Connections between some results on the generalized linear least squares problem

TECHNISCHE HOGESCHOOL EINDHOVEN
NEDERLAND
ONDERAFDELING DER WISKUNDE

TECHNOLOGICAL UNIVERSITY EINDHOVEN
THE NETHERLANDS
DEPARTMENT OF MATHEMATICS

Connections between some results on the generalized linear least squares problem

by

A.D. Poley

T.H.-Report 81-WSK-01
May 1981

Abstract

This paper deals with the generalized least squares problem

"find f which solves $\min\{\|Kf\|_2 \mid \|M(Lf - h)\|_2$ is minimal}."

It is shown that the solution can be written as $f = L^+_{MK}h$, where L^+_{MK} is a solution matrix such that, in case of nonuniqueness (i.e. $N(ML) \cap N(K) \neq \{\underline{0}\}$), f has minimal Euclidean norm. This L^+_{MK} is uniquely determined by Penrose-like conditions.

1980 AMS Mathematics Subject Classification. Primary 65F20; Secundary 62J05.

1. Introduction

In this paper we will consider the generalized least squares problem

(1.1.1) "find a vector f which minimizes $\|M(Lf - h)\|_2$."

If ML does not have full column rank, then f is not uniquely determined by (1.1.1) and we can prescribe additional conditions for f; for instance, we can consider the problem

(1.1.2) "find f which solves $\min\{\|Kf\|_2 \mid \|M(Lf - h)\|_2$ is minimal}."

In these problems K, M and L have to satisfy no other conditions than that their dimensions fit together. However, it is immediately clear that, if $N(ML) \cap N(K) \neq \{\underline{0}\}$, even problem (1.1.2) has no unique solution.

To start with, in §2 we consider a statistical problem similar to (1.1.1) and in solving that we follow the methods used by C.C. Paige [1].

In §3 we use similar methods to solve the more general problem (1.1.1). Afterwards we simplify the solution and obtain a good starting point for attacking problem (1.1.2) in §4. First we solve this problem under the condition $N(ML) \cap N(K) = \{\underline{0}\}$ (cf. Eldén [2]), next we consider the properties of the solution found without assuming that condition.

In §5 we discuss what Penrose-like conditions the solution matrix corresponding to problem (1.1.2) satisfies, and under which extra condition a matrix, which satisfies these conditions, is unique. Also we compare our results with those of Ben-Israel and Greville [3, Sec. 3.3].

Finally, in the last paragraph we consider the solution of problem (1.1.2) by use of Lagrange multipliers.

We list a few notations to be used throughout this report:

* R(A) and N(A) stand for the range and nullspace of a matrix A, respectively.

* $\|\cdot\|$ means the Euclidean vector norm.
* $\|A\|_F = (tr(A^HA))^{\frac{1}{2}}$ is the Frobenius matrix norm.
* $A \in RR$ or $A \in LR$ means A has full column rank or full row rank, respectively.
* A matrix U is called left-unitary when $U^HU = I$.
* $S^\perp$ means the orthogonal complement of a subspace S.
* Superscripts to a matrix refer to the Penrose conditions which that matrix satisfies. These conditions are

  (1) $AXA = A$, (3) $(AX)^H = AX$,

  (2) $XAX = X$, (4) $(XA)^H = XA$.

  For instance, if X satisfies (1) and (3), we write $X = A^{(1,3)}$. The pseudoinverse (or Moore-Penrose inverse) $A^{(1,2,3,4)}$ which satisfies all four conditions, will be denoted by A^+.
* The M,K-weighted pseudoinverse of L is defined by

$$L^+_{MK} = (I - (KE_0)^+K)(ML)^+M,$$

  where $E_0 = I - (ML)^+ML$, cf. Eldén [2, §2].

2. Paige's method of solution

Paige [1] considers the following stochastic model. Let W be a nonnegative definite Hermitian matrix and w a stochast with $\mathcal{E}(w) = 0$, $\mathcal{E}(ww^H) = \sigma^2W$. Let $y = Cx + w$, with C a known matrix and x a fixed but unknown parameter-vector. The problem now is to

(2.1.1) "estimate x from a realisation of y."

Let $W = BB^H$ with $B \in RR$, then (since first and second moments of w and Bv are equal) Paige's model can be reduced to the following equivalent form.

Let v be a stochast with $\mathcal{E}(v) = 0$ and $\mathcal{E}(vv^H) = \sigma^2 I$. Let $y = Cx + Bv$, with C a known matrix and x a parameter-vector. In stochastics it is shown (see Appendix) that problem (2.1.1) leads to

(2.1.2) "find to a given $y \in R((C|B))$ vectors x and v that minimize $\|v\|^2$ under the condition $y = Cx + Bv$."

The essence of Paige's solution method is to decompose C as

$$C = QR = (Q_1|Q_2)\left(\frac{R_1}{0}\right) = Q_1 R_1,$$

with Q unitary and $R_1 \in LR$, and to define

$$T = \left(\frac{T_1}{T_2}\right) := \left(\frac{Q_1^H}{Q_2^H}\right)B.$$

Then the condition $y = Cx + Bv$ is equivalent to

$$\text{(i)}\ Q_1^H y = T_1 v + R_1 x,$$
$$\text{(ii)}\ Q_2^H y = T_2 v.$$

Since R_1 has full row rank, there exists a solution x of (i) to every y and v. Hence (i) does not constrain v and the compatibility condition $y \in R(C) + R(B)$ is equivalent to $Q_2^H y \in R(T_2)$. As is well known, the minimum 2-norm vector v that satisfies the compatible system (ii) is

$$\hat{v} := T_2^+ Q_2^H y.$$

So x has to satisfy

(2.1.3) $R_1 x = (Q_1^H - T_1 T_2^+ Q_2^H)y = Q_1^H(I - B(Q_2^H B)^+ Q_2^H)y.$

The solution of (2.1.3) with minimal 2-norm can be written as

$$\hat{x} := R_1^+ Q_1^H (I - B(Q_2^H B)^+ Q_2^H)y = C^+(I - B(Q_2^H B)^+ Q_2^H)y.$$

The matrix $C^+(I - B(Q_2^H B)^+ Q_2^H)$ will be called the solution matrix to problem (2.1.1).

Remark. The condition $B \in RR$ is necessary for the reduction of Paige's model to its equivalent form, however, it does not play any role in the solution method of problem (2.1.2).

3. Application of Paige's method to problem (1.1.1)

3.1. Now we will derive the solution of problem (1.1.1) by a method similar to that used in §2. By defining $v := -M(Lf - h)$ we can formulate (1.1.1) as

(3.1.1) "find f and v that solve $\min_{f,v}\{\|v\| \mid MLf + v = Mh\}$."

So we can use the results of §2, with $B = I$, $C = ML = QR = Q_1R_1$ with Q unitary and $R_1 \in LR$. If we substitute this in (2.1.3), we find that f must satisfy

(3.1.2) $R_1f = Q_1{}^H(I - Q_2Q_2{}^H)Mh = Q_1{}^HMh.$

The solution with minimum 2-norm of (3.1.2) is

$$\hat{f} := R_1{}^+Q_1{}^HMh.$$

Since $R_1{}^+Q_1{}^H = (ML)^+$, we thus obtain the well-known minimum 2-norm solution of (1.1.1),

(3.1.3) $\hat{f} = (ML)^+Mh.$

3.2. If we want to obtain the solution matrix of problem (1.1.1) in a form similar to that of problem (2.1.2) in §2 (e.g. for reasons of symmetry, see §5.3), we observe that (1.1.1) is also equivalent to

(3.2.1) "find f and v that solve $\min_{f,v}\{\|v\| \mid M^+MLf + M^+v = M^+Mh\}$."

Using again the method of §2, now with $C = M^+ML = \tilde{Q}\tilde{R} = \tilde{Q}_1\tilde{R}_1$ with $\tilde{Q}$ unitary and $\tilde{R}_1 \in LR$, we find after substitution in (2.1.3) that f must satisfy

(3.2.2) $\tilde{R}_1 f = \tilde{Q}_1^H (I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+ Mh$.

The solution with minimal 2-norm of (3.2.2) (and so of (1.1.1)) is

(3.2.3) $\hat{f} := (M^+ ML)^+ (I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+ Mh$,

which should be equal to (3.1.3). Indeed, we have

(3.2.4) $(M^+ ML)^+ (I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+ = (ML)^+$,

which can be proved by verifying that the left-hand side of (3.2.4) satisfies the four $(ML)^+$-Penrose conditions:

(1) $ML\ (M^+ ML)^+ (I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+ ML$

$= MM^+ ML (M^+ ML)^+ M^+ ML = ML$ (since $\tilde{Q}_2^H M^+ ML = 0$) ,

(2) $(M^+ ML)^+ (I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+ ML (M^+ ML)^+ (I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+$

$= (M^+ ML)^+ M^+ ML (M^+ ML)^+ (I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+$,

(3) $MM^+ ML (M^+ ML)^+ (I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+$

$= M(I - \tilde{Q}_2 \tilde{Q}_2^H)(I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+$

$= MM^+ (I - (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H M^+) MM^+$ is Hermitian ,

(4) $(M^+ ML)^+ (I - M^+ (\tilde{Q}_2^H M^+)^+ \tilde{Q}_2^H) M^+ ML$

$= (M^+ ML)^+ M^+ ML$ is Hermitian.

Remark. In (3.2.1) we could, instead of M^+, take an arbitrary (1)-inverse $M^{(1)}$. Then the condition in (3.2.1) is equivalent to

$$v = M(h - Lf) + (I - MM^{(1)})z,$$

with z arbitrary. Under this condition, $\|v\|$ is minimal for any fixed f (and varying z) at $v = -M(Lf - h)$ if and only if

$$M^H (I - M^{(1)} M) = 0 ,$$

that is, if $M^{(1)}$ is also a 3-inverse of M.

Conclusion. Problem (1.1.1) is equivalent to "find f and v that solve

(3.2.1')
$$\min_{f,v}\{\|v\| \mid M^{(1,3)}MLf - M^{(1,3)}v = M^{(1,3)}Mh\}"$$

with $M^{(1,3)}$ an arbitrary (1,3)-inverse of M.

With $C = M^{(1,3)}ML = \tilde{Q}'\tilde{R}' = \tilde{Q}_1'\tilde{R}_1'$, $\tilde{Q}'$ unitary, $\tilde{R}_1' \in LR$, we find

$$(3.2.3')\, f = (M^{(1,3)}ML)^+(I - M^{(1,3)}(\tilde{Q}_2'^H M^{(1,3)})^+ \tilde{Q}_2'^H) M^{(1,3)}Mh \, .$$

Indeed, we have

$$(3.2.4')\, (M^{(1,3)}ML)^+(I - M^{(1,3)}(\tilde{Q}_2'^H M^{(1,3)})^+ \tilde{Q}_2'^H) M^{(1,3)}MM^{(1,3)} = (ML)^+ \, ,$$

which can be proved in the same way as (3.2.4).

4. Application of Paige's method to problem (1.1.2)

4.1. Now we consider the problem (1.1.2) under the extra condition $N(ML) \cap N(K) = \{\underset{\sim}{0}\}$. We remark that the vector f which minimizes (1.1.2), is that solution of (3.1.1) which minimizes $\|Kf\|$. So we can start from condition (3.1.2),

$$R_1 f = Q_1^H Mh$$

and reformulate the problem as

"find f that minimizes $\{\|Kf\| \mid R_1 f = Q_1^H Mh\}$."

Let $R_1 = \tilde{R}P^H = (\tilde{R}_1 \mid 0)\begin{pmatrix} P_1^H \\ \hline P_2^H \end{pmatrix} = \tilde{R}_1 P_1^H$, with P unitary and $\tilde{R}_1$ regular (this is possible since $R_1 \in LR$).

Now define $g = \begin{pmatrix} g_1 \\ \hline g_2 \end{pmatrix} = \begin{pmatrix} P_1^H \\ \hline P_2^H \end{pmatrix} f$.

Then $Kf = KP_1g_1 + KP_2g_2$ and $R_1 f = \tilde{R}_1 g_1 = Q_1^H Mh$, which implies $g_1 = \tilde{R}_1^{-1} Q_1^H Mh$. Then our problem becomes

(4.1.1) "find g_2 that minimizes $\|KP_1\tilde{R}_1^{-1}Q_1{}^H Mh + KP_2g_2\|$."

It is clear from the definition of R_1 that $R(ML) = R(P_2)$, hence the condition $N(ML) \cap N(K) = \{\underline{0}\}$ is equivalent to $KP_2 \in RR$.

So problem (4.1.1) has the unique solution

$$g_2 = -(KP_2)^+ KP_1\tilde{R}_1^{-1}Q_1{}^H Mh \ .$$

Thus we find under the condition $N(ML) \cap N(K) = \{\underline{0}\}$ the unique solution for problem (1.1.2) to be

$$f = P_1g_1 + P_2g_2 = L_{MK}^+ h,$$

where

(4.1.2) $L_{MK}^+ = (I - P_2(KP_2)^+K)P_1\tilde{R}_1^{-1}Q_1{}^H M = (I - P_2(KP_2)^+K)(ML)^+M$.

Since $ML = Q_1R_1 = Q_1\tilde{R}_1P_1{}^H$, we have

$$E_0 = (I - (ML)^+ML) = P_2P_2{}^H \ ,$$

$$(KE_0)^+ = (KP_2P_2{}^H)^+ = P_2(KP_2)^+ \ ,$$

which shows that (4.1.2) is another form of the M,K-weighted pseudo-inverse of L as defined by Eldén [2, §2].

4.2. The method of §4.1 can also be applied to problem (1.1.2) directly.

Let $ML = \tilde{L}P^H = (\tilde{L}_1 \mid 0)\begin{pmatrix} P_1{}^H \\ \hline P_2{}^H \end{pmatrix}$ with P unitary (so $R(P_2) = N(ML)$)

and $\tilde{L}_1 \in RR$ (P_1 is the same as in §4.1!).

Then define $P^H f =: g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$

so $\quad Kf = KP_1g_1 + KP_2g_2 \ , \quad MLf = \tilde{L}_1g_1$.

Since $\tilde{L}_1 \in RR$, $\|\tilde{L}g_1 - Mh\|$ is minimal iff $g_1 = \tilde{L}_1^{\;+}Mh$. Then our problem becomes

(4.2.1) "find g_2 that minimizes $\|KP_1\tilde{L}_1^{\;+}Mh + KP_2g_2\|$."

Again, $N(ML) \cap N(K) = \{\underline{0}\}$ implies $KP_2 \in RR$ and problem (4.1.1) has the unique solution

$$g_2 = -(KP_2)^+KP_1\tilde{L}_1^{\;+}Mh \ .$$

Since $ML = \tilde{L}_1P_1^{\;H}$, $P_1\tilde{L}_1^{\;+} = (ML)^+$, we again find, under the condition $N(ML) \cap N(K) = \{\underline{0}\}$, that problem (1.1.2) has the unique solution $f = L_{MK}^+h$, where L_{MK}^+ is given by (4.1.2).

4.3. If $N(ML) \cap N(K) \neq \{\underline{0}\}$, then L_{MK}^+h (with L_{MK}^+ defined by (4.1.2)) still solves problem (1.1.2), but it is not the unique solution, since the component of f in the intersection of the two null-spaces is arbitrary. We shall derive a special property that characterizes L_{MK}^+h among all solutions of problem (1.1.2). The first point where non-uniqueness occurs in the derivation of §4.1 is that the solutions of (4.1.1) are given by

$$g_2 = -(KP_2)^+KP_1\tilde{R}_1^{-1}Q_1^{\;H}Mh + (I - (KP_2)^+KP_2)z \ ,$$

with z arbitrary. This leads to the result

$$f = L_{MK}^+h + P_2(I - (KP_2)^+KP_2)z = L_{MK}^+h + \tilde{P}_2z \ ,$$

where $\tilde{P}_2 := P_2(I - (KP_2)^+KP_2)$. Now we remark that

$$(L_{MK}^+)^H\tilde{P}_2 = M^H(ML)^{+H}(I - K^H(KP_2)^{+H}P_2^{\;H})P_2(I - (KP_2)^+KP_2)$$

$$= M^H(ML)^{+H}(P_2 - K^H(KP_2)^{+H} - P_2(KP_2)^+KP_2 + K^H(KP_2)^{+H}(KP_2)^+KP_2)$$

$$= M^H(ML)^{+H}P_2(I - (KP_2)^+KP_2) = 0 \ ,$$

since $R(P_2) = N(ML)$. Therefore

$$\|f\|^2 = \|L_{MK}^+h\|^2 + \|P_2(I - (KP_2)^+KP_2)z\|^2 \ ,$$

which leads to the conclusion:

$f = L_{MK}^+h$ is that f that satisfies (1.1.2) and has minimal 2-norm.

5. Generalized Penrose conditions and their solutions

5.1. It follows from §4 that problem (1.1.2) admits of a solution matrix X in the following sense:

(5.1.1) "for each h, $f := Xh$ is a solution to problem (1.1.2)."

We shall now characterize X directly. If (for all h) $f := Xh$ minimizes $\|M(Lf - h)\|$, then for all h and all δf,

$$\|M(Lf - h)\|^2 \le \|M(LX - I)h + ML\,\delta f\|^2 \ ,$$

which is true iff

(5.1.2) $(ML)^H M(LX - I) = 0$.

If X satisfies (5.1.2) then f minimizes $\|M(Lf - h)\|$ iff $f = Xh + \delta f$ with $ML\,\delta f = 0$, hence, iff $f = Xh + (I - (ML)^+ML)z$, z arbitrary. Consequently, $f = Xh$ is (for all h) a solution to problem (1.1.2) iff for all h and all z,

$$\|KXh\|^2 \le \|KXh + K(I - (ML)^+ML)z\|^2$$

or, equivalently, iff

(5.1.3) $(KX)^H K(I - (ML)^+ML) = 0.$

Hence, X is a solution matrix to problem (1.1.2) iff X satisfies (5.1.2) and (5.1.3).

Since in the original problem L only occurs in the combination ML, we may assume without loss of generality that $R(L) \cap N(M) = \{\underline{0}\}$ (or equivalently $N(L) = N(ML)$).

<u>Lemma</u>. The conditions (5.1.2) and (5.1.3) are equivalent to (5.1.2) and

(5.1.3') $(KX)^H K(I - XL) = 0$.

<u>Proof</u>. i) If (5.1.3') holds, then

$$0 = (KX)^H K(I - XL)(I - (ML)^+ ML) = (KX)^H K(I - (ML)^+ ML),$$

since $N(L) = N(ML)$, i.e. (5.1.3) holds.

ii) If X satisfies (5.1.2), then

$$0 = (ML)^H M(I - LX)L = (ML)^H ML(I - XL) ,$$

hence also

$$ML(I - XL) = 0 .$$

So if (5.1.3) holds, then

$$0 = (KX)^H K(I - (ML)^+ ML)(I - XL) = (KX)^H K(I - XL) ,$$

i.e. (5.1.3') holds. □

Now we observe that

$$\text{(5.1.2) is equivalent to } \begin{cases} LXL = L , \\ M^H MLX \text{ is Hermitian ;} \end{cases}$$

$$\text{(5.1.3') is equivalent to } \begin{cases} K(XLX - X) = 0, \\ K^H KXL \text{ is Hermitian .} \end{cases}$$

So $\hat{X}$ satisfies (5.1.1) iff $\hat{X}$ satisfies the four Penrose-like conditions:

(5.1.4)

(1) $LXL = L$, (3) $M^H MLX$ is Hermitian ,

(2) $K(XLX - X) = 0$, (4) $K^H KXL$ is Hermitian .

5.2. Let us now search for a general solution of the conditions (5.1.4). First we remark that (5.1.4) is equivalent to (5.1.2) and (5.1.3). In §3.1 we had $ML = QR = Q_1R_1$ and in §4.1 we had $R_1 = \tilde{R}P^H = \tilde{R}_1P_1^H$. Consequently $N(ML) = N(P_1^H) = R(P_2)$ and

$$I - (ML)^+ML = P_2P_2^H \ .$$

If X_0 satisfies (5.1.2), then the general solution of (5.1.3) is

$$\hat{X} = X_0 + P_2Z \ ,$$

with Z arbitrary. Then (5.1.3) becomes

$$P_2P_2^HK^HK\hat{X} = 0 \ ,$$

or

$$(KP_2)^HK(X_0 + P_2z) = 0 \ .$$

This implies that

$$Z = -(KP_2)^+KX_0 + (I - (KP_2)^+KP_2)Z' \ ,$$

with Z' arbitrary. So the general solution of (5.1.2) and (5.1.3) is

$$\hat{X} = (I - P_2(KP_2)^+K)(X_0 + P_2Z') \ ,$$

with Z' arbitrary.

We can take $X_0 = (ML)^+M$, which satisfies (5.1.2). Then for the corresponding

$$\hat{X}_0 = (I - P_2(KP_2)^+K)(ML)^+M$$

we obtain

$$\hat{X}_0L\hat{X}_0 = (I - P_2(KP_2)^+K)P_1P_1^H\hat{X}_0 = \hat{X}_0 \ ,$$

which in general is not true for other solutions $\hat{X}$. Since the general solution $\hat{X}$ can be written as

$$\hat{X} = (I - P_2(KP_2)^+K)P_1P_1^H(ML)^+M + P_2(I - (KP_2)^+KP_2)Z' ,$$

and

$$(P_2(I - (KP_2)^+KP_2))^H(I - P_2(KP_2)^+K)P_1 = 0 ,$$

we see that the special $\hat{X}_0$ is the solution of (5.1.5) with the smallest F-norm. Regarding uniqueness, $\hat{X}$ is independent of Z' and therefore unique iff $P_2(I - (KP_2)^+KP_2) = 0$ or equivalently iff $N(KP_2) = N(P_2) = \{\underline{0}\}$. The latter is equivalent to $R(P_2) \cap N(K) = \{\underline{0}\}$, so to $N(ML) \cap N(K) = \{\underline{0}\}$.

<u>Conclusion</u>. If $N(ML) = N(L)$ and $N(L) \cap N(K) = \{\underline{0}\}$, then $\hat{X} = L^+_{MK}$ as defined by (4.1.2) is the unique matrix satisfying the conditions (5.1.5).

<u>Remarks</u>. 1. Without the assumption $R(L) \cap N(M) = \{\underline{0}\}$, we can maintain §5.1 if we replace L by $M^{(1)}ML$ in (5.1.3'), with $M^{(1)}$ an arbitrary (1)-inverse of M. Then

$$(5.1.2) \text{ is equivalent to } \begin{cases} M(LXL - L) = 0 , \\ M^HMLX \text{ is Hermitian} ; \end{cases}$$

$$(5.1.3') \text{ is equivalent to } \begin{cases} K(XM^{(1)}MLX - X) = 0 , \\ K^HKXM^{(1)}ML \text{ is Hermitian} ; \end{cases}$$

and (5.1.5) becomes

$$\begin{array}{ll} (1)\ M(LXL - L) = 0 , & (3)\ M^HMLX \text{ is Hermitian} , \\ (2)\ K(XM^{(1)}MLX - X) = 0 , & (4)\ K^HKXM^{(1)}ML \text{ is Hermitian} . \end{array}$$

2. Instead of looking for a solution matrix X for problem (1.1.2), we may also consider the problem

"find X that minimizes $\{ \|KX\|_F \mid \|M(I - LX)\|_F$ is minimal$\}$."

It is easily shown that X solves this problem iff X satisfies (5.1.2) and (5.1.3).

5.3. Ben-Israel and Greville [3, Sec. 3.3] find the same conditions (5.1.5) for the solution matrix of problem (1.1.2), although they restrict themselves to the case that $K^H K$ and $M^H M$ are regular, so K and M ϵ RR, which implies $N(L) = N(ML)$ and $N(ML) \cap N(K) = \{\underline{0}\}$. To construct the solution, they take the Cholesky factorisation of $K^H K$ and $M^H M$,

$$K^H K = R_K^H R_K \ , \ M^H M = R_M^H R_M \ ,$$

with R_K and R_M both regular matrices, and then find

$$L_{MK}^+ = R_K^{-1}(R_M L R_K^{-1})^+ R_M \ ,$$

see [3, Ex. 3.39]. By observing that

$$K = (U_1 \mid U_2)\left(\frac{R_K}{0}\right),\ M = (V_1 \mid V_2)\left(\frac{R_M}{0}\right),$$

with $U = (U_1 \mid U_2)$ and $V = (V_1 \mid V_2)$ both unitary, we can write

$$L_{MK}^+ = R_K^{-1} U_1^H U_1 (R_M L R_K^{-1})^+ V_1^H V_1 R_M$$

$$= R_K^{-1} U_1^H (V_1 R_M L R_K^{-1} U_1^H)^+ V_1 R_M = K^+(MLK^+)^+ M.$$

It is easy to verify that under the condition K and M ϵ RR,

$$L_{IK}^+ = K^+(LK^+)^+ = (I - P_2(KP_2)^+K)L^+$$

with P_2 left-unitary and $R(P_2) = N(L)$, and

$$L_{MI}^+ = (ML)^+M = L^+(I - M^+(Q_2^H M^+)^+ Q_2^H) \ ,$$

with Q_2 left-unitary and $R(Q_2) = N(L^H)$. Since under the same condition (even if only $N(ML) = N(L)$) $L^+_{MK} = L^+_{IK} L L^+_{MI}$, we can give four alternative expressions for L^+_{MK}:

(5.3.1) $L^+_{MK} = K^+(MLK^+)^+M$ (cf. [3, Ex. 3.39])

(5.3.2) $\qquad = (I - P_2(KP_2)^+K)(ML)^+M$ (cf. [2])

(5.3.3) $\qquad = K^+(LK^+)^+(I - M^+(Q_2{}^H M^+)^+ Q_2{}^H)$

(5.3.4) $\qquad = (I - P_2(KP_2)^+K)L^+(I - M^+(Q_2{}^H M^+)^+ Q_2{}^H)$,

where P_2 and Q_2 are left-unitary matrices with $R(P_2) = N(L)$ and $R(Q_2) = N(L^H)$. We now ask whether the expressions (5.3.1) - (5.3.4) still satisfy our Penrose conditions under less stringent conditions than $K \in RR$ and $M \in RR$. We already saw that this is true for (5.3.2) under the only condition $N(ML) = N(L)$. It is easily found that the Penrose conditions are satisfied by (5.3.1) and (5.3.3) if $K \in RR$, and by (5.3.4) if $M \in RR$.

We can settle the difficulty $M \notin RR$ (but still $K \in RR$), by not considering $Lf - h$, but $M^+M(Lf - h)$ instead. So by analogy we obtain the solution matrix

$$L^+_{MK} = (M^+ML)^+_{IK} M^+ML(M^+ML)^+_{MI} M^+M = (M^+ML)^+_{IK} M^+MLL^+_{MI} .$$

Thus (5.3.4) becomes

(5.3.4') $L^+_{MK} = (I - P_2(KP_2)^+K)(M^+ML)^+(I - M^+(Q_2{}^H M^+)^+ Q_2{}^H)$,

where now Q_2 and P_2 are left-unitary matrices, with $R(P_2) = N(M^+ML)$ and $R(Q_2) = N((M^+ML)^H)$. With these P_2 and Q_2, expressions (5.3.1) - (5.3.3) remain the same.

If also $K \notin RR$, we have a greater problem in finding alternative expressions.

If we restrict ourselves to $f \in R(K^+) = (N(K))^\perp$, then there is no problem, since the restriction of K to $R(K^+)$ is injective. Then we define $f := K^+Kz$, and solve the problem

"find z that minimizes $\{\|Kz\| \mid \|MLK^+Kz - Mh\|$ is minimal$\}$."

The solution of this problem is

$$z = (LK^+K)^+_{MK}h + Sy ,$$

with $R(S) = N(K)$ and y arbitrary. So $f = K^+Kz = K^+K(LK^+K)^+_{MK}h$ is unique. As before we have four expressions for the solution matrix:

$$(5.3.5)\quad K^+K(LK^+K)^+_{MK} = K^+(MLK^+)^+M$$

$$(5.3.6)\quad = K^+(I - KP_2(KP_2)^+)K(MLK^+K)^+M$$

$$(5.3.7)\quad = K^+(M^+MLK^+)^+(I - M^+(Q_2^HM^+)^+Q_2^H)M^+M$$

$$(5.3.8)\quad = K^+(I - KP_2(KP_2)^+)K(M^+MLK^+K)^+M^+(I - (Q_2^HM^+)^+Q_2^HM^+)M ,$$

where P_2 and Q_2 are left-unitary matrices, with $R(P_2) = N(M^+MLK^+K)$ and $R(Q_2) = N((M^+MLK^+K)^H)$

Without the restriction $f \in R(K^+)$, we were not able to find a generalization for the expressions (5.3.1) and (5.3.3).

6. Optimization theory

6.1. We now want to solve the problem (1.1.2) using some optimization theory. A general form for this problem is

"find $\underline{x}$ that minimizes $\{f(\underline{x}) \mid \underline{g}(\underline{x}) = \underline{b}\}$,"

where f and $\underline{g}$ are sufficiently smooth functions. The theory of Lagrange multipliers states that a solution $\hat{\underline{x}}$ of this problem corresponds with the $\underline{x}$-coordinates of a saddle point of the Lagrange functional

$$L(\underline{x},\underline{z}) = f(\underline{x}) + \underline{z}^H(\underline{g}(\underline{x}) - \underline{b}) \ .$$

We can find the saddle points from

$$\nabla_{\underline{x}} L = 0 : Df(\underline{x}) + \underline{z}^H Dg(\underline{x}) = 0 \ ,$$

$$\nabla_{\underline{z}} L = 0 : g(\underline{x}) = \underline{b} \ ,$$

and so we have to solve $\underline{x}$ from this system.

Now let us return to our problem. We saw in §3 that, using (3.1.2), problem (1.1.2) can be formulated as

(6.1.1) "find f that minimizes $\{\|Kf\|^2 \mid \|v\|^2$ is minimal $\wedge$ $MLf + v = Mh\}$."

Let us first consider the "inner" problem

$$\min_{f,v}\{\|v\|^2 \mid MLf + v = Mh\} \ ,$$

which is equivalent to

(6.1.2) $\min\{\|(0 \mid I)\begin{pmatrix} f \\ v \end{pmatrix}\|^2 \mid (ML \mid I)\begin{pmatrix} f \\ v \end{pmatrix} = Mh\}$.

We define $T := (0 \mid I)$, $g := \begin{pmatrix} f \\ v \end{pmatrix}$, $H := (ML \mid I)$, $b := Mh$, then problem (6.1.2) becomes

(6.1.3) $\min\{\|Tg\|^2 \mid Hg = b\}$.

The Lagrange functional corresponding to (6.1.3) is

(6.1.4) $L(g,z) = \frac{1}{2}g^H T^H Tg - z^H(Hg - b)$

and we find the saddle points from

$$\nabla_g L = 0 : g^H T^H T - z^H H = 0; \qquad (ML)^H z = 0, \ z = v;$$

$$\nabla_z L = 0 : Hg = b; \qquad MLf + v = Mh \ .$$

So the solution f of the system

$$(6.1.5)\quad \begin{pmatrix} I & ML \\ (ML)^H & 0 \end{pmatrix} \begin{pmatrix} z \\ f \end{pmatrix} = \begin{pmatrix} Mh \\ 0 \end{pmatrix}$$

also solves problem (6.1.2) and $v = z$.

Returning to our original problem (6.1.1), we define

$$p := \begin{pmatrix} z \\ f \end{pmatrix},\ C := \begin{pmatrix} I & ML \\ (ML)^H & 0 \end{pmatrix},\ S = (0 \mid K),\ k := \begin{pmatrix} Mh \\ 0 \end{pmatrix},$$

then (6.1.1) can be shortly written as

$$(6.1.6)\quad \min\{\|Sp\| \mid Cp = k\}\ .$$

The Lagrange functional is now

$$L(p,r) = \tfrac{1}{2}p^H S^H Sp + r^H(Cp - k),$$

and we find the saddle points from

$$\nabla_p L = 0 : p^H S^H S + r^H C = 0; \quad S^H Sp + C^H r = 0;$$

$$\nabla_r L = 0 : Cp = k\ .$$

Then the solution p of the system

$$\begin{pmatrix} S^H S & C^H \\ C & 0 \end{pmatrix} \begin{pmatrix} p \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ k \end{pmatrix}$$

also solves (6.1.6). After substitution of the expressions for S, C, p and k, this system becomes

$$\begin{bmatrix} 0 & 0 & I & ML \\ 0 & K^H K & (ML)^H & 0 \\ I & ML & 0 & 0 \\ (ML)^H & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z \\ f \\ r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ Mh \\ 0 \end{bmatrix}$$

where the vector $(z^H, r_1^H, r_2^H)^H$ is a vector of Lagrange multipliers. Now let U_2 be left-unitary and $R(U_2) = N(C)$, then it is easy to verify (cf. Eldén [2, corr. 3.3]) that

$$(6.1.7)\quad \begin{pmatrix} S^H S & C^H \\ C & 0 \end{pmatrix}^+ = \begin{pmatrix} U_2((SU_2)^H SU_2)^+ U_2{}^H & C_{IS}^+ \\ (C_{IS}^+)^H & -(SC_{IS}^+)^H SC_{IS}^+ \end{pmatrix},$$

where $C_{IS}^+ = (I - U_2(SU_2)^+S)C^+$.

In the same way as before, we find

$$(6.1.8)\quad C^+ = \begin{pmatrix} I & ML \\ (ML)^H & 0 \end{pmatrix}^+ = \begin{pmatrix} I - (ML)(ML)^+ & (ML)^{+H} \\ (ML)^+ & -(ML)^+(ML)^{+H} \end{pmatrix}.$$

We know that $U_2U_2{}^H = I - C^+C = I - CC^+$ (since C is Hermitian), and so

$$U_2U_2{}^H = \begin{pmatrix} 0 & 0 \\ 0 & I - (ML)^+ML \end{pmatrix} = \begin{pmatrix} 0 \\ P_2 \end{pmatrix} (0 \mid P_2{}^H),$$

with P_2 a left-unitary matrix such that $R(P_2) = N(ML)$.

So, we can take $U_2 = \begin{pmatrix} 0 \\ P_2 \end{pmatrix}$.

By combining the previous results we find for $\begin{pmatrix} S^H S & C^H \\ C & 0 \end{pmatrix}^+$ the formula displayed in fig. 1.

Consequently, the solution of problem (6.1.1) (and so of (1.1.2)) with minimum 2-norm is again found to be

$$f = (I - P_2(KP_2)^+K)(ML)^+Mh ,$$

where P_2 is left-unitary and $R(P_2) = N(ML)$.

$$\begin{pmatrix} \bar{S}^H S & C^H \\ C & 0 \end{pmatrix}^+ =$$

$$\begin{bmatrix} 0 & 0 & (C^+)_{11} & (C^+)_{12} \\ 0 & P_2((KP_2)^H KP_2)^+ P_2{}^H & (I - P_2(KP_2)^+K)(C^+)_{21} & (I - P_2(KP_2)^+K)(C^+)_{22} \\ (C^+)_{11} & (C^+)_{12}(I - P_2(KP_2)^+K)^H & -(C^+)_{12}K^H(I - KP_2(KP_2)^+)K(C^+)_{21} & -(C^+)_{12}K^H(I - KP_2(KP_2)^+)K(C^+)_{22} \\ (C^+)_{21} & (C^+)_{22}(I - P_2(KP_2)^+K)^H & -(C^+)_{22}K^H(I - KP_2(KP_2)^+)K(C^+)_{21} & -(C^+)_{22}K^H(I - KP_2(KP_2)^+)K(C^+)_{22} \end{bmatrix}$$

with $C^+ = \begin{pmatrix} I & ML \\ (ML)^H & 0 \end{pmatrix}^+ = \begin{pmatrix} I - ML(ML)^+ & (ML)^{+H} \\ ML^+ & -(ML)^+(ML)^{+H} \end{pmatrix}$

fig. 1

6.2. The method of Lagrange multipliers can also be applied directly to problem (1.1.2),

"find f that minimizes $\{\|Kf\| \mid \|M(Lf - h)\| \text{ is minimal}\}$."

The "inner" problem is easy to solve, since

$$\|M(Lf - h)\| \text{ is minimal iff } MLf = ML(ML)^{+}Mh.$$

Hence, problem (1.1.2) becomes

(6.2.1) "find f that minimizes $\{\|Kf\|^2 \mid MLf = ML(ML)^{+}Mh\}$."

The corresponding Lagrange functional is

$$L(f,r) = \tfrac{1}{2}f^H K^H Kf + r^H(MLf - ML(ML)^{+}Mh) .$$

Then the solution f of the system

$$\begin{pmatrix} K^H K & (ML)^H \\ ML & 0 \end{pmatrix} \begin{pmatrix} f \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ ML(ML)^{+}Mh \end{pmatrix}$$

satisfies problem (6.2.1).

Similar to (6.1.7) it is easy to verify that

$$\begin{pmatrix} K^H K & (ML)^H \\ ML & 0 \end{pmatrix}^{+} = \begin{pmatrix} -P_2((KP_2)^H KP_2)^{+}P_2 & (ML)^{+}_{IK} \\ ((ML)^{+}_{IK})^H & (K(ML)^{+}_{IK})^H K(ML)^{+}_{IK} \end{pmatrix},$$

where $(ML)^{+}_{IK} = (I - P_2(KP_2)^{+}K)(ML)^{+}$ and P_2 is left-unitary with $R(P_2) = N(ML)$. So again we find for the solution of problem (1.1.2) with minimum 2-norm,

$$f = (ML)^{+}_{IK}ML(ML)^{+}Mh = (I - P_2(KP_2)^{+}K)(ML)^{+}Mh.$$

Appendix. Equivalence of problems (2.1.1) and (2.1.2)

1. We start from problem (2.1.1), i.e.

"estimate x from a realisation of y = Cx + w."

Here C is a known matrix; w is a stochast with $\mathcal{E}(w) = 0$, $\mathcal{E}(ww^H) = \sigma^2 W$, where W is a nonnegative definite Hermitian matrix. As in §2 we set $W = BB^H$ with $B \in RR$, and $w = Bv$, then v is a stochast with $\mathcal{E}(v) = 0$, $\mathcal{E}(vv^H) = \sigma^2 I$.

Definition. A linear function $\varphi : x \to p^H x$ is called estimable if there is a linear function $\psi : y \to q^H y$ such that the stochastic vector $y = Cx + Bv$ satisfies $\mathcal{E}(q^H y) = p^H x$.
Then $\psi(y)$ is called linear unbiased estimator (LUE) of $\varphi(x)$.

Since for arbitrary q we have $\mathcal{E}(q^H y) = q^H Cx$, φ is estimable iff $p \in R(C^H)$, and $\psi(y) = q^H y$ is a LUE of $\varphi(x) = p^H x$ iff $C^H q = p$. For any q satisfying the latter condition we have

$$\mathcal{E}((q^H y - p^H x)^2) = \mathcal{E}((q^H Bv)^2) = \sigma^2 q^H BB^H q = \sigma^2 q^H Wq \ .$$

Definition. $\hat{\psi} : y \to \hat{q}^H y$ is called best linear unbiased estimator (BLUE) of $\varphi(x) = p^H x$ if

$$\hat{q}^H W\hat{q} = \min\{ q^H Wq \mid C^H q = p\} \ .$$

Using a Lagrange multiplier $\hat{\ell}$, we find that $\hat{\psi}$ is BLUE of φ iff $\hat{q}$ and $\hat{\ell}$ satisfy

$$\text{(A.1.1)} \quad \begin{pmatrix} W & C \\ C^H & 0 \end{pmatrix} \begin{pmatrix} \hat{q} \\ \hat{\ell} \end{pmatrix} = \begin{pmatrix} 0 \\ p \end{pmatrix} .$$

Remark. It is clear from (A.1.1) that $W\hat{q} \in R(C)$ and the system is compatible iff $p \in R(C^H)$.

Definition. An observation y is called compatible iff $y \in R((B|C))$, so iff $y \in R((W|C))$.

<u>Lemma</u>. Consider the matrix $A = \begin{pmatrix} W & C \\ C^H & 0 \end{pmatrix}$, where $W = BB^H$.

Then $\begin{pmatrix} r \\ x \end{pmatrix} \in N(A)$ iff $r \in N\left(\left(\dfrac{B^H}{C^H}\right)\right) \wedge x \in N(C)$,

and $\begin{pmatrix} y \\ z \end{pmatrix} \in R(A)$ iff $y \in R((B|C)) \wedge z \in R(C^H)$.

<u>Proof</u>. $\begin{pmatrix} W & C \\ C^H & 0 \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = 0$ is equivalent to $\begin{cases} BB^H r + Cx = 0, \\ r^H C = 0. \end{cases}$

This implies

$$0 = r^H Cx = -r^H BB^H r,$$

so $B^H r = 0$, and consequently $Cx = 0$.

Therefore

$$\begin{pmatrix} r \\ x \end{pmatrix} \in N(A) \Leftrightarrow [r \in N\left(\left(\dfrac{B^H}{C^H}\right)\right) = (R((B|C)))^{\perp}] \wedge [x \in N(C) = (R(C^H))^{\perp}],$$

which implies that

$$\begin{pmatrix} y \\ z \end{pmatrix} \in R(A) \Leftrightarrow y \in R((B|C)) \wedge z \in R(C^H) \ . \qquad \square$$

The lemma implies that the system

$$\text{(A.1.2)} \quad \begin{pmatrix} W & C \\ C^H & 0 \end{pmatrix} \begin{pmatrix} \hat{r} \\ \hat{x} \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}$$

has a solution iff y is compatible. If y is compatible, $\hat{x}$ is to be determined from (A.1.2) modulo N(C).

Now let y be compatible and let $\varphi : x \to p^H x$ be estimable, then $\hat{\Phi}$, $\hat{r}$ and $\hat{x}$ are determined as above, and we have

$$\hat{\Phi}(y) = \hat{q}^H y = \hat{q}^H (W\hat{r} + C\hat{x}) = -\hat{\ell}^H C^H \hat{r} + p^H \hat{x} = p^H \hat{x} = \varphi(\hat{x}) \ .$$

So for a given compatible y and with $\hat{x}$ as above, the BLUE for any estimable function $\varphi : x \to p^H x$, is $p^H \hat{x}$. In particular, Cx is estimable and its BLUE is $\hat{y} := C\hat{x}$.

2. Now we start from problem (2.1.2), i.e.

   "find to a given $y \in R((C|B))$ vectors x and v that minimize $\|v\|$ under the condition $y = Cx + Bv$."

By definition y is compatible. Using a Lagrange multiplier r, we find that v, r and x have to satisfy the system

$$\text{(A.2.1)} \quad \begin{pmatrix} -I & B^H & 0 \\ B & 0 & C \\ 0 & C^H & 0 \end{pmatrix} \begin{pmatrix} v \\ r \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix} .$$

This system is equivalent to

$$\text{(A.2.2)} \quad \begin{pmatrix} W & C \\ C^H & 0 \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}, \quad v = B^H r.$$

So the solutions x and v of problem (2.1.2) are $x = \hat{x}$ and $v = B^H\hat{r}$, with $\hat{x}$ and $\hat{r}$ as determined in the first section of this appendix.

Conclusion. Since both problems ((2.1.1) and (2.1.2)) lead to the same solution set for x, the equivalence of both is evident.

For a more professional look into these matters, see e.g. [4, Ch. 3].

Acknowledgement. The first draft of this paper has been written as a student's paper under the guidance of Prof. G.W. Veltkamp. I am grateful to him for leaving me some notes on the subject which were the start of this work, for the stimulating discussions and finally, for his comments on the manuscript.

References.

[1] C.C. Paige, Computer solution and perturbation analysis of generalised linear least squares problems. Math. Comp. 33 (1979), 171-183.

[2] L. Eldén, Perturbation theory for the least squares problem with linear equality constraints. Report LITH-MAT-R-1977-17, Dept. of Math., Linköping University, Linköping, 1977.

[3] A. Ben-Israel and T.N.E. Greville, Generalized Inverses: Theory and Applications. Wiley, New York, 1974.

[4] G.A. Seber, Linear Regression Analysis. Wiley, New York, 1977.