Hyperbolic Polynomials and Convex Analysis

Heinz H. Bauschke*, Osman Güler†, Adrian S. Lewis‡, and Hristo S. Sendov§

September 29, 1998

Abstract

A homogeneous polynomial $p(x)$ is *hyperbolic* with respect to a given vector d if the real polynomial $t \mapsto p(x + td)$ has all real roots for all vectors x. We show that any symmetric convex function of these roots is a convex function of x, generalizing a fundamental result of Gårding. Consequently we are able to prove a number of deep results about hyperbolic polynomials with ease. In particular, our result subsumes Davis's characterization of convex functions of the eigenvalues of Hermitian matrices, and von Neumann's classical result on unitarily invariant matrix norms. We then develop various convex-analytic tools for such symmetric functions, of interest in interior-point methods for optimization problems posed over related cones.

*Department of Combinatorics & Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Email: `bauschke@cecm.sfu.ca`. Research supported by an NSERC Postdoctoral Fellowship and by the Department of Combinatorics & Optimization, University of Waterloo.

†Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland 21228, USA. Email: `guler@math.umbc.edu`. Research partially supported by the National Science Foundation under grant DMS-9623135

‡Department of Combinatorics & Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Email: `aslewis@math.uwaterloo.ca`. Research supported by NSERC.

§Department of Combinatorics & Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Email: `hssendov@barrow.uwaterloo.ca`. Research supported by NSERC.

1991 M.R. Subject Classification.
Primary 90C25; Secondary 15A45, 52A41.
Key words and phrases. convex analysis, eigenvalue, Gårding's inequality, hyperbolic barrier function, hyperbolic polynomial, hyperbolicity cone, interior-point method, semidefinite program, singular value, symmetric function, unitarily invariant norm.

1 Introduction

A beautiful result of Davis [6] states that for any symmetric convex function f on $\mathbb{R}^n$, the function

$$Z \in S^n \mapsto f(\lambda(Z)) \tag{1}$$

is convex. We call a function *symmetric* if it is invariant under permutations of its arguments. For any matrix Z in S^n, the space of $n \times n$ real symmetric matrices, the vector $\lambda(Z)$ has components the eigenvalues of Z, arranged in decreasing order.
This convexity theorem has a strong resemblance to a famous result of von Neumann [31], characterizing unitarily invariant matrix norms as symmetric gauge functions of the singular values. Indeed, the analogy is not accidental: the paper [19] develops an axiomatic framework subsuming both models, and at a more sophisticated level, both results follow quickly from the Kostant convexity theorem in semisimple Lie theory [20].
The work we describe in this current paper also concerns the above type of convexity result, but with a very different and remarkably simple approach. To illustrate the key idea, consider the determinant as a function on S^n. This function is a homogeneous polynomial which is *hyperbolic* with respect to the identity matrix I: that is, the real polynomial

$$t \in \mathbb{R} \mapsto \det(Z - tI)$$

has all real roots, namely the eigenvalues $\lambda_i(Z)$. The properties of such polynomials play a significant role in the partial differential equations literature (see for example [13])), but we use just one, central result, due to Gårding [8]: the largest root $\lambda_1(\cdot)$ is always a convex function.
Working from Gårding's result, we show, just like Davis's theorem, that *any* symmetric convex function of the roots $\lambda_i(\cdot)$ is convex. The richness of the

class of hyperbolic polynomials then allows us to derive many elegant (and often classical) inequalities in a unified fashion. Examples include beautiful properties of the elementary symmetric functions. One particular hyperbolic polynomial leads back to von Neumann's result.

The second half of this paper is convex-analytic in character. Functions of the form (1) are fundamental in eigenvalue optimization and semidefinite programming [22]. They have an attractive duality theory: the Fenchel conjugate of the function (1) is described elegantly by the formula $(f \circ \lambda)^* = f^* \circ \lambda$ [18]. Analogously, von Neumann proved a similar result for unitarily invariant norms, useful in matrix approximation problems [14].

Hyperbolic polynomials offer a a unifying framework in which to study such convexity and duality results. They also have potential application in modern interior point methodology. Associated with any hyperbolic polynomial comes a closed convex *hyperbolicity cone* which, with the above notation, we can write

$$\{Z : \lambda_i(Z) \geq 0 \ \forall i\}.$$

For example, in the symmetric matrix case this is simply the cone of positive semidefinite matrices. Güler has shown how optimization problems over such cones are good candidates for interior point algorithms analogous to the dramatically successful techniques current in semidefinite programming [10]. With these aims in mind, we develop an attractive duality theory and convex-analytic tools for symmetric convex functions of the roots associated with general hyperbolic polynomials.

Notation

We write $\mathbb{R}^m_{++}$ (resp. $\mathbb{R}^m_+$) for the set $\{u \in \mathbb{R}^m : u_i > 0, \forall i\}$ (resp. $\{u \in \mathbb{R}^m : u_i \geq 0, \forall i\}$. The *closure* (resp. *boundary*, *convex hull*, *linear span*) of a set S is denoted $\operatorname{cl} S$ (resp. $\operatorname{bd} S$, $\operatorname{conv} S$, $\operatorname{span} S$). A *cone* is a nonempty set that contains every nonnegative multiple of all its members. If $u \in \mathbb{R}^m$, then $u_\downarrow$ is the vector u with its coordinates arranged decreasingly; also, $U_\downarrow := \{u_\downarrow : u \in U\}$, for every subset U of $\mathbb{R}^m$. The *transpose* of a matrix (or vector) A is denoted A^T. The *identity* matrix or map is written I. Suppose X is a Euclidean space with inner product $\langle \cdot, \cdot \rangle$ and $h : X \to [-\infty, +\infty]$ is convex, then h^* (resp. ∂h, ∇h, $\operatorname{dom} h$) stands for the *Fenchel conjugate* (resp. *subdifferential map*, *gradient map*, *domain*) of h. (Rockafellar's [30] is the standard reference for these notions from convex analysis.) Higher order derivatives are denoted by $\nabla^k h$. If $U \subseteq X$, then the *positive polar cone* is $U^{\oplus} := \{x \in X : \langle x, U \rangle \geq 0\}$. If A is a linear operator between Euclidean spaces, then its *conjugate* is written A^*. The *range* of a map λ is denoted by $\operatorname{ran} \lambda$. Finally, if A, B are two subsets of X, then $d(A, B) := \inf \|A - B\|$ is the *distance* between A and B.

2 Tools

We assume throughout the paper that

X is a finite-dimensional real vector space.

This section contains a selection of important facts on hyperbolic polynomials from Gårding's fundamental work [8], and a deep inequality on elementary symmetric functions.

Hyperbolic polynomials and eigenvalues

Definition 2.1 (homogeneous polynomial) Suppose p is a nonconstant polynomial on X and m is a positive integer. Then p is *homogeneous of degree* m, if $p(tx) = t^m p(x)$, for all $t \in \mathbb{R}$ and every $x \in X$.

Definition 2.2 (hyperbolic polynomial) Suppose that p is a homogeneous polynomial of degree m on X and $d \in X$ with $p(d) \neq 0$. Then p *is hyperbolic with respect to* d, if the polynomial $t \mapsto p(x + td)$ (where t is a scalar) has only real zeros, for every $x \in X$.

Definition 2.3 ("eigenvalues and trace") Suppose p is hyperbolic with respect to $d \in X$ of degree m. Then for every $x \in X$, we can write

$$p(x + td) = p(d) \prod_{i=1}^{m} (t + \lambda_i(x))$$

and assume without loss of generality that $\lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_m(x)$. The corresponding map $X \to \mathbb{R}^m_{\downarrow} : x \mapsto (\lambda_1(x), \ldots, \lambda_m(x))$ is denoted by λ and called the *eigenvalue map (with respect to p and d)*. We say that $\lambda_i(x)$

is the i^{th} *largest eigenvalue of x (with respect to p and d)* and define *the sum of the k largest eigenvalues* by $\sigma_k := \sum_{i=1}^{k} \lambda_i$, for every $1 \le k \le m$. The function σ_m is called the *trace.*

The eigenvalues $\{\lambda_i(x)\}$ are thus the roots of the polynomial $t \mapsto p(x - td)$. It follows that the trace σ_m is linear (see also the paragraph following Fact 2.13). Unless stated otherwise, we assume throughout the paper that

> p is a hyperbolic polynomial of degree m with respect to d, with eigenvalue map λ and $\sigma_k := \sum_{i=1}^{k} \lambda_k$,

for every $1 \le k \le m$. The notions "eigenvalues" and "trace" are well-motivated by the the following example.

The Hermitian matrices. Let X be the real vector space of the $m \times m$ Hermitian matrices and $p := \det$. Then p is hyperbolic of degree m with respect to $d := I$ and λ maps $x \in X$ to its eigenvalues, arranged decreasingly. Thus for every $1 \le k \le m$, the function σ_k is indeed the sum of the k largest eigenvalues and σ_m is the (ordinary) trace.

As we go, we will point out what some of the results become in the important case of the *Hermitian matrices.* Further examples are provided in Section 6. A simple way to generate new hyperbolic polynomials is differentiation:

Proposition 2.4 If $m > 1$, then $q(x) := \frac{d}{dt} p(x + td)\big|_{t=0} = \big(\nabla p(x)\big)(d)$ is hyperbolic with respect to d.

Proof. This is essentially Rolle's theorem; see also [8, Lemma 1]. ■

The following property of the eigenvalues is well-known ([8, Equation (2)]).

Fact 2.5 For all $r, s \in \mathbb{R}$ and every $1 \le i \le m$:

$$\lambda_i(rx + sd) = \begin{cases} r\lambda_i(x) + s, & \text{if } r \ge 0; \\ r\lambda_{m+1-i}(x) + s, & \text{otherwise.} \end{cases}$$

Hence the eigenvalue map λ is *positively homogeneous* ($\lambda(tx) = t\lambda(x)$, for all $t \ge 0$ and every $x \in X$) and continuous (use, for instance, [29, Appendix A]). Gårding showed that the largest eigenvalue map is *sublinear*, that is, positively homogeneous and convex.

Theorem 2.6 (Gårding) The largest eigenvalue map λ_1 is sublinear.

Proof. Positive homogeneity follows from Fact 2.5. Now Gårding showed that λ_m is concave [8, Theorem 2], which is equivalent to the convexity of λ_1, since $\lambda_1(-x) = -\lambda_m(x)$, for every $x \in X$. ■

The Hermitian matrices (continued). It is well-known that the largest eigenvalue map is convex in this case; see, for instance, [12].

Hyperbolicity cone

Definition 2.7 (hyperbolicity cone) The *hyperbolicity cone of p with respect to d*, written $C(d)$ or $C(p, d)$, is the set $\{x \in X : p(x + td) \neq 0, \forall t \geq 0\}$.

Fact 2.8 $C(d) = \{x \in X : \lambda_m(x) > 0\}$. Hence $C(d)$ is an open convex cone that contains d with closure $\operatorname{cl} C(d) = \{x \in X : \lambda_m(x) \geq 0\}$. If $c \in C(d)$, then p is hyperbolic with respect to c and $C(c) = C(d)$.

Proof. See Gårding's [8, Section 2]. ■

Definition 2.9 (complete hyperbolic polynomial) p is *complete* if

$$\{x \in X : \lambda(x) = 0\} = \{0\}.$$

Fact 2.10 Suppose p is hyperbolic with respect to d, with corresponding eigenvalue map λ and hyperbolicity cone $C(d)$. Then

$$\begin{aligned} \{x \in X : \lambda(x) = 0\} &= \{x \in X : x + C(d) = C(d)\} \\ &= \{x \in X : p(tx + y) = p(y), \forall y \in X, \forall t \in \mathbb{R}\}. \end{aligned}$$

Consequently, $\{x \in X : \lambda(x) = 0\} = \operatorname{cl} C(d) \cap (-\operatorname{cl} C(d))$.

Proof. See Gårding's [8, Section 3]. The "Consequently" part follows readily from the displayed equation and the openness of $C(d)$. ■

The Hermitian matrices (continued). The hyperbolicity cone of $p = \det$ with respect to $d = I$ is the set of all positive definite matrices. The polynomial $p = \det$ is complete, since every nonzero Hermitian matrix has at least one nonzero eigenvalue.

Elementary symmetric functions

Definition 2.11 (symmetric function) A function f on $\mathbb{R}^m$ is *symmetric*, if $f(u) = f(u_{\pi(i)})$, for all permutations π of $\{1, \ldots, m\}$ and every $u \in \mathbb{R}^m$.

Definition 2.12 (elementary symmetric functions) For any given integer $k = 1, 2, \ldots, m$, the map $E_k : \mathbb{R}^m \to \mathbb{R} : u \mapsto \sum_{i_1 < \cdots < i_k} \prod_{l=1}^{k} u_{i_l}$ is called the k^{th} *elementary symmetric function* on $\mathbb{R}^m$. We also set $E_0 := 1$.

Fact 2.13 For every $x \in X$ and all $t \in \mathbb{R}$,

$$p(x + td) = p(d) \prod_{i=1}^{m} (t + \lambda_i(x)) = p(d) \sum_{i=0}^{m} E_i(\lambda(x)) t^{m-i}$$

and for every $0 \le i \le m$,

$$p(d) E_i(\lambda(x)) = \frac{1}{(m-i)!} \nabla^{m-i} p(x) [\underbrace{d, d, \ldots, d}_{m-i \text{ times}}].$$

If $1 \le i \le m$, then $E_i \circ \lambda$ is hyperbolic with respect to d of degree i.

Proof. The first displayed equation is elementary while the second displayed equation is a consequence of Taylor's Theorem. The "If" part follows by employing Proposition 2.4 repeatedly. ■

Fact 2.13 gives a very transparent proof of the linearity of trace: $\sigma_m = E_1 \circ \lambda$ is a homogeneous (hyperbolic) polynomial of degree 1 and hence linear.

Also, the elementary symmetric functions themselves are hyperbolic:

Example 2.14 Let $X = \mathbb{R}^m$ and $d = (1, 1, \ldots, 1) \in \mathbb{R}^m$. Then for every $1 \le k \le m$, the k^{th} elementary symmetric function E_k is hyperbolic of degree k with respect to d.

Proof. Let $p := E_m$. It is straightforward to check that E_m is hyperbolic of degree m with respect to d with corresponding eigenvalue map $\lambda(x) = x_\downarrow$. Since each E_k is symmetric, the result now follows from Fact 2.13. ■

An inequality in elementary symmetric functions

The following inequality was discovered independently by McLeod [26] and by Bullen and Marcus [4, Theorem 3].

Fact 2.15 (McLeod, 1959; Bullen and Marcus, 1961) Suppose $1 \le k \le l \le m$ and $u, v \in \mathbb{R}^m_{++}$. Set $q := (E_l/E_{l-k})^{1/k}$. Then

$$q(u+v) > q(u) + q(v),$$

unless u and v are proportional or $k = l = 1$, in which case we have equality.

Bullen and Marcus's proof relies on an inequality by Marcus and Lopes ([23, Theorem 1], which is the case $k = 1$ in Fact 2.15. (Proofs can also be found in [2, Theorem 1.16], [5, Section V.4], and [28, Section VI.5].)
We record two interesting consequences of Fact 2.15.

Corollary 2.16 (Marcus and Lopes's [23, Theorem 2]) The function $-E_m^{1/m}$ is sublinear on $\mathbb{R}^m_+$, and it vanishes on $\operatorname{bd}\mathbb{R}^m_+$.

Proof. Set $k = l = m$ in Fact 2.15 and use continuity. ■

Recall that a function h is called *logarithmically convex*, if $\log \circ h$ is convex. The function q in Fact 2.15 is concave ("strictly modulo rays"), which yields logarithmic and strict convexity of $1/q$:

Proposition 2.17 Suppose q is a function defined on $\mathbb{R}^m_{++}$. Consider the following properties:
(i) the range of q is contained in $(0, +\infty)$;
(ii) $q(ru) = rq(u)$, for all $r > 0$ and every $u \in \mathbb{R}^m_{++}$;
(iii) $q(u+v) \ge q(u) + q(v)$, for all $u, v \in \mathbb{R}^m_{++}$;
(iv) if $u, v \in \mathbb{R}^m_{++}$ with $q(u+v) = q(u) + q(v)$, then $v = \rho u$, for some $\rho > 0$.
Suppose q satisfies (i)–(iii). Then $1/q$ is logarithmically convex. If furthermore (iv) holds, then $1/q$ is strictly convex.

Proof. The proof is straight forward. See [1] for details. ■

Corollary 2.18 Suppose $1 \le k \le l \le m$. Then the function $(E_{l-k}/E_l)^{1/k}$ is symmetric, positively homogeneous, and logarithmically convex. Moreover, the function is strictly convex on $\mathbb{R}^m_{++}$ unless $l = 1$ and $m \ge 2$.

Proof. Positive homogeneity and symmetry are clear. Log convexity follows by combining Proposition 2.17 and Fact 2.15; this even yields strict convexity unless $k = l = 1$. But if $k = l = 1$, then the function becomes $1/\sum_{i=1}^m u_i$, which is strictly convex exactly when $m = 1$. ■

3 Convexity

Sublinearity of the sum of the largest eigenvalues

Theorem 3.1 Suppose q is a homogeneous symmetric polynomial of degree n on $\mathbb{R}^m$, hyperbolic with respect to $e := (1, 1, \dots, 1) \in \mathbb{R}^m$, with eigenvalue map μ. Then

$$q \circ \lambda$$

is a hyperbolic polynomial of degree n with respect to d and its eigenvalue map is $\mu \circ \lambda$.

Proof. For simplicity, write $\tilde{p}$ for $q \circ \lambda$.
Step 1: $\tilde{p}$ is a polynomial on X.
Since $q(y)$ is a symmetric polynomial on $\mathbb{R}^m$, it is (by, e.g., [16, Proposition V.2.20.(ii)]) a polynomial in $E_1(y), \dots, E_m(y)$. On the other hand, by Fact 2.13, $E_i \circ \lambda$ is hyperbolic with respect to d of degree i, for $1 \leq i \leq m$. Altogether, $\tilde{p}(x) = q(\lambda(x))$ is a polynomial on X.
Step 2: $\tilde{p}$ is homogeneous of degree n.
Since q is symmetric and homogeneous, and in view of Fact 2.5, we obtain $\tilde{p}(tx) = q(\lambda(tx)) = t^n \tilde{p}(x)$, for all $t \in \mathbb{R}$ and every $x \in X$.
Step 3: $\tilde{p}(d) \neq 0$.
Again using Fact 2.5, we have $\tilde{p}(d) = q(\lambda(d)) = q(e) \neq 0$.
Step 4: $\tilde{p}$ is hyperbolic with respect to d.
Using once more Fact 2.5, we write for every $x \in X$ and all $t \in \mathbb{R}$:

$$\tilde{p}(x + td) = q(\lambda(x + td)) = q(\lambda(x) + te) = q(e) \prod_{k=1}^{n} (t + \mu_k(\lambda(x))). \quad \blacksquare$$

The next example is easy to check.

Example 3.2 Fix $1 \leq k \leq m$, set $e := (1, 1, \dots, 1) \in \mathbb{R}^m$, and let

$$q(u) := \prod_{1 \leq i_1 < i_2 < \cdots < i_k \leq m} \sum_{l=1}^{k} u_{i_l}.$$

Then q is a homogeneous symmetric polynomial on $\mathbb{R}^m$ of degree $\binom{m}{k}$, hyperbolic with respect to e, and its eigenvalues are $\{\frac{1}{k} \sum_{l=1}^{k} u_{i_l} : 1 \leq i_1 < i_2 < \cdots < i_k \leq m\}$. In particular, the largest eigenvalue of q is the arithmetic mean of the k largest components of u.

We now present our main result, the generalization of Theorem 2.6: the sum of the largest eigenvalues is sublinear. This readily implies Lipschitzness of the eigenvalue map.

Corollary 3.3 For every $1 \le k \le m$, the function σ_k is sublinear and λ_k is (globally) Lipschitz.

Proof. Fix $1 \le k \le m$, define q as in Example 3.2, and consider $\tilde{p} := q \circ \lambda$. By Theorem 3.1 and Example 3.2, the largest eigenvalue of $\tilde{p}$ is equal to $\frac{1}{k}\sigma_k(x)$. Now Theorem 2.6 yields the sublinearity of σ_k. Finally, recall that every sublinear finite function is globally Lipschitz (this follows from [30, Theorem 13.2, Corollary 13.2.2, and Corollary 13.3.3]); in particular, so is each σ_i. Thus λ_1 is Lipschitz. If $k \ge 2$, then $\lambda_k = \sigma_k - \sigma_{k-1}$ is — as the difference of two Lipschitz functions — Lipschitz as well. ■

The Hermitian matrices (continued). Here it is well known that the sum of the k largest eigenvalues is a convex function and that the k^{th} largest eigenvalue map is Lipschitz; see, for instance, [12].

Corollary 3.4 The function $w^T\lambda(\cdot)$ is sublinear, for every $w \in \mathbb{R}^m_{\downarrow}$.

Proof. Write $w^T\lambda = \sum_{i=1}^m w_i\lambda_i = w_m\sigma_m + \sum_{i=1}^{m-1}(w_i - w_{i+1})\sigma_i$ and then apply Corollary 3.3. ■

Note that we can rewrite Corollary 3.4 quite artificially as $w^T(\lambda(x+y) - \lambda(x)) \le w_{\downarrow}^T\lambda(y)$, for all $x, y \in X$ and $w \in \mathbb{R}^m_{\downarrow}$.
It would be interesting to find out about the following generalization:

Open Problem 3.5 (Lidskii's theorem) Decide whether or not

$$w^T(\lambda(x+y) - \lambda(x)) \le w_{\downarrow}^T\lambda(y), \quad \text{for all } x, y \in X \text{ and } w \in \mathbb{R}^m .$$

The condition means that the vector $\lambda(y)$ "majorizes" the vector $\lambda(x+y) - \lambda(x)$, for all $x, y \in X$; see [25, Proposition 4.B.8]. (The interested reader is referred to [25] for further information on majorization.) If this condition is satisfied, then we will simply say that "Lidskii's theorem holds".

The Hermitian matrices (continued). Lidskii's theorem does hold for the Hermitians. A recent and very complete reference is Bhatia's [3]; see also [21] for a new proof rooted in nonsmooth analysis.

In Section 6, we point out that Lidskii's theorem holds for further examples as well.

Convexity of composition

Fact 3.6 Suppose $f : \mathbb{R}^m \to [-\infty, +\infty]$ is convex and symmetric. Suppose further $u, v \in \mathbb{R}^m_\downarrow$ and $u - v \in (\mathbb{R}^m_\downarrow)^\oplus$. Then $f(u) \geq f(v)$. Moreover: if f is strictly convex on $\operatorname{conv}\{u_{\pi(i)} : \pi \text{ is a permutation of } \{1, \ldots, m\}\}$ and $u \neq v$, then $f(u) > f(v)$.

Proof. Imitate the proof of [19, Theorem 3.3] and consider [19, Example 7.1]. See also [25, 3.C.2.c on page 68]. ■

Theorem 3.7 (convexity) Suppose $x, y \in X$, $\alpha \in (0, 1)$, and $f : \mathbb{R}^m \to [-\infty, +\infty]$ is convex and symmetric. Then

$$f(\lambda(\alpha x + (1-\alpha)y)) \leq f(\alpha\lambda(x) + (1-\alpha)\lambda(y))$$

and hence $f \circ \lambda$ is convex. If f is strictly convex and $\alpha\lambda(x) + (1-\alpha)\lambda(y) \neq \lambda(\alpha x + (1-\alpha)y)$, then $f(\lambda(\alpha x + (1-\alpha)y)) < f(\alpha\lambda(x) + (1-\alpha)\lambda(y))$.

Proof. (See also [19, Proof of Theorem 4.3].) Fix an arbitrary $w \in \mathbb{R}^m_\downarrow$. Set $u := \alpha\lambda(x) + (1-\alpha)\lambda(y)$ and $v := \lambda(\alpha x + (1-\alpha)y)$. Then both u and v belong to $\mathbb{R}^m_\downarrow$. By Corollary 3.4, $w^T\lambda$ is convex on X. Therefore, $w^T\lambda(\alpha x + (1-\alpha)y) \leq \alpha w^T\lambda(x) + (1-\alpha)w^T\lambda(y)$; equivalently, $w^T(u - v) \geq 0$. It follows that $u - v \in (\mathbb{R}^m_\downarrow)^\oplus$. By Fact 3.6, $f(u) \geq f(v)$, which is the second displayed statement. The convexity of $f \circ \lambda$ follows. Finally, the "If" part is implied by the above and the "Moreover" part of Fact 3.6. ■

The Hermitian matrices (continued). In this case, the convexity of the composition is attributed to *Davis* [6]; see also [18, Corollary 2.7].

Another consequence is Gårding's inequality; see [10, Lemma 3.1].

Corollary 3.8 (Gårding's inequality) Suppose $p(d) > 0$. Then function $x \mapsto -(p(x))^{1/m}$ is sublinear on the hyperbolicity cone $C(d)$, and it vanishes on its boundary.

Proof. By Corollary 2.16, the function $-E_m^{1/m}$ is sublinear and symmetric on $\mathbb{R}^m_+$. Hence, by Theorem 3.7, the function $x \mapsto -(E_m(\lambda(x))^{1/m}$ is sublinear on $\{x \in X : \lambda(x) \geq 0\} = \operatorname{cl} C(d)$. The result follows, since $p(x) = p(d)E_m(\lambda(x))$, for every $x \in X$. ■

The Hermitian matrices (continued). Corollary 3.8 implies the *Minkowski Determinant Theorem*: $\sqrt[m]{\det(x+y)} \geq \sqrt[m]{\det x} + \sqrt[m]{\det y}$, whenever $x, y \in X$ are positive semi-definite.

Corollary 3.9 Suppose $x, y \in X$. Then:

(i) $||\lambda(x+y)|| \leq ||\lambda(x) + \lambda(y)||$.

(ii) $||\lambda(x+y)||^2 - ||\lambda(x)||^2 - ||\lambda(y)||^2 \leq 2\langle \lambda(x), \lambda(y) \rangle$.

Moreover, equality holds in (i) or (ii) if and only if $\lambda(x+y) = \lambda(x) + \lambda(y)$.

Proof. (i): Let $w := \lambda(x+y) \in \mathbb{R}^m_\downarrow$. Then, using Corollary 3.4 and the Cauchy-Schwarz inequality in $\mathbb{R}^m$, we estimate

$$\begin{aligned} ||\lambda(x+y)||^2 &= w^T \lambda(x+y) \leq w^T(\lambda(x) + \lambda(y)) \\ &\leq ||w|| ||\lambda(x) + \lambda(y)|| = ||\lambda(x+y)|| ||\lambda(x) + \lambda(y)||. \end{aligned}$$

The inequality follows. The condition for equality follows from the condition for equality in the Cauchy-Schwarz inequality.
(ii): The condition is equivalent to (i). ■

4 Making X Euclidean

Definition 4.1 Define $|| \cdot || : X \to [0, +\infty) : x \mapsto ||\lambda(x)||$ and

$$\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R} : (x, y) \mapsto \tfrac{1}{4}||x+y||^2 - \tfrac{1}{4}||x-y||^2.$$

Theorem 4.2 Suppose p is complete. Then X equipped with $\langle \cdot, \cdot \rangle$ is a Euclidean space with induced norm $|| \cdot ||$.

Proof. We have $||x||^2 = ||\lambda(x)||^2 = \sum_{i=1}^m \lambda_i(x)^2 = (E_1(\lambda(x)))^2 - 2E_2(\lambda(x))$. Facts 2.5 and 2.13 imply that $|| \cdot ||^2$ is a homogeneous polynomial of degree 2 on X. Since $|| \cdot || \geq 0$ and p is complete, the result now follows from the Polarization Identity. ■

Remark 4.3 The norm $||\cdot||$ defined in Definition 4.1 is precisely the *Hessian norm* used in interior point methods and thus well-motivated. To see this, assume that p is complete and recall that the *hyperbolic barrier function* is defined by $F(x) := -\ln(p(x))$. The Hessian norm at x is then given by

$$||x||_d^2 := \nabla^2 F(d)[x, x].$$

For t positive and sufficiently small, we have $p(tx+d) = p(d)\prod_{i=1}^m(1+t\lambda_i(x))$ and hence (after taking logarithms)

$$F(d+tx) = F(d) - \sum_{i=1}^m \ln(1+t\lambda_i(x)).$$

Expand the left (resp. right) side of this equation into a Taylor (resp. log) series. Then compare coefficients of t^2 to conclude $\nabla^2 F(d)[x,x]/2! = \|\lambda(x)\|^2/2$. Thus $\|\cdot\|_d = \|\cdot\|$. Further information can be found in [10]; see, in particular, [10, equation 16].

Proposition 4.4 (sharpened Cauchy-Schwarz) Suppose p is complete. Then

$$\langle x, y\rangle \leq \langle \lambda(x), \lambda(y)\rangle \leq \|x\|\|y\|, \qquad \text{for all } x, y \in X.$$

Proof. By the Cauchy-Schwarz inequality in $\mathbb{R}^m$ and Corollary 3.9.(ii), $2\langle\lambda(x),\lambda(y)\rangle \geq \|\lambda(x+y)\|^2 - \|\lambda(x)\|^2 - \|\lambda(y)\|^2 = \|x+y\|^2 - \|x\|^2 - \|y\|^2 = 2\langle x, y\rangle$. ■

The Hermitian matrices (continued). The inner product on the Hermitian matrices is precisely what one would expect: $\langle x, y\rangle = \operatorname{trace}(xy)$. The sharpening of the Cauchy-Schwarz inequality is essentially due to *von Neumann*; see [18, Theorem 2.2] and the discussion therein.

We can now refine Theorem 3.7.

Theorem 4.5 (strict convexity) Suppose p is complete and $f : \mathbb{R}^m \to [-\infty, +\infty]$ is strictly convex and symmetric. Then the composition $f \circ \lambda$ is strictly convex on X.

Proof. Fix $\alpha \in (0,1)$, $x, y \in X$ and set $\beta := 1-\alpha$. Suppose that $(f\circ\lambda)(\alpha x + \beta y) = \alpha(f\circ\lambda)(x) + \beta(f\circ\lambda)(y)$. We have to show that $x = y$. Using Theorem 3.7 and convexity of f, we estimate

$$\begin{aligned}\alpha(f\circ\lambda)(x) + \beta(f\circ\lambda)(y) &= (f\circ\lambda)(\alpha x + \beta y)\\ &\leq f(\alpha\lambda(x) + \beta\lambda(y))\\ &\leq \alpha(f\circ\lambda)(x) + \beta(f\circ\lambda)(y);\end{aligned}$$

hence equality must hold throughout. By strict convexity of f, we conclude that $\lambda(x) = \lambda(y)$. We also know that $\alpha\lambda(x)+\beta\lambda(y) = \lambda(\alpha x+\beta y)$ (otherwise, Theorem 3.7 would imply that the first displayed inequality is strict, which is a contradiction). Thus $\lambda(x) = \lambda(y) = \alpha\lambda(x) + \beta\lambda(y) = \lambda(\alpha x + \beta y)$. Since λ is norm preserving, we obtain $\|x\| = \|y\| = \|\alpha x + \beta y\|$. But $\|\cdot\|$ is induced by an inner product, whence $\|\cdot\|^2$ is strictly convex. Therefore, $x = y$ and the proof is complete. ■

Theorem 4.5 can be used to recover a recent result by Krylov (see [17, Theorem 6.4.(ii)]). Our proof appears to be more transparent than Krylov's.

Corollary 4.6 Suppose $p(d) > 0$. Then each of the following functions is convex on the hyperbolicity cone $C(d)$:

$$-\ln p, \quad \ln \frac{E_{m-1} \circ \lambda}{E_m \circ \lambda}, \quad \frac{E_{m-1} \circ \lambda}{E_m \circ \lambda}.$$

If p is complete, then each of these functions is strictly convex.

Proof. Define first $f(u) := -\ln p(d) - \sum_{i=1}^{m} \ln u_i$ on $\mathbb{R}^m_{++}$ and $F(x) := -\ln p(x)$ on $C(d)$. Then f is strictly convex and symmetric. Since $p(x) = p(d)E_m(\lambda(x))$, we have $F = f \circ \lambda$. It follows that F is convex (by Theorem 3.7), even strictly if p is complete (by Theorem 4.5). This proves the result for the first function. Now let $f := \ln(E_{m-1}/E_m)$ on $\mathbb{R}^m_{++}$ and $F := \ln \frac{E_{m-1}\circ\lambda}{E_m\circ\lambda}$ on $C(d)$. Then f is strictly convex by Corollary 2.18. By Theorem 3.7 (resp. Theorem 4.5), F is convex (resp. strictly convex, if p is complete). This yields the statement for the second function. Finally observe that the third function is obtained by taking the exponential of the second function. But this operation preserves (strict) convexity. ■

Krylov's result is closely related to parts of Güler's very recent work on hyperbolic barrier functions. We now give a simple proof of Güler's [10, Theorem 6.1]. The functions F and g below play a crucial role in interior point method, as they allow the construction of long-step interior-point methods using the hyperbolic barrier function F.

Corollary 4.7 Suppose $p(d) > 0$ and c belongs to the hyperbolicity cone $C := C(d)$. Define

$$F : C \to \mathbb{R} : x \mapsto -\ln(p(x)) \quad \text{and} \quad g : C \to \mathbb{R} : x \mapsto -(\nabla F(x))(c).$$

Then F and g are convex on C. If p is complete, then both F and g are strictly convex.

Proof. The statement on F is already contained in Corollary 4.6. Now let μ be the eigenvalue map corresponding to c. Then, by Fact 2.13, $p(x) = p(c)E_m(\mu(x))$ and $(\nabla p(x))(c) = p(c)E_{m-1}(\mu(x))$. Thus

$$g(x) = \frac{1}{p(x)}\big(\nabla p(x)\big)(c) = \frac{E_{m-1}(\mu(x))}{E_m(\mu(x))}.$$

Now argue as for the second function in the proof of Corollary 4.6. ■

The Hermitian matrices (continued). The statement on F corresponds to strict convexity of the function $x \mapsto -\ln\det(x)$ on the cone of positive semi-definite Hermitian matrices; this result is due to *Fan* [7].

We already pointed out that the trace σ_m is linear. With the notation introduced in Definition 4.1, we can express this more elegantly.

Proposition 4.8 (trace) $\sigma_m(x) = \langle d, x\rangle$, for every $x \in X$.

Proof. Fix $x \in X$. By Fact 2.5, $\|x \pm d\|^2 = \sum_{i=1}^m (\lambda_i(x \pm d))^2 = \sum_{i=1}^m (\lambda_i(x) \pm 1)^2 = \|x\|^2 \pm 2\sigma_m(x) + m$. So $4\langle x, d\rangle = \|x+d\|^2 - \|x-d\|^2 = 4\sigma_m(x)$. ■

5 Convex calculus

Definition 5.1 (isometric hyperbolic polynomial) We say p is *isometric (with respect to d)*, if for all $y, z \in X$, there exists $x \in X$ such that

$$\lambda(x) = \lambda(z) \quad \text{and} \quad \lambda(x+y) = \lambda(x) + \lambda(y).$$

It is clear that if p is isometric, then $\operatorname{ran}\lambda$ is a closed convex cone contained in $\mathbb{R}^m_{\downarrow}$. However, the range of λ need not be convex in general:

Example 5.2 (a hyperbolic polynomial that is not isometric) If the polynomial $p(x) = x_1x_2x_3$ is defined on $X = \operatorname{span}\big\{(1,1,1),(3,1,0)\big\}$, then p is hyperbolic of degree $m = 3$ with respect to $d = (1,1,1)$. Hence $\lambda(x) = x_{\downarrow}$ and p is complete. It follows that for all $\alpha, \beta \in \mathbb{R}$,

$$\lambda\big(\alpha(1,1,1) + \beta(3,1,0)\big) = \begin{cases} \alpha(1,1,1) + \beta(3,1,0), & \text{if } \beta \geq 0; \\ \alpha(1,1,1) + \beta(0,1,3), & \text{otherwise.} \end{cases}$$

Since $\lambda(3,1,0) + \lambda(-3,-1,0) = (3,0,-3) \notin \operatorname{ran}\lambda$, the set $\operatorname{ran}\lambda$ is a closed *nonconvex* cone in $\mathbb{R}^3_{\downarrow}$. In particular, p is not isometric.

Unless stated otherwise, we assume from now on that

p is complete, with corresponding inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$.

We chose the name "isometric" because of the equivalent condition (iii) in the following proposition.

Proposition 5.3 The following are equivalent:

(i) p is isometric.

(ii) $\max_{x:\lambda(x)=u}\langle x,y\rangle = \langle u,\lambda(y)\rangle$, for all $u \in \operatorname{ran}\lambda$ and every $y \in X$.

(iii) $d(u,\lambda(y)) = d(\lambda^{-1}(u),y)$, for all $u \in \operatorname{ran}\lambda$ and every $y \in X$.

Proof. This can be checked directly; see also [1, Proposition 5.4]. ■

The Hermitian matrices (continued). Clearly, $\operatorname{ran}\lambda = \mathbb{R}^m_\downarrow$ in this case. Isometricity can be seen as follows: Fix two Hermitian matrices y and z, and denote the corresponding diagonal matrices built from $\lambda(y),\lambda(z)$ by $\Lambda(y),\Lambda(z)$, respectively. Diagonalize $y = u^*\Lambda(y)u$, where u is some unitary matrix. Then set $x := u^*\Lambda(z)u$. It is easy to verify that $\lambda(x) = \lambda(z)$ and $\lambda(x+y) = \lambda(x)+\lambda(y)$, hence $p = \det$ is indeed isometric.

Theorem 5.4 (Fenchel conjugacy) Suppose that $f : \mathbb{R}^m \to (-\infty,+\infty]$ is symmetric. Then $(f\circ\lambda)^* \leq f^*\circ\lambda$. If p is isometric and $f(P_{\operatorname{ran}\lambda}u) \leq f(u)$, for every $u \in (\operatorname{dom} f)_\downarrow$, then $(f\circ\lambda)^* = f^*\circ\lambda$.

Proof. Fix an arbitrary $y \in X$. Then, using Proposition 4.4, symmetry of f, and the Hardy-Littlewood-Pólya inequality (see [11, Section 10.2]), the inequality follows from

$$\begin{aligned} f^*(\lambda(y)) &= \sup_{u\in\mathbb{R}^m}\big\{\langle u,\lambda(y)\rangle - f(u)\big\} = \sup_{u\in\mathbb{R}^m_\downarrow}\big\{\langle u,\lambda(y)\rangle - f(u)\big\} \\ &\geq \sup_{u\in\operatorname{ran}\lambda}\ \max_{x:\lambda(x)=u}\big\{\langle x,y\rangle - f(\lambda(x))\big\} = \sup_{x\in X}\big\{\langle x,y\rangle - (f\circ\lambda)(x)\big\} \\ &= (f\circ\lambda)^*(y). \end{aligned}$$

Now assume that p is isometric and $f(P_{\operatorname{ran}\lambda}u) \leq f(u)$, for every $u \in (\operatorname{dom} f)_\downarrow$. Fix momentarily an arbitrary $u \in \mathbb{R}^m_\downarrow$. Then, on the one hand, $f(P_{\operatorname{ran}\lambda}u) \leq$

$f(u)$ (if $u \in (\operatorname{dom} f)_\downarrow$, then the inequality follows by assumptions; otherwise, the inequality is trivial). Since $\operatorname{ran}\lambda$ is a closed convex cone that contains $\lambda(y) + P_{\operatorname{ran}\lambda} u$, a well-known property of projections yields on the other hand $\langle u - P_{\operatorname{ran}\lambda} u, \lambda(y)\rangle \leq 0$. Altogether, $\langle u, \lambda(y)\rangle - f(u) \leq \langle P_{\operatorname{ran}\lambda} u, \lambda(y)\rangle - f(P_{\operatorname{ran}\lambda} u)$. Therefore, using Proposition 5.3,

$$\begin{aligned} f^*(\lambda(y)) &= \sup_{u \in \mathbb{R}^m_\downarrow} \big\{\langle \lambda(y), u\rangle - f(u)\big\} \leq \sup_{u' \in \operatorname{ran}\lambda} \big\{\langle \lambda(y), u'\rangle - f(u')\big\} \\ &= \sup_{x \in X} \big\{\langle y, x\rangle - f(\lambda(x))\big\} = (f \circ \lambda)^*(y). \quad \blacksquare \end{aligned}$$

The assumption that $f(P_{\operatorname{ran}\lambda} u) \leq f(u)$, for every $u \in (\operatorname{dom} f)_\downarrow$ is important: in Example 6.1 below, we present an isometric hyperbolic polynomial and a convex symmetric function f with $(f \circ \lambda)^* \neq f^* \circ \lambda$.

Corollary 5.5 Suppose p is isometric and $f : \mathbb{R}^m \to (-\infty, +\infty]$ is symmetric. Then $(f \circ \lambda)^* = f^* \circ \lambda$ if any of the following conditions holds:

(i) $(\operatorname{dom} f)_\downarrow \subseteq \operatorname{ran}\lambda$.

(ii) $\operatorname{ran}\lambda = \mathbb{R}^m_\downarrow$.

(iii) f is convex and $P_{\operatorname{ran}\lambda} u \in \operatorname{conv}\{u_{\pi(i)} : \pi \text{ permutes } \{1, \ldots, m\}\}$, for every $u \in (\operatorname{dom} f)_\downarrow$.

Proof. (i) is clear from Theorem 5.4. (ii) is implied by (i). (iii): fix $u \in (\operatorname{dom} f)_\downarrow$ and write $P_{\operatorname{ran}\lambda} u = \sum_i \rho_i u^i$, where each ρ_i is nonnegative, $\sum_i \rho_i = 1$, and each u^i is some permutation of u. By convexity and symmetry of f, we conclude $f(P_{\operatorname{ran}\lambda} u) \leq f(u)$. Apply again Theorem 5.4. $\blacksquare$

Theorem 5.6 (subgradients) Suppose p is isometric, $\operatorname{ran}\lambda = \mathbb{R}^m_\downarrow$, and $f : \mathbb{R}^m \to (-\infty, +\infty]$ is convex and symmetric. Let $x, y \in X$. Then

$$y \in \partial(f \circ \lambda)(x) \text{ if and only if } \lambda(y) \in \partial f(\lambda(x)) \text{ and } \langle x, y\rangle = \langle \lambda(x), \lambda(y)\rangle.$$

Consequently, $\lambda\big[\partial(f \circ \lambda)(x)\big] = \partial f(\lambda(x))$.

Proof. Since $\operatorname{ran}\lambda = \mathbb{R}^m_\downarrow$, we have (Corollary 5.5.(ii)) $(f \circ \lambda)^* = f^* \circ \lambda$. In view of Proposition 4.4, the following equivalences hold true: $y \in \partial(f \circ \lambda)(x)$ $\Leftrightarrow (f \circ \lambda)(x) + (f \circ \lambda)^*(y) = \langle x, y\rangle \Leftrightarrow f(\lambda(x)) + f^*(\lambda(y)) = \langle \lambda(x), \lambda(y)\rangle$ and

$\langle x, y\rangle = \langle \lambda(x), \lambda(y)\rangle \Leftrightarrow \lambda(y) \in \partial f(\lambda(x))$ and $\langle x, y\rangle = \langle \lambda(x), \lambda(y)\rangle$. "Consequently": Clearly, by the above, $\lambda\big[\partial(f\circ\lambda)(x)\big] \subseteq \partial f(\lambda(x))$. Conversely, pick $v \in \partial f(\lambda(x))$. Then $f(\lambda(x)) + f^*(v) = \langle v, \lambda(x)\rangle$. By the assumption that $\operatorname{ran}\lambda = \mathbb{R}^m$ and Proposition 5.3.(ii), $\langle v, \lambda(x)\rangle = \langle y, x\rangle$, for some y with $\lambda(y) = v$. Hence $(f\circ\lambda)(x) + (f\circ\lambda)^*(y) = \langle y, x\rangle$ and so $y \in \partial(f\circ\lambda)(x)$, which implies $v = \lambda(y) \in \lambda\big[\partial(f\circ\lambda)(x)\big]$. ■

The Hermitian matrices (continued). Theorem 5.6 corresponds to [18, Theorem 3.2].

Corollary 5.7 (differentiability) Suppose p is isometric, $\operatorname{ran}\lambda = \mathbb{R}^m_\downarrow$, and $f : \mathbb{R}^m \to (-\infty, +\infty]$ is convex and symmetric. Let $x, y \in X$. Then $f\circ\lambda$ is differentiable at x and $y = \nabla(f\circ\lambda)(x)$ if and only if f is differentiable at $\lambda(x)$ and $\{y' \in X : \lambda(y') = \nabla f(\lambda(x)), \langle x, y'\rangle = \langle \lambda(x), \lambda(y')\rangle\} = \{y\}$.

Proof. Clear from Theorem 5.6. ■

Corollary 5.8 (variational description of σ_k) Let p be isometric, and suppose $\operatorname{ran}\lambda = \mathbb{R}^m_\downarrow$. Let $1 \le k \le m$. Then for every $x \in X$,

$$\sigma_k(x) = \max_{y:\lambda(y)\ge 0, \sigma_m(y)=k, \lambda_1(y)\le 1} \langle x, y\rangle$$

and $\partial\sigma_k(x) = \{y \in X : \langle x, y\rangle = \sigma_k(x), \lambda(y) \ge 0, \sigma_m(y) = k, \lambda_1(y) \le 1\}$.

Proof. Define $f(u) := \max_{i_1<\cdots<i_k} \sum_{l=1}^k u_{i_l}$. Then f is symmetric and convex on $\mathbb{R}^m$ and f^* is the indicator function of $\{u \in \mathbb{R}^m : \sum_i u_i = k \text{ and each } 0 \le u_i \le 1\}$. Now $\sigma_k = f\circ\lambda$ and so Corollary 5.5 yields $\sigma_k^* = f^*\circ\lambda$. Thus $y \in \partial\sigma_k(x) \Leftrightarrow x \in \partial\sigma_k^*(y) \Leftrightarrow \langle x, y\rangle = \sigma_k(x)$, $\lambda(y) \ge 0$, $\sigma_m(y) = k$, and $\lambda_1(y) \le 1$. ■

The Hermitian matrices (conclusion). Corollary 5.8 is a generalization of the variational formulations due to *Rayleigh* and *Ky Fan*; see also [12, Section 2].

6 Further examples

For complete details in these examples see the corresponding section in [1].

6.1 $\mathbb{R}^n$

Consider the vector space $X = \mathbb{R}^n$, the polynomial

$$p(x) = \prod_{i=1}^{n} x_i,$$

and the direction $d = (1, 1, ..., 1)$. Then p is hyperbolic and complete with eigenvalue map $\lambda(x) = x_\downarrow$. The induced norm and inner product in X are just the standard Euclidean ones in $\mathbb{R}^n$. We have ran$\lambda = \mathbb{R}^n_\downarrow$. It can be seen easily that p is isometric. In this case the sharpened Cauchy-Schwarz inequality (Proposition 4.4) reduces to the well-known Hardy-Littlewood-Pólya inequality (see [11, Chapter X])

$$x^T y \le x_\downarrow^T y_\downarrow.$$

Equality holds if and only if the vectors x and y can be simultaneously ordered with the same permutation. Since ran$\lambda = \mathbb{R}^n_\downarrow$, Corollary 5.5 shows that for every symmetric function $f : \mathbb{R}^n \to (-\infty, +\infty]$ we have $(f \circ \lambda)^* = f^* \circ \lambda$. Also Lidskii's Theorem holds, because $\lambda(x)$ is the ordered set of eigenvalues of the symmetric matrix Diag(x) (see [3, page 69]).

6.2 Singular values

Consider the vector space $M_{n,m}$ (of n by m real matrices). We assume $m \le n$ and denote the singular values of a matrix x in $M_{n,m}$ by $\sigma_1(x) \ge \sigma_2(x) \ge ... \ge \sigma_m(x)$. The Frobenius norm [14, page 291 & page 421] is defined by $||x||_F = ||\sigma(x)||$, where the last norm is the standard Euclidean norm in $\mathbb{R}^n$, and $\sigma(x) = (\sigma_1(x), \sigma_2(x), ..., \sigma_m(x))$. Now consider the vector space $X = M_{n,m} \times \mathbb{R}$, the polynomial

$$p(x, \alpha) = \det(\alpha^2 I_m - x^T x) \quad (x \in M_{n,m}, \alpha \in \mathbb{R}),$$

and the direction $d = (0, 1)$. Then p is hyperbolic and complete with eigenvalue map

$$\lambda(x, \alpha) = (\alpha + \sigma_1(x), \alpha + \sigma_2(x), ..., \alpha - \sigma_2(x), \alpha - \sigma_1(x)).$$

The induced norm and inner product are given by

$$||(x, \alpha)||^2 = 2m\alpha^2 + 2||x||_F^2,$$
$$\langle (x, \alpha), (x, \beta) \rangle = 2m\alpha\beta + 2\mathrm{tr}\, x^T y,$$

for (x, α) and (y, β) in X. With the help of the Singular Value Decomposition Theorem [14, Theorem 7.3.5] one can see that p is isometric. Notice that in this case the sharpened Cauchy-Schwarz inequality (Proposition 4.4) reduces to

$$\operatorname{tr} x^T y \le \sigma(x)^T \sigma(y).$$

Equality holds if and only if x and y have a simultaneous 'ordered' singular value decomposition (that is, there are unitary matrices u and v such that $x = u(\operatorname{Diag}\sigma(x))v$ and $y = u(\operatorname{Diag}\sigma(y))v$). This is the classical result known as 'von Neumann's Lemma' (see for example [15, page 182]). For a proof using results from this paper see [1]. We note that Lidskii's theorem holds too.
Note that when $m = 1$ we get the Lorentz Cone example which is discussed below. An analogous example can be obtained by considering the vector space $X = \mathbb{C}_{n,m} \times \mathbb{R}$.
We now show that for some functions in the singular value case we have $(f \circ \lambda)^* \ne f^* \circ \lambda$.

Example 6.1 Consider the symmetric function

$$f(u) = \max_{1 \le i \le m} u_i.$$

Its Fenchel conjugate is

$$f^*(v) = \begin{cases} 0, & \text{if } \sum_{i=1}^m v_i = 1,\ v_i \ge 0; \\ +\infty, & \text{else.} \end{cases}$$

If $m = 2$ and $y \in X$ is such that $\lambda(y) = \frac{1}{4}(3, 1, 1, -1)$, a short calculation [1] shows that $0 = (f \circ \lambda)^*(y) \ne (f^* \circ \lambda)(y) = +\infty$.

Remark 6.2 In 1937, von Neumann [31] gave a famous characterization of unitarily invariant matrix norms (that is, norms f on $\mathbb{C}^{m\times n}$ satisfying $f(uxv) = f(x)$ for all unitary matrices u and v and matrices x in $\mathbb{C}^{m\times n}$). His result states that such norms are those functions of the form $g \circ \sigma$, where g is a norm on $\mathbb{R}^m$, invariant under sign changes and permutations of components. Proof of this can be found also in [14, Theorem 7.4.24]. In [1] we show this theorem can also be derived in our framework.

6.3 Absolute reordering

Consider the vector space $X = \mathbb{R}^n \times \mathbb{R}$ Let the polynomial be

$$p(x, \alpha) = \prod_{i=1}^{n} (\alpha^2 - x_i^2),$$

and the direction be $d = (0, 1)$. Then p is hyperbolic and complete with eigenvalue map

$$\lambda(x, \alpha) = (|x|_{\downarrow}, (-|x|)_{\downarrow}) + \alpha e,$$

where $|x| = (|x_1|, |x_2|, \ldots, |x_n|)$, and $e = (1, 1, \ldots, 1) \in \mathbb{R}^{2n}$. Direct verification of the definition shows that p is isometric and furthermore that Lidskii's theorem holds. Note that the similarities with the previous example are not accidental. It corresponds to the subspace $(\mathrm{Diag}\ \mathbb{R}^n) \times \mathbb{R}$ of $M_{n,m} \times \mathbb{R}$.

6.4 Lorentz cone

Let the vector space be $X = \mathbb{R}^n$, and the polynomial be

$$p(x) = x^T A x = x_1^2 - x_2^2 - \cdots - x_n^2,$$

where $A = \mathrm{Diag}(1, -1, -1, ..., -1) \in M_n$ ($n \times n$ real matrices). Let the direction be $d = (d_1, d_2, ..., d_n) \in X$ such that $d_1^2 > d_2^2 + \cdots + d_n^2$. Then p is hyperbolic and complete with eigenvalue map

$$\lambda(x) = \left(\frac{x^T A d + \sqrt{D(x)}}{p(d)}, \frac{x^T A d - \sqrt{D(x)}}{p(d)} \right),$$

where $D(x) = (x^T A d)^2 - p(x)p(d)$ is the discriminant of $p(x + td)$ considered as a quadratic polynomial in t. (The fact that $D(x) \geq 0$ for each x, and so that $p(x)$ is hyperbolic, is the well-known Aczel inequality, see [27, p.57].) The induced norm and inner product are given by

$$\|x\|^2 = 2\frac{2(x^T A d)^2 - p(x)p(d)}{p(d)^2}, \quad \text{and}$$

$$\langle x, y \rangle = \frac{4(x^T A d)(y^T A d) - 2(x^T A y)p(d)}{p(d)^2},$$

for x and y in X. It is a bit trickier to see that p is isometric. Notice that in this case the sharpened Cauchy-Schwarz inequality (Proposition 4.4) becomes

$$(x^T Ad)(y^T Ad) - (x^T Ay)p(d) \leq \sqrt{D(x)D(y)},$$

and [1] gives the necessary and sufficient condition for equality. Lidskii's Theorem holds as well.

6.5 The degree 2 case

Let the vector space be $X = \mathbb{R}^n$. We assume that $p(x)$ is homogeneous polynomial of degree two. Without loss of generality, we write

$$p(x) = x^T Ax,$$

where $A \in S^n$. Fix a direction d in X with $p(d) \neq 0$. Then $p(x)$ is hyperbolic with respect to d if and only if the matrix $(d^T Ad)^{-1}A$ has exactly one positive eigenvalue (see [8, page 958]). Furthermore, p is complete if and only if A is nonsingular. Such a p is always isometric, and Lidskii's Theorem holds.

6.6 Antisymmetric tensor powers

Consider the function $p(x) = \det x$ on the vector space of $n \times n$ real symmetric (or Hermitian) matrices, and let $q = E_k$ be the elementary symmetric function of order k and $p_k(x) = E_k \circ \lambda(x)$. We saw earlier that p_k is a hyperbolic polynomial with respect to the identity matrix I (see Fact 2.13). We have

$$p_k(x) = \sum_{\alpha=(i_1<i_2<\cdots<i_k)} \det x[\alpha|\alpha] = \operatorname{tr}\left(\wedge^k x\right),$$

where $x[\alpha|\alpha]$ is the principal submatrix obtained from x by keeping its rows and columns $i_1, \ldots, i_k$, and the second equality above can be regarded as the definition of the symbol $\operatorname{tr}\left(\wedge^k x\right)$. For the first equality above, see [24], and justification for the use of the symbol $\operatorname{tr}\left(\wedge^k x\right)$ can be found in [9]. Now, from Corollary 3.8 (Gårding's inequality) and from the fact that $p_k(x) = \operatorname{tr}\left(\wedge^k x\right)$ is a homogeneous hyperbolic polynomial, it follows immediately that

$$\operatorname{tr}\left(\wedge^k (x+y)\right)^{1/k} \geq \operatorname{tr}\left(\wedge^k x\right)^{1/k} + \operatorname{tr}\left(\wedge^k y\right)^{1/k},$$

when x, y are symmetric and positive definite. This is one of the main results in [23].

Acknowledgment

It is our pleasure to thank Lars Gårding and Lars Hörmander for making some very valuable comments on an earlier version of this paper.

References

[1] H.H. BAUSCHKE, A.S. LEWIS, O. GÜLER, and H.S. SENDOV. Hyperbolic Polynomials and Convex Analysis. Research Report CORR 98-29, Department of Combinatorics and Optimization, University of Waterloo, June 1998.

[2] E.F. BECKENBACH and R. BELLMAN. *Inequalities*. Springer-Verlag, 1961.

[3] R. BHATIA. *Matrix Analysis*. Springer-Verlag, 1997.

[4] P. BULLEN and M. MARCUS. Symmetric means and matrix inequalities. *Proceedings of the American Mathematical Society*, 12:285–290, 1961.

[5] P.S. BULLEN, D.S. MITRINOVIĆ, and P.M. VASIĆ. *Means and their inequalities*. Mathematics and its applications. East European Series. D. Reidel Publishing Company, Dordrecht, Holland, 1987.

[6] C. DAVIS. All convex invariant functions of hermitian matrices. *Archiv der Mathematik*, 8:276–278, 1957.

[7] Ky FAN. On a theorem of Weyl concerning eigenvalues of linear transformations II. *Proceedings of the National Academy of Sciences of the United States of America*, 36:31–35, 1950.

[8] L. GÅRDING. An inequality for hyperpolic polynomials. *Journal of Mathematics and Mechanics*, 8(6):957–965, 1959.

[9] W. H. GREUB. *Multilinear Algebra*. Springer-Verlag, 1967.

[10] O. GÜLER. Hyperbolic polynomials and interior point methods for convex programming. *Mathematics of Operations Research*, 22(2):350–377, 1997.

[11] G.H. HARDY, J.E. LITTLEWOOD, and G. PÓLYA. *Inequalities.* Cambridge University Press, second edition, 1952.

[12] J.-B. HIRIART-URRUTY and D. YE. Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numerische Mathematik*, 70:45–72, 1992.

[13] L. HÖRMANDER. *Notions of convexity*, volume 127 of *Progress in mathematics.* Birkhäuser, Boston, 1994.

[14] R.A. HORN and C.R. JOHNSON. *Matrix analysis.* Cambridge University Press, second edition, 1985.

[15] R.A. HORN and C.R. JOHNSON. *Topics in matrix analysis.* Cambridge University Press, first edition, 1991. Paperback edition with corrections, 1994.

[16] T.W. HUNGERFORD. *Algebra*, volume 73 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1974.

[17] N.V. KRYLOV. On the general notion of fully nonlinear second-order elliptic equations. *Transactions of the American Mathematical Society*, 347(3):857–895, March 1995.

[18] A.S. LEWIS. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6(1):164–177, February 1996.

[19] A.S. LEWIS. Group invariance and convex matrix analysis. *SIAM Journal on Matrix Analysis*, 17(4):927–949, 1996.

[20] A.S. LEWIS. Convex analysis on Cartan subspaces, April 1997. Preprint.

[21] A.S. LEWIS. Lidskii's theorem via nonsmooth analysis. Research Report CORR 98-11, Department of Combinatorics and Optimization, University of Waterloo, April 1998.

[22] A.S. LEWIS and M.L. OVERTON. Eigenvalue optimization. *Acta Numerica*, 5:149–190, 1996.

[23] M. MARCUS and L. LOPES. Inequalities for symmetric functions and Hermitian matrices. *Canadian Journal of Mathematics*, 9:305–312, 1957.

[24] M. MARCUS and H. MINC. The invariance of symmetric functions of singular values. *Pacific Journal of Mathematics*, 12:327–332, 1962.

[25] A.W. MARSHALL and I. OLKIN. *Inequalities: Theory of Majorization and Its Applications*, volume 143 of *Mathematics in Science and Engineering*. Academic Press, New York, 1979.

[26] J.B. McLEOD. On four inequalities in symmetric functions. *Proceedings of the Edinburgh Mathematical Society (Series II)*, 11:211–219, 1958-1959.

[27] D.S. MITRINOVIĆ. *Analytic inequalities*. Springer-Verlag, New York, 1970. Die Grundlehren der mathematischen Wisenschaften, Band 1965.

[28] D.S. MITRINOVIĆ, J.E. PEČARIĆ, and A.M. FINK. *Classical and New Inequalities in Analysis*, volume 61 of *Mathematics and Its Applications (East European Series)*. Kluwer, 1993.

[29] A.M. OSTROWSKI. *Solution of equations in Euclidean and Banach spaces*. Academic Press, New York, third edition, 1973.

[30] R.T. ROCKAFELLAR. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[31] J. von NEUMANN. Some matrix inequalities and metrization of matric-space. *Tomsk University Review*, 1:286–300, 1937. In: *Collected Works*, Pergamon, Oxford, 1962, Volume IV, 205-218.