A Matrix Optimization Problem

Olaf Krafft
Technical University
Aachen, West Germany

Submitted by Ingram Olkin

ABSTRACT

A minimization problem for a matrix-valued matrix function is considered. A duality theorem is proved. Some examples illustrate its applicability.

Let Σ_n be the set of real symmetric $n\times n$ matrices. In Σ_n a partial ordering—the Löwner ordering—is defined by $A\geqslant B$ iff $A-B$ is nonnegative definite. Using this ordering, investigations of monotonicity, convexity, and extremum properties of matrix functions with values in Σ_n have recently aroused considerable interest, mainly due to applications in statistics and electrical network theory; see e.g. [7], [5], [2], [8]. In this paper we will derive a primal-dual relation for a quadratic minimization problem under a linear constraint, and we will show that some results from the literature can be subsumed under this problem.

We will use the following terminology: For a real $m\times n$ matrix A we write $A\in\mathbb{R}^{m\times n}$; A^T is the transpose of A, and A^+ the Moore-Penrose inverse of A. I_n is the unit matrix in $\mathbb{R}^{n\times n}$. If $A\in\mathbb{R}^{n\times n}$, $A\geqslant 0$, $B\in\mathbb{R}^{q\times n}$, then $\bar{B}_A$ denotes a minimum-A-seminorm g-inverse of B, i.e., $\bar{B}_A\in\mathbb{R}^{n\times q}$ satisfies

$$B\bar{B}_AB=B \quad \text{and} \quad \left(\bar{B}_AB\right)^TA=A\bar{B}_AB \tag{1}$$

[10, p. 46]. The matrix

$$S=S(A,B)=A\bar{B}_AB \tag{2}$$

is invariant under the choice of $\bar{B}_A$ and satisfies

$$S=B^T\left(\bar{B}_A\right)^TA\bar{B}_AB; \tag{3a}$$

52 Vanderbilt Ave., New York, NY 10017 0024-3795/83/$3.00

in particular

$$S \geqslant 0 \quad \text{and} \quad A \geqslant S \tag{3b}$$

(see [8]). S is called the shorted matrix A with respect to B.

Using the shorted operator S for the present problem is suggested by the fact that $\mathfrak{R}(A) \cap \mathfrak{R}(B^T) = \mathfrak{R}(S)$, where $\mathfrak{R}(D)$ denotes the range of the matrix D. This together with $S \geqslant 0$ allows us to treat the p-dimensional problem, $p \geqslant 2$, with Löwner ordering in the same way as the case $p = 1$.

THEOREM. *Let $A \in \mathbb{R}^{n \times n}$, $A \geqslant 0$, $B \in \mathbb{R}^{q \times n}$, $C \in \mathbb{R}^{q \times p}$ be given, and let S be defined by* (2). *Consider the following pair of primal and dual problems* (*with respect to the Löwner ordering*):

$$g(X) = XAX^T = \min$$

subject to $\quad BX^T = C, \qquad X \in \mathbb{R}^{p \times n} \quad (4)$

and

$$h(Y) = C^T B^{+T} Y^T + YB^+ C - YS^+ Y^T = \max$$

subject to $\quad (I_n - SS^+) Y^T = 0, \qquad Y \in \mathbb{R}^{p \times n}. \quad (5)$

If there exists a feasible solution of (4), *then:*

(*a*) *There exist feasible solutions $\hat{X}$ and $\hat{Y}$ of* (4) *and* (5), *respectively, such that for all feasible X and Y*

$$g(X) \geqslant g(\hat{X}) = C^T B^{+T} S B^+ C = h(\hat{Y}) \geqslant h(Y).$$

(*b*) *The general solution of* (4) *is*

$$\hat{X} = C^T \bar{B}_A^T + Z\left(I - (A - S)\bar{A}\right)\left(I - B^T \bar{B}_A^T\right),$$

where Z is arbitrary, $Z \in \mathbb{R}^{p \times n}$, $\bar{A}$ is an arbitrary g-inverse of A, and $\bar{B}_A$ is an arbitrary, but fixed, minimum-A-seminorm g-inverse of B.

(*c*) *The solution of* (5) *is unique, namely*

$$\hat{Y} = C^T B^{+T} S.$$

Proof. Note first that a solution X of $BX^T = C$ exists iff $C = B\bar{B}C$ for a g-inverse $\bar{B}$ of B.

(a): Let X and Y be feasible. Then using (2) and (3)

$$\begin{aligned} h(Y) &= C^T B^{+T} SS^+ Y^T + YS^+ SB^+ C - YS^+ Y^T \\ &= -(Y^T - SB^+ C)^T S^+ (Y^T - SB^+ C) + C^T B^{+T} SB^+ C \\ &\leqslant C^T B^{+T} SB^+ C = XB^T B^{+T} SB^+ BX^T = XSX^T \\ &\leqslant XSX^T + (X^T - A^+ SX^T)^T A (X^T - A^+ SX^T) = XAX^T = g(X). \end{aligned}$$

For $\hat{X} = C^T \bar{B}_A^T$ we have $B\hat{X}^T = B\bar{B}_A C = C$, i.e., $\hat{X}$ is feasible; and from (3)

$$g(\hat{X}) = C^T \bar{B}_A^T A \bar{B}_A C = C^T B^{+T} SB^+ C.$$

Further, $\hat{Y} = C^T B^{+T} S$ is feasible and

$$h(\hat{Y}) = 2C^T B^{+T} SB^+ C - C^T B^{+T} SS^+ SB^+ C = C^T B^{+T} SB^+ C.$$

(b): If $\bar{B}_A$ is fixed, the general solution of $BX^T = C$ is

$$X = C^T \bar{B}_A^T + U\left(I - B^T \bar{B}_A^T\right),$$

where $U \in \mathbb{R}^{p \times n}$ is arbitrary [10, p. 24]. The proof of (a) shows that we have $g(X) = g(\hat{X})$ iff $XAX^T = XSX^T$, i.e. $X(A - S) = 0$. Using $C = B\bar{B}_A C$, (1), and (2), this implies $U(A - S) = 0$; hence

$$U = Z\left(I - (A - S)\overline{(A - S)}\right),$$

where $Z \in \mathbb{R}^{p \times n}$ is arbitrary. Now from (1)–(3) it follows that $\bar{A}$ is a g-inverse of $A - S$. The converse is straightforward, and (c) can be proved similarly. ∎

Remark. The dual problem (5) has been obtained by a Lagrangian-multiplier approach to (4). Therefore, the result admits the following version as a saddle-point theorem: Let the operator $F: \mathbb{R}^{p \times n} \times \mathbb{R}^{p \times q} \to \mathbb{R}^{p \times p}$ be defined by

$$F(X, Y) = XAX^T - XB^T Y^T - YBX^T + YC + C^T Y^T.$$

Then for $\hat{X}=C^T\bar{B}_A^T$, $\hat{Y}=C^TB^{+T}SB^+$, and all $X\in\mathbb{R}^{p\times n}$, $Y\in\mathbb{R}^{p\times q}$ it holds that

$$F(\hat{X},Y)\leqslant F(\hat{X},\hat{Y})\leqslant F(X,\hat{Y}).$$

In the case $p=1$ this approach has been used in [6] to give a statistical interpretation of $\hat{Y}$, if the primal problem is that of finding a BLUE in the Gauss-Markoff model.

Let us consider now some special cases of the theorem.

EXAMPLE 1. If A is positive definite, then $\bar{B}_A=A^{-1}B^T(BA^{-1}B^T)^+$ satisfies (1); hence $S=B^T(BA^{-1}B^T)^+B$. From this one easily obtains that $\hat{X}$ is uniquely determined. If additionally $B=CC^T$, then $\bar{B}_A=A^{-1}C(C^TA^{-1}C)^+C^+$ and

$$\hat{X}=C^+C(C^TA^{-1}C)C^TA^{-1}.$$

$\hat{X}$ satisfies the conditions (3.3.5) in [10]. Thus we obtain a characterization of the I_q-norm A^{-1}-least-squares g-inverse of C as solution of a matrix optimization problem. In the special case $A=I_q$ we get with $\hat{X}=C^+$ the well-known characterization of the Moore-Penrose inverse; cf. [9] and [4].

EXAMPLE 2. Let $A_1, A_2\in\mathbb{R}^{n\times n}$, $A_1\geqslant 0$, $A_2\geqslant 0$; put

$$A=\begin{bmatrix}A_1 & 0\\ 0 & A_2\end{bmatrix},\qquad B=[I_n, I_n].$$

Then

$$\bar{B}_A=\frac{1}{2}\begin{bmatrix}I_n+(A_1+A_2)^+(A_1-A_2)\\ I_n+(A_1+A_2)^+(A_2-A_1)\end{bmatrix}$$

satisfies the condition (1); hence

$$S=\begin{bmatrix}D & D\\ D & D\end{bmatrix},$$

where

$$\begin{aligned}D&=A_1-A_1(A_1+A_2)^+A_1=A_2-A_2(A_1+A_2)^+A_2\\ &=A_1(A_1+A_2)^+A_2=A_2(A_1+A_2)^+A_1\end{aligned}$$

is the parallel sum of A_1 and A_2; cf. [1]. Thus the optimum value of the problem

$$\begin{aligned} & X_1A_1X_1^T + X_2A_2X_2^T = \min \\ \text{subject to} \quad & X_1^T + X_2^T = C, \qquad X_1, X_2 \in \mathbb{R}^{p\times n} \end{aligned} \tag{6}$$

is C^TDC. A solution is given by

$$\hat{X}_1 = C^T - C^TA_1(A_1 + A_2)^+, \qquad \hat{X}_2 = C^TA_1(A_1 + A_2)^+.$$

The solution of (5) is $\hat{Y} = [\hat{Y}_1, \hat{Y}_2] = [C^TD, C^TD]$. This is a generalization of problem (14) in [2]. Moreover, (6) provides a characterization of the parallel sum as the optimum value of (6) when $C = I_n$.

EXAMPLE 3. Considering the problem (4) for the case that $B = C$, we get $\min\{XAX^T: BX^T = B,\ X \in \mathbb{R}^{p\times n}\} = B^TB^{+T}SB^+B = S$. This is a characterization of the shorted matrix S which is different from those given in [2, Theorems 1, 5] and in [8, Theorem 2.2].

EXAMPLE 4. Let $p = q = 1$, A be positive definite, $B = e_k^T$, where e_k is the kth unit vector in $\mathbb{R}^n$, $C = 1$. In this case $\bar{B}_A = A^{-1}e_k(e_k^TA^{-1}e_k)^{-1}$ satisfies (1), and hence

$$S = S(A, e_k^T) = (e_k^TA^{-1}e_k)^{-1}e_ke_k^T;$$

see also [3, Theorem 7]. If A_k denotes the matrix obtained when deleting the kth row and column of A, and a^{kk} is the kth diagonal element of A^{-1}, then $e_k^TA^{-1}e_k = a^{kk} = \det A_k/\det A$. Thus we get from the theorem Bergström's inequality [7, p. 478],

$$\begin{aligned} \frac{\det(A^{(1)} + A^{(2)})}{\det(A_k^{(1)} + A_k^{(2)})} &= \min\{x^T(A^{(1)} + A^{(2)})x : e_k^Tx = 1\} \\ &\geqslant \min\{x^TA^{(1)}x : e_k^Tx = 1\} \\ &\quad + \min\{x^TA^{(2)}x : e_k^Tx = 1\} \\ &= \frac{\det A^{(1)}}{\det A_k^{(1)}} + \frac{\det A^{(2)}}{\det A_k^{(2)}} \end{aligned}$$

for any two positive definite matrices $A^{(1)}, A^{(2)} \in \mathbb{R}^{n \times n}$. Also from $A \geqslant S$ we get the well-known relation $a_{kk} \geqslant (a^{kk})^{-1}$.

I wish to thank S. K. Mitra and the associate editor for their help in improving the presentation and results.

REFERENCES

1 W. N. Anderson and R. J. Duffin, Series and parallel addition of matrices, *J. Math. Anal. Appl.* 26:576–594 (1969).
2 W. N. Anderson and G. E. Trapp, Shorted operators II, *SIAM J. Appl. Math.* 28:60–71 (1975).
3 R. Bott and R. J. Duffin, On the algebra of networks, *Trans. Amer. Math. Soc.* 74:99–109 (1953).
4 J. S. Chipman and M. M. Rao, The treatment of linear restrictions in regression analysis, *Econometrica* 32:198–209 (1964).
5 N. Gaffke and O. Krafft, Matrix inequalities in the Löwner ordering, in *Modern Applied Mathematics* (B. Korte, Ed.), North Holland, Amsterdam, 1982, pp. 595–622.
6 O. Krafft and S. K. Mitra, Least favorable parameters in the Gauss-Markoff model, *Sankhyā Ser. A* 44, to appear.
7 A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic, New York, 1979, Chapter 16E.
8 S. K. Mitra and M. L. Puri, Shorted operators and generalized inverses of matrices, *Linear Algebra Appl.* 25:45–56 (1979).
9 Penrose, R., On best approximate solutions of linear matrix equations, *Proc. Cambridge Philos. Soc.* 52:17–19 (1956).
10 C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and Its Applications*, Wiley, New York, 1971.

Received 29 June 1982; revised 4 August 1982