A geodesic interior-point method for linear optimization over symmetric cones

Frank Permenter

January 18, 2023

Abstract

We develop a new interior-point method (IPM) for symmetric-cone optimization, a common generalization of linear, second-order-cone, and semidefinite programming. In contrast to classical IPMs, we update iterates with a *geodesic* of the *cone* instead of the *kernel* of the *linear* constraints. This approach yields a primal-dual-symmetric, scale-invariant, and line-search-free algorithm that uses just half the variables of a standard primal-dual IPM. With elementary arguments, we establish polynomial-time convergence matching the standard $\mathcal{O}(\sqrt{n})$ bound. Finally, we prove global convergence of a long-step variant and provide an implementation that supports all symmetric cones. For linear programming, our algorithms reduce to central-path tracking in the log domain.

Introduction

Let $\mathcal{J}$ denote a Euclidean Jordan algebra [8] of rank n with multiplication operator $\circ : \mathcal{J} \times \mathcal{J} \to \mathcal{J}$, identity $e \in \mathcal{J}$, and trace inner-product $\langle u, v \rangle := \operatorname{tr} u \circ v$. This paper considers the following primal-dual pair of linear optimization problems formulated over the *cone-of-squares* $\mathcal{K} := \{u \circ u : u \in \mathcal{J}\}$

$$\begin{array}{llcll} \text{minimize} & \langle s_0, x \rangle & \qquad & \text{minimize} & \langle x_0, s \rangle \\ \text{subject to} & x \in \mathcal{K} \cap (x_0 + \mathcal{L}) & & \text{subject to} & s \in \mathcal{K} \cap (s_0 + \mathcal{L}^\perp), \end{array} \tag{1}$$

where (x, s) denotes the primal-dual decision variables, $(x_0, s_0) \in \mathcal{J} \times \mathcal{J}$ denotes fixed parameters and $\mathcal{L} \subseteq \mathcal{J}$ is a linear subspace with orthogonal complement $\mathcal{L}^\perp \subseteq \mathcal{J}$. This standard form [10] generalizes linear, second-order-cone, and semidefinite programs [5], which typically present $x_0 + \mathcal{L}$ as the solution set of linear equations and the dual constraint $s \in \mathcal{K} \cap (s_0 + \mathcal{L}^\perp)$ as a cone inequality. It is also called a *symmetric cone* program given that $\mathcal{K}$ is both *self-dual* and *homogeneous* [8].

This paper contributes to the theory of interior-point methods (IPMs), widely used algorithms for solving (1). For IPMs, the following assumption is standard.

Assumption 1. *The primal and dual satisfy Slater's condition, i.e.,* $\operatorname{int}(\mathcal{K}) \cap (x_0 + \mathcal{L}) \neq \emptyset$ *and* $\operatorname{int}(\mathcal{K}) \cap (s_0 + \mathcal{L}^\perp) \neq \emptyset$*, where* $\operatorname{int}(\mathcal{K}) \subseteq \mathcal{J}$ *denotes the interior of* $\mathcal{K}$*.*

We make this assumption throughout.

Interior-point methods A pair (x, s) is optimal if it satisfies the constraints of (1) and the additional *complementary slackness* condition $x \circ s = 0$. Primal-dual IPMs solve a perturbation of these constraints defined by $\mu > 0$:

$$x \in \mathcal{K} \cap (x_0 + \mathcal{L}), \qquad s \in \mathcal{K} \cap (s_0 + \mathcal{L}^\perp), \qquad x \circ s = \mu e. \tag{2}$$

A unique solution $(\hat{x}(\mu), \hat{s}(\mu))$ of (2) exists for all $\mu > 0$ under Assumption 1 [9, Theorem 2.2]. The set of solutions $\{(\hat{x}(\mu), \hat{s}(\mu)) : \mu > 0\}$ is called the *central path*. Primal-dual IPMs follow the central path to an optimal solution of (1), i.e., they solve (2) while gradually reducing μ to zero.

While there are a variety of primal-dual IPMs (e.g., [10, 34, 30]), they share a common feature. Specifically, when initialized at feasible points, they all produce iterates $\{x_i, s_i\}_{i=1}^N$ satisfying

$$x_{i+1} - x_i \in \mathcal{L}, \qquad s_{i+1} - s_i \in \mathcal{L}^\perp, \tag{3}$$

which implies that $x_i \in x_0 + \mathcal{L}$ and $s_i \in s_0 + \mathcal{L}^\perp$ for all i. These iterations reduce violation of the complementarity constraint $x_i \circ s_i = \mu e$ and are interleaved with reductions in μ. For fixed μ_0 and μ_f, IPMs can move from $(\hat{x}(\mu_0), \hat{s}(\mu_0))$ to $(\hat{x}(\mu_f), \hat{s}(\mu_f))$ in $\mathcal{O}(\sqrt{n})$ iterations, where n denotes the rank of $\mathcal{K}$.

Geodesic interior-point methods This paper introduces *geodesic interior-point methods*, a family of IPMs that views $\mathcal{K}$ as a Riemannian manifold [28, 22, 20, 8]. As indicated by (3), classical IPMs update (x, s) inside *subspaces* that preserve the *affine* constraints. In contrast, geodesic IPMs will update (x, s) along *geodesic curves* that preserve the *complementarity* constraint $x \circ s = \mu e$. In other words, rather than enforcing (3), they will take

$$x_{i+1} = g_{x_i}(t_i), \qquad s_{i+1} = g_{s_i}(t_i), \tag{4}$$

where $t_i \in \mathbb{R}$ is a chosen "step-size" and $g_{x_i} : \mathbb{R} \to \mathcal{K}$ and $g_{s_i} : \mathbb{R} \to \mathcal{K}$ are chosen geodesics satisfying

$$g_{x_i}(0) = x_i \qquad g_{s_i}(0) = s_i, \qquad g_{x_i}(t) \circ g_{s_i}(t) = \mu e \qquad \forall t. \tag{5}$$

These iterations will reduce violation of the affine constraints $x_i \in x_0 + \mathcal{L}$ and $s_i \in s_0 + \mathcal{L}^\perp$ and, like (3), will be interleaved with reductions in μ. Like classical IPMs, we will show that geodesic IPMs can trace the central path in $\mathcal{O}(\sqrt{n})$ iterations, with essentially identical per-iteration complexity. We note that while the Riemannian geometry of $\mathcal{K}$ has been used to analyze the central path [31, 28, 15], develop gradient methods [2, 4], and solve non-convex problems [1], to our knowledge no central-path following algorithm for (1) is based on (4)-(5).

Geodesics of symmetric cones A geodesic is the shortest path between two points as measured by a particular integral cost (made precise in Section 1). For this reason, tracing a geodesic curve typically requires solving an ordinary differential equation (ODE) that expresses this integral's optimality conditions. For symmetric cones, however, geodesics can be expressed in closed form. This in turn provides simple formulae for the update (4). Indeed, for linear programming (LP), i.e., when $\mathcal{J} = \mathbb{R}^n$ and $u \circ v$ denotes elementwise multiplication, the update (4) will take the form

$$x_{i+1} = x_i \circ \exp(t_i d_i), \qquad s_{i+1} = s_i \circ \exp(-t_i d_i) \tag{6}$$

for some $d_i \in \mathbb{R}^n$, where $\exp(u)$ denotes elementwise exponentiation. For semidefinite programming, i.e., when $\mathcal{J}$ denotes the symmetric matrices and $U \circ V = \frac{1}{2}(UV + VU)$, it will take the form

$$X_{i+1} = X_i^{1/2} \exp(t_i D_i) X_i^{1/2}, \qquad S_{i+1} = S_i^{1/2} \exp(-t_i D_i) S_i^{1/2} \tag{7}$$

for symmetric D_i, where $\exp(U)$ and $U^{1/2}$ denote the matrix exponential and the symmetric square root. Similar exponential parametrizations hold for arbitrary symmetric cones. This will allow us to state and analyze algorithms based on (4) using basic properties of Euclidean Jordan algebras.

Log-domain interpretation The LP update (6) is equivalent to addition in the log-domain. In fact, for LP, the proposed algorithms essentially reduce to Newton's method on a log-domain formulation of the central-path conditions, i.e., to nonlinear equations $f(z) = 0$ induced by

$$\sqrt{\mu}\exp(z) \in x_0 + \mathcal{L}, \qquad \sqrt{\mu}\exp(-z) \in s_0 + \mathcal{L}^{\perp}.$$

We expand on this log-domain formulation in [33], but it, surprisingly, seems otherwise unanalyzed. In fact, to our knowledge, *all* interior-point methods for LP—in order to satisfy (3)—operate in the Euclidean space $\mathcal{J}$ as opposed to the log-domain; see, e.g., [38, 32].

Manifold optimization interpretation Our algorithms can be stated using basic concepts from Riemannian geometry and manifold optimization. In particular, the key steps reduce to selection of a *tangent vector* and evaluation of the *Riemannian exponential map* associated with $\mathcal{K}$. This viewpoint is crucial to generalizing the presented techniques to arbitrary convex cones and is discussed in Section 2.2.

Relationship with IPMs of Nesterov and Todd As we will show, our approach has intimate connections with that of Nesterov and Todd [30]. At a high-level, both yield an algorithm with $\mathcal{O}(\sqrt{n})$ complexity that is *scale-invariant* and *primal-dual symmetric* [41]. At a deeper level, we can interpret our algorithms as [30, Section 6] modified to perform geodesic updates. Crucially, this modification removes line searches and computation of a *scaling point*, which requires eigenvalue decomposition. It also reduces the number of variables, as we can represent both x and s using $w \in \mathcal{K}$ satisfying $(x, s) = \sqrt{\mu}(w, w^{-1})$. As a trade-off, we must evaluate the exponential function, but this can be done using an assortment of techniques [24]. Indeed, early computational experiments show that our implementation competes with `sdtp3` [40], a widely used solver based on the Nesterov-Todd approach.

Outline This paper is organized as follows. Section 1 briefly reviews the Riemannian geometry of $\mathcal{K}$ and provides a general formula for the geodesic update (4). Section 2 gives an IPM based on geodesic updates and establishes its $\mathcal{O}(\sqrt{n})$ complexity, log-domain and manifold optimization interpretations, scale invariance, and relation to the Nesterov-Todd method. We also show that selection of (g_x, g_s), like selection of a search direction in classical IPMs, reduces to orthogonal projection. Since this procedure conservatively tracks the central path, we refer to it as our *short-step* algorithm [43]. In Section 3, we study connections between geodesic distance and symmetrized Kullback-Leibler divergence, proving key results invoked in our short-step analysis. Leveraging this study, we describe a less conservative *long-step* algorithm in Section 4 and prove its global convergence and scale invariance; we also discuss efficient computation of geodesic updates, construction of feasible points, and other implementation issues. Finally, Section 5 contains computational results and links to an implementation.

1 Geodesic updates for symmetric cones

The interior of a symmetric cone $\mathcal{K}$, denoted $\operatorname{int}\mathcal{K}$, can be viewed as a Riemannian manifold by equipping each $u \in \operatorname{int}\mathcal{K}$ with a local norm $\|\cdot\|_u$ using the *quadratic representation* $Q(u) : \mathcal{J} \to \mathcal{J}$, the self-adjoint, linear map induced by $u \in \mathcal{J}$ via the relation $Q(u)v := 2u \circ (u \circ v) - (u \circ u) \circ v$. For $u \in \operatorname{int}\mathcal{K}$, the map $Q(u)$ is also positive definite, leading to the definition $\|v\|_u := \|Q(u)^{-1/2}v\|$, where $\|w\| := \sqrt{\langle w, w \rangle}$. The local norm $\| \cdot \|_u$ in turn induces an arc-length $L(\gamma)$ for smooth curves $\gamma : [0,1] \to \operatorname{int} \mathcal{K}$ via

$$L(\gamma) := \int_0^1 \|\gamma'(t)\|_{\gamma(t)} dt.$$

We note that this Riemannian geometry is studied by [3, Chapter 6] for the cone of positive definite matrices and by [19, 20, 22, 8] for general symmetric cones.

For $u, v \in \operatorname{int} \mathcal{K}$, let $\delta(u,v)$ denote the infimum of $L(\gamma)$ over smooth curves $\gamma(t)$ satisfying $\gamma(0) = u$ and $\gamma(1) = v$. A curve of length $\delta(u,v)$ connecting u and v is called a *geodesic*. Useful properties are collected below, including explicit formulae for $\delta(u,v)$ and geodesic curves. These formulae employ the square root $u^{1/2}$ and inversion u^{-1} operations of the algebra $\mathcal{J}$, as well as its log and exponential functions.

Lemma 1.1 (e.g., [19, 20])**.** *The following statements hold:*

(a) $\delta(u,v)$ *is a metric on* $\operatorname{int} \mathcal{K}$.

(b) *Given* $u, v \in \operatorname{int} \mathcal{K}$, *let* $d := \log Q(u^{-1/2})v$ *and* $g(t) := Q(u^{1/2}) \exp(td)$. *The curve* $g(t)$ *is a geodesic from* u *to* v, *i.e.,*

$$g(0) = u, \qquad g(1) = v, \qquad L(g) = \delta(u,v).$$

Further, $\delta(u,v) = \|d\|$.

(c) $\delta(u,v) = \delta(Tu, Tv)$ *for all* $u, v \in \operatorname{int} \mathcal{K}$ *and for any automorphism* T *of* $\mathcal{K}$, *i.e., for any invertible, linear map* $T : \mathcal{J} \to \mathcal{J}$ *satisfying* $\{Tz : z \in \mathcal{K}\} = \mathcal{K}$.

(d) $\delta(u^{-1}, v^{-1}) = \delta(u,v)$ *for all* $u, v \in \operatorname{int} \mathcal{K}$

In light of (a), the function $\delta(u,v)$ is called *geodesic distance*. The vector d in (b) denotes *normal coordinates* of v at the point u. In light of (c), the inner-product $\langle v, w \rangle_u := \langle v, Q(u)^{-1} w \rangle$ associated with $\| \cdot \|_u$ is called a *scale-invariant* or *affine-invariant* metric for $\mathcal{K}$. Item (d) shows inversion is an *isometry*. Note that $g(0) = u$ and $g(1) = v$ in (b) is immediate from the identities $Q(u^{1/2}) = Q(u)^{1/2}$, $Q(u^{-1}) = Q(u)^{-1}$, and $Q(u^{1/2})e = u$; see Appendix A.

1.1 Complementary geodesics

Given $x, s \in \operatorname{int} \mathcal{K}$ satisfying $x \circ s = \mu e$, we wish to parametrize geodesics g_x and g_s starting at x and s that satisfy $g_x(t) \circ g_s(t) = \mu e$ for all t. Combining Lemma 1.1 with properties of the quadratic representation $Q(u)$ provides a parametrization in terms of $d \in \mathcal{J}$. We also express (x,s) using the point $w \in \mathcal{K}$ satisfying $(x,s) = \sqrt{\mu}(w, w^{-1})$.

Proposition 1.1. *For* $d \in \mathcal{J}$, $w \in \operatorname{int} \mathcal{K}$ *and* $\mu > 0$, *let* $(x,s) = \sqrt{\mu}(w, w^{-1})$ *and let*

$$g_x(t) = \sqrt{\mu} Q(w^{1/2}) \exp(td), \qquad g_s(t) = \sqrt{\mu} Q(w^{-1/2}) \exp(-td). \tag{8}$$

Then, $g_x(t)$ *and* $g_s(t)$ *are geodesics satisfying* $g_x(0) = x$, $g_s(0) = s$, *and* $g_x(t) \circ g_s(t) = \mu e$ *for all* $t \in \mathbb{R}$.

Proof. The condition $g_x(t) \circ g_s(t) = \mu e$ holds from the identity $[Q(u)v]^{-1} = Q(u^{-1})v^{-1}$, whereas $g_x(0) = x$ and $g_s(0) = s$ hold from the identities $\exp(0) = e$ and $Q(u^{1/2})e = u$; see Appendix A. Finally, that $g_x(t)$ and $g_s(t)$ are geodesic follows from Lemma 1.1 (b) and the identity $Q((\sqrt{c}u)^{1/2}) = \sqrt{c} Q(u^{1/2})$. ∎

Procedure `shortstep`(w_0, μ_0, μ_f)

$w \leftarrow w_0$, $\mu \leftarrow \mu_0$
while $\mu > \mu_f$ **do**
    $\mu \leftarrow \frac{1}{k}\mu$
    for $i = 1, 2, \ldots, m$ **do**
        $d \leftarrow d_N(w, \mu)$
        $w \leftarrow Q(w^{1/2}) \exp(d)$
    end
end
return (w, μ)

$\mathcal{K}$	Definition	rank
$\mathbb{R}^n_+$	$\{x \in \mathbb{R}^n : x_i \geq 0\}$	n
$\mathbb{S}^n_+$	$\{X^2 : X \in \mathbb{R}^{n \times n}, X = X^T\}$	n
$\mathbb{L}^{m+1}$	$\{(x_0, x_1) \in \mathbb{R} \times \mathbb{R}^m : x_0 \geq \lVert x_1 \rVert\}$	2

$\mathcal{K}$	$\exp(d)$	$Q(w^{1/2}) \exp(d)$
$\mathbb{R}^n_+$	element-wise exp.	$\exp(\log w + d)$
$\mathbb{S}^n_+$	matrix exponential	$W^{1/2} \exp(D) W^{1/2}$
$\mathbb{L}^{m+1}$	replace eigenvalues with $\exp(d_0 \pm \lVert d_1 \rVert)$	$(2zz^T - (\det z)R) \exp(d)$

Figure 1: Short-step algorithm (left) with parameters (k, m) and implementation details (right) for linear programs ($\mathbb{R}^n_+$), second-order-cone programs ($\mathbb{L}^{m+1}$), and semidefinite programs ($\mathbb{S}^n_+$). In the $\mathbb{L}^{m+1}$ row, the map R denotes $(u_0, u_1) \mapsto (u_0, -u_1)$, while $z = w^{1/2}$ and $\det z = z_0^2 - \lVert z_1 \rVert^2$.

1.2 Newton direction

Proposition 1.1 shows that the geodesic update of $(x, s) = \sqrt{\mu}(w, w^{-1})$ introduced by (4) is performed by selecting $d \in \mathcal{J}$ and evaluating (8) at some t. Since our goal is to decrease violation of the affine constraints $x \in x_0 + \mathcal{L}$ and $s \in s_0 + \mathcal{L}^\perp$, a natural choice for d is the *Newton direction*, which we define by substituting (8) into the central-path conditions (2) with the linearizations $\exp(d) \approx e + d$ and $\exp(-d) \approx e - d$.

Definition 1.1. *(Newton Direction) For $w \in \operatorname{int} \mathcal{K}$ and $\mu > 0$, the* Newton direction $d_N(w, \mu)$ *is the unique $d \in \mathcal{J}$ satisfying*

$$Q(w^{1/2})(e + d) \in \frac{1}{\sqrt{\mu}} x_0 + \mathcal{L}, \qquad Q(w^{-1/2})(e - d) \in \frac{1}{\sqrt{\mu}} s_0 + \mathcal{L}^\perp.$$

Uniqueness of $d_N(w, \mu)$ is proven later by Proposition 2.4, but essentially follows from invertibility of $Q(w^{1/2})$ and $Q(w^{-1/2})$. Geodesic updates using $d_N(w, \mu)$ are the basis of algorithms given in Section 2 and Section 4.

2 Short-step algorithm

We give a procedure `shortstep` (Figure 1) for tracking the central path that employs the geodesic updates described by Proposition 1.1. Per this proposition, it updates $w \in \operatorname{int} \mathcal{K}$ satisfying

$$x = \sqrt{\mu} w, \qquad s = \sqrt{\mu} w^{-1} \tag{9}$$

via $w \leftarrow Q(w^{1/2}) \exp(td)$, or, equivalently, via $w^{-1} \leftarrow Q(w^{-1/2}) \exp(-td)$. At each iteration, it sets d equal to the Newton direction $d_N(w, \mu)$ and the step-size t equal to one, i.e., it performs a full *Newton step*. By construction, each iterate w induces via (9) variables x and s satisfying $x \circ s = \mu e$. Each Newton step in turn aims to reduce the violation of the affine constraints $x \in x_0 + \mathcal{L}$ and $s \in s_0 + \mathcal{L}^\perp$.

The inputs are an initial $w_0 \in \operatorname{int} \mathcal{K}$ and centering parameters $\mu_0, \mu_f \in \mathbb{R}$ satisfying $\mu_0 > \mu_f > 0$. The output is an approximation of the centered point $\hat{w}(\mu)$ for $\mu \leq \mu_f$, where $\hat{w}(\mu)$ denotes the unique point satisfying $\sqrt{\mu}(\hat{w}(\mu), \hat{w}(\mu)^{-1}) = (\hat{x}(\mu), \hat{s}(\mu))$ for $(\hat{x}(\mu), \hat{s}(\mu))$ on the central path.

Behavior depends on a parameter k that controls how much μ decreases at each *outer iteration* and a parameter m that denotes the number of *inner iterations*. Like short-step IPMs [43], our analysis will choose k conservatively and assume that $w_0 = \hat{w}(\mu_0)$. A more aggressive algorithm that supports arbitrary initialization by using damped updates ($t < 1$) appears in Section 4.

To establish convergence results, we need two lemmas whose proofs we postpone to Section 3. They employ the function $q : \mathbb{R} \to \mathbb{R}_+$ and its nonnegative inverse $q^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$ defined via:

$$q(u) := 2(\cosh(u) - 1), \qquad q^{-1}(u) := \cosh^{-1}(1 + \frac{1}{2}u).$$

In passing, we observe that $q(u) \geq u^2$ for all $u \in \mathbb{R}$ and $\sqrt{u} \geq q^{-1}(u)$ for all $u \geq 0$. The first lemma bounds the geodesic distance between two centered points $\hat{w}(\mu_0)$ and $\hat{w}(\mu_1)$ using the rank of $\mathcal{K}$ (denoted by n) and the ratio k of the centering parameters.

Lemma 2.1 (μ-update)**.** *Let $\mu, k > 0$. Then,* $\frac{1}{n}\delta\left(\hat{w}(\mu), \hat{w}(\frac{1}{k}\mu)\right)^2 \leq q(\frac{1}{2}\log k)$.

The second lemma establishes a region of quadratic convergence of the sequence $w_0, w_1, \ldots, w_m$ generated by Newton steps (inner iterations).

Lemma 2.2 (Centering)**.** *For $\mu > 0$ and $w_0 \in \operatorname{int}\mathcal{K}$, recursively define w_i via the iterations $w_{i+1} = Q(w_i^{1/2})\exp(d_N(w_i, \mu))$. If $\delta\left(w_0, \hat{w}(\mu)\right) \leq q^{-1}(\beta)$ for $0 \leq \beta \leq \frac{1}{2}$, then $\delta(w_i, \hat{w}(\mu))^2 \leq \beta^{2^i}$.*

Our main result follows from these lemmas, the triangle inequality for geodesic distance δ, and an inequality relating the scalar functions q^{-1} and $\sqrt{u}$.

Theorem 2.1 (Main Result)**.** *Let* `shortstep` *(Figure 1) have parameters (k, m) that satisfy, for some $\frac{1}{2} \geq \beta > 0$ and $q^{-1}(\beta) > \epsilon > 0$, the conditions*

$$\beta^{2^m} \leq \epsilon^2, \qquad \frac{1}{2}\log k = q^{-1}(\frac{1}{n}\zeta^2), \tag{10}$$

where $\zeta := q^{-1}(\beta) - \epsilon$. Then, the following statements hold for `shortstep` *given input $(\hat{w}(\mu_0), \mu_0, \mu_f)$:*

(a) At most $m\lceil c^{-1}\sqrt{n}\log\frac{\mu_0}{\mu_f}\rceil$ Newton steps execute, where $c := 2q^{-1}(\zeta^2)$.

(b) The output (w, μ) satisfies $\delta(w, \hat{w}(\mu)) \leq \epsilon$ and $\mu \leq \mu_f$. Further,

$$\delta(\sqrt{\mu}w, \hat{x}(\mu)) \leq \epsilon, \qquad \delta(\sqrt{\mu}w^{-1}, \hat{s}(\mu)) \leq \epsilon,$$

where $(\hat{x}(\mu), \hat{s}(\mu))$ denotes the solution to the central-path conditions (2)*.*

Proof. Let $\gamma = (\log k)^{-1}\log\frac{\mu_0}{\mu_f}$. The number of outer iterations is at most $\lceil\gamma\rceil$. To upper bound $(\log k)^{-1}$, we first note that for any $a, b \in \mathbb{R}$ satisfying $0 \leq a \leq b$,

$$q^{-1}(a) \geq \frac{q^{-1}(b)}{\sqrt{b}}\sqrt{a},$$

since $q^{-1}(u)/\sqrt{u}$ is a decreasing function. Setting $a = \frac{1}{n}\zeta^2$ and $b = \zeta^2$ and using (10) gives

$$\frac{1}{2}\log k = q^{-1}(\frac{1}{n}\zeta^2) \geq \frac{q^{-1}(\zeta^2)}{\zeta}\frac{\zeta}{\sqrt{n}} = \frac{q^{-1}(\zeta^2)}{\sqrt{n}}.$$

Hence, $(\log k)^{-1} \leq c^{-1}\sqrt{n}$ for $c = 2q^{-1}(\zeta^2)$, proving the first statement.

We prove the next statement using induction on outer iterations, observing first that

$$\delta(\hat{w}(\mu), \hat{w}(k^{-1}\mu))^2 \leq n[q(\frac{1}{2}\log k)] = n\frac{1}{n}\zeta^2 = (q^{-1}(\beta) - \epsilon)^2 \tag{11}$$

by Lemma 2.1 and our choice of k. Now let w_i^μ denote w at the end of inner iteration i for the current μ and let $\mu' := k^{-1}\mu$. Make the inductive hypothesis that $\delta(w_m^\mu, \hat{w}(\mu)) \leq \epsilon$. Then,

$$\delta(w_m^\mu, \hat{w}(\mu')) \leq \delta(w_m^\mu, \hat{w}(\mu)) + \delta(\hat{w}(\mu), \hat{w}(\mu')) \leq \epsilon + q^{-1}(\beta) - \epsilon = q^{-1}(\beta),$$

where the second inequality follows from (11) and the first is the triangle inequality (Lemma 1.1(a)). This allows us to update μ to μ', restart inner iterations at $w_0^{\mu'} = w_m^\mu$, and use Lemma 2.2 to conclude that

$$\delta(w_m^{\mu'}, \hat{w}(\mu'))^2 \leq \beta^{2^m} \leq \epsilon^2,$$

where the second inequality follows from our choice of m. Hence we've show if $\delta(w_m^\mu, \hat{w}(\mu)) \leq \epsilon$, then $\delta(w_m^{\mu'}, \hat{w}(\mu')) \leq \epsilon$. The base case holds by identical argument using the assumption that $w_0 = \hat{w}(\mu_0)$. Hence, $\delta(w, \hat{w}(\mu)) \leq \epsilon$ holds at termination. That $\delta(\sqrt{\mu}w, \sqrt{\mu}\hat{w}(\mu)) \leq \epsilon$ and $\delta(\sqrt{\mu}w^{-1}, \sqrt{\mu}\hat{w}^{-1}(\mu)) \leq \epsilon$ follows from the invariance of δ under rescaling and inversion; see Lemma 1.1(c)-(d). □

The remainder of this section gives other properties of `shortstep`, namely, log-domain and manifold optimization interpretations, an orthogonal decomposition of the Newton direction, and scale invariance. We also discuss connections with an algorithm of Nesterov and Todd.

2.1 Log-domain interpretation

Suppose that (1) is a primal-dual pair of *linear programs*, i.e., that $\mathcal{K} = \mathbb{R}^n_+$. Under this assumption, the algebra $\mathcal{J}$ is associative. Hence, geodesic distance simplifies to $\delta(u, v) = \|\log u - \log v\|$, and the geodesic update $w \leftarrow Q(w^{1/2})\exp(d)$ satisfies

$$\log\left(Q(w^{1/2})\exp(d)\right) = \log(w \circ \exp(d)) = \log(w) + d, \tag{12}$$

i.e., it reduces to addition in the log domain. The Newton direction $d_N(w, \mu)$ also has a log-domain interpretation: it is precisely the direction one obtains by linearizing $x(z) := \sqrt{\mu}\exp z$ and $s(z) := \sqrt{\mu}\exp(-z)$ at $z = \log w$ and substituting into the central-path conditions (2).

Proposition 2.1. *Let $\mathcal{J}$ be associative. For $\mu > 0$ and $w \in \operatorname{int}\mathcal{K}$, let $d = d_N(w, \mu)$. Then,*

$$\exp(z) + J(z)d \in \frac{1}{\sqrt{\mu}}x_0 + \mathcal{L}, \qquad \exp(-z) - J(-z)d \in \frac{1}{\sqrt{\mu}}s_0 + \mathcal{L}^\perp,$$

where $z = \log w$ and $J(z) : \mathcal{J} \to \mathcal{J}$ is the Jacobian of $\exp(z)$.

Proof. Under our associativity assumption, we observe that $J(z)d = \exp(z) \circ d$ and

$$Q(w^{1/2})(e + d) = w + w \circ d, \qquad Q(w^{-1/2})(e - d) = w^{-1} - w^{-1} \circ d.$$

Substituting $w = \exp(z)$ and $w^{-1} = \exp(-z)$ and using Definition 1.1 proves the claim. □

In total, we can reinterpret the inner iterations of `shortstep` as simply Newton's method applied to the central-path conditions in the log domain. We elaborate on this interpretation (and extend it to quadratic optimization) in the paper [33].

Observe that when $\mathcal{J}$ is *not* associative, this interpretation fails because the identity (12) fails. For semidefinite programming, failure of (12) reduces to the fact that for matrices $W \succ 0$ and D,

$$\log\left(W^{1/2}\exp(D)W^{1/2}\right) \neq \log(W) + D,$$

since, in general, $\exp(A+B) \neq \exp(A)\exp(B)$ for the matrix exponential.

2.2 Manifold optimization interpretation

Geodesic updates can be alternatively described using the *Riemannian exponential map* of $\operatorname{int}\mathcal{K}$. This function, denoted $\operatorname{Exp}_u : \mathcal{J} \to \operatorname{int}\mathcal{K}$, maps *tangent vectors* $v \in \mathcal{J}$ to points on geodesics passing through $u \in \operatorname{int}\mathcal{K}$. Precisely, $\operatorname{Exp}_u(v) = g(1)$, where $g : [0,1] \to \operatorname{int}\mathcal{K}$ is the geodesic satisfying

$$g(0) = u, \qquad \dot{g}(0) = v,$$

where $\dot{g}(t) := \frac{d}{dt}g(t)$. For a general manifold $\mathcal{M} \subseteq \mathbb{R}^m$, evaluating this map requires solving a system of 2nd-order ODEs of the form

$$\ddot{g}_k + \sum_{i=1}^{m}\sum_{j=1}^{m}\Gamma_{ij}^{k}\dot{g}_i\dot{g}_j = 0, \quad k = 1, 2, \dots, m, \tag{13}$$

where $\Gamma_{ij}^k \in \mathbb{R}$ are the *Christoffel symbols* of $\mathcal{M}$. For symmetric cones, however, Exp_u has an explicit formula involving the exponential map of the *algebra* $\mathcal{J}$:

$$\operatorname*{Exp}_u(v) = Q(u^{1/2})\exp(Q(u^{-1/2})v).$$

Further, we can express geodesic updates (Proposition 1.1) of the primal-dual variables $(x,s) = \sqrt{\mu}(w, w^{-1})$ using Exp_x and Exp_s.

Proposition 2.2. *For $w \in \operatorname{int}\mathcal{K}$ and $\mu > 0$, let $d = d_N(w,\mu)$ and define*

$$x := \sqrt{\mu}w, \ \ s := \sqrt{\mu}w^{-1}, \ \ d_x := \sqrt{\mu}Q(w^{1/2})d, \ \ d_s := -\sqrt{\mu}Q(w^{-1/2})d. \tag{14}$$

The following statements hold.

- $\operatorname{Exp}_x(d_x) = \sqrt{\mu}Q(w^{1/2})\exp(d)$.
- $\operatorname{Exp}_s(d_s) = \sqrt{\mu}Q(w^{-1/2})\exp(-d)$.
- $\operatorname{Exp}_x(d_x) \circ \operatorname{Exp}_s(d_s) = \mu e$.

Proof. We first observe that

$$\sqrt{\mu}Q(w^{1/2}) = Q(\mu^{1/4}w^{1/2}) = Q((\sqrt{\mu}w)^{1/2}) = Q(x^{1/2}),$$

which implies that $Q(x^{-1/2}) = \frac{1}{\sqrt{\mu}}Q(w^{-1/2})$. Evaluating Exp_x at $d_x := \sqrt{\mu}Q(w^{1/2})d$ yields

$$\operatorname*{Exp}_x(d_x) = Q(x^{1/2})\exp(Q(x^{-1/2})\sqrt{\mu}Q(w^{1/2})d) = \sqrt{\mu}Q(w^{1/2})\exp(d).$$

The second statement follows by identical argument. The third follows using the first two statements and the identity $(Q(w^{1/2})\exp(d))^{-1} = Q(w^{-1/2})\exp(-d)$; see Lemma A.1. ∎

Recalling the definition of the local norm $\|v\|_u := \|Q(u)^{-1/2}v\|$, we can also characterize the tangent vectors d_x and d_s without reference to $d_N(w,\mu)$.

Proposition 2.3. *The tangent vectors d_x and d_s in* (14) *are the unique points in $\mathcal{J}$ satisfying*

$$x + d_x \in x_0 + \mathcal{L}, \quad s + d_s \in s_0 + \mathcal{L}^\perp, \quad d_s = -\mu Q(x)^{-1} d_x.$$

Further, $\|d\| = \|d_x\|_x$ and $\|d\| = \|d_s\|_s$.

Proof. The first two conditions are immediate from definition of (x, d_x), (s, d_s), and the Newton direction $d_N(w,\mu)$. To see that $d_s = -\mu Q(x)^{-1} d_x$, observe first that

$$\sqrt{\mu} d = Q(w^{-1/2}) d_x, \qquad -\sqrt{\mu} d = Q(w^{1/2}) d_s.$$

Hence, $-d_s = Q(w^{-1}) d_x = Q(\sqrt{\mu} x^{-1}) d_x = \mu Q(x^{-1}) d_x$. Uniqueness follows from the fact $Q(x)$ is invertible and the fact the first two conditions are equivalent to $\dim \mathcal{J}$ linearly independent constraints.

For the last statement, we have that

$$\|d_x\|_x^2 = \langle Q(x^{-1/2}) d_x, Q(x^{-1/2}) d_x \rangle.$$

But $Q(x^{1/2}) = \sqrt{\mu} Q(w^{1/2})$. Hence, $Q(x^{-1/2}) d_x = d$, proving that $\|d\| = \|d_x\|_x$. That $\|d\| = \|d_s\|_s$ follows by similar argument. ☐

These propositions suggest how `shortstep` generalizes to non-symmetric cones. Indeed, any cone with a *log-homogeneous, self-concordant barrier function* is equipped with a natural Riemannian geometry [31] that enables definition of Exp_x and Exp_s. The affine constraints characterizing the tangent vectors (Proposition 2.3) also generalize if one interprets $Q(u)^{-1}$ as the Hessian of the barrier function $\log\det u^{-1}$. One can also interpret $\mu Q(x^{-1})$ in Proposition 2.3 as the *parallel transport* operator from x to s, a canonical operation in Riemannian geometry. An obstruction to implementation, however, is evaluation of Exp_x and Exp_s, which, as mentioned, may require numerical solution of the ODE system (13). Our convergence analysis (Section 3) will also leverage the spectral theory of symmetric cones, and hence does not immediately generalize.

An alternative generalization, applicable to even symmetric cones, replaces Exp_u with a general *retraction* $R_u : \mathcal{J} \to \operatorname{int} \mathcal{K}$. Retractions are defined by relaxing the geodesic property of Exp_u. That is, a retraction R_u smoothly maps a tangent vector $v \in \mathcal{J}$ to a point $R_u(v) \in \operatorname{int}\mathcal{K}$ with the property that the curve $\gamma(t) := R_u(tv)$ satisfies $\gamma(0) = u$ and $\dot\gamma(0) = v$. The map Exp_u is a special case of a retraction for which $\gamma(t)$ is geodesic. See [1, Chapter 4.1] for more details.

2.3 Newton direction via orthogonal projection

We next derive an orthogonal, direct-sum decomposition of the Newton direction with respect to the subspaces $\mathcal{L}_w := \{Q(w^{-1/2})u : u \in \mathcal{L}\}$ and $\mathcal{L}_w^\perp = \{Q(w^{1/2})u : u \in \mathcal{L}^\perp\}$. This decomposition establishes both its claimed uniqueness (Definition 1.1) and a formula for its construction via orthogonal projection.

Proposition 2.4. *For $\mu > 0$ and $w \in \operatorname{int}\mathcal{K}$, let*

$$d_1 = \operatorname*{proj}_{\mathcal{L}_w^\perp}\left(Q(w^{-1/2})(\frac{1}{\sqrt{\mu}} x_0 - w) \right), \qquad d_2 = \operatorname*{proj}_{\mathcal{L}_w}\left(Q(w^{1/2})(\frac{1}{\sqrt{\mu}} s_0 - w^{-1}) \right).$$

Then the Newton direction $d_N(w,\mu)$ satisfies $d_N(w,\mu) = d_1 - d_2$.

Proof. Let $r_1 = Q(w^{-1/2})(\frac{1}{\sqrt{\mu}}x_0 - w)$ and $r_2 = Q(w^{1/2})(\frac{1}{\sqrt{\mu}}s_0 - w^{-1})$. By the identity $Q(z^{1/2})e = z$ (Lemma A.1), the conditions of Definition 1.1 are equivalent to

$$w + Q(w^{1/2})d \in \frac{1}{\sqrt{\mu}}x_0 + \mathcal{L}, \qquad w^{-1} - Q(w^{-1/2})d \in \frac{1}{\sqrt{\mu}}s_0 + \mathcal{L}^\perp.$$

Using $Q(z^{-1}) = Q(z)^{-1}$ (Lemma A.1), we conclude that $d \in r_1 + \mathcal{L}_w$ and $d \in -r_2 + \mathcal{L}_w^\perp$. Equivalently,

$$d \in (\operatorname*{proj}_{\mathcal{L}_w^\perp}(r_1) + \mathcal{L}_w) \bigcap (\operatorname*{proj}_{\mathcal{L}_w}(-r_2) + \mathcal{L}_w^\perp),$$

since any affine set $z_0 + \mathcal{S}$ satisfies $z_0 + \mathcal{S} = \operatorname{proj}_{\mathcal{S}^\perp}(z_0) + \mathcal{S}$. Hence, d has the following direct-sum decompositions with respect to $\mathcal{L}_w$ and $\mathcal{L}_w^\perp$:

$$d = \operatorname*{proj}_{\mathcal{L}_w^\perp}(r_1) + d_{\mathcal{L}_w}, \qquad d = \operatorname*{proj}_{\mathcal{L}_w}(-r_2) + d_{\mathcal{L}_w^\perp}.$$

Since such decompositions are unique, $d_{\mathcal{L}_w} = \operatorname{proj}_{\mathcal{L}_w}(-r_2)$, proving the claim. □

This decomposition has immediate practical implications: one can use any algorithm for orthogonal projection, e.g., the Gram-Schmidt process or a least-squares method, to find d_N. Section 4.2 gives an explicit linear system for performing this projection using this latter approach. Further, the size/structure of this linear system matches the size/structure of linear systems arising in classical IPMs.

2.4 Scale invariance

For an automorphism $T : \mathcal{J} \to \mathcal{J}$ of $\mathcal{K}$, consider the transformed primal-dual pair:

$$\begin{array}{llll} \text{minimize} & \langle (T^{-1})^* s_0, x\rangle & \text{minimize} & \langle T x_0, s\rangle \\ \text{subject to} & x \in \mathcal{K} \cap T(x_0 + \mathcal{L}) & \text{subject to} & s \in \mathcal{K} \cap (T^{-1})^*(s_0 + \mathcal{L}^\perp), \end{array} \tag{15}$$

where $(T^{-1})^* : \mathcal{J} \to \mathcal{J}$ denotes the adjoint of $T^{-1} : \mathcal{J} \to \mathcal{J}$. We next show the following: if `shortstep` maps input w_0 to output $\bar{w}$ for the primal-dual pair (1), then it maps input Tw_0 to output $T\bar{w}$ for the transformed pair (15). In other words, it is *scale invariant* in the sense of [41]. To show this, we first establish that the Newton direction $d_{N,T}(w, \mu)$ for the transformed problem satisfies $d_{N,T}(Tw, \mu) = M d_N(w, \mu)$ for an automorphism M, dependent on T and w, that is also *orthogonal*, i.e., $M^{-1} = M^*$. Scale invariance will follow, leveraging the fact that $\exp(Md) = M \exp(d)$ for any such M (Lemma A.2).

To give a formula for M and to establish its key properties, we use the decomposition $d_N(w, \mu) = d_1(w, \mu) - d_2(w, \mu)$ of the Newton direction from Proposition 2.4. We also decompose the transformed direction as $d_{N,T}(v, \mu) = d_{1,T}(v, \mu) - d_{2,T}(v, \mu)$ by applying Proposition 2.4 to the transformed problem (15).

Lemma 2.3. *Let $M = Q(Tw)^{-1/2} T Q(w)^{1/2}$ for $w \in \operatorname{int} \mathcal{K}$ and an automorphism $T : \mathcal{J} \to \mathcal{J}$ of $\mathcal{K}$. The following statements hold.*

(a) M is an orthogonal automorphism of $\mathcal{K}$.

(b) $M = Q(Tw)^{1/2}(T^{-1})^* Q(w)^{-1/2}$.

(c) *For all $\mu > 0$, the Newton directions satisfy $d_{N,T}(Tw,\mu) = Md_N(w,\mu)$. Further, their direct summands satisfy*

$$d_{1,T}(Tw,\mu) = Md_1(w,\mu), \qquad d_{2,T}(Tw,\mu) = Md_2(w,\mu).$$

Proof. That M is an automorphism follows because it is a composition of automorphisms. We next verify orthogonality, i.e., that $M^{-1} = M^*$:

$$M^*M = Q(w)^{1/2}T^*Q(Tw)^{-1}TQ(w)^{1/2} = Q(w)^{1/2}T^*(TQ(w)T^*)^{-1}TQ(w)^{1/2} = I,$$

where we've used the identities $Q(Tw) = TQ(w)T^*$ and $Q(w)^{1/2}Q(w)^{-1}Q(w)^{1/2} = I$ (Lemma A.1). Since by construction $M^*Q(Tw)^{1/2}(T^{-1})^*Q(w)^{-1/2} = I$, orthogonality implies the next statement.

By definition of M and the second property, we conclude that $MQ(w)^{-1/2} = Q(Tw)^{-1/2}T$ and $MQ(w)^{1/2} = Q(Tw)^{1/2}(T^{-1})^*$. Combining this with $Me = e$ (Lemma A.2) shows that both

$$d_{N,T}(Tw,\mu) \in M\Big(\frac{1}{\sqrt{\mu}}Q(w^{-1/2})x_0 - e + Q(w^{-1/2})\mathcal{L}\Big)$$

and

$$d_{N,T}(Tw,\mu) \in M\Big(e - Q(w^{1/2})\frac{1}{\sqrt{\mu}}s_0 + Q(w^{1/2})\mathcal{L}^{\perp}\Big).$$

Following the proof of Proposition 2.4, we conclude that $d_{N,T}(Tw,\mu) = M(d_1 - d_2)$, which implies $d_{1,T} = Md_1$ and $d_{2,T} = Md_2$. ☐

We use this lemma to show scale invariance of $w \leftarrow Q(w^{1/2})\exp(\alpha(d_1,d_2)d)$, where $d = d_N$ and $\alpha : \mathcal{J} \times \mathcal{J} \to \mathbb{R}$ is a step-size rule invariant under transformation by M.

Proposition 2.5. *Let $\alpha : \mathcal{J} \times \mathcal{J} \to \mathbb{R}$ be a function satisfying $\alpha(d_1,d_2) = \alpha(Md_1,Md_2)$ for any orthogonal automorphism $M : \mathcal{J} \to \mathcal{J}$. Then, for any automorphism $T : \mathcal{J} \to \mathcal{J}$, $w \in \operatorname{int}\mathcal{K}$, and $\mu > 0$,*

$$Q(\tilde{w}^{1/2})\exp\big(\alpha(\tilde{d}_1,\tilde{d}_2)\tilde{d}\big) = TQ(w^{1/2})\exp\big(\alpha(d_1,d_2)d\big),$$

where $\tilde{w} = Tw$, $d = d_N(w,\mu)$, $\tilde{d} = d_{N,T}(\tilde{w},\mu)$, $d_i = d_i(w,\mu)$, and $\tilde{d}_i = d_{i,T}(\tilde{w},\mu)$ for $i \in \{1,2\}$.

Proof. Let $M = Q(Tw)^{-1/2}TQ(w)^{1/2}$. By Lemma 2.3, M is an orthogonal automorphism. Hence, $\exp(Mx) = M\exp(x)$ for all x (Lemma A.2). Combining this with Lemma 2.3(c) yields

$$\begin{aligned} Q(\tilde{w}^{1/2})\exp(\alpha(\tilde{d}_1,\tilde{d}_2)\tilde{d}) &= Q(\tilde{w}^{1/2})\exp(\alpha(Md_1,Md_2)Md) \\ &= Q(\tilde{w}^{1/2})M\exp(\alpha(Md_1,Md_2)d). \end{aligned}$$

But $\alpha(Md_1,Md_2) = \alpha(d_1,d_2)$ by assumption and $Q(\tilde{w}^{1/2})M = TQ(w^{1/2})$ by definition of M and the identity $Q(u)^{1/2} = Q(u^{1/2})$; see Lemma A.1. ☐

Scale invariance of `shortstep` follows by invoking this result at each iteration with the step-size $\alpha(d_1,d_2) = 1$. We will use a nontrivial step-size rule in Section 4.

2.5 Comparison with the Nesterov-Todd algorithm

The celebrated algorithm of Nesterov and Todd (NT) [30, Section 6], which extends the linear programming algorithms of Kojima et al. [16] and Monteiro and Adler [26], shares key properties with `shortstep`: it is scale invariant, it executes $\mathcal{O}(\sqrt{n})$ iterations, it is primal-dual symmetric, and finding its search direction reduces to orthogonal projection. This suggests a fundamental connection with `shortstep`. In general, iterations of the NT algorithm do *not* satisfy $x = \mu s^{-1}$. However, if this relation holds, then the NT search direction coincides with our Newton direction. Further, its (x, s)-update is a first-order approximation of our geodesic update.

To see this, note that the NT direction is, in the framework of Jordan algebras [11, Section 3.2], the unique $(d_x, d_s) \in \mathcal{J} \times \mathcal{J}$ satisfying

$$x + \sqrt{\mu}Q(p^{1/2})d_x \in x_0 + \mathcal{L}, \quad s + \sqrt{\mu}Q(p^{-1/2})d_s \in s_0 + \mathcal{L}^{\perp}, \quad d_x + d_s = v^{-1} - v, \tag{16}$$

where p is the *scaling point*, defined as $Q(x^{1/2})(Q(x^{1/2})s)^{-1/2}$, and $v := \frac{1}{\sqrt{\mu}}Q(p^{-1/2})x$. Given (d_x, d_s), the NT algorithm updates (x, s) to (x', s'), where

$$x' := x + \sqrt{\mu}Q(p^{1/2})d_x, \qquad s' := s + \sqrt{\mu}Q(p^{-1/2})d_s. \tag{17}$$

Our result follows.

Proposition 2.6. *Let $x, s \in \operatorname{int}\mathcal{K}$ satisfy $x = \mu s^{-1}$ for $\mu > 0$. Let $w = \frac{1}{\sqrt{\mu}}x$ and $d = d_N(w, \mu)$. Then,*

(a) $p = w$, where p is the scaling point $Q(x^{1/2})(Q(x^{1/2})s)^{-1/2}$.

(b) $d_x = d$ and $d_s = -d$ where (d_x, d_s) is the NT direction (16)*.*

(c) $x' = \sqrt{\mu}Q(w^{1/2})(e + d)$ and $s' = \sqrt{\mu}Q(w^{-1/2})(e - d)$, where (x', s') is the NT update (17) *and $e + d$ and $e - d$ are the first-order Taylor-expansions of* $\exp(d)$ *and* $\exp(-d)$ *at $d = 0$.*

Proof. If $x = \mu s^{-1}$, then the definitions of w and the scaling point p easily imply that $p = w$ and $v = e$. We also conclude that $d_x + d_s = v^{-1} - v = 0$. Combining these identities with $w = Q(w^{1/2})e$, $w^{-1} = Q(w^{-1/2})e$, and (16) yields

$$\sqrt{\mu}Q(w^{1/2})(e + d_x) \in x_0 + \mathcal{L}, \qquad \sqrt{\mu}Q(w^{-1/2})(e - d_x) \in s_0 + \mathcal{L}^{\perp},$$

which are the defining conditions of $d_N(w, \mu)$ given by Definition 1.1. Hence, $d = d_x$. Finally, the claimed formula for (x', s') holds because $x = \sqrt{\mu}Q(w^{1/2})e$ and $s = \sqrt{\mu}Q(w^{-1/2})e$. ☐

Note with the stronger assumption that $x = s^{-1}$, we can similarly interpret algorithms based on the so-called H..K..M direction since, in this case, it coincides with the NT direction [39]. It was introduced independently by Helmberg et al. [14], Kojima et al. [17] and Monteiro [25]. Also note that even if $x = \mu s^{-1}$ fails, the scaling point p still has a Riemannian interpretation: it is precisely the midpoint of the geodesic connecting x and s^{-1}, or, equivalently, their geometric mean [21]. Finally, we note that the NT direction has an alternative derivation due to Sturm and Zhang [37]; see remarks in [36].

3 Geodesics and divergence

The goal of this section is to prove the μ-update and centering lemmas used in the analysis of `shortstep` (Figure 1). Towards this, we first study a proxy for geodesic distance $\delta(u, v)$ that is easier to bound during the course of Newton's method. This proxy generalizes the *symmetric Kullback-Leibler divergence* $h(U, V) := \mathrm{Tr}(UV^{-1} + U^{-1}V - 2I)$ of two zero-mean Gaussian distributions with covariance matrices U and V, also known as the *Jeffrey divergence* [12, 23]. We hence call this proxy *divergence.* We define it using the fact that $\operatorname{tr} e$ equals the rank of $\mathcal{K}$ (which we've denoted by n).

Definition 3.1. *Denote by $h(u, v)$ the* divergence *of $u, v \in \operatorname{int} \mathcal{K}$, defined as $h(u, v) := \langle u, v^{-1}\rangle + \langle u^{-1}, v\rangle - 2n$.*

Divergence is symmetric and non-negative, i.e., $h(u, v) = h(v, u)$ and $h(u, v) \geq 0$ for all $u, v \in \operatorname{int} \mathcal{K}$. Further, $h(u, v) = 0$ if and only if $u = v$. However, unlike geodesic distance $\delta(u, v)$, it is *not* a metric, as the triangle inequality can fail.

Recall from Lemma 1.1 that geodesic distance satisfies $\delta(u, v) = \| \log Q(v^{-1/2})u\|$. Equivalently, $\delta(u, v)^2 = \sum_{\lambda \in S} \lambda^2$, where S denotes the multiset of eigenvalues of $\log Q(v^{-1/2})u$. This formula holds for divergence if we replace λ^2 with the upper bound $q(\lambda) := 2(\cosh(\lambda) - 1)$ introduced in Section 2.

Lemma 3.1. *For all $u, v \in \operatorname{int} \mathcal{K}$, the divergence satisfies $h(u, v) = \sum_{\lambda \in S} q(\lambda)$, where S is the multiset of eigenvalues of $\log Q(v^{-1/2})u$.*

This enables us to prove the following bounds relating divergence to geodesic distance.

Lemma 3.2. *Let $u, v \in \operatorname{int} \mathcal{K}$. Then, $\delta(u, v)^2 \leq h(u, v) \leq q(\delta(u, v))$.*

Proof. Let $\lambda \in \mathbb{R}^n$ denote the vector of eigenvalues of $\log Q(v^{-1/2})u$. The lower bound follows from Lemma 3.1 and Lemma 1.1(b) given that $q(\lambda_i) \geq \lambda_i^2$. To prove the upper bound, it suffices to show that $\sum_{i=1}^{n}(\cosh(\lambda_i) - 1) \leq \cosh(\|\lambda\|) - 1$. To begin, consider the upper bound

$$\sum_{i=1}^{n}(\cosh(\lambda_i) - 1) \leq \sup_{\|z\|=\|\lambda\|} \sum_{i=1}^{n}(\cosh(z_i) - 1).$$

Let z achieve the supremum. Then it must be a critical point, which implies existence of $\gamma \in \mathbb{R}$ satisfying $\gamma z + \sinh(z) = 0$. We conclude that $z_i = 0$ or $|z_i| = c$ for a constant $c > 0$. We now claim that $z_i \neq 0$ and $z_j \neq 0$ implies $i = j$. Suppose otherwise. Then we don't change $\|z\|$ by setting $z_i = 0$ and $z_j = \sqrt{2}c$. Further, we increase $\sum_{i=1}^{n}(\cosh(z_i) - 1)$ given that $\cosh(\sqrt{2}c) - 1 > 2(\cosh(c) - 1)$, contradicting our assumption that z attains the supremum. $\square$

We also note that $h(u, v)$ shares the invariance properties of geodesic distance $\delta(u, v)$. It is symmetric with respect to inversion, i.e., $h(u, v) = h(u^{-1}, v^{-1})$. Hence, it measures the proximity of (w, w^{-1}) to the centered-points $(\hat{w}(\mu), \hat{w}(\mu)^{-1})$ in a primal-dual symmetric way, i.e., $h(w, \hat{w}(\mu)) = h(w^{-1}, \hat{w}(\mu)^{-1})$. It is also scale invariant, meaning $h(Tu, Tv) = h(u, v)$ for any automorphism T of $\mathcal{K}$.

Remark 1. *The quantity $h(v, v^{-1})$ where v is as defined in Section 2.5, is used to analyze a full-step Nesterov-Todd algorithm [11, Section 3.3].*

3.1 Divergence along the central path

Divergence has the following utility: we can calculate it *exactly* for two centered points $\hat{w}(\mu_0)$ and $\hat{w}(\mu_1)$ even if we do not know these points explicitly. Instead, all we need is the ratio of the centering parameters μ_0 and μ_1 and the rank of $\mathcal{K}$, denoted by n.

Theorem 3.1. *Let $\mu_0, \mu_1 > 0$. Then, $\frac{1}{n}h(\hat{w}(\mu_0), \hat{w}(\mu_1)) = q(\frac{1}{2}\log\frac{\mu_0}{\mu_1})$.*

Proof. Let $u = \hat{w}(\mu_0)$, $v = \hat{w}(\mu_1)$ and $\alpha = \sqrt{\frac{\mu_0}{\mu_1}}$. Since $\sqrt{\mu}_0(u, u^{-1})$ and $\sqrt{\mu}_1(v, v^{-1})$ are feasible,

$$v - \alpha u \in \mathcal{L}, \qquad v^{-1} - \alpha u^{-1} \in \mathcal{L}^{\perp}.$$

Hence, $0 = \langle v - \alpha u, v^{-1} - \alpha u^{-1}\rangle = (1+\alpha^2)n - \alpha\langle v, u^{-1}\rangle - \alpha\langle u, v^{-1}\rangle$. Rearranging shows that

$$\langle u, v^{-1}\rangle + \langle u^{-1}, v\rangle = n\frac{1+\alpha^2}{\alpha} = n(\alpha + \frac{1}{\alpha}) = 2n(\cosh(\log(\alpha))).$$

Hence, $h(u, v) = 2n(\cosh(\log(\alpha)) - 1)$. Using $q(t) := 2(\cosh(t) - 1)$ and $\log\alpha = \frac{1}{2}\log\frac{\mu_0}{\mu_1}$ yields:

$$\frac{1}{n}h(u, v) = q(\log\alpha) = q(\frac{1}{2}\log\frac{\mu_0}{\mu_1}).$$

□

Combining this theorem with the bounds relating divergence and geodesic distance (Lemma 3.2) lets us prove the μ-update lemma, which we reproduce below.

Lemma 2.1 (μ-update)**.** *Let $\mu, k > 0$. Then, $\frac{1}{n}\delta\left(\hat{w}(\mu), \hat{w}(\frac{1}{k}\mu)\right)^2 \leq q(\frac{1}{2}\log k)$.*

Proof. From Theorem 3.1, we conclude that $\frac{1}{n}h(\hat{w}(\mu), \hat{w}(\frac{1}{k}\mu)) = q(\frac{1}{2}\log k)$. Since $\delta(\hat{w}(\mu), \hat{w}(\frac{1}{k}\mu))^2 \leq h(\hat{w}(\mu), \hat{w}(\frac{1}{k}\mu))$ by Lemma 3.2, the claim follows. □

Remark 2. *Since geodesic distance is invariant under inversion and positive rescaling, we have, for $(x, s) = \sqrt{\mu}(w, w^{-1})$, that $\delta(x, \hat{x}(\mu)) = \delta(s, \hat{s}(\mu)) = \delta(w, \hat{w}(\mu))$. This implies that the lengths L_x and L_s of the primal and dual central paths also upper bound $\delta(\hat{w}(\mu_0), \hat{w}(\mu_1))$, where*

$$L_x := \int_{\mu_0}^{\mu_1} \|\frac{d}{d\mu}\hat{x}(\mu)\|_{\hat{x}(\mu)} d\mu, \qquad L_s := \int_{\mu_0}^{\mu_1} \|\frac{d}{d\mu}\hat{s}(\mu)\|_{\hat{s}(\mu)} d\mu,$$

and $\|v\|_u := \|Q(u)^{-1/2}v\|$. Bounds on L_x in terms of $\log(\mu_0/\mu_1)$ and *the (generally unknown) values of the barrier function* $\log\det z^{-1}$ *at $z = \hat{x}(\mu_0)$ and $z = \hat{x}(\mu_1)$ appear in [28, Lemma 4.1].*

3.2 Divergence along geodesics

Fix $\mu > 0$, $w \in \operatorname{int}\mathcal{K}$, and nonzero $d \in \mathcal{J}$, and define the function $f : \mathbb{R} \to \mathbb{R}$

$$f(t) = h\left(Q(w^{1/2})\exp(td), \hat{w}(\mu)\right).$$

That is, let $f(t)$ return the divergence between the centered point $\hat{w}(\mu)$ and points on the geodesic induced by (w, d). Though we don't know $\hat{w}(\mu)$ and hence cannot evaluate f, we can still establish crucial properties, such as its strict convexity.

Lemma 3.3. *The function f is strictly convex.*

Proof. Let $a := Q(w^{1/2})\hat{w}(\mu)^{-1}$ and let $\sum_{i=1}^{n} \lambda_i e_i$ denote the spectral decomposition of d. Then,

$$f(t) + 2n = \langle a, \exp(td)\rangle + \langle a^{-1}, \exp(-td)\rangle = \sum_{i=1}^{n} \exp(t\lambda_i)\langle a, e_i\rangle + \exp(-t\lambda_i)\langle a^{-1}, e_i\rangle.$$

But $\langle a, e_i\rangle > 0$ and $\langle a^{-1}, e_i\rangle > 0$ since $a, a^{-1} \in \operatorname{int}\mathcal{K}$ and $e_i \in \mathcal{K}$, proving the claim by strict convexity of the scalar exponential function. ☐

We can also ensure that $f(t) < f(0)$ for a piecewise step-size rule involving the *spectral norm* $\|d\|_\infty$ of d, defined as $\|d\|_\infty := \max_{\lambda \in S} |\lambda|$ where S denotes the multiset of eigenvalues of d. This rule also incorporates a parameter $\theta \in (0, 1)$ controlling the transition from full to damped Newton steps.

Theorem 3.2. *Let $d = d_N(w, \mu)$ and $\alpha = \max\{1, \frac{1}{2\theta}\|d\|_\infty^2\}$ for $\theta \in (0, 1)$. The following statements hold.*

- *(a) If $\alpha = 1$, then $f(1) \le \frac{1}{2}\|d\|_\infty^2 f(0) \le \theta f(0)$.*
- *(b) $f(1/\alpha) < f(0)$.*

To prove this theorem, we'll first provide the derivatives of $f(t)$ and a descent condition on t for arbitrary d. We then specialize results to the Newton direction $d_N(w, \mu)$.

Remark 3. *A function $p : \operatorname{int}\mathcal{K} \to \mathbb{R}$ is called* geodesically convex *if its restrictions to geodesics are convex in the usual sense, i.e., if $p(g(t))$ is a convex function of t for all curves $g(t)$ of the form $t \mapsto Q(w^{1/2})\exp(td)$. This convexity notion is a central concept in manifold optimization [35, 42, 7]. The convexity of $f(t)$ reflects the geodesic convexity of the divergence map $w \mapsto h(w, \hat{w}(\mu))$.*

3.2.1 Derivatives and descent condition

The derivatives $d^m f(t)/(dt)^m$, denoted $f^{(m)}$ for short, have a concise form given the role of the exponential function in the definition of f. Interpreting $f(t)$ as the trace of a particular point in $\mathcal{K}$ also allows us to bound *even* derivatives using just d and $f(t)$.

Lemma 3.4. *Let $a(t) = Q(\exp(td))^{1/2}Q(w^{1/2})\hat{w}(\mu)^{-1}$. The following hold for all $t \in \mathbb{R}$:*

- *(a) $f(t) = \operatorname{tr}(a(t) + a(t)^{-1} - 2e)$, where $a(t) + a(t)^{-1} - 2e \in \mathcal{K}$.*
- *(b) $f^{(m)}(t) = \langle a(t) + (-1)^m a(t)^{-1}, d^m\rangle$.*
- *(c) $f^{(2m)}(t) \le \|d\|_\infty^{2m} f(t) + 2\langle e, d^{2m}\rangle$.*

Proof. By definition of f and divergence (Definition 3.1), we have

$$f(t) = \langle Q(w^{1/2})\exp(td), \hat{w}(\mu)^{-1}\rangle + \langle Q(w^{-1/2})\exp(-td), \hat{w}(\mu)\rangle - 2n.$$

Substituting $\exp(td) = Q(\exp(td))^{1/2}e$ and $\exp(-td) = Q(\exp(-td))^{1/2}e$ yields

$$f(t) = \langle e, Q(\exp(td))^{1/2}Q(w^{1/2})\hat{w}(\mu)^{-1}\rangle + \langle e, Q(\exp(-td))^{1/2}Q(w^{-1/2})\hat{w}(\mu)\rangle - 2n.$$

Two applications of the identity $[Q(u)v]^{-1} = Q(u^{-1})v^{-1}$ shows that

$$a(t)^{-1} = Q(\exp(-td))^{1/2}Q(w^{-1/2})\hat{w}(\mu),$$

which proves the trace identity of the first statement. That $a(t) + a(t)^{-1} - 2e \in \mathcal{K}$ follows because each eigenvalue has form $\lambda + \frac{1}{\lambda} - 2$ for some $\lambda > 0$, which is always nonnegative.

For (b), we have that $\frac{d^m}{dt^m}\exp(td) = d^m \circ \exp(td) = Q(\exp(td))^{1/2} d^m$. This implies that

$$\begin{aligned}\frac{d^m}{dt^m}\langle e, a(t)\rangle &= \langle Q(w^{1/2})\hat{w}(\mu)^{-1}, \frac{d^m}{dt^m}\exp(td)\rangle \\ &= \langle Q(\exp(td))^{1/2} Q(w^{1/2})\hat{w}(\mu)^{-1}, d^m\rangle \\ &= \langle a(t), d^m\rangle.\end{aligned}$$

By similar argument, $\frac{d^m}{dt^m}\langle e, a(t)^{-1}\rangle = (-1)^m\langle a(t)^{-1}, d^m\rangle$. We conclude for all integers $m \geq 1$ that $f^{(m)}(t) = \langle a(t) + (-1)^m a(t)^{-1}, d^m\rangle$. For statement (c), we have, since $a(t) + a^{-1}(t) - 2e \in \mathcal{K}$,

$$\begin{aligned}f^{(2m)}(t) &= \langle a(t) + a^{-1}(t) - 2e, d^{2m}\rangle + 2\langle e, d^{2m}\rangle \\ &\leq \|a(t) + a^{-1}(t) - 2e\|_1 \|d\|_\infty^{2m} + 2\langle e, d^{2m}\rangle \\ &= \operatorname{tr}(a(t) + a^{-1}(t) - 2e)\|d\|_\infty^{2m} + 2\langle e, d^{2m}\rangle \\ &= \|d\|_\infty^{2m} f(t) + 2\langle e, d^{2m}\rangle.\end{aligned}$$

□

Let f' and f'' denote the first and second derivatives of f. Assuming $f'(0) < 0$, we now provide a descent condition on t, i.e., we establish an interval on which $f(t) \leq f(0)$. Our analysis rests on Taylor's theorem, convexity of f, and the Lemma 3.4 bound on $f''(t)$.

Lemma 3.5. *If $f'(0) < 0$ and $0 \leq t \leq \frac{-2f'(0)}{\|d\|_\infty^2 f(0) + 2\|d\|^2}$, then $f(t) \leq f(0)$.*

Proof. By Taylor's theorem, $f(t) = f(0) + f'(0)t + \frac{1}{2}f''(\zeta)t^2$ for some $\zeta \in [0, t]$. Further,

$$f''(\zeta) \leq \|d\|_\infty^2 f(\zeta) + 2\|d\|^2 \leq \max_{u\in\{0,t\}} \|d\|_\infty^2 f(u) + 2\|d\|^2,$$

where the first inequality is Lemma 3.4(c) and the second inequality uses convexity of $f(t)$. Hence,

$$f(t) \leq f(0) + f'(0)t + \frac{1}{2}\max_{u\in\{0,t\}}(\|d\|_\infty^2 f(u) + 2\|d\|^2)t^2. \tag{18}$$

Now, let $\hat{t}$ be the smallest $t > 0$ for which $f(\hat{t}) = f(0)$. Then

$$f(0) \leq f(0) + \hat{t}f'(0) + \frac{1}{2}(\|d\|_\infty^2 f(0) + 2\|d\|^2)\hat{t}^2,$$

which implies that

$$\hat{t} \geq \frac{-2f'(0)}{\|d\|_\infty^2 f(0) + 2\|d\|^2}.$$

Since $f(t) \leq f(0)$ for all $0 \leq t \leq \hat{t}$, the claim follows. □

3.2.2 Newton direction

Suppose now that $d = d_N(w, \mu)$. For this direction, we can bound $f'(0)$ using $f(0)$ and $\|d\|^2$ by applying the orthogonal, direct-sum decomposition of d from Proposition 2.4. Recall that this decomposition is with respect to $\mathcal{L}_w := \{Q(w^{-1/2})u : u \in \mathcal{L}\}$ and $\mathcal{L}_w^\perp = \{Q(w^{1/2})u : u \in \mathcal{L}^\perp\}$. This bound also provides an updated descent condition for t.

Lemma 3.6. *Suppose that $d = d_N(w,\mu)$. Then $f'(0) = -(f(0) + \|d\|^2)$. Further, $f(t) \le f(0)$ if*

$$0 \le t \le \frac{2(f(0) + \|d\|^2)}{\|d\|_\infty^2 f(0) + 2\|d\|^2}.$$

Proof. Let $r_1(t) = a(t)^{-1} - e$ and $r_2(t) = a(t) - e$, where $a(t)$ is as in Lemma 3.4. Then, by Lemma 3.4(b),

$$-f' = \langle a^{-1} - a, d\rangle = \langle a^{-1} - e + e - a, d\rangle = \langle r_1 - r_2, d\rangle.$$

Setting $t = 0$ and substituting $d = \operatorname{proj}_{\mathcal{L}_w^\perp} r_1(0) - \operatorname{proj}_{\mathcal{L}_w} r_2(0)$ using Proposition 2.4 gives

$$-f'(0) = \langle r_1 - r_2, d\rangle = -\langle r_1, r_2\rangle + \|\operatorname*{proj}_{\mathcal{L}_w^\perp} r_1\|^2 + \|\operatorname*{proj}_{\mathcal{L}_w} r_2\|^2 = -\langle r_1, r_2\rangle + \|d\|^2.$$

But $f(t) = -\langle r_1(t), r_2(t)\rangle$ by Lemma 3.4(a), proving the first claim. The descent condition (Lemma 3.5) specialized to the Newton direction $d = d_N$ proves the second claim. ☐

3.2.3 Proof of Theorem 3.2

We can now prove the properties of the step-size rule $t = \min\{1, \frac{2\theta}{\|d\|_\infty^2}\}$ claimed by Theorem 3.2 for the Newton direction $d = d_N(w,\mu)$ and parameter $\theta \in (0,1)$. Assume first that $t = 1$. Then $\|d\|_\infty^2 \le 2\theta < 2$, which, by Lemma 3.6, implies that $f(1) \le f(0)$. Combining this with the quadratic upper bound (18) and $f'(0) = -(f(0) + \|d\|^2)$ from Lemma 3.6 yields

$$f(1) \le f(0) - (f(0) + \|d\|^2) + \frac{1}{2}(\|d\|_\infty^2 f(0) + 2\|d\|^2) = \frac{1}{2}\|d\|_\infty^2 f(0) \le \theta f(0),$$

which is precisely the claim of Theorem 3.2-(a). Now suppose that $t = \frac{2\theta}{\|d\|_\infty^2} \le 1$. By Lemma 3.6 and strict convexity of f, we have $f(t) < f(0)$ if

$$\frac{2\theta}{\|d\|_\infty^2}(\|d\|_\infty^2 f(0) + 2\|d\|^2) < 2f(0) + 2\|d\|^2.$$

But this inequality follows since $0 < \theta < 1$ and $\frac{2\theta}{\|d\|_\infty^2} \le 1$. Hence, $f(t) < f(0)$, which is the claim of Theorem 3.2-(b).

3.3 Divergence bounds

Though the centered point $\hat{w}(\mu)$ is unknown, the Newton direction $d_N(w,\mu)$ can provide a lower bound h_{lb} of the divergence $h(w, \hat{w}(\mu))$ for any $w \in \operatorname{int}\mathcal{K}$ and $\mu > 0$. Under a norm condition, we can also obtain an upper bound h_{ub} and relative-error estimates; precisely, we can obtain h_{ub} and $\alpha \ge 1$ satisfying

$$h(w,\hat{w}(\mu)) \ge h_{lb} \ge \frac{1}{\alpha}h(w,\hat{w}(\mu)) \qquad h(w,\hat{w}(\mu)) \le h_{ub} \le \alpha h(w,\hat{w}(\mu)). \tag{19}$$

These bounds use the direct-sum decomposition $d_N = d_1 - d_2$ from Proposition 2.4 induced by the subspaces $\mathcal{L}_w := \{Q(w^{-1/2})u : u \in \mathcal{L}\}$ and $\mathcal{L}_w^\perp = \{Q(w^{1/2})u : u \in \mathcal{L}^\perp\}$.

Theorem 3.3. *For $\mu > 0$ and $w \in \operatorname{int}\mathcal{K}$, let $d = d_N(w,\mu)$, $d_1 = \operatorname{proj}_{\mathcal{L}_w^\perp} d$, and $d_2 = -\operatorname{proj}_{\mathcal{L}_w} d$. The following statements hold:*

(a) $h(w, \hat{w}(\mu)) \geq h_{lb}$ *for* $h_{lb} := \frac{\|d\|^2}{1+\|d_1+d_2\|_\infty}$.

(b) If $\|d_1 + d_2\|_\infty < 1$, *then* $h(w, \hat{w}(\mu)) \leq h_{ub}$ *for* $h_{ub} := \frac{\|d\|^2}{1-\|d_1+d_2\|_\infty}$. *Further, the relative-error estimates* (19) *hold for* $\alpha = \frac{1+\|d_1+d_2\|_\infty}{1-\|d_1+d_2\|_\infty}$.

Proof. Let $a = Q(w^{1/2})\hat{w}(\mu)^{-1}$, $z = a + a^{-1} - 2e$ and $g = a - a^{-1}$. Proposition 2.4 implies that $d_1 = \operatorname{proj}_{\mathcal{L}_w^\perp}(a^{-1} - e)$ and $d_2 = \operatorname{proj}_{\mathcal{L}_w}(a - e)$. From $d = d_1 - d_2$, we conclude

$$\operatorname*{proj}_{\mathcal{L}_w^\perp}(g + 2d) = \operatorname*{proj}_{\mathcal{L}_w^\perp}(a - a^{-1} + 2(a^{-1} - e)) = \operatorname*{proj}_{\mathcal{L}_w^\perp}(a + a^{-1} - 2e) = \operatorname*{proj}_{\mathcal{L}_w^\perp} z,$$

and, similarly, that $\operatorname{proj}_{\mathcal{L}_w}(g+2d) = -\operatorname{proj}_{\mathcal{L}_w} z$. This implies that $\langle g+2d, d\rangle = \langle z, d_1+d_2\rangle$. Hence,

$$-\|z\|_1\|d_1 + d_2\|_\infty \leq -\langle g + 2d, d\rangle \leq \|z\|_1\|d_1 + d_2\|_\infty.$$

But from Lemma 3.6, we also have that $-\langle g + 2d, d\rangle = h(w, \hat{w}(\mu)) - \|d\|^2$. Hence,

$$-\|z\|_1\|d_1 + d_2\|_\infty \leq h(w, \hat{w}(\mu)) - \|d\|^2 \leq \|z\|_1\|d_1 + d_2\|_\infty.$$

Using the fact that $\|z\|_1 = h(w, \hat{w}(\mu))$ from Lemma 3.4(a) and rearranging these inequalities gives

$$h(w, \hat{w}(\mu))(1 + \|d_1 + d_2\|_\infty) \geq \|d\|^2 \geq h(w, \hat{w}(\mu))(1 - \|d_1 + d_2\|_\infty).$$

Dividing by $1+\|d_1+d_2\|_\infty$ proves the formula and error estimate for h_{lb}. Dividing by $1-\|d_1+d_2\|_\infty$ proves the same for h_{ub}. ☐

Observe that we also obtain valid bounds by replacing $\|d_1 + d_2\|_\infty$ with $\|d_N(w, \mu)\|$ given that $\|d_1 + d_2\|_\infty \leq \|d_1 + d_2\| = \|d_1 - d_2\| = \|d_N(w, \mu)\|$. This in turn allows us to bound the size of Newton steps assuming bounds on divergence.

Corollary 3.1. *For* $\mu > 0$ *and* $w \in \operatorname{int}\mathcal{K}$, *suppose that* $h(w, \hat{w}(\mu)) \leq \frac{1}{2}$. *Then,* $\|d_N(w, \mu)\| \leq 1$.

Proof. Replacing $\|d_1 + d_2\|_\infty$ with $\|d_N(w, \mu)\|$ in the Theorem 3.3 lower bound yields

$$h(w, \hat{w}(\mu)) \geq \frac{\|d_N(w, \mu)\|^2}{1 + \|d_N(w, \mu)\|}, \tag{20}$$

which proves the claim. ☐

The inequalities of this section bear strong resemblance to inequalities [27, Theorems 4.1.7–8] derived for *self-concordant barrier functions*, standard objects in IPM analysis. We will elaborate on this connection in Section 3.5.

3.4 Quadratic convergence of Newton's method

We have seen that the Newton direction bounds the reduction in divergence (Theorem 3.2). Divergence in turn bounds the size of a full Newton step (Corollary 3.1). Combining these results proves quadratic convergence of the sequence $w_0, w_1, \ldots, w_m$ generated by Newton's method.

Theorem 3.4. *For* $\mu > 0$ *and* $w_0 \in \operatorname{int}\mathcal{K}$, *recursively define* w_i *via the iterations* $w_{i+1} = Q(w_i^{1/2})\exp(d_N(w_i, \mu))$. *If* $h(w_0, \hat{w}(\mu)) \leq \beta \leq \frac{1}{2}$, *then* $h(w_i, \hat{w}(\mu)) \leq \beta^{2^i}$.

Proof. Let $h_i = h(w_i, \hat{w}(\mu))$ and $d_i = d_N(w_i, \mu)$. Make the inductive hypothesis that $h_i \leq 1/2$. Then $\|d_i\| \leq 1$ by Corollary 3.1, implying $h_{i+1} \leq \frac{1}{2} h_i \|d_i\|_\infty^2$ by Theorem 3.2 (a), which shows $h_{i+1} \leq 1/2$. Since $h_0 \leq 1/2$ by assumption, we conclude that both $h_i \leq 1/2$ and $\|d_i\| \leq 1$ hold for all i. Further, for all i,

$$h_{i+1} \leq \frac{1}{2} h_i \|d_i\|_\infty^2 \leq \frac{1}{2} h_i \|d_i\|^2 \leq \frac{1}{2}(\|d_i\| + 1) h_i^2,$$

where the last inequality is (20). Since $\|d_i\| \leq 1$, we have $h_{i+1} \leq h_i^2$. Hence, $h_i \leq (h_0)^{2^i} \leq \beta^{2^i}$. □

Combining this with our previous bounds relating divergence and geodesic distance (Lemma 3.2) leads to a proof of the centering lemma, reproduced below.

Lemma 2.2 (Centering)**.** *For $\mu > 0$ and $w_0 \in \operatorname{int} \mathcal{K}$, recursively define w_i via the iterations $w_{i+1} = Q(w_i^{1/2}) \exp(d_N(w_i, \mu))$. If $\delta(w_0, \hat{w}(\mu)) \leq q^{-1}(\beta)$ for $0 \leq \beta \leq \frac{1}{2}$, then $\delta(w_i, \hat{w}(\mu))^2 \leq \beta^{2^i}$.*

Proof. By Lemma 3.2, we conclude that $h(w_0, \hat{w}(\mu)) \leq \beta \leq \frac{1}{2}$. By Theorem 3.4, this implies that $h(w_i, \hat{w}(\mu)) \leq \beta^{2^i}$, which, since $\delta(w_i, \hat{w}(\mu))^2 \leq h(w_i, \hat{w}(\mu))$, proves the claim. □

3.5 Energy interpretation and self-scaled barriers

We conclude this section by highlighting connections with the literature. This is strictly not needed for our analysis, but helps put our work into context. The main object of study is the *energy functional* $E(\gamma)$, defined on smooth curves $\gamma : [0, 1] \to \operatorname{int} \mathcal{K}$ via

$$E(\gamma(t)) := \int_0^1 \|\gamma'(t)\|_{\gamma(t)}^2 dt,$$

where $\|v\|_u^2 = \langle v, Q(u)^{-1} v \rangle$. In other words, energy is defined by replacing $\|\gamma'(t)\|_{\gamma(t)}$ with $\|\gamma'(t)\|_{\gamma(t)}^2$ in the arc-length integral (Section 1).

We next show that divergence $h(u, v)$ is precisely the energy of the line-segment connecting u and v. This immediately implies the Lemma 3.2 inequality $\delta(u, v)^2 \leq h(u, v)$ given that $\delta(u, v)^2 \leq E(\gamma)$ holds for any curve $\gamma(t)$ connecting u and v; see [6, Chapter 9, Lemma 2.3].

Proposition 3.1. *For $u, v \in \operatorname{int} \mathcal{K}$, let $\ell(t) := u + t(v - u)$. Then $h(u, v) = E(\ell(t))$, i.e.,*

$$h(u, v) = \int_0^1 \langle v - u, Q(\ell(t))^{-1}(v - u) \rangle dt. \tag{21}$$

Proof. By definition, $h(u, v) = \langle u, v^{-1} \rangle + \langle u^{-1}, v \rangle - 2n$. Rearranging shows $h(u, v) = \langle v - u, u^{-1} - v^{-1} \rangle$. Since $-Q(z)^{-1}$ is the Jacobian of the inverse map $z \mapsto z^{-1}$ [8, Proposition II.3.3], we can also write

$$v^{-1} - u^{-1} = \int_0^1 -Q(\ell(t))^{-1}(v - u) dt.$$

Hence, $h(u, v) = \langle v - u, u^{-1} - v^{-1} \rangle = \int_0^1 \langle v - u, Q(\ell(t))^{-1}(v - u) \rangle dt$, as claimed. □

In view of this result, we can bound $h(u, v)$ by bounding $Q(\ell(t))^{-1}$. For this, we use standard Hessian bounds for *self-scaled barrier functions* [30], which generalize $f(u) := \log \det u^{-1}$ and are central in IPM analysis over *self-scaled cones*. Specifically, we interpret $Q(u)^{-1}$ as the Hessian of $f(u)$ and invoke [29, Theorem 4.1]; see [29, 13] for the definition of self-scaled barriers and proof that $f(u)$ is self-scaled.

Proposition 3.2. *For $u, v \in \operatorname{int}\mathcal{K}$, let $\Delta = Q(u)^{-1/2}(u-v)$. If $\|\Delta\|_\infty < 1$, then*

$$\frac{\|\Delta\|^2}{1+\|\Delta\|_\infty} \le h(u,v) \le \frac{\|\Delta\|^2}{1-\|\Delta\|_\infty}. \tag{22}$$

Proof. Let $H(z) := Q(z)^{-1}$ and $\sigma_z(p) := \inf\{\beta \ge 0 : \beta z - p \in \mathcal{K}\}$. Then [29, Theorem 4.1] states

$$\frac{1}{(1+t\sigma_u(-p))^2} H(u) \preceq H(u-tp) \preceq \frac{1}{(1-t\sigma_u(p))^2} H(u)$$

for all $t \in [0, 1/\sigma_u(p))$, where we take $1/\sigma_u(p) = +\infty$ if $\sigma_u(p) = 0$. Taking $p = u - v$ and observing

$$\sigma_u(p) \le |\lambda_{\max}(Q(u)^{-1/2}p)| \le \|\Delta\|_\infty, \qquad \sigma_u(-p) \le |\lambda_{\min}(Q(u)^{-1/2}p)| \le \|\Delta\|_\infty,$$

gives, for all $t \in [0, 1/\|\Delta\|_\infty)$, the bounds

$$\frac{1}{(1+t\|\Delta\|_\infty)^2} H(u) \preceq H(u+t(v-u)) \preceq \frac{1}{(1-t\|\Delta\|_\infty)^2} H(u).$$

When $\|\Delta\|_\infty < 1$, we can substitute each bound into the energy integral (21) and apply the identities

$$\int_0^1 \frac{dt}{(1+t\|\Delta\|_\infty)^2} = \frac{1}{1+\|\Delta\|_\infty}, \qquad \int_0^1 \frac{dt}{(1-t\|\Delta\|_\infty)^2} = \frac{1}{1-\|\Delta\|_\infty},$$

to conclude that

$$\frac{\langle H(u)p, p\rangle}{1+\|\Delta\|_\infty} \le h(u,v) \le \frac{\langle H(u)p, p\rangle}{1-\|\Delta\|_\infty}.$$

Since $\|\Delta\|^2 = \langle H(u)p, p\rangle$, the claim follows. ∎

When applied to $h(w, \hat{w}(\mu))$, the bounds (22) are similar to those from Theorem 3.3, but not equivalent. In particular, (22) requires the unknown quantity $\hat{w}(\mu)$ to construct Δ, whereas Theorem 3.3 uses the Newton direction $d_N(w, \mu)$. Further, (22) does not preserve the symmetry $h(w, \hat{w}) = h(w^{-1}, \hat{w}^{-1})$, as replacing $\Delta = Q(w)^{-1/2}(w - \hat{w})$ with $\Delta = Q(w^{-1})^{-1/2}(w^{-1} - \hat{w}^{-1})$ leads to different bounds.

We also note that (22) still holds if $\|\Delta\|_\infty$ is replaced with $\|\Delta\|$. With this replacement, it is a special case of [27, Theorem 4.1.7–8], which holds for arbitrary *self-concordant* functions, a superset of self-scaled functions that are central in IPM analysis over general convex sets.

4 Long-step algorithm

When proving the convergence of `shortstep` (Figure 1), we established results that suggest an alternative algorithm. This alternative uses our divergence upper-bound (Theorem 3.3) to loosely track the central path and damped Newton steps to ensure that divergence strictly decreases (Theorem 3.2). We state this algorithm in Figure 2 using the following notation for the divergence upper-bound:

$$h_{ub}(w,\mu) = \begin{cases} \frac{\|d_N(w,\mu)\|^2}{1-\|d_1(w,\mu)+d_2(w,\mu)\|_\infty} & \|d_1+d_2\|_\infty < 1 \\ \infty & \textit{otherwise}, \end{cases}$$

(Here $d_1(w, \mu)$ and $d_2(w, \mu)$ denote the direct-summands of the Newton direction $d_N(w, \mu)$; see Proposition 2.4.) We name this algorithm `longstep` in reference to classical long-step IPMs [43], which also loosely track the central path.

Algorithm `longstep`$(w_0, \mu_0, \mu_f,\ \epsilon)$
$\mu \leftarrow \mu_0,\ w \leftarrow w_0$
while $\mu > \mu_f$ **do**
  $w \leftarrow$ `center`(w, μ, α)
  $\mu \leftarrow \inf\{\mu > 0 : h_{ub}(w, \mu) \leq \beta\}$
end
return `center`(w, μ, ϵ)

Procedure `center`(w_0, μ, ϵ)
$w \leftarrow w_0$
while $h_{ub}(w, \mu) > \epsilon$ **do**
  $d \leftarrow d_N(w, \mu)$
  $\gamma \leftarrow \max\{1, \frac{1}{2\theta}\|d\|_\infty^2\}$
  $w \leftarrow Q(w^{1/2}) \exp(\frac{1}{\gamma} d)$
end
return w

Figure 2: A long-step algorithm (left) and centering procedure (right). The parameters β and α control distance to the central path and θ the transition to damped Newton steps. The algorithms globally convergence on all inputs if $1 > \theta > 0$ and $\beta > \alpha > 0$.

The next theorem shows that `longstep` is *globally convergent*, i.e., it can be initialized arbitrarily. To prove this, we exploit the fact that the sublevel sets of divergence h are compact, which implies positive lower bounds on certain progress measures. This theorem also shows scale invariance (Section 2.4). This follows from Proposition 2.5 given that the step-size γ^{-1} and divergence bound h_{ub} depend only on the eigenvalues of $d_1 + d_2$ and $d_1 - d_2$.

Theorem 4.1. *If $1 > \theta > 0$ and $\beta > \alpha > 0$, then the algorithm* `longstep` *and its subroutine* `center` *(Figure 2) have the following properties.*

(a) For all inputs $w_0 \in \operatorname{int} \mathcal{K}$ and $(\mu_0, \mu_f, \epsilon) > 0$, `longstep` *terminates and returns w satisfying $h(w, \hat{w}(\mu)) \leq \epsilon$ for $\mu \leq \mu_f$. Further, it monotonically decreases μ.*

(b) For all inputs $w_0 \in \operatorname{int} \mathcal{K}$ and $(\mu, \epsilon) > 0$, `center` *terminates and returns w satisfying $h(w, \hat{w}(\mu)) \leq \epsilon$. Further, it monotonically decreases $h(w, \hat{w}(\mu))$.*

(c) Both `center` *and* `longstep` *are scale invariant.*

Proof. To prove statements (a)-(b), we first show compactness of the set

$$S(\zeta) := \{(w, \mu) : h(w, \hat{w}(\mu)) \leq \zeta, \mu_f \leq \mu \leq \mu_0\}.$$

It is closed because $(w, \mu) \mapsto h(w, \hat{w}(\mu))$ is continuous. To see it is bounded, note that the eigenvalues of $\hat{w}(\mu)$ and $\hat{w}(\mu)^{-1}$ are bounded below by some $c > 0$ on $\mu_f \leq \mu \leq \mu_0$, implying that

$$\zeta \geq h(w, \hat{w}(\mu)) \geq c\langle e, w + w^{-1}\rangle - 2n \geq c\|w\|_1 - 2n$$

when $(w, \mu) \in S(\zeta)$. Hence, if $(w, \mu) \in S(\zeta)$ then $\|w\|_1 + |\mu|$ is bounded, implying $S(\zeta)$ is compact.

To prove (b), let $\zeta = h(w_0, \hat{w}(\mu))$. Let $\Delta(w, \mu)$ denote the decrease in h after one Newton step from w, i.e.,

$$\Delta(w, \mu) = h(\hat{w}(\mu), w) - h(\hat{w}(\mu), w')$$

where $w' = Q(w^{1/2}) \exp(\frac{1}{\gamma} d)$. After N steps, $h(w, \hat{w}(\mu)) \leq h(w_0, \hat{w}(\mu)) - \Delta_* N$, where

$$\Delta_* := \inf_{w,\mu} \{\Delta(w, \mu) : h_{ub}(\hat{w}(\mu), w) \geq \epsilon, (w, \mu) \in S(\zeta)\}.$$

Compactness of $S(\zeta)$ implies Δ_* is attained, which implies $\Delta_* > 0$ by our step-size rule and Theorem 3.2. Since $h \geq 0$, we conclude that `center` must terminate before $\Delta_* N > h(w_0, \hat{w}(\mu))$.

To prove statement (a), note that $\mu \leq k_*^{-M}\mu_0$ after M iterations, where

$$k_* := \inf_{w,\mu,k}\{k \geq 1 : (w,\mu) \in S(\alpha), h_{ub}(w,\mu) \leq \alpha, h_{ub}(w,(1/k)\mu) = \beta\}.$$

Compactness of $S(\alpha)$ implies k_* is attained, which implies that $k_* > 1$ since $\beta > \alpha$. This implies $\mu < \mu_f$ eventually holds, implying termination of `longstep`.

Finally, statement (c) follows from Proposition 2.5 and Lemma 2.3, given that γ and h_{ub}, viewed as functions of d_1 and d_2, are invariant under transformation by an orthogonal automorphism M, i.e., $\gamma(d_1, d_2) = \gamma(Md_1, Md_2)$ and $h_{ub}(d_1, d_2) = h_{ub}(Md_1, Md_2)$. □

We close this section with practical matters related to implementation. Specifically, we show how to efficiently evaluate the divergence bound $h_{ub}(w,\mu)$ for fixed w, how to find the Newton direction using a least-squares technique, how to evaluate geodesic updates without computation of $w^{1/2}$, and how to construct feasible points for the primal-dual pair (1).

4.1 Evaluating divergence for μ-selection

For fixed w, the divergence bound $h_{w,ub}(\mu) := h_{ub}(w,\mu)$ has a simple formula that admits efficient selection of μ at each iteration of `longstep`. To evaluate the formula, we only need to know μ and quantities involving the vector

$$g_w := \operatorname*{proj}_{\mathcal{L}_w^\perp} Q(w^{-1/2})x_0 + \operatorname*{proj}_{\mathcal{L}_w} Q(w^{1/2})s_0,$$

where we recall that $\mathcal{L}_w := \{Q(w^{-1/2})u : u \in \mathcal{L}\}$ and $\mathcal{L}_w^\perp = \{Q(w^{1/2})u : u \in \mathcal{L}^\perp\}$.

Proposition 4.1. *For $w \in \operatorname{int}\mathcal{K}$, let g_w have minimum and maximum eigenvalues $\lambda_{\min}$ and $\lambda_{\max}$. Let $k(\mu) = \min(\frac{1}{\sqrt{\mu}}\lambda_{\min}, 2 - \frac{1}{\sqrt{\mu}}\lambda_{\max})$. Then, for all $\mu > 0$,*

$$h_{w,ub}(\mu) = \begin{cases} \frac{\frac{1}{\mu}\|g_w\|^2 - 2\frac{1}{\sqrt{\mu}}\operatorname{tr} g_w + n}{k(\mu)} & k(\mu) > 0 \\ \infty & \textit{otherwise.} \end{cases}$$

Proof. Let $d = d_N$ and let d_1 and d_2 be as in Proposition 2.4. Suppose that $h_{ub}(w,\mu)$ is finite, i.e., $1 - \|d_1 + d_2\|_\infty > 0$. Then we have that

$$h_{ub} = \frac{\|d\|^2}{1 - \|d_1 + d_2\|_\infty}, \qquad d_1 + d_2 = \frac{1}{\sqrt{\mu}}g_w - e.$$

Hence, $\|d_1 + d_2\|_\infty$ is the max of $1 - \frac{1}{\sqrt{\mu}}\lambda_{\min}$ and $\frac{1}{\sqrt{\mu}}\lambda_{\max} - 1$. The claimed denominator $k(\mu)$ follows using the identity

$$1 - \max(1-a, b-1) = 1 + \min(a-1, 1-b) = \min(a, 2-b).$$

The identity for $\|d\|^2$ follows by expanding $\|\frac{1}{\sqrt{\mu}}g_w - e\|^2$ and observing that $\|d\| = \|d_1 + d_2\|$. □

4.2 Newton direction via least squares

Interior-point methods typically find search directions by solving least-squares problem of the form

$$\text{minimize}_y \;\; \frac{1}{2}y^TA^*W(x,s)Ay - f^Ty \text{ subject to } By = g,$$

where $W(x,s)$ is a positive-definite weighting matrix induced by the current iterate (x,s) and (A,B,f,g) are parameters induced by the affine constraints $x_0+\mathcal{L}$ and $s_0+\mathcal{L}^\perp$. Equivalently, they solve linear systems of the form

$$\begin{bmatrix} A^*W(x,s)A & B^* \\ B & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

for which specialized algorithms exist (e.g., [18]). Such a system can also yield the Newton direction $d_N(w,\mu)$. This, of course, is not surprising given its construction via orthogonal projection (Proposition 2.4). Nevertheless, we give this system explicitly for affine constraints of the form:

$$s_0+\mathcal{L}^\perp = \{c - Ay : By = g, \ \ y \in \mathbb{R}^m\}, \qquad x_0 + \mathcal{L} = \{x \in \mathcal{J} : \exists z \in \mathbb{R}^d \ \ A^*x + B^*z = b\}, \quad (23)$$

where $(y,z) \in \mathbb{R}^m \times \mathbb{R}^d$ denote additional variables, $A : \mathbb{R}^m \to \mathcal{J}$ and $B : \mathbb{R}^m \to \mathbb{R}^d$ are linear maps with adjoint operators $A^* : \mathcal{J} \to \mathbb{R}^m$ and $B^* : \mathbb{R}^d \to \mathbb{R}^m$, and $(b,g,c) \in \mathbb{R}^m \times \mathbb{R}^d \times \mathcal{J}$ are fixed parameters.

In this notation, the Newton direction becomes the d that for some (y,z) solves

$$A^*(Q(w^{1/2})(e+d)) = \frac{1}{\sqrt{\mu}}b - B^*z, \quad Q(w^{-1/2})(e-d) = \frac{1}{\sqrt{\mu}}c - Ay, \quad By = \frac{1}{\sqrt{\mu}}g. \qquad (24)$$

Eliminating d and using $Q(w) = Q(w^{1/2})Q(w^{-1/2})^{-1}$ yields a system with the desired form. Note by modifying the right-hand-side of this system, we can also construct the direct-summands d_1 and d_2 of Proposition 2.4.

Proposition 4.2. *For $w \in \operatorname{int} \mathcal{K}$ and $\mu > 0$, let $(y,z) \in \mathbb{R}^m \times \mathbb{R}^d$ solve the least-squares system*

$$\begin{bmatrix} A^*Q(w)A & B^* \\ B & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\mu}}(b + A^*Q(w)c) - 2A^*w \\ \frac{1}{\sqrt{\mu}}g \end{bmatrix}.$$

Then, the Newton direction satisfies $d_N(w,\mu) = e - Q(w^{1/2})(\frac{1}{\sqrt{\mu}}c - Ay)$.

Proof. From the second equation of (24), we conclude that $d = e - Q(w^{1/2})(\frac{1}{\sqrt{\mu}}c - Ay)$. Substituting into the first equation yields

$$\begin{aligned} \frac{1}{\sqrt{\mu}}b - B^*z &= A^*Q(w^{1/2})\Big(2e - Q(w^{1/2})(\frac{1}{\sqrt{\mu}}c - Ay)\Big) \\ &= 2A^*w - \frac{1}{\sqrt{\mu}}A^*Q(w)c + A^*Q(w)Ay. \end{aligned}$$

Rearranging terms proves the claim. □

4.3 Evaluation of geodesic updates

For $w \in \operatorname{int} \mathcal{K}$ and $v \in \mathcal{J}$, let $g(w,v) := Q(w^{1/2}) \exp(Q(w^{1/2})v)$. By Proposition 4.2, we see that $d_N(w,\mu) = e + Q(w^{1/2})v$ for a particular point $v \in \mathcal{J}$. Hence, the geodesic update $Q(w^{1/2})\exp(d_N)$ satisfies, for particular $\kappa > 0$, the equation

$$Q(w^{1/2})\exp(d_N) = \frac{1}{\kappa}g(w,v). \qquad (25)$$

We next show that $g(w,v)$ can be computed without constructing the square root $w^{1/2}$. Letting $z = Q(w^{1/2})v$, the key idea is expressing the power series of $\exp(z)$ in terms of $Q(z)$ and applying the identity $Q(z) = Q(w^{1/2})Q(v)Q(w^{1/2})$ from Appendix A.

Proposition 4.3. *If $w \in \mathcal{K}$ and $v \in \mathcal{J}$, then $g(w,v) = \sum_{n=0}^{\infty} \frac{1}{(2n)!}(Q(w)Q(v))^n(w + \frac{1}{2n+1}Q(w)v)$.*

Proof. For arbitrary z, we have that $z^{2n} = Q(z)^n e$ and $z^{2n+1} = Q(z)^n z$, which implies that

$$Q(w^{1/2})\exp(z) = Q(w^{1/2})\sum_{i=0}^{\infty}\frac{1}{(2n)!}Q(z)^n(e + \frac{1}{2n+1}z). \tag{26}$$

For $z = Q(w^{1/2})v$, we have, using $Q(z)^n = (Q(w^{1/2})Q(v)Q(w^{1/2}))^n$, that

$$Q(w^{1/2})Q(z)^n = (Q(w)Q(v))^n Q(w^{1/2}). \tag{27}$$

Substituting (27) into (26) proves the claim. □

An alternative formula for $g(w,v)$ is available when $\mathcal{J}$ is *special*, i.e., if the product operation $x \circ y$ satisfies $x \circ y = \frac{1}{2}(xy + yx)$ for an associative product xy. An example of a special algebra is the set of symmetric matrices with product $\frac{1}{2}(XY + YX)$, where XY denotes ordinary matrix multiplication. For any special algebra, the quadratic representation satisfies $Q(x)y = xyx$ for all $x, y \in \mathcal{J}$. This fact allows us to compute $g(w,v)$ by evaluating $\exp(wv) := \sum_{d=0}^{\infty}\frac{1}{d!}(wv)^d$, i.e., the exponential map induced by the associative product, at the point wv.

Proposition 4.4. *If $\mathcal{J}$ is special then $g(w,v) = \exp(wv)w$ for all $v \in \mathcal{J}$ and $w \in \operatorname{int}\mathcal{K}$.*

Proof. Let $z = Q(w^{1/2})v$. Since $z = w^{1/2}vw^{1/2}$, we have that

$$w^{1/2}z^d w^{1/2} = w^{1/2}w^{1/2}v(wv)^{d-1}w^{1/2}w^{1/2} = (wv)^d w$$

Hence,

$$Q(w^{1/2})\exp(Q(w^{1/2})v) = w^{1/2}\left(\sum_{d=0}^{\infty}\frac{1}{d!}z^d\right)w^{1/2} = \left(\sum_{d=0}^{\infty}\frac{1}{d!}(wv)^d\right)w = \exp(wv)w.$$

□

Note that for symmetric matrices, $\exp(WV)$ is the usual matrix exponential evaluated at the matrix product WV. One can evaluate the matrix exponential using a power series or Pade approximation [24]. In total, the entire evaluation of $Q(w^{1/2})\exp(d_N)$ can be done without eigenvalue decomposition or square roots.

Remark 4. *The quantity $g(w,v) := Q(w^{1/2})\exp(Q(w^{1/2})v)$ can be written using the manifold exponential map (Section 2.2) as*

$$g(w,v) = \operatorname*{Exp}_w(Q(w)v).$$

For the algebra of symmetric matrices ($\mathcal{J} = \mathbb{S}^n$), it's known (e.g., [35]) that Exp_W satisfies

$$\operatorname*{Exp}_W(Z) = \exp(ZW^{-1})W.$$

This provides an alternative proof of Proposition 4.4 for the special case of $\mathcal{J} = \mathbb{S}^n$. Precisely, taking $Z = Q(W)V = WVW$, we deduce that

$$g(W,V) = \operatorname*{Exp}_W(Q(W)V) = \exp((WVW)W^{-1})W = \exp(WV)W,$$

as claimed.

4.4 Feasible points

Since the presented algorithms update w along geodesics, the point $\sqrt{\mu}(w, w^{-1})$ only satisfies the affine constraints of the primal-dual pair (1) in the limit. Nevertheless, under a norm condition, we can *always* produce a feasible (x, s) from the Newton direction $d_N(w, \mu)$.

Proposition 4.5. *For $w \in \operatorname{int} \mathcal{K}$ and $\mu > 0$, let $d = d_N(w, \mu)$ and*

$$x = \sqrt{\mu} Q(w^{1/2})(e + d), \qquad s = \sqrt{\mu} Q(w^{-1/2})(e - d).$$

If $\|d\|_\infty \leq 1$, then (x, s) is feasible for (1).

Proof. By definition of the Newton direction (Definition 1.1), it holds that $x \in x_0 + \mathcal{L}$ and $s \in s_0 + \mathcal{L}^\perp$. Further, since $\|d\|_\infty \leq 1$, we have that $e \pm d \in \mathcal{K}$. Finally, $x, s \in \mathcal{K}$ given that $Q(z)y \in \mathcal{K}$ for all $z \in \mathcal{J}$ and $y \in \mathcal{K}$. □

In light of Section 2.5, this proposition gives a sufficient condition for feasibility of a full Nesterov-Todd step when $x = \mu s^{-1}$. It can therefore be compared with [11, Lemma 3.3].

5 Computational results

We provide a series of computational experiments that illustrate key features of our algorithms and the performance of an implementation. First, we illustrate that `longstep` executes far fewer iterations than `shortstep`, despite its weaker theoretical guarantees. We then illustrate `longstep`-performance on a range of symmetric cones, including the exceptional cone and the psd Hermitian matrices with complex and quaternion entries; to our knowledge, these are the first computational results for the quaternion and the exceptional cone. We next demonstrate global convergence of the centering procedure `center`. Finally, we compare our `longstep`-implementation `conex` (pronounced CON-ex) to `sdtp3`, a widely used solver that is based on the Nesterov-Todd algorithm [40].

For each instance, the affine constraints are of the form (23) but with the equality constraints $By = g$ and associated dual variable z omitted. The operator $A : \mathbb{R}^m \to \mathcal{J}$ is randomly generated. Unless stated otherwise, the cone $\mathcal{K}$ is the set of psd matrices $\mathbb{S}^n_+$, the cost vectors are the identity, i.e., $x_0 = e$ and $s_0 = e$, and $m = 10$.

5.1 Algorithm comparison

We compare (Figure 3) the total number of Newton steps `longstep` and `shortstep` execute to update an initial centered point $\hat{w}(\mu)$ to $\hat{w}(\frac{1}{k}\mu)$ where $k = 25000^2$. For each n, we compute the average number of steps executed by `longstep` over twenty random problems semidefinite programs ($\mathcal{K} = \mathbb{S}^n_+$). The number of steps executed by `shortstep` is independent of the problem instance, so no averaging is necessary. As shown, `longstep` provides a significant improvement over `shortstep`. We also see that `longstep` enters a steady-state regime in which it reduces the centering parameter μ at a rate that is independent of n.

For `longstep`, we chose a divergence bound β that grows linearly with n; specifically, we took $\beta = 100n$ where 100 is chosen arbitrarily and the dependence on n is intended to compensate for the $\|d_N\|^2$-dependence of the divergence upper-bound h_{ub}. We chose a re-centering tolerance of $\alpha = 10$, and a final centering tolerance of $\epsilon = \frac{1}{200}$. For `shortstep`, we selected the centering-parameter update k and the number of inner iterations m using Theorem 2.1 with the parameter values $(\beta, \epsilon) = (\frac{1}{2}, \sqrt{\frac{1}{200}})$.

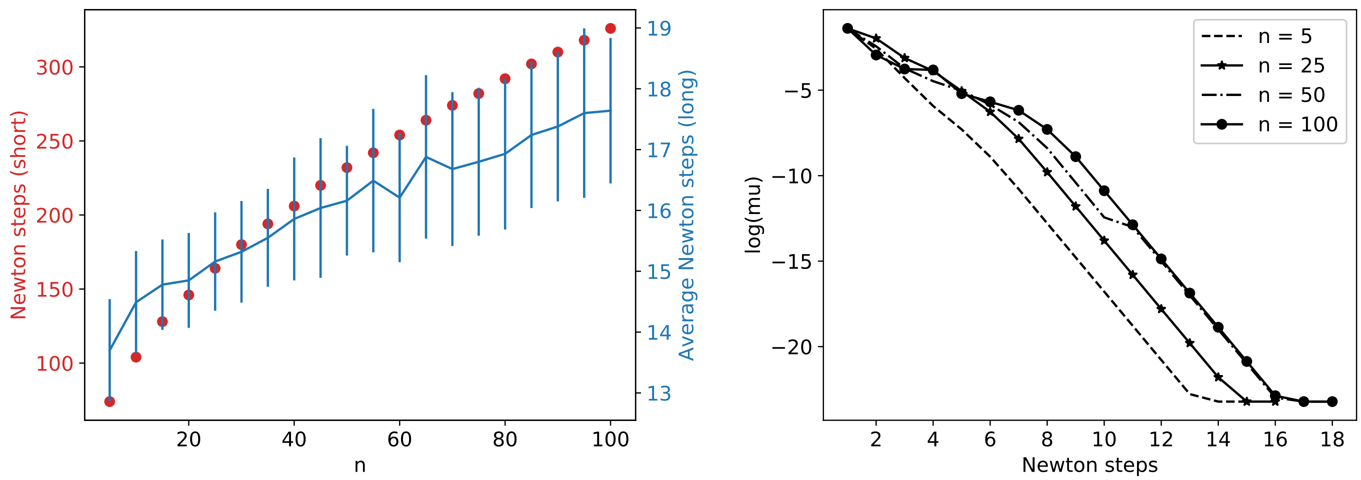

Figure 3: Total Newton steps vs n for `shortstep` and `longstep` (left) on random SDPs ($\mathcal{K} = \mathbb{S}^n_+$). (Note the different scales.) Typical decrease in centering parameter μ for `longstep` (right).

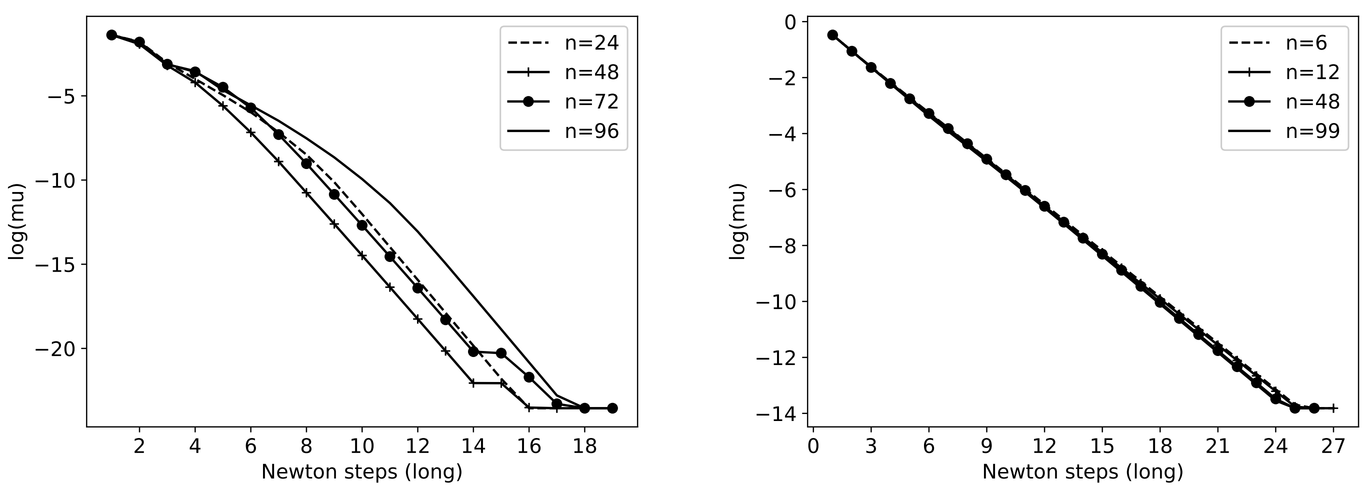

Figure 4: Typical decrease in centering parameter μ for `longstep` on a p-fold product of a special cone $\mathcal{K}_s$ (left) and p-fold product of an exceptional cone $\mathcal{K}_{ex}$ (right). The p-fold product of $\mathcal{K}_s$ has rank $n = 24p$ and and the product of $\mathcal{K}_{ex}$ has rank $n = 3p$.

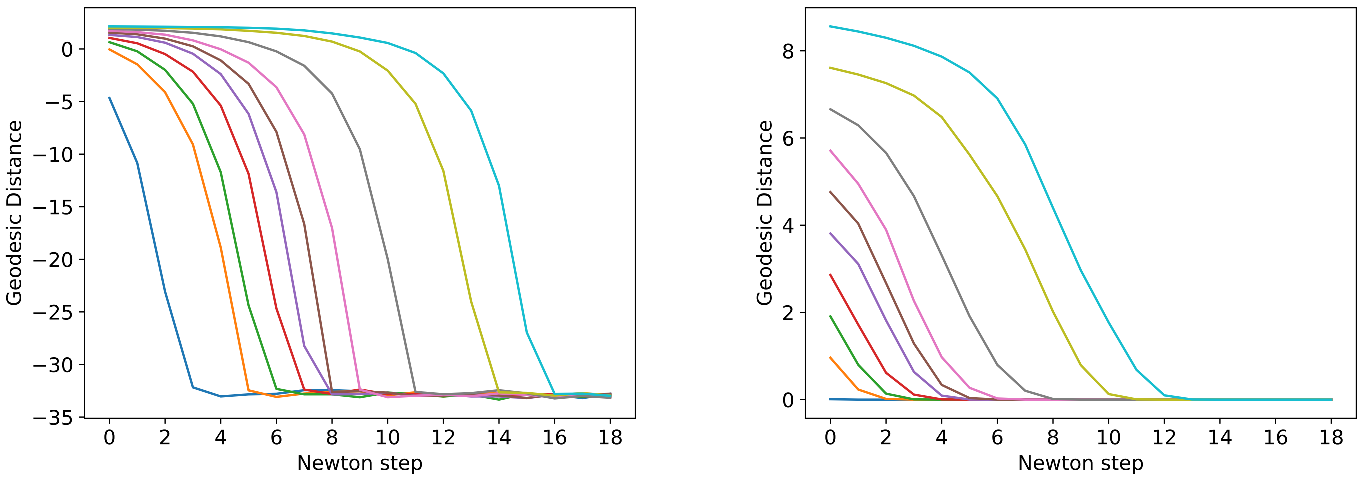

Figure 5: Convergence of the centering procedure `center` for different initialization points in log (left) and linear (right) scalings. Plotted is geodesic distance to the centered point $\hat{w}(\mu)$.

5.2 Products of special and exceptional cones

For a Euclidean-Hurwitz algebra $\mathbb{D}$, let $\mathbf{H}(\mathbb{D})^q$ denote the Hermitian matrices of order q endowed with multiplication $X \circ Y = \frac{1}{2}(XY + YX)$ and inner-product $\operatorname{Tr} X \circ Y$. If $\mathbb{D}$ is the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, or the quaternions $\mathbb{Q}$, then $\mathbf{H}(\mathbb{D})$ is a special Jordan algebra of rank q. When $\mathbb{D}$ is the octonions $\mathbb{O}$, then $\mathbf{H}(\mathbb{D})^3$ is an exceptional Jordan algebra of rank three. (Note that in this notation, $\mathbf{H}(\mathbb{R})^q$ refers to the symmetric matrices $\mathbb{S}^q$.)

Denoting the cone-of-squares by $\mathbf{H}(\mathbb{D})^q_+$, we consider the following cones

$$\mathcal{K}_s := \mathbf{H}(\mathbb{R})^q_+ \times \mathbf{H}(\mathbb{C})^q_+ \times \mathbf{H}(\mathbb{Q})^q_+, \qquad \mathcal{K}_{ex} := \mathbf{H}(\mathbb{O})^3_+.$$

Fixing $q = 8$, we plot (Figure 4) the progress of `longstep` on randomly generated instances formulated over p-fold products $\mathcal{K} = \mathcal{K}_s \times \mathcal{K}_s \times \cdots \times \mathcal{K}_s$. We similarly plot progress for products of $\mathcal{K}_{ex}$. For the special cone $\mathcal{K}_s$, we used a divergence upper-bound of $\beta = 100n$, where $n = 3qp$, the rank of $\mathcal{K}$. For the exceptional cone $\mathcal{K}_{ex}$, we used a much tighter tolerance of $\beta = \frac{1}{10}n$, where $n = 3p$. This tighter tolerance was necessary to avoid numerical errors we suspect are related to errors in the power-series approximation of the geodesic update (Proposition 4.3). Recall use of this approximation is necessary because $H(\mathbb{D})^3$ is not special and hence prevents use of the associative matrix exponential (Proposition 4.4). Note that this tighter tolerance leads to more iterations compared with the special cone $\mathcal{K}_s$, but also more regular updates of μ that are essentially independent of n.

5.3 Global convergence

The procedure `center` used by `longstep` globally converges. That is, it always returns $\hat{w}(\mu)$ given an *arbitrary* initial point $w_0 \in \operatorname{int} \mathcal{K}$ and centering parameter $\mu > 0$. For a fixed problem instance, we plot convergence behavior for different initial conditions (Figure 5). We observe that convergence rate is divided into an initial and quadratic phase. These phases are expected from Theorem 3.2.

Parameters (n, m)	Solver Time (sec) spdt3	conex	Dual Residual spdt3	conex	Duality Gap sdpt3	conex	$\lambda_{\min}(x)$ sdpt3	conex	$\lambda_{\min}(s)$ sdpt3	conex
(20, 20)	1.1e-01	4.1e-03	1.4e-12	3.9e-12	1.4e-09	8.9e-10	3.2e-10	2.2e-10	3.2e-10	2.2e-10
(50, 50)	7.0e-01	1.1e-01	1.0e-12	1.5e-12	1.1e-09	1.9e-09	1.2e-10	2.0e-10	1.2e-10	2.0e-10
(100, 100)	3.1e+00	9.8e-01	2.0e-12	3.9e-12	9.7e-10	2.4e-09	7.6e-11	1.9e-10	7.6e-11	1.9e-10
(20, 40)	1.4e-01	1.6e-02	6.9e-11	7.7e-13	4.6e-10	7.2e-10	1.2e-10	1.8e-10	1.2e-10	1.8e-10
(50, 250)	1.8e+00	5.6e-01	1.5e-11	9.8e-12	5.3e-09	6.6e-10	1.5e-09	1.9e-10	1.5e-09	1.9e-10
(100, 1000)	1.9e+01	1.4e+01	3.4e-11	3.1e-11	6.5e-10	6.9e-10	1.7e-10	1.9e-10	1.7e-10	1.9e-10

Table 5.4.1: Solver time and residual comparison between our implementation `conex` and `sdpt3`. The dual residual and duality gap refer to $k_1^{-1}\|A^*x - b\|$ and $k_2^{-1}|\langle c, x\rangle - b^T y|$ for $k_1 = 1 + \|b\|_\infty$ and $k_2 = 1 + |\langle c, x\rangle| + |b^T y|$. Instances use a random $A : \mathbb{R}^m \to \mathbb{S}^n$ with $c := e$, $b := A^*e$ and $\mathcal{K} = \mathbb{S}^n_+$.

5.4 Implementation

An implementation is available at `www.github.com/frankpermenter/conex`. All symmetric cones are directly supported, including the Hermitian psd matrices with quaternion entries, the exceptional cone, and generalized Lorentz cones of the form $\{(x, t) \in \mathbb{R}^n \times \mathbb{R} : \langle x, x\rangle \le t^2\}$ for arbitrary inner-products $\langle \cdot, \cdot \rangle$. Table 5.4.1 compares performance with `sdpt3` configured to use Nesterov-Todd steps. Computations were performed in Ubuntu 18.04 on a machine with an `Intel Xeon CPU E5-2687W v4 @ 3.00GHz` processor. Both solvers were configured to use the same linear algebra (BLAS) library. For `conex` error calculations, (x, s, y) was constructed from w and the final Newton step using Proposition 4.5. Results show that `conex` achieves the same accuracy in less time across a range of different problem sizes. We note that the relative timing difference is reduced for problems with $m > n$. For such problems, both solvers spend more time on an identical calculation: construction and solution of the linear system from Section 4.2; see also Section 2.5.

Acknowledgements

We thank Richard Y. Zhang and anonymous reviewers for helpful comments on an earlier draft.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[2] F. Alvarez, J. Bolte, and O. Brahic. Hessian Riemannian gradient flows in convex programming. *SIAM journal on control and optimization*, 43(2):477–501, 2004.

[3] R. Bhatia. *Positive definite matrices*. Princeton University Press, 2009.

[4] I. M. Bomze, P. Mertikopoulos, W. Schachinger, and M. Staudigl. Hessian barrier algorithms for linearly constrained optimization problems. *SIAM Journal on Optimization*, 29(3):2100–2127, 2019.

[5] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2009.

[6] M. P. Do Carmo and J. Flaherty Francis. *Riemannian geometry*. Birkhauser Boston, 1992.

[7] L. Duembgen and D. E. Tyler. Geodesic convexity and regularized scatter estimators. *arXiv preprint arXiv:1607.05455*, 2016.

[8] J. Faraut and A. Korányi. *Analysis on symmetric cones*. Oxford University Press, 1994.

[9] L. Faybusovich. Euclidean Jordan algebras and interior-point algorithms. *Positivity*, 1(4):331–357, 1997.

[10] L. Faybusovich. Linear systems in Jordan algebras and primal-dual interior-point algorithms. *Journal of computational and applied mathematics*, 86(1):149–175, 1997.

[11] G. Gu, M. Zangiabadi, and C. Roos. Full Nesterov–Todd step infeasible interior-point method for symmetric optimization. *European Journal of Operational Research*, 214(3):473–484, 2011.

[12] M. Harandi, M. Salzmann, and R. Hartley. Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods. *IEEE transactions on pattern analysis and machine intelligence*, 40(1): 48–62, 2017.

[13] R. A. Hauser and Y. Lim. Self-scaled barriers for irreducible symmetric cones. *SIAM Journal on Optimization*, 12(3):715–723, 2002.

[14] C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz. An interior-point method for semidefinite programming. *SIAM Journal on Optimization*, 6(2):342–361, 1996.

[15] N. Karmarkar. Riemannian geometry underlying interior-point methods for linear programming. *Contemporary Mathematics*, 114:51–75, 1990.

[16] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior point algorithm for linear programming. In *Progress in mathematical programming*, pages 29–47. Springer, 1989.

[17] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization*, 7(1):86–125, 1997.

[18] C. L. Lawson and R. J. Hanson. *Solving least squares problems*. SIAM, 1995.

[19] J. D. Lawson and Y. Lim. The geometric mean, matrices, metrics, and more. *The American Mathematical Monthly*, 108(9):797–812, 2001.

[20] H. Lee and Y. Lim. Metric and spectral geometric means on symmetric cones. *Kyungpook Mathematical Journal*, 47(1), 2007.

[21] Y. Lim. Geometric means on symmetric cones. *Archiv der Mathematik*, 75(1):39–45, 2000.

[22] Y. Lim. Riemannian and Finsler structures of symmetric cones. *Trends in Mathematics*, 4(2):111–118, 2001.

[23] M. Moakher and P. G. Batchelor. Symmetric positive-definite matrices: From geometry to applications and visualization. In *Visualization and Processing of Tensor Fields*, pages 285–298. Springer, 2006.

[24] C. Moler and C. Van Loan. Nineteen dubious ways to compute the exponential of a matrix. *SIAM review*, 20(4):801–836, 1978.

[25] R. D. Monteiro. Primal–dual path-following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7(3):663–678, 1997.

[26] R. D. Monteiro and I. Adler. Interior path following primal-dual algorithms. Part I: Linear programming. *Mathematical programming*, 44(1-3):27–41, 1989.

[27] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003.

[28] Y. Nesterov and A. Nemirovski. Primal central paths and Riemannian distances for convex sets. *Foundations of Computational Mathematics*, 8(5):533–560, 2008.

[29] Y. E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations research*, 22(1):1–42, 1997.

[30] Y. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on optimization*, 8(2):324–364, 1998.

[31] Y. E. Nesterov, M. J. Todd, et al. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Foundations of Computational Mathematics*, 2(4):333–361, 2002.

[32] J. Peng, C. Roos, and T. Terlaky. *Self-regularity: a new paradigm for primal-dual interior-point algorithms*, volume 22. Princeton University Press, 2009.

[33] F. Permenter. Log-domain interior-point methods for convex quadratic programming. *Optimization Letters*, pages 1–19, 2023.

[34] S. Schmieta and F. Alizadeh. Extension of primal-dual interior point algorithms to symmetric cones. *Mathematical Programming*, 96(3):409–438, 2003.

[35] S. Sra and R. Hosseini. Conic geometric optimization on the manifold of positive definite matrices. *SIAM Journal on Optimization*, 25(1):713–739, 2015.

[36] J. F. Sturm. Implementation of interior point methods for mixed semidefinite and second order cone optimization problems. *Optimization methods and software*, 17(6):1105–1154, 2002.

[37] J. F. Sturm and S. Zhang. Symmetric primal-dual path-following algorithms for semidefinite programming. *Applied Numerical Mathematics*, 29(3):301–315, 1999.

[38] M. J. Todd. A study of search directions in primal-dual interior-point methods for semidefinite programming. *Optimization methods and software*, 11(1-4):1–46, 1999.

[39] M. J. Todd, K.-C. Toh, and R. H. Tütüncü. On the Nesterov–Todd direction in semidefinite programming. *SIAM Journal on Optimization*, 8(3):769–796, 1998.

[40] K.-C. Toh, M. J. Todd, and R. Tütüncü. SDPT3 version 4.0: a MATLAB software for semidefinite-quadratic-linear programming. *http://www.math.nus.edu.sg/ mattohkc/sdpt3.html*, 2009.

[41] L. Tunçel. Primal-dual symmetry and scale invariance of interior-point algorithms for convex optimization. *Mathematics of Operations Research*, 23(3):708–718, 1998.

[42] A. Wiesel. Geodesic convexity and covariance estimation. *IEEE transactions on signal processing*, 60 (12):6182–6189, 2012.

[43] S. J. Wright. *Primal-dual interior-point methods*. SIAM, 1997.

A Appendix

This section contains background results about the Euclidean Jordan algebra $\mathcal{J}$ and cone-of-squares $\mathcal{K}$ that we referenced without proof. The first establishes properties of the quadratic representation $Q(u)v := 2u \circ (u \circ v) - (u \circ u) \circ v$.

Lemma A.1 ([8]). *The following statements hold.*

1. $Q(u)^{-1} = Q(u^{-1})$ *for all invertible* $u \in \mathcal{J}$.
2. $(Q(u)v)^{-1} = Q(u^{-1})v^{-1}$ *for all invertible* $u, v \in \mathcal{J}$.
3. $Q(Tu) = TQ(u)T^*$ *for any* $u \in \mathcal{J}$ *and automorphism* $T : \mathcal{J} \to \mathcal{J}$ *of* $\mathcal{K}$, *where* $T^* : \mathcal{J} \to \mathcal{J}$ *denotes the adjoint of* T.

4. *$Q(u)^2 = Q(u^2)$ for all $u \in \mathcal{J}$.*

5. *$Q(u)e = u^2$ for all $u \in \mathcal{J}$.*

6. *$Q(u)$ is self-adjoint, i.e., $\langle Q(u)v, w\rangle = \langle v, Q(u)w\rangle$ for all $u, v, w \in \mathcal{J}$.*

Proof. The first properties are Propositions II.3.1., II.3.3, III.5.2, p. 55, and p. 48 of [8]. The last is evident from the definition of $Q(u)$ and the fact that Jordan multiplication is self-adjoint, i.e., $\langle u \circ v, w\rangle = \langle v, u \circ w\rangle$. □

The next establishes properties of orthogonal automorphisms of $\mathcal{K}$. They trivially follow from the fact that such automorphisms are precisely the Jordan-algebra automorphisms of $\mathcal{J}$ given our use of the trace inner-product [8, p. 56].

Lemma A.2. *Let $M : \mathcal{J} \to \mathcal{J}$ be an orthogonal automorphism of $\mathcal{K}$. Then, the following statements hold for all $u \in \mathcal{J}$.*

1. *If u is an idempotent, i.e., $u \circ u = u$, then Mu is an idempotent.*

2. *If u has spectral decomposition $\sum_{i=1}^n \lambda_i e_i$, then Mu has spectral decomposition $\sum_{i=1}^n \lambda_i M e_i$.*

3. $\exp(Mu) = M \exp(u)$.

Further, $Me = e$.

Proof. By use of the trace inner-product, M is also an automorphism of $\mathcal{J}$ [8, p. 56] and hence satisfies $(Mx) \circ (My) = M(x \circ y)$. Hence, $(Mu) \circ (Mu) = M(u \circ u) = Mu$, showing the first statement. The second statement is immediate from the first: if u has spectral decomposition $\sum_{i=1}^n \lambda_i e_i$, then Mu has decomposition $\sum_{i=1}^n \lambda_i M e_i$, since the Me_i are idempotent and pairwise orthogonal, i.e., $\langle Me_i, Me_j\rangle = \langle e_i, M^* M e_j\rangle = \langle e_i, e_j\rangle = 0$. The third is immediate from the second:

$$\exp(Mu) = \sum_{i=1}^n \exp(\lambda_i) M e_i = \sum_{i=1}^n M \exp(\lambda_i) e_i = M \exp(u).$$

Finally, $Me = e$ given that $e = \exp(M0) = M \exp(0) = Me$. □