SIAM J. OPTIMIZATION
Vol. 6, No. 1, pp. 164–177, February 1996

CONVEX ANALYSIS ON THE HERMITIAN MATRICES*

A. S. LEWIS†

Abstract. There is growing interest in optimization problems with real symmetric matrices as variables. Generally the matrix functions involved are *spectral*: they depend only on the eigenvalues of the matrix. It is known that convex spectral functions can be characterized exactly as symmetric convex functions of the eigenvalues. A new approach to this characterization is given, via a simple Fenchel conjugacy formula. We then apply this formula to derive expressions for subdifferentials, and to study duality relationships for convex optimization problems with positive semidefinite matrices as variables. Analogous results hold for Hermitian matrices.

Key words. convexity, matrix function, Schur convexity, Fenchel duality, subdifferential, unitarily invariant, spectral function, semidefinite programming, quasi-Newton update

AMS subject classifications. 15A45, 49N15, 90C25, 65K10

1. Introduction.

A matrix norm on the $n \times n$ complex matrices is called *unitarily invariant* if it satisfies $\|XV\| = \|X\| = \|VX\|$ for all unitary V. A well-known result of von Neumann [30] states that if f is a symmetric gauge function on $\mathbb{R}^n$ then f induces a unitarily invariant norm, namely, $\|X\|_f = f(\sigma_1(X), \dots, \sigma_n(X))$, where $\sigma_1(X) \leq \cdots \leq \sigma_n(X)$ are the singular values of X. Conversely, every unitarily invariant norm can be written in this form. A good exposition may be found in [14]. More generally, a matrix norm is called *weakly unitarily invariant* if it satisfies $\|V^*XV\| = \|X\|$ for all unitary V.

The $n \times n$ complex Hermitian matrices may be regarded as a *real* inner product space $\mathcal{H}$, with inner product $\langle X, Y \rangle$ defined as traceXY. Let us now ask a similar question about general convex functions on $\mathcal{H}$: What can be said about unitarily invariant convex functions $F : \mathcal{H} \to (-\infty, +\infty]$, where now by *unitarily invariant* we mean $F(V^*XV) = F(X)$ whenever V lies in $\mathcal{U}$, the $n \times n$ unitary matrices? Such functions clearly depend only on the eigenvalues of X: they are sometimes called *spectral functions* (see [10]).

Observe first that if we write diag(λ) (given λ in $\mathbb{R}^n$) for the diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$, and define a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ by $f(\lambda) = F(\text{diag}(\lambda))$, then clearly f is convex and *symmetric*: $f(\lambda) = f(P\lambda)$ for all P in $\mathcal{P}$, the $n \times n$ permutation matrices. In fact the converse is also true: if $f : \mathbb{R}^n \to (-\infty, +\infty]$ is a symmetric convex function then it induces a unitarily invariant, convex matrix function $f_{\mathcal{H}} : \mathcal{H} \to (-\infty, +\infty]$, defined by

$$f_{\mathcal{H}}(X) = f(\lambda(X)), \tag{1.1}$$

where $\lambda(X) = (\lambda_1(X), \dots, \lambda_n(X))^T$ is the vector of eigenvalues of X in nondecreasing order. This result was first proved in [4], for everywhere finite f: the proof extends immediately to allow f to take the value $+\infty$.

As an example (see §4), if we take

$$f(\lambda) = \begin{cases} -\sum_{i=1}^n \log \lambda_i, & \text{if } \lambda > 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

* Received by the editors December 21, 1993; accepted for publication (in revised form) July 6, 1994. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada.

† Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1 (aslewis@orion.uwaterloo.ca).

then we obtain the well-known convex matrix function

$$f_{\mathcal{H}}(X) = \begin{cases} -\log\det X, & \text{if } X \text{ positive definite,} \\ +\infty, & \text{otherwise.} \end{cases}$$

The approach to the basic result in [4] is direct, using a technique also appearing in [18] (strict convexity is not discussed). An independent approach appears in [10], revealing the role of "Schur convexity." Let us define a convex cone

$$K = \{x \in \mathbb{R}^n \mid x_1 \geq x_2 \geq \cdots \geq x_n\},$$

which has dual cone $K^+ = \{y \mid y^T x \geq 0 \ \forall x \in K\}$ given by

$$K^+ = \left\{ y \in \mathbb{R}^n \mid \sum_1^j y_i \geq 0 \ (j = 1, 2, \ldots, n-1), \ \sum_1^n y_i = 0 \right\}.$$

We say that a function $f : K \to (-\infty, +\infty]$ is *Schur convex* if it is K^+-isotone: in other words, $f(x) \leq f(z)$ whenever $z - x \in K^+$. Implicit in the argument in [10] (which focuses on differentiable functions f) is the fact that the spectral function $f_{\mathcal{H}}$ is convex exactly when f restricted to K is convex and Schur convex.

In fact it is not difficult to see that convex, Schur convex functions are precisely restrictions to K of symmetric convex functions. Thus $f_{\mathcal{H}}$ is convex whenever f is convex and symmetric. By contrast with our approach, lower semicontinuity of f is not required in either [4] or [10]. On the other hand, these approaches give no insight into conjugacy or subdifferentials, which of course are of fundamental interest in an optimization context.

We will follow a new approach to the basic result via Fenchel conjugation: this approach is close in spirit to von Neumann's technique in the norm case (see [14]). For a function $f : \mathbb{R}^n \to (-\infty, +\infty]$, the *Fenchel conjugate* $f^* : \mathbb{R}^n \to [-\infty, +\infty]$ is the lower semicontinuous, convex function

$$f^*(y) = \sup\{x^T y - f(x) \mid x \in \mathbb{R}^n\}.$$

(We will make frequent use of ideas and notation from [26].) By analogy, for a matrix function $F : \mathcal{H} \to (-\infty, +\infty]$ we can define a *conjugate* matrix function $F^* : \mathcal{H} \to [-\infty, +\infty]$ (cf. [8]) by

$$F^*(Y) = \sup\{\mathrm{tr} XY - F(X) \mid X \in \mathcal{H}\}. \tag{1.2}$$

Exactly as in $\mathbb{R}^n$, because F^* is expressed as a supremum of (continuous) linear functions of Y, it must be convex and lower semicontinuous.

The idea of our key result is then rather simple. We will prove (Theorem 2.3) that if the function f is symmetric on $\mathbb{R}^n$ then the conjugate of the induced matrix function $f_{\mathcal{H}}$ defined in (1.1) is given by

$$(f_{\mathcal{H}})^* = (f^*)_{\mathcal{H}}. \tag{1.3}$$

Since every lower semicontinuous, convex function g (excepting $g \equiv +\infty$) can be written as a conjugate, $g = f^*$, it follows from this formula that the matrix function it induces, $g_{\mathcal{H}}$, is a conjugate function, and hence it is lower semicontinuous and convex: in fact, to be specific, $g_{\mathcal{H}} = ((g^*)_{\mathcal{H}})^*$. An analogous argument proves the corresponding result for real-orthogonally invariant, convex functions on the $n \times n$ real symmetric matrices. Using the conjugacy formula (1.3) it becomes straightforward to link strict convexity and differentiability properties of the underlying function f with those of the induced matrix function $f_{\mathcal{H}}$. Furthermore, (1.3) results in a simple expression for the subdifferential of $f_{\mathcal{H}}$ in terms of the subdifferential of f.

It is possible to follow an analogous route to the study of the real vector space of $m \times n$ complex matrices (with inner product $\langle X, Y \rangle = \mathrm{Re}(\mathrm{tr} XY^*)$) and (strongly) unitarily invariant functions F on this space (meaning $F(X) = F(UXV)$ for any unitary U and V). By analogy with (1.1), such functions have the form $F(X) = f(\sigma(X))$, where $\sigma_i(X)$ is the ith singular value of X (for $i = 1, 2, \ldots, q = \min\{m, n\}$), arranged in nondecreasing order, and f is symmetric and *absolute* ($f(\sigma) = f(|\sigma_1|, |\sigma_2|, \ldots, |\sigma_q|)$ for any σ in $\mathbb{R}^q$). In a fashion very similar to our present development we arrive at the analogue of formula (1.3) and hence expressions for subdifferentials. Details are deferred to a forthcoming note: we simply observe that such expressions have been the topic of a number of recent papers in the special case where f is a symmetric gauge function, and hence F is a unitarily invariant norm (see [36, 31, 32, 37, 5]). This approach will also yield characterizations of strict convexity and smoothness in this setting analogous to those in our present development: such results for unitarily invariant norms have appeared in [3, 36].

Studying convex matrix functions via their Fenchel conjugates is not a new idea. It is implicit for example in some of the techniques in [7], and was used explicitly in [8] to study the sum of the largest k eigenvalues of a real symmetric matrix, an approach also followed in [12] (see also [13]). The primary aim of these latter papers is to study sensitivity results via the subdifferential set. Various representations of this set were investigated in [22]–[24].

We present a number of well-known convex matrix functions, showing how their (strict) convexity follows easily. To conclude, we use the conjugacy formula to study duality relationships for various convex optimization problems posed over the cone of positive semidefinite, real symmetric matrices. Interest in matrix optimization problems (and duality in particular) has been growing in recent years (for instance [27, 23, 1, 33, 35, 28]). The examples we choose are of recent interest in applications of interior point methods (see for example [1, 15, 21, 2]), as well as for variational characterizations of certain quasi-Newton updates (see for example [9, 34]).

2. Conjugates of induced matrix functions. We begin with a technical lemma (cf. [11, Theorem 368]).

LEMMA 2.1. *Suppose that $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$ and $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$ are real numbers and that P is an $n \times n$ permutation matrix. Then $\alpha^T P \beta \leq \alpha^T \beta$, with equality if and only if there exists an $n \times n$ permutation matrix Q with $Q\alpha = \alpha$ and $QP\beta = \beta$.*

Proof. Consider permuting the components of $P\beta$ in the following fashion.

Phase 1. Whenever we find indices i and j with $\alpha_i < \alpha_j$ and $(P\beta)_i > (P\beta)_j$, swap $(P\beta)_i$ and $(P\beta)_j$, giving a new sum $\alpha^T P' \beta > \alpha^T P \beta$ (because $(\alpha_i - \alpha_j)((P\beta)_i - (P\beta)_j) < 0$). We repeat this procedure until it terminates, say with the sum $\alpha^T P'' \beta$. Notice that the sum $\alpha^T P' \beta$ increases strictly at each step, and can take only finitely many values.

Phase 2. Now partition $\{1, 2, \ldots, n\}$ into sets $I_1 \leq I_2 \leq \cdots \leq I_k$ so that $\alpha_i = \gamma_r$ for all i in I_r, where γ_r increases strictly with r. Finally choose a permutation with matrix Q, fixing each index set I_r, and permuting the components $\{(P''\beta)_i \mid i \in I_r\}$ into nondecreasing order for each r.

Now note that $Q\alpha = \alpha$, while since $(P''\beta)_i \leq (P''\beta)_j$ whenever $\alpha_i < \alpha_j$ we deduce that $QP''\beta = \beta$. Notice that $\alpha^T P''\beta = (Q\alpha)^T(QP''\beta) = \alpha^T\beta$. Hence we see that $\alpha^T P\beta \leq \alpha^T\beta$. If equality holds then Phase 1 must be vacuous, and hence $P = P''$. The converse is immediate. □

The basis of the following key result is fairly standard and due to von Neumann [30] (see, for example, [19, p. 248] and the discussion in [5]). The full result (including conditions for attainment) may be found in [29] via an algebraic approach. In keeping with the variational spirit of this paper, and for completeness, we present here an optimization-based proof, following ideas from [25]. The underlying variational problem originated once again with von Neumann (see [20]).

THEOREM 2.2. *For Hermitian matrices X and Y,*

$$\mathrm{tr}XY \leq \lambda(X)^T\lambda(Y), \tag{2.1}$$

with equality if and only if there exists a unitary matrix V with $V^*XV = \mathrm{diag}\lambda(X)$ and $V^*YV = \mathrm{diag}\lambda(Y)$.

Proof. Consider the optimization problem

$$\left\{\begin{array}{ll} \text{maximize} & \mathrm{tr}Z^*XZY \\ \text{subject to} & Z^*Z = I, \\ & Z \in \mathbb{C}^{n\times n}. \end{array}\right. \tag{2.2}$$

This problem is solvable by compactness. We can regard the constraint as a linear map between two real vector spaces, $\Phi : \mathbb{C}^{n\times n} \to \mathcal{H}$, with a nonsingular derivative at any feasible point. Thus corresponding to any optimal solution Z_0 there will exist a Lagrange multiplier Λ in $\mathcal{H}$, so that

$$\nabla_Z(\mathrm{tr}Z^*XZY - \mathrm{tr}Z^*Z\Lambda)|_{Z_0} = 0.$$

Thus for all W in $\mathbb{C}^{n\times n}$,

$$\begin{aligned} 0 &= \lim_{t\to 0} t^{-1}(\mathrm{tr}(Z_0 + tW)^*(X(Z_0+tW)Y - (Z_0+tW)\Lambda) \\ &\quad - Z_0^*(XZ_0Y - Z_0\Lambda)) \\ &= \mathrm{tr}Z_0^*XWY + \mathrm{tr}W^*XZ_0Y - \mathrm{tr}Z_0^*W\Lambda - \mathrm{tr}W^*Z_0\Lambda \\ &= \mathrm{tr}(YZ_0^*X - \Lambda Z_0^*)W + \mathrm{tr}W^*(XZ_0Y - Z_0\Lambda). \end{aligned}$$

Choosing $W = XZ_0Y - Z_0\Lambda$ shows that $XZ_0Y = Z_0\Lambda$, and hence

$$Z_0^*XZ_0Y = \Lambda = \Lambda^* = YZ_0^*XZ_0.$$

Thus Y commutes with $Z_0^*XZ_0$, so there is a unitary matrix U diagonalizing Y and $Z_0^*XZ_0$. In other words

$$\begin{aligned} U^*YU &= \mathrm{diag}(P_1\lambda(Y)) \quad \text{and} \\ U^*Z_0^*XZ_0U &= \mathrm{diag}(P_2\lambda(X)) \end{aligned} \tag{2.3}$$

for some permutation matrices P_1 and P_2. Now we have

$$\begin{aligned} \mathrm{tr}XY &\leq \mathrm{tr}Z_0^*XZ_0Y = \mathrm{tr}(U^*Z_0^*XZ_0U)(U^*YU) \\ &= (P_1\lambda(Y))^T(P_2\lambda(X)) = \lambda(Y)^T(P_1^*P_2)\lambda(X) \\ &\leq \lambda(Y)^T\lambda(X) \end{aligned}$$

by Lemma 2.1.

If equality holds in (2.1) then $Z_0 = I$ is optimal for (2.2), and equality holds above. Again by Lemma 2.1 there is a permutation matrix Q with $Q\lambda(Y) = \lambda(Y)$ and $QP_1^*P_2\lambda(X) = \lambda(X)$. From (2.3) we know that $P_1^*U^*YUP_1 = \text{diag}\lambda(Y)$, so

$$QP_1^*U^*YUP_1Q^* = \text{diag}(Q\lambda(Y)) = \text{diag}\lambda(Y).$$

Also from (2.3) we have

$$QP_1^*U^*XUP_1Q^* = \text{diag}(QP_1^*P_2\lambda(X)) = \text{diag}\lambda(X),$$

and the result follows if we choose $V = UP_1Q^*$. □

We can now prove the main result.

THEOREM 2.3. *Suppose that the function* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *is symmetric. Then* $(f_{\mathcal{H}})^* = (f^*)_{\mathcal{H}}$.

Proof. For a Hermitian matrix Y we have

$$\begin{aligned}(f_{\mathcal{H}})^*(Y) &= \sup\{\text{tr}XY - f(\lambda(X)) \mid X \in \mathcal{H}\}\\ &= \sup\{\text{tr}XY - f(P\lambda(X)) \mid X \in \mathcal{H},\ P \in \mathcal{P}\}\\ &= \sup\{\text{tr}XY - f(\lambda) \mid X \in \mathcal{H},\ P \in \mathcal{P},\ \lambda \in \mathbb{R}^n,\ P\lambda(X) = \lambda\}.\end{aligned}$$

Now considering the supremum over λ first, we can rewrite this as

$$\begin{aligned}(f_{\mathcal{H}})^*(Y) &= \sup_{\lambda \in \mathbb{R}^n}\{\sup\{\text{tr}XY \mid X \in \mathcal{H},\ P \in \mathcal{P},\ P\lambda(X) = \lambda\} - f(\lambda)\}\\ &= \sup_{\lambda \in \mathbb{R}^n}\{\sup\{\lambda(Y)^T(Q\lambda) \mid Q \in \mathcal{P}\} - f(\lambda)\}\\ &= \sup\{\lambda(Y)^T(Q\lambda) - f(Q\lambda) \mid \lambda \in \mathbb{R}^n,\ Q \in \mathcal{P}\}\\ &= f^*(\lambda(Y)) = (f^*)_{\mathcal{H}}(Y),\end{aligned}$$

where we used Lemma 2.1 and Theorem 2.2 in the first step to see that the inner suprema are equal. □

COROLLARY 2.4. *If the function* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *is symmetric, convex, and lower semicontinuous then the matrix function* $f_{\mathcal{H}} : \mathcal{H} \to (-\infty, +\infty]$ *is convex and lower semicontinuous.*

Proof. We can assume that f is somewhere finite. Then since f^* is nowhere $-\infty$, with $f^{**} = f$ [26, Theorem 12.2], and since f^* is symmetric [26, Corollary 12.3.1], we have $f_{\mathcal{H}} = ((f^*)^*)_{\mathcal{H}} = ((f^*)_{\mathcal{H}})^*$. Thus $f_{\mathcal{H}}$ is a conjugate function, so is convex and lower semicontinuous. □

Exactly analogous results for functions on the real symmetric matrices may be derived by replacing unitary by real orthogonal matrices throughout.

3. Subgradients, differentiability, and strict convexity. Suppose that $\mathcal{X}$ is a finite-dimensional, real inner product space. The *conjugate* of a function $F : \mathcal{X} \to [-\infty, +\infty]$ is the function $F^* : \mathcal{X} \to [-\infty, +\infty]$ defined by

$$F^*(Y) = \sup_{X \in \mathcal{X}}\{\langle X, Y\rangle - F(X)\}.$$

Since $\mathcal{X}$ is isomorphic to $\mathbb{R}^n$ with its usual inner product, convex-analytic results on $\mathbb{R}^n$ can be translated directly.

The *domain* of F is the set $\text{dom}F = \{X \in \mathcal{X} \mid F(X) < +\infty\}$. If this set is nonempty and F never takes the value $-\infty$, then we say that F is *proper*. By [26, Theorem 12.2], if F is proper and convex, then F^* is proper, convex, and lower semicontinuous. For proper F with X in dom$F$, we can define the *(convex) subdifferential* of F at X as the convex set

$$\partial F(X) = \{Y \in \mathcal{X} \mid F(X) + F^*(Y) = \langle X, Y \rangle\}, \tag{3.1}$$

and when F is also convex this set is a singleton $\{Y\}$ exactly when F is differentiable at X, with gradient $\nabla F(X) = Y$ [26, Theorem 25.1].

We say that the proper, convex function F is *essentially smooth* if it is differentiable on the interior of dom$F$ (assumed nonempty), with $\|\nabla F(X^r)\| \to +\infty$ whenever X^r approaches a boundary point of dom$F$. We say that F is *essentially strictly convex* if F is strictly convex on any convex subset of $\{X \in \text{dom}F \mid \partial F(X) \neq \emptyset\}$ (and hence in particular on the interior of dom$F$) [26, Chapter 26]. A lower semicontinuous, proper, convex function F satisfies $F = F^{**}$ [26, Theorem 12.2], and F^* is essentially strictly convex if and only if F is essentially smooth [26, Theorem 26.3]: this is the case exactly when $\partial F(X)$ is single-valued when nonempty [26, Theorem 26.1].

THEOREM 3.1. *Suppose that the function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is symmetric. Then $Y \in \partial f_{\mathcal{H}}(X)$ if and only if $\lambda(Y) \in \partial f(\lambda(X))$ and there exists a unitary matrix V with $V^*XV = \text{diag}\lambda(X)$ and $V^*YV = \text{diag}\lambda(Y)$.*

Proof. For Hermitian matrices X and Y, Y lies in $\partial f_{\mathcal{H}}(X)$ exactly when

$$\text{tr}XY = f_{\mathcal{H}}(X) + (f_{\mathcal{H}})^*(Y) = f(\lambda(X)) + f^*(\lambda(Y)) \geq \lambda(X)^T\lambda(Y) \geq \text{tr}XY,$$

by Theorem 2.3, and the result follows by Theorem 2.2. □

COROLLARY 3.2. *Suppose that the function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is symmetric, convex, and lower semicontinuous. Then the function $f_{\mathcal{H}} : \mathcal{H} \to (-\infty, +\infty]$ is essentially smooth if and only if f is essentially smooth. In this case, for any Hermitian X in* int(dom$f_{\mathcal{H}}$) *we have that*

$$\nabla f_{\mathcal{H}}(X) = V\text{diag}(\nabla f(\lambda(X)))V^* \tag{3.2}$$

for any unitary V satisfying $V^*XV = \text{diag}\lambda(X)$.

Proof. Suppose that f is essentially smooth (the converse is straightforward by restricting to diagonal matrices). If $\partial f_{\mathcal{H}}(X)$ is nonempty then by Theorem 3.1 it is exactly

$$\{V\text{diag}(\nabla f(\lambda(X)))V^* \mid V^*XV = \text{diag}\lambda(X),\ V \in \mathcal{U}\}.$$

Therefore every element of the convex set $\partial f_{\mathcal{H}}(X)$ has the same Frobenius norm, $\|\nabla f(\lambda(X))\|_2$, and hence this set is a singleton (because the Frobenius norm is strict). Thus $f_{\mathcal{H}}$ is essentially smooth. □

Some additional comments are warranted in regard to this corollary. Notice that the proof above actually shows that if the function f is symmetric, lower semicontinuous, and convex then the function $f_{\mathcal{H}}$ is differentiable at X whenever f is differentiable at $\lambda(X)$, with gradient given by (3.2). Furthermore, using Davis's result [4] in place of Corollary 2.4 allows us to dispense with the assumption of lower semicontinuity in this observation. In fact a completely different approach [17] shows that convexity is not needed for the gradient formula (3.2): this paper also derives a result analogous to Theorem 3.1 for the Clarke generalized derivative.

Taking conjugates gives the following result.

COROLLARY 3.3. *Suppose that the function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is symmetric, convex and lower semicontinuous. Then the function $f_{\mathcal{H}} : \mathcal{H} \to (-\infty, +\infty]$ is essentially strictly convex if and only if f is essentially strictly convex.*

Again, exactly parallel arguments show the corresponding results for real symmetric matrices.

4. Examples. In this section we will see that many of the classically known convex functions on the Hermitian matrices can be derived from our main result. Any symmetric convex function is Schur-convex [19, Proposition 3.C.2], and not surprisingly, many of the standard Schur-convex functions are symmetric, convex, and lower semicontinuous. We will simply illustrate a variety of examples.

The simplest class of examples are functions of $\lambda \in \mathbb{R}^n$ of the form

$$\sum_{i=1}^{n} g(\lambda_i) \text{ for } g : \mathbb{R} \to (-\infty, +\infty] \text{ convex, lower semicontinuous.} \tag{4.1}$$

In particular, in (4.1) we could take

$$g(\mu) = \begin{cases} 0, & \text{if } \mu \geq 0, \\ +\infty, & \text{if } \mu < 0, \end{cases} \tag{4.2}$$

$$g(\mu) = \begin{cases} 1/\mu, & \text{if } \mu > 0, \\ +\infty, & \text{if } \mu \leq 0, \end{cases} \tag{4.3}$$

$$g(\mu) = \begin{cases} -\log \mu, & \text{if } \mu > 0, \\ +\infty, & \text{if } \mu \leq 0. \end{cases} \tag{4.4}$$

More generally, we could consider

$$\sum_{P \in \mathcal{P}} h(P\lambda) \text{ for } h : \mathbb{R}^n \to (-\infty, +\infty] \text{ convex, lower semicontinuous.} \tag{4.5}$$

This will encompass such functions as $\sum_i |\lambda_i - \bar{\lambda}|$ (where $\bar{\lambda} = \sum_i \lambda_i / n$) and $\sum_{i,j} |\lambda_i - \lambda_j|$ for example.

For any symmetric set $C \subset \mathbb{R}^n$ (in other words, closed under coordinate permutations) the support function $\sup\{\phi^T \lambda \mid \phi \in C\}$ will be convex, lower semicontinuous and symmetric. In this way we obtain the examples (for $k = 1, 2, \ldots, n$)

$$\text{sum of the } k \text{ largest elements of } \{\lambda_1, \lambda_2, \ldots, \lambda_n\} \tag{4.6}$$

(by taking $C = \{\phi \mid 0 \leq \phi_i \leq 1 \ (i = 1, 2, \ldots, n),\ \sum \phi_i = k\}$), and similarly

$$-\text{ sum of the } k \text{ smallest elements of } \{\lambda_1, \lambda_2, \ldots, \lambda_n\}. \tag{4.7}$$

In particular, any symmetric gauge function will be symmetric, convex, and continuous (see [14, p. 438]). Examples are $\|\lambda\|_p$ for $1 \leq p \leq +\infty$, and

$$\text{sum of the } k \text{ largest elements of } \{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|\}. \tag{4.8}$$

For $k = 1, 2, \ldots, n$, the elementary symmetric function $S_k(\lambda)$ and the complete symmetric function $C_k(\lambda)$ have the property that $-(S_k(\lambda))^{1/k}$ and $(C_k(\lambda))^{1/k}$ are both symmetric, convex, and continuous on the nonnegative orthant $\mathbb{R}^n_+$ [19, 3.F.2 and 3.F.5]. A particular example is

$$\begin{cases} -(\lambda_1 \lambda_2 \ldots \lambda_n)^{1/n}, & \text{if } \lambda \geq 0, \\ +\infty, & \text{otherwise} \end{cases} \tag{4.9}$$

Furthermore, for any strictly positive real a, the function

$$\begin{cases} S_k(\lambda_1^{-a}, \lambda_2^{-a}, \ldots, \lambda_n^{-a}), & \text{if } \lambda > 0, \\ +\infty, & \text{otherwise} \end{cases} \tag{4.10}$$

is symmetric, convex, and lower semicontinuous [19, 3.G.1.m]. Somewhat analogously to (4.5), we could consider

$$\max_{P \in \mathcal{P}} h(P\lambda) \text{ for } h : \mathbb{R}^n \to (-\infty, +\infty] \text{ convex, lower semicontinuous.}$$

Examples are (4.6),(4.7) and (on the domain $\mathbb{R}^n_+$)

$$-\left(\text{product of the } k \text{ largest elements of } \{\lambda_1, \lambda_2, \ldots, \lambda_n\}\right)^{1/k}. \tag{4.11}$$

For $n \times n$ Hermitian matrices X and Y we write $X \preceq Y$ if $Y - X$ is positive semidefinite, and $X \prec Y$ if $Y - X$ is positive definite. We denote the identity matrix by I. Now by Corollary 2.4, each of the examples above induces a lower semicontinuous convex function on the Hermitian matrices, and Theorem 2.3 gives a formula for the conjugate. Thus example (4.2) induces the indicator function of the cone of positive semidefinite matrices $\{X \succeq 0\}$, which is thus a closed, convex cone, and computing the conjugate shows that this cone is self-dual (Fejer's Theorem; see [14]):

$$\text{tr}XY \geq 0 \text{ for all } X \succeq 0 \quad \Leftrightarrow \quad Y \succeq 0.$$

The functions (4.3), (4.4), (4.6), and (4.7) (whose conjugates may be computed directly) induce, respectively, the lower semicontinuous, convex, matrix functions

$$\begin{cases} \text{tr}X^{-1}, & \text{if } X \succ 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

$$\begin{cases} -\log\det X, & \text{if } X \succ 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

$$\sum_{i=n-k+1}^{n} \lambda_i(X) \text{ and} \tag{4.12}$$

$$-\sum_{i=1}^{k} \lambda_i(X),$$

and by applying Theorem 2.3 we see that the corresponding conjugate functions are

$$\begin{cases} -2\text{tr}(-Y)^{1/2}, & \text{if } Y \prec 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

$$\begin{cases} -n - \log\det(-Y), & \text{if } Y \prec 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

$$\begin{cases} 0, & \text{if } 0 \preceq Y \preceq I, \text{ with tr}Y = k, \\ +\infty, & \text{otherwise, and} \end{cases} \tag{4.13}$$

$$\begin{cases} 0, & \text{if } -I \preceq Y \preceq 0, \text{ with } \mathrm{tr}Y = n-k, \\ +\infty, & \text{otherwise.} \end{cases}$$

The function $\|\lambda\|_p$ induces the Schatten p-norm, special cases being the trace norm ($p = 1$), the Frobenius norm ($p = 2$), and the spectral norm ($p = \infty$). The function (4.8) induces the Ky Fan k-norm $\sum_{i=n-k+1}^{n} |\lambda_i(X)|$. The functions (4.9), (4.10), and (4.11) induce the matrix functions

$$\begin{cases} -(\det X)^{1/n}, & \text{if } X \succeq 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

$$\begin{cases} S_k(\lambda_1(X)^{-a}, \lambda_2(X)^{-a}, \ldots, \lambda_n(X)^{-a}), & \text{if } X \succ 0, \\ +\infty, & \text{otherwise, and} \end{cases}$$

$$\begin{cases} -(\prod_{i=n-k+1}^{n} \lambda_i(X))^{1/k}, & \text{if } X \succeq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

All of these examples may be found in [19, 16.F] for example. Many are easily seen to be strictly convex with the help of Corollary 3.3.

As a final example, suppose that the set $C \subset \mathbb{R}^n$ is closed, convex, and symmetric. By applying Corollary 2.4 with f the indicator function of C we see immediately that the set of Hermitian matrices X with $\lambda(X) \in C$ is a closed, convex set (cf. [16]).

Theorem 3.1 can be used to calculate subdifferentials. Consider for example the sum of the k largest eigenvalues of X (example (4.12)). The problem of deriving expressions for the subdifferential of this function is considered via the computation of the conjugate function (4.13) in [8, 12, 24]. If the function $f(\lambda)$ is given by (4.6) then at any point λ in $\mathbb{R}^n$ satisfing $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ it is a straightforward calculation to check that $\mu \in \partial f(\lambda)$ if and only if

$$\mu_i \begin{cases} = & 0 & \text{if } \lambda_i < \lambda_{n-k+1}, \\ \in & [0,1] & \text{if } \lambda_i = \lambda_{n-k+1}, \\ = & 1 & \text{if } \lambda_i > \lambda_{n-k+1}, \end{cases}$$

with $\sum \mu_i = k$. Using Theorem 3.1 we see that the subdifferential of X is exactly the set of matrices $V\mathrm{diag}(\mu)V^*$ with unitary V satisfying $V^*XV = \mathrm{diag}\lambda(X)$ and μ in $\mathbb{R}^n$ satisfying

$$\mu_i \begin{cases} = & 0 & \text{if } \lambda_i(X) < \lambda_{n-k+1}(X), \\ \in & [0,1] & \text{if } \lambda_i(X) = \lambda_{n-k+1}(X), \\ = & 1 & \text{if } \lambda_i(X) > \lambda_{n-k+1}(X), \end{cases}$$

and $\sum \mu_i = k$. In particular, for example, for the maximum eigenvalue function $\lambda_n(X)$ (which is the case $k = 1$) we obtain the well-known result

$$\partial \lambda_n(X) = \mathrm{conv}\{vv^* \mid \|v\| = 1, \ Xv = \lambda_n(X)v\}.$$

A similar expression can be obtained for the subdifferential of the Ky Fan k-norm.

5. Fenchel duality, semidefinite programming, and quasi-Newton updates. In this section we illustrate how the conjugacy formula derived in §2 can be used to study duality properties of optimization problems involving real symmetric matrices. In particular we can study analogues of linear programming over the cone of positive semidefinite matrices (see [27, 23, 1, 33, 35, 2]), penalized versions of such problems (see, for example, [1, 15, 21, 2]), and convex optimization problems leading to well-known quasi-Newton formulae for minimization algorithms [9, 34].

Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are finite-dimensional inner-product spaces. For functions $F : \mathcal{X} \to (-\infty, +\infty]$ and $G : \mathcal{Y} \to (-\infty, +\infty]$, and a linear map $A : \mathcal{X} \to \mathcal{Y}$, consider the optimization problem

$$\alpha = \inf_{X \in \mathcal{X}} \{F(X) + G(AX)\}. \tag{5.1}$$

If we define the *adjoint map* $A^T : \mathcal{Y} \to \mathcal{X}$ by

$$\langle AX, Y \rangle = \langle X, A^T Y \rangle \quad \text{for all } X \in \mathcal{X}, Y \in \mathcal{Y}, \tag{5.2}$$

then we can associate with the *primal* problem (5.1) a *dual* problem

$$\beta = \sup_{Y \in \mathcal{Y}} \{-F^*(A^T Y) - G^*(-Y)\}. \tag{5.3}$$

The *weak duality inequality* $\alpha \geq \beta$ is trivial to check. Fenchel duality results give conditions ensuring that $\alpha = \beta$. We consider one such particular result. We say that the function G is *polyhedral* if its *epigraph* $\text{epi}G = \{(Y, r) \in \mathcal{Y} \times \mathbb{R} \mid r \geq G(Y)\}$ is a polyhedron. We denote the interior of a convex set $C \subset \mathcal{X}$ with respect to its affine span by $\text{ri}C$. The various parts of the following result (stated for $\mathcal{X} = \mathbb{R}^n$) may be found in [26].

THEOREM 5.1. *Suppose in problem* (5.1) *that the functions F and G are convex, with G polyhedral. Then providing that there exists an X in* $\text{ri}(\text{dom}F)$ *with AX in* $\text{dom}G$*, the primal and dual values* (5.1) *and* (5.3) *are equal, and the dual value is attained when finite. In this case, X_0 and Y_0 are optimal for the primal and dual problems respectively if and only if $-Y_0 \in \partial G(AX_0)$ and $A^T Y_0 \in \partial F(X_0)$. In particular, if F is lower semicontinuous and F^* is differentiable at $A^T Y_0$ then the unique primal optimal solution is $X_0 = \nabla F^*(A^T Y_0)$.*

As an example, consider the semidefinite programming problem (cf. [21]):

$$\begin{cases} \inf & \text{tr}EX \\ \text{subject to} & X \in B + \mathcal{L}, \\ & 0 \preceq X \in \mathcal{S}, \end{cases} \tag{5.4}$$

where $\mathcal{S}$ denotes the $n \times n$ real symmetric matrices, E and B are given symmetric matrices, and $\mathcal{L}$ is a given subspace of $\mathcal{S}$. Observe that for any function $H : \mathcal{X} \to (-\infty, +\infty]$,

$$(H + \langle E, \cdot \rangle)^*(Y) = H^*(Y - E). \tag{5.5}$$

If we choose spaces $\mathcal{X} = \mathcal{Y} = \mathcal{S}$, the map A to be the identity,

$$F(X) = \begin{cases} \text{tr}EX, & \text{if } X \succeq 0, \\ +\infty, & \text{otherwise, and} \end{cases}$$

$$G(X) = \begin{cases} 0, & \text{if } X \in B + \mathcal{L}, \\ +\infty, & \text{otherwise,} \end{cases}$$

then it is easy to calculate directly (or using (4.2) with (5.5)) that

$$F^*(Y) = \begin{cases} 0, & \text{if } Y \preceq E, \\ +\infty, & \text{otherwise, and} \end{cases}$$

$$G^*(Y) = \begin{cases} \mathrm{tr}BY, & \text{if } Y \in \mathcal{L}^\perp, \\ +\infty, & \text{otherwise,} \end{cases}$$

where the orthogonal complement $\mathcal{L}^\perp$ is the subspace of symmetric matrices Y satisfying $\mathrm{tr}XY = 0$ whenever $X \in \mathcal{L}$. Hence the dual problem is (cf. [27])

$$\text{(5.6)} \qquad \begin{cases} \sup & \mathrm{tr}BY \\ \text{subject to} & Y \in \mathcal{L}^\perp, \\ & E \succeq Y \in \mathcal{S}. \end{cases}$$

We can emphasize the symmetry with the primal problem (5.4) by setting $Z = E - Y$, if so desired. Now Theorem 5.1 shows that providing there exists a positive definite X in $B + \mathcal{L}$, the primal and dual values are equal, with attainment in the dual if it is feasible. In this case complementary slackness results are also straightforward to derive: feasible X_0 and Y_0 are respectively primal and dual optimal if and only if $\mathrm{tr}X_0(E - Y_0) = 0$.

A related problem, arising for example when we replace the constraint $X \succeq 0$ in problem (5.4) by adding a penalty function to the objective function, is

$$\text{(5.7)} \qquad \begin{cases} \inf & \mathrm{tr}EX + \delta f(\lambda(X)) \\ \text{subject to} & X \in B + \mathcal{L}, \end{cases}$$

where $\delta > 0$ is a small penalty parameter and the function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is lower semicontinuous and convex with $\mathrm{cl}(\mathrm{dom}f) = \mathbb{R}^n_+$. An example is (4.4), giving the "logarithmic barrier" penalized problem

$$\text{(5.8)} \qquad \begin{cases} \inf & \mathrm{tr}EX - \delta \log\det X \\ \text{subject to} & X \in B + \mathcal{L}, \\ & 0 \prec X \in \mathcal{S}. \end{cases}$$

The dual problem for (5.7), using the real symmetric version of Theorem 2.3, is

$$\text{(5.9)} \qquad \begin{cases} \inf & \mathrm{tr}BY - \delta f^*(\delta^{-1}\lambda(Y - E)) \\ \text{subject to} & Y \in \mathcal{L}^\perp. \end{cases}$$

Again, providing there is a positive definite matrix X in $B + \mathcal{L}$, Theorem 5.1 shows that the primal and dual values are equal, with attainment in (5.9) when it is feasible. For the primal problem (5.8) we obtain the dual problem

$$\begin{cases} \inf & \mathrm{tr}BY - \delta \log\det(E - Y) + \delta n(\log\delta - 1) \\ \text{subject to} & Y \in \mathcal{L}^\perp, \\ & E \succ Y, \end{cases}$$

which is just the logarithmic barrier penalized version of the original dual problem (5.6). Semidefinite programming problems involving other objective functions can also be studied using these techniques. The maximum eigenvalue is an example [22].

To conclude, we consider problems of the form

$$(5.10)\qquad \begin{cases} \inf & \mathrm{tr}EX + f(\lambda(X)) \\ \text{subject to} & Xs = y, \\ & X \in \mathcal{S}, \end{cases}$$

where again the function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is lower semicontinuous and convex with $\mathrm{cl}(\mathrm{dom} f) = \mathbb{R}^n_+$, and the vectors s and y in $\mathbb{R}^n$ are given. Such problems arise in the context of characterizing quasi-Newton Hessian updates satisfying the "secant equation" $Xs = y$ (see, for example, [9]). The given real symmetric matrix E is derived from the old Hessian approximation. Once again a good example comes from (4.4), which gives the problem

$$(5.11)\qquad \begin{cases} \inf & \mathrm{tr}EX - \log\det X \\ \text{subject to} & Xs = y, \\ & 0 \prec X \in \mathcal{S}. \end{cases}$$

The adjoint of the linear map $A : \mathcal{S} \to \mathbb{R}^n$ defined by $AX = Xs$ is easily computed to be given by $A^T z = (zs^T + sz^T)/2$ for z in $\mathbb{R}^n$. If we choose $F : \mathcal{S} \to (-\infty, +\infty]$ to be given by $F(X) = \mathrm{tr}EX + f(\lambda(X))$ and $G : \mathbb{R}^n \to (-\infty, +\infty]$ defined by $G(w) = 0$ if $w = y$ and $+\infty$ otherwise, then applying Theorem 2.3 gives the dual problem

$$(5.12)\qquad \sup_{z \in \mathbb{R}^n} \{y^T z - f^*(\lambda(-E + (zs^T + sz^T)/2))\}.$$

If $s^T y > 0$ then it is well known that there exists a positive definite matrix X satisfying the secant equation $Xs = y$ (for this and other standard theory of quasi-Newton updates, see [6]). Hence Theorem 5.1 applies to show that the primal and dual values are equal, and that (5.12) is attained when finite. Furthermore, if z_0 solves (5.12) and we denote the matrix $-E + (z_0 s^T + s z_0^T)/2$ by E_0, and if f^* is differentiable at $\lambda(E_0)$, then by the comment after Corollary 3.2 the unique optimal solution of (5.10) is $X_0 = V \nabla f^*(\lambda(E_0)) V^*$ for any unitary matrix V satisfying $V^* E_0 V = \mathrm{diag}\lambda(E_0)$.

In particular, the dual problem for (5.11) becomes

$$(5.13)\qquad \sup_{z \in \mathbb{R}^n} \{y^T z + \log\det(\lambda(E - (zs^T + sz^T)/2))\} + n.$$

This is straightforward to solve explicitly, using the fact that $\nabla \log\det X = X^{-1}$, and assuming that E is positive definite (which ensures that (5.13) has the feasible solution $z = 0$). The resulting optimal solution X_0 of (5.11) is the "BFGS update" of E^{-1} (see [6, p. 205] and [9]).

Acknowledgments. The author wishes to thank O. Güler, M. Teboulle, and H. Wolkowicz for drawing his attention to [4, 18, 10], and an anonymous referee for many helpful suggestions, including a number of references to literature on matrix norms.

REFERENCES

[1] F. Alizadeh, *Optimization over the positive definite cone: interior point methods and combinatorial applications*, in Advances in Optimization and Parallel Computing, P. Pardolos, ed., North-Holland, Amsterdam, 1992, pp. 1–25.

[2] ———, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, SIAM J. Optim., 5 (1995), pp. 13–51.

[3] J. ARAZY, *On the geometry of the unit ball of unitary matrix spaces*, Integral Equations Operator Theory, 4 (1981), pp. 151–171.
[4] C. DAVIS, *All convex invariant functions of hermitian matrices*, Archiv der Mathematik, 8 (1957), pp. 276–278.
[5] E.M. DE SÁ, *Exposed faces and duality for symmetric and unitarily invariant norms*, Linear Algebra Appl., 197,198 (1994), pp. 429–450.
[6] J. E. DENNIS AND R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[7] P. A. FILLMORE AND J. P. WILLIAMS, *Some convexity theorems for matrices*, Glasgow Math. J., 12 (1971), pp. 110–117.
[8] R. FLETCHER, *Semi-definite matrix constraints in optimization*, SIAM J. Control Optim., 23 (1985), pp. 493–513.
[9] ———, *A new variational result for quasi-Newton formulae*, SIAM J. Optim., 1 (1991), pp. 18–21.
[10] S. FRIEDLAND, *Convex spectral functions*, Linear Multilinear Algebra, 9 (1981), pp. 299–316.
[11] G. H. HARDY, J. E. LITTLEWOOD, AND G. PÓLYA, *Inequalities*, Cambridge University Press, Cambridge, UK, 1952.
[12] J.-B. HIRIART-URRUTY, A. SEEGER, AND D. YE, *Sensitivity analysis for a class of convex functions defined over a space of symmetric matrices*, Vol. 382, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, New York, 1992, pp. 133–154.
[13] J.-B. HIRIART-URRUTY AND D. YE, *Sensitivity Analysis of all Eigenvalues of a Symmetric Matrix*, Tech. report, Laboratoire d'analyse numerique, Université Paul Sabatier, Toulouse, France, 1992.
[14] R. A. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[15] F. JARRE, *An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices*, SIAM J. Control Optim., 31 (1993), pp. 1360–1377.
[16] F. JOHN, *On symmetric matrices whose eigenvalues satisfy linear inequalities*, Proc. American Mathematical Society, 17 (1966), pp. 1140–1146.
[17] A. S. LEWIS, *Derivatives of Spectral Functions*, Tech. Report CORR 94-04, University of Waterloo, 1994; Math. Oper. Res., to appear.
[18] M. MARCUS, *Convex functions of quadratic forms*, Duke Math. J., 24 (1957), pp. 321–326.
[19] A. W. MARSHALL AND I. OLKIN, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York, 1979.
[20] L. MIRSKY, *On the trace of matrix products*, Math. Nach., 20 (1959), pp. 171–174.
[21] Y. E. NESTEROV AND A. S. NEMIROVSKY, *Interior Point Polynomial Methods in Convex Programming*, Society for Industrial and Applied Mathematics, Philadelphia, 1994.
[22] M.L. OVERTON, *On minimizing the maximum eigenvalue of a symmetric matrix*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 256–268.
[23] ———, *Large-scale optimization of eigenvalues*, SIAM J. Optim., 2 (1992), pp. 88–120.
[24] M.L. OVERTON AND R.S. WOMERSLEY, *Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices*, Math. Programming, Series B, 62 (1993), pp. 321–357.
[25] F. RENDL AND H. WOLKOWICZ, *Applications of parametric programming and eigenvalue maximization to the quadratic assignment problem*, Math. Programming, 53 (1992), pp. 63–78.
[26] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[27] A. SHAPIRO, *Extremal problems on the set of nonnegative definite matrices*, Linear Algebra Appl., 67 (1985), pp. 7–18.
[28] A. SHAPIRO AND M.K.H. FAN, *On eigenvalue optimization*, SIAM J. Optim., 5 (1995), pp. 552–569.
[29] C.M. THEOBALD, *An inequality for the trace of the product of two symmetric matrices*, Math. Proc. Cambridge Philosophical Soc., 77 (1975), p. 265.
[30] J. VON NEUMANN, *Some matrix inequalities and metrization of matric-space*, Tomsk University Rev., 1 (1937), pp. 286–300. In Collected Works, Vol. IV, Pergamon Press, Oxford, 1962, pp. 205–218.
[31] G. A. WATSON, *Characterization of the subdifferential of some matrix norms*, Linear Algebra Appl., 170 (1992), pp. 33–45.
[32] ———, *On matrix approximation problems with Ky Fan k norms*, Numer. Algorithms, 5 (1993), pp. 263–272.
[33] H. WOLKOWICZ, *Explicit Solutions for Interval Semidefinite Linear Programs*, Tech. Report CORR 93-29, Department of Combinatorics and Optimization, University of Waterloo, 1993.

[34] H. WOLKOWICZ AND Q. ZHAO, *An Efficient Region of Optimal Updates for Least Change Secant Methods*, Tech. Report CORR 92-27, Department of Combinatorics and Optimization, University of Waterloo, 1992.

[35] B. YANG AND R.J. VANDERBEI, *The Simplest Semidefinite Programs are Trivial*, Tech. report, Program in Statistics and Operations Research, Princeton University, Princeton, NJ, 1993.

[36] K. ZIĘTAK, *On the characterization of the extremal points of the unit sphere of matrices*, Linear Algebra Appl., 106 (1988), pp. 57–75.

[37] ———, *Subdifferentials, faces and dual matrices*, Linear Algebra Appl., 185 (1993), pp. 125–141.