n = F\left[A^gE(\alpha Y_1 + \beta Y_2)\right] - \alpha FA^gE(Y_1)$$
$$- \beta FA^gE(Y_2)$$

showing that F is linear over $R(A^g)$, for all generalized inverses of A by Theorem 2.1. Now consider the set of matrices defined by

$$H(Z) = A^+ + (I - A^+A)Z$$

where Z is an arbitrary $n \times m$ matrix with the restriction that Z maps $R(A)$ into $N(A)$. Now $AH(Z)A = A$ showing $H(Z)$ is a generalized inverse of A so that F is linear over $R[H(Z)]$ for all Z. Since Z is an arbitrary map from $R(A)$ into $N(A)$, then for each $Y_0 \varepsilon R(A)$ and $X_0 \varepsilon N(A)$ there is some Z so that

$$H(Z)Y_0 = A^+Y_0 + X_0 .$$

Allowing Y_0 to vary over $R(A)$, we see that F is linear over $R(A^+) \oplus N(A) = V_n$.

COROLLARY 2.5 [37]: *Let* $V_n = \ell^q(n)$, $V_m = \ell^p(m)$ *for* $m \geq 3$, $n \geq 3$. *Then* B *is linear for every* $m \times n$ *matrix* A *if and only if* $p = q = 2$.

COROLLARY 2.6 [37]: *If* A *is an* $(n+1) \times n$ *matrix of rank* n, *then the norm generalized inverse* B *is linear.*

One property of the p-q generalized inverse when $p = q = 2$ (the 2-2g.i.) is that the q-p g.i. of the p-q g.i. is $(A^+)^+ = A$, a property of the symmetry of A and A^+. An interesting problem is to define a norm g.i. C of a norm g.i. B of A and determine if $C = A$. In general, as was seen in Theorem 2.4 and Corollary 2.5, B is nonlinear.

Observe that in the fundamental existence theorem, Theorem 2.1, the linearity of A was required to make $N(A)$ and $R(A)$ finite dimensional subspaces. That $N(A)$ and $R(A)$ be subspaces was required to associate $N(A)$ and $R(A)$ with the finite-dimensional subspace M of Theorem 1.3. By the remark after Theorem 1.3, observe that the only properties of M required are that M be closed, complete, and convex (convex for uniqueness). A more general theorem is therefore

> *For each* $Y \varepsilon V_m$ *and every pair of strictly convex norms, there exists a best approximate* $X_0 \varepsilon V_n$ *of* $A(X) = Y$, *if* $R(A)$ *and* $N(A)$ *are closed and complete subsets of* V_m *and* V_n *respectively. If* $R(A)$ *and* $N(A)$ *are also convex,* X_0 *is unique.*

Consequently, in order to answer the question posed in this remark, it must be determined whether $R(B)$ and $N(B)$ each satisfy the hypotheses of the theorem.

<u>LEMMA 2.6</u>: *For any norm generalized inverse* B *of* A , $R(B) = N(F)$.

Proof: Let $Y \varepsilon V_m$ and $B = (I - F)A^gE$. Then $E(Y) \varepsilon R(A)$ so that $E(V_m) = R(A)$, since $E(Y) = Y$ if $Y \varepsilon R(A)$. Therefore, letting $Z = A^gE(Y) - A^+E(Y)$ and noticing again that $E(Y) \varepsilon R(A)$,

$$AZ = AA^gE(Y) - AA^+E(Y)$$
$$= E(Y) - E(Y)$$
$$= \phi_m$$

since AA^g and AA^+ are both projections onto $R(A)$ by Theorem 1.8. Consequently,

$$\begin{aligned}(I - F)A^gE(Y) &= A^+E(Y) + Z - F\left(A^+E(Y) + Z\right)\\ &= A^+E(Y) + Z - F\left(A^+E(Y)\right) - Z\\ &= (I - F)A^+E(Y) \qquad (2.2)\end{aligned}$$

so that

$$\begin{aligned}R(B) &= R\left((I - F)A^+E\right)\\ &= (I - F)R(A^+E)\\ &= (I - F)R(A^+) \quad .\end{aligned}$$

To show $(I - F)R(A^+) = N(F)$, first let $X \varepsilon R(B)$. Then there is a $Z \varepsilon R(A^+)$ such that $X = (I - F)Z$ and

$$\begin{aligned}F(X) &= F[(I - F)(Z)]\\ &= (F - F^2)(z)\\ &= \phi_n\end{aligned}$$

and therefore $X \varepsilon N(F)$ and $N(B) \subset N(F)$.

Suppose $X \varepsilon N(F)$. Then $F(X) = \phi_n$ and $(I - F)(X) = X$. To show $X \varepsilon R(B)$, it need only be shown that $X \varepsilon R(A^+)$. Since $X \varepsilon V_n$, X can be written as $X = X_1 + X_2$ where $X_1 \varepsilon N(A) = R(F)$ and $X_2 \varepsilon R(A^+)$, by Theorem 1.8 e). If $X = \phi_n$, then $X \varepsilon R(A^+)$ and the lemma is complete. If $X \neq \phi_n$, then $X \not\varepsilon R(F)$, for that would imply $X = F(X) = \phi_n$. Therefore, in either case, $X_1 = \phi_n$ implying that $X = X_2 \varepsilon R(A^+)$ and that $N(F) = R(B)$.

<u>LEMMA 2.7</u>: *For any norm generalized inverse* B *of* A , $N(B) = N(E)$.

Proof: Now $N(B) = \left\{Z : (I - F)A^gE(Z) = \phi_m\right\}$. Since A^g is linear and $F(\phi_n) = \phi_n$, then $N(E) \subset N(B)$. Let $Z \varepsilon N(B)$. Since $Z \varepsilon V_m$ and E is a metric-projection, Z can be written as $Z = Z_0 + Z_1$, where $E(Z_0) = \phi_m$, $E(Z_1) = Z_1$, by Corollary 1.7 c).

Thus

$$\begin{aligned}
\phi_n &= B(Z) \\
&= (I - F)A^gE(Z_0 + Z_1) \\
&= (I - F)A^g(Z_1) \\
&= (I - F)A^+(Z_1)
\end{aligned}$$

by the properties of E and Equation (2.2). Now $(I - F)A^+(Z_1) = \phi_n$, or $FA^+(Z_1) = A^+Z_1$ is true if and only if $A^+Z_1 \in N(A)$ is true since F is a projection on $N(A)$. But $A^+Z_1 \in R(A^+)$ so that $A^+Z_1 = \phi_n$ since $V_n = N(A) \oplus R(A^+)$ by Theorem 1.8 e).

Since E is a projection on $R(A)$ and $E(Z_1) = Z_1$, then $Z_1 \in R(A)$ so that $Z_1 = AA^+Z_1 = A\phi_n = \phi_m$. Thus $Z = Z_0 + Z_1 = Z_0 \in N(E)$. Therefore, $N(B) = N(E)$.

THEOREM 2.8: *For any norm generalized inverse* B *of* A, $N(B)$ *and* $R(B)$ *are closed sets.*

Proof: Observe that

$$\begin{aligned} N(B) &= N(E) \\ &= \{Z : E(Z) = \phi_m\} \end{aligned}$$

from Lemma 2.7 and

$$\begin{aligned} R(B) &= N(F) \\ &= \{Z : F(Z) = \phi_n\} \end{aligned}$$

from Lemma 2.6. Now E and F are both continuous by Corollary 1.13. Since the inverse image of a closed set is closed when the function is continuous, $N(B)$

and R(B) are closed since $E^{-1}\left(\left\{\phi_m\right\}\right) = N(B)$ and $F^{-1}\left(\left\{\phi_n\right\}\right) = N(B)$, where $^{-1}$ denotes an inverse opertor.

These results can be summarized in

<u>THEOREM 2.9</u>: *The norm generalized inverse* C *of the norm generalized inverse* B *of a matrix* A *always exists. The solution set* S *is a unique point if* $p = q = 2$ *or* $n,m \leq 2$ *or if* A *is an* $(n+1)\times n$ *matrix of rank* n, *(i.e., if* B *is linear).*

Proof: By the generalized existence theorem a best approximate solution Y_0 to the equation

$$B(X) = Y$$

exists for each $X \varepsilon V_n$ if R(B) and N(B) are closed and complete subsets of V_n and V_m, respectively. By Theorem 2.8, R(B) and N(B) are closed sets showing the existence of the norm generalized inverse C of B defined by $C(X) = S_X$ where S_X is the set of best approximate solutions Y_0 to $B(X) = Y_0$, for every $X \varepsilon V_n$.

By the generalized existence theorem, S_X is a unique point for every X if R(B) and N(B) are convex, which is true, if B is linear, which is true for all

A if and only if $p = q = 2$ (when $m,n \geq 3$) or $n,m \leq 2$ by Corollary 2.5 or A is an $(n+1)\times n$ matrix of rank n by Corollary 2.6.

For the norm generalized inverse C of the norm generalized inverse B of an $m\times n$ matrix A to equal A, then C must be expressible as a matrix and, as such, must have a unique image for each $X \in V_n$. This is the case if and only if B is linear.

CHAPTER III

CALCULATION OF THE p-q GENERALIZED INVERSE FOR ℓ^p AND ℓ^q SPACES

3.1 Preliminary Results

Prior to defining and proving an algorithm for the calculation of the p-q generalized inverse, which is defined in Section 1.1, several preliminary results are necessary.

THEOREM 1.1: *Let* V_{11}, V_{12}, V_{21}, *and* V_{22} *be real* n×n, n×m, m×n, *and* m×m *matrices, respectively. Define*

1. $R_{11} = \left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]^{+}$
2. $R_{12} = -R_{11}V_{12}V_{22}^{+}$
3. $R_{21} = -V_{22}^{+}V_{21}R_{11}$
4. $R_{22} = V_{22}^{+} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}$
5. $R_{11}^{*} = V_{11}^{+} + V_{11}^{+}V_{12}R_{22}^{*}V_{21}V_{11}^{+}$
6. $R_{12}^{*} = -V_{11}^{+}V_{12}R_{22}^{*}$
7. $R_{21}^{*} = -R_{22}^{*}V_{21}V_{11}^{+}$
8. $R_{22}^{*} = \left[V_{22} - V_{21}V_{11}^{+}V_{12}\right]^{+}$

and let

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}; \quad R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}; \quad R^* = \begin{pmatrix} R^*_{11} & R^*_{12} \\ R^*_{21} & R^*_{22} \end{pmatrix} .$$

Then $R = V^+$ *if and only if*

1'. $V^+_{22}V_{21}\left[I - R_{11}R^+_{11}\right] = \left(R_{11}V_{12}\left[I - V^+_{22}V_{22}\right]\right)^T$

2'. $\left[I - V_{22}V^+_{22}\right]V_{21}R_{11} = \left(\left[I - R^+_{11}R_{11}\right]V_{12}V^+_{22}\right)^T$

3'. $V_{21}\left[I - R_{11}R^+_{11}\right] = \phi$

4'. $\left[I - R^+_{11}R_{11}\right]V_{12} = \phi$

and $R^* = V^+$ *if and only if*

5'. $V^+_{11}V_{12}\left[I - R^*_{22}R^{*+}_{22}\right] = \left(R^*_{22}V_{21}\left[I - V^+_{11}V_{11}\right]\right)^T$

6'. $\left[I - V_{11}V^+_{11}\right]V_{12}R^*_{22} = \left(\left[I - R^{*+}_{22}R^*_{22}\right]R_{21}V^+_{11}\right)^T$

7'. $V_{12}\left[I - R^*_{22}R^{*+}_{22}\right] = \phi$

8'. $\left[I - R^{*+}_{22}R^*_{22}\right]V_{21} = \phi$

which then implies

$$1''.\quad \left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]^{+} = V_{11}^{+} + V_{11}^{+}V_{12}\left[V_{22} - V_{21}V_{11}^{+}V_{12}\right]^{+}V_{21}V_{11}^{+}$$

$$2''.\quad \left[V_{22} - V_{21}V_{11}^{+}V_{12}\right]^{+} = V_{22}^{+} + V_{22}^{+}V_{21}\left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]^{+}V_{12}V_{22}^{+}$$

$$3''.\quad \left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]^{+}V_{12}V_{22}^{+} = V_{11}^{+}V_{12}\left[V_{22} - V_{21}V_{11}^{+}V_{12}\right]^{+}$$

$$4''.\quad V_{22}^{+}V_{21}\left[V_{11} - V_{12}V_{22}^{+}V_{21}\right] = \left[V_{22} - V_{21}V_{11}^{+}V_{12}\right]^{+}V_{21}V_{11}^{+} \quad .$$

Proof: If all eight conditions hold, then $R = R^*$ by the uniqueness of V^+, giving the equalities 1", 2", 3", and 4" by equating corresponding submatrices of R and R^*. Thus, it need only be shown that

a) $R = V^+$ if and only if 1', 2', 3', and 4' hold.

b) $R^* = V^+$ if and only if 5', 6', 7' and 8' hold.

But the only difference between the definitions of R and R^* and between the conditions 1' through 8' is an interchange of the symbols 1 and 2 so that using a symmetrical argument we need only prove statement a) above.

To prove a), it will be shown that the four Moore-Penrose equations of Definition 2.1e in Chapter I

1. $VRV = V$

2. $RVR = R$

3. $(VR)^T = VR$

4. $(RV)^T = RV$

are satisfied. Therefore, consider

1.

$$S = VRV = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} .$$

Now

$$\begin{aligned}
S_{11} &= V_{11}R_{11}V_{11} + V_{12}R_{21}V_{11} + V_{11}R_{12}V_{21} + V_{12}R_{22}V_{21} \\
&= V_{11}R_{11}V_{11} - V_{12}V_{22}^{+}V_{21}R_{11}V_{11} - V_{11}R_{11}V_{12}V_{22}^{+}V_{21} \\
&\quad + V_{12}\left[V_{22}^{+} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}\right]V_{21} \\
&= V_{11}R_{11}R_{11}^{+} - V_{12}V_{22}^{+}V_{21}R_{11}\left[V_{11} - V_{12}V_{22}^{+}V_{21}\right] \\
&\quad + V_{12}V_{22}^{+}V_{21} \\
&= \left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]R_{11}R_{11}^{+} + V_{12}V_{22}^{+}V_{21} \\
&= R_{11}^{+} + V_{12}V_{22}^{+}V_{21} \\
&= V_{11}
\end{aligned}$$

$$S_{21} = V_{21}R_{11}V_{11} + V_{22}R_{21}V_{11} + V_{21}R_{12}V_{21} + V_{22}R_{22}V_{21}$$

$$= V_{21}R_{11}V_{11} - V_{22}V_{22}^{+}V_{21}R_{11}V_{11} - V_{21}R_{11}V_{12}V_{22}^{+}V_{21}$$

$$+ V_{22}\left[V_{22}^{+} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}\right]V_{21}$$

$$= V_{21}R_{11}\left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]$$

$$+ V_{22}V_{22}^{+}V_{21}\left[I - R_{11}V_{11} + R_{11}V_{12}V_{22}^{+}V_{21}\right]$$

$$= V_{21} - V_{21} + V_{21}R_{11}R_{11}^{+} + V_{22}V_{22}^{+}V_{21}\left[I - R_{11}R_{11}^{+}\right]$$

$$= V_{21} - V_{21}\left[I - R_{11}R_{11}^{+}\right] + V_{22}V_{22}^{+}V_{21}\left[I - R_{11}R_{11}^{+}\right]$$

$$= V_{21} - \left[I - V_{22}V_{22}^{+}\right]V_{21}\left[I - R_{11}R_{11}^{+}\right]$$

and therefore $S_{21} = V_{21}$ if and only if
$\left[I - V_{22}V_{22}^{+}\right]V_{21}\left[I - R_{11}R_{11}^{+}\right] = \phi$ (i.e., that
$-V_{21}\left[I - R_{11}R_{11}^{+}\right] + V_{22}V_{22}^{+}V_{21}\left[I - R_{11}R_{11}^{+}\right] = \phi$).
If conditions 1' and 3' are used, observe that

$$-V_{21}\left[I - R_{11}R_{11}^{+}\right] + V_{22}V_{22}^{+}V_{21}\left[I - R_{11}R_{11}^{+}\right]$$

$$= \phi + V_{22}\left[I - V_{22}^{+}V_{22}\right]V_{12}^{T}R_{11}^{T} = \phi$$

so that conditions 1' and 3' imply $S_{21} = V_{21}$

$$S_{12} = V_{11}R_{11}V_{12} + V_{12}R_{21}V_{12} + V_{11}R_{12}V_{22} + V_{12}R_{22}V_{22}$$

$$= V_{11}R_{11}V_{12} - V_{12}V_{22}^{+}V_{21}R_{11}V_{12} - V_{11}R_{11}V_{12}V_{22}^{+}V_{22}$$

$$+ V_{12}\left[V_{22}^{+} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}\right]V_{22}$$

$$= \left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]R_{11}V_{12}$$

$$+ \left[I - \left\{V_{11} - V_{12}V_{22}^{+}V_{21}\right\}R_{11}\right]V_{12}V_{22}^{+}V_{22}$$

$$= V_{12} - \left[I - R_{11}^{+}R_{11}\right]V_{12} + \left[I - R_{11}^{+}R_{11}\right]V_{12}V_{22}^{+}V_{22}$$

$$= V_{12} - \left[I - R_{11}^{+}R_{11}\right]V_{12}\left[I - V_{22}^{+}V_{22}\right]$$

and therefore $S_{12} = V_{12}$ if and only if

$\left[I - R_{11}^{+}R_{11}\right]V_{12}\left[I - V_{22}^{+}V_{22}\right] = \phi$.

Using conditions 2' and 4',

$$-\left[I - R_{11}^{+}R_{11}\right]V_{12} + \left[I - R_{11}^{+}R_{11}\right]V_{12}V_{22}^{+}V_{22}$$

$$= \phi + R_{11}^{T}V_{21}^{T}\left[I - V_{22}V_{22}^{+}\right]V_{22} = \phi$$

which implies $S_{12} = V_{12}$.

$$
\begin{aligned}
S_{22} &= V_{21}R_{11}V_{12} + V_{22}R_{21}V_{12} + V_{21}R_{12}V_{22} + V_{22}R_{22}V_{22} \\
&= V_{21}R_{11}V_{12} - V_{22}V_{22}^{+}V_{21}R_{11}V_{12} - V_{21}R_{11}V_{12}V_{22}^{+}V_{22} \\
&\quad + V_{22}\left[V_{22}^{+} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}\right]V_{22} \\
&= V_{22} + \left[I - V_{22}V_{22}^{+}\right]V_{21}R_{11}V_{12} \\
&\quad - \left[I - V_{22}V_{22}^{+}\right]V_{21}R_{11}V_{12}V_{22}^{+}V_{22} \\
&= V_{22} + \left[I - V_{22}V_{22}^{+}\right]V_{21}R_{11}V_{12}\left[I - V_{22}^{+}V_{22}\right] \\
&= V_{22} + \left[I - V_{22}V_{22}^{+}\right]V_{21}R_{11}R_{11}^{+}R_{11}V_{12}\left[I - V_{22}^{+}V_{22}\right]
\end{aligned}
$$

and therefore $S_{22} = V_{22}$ if and only if

$$\left[I - V_{22}V_{22}^{+}\right]V_{21}R_{11}R_{11}^{+}R_{11}V_{12}\left[I - V_{22}^{+}V_{22}\right] = \phi \ .$$

If conditions 1' and 2' are used, then

$$
\begin{aligned}
&\left[I - V_{22}V_{22}^{+}\right]V_{21}R_{11}R_{11}^{+}R_{11}V_{12}\left[I - V_{22}^{+}V_{22}\right] \\
&\qquad = V_{22}^{+T}V_{12}^{T}\left[I - R_{11}^{+}R_{11}\right]R_{11}^{+}\left[I - R_{11}R_{11}^{+}\right]V_{21}^{T}V_{22}^{+T} \\
&\qquad = \phi \ .
\end{aligned}
$$

2.

$$S = RVR = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} .$$

Now

$$\begin{aligned} S_{11} &= R_{11}V_{11}R_{11} + R_{12}V_{21}R_{11} + R_{11}V_{12}R_{21} + R_{12}V_{22}R_{21} \\ &= R_{11}V_{11}R_{11} - R_{11}V_{12}V_{22}^{+}V_{21}R_{11} - R_{11}V_{12}V_{22}^{+}V_{21}R_{11} \\ &\quad + R_{11}V_{12}V_{22}^{+}V_{22}V_{22}^{+}V_{21}R_{11} \\ &= R_{11}\left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]R_{11} \\ &= R_{11} \end{aligned}$$

$$\begin{aligned} S_{21} &= R_{21}V_{11}R_{11} + R_{22}V_{21}R_{11} + R_{21}V_{12}R_{21} + R_{22}V_{22}R_{21} \\ &= -V_{22}^{+}V_{21}R_{11}V_{11}R_{11} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}V_{21}R_{11} \\ &\quad + \left[V_{22} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}\right]\left[V_{21}R_{11} - V_{22}V_{22}^{+}V_{21}R_{11}\right] \\ &= -V_{22}^{+}V_{21}R_{11}\left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]R_{11} \\ &\quad + \left[I + V_{22}^{+}V_{21}R_{11}V_{12}\right]V_{22}^{+}\left[I - V_{22}V_{22}^{+}\right]V_{21}R_{11} \\ &= R_{21} \end{aligned}$$

$$
\begin{aligned}
S_{12} &= R_{11}V_{11}R_{12} + R_{12}V_{21}R_{12} + R_{11}V_{12}R_{22} + R_{12}V_{22}R_{22} \\
&= -R_{11}V_{11}R_{11}V_{12}V_{22}^{+} + R_{11}V_{12}V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+} \\
&\quad + \left[R_{11}V_{12} - R_{11}V_{12}V_{22}^{+}V_{22}\right]\left[V_{22}^{+} + V_{22}V_{21}R_{11}V_{12}V_{22}^{+}\right] \\
&= -R_{11}\left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]R_{11}V_{12}V_{22}^{+} \\
&\quad + R_{11}V_{12}\left[I - V_{22}^{+}V_{22}\right]V_{22}^{+}\left[I - V_{21}R_{11}V_{12}V_{22}^{+}\right] \\
&= R_{12}
\end{aligned}
$$

$$
\begin{aligned}
S_{22} &= R_{21}V_{11}R_{12} + R_{22}V_{21}R_{12} + R_{21}V_{12}R_{22} + R_{22}V_{22}R_{22} \\
&= V_{22}^{+}V_{21}R_{11}V_{11}R_{11}V_{12}V_{22}^{+} \\
&\quad - \left[V_{22}^{+} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}\right]V_{21}R_{11}V_{12}V_{22}^{+} \\
&\quad + \left\{\left[V_{22}^{+} + V_{22}V_{21}R_{11}V_{12}V_{22}\right]V_{22} - V_{22}^{+}V_{21}R_{11}V_{12}\right\} \\
&\quad \cdot V_{22}^{+}\left[I + V_{21}R_{11}V_{12}V_{22}^{+}\right] \\
&= V_{22}^{+}V_{21}R_{11}\left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]R_{11}V_{12}V_{22}^{+} + V_{22}^{+} \\
&= V_{22}^{+} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+} \\
&= R_{22}
\end{aligned}
$$

3.

$$S = VR \quad .$$

Now

$$S_{11} = V_{11}R_{11} + V_{12}R_{21}$$

$$= V_{11}R_{11} - V_{12}V_{22}^{+}V_{21}R_{11}$$

$$= \left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]R_{11}$$

$$= R_{11}^{+}R_{11}$$

$$= S_{11}^{T}$$

$$S_{21} = V_{21}R_{11} + V_{22}R_{21}$$

$$= V_{21}R_{11} - V_{22}V_{22}^{+}V_{21}R_{11}$$

$$= \left(I - V_{22}V_{22}^{+}\right)V_{21}R_{11}$$

$$S_{12} = V_{11}R_{12} + V_{12}R_{22}$$

$$= -V_{11}R_{11}V_{12}V_{22}^{+} + V_{12}V_{22}^{+} + V_{12}V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}$$

$$= V_{12}V_{22}^{+} - \left[V_{11} - V_{12}V_{22}^{+}V_{21}\right]R_{11}V_{12}V_{22}^{+}$$

$$= \left[I - R_{11}^{+}R_{11}\right]V_{12}V_{22}^{+} \quad .$$

Consequently $S_{12} = S_{21}^T$ if and only if

$$\left[I - V_{22}V_{22}^+\right]V_{21}R_{11} = \left(\left[I - R_{11}^+R_{11}\right]V_{12}V_{22}^+\right)^T$$

which is condition 2'.

$$\begin{aligned} S_{22} &= V_{21}R_{12} + V_{22}R_{22} \\ &= -V_{21}R_{11}V_{12}V_{22}^+ + V_{22}\left[V_{22}^+ + V_{22}^+V_{21}R_{11}V_{12}V_{22}^+\right] \\ &= V_{22}V_{22}^+ - \left[I - V_{22}V_{22}^+\right]V_{21}R_{11}V_{12}V_{22}^+ \end{aligned}$$

and therefore $S_{22} = S_{22}^T$ if and only if

$$\left(\left[I - V_{22}V_{22}^+\right]V_{21}R_{11}V_{12}V_{22}^+\right)^T = \left[I - V_{22}V_{22}^+\right]V_{21}R_{11}V_{12}V_{22}^+ .$$

Given conditions 2' and 4',

$$\begin{aligned} \left[I - V_{22}V_{22}^+\right]V_{21}R_{11}V_{12}V_{22}^+ &= \left(\left[I - R_{11}^+R_{11}\right]V_{12}V_{22}^+\right)^T V_{12}V_{22}^+ \\ &= \left(\phi \times V_{22}^+\right)^T V_{12}V_{22}^+ \\ &= \phi \end{aligned}$$

which implies $S_{22} = S_{22}^T$.

4.

$$S = RV \quad .$$

Now

$$\begin{aligned} S_{11} &= R_{11}V_{11} + R_{12}V_{21} \\ &= R_{11}V_{11} + R_{11}V_{12}V_{22}^{+}V_{21} \\ &= R_{11}R_{11}^{+} \\ &= S_{11}^{T} \end{aligned}$$

$$\begin{aligned} S_{21} &= R_{21}V_{11} + R_{22}V_{21} \\ &= -V_{22}^{+}V_{21}R_{11}V_{11} + \left[V_{22}^{+} + V_{22}^{+}V_{21}R_{11}V_{12}V_{22}^{+}\right]V_{21} \\ &= V_{22}^{+}V_{21} - V_{22}^{+}V_{21}R_{11}\left[V_{11} - V_{12}V_{22}^{+}V_{21}\right] \\ &= V_{22}^{+}V_{21}\left[I - R_{11}R_{11}^{+}\right] \end{aligned}$$

$$\begin{aligned} S_{12} &= R_{11}V_{12} + R_{12}V_{22} \\ &= R_{11}V_{12} - R_{11}V_{12}V_{22}^{+}V_{22} \\ &= R_{11}V_{12}\left[I - V_{22}^{+}V_{22}\right] \end{aligned}$$

and therefore $S_{12} = S_{21}^T$ if and only if

$$V_{22}^+ V_{21}\left[I - R_{11}R_{11}^+\right] = \left(R_{11}V_{12}\left[I - V_{22}^+ V_{22}\right]\right)^T$$

which is condition 1'.

$$\begin{aligned} S_{22} &= R_{21}V_{12} + R_{22}V_{22} \\ &= -V_{22}^+ V_{21}R_{11}V_{12} + \left[V_{22}^+ + V_{22}^+ V_{21}R_{11}V_{12}V_{22}^+\right]V_{22} \\ &= V_{22}^+ V_{22} - V_{22}^+ V_{21}R_{11}V_{12}\left[I - V_{22}^+ V_{22}\right] \end{aligned}$$

and therefore $S_{22} = S_{22}^T$ if and only if

$$\left(V_{22}^+ V_{21}R_{11}V_{12}\left[I - V_{22}^+ V_{22}\right]\right)^T = V_{22}^+ V_{21}R_{11}V_{12}\left[I - V_{22}^+ V_{22}\right] .$$

Given conditions 1' and 4',

$$\begin{aligned} V_{22}^+ V_{21}R_{11}V_{12}\left[I - V_{22}^+ V_{22}\right] &= V_{22}^+ V_{21}\left(V_{22}^+ V_{21}\left[I - R_{11}R_{11}^+\right]\right)^T \\ &= V_{22}^+ V_{21}\left(V_{22}^+ \times \phi\right)^T \\ &= \phi \end{aligned}$$

which implies $S_{22} = S_{22}^T$.

In each of the above calculations, it has been shown that conditions 1', 2', 3', and 4' imply that $R = V^+$. Observe also that $R = V^+$ explicitly implied condition 1' in Penrose equation 4, and condition 2' in Penrose equation 3. To show the implication of 3', first premultiply 1' by V_{22} to obtain

$$V_{22}V_{22}^{+}V_{21}\left[I - R_{11}R_{11}^{+}\right] = V_{22}\left[I - V_{22}^{+}V_{22}\right]V_{12}^{T}R_{11}^{T} = \phi .$$

This shows that in Penrose equation 1, $S_{21} = V_{21}$ if and only if

$$-V_{21}\left[I - R_{11}R_{11}^{+}\right] + V_{22}V_{22}^{+}V_{21}\left[I - R_{11}R_{11}^{+}\right] = \phi$$

which is

$$-V_{22}\left[I - R_{11}R_{11}^{+}\right] = \phi$$

or condition 3'. All the conditions implied by $R = V^+$ are satisfied simultaneously (note that $R = V^+$ implies that condition 1' does not require condition 3' to be satisfied). To show the implication of 4', postmultiply the transpose by V_{22} to obtain

$$\left[I - R_{11}^{+}R_{11}\right]V_{12}V_{22}^{+}V_{22} = R_{11}^{T}V_{21}^{T}\left[I - V_{22}V_{22}^{+}\right]V_{22} = \phi$$

so that $S_{12} = V_{12}$ in Penrose equation 1 if and only if

$$\phi = \left[I - R_{11}^{+}R_{11}\right]V_{12} + \left[I - R_{11}^{+}R_{11}\right]V_{12}V_{22}^{+}V_{22}$$

$$= \left[I - R_{11}^{+}R_{11}\right]V_{12}$$

which is condition 4'.

<u>DEFINITION 1.2</u> [22]: *A real* $n \times n$ *matrix* A *is called* EPr *if and only if it satisfies the conditions:*

1. A *has rank* r .
2. $\sum_{i=1}^{n} X_i A_i = 0$ *if and only if* $\sum_{i=1}^{n} X_i A^i = 0$

for all real X_i *where* A^i *is the ith row and* A_i *is the ith column of* A .

Condition 2 can be written in matrix notation as

$$AX = \phi \text{ if and only if } A^T X = \phi$$

or that

$$\left\{X : AX = \phi\right\} = \left\{X : A^T X = \phi\right\}$$

which is $N(A) = N(A^T)$. If the rank is understood, an EPr matrix will be referred to as an EP matrix.

LEMMA 1.3: *If* V_1, V_2, *and* V_3 *are subspaces of the n-dimensional real space* V_n *such that* $V_n = V_1 \oplus V_2 = V_1 \oplus V_3$, *and if* V_1 *is orthogonal to both* V_2 *and* V_3 , *then* $V_2 = V_3$.

Proof: Suppose that there is an $X \in V_2$. Since $X \in V_n$, and $V_n = V_1 \oplus V_3$ there are vectors $v_1 \in V_1$, $v_3 \in V_3$ such that $X = v_1 + v_3$. Let $\left\{b_{1i}\right\}_{i=1}^{r}$ be an orthogonal basis for V_1 and $\left\{b_{3i}\right\}_{i=1}^{n-r}$ be an orthogonal basis for V_3 . Then by the orthogonality hypothesis, $b_{1i}^T b_{3j} = 0$ for $i = 1,\cdots,r$; $j = 1,\cdots,n-r$ [17, p. 24, p. 34]. Therefore, there are constants $\left\{\alpha_{1i}\right\}_{i=1}^{r}$ and $\left\{\alpha_{3i}\right\}_{i=1}^{n-r}$ such that

$$X = \sum_{i=1}^{r} \alpha_{1i} b_{1i} + \sum_{i=1}^{n-r} \alpha_{3i} b_{3i} .$$

Since $X \in V_2$ and $V_n = V_1 \oplus V_2$, $X^T v_1 = 0$ for all $v_1 \in V_1$ so that for $j = 1,\cdots,r$,

$$\begin{aligned} 0 &= b_{1j}^T X \\ &= \sum_{i=1}^{r} \alpha_{1i} b_{1j}^T b_{1i} + \sum_{i=1}^{n-r} \alpha_{3i} b_{1j}^T b_{3i} \\ &= \alpha_{1j} \end{aligned}$$

which implies

$$X = \sum_{i=1}^{n-r} \alpha_{3i} b_{3i} \in V_3$$

or that $V_2 \subset V_3$. By exchanging V_2 for V_3 and basis b_{2i} for b_{3i} in the above proof, $V_3 \subset V_2$ which implies $V_2 = V_3$.

THEOREM 1.4: *A real* n×n *matrix* A *is* EPr *if and only if* A^+ *is* EPr. *Further,* A *is* EPr *if and only if* $R(A) = R(A^T)$.

Proof: Note that

$$\begin{aligned}
N(A^+) &= \{X : A^+X = 0\} \\
&= \{X : (I - AA^+)Z = X \text{ for some } Z\} \\
&= \{X : (I - AA^+)(I - AA^+)Z = X, \text{ for some } Z\} \\
&= \{X : (I - AA^+)X = X\} \\
&= \{X : AA^+X = \phi\} \\
&= \{X : (AA^+)^T X = \phi\} \\
&= \{X : A^{+T}A^T X = \phi\} \\
&= \{X : (I - A^{+T}A^T)(I - A^{+T}A^T)Z = X, \text{ for all } Z\} \\
&= \{X : A^T X = 0\} \\
&= N(A^T) .
\end{aligned}$$

Similarly by replacing A^+ with A^{+T} and A with A^T, then $N(A^{+T}) = N(A)$. Therefore, $N(A) = N(A^T)$ if and only if $N(A^{+T}) = N(A^+)$, which is to say, A is EPr if and only if A^+ is EPr.

For the second part, note that V_n, the real n-dimensional vector space, is the direct sum of

$$V_n = N(A^+) \oplus R(A)$$
$$= N(A^{+T}) \oplus R(A^T)$$

by Theorem 2.8e in Chapter II. Now A is EPr if and only if A^+ is EPr if and only if $N(A^+) = N(A^{+T})$ which implies $R(A) = R(A^T)$ by Lemma 1.3 since $R(A)$ is orthogonal to $N(A^+)$ and $R(A^T)$ is orthogonal to $N(A^{+T})$. Also by Lemma 1.3, if $R(A) = R(A^T)$, then $N(A^+) = N(A^T)$ and A is EPr.

<u>COROLLARY 1.5</u>: *Given* V, R, *and* R^* *as in Theorem* 1.1.

a) *If* V_{22} *is nonsingular, conditions* 1' *and* 2' *reduce to* 3' *and* 4'.

b) *If* V_{11} *is nonsingular, conditions* 5' *and* 6' *reduce to* 7' *and* 8'.

c) *If* R_{11} *and* V_{22} *are EP matrices and* $V_{12} = V_{21}^T$, *conditions* 2' *and* 4' *reduce to* 1' *and* 3'.

d) *If* R^*_{22} *and* V_{11} *are EP matrices and* $V_{12} = V^T_{21}$, *conditions* 6' *and* 8' *reduce to* 5' *and* 7'.

e) *If* R_{11} *is nonsingular, conditions* 1', 2', 3', *and* 4' *reduce to*

$$9'. \quad V_{12}\left[I - V^+_{22}V_{22}\right] = \phi$$

$$10'. \quad \left[I - V_{22}V^+_{22}\right]V_{21} = \phi .$$

f) *If* R^*_{22} *is nonsingular, conditions* 5', 6', 7', *and* 8' *reduce to*

$$11'. \quad V_{21}\left[I - V^+_{11}V_{11}\right] = \phi$$

$$12'. \quad \left[I - V_{11}V^+_{11}\right]V_{12} = \phi .$$

g) *If* a), c), *and* d) *hold, then the eight conditions reduce to* 3', 5', *and* 7'.

h) *If* b), c), *and* d) *hold, then the eight conditions reduce to* 1', 3', *and* 7'.

i) *If* a), c), d), *and* f) *hold, then the eight conditions reduce to* 3' *and* 11'.

j) *If* b), c), d), *and* e) *hold, then the eight conditions reduce to* 7' *and* 9'.

k) *If* a), b), e), *and* f) *hold, no conditions exist, as this is a full rank case.*

Proof: The proofs of statements b), d), f), h), and j) are analogous to the proofs of a), c), e), g), and i),

respectively, by interchanging subscripts. Therefore, only statements a), c), e), g), i), and k) will be proven.

a) If V_{22} is nonsingular, $V_{22}^{+} = V_{22}^{-1}$ and $\left(I - V_{22}V_{22}^{+}\right) = \phi$ so that conditions 1' and 2' are simply

1'. $V_{22}^{-1}V_{21}\left[I - R_{11}R_{11}^{+}\right] = \phi$

2'. $\left[I - R_{11}^{+}R_{11}\right]V_{12}V_{22}^{-1} = \phi$.

Premultiplying 1' and postmultiplying 2' by V_{22} yield 3' and 4', respectively.

c) If condition 3' is transposed, the condition becomes

$$\left[I - R_{11}^{+T}R_{11}^{T}\right]V_{12} = \phi$$

or that $V_{12} \not\in N\left(R_{11}^{T}\right) = N(R_{11})$ by the EP hypothesis on R_{11} . Therefore, $\left[I - R_{11}^{+}R_{11}\right]V_{12} = \phi$ which is condition 4'. To show that 1' and 2' are equivalent, note that $R\left(R_{11}^{T}\right) = R(R_{11})$ since R_{11} is EP and therefore

$$R\left(V_{21}R_{11}^{T}\right) = R\left(V_{21}R_{11}\right) . \qquad (1.1)$$

Since V_{22} is EP, $N\left(V_{22}^{+}\right) = N\left(V_{22}^{+T}\right)$. From Theorem 1.8 of Chapter II, when $X \varepsilon V_n$, then X can be written uniquely as $X = U + W$, where $U \varepsilon N\left(V_{22}^{+}\right) = N\left(V_{22}^{+T}\right)$ and $W \varepsilon R\left(V_{22}\right) = R\left(V_{22}^{T}\right)$ by Theorem 1.4. Since $\left(I - V_{22}V_{22}^{+}\right)$ is a projection onto $N\left(V_{22}^{+}\right)$ and $\left(I - V_{22}^{T}V_{22}^{+T}\right)$ is a projection onto $N\left(V_{22}^{+T}\right)$,

$$\begin{aligned} \left(I - V_{22}V_{22}^{+}\right)X &= \left(I - V_{22}V_{22}^{+}\right)(U + W) \\ &= \left(I - V_{22}V_{22}^{+}\right)U \\ &= U \\ &= \left(I - V_{22}^{T}V_{22}^{+T}\right)U \\ &= \left(I - V_{22}^{T}V_{22}^{+T}\right)X \end{aligned}$$

for all $X \varepsilon V_n$ so that

$$\left(I - V_{22}V_{22}^{+}\right) = \left(I - V_{22}^{T}V_{22}^{+T}\right) .$$

Considering 2' and substituting in 4',

$$\begin{aligned} \left[I - V_{22}V_{22}^{+}\right]V_{21}R_{11} &= \left(\left[I - R_{11}^{+}R_{11}\right]V_{12}V_{22}^{+}\right)^{T} \\ &= \phi \end{aligned}$$

showing that $R\left(V_{21}R_{11}\right) \not\subset N\left(V_{22}^{+}\right)$. From equation (1.1) then $R\left(V_{21}R_{11}^{T}\right) \not\subset N\left(V_{22}^{+}\right) = N\left(V_{22}^{+T}\right)$ so that

$$\phi = \left[I - V_{22}^{T}V_{22}^{+T}\right]V_{21}R_{11}^{T}$$

$$= \left(R_{11}V_{12}\left[I - V_{22}^{+}V_{22}\right]\right)^{T}$$

which is condition 1' if 3' is substituted.

e) If R_{11} is nonsingular, $R_{11}^{+} = R_{11}^{-1}$ so that

$$\left[I - R_{11}R_{11}^{+}\right] = \phi$$

$$\left[I - R_{11}^{+}R_{11}\right] = \phi$$

and therefore conditions 3' and 4' are automatically satisfied along with the left side of 1' and right side of 2', reducing 1' and 2' to the conditions

1'. $R_{11}V_{12}\left[I - V_{22}^{+}V_{22}\right] = \phi$

2'. $\left[I - V_{22}V_{22}^{+}\right]V_{21}R_{11} = \phi$

which are conditions 9' and 10' after premultiplication of 1' and postmultiplication of 2' by R_{11}^{-1} .

g) If a) holds, then 1', 2', 3', and 4' are reduced to 3' and 4' which are further reduced to 3' if c) holds. If d) holds then 5', 6', 7' and 8' are reduced to 5' and 7'.

i) If a), c), and d) hold, then 1', 2', 3', 4' are reduced to 3' by g). If d) holds, then 5', 6', 7', and 8' are reduced to 5' and 7'. If f) holds, then 5' is reduced to 11', and 7' is automatically satisfied. (See e).)

k) If $V_{22}^{+} = V_{22}^{-1}$, $V_{11}^{+} = V_{11}^{-1}$, $R_{11}^{+} = R_{11}^{-1}$, and $R_{22}^{*+} = R_{22}^{*-1}$, then

$$\left[I - V_{22}V_{22}^{+}\right] = \phi = \left[I - V_{22}^{+}V_{22}\right]$$

$$\left[I - V_{11}V_{11}^{+}\right] = \phi = \left[I - V_{11}^{+}V_{11}\right]$$

$$\left[I - R_{11}R_{11}^{+}\right] = \phi = \left[I - R_{11}^{+}R_{11}\right]$$

$$\left[I - R_{22}^{*}R_{22}^{*+}\right] = \phi = \left[I - R_{22}^{*+}R_{22}^{*}\right]$$

which automatically satisfies all eight conditions. Note then that

$$1''. \quad \left[V_{11} - V_{12}V_{22}^{-1}V_{21}\right]^{-1} = V_{11}^{-1} + V_{11}^{-1}V_{12}R_{22}^{*-1}V_{21}V_{11}^{-1}$$

$$2''. \quad \left[V_{22} - V_{21}V_{11}^{-1}V_{12}\right]^{-1} = V_{22}^{-1} + V_{22}^{-1}V_{21}R_{11}^{-1}V_{12}V_{22}^{-1}$$

which is the "inside-out" rule [14, pp. 45-49], [27, pp. 24-26], [43, p. 29].

The above corollary can be used to obtain Theorem 6 (iii) of Lewis and Newman [26].

LEMMA 1.6: *If* A *is an* $n \times n$ *positive semidefinite matrix and* B *is* $m \times n$, *then*

$$N\left(A + B^T B\right) = N(A) \cap N(B) .$$

Proof: If $X \varepsilon N(A + B^T B)$, then $(A + B^T B)X = \phi$ or that

$$-AX = B^T BX . \tag{1.2}$$

Further since $X^T B^T BX = (BX)^T BX \geq 0$,

$$\begin{aligned} 0 &= X^T(A + B^T B)X \\ &= X^T AX + X^T B^T BX \\ &\geq 0 \end{aligned} \tag{1.3}$$

so that equation (1.2) becomes $-X^T AX = X^T B^T BX$ which is true if and only if

$$\begin{aligned} X^T AX &= 0 \\ &= X^T B^T BX \\ &= (BX)^T BX \end{aligned}$$

since A and $B^T B$ are positive semidefinite.

Consequently, $(A + B^TB)X = \phi$ if and only if $BX = \phi$ so that $AX = \phi$ from (1.2). Therefore $X \varepsilon N(A)$ and $X \varepsilon N(B)$ which is $X \varepsilon N(A) \cap N(B)$. If $X \varepsilon N(A) \cap N(B)$, then $X \varepsilon N(A + B^TB)$.

<u>LEMMA 1.7</u> [26]: *If* A *is positive semidefinite, then* A *is* EP.

<u>COROLLARY 1.8</u> [26]: *If* A *is a positive semidefinite* $n \times n$ *matrix,* C *is an* $r \times n$ *matrix and* $\overline{A} = A + C^TC$, *then*

$$\overline{A}^+ = A^+ - A^+C^T\left(I + CA^+C^T\right)^{-1}CA^+$$

if and only if $N(A) \subset N(C)$.

Proof: Let $V_{11} = A$, $V_{21} = C = V_{12}^T$, and $V_{22} = -I$. Notice that $R_{11}^+ = A + C^TC = \overline{A}$ is positive semidefinite by the proof of Lemma 1.6, equation (1.3) and therefore $\overline{A}$ is EP as is V_{11} by Lemma 1.7, implying that R_{11} is EP by Theorem 1.4. Now

$$\begin{aligned} R_{22}^* &= \left[V_{22} - V_{21}V_{11}V_{12}\right]^+ \\ &= -\left[I + CA^+C^T\right]^+ . \end{aligned}$$

Since A^+ is positive semidefinite, CA^+C^T is positive semidefinite. Letting CAC^T be "A" and I be "B" of Lemma 1.7, we have

$$\begin{aligned} r(I + CAC^T) &= n - \mathit{dim}\ N(I + CAC^T) \\ &= n - \mathit{dim}\left(N(I) \cap N(CAC^T)\right) \\ &= n \end{aligned}$$

showing that $R^*_{22} = -\left[I + CA^+C^+\right]^+$ is nonsingular where $r(\cdot)$ stands for the rank of a matrix. Therefore, all of the hypotheses of Corollary 1.5 i) are satisfied showing that

$$\begin{aligned} \left(A + C^TC\right)^+ &= A^+ + A^+C^T\left[-\left(I + CA^+C^T\right)^{-1}\right]CA^+ \\ &= \overline{A}^+ \end{aligned}$$

if and only if 3' and 11' hold, which are

3'. $C\left[I - \overline{A}\overline{A}^+\right] = \phi$

11'. $C\left[I - AA^+\right] = \phi$.

Now 3' holds if and only if $R\left[I - AA^+\right] \subset N(C)$. However, $I-AA^+$ is the projection onto $N(\overline{A})$ so that 3' holds if and only if $N(\overline{A}) \subset N(C)$. But $N(\overline{A}) = N(A) \cap N(C) \subset N(C)$

by Lemma 1.6 so that 3' will always hold. The only remaining condition is 11' which holds if and only if $R\left[(I - AA^+)\right] \subset N(C)$ which is if and only if $N(A) \subset N(C)$.

DEFINITION 1.9: *For* D *a diagonal matrix with diagonal elements* $d_1, d_2, \cdots, d_n$ *, denoted by* $D = diag\left[d_1,\cdots,d_n\right] = diag\left[d_i\right]$ *, and any real number* r *, let*

$$D^r = diag\left[d_1',d_2',\cdots,d_n'\right]$$

where $d_i' = d_i^r$ *if* $d_i \neq 0$ *or* $d_i' = 0$ *if* $d_i = 0$.

Observe that

$$\begin{aligned} D_1^r D_2^r &= diag\left[d_{1i}^r d_{2i}^r\right] \\ &= \left(D_1 D_2\right)^r \end{aligned}$$

$$D^r D^s = diag\left[d_i^r d_i^s\right] = D^{r+s}$$

with equality also holding if $d_i = 0$ for some i . If some diagonal elements are zero, ambiguity occurs for negative r . To examine this, suppose D is an $n \times n$ dimensional diagonal matrix such that

$$D = diag\left[d_1,\cdots,d_m,0,\cdots,0\right] .$$

Then by the above definition,

$$D^{-1} = diag\left[1/d_1,\cdots,1/d_m,0,\cdots,0\right] \quad .$$

so that D^{-1} is really the pseudoinverse of D and should probably be denoted by D^+ . This makes for clumsy notation; for example,

$$\begin{aligned} D^{-1/2} &= diag\left[1/d_1^{1/2},\cdots,1/d_m^{1/2},0,\cdots,0\right] \\ &= \left(D^+\right)^{1/2} \end{aligned}$$

so that, with this qualification, D^r will be used for all real r whether D is nonsingular or not.

For the next two theorems, consider the linear model

$$Y = AX$$

and the weighted least squares solution $\hat{X}$ which is such that

$$(Y - A\hat{X})^T W(Y - A\hat{X}) \leq \left(Y - AX^*\right)^T W(Y - AX^*)$$

for all other X^* and positive semidefinite weight matrix W . Since the weight function only will appear

in a quadratic form, W can be assumed to be symmetric, without loss of generality. Letting $W = W^{(1/2)T}W^{1/2}$ be a factorization of W [16, p. 4], then

$$(Y - A\hat{X})^T W(Y - A\hat{X}) = (Y - A\hat{X})^T W^{(1/2)T}W^{1/2}(Y - A\hat{X})$$

$$= \left(W^{1/2}Y - W^{1/2}A\hat{X}\right)^T$$

$$\cdot \left(W^{1/2}Y - W^{1/2}A\hat{X}\right)$$

and the least squares solution is

$$\hat{X} = \left(W^{1/2}A\right)^+ W^{1/2}Y$$

$$= \left(A^T W^{(1/2)T}W^{1/2}A\right)^+ A^T W^{(1/2)T}W^{1/2}Y$$

$$= \left(A^T W A\right)^+ A^T W Y \tag{1.4}$$

[16], [28], [42], [43], [44].

<u>THEOREM 1.10</u>: *Suppose* Y, A, *and* W *are partitioned as*

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \quad A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad W = \begin{bmatrix} W_1 & \phi \\ \phi & W_2 \end{bmatrix}$$

with dimensions of Y_1, A_1, *and* W_1 *as well as* Y_2, A_2, *and* W_2 *corresponding. Further suppose that* W_2 *is nonsingular. Then if the weighted least squares solution is such that* $Y_2 - A_2\hat{X} = \phi$, *then* $\hat{X} = \hat{X}_1$, *where* $\hat{X}_1$ *is the weighted least squares solution using the model* $Y_1 = A_1X$ *and weights* W_1.

Proof: Observe first that

$$\begin{aligned} A^TWA &= \left(A_1^T \vdots A_2^T\right)\begin{pmatrix} W_1 & \phi \\ \phi & W_2 \end{pmatrix}\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \\ &= A_1^TW_1A_1 + A_2^TW_2A_2 \\ &= A_1^TW_1A_1 + A_2^TW_2^{(1/2)T}W_2^{1/2}A_2 \\ &= A_1^TW_1A_1 + \left(W_2^{1/2}A_2\right)^T\left(W_2^{1/2}A_2\right) \end{aligned}$$

so that

$$N(A^TWA) = N\left(A^TW_1A_1\right) \cap N\left(W_2^{1/2}A_2\right) \subset N\left(W_2^{1/2}A_2\right)$$

by Lemma 1.6, so that Corollary 1.8 will apply to A^TWA, and

$$\left(A^TWA\right)^+ = \left[A_1^TW_1A_1 + \left(W_2^{1/2}A_2\right)^T\left(W_2^{1/2}A_2\right)\right]^+$$

$$= \left(A_1^TW_1A_1\right)^+ - \left(A_1^TW_1A_1\right)^+\left(W_2^{1/2}A_2\right)^T$$

$$\cdot \left[I + \left(W_2^{1/2}A_2\right)\left(A_1^TW_1A_1\right)^+\left(W_2^{1/2}A_2\right)^T\right]^{-1}$$

$$\cdot \left(W_2^{1/2}A_2\right)\left(A_1^TW_1A_1\right)^+$$

$$= \left(A_1^TW_1A_1\right)^+ - \left(A_1^TW_1A_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}$$

$$\cdot A_2\left(A_1^TW_1A_1\right)^+$$

since W_2 is nonsingular and therefore $W_2^{1/2}$ is nonsingular.

Consequently, consider the condition

$$
\begin{aligned}
\phi &= Y_2 - A_2\hat{X} \\
&= Y_2 - A_2\Big\{\left(A_1^TW_1A_1\right)^+ \\
&\quad - \left(A_1^TW_1A_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}A_2\left(A_1^TW_1A_1\right)^+\Big\}\left[A_1^TW_1Y_1 + A_2^TW_2Y_2\right] \\
&= Y_2 - A_2\Big\{\hat{X}_1 - \left(A_1^TWA_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}A_2\left(A_1^TW_1A_1\right)^+A_1^TW_1Y_1 \\
&\quad + \left(A_1^TW_1A_1\right)^+A_2^TW_2Y_2 \\
&\quad - \left(A_1^TW_1A_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}A_2\left(A_1^TW_1A_1\right)^+A_2^TW_2Y_2\Big\} \\
&= Y_2 - A_2\hat{X}_1 + A_2\left(A_1^TWA_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}A_2\left(A_1^TW_1A_1\right)^+A_1^TW_1Y_1 \\
&\quad - A_2\left(A_1^TW_1A_1\right)^+A_2^TW_2Y_2 + A_2\left(A_1^TW_2Y_2\right) \qquad\qquad (1.5) \\
&\quad + A_2\left(A_1^TW_1A_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}\left[A_2\left(A_1^TW_1A_1\right)^+A_2^TW_2 + I\right]Y_2 \\
&\quad - A_2\left(A_1^+W_1A_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}Y_2 \\
&= Y_2 - A_2\hat{X}_1 + A_2\left(A_1^TWA_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}A_2\left(A_1^TW_1A_1\right)^+A_1^TW_1Y_1 \\
&\quad - A_2\left(A_1^TW_1A_1\right)^+A_2^TW_2Y_2 + A_2\left(A_1^TW_1A_1\right)^+A_2^TWY_2 \\
&\quad - A_2\left(A_1^+W_1A_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}Y_2 \\
&= Y_2 - A_2\hat{X}_1 - A_2\left(A_1^TW_1A_1\right)^+A_2^T\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}\left[Y - A_2\hat{X}_1\right] \\
&= \left[I - A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]\left[W_2^{-1} + A_2\left(A_1^TW_1A_1\right)^+A_2^T\right]^{-1}\left[Y - A_2\hat{X}_1\right] \\
&= W_2\left[Y_2 - A_2\hat{X}_1\right] .
\end{aligned}
$$

Therefore, since W_2 is nonsingular, $Y_2 - A_2\hat{X}_1 = \phi$.

To conclude the theorem, calculate

$$
\begin{aligned}
\hat{X} &= \left(A^{T}WA\right)^{+}A^{T}WY \\
&= \Big\{\left(A_1^{T}W_1A_1\right)^{+} - \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\left[W_2^{-1} + A_2\left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\right]^{-1} \\
&\qquad \cdot A_2\left(A_1^{T}W_1A_1\right)^{+}\Big\}\left[A_1^{T}W_1Y_1 + A_2^{T}W_2Y_2\right] \\
&= \hat{X}_1 + \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}W_2Y_2 \\
&\qquad - \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\left[W_2^{-1} + A_2\left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\right]^{-1}A_2\left(A_1^{T}W_1A_1\right)^{+}A_2^{T}W_2Y_2 \\
&\qquad - \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\left[W_2^{-1} + A_2\left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\right]^{-1}A_2\hat{X}_1 \\
&= \hat{X}_1 + \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}W_2Y_2 \\
&\qquad - \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\left[W_2^{-1} + A_2\left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\right]^{-1}A_2\left(A_1^{T}W_1A_1\right)^{+} \\
&\qquad \cdot A_2^{T}W_2Y_2 - \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\left[W_2^{-1} + A_2\left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\right]^{-1}Y_2 \\
&= \hat{X}_1 + \left(A_1^{+}W_1A_1\right)^{+}A_2^{T}W_2Y_2 \\
&\qquad - \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\left[W_2^{-1} + A_2\left(A_1^{T}W_1A_1\right)^{+}A_2^{T}\right]^{-1} \\
&\qquad \cdot \left[A_2\left(A_1^{T}W_1A_1\right)^{+}A_2^{T} + W_2^{-1}\right]W_2Y_2 \\
&= \hat{X}_1 + \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}W_2Y_2 - \left(A_1^{T}W_1A_1\right)^{+}A_2^{T}W_2Y_2 \\
&= \hat{X}_1 \quad .
\end{aligned}
$$

LEMMA 1.11: *Let* $g = (Y - A\hat{X})/\left[(Y - A\hat{X})\ (Y - A\hat{X})\right]^{1/2}$ *where* $\hat{X}$ *is the least squares estimate for a nonweighted model* $(W = I)$. *Then*

a) $g^T g = 1$;

b) $A^T g = \phi$;

c) *for any* h *such that* h *satisfies* a) *and* b), *then* $Y^T h \leq Y^T g$

d) $\left[(Y-A\hat{X})^T (Y-A\hat{X})\right]^{1/2} = Y^T(Y-A\hat{X})/\left[(Y-A\hat{X})^T(Y-A\hat{X})\right]^{1/2}$.

Proof:

a)
$$g^T g = \frac{(Y-A\hat{X})^T}{\left[(Y-A\hat{X})^T(Y-A\hat{X})\right]^{1/2}} \cdot \frac{(Y-A\hat{X})}{\left[(Y-A\hat{X})^T(Y-A\hat{X})\right]^{1/2}}$$
$$= 1 .$$

b) Observe that $\hat{X} = (A^T A)^{+} A^T Y$ or that $(A^T A)\hat{X} = A^T Y$ since $A^T Y \subset R(A^T) = R(A^T A)$ and $(A^T A)\left(A^T A\right)^{+}$ is the projection onto $R(A^T A)$. Then

$$\begin{aligned} A^T g &= A^T(Y - A\hat{X})/\left[(Y - A\hat{X})^T(Y - A\hat{X})\right]^{1/2} \\ &= (A^T Y - A^T A\hat{X})/\left[(Y - AX)^T(Y - AX)\right]^{1/2} \\ &= \phi . \end{aligned}$$

c) If $A^T h = \phi$, then $h \varepsilon N(A^T) = N(A^+)$, so that $(I - AA^+)h = h$, and h can be written as $(I-AA^+)Z$, for Z any element of the whole space. If $h^T h = 1$, then h must be in the form

$$h = (I - AA^+)Z/\left\{\left[(I - AA^+)Z\right]^T\left[(I - AA^+)Z\right]\right\}$$

$$= (I - AA^+)Z/\left[Z^T(I - AA^+)Z\right] .$$

The Cauchy-Schwartz inequality is

$$u^T v \leq \left(u^T u\right)^{1/2}\left(v^T v\right)^{1/2}$$

so that with $u = (I - AA^+)Y$ and $v = (I - AA^+)Z$, the inequality becomes

$$Y^T(I - AA^+)(I - AA^+)Z$$

$$\leq \left[Y^T(I - AA^+)Y\right]^{1/2}\left[Z^T(I - AA^+)Z\right]^{1/2}$$

which is

$$Y^T(I - AA^+)Z/\left[Z^T(I - AA^+)Z\right]^{1/2} \leq \left[Y^T(I - AA^+)Y\right]^{1/2}$$

or

$$
\begin{aligned}
Y^T h &\leq \left[Y^T(I - AA^+)Y\right]^{1/2} \\
&= Y^T(I - AA^+)Y/\left[Y^T(I - AA^+)Y\right]^{1/2} \\
&= Y^T(Y - A\hat{X})/\left[Y^T(I - AA^+)Y\right]^{1/2} \\
&= Y^T g
\end{aligned}
$$

since $A^+Y = \hat{X}$.

d) $$
\begin{aligned}
(Y - A\hat{X})^T(Y - A\hat{X}) &= Y^T(Y - A\hat{X}) - \hat{X}^T A^T(Y - A\hat{X}) \\
&= Y^T(Y - A\hat{X}) - \hat{X}^T(A^T Y - A^T A\hat{X}) \\
&= Y^T(Y - A\hat{X}) \ .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\left[(Y - A\hat{X})^T(Y - A\hat{X})\right]^{1/2} &= \frac{(Y - A\hat{X})^T(Y - A\hat{X})}{\left[(Y - A\hat{X})^T(Y - A\hat{X})\right]^{1/2}} \\
&= \frac{Y^T(Y - A\hat{X})}{\left[(Y - A\hat{X})^T(Y - A\hat{X})\right]^{1/2}} \ .
\end{aligned}
$$

<u>LEMMA 1.12</u>: *(Holder Inequality in matrix trace notation):* *For real numbers* $a_i \geq 0$, $b_i \geq 0$, $i = 1,\cdots,n$ *and* r *and* s *such that* $(1/r) + (1/s) = 1$, *then if* $D_1 = diag(a_i)$, $D_i = diag(b_i)$, *then*

$$tr(D_1D_2) \leq tr\left(D_1^r\right)^{1/r} tr\left(D_i^s\right)^{1/s}$$

with equality if and only if there is a real constant α *such that*

$$D_1^r = \alpha D_2^s$$

where $tr(D) = \sum_{i=1}^{n} d_i$ *for any diagonal matrix* D .

Proof: $tr(D_1D_2) = \sum_{i=1}^{n} a_ib_i$

$$\leq \left(\sum_{i=1}^{n} a_i^r\right)^{1/r}\left(\sum_{i=1}^{n} b_i^s\right)^{1/s}$$

$$= tr\left(D_1^r\right)^{1/r} tr\left(D_2^s\right)^{1/s}$$

with equality if and only if there is a real constant α such that

$$a_i^r = \alpha b_i^s$$

for $i = 1,\cdots,n$, which is true if and only if

$$D_1^r = \alpha D_2^s$$

by the discrete Holder Inequality [18, pp. 21-26].

LEMMA 1.13: *If a sequential process is defined by* $x_{n+1} = F(y_n)$ *with* F *continuous, then if*

$$\lim_{n\to\infty} x_n = x \quad ; \quad \lim_{n\to\infty} y_n = y$$

then $F(y) = x$.

Proof: F continuous, $\lim x_n = x$, and $\lim y_n = y$ imply that for every $\varepsilon > 0$ there are constants N_1, N_2, and N_3 such that

a) if $n > N_1$, $|F(y_n) - F(y)| < \varepsilon/3$,

b) if $n > N_2$, $|x_{n+1} - x_n| < \varepsilon/3$,

c) if $n > N_3$, $|x_n - x| < \varepsilon/3$.

Pick $N = \max[N_1,N_2,N_3]$ so that for $n > N$,

$$\begin{aligned} |F(y) - x| &= |F(y) - F(y_n) + F(y_n) - x_n + x_n - x| \\ &\le |F(y) - F(y_n)| + |F(y_n) - x_n| + |x_n - x| \\ &= |F(y) - F(y_n)| + |x_{n+1} - x_n| + |x_n - x| \\ &< \varepsilon \end{aligned}$$

for all ε so that $F(y) = x$ since $|F(y)-x|$ is independent of n.

LEMMA 1.14: *The function*

$$F(x) = \left[1 + (x/c)^r\right]^s - 1$$

approaches zero more rapidly than does x *when* $r > 1 > s$ *and* c *is a positive constant. In fact the rate is* x^{r-1}.

Proof: For x close to zero and since $s < 1$, the infinite binomial series for $F(x)$ converges to yield

$$F(x) = 1 + s(x/c)^r + s(s-1)(x/c)^{r^2}/2! + \cdots - 1$$

$$= (x/c)\left[s(x/c)^{r-1} + s(s-1)(x/c)^{2r-1}/2! \cdots\right]/c$$

so that

$$\lim_{x\to 0} F(x)/x = \lim_{x\to 0}\left[s(x/c)^{r-1} + h(x^{2r-1})\right]$$

$$= 0$$

where $h(x^{2r-1})$ are higher order terms of the order x^{2r-1} or higher. Consequently, $F(x)$ approaches zero more rapidly than x and at a rate of x^{r-1}.

3.2 Calculation of ℓ^p Approximations, $p \geq 2$

In this section an algorithm is presented to obtain the ℓ^p approximations to the degenerate system

$$Y = AX$$

which is that value (or values) of the $n \times 1$ dimensional vector X, say $\hat{X}^p$, such that

$$\mathrm{tr}\left[S\left(\hat{X}^p\right)^p\right]^{1/p} = \left[\sum_{i=1}^{m} |Y_i - A^i\hat{X}^p|^p\right]^{1/p}$$

$$\leq \left[\sum_{i=1}^{m} |Y_i - A^i X^*|^p\right]^{1/p}$$

$$= \mathrm{tr}\left[S\left(X^*\right)^p\right]^{1/p}$$

where X^* is any value of X, and $S(X) = diag\left[|Y_i - A^i X|\right]$ with A^i the ith row of A.

The solutions $\hat{X}^p$ will be characterized by selecting a weighting matrix $W = diag\left[w_i \geq 0\right]$, $\mathrm{tr}[W] = 1$ so that $\hat{X}^p$ is the solution of

$$\mathrm{tr}\left[WS\left(\hat{X}^p\right)^2\right]^{1/2} = \left[\sum_{i=1}^{m} w_i\left(Y_i - A^i\hat{X}^p\right)^2\right]^{1/2}$$

$$= \left[\left(Y - A\hat{X}^p\right)^T W\left(Y - A\hat{X}^p\right)\right]^{1/2}$$

$$= \left[\left(W^{1/2}Y - W^{1/2}A\hat{X}^p\right)^T \cdot \left(W^{1/2}Y - W^{1/2}A\hat{X}^p\right)\right]^{1/2}$$

$$\leq \left[\left(Y - AX^*\right)^T W\left(Y - AX^*\right)\right]^{1/2}$$

$$= \mathrm{tr}\left[WS\left(X^*\right)\right]^{1/2} \qquad (2.1)$$

for all X^*. The solutions to (2.1) are least squares solutions to

$$W^{1/2}Y = W^{1/2}AX$$

which are

$$\hat{X}^p = \left(W^{1/2}A\right)^+ W^{1/2}Y + \left[I - \left(W^{1/2}A\right)^+\left(W^{1/2}A\right)\right]Z$$

$$= \left(A^TWA\right)^+ A^TWY + \left[I - \left(W^{1/2}A\right)^+\left(W^{1/2}A\right)\right]Z \qquad (2.2)$$

where Z is an arbitrary $n\times 1$ vector [9], [27], [41], [42], [44].

To illustrate the heuristics behind the algorithm, consider the following theorem.

THEOREM 2.1: *If* $\hat{X}^p$ *is a weighted* ℓ^p *approximation, then it is a weighted* ℓ^r *approximation for* $p, r > 0$.

Proof: $\hat{X}^p$ is a weighted ℓ^p approximation if and only if for the weights $W = diag[w_i]$, let $U = WS\left(\hat{X}^p\right)^{p-r}$, so that

$$\begin{aligned}
Q_u^r(\hat{X}^p) &= tr\left(US(\hat{X}^p)^r\right) \\
&= tr\left[\left(WS(\hat{X}^p)^{p-r}\right)S(\hat{X}^p)^r\right] \\
&= tr\left(WS(\hat{X}^p)^p\right) \\
&\leq tr\left(WS(X^*)^p\right) \\
&= tr\left(US(X^*)^r\right)
\end{aligned}$$

for all X^* , so that $\hat{X}^p$ is the weighted ℓ^r approximation with weights U . Notice that if $r > p$, U has points of possible singularity at $Y_i - A^i\hat{X}^p$. If such a singularity occurs for some subset J_2 of $\{1,\cdots,m\}$, set $U_2 = \phi$. With $U_1 = W_1 S_1\left(\hat{X}^p\right)^{p-r}$ now restricted to the nonzero subset, then

$$
\begin{aligned}
Q_w^p(\hat{X}^p) &= \mathrm{tr}\left(WS(X^p)^p\right) \\
&= \mathrm{tr}\left(W_1S_1(X^p)^p\right) \\
&= \mathrm{tr}\left(U_1S_1(X^p)^r\right) \\
&= \mathrm{tr}\left(U_1S_1(X^p)^r\right) + \mathrm{tr}\left(U_2S_2(X^p)^r\right) \\
&= \mathrm{tr}\left(US(X^p)^r\right) \\
&= Q_u^r(X^p)
\end{aligned}
$$

since $U_2 = \phi = S_2\left(\hat{X}^p\right)^r$ if $r > 0$.

For the discussion in this paper, consider only ℓ^p spaces for $1 < p < +\infty$. Algorithms for $p = +\infty$ can be found in references [11], [21], [23], [31], and [32] and for $-\infty \leq p < 1$ in [12] and for $p = 1$ in [11] and [52]. The algorithm defined below is for $2 < p < \infty$.

<u>DEFINITION 2.2</u>: *Let* $W_1 = diag\left[W_1^1 > 0\right]$ *be arbitrary except subject to* $\mathrm{tr}\left[W_1\right] = 1$. *Define*

1. $$W_{k+1} = \frac{\left[W_kS\left(\hat{X}_k^p\right)\right]^{(p-2)/(p-1)}}{\mathrm{tr}\left\{\left[W_kS\left(\hat{X}_k^p\right)\right]^{(p-2)/(p-1)}\right\}}$$

2. $$\sigma^k = tr\left[W_k S\left(\hat{X}_k^p\right)^2\right]^{1/2} \Big/ tr\left[W_k^{p/(p-2)}\right]^{(p-2)/2p}$$

$$= \left[Q^2_{W_k}\right]^{1/2} \Big/ tr\left[W_k^{p/(p-2)}\right]^{(p-2)/2p}$$

3. $J_k = \left\{i : w_i^k > 0\right\}$ *where* $\hat{X}_k^p$ *is the weighted least squares solution with weights* W_k *minimizing* $Q^2_{W_k}(X)$.

4. *The algorithm, step by step, is then to*

a) *Calculate, for* $k = 1$,

i) $\hat{X}_1^p$

ii) $\eta_1 = tr\left[W_1 S\left(\hat{X}_1^p\right)^2\right]$

iii) *If* $\eta_1 = 0$, *terminate the algorithm; otherwise calculate*

$$\delta_1 = tr\left[W_k^{p/(p-2)}\right]^{(p-2)/2p}$$

iv) σ^1

b) *Calculate, for the kth step,* $k > 1$

i) W_{k+1}

ii) $\hat{X}_k^p$

iii) $\eta_k = tr\left[W_k S\left(X_k^p\right)^2\right]$

iv) *If* $\eta_k = 0$, *terminate the algorithm; otherwise calculate*

$$\delta_k = \mathrm{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/2p}$$

v) σ^k

vi) *If* $|\sigma^k - \sigma^{k-1}| < \varepsilon$, *a predetermined constant, terminate the algorithm; otherwise return to b)i).*

One value for W_1 is *diag*$\{1/m,\cdots,1/m\}$. Since the elements of the two diagonal matrices W_k and $S\left(\hat{X}_k^p\right)$ are nonzero, then if $\eta_k = 0$,

$$w_i|Y_i - A^i\hat{X}_k^p|^2 = 0 \tag{2.3}$$

for $i = 1,2,\cdots,m$, and since $w_i > 0$ for $i \varepsilon J_k$, then $Y_i - A^i\hat{X}_k^p = 0$ for all nonzero weights. Notice that if $\eta_k \neq 0$, then σ^k is well defined, and there is at least one term $w_i|Y_i - A^i\hat{X}_k^p|$ nonzero implying that W_{k+1} is well defined. Observe that if $w_{k+1}^i = 0$ for some i , then the corresponding term $\left(w_k^i|Y_i - A^i\hat{X}_k^p|\right)^{(p-1)/(p-2)}$ is zero. This forces either $w_k^i = 0$ or $|Y_i - A^i\hat{X}_k^p| = 0$ for $p > 2$, impling that if $w_k^i = 0$ then $w_{k+1}^i = 0$. The algorithm is therefore a well-defined procedure. It will be shown that if $\eta_1 > 0$,

1. $\lim_{k\to\infty} \sigma^k = \left[Q_I^p(\hat{X}^p)\right]^{1/p}$

2. $\lim_{k\to\infty} \hat{X}_k^p = \hat{X}^p$

3. $\lim_{k\to\infty} J_k = J$

where $\hat{X}^p$ is the nonweighted ℓ^p solution on J . If $J \neq \{1,\cdots,m\}$, then a procedure will be presented to increase J in a finite number of steps to $\{1,\cdots,m\}$. This algorithm was developed from the algorithms of references [23] and [46] and appears in another version for L^p spaces, p even, in reference [21].

LEMMA 2.3: *If* $\eta_1 > 0$, *then* $\eta_k > 0$, *for all* k .

Proof: By induction, since $\eta_1 > 0$ is given, assume that $\eta_k > 0$ which implies that there is a $w_k^i > 0$ and therefore J_k is nonempty. Two possibilities occur: either $J_{k+1} = J_k$ or $J_k - J_{k+1}$ is nonempty. The third possibility $J_{k+1} - J_k$ nonempty implies there is an i such that $w_i^k = 0$, but $w_i^{k+1} > 0$, which is impossible from the remarks after Definition 2.2.

a) If $J_{k+1} = J_k$, then $\eta^{k+1} = 0$ implies from the remarks after Definition 2.2 that

$$w_i \left| Y_i - A^i \hat{X}_{k+1}^p \right|^2 = 0$$

for all i, so that since $w_{k+1}^{i} > 0$ for $i \in J_{k+1}$,

$$|Y_i - A^i \hat{X}_{k+1}^{p}| = 0 .$$

Define the submatrices

$$S_1(X) = diag\left[s_j = |Y_{i(j)} - A^{i(j)}X| \ , \ i(j) \in J_{k+1}\right]$$

$$S_2(X) = diag\left[s_j = |Y_{i(j)} - A^{i(j)}X| \ , \ i(j) \notin J_{k+1}\right]$$

$$W_{k+1,1} = diag\left[w_{k+1,1}^{j} = w_{k+1}^{i(j)} \ , \ i(j) \in J_{k+1}\right]$$

$$W_{k+1,2} = diag\left[w_{k+1,2}^{j} = w_{k+1}^{i(j)} \ , \ i(j) \notin J_{k+1}\right]$$

where $i(1)$ is the smallest integer in the set, $i(n)$ is the *nth* smallest integer in the set. Then $S_1\left(\hat{X}_{k+1}^{p}\right) = \phi$, $W_{k+1,2} = \phi$, and since $J_{k+1} = J_k$, $W_{k,2} = \phi$ also, making

$$\phi = W_{k+1,1}S_1\left(\hat{X}_{k+1}^{p}\right) = W_{k,1}S_1\left(\hat{X}_{k+1}^{p}\right)$$

$$\phi = W_{k+1,2}S_2\left(\hat{X}_{k+1}^{p}\right) = W_{k,2}S_2\left(\hat{X}_{k+1}^{p}\right) .$$

Since $\hat{X}_k^p$ is the least squares estimate with weights W_k,

$$\begin{aligned}
\eta_k &= tr\left[W_k S\left(\hat{X}_k^p\right)\right] \\
&\leq tr\left[W_k S\left(\hat{X}_{k+1}^p\right)\right] \\
&= tr\left[W_{k,1} S_1\left(\hat{X}_{k+1}^p\right)\right] + tr\left[W_{k,2} S_2\left(\hat{X}_{k+1}^p\right)\right] \\
&= 0
\end{aligned}$$

which contradicts the assumption that $\eta_k > 0$.

b) If $J_k - J_{k+1}$ is nonempty and if $i \varepsilon J_k - J_{k+1}$, then $w_k^i > 0$ and $w_{k+1}^i = 0$. Define the submatrices

$$A_1 = \begin{pmatrix} A^{i(1)} \\ \vdots \\ A^{i(n_1)} \end{pmatrix} , \quad Y_1 = \begin{pmatrix} Y_{i(1)} \\ \vdots \\ Y_{i(n_1)} \end{pmatrix}$$

$$W_{k,1} = diag\left[w_{k,1}^i = w_k^{i(j)} , \quad i(j) \not\varepsilon J_k - J_{k+1}\right]$$

for $i(j) \not\varepsilon J_k - J_{k+1}$ and i(j) the ordering as in part a) with n_1 the number of indices not in $J_k - J_{k+1}$. Define in a similar manner, for $i(j) \varepsilon J_k - J_{k+1}$, and n_2 the number of indices in $J_k - J_{k+1}$,

$$A_2 = \begin{pmatrix} A^{i(1)} \\ \vdots \\ A^{i(n_2)} \end{pmatrix} , \quad Y_1 = \begin{pmatrix} Y_{i(1)} \\ \vdots \\ Y_{i(n_2)} \end{pmatrix}$$

$$W_{k,2} = diag\left[w^j_{k,2} = w_k^{i(j)} > 0 , \quad i(j) \in J_k - J_{k+1}\right] .$$

From the remarks after Definition 2.2, $w^i_{k+1} = 0$ implies that either $w^i_k = 0$ or $|Y_i - A^i\hat{X}^p_k| = 0$ so that for $i \in J_k - J_{k+1}$, $|Y_i - A^i\hat{X}^p_k| = 0$ since $w^i_k > 0$. Therefore with S_1 and S_2 partitioned compatible with A_1 and A_2, $Y_2 - A_2\hat{X}^p_k = 0$ and

$$S_2\left(\hat{X}^p_k\right) = diag\left[S_j = |Y_{j(i)} - A^{j(i)}\hat{X}^p_k| , \quad i \in J_k - J_{k+1}\right]$$

$$= \phi .$$

Since the diagonal elements of $W_{k,2}$ are positive, $W_{k,2}$ is nonsingular and the system satisfies the hypotheses of Theorem 1.10 implying that $\hat{X}^p_k$ is also the least squares estimate over the reduced system $Y_1 = A_1X$. Therefore, similar to part a),

$$\begin{aligned}
\eta_k &= \operatorname{tr}\left[W_k S\left(\hat{X}_k^p\right)\right] \\
&= \operatorname{tr}\left[W_{k,1} S_1\left(\hat{X}_k^p\right)\right] + \operatorname{tr}\left[W_{k,2} S_2\left(\hat{X}_k^p\right)\right] \\
&= \operatorname{tr}\left[W_{k,1} S_1\left(\hat{X}_k^p\right)\right] \\
&\leq \operatorname{tr}\left[W_{k,1} S_1\left(\hat{X}_{k+1}^p\right)\right] + \operatorname{tr}\left[W_{k+1,2} S_2\left(\hat{X}_{k+1}^p\right)\right]
\end{aligned}$$

since $W_{k+1,2} = \phi$. Observe that the diagonal elements of $W_{k,1}$ were obtained from the complement of $J_k - J_{k+1}$. Since the only nonzero weights at the *kth* iteration are indexed by J_k , the nonzero diagonal elements of $W_{k,1}$ must also be indexed by J_{k+1} . If $\eta_{k+1} = 0$, then, similar to a),

$$w_i \left|Y_i - A^i \hat{X}_{k+1}^p\right|^2 = 0$$

for all i . Since $w_{k+1}^i > 0$ for $i \in J_{k+1}$, then $\left|Y_i - A^i \hat{X}_{k+1}^p\right| = 0$, implying that $S_1\left(\hat{X}_{k+1}^p\right) = 0$. Thus

$$\begin{aligned}
\eta_k &\leq \operatorname{tr}\left[W_{k,1} S_1\left(\hat{X}_{k+1}^p\right)\right] + \operatorname{tr}\left[W_{k+1,2} S_2\left(\hat{X}_{k+1}^p\right)\right] \\
&= 0
\end{aligned}$$

contradicting that $\eta_k > 0$. Therefore $\eta_{k+1} > 0$.

Lemma 2.3 implies that the J_k are nested; that is, that $J_1 \supset J_2 \supset \cdots \supset J_k \supset J_{k+1} \supset \cdots$. Also, it was shown that if $\eta_k = 0$, then $|Y_i - A^i\hat{X}_k^p| = 0$ for all nonzero weights. This lemma then implies that if the algorithm does not exactly fit (interpolate) the vector X with the original weights, the algorithm will not exactly fit X at any subsequent step. For the full rank case $\left(r(A) = m\right)$ where every m×m submatrix is of rank m (the Haar Condition [6, p. 74]), $\cap J_k$ must contain at least m+1 points since any m×m submatrix of rank m will interpolate X . Since the J_k are nested with a finite number of elements, then

$$\bigcap_{i=1}^{\ell} J_k = J_\ell$$

so that

$$\lim_{\ell\to\infty} \text{num}\left(\bigcap_{i=1}^{\ell} J_k\right) = \lim_{\ell\to\infty} \text{num}(J_\ell)$$

where num(•) stands for the number of elements in the set. Therefore the limit must be attained. The following lemma proves that $\sigma^k > 0$ is a strictly monotonically increasing sequence so that $\cap J_k$ is nonempty, since $\cap J_k$ empty implies $\eta_k = 0$, for large k .

LEMMA 2.4: *If* $W_{k+1} = W_k$, *then* $\sigma^{k+1} = \sigma^k$. *Otherwise* $\sigma^{k+1} > \sigma^k$.

Proof: From equation (2.1), if $W_k = W_{k+1}$, then the values $\hat{X}_k^p$ equal the values $\hat{X}_{k+1}^p$. Consequently, from Definition 2.2, $\sigma^k = \sigma^{k+1}$.

Suppose then that $W_k \neq W_{k+1}$ and consider

$$g = \frac{W_{k+1}^{-1/2}\left(W_{k+1}^{1/2}Y - W_{k+1}^{1/2}A\hat{X}_{k+1}^p\right)}{\left[\left(Y - A\hat{X}_{k+1}^p\right)^T W_{k+1}\left(Y - A\hat{X}_{k+1}^p\right)\right]^{1/2}} .$$

$$= \frac{W_{k+1}^{-1/2}\left(Y_{k+1}^* - A_{k+1}^*\hat{X}_{k+1}^p\right)}{\left[\left(Y_{k+1}^* - A_{k+1}^* X_{k+1}^p\right)^T\left(Y_{k+1}^* - A_{k+1}^*\hat{X}_{k+1}^p\right)\right]^{1/2}}$$

where $Y_{k+1}^* = W_{k+1}^{1/2}Y$, $A_{k+1}^* = W_{k+1}^{1/2}A$ and $W_{k+1}^{1/2}$ is defined in Definition 1.9. Note also that

$$\hat{X}_{k+1}^p = \left(A^T W_{k+1} A\right)^+ A^T W_{k+1} Y + \left[I - \left(W^{1/2}A\right)^+\left(W^{1/2}A\right)\right] Z$$

$$= \left(A_{k+1}^{*T} A_{k+1}^*\right)^+ A_{k+1}^{*T} Y_{k+1}^* + \left[I - A^{*+}A^*\right] Z$$

so that the values $\hat{X}_{k+1}^p$ are the least squares estimates for $Y_{k+1}^* = A_{k+1}^* X$. From Lemma 1.11

a)

$$g^T W_{k+1} g = \frac{\left(Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p\right)^T W_{k+1}^{-1/2} W_{k+1} W_{k+1}^{-1/2} \left(Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p\right)}{\left(Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p\right)^T \left(Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p\right)}$$

$$= 1$$

since $Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p = W^{1/2}\left(Y - A\hat{X}_{k+1}^p\right)$, so that

$$\left(W_{k+1}^{1/2} W_{k+1}^{-1/2}\right) W_{k+1}^{1/2}\left(Y - A\hat{X}_{k+1}^p\right) = W_{k+1}^{1/2}\left(Y - A\hat{X}_{k+1}^p\right)$$

as $W_{k+1}^{-1/2}$ is the pseudoinverse of $W^{1/2}$ by the remarks after Definition 1.9.

b)

$$A^T W_{k+1} g = \frac{A^T W_{k+1}^{1/2} W_{k+1}^{1/2} W_{k+1}^{-1/2}\left(Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p\right)}{\left[\left(Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p\right)^T \left(Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p\right)\right]^{1/2}}$$

$$= \frac{A_{k+1}^{*T}\left(Y_{k+1}^* - A_{k+1}^* X_{k+1}^p\right)}{\left[\left(Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p\right)^T \left(Y_{k+1}^* - A_{k+1}^* \hat{X}_{k+1}^p\right)\right]^{1/2}}$$

$$= \phi$$

showing g is orthogonal with weight W_{k+1} to the space of columns of A .

c) For any h orthogonal with weight W_{k+1} to the columns of A and $h^T W_{k+1} h = 1$, then defining $h^*_{k+1} = W^{1/2}_{k+1} h$, $g^*_{k+1} = W^{1/2}_{k+1} g$, we find

$$Y^T W_{k+1} h \;=\; Y^{*T}_{k+1} h^*_{k+1} \;\le\; Y^{*T}_{k+1} g^*_{k+1} \;=\; Y^T W_{k+1} g$$

by Lemma 1.11 so that g maximizes $Y^T W_{k+1} h$, for $h^T W_{k+1} h = 1$ and $A^T W_{k+1} h = \phi$.

Now consider

$$h \;=\; \frac{W^{-1}_{k+1} W_k \left(Y - A\hat{X}^p_k\right)}{\left[\left(Y - A\hat{X}^p_k\right)^T W_k W^{-1}_{k+1} W_k \left(Y - A\hat{X}^p_k\right)\right]^{1/2}}$$

which satisfies

a)

$$h^T W_{k+1} h \;=\; \frac{\left(Y - A\hat{X}^p_k\right)^T W_k W^{-1}_{k+1} W_{k+1} W^{-1}_{k+1} W_{k+1} \left(Y - A\hat{X}^p_k\right)}{\left(Y - A\hat{X}^p_k\right)^T W_k W^{-1}_{k+1} W_k \left(Y - A\hat{X}^p_k\right)}$$

$$= \;1$$

since W^{-1}_{k+1} is the pseudoinverse of W_{k+1} by the remarks after Definition 1.9.

b) Using Definition 1.9, note that the *ith* diagonal element of $W_{k+1} W^{-1}_{k+1}$ is 1, if $w^i_{k+1} > 0$

and 0 if $w_{k+1}^{i} = 0$. Thus the *ith* diagonal element of $W_{k+1}W_{k+1}^{-1}W_k$ is either w_k^i if $w_{k+1}^i > 0$ or 0 if $w_{k+1}^i = 0$. Then

$$A^T W_k\left(Y - A\hat{X}_k^p\right) = A^{*T}\left(Y^* - A^*\hat{X}_k^p\right)$$

$$= \phi$$

by Lemma 1.11 d, and since the *ith* element of $W_k\left(Y - A\hat{X}_k^p\right)$ is simply $w_k^i\left(Y_i - A^i X_k^p\right)$, which is the *ith* element of $W_{k+1}W_{k+1}^{-1}\left(Y - A\hat{X}_k^p\right)$ if $w_{k+1}^i > 0$. If $w_k^i = 0$, then the corresponding elements are trivially equal for any value of w_{k+1}^i as $w_k^i\left(Y_i - A^i\hat{X}_k^p\right) = 0$. If $w_k^i > 0$ but $w_{k+1}^i = 0$, then by the remarks after Definition 2.2, $Y_i - A^i\hat{X}_k^p = 0$ so that the corresponding elements are equal. Therefore

$$W_{k+1}W_{k+1}^{-1}W_k\left(Y - A\hat{X}_k^p\right) = W_k\left(Y - A\hat{X}_k^p\right)$$

making

$$A^T W_{k+1}W_{k+1}^{-1}W_k\left(Y - A\hat{X}_k^p\right) = 0$$

and

$$A^T W_{k+1} h = \frac{A^T W_{k+1} W_{k+1}^{-1} W_k \left(Y - A\hat{X}_k^p\right)}{\left[\left(Y - A\hat{X}_k^p\right)^T W_k W_{k+1}^{-1} W_k \left(Y - A\hat{X}_k^p\right)\right]^{1/2}}$$

$$= 0$$

so that $Y^T W_{k+1} h \leq Y^T W_{k+1} g$.

Let

$$D_1 = \left[S\left(\hat{X}_k^p\right)^2 W_k\right]^{p/2(p-1)}$$

$$D_2 = W_k^{p/2(p-1)}$$

$$r = 2(p - 1)/p$$

$$s = 2(p - 1)/(p - 2)$$

and noting that $(1/r) + (1/s) = 1$, calculate the following expressions using the definition of W_{k+1} :

a) $$\mathrm{tr}\left[W_{k+1}^{p/(p-2)}\right]^{(p-2)/2p}$$

$$= \frac{\mathrm{tr}\left\{\left[W_k S\left(\hat{X}_k^p\right)\right]^{p/(p-1)}\right\}^{(p-2)/2p}}{\mathrm{tr}\left\{\left[W_k S\left(\hat{X}_k^p\right)\right]^{(p-2)/(p-1)}\right\}^{1/2}}$$

b) $$\left[\left(Y - A\hat{X}_k^p\right)^T W_k W_{k+1}^{-1} W_k \left(Y - A\hat{X}_k^p\right)\right]^{1/2}$$

$$= \mathrm{tr}\left[W_{k+1}^{-1} W_k^2 S\left(\hat{X}_k^p\right)^2\right]^{1/2}$$

$$= \mathrm{tr}\left\{W_k^2 S\left(\hat{X}_k^p\right)^2 \left[W_k S\left(\hat{X}_k^p\right)\right]^{-(p-2)/p-1}\right.$$

$$\left.\cdot\ \mathrm{tr}\left[\left(W_k S\left(\hat{X}_k^p\right)\right)^{(p-2)/(p-1)}\right]\right\}^{1/2}$$

$$= \mathrm{tr}\left\{\left[W_k S\left(\hat{X}_k^p\right)\right]^{p/(p-1)}\right\}^{1/2}$$

$$\cdot\ \mathrm{tr}\left\{\left[W_k S\left(\hat{X}_k^p\right)\right]^{(p-2)/(p-1)}\right\}^{1/2}$$

c) $$\left[\left(Y - A\hat{X}_k^p\right)^T W_k W_{k+1}^{-1} W_k \left(Y - A\hat{X}_k^p\right)\right]^{1/2}$$

$$\cdot\ \mathrm{tr}\left[W_{k+1}^{p/(p-2)}\right]^{(p-2)/2p}$$

$$= \mathrm{tr}\left\{\left[W_k S\left(\hat{X}_k^p\right)\right]^{p/(p-1)}\right\}^{1/2}$$

$$\cdot\ \mathrm{tr}\left\{\left[W_k S\left(\hat{X}_k^p\right)\right]^{p/(p-1)}\right\}^{(p-2)/2p}$$

$$= \mathrm{tr}\left\{\left[W_k S\left(\hat{X}_k^p\right)\right]^{p/(p-1)}\right\}^{(p-1)/p}$$

$$= \mathrm{tr}\left\{\left[S\left(\hat{X}_k^p\right)^{p/(p-1)} W_k^{p/2(p-1)}\right]\right.$$

$$\left.\cdot\ \left[W_k^{p/2(p-1)}\right]\right\}^{(p-1)/p}$$

$$= \mathrm{tr}\left\{D_1 D_2\right\}^{(p-1)/p}$$

$$\leq \left\{\mathrm{tr}\left[D_1^r\right]^{1/r} \mathrm{tr}\left[D_2^s\right]^{1/s}\right\}^{(p-1)/p}$$

$$= \left\{\mathrm{tr}\left[S\left(\hat{X}_k^p\right)^2 W_k\right]^{p/2(p-1)}\right.$$

$$\left.\cdot\ \mathrm{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/2(p-1)}\right\}^{(p-1)/p}$$

$$= \mathrm{tr}\left[W_k S\left(\hat{X}_k^p\right)^2\right]^{1/2} \mathrm{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/2p}$$

by Lemma 1.12 (Holder Inequality in matrix trace notation). By the same lemma, equality holds if and only if there is an α such that

$$D_1^r = \alpha D_2^s$$

or that

$$W_k S\left(\hat{X}_k^p\right)^2 = \alpha W_k^{p/(p-2)}$$

or

$$S\left(\hat{X}_k^p\right) = \alpha^{1/2} W_k^{1/(p-2)} \quad . \tag{2.4}$$

But then

$$W_{k+1} = \left[W_k S\left(\hat{X}_k^p\right)\right]^{(p-2)/(p-1)} \Big/ tr\left[W_k S\left(\hat{X}_k^p\right)\right]^{(p-2)/(p-1)}$$

$$= W_k / tr\left\{W_k\right\}$$

$$= W_k$$

since

$$\operatorname{tr}\left[W_{k+1}\right] = \operatorname{tr}\left\{\left[W_k S\left(\hat{X}_k^p\right)\right]^{(p-2)/(p-1)} \Big/ \operatorname{tr}\left[\left(W_k S\left(\hat{X}_k^p\right)^{(p-2)/(p-1)}\right)\right]\right\}$$

$$= 1 .$$

But equality contradicts the hypothesis of this case which is $W_{k+1} \neq W_k$. Consequently the inequality is strict.

Therefore, using Lemma 1.11 d) and the above results, we have

$$\begin{aligned}
\sigma^{k+1} &= \left[\left(Y - A\hat{X}_{k+1}^p\right)^T W_{k+1}\left(Y - A\hat{X}_{k+1}^p\right)\right]^{1/2} \Big/ \operatorname{tr}\left[W_{k+1}^{p/(p-2)}\right]^{(p-2)/2p} \\
&= \left[\left(Y_{k+1}^* - A_{k+1}^*\hat{X}_{k+1}^p\right)^T\left(Y_{k+1}^* - A_{k+1}^*\hat{X}_{k+1}^p\right)\right]^{1/2} \Big/ \operatorname{tr}\left[W_{k+1}^{p/(p-2)}\right]^{(p-2)/2p} \\
&= \frac{Y_{k+1}^{*T}\left(Y_{k+1}^* - A_{k+1}^*\hat{X}_{k+1}^p\right)}{\left[\left(Y_{k+1}^* - A_{k+1}^*\hat{X}_{k+1}^p\right)^T\left(Y_{k+1}^* - A_{k+1}^*\hat{X}_{k+1}^p\right)\right]^{1/2} \operatorname{tr}\left[W_{k+1}^{p/(p-2)}\right]^{(p-2)/2p}} \\
&= Y_{k+1}^{*T} g_{k+1}^* \Big/ \operatorname{tr}\left[W_{k+1}^{p/(p-2)}\right]^{(p-2)/2p} \\
&\geq Y_{k+1}^{*T} h_{k+1}^* \Big/ \operatorname{tr}\left[W_{k+1}^{p/(p-2)}\right]^{(p-2)/2p} \\
&= \frac{Y^T W_{k+1} W_{k+1}^{-1} W_k\left(Y - A\hat{X}_k^p\right)}{\left[\left(Y - A\hat{X}_k^p\right)^T W_k W_{k+1}^{-1} W_k\left(Y - A\hat{X}_k^p\right)\right]^{1/2} \operatorname{tr}\left[W_{k+1}^{p/(p-2)}\right]^{(p-2)/2p}} \\
&> \frac{Y^T W_k\left(Y - A\hat{X}_k^p\right)}{\operatorname{tr}\left[W_k S\left(\hat{X}_k^p\right)^2\right]^{1/2} \operatorname{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/2p}}
\end{aligned}$$

and using calculation c),

$$\sigma^{k+1} > \frac{Y^T W_k\left(Y - A\hat{X}_k^p\right)}{\left[\left(Y - A\hat{X}_k^p\right)^T W_k \left(Y - A\hat{X}_k^p\right)\right]^{1/2} tr\left[W_k^{p/(p-2)}\right]^{(p-2)/2p}}$$

$$= \left[\left(Y - A\hat{X}_k^p\right)^T W_k \left(Y - A\hat{X}_k^p\right)\right]^{1/2} \Big/ tr\left[W_k^{p/(p-2)}\right]^{(p-2)/2p}$$

$$= \sigma^k \quad . \tag{2.5}$$

LEMMA 2.5: *Let* $\hat{X}^p$ *be the best* ℓ^p *approximation to* X *on* A *. Then*

$$\sigma^k \leq \xi^p = tr\left[S\left(\hat{X}^p\right)^p\right]^{1/p} \quad .$$

Proof: Since $\hat{X}_k^p$ is the best least squares approximation with weights W_k , then

$$tr\left[W_k S\left(\hat{X}_k^p\right)^2\right] \leq tr\left[W_k S\left(\hat{X}^p\right)^2\right] \quad .$$

Letting

$$D_1 = W_k$$

$$D_2 = S\left(\hat{X}^p\right)^2$$

$$r = p/(p - 2)$$

$$s = p/2$$

and noting that $(1/r) + (1/s) = 1$, we find

$$\begin{aligned}
\left(\sigma^k\right)^2 &= \operatorname{tr}\left[W_k S\left(\hat{X}_k^p\right)^2\right] \Big/ \operatorname{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/p} \\
&\le \operatorname{tr}\left[W_k S\left(\hat{X}^p\right)^2\right] \Big/ \operatorname{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/p} \\
&= \operatorname{tr}\left[D_1 D_2\right] \Big/ \operatorname{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/p} \\
&\le \operatorname{tr}\left[D_1^r\right]^{1/r} \operatorname{tr}\left[D_2^s\right]^{1/s} \Big/ \operatorname{tr}\left[W_k^{p/(p-2)}\right]^{(p-1)/p} \\
&= \operatorname{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/p} \\
&\quad \cdot \operatorname{tr}\left\{\left[S\left(\hat{X}^p\right)^2\right]^{p/2}\right\}^{2/p} \Big/ \operatorname{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/p} \\
&= \operatorname{tr}\left[S\left(\hat{X}^p\right)^p\right]^{2/p} \\
&= \left(\xi^p\right)^2 .
\end{aligned}$$

Since $\{\sigma^k\}$ is bounded above and monotone, it has a limit. Denote this by

$$\sigma^* = \lim_{k\to\infty} \sigma^k$$

and since $\phi \le W_k \le I$, elementwise for all k , the sequence $\{W_k\}$ is also bounded, so that there exists a sub-sequence of $\{W_k\}$, say $W_{k(i)}$, which converges, say to W_0 . Also the sequence of index sets $\{J_k\}$ is

monotonically decreasing as in the remarks prior to Lemma 2.4 and hence converges, say to J_0 . Note that J_0 is not null since the sequence σ^k is monotonically increasing, and J_0 null would imply $\sigma^* = 0$. Furthermore, since $J_k \supset J_{k+1}$ for all k , each convergent sub-sequence must have the same limiting index set J_0 .

<u>LEMMA 2.6</u>: *Let* $\hat{X}^p_0$ *be the best least squares approximation to* X *with weights* W_0 . *Then* $\sigma^0 > 0$ *and*

$$\lim_{k(i)\to\infty} \hat{X}^p_{k(i)} = \hat{X}^p_0 \quad . \qquad (2.6)$$

Proof: As was mentioned in the remarks prior to Lemma 2.4, there must exist a number K such that for all $k > K$, $J_k = J_0$. Thus

$$\begin{aligned} r\left(A^T W_k A\right) &= r\left[\left(A^T W_k A\right)^T \left(A^T W_k A\right)\right] \\ &= r\left[\left(A^T W_k A\right)\left(A^T W_k A\right)\right] \\ &= r \end{aligned}$$

for $k > K$, where $r(\cdot)$ is the rank of the matrix.

Further

$$
\begin{aligned}
\left(A^TWA\right)\left(A^TWA\right)^+ &= \left[\left(A^TWA\right)\left(A^TWA\right)^+\right]^T \\
&= \left(A^TWA\right)^{+T}\left(A^TWA\right)^T \\
&= \left(A^TWA\right)^+\left(A^TWA\right)
\end{aligned}
$$

so that $\left(A^TWA\right)^+$ commutes with (A^TWA). Thus for any W_k such that $\{i:w^i > 0\} = J_0$, and since A^TWA is differentiable in each w^i, then $\left(A^TWA\right)^+$ is also differentiable showing $\left(A^TWA\right)^+$ continuous in the weights W [19]. Consequently, $\left(A^TW_{k(i)}A\right)^+$ is continuous in the weights for $k(i) > K$ and therefore

$$
\begin{aligned}
\hat{X}^p_{k(i)} &= \left(A^TW_{k(i)}A\right)^+A^TW_{k(i)}Y \\
&\quad + \left[I - \left(W^{1/2}_{k(i)}A\right)\left(W^{1/2}_{k(i)}A\right)^+\right]Z
\end{aligned}
$$

is continuous in the weights for $k(i) > K$ implying equation (2.6). Also, $Y - A\hat{X}^p_{k(i)}$ is continuous in the weights so that $\sigma^{k(i)}$ is continuous in the weights by Definition 2.2. Therefore

$$
\begin{aligned}
0 < \sigma^1 < \sigma^* &= \lim_{k\to\infty} \sigma^k \\
&= \lim_{k(i)\to\infty} \sigma^{k(i)} = \sigma^0 \ .
\end{aligned}
$$

LEMMA 2.7: $\lim_{k\to\infty} W_k = W_0$

Proof: $\lim_{k\to\infty} \sigma^k = \sigma^0$ implies

$\lim_{k\to\infty} (\sigma^{k+1} - \sigma^k) = 0$. Now observing the proof of Lemma 2.4, it must be true that the two inequalities of equation (2.5) approach equality. Now the second inequality is a result of the Holder Inequality

$$\mathrm{tr}\left\{\left[S\left(\hat{X}_k^p\right)^{p/(p-1)} W^{p/2(p-1)}\right]\left[W^{p/2(p-1)}\right]\right\}$$

$$\leq \mathrm{tr}\left[S\left(\hat{X}_k^p\right)^2 W_k\right]^{p/2(p-1)} \mathrm{tr}\left[W_k^{p/(p-2)}\right]^{(p-2)/2p}$$

or using D_1, D_2, r , and s as defined in the proof, we have

$$\mathrm{tr}\left\{D_1 D_2\right\} \leq \mathrm{tr}\left\{D_1^{2(p-1)/p}\right\}^{p/2(p-1)}$$

$$\cdot\ \mathrm{tr}\left\{D_2^{s(p-1)/(p-2)}\right\}^{(p-2)/2(p-1)}$$

$$= \mathrm{tr}\left\{D_1^r\right\}^{1/r} \mathrm{tr}\left\{D_2^s\right\}^{1/s} \quad .$$

Therefore, $\lim (\sigma^{k+1} - \sigma^k) = 0$ implies

$$\lim\left[\mathrm{tr}\left\{D_1 D_2\right\} - \mathrm{tr}\left\{D_1^r\right\}^{1/r} \mathrm{tr}\left\{D_2^s\right\}^{1/s}\right] = 0$$

which further implies the equality condition

$$\lim \left[D_1^r - \alpha D_2^s \right] = \phi$$

for a real constant α , which is

$$\lim \left[S\left(\hat{X}_k^p\right) - \alpha^{1/2} W_k^{1/(p-2)} \right] = \phi$$

from equation (2.4). Consequently recalling that $tr\{W_k\} = 1$, we obtain

$$\lim\left\{W_{k+1} - W_k\right\} = \lim \left\{ \frac{\left[W_k S\left(\hat{X}_k^p\right)\right]^{(p-2)/(p-1)}}{tr\left\{\left[W_k S\left(\hat{X}_k^p\right)\right]^{(p-2)/(p-1)}\right\}} - W_k \right\}$$

$$= \lim \left\{ \frac{\left[W_k \alpha^{1/2} W_k^{1/(p-2)}\right]^{(p-2)/(p-1)}}{tr\left\{\left[W_k \alpha^{1/2} W^{1/(p-2)}\right]^{(p-2)/(p-1)}\right\}} - W_k \right\}$$

$$= \lim \left\{ \frac{\alpha^{1/2}\left[W_k^{(p-1)/(p-2)}\right]^{(p-2)/(p-1)}}{\alpha^{1/2} tr\left\{\left[W_k^{(p-1)/(p-2)}\right]^{(p-2)/(p-1)}\right\}} - W_k \right\}$$

$$= \lim\left\{W_k / tr\left\{W_k\right\} - W_k\right\} = \phi \quad .$$

This implies that $\{W_k\}$ is a Cauchy sequence, and since one sub-sequence converges to W_0, then $\lim W_k = W_0$.

<u>LEMMA 2.8</u>: *The least squares estimate* $\hat{X}_0^p$ *is also the best* ℓ^p *approximation to* X *with indices in* J_0 *and*

$$\sigma^0 = \operatorname{tr}\left[I_0 S\left(\hat{X}_0^p\right)^p\right]^{1/p} \qquad (2.7)$$

where $I_0 = \mathit{diag}\left[d_i = 1 \text{ if } i \varepsilon J_0 \,;\ d_i = 0 \text{ otherwise}\right]$.

Proof: Observe from Lemma 2.6 that σ^{k+1} is a continuous function of the weights W_k, and denote this fact by $F(W_k) = \sigma^{k+1}$. Let $\{W_k\}$ correspond with $\{y_n\}$ and $\{\sigma^k\}$ with $\{x_n\}$ of Lemma 1.13. Recall that $\lim \sigma^k = \sigma^0$ and $\lim W_k = W_0$, so that $F(W_0) = \sigma^0$ by Lemma 1.13.

Now suppose the algorithm were restarted with W_0 as the initial weights. Then $\hat{X}_0^p$ and σ^0 are the estimates for the weights W_0. For the first actual iteration using the weights,

$$W_{1,1} = \left[W_0 S\left(\hat{X}_0^p\right)\right]^{(p-2)/(p-1)} \Bigg/ \operatorname{tr}\left\{\left[W_0 S\left(\hat{X}_0^p\right)\right]^{(p-2)/(p-1)}\right\}$$

with $\hat{X}^p_{1,1}$ and σ^1_1 corresponding to these weights. Since the function F relating the weights at iteration k with σ^{k+1} is the same as for the first sequence $\{W_k\}$, then $\sigma^1_1 = F(W_0) = \sigma_0$, and therefore $W_{1,1} = W_0$ by Lemma 2.4. Consequently

$$W_0 = \left[W_0 S\left(X^p_0\right)\right]^{(p-2)/(p-1)} \Big/ \operatorname{tr}\left\{\left[W_0 S\left(X^p_0\right)\right]^{(p-2)/(p-1)}\right\}$$

$$= W_0^{(p-2)/(p-1)} S\left(X^p_0\right)^{(p-2)/(p-1)} \Big/ \operatorname{tr}\left\{\left[W_0 S\left(X^p_0\right)\right]^{(p-2)/(p-1)}\right\}$$

so that multiplying both sides by $W^{-(p-2)/(p-1)}$, noting that

$$W^{(p-2)/(p-1)} W^{-(p-2)/(p-1)} = I_0$$

and using the remarks after Definition 1.9, we have

$$W_0^{1/(p-1)} = I_0 S\left(\hat{X}^p_0\right)^{(p-2)/(p-1)} \operatorname{tr}\left\{\left[W_0 S\left(\hat{X}^p_0\right)\right]^{(p-2)/(p-1)}\right\}$$

or

$$W_0 = I_0 S\left(\hat{X}^p_0\right)^{(p-2)} \operatorname{tr}\left\{\left[W_0 S\left(\hat{X}^p_0\right)\right]^{(p-2)/(p-1)}\right\} \quad (2.8)$$

and since, by Lemma 1.11 d), $\hat{X}^p_0$ is the best least squares approximation with weights W_0, the normal equations for any X are

$$
\begin{aligned}
(AX)^T W_0\left(Y - A\hat{X}_0^p\right) &= X^T A^T W_0^{1/2} W_0^{1/2}\left(Y - A\hat{X}_0^p\right) \\
&= X^T A^{*T}\left(Y^* - A^*\hat{X}_0^p\right) \\
&= X^T\left(A^{*T}Y - A^{*T}A^*\hat{X}_0^p\right) \\
&= \phi
\end{aligned}
$$

so that using equation (2.8) and multiplying out the denominator, we obtain

$$
\begin{aligned}
0 &= \left(A\hat{X}_0^p\right)^T I_0 S\left(\hat{X}_0^p\right)^{(p-2)}\left(Y - A\hat{X}_0^p\right) \\
&= \left(A^1\hat{X}_0^p, \cdots, A^m\hat{X}_0^p\right) I_0 \begin{pmatrix} |Y_1 - A^1\hat{X}_0^p|^{p-2} \cdot \left(Y_1 - A^1\hat{X}_0^p\right) \\ \vdots \\ |Y_m - A^m\hat{X}_0^p|^{p-2} \cdot \left(Y_1 - A^m\hat{X}_0^p\right) \end{pmatrix} \\
&= \sum_{i \varepsilon J_0} |Y_i - A^i\hat{X}_0^p|^{(p-2)}\left(Y_i - A^i\hat{X}_0^p\right) A^i\hat{X}_0^p \\
&= \sum_{i \varepsilon J_0} |Y_i - A^i\hat{X}_0^p|^{(p-1)} \operatorname{sgn}\left(Y_i - A^i\hat{X}_0^p\right) A^i\hat{X}_0^p \\
&= -\sum_{i \varepsilon J_0} \frac{\partial}{\partial X} |Y_i - A^i X|^p \Bigg|_{X = X_0^p}
\end{aligned} \tag{2.9}
$$

where $\operatorname{sgn}(x) = 1$ if $x \geq 0$ and -1 if $x < 0$ is the signum function. Therefore $\hat{X}_0^p$ minimizes

$$\sum_{i \varepsilon J_0} |Y_i - A^i X|^p = \operatorname{tr}\left[S(X)^p I_0\right]$$

and consequently $\hat{X}_0^p$ is the best ℓ^p estimate with indices in J_0 .

To establish equation (2.7), use (2.8) and calculate

$$\sigma^0 = \operatorname{tr}\left[W_0 S\left(\hat{X}_0^p\right)^2\right]^{1/2} \Big/ \operatorname{tr}\left[W_0^{p/(p-2)}\right]^{(p-2)/2p}$$

$$= \frac{\operatorname{tr}\left[I_0 S\left(\hat{X}_0^p\right)^p\right]^{1/2} \Big/ \operatorname{tr}\left\{\left[W_0 S\left(\hat{X}_0^p\right)\right]^{(p-2)/(p-1)}\right\}^{1/2}}{\operatorname{tr}\left[I_0 S\left(\hat{X}_0^p\right)^p\right]^{(p-2)/2p} \Big/ \operatorname{tr}\left\{\left[W_0 S\left(\hat{X}_0^p\right)^{(p-2)/(p-1)}\right\}^{1/2}}$$

$$= \operatorname{tr}\left[I_0 S\left(\hat{X}_0^p\right)^p\right]^{1/2-(p-2)/2p}$$

$$= \operatorname{tr}\left[I_0 S\left(\hat{X}_0^p\right)\right]^{1/p} .$$

It should be noted here that the proof in reference [46] of Lemma 2.7 is incorrect and of Lemma 2.8 is incomplete and incorrect as stated, even for the considerably greater hypotheses which they impose upon the model. Lemma 2.7 shows that the

convergent sub-sequence $\{W_{k(i)}\}$ is in reality the whole sequence so that Lemma 2.6 implies $\lim \hat{X}^p_k = \hat{X}^p_0$. The lemmas and this fact can be summarized for easy reference in the theorem which follows.

THEOREM 2.9: *The algorithm defined in Definition 2.2 has the following properties:*

a) $J_k \supset J_{k+1}$ *for all* k *and* $\lim J_k = J_0$, *a nonempty set*

b) $\lim W_k = W_0$, $\operatorname{tr} W_0 = 1$

c) $\lim \hat{X}^p_k = \hat{X}^p_0$, *the best* ℓ^p *approximation on the set of nonzero weights* J_0

d) $\lim \sigma^k = \operatorname{tr}\left[I_0 S\left(\hat{X}^p_0\right)^p\right]^{1/p}$.

Observe that the above theorem merely states that $\hat{X}^p_0$ is the best ℓ^p estimate on J_0 , not $\{1,2,\cdots,m\} = M$. If $J_0 = M$, there is no difficulty. If $J_0 \neq M$, then $J_0 \subset M$, and $\sigma^0 < \sigma^0_p$ where σ^0_p is the value of σ for the ℓ^p approximation on M , since the approximation over a submatrix of A will have a smaller error than an approximation over all of A . This suggests restarting the algorithm with an index set $J_{1,1} \supset J_0$ and a $\sigma^1_1 > \sigma^0$.

LEMMA 2.10: *If* $Y_{i*} = A^{i*}\hat{X}^P(\lambda)$ *for some* $\lambda = \lambda_1 \neq 0,1$, *then the equality holds for all* λ *where* $\hat{X}^P(\lambda)$ *is the least squares solution to* $Y = AX$ *with weight matrix*

$$W(\lambda,i^*) = (1-\lambda)W_0 + \lambda U(i^*)$$

where $U(i^*) = diag\,[U_{i*} = 1$ *for* $i^* \in M - J_0$ *and zero otherwise*].

Proof: Partition the system of equations $Y = AX$ into two parts

$$Y_{(1)} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} \quad ; \quad Y_{(2)} = Y_{i*}$$

$$A_{(1)} = \begin{bmatrix} A^1 \\ A^2 \\ \vdots \\ A^m \end{bmatrix} \quad ; \quad A_{(2)} = A^{i*}$$

$$W_{(1)} = (1-\lambda)diag[W_{0i}, i \neq i^*]$$

$$W_{(2)} = \lambda$$

where the partitions $Y_{(1)}$ and $A^{(1)}$ do not include Y_{i*} and A^{i*} , respectively.

This new system

$$\begin{bmatrix} Y_{(1)} \\ Y_{(2)} \end{bmatrix} = \begin{bmatrix} A^{(1)} \\ A^{(2)} \end{bmatrix} X$$

is now in the partitioned form required in Theorem 1.10. Equation 1.5 shows that $Y_{(2)} - A_{(2)}\hat{X} = \phi$ if and only if $Y_{(2)} - A_{(2)}\hat{X}_1 = \phi$, where $\hat{X}$ is the least squares solution over the complete system and $\hat{X}_1$ is the least squares solution over $Y_{(1)} = A^{(1)}X$ with weights $W_{(1)}$. The only requirement is that $W_{(2)}$ is nonsingular which is true if and only if $\lambda \neq 0$. Now

$$\begin{aligned} \hat{X}_1(\lambda) &= \left(A^T_{(1)} W_{(1)} A_{(1)}\right)^+ A^T_{(1)} W_{(1)} Y_{(1)} \\ &= \left(A^T(1 - \lambda) W_0 A\right)^+ A^T(1 - \lambda) W_0 Y \\ &= \left(A^T W_0 A\right)^+ A^T W_0 Y \\ &= \hat{X}^P_0 \end{aligned}$$

if $\lambda \neq 1$. Since $Y_{i*} = A^{i*}\hat{X}^P(\lambda_1)$, then $Y_{i*} - A^{i*}\hat{X}_1 = \phi$ for all λ so that $Y_{i*} = A^{i*}\hat{X}^P(\lambda)$ for all $\lambda \neq 1,0$. Observe also that $W_{(1)} = \phi$ when $\lambda = 1$ implying that $\hat{X} = (A^{i*})^+ Y_{i*}$ is such that $Y_{i*} = A^{i*}\hat{X}$.

THEOREM 2.11: *If* J_0 *is a proper subset of* M , *let*

$$W_1(\lambda, i^*) = (1 - \lambda)W_0 + \lambda U(i^*)$$

where $U(i^*) = diag\ [U_{i*} = 1$ *for* $i^* \varepsilon M - J_0$ *and zero otherwise*]. *If it is true that*

$$\mathrm{tr}\left[U(i^*)S\left(\hat{X}^P(\lambda)\right)\right] > 0 \ , \qquad (2.10)$$

for some λ_1, $0 < \lambda_1 \leq \lambda_0 < 1$ and $\hat{X}^P(\lambda)$ *the least squares solution using weights* $W_1(\lambda, i^*)$, *then the algorithm may be restarted with the weights* W_1 . *For* $\lambda_1 \leq \lambda \leq \lambda_0$, *then*

$$\sigma(\lambda) > \sigma^0$$

where

$$\sigma(\lambda) = \frac{\mathrm{tr}\left[W_1(\lambda, i^*)S\left(\hat{X}^P(\lambda)\right)^2\right]^{1/2}}{\mathrm{tr}\left[W_1(\lambda, i^*)^{P/(P-2)}\right]^{(P-2)/P}}$$

If equation (2.10) *holds for all* $i^* \varepsilon M - J_0$, *then the best* ℓ^P *approximation to* X *on* M *may be obtained after a finite number of restarts.*

Proof: Observe first that $\mathrm{tr}|W_0 S(\hat{X}^P_0)|$ is constant in λ and that

$$\lim_{\lambda \to 0^+} |Y_{i*} - A^{i*}\hat{X}^P(\lambda)| = |Y_{i*} - A^{i*}\hat{X}^P(0)|$$

$$= \mathrm{tr}\left[U(i^*)S(\hat{X}^P)\right]$$

and is therefore bounded since $\hat{X}^P(\lambda)$ is continuous in λ, say for $0 < \lambda < \lambda_m$. If $\mathrm{tr}[U(i^*)S(\hat{X}^P_0)] > 0$, let b be this bound. If $\mathrm{tr}[U(i^*)S(\hat{X}^P_0)] = 0$, then restrict $0 < \lambda_1 \le \lambda \le \lambda_0$ and set

$$b = \inf_{\lambda_1 < \lambda < \lambda_0} \mathrm{tr}\left[U(i^*)S\left(\hat{X}^P(\lambda)\right)\right]$$

Observe that since $\mathrm{tr}[U(i^*)S(\hat{X}^P(\lambda))]$ is nonzero for some λ_1, $0 < \lambda_1 \le \lambda_0$, then it is nonzero over the whole interval by the contrapositive of Lemma 2.10. Further since $(A^T W_1(\lambda, i^*)A)^+$ is continuous in λ by the proof of Lemma 2.6; then, in turn, $\hat{X}^P(\lambda)$, $S(\hat{X}^P(\lambda))$ and $\mathrm{tr}[U(i^*)S(\hat{X}^P(\lambda))]$ are continuous in λ implying that the infimum $b(\lambda_1)$ over the closed interval $[\lambda_1, \lambda_0]$ is attained and therefore $b(\lambda_1) > 0$.

By Lemma 1.14, if

$$x = \lambda/(1 - \lambda)$$

$$c = \mathrm{tr}\left(W_0^{p/(p-2)}\right)$$

$$r = p/(p - 2)$$

$$s = (p - 2)/p$$

then $F(x)$ approaches zero more rapidly than x^r , or

$$b(\lambda_1)x - \mathrm{tr}\left[W_0 S\ \hat{X}_0^p\right]F(x) > 0 \qquad (2.11)$$

If x is sufficiently small, which is for $0 < \lambda < \lambda_2$, say. Observe that equation (2.10) converges to zero slower than any positive power of x . Consequently, we can find a $\lambda < \lambda_2$ so that if $1 < \lambda < \min(\lambda_0, \lambda_2)$, equation (2.11) becomes

$$0 < \frac{b[x] - \mathrm{tr}\left[W_0 S\left(\hat{X}_0^p\right)\right]\left\{\left[1 + [x]^{p/(p-2)}\Big/\mathrm{tr}\left(W_0^{p/(p-2)}\right)\right]^{(p-2)/p} - 1\right\}}{\left[1 + [x]^{p/(p-2)}\Big/\mathrm{tr}\left(W_0^{p/(p-2)}\right)\right]^{(p-2)/p}}$$

$$< \frac{\mathrm{tr}\left(US\left(\hat{X}^p(\lambda)\right)^2\right) + (1 - \lambda)\mathrm{tr}\left(W_0 S\left(\hat{X}_0^p\right)^2\right)}{\left[(1 - \lambda)\mathrm{tr}\left(W_0^{p/(p-2)}\right) + \lambda^{p/(p-2)}\right]^{(p-2)/p}} - \frac{\mathrm{tr}\left(W_0 S\left(\hat{X}_0^p\right)^2\right)}{\mathrm{tr}\left[W_0^{p/(p-2)}\right]^{(p-2)/p}}$$

$$\leq \frac{\mathrm{tr}\left(US\left(\hat{X}^p(\lambda)\right)^2\right) - (1 - \lambda)\mathrm{tr}\left(W_0 S\left(\hat{X}^p(\lambda)\right)^2\right)}{\left[(1 - \lambda)\mathrm{tr}\left(W_0^{p/(p-2)}\right) + \lambda^{p/(p-2)}\right]^{(p-2)/p}} - \left(\sigma^0\right)^2$$

$$= \left(\sigma^1\right)^2 - \left(\sigma^0\right)^2$$

so that $\sigma^1 > \sigma^0$, since $\hat{X}_0^p$ is the least squares solution on W_0 , and therefore

$$\mathrm{tr}\left(W_0 S\left(\hat{X}^p(\lambda)\right)^2\right) \geq \mathrm{tr}\left(W_0 S\left(\hat{X}_0^p\right)^2\right)$$

for the third inequality. The restarted sequence σ^k is monotonically increasing so that it cannot converge to σ^0 . It must then converge on a set $J_0(\lambda) \supset J_0$ so that in a finite number of restarts the best ℓ^p approximation on M will be attained.

Rice [45] states that examples when restarts are required for his hypotheses are very difficult to construct. If a restart is necessary, one can determine whether a solution will occur either by reaching M as a nonzero index set or by the sequence σ^k diverging since it will no longer be bounded above by ξ^p of Lemma 2.5.

The case not yet discussed occurs when

$$Y_{i*} = A^{i*}\hat{X}^p(\lambda) \qquad (2.12)$$

for some $\lambda \neq 0,1$ and some $i^* \in M - J_0$. This may occur when either

a) the best ℓ^p approximation to X on M occurs when equation (2.12) holds. Now the

best ℓ^P approximation to X on M is a least squares approximation with appropriate weights. Without loss of generality, since (2.12) holds, let the i^* weight be nonzero. Therefore the weighted least square approximate (and therefore the best ℓ^P approximate) is the weighted least squares approximate over $M - \{i^*\}$ by Theorem 1.10. Therefore, $\hat{X}_0^P$ is the best ℓ^P approximation on M if equation (2.12) holds for all $i^* \varepsilon M - J_0$. Otherwise restart until equation (2.12) holds for all non-zero weight indices.

b) the approximate $\hat{X}_0^P$ is an local best ℓ^P approximation. Recall that the weight matrices W_k are a function of the initial weights W_1. It may be true that an intermediate or local solution which satisfies equation (2.12) exists. The only known method is to restart the algorithm with completely new weights. A method of logically choosing these weights is unknown.

3.3 Calculation of the p-q Generalized Inverse, $p, q \geq 2$

In this section an algorithm will be presented which calculates the p-q generalized inverse

$$B = (I - F)A^g E$$

of a matrix A, where E and F are metric projections onto $R(A)$ and $N(A)$ (see Section 2.3), respectively, for a degenerate linear model (see Section 1.1)

$$Y = AX$$

and where $Y \varepsilon U_m$, $X \varepsilon V_n$, $V_n = \ell^q(n)$, $V_m = \ell^p(m)$, and $p,q \geq 2$. The best approximate vector B(Y) (unique, since A is linear and ℓ^p spaces are strictly convex for $1 < p,q < \infty$ from Theorem 2.1 in Chapter II) satisfies

1. $\mathrm{tr}\left[S\left(B(Y)\right)^p\right]^{1/p} < \mathrm{tr}\left[S(X)^p\right]^{1/p}$

2. If there is an X such that

$$\mathrm{tr}\left[S\left(B(Y)\right)^p\right]^{1/p} = \mathrm{tr}\left[S(X)^p\right]^{1/p}$$

then

$$\mathrm{tr}\left[D\left(B(Y)\right)^q\right]^{1/q} \leq \mathrm{tr}\left[D(X)^q\right]^{1/q}$$

for all $X \varepsilon V_n$ and $D(X) = \mathit{diag}\left[X_i\right.$, the *ith* element of $\left.X\right]$.

<u>LEMMA 3.1</u>: $E(Y) = A\hat{X}^p$ *where* $\hat{X}^p$ *is the best* ℓ^p *approximation to* X.

Proof: If $\hat{X}^p$ is the best ℓ^p approximation to X, then for any $X^* \varepsilon V_n$,

$$\left[\sum_{i=1}^{m}\left(Y_i - A^i\hat{X}_i^p\right)^p\right]^{1/p} = tr\left[S\left(\hat{X}^p\right)^p\right]^{1/p}$$

$$\leq tr\left[S\left(X^*\right)^p\right]^{1/p}$$

$$= \left[\sum_{i=1}^{m}\left(Y_i - A^iX_i^*\right)^p\right]^{1/p} .$$

Now $\{Y:Y = AX,\ X \varepsilon V_n\} = R(A)$, so the above inequality shows that the vector $A\hat{X}^p$ is the vector in V_m which minimizes the ℓ^p norm over all vectors in $R(A)$. Therefore $A\hat{X}^p$ is the ℓ^p or metric projection of Y on $R(A)$, and consequently $E(Y) = A\hat{X}^p$.

LEMMA 3.2: $\{X:AX = E(Y)\} = \{X:X = \hat{X}^p - N,\ N \varepsilon N(A)\}$ $= \left\{X:X = \left(A^TW_0A\right)^+A^TW_0Y - N,\ N \varepsilon N(A)\right\}$.

Proof: Observe that since $A\hat{X}^p = E(Y)$, $\hat{X}^p \varepsilon \{X:AX = E(Y)\} = P(Y)$. If $X \varepsilon P(Y)$, write $X = \hat{X}^p - N$, then

$$AX = A\hat{X}^p - AN$$

$$= E(Y)$$

so that $AN = \phi$, and therefore $N \varepsilon N(A)$. Now if W_0 is the diagonal matrix of weights and $\hat{X}^p$ is the weighted least squares approximation as in Section 3.2, then

$$\hat{X}^p = \left(A^T W_0 A\right)^+ A^T W_0 Y + \left[I - \left(W_0^{1/2} A\right)^+ \left(W_0^{1/2} A\right)\right] Z$$

by equation (2.2) where Z is an arbitrary $n \times 1$ vector. Let $Z = \phi$; then $A\hat{X}^p = E(Y)$ still, and the proof above holds for this case.

Observe from Theorem 2.1 in Chapter II that the calculation $A^g E(Y)$ is to obtain an element X of V_n such that $AX = E(Y)$. Clearly $\left(A^T W_0 A\right)^+ A^T W_0 Y$ is such a vector with the set of all possible solutions being of the form $\left(A^T W_0 A\right)^+ A^T W_0 Y - N$, $N \varepsilon N(A)$.

LEMMA 3.3: *If* $r\left(N(A)\right) = r$, *then there exists an* $n \times r$ *matrix* C *such that for any* $N \varepsilon N(A)$ *there is a* $Z \varepsilon V_r$ *such that* $N = CZ$.

Proof: Let $\{b_i\}_{i=1}^r$ be a basis for $N(A)$. Then for any $N \varepsilon N(A)$, there are constants $\{Z_i\}_{i=1}^r$ such that

$$N = Z_1 b_1 + Z_2 b_2 + \cdots + Z_r b_r$$

$$= Z_1 (b_{11}, \cdots, b_{m1}) + \cdots + Z_r (b_{1r}, \cdots, b_{mr})$$

$$= \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mr} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_r \end{pmatrix}$$

$$= CZ \ .$$

Now from the set of all possible solutions,
$X = \left(A^T W_0 A\right)^+ A^T W_0 Y - CZ$, $Z \in V_r$, select that
$X_0 = \left(A^T W_0 A\right)^+ A^T W_0 Y - C\hat{Z}^q$ such that

$$\operatorname{tr}\left[D\left(X_0\right)^q\right]^{1/q} = \operatorname{tr}\left\{D\left[\left(A^T W_0 A\right)^+ A^T W_0 Y - C\hat{Z}^q\right]^q\right\}^{1/q}$$

$$\leq \operatorname{tr}\left\{D\left[\left(A^T W_0 A\right)^+ A^T W_0 Y - CZ^*\right]^q\right\}^{1/q}$$

$$= \operatorname{tr}\left[D\left(X^*\right)^q\right]^{1/q}$$

for any $Z^* \in V_r$. But this is simply the ℓ^p problem calculated in Section 3.2 with q replacing p , C replacing A , Z replacing X , and

$\left(A^T W_0 A\right)^+ A^T W_0 Y$ replacing Y. Consequently, the algorithm developed in Section 3.2 will work for this calculation yielding a $\hat{Z}^q$ $\left(\text{since } N(C) = \phi\right)$, the best ℓ^q estimate for Z and the best least squares estimate with weights, say U_0. Therefore

$$\hat{Z}^q = \left(C^T U_0 C\right)^+ C^T U_0 \left(A^T W_0 A\right)^+ A^T W_0 Y$$

and the solution

$$\begin{aligned} X_0 &= \left(A^T W_0 A\right)^+ A^T W_0 Y - C\hat{Z}^q \\ &= \left(A^T W_0 A\right)^+ A^T W_0 Y - C\left(C^T U_0 C\right)^+ C^T U_0 \left(A^T W_0 A\right)^+ A^T W_0 Y \\ &= \left[I - C\left(C^T U_0 C\right)^+ C^T U_0\right] \left(A^T W_0 A\right)^+ A^T W_0 Y \quad . \end{aligned}$$

The above results can be summarized in the following theorem.

<u>THEOREM 3.4</u>: *For* C *defined as above,*

a) $E(Y) = A\left(A^T W_0 A\right)^+ A^T W_0 Y$, *where* W_0 *is the diagonal matrix of least squares weights associated with the best* ℓ^p *approximation of* X *in the linear approximation* $Y = AX$.

b) $F(X) = C\left(C^T U_0 C\right)^+ C^T U_0 X$, *where* U_0 *is the diagonal matrix of least squares weights associated*

with the ℓ^q *approximation of* Z *in the linear approximation* $X = CZ$.

c) *The p-q generalized inverse* , B *of* A *is*

$$B(Y) = \left[I - C\left(C^T U_0 C\right)^+ C^T U_0\right]\left(A^T W_0 A\right)^+ A^T W_0 Y \tag{3.1}$$

where W_0 *is as in* a) *and* U_0 *is as in* b) *with* $X = \left(A^T W_0 A\right)^+ A^T W_0 Y$.

It should be observed here that B is not necessarily a linear operator since both U_0 and W_0 depend on the vector approximated. If $U_0(\cdot)$ and $W_0(\cdot)$ denote this dependence, then equation (3.1) could be written in functional notation as

$$\begin{aligned} B(\cdot) = & \left\{I - C\left(C^T U_0\left[\left(A^T W_0(\cdot)A\right)^+ A^T W_0(\cdot)\right]C\right)^+ \times \right. \\ & \left. C^T U_0\left[\left(A^T W_0(\cdot)A\right)^+ A^T W_0(\cdot)\cdot\right]\right\} \times \\ & \left(A^T W_0(\cdot)A\right)^+ A^T W_0(\cdot) \ . \end{aligned}$$

It should also be observed that any techniques used to calculate the ℓ^p and ℓ^q approximations of Theorem 3.4 a) and b) can be used to calculate c).

BIBLIOGRAPHY

1. A. Ben-Israel and A. Charnes, *Contributions to the Theory of Generalized Inverses*, J. Soc. Indust. Appl. Math. 11 (1963), 667-699.

2. J. Blatter, P. D. Morris, and D. E. Wulbert, *Continuity of the Set-Valued Metric Projection* Math. Annalen 178 (1968), 12-24.

3. T. L. Bouillion and P. L. Odell, *Generalized Matrix Invertibility*, Wiley — Innerscience, New York, 1971.

4. A. L. Brown, *Best n-Dimensional Approximation to Sets of Functions*, Proc. London Math. Soc. 14 (1964), 577-594.

5. R. Creighton Buck, "Linear Spaces and Approximation Theory," in *On Numerical Approximation*, ed. by Rudolph E. Langer, U. of Wisconsin Press, Madison, 1959, 11-24.

6. E. W. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, New York, 1966.

7. ______________ and D. E. Wulbert, *The Existence and Unicity of Best Approximations*, Math. Scand., (to appear).

8. R. E. Cline, *Inverses of Rank Invariant Powers of a Matrix*, Soc. Indust. Appl. Math. Numer. Anal. 5 (1968), 182-197.

9. H. P. Decell, Jr., and P. L. Odell, *A Note Concerning a Generalization of the Gauss-Markov Theorem*, Texas Journal of Science 27 (1966), 21-24.

10. C. A. Desoer and B. H. Whalen, *A Note on Pseudoinverses*, J. Soc. Indust. Appl. Math. 11 (1963), 442-447.

11. C. S. Duris and Sreedharan, *Chebyshev and ℓ^1-Solutions of Linear Equations Using Least Squares Solutions*, Soc. Indust. Appl. Math. Numer. Anal. 5 (1968), 491-505.

12. H. Ekblom and S. Henriksson, *L_p-Criteria for the Estimation of Location Parameters*, J. Soc. Indust. Appl. Math. 17 (1969), 1130-1140.

13. J. S. Frame, *Matrix Functions and Applications. I. Matrix Operations and Generalized Inverses*, IEEE Spectrum 1 (1964), 209-220.

14. F. R. Gantmacher, *The Theory of Matrices*, v. 1, K. A. Hirsch, translator, Chelsea, New York, 1959.

15. A. J. Goldman and M. Zelen, *Weak Generalized Inverses and Minimum Variance Linear Unbiased Estimation*, J. Res. Nat. Bur. Standards Sect. B 68 (1964), 151-172.

16. F. A. Graybill, *An Introduction to Linear Statistical Models* v. 1, McGraw-Hill, New York, 1961.

17. W. H. Greub, *Linear Algebra*, 3rd ed., Springer-Verlag, New York, 1967.

18. G. H. Hardy, J. E. Littlewood and G. Polya, *Inequalities*, 2nd ed., Cambridge University Press, Cambridge, 1952.

19. John Z. Hearon, and John W. Evans, *Differentiable Generalized Inverses*, J. Res. Nat. Bur. Standards Sect. B 72 (1968), 109-113.

20. Edwin Hewitt and Karl Stromberg, *Real and Abstract Analysis*, Springer-Verlag, New York, 1965.

21. L. A. Karlovitz, *Construction of Nearest Points in the* L^p, p *Even, and* L^∞ *Norms 1*, J. of Approx. Theory 3 (1970), 123-127.

22. I. J. Katz and M. H. Pearl, *On* EPr *and Normal* EPr *Matrices*, J. Res. Nat. Bur. Standards Sect. B 70 (1966), 47-77.

23. Charles L. Lawson, *Contributions to the Theory of Least Maximum Approximation*, Ph.D. Thesis, UCLA, 1961.

24. A. J. Lazar, P. D. Morris, and D. E. Wulbert, *Continuous Selections for Metric Projections*, J. Funct. Anal. 3 (1969), 193-216.

25. Joram Lindenstrauss, *On Nonlinear Projections in Banach Spaces*, Michigan Math. J. 11 (1964), 263-287.

26. T. O. Lewis and T. G. Newman, *Pseudoinverses of Positive Semidefinite Matrices*, J. Soc. Indust. Appl. Math. 16 (1968), 701-703.

27. ________________ and P. L. Odell, *A Generalization of the Gauss-Markov Theorem*, J. Amer. Statis. Assoc. 6 (1967), 1063-1066.

28. ________________ and ________________, *A Theory of Linear Estimation*, 2nd ed., Texas Center for Research, Austin, Texas, 1970. To be published by Prentice-Hall.

29. G. G. Lorentz, *Approximation of Functions*, Holt, Rinehart and Winston, New York, 1966.

30. Marcel Meicler, *Weighted Generalized Inverses with Minimal* p *and* q *Norms*, Ph.D. Thesis, The Univ. of Texas, 1966.

31. ________________, *Chevyshev Solution of an Inconsistent System of* n+1 *Linear Equations in* n *Unknowns in Terms of its Least-Squares Solution*, SIAM Rev. 10 (1968), 373-375.

32. ________________, *A Steepest Ascent Method for the Chevyschev Problem*, Math. Comp. 23 (1969), 813-817.

33. ________________ and P. L. Odell, *q-p Generalized Inverses*, Report TR-13, Dept. 672-20, Lockheed Electronics Co., Houston Aerospace Systems Div., Houston, 1967.

34. Gunter Meinardus, *Approximation of Functions: Theory and Numerical Methods*, Springer-Verlag, New York, 1967.

35. Peter D. Morris, *Metric Projections onto Subspaces of Finite Codimension*, Duke Math. J. **35** (1968), 799-808.

36. T. G. Newman, M. Meicler, and P. L. Odell, "On the Concept of a p-q Generalized Inverse of a Matrix," in *Theory and Application of Generalized Inverses of Matrices*, Math. Series 4, T. L. Boullion and P. L. Odell, ed., Texas Technological College Press, Lubbock, Texas, 1968.

37. ________________ and P. L. Odell, *On the Concept of a* p-q *Generalized Inverse of a Matrix*, J. Soc. Indust. Appl. Math. **17** (1969), 520-525.

38. R. Penrose, *A Generalized Inverse for Matrices*, Proc. Camb. Phil. Soc. **51** (1955), 406-413.

39. ________________, *On Best Approximate Solutions of Linear Matrix Equations*, Proc. Camb. Phil. Soc. **52** (1956), 17-19.

40. R. R. Phelps, *Cebysev Subspaces of Finite Codimension in* C(X), Pacific J. Math. **13** (1963), 647-655.

41. C. M. Price, *The Matrix Pseudoinverse and Minimal Variance Estimates*, SIAM Rev. **6** (1964), 115-120.

42. C. R. Rao, *A Note on a Generalized Inverse of a Matrix with Applications to Problems of Mathematical Statistics*, J. Roy. Statis. Soc. B, **24** (1962), 152-158.

43. ________________, *Linear Statistical Inference and Its Applications*, John Wiley, New York, 1965.

44. ________________, "Generalized Inverse for Matrices and Its Applications in Mathematical Statistics," in *Festschrift for J. Neyman: Research Papers in Statistics*, F. N. David, ed., John Wiley, London, 1966, 263-279.

45. J. R. Rice, Private Communication, 2 July 1970.

46. ______________ and Karl H. Usow, *The Lawson Algorithm and Extensions*, Math. Comp. 22 (1968), 118-127.

47. C. A. Rohde, *Some Results on Generalized Inverses*, SIAM Rev. 8 (1966), 201-205.

48. Ivan Singer, *Cea Mai Buna Aproximare in Spatii Vectoriale Normate Prin Elementi din Subspatii Vectoraile*, (Romanian), MR 38, #3677. English translation: *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces*, Radu Georgescu, translator, Springer-Verlag, New York, 1970.

49. Angus E. Taylor, *Introduction to Functional Analysis*, John Wiley, New York, 1958.

50. John Todd, *Introduction to the Constructive Theory Functions*, Academic Press, New York, 1963.

51. Karl H. Usow, *On* L_1 *Approximation* I: *Computation for Continuous Functions and Continuous Dependence*, Soc. Indust. Appl. Math. Numer. Anal. 4 (1967), 70-88.

52. ____________, *On* L_1 *Approximation* II: *Computation for Discrete Functions and Discretization Effects*, Soc. Indust. Appl. Math. Numer. Anal. 4 (1967), 233-244.

53. D. E. Wulbert, *Continuity of Metric Projections*, Trans. Amer. Math. Soc. 134 (1968), 335-342.