

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

# Conic optimization with spectral functions on Euclidean Jordan algebras

Chris Coey, Lea Kapelevich

Operations Research Center, MIT, Cambridge, MA 02142, [coey@mit.edu](mailto:coey@mit.edu), [lkap@mit.edu](mailto:lkap@mit.edu)

Juan Pablo Vielma

Google Research and MIT Sloan School of Management, Cambridge, MA 02142, [jvielma@google.com](mailto:jvielma@google.com)

Spectral functions on Euclidean Jordan algebras arise frequently in convex optimization models. Despite the success of primal-dual conic interior point solvers, there has been little work on enabling direct support for *spectral cones*, i.e. proper nonsymmetric cones defined from epigraphs and perspectives of spectral functions. We propose simple logarithmically homogeneous barriers for spectral cones and we derive efficient, numerically stable procedures for evaluating barrier oracles such as inverse Hessian operators. For two useful classes of spectral cones - the *root-determinant cones* and the *matrix monotone derivative cones* - we show that the barriers are self-concordant, with nearly optimal parameters. We implement these cones and oracles in our open source solver Hypatia, and we write simple, natural formulations for four applied problems. Our computational benchmarks demonstrate that Hypatia often solves the natural formulations more efficiently than advanced solvers such as MOSEK 9 solve equivalent extended formulations written using only the cones these solvers support.

*Key words:* conic optimization, interior point methods, spectral functions, Euclidean Jordan algebras, logarithmically homogeneous self-concordant barriers

*MSC2000 subject classification:* 90C51

*OR/MS subject classification:* Primary: programming; nonlinear: convex; secondary: mathematics: functions

*History:*

**1. Introduction.** A conic problem minimizes a linear function over the intersection of an affine subspace with a Cartesian product of primitive proper cones (i.e. irreducible, closed, convex, pointed, and full-dimensional conic sets). Any convex problem can be represented in conic form. An advantage of using conic form is that under mild assumptions, a conic problem has a simple and easily checkable certificate of optimality, primal infeasibility, or dual infeasibility [35].

In convex optimization applications, we frequently encounter spectral functions on Euclidean Jordan algebras such as the real vectors and real symmetric or complex Hermitian matrices. In this context, a spectral function is a real-valued symmetric function of the (real) eigenvalues. Examples include the geometric mean (or root-determinant), the entropy (e.g. von Neumann entropy), and the trace of the inverse (e.g. the A-optimal design criterion). Indeed, many disciplined convex programming (DCP) functions are spectral functions [20, 19]. We define a *spectral cone* as a proper cone that is the epigraph (or hypograph) set of either a homogeneous spectral function or the perspective function of a nonhomogeneous spectral function. These cones allow simple, natural conic reformulations of a wide range of convex optimization problems. Despite this, to our knowledge

there has been little prior work on enabling direct support for various spectral cones in primal-dual conic solvers.

Many conic solvers (such as CSDP [4], CVXOPT [1], ECOS [12, 38], MOSEK [25], SDPA [44]) implement primal-dual interior point methods (PDIPMs). Conic PDIPMs require easily computable oracles for a logarithmically homogeneous self-concordant barrier (LHSCB) function for the proper cone of the conic problem. Complexity analysis of idealized PDIPMs shows that they converge to  $\varepsilon$  tolerance in  $\mathcal{O}(\sqrt{\nu} \log(1/\varepsilon))$  iterations, where  $\nu$  is the barrier parameter of the LHSCB.

Currently, most PDIPMs are specialized for the nonnegative, second order, and positive semidefinite (PSD) cones, which are cones of squares of Euclidean Jordan algebras. However, these symmetric cones limit modeling generality and often require the construction of large extended formulations (EFs) [9]. For many spectral cones (e.g. for the negative entropy function), equivalent EFs using only symmetric cones do not exist, and when they do, they can be impractically large. Nonsymmetric conic PDIPMs can handle a much broader class of cones [30, 27, 39]. In Coey et al. [8], we generalize the method by Skajaa and Ye [39] and enhance its practical performance, and we test our implementations in our open-source conic solver Hypatia [9].

A key feature of Hypatia is its generic cone interface, which allows specifying a proper cone  $\mathcal{K}$  by implementing a small list of oracles. Once  $\mathcal{K}$  is defined, both  $\mathcal{K}$  and its dual cone  $\mathcal{K}^*$  may be used in any combination with other recognized cones in Hypatia to construct conic models. The oracles to implement are: an initial interior point  $t \in \text{int}(\mathcal{K})$ , a feasibility test for the cone interior  $\text{int}(\mathcal{K})$  (and optionally for the dual cone interior  $\text{int}(\mathcal{K}^*)$ ), and several derivative oracles for an LHSCB for the cone. The LHSCB oracles needed for ideal performance are the gradient, the Hessian operator (i.e. the second order directional derivative applied once to a given direction), the inverse of the Hessian operator, and the third order directional derivative (applied twice to a given direction). Fast and numerically stable procedures for evaluating oracles are crucial for practical performance in conic PDIPM solvers such as Hypatia. Note that these oracles may be useful in other algorithms and solvers - for example, application of the inverse Hessian operator is typically a bottleneck in computing Newton directions.

Our first main contribution is to define simple logarithmically homogeneous barriers for spectral cones and derive efficient and numerically stable barrier oracle procedures. For example, for the case where the spectral function of the cone is separable, we show how to apply the inverse Hessian operator of the barrier function very cheaply using a closed-form formula, without the need to compute or factorize an explicit Hessian matrix (which can be expensive and prone to numerical issues). Similarly, for the negative log-determinant and root-determinant spectral cones, we derive highly-efficient specialized oracle procedures. The Jordan algebra abstraction allows us to derive oracles for a large class of cones in a unified way without needing to specialize on the cone of squares domain.

Our second main contribution is to show that for two important classes of spectral cones - the root-determinant cones and the matrix monotone derivative (MMD) cones - the barriers we propose are LHSCBs. Again, the Jordan algebra abstraction allows us to write one simple proof per class, rather than separate proofs for every possible cone of squares domain of interest. Our new LHSCBs have parameters that are only a small additive increment of one larger than the parameter of the standard LHSCB for the cone of squares domain of the cone, hence the parameters are near-optimal. MMD cones allow modeling epigraphs of a variety of useful separable spectral functions, in particular the trace (or sum) of the negative logarithm (i.e. the negative log-determinant), negative entropy, and certain power functions. Furthermore, the dual cones of the MMD cones allow modeling epigraphs of even more separable spectral functions, such as the trace of the inverse, exponential, and more power functions.

These two contributions enable efficient and numerically stable implementations of the MMD cone and the log-determinant and root-determinant cones in nonsymmetric conic PDIPM solvers

such as Hypatia. We define these cones through Hypatia’s cone interface. Our MMD cone implementation is parametrized by both a Jordan algebra domain and an MMD function, allowing the user to define new domains and MMD functions. An MMD function is easily specified by implementing a small set of oracles for its univariate form: the function itself, its first three derivatives, and its convex conjugate, as well as an interior point for the corresponding MMD cone. We predefine five common MMD functions and three typical Jordan algebra domains: the real vectors, real symmetric matrices, and complex Hermitian matrices.

We use these new spectral cones in Hypatia to formulate example problems from distribution estimation, experiment design, quantum information science, and polynomial optimization. The natural formulations (NFs) using these cones are simpler and smaller than equivalent EFs written in terms of the handful of standard cones recognized by either ECOS or MOSEK 9 (i.e. the common symmetric cones and the three-dimensional exponential and power cones). Our computational experiments demonstrate that, across a wide range of sizes and spectral functions, Hypatia can solve the NFs faster than Hypatia, MOSEK, or ECOS can solve the equivalent EFs. Furthermore, to illustrate the practical impact of our efficient oracle procedures, we show that our closed-form formula for the MMD cone inverse Hessian product is faster and more numerically reliable than a naive direct solve using an explicit Hessian factorization.

**1.1. Overview.** We describe relevant aspects of Euclidean Jordan algebras, cones of squares, and spectral decompositions in Section 2. In Section 3, we define spectral functions on Euclidean Jordan algebras and give expressions for gradients and second and third order directional derivatives of spectral functions. We also specialize these formulae for separable spectral functions and the log-determinant case. We use these results in later sections to derive expressions for barrier oracles.

In Section 4, we define spectral function cones (and their dual cones) from epigraphs of homogenized convex spectral functions on cones of squares. We propose simple logarithmically homogeneous barriers for these cones and describe the additional properties that must be satisfied by an LHSCB. We also define several barrier oracles needed by Hypatia’s PDIPM. Then in Section 5, we describe fast and numerically stable procedures for these oracles, using the derivative results from Section 3. We specialize the oracle procedures for cones defined from separable spectral functions and the log-determinant function.

In Section 6, we define the MMD cone and its dual cone, and we give useful examples of MMD functions. We show that for the MMD cone, our barrier function is an LHSCB with near-optimal parameter. In Section 7, we define the root-determinant cone and its dual cone, prove that our barrier is an LHSCB with near-optimal parameter, and derive efficient oracle procedures.

Finally, in Section 8, we describe four applied examples over the new root-determinant, log-determinant, and MMD cones and their dual cones. We perform computational testing to demonstrate the advantages of solving these NFs directly with Hypatia and to exemplify the impact of our efficient oracle procedures.

**1.2. Notation.** We often write equation references above relation symbols to indicate the use of earlier results. For a natural number  $d$ , we define the index set  $\llbracket d \rrbracket = \{1, 2, \dots, d\}$ . For a set  $\mathcal{C}$ ,  $\text{cl}(\mathcal{C})$  and  $\text{int}(\mathcal{C})$  denote the closure and interior.  $\mathbb{R}$  is the space of reals,  $\mathbb{R}_{\geq}$  is the cone of nonnegative reals,  $\mathbb{R}_{>} = \text{int}(\mathbb{R}_{\geq})$  is the positive reals, and  $\mathbb{R}^d$  is the  $d$ -dimensional real vectors.  $\mathbb{S}^d$  and  $\mathbb{H}^d$  are the real symmetric and complex Hermitian matrices with side dimension  $d$ , and  $\mathbb{S}_{\geq}^d$  and  $\mathbb{H}_{\geq}^d$  are the corresponding cones of positive semidefinite matrices.

The inner product of  $w$  and  $r$  is  $\langle w, r \rangle$ . The  $k$ th derivative of a function  $f$  evaluated at a point  $w$  is  $\nabla^k f(w)$ , which may be interpreted as an operator. For example, the second directional derivative of  $f$  at  $w$  in the direction  $r, r$  is  $\nabla^2 f(w)[r, r] = \langle \nabla^2 f(w)[r], r \rangle$ . Often we omit the point at which the derivative is evaluated if this is clear from context. We use subscripts for partial derivatives, for example  $\nabla_w f$ .

**2. Jordan algebras.** Jordan algebraic concepts provide a useful and straightforward abstraction for spectral functions, spectral cones, and our barrier results in later sections. We follow the notation of Faraut and Koranyi [13, Chapter 2] where possible.

An algebra over the real or complex numbers is a vector space  $V$  equipped with a bilinear product  $\circ : V \times V \rightarrow V$ . For  $w \in V$ ,  $w^2 := w \circ w$ . We refer to  $V$  as a *Jordan algebra* if for all  $w_a, w_b \in V$ :

$$w_a \circ w_b = w_b \circ w_a, \quad (1a)$$

$$w_a \circ (w_a^2 \circ w_b) = w_a^2 \circ (w_a \circ w_b). \quad (1b)$$

For example, for  $V = \mathbb{R}^d$ , we can define  $\circ$  as an elementwise multiplication, or for  $V = \mathbb{S}^d$  and  $V = \mathbb{H}^d$ , we can let  $w_a \circ w_b = \frac{1}{2}(w_a w_b + w_b w_a)$ .

Given  $w_a \in V$ , we define the linear map  $L(w_a) : V \rightarrow V$  satisfying:

$$L(w_a)w_b = w_a \circ w_b \quad \forall w_b \in V. \quad (2)$$

Given  $w \in V$ , we define the linear map  $P(w) : V \rightarrow V$  satisfying:

$$P(w) = 2L(w)^2 - L(w^2). \quad (3)$$

$P$  is called the *quadratic representation* of  $V$ . In general,  $P(w) \neq L(w)^2 \neq L(w^2)$  because  $\circ$  need not be associative. For example, for  $V = \mathbb{S}^d$ , we have  $L(w_a^2)w_b = \frac{1}{2}(w_a^2 w_b + w_b w_a^2)$ ,  $L(w_a)^2 w_b = \frac{1}{2}(L(w_a)^2 w_b + w_a w_b w_a)$ , and  $P(w_a)w_b = w_a w_b w_a$ .

For any positive integer  $k$ , we have [42, Corollary 2.3.9]:

$$P(w)^k = P(w^k). \quad (4)$$

It is standard to assume the existence of a multiplicative identity  $e$ . Note that  $P(e)w = w$  and  $P(w)e = w^2$ . A point  $w \in V$  is *invertible* if and only if  $L(w)$  is invertible, and the inverse of  $w$  is the element  $w^{-1} \in V$  such that  $w^{-1} = L(w)^{-1}e$  [13, Proposition II.2.2]. Equation (4) also holds for  $k = -1$  if  $w$  is invertible [13, Proposition II.3.1].

Henceforth we consider only the finite dimensional *Euclidean* Jordan algebras. A Jordan algebra  $V$  is Euclidean if  $\langle w_a \circ w_b, w_c \rangle = \langle w_b, w_a \circ w_c \rangle$  for all  $w_a, w_b, w_c \in V$ .

We call  $\mathcal{Q}$  a *cone of squares* on  $V$  if  $\mathcal{Q} = \{w \circ w : w \in V\}$ . The cone  $\mathcal{Q}$  is proper (closed, convex, pointed, and solid) because  $V$  is Euclidean (and therefore formally real); see Papp and Alizadeh [32, Theorem 3.3] and Faraut and Koranyi [13, Section III.1 and Proposition VIII.4.2]. In addition,  $\mathcal{Q}$  is *self-dual* and *homogeneous*; see Vieira [42, Proposition 2.5.8] and Faraut and Koranyi [13, Theorem III.2.1]. For example, for  $V = \mathbb{S}^d$ , the cone of squares is  $\mathcal{Q} = \mathbb{S}_{\geq}^d$ .

For convenience, we often write  $a \succeq b$  instead of  $a - b \in \mathcal{Q}$ , or  $a \succ b$  instead of  $a - b \in \text{int}(\mathcal{Q})$ , where  $\mathcal{Q}$  is clear from context. If  $w \succ 0$ , then  $w$  is invertible [13, Theorem III.2.1]. Furthermore  $w \succ 0$  implies that  $w^{1/2}$  is well-defined and invertible, and  $P(w^{1/2}) = P(w)^{1/2}$  Vieira [42, Proposition 2.5.11]. This also implies by Equation (4) (with  $k = -1$ ) that  $P(w^{-1/2}) = P(w)^{-1/2}$ .

**2.1. Spectral decomposition.** In a Euclidean Jordan algebra  $V$ , an *idempotent* is an element  $c \in V$  such that  $c^2 = c$ . Two idempotents  $c_1, c_2$  are *orthogonal* if  $c_1 \circ c_2 = 0$ . Let  $d$  be the rank of  $V$ .  $c_1, \dots, c_d$  is a *complete* system of orthogonal idempotents if  $c_1, \dots, c_d$  are all idempotents, pairwise orthogonal, and  $\sum_{i \in [d]} c_i = e$ . An idempotent is *primitive* if it is non-zero and cannot be written as the sum of two orthogonal non-zero idempotents. A *Jordan frame* is a complete system of orthogonal idempotents, where each idempotent is primitive. The number of elements in any Jordan frame is called the *rank* of  $V$ . For example, the rank of  $\mathbb{R}^d$ ,  $\mathbb{S}^d$ , or  $\mathbb{H}^d$  is  $d$ .

For any  $w \in V$ , there exist unique real numbers (not necessarily distinct)  $w_1, \dots, w_d$  and a unique Jordan frame  $c_1, \dots, c_d$  such that  $w$  has the *spectral decomposition* [13, Theorem III.1.2]:

$$w = \sum_{i \in \llbracket d \rrbracket} w_i c_i. \quad (5)$$

We call  $w_1, \dots, w_d$  the *eigenvalues* of  $w$ . The *determinant* is  $\det(w) = \prod_{i \in \llbracket d \rrbracket} w_i$  and the *trace* is  $\text{tr}(w) = \sum_{i \in \llbracket d \rrbracket} w_i$  [13, Section II.2, page 29]. For example, for  $V = \mathbb{R}^d$ , the Jordan frame is the standard unit vectors and  $w$  is its own vector of eigenvalues. For  $V = \mathbb{S}^d$ , we can think of the Jordan frame as the rank one PSD matrices from a full symmetric eigendecomposition.

Henceforth, we define the inner product on  $V$  as  $\langle w_a, w_b \rangle = \text{tr}(w_a \circ w_b)$ . Under this inner product,  $P(w)$  is self-adjoint [42, Page 27]. Thus for  $w \in \text{int}(\mathcal{Q})$  and  $r_1, r_2 \in V$ , we have:

$$\langle P(w)r_1, r_2 \rangle = \langle P(w^{1/2})r_1, P(w^{1/2})r_2 \rangle = \langle r_1, P(w)r_2 \rangle. \quad (6)$$

**2.2. Peirce decomposition.** We let  $c_1, \dots, c_d$  be a Jordan frame for  $V$ , and define for  $i, j \in \llbracket d \rrbracket$ :

$$V(c_i, \lambda) := \{w : c_i \circ w = \lambda w\}, \quad (7a)$$

$$V_{i,i} := V(c_i, 1) = \{t c_i : t \in \mathbb{R}\}, \quad (7b)$$

$$V_{i,j} := V(c_i, \frac{1}{2}) \cap V(c_j, \frac{1}{2}). \quad (7c)$$

$V$  has the direct sum decomposition  $V = \bigoplus_{i,j \in \llbracket d \rrbracket : i \leq j} V_{i,j}$  [13, Theorem IV.1.3]. For example, for  $V = \mathbb{S}^d$ , let  $E_{i,j}$  be a matrix of zeros except in the  $(i, j)$ th position, and let  $c_i = E_{i,i}$ ; then  $V_{i,i} = \{t E_{i,i} : t \in \mathbb{R}\}$  and  $V_{i,j} = \{t(E_{i,j} + E_{j,i}) : t \in \mathbb{R}\}$ .

The *Peirce decomposition* allows us to write any  $r \in V$  as:

$$r = \sum_{i,j \in \llbracket d \rrbracket : i \leq j} r_{i,j} = \sum_{i,j \in \llbracket d \rrbracket : i < j} r_{i,j} + \sum_{i \in \llbracket d \rrbracket} r_i c_i, \quad (8)$$

where  $r_i = \langle r, c_i \rangle \in \mathbb{R}$  and  $r_{i,j} \in V_{i,j}$  for all  $i, j \in \llbracket d \rrbracket$ . Each  $r_{i,j}$  is a projection of  $r$  onto  $V_{i,j}$ , where:

$$r_{i,i} = r_i c_i = P(c_i)r \quad \forall i \in \llbracket d \rrbracket, \quad (9a)$$

$$r_{i,j} = 4L(c_i)L(c_j)r = 4c_i \circ (c_j \circ r) \quad \forall i, j \in \llbracket d \rrbracket : j \neq i. \quad (9b)$$

Note that  $r_{i,j} = r_{j,i}$ , since  $L(c_i)$  and  $L(c_j)$  commute [13, Lemma IV.1.3]. For example, let  $c_1, \dots, c_d$  be a Jordan frame for  $V = \mathbb{S}^d$  and let  $r \in V$ ; then  $r_i = c_i r c_i$  and  $r_{i,j} = c_i r c_j + c_j r c_i$ , for  $i, j \in \llbracket d \rrbracket$ .

We list some facts relating to compositions of projection operators (see Faraut and Koranyi [13, Theorem IV.2.2] and Sun and Sun [40, page 430]):

$$L(c_i)L(c_j)L(c_k)L(c_l) = 0 \quad \forall i, j, k, l \in \llbracket d \rrbracket : i \neq j, k \neq l, (i, j) \neq (k, l), \quad (10a)$$

$$L(c_i)L(c_j)P(c_k) = P(c_k)L(c_i)L(c_j) = 0 \quad \forall i, j, k \in \llbracket d \rrbracket : i \neq j, \quad (10b)$$

$$(4L(c_i)L(c_j))^2 = 4L(c_i)L(c_j) \quad \forall i, j \in \llbracket d \rrbracket, \quad (10c)$$

$$P(c_i)^2 = P(c_i) \quad \forall i \in \llbracket d \rrbracket, \quad (10d)$$

$$\sum_{i,j \in \llbracket d \rrbracket : i < j} 4L(c_i)L(c_j) + \sum_{i \in \llbracket d \rrbracket} P(c_i) = L(e). \quad (10e)$$

Given  $\lambda_{i,j} \neq 0 \in \mathbb{R}$  for  $i, j \in \llbracket d \rrbracket$ , consider an operator  $\Lambda : V \rightarrow V$  of the form:

$$\Lambda := \sum_{i,j \in \llbracket d \rrbracket : i < j} 4\lambda_{i,j}L(c_i)L(c_j) + \sum_{i \in \llbracket d \rrbracket} \lambda_{i,i}P(c_i). \quad (11)$$

The inverse operator  $\Lambda^{-1}$  is given by:

$$\Lambda^{-1} = \sum_{i,j \in \llbracket d \rrbracket: i < j} 4\lambda_{i,j}^{-1}L(c_i)L(c_j) + \sum_{i \in \llbracket d \rrbracket} \lambda_{i,i}^{-1}P(c_i). \quad (12)$$

It can be verified using Equation (10) that for any  $r \in V$ ,  $\Lambda\Lambda^{-1}r = \Lambda^{-1}\Lambda r = r$ . For example, let  $w = \sum_{i \in \llbracket d \rrbracket} w_i c_i \in V$  be invertible and suppose that  $\lambda_{i,j} = w_i w_j$  for  $i, j \in \llbracket d \rrbracket$ ; then:

$$\Lambda = \sum_{i,j \in \llbracket d \rrbracket: i < j} 4w_i w_j L(c_i)L(c_j) + \sum_{i \in \llbracket d \rrbracket} w_i^2 P(c_i) = P(w), \quad (13a)$$

$$\Lambda^{-1} = \sum_{i,j \in \llbracket d \rrbracket: i < j} 4w_i^{-1}w_j^{-1}L(c_i)L(c_j) + \sum_{i \in \llbracket d \rrbracket} w_i^{-2}P(c_i) = P(w^{-1}). \quad (13b)$$

**3. Spectral functions and derivatives.** Let  $V$  be a Jordan algebra of rank  $d$ . A real-valued function  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  is *symmetric* if it is invariant to the order of its inputs. A symmetric function  $f$  composed with an eigenvalue map  $\lambda: V \rightarrow \mathbb{R}^d$  induces a *spectral function*  $\varphi: V \rightarrow \mathbb{R}$  such that  $\varphi(w) = f(\lambda(w))$ , where  $\lambda(w) = (w_1, \dots, w_d)$  is the eigenvalue vector of  $w$  [2, Definition 8]. Note that  $\varphi$  is convex if and only if  $f$  is convex [10].

In this section, we give expressions for (directional) derivatives of the spectral function  $\varphi$ . We express these derivatives at a point  $w \in V$  (satisfying certain assumptions as necessary below) with spectral decomposition Equation (5). The derivatives are taken in direction  $r \in V$  with Peirce decomposition Equation (8). The gradient is  $\nabla\varphi(w) \in V$  and the second and third order directional derivatives are  $\nabla^2\varphi(w)[r] \in V$  and  $\nabla^3\varphi(w)[r, r] \in V$ . These derivative expressions are needed for the barrier oracles we derive in Section 5 and for our LHSCB proofs in Sections 6 and 7.

We begin with the general nonseparable spectral function case in Section 3.1, before specializing for separable spectral functions in Section 3.2. The root-determinant function, from which we define the root-determinant cone in Section 7, is nonseparable. The class of matrix monotone derivative (MMD) trace functions that we use to define the MMD cone in Section 6 are separable. Finally we specialize for the important case of the negative log-determinant function in Section 3.3; as we discuss in Section 4.4, this is the standard LHSCB for a cone of squares of a Jordan algebra, and it appears as an additive component in the barrier functions we propose.

**3.1. The nonseparable case.** Let  $\nabla f$ ,  $\nabla^2 f$ , and  $\nabla^3 f$  denote the derivatives of  $f$  evaluated at  $\lambda(w)$ . We use subindices to denote particular components of these derivatives. According to Baes [2, Theorem 38] and Sun and Sun [40, Theorem 4.1], the gradient of  $\varphi$  at  $w$  is:

$$\nabla\varphi(w) = \sum_{i \in \llbracket d \rrbracket} (\nabla f)_i c_i. \quad (14)$$

Henceforth we assume the eigenvalues of  $w$  are all distinct for simplicity. The second order directional derivative of  $\varphi$  in direction  $r$  is [40, Theorem 4.2]:

$$\nabla^2\varphi(w)[r] = \sum_{i,j \in \llbracket d \rrbracket: i < j} \frac{(\nabla f)_i - (\nabla f)_j}{w_i - w_j} r_{i,j} + \sum_{i,j \in \llbracket d \rrbracket} (\nabla^2 f)_{i,j} r_i c_j. \quad (15)$$

Sun and Sun [40, Theorem 4.2] also generalize this expression to allow for non-distinct eigenvalues.

To derive an expression for the third order directional derivative  $\nabla^3\varphi(w)[r, r]$ , we let:

$$w(t) := w + tr = \sum_{i \in \llbracket d \rrbracket} w_i(t) c_i(t), \quad (16)$$

where  $w_i(t)$  is the  $i$ th eigenvalue of  $w(t)$ . Note that  $\nabla^2\varphi(w)[r] = \frac{d}{dt}\nabla\varphi(w(t))|_{t=0}$  and  $\nabla^3\varphi(w)[r, r] = \frac{d^2}{dt^2}\nabla\varphi(w(t))|_{t=0}$ . We let  $\nabla f(t)$ ,  $\nabla^2 f(t)$ , and  $\nabla^3 f(t)$  denote the derivatives of  $f$  evaluated at  $\lambda(w(t))$ . Due to the chain rule and Equation (15):

$$\frac{d}{dt}\nabla\varphi(w(t)) = \nabla^2\varphi(w(t))[r] \tag{17a}$$

$$= \sum_{i,j \in \llbracket d \rrbracket : i < j} \frac{(\nabla f(t))_i - (\nabla f(t))_j}{w_i(t) - w_j(t)} r_{i,j}(t) + \sum_{i,j \in \llbracket d \rrbracket} (\nabla^2 f(t))_{i,j} r_i(t) c_j(t). \tag{17b}$$

We differentiate Equation (17) once more. From Vieira [43, Corollary 1 and Theorem 3.3] and Sun and Sun [40, Equation 37], for  $i \in \llbracket d \rrbracket$  we have:

$$\frac{d}{dt}w_i(t) = r_i(t), \tag{18a}$$

$$\frac{d}{dt}c_i(t) = s_i(t) := \sum_{j \in \llbracket d \rrbracket : j \neq i} \frac{r_{i,j}(t)}{w_i(t) - w_j(t)}. \tag{18b}$$

Using the chain rule, Equation (18a) implies:

$$\frac{d}{dt}(\nabla f(t))_i = \sum_{k \in \llbracket d \rrbracket} (\nabla^2 f(t))_{i,k} r_k(t) \quad \forall i \in \llbracket d \rrbracket, \tag{19a}$$

$$\frac{d}{dt}(\nabla^2 f(t))_{i,j} = \sum_{k \in \llbracket d \rrbracket} (\nabla^3 f(t))_{i,j,k} r_k(t) \quad \forall i, j \in \llbracket d \rrbracket. \tag{19b}$$

Applying the chain and product rules, we have:

$$\frac{d}{dt} \frac{1}{w_i(t) - w_j(t)} = \frac{r_j(t) - r_i(t)}{(w_i(t) - w_j(t))^2} \quad \forall i, j \in \llbracket d \rrbracket : i \neq j, \tag{20a}$$

$$\frac{d}{dt}r_{i,j}(t) = \frac{d}{dt}(4c_i(t) \circ (c_j(t) \circ r)) \tag{20b}$$

$$= 4c_i(t) \circ (s_j(t) \circ r) + 4s_i(t) \circ (c_j(t) \circ r) \quad \forall i, j \in \llbracket d \rrbracket : i \neq j, \tag{20c}$$

$$\frac{d}{dt}\langle c_i(t), r \rangle c_j(t) = \langle s_i(t), r \rangle c_j(t) + r_i(t) s_j(t) \quad \forall i, j \in \llbracket d \rrbracket. \tag{20d}$$

Finally, letting  $s_i := s_i(0)$  for all  $i \in \llbracket d \rrbracket$ , these results imply that:

$$\nabla^3\varphi[r, r] = \frac{d^2}{dt^2}\nabla\varphi(w(t))|_{t=0} \tag{21a}$$

$$= \sum_{i,j \in \llbracket d \rrbracket : i < j} \frac{(\nabla f)_i - (\nabla f)_j}{w_i - w_j} \left( 4c_i \circ (s_j \circ r) + 4s_i \circ (c_j \circ r) - \frac{r_i - r_j}{w_i - w_j} r_{i,j} \right) + \sum_{i,j,k \in \llbracket d \rrbracket : i < j} \frac{(\nabla^2 f)_{i,k} - (\nabla^2 f)_{j,k}}{w_i - w_j} r_k r_{i,j} + \sum_{i,j \in \llbracket d \rrbracket} (\nabla^2 f)_{i,j} (\langle s_i, r \rangle c_j + r_i s_j) + \sum_{i,j,k \in \llbracket d \rrbracket} (\nabla^3 f)_{i,j,k} r_i r_k c_j \tag{21b}$$

$$= \sum_{i,j \in \llbracket d \rrbracket : i < j} \frac{(\nabla f)_i - (\nabla f)_j}{w_i - w_j} \left( 4c_i \circ (s_j \circ r) + 4s_i \circ (c_j \circ r) - \frac{r_i - r_j}{w_i - w_j} r_{i,j} \right) + \sum_{i,j \in \llbracket d \rrbracket} (\nabla^2 f)_{i,j} (2r_j s_i + \langle s_i, r \rangle c_j) + \sum_{i,j,k \in \llbracket d \rrbracket} (\nabla^3 f)_{i,j,k} r_i r_k c_j. \tag{21c}$$

The derivative expressions simplify significantly for  $V = \mathbb{R}^d$ . For  $V = \mathbb{S}^d$ , the form of Equation (15) is well-known [16] and the form of Equation (21c) appears in Sendov [37].

**3.2. The separable case.** The spectral function  $\varphi$  induced by  $f$  is *separable* if  $f$  is a separable function, i.e.  $f(\lambda) = \sum_{i \in [d]} h(\lambda_i)$  for  $\lambda \in \mathbb{R}^d$  and some function  $h : \mathbb{R} \rightarrow \mathbb{R}$ . For convenience, if  $w \in V$ , we also define  $h : V \rightarrow V$  as  $h(w) := \sum_{i \in [d]} h(w_i)c_i$ . This allows us to write  $\varphi(w) = \text{tr}(h(w)) = \sum_{i \in [d]} h(\lambda_i)$ . Note that  $\varphi$  is convex if and only if  $h$  is convex. For example, if  $h(w) = -\log(w)$  then  $\varphi(w) = \text{tr}(-\log(w)) = -\log \det(w)$ ; we consider this special case in Section 3.3.

We specialize the derivatives from Equations (14), (15) and (21c), maintaining the simplifying assumption of distinct eigenvalues. Since  $(\nabla^2 f)_{i,j} = (\nabla^3 f)_{i,j,k} = 0$  unless  $i = j = k$ , we have:

$$\nabla \varphi(w) = \sum_{i \in [d]} \nabla h(w_i)c_i, \quad (22a)$$

$$\nabla^2 \varphi(w)[r] = \sum_{i,j \in [d]: i < j} \frac{\nabla h(w_i) - \nabla h(w_j)}{w_i - w_j} 4c_i \circ (c_j \circ r) + \sum_{i \in [d]} \nabla^2 h(w_i) P(c_i)r, \quad (22b)$$

$$\begin{aligned} \nabla^3 \varphi(w)[r, r] = & \sum_{i,j \in [d]: i < j} \frac{\nabla h(w_i) - \nabla h(w_j)}{w_i - w_j} \left( 4c_i \circ (s_j \circ r) + 4s_i \circ (c_j \circ r) - \frac{r_i - r_j}{w_i - w_j} r_{i,j} \right) + \\ & \sum_{i \in [d]} \nabla^2 h(w_i) (2r_i s_i + \langle s_i, r \rangle c_i) + \sum_{i \in [d]} \nabla^3 h(w_i) r_i^2 c_i. \end{aligned} \quad (22c)$$

**3.3. The negative log-determinant case.** The negative log-determinant function  $\varphi(w) = -\log \det(w)$  is a separable spectral function. We let  $w \succ 0$  and drop the simplifying assumption of distinct eigenvalues. For convenience, we let  $\hat{r} := P(w^{-1/2})r \in V$ . First, note that (similar to Vieira [42, Lemma 3.3.4]):

$$\langle w^{-1}, r \rangle = \langle P(w^{-1/2})e, r \rangle \stackrel{6}{=} \langle e, P(w^{-1/2})r \rangle = \text{tr}(\hat{r}). \quad (23)$$

Due to Faraut and Koranyi [13, Proposition II.2.3]:

$$\nabla_w (\text{tr}(\hat{r})) = \nabla_w (w^{-1})[r] = -P(w^{-1})r. \quad (24)$$

Adapting the result in Faybusovich and Tsuchiya [15, Lemma 3.4]:

$$\nabla_w (P(w^{-1})r)[r] = -2P(w^{-1/2})\hat{r}^2. \quad (25)$$

Now, the gradient of  $\varphi$  is [13, Propositions III.4.2(ii)]:

$$\nabla \varphi(w) = -w^{-1}, \quad (26)$$

so the second and third order directional derivatives are:

$$\nabla^2 \varphi(w)[r] \stackrel{24}{=} P(w^{-1})r, \quad (27a)$$

$$\nabla^3 \varphi(w)[r, r] \stackrel{25}{=} -2P(w^{-1/2})(P(w^{-1/2})r)^2. \quad (27b)$$

Note that unlike the separable spectral function case in Section 3.2, here we do not need the explicit eigenvalues of  $w$ .

**4. Cones and barrier functions.** In this paper we are concerned with a class of proper cones that can be characterized as follows:

$$\mathcal{K} := \text{cl}\{\tilde{u} \in \mathcal{E} : \zeta(\tilde{u}) \geq 0\} \subset \tilde{V}, \quad (28)$$

where  $\zeta : \mathcal{E} \rightarrow \mathbb{R}$  is a concave, (degree one) homogeneous function and  $\mathcal{E}$  is some convex cone in the space  $\tilde{V}$ . In particular, we define  $\zeta$  in terms of a  $C^3$ -smooth spectral function  $\varphi$  that is defined on the interior of a cone of squares  $\mathcal{Q}$  of a Jordan algebra  $V$  of rank  $d$ .

In Sections 4.1 and 4.2, we define two types of spectral function cones, and in Section 4.3 we define their dual cones. Then in Section 4.4, we propose simple logarithmically homogeneous barriers for these cones and describe the additional properties that must be satisfied by an LHSCB. We also define the barrier oracles needed for ideal performance of Hypatia's PDIPM.



**4.1. The homogeneous case.** First, we suppose that  $\varphi$  is convex and homogeneous. Then  $\zeta(u, w) := u - \varphi(w)$  is concave and homogeneous, and we let  $\mathcal{E} := \mathbb{R} \times \text{int}(\mathcal{Q})$  and  $\tilde{V} := \mathbb{R} \times V$ . This defines a convex cone that is the closure of the epigraph set of  $\varphi$ :

$$\mathcal{K}_h := \text{cl}\{(u, w) \in \mathbb{R} \times \text{int}(\mathcal{Q}) : u \geq \varphi(w)\}. \quad (29)$$

Note that if  $\varphi$  is concave, we can analogously define a cone from the hypograph set of  $\varphi$ . In Section 7, we consider the root-determinant cone, which is the hypograph of the concave root-determinant function. To check membership in  $\text{int}(\mathcal{K}_h)$ , we first determine whether  $w \in \text{int}(\mathcal{Q})$  (which is equivalent to positivity of the eigenvalues), and if so, whether  $\zeta(\tilde{u}) > 0$ .

**4.2. The non-homogeneous case.** Now we suppose that  $\varphi$  is convex and non-homogeneous. We define the perspective function of  $\varphi$ , per  $\varphi : \mathbb{R}_{>} \times \text{int}(\mathcal{Q}) \rightarrow \mathbb{R}$ , as  $(\text{per } \varphi)(v, w) := v\varphi(v^{-1}w)$ . This is a homogeneous and convex function [5, Section 3.2.6]. We let  $\zeta(u, v, w) := u - (\text{per } \varphi)(v, w)$ , with  $\mathcal{E} := \mathbb{R} \times \mathbb{R}_{>} \times \text{int}(\mathcal{Q})$  and  $\tilde{V} := \mathbb{R} \times \mathbb{R} \times V$ . This defines a convex cone that is the closure of the epigraph set of the perspective function of  $\varphi$ :

$$\mathcal{K}_p := \text{cl}\{(u, v, w) \in \mathbb{R} \times \mathbb{R}_{>} \times \text{int}(\mathcal{Q}) : u \geq v\varphi(v^{-1}w)\}. \quad (30)$$

Equivalently, we can view  $\mathcal{K}_p$  as the closed conic hull of the epigraph set of  $\varphi$  [29, Chapter 5]. In Section 6, we consider the special case where  $\varphi$  is a separable spectral function with matrix monotone first derivative. The membership check for  $\text{int}(\mathcal{K}_p)$  is similar to that of  $\text{int}(\mathcal{K}_h)$  except we first check whether  $v > 0$ .

**4.3. Dual cones.** A proper cone is *primitive* if it cannot be written as a Cartesian product of two or more lower dimensional proper cones. The dual cone of a primitive, proper cone  $\mathcal{K}$  is another primitive, proper cone:

$$\mathcal{K}^* := \{z : \langle s, z \rangle \geq 0, \forall s \in \mathcal{K}\}. \quad (31)$$

Recall that when  $\mathcal{K}$  is defined through Hypatia’s generic cone interface, both  $\mathcal{K}$  and  $\mathcal{K}^*$  become available for constructing conic models.

We assume that  $\varphi$  is convex, and we derive the dual cones of the epigraph cones  $\mathcal{K}_h$  and  $\mathcal{K}_p$  (these steps can be adapted for analogous hypograph cones if  $\varphi$  is concave). We define the convex conjugate function  $\varphi^* : V \rightarrow \mathbb{R} \cup \infty$  of  $\varphi$  as the modified Legendre-Fenchel transformation (similar to Zhang [45, page 483]):

$$\varphi^*(r) = \sup_{w \in \text{dom}(\varphi)} \{-\langle w, r \rangle - \varphi(w)\}, \quad (32)$$

which is a convex function. The conjugate of a symmetric function is also a symmetric function [2, Lemma 29], and the conjugate of a spectral function induced by a symmetric function  $f$  is the spectral function induced by  $f^*$  [2, Theorem 30]. Thus for  $\varphi(w) = f(\lambda(w))$  we have the conjugate function  $\varphi^*(w) = f^*(\lambda(w))$ .

For the epigraph-perspective cone  $\mathcal{K}_p$  in Equation (30), Zhang [45, Theorem 3.2] and Rockafellar [36, Theorem 14.4] derive the dual cone  $\mathcal{K}_p^*$ :

$$\mathcal{K}_p^* = \text{cl}\{(u, v, w) \in \mathbb{R}_{>} \times \mathbb{R} \times V : v \geq u\varphi^*(u^{-1}w)\}. \quad (33)$$

We can view  $\mathcal{K}_p^*$  as the epigraph set of the perspective function of the conjugate of  $\varphi$ , but with the epigraph and perspective components swapped (compare to Equation (30)). Depending on the natural domain of  $\varphi^*$ , the  $w$  component of  $\mathcal{K}_p^*$  is not necessarily restricted to lie in  $\mathcal{Q}$ ; in Section 6.2 we discuss several example spectral functions, some of which have conjugates defined on all  $V$  and others only on  $\text{int}(\mathcal{Q})$ .

For  $\mathcal{K}_h$  in Equation (29), we derive the dual cone  $\mathcal{K}_h^*$  as follows. Since  $\varphi$  is homogeneous in this case,  $(\text{per } \varphi)(v, w) = v\varphi(v^{-1}w) = \varphi(w)$ . Therefore the corresponding perspective cone  $\mathcal{K}_p$  for  $\varphi$  is not a primitive cone, as it can be written as a (permuted) Cartesian product of  $\mathbb{R}_{\geq}$  and  $\mathcal{K}_h$ :

$$\mathcal{K}_p = \text{cl}\{(u, v, w) \in \tilde{V} : v \in \mathbb{R}_{\geq}, (u, w) \in \mathcal{K}_h\}. \quad (34)$$

Since the dual cone of a Cartesian product of cones is the Cartesian product of their dual cones, we have (since  $\mathbb{R}_{\geq}^* = \mathbb{R}_{\geq}$ ):

$$\mathcal{K}_p^* = \text{cl}\{(u, v, w) \in \tilde{V} : v \in \mathbb{R}_{\geq}, (u, w) \in \mathcal{K}_h^*\}. \quad (35)$$

By [23, Theorem 2.1], the homogeneity of  $\varphi$  implies that  $\varphi^*$  can only take the values zero or infinity. Hence by Equation (33), we know:

$$\mathcal{K}_p^* = \text{cl}\{(u, v, w) \in \mathbb{R}_{>} \times \mathbb{R} \times V : v \geq 0, u\varphi^*(u^{-1}w) < \infty\}. \quad (36)$$

Since Equations (35) and (36) describe the same cone, we can conclude that the dual cone of  $\mathcal{K}_h$  is:

$$\mathcal{K}_h^* = \text{cl}\{(u, w) \in \mathbb{R}_{>} \times V : \varphi^*(u^{-1}w) < \infty\}. \quad (37)$$

**4.4. Barrier functions and oracles.** A logarithmically homogeneous barrier (LHB) function  $\Gamma$  for a proper cone  $\mathcal{K} \subset \tilde{V}$  is  $C^2$ -smooth and satisfies  $\Gamma(\tilde{u}_i) \rightarrow \infty$  along every sequence  $\tilde{u}_i \in \text{int}(\mathcal{K})$  converging to the boundary of  $\mathcal{K}$ , and:

$$\Gamma(\theta\tilde{u}) = \Gamma(\tilde{u}) - \nu \log(\theta) \quad \forall \tilde{u} \in \text{int}(\mathcal{K}), \theta \in \mathbb{R}_{>}, \quad (38)$$

for some  $\nu \geq 0$  [29, Definition 2.3.2]. If  $\Gamma$  is also self-concordant, then it is an LHSCB with parameter  $\nu \geq 1$  (or a  $\nu$ -LHSCB) for  $\mathcal{K}$ . For self-concordance,  $\Gamma$  must be  $C^3$ -smooth and satisfy [29, Definition 2.1.1]:

$$|\nabla^3\Gamma(\tilde{u})[\tilde{p}, \tilde{p}, \tilde{p}]| \leq 2(\nabla^2\Gamma(\tilde{u})[\tilde{p}, \tilde{p}])^{3/2} \quad \forall \tilde{u} \in \text{int}(\mathcal{K}), \tilde{p} \in \tilde{V}. \quad (39)$$

The best known PDIPMs need at most  $O(\sqrt{\nu} \log(1/\varepsilon))$  iterations to converge to a solution within  $\varepsilon$  tolerance [31]. This includes the algorithm by Skajaa and Ye [39], Papp and Yıldız [33], which Hypatia is based on [8].

The LHB we consider for a cone of the form Equation (28) is:

$$\Gamma(\tilde{u}) := -\log(\zeta(\tilde{u})) + \Psi(\tilde{u}), \quad (40)$$

where  $\Psi$  can be thought of as an LHSCB for the domain of  $\zeta$  or  $\text{cl}(\mathcal{E})$ . The negative logarithm function  $-\log$  is the standard LHSCB for  $\mathbb{R}_{\geq}$ , with parameter  $\nu = 1$ . Similarly, the spectral function  $-\log\det$  (see Section 3.3) is the standard LHSCB for a cone of squares  $\mathcal{Q}$  of  $V$ , with  $\nu = d$  (the rank of  $V$ ). For  $\mathcal{K}_h$  we let  $\Psi(\tilde{u}) = -\log\det(w)$ , hence  $\Gamma$  has parameter  $\nu = 1 + d$ . Since an LHSCB for a Cartesian product of cones is the sum of LHSCBs for the primitive cones, for  $\mathcal{K}_p$  we let  $\Psi(\tilde{u}) = -\log(v) - \log\det(w)$ , hence  $\Gamma$  has parameter  $\nu = 2 + d$ . Note that although  $\Gamma$  is an LHB, it is not necessarily self-concordant; in Sections 6.4 and 7.3 we prove that  $\Gamma$  is an LHSCB for some useful special cases.

We now define four barrier oracles that Hypatia's PDIPM uses; for ideal performance, these oracle implementations should be efficient and numerically stable. For an interior point  $\tilde{u} \in \text{int}(\mathcal{K})$

and a direction  $\tilde{p} \in \tilde{V}$ , the gradient  $g$ , the Hessian product  $H$ , the inverse Hessian product  $\bar{H}$ , and the third order directional derivative  $T$  are:

$$g := \nabla \Gamma(\tilde{u}), \quad (41a)$$

$$H := \nabla^2 \Gamma(\tilde{u})[\tilde{p}], \quad (41b)$$

$$\bar{H} := (\nabla^2 \Gamma(\tilde{u}))^{-1}[\tilde{p}], \quad (41c)$$

$$T := \nabla^3 \Gamma(\tilde{u})[\tilde{p}, \tilde{p}]. \quad (41d)$$

Note  $g, H, \bar{H}, T \in \tilde{V}$ . In later sections, we use subscripts to refer to subcomponents of these oracles, for example the  $w$  component of the gradient oracle is  $g_w := \nabla_w \Gamma(\tilde{u}) \in V$ . Ideally,  $H$  applies the positive definite linear operator  $\nabla^2 \Gamma(\tilde{u}) : \tilde{V} \rightarrow \tilde{V}$  without constructing an explicit Hessian, and similarly,  $\bar{H}$  applies the (unique) inverse operator  $(\nabla^2 \Gamma(\tilde{u}))^{-1} : \tilde{V} \rightarrow \tilde{V}$  without constructing or factorizing an explicit Hessian.

We note that for the standard LHSCB  $\Psi$  for a cone of squares, efficient and numerically stable procedures for these four oracles are well-known. The same cannot be said for the LHB  $\Gamma$  currently. In Section 5, we derive these oracles for  $\mathcal{K}_p$  (noting that they can be adapted easily for  $\mathcal{K}_h$ ). In the special cases for which we show  $\Gamma$  is an LHSCB, the oracles can be computed particularly efficiently.

**5. Barrier oracles for epigraph-perspective cones.** In this section, we derive expressions and evaluation procedures for the  $g, H, T$ , and  $\bar{H}$  oracles (defined in Section 4.4) corresponding to the LHB  $\Gamma$  for the epigraph-perspective cone  $\mathcal{K}_p$  (defined in Equation (30)). Recall that we let  $\tilde{p} = (p, q, r) \in \mathbb{R} \times \mathbb{R} \times V$  and  $\tilde{u} = (u, v, w) \in \text{int}(\mathcal{K}_p)$ , and we define  $\zeta$  and  $\Gamma : \text{int}(\mathcal{K}_p) \rightarrow \mathbb{R}$  from Equation (40) as:

$$\zeta(\tilde{u}) := u - v\varphi(v^{-1}w), \quad (42a)$$

$$\Gamma(\tilde{u}) := -\log(\zeta(\tilde{u})) - \log(v) - \log \det(w). \quad (42b)$$

We note that the oracles for  $\mathcal{K}_h$  in Equation (29) are simpler because no perspective operation is needed for a homogeneous  $\varphi$ ; they can be obtained by fixing  $v = 1$  and  $q = 0$  and ignoring the  $v$  components in the oracle expressions in this section.

Without assuming any particular form for  $\varphi$ , we write  $g, H$ , and  $T$  in Section 5.1 and  $\bar{H}$  in Section 5.2 in terms of the derivatives of  $\varphi$ . If  $\varphi$  is a spectral function, these derivatives can be computed using the expressions from Section 3. In the case that  $\varphi$  is a separable spectral function (see Section 3.2), we derive in Section 5.3 a more specialized procedure for  $\bar{H}$ , which is no more expensive than  $H$ . Finally, in Section 5.4, we specialize the four oracles for the negative log-determinant function (i.e.  $\varphi(w) = -\log \det(w)$ ; see Section 3.3) and discuss implementations. In Section 8.1, we describe our implementations of these oracles in Hypatia and link to the high-level Julia code.

**5.1. Derivatives.** First, we express the derivatives of  $\zeta$  in terms of those of  $\varphi$ . We define the function  $\mu : \mathbb{R}_{>} \times \text{int}(\mathcal{Q}) \rightarrow \text{int}(\mathcal{Q})$  and its first directional derivative  $\xi \in V$  in the direction  $(q, r)$  as:

$$\mu(v, w) := v^{-1}w, \quad (43a)$$

$$\xi := \nabla \mu(v, w)[(q, r)] = \nabla_v \mu(v, w)q + \nabla_w \mu(v, w)[r] = v^{-1}(r - q\mu(v, w)). \quad (43b)$$

For convenience, we fix the constants  $\mu := \mu(v, w)$ ,  $\varphi := \varphi(\mu)$ , and  $\zeta := \zeta(\tilde{u})$ . Let  $\nabla \varphi$ ,  $\nabla^2 \varphi$ , and  $\nabla^3 \varphi$  be the derivatives of  $\varphi$  evaluated at  $\mu$ , and let  $\nabla \zeta$ ,  $\nabla^2 \zeta$ , and  $\nabla^3 \zeta$  be the derivatives of  $\zeta$  evaluated at  $\tilde{u}$ . Using Equation (43), the directional derivatives of  $\zeta$  can be written compactly as:

$$\nabla_u \zeta = 1, \quad (44a)$$

$$\nabla_v \zeta = -\varphi + \nabla\varphi[\mu], \quad (44b)$$

$$\nabla_w \zeta = -\nabla\varphi, \quad (44c)$$

$$\nabla\zeta[\tilde{p}] = p - q\varphi - v\nabla\varphi[\xi], \quad (44d)$$

$$(\nabla^2\zeta[\tilde{p}])_v = \nabla^2\varphi[\xi, \mu], \quad (44e)$$

$$(\nabla^2\zeta[\tilde{p}])_w = -\nabla^2\varphi[\xi], \quad (44f)$$

$$\nabla^2\zeta[\tilde{p}, \tilde{p}] = -v\nabla^2\varphi[\xi, \xi], \quad (44g)$$

$$(\nabla^3\zeta[\tilde{p}, \tilde{p}])_v = \nabla^3\varphi[\xi, \xi, \mu] + \nabla^2\varphi[\xi, \xi] - 2v^{-1}q\nabla^2\varphi[\xi, \mu], \quad (44h)$$

$$(\nabla^3\zeta[\tilde{p}, \tilde{p}])_w = 2v^{-1}q\nabla^2\varphi[\xi] - \nabla^3\varphi[\xi, \xi], \quad (44i)$$

$$\nabla^3\zeta[\tilde{p}, \tilde{p}, \tilde{p}] = -v\nabla^3\varphi[\xi, \xi, \xi] + 3q\nabla^2\varphi[\xi, \xi]. \quad (44j)$$

Using Equation (44), we now derive the directional derivatives of  $\Gamma$ . For convenience, we let  $\nabla\Gamma$ ,  $\nabla^2\Gamma$ ,  $\nabla^3\Gamma$  be the derivatives of  $\Gamma$  evaluated at  $\tilde{u}$ . We define:

$$\sigma := -\nabla_v \zeta = \varphi - \nabla\varphi[\mu] \in \mathbb{R}. \quad (45)$$

The components of the gradient  $g$  of  $\Gamma$  are:

$$g_u = -\zeta^{-1}, \quad (46a)$$

$$g_v = \zeta^{-1}\sigma - v^{-1}, \quad (46b)$$

$$g_w \stackrel{26}{=} \zeta^{-1}\nabla\varphi - w^{-1}. \quad (46c)$$

Differentiating Equation (46), the Hessian components are:

$$\nabla_{u,u}^2\Gamma = \zeta^{-2} > 0, \quad (47a)$$

$$\nabla_{v,u}^2\Gamma = -\zeta^{-2}\sigma \in \mathbb{R}, \quad (47b)$$

$$\nabla_{w,u}^2\Gamma = -\zeta^{-2}\nabla\varphi \in V, \quad (47c)$$

$$\nabla_{v,v}^2\Gamma = v^{-2} + \zeta^{-2}\sigma^2 + v^{-1}\zeta^{-1}\nabla^2\varphi[\mu, \mu] > 0, \quad (47d)$$

$$\nabla_{w,v}^2\Gamma = \zeta^{-2}\sigma\nabla\varphi - v^{-1}\zeta^{-1}\nabla^2\varphi[\mu] \in V. \quad (47e)$$

Differentiating Equation (46c) in the direction  $r$ :

$$\nabla_{w,w}^2\Gamma[r] \stackrel{24}{=} \zeta^{-2}\nabla\varphi[r]\nabla\varphi + v^{-1}\zeta^{-1}\nabla^2\varphi[r] + P(w^{-1})r \in V. \quad (48)$$

Let:

$$\chi := \zeta^{-1}(p - q\sigma - \nabla\varphi[r]) \in \mathbb{R}. \quad (49)$$

The components of the Hessian product  $H$  are:

$$H_u = \zeta^{-1}\chi, \quad (50a)$$

$$H_v = -\zeta^{-1}\sigma\chi - \zeta^{-1}\nabla^2\varphi[\xi, \mu] + v^{-2}q, \quad (50b)$$

$$H_w = -\zeta^{-1}\chi\nabla\varphi + \zeta^{-1}\nabla^2\varphi[\xi] + P(w^{-1})r. \quad (50c)$$

Let:

$$\kappa := 2\zeta^{-1}(\chi + v^{-1}q)\nabla^2\varphi[\xi] - \zeta^{-1}\nabla^3\varphi[\xi, \xi] \in V. \quad (51)$$

The components of the third order directional derivative  $T$  are:

$$T_u = -2\zeta^{-1}\chi^2 - v\zeta^{-2}\nabla^2\varphi[\xi, \xi], \quad (52a)$$

$$T_v = -T_u\sigma + \langle \kappa, \mu \rangle - \zeta^{-1}\nabla^2\varphi[\xi, \xi] - 2q^2v^{-3}, \quad (52b)$$

$$T_w \stackrel{25}{=} -T_u\nabla\varphi - \kappa - 2P(w^{-1/2})(P(w^{-1/2})r)^2. \quad (52c)$$

**5.2. Inverse Hessian operator.** The Hessian of  $\Gamma$  at any point  $\tilde{u} \in \text{int}(\mathcal{K}_p)$  is a positive definite linear operator and hence invertible. By treating the components of the Hessian in Equations (47) and (48) analogously to blocks of a positive definite matrix, we derive the inverse operator. For convenience, we let:

$$Y_u := (\nabla_{w,w}^2 \Gamma)^{-1} \nabla_{w,u}^2 \Gamma, \quad (53a)$$

$$Y_v := (\nabla_{w,w}^2 \Gamma)^{-1} \nabla_{w,v}^2 \Gamma, \quad (53b)$$

$$Z_{u,u} := \nabla_{u,u}^2 \Gamma - \langle \nabla_{w,u}^2 \Gamma, Y_u \rangle, \quad (53c)$$

$$Z_{v,u} := \nabla_{v,u}^2 \Gamma - \langle \nabla_{w,u}^2 \Gamma, Y_v \rangle, \quad (53d)$$

$$Z_{v,v} := \nabla_{v,v}^2 \Gamma - \langle \nabla_{w,v}^2 \Gamma, Y_v \rangle. \quad (53e)$$

Note  $Y_u, Y_v \in V$ . We let  $Z$  be:

$$Z := \begin{bmatrix} Z_{u,u} & Z_{v,u} \\ Z_{v,u} & Z_{v,v} \end{bmatrix} \in \mathbb{S}_{\zeta}^2, \quad (54)$$

and its inverse is:

$$\bar{Z} := Z^{-1} = \frac{1}{Z_{u,u}Z_{v,v} - Z_{v,u}^2} \begin{bmatrix} Z_{v,v} & -Z_{v,u} \\ -Z_{v,u} & Z_{u,u} \end{bmatrix} \in \mathbb{S}_{\zeta}^2. \quad (55)$$

It can be verified (for example, by analogy to the block symmetric matrix inverse formula) that the inverse Hessian product oracle  $\bar{H}$  in Equation (41c) is:

$$\bar{H}_u = \bar{Z}_{u,u}(p - \langle Y_u, r \rangle) + \bar{Z}_{v,u}(q - \langle Y_v, r \rangle), \quad (56a)$$

$$\bar{H}_v = \bar{Z}_{v,u}(p - \langle Y_u, r \rangle) + \bar{Z}_{v,v}(q - \langle Y_v, r \rangle), \quad (56b)$$

$$\bar{H}_w = -\bar{H}_u Y_u - \bar{H}_v Y_v + (\nabla_{w,w}^2 \Gamma)^{-1} r. \quad (56c)$$

Hence computing  $\bar{H}$  is essentially only as difficult as applying the positive definite linear operator  $(\nabla_{w,w}^2 \Gamma)^{-1}$ . We are not aware of a simple expression for  $(\nabla_{w,w}^2 \Gamma)^{-1}$  in general, but we explore the special cases of separable spectral functions in Section 5.3, the negative log-determinant function in Section 5.4, and the root-determinant function in Section 7.4.

**5.3. Inverse Hessian operator for the separable spectral case.** Suppose  $w \succ 0$  has the spectral decomposition Equation (5), i.e.  $w$  has the eigenvalues  $w_1, \dots, w_d > 0$  and the Jordan frame  $c_1, \dots, c_d$ . As in Section 3.2, we assume distinct eigenvalues for simplicity. In the special case where  $\varphi$  is a convex separable spectral function, i.e.  $\varphi(w) = \sum_{i \in [d]} h(w_i)$  for some convex  $h: \mathbb{R}_{>} \rightarrow \mathbb{R}$ , we show how to compute  $\bar{H}$  as efficiently as Hessian product oracle  $H$ . For all  $i \in [d]$ , we let  $h_i$ ,  $(\nabla h)_i$ ,  $(\nabla^2 h)_i$ , and  $(\nabla^3 h)_i$  denote the value and derivatives of  $h$  evaluated at  $\mu$ . We define  $m_{i,j} \in \mathbb{R}$  for  $i, j \in [d]$  as:

$$m_{i,j} := \begin{cases} \zeta^{-1} \frac{(\nabla h)_i (\nabla h)_j}{w_i - w_j} + w_i^{-1} w_j^{-1} & i \neq j, \\ \zeta^{-1} w^{-1} (\nabla^2 h)_i + w_i^{-2} & i = j. \end{cases} \quad (57)$$

Since  $h$  is convex,  $m_{i,j} > 0, \forall i, j \in [d]$ . Let  $M: V \rightarrow V$  be the self-adjoint linear operator:

$$M := v^{-1} \zeta^{-1} \nabla^2 \varphi + P(w^{-1}) = \sum_{i,j \in [d]: i < j} 4m_{i,j} L(c_i) L(c_j) + \sum_{i \in [d]} m_{i,i} P(c_i). \quad (58)$$

Using Equation (12), we have the self-adjoint inverse operator of  $M$ :

$$M^{-1} = \sum_{i,j \in [d]: i < j} 4m_{i,j}^{-1} L(c_i) L(c_j) + \sum_{i \in [d]} m_{i,i}^{-1} P(c_i). \quad (59)$$

Substituting Equation (58) into Equation (48), we have:

$$\nabla_{w,w}^2 \Gamma[r] = \zeta^{-2} \nabla \varphi[r] \nabla \varphi + Mr. \quad (60)$$

Note that the first term in Equation (60) is analogous to the application (to  $r$ ) of a low-rank update to  $M$ , and that  $M^{-1}$  in Equation (59) is easy to apply. By analogy to the Sherman-Morrison-Woodbury formula [11, Theorem 1.1], we can derive a simple expression for the inverse operator  $(\nabla_{w,w}^2 \Gamma)^{-1}r$ .

We let:

$$\alpha := M^{-1} \nabla \varphi = \sum_{i \in [d]} m_{i,i}^{-1} (\nabla h)_i c_i \in V, \quad (61a)$$

$$\gamma := v^{-2} \zeta^{-1} M^{-1} \nabla^2 \varphi[w] = v^{-2} \zeta^{-1} \sum_{i \in [d]} m_{i,i}^{-1} (\nabla^2 h)_i w_i c_i \succ 0. \quad (61b)$$

Noting that  $\gamma, w^{-1} \succ 0$  implies  $\langle \gamma, w^{-1} \rangle > 0$ , we define the scalar constants:

$$k_1 := \zeta^2 + \langle \nabla \varphi, \alpha \rangle > 0, \quad (62a)$$

$$k_2 := \sigma + \langle \nabla \varphi, \gamma \rangle = \sigma + v^{-2} \zeta^{-1} \langle \nabla^2 \varphi[w], \alpha \rangle, \quad (62b)$$

$$k_3 := v^{-2} + v^{-2} \zeta^{-1} \langle \nabla^2 \varphi[w], \mu - \gamma \rangle = v^{-2} + v^{-1} \langle \gamma, w^{-1} \rangle > 0. \quad (62c)$$

Now using the Sherman-Morrison-Woodbury formula:

$$(\nabla_{w,w}^2 \Gamma)^{-1}r = M^{-1}r - \frac{\zeta^{-2} \langle M^{-1} \nabla \varphi, r \rangle}{1 + \zeta^{-2} \langle M^{-1} \nabla \varphi, \nabla \varphi \rangle} M^{-1} \nabla \varphi \quad (63a)$$

$$= M^{-1}r - k_1^{-1} \langle \alpha, r \rangle \alpha. \quad (63b)$$

Substituting Equations (47) and (63) into Equation (53):

$$Y_u = (\nabla_{w,w}^2 \Gamma)^{-1} (-\zeta^{-2} \nabla \varphi) \quad (64a)$$

$$= -\zeta^{-2} \alpha + \zeta^{-2} k_1^{-1} \langle \alpha, \nabla \varphi \rangle \alpha \quad (64b)$$

$$= -k_1^{-1} \alpha, \quad (64c)$$

$$Y_v = (\nabla_{w,w}^2 \Gamma)^{-1} (\zeta^{-2} \sigma \nabla \varphi - v^{-1} \zeta^{-1} \nabla^2 \varphi[w]) \quad (64d)$$

$$= -\sigma Y_u - v^{-2} \zeta^{-1} (\nabla_{w,w}^2 \Gamma)^{-1} \nabla^2 \varphi[w] \quad (64e)$$

$$= \sigma k_1^{-1} \alpha - \gamma + v^{-2} \zeta^{-1} k_1^{-1} \langle \alpha, \nabla^2 \varphi[w] \rangle \alpha \quad (64f)$$

$$= k_1^{-1} k_2 \alpha - \gamma, \quad (64g)$$

$$Z_{u,u} = \zeta^{-2} - \langle \nabla_{w,u}^2 \Gamma, Y_u \rangle \quad (64h)$$

$$= \zeta^{-2} - \zeta^{-2} k_1^{-1} \langle \nabla \varphi, \alpha \rangle \quad (64i)$$

$$= k_1^{-1}, \quad (64j)$$

$$Z_{v,u} = -\zeta^{-2} \sigma - \langle \nabla_{w,u}^2 \Gamma, Y_v \rangle \quad (64k)$$

$$= -\zeta^{-2} (\sigma - \langle \nabla \varphi, k_1^{-1} k_2 \alpha - \gamma \rangle) \quad (64l)$$

$$= -\zeta^{-2} (\sigma - k_1^{-1} k_2 (k_1 - \zeta^2) + k_2 - \sigma) \quad (64m)$$

$$= -k_1^{-1} k_2, \quad (64n)$$

$$Z_{v,v} = \nabla_{v,v}^2 \Gamma - \langle \nabla_{w,v}^2 \Gamma, Y_v \rangle \quad (64o)$$

$$= \nabla_{v,v}^2 \Gamma + \sigma \langle \nabla_{w,u}^2 \Gamma, Y_v \rangle + v^{-2} \zeta^{-1} \langle \nabla^2 \varphi[w], Y_v \rangle \quad (64p)$$

$$\stackrel{64n}{=} \nabla_{v,v}^2 \Gamma + \sigma (k_1^{-1} k_2 - \zeta^{-2} \sigma) + v^{-2} \zeta^{-1} \langle \nabla^2 \varphi[w], Y_v \rangle \quad (64q)$$

$$= v^{-2} + v^{-3} \zeta^{-1} \nabla^2 \varphi[w, w] + \sigma k_1^{-1} k_2 + v^{-2} \zeta^{-1} \langle \nabla^2 \varphi[w], Y_v \rangle \quad (64r)$$

$$= v^{-2} + v^{-3} \zeta^{-1} \nabla^2 \varphi[w, w] + \sigma k_1^{-1} k_2 + k_1^{-1} k_2 (k_2 - \sigma) - v^{-2} \zeta^{-1} \langle \nabla^2 \varphi[w], \gamma \rangle \quad (64s)$$

$$= v^{-2} + k_1^{-1} k_2 + v^{-2} \zeta^{-1} \langle \nabla^2 \varphi[w], \mu - \gamma \rangle \quad (64t)$$

$$= k_3 + k_1^{-1} k_2^2. \quad (64u)$$

For  $Z$  in Equation (54), we have  $\det(Z) = k_1^{-1}k_3$ , so its inverse  $\bar{Z}$  in Equation (55) is:

$$\bar{Z}_{u,u} = k_1(k_3 + k_1^{-1}k_2^2)k_3^{-1} = k_1 + k_2^2k_3^{-1}, \quad (65a)$$

$$\bar{Z}_{u,v} = k_2k_3^{-1}, \quad (65b)$$

$$\bar{Z}_{v,v} = k_3^{-1}. \quad (65c)$$

Finally, we substitute Equations (64) and (65) into Equation (56) to derive the inverse Hessian product  $\bar{H}$ . We let:

$$c_1 := p - \langle Y_u, r \rangle \stackrel{64}{=} p + k_1^{-1}\langle \alpha, r \rangle, \quad (66a)$$

$$c_2 := q - \langle Y_v, r \rangle \stackrel{64}{=} q - k_1^{-1}k_2\langle \alpha, r \rangle + \langle \gamma, r \rangle. \quad (66b)$$

For convenience, we derive  $\bar{H}_v$  before  $\bar{H}_u$  and  $\bar{H}_w$ :

$$\bar{H}_v = \bar{Z}_{u,v}c_1 + \bar{Z}_{v,v}c_2 \quad (67a)$$

$$= k_3^{-1}(k_2c_1 + c_2) \quad (67b)$$

$$= k_3^{-1}(k_2p + q + \langle \gamma, r \rangle), \quad (67c)$$

$$\bar{H}_u = \bar{Z}_{u,u}c_1 + \bar{Z}_{u,v}c_2 \quad (67d)$$

$$\stackrel{67b}{=} (\bar{Z}_{u,u} - \bar{Z}_{u,v}k_2)c_1 + \bar{Z}_{u,v}k_3\bar{H}_v \quad (67e)$$

$$= k_1p + k_2\bar{H}_v + \langle \alpha, r \rangle, \quad (67f)$$

$$\bar{H}_w = -\bar{H}_uY_u - \bar{H}_vY_v + (\nabla_{w,w}^2)^{-1}r \quad (67g)$$

$$\stackrel{63}{=} \bar{H}_uk_1^{-1}\alpha - \bar{H}_v(k_1^{-1}k_2\alpha - \gamma) + M^{-1}r - k_1^{-1}\langle \alpha, r \rangle\alpha \quad (67h)$$

$$= p\alpha + \bar{H}_v\gamma + M^{-1}r. \quad (67i)$$

In Section 8.5, we compare the efficiency and numerical performance of the closed-form formula for  $\bar{H}$  in Equation (67) against a naive approach to computing  $\bar{H}$  that performs a Cholesky factorization of an explicit Hessian matrix and uses a direct linear solve. The closed-form formula is faster and more scalable, more memory-efficient, more reliable to compute (as the Cholesky decomposition can fail), and more numerically accurate.

**5.4. Oracles for the log-determinant case.** We now specialize the oracles derived in Sections 5.1 and 5.3 for the separable spectral function  $\varphi(w) = -\log\det(w) = -\sum_{i \in [d]} \log(w_i)$ . In Section 6, we show that  $\Gamma$  is an LHSCB in this case. We let:

$$\hat{\xi} := P(w^{-1/2})\xi = v^{-1}P(w^{-1/2})(-v^{-1}qw + r) = v^{-1}(-v^{-1}qe + \hat{r}) \in V. \quad (68)$$

We have:

$$\nabla\varphi \stackrel{26}{=} -\mu^{-1} = -vP(w^{-1/2})e, \quad (69a)$$

$$\nabla^2\varphi[\xi] \stackrel{27a}{=} v^2P(w^{-1})\xi = v^2P(w^{-1/2})\hat{\xi}, \quad (69b)$$

$$\nabla^3\varphi[\xi, \xi] \stackrel{27b}{=} -2v^3P(w^{-1/2})\hat{\xi}^2. \quad (69c)$$

The constants from Section 5.1 have the form:

$$\sigma \stackrel{45}{=} \varphi + d, \quad (70a)$$

$$\chi \stackrel{49}{=} \zeta^{-1}(p - q\sigma + v\text{tr}(\hat{r})). \quad (70b)$$

From Equation (46), the  $w$  component of the gradient is:

$$g_w = -(1 + v\zeta^{-1})w^{-1}. \quad (71a)$$

From Equation (50), the  $v$  and  $w$  components of the Hessian product are:

$$H_v = -\zeta^{-1}\sigma\chi - v\zeta^{-1}\text{tr}(\hat{\xi}) + v^{-2}q, \quad (72a)$$

$$H_w = P(w^{-1/2})(v\zeta^{-1}\chi e + v^2\zeta^{-1}\hat{\xi} + \hat{r}). \quad (72b)$$

From Equation (52), the third order directional derivative is:

$$T_u = -2\zeta^{-1}\chi^2 - v^3\zeta^{-2}\text{tr}(\hat{\xi}^2), \quad (73a)$$

$$T_v = -T_u\sigma + v\zeta^{-1}(2(\chi + v^{-1}q)\text{tr}(\hat{\xi}) + v\text{tr}(\hat{\xi}^2)) - 2v^{-3}q^2, \quad (73b)$$

$$T_w = P(w^{-1/2})(T_u v e - 2\zeta^{-1}v^2((\chi + v^{-1}q)\hat{\xi} + v\hat{\xi}^2) - 2\hat{r}^2). \quad (73c)$$

We derive the inverse Hessian product  $\bar{H}$  by specializing the separable case in Equation (67). We let:

$$\check{r} := P(w^{1/2})r \in V, \quad (74a)$$

$$\theta := v^2(\zeta + (1+d)v)^{-1}. \quad (74b)$$

From Equations (61) and (62), we have:

$$M^{-1} = \zeta(\zeta + v)^{-1}P(w), \quad (75a)$$

$$\alpha = -v\zeta(\zeta + v)^{-1}w, \quad (75b)$$

$$\gamma = (\zeta + v)^{-1}w, \quad (75c)$$

$$k_1 = \zeta^2 + dv^2\zeta(\zeta + v)^{-1}, \quad (75d)$$

$$k_2 = \varphi + d\zeta(\zeta + v)^{-1}, \quad (75e)$$

$$k_3^{-1} = (\zeta + v)\theta. \quad (75f)$$

For convenience, we derive  $\bar{H}_v$  before  $\bar{H}_u$  and  $\bar{H}_w$  as in Equation (67):

$$\bar{H}_v = k_3^{-1}(k_2 p + q + \langle \gamma, r \rangle) \quad (76a)$$

$$= (\zeta + v)\theta((\varphi + d\zeta(\zeta + v)^{-1})p + q + \langle r, w \rangle(\zeta + v)^{-1}) \quad (76b)$$

$$= \theta((\zeta + v)(\varphi p + q) + d\zeta p + \text{tr}(\check{r})), \quad (76c)$$

$$\bar{H}_u = k_1 p + (\varphi + d\zeta(\zeta + v)^{-1})\bar{H}_v + \langle \alpha, r \rangle \quad (76d)$$

$$= (\zeta^2 + dv^2\zeta(\zeta + v)^{-1})p + (\varphi + d\zeta(\zeta + v)^{-1})\bar{H}_v + \langle -v\zeta(\zeta + v)^{-1}w, r \rangle \quad (76e)$$

$$= \zeta(\zeta + v)^{-1}(dv^2 p + d\bar{H}_v - v\langle w, r \rangle) + \zeta^2 p + \varphi\bar{H}_v, \quad (76f)$$

$$\bar{H}_w = p\alpha + \bar{H}_v\gamma + M^{-1}r \quad (76g)$$

$$= -pv\zeta(\zeta + v)^{-1}w + \bar{H}_v(\zeta + v)^{-1}w + \zeta(\zeta + v)^{-1}P(w)r \quad (76h)$$

$$= (\zeta + v)^{-1}P(w^{1/2})((-\zeta p v + \bar{H}_v)e + \zeta\check{r}). \quad (76i)$$

Recall that in Section 5.3, we used the simplifying assumption of distinct eigenvalues, but for the negative log-determinant case this is not necessary. Note that if it is possible to apply  $P(w^{1/2})$  and  $P(w^{-1/2})$  without accessing the eigenvalues of  $w$ , then all oracles can be computed without an explicit eigendecomposition. For example, in our implementation for  $V = \mathbb{S}^d$  and  $V = \mathbb{H}^d$ , only a Cholesky factorization of  $w$  is needed. This is unlike the more general separable spectral function case, where the explicit eigenvalues of  $w$  are needed.

**6. Matrix monotone derivative cones.** After defining the matrix monotone property of a function in Section 6.1, we introduce the matrix monotone derivative cone  $\mathcal{K}_{\text{MMD}}$  in Section 6.2.  $\mathcal{K}_{\text{MMD}}$  is a special case of the epigraph-perspective cone  $\mathcal{K}_p$  with a separable spectral function  $\varphi$ . In Section 6.4, we prove that our barrier function  $\Gamma$  for  $\mathcal{K}_{\text{MMD}}$  is an LHSCB, using derivative results we obtain in Section 6.3.



**6.1. Matrix monotonicity.** A function  $f$  is *matrix monotone* (or operator monotone) if  $w_a \succeq w_b \succeq 0$  implies  $f(w_a) \succeq f(w_b)$  for all  $w_a, w_b \in \mathbb{S}^d$  for all integers  $d$ . The following integral representation result is attributed to Löwner [24] (see e.g. Kwong [22, Theorem 1] and Furuta [18, Theorem L]). A function  $f : \mathbb{R}_{>} \rightarrow \mathbb{R}$  is matrix monotone in  $\mathbb{R}_{>}$  if and only if it has the representation:

$$f(x) = \alpha + \beta x + \int_0^\infty \frac{x}{x+t} \, d\rho(t) = \alpha + \beta x + \int_0^\infty \left(1 - \frac{t}{x+t}\right) \, d\rho(t), \quad (77)$$

where  $\alpha \in \mathbb{R}, \beta \in \mathbb{R}_{\geq}$  and  $\rho$  is a positive measure on  $\mathbb{R}_{>}$  such that  $\int_0^\infty (1+t)^{-1} \, d\rho(t) < \infty$ .

This result implies that, for a cone of squares  $\mathcal{Q}$  of a Jordan algebra, and for  $w \in \text{int}(\mathcal{Q})$  with the spectral decomposition  $w = \sum_{i \in [d]} w_i c_i$ , we have:

$$f(w) = \sum_{i \in [d]} f(w_i) c_i \quad (78a)$$

$$= \alpha e + \beta w + \int_0^\infty \sum_{i \in [d]} \left(1 - \frac{t}{w_i + t}\right) c_i \, d\rho(t) \quad (78b)$$

$$= \alpha e + \beta w + \int_0^\infty (e - t(w + te)^{-1}) \, d\rho(t). \quad (78c)$$

This is similar to the representation in Faybusovich and Tsuchiya [15, page 1520].

**6.2. Cone definition.** Let  $h : \mathbb{R}_{>} \rightarrow \mathbb{R}$  be a convex  $C^3$ -smooth function. We assume that the first derivative of  $h$ ,  $\nabla h$ , is a matrix monotone function. This also implies that  $h$  is convex. We call such functions *matrix monotone derivative* (MMD) functions. In Table 1, we give some common examples of MMD functions, with abbreviated names in the first column. We also give  $\nabla h$ , the domain of the convex conjugate  $h^*$  (defined in Equation (32)), and a closed-form formula for  $h^*$ .

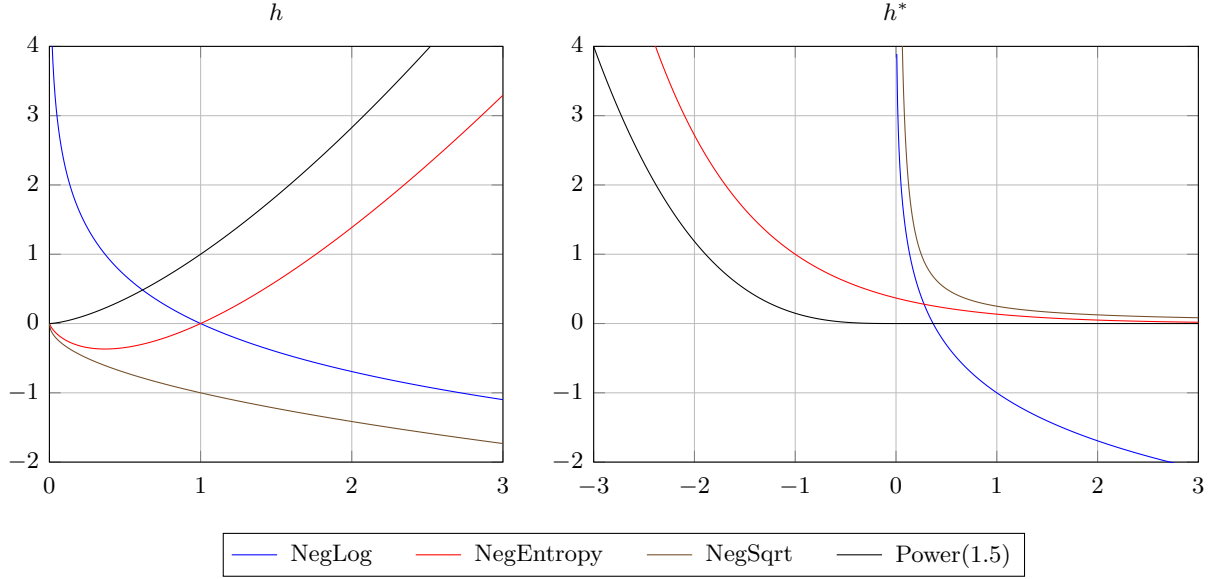
Due to Carlen [6, Theorem 2.6 (Löwner-Heinz Theorem)], the functions  $x \rightarrow \log(x)$ ,  $x \rightarrow -x^p$  for  $p \in [-1, 0]$ , and  $x \rightarrow x^p$  for  $p \in [0, 1]$  are matrix monotone. This implies that in Table 1, each function in the  $\nabla h$  column is matrix monotone. Note that NegSqrt is equivalent to NegPower for  $p = 1/2$ ; we highlight NegSqrt as an interesting case, for which the conjugate  $h^*$  is a positive rescaling of the inverse function. Note we exclude the case  $p = 1$  in NegPower and Power because it is homogeneous ( $h$  is linear).

More examples of matrix monotone functions (i.e. possible  $\nabla h$  functions) can be found in Kwong [22], Furuta [18]. Some of these functions may correspond to  $h$  or  $h^*$  functions that are directly useful for applied optimization modeling.

TABLE 1. Examples of MMD functions.

	$h$	$\nabla h$	$\text{dom}(h^*)$	$h^*$
NegLog	$-\log(x)$	$-x^{-1}$	$\mathbb{R}_{>}$	$-1 - \log(x)$
NegEntropy	$x \log(x)$	$1 + \log(x)$	$\mathbb{R}$	$\exp(-1 - x)$
NegSqrt	$-\sqrt{x}$	$-\frac{1}{2}x^{-1/2}$	$\mathbb{R}_{>}$	$\frac{1}{4}x^{-1}$
NegPower, $p \in (0, 1)$	$-x^p$	$-px^{p-1}$	$\mathbb{R}_{\geq}$	$-(p-1)(x/p)^q$
Power, $p \in (1, 2]$	$x^p$	$px^{p-1}$	$\mathbb{R}$	$(p-1)(x_-/p)^q$

We let  $q := p/(p-1)$ , which gives  $q \in (-\infty, 0)$  for  $p \in (0, 1)$  in the NegPower case, or  $q \in [2, \infty)$  for  $p \in (1, 2]$  in the Power case. We also let  $x_- := \max(-x, 0)$  in the Power case.

FIGURE 1. Graphs of some MMD functions ( $h$ , left) and their conjugates ( $h^*$ , right) from Table 1.

Suppose  $\mathcal{Q}$  is the cone of squares of a Jordan algebra  $V$  with rank  $d$ . Let  $\varphi : \text{int}(\mathcal{Q}) \rightarrow \mathbb{R}$  be the  $C^3$ -smooth function  $\varphi(w) = \text{tr}(h(w)) = \sum_{i \in [d]} h(w_i)$ , which is a convex separable spectral function (see Section 3.2). As in Section 4.2, we let  $\tilde{u} = (u, v, w) \in \mathcal{E} = \mathbb{R} \times \mathbb{R}_{>} \times \text{int}(\mathcal{Q})$ . The function  $\zeta : \mathcal{E} \rightarrow \mathbb{R}$  has the form:

$$\zeta(\tilde{u}) := u - v \text{tr}(h(v^{-1}w)). \quad (79)$$

We define the matrix monotone derivative cone  $\mathcal{K}_{\text{MMD}}$ , a special case of the epigraph-perspective cone  $\mathcal{K}_p$  in Equation (30), as:

$$\mathcal{K}_{\text{MMD}} := \text{cl}\{\tilde{u} \in \mathcal{E} : u \geq v \text{tr}(h(v^{-1}w))\}, \quad (80)$$

which is a proper cone. For the separable case, the convex conjugate  $\varphi^* : V \rightarrow \mathbb{R} \cup \infty$  (see Equation (32)) of  $\varphi$  is  $\varphi^*(r) = \text{tr}(h^*(r))$ . So from Equation (33), the dual cone is:

$$\mathcal{K}_{\text{MMD}}^* := \text{cl}\{\tilde{u} \in \mathbb{R}_{>} \times \mathbb{R} \times V : v \geq u \text{tr}(h^*(u^{-1}w))\}. \quad (81)$$

**6.3. Derivatives of the MMD trace.** We derive some expressions needed for our LHSCB proof in Section 6.4. Suppose  $w \succ 0$ . Since  $\varphi(w) = \text{tr}(h(w))$  and  $\nabla h$  is matrix monotone, from Equation (78) we can write the gradient:

$$\nabla \varphi(w) = \nabla h(w) = \alpha e + \beta w + \int_0^\infty (e - t(w + te)^{-1}) \, d\rho(t). \quad (82)$$

Note that  $w + te \succ 0$  for  $t \geq 0$ , so  $(w + te)^{-1}$  is well-defined. Differentiating Equation (82) in the direction  $r \in V$ , we have the second order directional derivative:

$$\nabla^2 \varphi(w)[r] \stackrel{24}{=} \beta r + \int_0^\infty t P((w + te)^{-1}) r \, d\rho(t), \quad (83)$$

and the third order directional derivative:

$$\nabla^3 \varphi(w)[r, r] \stackrel{25}{=} -2 \int_0^\infty t P((w + te)^{-1/2}) (P((w + te)^{-1/2}) r)^2 \, d\rho(t). \quad (84)$$

**6.4. Self-concordant barrier.** For  $\mathcal{K}_{\text{MMD}}$ , the LHB  $\Gamma : \text{int}(\mathcal{K}_{\text{MMD}}) \rightarrow \mathbb{R}$  from Equation (40) has the form:

$$\Gamma(\tilde{u}) := -\log(\zeta(\tilde{u})) - \log(v) - \log \det(w). \quad (85)$$

We describe easily-computable oracles for this  $\Gamma$  in Section 5, including an inverse Hessian product  $\bar{H}$  in Section 5.3 that is as easy to compute as the Hessian product  $H$  (since  $\varphi$  is a separable spectral function).

We note that Faybusovich and Tsuchiya [15] derive a  $(1 + d)$ -self-concordant barrier for the related convex (but not conic) set  $\mathcal{S}$ :

$$\mathcal{S} := \text{cl}\{(u, w) \in \mathbb{R} \times \text{int}(\mathcal{Q}) : u - \varphi(w) \geq 0\}. \quad (86)$$

$\mathcal{K}_{\text{MMD}}$  is the conic hull of  $\mathcal{S}$ . In Proposition 1, we prove that our barrier  $\Gamma$  in Equation (85) is self-concordant, hence it is an LHSCB for  $\mathcal{K}_{\text{MMD}}$  with parameter  $2 + d$ . This small additive increment of one in the barrier parameter is in sharp contrast to generic conic hull results, which give barriers with a multiplicative factor in the parameter. For example, Nesterov [26] proposes a barrier with parameter  $(3.08\sqrt{1+d} + 3.57)^2$ , and Freund et al. [17] use the result of Nesterov and Nemirovskii [29, Proposition 5.1.4] to propose a barrier with parameter no less than  $9.48(1 + d)$ .

**PROPOSITION 1.**  $\Gamma$  in Equation (85) is a  $(2 + d)$ -LHSCB for  $\mathcal{K}_{\text{MMD}}$  in Equation (80).

*Proof.* We show that  $\zeta$  in Equation (79) is  $(\mathbb{R}_{\geq}, 1)$ -compatible with the domain  $\mathcal{E}$ , in the sense of Nesterov and Nemirovskii [29, Definition 5.1.1]. This follows if (i)  $\zeta$  is  $C^3$ -smooth on  $\mathcal{E}$ , (ii)  $\zeta$  is concave with respect to  $\mathbb{R}_{\geq}$ , (iii) for any point  $\tilde{u} \in \text{int}(\mathcal{K}_{\text{MMD}})$  and direction  $\tilde{p} = (p, q, r) \in \mathbb{R} \times \mathbb{R} \times V$  satisfying  $v \pm q \geq 0$  and  $w \pm r \succeq 0$  it holds that:

$$\nabla^3 \zeta(\tilde{u})[\tilde{p}, \tilde{p}, \tilde{p}] \leq -3\nabla^2 \zeta(\tilde{u})[\tilde{p}, \tilde{p}]. \quad (87)$$

Suppose  $v \pm q \geq 0$  and  $w \pm r \succeq 0$ . As in Equation (43), we let  $\mu := \mu(v, w) = v^{-1}w \succ 0$  and  $\xi := v^{-1}(r - q\mu)$ . From Equation (44), the second and third order directional derivatives of  $\zeta$  at  $\tilde{u}$  in direction  $\tilde{p}$  are:

$$\nabla^2 \zeta(\tilde{u})[\tilde{p}, \tilde{p}] = -v\nabla^2 \varphi(\mu)[\xi, \xi], \quad (88a)$$

$$\nabla^3 \zeta(\tilde{u})[\tilde{p}, \tilde{p}, \tilde{p}] = -v\nabla^3 \varphi(\mu)[\xi, \xi, \xi] + 3q\nabla^2 \varphi(\mu)[\xi, \xi]. \quad (88b)$$

Since  $\varphi$  is convex and  $C^3$ -smooth on  $\text{int}(\mathcal{Q})$  by assumption, Equations (79) and (88) imply that  $\zeta$  is concave and  $C^3$ -smooth on  $\mathcal{E}$ . It remains to show that Equation (87) holds.

For  $t \geq 0$ , let:

$$a(t) := \mu + te \succ 0, \quad (89a)$$

$$\bar{a}(t) := a(t)^{-1/2} \succ 0, \quad (89b)$$

$$\hat{\xi}(t) := P(\bar{a}(t))\xi. \quad (89c)$$

By the integral representation result from Section 6.1 [24], there exists a positive measure  $\rho$  and  $\beta \geq 0$  such that the directional derivatives of  $\varphi$  are:

$$\nabla^2 \varphi(\mu)[\xi, \xi] \stackrel{83}{=} \beta \text{tr}(\xi^2) + \int_0^\infty t \text{tr}(\hat{\xi}(t)^2) d\rho(t), \quad (90a)$$

$$\nabla^3 \varphi(\mu)[\xi, \xi, \xi] \stackrel{84}{=} -2 \int_0^\infty t \text{tr}(\hat{\xi}(t)^3) d\rho(t). \quad (90b)$$

From Equations (88) and (90), the compatibility condition Equation (87) is equivalent to nonnegativity of:

$$-3\nabla^2\zeta(\tilde{u})[\tilde{p}, \tilde{p}] - \nabla^3\zeta(\tilde{u})[\tilde{p}, \tilde{p}, \tilde{p}] \quad (91a)$$

$$= 3(v-q)\beta \operatorname{tr}(\xi^2) + \int_0^\infty t(3(v-q) \operatorname{tr}(\hat{\xi}(t)^2) - 2v \operatorname{tr}(\hat{\xi}(t)^3)) \, d\rho(t). \quad (91b)$$

Since  $v \geq q$ , the first term in Equation (91b) is nonnegative. The second term (the integral) is nonnegative if for all  $t \geq 0$ , the following inner term is nonnegative:

$$3(v-q) \operatorname{tr}(\hat{\xi}(t)^2) - 2v \operatorname{tr}(\hat{\xi}(t)^3) = \langle \hat{\xi}(t)^2, 3(v-q)e - 2v\hat{\xi}(t) \rangle. \quad (92)$$

By self-duality of  $\mathcal{Q}$ , Equation (92) is nonnegative if  $3(v-q)e - 2v\hat{\xi}(t) \succeq 0$ , which we now prove. For  $t \geq 0$ , let:

$$b(t) := (1 - v^{-1}q)a(t) - \xi = v^{-1}(w - r) + t(1 - v^{-1}q)e. \quad (93)$$

Since  $w \succeq r$  and  $1 - v^{-1}q \geq 0$ , we have  $b(t) \succeq 0$ . Hence we have:

$$(1 - v^{-1}q)e - \hat{\xi}(t) \stackrel{93}{=} P(\bar{a}(t))b(t) \succeq 0, \quad (94)$$

since  $\bar{a}(t) \succ 0$  implies  $P(\bar{a}(t))$  is an automorphism on  $\mathcal{Q}$  (see Faraut and Koranyi [13, Page 48]). Therefore:

$$3(v-q)e - 2v\hat{\xi}(t) \stackrel{94}{\succeq} 3(v-q)e - 2v(1 - v^{-1}q)e = (v-q)e \succeq 0. \quad (95)$$

So Equation (92) is nonnegative, which implies the integral term in Equation (91b) is nonnegative.

Thus Equation (91b) is nonnegative, so Equation (87) holds and compatibility is proved. Now by Nesterov and Nemirovskii [29, Proposition 5.1.7],  $\Gamma$  is an LHSCB for  $\mathcal{K}_{\text{MMD}}$  with parameter  $2 + d$ .  $\square$

Note that the negative log-determinant cone, whose barrier function we examined in Section 5.4, is a special case of  $\mathcal{K}_{\text{MMD}}$  where  $h$  is the NegLog function from Table 1. When  $d = 1$  (i.e.  $\mathcal{Q}$  is just the nonnegative reals), this cone is equivalent to the standard three-dimensional exponential cone, and our barrier  $\Gamma$  matches the usual 3-LHSCB for this cone from Nesterov [28, Section 5.4.7, Example 2]. Furthermore for  $d = 1$ ,  $\mathcal{K}_{\text{MMD}}$  with NegEntropy is an invertible linear transformation of the exponential cone, and again our barrier  $\Gamma$  matches the 3-LHSCB derived through the linear transformation. When  $d = 1$  and  $h$  is the NegPower function,  $\mathcal{K}_{\text{MMD}}$  is equivalent to a special case of Nesterov [28, Section 5.4.7, Example 3] and  $\Gamma$  matches the proposed 3-LHSCB therein (see also case 3 of Hildebrand [21, Theorem 2]).

**7. Root-determinant cones.** In Section 7.1, we define the root-determinant cone  $\mathcal{K}_{\text{rtdet}}$ , which is the hypograph of the homogeneous nonseparable spectral root-determinant function. After expressing the derivatives of this function in Section 7.2, we prove that our barrier function  $\Gamma$  for  $\mathcal{K}_{\text{rtdet}}$  is an LHSCB in Section 7.3, and we derive easily-computable barrier oracles in Section 7.4.

**7.1. Cone definition.** Suppose  $\mathcal{Q}$  is a cone of squares of a Jordan algebra  $V$  with rank  $d$ . Let  $\varphi: \mathcal{Q} \rightarrow \mathbb{R}_{\geq}$  denote the root-determinant function (or the geometric mean of the eigenvalues):

$$\varphi(w) := \det(w)^{1/d} = \prod_{i \in [d]} w_i^{1/d}, \quad (96)$$

which is a concave, homogeneous nonseparable spectral function (see Section 3.1). We let  $\tilde{u} := (u, w) \in \mathcal{E} = \mathbb{R} \times \mathcal{Q}$ . The function  $\zeta: \mathcal{E} \rightarrow \mathbb{R}$  has the form:

$$\zeta(\tilde{u}) := \varphi(w) - u. \quad (97)$$

We define the root-determinant cone  $\mathcal{K}_{\text{rtdet}}$  and its dual cone as:

$$\mathcal{K}_{\text{rtdet}} := \{(u, w) \in \mathbb{R} \times \mathcal{Q} : u \leq \det(w)^{1/d}\}, \quad (98a)$$

$$\mathcal{K}_{\text{rtdet}}^* := \{(u, w) \in \mathbb{R}_{\leq} \times \mathcal{Q} : -d^{-1}u \leq \det(w)^{1/d}\}. \quad (98b)$$

We note that  $\mathcal{K}_{\text{rtdet}}$  is a hypograph modification of the epigraph cone  $\mathcal{K}_h$  in Equation (29), and it is a primitive proper cone.  $\mathcal{K}_{\text{rtdet}}^*$  can be derived by modifying the steps we use to derive  $\mathcal{K}_h^*$  in Equation (37) and using the convex conjugate of the negative root-determinant function.

**7.2. Derivatives of root-determinant.** We derive some expressions needed for our LHSCB proof in Section 7.3. Suppose  $w \succ 0$ . Since  $\varphi(w) = \exp(d^{-1} \log \det(w))$ , applying the chain rule and using Equation (26) gives us the gradient:

$$\nabla \varphi(w) = d^{-1} \varphi(w) w^{-1}. \quad (99)$$

Let  $r \in V$  and  $\hat{r} := P(w^{-1/2})r \in V$ . Using the product rule on Equation (99), we have the second order directional derivative:

$$\nabla^2 \varphi(w)[r] = d^{-2} \langle w^{-1}, r \rangle \nabla \varphi(w) + d^{-1} \varphi(w) \nabla_w(w^{-1})[r] \quad (100a)$$

$$\stackrel{24}{=} d^{-1} \varphi(w) \text{tr}(\hat{r}) w^{-1} - d^{-1} \varphi(w) P(w^{-1})r \quad (100b)$$

$$= d^{-1} \varphi(w) (d^{-1} \text{tr}(\hat{r}) w^{-1} - P(w^{-1})r) \quad (100c)$$

$$= d^{-1} \varphi(w) P(w^{-1/2}) (d^{-1} \text{tr}(\hat{r}) e - \hat{r}). \quad (100d)$$

Finally, using the product rule on Equation (100c), we have the third order directional derivative:

$$\nabla^3 \varphi(w)[r, r] = d^{-1} (d^{-1} \text{tr}(\hat{r}) w^{-1} - P(w^{-1})r, r) \nabla \varphi(w) + d^{-1} \varphi(w) ( \quad (101a)$$

$$d^{-1} (\langle w^{-1}, r \rangle \nabla_w (\langle w^{-1}, r \rangle) + \text{tr}(\hat{r}) \nabla_w(w^{-1})[r]) - \nabla_w(P(w^{-1})r)[r] ) \quad (101b)$$

$$\stackrel{25}{=} d^{-2} \varphi(w) (d^{-1} \text{tr}(\hat{r})^2 - \text{tr}(\hat{r}^2)) w^{-1} + d^{-1} \varphi(w) ( \quad (101c)$$

$$-2d^{-1} \text{tr}(\hat{r}) P(w^{-1})r + 2P(w^{-1/2})(P(w^{-1/2})r)^2 )$$

$$= d^{-1} \varphi(w) P(w^{-1/2}) (d^{-1} (d^{-1} \text{tr}(\hat{r})^2 - \text{tr}(\hat{r}^2)) e - 2d^{-1} \text{tr}(\hat{r}) \hat{r} + 2\hat{r}^2).$$

**7.3. Self-concordant barrier.** For  $\mathcal{K}_{\text{rtdet}}$ , the LHB  $\Gamma : \text{int}(\mathcal{K}_{\text{rtdet}}) \rightarrow \mathbb{R}$  from Equation (40) has the form:

$$\Gamma(\tilde{u}) := -\log(\zeta(\tilde{u})) - \log \det(w). \quad (102)$$

In Proposition 2 we show that  $\Gamma$  is self-concordant with parameter  $1 + d$ . Since the optimal barrier parameter for  $\mathcal{E}$  is  $d$ , our parameter cannot be reduced by more than one.

PROPOSITION 2.  $\Gamma$  in Equation (102) is a  $(1 + d)$ -LHSCB for  $\mathcal{K}_{\text{rtdet}}$  in Equation (98a).

*Proof.* Note  $\Psi(\tilde{u}) := -\log \det(w)$  is a  $d$ -LHSCB for  $\mathcal{E}$ . We show that  $\zeta$  in Equation (97) is  $(\mathbb{R}_{\geq}, 1)$ -compatible with the barrier  $\Psi$  in the sense of Nesterov and Nemirovskii [29, Definition 5.1.2]. Compatibility follows if (i)  $\zeta$  is  $C^3$ -smooth on  $\text{int}(\mathcal{E})$ , (ii) concave with respect to  $\mathbb{R}_{\geq}$ , (iii) for any point  $\tilde{u} \in \text{int}(\mathcal{K}_{\text{rtdet}})$  and direction  $\tilde{p} = (p, r) \in \mathbb{R} \times V$  it holds that:

$$\nabla^3 \zeta(\tilde{u})[\tilde{p}, \tilde{p}, \tilde{p}] \leq -3(\nabla^2 \Psi(\tilde{u})[\tilde{p}, \tilde{p}])^{1/2} \nabla^2 \zeta(\tilde{u})[\tilde{p}, \tilde{p}]. \quad (103)$$

Suppose  $\tilde{u} \in \text{int}(\mathcal{K}_{\text{rtdet}})$ . From Equation (97), we have:

$$\nabla^2 \zeta(\tilde{u})[\tilde{p}, \tilde{p}] = \nabla^2 \varphi(w)[r, r], \quad (104a)$$

$$\nabla^3 \zeta(\tilde{u})[\tilde{p}, \tilde{p}, \tilde{p}] = \nabla^3 \varphi(w)[r, r, r]. \quad (104b)$$

Since  $\varphi$  is concave and  $C^3$ -smooth on  $\text{int}(\mathcal{Q})$ , Equation (104) implies  $\zeta$  is concave and  $C^3$ -smooth on  $\text{int}(\mathcal{E})$ . It remains to show that Equation (103) holds.

Let  $\sigma \in \mathbb{R}^d$  be the eigenvalues of  $\hat{r} := P(w^{-1/2})r$ . Then:

$$(\nabla^2 \Psi(\tilde{u})[\tilde{p}, \tilde{p}])^{1/2} \stackrel{27a}{=} \text{tr}(\hat{r}^2)^{1/2} = \|\sigma\|. \quad (105)$$

Let  $m_k := d^{-1} \text{tr}(\hat{r}^k), \forall k \in \llbracket 3 \rrbracket$ , and let  $\delta_i := \sigma_i - m_1, \forall i \in \llbracket d \rrbracket$ . By the formulae for variance and skewness, we have:

$$m_2 - m_1^2 = d^{-1} \sum_{i \in \llbracket d \rrbracket} \delta_i^2, \quad (106a)$$

$$m_3 - 3m_1m_2 + 2m_1^3 = d^{-1} \sum_{i \in \llbracket d \rrbracket} \delta_i^3. \quad (106b)$$

For convenience, let  $\varphi := \varphi(w) > 0$  be a constant. We have:

$$\nabla^2 \varphi(w)[r, r] \stackrel{100}{=} d^{-1} \varphi \langle P(w^{-1/2})(d^{-1} \text{tr}(\hat{r})e - \hat{r}), r \rangle \quad (107a)$$

$$= -\varphi(d^{-1} \text{tr}(\hat{r}^2) - d^{-2} \text{tr}(\hat{r})^2) \quad (107b)$$

$$= -\varphi(m_2 - m_1^2) \quad (107c)$$

$$\stackrel{106a}{=} -d^{-1} \varphi \sum_{i \in \llbracket d \rrbracket} \delta_i^2 \leq 0. \quad (107d)$$

Similarly:

$$\nabla^3 \varphi(w)[r, r, r] \stackrel{101}{=} d^{-1} \varphi \langle d^{-1}(d^{-1} \text{tr}(\hat{r})^2 - \text{tr}(\hat{r}^2))e - 2d^{-1} \text{tr}(\hat{r})\hat{r} + 2\hat{r}^2, \hat{r} \rangle \quad (108a)$$

$$= d^{-1} \varphi(d^{-1}(d^{-1} \text{tr}(\hat{r})^2 - \text{tr}(\hat{r}^2)) \text{tr}(\hat{r}) - 2d^{-1} \text{tr}(\hat{r}) \text{tr}(\hat{r}^2) + 2 \text{tr}(\hat{r}^3)) \quad (108b)$$

$$= \varphi(m_1^3 - 3m_1m_2 + 2m_3) \quad (108c)$$

$$= \varphi(3m_1(m_2 - m_1^2) + 2(m_3 - 3m_1m_2 + 2m_1^3)) \quad (108d)$$

$$\stackrel{106a}{=} d^{-1} \varphi \sum_{i \in \llbracket d \rrbracket} (3m_1\delta_i^2 + 2\delta_i^3) \quad (108e)$$

$$= d^{-1} \varphi \sum_{i \in \llbracket d \rrbracket} \delta_i^2 (m_1 + 2\sigma_i). \quad (108f)$$

Finally, using Equations (105), (107) and (108) the compatibility condition Equation (103) is equivalent to nonnegativity of:

$$-\nabla^3 \zeta(\tilde{u})[\tilde{p}, \tilde{p}, \tilde{p}] - 3(\nabla^2 \Psi(\tilde{u})[\tilde{p}, \tilde{p}])^{1/2} \nabla^2 \zeta(\tilde{u})[\tilde{p}, \tilde{p}] \quad (109a)$$

$$= -d^{-1} \varphi \sum_{i \in \llbracket d \rrbracket} \delta_i^2 (m_1 + 2\sigma_i) + 3\|\sigma\| d^{-1} \varphi \sum_{i \in \llbracket d \rrbracket} \delta_i^2 \quad (109b)$$

$$= d^{-1} \varphi \sum_{i \in \llbracket d \rrbracket} \delta_i^2 (\|\sigma\| - m_1 + 2(\|\sigma\| - \sigma_i)). \quad (109c)$$

Clearly,  $d^{-1} \varphi \delta_i^2 \geq 0$  and  $\sigma_i \leq \|\sigma\|$  for all  $i \in \llbracket d \rrbracket$ . We have  $m_1 \leq d^{-1} \|\sigma\|_1 \leq d^{-1/2} \|\sigma\| \leq \|\sigma\|$ . Hence Equation (109c) is nonnegative.

Thus Equation (103) holds and compatibility is proved. Now by Nesterov and Nemirovskii [29, Proposition 5.1.7],  $\Gamma$  is a  $(1+d)$ -LHSCB for  $\mathcal{K}_{\text{rtdet}}$ .  $\square$

**7.4. Evaluating barrier oracles.** Using the derivatives of  $\varphi$  from Section 7.2, we derive easily-computable oracles for the LHSCB Equation (102). In Section 8.1, we describe our implementations of these oracles in Hypatia and link to the high-level Julia code.

Let  $\tilde{u} \in \text{int}(\mathcal{K}_{\text{rtdet}})$  and  $\tilde{p} = (p, r) \in \mathbb{R} \times V$ . For convenience, let  $\varphi := \varphi(w) > 0$  be a constant. We define the scalar constants:

$$\eta := d^{-1}\varphi\zeta^{-1}, \quad (110a)$$

$$\theta := 1 + \eta, \quad (110b)$$

$$\chi := -\zeta^{-1}p + \eta \text{tr}(\hat{r}), \quad (110c)$$

$$\tau := \chi - d^{-1} \text{tr}(\hat{r}), \quad (110d)$$

$$v := \text{tr}(\hat{r}^2) - d^{-1} \text{tr}(\hat{r})^2. \quad (110e)$$

Note that:

$$\nabla_u(\zeta^{-1}) = \zeta^{-2}, \quad (111a)$$

$$\nabla_u\eta = \zeta^{-1}\eta, \quad (111b)$$

$$\nabla_u\chi = \zeta^{-1}\chi, \quad (111c)$$

$$\nabla_w(\zeta^{-1}) = -\zeta^{-2}\nabla\varphi(w) \stackrel{99}{=} -\zeta^{-1}\eta w^{-1}, \quad (111d)$$

$$\nabla_w\eta = \eta(d^{-1} - \eta)w^{-1}, \quad (111e)$$

$$\nabla_w\chi \stackrel{24}{=} \zeta^{-1}\eta p w^{-1} + \eta(d^{-1} - \eta) \text{tr}(\hat{r})w^{-1} - \eta P(w^{-1})r \quad (111f)$$

$$= -\eta(\tau w^{-1} + P(w^{-1})r). \quad (111g)$$

The gradient of  $\Gamma$  in Equation (102) is:

$$g_u = \zeta^{-1}, \quad (112a)$$

$$g_w = -\zeta^{-1}\nabla\varphi(w) - w^{-1} \quad (112b)$$

$$= -\theta w^{-1}. \quad (112c)$$

Differentiating Equation (112), the Hessian product is:

$$H_u = \nabla_u g_u p + \nabla_u g_w[r] \quad (113a)$$

$$= \zeta^{-2}p - \zeta^{-1}\eta \text{tr}(\hat{r}) \quad (113b)$$

$$= -\zeta^{-1}\chi, \quad (113c)$$

$$H_w = \nabla_w g_u p + \nabla_w g_w[r] \quad (113d)$$

$$\stackrel{24}{=} -\zeta^{-1}\eta p w^{-1} - \text{tr}(\hat{r})\eta(d^{-1} - \eta)w^{-1} + \theta P(w^{-1})r \quad (113e)$$

$$= P(w^{-1/2})(\eta\tau e + \theta\hat{r}). \quad (113f)$$

Differentiating Equation (113), the third order directional derivative is:

$$T_u = \nabla_u H_u p + \nabla_u H_w[r] \quad (114a)$$

$$= -2\zeta^{-2}p\chi + \zeta^{-1}\eta(\tau \text{tr}(\hat{r}) + \text{tr}(\hat{r}^2)) + \zeta^{-1}\eta \text{tr}(\hat{r})\chi \quad (114b)$$

$$= \zeta^{-1}(2\chi^2 + \eta v), \quad (114c)$$

$$T_w = \nabla_w H_u p + \nabla_w H_w[r] \quad (114d)$$

$$\stackrel{25}{=} \zeta^{-1}\eta p \chi w^{-1} + \zeta^{-1}\eta p(\tau w^{-1} + P(w^{-1})r) - \eta\tau P(w^{-1})r + \text{tr}(\hat{r})\tau\eta(d^{-1} - \eta)w^{-1} + \eta(-\eta(\tau \text{tr}(\hat{r}) + \text{tr}(\hat{r}^2)) + d^{-1} \text{tr}(\hat{r}^2))w^{-1} + \eta(d^{-1} - \eta) \text{tr}(\hat{r})P(w^{-1})r - 2\theta P(w^{-1/2})\hat{r}^2 \quad (114e)$$

$$= \eta(-\chi\tau + \zeta^{-1}p\chi + (d^{-1} - \eta)(\text{tr}(\hat{r})\tau + \text{tr}(\hat{r}^2)))w^{-1} + \eta(-\tau + \zeta^{-1}p + (d^{-1} - \eta) \text{tr}(\hat{r}))P(w^{-1})r - 2\theta P(w^{-1/2})\hat{r}^2 \quad (114f)$$

$$= P(w^{-1/2})(\eta(-2\chi\tau + (d^{-1} - \eta)v)e - 2\eta\tau\hat{r} - 2\theta\hat{r}^2). \quad (114g)$$

In Lemma 1 below, we give a closed-form inverse Hessian product operator. This operator Equation (115) a similar structure to the Hessian product operator Equation (113), except that it applies  $P(w^{1/2})$  instead of  $P(w^{-1/2})$ .

LEMMA 1. *Letting  $\check{r} := P(w^{1/2})r \in V$ , the inverse Hessian product is:*

$$\bar{H}_u = (\zeta^2 + d^{-1}\varphi^2)p + d^{-1}\varphi \operatorname{tr}(\check{r}), \quad (115a)$$

$$\bar{H}_w = P(w^{1/2})(d^{-1}(\varphi p + \eta\theta^{-1} \operatorname{tr}(\check{r}))e + \theta^{-1}\check{r}). \quad (115b)$$

*Proof.* Note that the Hessian operator Equation (113) is a positive definite linear operator, so it has a unique inverse linear operator. We show that  $(\nabla^2\Gamma)^{-1}(\nabla^2\Gamma[\tilde{p}]) = \tilde{p}$ . Into Equation (115), we substitute the values from Equation (113) i.e.  $p = H_u = -\zeta^{-1}\chi$  and  $r = H_w = P(w^{-1/2})(\eta\tau e + \theta\hat{r})$ . Since  $P(w^{1/2}) = P(w^{-1/2})^{-1}$ , we have:

$$\check{r} = P(w^{1/2})H_w = \eta\tau e + \theta\hat{r}, \quad (116a)$$

$$\operatorname{tr}(\check{r}) = d\eta\tau + \theta \operatorname{tr}(\hat{r}). \quad (116b)$$

We have:

$$\bar{H}_u = (\zeta^2 + d^{-1}\varphi^2)(-\zeta^{-1}\chi) + d^{-1}\varphi(d\eta\tau + \theta \operatorname{tr}(\hat{r})) \quad (117a)$$

$$= -\zeta\chi + \varphi\eta(\tau - \chi) + d^{-1}\varphi\theta \operatorname{tr}(\hat{r}) \quad (117b)$$

$$= -\zeta(\chi - \eta\operatorname{tr}(\hat{r})) \quad (117c)$$

$$= p, \quad (117d)$$

and:

$$\bar{H}_w = P(w^{1/2})(d^{-1}(-\varphi\zeta^{-1}\chi + \eta\theta^{-1}(d\eta\tau + \theta \operatorname{tr}(\hat{r})))e + \theta^{-1}(\eta\tau e + \theta\hat{r})) \quad (118a)$$

$$= P(w^{1/2})(\eta(-\tau + \eta\theta^{-1}\tau + \theta^{-1}\tau)e + \hat{r}) \quad (118b)$$

$$= P(w^{1/2})(\hat{r}) \quad (118c)$$

$$= r. \quad (118d)$$

Hence Equation (115) is the unique inverse operator of Equation (113).  $\square$

We note the polynomial-like structure of the oracles. In particular, the  $w$  components of the  $g$ ,  $H$ , and  $T$  oracles are computed by applying  $P(w^{-1/2})$  to a polynomial in  $\hat{r}$ , of degree zero for  $g$ , degree one for  $H$ , and degree two for  $T$ . Analogously to  $H$ , its inverse  $\bar{H}$  is computed by applying  $P(w^{1/2})$  to a polynomial of degree one in  $\check{r}$ . This structure leads to simple, efficient, and numerically-stable implementations. We also note the structural similarity (ignoring constants) between the  $u$  and  $w$  components of these oracles and those of the negative log-determinant barrier in Section 5.4. In both cases, the oracles can be computed without an explicit eigendecomposition if it is possible to apply  $P(w^{1/2})$  and  $P(w^{-1/2})$  directly. For example for  $V = \mathbb{S}^d$  and  $V = \mathbb{H}^d$ , only a Cholesky factorization of  $w$  is needed.

**8. Examples and computational testing.** In Coey et al. [8], we describe a performance-oriented PDIPM and we implement this algorithm in our open-source conic solver, Hypatia. In Section 8.1, we outline our implementations of the MMD cone and the log-determinant and root-determinant cones in Hypatia. Then in Sections 8.4.1 to 8.4.4, we present example problems with simple, natural formulations (NFs) in terms of these cones. Using techniques we describe in Section 8.2, we construct equivalent extended formulations (EFs) that can be recognized by MOSEK 9 or ECOS.

Our computational benchmarks follow the methodology we describe in Section 8.3 and show that Hypatia often solves the NFs much more efficiently than Hypatia, MOSEK, or ECOS solve the EFs. Finally, in Section 8.5, we exemplify the computational impact of efficient oracle procedures by comparing the performance of our closed-form inverse Hessian product formula in Equation (67) with that of a naive direct solve using the explicit Hessian matrix.



**8.1. Hypatia solver.** Hypatia’s generic cone interface (see Coey et al. [8, Section 3]) allows specifying a *vectorized* proper cone  $\mathcal{K} \subset \mathbb{R}^q$  for some dimension  $q$ . The nonnegative cone is already in vectorized format. For the real symmetric PSD cone  $\mathbb{S}_{\geq}^d$ , we use the standard *svec* transformation, which rescales and stores only the elements of the matrix triangle in a vector of dimension  $d(d+1)/2$ . For the complex Hermitian PSD cone  $\mathbb{H}_{\geq}^d$ , we perform a modified *svec* transformation to a  $d^2$ -dimensional vector, storing each real diagonal element as a single element and each complex off-diagonal element in the triangle as two (rescaled) consecutive real elements (the real part followed by the imaginary part). These transformations preserve inner products and the self-duality of cones of squares.

We adapt these transformations to enable vectorization of spectral cones. For example, for the epigraph-perspective cone  $\mathcal{K}_p$  in Equation (30), the vectorization is  $(u, v, \text{vec}(w)) \in \mathbb{R}^{2+q}$ , where  $\text{vec}(w) \in \mathbb{R}^q$  is the appropriate vectorization of  $w \in \mathcal{Q}$ . Fortunately, the dual cone of this vectorized cone is the analogous vectorization of the dual cone  $\mathcal{K}_p^*$  in Equation (33).

Hypatia’s primal general conic form (see Coey et al. [8, Section 4]) over a vector of variables  $x$  minimizes a linear function  $\langle c, x \rangle$  subject to linear equality constraints  $b - Ax = 0$  and conic constrains  $h - Gx \in \mathcal{K}$ . Here, the vectorized proper cone  $\mathcal{K}$  is specified as a Cartesian product  $\mathcal{K}_1 \times \dots \times \mathcal{K}_K$  of  $K$  proper cones recognized by Hypatia. This means for each  $\mathcal{K}_i$ , we must have either  $\mathcal{K}_i$  or  $\mathcal{K}_i^*$  defined through Hypatia’s cone interface.

For the domains  $\mathbb{R}^d$ ,  $\mathbb{S}^d$ , and  $\mathbb{H}^d$ , we implement vectorizations of the MMD cone  $\mathcal{K}_{\text{MMD}}$ , the log-determinant cone  $\mathcal{K}_{\log\det}$ , and the root-determinant cone  $\mathcal{K}_{\text{rt}\det}$ .<sup>1</sup> This allows the user to model with these cones or their dual cones in Hypatia. As we discuss at the end of Sections 5.4 and 7.4, for  $\mathcal{K}_{\log\det}$  and  $\mathcal{K}_{\text{rt}\det}$  the oracle procedures are quite specialized, for example we compute a Cholesky factorization rather than an eigendecomposition for the  $\mathbb{S}^d$  and  $\mathbb{H}^d$  domains.

For  $\mathcal{K}_{\text{MMD}}$ , we predefine the MMD functions in Table 1. The user can easily add support for any MMD function by implementing a list of simple oracles.<sup>2</sup> Recall that  $\mathcal{K}_{\text{MMD}}^*$  in Equation (81) is defined using the convex conjugate of the MMD function; in the examples below we suffix the MMD function names with *Conj* (e.g. *NegEntropyConj*) to indicate use of the convex conjugate function and  $\mathcal{K}_{\text{MMD}}^*$ . We write *NegLogdet* or *NegRtdet* for the negative log-determinant or negative root-determinant function, the epigraph of which we represent using  $\mathcal{K}_{\log\det}$  or  $\mathcal{K}_{\text{rt}\det}$ . In our examples, we choose not to use  $\mathcal{K}_{\text{rt}\det}^*$  or  $\mathcal{K}_{\log\det}^*$  (or equivalently,  $\mathcal{K}_{\text{MMD}}^*$  with *NegLogConj*), because these particular dual cones provide little additional modeling power over their primal cones.

**8.2. Natural and extended formulations.** To assess the computational value of our new cones and efficient oracles, we compare the performance of Hypatia on NFs over  $\mathcal{K}_{\text{MMD}}$ ,  $\mathcal{K}_{\log\det}$ , and  $\mathcal{K}_{\text{rt}\det}$  against that of other conic PDIPM solvers on equivalent EFs. ECOS [12] is another open-source conic PDIPM solver, but it only supports nonnegative, second-order, and three-dimensional exponential cones. MOSEK version 9 [25] is a commercial conic PDIPM solver that supports the same cones as ECOS as well as three-dimensional power cones and real symmetric PSD cones. We call the cones supported by MOSEK 9 the *standard cones*. Hypatia currently supports around two

<sup>1</sup> Our  $\mathcal{K}_{\log\det}$  implementation is for the hypograph of the log-determinant, rather than the epigraph of negative log-determinant considered in Section 5.4; this only requires minor changes to the oracle derivations and the LHSCB proof from Section 6.4 for validity.  $\mathcal{K}_{\text{MMD}}$  is implemented at <https://github.com/chriscoey/Hypatia.jl/tree/master/src/Cones/epipersepspectral>, and  $\mathcal{K}_{\log\det}$  and  $\mathcal{K}_{\text{rt}\det}$  are implemented at <https://github.com/chriscoey/Hypatia.jl/blob/master/src/Cones/hypoperlogdettri.jl> and <https://github.com/chriscoey/Hypatia.jl/blob/master/src/Cones/hyporootdettri.jl> for symmetric/Hermitian matrix domains and <https://github.com/chriscoey/Hypatia.jl/blob/master/src/Cones/hypoperlog.jl> and <https://github.com/chriscoey/Hypatia.jl/blob/master/src/Cones/hypogeomean.jl> for vector domains.

<sup>2</sup> The MMD function interface is described and implemented at <https://github.com/chriscoey/Hypatia.jl/blob/master/src/Cones/epipersepspectral/sepspectralfun.jl>.

dozen cone types (not counting dual cones), including  $\mathcal{K}_{\text{MMD}}$ ,  $\mathcal{K}_{\text{logdet}}$ ,  $\mathcal{K}_{\text{rtdet}}$ , and (generalizations of) the standard cones [8]. We generate the EFs automatically using formulation techniques discussed in Coey et al. [9] and Coey [7, Section 2.2.3].<sup>3</sup>

For  $V = \mathbb{R}^d$ , our EFs are constructed as follows. The EFs for *NegLog*, *NegEntropy*, *NegEntropyConj*, and *NegRtdet* use  $d$  exponential cones. The EFs for *NegSqrt* and *NegSqrtConj* use  $d$  three-dimensional second-order cones. The EFs for *Power*, *NegPower*, *PowerConj*, and *NegPowerConj* use  $d$  power cones. The example in Section 8.4.1 uses  $V = \mathbb{R}^d$ .

For  $V = \mathbb{S}^d$ , our EFs are constructed as follows. For most spectral functions, we adapt the EF from Ben-Tal and Nemirovski [3, Proposition 4.2.1], which requires using an EF from the  $V = \mathbb{R}^d$  case for the corresponding spectral function, and adding constraints on the sum of the  $i$  largest eigenvalues of a matrix for each  $i \in \llbracket d \rrbracket$ . This is a large formulation with many additional variables and PSD constraints. For *NegLog* and *NegRtdet*, we use a much simpler EF from Ben-Tal and Nemirovski [3, Example 18.d]. Since *NegSqrtConj* is a scaling of the inverse function (see Table 1), a Schur complement representation allows us to use an EF with one PSD cone constraint. For  $V = \mathbb{H}^d$ , we reformulate any complex PSD cone constraint to a real PSD cone constraint with twice the side dimension [25, Section 6.2.7]. The examples in Sections 8.4.2 and 8.4.3 use  $V = \mathbb{S}^d$  and the example in Section 8.4.4 uses  $V = \mathbb{H}^d$ .

**8.3. Computational methodology.** We perform all instance generation, computational experiments, and results analysis with Ubuntu 21.10, Julia 1.8.0-DEV.862, and Hypatia 0.5.3.<sup>4</sup> We use dedicated hardware with an AMD Ryzen 9 3950X 16-core processor (32 threads) and 128GB of RAM. For each example problem in Sections 8.4.1 to 8.4.4, we generate random instances of a range of sizes, using JuMP 0.21.10 and MathOptInterface v0.9.22. All instances are primal-dual feasible, so we expect solvers to return optimality certificates.

We use the conic PDIPM solvers in MOSEK version 9 and ECOS version 2.0.5 (with no features disabled). Hypatia uses a particular default algorithmic implementation that we describe in Coey et al. [8] (the combined directions method with the QR-Cholesky linear system procedure). We limit each solver to 16 threads and set a solve time limit of 1800 seconds. We set relative feasibility and optimality gap tolerances to  $10^{-7}$  and absolute optimality gap tolerances to  $10^{-10}$ .

For each instance, the relative difference between the objective values of the formulation/solver combinations that converge never exceeds  $10^{-4}$ . For each instance/formulation/solver combination that returns a solution, we measure the maximum violation  $\epsilon$  of the primal-dual optimality conditions in Coey et al. [9, Equation 19]. In Figures 2 to 5, we plot solve times in seconds against an instance size parameter, excluding solves for which  $\epsilon > 10^{-5}$ . Hypatia-NF (i.e. Hypatia solving the NF) is faster than any EF solver (Hypatia-EF, MOSEK-EF, ECOS-EF) across all instance sizes and spectral functions tested for each example, and always scales to larger sizes.

## 8.4. Examples and results.

**8.4.1. Nonparametric distribution estimation.** Suppose we have a random variable  $X$  taking values in the finite set  $\{\alpha_i\}_{i \in \llbracket d \rrbracket}$ . We seek a probability distribution  $\rho \in \mathbb{R}^d$  that minimizes a convex spectral function  $\varphi$ , given some prior information expressed with  $d/2$  linear equality constraints. Adapting Boyd et al. [5, Section 7.2], the problem is:

$$\min_{\rho \in \mathbb{R}^d} \varphi(\rho) : \tag{119a}$$

$$\text{tr}(\rho) = d, \tag{119b}$$

$$A\rho = b. \tag{119c}$$

<sup>3</sup> See [https://github.com/chriscoey/Hypatia.jl/blob/master/examples/spectral\\_functions\\_JuMP.jl](https://github.com/chriscoey/Hypatia.jl/blob/master/examples/spectral_functions_JuMP.jl).

<sup>4</sup> Benchmark scripts and instructions for reproducing and analyzing results are available at <https://github.com/chriscoey/Hypatia.jl/tree/master/benchmarks/natvsext>. A raw output CSV file and detailed results tables are at <https://github.com/chriscoey/Hypatia.jl/wiki>.

For four spectral functions  $\varphi$  on  $\mathbb{R}_{\geq}^d$  (with EFs that ECOS can recognize) and a range of sizes  $d$ , we build random instances of Equation (119). The solver timings are summarized in Figure 2. Note that for *NegRtdet*, no solve times are plotted for MOSEK-EF because the optimality condition violations  $\epsilon$  are too large (see Section 8.3); tightening MOSEK’s tolerance options improves these violations, though in either case MOSEK-EF is significantly slower than Hypatia-NF. We do not plot results for *NegLogdet* (the  $\mathcal{K}_{\log\det}$  formulation using the specialized oracles from Section 5.4) as they are nearly identical to the results for  $\mathcal{K}_{\text{MMD}}/\text{NegLog}$ ; however, the efficiency benefits of *NegLogdet* are realized for the matrix domain in Section 8.4.2.

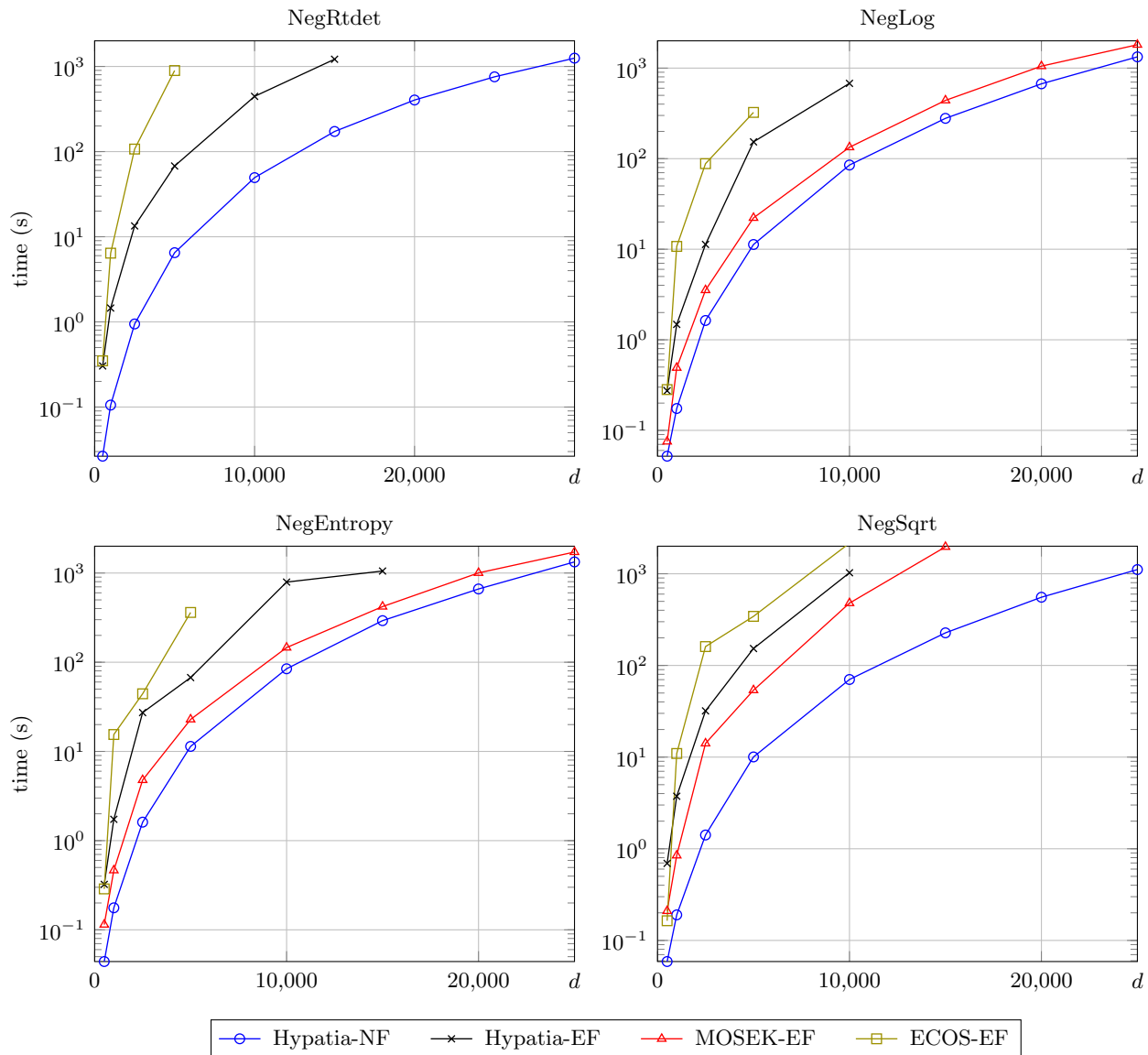


FIGURE 2. Nonparametric distribution estimation solver performance.

**8.4.2. Experiment design.** We formulate a continuous relaxation of the experiment design problem, similar to Boyd et al. [5, Section 7.5]. The variable  $\rho \in \mathbb{R}^{2d}$  is the number of trials to run for each of  $2d$  experiments that are useful for estimating a vector in  $\mathbb{R}^d$ . The experiments are

described by the columns of  $V \in \mathbb{R}^{d \times 2d}$  and we require that  $2d$  experiments are performed. We minimize a convex spectral function of the information matrix:

$$\min_{\rho \in \mathbb{R}^{2d}} \varphi(V \text{Diag}(\rho)V') : \quad (120a)$$

$$\text{tr}(\rho) = 2d, \quad (120b)$$

$$\rho \geq 0, \quad (120c)$$

where  $V'$  is the transpose of  $V$  and  $\text{Diag}(\rho)$  is the diagonal matrix of  $\rho$ . For four different  $\varphi$  on  $\mathbb{S}_{\geq}^d$  and various  $d$ , we build random instances of Equation (120). The solver timings are summarized in Figure 3. Since ECOS does not support  $\mathbb{S}_{\geq}^d$ , we only compare with MOSEK. The *Hypatia-NegLogdet* curve indicates that Hypatia with  $\mathcal{K}_{\log\text{det}}$  is somewhat more efficient than Hypatia with the equivalent  $\mathcal{K}_{\text{MMD}}/\text{NegLog}$  formulation; this is due to our oracle specializations in Section 5.4 and our implementation using a Cholesky factorization rather than an eigendecomposition.

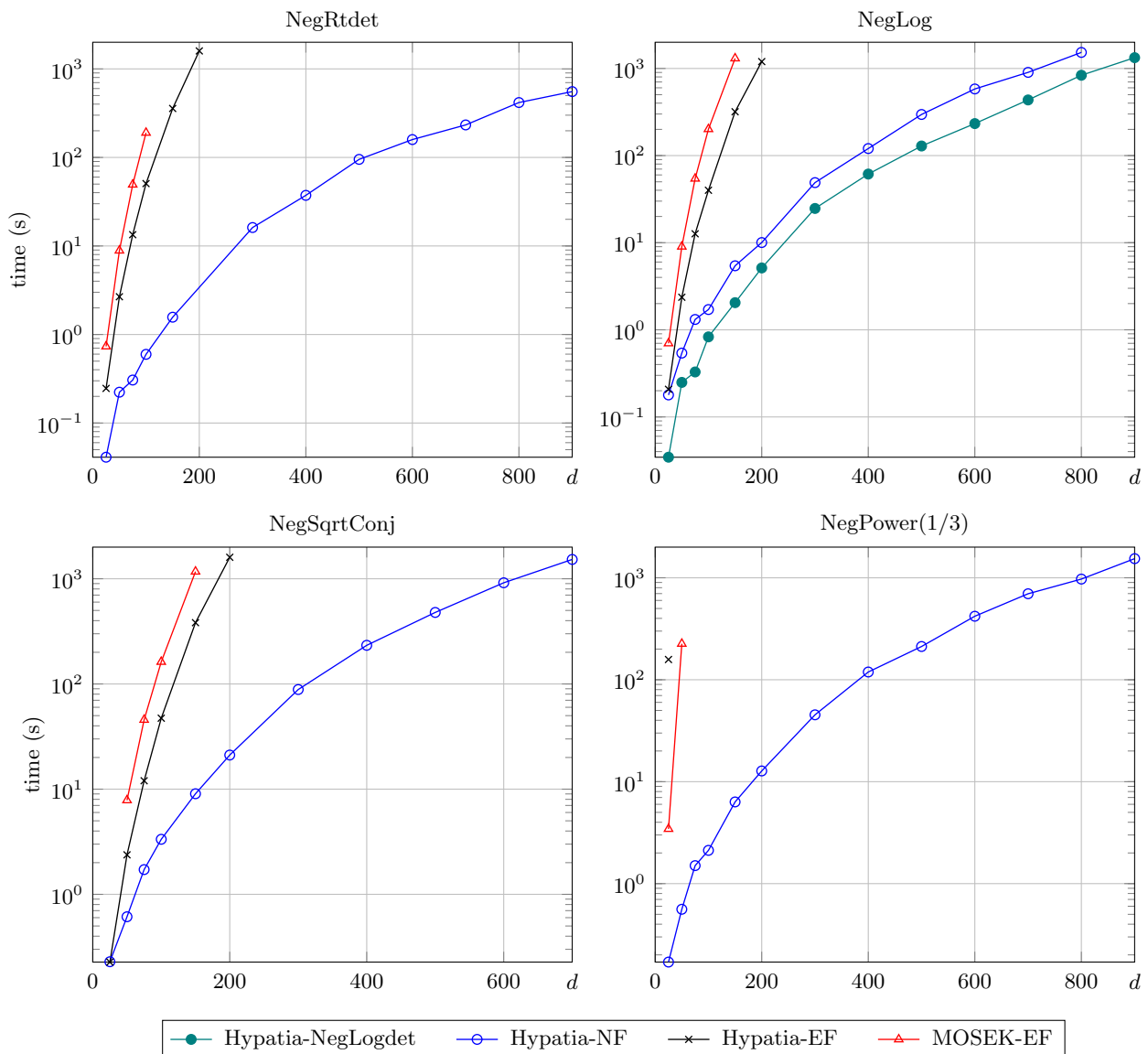


FIGURE 3. Experiment design solver performance.

**8.4.3. Central polynomial Gram matrix.** Suppose we have a polynomial of degree  $2k$  in  $m$  variables. Let  $L = \binom{m+k}{m}$  and  $U = \binom{m+2k}{m}$ , and let  $b \in \mathbb{R}^U$  be the monomial coefficients of the polynomial. We seek a Gram matrix  $\rho \in \mathbb{S}^L$  corresponding to  $b$  [34, Lemma 3.33] that minimizes a convex spectral function  $\varphi$ :

$$\min_{\rho \in \mathbb{S}^L} \varphi(\rho) : \tag{121a}$$

$$C \operatorname{vec}(\rho) = b, \tag{121b}$$

where the matrix  $C \in \mathbb{R}^{U \times L(L+1)/2}$  maps the Gram matrix to the (lower-dimensional) polynomial coefficient space. We build random instances of Equation (121), varying  $m \in \{1, 4\}$  and  $k$  (depending on  $m$ ). Recall from Table 1 that *ConjNegEntr* and *ConjPower-1.5* are defined on  $\mathbb{S}^d$ , but *NegEntr* and *MatPower12(1.5)* are only defined on  $\mathbb{S}_{\geq}^d$ , which implicitly requires that  $b$  be a sum of squares polynomial and hence globally nonnegative. The solver timings are summarized in Figure 4 (a log-log plot).

**8.4.4. Classical-quantum channel capacity.** We compute the capacity of a classical-quantum channel, adapting the formulation from Sutter et al. [41, Example 2.16] and Fawzi and Fawzi [14, Section 3.1]. The variable  $\rho \in \mathbb{R}^d$  is a probability distribution on the  $d$ -dimensional input alphabet. For  $i \in \llbracket d \rrbracket$ , let  $P_i \in \mathbb{H}_{\geq}^d$  be fixed density matrices satisfying  $\operatorname{tr}(P_i) = 1$ . Letting  $\varphi$  represent the trace of *NegEntropy* on  $\mathbb{H}_{\geq}^d$ , the formulation is:

$$\min_{\rho \in \mathbb{R}^d} \varphi\left(\sum_{i \in \llbracket d \rrbracket} \rho_i P_i\right) - \sum_{i \in \llbracket d \rrbracket} \rho_i \varphi(P_i) : \tag{122a}$$

$$\operatorname{tr}(\rho) = 1, \tag{122b}$$

$$\rho \geq 0. \tag{122c}$$

We generate random instances of Equation (122), varying  $d$ . The solver timings are summarized in Figure 5.

**8.5. Inverse Hessian product oracle.** To illustrate the importance of our efficient and numerically stable oracle procedures, we compare the performance of two different approaches to computing the inverse Hessian product oracle  $\bar{H}$  in Equation (41c) for  $\mathcal{K}_{\text{MMD}}$  cones. The naive approach is to compute the explicit Hessian matrix, perform a Cholesky factorization, and use a direct linear solve. Alternatively, we derive a closed-form formula for  $\bar{H}$  in Equation (67), since  $\mathcal{K}_{\text{MMD}}$  is a special case of  $\mathcal{K}_p$  with a separable spectral function. This formula is essentially as easy to compute as the Hessian product oracle  $H$  in Equation (50) (which does not use an explicit Hessian matrix). In Table 2, we compare the worst-case memory and time complexities for these procedures.

TABLE 2. Cone dimension and worst-case complexities for the two inverse Hessian product procedures.

$V$	$\dim(\mathcal{K}_{\text{MMD}})$	closed-form formula		factorize and solve	
		memory	time	memory	time
$\mathbb{R}^d$	$O(d)$	$O(d)$	$O(d)$	$O(d^2)$	$O(d^3)$
$\mathbb{S}^d$ or $\mathbb{H}^d$	$O(d^2)$	$O(d^2)$	$O(d^3)$	$O(d^4)$	$O(d^6)$

To compare the practical performance of these procedures, we perform computational experiments using Hypatia. We first solve NF instances of a range of sizes for the examples from Section 8.4.1 (with  $V = \mathbb{R}^d$ ) and Section 8.4.2 (with  $V = \mathbb{S}^d$ ), using  $\mathcal{K}_{\text{MMD}}$  with the *NegEntropy* function. For each instance, at Hypatia’s final PDIPM iterate, we take the direction  $\tilde{p} = g$  (i.e.

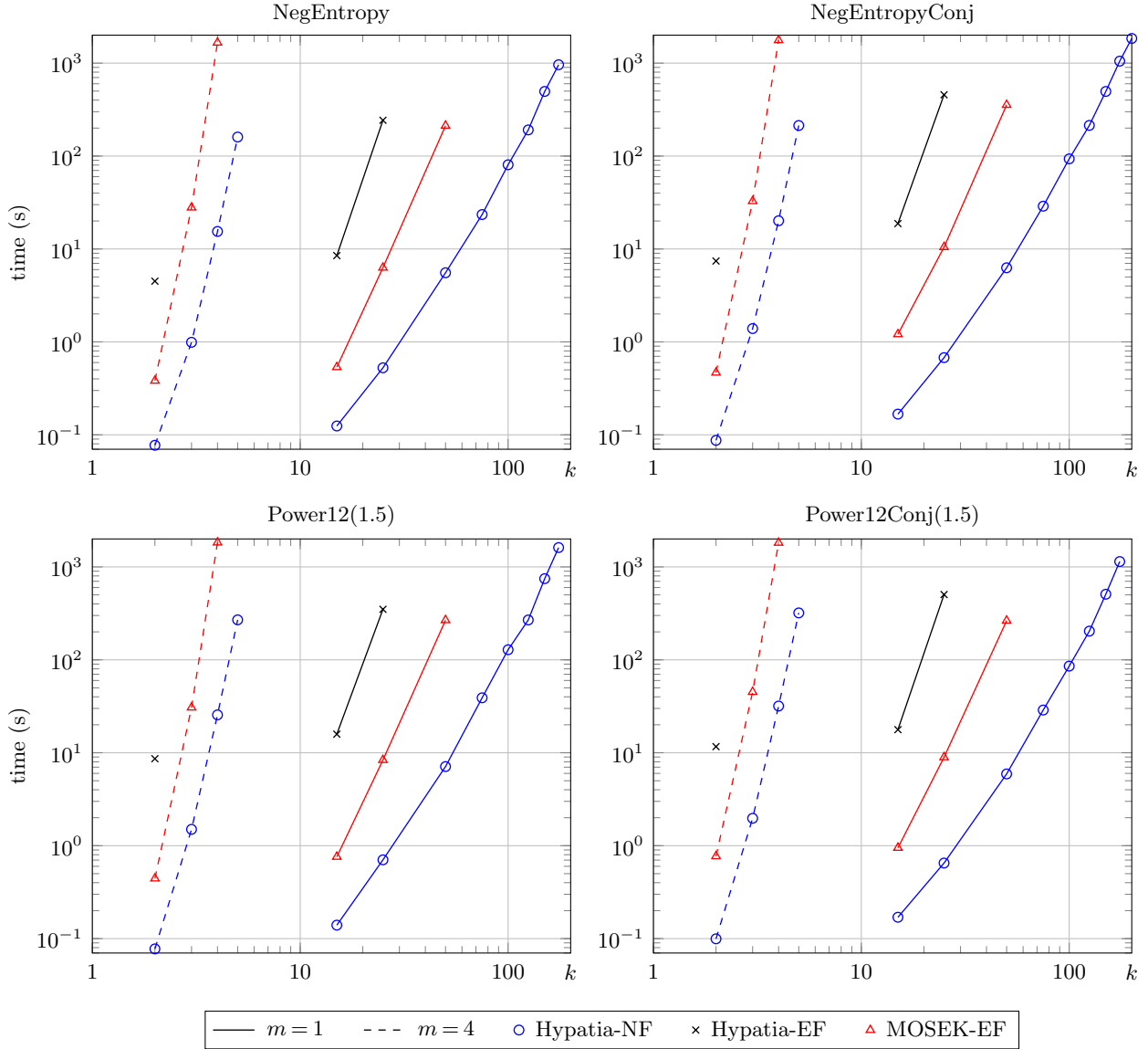


FIGURE 4. Central polynomial Gram matrix solver performance.

the gradient oracle in Equation (41a) at the iterate) and compute  $\bar{H}$  for this direction using both procedures. To measure the numerical accuracy of each procedure, we compute  $\epsilon := |1 - \nu^{-1} \langle \bar{H}, g \rangle|$ , which measures the violation of a particular identity [31, Equation 2.5] satisfied by a logarithmically homogeneous function such as the LHSCB  $\Gamma$  with parameter  $\nu = 2 + d$ . We also time each procedure, excluding Hessian memory allocation time for the factorization-based procedure.

Our results are displayed in Figure 6. The Cholesky factorization fails at  $d = 3000$  for the  $\mathbb{R}^d$  example and at  $d = 20, 50, 200$  for the  $\mathbb{S}^d$  example; when this occurs, Hypatia uses a Bunch-Kaufman factorization as a fallback (note Julia calls performant OpenBLAS routines for the Cholesky and Bunch-Kaufman factorizations). Note that we loosen the convergence tolerances specified in Section 8.3 by a factor of 100, so that the factorization-based procedure fails less often. These comparisons demonstrate that our closed-form formula generally allows computing  $\bar{H}$  faster and with greater numerical accuracy. Also, the closed-form procedure is much more memory efficient than the factorization-based procedure, as it never forms an explicit Hessian matrix.

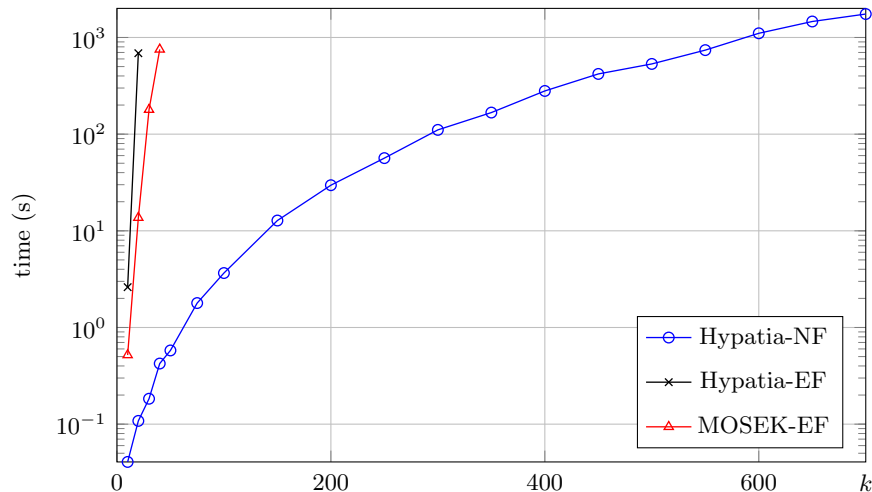


FIGURE 5. Classical-quantum channel capacity solver performance.

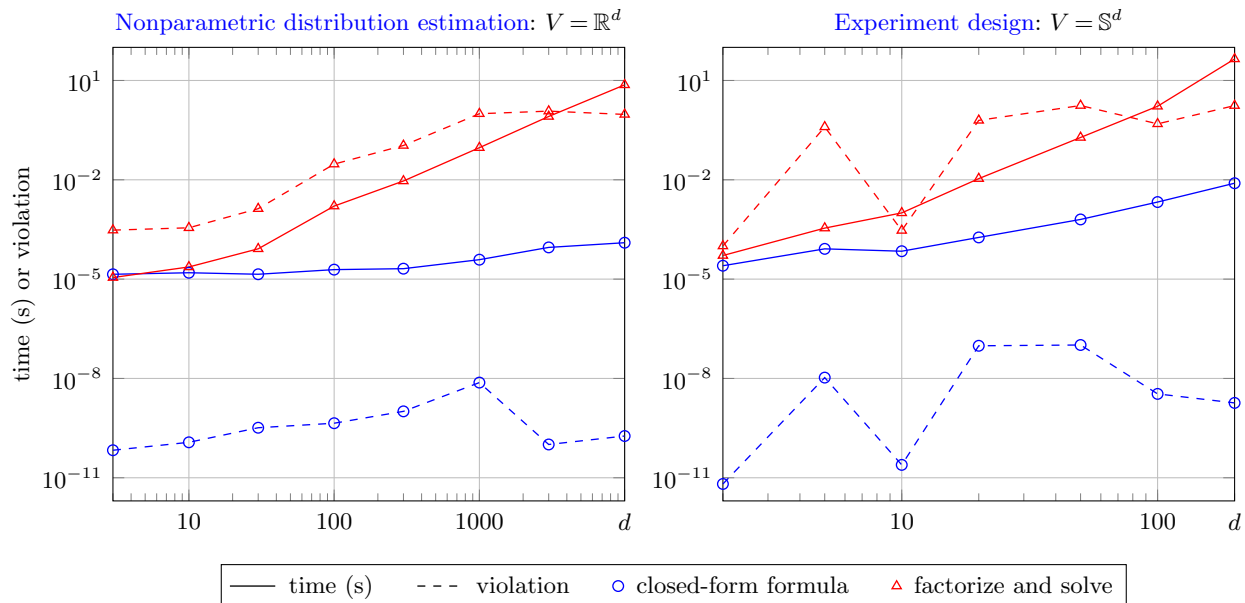


FIGURE 6. Speed and logarithmic homogeneity condition violation.

**Acknowledgments.** The authors thank Arkadi Nemirovski for helpful suggestions leading to barriers for root-determinant cones, and the anonymous referees for their comments and suggestions. This work has been partially funded by the National Science Foundation under grant OAC-1835443 and the Office of Naval Research under grant N00014-18-1-2079.

## References

- [1] Andersen M, Dahl J, Liu Z, Vandenberghe L, Sra S, Nowozin S, Wright S (2011) Interior-point methods for large-scale cone programming. Sra S, Wright SJ, Nowozin S, eds., *Optimization for Machine Learning*, volume 5583 (MIT Press Cambridge, MA).
- [2] Baes M (2007) Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras. *Linear Algebra and Its Applications* 422(2-3):664–700.
- [3] Ben-Tal A, Nemirovski A (2001) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. MOS-SIAM Series on Optimization (SIAM).

- [4] Borchers B (1999) CSDP, a C library for semidefinite programming. *Optimization Methods and Software* 11(1-4):613–623.
- [5] Boyd S, Boyd SP, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press).
- [6] Carlen E (2010) Trace inequalities and quantum entropy: an introductory course. Sims R, Ueltschi D, eds., *Entropy and the Quantum*, volume 529 of *Contemporary Mathematics*, 73–140 (American Mathematical Society).
- [7] Coey C (2022) *Interior point and outer approximation methods for conic optimization*. Ph.D. thesis, Massachusetts Institute of Technology.
- [8] Coey C, Kapelevich L, Vielma JP (2022) Performance enhancements for a generic conic interior point algorithm. *Mathematical Programming Computation* .
- [9] Coey C, Kapelevich L, Vielma JP (2022) Solving natural conic formulations with Hypatia. *INFORMS Journal on Computing* 34(5):2686–2699.
- [10] Davis C (1957) All convex invariant functions of Hermitian matrices. *Archiv der Mathematik* 8(4):276–278.
- [11] Deng CY (2011) A generalization of the Sherman–Morrison–Woodbury formula. *Applied Mathematics Letters* 24(9):1561–1564.
- [12] Domahidi A, Chu E, Boyd S (2013) ECOS: an SOCP solver for embedded systems. *2013 European Control Conference (ECC)*, 3071–3076 (IEEE).
- [13] Faraut J, Koranyi A (1994) Oxford mathematical monographs. Ball JM, Friedlander EM, Macdonald IG, Nirenberg L, Penrose R, Stuart JT, eds., *Analysis on Symmetric Cones* (Clarendon Press).
- [14] Fawzi H, Fawzi O (2018) Efficient optimization of the quantum relative entropy. *Journal of Physics A: Mathematical and Theoretical* 51(15):154003.
- [15] Faybusovich L, Tsuchiya T (2017) Matrix monotonicity and self-concordance: how to handle quantum entropy in optimization problems. *Optimization Letters* 11(8):1513–1526.
- [16] Faybusovich L, Zhou C (2021) Long-step path-following algorithm for quantum information theory: Some numerical aspects and applications. *Numerical Algebra, Control & Optimization* 12(2):445–467.
- [17] Freund RW, Jarre F, Schaible S (1996) On self-concordant barrier functions for conic hulls and fractional programming. *Mathematical Programming* 74(3):237–246.
- [18] Furuta T (2008) Concrete examples of operator monotone functions obtained by an elementary method without appealing to Löwner integral representation. *Linear Algebra and Its Applications* 429(5-6):972–980.
- [19] Grant M, Boyd S (2014) CVX: MATLAB software for disciplined convex programming, version 2.1. URL <http://cvxr.com/cvx/>, accessed 24-Aug-2022.
- [20] Grant M, Boyd S, Ye Y (2006) Disciplined convex programming. Liberti L, Maculan N, eds., *Global Optimization: From Theory to Implementation*, volume 84 of *Nonconvex Optimization and Its Applications*, 155–210 (Springer).
- [21] Hildebrand R (2014) Analytic formulas for complete hyperbolic affine spheres. *Beiträge zur Algebra und Geometrie/Contributions to Algebra and Geometry* 55(2):497–520.
- [22] Kwong MK (1989) Some results on matrix monotone functions. *Linear Algebra and Its Applications* 118:129–153.
- [23] Lasserre JB (1998) Homogeneous functions and conjugacy. *Journal of Convex Analysis* 5(2):397–404.
- [24] Löwner K (1934) Über monotone matrixfunktionen. *Mathematische Zeitschrift* 38(1):177–216.
- [25] MOSEK ApS (2020) Modeling Cookbook Release 3.3.0. URL <https://docs.mosek.com/modeling-cookbook/index.html>, accessed 24-Aug-2022.
- [26] Nesterov Y (2006) Constructing self-concordant barriers for convex cones. *CORE discussion paper* .
- [27] Nesterov Y (2012) Towards non-symmetric conic optimization. *Optimization Methods and Software* 27(4-5):893–917.



- [28] Nesterov Y (2018) *Lectures on Convex Optimization*, volume 137 of *Springer Optimization and Its Applications* (Springer Cham).
- [29] Nesterov Y, Nemirovskii A (1994) *Interior-Point Polynomial Algorithms in Convex Programming*. Studies in Applied Mathematics (Society for Industrial and Applied Mathematics).
- [30] Nesterov Y, Todd MJ, Ye Y (1999) Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Mathematical Programming* 84(2):227–267.
- [31] Nesterov YE, Todd MJ (1997) Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations Research* 22(1):1–42.
- [32] Papp D, Alizadeh F (2013) Semidefinite characterization of sum-of-squares cones in algebras. *SIAM Journal on Optimization* 23(3):1398–1423.
- [33] Papp D, Yıldız S (2017) On “A homogeneous interior-point algorithm for non-symmetric convex conic optimization”. *arXiv preprint arXiv:1712.00492* .
- [34] Parrilo PA (2012) Chapter 3: Polynomial optimization, sums of squares, and applications. Blekherman G, Parrilo PA, Thomas RR, eds., *Semidefinite Optimization and Convex Algebraic Geometry*, volume 13 of *MOS-SIAM Series on Optimization*, 47–157 (SIAM).
- [35] Permenter F, Friberg HA, Andersen ED (2017) Solving conic optimization problems via self-dual embedding and facial reduction: a unified approach. *SIAM Journal on Optimization* 27(3):1257–1282.
- [36] Rockafellar RT (1970) *Convex Analysis*, volume 28 of *Princeton Mathematical Series* (Princeton University Press).
- [37] Sendov HS (2007) The higher-order derivatives of spectral functions. *Linear Algebra and Its Applications* 424(1):240–281.
- [38] Serrano SA (2015) *Algorithms for unsymmetric cone optimization and an implementation for problems with the exponential cone*. Ph.D. thesis, Stanford University.
- [39] Skajaa A, Ye Y (2015) A homogeneous interior-point algorithm for nonsymmetric convex conic optimization. *Mathematical Programming* 150(2):391–422.
- [40] Sun D, Sun J (2008) Löwner’s operator and spectral functions in Euclidean Jordan algebras. *Mathematics of Operations Research* 33(2):421–445.
- [41] Sutter D, Sutter T, Esfahani PM, Renner R (2015) Efficient approximation of quantum channel capacities. *IEEE Transactions on Information Theory* 62(1):578–598.
- [42] Vieira MV (2007) *Jordan Algebraic approach to symmetric optimization*. Ph.D. thesis, TU Delft.
- [43] Vieira MV (2016) Derivatives of eigenvalues and Jordan frames. *Numerical Algebra, Control & Optimization* 6(2):115–126.
- [44] Yamashita M, Fujisawa K, Kojima M (2003) Implementation and evaluation of SDPA 6.0 (semidefinite programming algorithm 6.0). *Optimization Methods and Software* 18(4):491–505.
- [45] Zhang S (2004) A new self-dual embedding method for convex programming. *Journal of Global Optimization* 29(4):479–496.