Extremal Problems on the Set of Nonnegative Definite Matrices

Alexander Shapiro
Department of Mathematics and Applied Mathematics
University of South Africa
P.O. Box 392
Pretoria 0001, South Africa

Submitted by Ingram Olkin

ABSTRACT

An optimization problem of minimizing a real-valued function of certain elements of a symmetric matrix subject to this matrix being nonnegative definite is considered. Optimality conditions are proposed. The duality result of Olkin and Pukelsheim (1982) is extended to a wide class of such problems. Applications are discussed.

1. INTRODUCTION

Let τ be a symmetric subset of the index set

$$\mathbb{N}_n = \{1, \ldots, n\} \times \{1, \ldots, n\},$$

and $\Omega(\tau)$ denote the corresponding space of $n \times n$ real symmetric matrices $X = [x_{ij}]$, such that $x_{ij} = 0$ whenever $(i, j) \notin \tau$. In this paper we study the following optimization problem:

(P) $$\underset{X \in \Omega(\tau)}{\text{minimize}} \, f(X)$$

$$\text{subject to} \quad S + X \geqslant 0,$$

where S is a specified symmetric matrix and f is a real-valued function of X. (We write $S \geqslant 0$ and $S > 0$ to denote that a symmetric matrix S is nonnegative definite and positive definite respectively.) Such optimization problems come up in many practical situations (see e.g., [1] and [6]), in particular in statistics, where $S + X$ is usually a covariance matrix with varying elements

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

and f is a linear function of X (see [2], [4], [5], [9], and [13]). As will be shown, the linear case has especially nice features, which we shall exploit in Sections 3 and 4.

We examine Problem (P) utilizing techniques and methods of nonlinear programming and convex analysis. The organization of this paper will be as follows. In Section 2 we present first-order optimality conditions for Problem (P). Here the development is based on the theory of optimality conditions for semiinfinite programs, and we briefly describe the required results. In Section 3 we show that there exists a certain duality between different problems of the form mentioned above. This will extend the result of Olkin and Pukelsheim [9] for a wide class of such problems. In Section 4 we consider applications to specific problems, and in Section 5 we give some uniqueness results.

The following notation will be used. The transpose of a column vector $z \in \mathbb{R}^n$ is z^T, and $|z| = (z^T z)^{1/2}$ is the Euclidean norm of z. The gradient vector of a function $f: \mathbb{R}^m \to \mathbb{R}$ at x is denoted by $\nabla f(x)$.

2. SEMIINFINITE PROGRAMMING AND OPTIMALITY CONDITIONS

A semiinfinite program is a mathematical problem of the form

$$\text{(Q)} \qquad \begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g(x, z) \geqslant 0, \quad z \in \mathscr{Z} \end{aligned}$$

where $x \in \mathbb{R}^m$ and $\mathscr{Z}$ is an infinite set. One can formulate Problem (P) as a semiinfinite program by writing the constraint $S + X \geqslant 0$ in the equivalent form as follows:

$$z^T(S + X)z \geqslant 0, \qquad z \in \mathscr{Z} = \{ y \in \mathbb{R}^n : |y| = 1 \}. \tag{2.1}$$

Since the pioneer work of Fritz John [8], semiinfinite programs have been studied by many authors (see, e.g., [7] and references therein). In the following theorem we formulate optimality conditions for Problem (Q) (see [10, Theorem 5.1, Corollary 1]). In what follows we suppose that x^* is a feasible point for Problem (Q), i.e. $g(x^*, z) \geqslant 0$ for all $z \in \mathscr{Z}$; the functions $f(\cdot)$ and $g(\cdot, z)$, $\forall z \in \mathscr{Z}$, are continuously differentiable; $\mathscr{Z}$ is a compact topological space; and $g(x, z)$ and $\nabla_x g(x, z)$ are continuous on $\mathbb{R}^m \times \mathscr{Z}$. We define

$$\mathscr{Z}(x) = \{z \in \mathscr{Z} : g(x,z) = 0\}$$

and remind the reader that Problem (Q) is a convex program if $f(\cdot)$ is convex and $g(\cdot, z)$ is concave for all $z \in \mathscr{Z}$, and that the Slater condition means the existence of a point $x \in \mathbb{R}^m$ such that $g(x,z) > 0$ for all $z \in \mathscr{Z}$.

THEOREM 2.1. *In order for x^* to be a solution to Problem* (Q) *it is necessary that there exist nonnegative multipliers $\lambda_0, \lambda_1, \dots, \lambda_k$, not all of them zero, and points $z_i \in \mathscr{Z}(x^*)$, $i = 1, \dots, k$, such that*

$$\lambda_0 \nabla f(x^*) - \sum_{i=1}^{k} \lambda_i \nabla_x g(x^*, z_i) = 0. \tag{2.2}$$

Moreover, if Problem (Q) *is a convex program and the Slater condition holds, then* (2.2) *is also a sufficient condition for x^* to be a solution of Problem* (Q) *and one can take $\lambda_0 = 1$.*

REMARK 2.1. If the set $\mathscr{Z}(x^*)$ is empty, i.e. $g(x^*, z) > 0$ for all $z \in \mathscr{Z}$, and hence x^* is an interior point of the feasible region, then (2.2) becomes

$$\nabla f(x^*) = 0,$$

which is the standard optimality condition for an unconstrained local optimum.

Now we apply the optimality conditions of Theorem 2.1 to Problem (P). As has been mentioned, Problem (P) can be considered as a semiinfinite program with

$$g(X,z) = z^T(S+X)z.$$

We denote by $\mathscr{P}_\tau(B)$ the projection of a symmetric matrix $B = [b_{ij}]$ onto the space $\Omega(\tau)$, i.e., the (i,j) element of $\mathscr{P}_\tau(B)$ is b_{ij} whenever $(i,j) \in \tau$ and is zero otherwise. Since

$$g(X,z) = \operatorname{tr}(S+X)zz^T,$$

the gradient of g, with respect to X, is $\mathscr{P}_\tau(zz^T)$. [Here the gradient of a function of X is defined with respect to the scalar product $A \circ B = \operatorname{tr} AB$ on the space of symmetric matrices, and thus this gradient is given by a symmetric matrix from the space $\Omega(\tau)$.] We also remark that for Problem (P), the Slater condition means the existence of a matrix $X \in \Omega(\tau)$ such that $S + X > 0$.

THEOREM 2.2. *Let the function $f(X)$ be convex, X^* be a feasible point for Problem* (P), *and the Slater condition hold. Then X^* is a solution to Problem* (P) *iff there exists a symmetric matrix $B \geqslant 0$ such that*

$$(S + X^*)B = 0 \tag{2.3}$$

and

$$\nabla f(X^*) = \mathscr{P}_\tau(B). \tag{2.4}$$

Proof. From Theorem 2.1 we know that X^* is a solution to Problem (P) iff there exist vectors $z_1, \ldots, z_k$, $|z_i| = 1$, $z_i^T(S + X^*)z_i = 0$, $i = 1, \ldots, k$, and nonnegative numbers $\lambda_1, \ldots, \lambda_k$ such that

$$\nabla f(X^*) - \sum_{i=1}^{k} \lambda_i \mathscr{P}_\tau\left(z_i z_i^T\right) = 0.$$

It remains to define the matrix B by

$$B = \sum_{i=1}^{k} \lambda_i z_i z_i^T. \qquad \blacksquare$$

REMARK 2.2. If $S + X^* > 0$, then $B = 0$ and (2.4) becomes $\nabla f(X^*) = 0$ (see Remark 2.1).

If $f(X)$ is a linear function given by

$$f(X) = \operatorname{tr} AX, \tag{2.5}$$

then

$$\nabla f(X) = \mathscr{P}_\tau(A)$$

and the solution, if it exists, is always attained on the boundary of the feasible region, i.e., $S + X^*$ is singular.

3. DUALITY

Let us consider the Lagrangian function

$$L(X, \Lambda) = f(X) - \operatorname{tr} \Lambda(S + X) \tag{3.1}$$

corresponding to Problem (P). It can be seen that the function

$$p(X) = \max_{\Lambda \geqslant 0} L(X, \Lambda) \tag{3.2}$$

is equal to $f(X)$ whenever $S + X \geqslant 0$ and is $+\infty$ otherwise. Therefore the problem of minimizing $p(X)$ over $\Omega(\tau)$ is equivalent to the primary Problem (P). We define the dual problem as follows:

(D) $$\begin{aligned} &\text{maximize} \quad h(\Lambda) \\ &\text{subject to} \quad \Lambda \geqslant 0, \end{aligned}$$

where

$$h(\Lambda) = \min_{X \in \Omega(\tau)} L(X, \Lambda).$$

It is well known that the optimal value of Problem (D) is less than or equal to the optimal value of Problem (P). Moreover, if (X^*, Λ^*) is a *saddle* point of L, then

$$h(\Lambda^*) = p(X^*)$$

and X^* solves Problem (P) while Λ^* solve the dual problem (D) (e.g., [14, p. 45]). We remind the reader that a point (X^*, Λ^*), with $X^* \in \Omega(\tau)$ and $\Lambda^* \geqslant 0$, is called a saddle point of the Lagrangian L iff

$$L(X^*, \Lambda) \leqslant L(X^*, \Lambda^*) \leqslant L(X, \Lambda^*) \tag{3.3}$$

for all $X \in \Omega(\tau)$ and $\Lambda \geqslant 0$.

THEOREM 3.1. *Let the function $f(X)$ be convex, and suppose that Problem* (P) *attains its solution at a point X^* and the Slater condition for Problem* (P) *holds. Then Problems* (P) *and* (D) *have the same optimal value,* *and* $\Lambda^* = B$ *solves Problem* (D), *where* B *is a nonnegative definite matrix satisfying the optimality conditions* (2.3) *and* (2.4).

Proof. Since Problem (P) is a convex program and the Slater condition holds, there exists a nonnegative definite matrix B satisfying the optimality conditions (2.3) and (2.4). We only have to show that (X^*, Λ^*), where $\Lambda^* = B$, is a saddle point. It follows from (2.3) that

$$L(X^*, \Lambda^*) = f(X^*) \geqslant L(X^*, \Lambda)$$

for every $\Lambda \geqslant 0$. On the other hand, we have that the function $L(\cdot, \Lambda^*)$ is convex and its gradient at a point X is

$$\nabla f(X) - \mathscr{P}_\tau(\Lambda^*).$$

From (2.4) this gradient is zero at X^*, and hence the function $L(\cdot, \Lambda^*)$ attains its minimum at X^*. This proves (3.3). ■

Now we apply the general result of Theorem 3.1 to the case of linear function $f(X)$. Consider the following (linear) problem

(LP)
$$\underset{X \in \Omega(\tau)}{\text{minimize}} \quad \operatorname{tr} AX$$
$$\text{subject to} \quad S + X \geqslant 0.$$

The corresponding Lagrangian is

$$L(X, \Lambda) = \operatorname{tr} AX - \operatorname{tr} \Lambda(S + X)$$

and

$$h(\Lambda) = \min_{X \in \Omega(\tau)} \operatorname{tr}(A - \Lambda)X - \operatorname{tr} \Lambda S.$$

It can be seen that

$$h(\Lambda) = -\operatorname{tr} \Lambda S$$

whenever $\mathscr{P}_\tau(A - \Lambda) = 0$, and $h(\Lambda)$ is $-\infty$ otherwise. In other words, $h(\Lambda)$ is $-\infty$ if $\Lambda - A \notin \Omega(\sigma)$, where $\sigma = \mathbb{N}_n \setminus \tau$. Denoting $Y = \Lambda - A$, we come to the following dual problem:

$$\text{(LD)} \qquad \underset{Y \in \Omega(\sigma)}{\text{maximize}} \quad -\operatorname{tr} S(A+Y)$$
$$\text{subject to} \quad A + Y \geqslant 0,$$

and Theorem 3.1 implies the following result.

THEOREM 3.2. *Suppose that Problem* (LP) *attains its solution at a point* X^* *and the Slater condition for Problem* (LP) *holds. Then Problems* (LP) *and* (LD) *have the same optimal value, and* $Y^* = B - A$ *solves Problem* (LD), *where* B *is a nonnegative definite matrix satisfying the optimality conditions of Theorem* 2.2.

As we shall see in the next section, a particular case of this duality has been discovered by Olkin and Pukelsheim [9].

4. APPLICATIONS

Consider the following problem:

$$\text{(AP)} \qquad \text{minimize} \quad -2\operatorname{tr} X_{12}$$
$$\text{subject to} \quad S + X \geqslant 0,$$

where

$$S = \left[\begin{array}{c|c} S_{11} & 0 \\ \hline 0 & S_{22} \end{array}\right], \qquad X = \left[\begin{array}{c|c} 0 & X_{12} \\ \hline X_{12}^T & 0 \end{array}\right],$$

S_{11} and S_{22} are specified $n \times n$ positive definite matrices, and X_{12} is an $n \times n$ variable matrix. This problem arose in connection with finding the minimal L_2-distance between two random vectors with covariance matrices S_{11} and S_{22} respectively (see [5], [9]).

Problem (AP) takes the form of Problem (LP) if we define

$$A = \left[\begin{array}{c|c} 0 & -I \\ \hline -I & 0 \end{array}\right]$$

and $\tau = \{(i, j) : |i - j| \geqslant n;\ 1 \leqslant i, j \leqslant 2n\}$. Applying the optimality conditions of Theorem 2.2, we obtain that X^*, with $S + X^* \geqslant 0$, is a solution to Problem (AP) iff there exists a nonnegative definite matrix

$$B = \left[\begin{array}{c|c} Y_{11} & -I \\ \hline -I & Y_{22} \end{array}\right]$$

such that

$$(S + X^*)B = 0. \tag{4.1}$$

The equations (4.1) can be written in the following form:

$$\begin{aligned} S_{11}Y_{11} - X_{12} &= 0, \\ -S_{11} + X_{12}Y_{22} &= 0, \\ X_{12}^T Y_{11} - S_{22} &= 0, \\ -X_{12}^T + S_{22}Y_{22} &= 0. \end{aligned} \tag{4.2}$$

The first two equations of (4.2) imply that

$$S_{11}Y_{11}Y_{22} = S_{11},$$

and consequently the matrices Y_{11} and Y_{22} are nonsingular and inverse to each other. Therefore, due to Theorem 3.2, a dual of Problem (AP) is

(AD) $\quad$ maximize $\quad -\operatorname{tr}\left(S_{11}Y_{11} + S_{22}Y_{11}^{-1}\right)$

subject to $\quad A + Y \geqslant 0,$

where

$$Y = \left[\begin{array}{c|c} Y_{11} & 0 \\ \hline 0 & Y_{11}^{-1} \end{array}\right].$$

One can note that $A + Y \geqslant 0$ iff $Y_{11} > 0$. Indeed, if $Y_{11} > 0$ then $A + Y = Z^T Z$, where $Z = [-Y_{11}^{1/2}; Y_{11}^{-1/2}]$, and hence $A + Y \geqslant 0$. On the other hand it is clear that $A + Y \geqslant 0$ implies $Y_{11} > 0$.

The duality between Problems (AP) and (AD) has been discussed in [9]. Finally it can be easily verified that

$$X_{12}^{*} = S_{11} S_{22}^{1/2} \left(S_{22}^{1/2} S_{11} S_{22}^{1/2} \right)^{-1/2} S_{22}^{1/2}$$

and

$$Y_{11}^{*} = Y_{22}^{*\,-1} = S_{22}^{1/2} \left(S_{22}^{1/2} S_{11} S_{22}^{1/2} \right)^{-1/2} S_{22}^{1/2}$$

solve the equations (4.2) and that $S + X^* \geqslant 0$, $Y_{11}^* > 0$. Consequently X^* solves Problem (AP) and Y^* solves Problem (AD) (cf. [1], [5], [6], [9]).

As an another example we consider the following problem:

$$\text{(BP)} \qquad \begin{aligned} &\text{minimize} \quad \operatorname{tr} AX \\ &\text{subject to} \quad S + X \geqslant 0, \end{aligned}$$

where A is an $n \times n$ nonnegative definite diagonal matrix, S is an $n \times n$ symmetric matrix, and X is an $n \times n$ diagonal variable matrix. This problem has arisen in investigating the so-called minimum-trace factor analysis (see [2], [4], [11], [13]). A dual to Problem (BP) is

$$\text{(BD)} \qquad \begin{aligned} &\text{maximize} \quad -\operatorname{tr} S(A+Y) \\ &\text{subject to} \quad A + Y \geqslant 0, \end{aligned}$$

where Y is an $n \times n$ variable symmetric matrix with zero diagonal elements.

If $A \geqslant 0$, then the feasible region of Problem (BD) is nonempty and compact, and hence Problem (BD) possesses a solution Y^*. Furthermore, if $A > 0$, then the Slater condition for Problem (BD) holds, and consequently the two problems (BP) and (BD) are equivalent to each other and Problem (BP) has a solution. However, if A is *singular*, then it may occur that Problem (BP) does not possess a solution although the optimal value of (BP) is still finite.

The duality between (BP) and (BD) has been utilized in [2] to construct an efficient numerical algorithm for Problem (BP). A rigorous proof of this duality has been given in [13].

5. UNIQUENESS RESULTS

In this section we investigate uniqueness properties of Problem (LP). It is not true that Problem (LP) always has at most one solution. However, it is remarkable that "usually" Problem (LP) attains its solution at a unique point. Note that in Problems (LP) and (LD) it is not a restriction to take $A \in \Omega(\tau)$.

THEOREM 5.1. *Let $\Omega(\tau)$ be a space of block-diagonal, symmetric matrices, and let $A > 0$, $A \in \Omega(\tau)$. Then Problem* (LP) *has a unique solution.*

Proof. First we note that, since $\Omega(\tau)$ is a space of block-diagonal matrices, the set τ includes the diagonal elements of $\mathbb{N}_n$, i.e., $(i, i) \in \tau$ for $i = 1, \dots, n$. It follows that the feasible region of the dual problem is compact and hence the dual problem has a solution. Moreover, since $A > 0$, the Slater condition for the dual problem holds. Consequently the primary problem (LP) has a solution as well, and the optimality conditions of Theorem 2.2 must be satisfied.

Now let us assume that there are two solutions X^* and X_0. It follows from the convexity of Problem (LP) that $X_1 = (X^* + X_0)/2$ is also a solution. By the optimality conditions there exists a matrix $B \geqslant 0$ such that

$$(S + X_1)B = 0 \tag{5.1}$$

and

$$\mathscr{P}_\tau(B) = A. \tag{5.2}$$

From (5.1) we obtain

$$(S + X^*)B + (S + X_0)B = 0$$

and then

$$B(S + X^*)B + B(S + X_0)B = 0. \tag{5.3}$$

The two matrices in the left side of (5.3) are nonnegative definite and, since their sum is zero, each of them must be zero. This implies that

$$(S + X^*)B = 0$$

and

$$(S + X_0)B = 0.$$

Then we have

$$(X^* - X_0)B = 0. \tag{5.4}$$

From (5.2), (5.4) and since $X^* - X_0$ and A are block-diagonal, we obtain that

$$(X^* - X_0)A = 0.$$

Then, since $A > 0$, we have

$$X^* - X_0 = 0. \qquad \blacksquare$$

This proof is a modification of the proofs in [4, Theorem 4] and in [13, Lemma 2.4].

THEOREM 5.2. *Let S and τ be fixed, and let the feasible region be nonempty and compact. Then for almost every A Problem* (LP) *has a unique solution.*

Proof. Let us consider the optimal value of Problem (LP) as a function $F(A)$ of the matrix A. Since the feasible region is compact, $F(A)$ is finite for all A. In the following lemma we show that F is differentiable at A iff the corresponding Problem (LP) has a unique solution. Now, since the function $\operatorname{tr} AX$ is linear in A and the set of feasible X is compact, we have that the function $F(A)$ is locally Lipschitz. By the well-known theorem of Rademacher (e.g., Stein [12]) a locally Lipschitz function is differentiable almost everywhere. The theorem follows. ■

LEMMA 5.1. *The optimal value function F is differentiable at A iff the corresponding Problem* (LP) *has a unique solution.*

Proof. It follows from the well-known result of Danskin [3] (also [10, Theorem 3.2]) that if $f(a, x)$ is a function of two variables $a \in \mathbb{R}^m$ and $x \in \Phi$, the set Φ is compact, $f(a, x)$ is continuously differentiable in a for every $x \in \Phi$, and $f(a, x)$ and $\nabla_a f(a, x)$ are continuous on $\mathbb{R}^m \times \Phi$, then the min function

$$F(a) = \min_{x \in \Phi} f(a, x)$$

is differentiable at a iff the gradients $\nabla_a f(a, x)$ are identical for all $x \in \Phi(a)$, where

$$\Phi(a) = \{x \in \Phi : F(a) = f(a, x)\}.$$

It remains to apply this result with

$$f(a, x) = \operatorname{tr} AX$$

and Φ the feasible region of Problem (LP). ■

REMARK 5.1. It can be seen that the feasible region of Problem (LP) is compact iff the set τ does not include a diagonal element of $\mathbb{N}_n$.

REFERENCES

1 T. W. Anderson and I. Olkin, An extremal problem for positive definite matrices, *Linear and Multilinear Algebra* 6:257–262 (1978).

2 P. M. Bentler, A lower-bound method for the dimension-free measurement of internal consistency, *Social Sci. Res.* 1:343–357 (1972).

3 J. M. Danskin, *The Theory of* max-min *and Its Application to Weapons Allocation Problems*, Springer, New York, 1967.

4 G. Della Riccia and A. Shapiro, Minimum rank and minimum trace of covariance matrices, *Psychometrika* 47:443–448 (1982).

5 D. C. Dowson and B. V. Landau, The Fréchet distance between multivariate normal distributions, *J. Multivariate Anal.* 12:450–455 (1982).

6 H. Flanders, An extremal problem in the space of positive definite matrices, *Linear and Multilinear Algebra* 3:33–39 (1975).

7 K. R. Gehner, Necessary and sufficient conditions for the Fritz John problem with linear equality constraints, *SIAM J. Control Optim.* 12:140–149 (1974).

8 F. John, Extremum problems with inequalities as subsidiary conditions, in *Studies and Essays, Courant Anniversary Volume*, Wiley, New York, 1948, pp. 187–204.

9 I. Olkin and F. Pukelsheim, The distance between two random vectors with given dispersion matrices, *Linear Algebra Appl.* 48:257–263 (1982).

10 B. N. Pshenichnyi, *Necessary Conditions for an Extremum*, Marcel Dekker, New York, 1971.

11 A. Shapiro, Weighted minimum trace factor analysis, *Psychometrika* 47:243–264 (1982).

12 E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton U.P., Princeton, N.J., 1970.

13 J. M. F. ten Berge, T. A. B. Snijders, and F. E. Zegers, Computational aspects of the greatest lower bound to the reliability and constrained minimum trace factor analysis, *Psychometrika* 46:201–213 (1981).

14 W. J. Zangwill, *Nonlinear Programming, a Unified Approach*, Prentice-Hall, 1967.

Received 28 June 1983; revised 12 January 1984