Journal of Mathematical Modeling
Vol. 2, No. 2, 2015, pp. 132-142

JMM

On the optimal correction of inconsistent matrix equations $AX = B$ and $XC = D$ with orthogonal constraint

Saeed Ketabchi, Elham Samadi and Hossein Aminikhah*

Department of Applied Mathematics, School of Mathematical Sciences, University of Guilan, P.O. Box 1914, Rasht, Iran
Emails: sketabchi@guilan.ac.ir, elhamsamadi65@yahoo.com, aminikhah@guilan.ac.ir

Abstract. This work focuses on the correction of both the coefficient and the right hand side matrices of the inconsistent matrix equations $AX = B$ and $XC = D$ with orthogonal constraint. By optimal correction approach, a general representation of the orthogonal solution is obtained. This method is tested on two examples to show that the optimal correction is effective and highly accurate.

Keywords : Matrix equations, optimal correction, generalized Lagrange function, orthogonal constraint, Kuhn-Tucker conditions.
AMS Subject Classification : 90C05, 90C51.

1 Introduction

Solving the matrix equations

$$AX = B,\ XC = D, \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times p}$, $C \in \mathbb{R}^{p\times l}$, $D \in \mathbb{R}^{n\times l}$, is one of the important study field in linear algebra. The matrix equations (1) arise in

*Corresponding author.
Received: 3 February 2014 / Revised: 5 July 2014 / Accepted: 27 July 2014.

engineering and in some special matrix inverse problems [4, 6]. Many authors have worked on this problem, and a series of useful results have been obtained. For example, Yun [7] has presented an explicit representation of least squares solutions by the matrix differentiation and the singular value decompositions (SVD). In [3] Qiu considered the least squares solutions to the matrix equations (1) with some constraints such as orthogonality, symmetric orthogonality and symmetric. Chu [1] and Mitra [2] considered the necessary and sufficient conditions for the solvability and general solution by using SVD and generalized inverse of a matrix, respectively. The main purpose of this article is the optimal correction of inconsistent matrix equations with changes in the coefficients matrices and right-hand side matrices.

In this paper, we shall adopt the following notation. $\mathbb{R}^{m\times n}$ denotes the set of all $m \times n$ real matrices, $\mathbb{OR}^{n\times n}$ denotes the set of all orthogonal matrices in $\mathbb{R}^{n\times n}$. A^T, $tr(A)$ and $\|A\|$ stand for the transpose, the trace and the Frobenius norm of a real matrix A, respectively. For $A, B \in \mathbb{R}^{m\times n}$, we define the inner product in $\mathbb{R}^{m\times n}$: $\langle A, B\rangle = tr(B^T A)$. It is known that the matrix norm $\|.\|$ induced by this inner product is the Frobenius norm.

The rest of the paper is organized as follows. Theoretical considerations are discussed in Sections 3. Section 4 is devoted to numerical experiments. The paper is ended by some concluding remarks in Section 5.

2 Changes in the coefficients matrices of the matrix equations (1) with orthogonality constraints

We consider the matrix equations (1) with orthogonal constraint and apply the changes in the coefficient matrices. The simplest kind of changes occurs when the coefficient matrices are give with some errors. The goal of this section is to find the matrices E_1^* and E_2^* with minimum norm that matrix equations $(A + E_1)X = B$ and $X(C + E_2) = D$ with orthogonal constraint are consistent. In other words, we are faced with the following constraint optimization problem

$$\min_{X,E_1,E_2} \|E_1\|^2 + \|E_2\|^2 \tag{2}$$

$$(A + E_1)X = B,$$

$$X(C + E_2) = D,$$

$$X^T X = XX^T = I.$$

The optimization problem (2) is equivalent to

$$\min_{X} \min_{E_1,E_2} \|E_1\|^2 + \|E_2\|^2 \tag{3}$$
$$(A + E_1)X = B,$$
$$X(C + E_2) = D,$$
$$X^T X = XX^T = I.$$

So with the help of generalized Lagrange function, we gain the optimal solution of the internal and external problems. Therefore, we introduce the following Lemmas.

Lemma 1. *([5]) Let A, W and U be three real-valued matrices and $\widetilde{A}$ be an unknown variable matrix, then*

1. $\frac{\partial tr(W^T A^T U^T U \widetilde{A} W)}{\partial \widetilde{A}} = U^T UAWW^T.$
2. $\frac{\partial tr(W^T \widetilde{A}^T U^T UAW)}{\partial \widetilde{A}} = U^T UAWW^T.$
3. $\frac{\partial tr(W^T \widetilde{A}^T U^T U \widetilde{A} W)}{\partial \widetilde{A}} = 2U^T U\widetilde{A}WW^T.$

Lemma 2. *([3]) Let $W = A^T B + DC^T$ and the singular value decomposition of W be $W = U diag(\Sigma_r, O_{n-r})V^T$, where*

$$\Sigma_r = diag(\delta_1, \ldots, \delta_r), \quad 0 < \delta_r < \cdots < \delta_1,$$
$$r = rank(W), \quad U, V \in \mathbb{OR}^{n\times n}.$$

Then, the orthogonal solution for the problem $\min_X \|AX - B\|^2 + \|XC - D\|^2$ *can be expressed as*

$$X = U \begin{bmatrix} I_r & 0 \\ 0 & G \end{bmatrix} V^T,$$

where the matrix $G \in \mathbb{OR}^{n-r}$ is arbitrary.

Theorem 1. *Consider the following constrained optimization problem*

$$\min_{E_1,E_2} \|E_1\|^2 + \|E_2\|^2 \tag{4}$$
$$(A + E_1)X = B,$$
$$X(C + E_2) = D,$$
$$X^T X = XX^T = I.$$

Then, the optimal solutions of (4) can be expressed as

$$E_1^* = BX^T - A, \qquad E_2^* = X^T D - C.$$

Proof. Clearly, the problem (4) is convex. Now, with the help of the Kuhn-Tucker conditions we obtain the optimal solution of (4) and the generalized Lagrange function will be as the following relation:

$$\begin{aligned} L(E_1, E_2, \lambda_1, \lambda_2) &= \|E_1\|^2 + \|E_2\|^2 + < \lambda_1^T, (A + E_1)X - B > \\ &\quad + < \lambda_2^T, X(C + E_2) - D >, \end{aligned}$$

where $\lambda_1 \in \mathbb{R}^{p\times m}$ and $\lambda_2 \in \mathbb{R}^{l\times n}$. According to the definition of the inner product and Lemma 1, the Kuhn-Tucker conditions are as follows:

$$\nabla_{E_1} L(E_1^*, E_2^*, \lambda_1^*, \lambda_2^*) = 2E_1^* + (X\lambda_1^*)^T = 0, \tag{5}$$

$$\nabla_{E_2} L(E_1^*, E_2^*, \lambda_1^*, \lambda_2^*) = 2E_2^* + (\lambda_2^* X)^T = 0, \tag{6}$$

$$\nabla_{\lambda_1} L(E_1^*, E_2^*, \lambda_1^*, \lambda_2^*) = ((A + E_1^*)X - B)^T = 0, \tag{7}$$

$$\nabla_{\lambda_2} L(E_1^*, E_2^*, \lambda_1^*, \lambda_2^*) = (X(C + E_2^*) - D)^T = 0. \tag{8}$$

Substituting (5) and $X^T X = XX^T = I$ in (7) yields the following results:

$$(A - \frac{1}{2}\lambda_1^{*^T} X^T)X = B \quad \Rightarrow \quad \lambda_1^{*^T} = 2(AX - B), \quad E_1^* = BX^T - A.$$

From (6) and (8), we have

$$X(C - \frac{1}{2}X^T \lambda_2^{*^T}) = D \quad \Rightarrow \quad \lambda_2^{*^T} = 2(XC - D), \quad E_2^* = X^T D - C.$$

Therefore, the proof is completed. □

The optimization problem (3) can be written as

$$\min_X \|BX^T - A\|^2 + \|X^T D - C\|^2 \tag{9}$$
$$X^T X = XX^T = I.$$

Now, we can find the least squares orthogonal solution of (9) by applying Lemma 2.

Theorem 2. *Let the matrix equations (1) with orthogonal constraint are not consistent then the orthogonal solution of matrix equations*

$$(A + E_1^*)X = B \quad \textit{and} \quad X(C + E_2^*) = D,$$

where

$$E_1^* = BX^T - A \quad \textit{and} \quad E_2^* = X^T D - C,$$

can be expressed as

$$X = V \begin{bmatrix} I_r & 0 \\ 0 & G^T \end{bmatrix} U^T,$$

where the matrix $G \in \mathbb{OR}^{n-r}$ *is arbitrary.*

Proof. Usnig the Lemma 2.2 and theorem 2.3, the above theorem can be easily proved. □

3 Change in the coefficient and right hand side matrices of the matrix equations (1)

Assume that the matrix equations (1) are not consistent, then we can simultaneously apply the suitable changes in the entries of the coefficient and the right hand side matrices. The main goal of this section is to find the matrices E_1^*, E_2^*, R_1^* and R_2^* with minimum norm such that matrix equations $(A+E_1)X = B+R_1$ and $X(C+E_2) = D+R_2$ be consistent. In other words, we will consider the following optimization problem,

$$\min_{X,E_1,E_2,R_1,R_2} \|E_1\|^2 + \|E_2\|^2 + \|R_1\|^2 + \|R_2\|^2 \tag{10}$$

$$\begin{aligned}(A+E_1)X &= B+R_1,\\ X(C+E_2) &= D+R_2.\end{aligned}$$

The optimization problem (10) is equivalent to

$$\min_{X} \min_{E_1,E_2,R_1,R_2} \|E_1\|^2 + \|E_2\|^2 + \|R_1\|^2 + \|R_2\|^2 \tag{11}$$

$$\begin{aligned}(A+E_1)X &= B+R_1,\\ X(C+E_2) &= D+R_2.\end{aligned}$$

By the Kuhn-Tucker conditions, we obtain the optimal solution of the internal problem. Then we solve the external problem.

Theorem 3. *Consider the following constrained optimization problem*

$$\min_{E_1,E_2,R_1,R_2} \|E_1\|^2 + \|E_2\|^2 + \|R_1\|^2 + \|R_2\|^2 \tag{12}$$

$$\begin{aligned}(A+E_1)X &= B+R_1,\\ X(C+E_2) &= D+R_2.\end{aligned}$$

The optimal solution of (12) can be expressed as

$$\begin{aligned}E_1^* &= -(AX-B)(X^TX+I)^{-1}X^T,\\ E_2^* &= -X^T(XX^T+I)^{-1}(XC-D),\\ R_1^* &= (AX-B)(X^TX+I)^{-1},\\ R_2^* &= (XX^T+I)^{-1}(XC-D).\end{aligned}$$

Proof. Clearly, the objective function and the constraints are differentiable and convex. Now, we apply the Kuhn-Tucker conditions to obtain the optimal solution of (12). The generalized Lagrange function is as follows:

$$L(E_1, E_2, R_1, R_2, \lambda_1, \lambda_2) = \|E_1\|^2 + \|E_2\|^2 + \|R_1\|^2 + \|R_2\|^2$$

$$+ < \lambda_1^T, (A + E_1)X - (B + R_1) > + < \lambda_2^T, X(C + E_2) - (D + R_2) >,$$

where $\lambda_1 \in \mathbb{R}^{n\times m}$ and $\lambda_2 \in \mathbb{R}^{l\times n}$. According to the definition of the inner product and Lemma 1, the Kuhn-Tucker conditions are as follows:

$$\nabla_{E_1} L(E_1^*, E_2^*, R_1^*, R_2^*, \lambda_1^*, \lambda_2^*) = 2E_1^* + (X\lambda_1^*)^T = 0 \Rightarrow E_1^* = -\frac{1}{2}{\lambda_1^*}^T X^T, \quad (13)$$

$$\nabla_{E_2} L(E_1^*, E_2^*, R_1^*, R_2^*, \lambda_1^*, \lambda_2^*) = 2E_2^* + (\lambda_2^* X)^T = 0 \Rightarrow E_2^* = -\frac{1}{2}X^T {\lambda_2^*}^T, \quad (14)$$

$$\nabla_{R_1} L(E_1^*, E_2^*, R_1^*, R_2^*, \lambda_1^*, \lambda_2^*) = 2R_1^* - {\lambda_1^*}^T = 0 \Rightarrow R_1^* = \frac{1}{2}{\lambda_1^*}^T, \quad (15)$$

$$\nabla_{R_2} L(E_1^*, E_2^*, R_1^*, R_2^*, \lambda_1^*, \lambda_2^*) = 2R_2^* - {\lambda_2^*}^T = 0 \Rightarrow R_2^* = \frac{1}{2}{\lambda_2^*}^T, \quad (16)$$

$$\nabla_{\lambda_1} L(E_1^*, E_2^*, R_1^*, R_2^*, \lambda_1^*, \lambda_2^*) = ((A + E_1^*)X - (B + R_1^*))^T = 0, \quad (17)$$

and

$$\nabla_{\lambda_2} L(E_1^*, E_2^*, R_1^*, R_2^*, \lambda_1^*, \lambda_2^*) = (X(C + E_2^*) - (D + R_2^*))^T = 0. \quad (18)$$

Substituting (13) and (15) into (17), the following results yield

$$\begin{aligned} \lambda_1^* &= 2(AX - B)(X^T X + I)^{-1}, \\ E_1^* &= -(AX - B)(X^T X + I)^{-1} X^T, \\ R_1^* &= (AX - B)(X^T X + I)^{-1}. \end{aligned}$$

From (14), (16) and (18) , we have

$$\begin{aligned} \lambda_2^* &= 2(XX^T + I)^{-1}(XC - D), \\ E_2^* &= -X^T(XX^T + I)^{-1}(XC - D), \\ R_2^* &= (XX^T + I)^{-1}(XC - D), \end{aligned}$$

which completes the proof. □

Suppose that

$$\begin{aligned} E_1^* &= -(AX - B)(X^T X + I)^{-1} X^T, \\ R_1^* &= (AX - B)(X^T X + I)^{-1}, \\ E_2^* &= -X^T(XX^T + I)^{-1}(XC - D), \\ R_2^* &= (XX^T + I)^{-1}(XC - D), \end{aligned}$$

then the optimization problem of (11) is equivalent to

$$\min_X \|(AX-B)(X^TX+I)^{-1}X^T\|^2 + \|X^T(XX^T+I)^{-1}(XC-D)\|^2 + \tag{19}$$

$$\|(AX-B)(X^TX+I)^{-1}\|^2 + \|(XX^T+I)^{-1}(XC-D)\|^2.$$

Theorem 4. *Let the matrix equations (1) with orthogonal constraint are not consistent. Then the orthogonal solution of matrix equations* $(A+E_1^*)X = B+R_1^*$ *and* $X(C+E_2^*) = D+R_2^*$ *where,*

$$E_1^* = \frac{BX^T-A}{2}, \quad E_2^* = \frac{X^TD-C}{2}, \quad R_1^* = \frac{AX-B}{2} \quad \textit{and} \quad R_2^* = \frac{XC-D}{2}.$$

can be expressed as

$$X = U\begin{bmatrix} I_r & 0 \\ 0 & G \end{bmatrix}V^T,$$

where the matrix $G \in \mathbb{OR}^{n-r}$ *is arbitrary.*

Proof. Since the solution of the matrix equations (1) with orthogonal constraint are not consistent, thus the purpose is to find the matrices E_1^*, E_2^*, R_1^* and R_2^* with minimum norm such that matrix equations $(A+E_1)X = B+R_1$ and $X(C+E_2) = D+R_2$ with orthogonal constraint be consistent. In other words, we will consider the following optimization problem

$$\min_X \min_{E_1,E_2,R_1,R_2} \|E_1\|^2 + \|E_2\|^2 + \|R_1\|^2 + \|R_2\|^2 \tag{20}$$

$$(A+E_1)X = B+R_1,$$

$$X(C+E_2) = D+R_2,$$

$$X^TX = XX^T = I.$$

From theorem 4 and $X^TX = XX^T = I$, we easily obtain

$$E_1^* = \frac{BX^T-A}{2}, \quad E_2^* = \frac{X^TC-D}{2}, \quad R_1^* = \frac{AX-B}{2} \quad \text{and} \quad R_2^* = \frac{XC-D}{2}.$$

Thus the optimization problem of (20) is equivalent to

$$\min_X \frac{\|BX^T-A\|^2 + \|X^TD-C\|^2 + \|AX-B\|^2 + \|XC-D\|^2}{4} \tag{21}$$

$$X^TX = XX^T = I,$$

since

$$f(X) = \|A_1X-B_1\|^2 + \|XC_1-D_1\|^2,$$

where

$$A_1 = \begin{bmatrix} A \\ D^T \end{bmatrix}, \quad B_1 = \begin{bmatrix} B \\ C^T \end{bmatrix}, \quad C_1 = \begin{bmatrix} B^T & C \end{bmatrix} \text{ and } D_1 = \begin{bmatrix} A^T & D \end{bmatrix}.$$

Then we can find the orthogonal solution to (21) only by applying Lemma 2. Let the singular value decomposition of the matrix $W = A_1^T B_1 + D_1 C_1^T$ be $W = U\mathrm{diag}(\Sigma_r, O_{n-r})V^T$, where

$$\begin{aligned} \Sigma &= \mathrm{diag}(\delta_1, \ldots, \delta_r), \quad 0 < \delta_r < \cdots < \delta_1, \\ r &= \mathrm{rank}(W), \quad U, V \in \mathbb{OR}^{n\times n}. \end{aligned}$$

Then the orthogonal solution for problem (21) can be expressed as

$$X = U \begin{bmatrix} I_r & 0 \\ 0 & G \end{bmatrix} V^T,$$

where the matrix $G \in \mathbb{OR}^{n-r}$ is arbitrary. □

4 Numerical Examples

In order to show the effectiveness of the theory which discussed in Sections 1 and 2, we present some numerical examples. We use MATLAB 7.9.0 on a *Core 3 Duo 2.40 GHz* with main memory *4 GB*.

Example 1. Let $m = 4, n = 3, l = 3$. The matrices A, B, C and D are given by

$$A = \begin{bmatrix} 2.1835 & -0.3814 & 2.1835 \\ 2.9956 & -1.3592 & 2.9956 \\ -2.3806 & 0.6847 & -2.3806 \\ 2.1835 & -0.3814 & 2.1835 \end{bmatrix},$$

$$B = \begin{bmatrix} 2.9702 & 4.2476 & 2.9702 \\ -5.5247 & 5.2106 & -5.5247 \\ 0.6930 & -0.6821 & 0.6930 \\ 2.9702 & 4.2476 & 2.9702 \end{bmatrix},$$

$$C = \begin{bmatrix} 4.2335 & -0.4580 & 4.2335 \\ 1.7921 & 2.5081 & 1.7921 \\ 4.2335 & -0.4580 & 4.2335 \end{bmatrix},$$

The Matlab code of this paper is available from the authors upon request.

and

$$D = \begin{bmatrix} 0.9274 & -0.6481 & 0.9274 \\ -4.0500 & 1.6681 & -4.0500 \\ 0.9274 & -0.6481 & 0.9274 \end{bmatrix}.$$

According to Theorem 4, we obtain the minimum-norm least squares solution of Example 1 as follows

$$E_1^* = \begin{bmatrix} 0.3065 & -4.4442 & 0.3065 \\ 1.5249 & 8.2387 & 1.5249 \\ 1.6680 & -1.5143 & 1.6680 \\ 0.3065 & -4.4442 & 0.3065 \end{bmatrix},$$

$$E_2^* = \begin{bmatrix} -1.5799 & 0.6980 & -1.5799 \\ 0.1501 & -3.6795 & 0.1501 \\ -1.5799 & 0.6980 & 0.4201 \end{bmatrix},$$

and

$$X = \begin{bmatrix} -0.5799 & 0.6980 & 0.4201 \\ -0.6980 & -0.1599 & -0.6980 \\ 0.4201 & 0.6980 & -0.5799 \end{bmatrix}.$$

Therefore, we have

$$f = \|(A + E_1^*)X - B\|^2 + \|X(C + E_2^*) - D\|^2 = 8.9970 \times 10^{-30},$$

$$\|XX^T - I\| = 2.8201 \times 10^{-16}.$$

Example 2. Let $m = 4, n = 3, l = 3$. The matrices A, B, C and D are given by

$$A = \begin{bmatrix} 0.6141 & 0.5228 & 0.6141 \\ 0.2703 & -1.8830 & 0.2703 \\ 1.0305 & -0.3247 & 1.0305 \\ 0.6141 & 0.5228 & 0.6141 \end{bmatrix},$$

$$B = \begin{bmatrix} 4.5902 & 4.8104 & 4.5902 \\ -0.7471 & 3.7040 & -0.7471 \\ -3.1926 & -0.6889 & -3.1926 \\ 4.5902 & 4.8104 & 4.5902 \end{bmatrix},$$

$$C = \begin{bmatrix} -3.0123 & -0.3119 & -3.0123 \\ -0.8893 & 0.5521 & -0.8893 \\ -3.0123 & -0.3119 & -3.0123 \end{bmatrix},$$

and

$$D = \begin{bmatrix} -0.7835 & 1.1757 & -0.7835 \\ 1.3303 & -3.7520 & 1.3303 \\ -0.7835 & 1.1757 & -0.7835 \end{bmatrix}.$$

According to Theorem 2, we obtain the minimum-norm least squares solution of Example 2 as follows

$$E_1^* = \begin{bmatrix} 2.5425 & -0.5436 & 2.5425 \\ 0.2542 & -0.9040 & 0.2542 \\ -1.9921 & -0.7614 & -1.9921 \\ 2.5425 & -0.5436 & 2.5425 \end{bmatrix},$$

$$R_1^* = \begin{bmatrix} -1.9368 & -2.3920 & -1.9368 \\ 0.1295 & -0.9555 & 0.1295 \\ 1.9690 & 0.8732 & 1.9690 \\ -1.9368 & -2.3920 & -1.9368 \end{bmatrix},$$

$$E_2^* = \begin{bmatrix} 1.4288 & -0.0619 & 1.4288 \\ -0.4141 & 1.7526 & -0.4141 \\ 1.4288 & -0.0619 & 1.4288 \end{bmatrix},$$

$$R_2^* = \begin{bmatrix} -1.0471 & -0.6143 & -1.0471 \\ -1.4358 & 1.5247 & -1.4358 \\ -1.0471 & -0.6143 & -1.0471 \end{bmatrix},$$

and

$$X = \begin{bmatrix} -0.0785 & 0.3803 & 0.9215 \\ 0.3803 & -0.8431 & 0.3803 \\ 0.9215 & 0.3803 & -0.0785 \end{bmatrix}.$$

Therefore, we have

$$f = \|(A+E_1^*)X-(B+R_1^*)\|^2+\|X(C+E_2^*)-(D+R_2^*)\|^2 = 9.8219\times 10^{-30}.$$

$$\|XX^T - I\| = 6.0531 \times 10^{-16}.$$

5 Conclusion

In this work, we considered an optimal correction of inconsistent matrix equations with changes in the coefficients matrices and right-hand side matrices. We have discussed the methodology for the construction of these schemes and studied their performance on two test problems. We showed that the present approach is efficient and highly accurate.

Acknowledgments

We are very grateful to anonymous referees for their careful reading and valuable comments which led to the improvement of this paper.

References

[1] K. Chu, *Singular value and generalized singular value decomposition and the solution of linear matrix equations*, Linear Algebra Appl. **88/89** (1987) 83-98.

[2] S. K. Mitra, *A pair of simultaneos linear matrix equations* $A_1XB_1 = C_1, A_2XB_2 = C_2$ *and a matrix programming problem*, Linear Algebra Appl. **131** (1990) 107-123.

[3] Y. Qiu and A. Wang, *Least-squares solutions to the equations AX=B, XC=D with some constraint*, Appl. Math. Comput. **204** (2008) 872-880.

[4] F. Trench, *Hermitian R-symmetric, and hermitian R-skew symmetric Procrustes problems*, Linear Algebra Appl. **387** (2004) 83-98.

[5] Y. Wei, N. Zhang W. Xu and M.K. Ng, *Tikhonov regulaization for weighted total Least-squares problems*, Appl. Math. Lett. **20** (2007) 82-87.

[6] D. Xie, X. Hu and Z. Zhang, *The solvability conditions for the inverse eigenproblems of symmetric and anti-persymmetric matrices and its approximation*, Linear Algebra Appl. **10** (2003) 223-234.

[7] Y. Yuan, *Least-squares solutions to the matrix equations AX=B, XC=D* , Appl. Math. Comput. **216** (2010) 3120-3125.