On a Particular Case of the Inconsistent Linear Matrix Equation $AX + YB = C$

K. Zietak
Institute of Computer Science
University of Wrocław
Ul. Przesmyckiego 20
51-151 Wrocław, Poland

Submitted by Peter Lancaster

ABSTRACT

We consider the linear matrix equation $AX + YB = C$ where A, B, and C are given matrices of dimensions $(r+1)\times r$, $s\times(s+1)$, and $(r+1)\times(s+1)$, respectively, and rank $A = r$, rank $B = s$. We give a connection between the least-squares solution and the solution which minimizes an arbitrary norm of the residual matrix $C - AX - YB$.

1. INTRODUCTION

Let $\mathscr{M}_{mn}$ denote the space of real $m\times n$ matrices. We consider the linear matrix equation

$$AX + YB = C, \tag{1.1}$$

where $A\in\mathscr{M}_{mr}$, $B\in\mathscr{M}_{sn}$ and $C=(c_{ij})\in\mathscr{M}_{mn}$ are given. We may write the equation (1.1) in the form

$$Dx = d \tag{1.2}$$

with $D=(I_n\otimes A, B^T\otimes I_m)$, $D\in\mathscr{M}_{mn,rn+sm}$, and appropriate definitions of the vectors x and d, $x\in R^{rn+sm}$, $d\in R^{mn}$, where $\otimes$ denotes the Kronecker product and I_n is the identity matrix of order n. The equation (1.1) has a solution X and Y if and only if [1]

$$(I - AA^-)C(I - B^-B) = 0, \tag{1.3}$$

52 Vanderbilt Ave., New York, NY 10017 0024-3795/85/$3.30

where A^- and B^- are any g-inverses of A and B, respectively, i.e., $AA^-A = A$ and $BB^-B = B$. We denote

$$P_g = AA^-, \qquad Q_g = B^-B.$$

If the condition (1.3) is satisfied, then the general solution of (1.1) has the form [1]

$$X = A^-C - A^-ZB + (I - A^-A)W,$$
$$Y = (I - AA^-)CB^- + Z - (I - AA^-)ZBB^- \tag{1.4}$$

with $W \in \mathcal{M}_{rn}$ and $Z \in \mathcal{M}_{ms}$ arbitrary.

In the paper we assume that the condition (1.3) is not satisfied and we find a solution of (1.1) which minimizes an arbitrary norm of the residual matrix

$$R(X;Y) = C - AX - YB.$$

In particular, we may choose the l_p-norm for $1 \leqslant p \leqslant \infty$. Then the matrices X_p and Y_p are the l_p-solution of (1.1) if

$$\|C - AX_p - Y_pB\|_p = \delta_p = \min_{X,Y} \|C - AX - YB\|_p,$$

where

$$\|C\|_p = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |c_{ij}|^p \right)^{1/p} \qquad (1 \leqslant p \leqslant \infty).$$

We denote

$$R_p = \left(r_{ij}^{(p)}\right) = R(X_p; Y_p).$$

The least-squares solution and the Chebyshev solution correspond to the values $p = 2$ and $p = \infty$, respectively. The properties of the Chebyshev solution and the l_p-solution for $1 < p < \infty$ and for $m > r$, $n > s$ were investigated in [10] and [9], respectively.

We may reduce the number of unknowns in (1.1) imposing additional conditions on some of the unknowns in (1.1) which do not change the residual matrix $R(X;Y)$. Thus, for $m > r$ and $n > s$ the number of single equations in (1.1) is greater than the number of the remaining unknowns (see [10]). In this paper, however, we impose no additional conditions to reduce the number of unknowns in (1.1).

The main purpose of this paper is to present the relations between the least-squares solution and the solution which minimizes an arbitrary norm of the residual matrix under the assumptions

$$m = r+1, \quad n = s+1, \quad \operatorname{rank} A = r, \quad \operatorname{rank} B = s. \tag{1.5}$$

This special case of (1.1) plays an important role in studying the properties of Chebyshev solution of (1.1) with arbitrary m and n (see [10]).

2. MAIN RESULT

The matrix C may be interpreted as an element c of the vector space $\mathscr{R}^{mn}$:

$$c = \left(c_{11},\dots,c_{1n},\dots,c_{m1},\dots,c_{mn}\right)^T.$$

Let the space $\mathscr{R}^{mn}$ be equipped with an arbitrary vector norm $\|\cdot\|$, and let $\|C\|$ be equal to the norm of the vector c. Together with the norm $\|\cdot\|$, we consider also the dual norm $\|\cdot\|^*$ determined in the following way:

$$\|C\|^* = \max_{\substack{W \\ \|W\|=1}} \langle C, W\rangle,$$

where $W = (w_{ij})$, $W \in \mathscr{M}_{mn}$, and

$$\langle C, W\rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} w_{ij}.$$

If $1/p + 1/q = 1$, then the l_q-norm is the dual norm of the l_p-norm ($1 \leqslant p \leqslant \infty$). The matrix C^*, $C^* \in \mathscr{M}_{mn}$, such that $\|C^*\| = 1$ and $\langle C^*, C\rangle = \|C\|^*$ is called a dual matrix to $C \neq 0$.

We consider the following problem. For the given matrices A, B, and C and for a given vector norm $\|\cdot\|$, find matrices $\tilde{X}$ and $\tilde{Y}$ such that

$$\|R(\tilde{X};\tilde{Y})\| = \tilde{\delta} = \min_{X,Y} \|R(X;Y)\|. \tag{2.1}$$

Since we assume that the condition (1.3) is not satisfied, the error $\tilde{\delta}$ is nonzero: $\tilde{\delta} > 0$. The solution of (2.1) is not unique.

The problem (2.1) is related to the discrete approximation of a function $f(\xi, \eta)$ in two variables over a discrete point set

$$\{(\xi_i, \eta_j) : i = 1, \dots, m;\ j = 1, \dots, n\}$$

by functions of the form

$$\sum_{k=1}^{r} a_k(\xi) x_k(\eta) + \sum_{l=1}^{s} y_l(\xi) b_l(\eta),$$

where $a_k(\xi)$ and $b_l(\eta)$ are given functions. When $m = r + 1$ and $n = s + 1$ we have the simplest case of such approximation.

We define the following set of matrices (the subdifferential of $\|R(X;Y)\|$):

$$\mathscr{V}(\|R(X;Y)\|) = \{W : W \in \mathscr{M}_{mn}, \|W\|^* \leq 1, \|R(X;Y)\| = \langle R(X;Y), W \rangle\}.$$

Now we formulate the theorem which states the characterization of the solution of the problem (2.1).

THEOREM 2.1. *The matrices X and Y are a solution of the problem* (2.1) *if and only if there exists a matrix* $U \in \mathscr{V}(\|R(X;Y)\|)$ *such that*

$$U^T A = 0, \qquad UB^T = 0. \tag{2.2}$$

We omit the proof of the theorem because it follows immediately from Theorem 1.7 given in [8, p. 16] and applied to the equation (1.2). Theorem 2.1 generalizes the characterizations of the Chebyshev solution and the l_p-solution ($1 < p < \infty$) of the equation (1.1), which were given in [10] and [9], respectively.

For arbitrary m and n the l_p-solution of (1.1) is given explicitly only for $p = 2$. The matrices X_2 and Y_2 are the least-squares solution of (1.1) if and only if (see [9]; compare [6])

$$R_2 = (I - P)C(I - Q), \tag{2.3}$$

where $P = AA^-$, $Q = B^T(B^T)^-$ and A^-, $(B^T)^-$ are symmetric g-inverses of A and B^T, respectively. This means that P and Q are symmetric and

$$A^T(I - P) = 0, \qquad B(I - Q) = 0. \tag{2.4}$$

LEMMA 2.1. *Let the assumptions* (1.5) *be satisfied, and let the vectors* w *and* u *satisfy*

$$w^T w = 1, \qquad u^T u = 1, \tag{2.5}$$

$$A^T w = 0, \qquad Bu = 0. \tag{2.6}$$

Then

$$I - P = ww^T, \qquad I - Q = uu^T, \tag{2.7}$$

$$R_2 = \gamma w u^T, \tag{2.8}$$

$$\delta_2 = |\gamma|, \tag{2.9}$$

where

$$\gamma = w^T C u,$$

and there exist nonzero vectors v *and* z *such that*

$$I - P_g = v w^T, \qquad I - Q_g = u z^T. \tag{2.10}$$

Proof. Since rank $A = r$ and rank $B = s$, it follows that $\ker(A^T)$ and $\ker(B)$ are one-dimensional and are spanned by the vectors w and u, respectively [see (2.6)]. Moreover, we have (see [2, p. 16])

$$\operatorname{rank} P = \operatorname{rank} P_g = r,$$

$$\operatorname{rank} Q = \operatorname{rank} Q_g = s,$$

and consequently

$$\operatorname{rank}(I - P) = \operatorname{rank}(I - Q) = \operatorname{rank}(I - P_g) = \operatorname{rank}(I - Q_g) = 1.$$

Therefore the expressions (2.7) are true, because P and Q are symmetric and the relations (2.4) and (2.5) hold.

The formula (2.8) follows immediately from (2.3) and (2.7). From (2.5) and (2.8) we obtain

$$\delta_2 = \|R_2\|_2 = \left[\gamma^2 (wu^T)^T wu^T\right]^{1/2} = |\gamma|,$$

so (2.9) holds.

Now, by the definitions of P_g and Q_g we have

$$(I - P_g)A = 0, \qquad B(I - Q_g) = 0.$$

Thus the rows of $I - P_g$ belong to $\ker(A^T)$ and the columns of $I - Q_g$ belong to $\ker(B)$. Since the matrices $I - P_g$ and $I - Q_g$ have rank 1, there exist vectors v and z such that (2.10) holds, which completes the proof. ■

From Lemma 2.1 we obtain that if the assumptions (1.5) are satisfied then

$$\langle R(X;Y), R_2\rangle = \langle C, R_2\rangle. \tag{2.11}$$

LEMMA 2.2. *Let* $F = (f_{ij}) \in \mathcal{M}_{mn}$, *and let the assumptions* (1.5) *be satisfied. If the condition* (1.3) *is not satisfied and*

$$\langle F, R_2\rangle = 0, \tag{2.12}$$

then the equation $AX + YB = F$ *has a solution.*

Proof. Let the vectors w, u, v, and z satisfy (2.5), (2.6), and (2.10). Then

$$(I - P_g)F(I - Q_g) = vw^TFuz^T = \alpha vz^T, \tag{2.13}$$

where $\alpha = w^TFu$. From (2.8) we obtain

$$\langle F, R_2\rangle = \gamma w^TFu = \gamma\alpha. \tag{2.14}$$

Because the condition (1.3) is not satisfied, we have $R_2 \neq 0$ and consequently $\gamma \neq 0$ [see (2.9)]. From (2.12) and (2.14) it follows that $\alpha = 0$. Therefore for the equation $AX + YB = F$ the condition (1.3) is satisfied [see (2.13)], which completes the proof. ■

Now we prove the theorem which determines the connection between the solution of the problem (2.1) for an arbitrary vector norm and the least-squares solution under the assumption (1.5). This theorem is an extension of Sreedharan's theorem concerning an overdetermined system of $n + 1$ linear equations in n unknowns (see [7], [5]).

THEOREM 2.2. *Let the assumptions* (1.5) *be satisfied, and assume that the condition* (1.3) *does not hold. Then the equation*

$$AX + YB = C - \frac{\langle C, R_2 \rangle}{\|R_2\|^*} R_2^* \tag{2.15}$$

has a solution, and any solution of (2.15) *is a solution of the problem* (2.1). *Moreover, the error* $\tilde{\delta}$ *is equal to*

$$\tilde{\delta} = \frac{\langle C, R_2 \rangle}{\|R_2\|^*}. \tag{2.16}$$

REMARK. If the condition (1.3) holds, then the equation $AX + YB = C$ has a solution which is also a solution of (2.1).

Proof. Since the condition (1.3) is not satisfied, we have $R_2 \neq 0$. First we show that the equation (2.15) has a solution. For this purpose we apply Lemma 2.2. From the definition of the dual matrix we have

$$\|R_2\|^* = \langle R_2^*, R_2 \rangle \quad \text{and} \quad \|R_2^*\| = 1.$$

Therefore

$$\left\langle C - \frac{\langle C, R_2 \rangle}{\|R_2\|^*} R_2^*, R_2 \right\rangle = 0.$$

Thus the condition (2.12) is satisfied for

$$F = C - \frac{\langle C, R_2 \rangle}{\|R_2\|^*} R_2^*,$$

which means that the equation (2.15) has a solution.

Now, we verify that each solution of (2.15) is a solution of the problem (2.1). Let

$$\rho = \frac{\langle C, R_2 \rangle}{\|R_2\|^*}.$$

From (2.8) we obtain that $\rho > 0$, because $\langle C, R_2\rangle = \gamma^2$. Let X and Y be arbitrary, $X \in \mathcal{M}_{r,s+1}$ and $Y \in \mathcal{M}_{r+1,s}$. Then from (2.11) and by the definition of the dual norm (we recall that $\|R_2^*\| = 1$) we have

$$\|R(X;Y)\| \geqslant \frac{\langle R(X;Y), R_2\rangle}{\|R_2\|^*} = \frac{\langle C, R_2\rangle}{\|R_2\|^*} = \rho. \qquad (2.17)$$

Therefore

$$\tilde{\delta} = \min_{X,Y} \|R(X;Y)\| \geqslant \rho.$$

For the matrices X and Y, which are the solution of (2.15), we obtain equality in (2.17). So they are the solution of the problem (2.1), and the error $\tilde{\delta}$ is equal to ρ, which completes the proof. ■

In Sreedharan's theorem, the matrix of an overdetermined system of $n+1$ linear equations in n unknowns is assumed to have rank n. In Theorem 2.2, however, we only assume that A and B have full rank.

Let the assumptions of Theorem 2.2 be satisfied. Then for the solution $\tilde{X}$ and $\tilde{Y}$ of the problem (2.1) we obtain [see (2.16) and (2.11)]

$$\|R(\tilde{X};\tilde{Y})\| = \frac{\langle C, R_2\rangle}{\|R_2\|^*} = \frac{\langle R(\tilde{X};\tilde{Y}), R_2\rangle}{\|R_2\|^*}.$$

Therefore the following corollary is valid.

COROLLARY 2.1. *Let the assumptions of Theorem 2.2 be satisfied. Then the matrix*

$$U = \frac{1}{\|R_2\|^*} R_2$$

belongs to $\mathcal{V}(\|R(\tilde{X};\tilde{Y})\|)$, *where the matrices* $\tilde{X}$ *and* $\tilde{Y}$ *are the solution of* (2.1).

3. CONCLUSIONS

Now, we consider the l_∞-norm. Then the l_1-norm is the dual norm. Let $x = (x_1, \dots, x_n)^T \in \mathcal{R}^n$. Then the vector $x^* = (x_1^*, \dots, x_n^*)^T$ defined by $x_i^* = \operatorname{sign} x_i$ is the dual vector to the vector x. From Theorem 2.2 we have the following corollary for the l_∞-norm, i.e. for the Chebyshev norm.

COROLLARY 3.1. *Let the assumptions of Theorem 2.2 be satisfied. Then the equation*

$$AX + YB = C - \rho S, \tag{3.1}$$

where $S = (s_{ij})$, $s_{ij} = \operatorname{sign}(r_{ij}^{(2)})$, $\rho = \langle C, R_2 \rangle / \|R_2\|_1$, *has a solution, and any solution of* (3.1) *is a Chebyshev solution of* $AX + YB = C$ *and* $\delta_\infty = \rho$.

A similar corollary may be formulated for the l_p-norm for $1 \leqslant p < \infty$. We can compute the Chebyshev solution of $AX + YB = C$ under the assumptions (1.5) by means of the formulae (1.4) applied to the equation (3.1).

We introduce auxiliary vectors $\hat{w} = (\hat{w}_1, \dots, \hat{w}_{r+1})^T$ and $\hat{u} = (\hat{u}_1, \dots, \hat{u}_{s+1})^T$ with

$$\hat{w}_i = (-1)^i \det A_i, \qquad \hat{u}_j = (-1)^j \det B_j,$$

where A_i and B_j are obtained from A and B by deletion of the ith row and the jth column, respectively. Then there exist scalars α and β such that (see [4])

$$w = \alpha \hat{w}, \qquad u = \beta \hat{u}, \tag{3.2}$$

where w and u are determined as in Lemma 2.1. Let the assumptions of Theorem 2.2 be satisfied. Let $\hat{\gamma} = \langle C, \hat{w}\hat{u}^T \rangle / \|\hat{w}\hat{u}^T\|_1$. We know (see [10]) that $\delta_\infty = |\hat{\gamma}|$ and the matrices X_∞ and Y_∞ are a Chebyshev solution of (1.1) if and only if

$$r_{ij}^{(\infty)} = \operatorname{sign}(\hat{w}_i \hat{u}_j \hat{\gamma}) \|R(X_\infty; Y_\infty)\|_\infty \tag{3.3}$$

for all pairs (i, j) such that $\hat{w}_i \hat{u}_j \neq 0$. The expression on the right side of (3.3) is equal [see (3.2) and Lemma 2.1] to

$$\operatorname{sign}\left(r_{ij}^{(2)}\right) \rho,$$

where ρ is determined as in (3.1). Hence, we have the following corollary.

COROLLARY 3.2. *Let the assumptions of Theorem 2.2 be satisfied. Then for each Chebyshev solution of* $AX + YB = C$ *we have*

$$r_{ij}^{(\infty)} = \operatorname{sign}\left(r_{ij}^{(2)}\right) \delta_\infty$$

for (i, j) *such that* $r_{ij}^{(2)} \neq 0$.

The solution of (3.1) is a strict Chebyshev solution of (1.1) under the assumption (1.5) (see e.g. [3] for the definition of the strict Chebyshev solution). We can obtain the other Chebyshev solutions of (1.1) modifying the definition of S in (3.1) in the following way:

$$s_{ij} = \begin{cases} \operatorname{sign}\left(r_{ij}^{(2)}\right) & \text{for} \quad r_{ij}^{(2)} \neq 0, \\ h_{ij} & \text{for} \quad r_{ij}^{(2)} = 0, \end{cases} \tag{3.4}$$

where $|h_{ij}| \leqslant 1$. The matrix S determined as in (3.4) is the dual matrix to R_2. Therefore each solution of (3.1) in this case is a Chebyshev solution of (1.1) (see Theorem 2.2).

The author is indebted to Professor P. Lancaster from the University of Calgary for his comments and suggestions, which improved the presentation.

REFERENCES

1 J. K. Baksalary and R. Kala, The matrix equation $AX - YB = C$, *Linear Algebra Appl.* 25:41–43 (1979).
2 A. Ben-Israel and Th. N. E. Greville, *Generalized Inverses. Theory and Applications*, Wiley-Interscience, New York, 1973.
3 J. Descloux, Approximation in L^p and Chebyshev approximations, *J. Soc. Indust. Appl. Math.* 11:1017–1026 (1963).
4 C. S. Duris and V. P. Sreedharan, Chebyshev and l_1-solutions of linear equations using least squares solutions, *SIAM J. Numer. Anal.* 5:491–505 (1968).
5 M. L. Levitan and R. Y. S. Lynn, An overdetermined linear system, *J. Approx. Theory* 18:264–277 (1976).
6 C. Radakrishna Rao, Matrix approximation and reduction of dimensionality in multivariate statistical analysis, in *Multivariate Analysis V* (P. R. Krishnaiah, Ed.), North-Holland, Amsterdam, 1980.
7 V. P. Sreedharan, Solutions of overdetermined linear equations which minimize error in an abstract norm, *Numer. Math.* 13:146–151 (1969).
8 G. A. Watson, *Approximation Theory and Numerical Methods*, Wiley, London, 1980.
9 K. Ziętak, The l_p-solution of the linear matrix equation $AX + YB = C$, *Computing*, to appear.
10 K. Ziętak, The Chebyshev solution of the linear matrix equation $AX + YB = C$, submitted to *Numer. Math.*; see also Report N-121, Institute of Computer Science, Univ. of Wrocław, Wrocław, 1983.

Received 5 December 1983; revised 17 April 1984