Sankhyā : The Indian Journal of Statistics
1975, Volume 37, Series A, Pt. 4, pp. 550–563.

OPTIMAL INVERSE OF A MATRIX[1]

By SUJIT KUMAR MITRA
Indian Statistical Institute

SUMMARY. An optimal approximate solution (x) of the possibly inconsistent equation $Ax = y$ minimizes the norm of $\begin{pmatrix} Ax-y \\ x \end{pmatrix}$ considered a vector in the appropriate product space. Such a solution is computed as $x = Gy$ through an optimal inverse G of A. This definition generalizes the earlier work of Foster (1961). Properties of optimal inverse are studied for the first time and some applications are discussed.

1. Best approximate solution

For (column) vectors y, v in $\mathcal{E}^m$ (the vector space of complex m-tuples) let the inner product be defined by

$$(y, v)_m = v^* My \qquad \ldots \quad (1.1)$$

where M is a given positive definite (p.d.) matrix of order $m \times m$ and $*$ on a vector or a matrix indicates complex conjugate transpose[2]. Let the inner product in $\mathcal{E}^n$ denoted by $(\ ,\ )_n$ be similarly defined through a p.d. matrix N of order $n \times n$. The corresponding vector norms induced by these inner products are denoted by $\|\ \|_m$ and $\|\ \|_n$ respectively. Let A be a complex matrix of order $m \times n$. A matrix G is called minimum N-norm g-inverse of A and denoted by $A^-_{m(N)}$ if for every $y \in \mathcal{M}(A)$ the column span of A, $x = Gy$ is a solution of the equation $Ax = y$, with the least N-norm. For some $y \in \mathcal{E}^m$, if the equation $Ax = y$ is inconsistent no $x \in \mathcal{E}^n$ exists such that $Ax = y$. In such a case the most that can be achieved is a choice of x that makes Ax as close as possible to y in some acceptable sense. . An approximate solution in the sense of minimizing $\| Ax - y \|_m$ requires the orthogonal projection of y onto $\mathcal{M}(A)$. Accordingly a matrix G is called a M-least squares g-inverse of A and denoted by the symbol $A^-_{l(M)}$ if $P = AG$ is the orthogonal projector onto $\mathcal{M}(A)$. In such a case $x = Gy$ is also referred to as a M-least squares solution of the inconsistent equation $Ax = y$. The best approximate solution of Penrose (1955) (also called extremal virtual solution by Tseng (1949))is the

---

Paper presented at the International Symposium on Recent Trends of Research in Statistics held at the Indian Statistical Institute, Calcutta, December 16–27, 1974.

[1] Research partly supported by National Science Foundation Grant GP 32822 while the author was visiting Indiana University.

[2] For the sake of uniformity we shall use $*$ also on real vectors and matrices. Here the $*$ operation leads to a transpose.

M-least squares solution of $Ax = y$ with the least N-norm $\| \|_n$. A matrix G of order $n \times m$ is called the minimum N-norm M-least squares inverse of A and denoted by A^+_{MN} if for each $y \in \mathcal{E}^m$, $x = Gy$ is the best approximate solution of $Ax = y$. The suffixes M and N in the notation A^+_{MN} are generally suppressed when both M and N are identity matrices.

2. Another formulation of the best : optimal approximate solution

If the intention is to make both $Ax-y$ and x small in some sense an alternative approach would be to consider $\begin{pmatrix} Ax-y \\ x \end{pmatrix}$ a vector in the product space $\mathcal{E}^m \times \mathcal{E}^n$ and minimize a suitable norm of this vector not necessarily a product norm. If the inner product in this product space and the corresponding vector norm are induced by the p.d. matrix Λ of order $(m+n)\times(m+n)$ the optimal approximate solution in this sense would require the orthogonal projection (under this inner product) of the vector $\begin{pmatrix} y \\ 0 \end{pmatrix}$ in $\mathcal{E}^m \times \mathcal{E}^n$ onto a subspace $\mathcal{S}$ of $\mathcal{E}^m \times \mathcal{E}^n$ where $\mathcal{S}$ consists of all vectors $\begin{pmatrix} b \\ a \end{pmatrix}$ with a arbitrary and $b = Aa$. Using the explicit representation of the orthogonal projector given in Rao and Mitra [(1971), p. 111] the unique OAS is seen to be given by

$$x = A^\dagger_\Lambda y \qquad \dots \quad (2.1)$$

where

$$A^\dagger_\Lambda = (A^*\Lambda_{11}A + A^*\Lambda_{12} + \Lambda^*_{12}A + \Lambda_{22})^{-1}(A^*\Lambda_{11} + \Lambda^*_{12}) \qquad \dots \quad (2.2)$$

and

$$\begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda^*_{12} & \Lambda_{22} \end{pmatrix}$$

is the relevant partitioned form of Λ.

With a product norm on $\mathcal{E}^m \times \mathcal{E}^n$, Λ is given by the direct sum of M and N. In symbols, $\Lambda = M \oplus N$. Here the OAS minimizes

$$(Ax-y)^*M(Ax-y) + x^*Nx \qquad \dots \quad (2.3)$$

and is given by $x = A^\dagger_{M\oplus N}\, y$ where

$$A^\dagger_{M\oplus N} = (A^*MA+N)^{-1}A^*M. \qquad \dots \quad (2.4)$$

Minimization of a criterion as in (2.3) was considered earlier in smoothing procedures for solving an ill-conditioned system of linear equations

$$\bar{A}x = \bar{b} \qquad \dots \quad (2.5)$$

where both the coefficient matrix $\bar{A}$ and the vector $\bar{b}$ are corrupted by random noise (Tikhonov, 1965). A closely related problem is the numerical solution of the Fredholm integral equation of the first kind

$$\int K(s, t)x(t)dt = b(s) \tag{2.6}$$

with a singular or ill-conditioned kernel $K(s, t)$ and some disturbances in the right hand side $b(s)$ (Phillips, 1962). This type of equations appears in many branches of physical sciences, typically in the experimental sciences where physical data are measured by indirect sensing devices. For an excellent bibliography on smoothing procedures the reader is referred to Tanabe (1974).

The ridge regression estimates of the regression parameters (Hoerl and Kennard, 1970a and b) in a linear regression model are obtained through a modification of the usual method of least squares of the same type as envisaged in the minimization of (2.3).

Foster (1961) uses the matrix $(A^*)^{\dagger}_{N\oplus M}$ in arriving at an optimal estimate of the signal x from the observed random vector y assumed related to x through a linear structural equation of the type

$$y = Ax+e. \quad \dots \quad (2.7)$$

Consider a linear estimator

$$\hat{x} = By+b \quad \dots \quad (2.8)$$

and the optimality criterion

$$\phi(B, b) = \text{tr}\Phi\Delta \quad \dots \quad (2.9)$$

where Φ is a given nonnegative definite (n.n.d.) matrix and Δ is the expected error square and product matrix

$$\Delta = E(x-\hat{x})(x-\hat{x})^*. \quad \dots \quad (2.10)$$

Under the assumption that the signal and noise vectors x and e are uncorrelated with mean values a and 0 and dispersion matrices N and M respectively, it is shown that the optimal estimate of x in the sense of minimizing $\phi(B, b)$ is independent of Φ and is given by

$$B = [(A^*)^{\dagger}_{N\oplus M}]^*, \; b = [I-BA]a. \quad \dots \quad (2.11)$$

In Theorem 2.1 we shall establish a similar optimality property of the more general matrix functions of the type (2.2).

Consider the partitioned form of Λ used in the derivation of (2.2). For any such matrix Λ define

$$\tilde{\Lambda} = K^*\Lambda K \quad \dots \quad (2.12)$$

where

$$K = \begin{pmatrix} 0 & I_m \\ I_n & 0 \end{pmatrix}. \qquad \dots \quad (2.13)$$

Theorem 2.1 : *Let* $B_0 = [(A^*)^{\dagger}_{\tilde{\Lambda}}]^*$, $b_0 = (I-B_0A)a$. *If the noise and signal vectors* e *and* x *are correlated and* Λ *is the dispersion matrix* $D\begin{pmatrix} e \\ x \end{pmatrix}$ *of* e *and* x, *then*

$$\phi(B_0, b_0) = \inf \phi(B, b). \qquad \dots \quad (2.14)$$

Proof: Observe that

$$x - \hat{x} = (I-BA)(x-a) - Be - [b-(I-BA)a]$$

Hence

$$\begin{aligned} \Delta = & (I-BA)\Lambda_{22}(I-BA)^* + B\Lambda_{11}B^* \\ & - B\Lambda_{12}(I-BA)^* - (I-BA)\Lambda_{12}^*B^* \\ & + [b-(I-BA)a][b-(I-BA)a]^*. \end{aligned} \qquad \dots \quad (2.15)$$

Since the only term in Δ involving b is nonnegative definite, it is clear that for any given B an optimal choice is to fix $b = (I-BA)a$. Observe that since Φ is n.n.d. it can be represented in the form $\Phi = \Sigma \eta_i \eta_i^*$. Hence with this choice of b we have

$$\begin{aligned} \phi(B, b) &= \operatorname{tr} \Phi\Delta = \Sigma \eta_i^* \Delta \eta_i \\ &= \Sigma \begin{pmatrix} A^*B^*\eta_i - \eta_i \\ B^*\eta_i \end{pmatrix}^* \tilde{\Lambda} \begin{pmatrix} A^*B^*\eta_i - \eta_i \\ B^*\eta_i \end{pmatrix} \end{aligned} \qquad \dots \quad (2.16)$$

Following the argument which led to the derivation of (2.2) it is seen that the minimum of this expression is attained at $B = B_0$ where B_0 is given by $B_0^* = (A^*)^{\dagger}_{\tilde{\Lambda}}$.

The matrix $[(A^*)^{\dagger}_{N \oplus M}]^*$ is designated by Foster (1961) the optimum inverse of A. Chipman (1969) gives an interesting use of the same matrix in the Bayesian analysis of a linear regression model. See also the section on best linear estimation in Rao (1971) in this connection. We shall however take the liberty of calling $A^{\dagger}_{M \oplus N}$ the optimal inverse of A and in fact propose to extend the terminology to cover the more general form $A^{\dagger}_{\Lambda}$ (henceforth to be referred to as the Λ-optimal inverse of A). One reason for this is that the optimal inverse in this sense is indeed a natural generalization of the Moore-Penrose inverse A^+_{MN} as was noticed in the introductory discussions in the present section and does in fact have many properties which are strikingly similar to that of the Moore-Penrose inverse. Theorem 3.1 which is the first of several theorems we shall prove in this direction shows that there are no basic conflicts in our terminology and the one introduced earlier by Foster, since

$$[(A^*)^{\dagger}_{N\oplus M}]^* = A^{\dagger}_{M^{-1}\oplus N^{-1}}. \quad \ldots \quad (2.17)$$

3. Optimal inverse versus Moore-Penrose inverse

We shall now establish Lemmas 3.1 and 3.2 which we need in the proof of Theorem 3.1.

Lemma 3.1 : *Let B be a complex matrix of order $t \times n$ and rank r and P be the orthogonal projector onto $\mathcal{M}(B)$ under the inner product*

$$(x, y)_t = y^* T x. \quad \ldots \quad (3.1)$$

Then $I - P^*$ is the orthogonal projector onto $\mathcal{N}(B^*)$ the null space of B^*, under the dual inner product

$$(x, y)_t = y^* T^{-1} x. \quad \ldots \quad (3.2)$$

Proof : Lemma 3.1 follows from wellknown properties of the orthogonal projector (see e.g. Theorem 5.2.1. in Rao and Mitra (1971)) once it is noted that

$$\mathcal{M}(I - P^*) = \mathcal{N}(B^*).$$

We leave the details to the reader.

Lemma 3.2 : *For a scalar $\lambda \neq 0$*

$$(\lambda A)^{\dagger}_{\Lambda} = \lambda^{-1}(A)^{\dagger}_{\Lambda_\lambda} \quad \ldots \quad (3.3)$$

where

$$\Lambda_\lambda = \begin{pmatrix} \lambda\bar{\lambda}\,\Lambda_{11} & \bar{\lambda}\Lambda_{12} \\ \lambda\Lambda^*_{12} & \Lambda_{22} \end{pmatrix}. \quad \ldots \quad (3.4)$$

Proof : Lemma 3.2 is a simple consequence of the definition of the optimal inverse.

Theorem 3.1 :

$$(A^{\dagger}_{\Lambda})^* = (A^*)^{\dagger}_{(\tilde{\Lambda}_{-1})^{-1}} \quad \ldots \quad (3.5)$$

where $\tilde{\Lambda}$ is as defined in (2.12).

Proof: Observe that the subspace $\mathcal{S}$ of $\mathcal{E}^m \times \mathcal{E}^n$ could be identified with $\mathcal{M}(B)$, where $B^* = (A^* : I)$. If P is the orthogonal projector onto $\mathcal{M}(B)$ under the inner product induced by Λ, by Lemma 3.1, $(I-P^*)$ is the orthogonal projector onto $\mathcal{M}(C)$ under the dual inner product where $\mathcal{M}(C) = \mathcal{N}(B^*)$. Here one choice of C is given by

$$C = \begin{pmatrix} I \\ -A^* \end{pmatrix}. \qquad \dots \quad (3.6)$$

If

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \qquad \dots \quad (3.7)$$

is the relevant partitioned form of P

$$I-P^* = \begin{pmatrix} I-P_{11}^* & -P_{12}^* \\ -P_{21}^* & I-P_{22}^* \end{pmatrix}. \qquad \dots \quad (3.8)$$

The OAS of the equation $(-A^*)y = x$ under the norm induced by

$$\Omega = \begin{pmatrix} 0 & I_m \\ I_n & 0 \end{pmatrix} \Lambda^{-1} \begin{pmatrix} 0 & I_m \\ I_n & 0 \end{pmatrix} \qquad \dots \quad (3.9)$$

is therefore obtained by applying the projector $(I-P^*)$ on $\begin{pmatrix} 0 \\ x \end{pmatrix}$. The required OAS is

$$y = (-A^*)^\dagger_\Omega \quad x = -P_{21}^* x. \qquad \dots \quad (3.10)$$

Since $P_{21} = A^\dagger_\Lambda$, Theorem 3.1 follows by a straightforward application of Lemma 3.2 with $\lambda = -1$.

Of special interest is the case where $\Lambda = M \oplus N$. $A^\dagger_{M \oplus N}$ is not a generalized inverse of A since $A - AA^\dagger_{M \oplus N} A = A(A^*MAN^{-1}+I)^{-1} \neq 0$ unless $A = 0$. Also if $B = A^\dagger_{M \oplus N}$, $B^\dagger_{N \oplus M} \neq A$, since equality here would imply

$$\begin{aligned}
(B^*NB+M)A &= B^*N \\
&\Longrightarrow MA(A^*MA+N)^{-1}N(A^*MA+N)^{-1}A^*MA+MA \\
&\qquad = MA(A^*MA+N)^{-1}N \\
&\Longrightarrow A^*MA(A^*MA+N)^{-1}N(A^*MA+N)^{-1}A^*MA \\
&\qquad = -A^*MA(A^*MA+N)^{-1}A^*MA \\
&\Longrightarrow A^*MA(A^*MA+N)^{-1}A^*MA = 0 \\
&\Longrightarrow B = 0.
\end{aligned}$$

Theorem 3.2 and 3.3 show that almost in every other respect the properties of $A^{\dagger}_{M\oplus N}$ are similar to that of A^{+}_{MN}. See for example the properties of A^{-}_{MN} listed in Chapter 3 of Rao and Mitra (1971). The uniqueness of $A^{\dagger}_{M\oplus N}$ was noted earlier in Section 2.

Theorem 3.2 : *If G be $A^{\dagger}_{M\oplus N}$, then*

(i) *Rank G = Rank A*

(ii) $(NGA)^{*} = NGA$

(iii) $(MAG)^{*} = MAG$.

(iv) *AG and GA are semisimple matrices with rank equal to rank of A and spectrum wholly contained in the half open interval* (0, 1).

(v) *The row and column spans of G are same as that of $A^{\#}$. $A^{\#}$ the adjoint of A is defined by the equation*

$$(Ax, y)_m = (x, A^{\#}y)_n. \qquad \ldots \quad (3.11)$$

Proof: (i), (iii) and (iv) are almost immediate from the representation of $A^{\dagger}_{M\oplus N}$ given in (2.4). NGA is the parallel sum of N and $A^{*}MA$ as defined by Anderson and Duffin (1969). (ii) follows from the corresponding property of the parallel sum established by these authors.

Since $A^{\#} = N^{-1}A^{*}M$ the claim made in (v) about the rowspan of $A^{\dagger}_{M\oplus N}$ follows from (2.4). The claim about the column span is easily established using (2.17).

Theorem 3.3 : (i) *If U and V are isometrics in $\mathcal{E}^m$ and $\mathcal{E}^n$, that is*

$$\|Uy\|_m = \|y\|_m, \ \|Vx\|_n = \|x\|_n, \ \forall x \in \mathcal{E}^n, \quad y \in \mathcal{E}^m,$$

then

$$V(UAV)^{\dagger}_{M\oplus N}U = A^{\dagger}_{M\oplus N}. \qquad \ldots \quad (3.12)$$

(ii) *If A is square and normal, that is $A^{*}MA = AMA^{*}$, then*

$$\left[A^{\dagger}_{M\oplus M_0}\right]A = \left[(A^{*})^{\dagger}_{M\oplus M_0}\right]A^{*} \qquad \ldots \quad (3.13)$$

where M_0 is a positive definite matrix of order $m\times m$ possibly different from M.

Proof: (i) follows from the definition of the optimal inverse, (ii) from (2.4).

4. Optimal inverse under a seminorm

The theory we have presented earlier for the case where Λ is p.d. extends itself with only minor modification to the positive semidefinite (p.s.d.) case. Perhaps the only change needed is that one considers here a projector under the Λ-seminorm in place of the orthogonal projector onto $\mathcal{M}\begin{pmatrix} A \\ \cdots \\ I \end{pmatrix}$. Such projectors however need not be unique (Mitra and Rao, 1974) and this could result in nonuniqueness of a Λ-optimal inverse. We recall here the definition of a projector under a seminorm and present in Lemma 4.1 some of the known properties of such a projector. For a proof of Lemma 4.1 the reader is referred to Mitra and Rao (1974).

Definition : Let B be a complex matrix of order $t\times n$ and T a p.s.d. matrix of order $t\times t$. A matrix P is a projector into $\mathcal{M}(B)$ under the seminorm induced by T if $\mathcal{M}(P)\subset\mathcal{M}(B)$ and

$$(y-Py)^*T(y-Py) \leqslant (y-Bx)^*T(y-Bx)$$

for every $y\in\mathcal{E}^t$ and $x\in\mathcal{E}^n$.

Lemma 4.1 : (i) *For P to be a projector into $\mathcal{M}(B)$ under the seminorm induced by T it is necessary and sufficient that the following holds :*

(a) $\mathcal{M}(P)\subset\mathcal{M}(B)$

(b) $P^*TP = TP$

(c) $TPB = TB$.

(ii) *The matrix TP is unique with respect to the choice of a projector P and is positive semidefinite.*

Let a projector P into $\mathcal{M}\begin{pmatrix} A \\ \cdots \\ I \end{pmatrix}$ under the Λ seminorm be partitioned as

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

where P_{11} is of order $m\times m$. Then $P_{21} = G$ is one choice of A_Λ^+ a Λ-optimal inverse of A. It is also not difficult to see that the entire class $\{A_\Lambda^+\}$ of Λ-optimal inverses of A is determined by this process from the class of all matrices P which are projectors into $\mathcal{M}\begin{pmatrix} A \\ \cdots \\ I \end{pmatrix}$ under the Λ-seminorm. The following theorem is easily deduced from Lemma 4.1.

Theorem 4.1 : *If* $G \in \{A^{\dagger}_{\Lambda}\}$, *then*

(a) $\Lambda_{11}AG+\Lambda_{12}G$ *is unique and n.n.d.*

(b) $\Lambda^{*}_{12}AG+\Lambda_{22}G$ *is unique*

(c) $\Lambda_{11}-\Lambda_{11}AG-\Lambda_{12}G$ *is n.n.d.*

(d) *For any* $y \in \mathcal{E}^m$

$$\min_{x \in \mathcal{E}^n} \left| \begin{matrix} Ax-y \\ x \end{matrix} \right|^2 = y^{*}(\Lambda_{11}-\Lambda_{11}AG-\Lambda_{12}G)y \qquad \ldots \quad (4.1)$$

(e) $\Lambda^{*}_{12}(AGA-A)+\Lambda_{22}GA$ *is n.n.d. and also is* $\Lambda_{22}-\Lambda_{22}GA-\Lambda^{*}_{12}(AGA-A)$.

(f) *For any* $u \in \mathcal{E}^n$

$$\min_{x \in \mathcal{E}^n} \left\| \begin{matrix} A\ x \\ x-u \end{matrix} \right\|^2 = u^{*}\{\Lambda^{*}_{12}(AGA-A)+\Lambda_{22}GA\}u. \qquad \ldots \quad (4.2)$$

The following theorem is also easily established. We omit its proof.

Theorem 4.2 : (a) *For a matrix G to be a Λ-optimal inverse of A it is necessary and sufficient that*

$$(A^{*}\Lambda_{11}A+A^{*}\Lambda_{12}+\Lambda^{*}_{12}A+\Lambda_{22})G = A^{*}\Lambda_{11}+\Lambda^{*}_{12}. \qquad \ldots \quad (4.3)$$

(b) *A particular solution to a Λ-optimal inverse of A is*

$$G_0 = \Delta^{-}_{22}(A^{*}\Lambda_{11}+\Lambda^{*}_{12}) \qquad \ldots \quad (4.4)$$

where Δ^{-}_{22} is any g-inverse of

$$\Delta_{22} = (A^{*}\Lambda_{11}A+A^{*}\Lambda_{12}+\Lambda^{*}_{12}A+\Lambda_{22}). \qquad \ldots \quad (4.5)$$

(c) *A general solution is*

$$G = G_0+(I-\Delta^{-}_{22}\Delta_{22})U \qquad \ldots \quad (4.6)$$

where U is arbitrary.

(d) *A Λ-optimal inverse of A is unique iff Δ_{22} is p.d.*

The special case $\Lambda = M \oplus N$ makes an interesting reading as it offers an immediate opportunity of comparison with A_{MN} a minimum N seminorm M semileast square inverse of A. The definition of such an inverse is similar to that of A^{+}_{MN} in Section 1 except that the matrices M and N inducing the seminorms are p.s.d. matrices. The inverses A_{MN} are studied in details in Mitra and Rao (1974).

Corollary 4.1 : *If* $G \in \{A^{\dagger}_{M \oplus N}\}$, *then*

(a) *MAG is unique and n.n.d. and so is* $M-MAG$.

(b) *For any* $y \in \mathcal{E}^m$

$$\min_{x \in \mathcal{E}^n} \left\| \begin{matrix} Ax-y \\ x \end{matrix} \right\|^2_{M \oplus N} = y^*(M-MAG)\,y. \qquad \ldots \quad (4.7)$$

(c) *NG is unique.*

(d) *NGA is unique and n.n.d. and so is* $N-NGA$.

(e) *For any* $u \in \mathcal{E}^n$

$$\min_{x \in \mathcal{E}^n} \left\| \begin{matrix} Ax \\ x-u \end{matrix} \right\|^2_{M \oplus N} = u^* NGA\, u. \qquad \ldots \quad (4.8)$$

Corollary 4.2 : (a) *For a matrix G to belong to* $\{A^{\dagger}_{M \oplus N}\}$ *it is necessary and sufficient that* $N_0 G = A^*M$, *where*

$$N_0 = A^*MA + N. \qquad \ldots \quad (4.9)$$

(b) *A particular solution to* $A^{\dagger}_{M \oplus N}$ *is*

$$G_0 = N_0^- A^* M, \qquad \ldots \quad (4.10)$$

and a general solution is

$$G_0 + (I - N_0^- N_0)\,U \qquad \ldots \quad (4.11)$$

where U is arbitrary.

(c) *If* $\epsilon\, G\ \{A^{\dagger}_{M \oplus N}\}$ *then*

$$\mathcal{M}(MAG) = \mathcal{M}(MA) \qquad \ldots \quad (4.12)$$

$$\mathcal{M}(NGA) = \mathcal{M}(N) \cap \mathcal{M}(A^*MA). \qquad \ldots \quad (4.13)$$

5. Relations with the minimum seminorm semileast squares inverse

For a positive scalar λ, consider $(A^*MA+\lambda N)^- A^*M$ which is one choice of $A^{\dagger}_{M \oplus \lambda N}$. Theorem 5.1 shows that for each λ there is a determination of $(A^*MA+\lambda N)^-$ such that as $\lambda \downarrow 0$ the corresponding sequence of optimal inverses $A^{\dagger}_{M \oplus \lambda N}$ converges to A_{MN}, a minimum N seminorm M semileast squares inverse of A. $(A^*MA+\lambda N)^+$ is one such determination. Theorem 5.1 thus extends the corresponding result for A^+ due to Den Broeder and Charnes (1957). A more accessible account of this work is given in Ben Israel and Charnes (1963). These authors in their proof use Autonne's Theorem (also known as singular value decomposition) which states that every complex matrix A can always be represented as $A = VDW$ where V and W are unitary matrices and D is diagonal. The approach we take is somewhat different.

Let A^*MA+N be of rank s and S be a matrix of order $n \times s$ and rank s such that $S^*(A^*MA+N)S$ is p.d. Observe that for each positive λ, $S^*(A^*MA+\lambda N)S$ is also positive definite. By Theorem 2.1 of Mitra (1968) therefore,

$$S\{S^*(A^*MA+\lambda N)S\}^{-1}S^* \qquad \ldots \quad (5.1)$$

is a g-inverse of $A^*MA+\lambda N$ with its column span contained in that of S, in fact the unique Hermitian g-inverse with this property.

Theorem 5.1 : *For the choice of* $(A^*MA+\lambda N)^-$ *as in* (5.1) *with* S *fixed independently of* λ, $\lim_{\lambda \downarrow 0} (A^*MA+\lambda N)^- A^*M$ *exists and is one choice for* A_{MN}.

Proof: Without any loss of generality one may assume that

$$S^*(A^*MA+N)S = I \qquad \ldots \quad (5.2a)$$

and

$$S^*(A^*MA)S = D_a^{\dagger} \qquad \ldots \quad (5.2b)$$

where D_a is diagonal, for if it be not so, a matrix S for which $S^*(A^*MA+N)S$ is p.d. could always be replaced by SL so that SL satisfies (5.2a and b). This only requires a cogredient transformation L which will simultaneously reduce the p.d. matrix $S^*(A^*MA+N)S$ to I and the n.n.d. matrix S^*A^*MAS to a diagonal matrix D_a.

Observe that $\mathcal{M}(S) = \mathcal{M}(SL)$. Hence one arrives at the same g-inverse for $A^*MA+\lambda N$ irrespective of whether one considers S or SL in the formula (5.1). Let U be a matrix such that $\mathcal{M}(U) = \mathcal{N}(A^*MA+N)$. If S satisfies (5.2a and b) for a g-inverse of $A^*MA+\lambda N$ determined as in (5.1) we have $(A^*MA+\lambda N)^- A^*MAS = S[D_a+\lambda(I-D_a)]^{-1} D_a$ which as $\lambda \downarrow 0$ converges to a matrix SD where the diagonal matrix D is same as D_a except that in those diagonal positions where D_a has a nonnull entry the corresponding entry in D is strictly equal to 1. Since $A^*MAU = 0$ it follows that $\lim_{\lambda \downarrow 0} (A^*MA+\lambda N)^-$ $A^*MA(S : U)$ exists and so does $\lim_{\lambda \downarrow 0} (A^*MA+\lambda N)^- A^*M$.

† Note that the diagonal entries in D_a are the proper eigen values of A^*MA with respect to A^*MA+N and the columns of S constitutes one determination for the corresponding set of proper eigen vectors (Mitra and Rao, 1968).

The rest of Theorem 5.1 follows from a variational technique outlined in Lemma 5.1.

Lemma 5.1 : *For $\lambda > 0$ and p.s.d. matrices M and N, let $x_\lambda = A^{\dagger}_{M\oplus\lambda N}y$ denote an optimal approximate solution of the equation $Ax = y$. If $\lim\limits_{\lambda \downarrow 0} x_\lambda = \hat{x}$ exists, then $\hat{x}$ is a minimum N seminorm M semileast squares solution of $Ax = y$.*

Proof: By definition

$$(Ax_\lambda - y)^* M(Ax_\lambda - y) + \lambda x_\lambda^* N x_\lambda \leqslant (Ax-y)^* M(Ax-y) + \lambda x^* N x \quad \ldots \quad (5.3)$$

Letting $\lambda \downarrow 0$ on both sides we establish the semileast squares property of $\hat{x}$. To complete the proof of Lemma 5.1 consider x^0 another M semileast squares solution of the equation $Ax = y$. Since

$$(Ax_\lambda - y)^* M(Ax_\lambda - y) \geqslant (Ax^0 - y)^* M(Ax^0 - y),\ (5.3) \Longrightarrow x_\lambda^* N x_\lambda \leqslant (x^0)^* N x^0. \quad \ldots \quad (5.4)$$

Letting $\lambda \downarrow 0$ in (5.4) we establish the minimum seminorm semileast squares property of $\hat{x}$. This concludes the proofs of Lemma 5.1 and Theorem 5.1.

Corollary 5.1 : *$\lim\limits_{\lambda \downarrow 0} (A^* MA + \lambda N)^+ A^* M$ exists and one is choice for A_{MN}.*

Proof: One takes S to be the matrix formed by s independent columns of $A^* MA + N$. With this choice of S, the g-inverse determined by (5.1) is easily seen to be $(A^* MA + \lambda N)^+$.

In Theorem 5.2 we have a generalization of another similar result also due to Den Broeder and Charnes (1957) for the Moore Penrose inverse A^+ of a complex matrix A.

Theorem 5.2 : *For a g-inverse of $(A^* MA + N)$ determined as in (5.1),*

$$\lim_{p \to \infty} \sum_{k=1}^{p} \{N(A^* MA + N)^-\}^k A^* M \text{ exists and}$$

$$NA_{MN} = \sum_{k=1}^{\infty} \{N(A^* MA + N)^-\}^k A^* M \quad \ldots \quad (5.5)$$

(where $A^* M$ may not be removed as a factor from the series).

Proof: As before without any loss of generality we assume that S satisfies (5.2a and b). Then

$$\{N(A^* MA + N)^-\}^k = (NSS^*)^k$$
$$= NS(I - D_a)^{k-1} S^*.$$

Hence,

$$\sum_{k=1}^{\infty}\{N(A^*MA+N)^-\}^k A^*MAS$$
$$= NSD$$

where D is defined as in the proof of Theorem 5.1. This in turn was shown to be equal to $\lim_{\lambda\downarrow 0} N(A^*MA+\lambda N)^- A^*MAS = NA_{MN}AS$ in the same proof.

Also since $A^*MAU = 0$

$$\sum_{k=1}^{\infty}\{N(A^*MA+N)^-\}^k A^*MA(S:U) = NA_{MN}A(S:U)$$

$$\Longrightarrow \sum_{k=1}^{\infty}\{N(A^*MA+N)^-\}^k A^*MA = NA_{MN}A$$

$$\Longrightarrow (5.5).$$

The uniqueness of NA_{MN} is shown in Mitra and Rao (1974). Theorem 5.3 resembles the Neumann type series expansion for A^+ given by Ben Israel and Charnes (1963).

Let d be the maximum of the diagonal elements of $D_a\{(I-D_a)^+\}^2(I-D_a)$ where D_a is as defined in (5.2a and b) or in other words let d be the maximum proper eigen value of A^*MA with respect to N (Mitra and Rao, 1968). Put $N^- = S(I-D_a)^+S^*$.

Theorem 5.3: *For a real number* α *with,*

$$0 < \alpha < \frac{2}{d}$$

the series

$$\alpha \sum_{k=0}^{p}\{(N-\alpha A^*MA)N^-\}^k A^*M$$

converges and

$$NA_{MN} = \alpha\sum_{k=0}^{\infty}\{(N-\alpha A^*MA)N^-\}^k A^*M \qquad \ldots \quad (5.6)$$

(where A^*M *may not be removed as a factor from the series).*

Proof: Theorem 5.3 can be proved on the same lines as in Theorem 5.2. We omit the proof.

References

Anderson, W. N. Jr and Duffin, R. J. (1969): Series and parallel addition of matrices. *J. Math. Anal. Appl.*, 26, 576–594.

Ben Israel, A. and Charnes, A. (1963): Contributions to the theory of generalized inverses. *SIAM J. Appl. Math.*, 11, 667–699.

Chipman, J. S. (1969): Specification problems in regression analysis. *Theory and Applications of Generalized Inverses of Matrices.* Symposium proceedings, Texas Tech. University, Lubbock, Texas, 114–176.

Den Broeder, G. G., Jr. and Charnes, A. (1957): Contributions to the theory of generalized inverses for matrices Purdue University, Lafayette, Indiana. Republished as ONR Research Memo No. 39, Northwestern University, The Technological Institute, Evanston, Illinois (1962).

Foster, M. (1961): An application of the Wiener-Kolmogorov smoothing theory to matrix inversion. *SIAM J. Appl. Math.*, 9, 387–392.

Hoerl, A. E. and Kennard, R. W. (1970a): Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*, 12, 55–68.

——— (1970b): Ridge regression: Applications to non-orthogonal problems. *Technometrics*, 12, 69–82.

Mitra, S. K. (1968): A new class of g-inverse of square matrices. *Sankhyā*, A, 30,323–330.

Mitra, S. K. and Rao, C. R. (1968): Simultaneous reduction of a pair of quadratic forms, *Sankhyā*, A, 30, 313–322.

——— (1974): Projections under seminorms and generalized Moore Penrose inverses. *Linear Algebra and Appl.*, 9, 155–167.

Penrose, R. (1955): A generalized inverse for matrices. *Proc. Cambridge Philos. Soc.*, 51, 406–413.

——— (1956): On best approximate solution of linear matrix equations. *Proc. Cambridge Philos. Soc.*, 52, 17–19.

Phillips, D. L. (1962): A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9, 84–97.

Rao, C. R. (1971): Unified theory of linear estimation. *Sankhyā*, A, 33, 371–394.

Rao, C. R. and Mitra, S. K. (1971): *Generalized Inverse of Matrices and its Applications*, Wiley, New York.

Sibuya, M. (1970): Subclasses of generalized inverse of matrices. *Ann. Inst. Statist. Math.*, 22, 543–556.

Tanabe, K. (1974): Statistical regularization of a noisy ill-conditioned system of linear equations by Akaike's information criterion, (to be published).

Tikhonov, A. N. (1965): The stability of algorithms for the solution of degenerate systems of linear algebraic equations. *USSR Comput. Math. and Math. Physics*, 5, 181–188.

Tseng, Yu. Ya. (1949): Sur les solutions des equations operatrices fonctionelles entre les espaces unitaries. *C.R. Acad. Sci. Paris*, 228, 640–641.

——— (1956): Virtual solutions and general inversions. *Uspehi. Mat. Nauk.* (New Series), 11, 213–215.