Interior-point methods on manifolds: theory and applications

Hiroshi Hirai*   Harold Nieuwboer†   Michael Walter‡

Abstract

Interior-point methods offer a highly versatile framework for convex optimization that is effective in theory and practice. A key notion in their theory is that of a self-concordant barrier. We give a suitable generalization of self-concordance to Riemannian manifolds and show that it gives the same structural results and guarantees as in the Euclidean setting, in particular local quadratic convergence of Newton's method. We analyze a path-following method for optimizing compatible objectives over a convex domain for which one has a self-concordant barrier, and obtain the standard complexity guarantees as in the Euclidean setting. We provide general constructions of barriers, and show that on the space of positive-definite matrices and other symmetric spaces, the squared distance to a point is self-concordant. To demonstrate the versatility of our framework, we give algorithms with state-of-the-art complexity guarantees for the general class of scaling and non-commutative optimization problems, which have been of much recent interest, and we provide the first algorithms for efficiently finding high-precision solutions for computing minimal enclosing balls and geometric medians in nonpositive curvature.

Contents

*Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, 113-8656, Japan, hirai@mist.i.u-tokyo.ac.jp

†Korteweg-de Vries Institute for Mathematics and QuSoft, University of Amsterdam, The Netherlands and Faculty of Computer Science, Ruhr University Bochum, Germany, h.a.nieuwboer@uva.nl

‡Faculty of Computer Science, Ruhr University Bochum, Germany, michael.walter@rub.de

1   Introduction and summary of results

The development of interior-point methods is one of the greatest successes in convex optimization, and by now has a long history dating back to the works of Frisch [Fri55], Karmarkar [Kar84a, Kar84b], Gill et al. [GMS$^+$86] and many others. It led to one of the first polynomial-time algorithms for linear programming (in contrast with the simplex algorithm due to Dantzig [Dan63]), the other being the ellipsoid method due to Khachiyan [Kha80]. In the seminal work of Nesterov and Nemirovskii [NN94], it was shown that the key property to the analysis of interior-point methods is the notion of *self-concordance*. Essentially every convex programming problem is in principle amenable to interior-point methods, which follows from constructions of self-concordant barriers for arbitrary (bounded) convex domains, cf. [NN94, Hil14, Fox15, BE19, Che21]. Furthermore, interior-point methods are eminently practical, and currently give the best algorithms for linear programming [LS20, vdB19].

So far, these successes have been restricted to convex optimization on Euclidean space. While there is a strong connection between self-concordance-based interior-point methods and Riemannian geometry [Dui99, NT02, NN08], the framework of interior-point methods has not yet been generalized to objectives which are *geodesically convex*, i.e., convex on Riemannian manifolds. Indeed, while there have been previous attempts at extending interior-point methods to this setting [Udr97, Ji07, JMJ07], a satisfactory generalization of the Euclidean theory had still been elusive – in particular, the natural quadratic convergence analysis of Newton's method for self-concordant functions, which in turn enables efficient path-following methods with global guarantees.

Instead, research on Riemannian optimization has so far largely focused on different approaches. There is extensive literature on first- and second-order methods for convex and non-convex optimization, see e.g. [Udr94, AMS09, Sat21, Bou23] for comprehensive overviews and [FS02, DPM03,

ABM08, SH15, ZS16, AS20, WS22, SW22]. Recently, [LY22] gave a path-following method for non-convex constrained manifold optimization which does not use self-concordance. In another direction, geodesic updates can also be useful for Euclidean convex optimization problems [Per20, Per22].

> **Structural results:** In this work, we extend the interior-point method framework to Riemannian manifolds. We generalize the key notion of self-concordance, and show that (unlike prior definitions) it gives the same structural results and guarantees as in the Euclidean setting, in particular local quadratic convergence of Newton's method. This allows us to give a path-following method for optimizing suitable objective functions over domains for which a self-concordant barrier is available, and we give complexity guarantees that match the Euclidean ones.

These results are already very interesting from a theoretical perspective. To put the framework to use, however, one still has to find explicit self-concordant barriers. To this end, we give general constructions of barriers, as well as several explicit examples, and we show that our framework is indeed applicable to a wide variety of problems, including (but not limited to):

(A) *Geometry:* Given points $p_1, \ldots, p_m$ on a Riemannian manifold, what is the minimum radius ball that contains all these points? What is their geometric median, i.e., the point that minimizes the sum of distances to each p_i?

(B) *Quantum marginals:* Given density matrices $\rho_1, \ldots, \rho_k$, each describing the quantum state of one party, does there exist a k-party pure quantum state with marginals equal to the ρ_k?

(C) *Tensor networks:* Given a $(2d + 1)$-leg tensor, does it ever define a non-zero tensor network state of PEPS type? And how can one efficiently compute a canonical form?

(D) *Brascamp–Lieb inequalities:* Given linear maps $L_k \colon \mathbb{R}^m \to \mathbb{R}^{m_k}$ and numbers $q_k > 0$ for $k \in [n]$, what is the optimal constant $C > 0$ such that $\int_{\mathbb{R}^m} \prod_{k=1}^n f_k(L_k x)\, dx \leq C \prod_{k=1}^n \|f_k\|_{1/q_k}$ for all non-negative functions f_k on $\mathbb{R}^{m_k}$? Many classical integral inequalities fall into this setting, such as the Hölder, Young, Loomis–Whitney, and certain hyper-contractivity inequalities.

The first question in Problem (A) on finding a minimum enclosing ball has been studied before in the Riemannian setting [AN13, NH15], and [NH15] gave an algorithm for the specific case of hyperbolic space, yielding a ball with radius at most a factor $1 + \delta$ larger than the optimal radius in $\mathcal{O}(1/\delta^2)$ iterations. The geometric median problem has been studied in [FVJ09, Yan10], and [Yan10] gave an explicit subgradient algorithm on general manifolds, finding a point whose squared distance to the point achieving minimal sum of distances to the p_i is at most ε in $\mathcal{O}(1/\varepsilon)$ iterations. Interestingly, the other problems are not even obviously related to geodesically convex optimization in the first place. Problem (B) is not currently known to be solvable in polynomial time in all parameters, although partial results are known [BFG$^+$18], and there is complexity-theoretic evidence that polynomial-time algorithms might exist, as it is in NP $\cap$ coNP [BCMW17]. Problem (C) arose very recently in quantum information [AMN$^+$22], and again no algorithms are known that run in polynomial time in all parameters, except for $d = 1$. Problem (D) was studied in [GGOW17], but current methods have an exponential dependence on the bit complexity of the coefficients q_j.

What connects problems (B)–(D) to each other, and to geodesically convex optimization, is that they all belong to the broad class of *scaling problems.* In particular, problem (B) can be reduced to *tensor scaling* [BFG$^+$18], and problems (C) and (D) generalize respectively reduce to *operator scaling* (but not efficiently so) [AMN$^+$22, GGOW17]. There has been much recent progress on this class of problems [LSW00, Gur04, WDGC13, Wal14, GGOW20, GGOW16, GGOW17, CMTV17,

ALOW17, AGL$^+$18, BFG$^+$18, BGO$^+$18, BFG$^+$19], and the quest of finding better algorithms is one key motivation for our work. Scaling problems have strong connections to many different areas in mathematics and theoretical computer science beyond those mentioned above: they are related to approximating permanents [LSW00], non-commutative rational identity testing [GGOW16], Horn's problem on spectra of sums of Hermitian matrices [Fra18], the Paulsen problem [KLLR18, HM21b], canonical forms and zero-testing of tensor networks [AMN$^+$22], strengthening the Sylvester–Gallai theorem [BDWY12, DSW14, DGOS18], approximating optimal transport plans in machine learning [Cut13], maximum-likelihood estimation in statistics [AKRS21b, AKRS21a, FORW21], the asymptotic non-vanishing of Kronecker coefficients in representation theory [IMW17, BFG$^+$18], and geometric invariant theory [KN79, NM84, MFK94]. As elucidated in a long sequence of works, see [BFG$^+$19], these are all related to a *norm minimization* problem: given a linear action of a nice (complex reductive) Lie group G on a vector space V, and a vector $v \in V$, the goal is to minimize the norm over the orbit $G \cdot v$ (see Section 1.4). When G is commutative, such as in the case of matrix scaling, these problems reduce to geometric programming (a well-known generalization of linear programming) and hence they can be solved efficiently [CMTV17, ALOW17, SV14, BLNW20]. In the most difficult situations, however, including in most of the mentioned applications, the group G is non-commutative; hence this class of problems has also been called *non-commutative (group) optimization* problems. In this case, efficient algorithms are known only in special cases, which have recently been understood to all satisfy a certain total unimodularity [GGOW20, AGL$^+$18, BFG$^+$19]. For general non-commutative optimization, and in particular for problems (B)–(D), there are currently no algorithms that run in time polynomial in all parameters.

Algorithmic applications: For problem (A), our framework gives (to the best of our knowledge) the first algorithms for efficiently finding high-precision solutions in nonpositive curvature. For the entire class of *scaling* or *non-commutative optimization problems*, and in particular for problems (B)–(D), our framework yields new algorithms that match the complexity guarantees of the state-of-the-art algorithms [BFG$^+$19], while not obviously suffering from the same obstructions as those methods, opening up a new avenue for future research.

Indeed, the current state-of-the-art methods are fundamentally incapable of providing algorithms that run in polynomial time in all parameters for the general scaling problem, and in particular for problems (B)–(D). The main reason that we lack the kind of sophisticated optimization methods that are known in the Euclidean setting, as reviewed earlier, is due to the geometry of the spaces that one has to optimize over, which poses fundamental new challenges and obstructions.

To make this more concrete, we consider the quantum marginal problem (B). For simplicity, we take $k = 3$ parties of the same dimension $n \geq 2$. Let $G = \mathrm{SL}(n, \mathbb{C}) \times \mathrm{SL}(n, \mathbb{C}) \times \mathrm{SL}(n, \mathbb{C})$ act on 3-tensors in $V = \mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n$ by simultaneous base change or *tensor scaling*, i.e., $(g_1, g_2, g_3) \cdot v = (g_1 \otimes g_2 \otimes g_3)v$ for $g_j \in \mathrm{SL}(n, \mathbb{C})$ and $v \in V$. Then the relevant optimization problem amounts to minimizing the (for convenience squared) ℓ^2-norm over all such scalings:[1]

$$\inf_{g_1, g_2, g_3 \in \mathrm{SL}(n)} \|(g_1 \otimes g_2 \otimes g_3)v\|_2^2. \tag{1.1}$$

---

[1]To see that this is related to quantum marginals, consider the pure quantum state $\rho = ww^*$ with $w := g \cdot v / \|g \cdot v\|$. Then the (logarithmic) gradient of the objective is given by $(\rho_1 - I/n, \rho_2 - I/n, \rho_3 - I/n)$, where ρ_1, ρ_2, ρ_3 are the one-body reduced density matrices or marginals of ρ. Therefore, minimizers of Eq. (1.1) correspond to quantum state with maximally mixed marginals. The general quantum marginal problem amounts to characterizing the set of possible *gradients* for generic v. See [BFG$^+$18] for more detail.

We can reduce the optimization to $M = \mathrm{SPD}(n) \times \mathrm{SPD}(n) \times \mathrm{SPD}(n)$, where $\mathrm{SPD}(n)$ denotes the complex positive-definite matrices of unit determinant. Indeed, since $P_j := g_j^* g_j$ is an arbitrary matrix in $\mathrm{SPD}(n)$, we see that Eq. (1.1) is equivalent to:

$$\inf_{P_1, P_2, P_3 \in \mathrm{SPD}(n)} \langle v | P_1 \otimes P_2 \otimes P_3 | v \rangle \tag{1.2}$$

Unfortunately, the domain is non-convex as a subset of the Euclidean space of triples of Hermitian matrices, and in any case the objective is not a convex function of the variables.

However, a key observation is that the objective becomes convex when $\mathrm{SPD}(n)$ and hence M is given a natural non-Euclidean geometry, namely the so-called *affine-invariant* metric, which also appears as the *Fisher-Rao metric* for Gaussian covariance matrices in statistics (see Section 1.3 for a precise definition). Then the straight lines of Euclidean space get replaced by the geodesics of the new metric, which take the form $P_j(t) = \sqrt{P_j} e^{H_j t} \sqrt{P_j}$ for traceless Hermitian matrices H_j and clearly remain in $\mathrm{SPD}(n)$. It is easy to verify that the objective in Eq. (1.2) is *convex along such geodesics* (in fact, log-convex). The same phenomenon occurs for any scaling or non-commutative optimization problem; while non-convex in the Euclidean sense, these problems become convex when formulated appropriately [BFG$^+$19]. In most applications, the domain is given by the positive-definite matrices $\mathrm{PD}(n)$, by $\mathrm{SPD}(n)$, or by products of these spaces.

A key property of these domains is that they have *non-positive curvature*, in contrast with Euclidean space, which has zero curvature. This gives rise to significant geometric challenges for optimization algorithms. For example, Rusciano [Rus19] gave a (non-constructive) cutting-plane method in non-positive curvature, with a logarithmic dependence on the volume of the domain. Unfortunately, the volume of balls in manifolds of non-positive curvature grows *exponentially* with the radius (in constant dimension). In a black-box setting, where one can make queries to a function- and gradient oracle, the same geometric fact implies that any algorithm that wants to find an approximate minimizer must make a number of queries that is *linear* in the distance to the approximate minimizer [HM21a, CB22]. This again suggests that efficient algorithms for geodesic convex optimization in non-positive curvature in general, and for non-commutative optimization problems in particular, must make use of additional structure beyond diameter bounds, as the distance to an approximate minimizer is in general exponential in the input size [FR21]. The best current algorithms for non-commutative optimization [BFG$^+$19] also only have a linear dependence on these diameter bounds. The reason is that they are *box-constrained Newton methods*, i.e., a Newton-type method where the steps are constrained to a subdomain of essentially fixed size. To traverse the in general exponentially large distance to the approximate minimizer, such algorithms must perform exponentially many iterations.

> To overcome these challenges and obstructions, it is natural to resort to methods which are capable of better exploiting the structure of the optimization problem at hand. Interior-point methods offer a powerful such framework in the Euclidean case, and they have already proved successful for commutative scaling problems [CMTV17, BLNW20]. With this work, we hope to contribute a first clear step towards generalizing this powerful framework to the manifold setting.

Indeed, we believe that our results suggest several interesting directions for follow-up research. For instance, does every convex domain admit a self-concordant barrier, as is the case in the Euclidean setting? Do there exist self-concordant barriers with better barrier parameters which can be used for these applications, leading to better algorithms? Alternatively, can it be shown that our constructions

are essentially optimal? Can interior-point methods on manifold always be initialized efficiently, and is there a suitable notion of duality?[2] We discuss these questions in more detail in Section 7.

1.1 Self-concordance and Newton's method on manifolds

In the remainder of this introduction, we give a more detailed overview of our results, starting with our proposed notion of self-concordance. Throughout, $f : D \to \mathbb{R}$ is a smooth function defined on a convex subset $D \subseteq M$ of a connected, geodesically complete Riemannian manifold M. Then f is called *convex* if it is convex along geodesics. Let ∇ denote the *covariant derivative* (or Levi–Civita connection), which allows taking derivatives of vector and tensor fields, and in particular to define Hessians $\nabla^2 f$ and higher derivatives (we review the required Riemannian geometry in Section 2). Then our proposed generalization of self-concordance to possibly curved manifolds is as follows.

Definition 1.1 (Self-concordance). *For $\alpha > 0$, a convex function f is called α-self-concordant if, for all $p \in D$ and for all tangent vectors $u, v, w \in T_p M$, we have*

$$|(\nabla^3 f)_p(u, v, w)| \leq \frac{2}{\sqrt{\alpha}} \sqrt{(\nabla^2 f)_p(u, u)} \sqrt{(\nabla^2 f)_p(v, v)} \sqrt{(\nabla^2 f)_p(w, w)}. \tag{1.3}$$

If f is closed convex, meaning its epigraph is closed, then f is called strongly α-self-concordant.

Self-concordance can be interpreted as giving a bound on the norm of the third derivative $(\nabla^3 f)_p$, that is, on the change of the Hessian $(\nabla^2 f)_p$, with respect to the (possibly degenerate) inner product defined by the Hessian itself. We say that f is *α-self-concordant along geodesics* if one requires the above bound only for $u = v = w$, that is, if for all $p \in D$ and for all $u \in T_p M$, we have

$$|(\nabla^3 f)_p(u, u, u)| \leq \frac{2}{\sqrt{\alpha}} ((\nabla^2 f)_p(u, u))^{3/2}. \tag{1.4}$$

When $M = \mathbb{R}^n$, the third derivative is a symmetric tensor and hence the two notions coincide. However, in general, the third derivative is *not* symmetric in all its arguments, and indeed its asymmetry is precisely related to the manifold's *curvature* via the Ricci identity [Lee18, Thm. 7.14], as we discuss in Section 3. Prior work only considered self-concordance along geodesics [Ji07] (which suffices for a damped Newton method) and did not take the asymmetry into account [Udr97, JMJ07].

Here we show explicitly that self-concordance is in general strictly stronger than self-concordance along geodesics (cf. Section 1.3), and it is the stronger notion that allows for the desired quadratic convergence of Newton's method – a cornerstone of the interior point theory. Assume for simplicity that the Hessian $(\nabla^2 f)_p$ is positive definite for all $p \in D$. Then the *Newton iterate* of f at $p \in D$ is defined by minimizing the local quadratic approximation:

$$p_{f,+} := \mathrm{Exp}_p(u^*), \quad u^* = \underset{u \in T_p M}{\mathrm{argmin}} \left(f(p) + df_p(u) + \frac{1}{2}(\nabla^2 f)_p(u, u) \right).$$

The progress is quantified in terms of the *Newton decrement*, which is directly related to the gap between the original function value and the minimum of the local quadratic approximation. It is defined for any $\alpha > 0$ and $p \in D$ as

$$\lambda_{f,\alpha}(p) = \sup_{0 \neq u \in T_p M} \frac{|df_p(u)|}{\sqrt{\alpha(\nabla^2 f)_p(u, u)}}. \tag{1.5}$$

Then we prove following result on general Riemannian manifolds in Theorem 3.16:

---

[2]The lack of nontrivial linear functions in the presence of curvature poses significant challenges.

Theorem 1.2 (Quadratic convergence). *Let $f\colon D \to \mathbb{R}$ be a strongly α-self-concordant function defined on an open convex set $D \subseteq M$, with positive definite Hessian. Let $p \in D$ be a point such that $\lambda_{f,\alpha}(p) < 1$. Then the Newton iterate remains in the domain, i.e., $p_{f,+} \in D$, and moreover*

$$\lambda_{f,\alpha}(p_{f,+}) \leq \left(\frac{\lambda_{f,\alpha}(p)}{1 - \lambda_{f,\alpha}(p)} \right)^2.$$

To relate the Newton decrements at p and $p_{f,+}$, we control the change in the Hessian of f along the geodesic from p to $p_{f,+}$. This crucially uses the notion of self-concordance of Eq. (1.3), rather than the weaker definition along geodesics as in Eq. (1.4). This is because there are two directions involved: the one of the geodesic, and the one corresponding to the subsequent Newton decrement.

1.2 Barriers and a path-following method on manifolds

Interior-point methods provide a natural and modular approach for minimizing an objective f constrained to a bounded convex domain $D \subseteq M$. The key idea is to rather minimize, for $t > 0$,

$$F_t \colon D \to \mathbb{R}, \quad F_t := tf + F,$$

where F is a self-concordant "barrier" that is finite on D and diverges to ∞ on its boundary.[3] This automatically ensures the constraint, as F_t is finite only on D, and for large t the objective dominates. One then starts with an approximate minimizer of F and $t \approx 0$, and follows the *central path* $z(t) := \operatorname{argmin}_{p \in D} F_t(p)$ by iteratively performing two steps: increase t to some t' such that the current point is still not too far from $z(t')$, and then take a Newton step for $F_{t'}$ to move closer to it. For large enough $t > 0$, we arrive at an approximate minimizer of f on $D \subseteq M$.

More precisely, the function $F \colon D \to \mathbb{R}$ is required to be a *(non-degenerate strongly self-concordant) barrier* for D, with barrier parameter $\theta \geq 0$, which means that F is strongly 1-self-concordant, has positive definite Hessian, and $\lambda_F(p)^2 \leq \theta$ for all $p \in D$. The barrier parameter θ controls how rapidly t can be increased in every iteration.

In order to guarantee that Newton's method indeed moves closer to the central path, we are interested in conditions on f that ensure that the functions F_t are self-concordant for every $t > 0$, with a constant independent of t. One way to guarantee this is to assume that the objective $f \colon D \to \mathbb{R}$ is *compatible* with the barrier F in the following sense: there are constants $\beta_1, \beta_2 \geq 0$ such that, for all $p \in D$ and $u, v \in T_p M$,

$$\begin{aligned}
|(\nabla^3 f)_p(u, v, v)| \leq\ & 2\beta_1 \sqrt{(\nabla^2 F)_p(u, u)}(\nabla^2 f)_p(v, v) \\
& + 2\beta_2 \sqrt{(\nabla^2 F)_p(v, v)} \sqrt{(\nabla^2 f)_p(u, u)} \sqrt{(\nabla^2 f)_p(v, v)}.
\end{aligned}$$

In particular, linear and quadratic functions are compatible with arbitrary self-concordant barriers, but these are not the only examples, and we crucially use this level of generality to give algorithms for the general scaling or non-commutative optimization problem. We expand on compatibility in Section 4.2, and show that it is also useful for constructing new self-concordant barriers, for instance for the epigraph of a function compatible with a self-concordant barrier (Theorem 4.11).[4]

---

[3]In the Euclidean setting, the barrier $F(x) = -\log x$ models the constraint that $x > 0$, and $F(X) = -\log \det X$ defines the constraint that X is a positive-definite matrix [NN94, Ren01]. Constraints are combined simply by adding the respective barriers. In the manifold setting, barriers are much harder to come by, but we give general constructions and concrete examples in Sections 4 to 6.

[4]While optimizing a function f on a domain D can always be reduced to optimizing a linear function over its epigraph $\{(p, t) \in D \times \mathbb{R} : f(p) < t\}$, this requires a barrier for the epigraph. We construct such a barrier precisely when f is compatible with F. However, it may be more difficult to initialize the path-following method on the epigraph rather than directly on D, so it can be advantageous to optimize f directly. See Section 6.1.

Our notion of compatibility is inspired by a similar notion in the Euclidean setting, as is our analysis of the path-following method [NN94]. Its precise guarantees match those from the Euclidean setting, and are given in the following theorem, which we prove in Theorem 4.17:

Theorem 1.3 (Path-following method). *Let $D \subseteq M$ be an open, bounded, and convex domain, and let $f, F \colon D \to \mathbb{R}$ be smooth convex functions, such that F is a self-concordant barrier with barrier parameter $\theta \geq 0$ and f has a closed convex extension. Let $\alpha > 0$ be such that $F_t := tf + F$ is α-self-concordant for all $t \geq 0$. Let $p \in D$ be such that $\lambda_F(p) \leq \frac{\sqrt{\alpha}}{8}$, and let $\varepsilon > 0$. Then, using*

$$\mathcal{O}\left(\left(1 + \sqrt{\frac{\theta}{\alpha}}\right) \log\left(\frac{(\theta + \alpha)\|df_p\|^*_{F,p}}{\varepsilon\sqrt{\alpha}}\right)\right)$$

Newton iterations, one can find a point $p_\varepsilon \in D$ such that

$$f(p_\varepsilon) - \inf_{q \in D} f(q) \leq \varepsilon.$$

The quantity $\|df_p\|^*_{F,p}$ is a lower bound on the variation $\sup_{q \in D} f(q) - \inf_{q \in D} f(q)$ of f over D (Lemma 4.18), and hence imposes a natural notion of scale in the complexity bound.

1.3 Examples of self-concordance: Squared distance in non-positive curvature

Self-concordance on manifolds is much more difficult to verify than for Euclidean space, and this begs the question whether nontrivial examples even exist. A natural candidate is $f(p) = d(p, p_0)^2$, the *squared distance* function to some point $p_0 \in M$. On Euclidean space, f is trivially self-concordant, as its third derivative vanishes identically. In the presence of curvature the third derivative can be nonzero. Nevertheless, we prove that the squared distance is self-concordant on $\mathrm{PD}(n)$ and, as a corollary, also on a broad class of manifolds with non-positive curvature.

We now discuss this in more detail. As in the introduction, we denote by $\mathrm{PD}(n) = \mathrm{PD}(n, \mathbb{C})$ the complex positive-definite matrices, endowed with the well-known *affine-invariant* Riemannian metric, which is given as follows. Since $\mathrm{PD}(n)$ is an open subset of $\mathrm{Herm}(n)$, the Hermitian $n \times n$-matrices, we can identify the tangent space $T_P\mathrm{PD}(n)$ at every $P \in \mathrm{PD}(n)$ with $\mathrm{Herm}(n)$. Then the Riemannian metric is defined as follows: for any two tangent vectors $U, V \in T_P\mathrm{PD}(n)$, their inner product is

$$\langle U, V \rangle_P = \mathrm{Tr}\left[P^{-1}UP^{-1}V\right].$$

With this metric, $\mathrm{PD}(n)$ is a Hadamard manifold, i.e., a simply connected geodesically complete Riemannian manifold with non-positive curvature. Its geodesics, parallel transport, covariant derivatives, and so forth all have well-known closed-form expressions, which are amenable to tools from matrix analysis. For example, the geodesics through $P \in \mathrm{PD}(n)$ are of the form $t \mapsto \sqrt{P}e^{tH}\sqrt{P}$ for $H \in \mathrm{Herm}(n)$, and geodesic midpoints are the same as operator geometric means. The distance between two matrices $P, Q \in \mathrm{PD}(n)$, defined as the minimum length of any path connecting them, is

$$d(P, Q) = \|\log(P^{-1/2}QP^{-1/2})\|_{\mathrm{HS}},$$

where $\|\cdot\|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt (i.e., Frobenius) norm. In Theorem 5.15 we show:

Theorem 1.4 (Self-concordance of squared distance). *For any $P_0 \in \mathrm{PD}(n)$, the squared distance $f \colon \mathrm{PD}(n) \to \mathbb{R}$ to P_0, defined by $f(P) = d(P, P_0)^2$, is 2-self-concordant.*

We conjecture that the squared distance is actually 8-self-concordant, see Remark 5.14. Self-concordance on $\mathrm{PD}(n, \mathbb{C})$ implies the same result for the squared distance on any convex subset of it. Therefore, the self-concordance holds on any Hadamard manifold that is also a so-called symmetric space;[5] we will call this a *Hadamard symmetric space*. In particular, using [BH99, Prop. 10.58] we obtain the following result, which covers most non-positively curved spaces of import in applications, including the general scaling or non-commutative optimization problem (cf. Section 6.1):

Corollary 1.5. *Let $G \subseteq \mathrm{GL}(n, \mathbb{R})$ be an algebraic subgroup[6] such that $g^T \in G$ for every $g \in G$. Set $M := \{g^T g : g \in G\} \subseteq \mathrm{PD}(n, \mathbb{R})$. Then $M \subseteq \mathrm{PD}(n, \mathbb{R})$ is a convex subset, and for every $p_0 \in M$, the function $f \colon M \to \mathbb{R}, f(p) = d(p, p_0)^2$ is 2-self-concordant.*

Hyperbolic space $\mathbb{H}^n$ is a paradigmatic example of a manifold with non-positive curvature in this class. Corollary 1.5 implies that the squared distance function to a point in $\mathbb{H}^n$ is 1-self-concordant, as one has to rescale the curvature by a factor 2 to obtain an isometric embedding into $\mathrm{PD}(n, \mathbb{C})$. Similarly, the conjectured 8-self-concordance on $\mathrm{PD}(n, \mathbb{C})$ would imply 4-self-concordance on $\mathbb{H}^n$.

We are able to prove the stronger result that the squared distance on $\mathbb{H}^n$ is in fact 8-self-concordant, and that this is optimal, see Theorem 5.16. In contrast, the squared distance on hyperbolic space is $\frac{27}{2}$-self-concordant along geodesics, as was shown previously in [Ji07, Lem. 11].[7] It is an interesting open question whether there exists a universal constant $C > 0$ such that if M is a Hadamard manifold with all sectional curvatures in $[-\kappa, 0]$, then for every $p_0 \in M$, $f(p) = d(p, p_0)^2$ is C/κ-self-concordant.

Using the self-concordance of the squared distance, it is easy to construct a self-concordant barrier for its epigraph (cf. Theorem 4.11). To this end we provide the following result, which applies in particular to $\mathrm{PD}(n)$, hyperbolic space, and all other Hadamard symmetric spaces.

Theorem 1.6 (Epigraph barrier)**.** *Let M be a Hadamard manifold, and let $p_0 \in M$. Assume that the function $f \colon M \to \mathbb{R}$, $f(p) = d(p, p_0)^2$ is α-self-concordant. Let $D = \{(p, S) \in M \times \mathbb{R} : f(p) < S\}$. Then, the function $F \colon D \to \mathbb{R}$ defined by*

$$F(p, S) = -\log\big(S - d(p, p_0)^2\big) + \frac{1}{\alpha} d(p, p_0)^2 \tag{1.6}$$

is strongly 1-self-concordant, and $\lambda_F(p, S)^2 \le 1 + \frac{2}{\alpha} d(p, p_0)^2$.

The reason that the proposition does *not* state that F is a barrier is that the Newton decrement $\lambda_F(p, S)$ is not bounded by a constant, but rather depends on the distance to the point p_0. To obtain a barrier, on needs to impose an additional constraint on the domain to force it to be bounded, for instance by requiring that $S < S_0$, which can be implemented by adding a logarithmic barrier term $-\log(S_0 - S)$ to F. The dependence of the Newton decrement on the distance to p_0 is caused by the term $\frac{1}{\alpha} d(p, p_0)^2$ in Eq. (1.6), but without this term the function would not be self-concordant. See also Theorem 4.14, where we construct a barrier for the sublevel set of a self-concordant function, with barrier parameter depending on the gap in function value.

We also provide a strengthening of the above theorem for hyperbolic space (see Theorem 5.22):

---

[5] Any such space is the product of a symmetric space of non-compact type and a Euclidean space [Hel79, Prop. V.4.2], and embeds, possibly after rescaling the metric on each of its de Rham factors, as a complete convex submanifold of $\mathrm{PD}(n, \mathbb{R})$ for some $n \ge 1$, and hence also in $\mathrm{PD}(n, \mathbb{C})$ [Ebe97, Thm. 2.6.5]. See [Hel79] for more background.

[6] This means that G is a subset of $\mathrm{GL}(n, \mathbb{R})$ determined by polynomial equations in the matrix entries.

[7] They prove that $M_f = \sqrt{16/27}$, where the constant M_f is related to the constant α in our definition of self-concordance along geodesics by $M_f = 2/\sqrt{\alpha}$.

Theorem 1.7. *Let* $M = \mathbb{H}^n$, $p_0 \in M$, *and define* $f \colon M \to \mathbb{R}$ *by* $f(p) = d(p, p_0)^2$. *Let* $D = \{(p, R, S) \in M \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \; : \; RS - f(p) > 0\}$. *Then the function* $F \colon D \to \mathbb{R}$ *by*

$$F(p, R, S) = -\log(RS - f(p)) + f(p)$$

is strongly $\frac{1}{2}$-*self-concordant. Furthermore,* $\lambda_{F, \frac{1}{2}}(p, R, S)^2 \leq 4 + 4f(p)$.

The significance of this result is that it can be used to construct a barrier for the epigraph of the *distance* to a point, rather than the squared distance, by restricting to the subspace defined by the equation $S = R$. This is essential for applying the framework to the geometric median problem, see Section 1.6. In the Euclidean setting, the additional f-term is unnecessary; see for instance the proof of [NN94, Prop. 5.4.3]. In our setting the proof is more complicated, as it involves a strengthening of the self-concordance estimate on the third derivative of the squared distance. The key estimates which enable our proof of the above theorem are given in Theorem 5.16.

1.4 Application I: Non-commutative optimization and scaling problems

Our first application is the one which motivated us to extend the framework in the first place. To state our result in full generality requires a bit of setup [BFG+19]. Let $G \subseteq \mathrm{GL}(n, \mathbb{C})$ be a connected algebraic subgroup such that $g^* \in G$ for all $g \in G$. Let $\pi \colon G \to \mathrm{GL}(V)$ be a rational representation on a finite-dimensional complex vector space V. Assume V is endowed with an inner product such that the unitary matrices in G act unitarily. The general *norm minimization* problem asks to minimize the norm over the orbit of a given vector $v \in V$. That is, we wish to minimize $\|\pi(g)v\|$ over $g \in G$. As described earlier, this problem subsumes the class of *non-commutative optimization* or *scaling problems* that have been of much interest in the recent literature. Note that $\|\pi(g)v\|^2 = \langle v | \pi(g^*g) | v \rangle$. Accordingly, it suffices to minimize the so-called *log-norm* or *Kempf–Ness function* defined by

$$\phi_v \colon M \to \mathbb{R}, \quad \phi_v(p) = \log \langle v | \pi(p) | v \rangle$$

over $M = \{g^*g : g \in G\} = G \cap \mathrm{PD}(n)$. This function is convex along the geodesics of M. It is also $N(\pi)^2$-smooth in the convexity sense, where $N(\pi)$ is the so-called *weight norm* of the action, which is easy to compute and at most polynomially large (see Section 6.1 and [BFG+19] for details). Therefore, if ϕ_v is bounded from below, a simple gradient descent algorithm can be used to find a point $p \in M$ such that $\|\mathrm{grad}(\phi_v)_p\| \leq \delta$ within $\mathcal{O}(N(\pi)^2 [\phi_v(I) - \inf_{q \in M} \phi_v(q)]/\delta^2)$ iterations [BFG+19, Thm. 4.2]. A more sophisticated box-constrained Newton method is able to find an ε-approximate minimizer p_ε of ϕ_v within $\mathcal{O}((1 + R_0) N(\pi) \log[(\phi_v(I) - \inf_{q \in M} \phi_v(q))/\varepsilon])$ iterations, where $R_0 > 0$ is an upper bound on the distance to such a minimizer [BFG+19, Thms. 5.1 & 5.7]. Using our interior-point path-following method we prove the following result in Theorem 6.9:

Theorem 1.8 (Non-commutative optimization). *Let* $0 \neq v \in V$ *and* $R_0, \varepsilon > 0$. *Let* $M = \{g^*g : g \in G\} \subseteq \mathrm{PD}(n)$ *and* $D = \{p \in M : d(p, p_0) \leq R_0\}$, *and define* $\phi_v \colon M \to \mathbb{R}$ *by* $\phi_v(p) = \log \langle v | \pi(p) | v \rangle$. *Then there is an algorithm that within* $\mathcal{O}((1 + R_0) N(\pi) \log(N(\pi) R_0 / \varepsilon))$ *iterations of the path-following method finds* $p_\varepsilon \in D$ *such that*

$$\phi_v(p_\varepsilon) - \inf_{p \in D} \phi_v(p) \leq \varepsilon.$$

This essentially matches the complexity of the box-constrained Newton method mentioned above, which is currently the state-of-the-art. There is a small difference, in that our complexity has $N(\pi) R_0$ in the logarithm, rather than the potential gap $\phi_v(I) - \inf_{q \in M} \phi_v(q)$; these are related since ϕ_v is $N(\pi)$-Lipschitz. The approach we take to obtain this result is to use the barrier on M which

10

arises from Corollary 1.5 and Theorem 1.6, and to show that the Kempf–Ness function is compatible with the squared distance function, which is enough to implement the path-following method, as explained earlier. It would be very interesting to find a suitable barrier for this problem with a smaller barrier parameter (or prove that no such barrier exists).

1.5 Application II: Minimum-enclosing ball problem on $\mathrm{PD}(n)$

Next we consider the *minimum enclosing ball (MEB)* problem: given distinct points $p_1, \ldots, p_m \in M$, find p such that $R(p) := \max_i d(p, p_i)$ is minimal. When $M = \mathbb{R}^n$ is Euclidean space, this is a well-studied problem in computational geometry. There, it can be formulated as a second-order cone problem, to which interior-point methods are applicable (see, e.g., [KMY04]).

When M is a Hadamard manifold, the distance to a point is convex, and hence the MEB problem is a convex optimization problem. In particular, for hyperbolic space $M = \mathbb{H}^n$, there has been previous work on the MEB problem [AN13, NH15]. The only algorithm with explicit complexity bounds that we are aware of is due to Nielsen and Hadjeres [NH15]. If R_* is the minimal radius of an MEB and $\delta > 0$, then they can find a point $p \in \mathbb{H}^n$ such that $\max_i d(p, p_i) \leq (1 + \delta)R_*$ within $\mathcal{O}(1/\delta^2)$ iterations of an algorithm, each of which is simple to implement.

To find MEBs using interior-point methods, it is sufficient to have a barrier for the epigraph of the squared distance. In particular, the barrier constructed using Theorems 1.4 and 1.6 can be used to solve this problem on $\mathrm{PD}(n)$, and we prove the following result in Theorem 6.15

Theorem 1.9 (Minimum enclosing ball). *Let* $p_1, \ldots, p_m \in \mathrm{PD}(n)$ *be* $m \geq 3$ *points, and set* $R_0 = \max_{i \neq j} d(p_i, p_j)$. *Let* $R(p) = \max_i d(p, p_i)$, *set* $R_* = \inf_{p \in M} R(p)$, *and let* $\varepsilon > 0$. *Then with* $\mathcal{O}((m+1)R_0^2)$ *iterations of a damped Newton method and*

$$\mathcal{O}\left(\sqrt{1 + m(R_0^2 + 1)} \log \left(\frac{m(R_0^2 + 1)}{\varepsilon} \right) \right)$$

iterations of the path following method, one can find $p_\varepsilon \in \mathrm{PD}(n)$ *such that*

$$R(p_\varepsilon) - R_* \leq \varepsilon.$$

A similar result can be obtained on arbitrary Hadamard symmetric spaces. We also note that the optimal radius R_* satisfies $R_0 \leq 2R_*$ (Lemma 6.12), so that the above also yields a multiplicative error guarantee. Compared to the results of [NH15], we have a logarithmic dependence on the precision ε, but a linear dependence on R_0 (as opposed to no dependence).

1.6 Application III: Geometric median on hyperbolic space

Our last application is the *geometric median problems*. In the Euclidean setting this is also known as the *Fermat–Weber problem* [CLM+16]. It is formally defined as follows: given points $p_1, \ldots, p_m \in M$, not all contained in a single geodesic, find $p_0 \in M$ such that

$$p_0 \in \operatorname*{argmin}_{p \in \mathbb{H}^n} s(p) := \sum_{j=1}^m d(p, p_j).$$

The objective function s is convex on Hadamard manifolds M. In contrast with the *geometric mean (or barycenter) problem*, which is to find the minimizer of $\sum_{j=1}^m d(p, p_j)^2$, finding the geometric median is non-trivial even on $M = \mathbb{R}^n$. The first and one of the best-known algorithms for this problem on Euclidean space is Weiszfeld's algorithm [Wei37], which is a simple iterative procedure

based on solving the first-order optimality condition $\text{grad}(s)_p = \sum_{j=1}^{m}(p-p_j)/d(p,p_j) = 0$ for p, while treating the $d(p,p_j)$ as constants. Unfortunately, the update rule is not well-defined when p is one of the p_j's (which can be fixed, see e.g. [Ost78]), and it may converge very slowly in general. In [XY97] it was observed that one can also apply interior-point methods, by viewing the geometric median problem as a second-order cone program. More recent work [CLM$^+$16] has shown that a specialized long-step interior-point method is capable of solving the geometric median problem on $\mathbb{R}^n$ in nearly-linear time, and we refer the reader to their paper for a broader literature review. Weiszfeld's approach has been generalized to the Riemannian setting [FVJ09]. A sub-gradient approach [Yan10] can find a point with squared distance to the minimizer of s at most ε in $\mathcal{O}(1/\varepsilon)$ iterations; however, in the negatively curved setting, it suffers from an exponential dependence on the quantity $R_0 = \max_{i \neq j} d(p_i, p_j)$.

We can solve the geometric median problem on hyperbolic space $\mathbb{H}^n$ by using our interior-point framework and our barrier for the epigraph of the *distance* constructed using Theorem 1.7, which serve as analogs of the second-order cone and the associated barrier. In Theorem 6.20 we prove:

Theorem 1.10 (Geometric median). *Let $p_1, \ldots, p_m \in \mathbb{H}^n$ be $m \geq 3$ points, not all on one geodesic, and set $R_0 = \max_{i \neq j} d(p_i, p_j)$. Define $s \colon \mathbb{H}^n \to \mathbb{R}$ by $s(p) = \sum_{j=1}^{m} d(p, p_j)$, and let $\varepsilon > 0$. Then with $\mathcal{O}((m+1)R_0^2)$ iterations of a damped Newton method and*

$$\mathcal{O}\left(\sqrt{m(R_0^2+1)} \log\left(\frac{mR_0(R_0^2+1)}{\varepsilon}\right)\right)$$

iterations of the path following method, one can find $p_\varepsilon \in \mathbb{H}^n$ such that

$$s(p_\varepsilon) - \inf_{q \in \mathbb{H}^n} s(q) \leq \varepsilon.$$

For not too small ε, the cost is dominated by the damped Newton method, which we use to find a good starting point for the path-following method. We leave it as an open problem as to whether this can be avoided. Furthermore, the above applies only to $\mathbb{H}^n$ rather than to $\text{PD}(n)$: it relies on the barrier constructed using Theorem 1.7, which uses a non-trivial strengthening of the self-concordance estimates for the squared distance. We expect that such a strengthening can also be obtained more generally, and this would immediately generalize the algorithmic result from Theorem 1.10 to these spaces; we also leave this as a problem for future work.

1.7 Organization of the paper

In Section 2, we review standard concepts from Riemannian geometry and convexity that we use later. In Section 3 we define self-concordance and analyze Newton's method, showing its quadratic convergence for self-concordant functions. In Section 4, we define self-concordant barriers and the notion of compatibility, discuss how to construct new self-concordant functions out of old ones, and analyze a path-following method. In Section 5 we describe general properties of the distance function on Hadamard manifolds, show that the squared distance is self-concordant on $\text{PD}(n)$, and give refinements of the self-concordance estimate on the model spaces for constant negative sectional curvature, which are used to construct a barrier for the epigraph of the distance function. In Section 6, we discuss our applications: the first is on norm minimization and noncommutative optimization, the second is on computing the minimum enclosing ball on Hadamard symmetric spaces, and the third is on computing geometric medians on model spaces. We also briefly discuss an application to the Riemannian barycenter problem. We conclude in Section 7, where we mention interesting open problems and future research directions.

2 Preliminaries in Riemannian geometry

In this section, we recall and fix our notation for some basic concepts in Riemannian geometry that we will need in the remainder. We follow the conventions of [Lee18]. See [Lee18, BH99] for comprehensive introductions to Riemannian geometry and non-positive curvature, respectively.

2.1 Metric, lengths, distances

Throughout this paper, we let M denote a connected Riemannian manifold. Unless specified otherwise, all differential geometric objects (manifolds, functions, sections, etc.) are assumed to be C^∞-smooth. We write T_pM and T_p^*M for the tangent and cotangent space at a point $p \in M$, and write TM and T^*M the tangent and cotangent bundle of M, respectively. The space of sections of a vector bundle E on M is denoted by $\Gamma(E)$. Sections of the (co)tangent bundle are called (co)vector fields. Given a function f, we write df for its differential, which is a covector field. Then $Xf = df(X)$ is the directional derivative of f in direction X for any vector field X. The Lie bracket of two vector fields X and Y is the vector field $[X, Y]$ that acts as $[X, Y]f = X(Yf) - Y(Xf)$ on any function f. More generally, for $k, l \geq 0$, a (k, l)-*tensor field* is by definition a section of the bundle $T^{(k,l)}M :=$ $(TM)^{\otimes k} \otimes (T^*M)^{\otimes l}$ or, equivalently, a $C^\infty(M)$-multilinear map $\Gamma(T^*M)^k \times \Gamma(TM)^l \to C^\infty(M)$; when $k = 1$ we can also think of it as a $C^\infty(M)$-multilinear map $\Gamma(TM)^l \to \Gamma(TM)$.

The Riemannian metric on M is a smoothly varying family of inner products on the tangent spaces, i.e., for every $p \in M$ we have an inner product $\langle \cdot, \cdot \rangle_p$ on T_pM such that the map $p \mapsto \langle \cdot, \cdot \rangle_p$ is a section of the bundle $T^{(0,2)}M$. The induced norm on T_pM is denoted by $\|\cdot\|_p$. We write $\langle X, Y \rangle$ and $\|X\|$ for the functions computing the pointwise inner product and norm, respectively, of vector fields X, Y.

Using the Riemannian metric, we can define the *length* of a piecewise regular (meaning smooth and non-zero derivative) curve by $L(\gamma) = \int_a^b \|\dot\gamma(t)\|_{\gamma(t)}dt$. This is independent of the parameterization. In particular, we may always reparameterize such that the curve has unit speed, i.e., $\|\dot\gamma(t)\| = 1$, except for finitely many points; in this case the length is $L(\gamma) = b - a$. Given a notion of length, we define the *Riemannian distance* $d(p, q)$ between any two points $p, q \in M$ as the infimum of the lengths of all piecewise regular curves from p to q. In this way, M becomes a metric space. Its topology is the same as the original topology of the manifold M.

2.2 Covariant derivative and curvature

The Riemannian metric determines the *Levi-Civita connection* ∇. It assigns to any two vector fields X and Y the *covariant derivative* $\nabla_X Y$ of Y along X, which is again a vector field, and is determined uniquely by being a connection on the tangent bundle (meaning it is C^∞-linear in X, $\mathbb{R}$-linear in Y, and satisfies the product rule $\nabla_X(fY) = f\nabla_X Y + (Xf)Y$ for all functions f) which is *compatible with the metric* in the sense that $X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$ and *symmetric*, meaning $\nabla_X Y - \nabla_Y X = [X, Y]$, where $[X, Y]$ denotes the Lie bracket. The $C^\infty(M)$-linearity in X implies that $\nabla_X Y\big|_p$ depends only on the tangent vector $v := X_p$ at the point $p \in M$ and the values of Y in an arbitrarily small neighbourhood of p; accordingly we will also write $\nabla_v Y$. Moreover, $X \mapsto \nabla_X Y$ defines a (1,1)-tensor field, called the *total covariant derivative* ∇Y of Y.

One can uniquely extend the above to define connections and covariant derivatives for all tensor bundles $T^{(k,l)}M$ by demanding that for functions it agrees with the differential, that it satisfies a product rule with respect to tensor products, $\nabla_X(T \otimes S) = (\nabla_X T) \otimes S + T \otimes (\nabla_X S)$ for all vector

13

fields X and tensor fields T, S, and that it commutes with all contractions. As a consequence,

$$
\begin{aligned}
X(T(\omega_1,\ldots,\omega_k,Z_1,\ldots,Z_l)) &= (\nabla_X T)(\omega_1,\ldots,\omega_k,Z_1,\ldots,Z_l) \\
&+ T(\nabla_X \omega_1,\omega_2,\ldots,\omega_k,Z_1,\ldots,Z_l) + \ldots + T(\omega_1,\ldots,\omega_k,Z_1,\ldots,Z_{l-1},\nabla_X Z_l)
\end{aligned}
\tag{2.1}
$$

for any (k,l)-tensor field T, vector fields X, Z_1, $\ldots$, Z_l, and covector fields $\omega_1,\ldots,\omega_k$. Again, we write $\nabla_v T := (\nabla_X T)_p$ as this only depends on the tangent vector $v := X_p$ at the point $p \in M$. For any (k,l)-tensor field T, the map $(\omega_1,\ldots,\omega_k,X,Z_1,\ldots,Z_l) \mapsto (\nabla_X T)(\omega_1,\ldots,\omega_k,Z_1,\ldots,Z_l)$ defines a $(k,1+l)$-tensor field, called the *total covariant derivative* and denoted by ∇T. We note that [Lee18] uses a different convention. In particular, we can define the *Hessian* of a function f as $\nabla^2 f = \nabla(\nabla f)$, which is a $(0,2)$-tensor field that turns out to be symmetric for the Levi-Civita connection; see Section 2.4.

Let $\tilde{M} \subseteq M$ be an embedded submanifold, equipped with the induced metric, and let $\tilde{\nabla}$ denote its Levi-Civita connection. If X,Y are vector fields on $\tilde{M}$ that are extended arbitrarily to a neighborhood of $\tilde{M}$ in M, then the *Gauss formula* holds on $\tilde{M}$:

$$
\nabla_X Y = \tilde{\nabla}_X Y + \mathrm{I\!I}(X,Y),
\tag{2.2}
$$

where $\mathrm{I\!I}(X,Y) := \pi^\perp(\nabla_X Y)$ is the *shape tensor* or *second fundamental form* $\mathrm{I\!I}$ of $\tilde{M}$, with $\pi^\perp \colon TM|_{\tilde{M}} \to (T\tilde{M})^\perp$ the orthogonal projection [Lee18, Thm. 8.2].

While the covariant derivative itself is not a tensor field, it can be used to define the so-called *Riemann curvature tensor* which is a fundamental local invariant of Riemannian manifolds. Given vector fields X, Y, Z, we can define the vector field

$$
R(X,Y)Z := \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z.
$$

We may think of $R(X,Y)$ as a C^∞-linear operator on the tangent bundle; hence R is a $(1,3)$-tensor field. The operator $R(X,Y)$ is skew-symmetric, and it is a skew-symmetric function of X and Y. It further satisfies the algebraic Bianchi identity $R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0$. It can also be useful to define $R(X,Y,Z,W) := \langle R(X,Y)Z,W \rangle$, which is a $(0,4)$-tensor field.

A closely related object is the *sectional curvature*, which given two linearly independent tangent vectors $v,w \in T_pM$ at the same point $p \in M$ is defined by

$$
K(v,w) = \frac{\langle R(v,w)w,v \rangle_p}{\langle v,v \rangle_p \langle w,w \rangle_p - \langle v,w \rangle_p^2}.
$$

It only depends on the two-dimensional tangent plane spanned by v and w. The sectional curvature determines the Riemann curvature tensor uniquely. Its sign is an important characteristic of a Riemannian manifold. We say that M has *non-positive (sectional) curvature* if $K(v,w) \leq 0$ for all $v,w \in T_pM$ and $p \in M$. The next lemma records how these notions behave under rescaling of the Riemannian metric.

Lemma 2.1. *Let M be a Riemannian manifold with Riemannian metric $\langle \cdot,\cdot \rangle$, and let $c > 0$. Let M' be the same manifold but with Riemannian metric given by $\langle \cdot,\cdot \rangle' = c \langle \cdot,\cdot \rangle$. Then M' has the same Levi–Civita connection as M, and hence the same $(1,3)$-curvature tensor. For every $p,q \in M$, one has $d_{M'}(p,q) = \sqrt{c}\, d_M(p,q)$. Furthermore, for all $p \in M$ and linearly independent $v,w \in T_pM = T_pM'$, the sectional curvature satisfies $K_{M'}(v,w) = K_M(v,w)/c$.*

14

2.3 Parallel transport, geodesics, completeness

All definitions given so far restrict naturally to open subsets. However, it is often useful to restrict to curves in a manifold and differentiate a vector or tensor field along it. If γ is a curve defined on an interval $I \subseteq \mathbb{R}$, then a (k,l)-*tensor field along* γ is a function $Y \colon I \to T^{(k,l)}M$ such that $Y(t) \in T^{(k,l)}_{\gamma(t)}M$ for every $t \in I$, i.e., a section of the pullback bundle $\gamma^* T^{(k,l)}$. Then there is a unique $\mathbb{R}$-linear operator D_t, called the *covariant derivative along* γ, that satisfies the product rule $D_t(fY) = \dot{f}Y + fD_tY$ for $f \in C^\infty(I)$ for $f \in C^\infty(I)$, and which agrees with $\nabla_{\dot\gamma(t)}$ for every tensor field that extends to a neighborhood of γ.

A vector or tensor field Y along a curve γ is called *parallel* if its covariant derivative along γ vanishes identically, i.e., $D_tY \equiv 0$. For any curve $\gamma \colon I \to M$, $0 \in I$, and any tensor $y_0 \in T^{(k,l)}_{\gamma(0)}M$, standard results in ordinary differential equations imply that there always exists a unique parallel tensor field Y along γ such that $\gamma(0) = y_0$, called the *parallel transport* of y_0 along γ. For any $t \in I$, we get a linear isomorphism $\tau_{\gamma,t} \colon T^{(k,l)}_{\gamma(0)}M \to T^{(k,l)}_{\gamma(t)}M$ by setting $\tau_{\gamma,t}(y_0) = Y(t)$ called a *parallel transport map*. This is useful to compute covariant derivatives: if T is a (k,l)-tensor field then for all $p \in M$, $v \in T_pM$, $\eta_1, \ldots, \eta_k \in T_p^*M$, and $w_1, \ldots, w_l \in T_pM$ we have

$$\nabla_v T(\eta_1, \ldots, \eta_k, w_1, \ldots, w_l) = \partial_{t=0} T_{\gamma(t)}(\tau_{\gamma,t}\eta_1, \ldots, \tau_{\gamma,t}\eta_k, \tau_{\gamma,t}w_1, \ldots, \tau_{\gamma,t}w_l), \qquad (2.3)$$

where γ is an arbitrary curve such that $\gamma(0) = p$ and $\dot\gamma(0) = v$. We are often interested in parallel transport along the manifold's geodesics, which we introduce next.

A curve γ is called a *geodesic* if it is parallel to its own tangent vector field, i.e., $D_t\dot\gamma \equiv 0$. For every $p \in M$ and $v \in T_pM$, there is a unique geodesic $\gamma \colon I \to M$ with $\gamma(0) = p$ and $\dot\gamma(0) = v$, defined on some maximal open interval I containing 0. Note that $\dot\gamma(t) = \tau_{\gamma,t}(\dot\gamma(0))$ for all $t \in I$. If $1 \in I$, we define $\mathrm{Exp}_p(v) := \gamma_v(1)$. We call M *geodesically complete* if $I = \mathbb{R}$, i.e., if geodesics with arbitrary initial data exist for arbitrary times. Then the exponential map is defined on the whole tangent space, $\mathrm{Exp}_p \colon T_pM \to M$. The Hopf–Rinow theorem states that if M is connected, geodesic completeness is equivalent to completeness with respect to the Riemannian distance function, as well as to the Heine–Borel property (bounded closed subsets are compact).

Any length-minimizing curve is a geodesic when parameterized with unit speed. In general, geodesics are only locally length-minimizing, but when M is connected and complete then any two points $p, q \in M$ are connected by a length-minimizing geodesic, although there may be many other geodesics. However, if M is not only complete but also has non-positive sectional curvature, then by the Cartan–Hadamard theorem the exponential map at each point is a covering map. In particular, if M also is simply connected, then the exponential map is a diffeomorphism, so there is a *unique* (up to reparameterization) geodesic connecting any two points p and q. We will denote the corresponding parallel transport by $\tau_{p \to q}$. Manifolds that are simply connected, geodesically complete, and have non-positive sectional curvature are called *Hadamard manifolds*. This includes a great variety of spaces of import in applications, such as Euclidean and hyperbolic spaces, the positive definite matrices, and other symmetric spaces with non-positive curvature (see Sections 5 and 6).

2.4 Gradient and Hessian

Given a function $f \colon D \to \mathbb{R}$ defined on an open subset $D \subseteq M$, we define its *gradient* as the vector field $\mathrm{grad}(f)$ that is dual to its differential. That is, for all vector fields X we have

$$\langle \mathrm{grad}(f), X \rangle = df(X) = Xf.$$

The *Hessian* of f is defined as the second covariant derivative $\nabla^2 f = \nabla(\nabla f) = \nabla df$, which is a $(0,2)$-tensor field, that is, a smoothly varying family of bilinear forms. By definition and using

, we have for any two vector fields X and Y that

$$(\nabla^2 f)(X, Y) = (\nabla_X df)(Y) = X(df(Y)) - df(\nabla_X Y) = X(Yf) - (\nabla_X Y)f, \qquad (2.4)$$

which implies that Hessian is a *symmetric* tensor, by the symmetry of the Levi-Civita connection. Since the Hessian is a symmetric tensor, it is determined by the associated quadratic form. The latter can be conveniently calculated in terms of geodesics: for any $p \in M$ and $v \in T_p M$,

$$(\nabla^2 f)_p(v, v) = \partial_{t=0}^2 f(\mathrm{Exp}_p(tv)). \qquad (2.5)$$

Using metric compatibility, one can write $(\nabla^2 f)(X, Y) = \langle \nabla_X \mathrm{grad}(f), Y \rangle$, which shows that the $(1,1)$-tensor field $\mathrm{Hess}(f) := \nabla \mathrm{grad}(f)$ is the natural operator definition of the Hessian.

One can similarly consider higher covariant derivatives, but these need no longer be symmetric as a consequence of the non-vanishing of the curvature tensor. In particular, the third covariant derivative is no longer captured by its diagonal $(\nabla^3 f)_p(v, v, v) = \partial_{t=0}^3 f(\mathrm{Exp}_p(tv))$. This complicates the theory of self-concordance, as we will discuss in Section 3.

2.5 Convexity

Finally we recall here some basic notions of convexity on Riemannian manifolds. We first discuss convexity of subsets and then turn to convexity of functions. We assume that M is connected and geodesically complete, so that any two points are connected by a (length-minimizing) geodesic.

A subset $D \subseteq M$ is called *(totally) convex* if for every geodesic $\gamma \colon [0, 1] \to M$ with $\gamma(0) \in D$ and $\gamma(1) \in D$, it holds that $\gamma(t) \in D$ for all $t \in [0, 1]$. We remark that, in general, two points can be connected by more than one geodesic; accordingly there is more than one natural definition of convexity. We are primarily interested in applications to Hadamard spaces, where any two points are connected by a unique geodesic, just like in Euclidean space.

A (not necessarily continuous) function $f \colon D \to \mathbb{R}$ defined on a convex subset $D \subseteq M$ is called *convex* if for every geodesic $\gamma \colon [0, 1] \to M$ with $\gamma(0) \in D$ and $\gamma(1) \in D$, it holds that $f \circ \gamma \colon [0, 1] \to \mathbb{R}$ is convex. That is, f is convex along all geodesics in its domain. Equivalently, f is convex if and only if its epigraph

$$E_f = \{(p, t) \in D \times \mathbb{R} : f(p) \leq t\} \qquad (2.6)$$

is a convex subset of $M \times \mathbb{R}$. If the epigraph is also closed as a subset of $M \times \mathbb{R}$, then f is called *closed convex*. This useful condition controls the behavior of a convex function at its boundary, as in the following lemma, which thanks to the Hopf-Rinow theorem can be proved just like in the Euclidean case [Nes18, Thm. 3.1.4]. In particular, any *continuous* convex function on a *closed* domain is closed convex. Parts (i) and (ii) state that any closed convex function $f \colon D \to \mathbb{R}$ is lower semicontinuous, also if we extend it to M by setting $f(p) = \infty$ for $p \notin D$ (in fact, this characterizes when a convex function is closed, but we will not need this).

Lemma 2.2. *Let $f \colon D \to \mathbb{R}$ be a (not necessarily continuous) closed convex function defined on a convex subset $D \subseteq M$. Then:*

(i) *If $(p_k) \subseteq D$ is a sequence such that $p_\infty := \lim_{k \to \infty} p_k \in D$, then $\liminf_{k \to \infty} f(p_k) \geq f(p_\infty)$.*

(ii) *If $(p_k) \subseteq D$ is a sequence such that $\lim_{k \to \infty} p_k \notin D$, then $\lim_{k \to \infty} f(p_k) = \infty$.*

(iii) *If for some $L \in \mathbb{R}$ the level set $\mathcal{L} = \{p \in D : f(p) \leq L\}$ is non-empty and bounded, then f attains its minimum.*

Proof. (i) We need to show: for any subsequence (p_{k_j}) such that $\lim_{j\to\infty} f(p_{k_j}) = f_\infty$ for some $f_\infty \in \mathbb{R} \cup \{\pm\infty\}$, we have that $f_\infty \geq f(p_\infty)$. If $f_\infty = \infty$ there is nothing to show. If $f_\infty \in \mathbb{R}$ then we have $\lim_{j\to\infty}(p_{k_j}, f(p_{k_j})) = (p_\infty, f_\infty) \in E_f$, since the epigraph is closed, and hence $f_\infty \geq f(p_\infty)$. Finally, we note $f_\infty = -\infty$ cannot occur. Indeed, if $f_\infty = -\infty$ then $f(p_{k_j}) \leq f(p_\infty) - 1$ for j large enough, hence $(p_{k_j}, f(p_\infty) - 1) \in E_f$ for j large enough and hence $\lim_{j\to\infty}(p_{k_j}, f(p_\infty) - 1) = (p_\infty, f(p_\infty) - 1) \in E_f$, which is a contradiction.

(ii) Assume this is not so. Then there are a subsequence (p_{k_j}) and $L \in \mathbb{R}$ such that $f(p_{k_j}) \leq L$ for all j. Now, $\lim_{j\to\infty}(p_{k_j}, L) = (p_\infty, L)$, where $p_\infty := \lim_{k\to\infty} p_k$, but each (p_{k_j}, L) is contained in the epigraph, and hence the same must be true for the limit. It follows that $p \in D$, which is a contradiction.

(iii) Since the level set $\mathcal{L}$ is non-empty, it contains a sequence (p_k) such that $\lim_{k\to\infty} f(p_k) = f_* := \inf_{p\in D} f(p)$. Because the epigraph is a closed subset of $M \times \mathbb{R}$, the same is true for $\mathcal{L} \times \{L\} = E_f \cap (M \times \{L\})$, and hence $\mathcal{L}$ is a closed subset of M. It is also bounded by assumption. By the Hopf–Rinow theorem, which is applicable because we assume that M is geodesically complete, it follows that $\mathcal{L}$ is compact. After passing to a subsequence, we may therefore assume that $p_\infty := \lim_{k\to\infty} p_k$ exists and is in $\mathcal{L} \subseteq D$. For continuous f, we then have $f(p_\infty) = f_*$ and this concludes the proof. If f is not continuous then we can proceed as follows. First suppose that $f_* = -\infty$. Fix any $p_0 \in \mathcal{L}$. Because M is geodesically complete and $\mathcal{L}$ is bounded, there exists a constant $C > 0$ such that we can write $p_k = \mathrm{Exp}_{p_0}(u_k)$ for some $u_k \in T_{p_0}M$ such that $\|u_k\|_{p_0} = d(p_0, p_k) \leq C$ for all k. Then we can choose $\alpha_k \in (0, 1)$ such that $\alpha_k \to 0$ and $\alpha_k f(p_k) \to -\infty$. Then the points $q_k := \mathrm{Exp}_{p_0}(\alpha_k u_k)$ satisfy

$$f(q_k) \leq (1 - \alpha_k)f(p_0) + \alpha_k f(p_k) = f(p_0) + \alpha_k(f(p_k) - f(p_0)) \to -\infty,$$

where the first inequality holds by geodesic convexity. In particular, there is some constant $K \in \mathbb{R}$ such that $f(q_k) \leq K < f(p_0)$ for large enough k. Now, (q_k, K) is in the epigraph and converges to (p_0, K), because $\alpha_k \to 0$ and $\|u_k\|_{p_0} \leq C$ for all k. But $f(p_0) > K$, so (p_0, K) is not in the epigraph. This contradicts the assumption that the epigraph is closed. Thus we must have that $f_* > -\infty$. Then, $\lim_{k\to\infty}(p_k, f(p_k)) = (p_\infty, f_*)$ and since the epigraph is closed, it must contain the latter, meaning that $f(p_\infty) \leq f_*$ and hence $f(p_\infty) = f_*$. $\qquad\square$

We will later (in Section 4) be in the situation that $D \subseteq M$ is open and we are interested in smooth objective functions $f\colon D \to \mathbb{R}$ that have a *closed convex extension*, meaning that f extends to a closed convex function on some convex superset of D. This is the case in particular if f extends to a continuous convex function on the closure $\overline{D}$.

Just like in the Euclidean setting [Nes18, Thm. 3.1.5], one can see that the sum of two closed convex functions is again closed convex.

Lemma 2.3. *Let $f_1\colon D_1 \to \mathbb{R}$, $f_2\colon D_2 \to \mathbb{R}$ be closed convex functions defined on convex subsets $D_1, D_2 \subseteq M$. Then the function $f_1 + f_2$ is a closed convex function on $D_1 \cap D_2$.*

Proof. It is clear that $f_1 + f_2$ is a convex function on $D := D_1 \cap D_2$. To see that it is closed, consider an arbitrary convergent sequence (p_k, t_k) in $E_{f_1+f_2}$, with limit point $(p_\infty, t_\infty) \in M \times \mathbb{R}$. By Lemma 2.2, since f_1 and f_2 are closed convex, we have

$$\liminf_{k\to\infty} f_1(p_k) \geq f_1(p_\infty) \quad \text{and} \quad \liminf_{k\to\infty} f_2(p_k) \geq f_2(p_\infty),$$

and hence

$$t_\infty = \lim_{k\to\infty} t_k \geq \liminf_{k\to\infty} f_1(p_k) + \liminf_{k\to\infty} f_2(p_k) \geq f_1(p_\infty) + f_2(p_\infty),$$

which means that $(p_\infty, t_\infty) \in E_{f_1+f_2}$. Hence $f_1 + f_2$ is closed. $\qquad\square$

As in the Euclidean setting, one can also characterize convexity differentially. In particular, a C^2-smooth function $f \colon D \to \mathbb{R}$ defined on an open convex subset $D \subseteq M$ is convex if and only if the quadratic forms defined by the Hessian are positive semidefinite, i.e.,

$$(\nabla^2 f)_p(v, v) \geq 0 \tag{2.7}$$

for all $v \in T_p M$ and $p \in D$. We discuss two refinements of the notion of convexity (for simplicity only in the C^2-smooth setting): If f is strictly convex along any geodesic in the domain, then f is called strictly convex. A sufficient condition for strict convexity is the following: for every $p \in D$, the Hessian $(\nabla^2 f)_p$ is positive definite, i.e., Eq. (2.7) holds with equality only for $v = 0 \in T_p M$. Similarly, we say that f is μ-strongly convex for some $\mu > 0$ if it is so along any unit-speed geodesic in the domain. This is the case if and only if, for all $v \in T_p M$ and $p \in D$,

$$(\nabla^2 f)_p(v, v) \geq \mu \|v\|_p^2.$$

In convex optimization, upper bounds on the Hessian of a convex function are often also useful. We say that f is ν-smooth (not to be confused with smoothness in the sense of C^∞) if it is so along any unit-speed geodesic in the domain, that is, if and only if

$$(\nabla^2 f)_p(v, v) \leq \nu \|v\|_p^2$$

for all $v \in T_p M$ and $p \in D$. When M is a Hadamard space then it is well-known that the distance $d(\cdot, p_0)$ to any fixed point $p_0 \in M$ is convex, and that $\frac{1}{2} d^2(\cdot, p_0)$ is 1-strongly convex, just like in Euclidean space. However, the latter will in general no longer be smooth. We discuss these important functions in Section 5.

Let $D \subseteq M$ be a convex subset (not necessarily open) that is also an embedded submanifold. Equip D with the induced metric and let $\tilde{\nabla}$ denote its Levi-Civita connection. Then D is a totally geodesic submanifold, so its shape tensor $\mathbb{II}$ vanishes [Lee18, Prop. 8.12]. Now let T be a $(0, l)$-tensor field on D that is extended arbitrarily to a neighborhood of D in M. Then by Eqs. (2.1) and (2.2) we find that $\tilde{\nabla} T = \nabla T|_{(TD)^{\otimes(1+l)}}$, where the right-hand side notation means that we restrict ∇T to a $(0, 1+l)$-tensor field on D. In particular, we inductively see that for every function $f \colon M \to \mathbb{R}$ and every $l \geq 0$, the following holds on D:

$$\tilde{\nabla}^l \tilde{f} = \nabla^l f|_{(TD)^{\otimes l}}. \tag{2.8}$$

3 Self-concordance and Newton's method on manifolds

In this section we generalize the notion of self-concordance and the corresponding analysis of Newton's method from the Euclidean setting to the Riemannian setting, and we comment on complications incurred by curvature. For expositions of the Euclidean theory of self-concordance and interior-point methods we refer to [NN94, Nes18, Ren01]. Throughout this section we assume that M is a connected and geodesically complete Riemannian manifold.

3.1 Self-concordance

Let $f \colon D \to \mathbb{R}$ be a convex function defined on an open convex subset $D \subseteq M$. Then the Hessian is positive semidefinite, by Eq. (2.7), hence induces a (semi-)norm at each point. The rate of change of the Hessian is captured by the third covariant derivative, $\nabla^3 f = \nabla(\nabla(\nabla f)) = \nabla(\nabla^2 f)$. A function is called self-concordant if the latter can be bounded in terms of the former, as follows:

Definition 3.1 (Self-concordance)**.** *Let* $f\colon D \to \mathbb{R}$ *be a convex function defined on an open convex subset* $D \subseteq M$, *and let* $a > 0$. *We say that* f *is* α-self-concordant *if, for all* $p \in D$ *and for all* $u, v, w \in T_p M$, *we have*

$$|(\nabla^3 f)_p(u, v, w)| \leq \frac{2}{\sqrt{\alpha}} \sqrt{(\nabla^2 f)_p(u, u)} \sqrt{(\nabla^2 f)_p(v, v)} \sqrt{(\nabla^2 f)_p(w, w)}, \tag{3.1}$$

It is called strongly α-self-concordant *if is not just convex but closed convex, that is, if its epigraph* (2.6) *is a closed subset of* $M \times \mathbb{R}$.

Here we follow the conventions of [NN94]. To interpret the definition, let us for a convex function f, a point p in its domain, and $\alpha > 0$ define the positive semidefinite bilinear form and seminorm

$$\langle v, w \rangle_{f,p,\alpha} = \frac{(\nabla^2 f)_p(v, w)}{\alpha} \quad \text{and} \quad \|u\|_{f,p,\alpha} = \sqrt{\frac{(\nabla^2 f)_p(u, u)}{\alpha}}. \tag{3.2}$$

When the Hessian is positive definite (as is the case, e.g., when f is strongly convex), these endow M with a new Riemannian metric. In convex optimization, $\langle \cdot, \cdot \rangle_{f,p,\alpha}$ is called the "local inner product" and $\|\cdot\|_{f,p,\alpha}$ the "local norm", but we will refrain from using this terminology as it is ambiguous in the Riemannian setting. For $\alpha = 1$, we will usually abbreviate $\langle \cdot, \cdot \rangle_{f,p} := \langle \cdot, \cdot \rangle_{f,p,1}$ and $\|\cdot\|_{f,p} := \|\cdot\|_{f,p,1}$. We can now rewrite Eq. (3.1) as follows:

$$|(\nabla^3 f)_p(u, v, w)| \leq 2\alpha \|u\|_{f,p,\alpha} \|v\|_{f,p,\alpha} \|w\|_{f,p,\alpha}. \tag{3.3}$$

Thus self-concordance can be interpreted as a boundedness of the third covariant derivatives at each point with respect to the seminorms defined by the Hessian.

We record some basic properties. Recall that self-concordant functions are defined on an open and convex domain, by definition.

Lemma 3.2. *(i) Let* f *be a (strongly)* α-self-concordant function and let $c > 0$. *Then* cf *is (strongly)* $c\alpha$-self-concordant.

(ii) *Let* $f_k\colon D_k \to \mathbb{R}$ *be* α_k-self-concordant functions for $k = 1, 2$, *and suppose* $D := D_1 \cap D_2$ *is non-empty. Then* $f := f_1 + f_2\colon D \to \mathbb{R}$ *is* α-self-concordant, with $\alpha := \min(\alpha_1, \alpha_2)$. *If the functions* f_k *are strongly* α_k-self-concordant, then f *is strongly* α-self-concordant.

(iii) *Let* $f_k\colon D_k \to \mathbb{R}$ *be* α-self-concordant functions for $k = 1, 2$. *Then the function* $f\colon D_1 \times D_2 \to \mathbb{R}$ *defined by* $f(p_1, p_2) := f_1(p_1) + f_2(p_2)$ *is* α-self-concordant. *If both functions* f_k *are strongly* α-self-concordant, then so is f.

Property (i) follows from the definition, and (iii) follows from (ii). Before we prove (ii), we give a simpler characterization of self-concordance. As the Hessian is symmetric, third covariant derivatives are symmetric in the last two arguments. This can also be seen explicity from the following formula for the third covariant derivative $\nabla^3 f$, which follows from Eq. (2.1) and holds for any three vector fields X, Y, Z:

$$(\nabla^3 f)(X, Y, Z) = X\big((\nabla^2 f)(Y, Z)\big) - (\nabla^2 f)(\nabla_X Y, Z) - (\nabla^2 f)(Y, \nabla_X Z). \tag{3.4}$$

This leads to the following simplification:

Lemma 3.3. *A convex function $f : D \to \mathbb{R}$ defined on an open convex subset $D \subseteq M$ is α-self-concordant if, and only if, for all $p \in M$ and $u, v \in T_p M$, we have*

$$|(\nabla^3 f)_p(u, v, v)| \leq \frac{2}{\sqrt{\alpha}} \sqrt{(\nabla^2 f)_p(u, u)} \; (\nabla^2 f)_p(v, v) \tag{3.5}$$

or, equivalently,

$$|(\nabla^3 f)_p(u, v, v)| \leq 2\alpha \|u\|_{f, p, \alpha} \|v\|_{f, p, \alpha}^2. \tag{3.6}$$

However, third covariant derivatives are *not* symmetric when M is a curved manifold, as follows from the Ricci identity [Lee18, Thm. 7.14]. To see this, we combine Eqs. (2.4) and (3.4) to see that for any three vector fields X, Y, Z:

$$(\nabla^3 f)(X, Y, Z) = X(Y(Zf)) - X((\nabla_Y Z)f) - (\nabla_X Y)(Zf) + (\nabla_{\nabla_X Y} Z)f$$
$$- Y((\nabla_X Z)f) + (\nabla_Y(\nabla_X Z))f.$$

Using symmetry of the Levi-Civita connection, one finds that

$$(\nabla^3 f)(X, Y, Z) - (\nabla^3 f)(Y, X, Z) = -(R(X, Y)Z)f = -\langle R(X, Y)Z, \operatorname{grad}(f) \rangle \tag{3.7}$$

Accordingly, the third covariant derivative is in general *not* symmetric. Indeed, the asymmetry is precisely related to the nonvanishing of the Riemann curvature tensor!

Due to this asymmetry, to establish self-concordance, we have to show Eq. (3.5) for possibly different $u, v \in T_p M$, whereas we could assume $u = v$ in the Euclidean case; see Section 3.2 for more details. The following proof of Lemma 3.2(ii) is a generalization of [Nes18, Thm. 5.1.1] to our setting.

Proof of Lemma 3.2(ii). For $p \in D = D_1 \cap D_2$ and $u, v \in T_p M$, we have

$$\frac{|(\nabla^3 f)_p(u, v, v)|}{2\sqrt{(\nabla^2 f)_p(u, u)}(\nabla^2 f)_p(v, v)} \leq \frac{|(\nabla^3 f_1)_p(u, v, v)| + |(\nabla^3 f_2)_p(u, v, v)|}{2\sqrt{(\nabla^2 f_1)_p(u, u) + (\nabla^2 f_2)_p(u, u)}((\nabla^2 f_1)_p(v, v) + (\nabla f_2)_p(v, v))}$$
$$\leq \frac{x_1 \omega_1 / \sqrt{\alpha_1} + x_2 \omega_2 / \sqrt{\alpha_2}}{\sqrt{x_1^2 + x_2^2}(\omega_1 + \omega_2)}, \tag{3.8}$$

where we let $x_i := \sqrt{(\nabla^2 f_i)_p(u, u)}$ and $\omega_i := (\nabla^2 f_i)_p(v, v)$ for $i = 1, 2$, and for the last estimate we used α_i-self-concordance of f_i. We now upper bound the quantity in Eq. (3.8). Observing invariance under the change $(x_1, x_2, \omega_1, \omega_2) \to (sx_1, sx_2, t\omega_1, t\omega_2)$ for $s, t > 0$, we may consider the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & \omega_1 x_1 / \sqrt{\alpha_1} + \omega_2 x_2 / \sqrt{\alpha_2} \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 1, \; \omega_1 + \omega_2 = 1, \\ & x_1, x_2, \omega_1, \omega_2 \geq 0. \end{aligned}$$

First we fix ω_i, and maximize over the choice of x_i. This is a linear maximization over the intersection of the unit circle with the positive orthant, with objective given by $(\omega_1 / \sqrt{\alpha_1}, \omega_2 / \sqrt{\alpha_2})$, which is itself in the positive orthant. Therefore the maximum is attained at

$$(x_1, x_2) = \frac{(\omega_1 / \sqrt{\alpha_1}, \omega_2 / \sqrt{\alpha_2})}{\sqrt{\omega_1^2 / \alpha_1 + \omega_2^2 / \alpha_2}},$$

where the value of the objective is $\sqrt{\omega_1^2/\alpha_1 + \omega_2^2/\alpha_2}$. This reduces the problem to

$$\text{maximize } \sqrt{\omega_1^2/\alpha_1 + \omega_2^2/\alpha_2} \ \text{ s.t. } \ \omega_1 + \omega_2 = 1, \ \omega_1, \omega_2 \geq 0.$$

By convexity of the objective, the maximum is attained at $(\omega_1, \omega_2) = (1, 0)$ or $(\omega_1, \omega_2) = (0, 1)$. Therefore Eq. (3.8) is at most $\max(1/\sqrt{\alpha_1}, 1/\sqrt{\alpha_2})$, and f is α-self-concordant for $\alpha = \min(\alpha_1, \alpha_2)$. The claim that f is strongly α-self-concordant whenever the f_i are strongly α_i-self-concordant then follows from Lemma 2.3. $\qquad\square$

We now state a key property that is required for the analysis of Newton's method of self-concordant functions. It quantifies the change of the Hessian or local norm as a function of the distance, measured with respect to the norm (3.2), providing a finitary version of Definition 3.1. Then the following result is a direct translation of the Euclidean argument in [NN94, Thm. 2.1.1] along with the notion of self-concordance from Definition 3.1.

Theorem 3.4 (Stability of Hessians). *Let $f : D \to \mathbb{R}$ be an α-self-concordant function defined on an open convex subset $D \subseteq M$, and let $p \in D$. Let $u \in T_p M$ be such that $r := \|u\|_{f,p,\alpha} < 1$. If $q := \mathrm{Exp}_p(u) \in D$, then we have the following estimate: for all $v \in T_p M$,*

$$(1 - r)^2 \, (\nabla^2 f)_p(v, v) \leq (\nabla^2 f)_q(\tau_{\gamma,1} v, \tau_{\gamma,1} v) \leq \frac{1}{(1 - r)^2} \, (\nabla^2 f)_p(v, v), \tag{3.9}$$

or, equivalently,

$$(1 - r)^2 \, (\nabla^2 f)_p \preceq \tau_{\gamma,1}^* (\nabla^2 f)_q \preceq \frac{1}{(1 - r)^2} \, (\nabla^2 f)_p,$$

where $\tau_{\gamma,1}$ denotes the parallel transport along the geodesic $\gamma(t) := \mathrm{Exp}_p(tu)$ from p to q.

Proof. Since the domain is convex, we know that $\gamma(t) = \mathrm{Exp}_p(tu) \in D$ for all $t \in [0, 1]$. Consider the following two functions:

$$\phi : [0, 1] \to \mathbb{R}, \quad \phi(t) = (\nabla^2 f)_{\gamma(t)}(\tau_{\gamma,t} v, \tau_{\gamma,t} v),$$
$$\psi : [0, 1] \to \mathbb{R}, \quad \psi(t) = (\nabla^2 f)_{\gamma(t)}(\tau_{\gamma,t} u, \tau_{\gamma,t} u).$$

Using Eq. (2.3), with $T = \nabla^2 f$ and using that $\dot{\gamma}(t) = \tau_{\gamma,t} u$, we have

$$\dot{\phi}(t) = \big(\nabla_{\dot{\gamma}(t)}(\nabla^2 f)\big)(\tau_{\gamma,t} v, \tau_{\gamma,t} v) = (\nabla^3 f)(\tau_{\gamma,t} u, \tau_{\gamma,t} v, \tau_{\gamma,t} v).$$

Hence, using α-self-concordance as in Eq. (3.1),

$$|\dot{\phi}(t)| \leq \frac{2}{\sqrt{\alpha}} \sqrt{\psi(t)} \, \phi(t). \tag{3.10}$$

Similarly,

$$\dot{\psi}(t) = \big(\nabla_{\dot{\gamma}(t)}(\nabla^2 f)\big)(\tau_{\gamma,t} u, \tau_{\gamma,t} u) = (\nabla^3 f)(\tau_{\gamma,t} u, \tau_{\gamma,t} u, \tau_{\gamma,t} u).$$

and hence using only α-self-concordance along the geodesic γ, as in Eq. (3.13), we find that

$$|\dot{\psi}(t)| \leq \frac{2}{\sqrt{\alpha}} \psi(t)^{3/2}. \tag{3.11}$$

21

With these estimates in place we can proceed as in the proof of [NN94, Thm. 2.1.1]. By Grönwall's inequality, there are two cases: either ψ vanishes identically on the interval $[0,1]$, or it is everwhere positive. In the former case, Eq. (3.10) implies that ϕ is constant and hence $\phi(1) = \phi(0)$, which in turn implies the claim. In the latter case, we can write Eq. (3.11) as

$$\left| \partial_t \psi(t)^{-1/2} \right| = \frac{1}{2} \frac{|\dot\psi(t)|}{\psi(t)^{3/2}} \leq \frac{1}{\sqrt{\alpha}}, \tag{3.12}$$

from which it follows that

$$\psi(t)^{-1/2} \geq \psi(0)^{-1/2} - \frac{t}{\sqrt{\alpha}} = \frac{1}{\sqrt{\alpha}\|u\|_{f,p,\alpha}} - \frac{t}{\sqrt{\alpha}} = \frac{1 - rt}{r\sqrt{\alpha}}$$

and hence, since $r < 1$,

$$\sqrt{\psi(t)} \leq \frac{r\sqrt{\alpha}}{1 - rt}.$$

Thus Eq. (3.10) implies

$$|\dot\phi(t)| \leq \frac{2r}{1 - rt}\,\phi(t).$$

Similarly to the above, either ϕ vanishes identically on $[0,1]$, in which case there is nothing to prove, or it is everywhere positive, in which case we have

$$|\partial_t \log \phi(t)| \leq \frac{2r}{1 - rt}$$

and hence

$$\left| \log \frac{\phi(t)}{\phi(0)} \right| \leq 2 \log \frac{1}{1 - rt}.$$

For $t = 1$ this yields the desired inequality. $\qquad\square$

3.2 Self-concordance along geodesics

When $M = \mathbb{R}^n$ is a Euclidean space, then the third derivative is symmetric in all three arguments, and standard results on trilinear forms [Ban38] imply that the above is equivalent to $\left| \partial_{t=0}^3 f(p + tv) \right| = |(\nabla^3 f)_p(v, v, v)]| \leq 2\alpha \|v\|_{f,p,\alpha}^3$ for all $p, v \in \mathbb{R}^n$, which shows that self-concordance is equivalent to *self-concordance along the geodesics* of Euclidean space. This characterization is highly useful for showing that functions are self-concordant. The richness of the family of self-concordant functions is a key reason for the wide applicability of interior-point methods [NN94, Hil14, Fox15, BE19, Che21].

This notion can also be generalized naturally to the Riemannian setting:

Definition 3.5 (Self-concordance along geodesics)**.** *Let $f \colon D \to \mathbb{R}$ be a convex function defined on an open convex subset $D \subseteq M$, and let $\alpha > 0$. We say that f is α-self-concordant along geodesics if, for all $p \in D$ and for all $v \in T_pM$, we have*

$$\left| \partial_{t=0}^3 f(\mathrm{Exp}_p(tv)) \right| = |(\nabla^3 f)_p(v, v, v)]| \leq \frac{2}{\sqrt{\alpha}} \left((\nabla^2 f)_p(v, v) \right)^{3/2} \tag{3.13}$$

or, equivalently,

$$\left| \partial_{t=0}^3 f(\mathrm{Exp}_p(tv)) \right| = |(\nabla^3 f)_p(v, v, v)]| \leq 2\alpha \|v\|_{f,p,\alpha}^3. \tag{3.14}$$

It is called strongly α-self-concordant along geodesics *if is not just convex but closed convex, that is, if its epigraph* (2.6) *is a closed subset of $M \times \mathbb{R}$.*

In other words, f is (strongly) α-self-concordant along geodesics if and only if for every geodesic $\gamma\colon \mathbb{R} \to M$, the function $f \circ \gamma\colon I \to \mathbb{R}$ is (strongly) α-self-concordant on $I := \gamma^{-1}(D)$. There is also a version of Lemma 3.2 as a direct consequence of the Euclidean result.

Definition 3.5 had been proposed in [Ji07, JMJ07] as a suitable notion of self-concordance in the Riemannian setting. Clearly, any (strongly) self-concordant function is also (strongly) self-concordant along geodesics. However, since third covariant derivatives are *not* symmetric in all arguments when M is a curved manifold, as we saw in Eq. (3.7), self-concordance along geodesics need *not* imply self-concordance in the stronger sense of Definition 3.1, in contrast to what was suggested in [JMJ07, Eq. (3) and Prop. 1]. While self-concordance along geodesics already allows lifting several useful results from the Euclidean theory, it is the stronger notion of Definition 3.1 that is required to prove the fundamental Theorem 3.4, which underpins the analysis of the Newton method in the quadratic convergence regime in Theorem 3.16. We give non-trivial examples of self-concordant functions on curved spaces in Sections 5 and 6.

In the remainder of this section we discuss a number of useful results for functions that are self-concordant along geodesics. These follow directly from the Euclidean theory. While some of these were already proved in [Ji07, JMJ07], we give all proofs to keep the exposition self-contained. We start with a version of [Nes18, Thm. 5.1.5].

Proposition 3.6 (Stability of second derivative along geodesic). *Let $f\colon D \to \mathbb{R}$ be α-self-concordant along geodesics, with $D \subseteq M$ open and convex, and let $p \in D$. Consider any geodesic $\gamma(t) = \mathrm{Exp}_p(tu)$ such that $\gamma(1) \in D$, and set $r := \|u\|_{f,p,\alpha}$. Then the α-self-concordant function $g(t) := f(\gamma(t))$ for $t \in [0,1]$ satisfies the lower bound*

$$\ddot{g}(t) \geq \frac{\ddot{g}(0)}{(1+tr)^2} = \frac{\alpha r^2}{(1+tr)^2}, \tag{3.15}$$

and if $rt < 1$ also the upper bound

$$\ddot{g}(t) \leq \frac{\ddot{g}(0)}{(1-tr)^2} = \frac{\alpha r^2}{(1-tr)^2}. \tag{3.16}$$

Proof. As in the proof of Theorem 3.4, we consider the function

$$\psi\colon [0,1] \to \mathbb{R}, \quad \psi(t) = \ddot{g}(t),$$

and find from Eq. (3.11) that it either vanishes identically on $[0,1]$, in which case the claim holds trivially, or it is everywhere positive, in which case Eq. (3.12) holds, namely for all $t \in [0,1]$,

$$\left| \partial_t \psi(t)^{-1/2} \right| \leq \frac{1}{\sqrt{\alpha}}.$$

Accordingly,

$$\psi(0)^{-1/2}(1-tr) = \psi(0)^{-1/2} - \frac{t}{\sqrt{\alpha}} \leq \psi(t)^{-1/2} \leq \psi(0)^{-1/2} + \frac{t}{\sqrt{\alpha}} = \psi(0)^{-1/2}(1+tr),$$

which implies both bounds. $\qquad\qquad\square$

The lower bound strengthens the one in Eq. (3.9) in the special case that $v = u$. The upper bound implies that any function that is strongly self-concordant along geodesics must contain a certain region in its domain. We first define the region and then state the result.

Definition 3.7 (Dikin ellipsoid)**.** *Let* $f \colon D \to \mathbb{R}$ *be a convex function defined on an open convex subset* $D \subseteq M$, *and let* $\alpha > 0$. *Then the* (open) *Dikin ellipsoid of radius* $r > 0$ *at* $p \in M$ *is*

$$B^{\circ}_{f,p,\alpha}(r) = \big\{\mathrm{Exp}_p(u) : u \in T_pM, \ \|u\|_{f,p,\alpha} < r\big\}.$$

For $\alpha = 1$, *we abbreviate* $B^{\circ}_{f,p} := B^{\circ}_{f,p,1}$.

The following result is easily generalized from the Euclidean setting. The proof is essentially the same as in [NN94, Thm. 2.1.1].

Corollary 3.8 (Dikin inclusion)**.** *Let* $f \colon D \to \mathbb{R}$ *be strongly* α-*self-concordant along geodesics, defined on an open convex subset* $D \subseteq M$. *Then* $B^{\circ}_{f,p,\alpha}(1) \subseteq D$ *for every* $p \in D$.

Proof. Take any $v \in T_pM$ such that $r := \|v\|_{f,p,\alpha} < 1$. Let σ be the supremum of those $s \geq 0$ such that $\gamma(s) := \mathrm{Exp}_p(sv) \in D$. Since $p \in D$ and D is open, we know that $\sigma > 0$, and since D is convex, we know that $\gamma(s) \in D$ for all $s \in [0, \sigma)$.

We need to show that $\gamma(1) \in D$ and claim that in fact $\sigma > 1/r > 1$ (with $1/0 = \infty$). For sake of finding a contradiction, assume that this is not so, i.e., that $\sigma \leq 1/r$. For every $s \in [0, \sigma)$ we can apply Proposition 3.6 with $u := sv$, which satisfies $\|u\|_{f,p,\alpha} = sr < \sigma r \leq 1$. Then the upper bound in Eq. (3.16) gives

$$\ddot{g}(s) \leq \frac{1}{(1 - sr)^2} \ddot{g}(0),$$

where $g(s) = f(\gamma(s))$. Accordingly, the function g has bounded derivative on $[0, \sigma)$, thus it is itself bounded on this interval, say $g(s) \leq L$ for some $L \in \mathbb{R}$. As f is strongly self-concordant, the level set $\{q \in D : f(q) \leq L\}$ is closed in M, and hence it must contain $\gamma(\sigma) = \lim_{s \uparrow \sigma} \gamma(s)$. But D is open, so this in turn implies there must also exist some $t > \sigma$ such that $\gamma(t) \in D$, contradicting the definition of σ. $\qquad\square$

In other words, for any $p \in D$ and $u \in T_pM$ such that $\|u\|_{f,p,\alpha} < 1$ it is automatically true that $\mathrm{Exp}_p(u) \in D$, so we do not have to assume this in Theorem 3.4 and Proposition 3.6.

The above also implies that a strongly-self-concordant function can only have a degenerate Hessian if its domain contains a geodesic.

Corollary 3.9 (Domain)**.** *If a strongly* α-*self-concordant function* $f \colon D \to \mathbb{R}$ *contains no* (infinite) *geodesic in its domain, then* $(\nabla^2 f)_p$ *is positive definite for all* $p \in D$. *In particular, this is the case if* M *is a Hadamard manifold and the domain is bounded.*

Proof. If $(\nabla^2 f)_p(u, u) = 0$ for some $p \in D$ and $u \in T_pM$, then $\mathrm{Exp}_p(\mathbb{R}u) \subseteq B^{\circ}_{f,p,\alpha}(1)$. Thus Corollary 3.8 shows that D contains the geodesic $\gamma(t) = \mathrm{Exp}_p(tu)$ for $t \in \mathbb{R}$. $\qquad\square$

The following results bound a self-concordant function in terms of its linear approximation at some arbitrary point, in terms of the quantity

$$\rho \colon (-\infty, 1) \to \mathbb{R}, \quad \rho(r) = -r - \log(1 - r), \tag{3.17}$$

which is $\rho(r) = \frac{1}{2}r^2 + \mathcal{O}(r^3)$ for small r. The first result lifts [Nes18, Thm. 5.1.8] to the geodesic setting and follows directly by integrating the lower bound in Proposition 3.6.

Corollary 3.10 (Lower bound). *Let $f\colon D \to \mathbb{R}$ be α-self-concordant along geodesics, defined on an open convex subset $D \subseteq M$, and let $p \in D$. Then, for every $u \in T_p M$ such that $q := \mathrm{Exp}_p(u) \in D$, we have*

$$df_q(\tau_{\gamma,t} u) - df_p(u) \geq \frac{\alpha t r^2}{1 + tr} \tag{3.18}$$

where $r := \|u\|_{f,p,\alpha}$ and $\tau_{\gamma,t}$ denotes the parallel transport along the geodesic $\gamma(t) := \mathrm{Exp}_p(tu)$ from p to q, and

$$f(q) \geq f(p) + df_p(u) + \alpha\rho(-r).$$

Proof. By Proposition 3.6, we see that $g(t) := f(\mathrm{Exp}_p(tu))$ satisfies

$$\ddot{g}(t) \geq \frac{\alpha r^2}{(1 + tr)^2}$$

for all $t \in [0, 1]$. By integrating,

$$\dot{g}(t) - \dot{g}(0) \geq \int_0^t \frac{\alpha r^2}{(1 + sr)^2} ds = \frac{\alpha t r^2}{1 + tr}.$$

Since $\dot{g}(0) = df_p(u)$ and $\dot{g}(1) = df_q(\tau_{\gamma,1} u)$, this proves the first bound. One more integral yields

$$g(1) - g(0) - \dot{g}(0) \geq \int_0^1 \frac{\alpha s r^2}{1 + sr} ds = \alpha(r - \log(1 + r)) = \alpha\rho(-r). \qquad \square$$

The second result generalizes [Nes18, Thm. 5.1.9] to the geodesic setting and follows by similarly integrating the upper bound in Proposition 3.6.

Corollary 3.11 (Upper bound). *Let $f\colon D \to \mathbb{R}$ be α-self-concordant along geodesics, defined on an open convex subset $D \subseteq M$, and let $p \in D$. Then, for every $u \in T_p M$ such that $q := \mathrm{Exp}_p(u) \in D$ and $r := \|u\|_{f,p,\alpha} < 1$, we have*

$$df_q(\tau_{\gamma,t} u) - df_p(u) \leq \frac{\alpha t r^2}{1 - rt},$$

where $\tau_{\gamma,t}$ denotes the parallel transport along the geodesic $\gamma(t) = \mathrm{Exp}_p(tu)$ from p to q, and

$$f(q) \leq f(p) + df_p(u) + \alpha\rho(r).$$

If f is strongly α-self-concordant along geodesics, then the requirement that $q \in D$ is automatic (by Corollary 3.8).

Proof. Similarly to the proof of Corollary 3.10, we can apply Proposition 3.6 to see that the function $g(t) := f(\mathrm{Exp}_p(tu))$ satisfies

$$\ddot{g}(t) \leq \frac{\alpha r^2}{(1 - tr)^2}$$

for all $t \in [0, 1]$. By integration,

$$\dot{g}(t) - \dot{g}(0) \leq \int_0^t \frac{\alpha r^2}{(1 - sr)^2} ds = \frac{\alpha t r^2}{1 - tr}$$

and

$$g(1) - g(0) - \dot{g}(0) \leq \int_0^1 \frac{\alpha s r^2}{1 - sr} ds = \alpha(-r - \log(1 - r)) = \alpha\rho(r). \qquad \square$$

3.3 Newton's method

We are now ready to give an analysis of Newton's method for self-concordant functions. In particular, as in the Euclidean case, we are able to provide quadratic guarantees on the changes in the so-called Newton decrement (Theorem 3.16). This key result requires self-concordance. Afterwards we also recall some useful results due to [Ji07, JMJ07] which only rely on self-concordance along geodesics.

Recall Newton's method (cf. [Udr94, §7.5]): given a convex function f and a point p in its domain, consider its local quadratic approximation

$$f(\mathrm{Exp}_p(v)) \approx f(p) + df_p(v) + \frac{1}{2}(\nabla^2 f)_p(v, v)$$

and minimize the right-hand side over all $v \in T_p M$. If $(\nabla^2 f)_p$ is non-degenerate and hence positive definite, as we will assume for convenience, there is a unique minimizer called the Newton step.

Definition 3.12 (Newton step and Newton iterate). *Let $f \colon D \to \mathbb{R}$ be a convex function defined on an open convex set $D \subseteq M$, and let $p \in D$ be a point such that $(\nabla^2 f)_p$ is positive definite. Then we define the* Newton step *of f at p as the unique vector $n_{f,p} \in T_p M$ such that*

$$(\nabla^2 f)_p(n_{f,p}, \cdot) = -df_p \tag{3.19}$$

and the Newton iterate *of f at p is defined as*

$$p_{f,+} := \mathrm{Exp}_p(n_{f,p}) \in M,$$

which need not be in D. We can also write

$$n_{f,p} = -\mathrm{Hess}(f)_p^{-1} \mathrm{grad}(f)_p \quad and \quad p_{f,+} = \mathrm{Exp}_p(-\mathrm{Hess}(f)_p^{-1} \mathrm{grad}(f)_p).$$

in terms of the gradient vector and Hessian operator (cf. Section 2.4).

The gap between the function value and the minimum of the quadratic approximation is

$$\frac{1}{2}(\nabla^2 f)_p(n_{f,p}, n_{f,p}) = \frac{\alpha}{2}\|n_{f,p}\|_{f,p,\alpha}^2 = \frac{\alpha}{2}\lambda_{f,\alpha}(p)^2,$$

where $\lambda_{f,\alpha}$ is the so-called Newton decrement, which we define next.

Definition 3.13 (Newton decrement). *Let $f \colon D \to \mathbb{R}$ be a convex function defined on an open convex set $D \subseteq M$, let $p \in D$ be a point such that $(\nabla^2 f)_p$ is positive definite, and let $\alpha > 0$. Then we define the* Newton decrement *of f at p by*

$$\lambda_{f,\alpha}(p) := \|n_{f,p}\|_{f,p,\alpha} = \frac{1}{\alpha}\|df_p\|_{f,p,\alpha}^* = \max_{0 \neq v \in T_p M} \frac{|df_p(v)|}{\alpha\|v\|_{f,p,\alpha}} = \max_{0 \neq v \in T_p M} \frac{|df_p(v)|}{\sqrt{\alpha(\nabla^2 f)_p(v, v)}},$$

where $\|\omega\|_{f,p,\alpha}^* := \max_{0 \neq v \in T_p M} \frac{|\omega(v)|}{\|v\|_{f,p,\alpha}}$ is the dual norm on $T_p^* M$ induced by $\|\cdot\|_{f,p,\alpha}$. That is,[8]

$$\lambda_{f,\alpha}(p) = \min\{\lambda \geq 0 : df_p \otimes df_p \preceq \lambda^2 \alpha (\nabla^2 f)_p\} \tag{3.20}$$

$$= \min\{\lambda \geq 0 : -df_p(u) - \frac{1}{2}(\nabla^2 f)_p(u, u) \leq \frac{\lambda^2 \alpha}{2} \ \forall u \in T_p M\}. \tag{3.21}$$

For $\alpha = 1$, we abbreviate $\lambda_f := \lambda_{f,1}$ and $\|\cdot\|_{f,p}^* := \|\cdot\|_{f,p,1}^*$.

---

[8]To see the second equality, replace u by tu for $t \in \mathbb{R}$, and maximize over t.

The Newton decrement is invariant under rescaling f in the sense that $\lambda_{f,\alpha} = \lambda_{cf,c\alpha}$ for any constant $c > 0$ (cf. Lemma 3.2). When $(\nabla^2 f)_p$ is degenerate, the Newton decrement can still be defined as $\lambda_{f,\alpha}(p) = \inf\{c \geq 0 : |df_p(v)| \leq \alpha c \|v\|_{f,p,\alpha} \; \forall v \in T_pM\}$, which has the same interpretation as explained above; but we will mostly not need this.

Just like in the Euclidean case the Newton decrement provides a certificate for the existence of minimizers and the function gap. This essentially follows from the Euclidean argument [Nes18, Thm. 5.1.13].

Proposition 3.14 (Existence of minimizers). *Let $f\colon D \to \mathbb{R}$ be α-self-concordant along geodesics, defined on an open convex subset $D \subseteq M$. If $p \in D$ is such that $\lambda_{f,\alpha}(p) < 1$, then f is bounded from below: we have*

$$f_* := \inf_{q \in D} f(q) \geq f(p) - \alpha\rho(\lambda_{f,\alpha}(p)), \tag{3.22}$$

where ρ is the quantity defined in Eq. (3.17). If in addition f is strongly α-self-concordant along geodesics and $(\nabla^2 f)_p$ is positive definite, then the function attains its minimum at some $p_* \in D$.

Proof. We abbreviate $\lambda := \lambda_{f,\alpha}(p)$ and $r := \|u\|_{f,p,\alpha}$. For every $q = \mathrm{Exp}_p(u) \in D$, we have using Corollary 3.10 and the definition of the Newton decrement the lower bound

$$f(q) - f(p) \geq df_p(u) + \alpha\rho(-r) \geq -\alpha r\lambda + \alpha\rho(-r) = \alpha\delta(r), \tag{3.23}$$

where

$$\delta(r) = r(1 - \lambda) - \log(1 + r).$$

If $\lambda < 1$, $\delta(r)$ is minimized at $r = \lambda/(1 - \lambda)$, and we obtain

$$f(q) - f(p) \geq \alpha(\lambda + \log(1 - \lambda)) = -\alpha\rho(\lambda).$$

This implies Eq. (3.22).

On the other hand, $\delta(r) \to \infty$ as $r \to \infty$, so Eq. (3.23) shows that the level set $\{q \in D : f(q) \leq f(p)\}$ is contained in a Dikin ellipsoid of some suitable radius. If we assume that $(\nabla^2 f)_p$ is positive definite then Dikin ellipsoids are bounded. Thus if f is also α-strongly self-concordant along geodesics then Lemma 2.2 (iii) shows that f attains its minimum at some $p_* \in D$. $\qquad\square$

The minimizer in Proposition 3.14 is unique assuming strict convexity, as follows, e.g., if $\nabla^2 f$ is positive definite throughout the domain. The Newton decrement also certifies closeness to minimizers if they exist:

Lemma 3.15. *Let $f\colon D \to \mathbb{R}$ be α-self-concordant along geodesics, defined on an open convex subset $D \subseteq M$, and let $p \in D$ be such that $\lambda_{f,\alpha}(p) < 1$. If f attains a minimum at $p_* = \mathrm{Exp}_p(u)$ for $u \in T_pM$, then*

$$\|u\|_{f,p,\alpha} \leq \frac{\lambda_{f,\alpha}(p)}{1 - \lambda_{f,\alpha}(p)}.$$

Proof. Consider the geodesic $\gamma(t) = \mathrm{Exp}_p(tu)$ from p to p_*. Then by Corollary 3.10, we have

$$\frac{\alpha r^2}{1 + r} \leq df_{p_*}(\tau_{\gamma,1}u) - df_p(u) = -df_p(u) \leq |df_p(u)| \leq \alpha r\lambda_{f,\alpha}(p),$$

where $r := \|u\|_{f,p,*}$; the equality follows because $df_{p_*} = 0$ because p_* is a minimizer of f. Thus we have

$$\frac{r}{1+r} \leq \lambda_{f,\alpha}(p)$$

and for $\lambda_{f,\alpha}(p) < 1$ this implies the desired bound. $\qquad\square$

The following theorem is key to the analysis of Newton's method for self-concordant functions. It bounds the Newton decrement after one Newton step quadratically in terms of the original Newton decrement. This requires self-concordance in the sense of Definition 3.1, rather than the weaker notion along geodesics, as its proof involves comparing the length of the new Newton step transported along the geodesics given by the previous Newton step, i.e., there are two natural directions. The proof adapts the Euclidean argument in [Ren01, Thm. 2.2.4].

Theorem 3.16. *Let $f: D \to \mathbb{R}$ be a strongly α-self-concordant function defined on an open convex set $D \subseteq M$, with positive definite Hessian. Let $p \in D$ be a point such that $\lambda_{f,\alpha}(p) < 1$. Then the Newton iterate remains in the domain, i.e., $p_{f,+} \in D$, and moreover*

$$\lambda_{f,\alpha}(p_{f,+}) \leq \left(\frac{\lambda_{f,\alpha}(p)}{1 - \lambda_{f,\alpha}(p)} \right)^2.$$

Proof. We abbreviate the Newton step, iterate, and increment by $n_p := n_{f,p}$, $p_+ := p_{f,+}$, and $\lambda := \lambda_{f,\alpha}(p)$, respectively. Corollary 3.8 along with the definitions shows that $p_+ \in D$. Then the entire geodesic segment $\gamma(t) := \mathrm{Exp}_p(t n_p)$ for $t \in [0,1]$ is contained in the domain D. We now prove the desired estimate, starting with Theorem 3.4, which gives the upper bound

$$\lambda_{f,\alpha}(p_+) = \max_{w \in T_{p_+}M} \frac{|df_{p_+}(w)|}{\alpha\|w\|_{f,p_+,\alpha}} = \max_{v \in T_pM} \frac{|df_{p_+}(\tau_{\gamma,1}v)|}{\alpha\|\tau_{\gamma,1}v\|_{f,p_+,\alpha}} \leq \frac{1}{1-\lambda} \max_{v \in T_pM} \frac{|df_{p_+}(\tau_{\gamma,1}v)|}{\alpha\|v\|_{f,p,\alpha}}, \quad (3.24)$$

where $\tau_{\gamma,1}$ denotes parallel transport along the geodesic γ from p to p_+. Next, we observe that by the fundamental theorem of calculus, Eq. (2.3), and Eq. (3.19), for all $v \in T_pM$,

$$
\begin{aligned}
df_{p_+}(\tau_{\gamma,1}v) &= df_{p_+}(\tau_{\gamma,1}v) - df_p(v) + df_p(v) \\
&= \int_0^1 \partial_t df_{\gamma(t)}(\tau_{\gamma,t}v)\, dt + df_p(v) \\
&= \int_0^1 (\nabla_{\dot\gamma(t)} df)_{\gamma(t)}(\tau_{\gamma,t}v)\, dt + df_p(v) \\
&= \int_0^1 (\nabla^2 f)_{\gamma(t)}(\tau_{\gamma,t}n_p, \tau_{\gamma,t}v)\, dt + df_p(v) \\
&= \int_0^1 [(\nabla^2 f)_{\gamma(t)}(\tau_{\gamma,t}n_p, \tau_{\gamma,t}v) - (\nabla^2 f)_p(n_p, v)]\, dt \\
&= \beta(n_p, v), \quad (3.25)
\end{aligned}
$$

where we have introduced the symmetric bilinear form

$$\beta: T_pM \times T_pM \to \mathbb{R}, \quad \beta(u,v) = \int_0^1 [(\nabla^2 f)_{\gamma(t)}(\tau_{\gamma,t}u, \tau_{\gamma,t}v) - (\nabla^2 f)_p(u,v)]\, dt.$$

By Theorem 3.4 and using $\|t n_p\|_{f,p,\alpha} = t\lambda$, we have, for all $v \in T_pM$,

$$\left[(1-t\lambda)^2 - 1\right](\nabla^2 f)_p(v,v) \leq (\nabla^2 f)_{\gamma(t)}(\tau_{\gamma,t}v, \tau_{\gamma,t}v) - (\nabla^2 f)_p(v,v) \leq \left[\frac{1}{(1-t\lambda)^2} - 1\right](\nabla^2 f)_p(v,v).$$

By integrating the lower and upper bounds from $t = 0$ to $t = 1$,

$$-\left(\lambda - \frac{\lambda^2}{3}\right)(\nabla^2 f)_p(v,v) \leq \beta(v,v) \leq \left(\frac{\lambda}{1-\lambda}\right)(\nabla^2 f)_p(v,v).$$

One may verify that $\max\{\lambda - \lambda^2/3, \lambda/(1-\lambda)\} = \lambda/(1-\lambda)$ as $\lambda < 1$. Together with the Cauchy-Schwarz inequality, this implies that for all $u, v \in T_p M$,

$$|\beta(u,v)| \leq \frac{\lambda}{1-\lambda}\sqrt{(\nabla^2 f)_p(u,u)}\sqrt{(\nabla^2 f)_p(v,v)} = \frac{\alpha\lambda}{1-\lambda}\|u\|_{f,p,\alpha}\|v\|_{f,p,\alpha}.$$

Together with Eqs. (3.24) and (3.25), we obtain the upper bound

$$\lambda_{f,\alpha}(p_+) \leq \frac{1}{1-\lambda}\max_{v \in T_p M}\frac{|\beta(n_p, v)|}{\alpha\|v\|_{f,p,\alpha}} \leq \frac{\lambda}{(1-\lambda)^2}\|n_p\|_{f,p,\alpha} = \frac{\lambda^2}{(1-\lambda)^2}. \qquad \square$$

Theorem 3.16 implies that the Newton method converges quadratically for sufficiently small λ. For example, suppose that $\lambda \leq \lambda_* := 1 - \frac{1}{\sqrt{2}}$. Then we have

$$\left(\frac{\lambda}{1-\lambda}\right)^2 \leq \left(\frac{\lambda}{1-\lambda_*}\right)^2 = 2\lambda^2 \leq \lambda_*, \tag{3.26}$$

meaning the Newton decrement decreases quadratically and stays below λ_*, so we can iterate. This implies the following result (cf. [NN94, Thm. 2.2.3]):

Theorem 3.17 (Quadratic convergence of the Newton method). *Let $f\colon D \to \mathbb{R}$ be a strongly α-self-concordant function defined on an open convex set $D \subseteq M$, with positive definite Hessian. Let $p_0 \in D$ be a point such that $\lambda_{f,\alpha}(p_0) \leq \lambda_* := 1 - 1/\sqrt{2} \approx 0.293$. Then the Newton iterations*

$$p_{t+1} := \mathrm{Exp}_{p_t}(n_{f,p_t})$$

are well-defined for all $t \in \mathbb{N}$ (i.e., each $p_t \in D$) and we have

$$\lambda_{f,\alpha}(p_t) \leq \frac{1}{2}(2\lambda_{f,\alpha}(p_0))^{2^t} \leq \frac{1}{2}(2\lambda_*)^{2^t}.$$

In particular, $\mathcal{O}(\log\log\frac{\alpha}{\varepsilon})$ Newton iterations suffice to find a point p_t such that $f(p_t) \leq f_* + \varepsilon$, for $\varepsilon < \alpha/e$.

Proof. We abbreviate $\lambda_t := \lambda_{f,\alpha}(p_t)$. By Theorem 3.16 and Eq. (3.26), one can see inductively that $p_t \in D$ is well-defined for all $t \in \mathbb{N}$ and that we have $\lambda_t \leq \lambda_*$ and

$$2\lambda_t \leq (2\lambda_{t-1})^2 \leq \ldots \leq (2\lambda_0)^{2^t} \leq (2\lambda_*)^{2^t},$$

as claimed. This also implies the last statement, since to achieve $f(p_t) \leq f_* + \varepsilon$ it suffices to have $\rho(\lambda_t) \leq \varepsilon/\alpha$, by Proposition 3.14, and we have $\rho(\lambda_t) \leq \lambda_t^2$ for $\lambda_t \leq \lambda_*$. $\qquad \square$

What if we have a starting point such that the Newton decrement does *not* guarantee quadratic convergence? In this case it is well-known that one can employ a *damped* Newton method, with a step size that ensures that one stays inside the Dikin ellipsoid (and hence in the domain) at each step. This works just the same in the Riemannian setting and only requires self-concordance along geodesics (cf. [Nes18, Thm. 5.1.15]):

Theorem 3.18 (Damped Newton method)**.** *Let $f\colon D \to \mathbb{R}$ be strongly α-self-concordant along geodesics, defined on an open convex set $D \subseteq M$, with positive definite Hessian. Let $p_0 \in D$ be an arbitrary starting point. Then the damped Newton iterations*

$$p_{t+1} := \mathrm{Exp}_{p_t}(u_t) \quad \text{where} \quad u_t := \frac{1}{1 + \lambda_{f,\alpha}(p_t)} n_{f,p_t}$$

are well-defined for all $t \in \mathbb{N}$ (i.e., each $p_t \in D$) and we have

$$f(p_{t+1}) \leq f(p_t) - \alpha\rho(-\lambda_t),$$

where ρ is the quantity defined in [Eq. (3.17)](). In particular, if f is bounded from below and we set $f_* := \inf_{p \in D} f(p)$, then $\mathcal{O}((f(p_0) - f_*)/\alpha)$ damped Newton iterations suffice to find a point p_t such that $\lambda_{f,\alpha}(p_t) \leq \lambda_*$ (or any other constant).

Proof. We abbreviate $\lambda_t := \lambda_{f,\alpha}(p_t)$. Using [Corollary 3.11]() one can see inductively that $r := \|u_t\|_{f,p,\alpha} = \lambda_t/(1 + \lambda_t) < 1$ and $p_t \in D$ is well-defined for all $t \in \mathbb{N}$. Moreover,

$$\begin{aligned}
f(p_{t+1}) &\leq f(p_t) + df_{p_t}(u_t) + \alpha\rho(r) \\
&= f(p_t) - (\nabla^2 f)_p(n_{f,p_t}, u_t) + \alpha\rho(r) \\
&= f(p_t) - \alpha\left(\frac{\lambda_t^2}{1 + \lambda_t} - \rho(r)\right) \\
&= f(p_t) - \alpha(\lambda_t - \log(1 + \lambda_t)) \\
&= f(p_t) - \alpha\rho(-\lambda_t). \qquad \square
\end{aligned}$$

In particular, [Theorem 3.18]() and [Corollary 3.8]() have the following structural consequence.

Corollary 3.19. *Let $f\colon D \to \mathbb{R}$ be strongly α-self-concordant along geodesics, defined on an open convex set $D \subseteq M$, with positive definite Hessian. Then f is bounded from below if and only if it attains its minimum (necessarily at a unique minimizer, by strict convexity).*

By combining [Theorems 3.17]() and [3.18](), we see that we can approximately minimize any strongly α-self-concordant function with positive definite Hessian by first using damped Newton steps from an arbitrary starting point p_0 until we arrive at point with Newton decrement $\leq \lambda_*$; then we are in the quadratic convergence regime and we can take ordinary Newton steps until we arrive at a point p_t with $\rho(\lambda_{f,\alpha}(p_t)) \leq \varepsilon/\alpha$, so that p_t is an ε-approximate minimizer. This requires $\mathcal{O}((f(p_0) - f_*)/\alpha + \log\log(\alpha/\varepsilon))$ Newton iterations.

4 Barriers, compatibility, path-following method on manifolds

The methods developed in [Section 3]() are sufficient to optimize strongly self-concordant functions. However, it is difficult to guarantee that one starts in the quadratic convergence regime for Newton's method, and the damped Newton method has a worst-case complexity which depends on the gap in function value. Moreover, most convex optimization problems do *not* take the form of a minimization of a strongly self-concordant function over its natural domain. Rather, one is given a convex objective f and a domain D and wants to minimize the former over the latter.

In this section, we show how to circumvent these two issues, assuming one has a *self-concordant barrier* for the domain over which one optimizes. To this end, we generalize the analysis of so-called *path-following (interior point) methods* [NN94] from the Euclidean to the Riemannian setting. We

treat not only the case of geodesically linear objectives, but the more general class of objectives that are *compatible* with the given self-concordant barrier. This will be useful for the applications discussed in Section 6. Throughout this section we assume that M is a connected and geodesically complete Riemannian manifold.

4.1 Self-concordant barriers

We first define the notion of a self-concordant barrier. The estimates in this section only require the self-concordance to be along geodesics, and we make explicit whenever this is the case. However, the path-following method presented in Section 4.3 requires the stronger notion.

Definition 4.1 (Barrier). *Let $D \subseteq M$ be an open and convex subset, and let $\theta \geq 0$. We say that a function $F: D \to \mathbb{R}$ is a non-degenerate strongly self-concordant barrier with parameter θ, or in short a θ-barrier, if F is a strongly 1-self-concordant function with positive definite Hessian such that $\lambda_F(p) \leq \sqrt{\theta}$ for all $p \in D$, with $\lambda_F = \lambda_{F,1}$ the Newton decrement (Definition 3.13). We say that F is a θ-barrier along geodesics if it is only strongly 1-self-concordant along geodesics.*

The parameter of a barrier plays an important role in the complexity analysis of the path-following method that we discuss in Section 4.3. The following lemma follows readily from the definition:

Lemma 4.2. *Let $F_1: D_1 \to \mathbb{R}$ be a θ_1-barrier and let $F_2: D_2 \to \mathbb{R}$ be a θ_2-barrier. Then $F_1 + F_2$ is a $(\theta_1 + \theta_2)$-barrier for $D := D_1 \cap D_2$, assuming D is non-empty.*

Next, we prove an important inequality which involves the barrier parameter. To state the result, we define a Riemannian version of the so-called Minkowski function(al) or gauge function. It measures the inverse distance from a point to the boundary of the domain.

Definition 4.3 (Minkowski functional). *Let $D \subseteq M$ be an open convex subset. For $p \in D$, we define the Minkowski functional by*

$$\pi_{D,p}: T_pM \to \mathbb{R}_{\geq 0}, \quad \pi_{D,p}(u) = \inf\{s \geq 0 : \mathrm{Exp}_p(\tfrac{1}{s}u) \in D\}.$$

This is well-defined since D is open and hence $\pi_{D,p}(u) < \infty$ for every $u \in T_pM$. Note that if $s := \pi_{D,p}(u) = 0$, then the entire infinite geodesic ray $\gamma(t) = \mathrm{Exp}_p(tu)$ is contained in the domain, while if $s > 0$ then $\mathrm{Exp}_p(\tfrac{1}{s}u)$ is a point in its boundary $\partial D = \overline{D} \setminus D$. Moreover, if $u \in T_pM$ is such that $\mathrm{Exp}_p(u) \in \overline{D}$, then $\pi_p(u) \leq 1$.

Then we have the following result, which can be deduced directly from its Euclidean version [NN94, §2.3.2]. We provide a self-contained proof for convenience.

Proposition 4.4. *Let $D \subseteq M$ be open and convex, and let $F: D \to \mathbb{R}$ be a θ-barrier along geodesics. Then one has, for all $p \in D$ and $u \in T_pM$,*

$$dF_p(u) \leq \theta \, \pi_{D,p}(u).$$

In particular, if $q = \mathrm{Exp}_p(u) \in \overline{D}$ then

$$dF_p(u) \leq \theta.$$

Proof. The second statement follows from the first by the preceding discussion. To prove the first, let $p \in D$ and $u \in T_pM$. If $dF_p(u) \leq 0$ then there is nothing to prove, so we assume that $dF_p(u) > 0$. Define

$$g(t) := F(\mathrm{Exp}_p(tu)).$$

31

Then g is well-defined on the interval $I = [0, \pi_{D,p}(u)^{-1})$, where we interpret $0^{-1} = \infty$. By definition of the Newton decrement and recalling that $\ddot{g}(t) > 0$ as F has positive definite Hessian, we have

$$\frac{(\dot{g}(t))^2}{\ddot{g}(t)} \leq \lambda_F^2(p) = \theta.$$

Since we assumed that $\dot{g}(0) = dF_p(u) > 0$, we find that $\theta > 0$, as well as $\dot{g}(t) > 0$ for all $t \in I$, by convexity. Accordingly, we can write the above as

$$\partial_t\left(\frac{1}{\dot{g}(t)}\right) = -\frac{\ddot{g}(t)}{(\dot{g}(t))^2} \leq -\frac{1}{\theta},$$

which implies that

$$\frac{1}{\dot{g}(t)} = \frac{1}{\dot{g}(0)} + \int_0^t \partial_t\left(\frac{1}{\dot{g}(t)}\right) \leq \frac{1}{\dot{g}(0)} - \frac{t}{\theta},$$

and hence

$$\dot{g}(t) \geq \frac{1}{\frac{1}{\dot{g}(0)} - \frac{t}{\theta}} = \frac{\theta\dot{g}(0)}{\theta - t\dot{g}(0)}.$$

As the right-hand side diverges as t approaches $\theta/\dot{g}(0)$, we must have $t < \theta/\dot{g}(0)$ for all $t \in I$. Hence

$$\pi_{D,p}(u)^{-1} \leq \frac{\theta}{\dot{g}(0)},$$

which is the desired bound. $\qquad\square$

As a consequence, non-trivial barriers must have positive parameter:

Corollary 4.5. *Let $D \subseteq M$ be open and convex, and let $F\colon D \to \mathbb{R}$ be a θ-barrier along geodesics with $\theta = 0$. Then F is constant and $D = M$.*

Proof. Proposition 4.4 shows that $dF = 0$, hence F is locally constant and $\nabla^2 F = 0$. Because F is strongly self-concordant, we may apply Corollary 3.8 to conclude that $\mathrm{Exp}_p(T_pM) \subseteq D$ and hence $D = M$, since M is connected and geodesically complete. $\qquad\square$

The minimizer of a barrier, which if it exists is necessarily unique (recall that barriers have positive definite Hessians by definition), plays a special role in the theory.

Definition 4.6 (Analytic center). *Let $D \subseteq M$ be open and convex, and let $F\colon D \to \mathbb{R}$ be a θ-barrier along geodesics. If F attains its minimum, then the unique minimizer is called the* analytic center *of D.*

Recall that a barrier attains its minimum if and only if it is bounded from below (Corollary 3.19). The following result shows that the domain is necessarily enclosed in a Dikin ellipsoid about the analytic center, with radius given by the barrier's parameter. It adapts the Euclidean argument (cf. [Nes18, Thm. 5.3.9], [NN94, Prop. 2.3.2 (iii)]) to the Riemannian setting.

Proposition 4.7 (Enclosing Dikin ellipsoid). *Let $D \subseteq M$ be open and convex, and let $F\colon D \to \mathbb{R}$ be a θ-barrier along geodesics. If $\theta > 0$ and F is bounded from below, with analytic center $p_* \in D$, then*

$$D \subseteq B_{F,p_*}^{\circ}(2\theta + 1),$$

where $B_{F,p_*}^{\circ} = B_{F,p_*,1}^{\circ}$ denotes the Dikin ellipsoid (Definition 3.7). That is, the domain is contained in the Dikin ellipsoid with radius $2\theta + 1$ about p_*.

Proof. Let $u \in T_{p_*}M$ be such that $\|u\|_{F,p_*} = 1$, and let $\gamma(t) := \mathrm{Exp}_{p_*}(tu)$. By Corollary 3.8, we know that $B^\circ_{F,p_*}(1) \subseteq D$, hence $g(t) := F(\gamma(t))$ is well-defined for $t \in [0,1)$.

To show that $D \subseteq B^\circ_{F,p_*}(2\theta + 1)$, by convexity of D it suffices to show that $\gamma(1 + 2\theta) \notin D$. From Eq. (3.18) in Corollary 3.10 and p_* being a minimizer of F, it follows that, for $t \in [0,1)$,

$$dF_{\gamma(t)}(\tau_{\gamma,t}u) = dF_{\gamma(t)}(\tau_{\gamma,t}u) - dF_{p_*}(u) = \dot{g}(t) - \dot{g}(0) \geq \frac{t}{1+t}.$$

Proposition 4.4 on the other hand implies that for

$$dF_{\gamma(t)}(\tau_{\gamma,t}u) \leq \theta\,\pi_{D,\gamma(t)}(\tau_{\gamma,t}u).$$

Together, we obtain that, for every $t \in [0,1)$,

$$\theta\,\pi_{D,\gamma(t)}(\tau_{\gamma,t}u) \geq \frac{t}{1+t}.$$

Now, by the definition of the Minkowski functional, for every $s \in [0, \pi_{D,\gamma(t)}(\tau_{\gamma,t}(u)))$, we have

$$\gamma\left(t + \tfrac{1}{s}\right) = \mathrm{Exp}_{p_*}\left(\left(t + \tfrac{1}{s}\right)u\right) = \mathrm{Exp}_{\gamma(t)}\left(\tfrac{1}{s}\tau_{\gamma,t}u\right) \notin D.$$

Therefore, for every $t \in [0,1)$ and $s \in [0, \frac{t}{\theta(1+t)})$, we have

$$\gamma\left(t + \tfrac{1}{s}\right) \notin D.$$

Letting $t \to 1$ and $s \to 1/(2\theta)$ gives that $\gamma(1 + 2\theta) \notin D$, since $M \setminus D$ is closed. $\qquad\square$

4.2 Compatibility

Given a barrier F, for which convex functions f is it the case that $tf + F$ is self-concordant for all $t \geq 0$, with parameter independent of t? This is clearly the case if f is (affine) linear or quadratic in the sense that the third covariant derivative $\nabla^3 f$ vanishes. We now define the more general notion of *compatibility*, which suffices for this, as shown in Proposition 4.10 below.

Definition 4.8 (Compatibility). *Let $D \subseteq M$ be open and convex, let $f, F\colon D \to \mathbb{R}$ be convex functions. For $\beta_1, \beta_2 \geq 0$, we say that f is (β_1, β_2)-compatible with F if for all $p \in D$ and $u, v \in T_pM$, one has*

$$\begin{aligned}
|(\nabla^3 f)_p(u,v,v)| \leq\ & 2\beta_1\sqrt{(\nabla^2 F)_p(u,u)(\nabla^2 f)_p(v,v)} \\
& + 2\beta_2\sqrt{(\nabla^2 F)_p(v,v)}\sqrt{(\nabla^2 f)_p(u,u)}\sqrt{(\nabla^2 f)_p(v,v)}.
\end{aligned} \tag{4.1}$$

For $\beta \geq 0$, we say that f is β-compatible with F along geodesics if for all $p \in D$ and $v \in T_pM$,

$$|(\nabla^3 f)_p(v,v,v)| \leq 2\beta\sqrt{(\nabla^2 F)_p(v,v)(\nabla^2 f)_p(v,v)}. \tag{4.2}$$

Clearly, if f is a linear or a convex quadratic function, in the sense that its second or third covariant derivative vanishes, then it is clearly automatically compatible with any convex F. Moreover, any α-self-concordant function is (β_1, β_2)-compatible with itself, for $\beta_1 + \beta_2 = 1/\sqrt{\alpha}$. As we show in Proposition 4.10, given a barrier F for a domain D and a convex objective function f, compatibility guarantees that $tf + F$ is self-concordant for all $t \geq 0$, with a parameter independent of t, and hence one can use the path-following method presented in Section 4.3 below to optimize f over D. We apply this theory in Section 6.

Compatibility along geodesics reduces to the well-known Euclidean notion, see [NN94, Def. 3.2.1] or [Nes18, Def. 5.4.2]. In these works it is also explained how to generalize the notion of compatibility to vector-valued functions f, which is useful for constructing new barriers out of old ones; see [NN94, §5.1.2] or [Nes18, §5.4.6] for details. We do not provide such a generalization here. Clearly, if f is (β_1, β_2)-compatible with F then it is also β-compatible with F along geodesics for $\beta := \beta_1 + \beta_2$. Yet the latter does not imply the former, even in the Euclidean setting.

We may equivalently write Eqs. (4.1) and (4.2) as follows in terms of the seminorms $\|\cdot\|_{g,p} = \|\cdot\|_{g,p,1}$ induced by the inner products $\langle \cdot, \cdot \rangle_{g,p} = \langle \cdot, \cdot \rangle_{g,p,1}$ defined in Eq. (3.2):

$$|(\nabla^3 f)_p(u, v, v)| \leq 2\beta_1 \|u\|_{F,p} \|v\|_{f,p}^2 + 2\beta_2 \|v\|_{F,p} \|u\|_{f,p} \|v\|_{f,p} \tag{4.3}$$

and

$$|(\nabla^3 f)_p(v, v, v)| \leq 2\beta \|v\|_{F,p} \|v\|_{f,p}^2. \tag{4.4}$$

We now state some basic properties of compatibility. The following result holds analogously for compatibility along geodesics.

Lemma 4.9. *Let $D \subseteq M$ be open and convex, $F \colon D \to \mathbb{R}$ a convex function, and $\beta \in \mathbb{R}_{\geq 0}^2$.*

    (i) *Let $f \colon D \to \mathbb{R}$ be a convex function that is β-compatible with F and let $c \geq 0$. Then cf is β-compatible with F.*

    (ii) *Let $f_1, f_2 \colon D \to \mathbb{R}$ be two convex functions that are each β-compatible with F. Then their sum $f_1 + f_2$ is β-compatible with F.*

Proof. Property (i) is clear from the definition, as both sides of Eq. (4.1) are positively homogeneous in f. To prove property (ii), we note that for every $p \in D$ and $u, v \in T_p M$,

$$|(\nabla^3 (f_1 + f_2))_p(u, v, v)| \leq |(\nabla^3 f_1)_p(u, v, v)| + |(\nabla^3 f_2)_p(u, v, v)|$$
$$\leq 2\beta_1 \sqrt{(\nabla^2 F)_p(u, u)(\nabla^2 f_1)_p(v, v)} + 2\beta_1 \sqrt{(\nabla^2 F)_p(u, u)(\nabla^2 f_2)_p(v, v)}$$
$$+ 2\beta_2 \sqrt{(\nabla^2 F)_p(v, v)} \left(\sqrt{(\nabla^2 f_1)_p(u, u)} \sqrt{(\nabla^2 f_1)_p(v, v)} + \sqrt{(\nabla^2 f_2)_p(u, u)} \sqrt{(\nabla^2 f_2)_p(v, v)} \right)$$
$$\leq 2\beta_1 \sqrt{(\nabla^2 F)_p(u, u)(\nabla^2 f_1)_p(v, v)} + 2\beta_1 \sqrt{(\nabla^2 F)_p(u, u)(\nabla^2 f_2)_p(v, v)}$$
$$+ 2\beta_2 \sqrt{(\nabla^2 F)_p(v, v)} \sqrt{(\nabla^2 f_1)_p(u, u) + (\nabla^2 f_2)_p(u, u)} \sqrt{(\nabla^2 f_1)_p(v, v) + (\nabla^2 f_2)_p(v, v)}$$
$$= 2\beta_1 \sqrt{(\nabla^2 F)_p(u, u)(\nabla^2 (f_1 + f_2))_p(v, v)}$$
$$+ 2\beta_2 \sqrt{(\nabla^2 F)_p(v, v)} \sqrt{(\nabla^2 (f_1 + f_2))_p(u, u)} \sqrt{(\nabla^2 (f_1 + f_2))_p(v, v)}.$$

The first inequality holds by compatibility of f_1 and of f_2 with F, and the second inequality uses the Cauchy-Schwarz inequality. $\qquad\square$

We now show if a convex function f is compatible with a self-concordant function F (e.g., a barrier), then $tf + F$ is self-concordant for every $t \geq 0$, with a self-concordance constant that is independent of t. We emphasize that it is not necessary for f itself to be self-concordant. The proof is inspired by [NN94, Prop. 3.2.2] in the Euclidean setting. The result holds analogously if we use compatibility and self-concordance along geodesics in the hypothesis and conclusion.

Proposition 4.10. *Let $D \subseteq M$ be open and convex and let $f, F \colon D \to \mathbb{R}$ be convex functions. Suppose that f is (β_1, β_2)-compatible with F and F is 1-self-concordant. Then $tf + F \colon D \to \mathbb{R}$ is α-self-concordant for every $t \geq 0$, with*

$$
\alpha := \begin{cases}
\frac{4\left(\beta_2^2 - (\beta_1 - 1)^2\right)}{\beta_2^2(\beta_2^2 + 4\beta_1)} & \text{if } \beta_2^2 > 2\max\{\beta_1(\beta_1 - 1), 1 - \beta_1\}, \\
\frac{1}{\max\{\beta_1^2, 1\}} & \text{otherwise.}
\end{cases}
$$

If in addition F is strongly 1-self-concordant and f has a closed convex extension, then $tf + F \colon D \to \mathbb{R}$ is strongly α-self-concordant for every $t \geq 0$.

Proof. We abbreviate $F_t := tf + F$. Clearly, F_t is convex for every $t \geq 0$, so it remains to prove the self-concordance estimate. For any $p \in D$ and $u, v \in T_p M$, using Eqs. (3.6) and (4.3),

$$
\begin{aligned}
|(\nabla^3 F_t)_p(u, v, v)| &\leq t|(\nabla^3 f)_p(u, v, v)| + |(\nabla^3 F)_p(u, v, v)| \\
&\leq 2t\beta_1 \|u\|_{F,p}\|v\|_{f,p}^2 + 2t\beta_2\|v\|_{F,p}\|u\|_{f,p}\|v\|_{f,p} + 2\|u\|_{F,p}\|v\|_{F,p}^2 \\
&= 2\left(\sqrt{t}\|u\|_{f,p}\left(\sqrt{t}\beta_2\|v\|_{F,p}\|v\|_{f,p}\right) + \|u\|_{F,p}\left(t\beta_1\|v\|_{f,p}^2 + \|v\|_{F,p}^2\right)\right) \\
&\leq 2\sqrt{t\|u\|_{f,p}^2 + \|u\|_{F,p}^2}\sqrt{\left(\sqrt{t}\beta_2\|v\|_{F,p}\|v\|_{f,p}\right)^2 + \left(t\beta_1\|v\|_{f,p}^2 + \|v\|_{F,p}^2\right)^2} \\
&= 2\|u\|_{F_t,p}\sqrt{t\beta_2^2\|v\|_{F,p}^2\|v\|_{f,p}^2 + \left(t\beta_1\|v\|_{f,p}^2 + \|v\|_{F,p}^2\right)^2}
\end{aligned}
$$

using the Cauchy-Schwarz inequality in the second-to-last step. To show that F_t is α-self-concordant, by Eq. (3.5) it therefore suffices to show that (note we use $\|\cdot\|_{g,p,1}$ rather than $\|\cdot\|_{g,p,\alpha}$!)

$$
\sqrt{t\beta_2^2\|v\|_{F,p}^2\|v\|_{f,p}^2 + \left(t\beta_1\|v\|_{f,p}^2 + \|v\|_{F,p}^2\right)^2} \leq \frac{1}{\sqrt{\alpha}}\|v\|_{F_t,p}^2. \tag{4.5}
$$

Without loss of generality, we can assume that $\|v\|_{F_t,p}^2 = 1$. Writing $x := \|v\|_{f,p}^2$ and $y := \|v\|_{F,p}^2$, we see that Eq. (4.5) holds provided we can prove that

$$
\beta_2^2 txy + (\beta_1 tx + y)^2 \leq \frac{1}{\alpha} \tag{4.6}
$$

for all $x, y \geq 0$ subject to the constraint $tx + y = 1$. Eliminating t and x using this constraint, the left-hand side can be written as

$$
\begin{aligned}
q(y) &:= \beta_2^2(1 - y)y + (\beta_1(1 - y) + y)^2 \\
&= \left((1 - \beta_1)^2 - \beta_2^2\right)y^2 + \left(2\beta_1 - 2\beta_1^2 + \beta_2^2\right)y + \beta_1^2,
\end{aligned}
$$

so we wish to show that $q(y) \leq 1/\alpha$ for all $y \in [0, 1]$. Note that $q(y)$ is a quadratic polynomial. We distinguish two cases:

If $(1 - \beta_1)^2 < \beta_2^2$, then q is strictly concave and attains its maximum on $\mathbb{R}$ at

$$
y_* = \frac{2\beta_1 - 2\beta_1^2 + \beta_2^2}{2\left(\beta_2^2 - (1 - \beta_1)^2\right)}.
$$

Note that $y_* \in (0, 1)$ if and only if

$$
0 < 2\beta_1 - 2\beta_1^2 + \beta_2^2 < 2\left(\beta_2^2 - (1 - \beta_1)^2\right),
$$

35

which is equivalent to

$$\beta_2^2 > 2 \max\{\beta_1(\beta_1 - 1), 1 - \beta_1\}.$$

If $y_* \in (0, 1)$, then the maximum of $q(y)$ on $[0, 1]$ is given by

$$q(y_*) = \frac{(2\beta_1 - 2\beta_1^2 + \beta_2^2)^2}{4(\beta_2^2 - (1 - \beta_1)^2)} + \beta_1^2 = \frac{\beta_2^4 + 4\beta_1\beta_2^2}{4(\beta_2^2 - (1 - \beta_1)^2)},$$

while otherwise it is attained at the boundary, where $q(0) = \beta_1^2$ and $q(1) = 1$.

If $(1 - \beta_1)^2 \geq \beta_2^2$, then $q(y)$ is convex and hence attains its maximum always at the boundary. Summarizing both cases, we find that

$$\max_{y \in [0,1]} q(y) = \begin{cases} \frac{\beta_2^4 + 4\beta_1\beta_2^2}{4(\beta_2^2 - (1 - \beta_1)^2)} & \text{if } (\beta_1 - 1)^2 < \beta_2^2 \text{ and } \beta_2^2 > 2\max\{\beta_1(\beta_1 - 1), 1 - \beta_1\}, \\ \max\{\beta_1^2, 1\} & \text{otherwise.} \end{cases}$$

The condition of the first case is equivalent to

$$\beta_2^2 > 2 \max\{\beta_1(\beta_1 - 1), 1 - \beta_1\},$$

and hence we have confirmed Eq. (4.6). Thus we have proved that $F_t = tf + F$ is an α-self-concordant function on D. Finally, the last claim follows from Lemma 2.3 $\qquad\square$

Finally, we construct a self-concordant barrier for the epigraph of any function compatible with a barrier for its domain. This result generalizes the Euclidean result [Nes18, Thm. 5.3.5], which constructs a self-concordant barrier for the open epigraph

$$E_f^\circ := \{(p, t) \in D \times \mathbb{R} : f(p) < t\} \tag{4.7}$$

of a self-concordant barrier. As before, it holds analogously if we use the notions along geodesics in the hypothesis and conclusion.

Theorem 4.11 (Barriers for epigraphs). *Let $D \subseteq M$ be open and convex and let $f, F \colon D \to \mathbb{R}$ be convex functions. Suppose that f is (β_1, β_2)-compatible with F and F is 1-self-concordant. Then, the function*

$$G \colon E_f^\circ \to \mathbb{R}, \quad G(p, t) = -\log(t - f(p)) + F(p)$$

defined on the open epigraph E_f°, see Eq. (4.7), is convex and α-self-concordant, with

$$\alpha := \frac{1}{\max\{1 + \beta_1^2, \beta_1 + \frac{1}{2}\beta_2^2, \frac{2}{3}\beta_2^2\}}. \tag{4.8}$$

Furthermore, for every $(p, t) \in E_f^\circ$ one has

$$\lambda_{G,\alpha}(p, t)^2 = \frac{\lambda_G(p, t)^2}{\alpha} \leq \frac{1 + \lambda_F(p)^2}{\alpha}. \tag{4.9}$$

If in addition F is strongly 1-self-concordant and f has a closed convex extension, then G is strongly α-self-concordant. In particular, if F is a θ-barrier for D and f has a closed convex extension, then G/α is a $(1 + \theta)/\alpha$-barrier for E_f°.

Proof. We identify $v \in T_{(p,t)}E_f^\circ \cong T_pD \oplus \mathbb{R}$ and write $v = (v_p, v_t)$, with $v_p \in T_pD$ and $v_t \in \mathbb{R}$. Then the differential of G is given by

$$dG_{(p,t)}(v) = -\frac{1}{t - f(p)}(v_t - df_p(v_p)) + dF_p(v_p) \tag{4.10}$$

and the Hessian of G by

$$(\nabla^2 G)_{(p,t)}(v, v) = \underbrace{\frac{1}{(t - f(p))^2}(v_t - df_p(v_p))^2}_{=:A_v^2} + \underbrace{\frac{1}{t - f(p)}(\nabla^2 f)_p(v_p, v_p)}_{=:B_v^2} + \underbrace{(\nabla^2 F)_p(v_p, v_p)}_{=:C_v^2}. \tag{4.11}$$

The underbraced terms are all non-negative as $t > f(p)$ and both f and F are convex, hence we can write them as squares of real numbers A_v, B_v, C_v. This also shows that G is convex. We now prove that G is self-concordant. The third covariant derivative can be computed as follows: for all $u, v \in T_{(p,t)}E_f^\circ$, we have

$$
\begin{aligned}
(\nabla^3 G)_{(p,t)}(u, v, v) &= -\frac{2}{(t - f(p))^3}(u_t - df_p(u_p))(v_t - df_p(v_p))^2 \\
&\quad - \frac{2}{(t - f(p))^2}(v_t - df_p(v_p))(\nabla^2 f)_p(u_p, v_p) \\
&\quad - \frac{1}{(t - f(p))^2}(u_t - df_p(u_p))(\nabla^2 f)_p(v_p, v_p) \\
&\quad + \frac{1}{t - f(p)}(\nabla^3 f)_p(u_p, v_p, v_p) + (\nabla^3 F)_p(u_p, v_p, v_p) \\
&= -2A_u A_v^2 - 2A_v \frac{1}{t - f(p)}(\nabla^2 f)_p(u_p, v_p) - A_u B_v^2 \\
&\quad + \frac{1}{t - f(p)}(\nabla^3 f)_p(u_p, v_p, v_p) + (\nabla^3 F)_p(u_p, v_p, v_p).
\end{aligned}
$$

Now, we have

$$\frac{1}{t - f(p)}(\nabla^2 f)_p(u_p, v_p) \le B_u B_v$$

by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
\frac{1}{t - f(p)}\left|(\nabla^3 f)_p(u_p, v_p, v_p)\right| &\le \frac{2}{t - f(p)}\left(\beta_1 \|u\|_{F,p}\|v\|_{f,p}^2 + \beta_2 \|v\|_{F,p}\|u\|_{f,p}\|v\|_{f,p}\right) \\
&= 2\left(\beta_1 B_v^2 C_u + \beta_2 B_u B_v C_v\right)
\end{aligned}
$$

by compatibility of f with F as in Eq. (4.3), and finally

$$\left|(\nabla^3 F)_p(u_p, v_p, v_p)\right| \le 2C_u C_v^2$$

by 1-self-concordance of F (Eq. (3.5)). Combining these estimates, we can upper bound the third covariant derivative of G in absolute value as

$$
\begin{aligned}
\left|(\nabla^3 G)_{(p,t)}(u, v, v)\right| &\le 2A_u A_v^2 + 2A_v B_u B_v + A_u B_v^2 + 2\left(\beta_1 B_v^2 C_u + \beta_2 B_u B_v C_v\right) + 2C_u C_v^2 \\
&= A_u(2A_v^2 + B_v^2) + B_u(2A_v B_v + 2\beta_2 B_v C_v) + C_u(2\beta_1 B_v^2 + 2C_v^2)
\end{aligned}
$$

37

$$\leq \sqrt{A_u^2 + B_u^2 + C_u^2}\sqrt{(2A_v^2 + B_v^2)^2 + (2A_vB_v + 2\beta_2 B_vC_v)^2 + (2\beta_1 B_v^2 + 2C_v^2)^2}$$

$$\leq 2\sqrt{(\nabla^2 G)_{(p,t)}(u,u)}\sqrt{\max\{1+\beta_1^2, \beta_1 + \tfrac{1}{2}\beta_2^2, \tfrac{2}{3}\beta_2^2\}}(\nabla^2 G)_{(p,t)}(v,v)$$

$$= \frac{2}{\sqrt{\alpha}}\sqrt{(\nabla^2 G)_{(p,t)}(u,u)}\,(\nabla^2 G)_{(p,t)}(v,v),$$

where the last inequality holds due to $2xy \leq x^2 + y^2$, as in

$$\frac{1}{4}\big[(2A_v^2 + B_v^2)^2 + (2A_vB_v + 2\beta_2 B_vC_v)^2 + (2\beta_1 B_v^2 + 2C_v^2)^2\big]$$
$$= A_v^4 + \big(\tfrac{1}{4} + \beta_1^2\big)B_v^4 + C_v^4 + 2A_v^2B_v^2 + 2\big(\beta_1 + \tfrac{1}{2}\beta_2^2\big)B_v^2C_v^2 + 2\beta_2 A_vB_v^2C_v$$
$$= A_v^4 + \big(\tfrac{1}{4} + \beta_1^2\big)B_v^4 + C_v^4 + 2A_v^2B_v^2 + 2\big(\beta_1 + \tfrac{1}{2}\beta_2^2\big)B_v^2C_v^2 + 2\big(\tfrac{\sqrt{3}}{2}B_v^2\big)\big(\tfrac{2}{\sqrt{3}}\beta_2 A_vC_v\big)$$
$$\leq A_v^4 + \big(\tfrac{1}{4} + \beta_1^2\big)B_v^4 + C_v^4 + 2A_v^2B_v^2 + 2\big(\beta_1 + \tfrac{1}{2}\beta_2^2\big)B_v^2C_v^2 + \tfrac{3}{4}B_v^4 + \tfrac{4}{3}\beta_2^2 A_v^2C_v^2$$
$$= A_v^4 + \big(1 + \beta_1^2\big)B_v^4 + C_v^4 + 2A_v^2B_v^2 + 2\big(\beta_1 + \tfrac{1}{2}\beta_2^2\big)B_v^2C_v^2 + 2\tfrac{2}{3}\beta_2^2 A_v^2C_v^2$$
$$\leq \max\{1+\beta_1^2, \beta_1 + \tfrac{1}{2}\beta_2^2, \tfrac{2}{3}\beta_2^2\}\big(A_v^2 + B_v^2 + C_v^2\big)^2.$$

We conclude that G is indeed α-self-concordant with α as in Eq. (4.8).

Next, we prove the bound on the differential. Using Eq. (4.10) and with A_v, B_v as in Eq. (4.11), we have

$$|dG_{(p,t)}(v)| \leq A_v + |dF_p(v_p)| \leq A_v + \lambda_F(p)C_v$$
$$\leq \sqrt{1 + \lambda_F(p)^2}\sqrt{A_v^2 + C_v^2} \leq \sqrt{1 + \lambda_F(p)^2}\sqrt{(\nabla^2 G)_{(p,t)}(v,v)},$$

by definition of the Newton decrement and the Cauchy-Schwarz inequality. Thus we find that

$$\lambda_{G,\alpha}(p,t) \leq \sqrt{\frac{1 + \lambda_F(p)^2}{\alpha}}.$$

which establishes Eq. (4.9).

Finally, if F is strongly 1-self-concordant, hence closed convex on D, and if f has a closed convex extension then it is easy to see that G is closed convex on E_f°, using that $(s,t) \mapsto -\log(t-s)$ is closed convex on $\{(s,t) \in \mathbb{R}^2 : s < t\}$. $\qquad\square$

In particular, we can apply this construction to any self-concordant function:

Corollary 4.12. *Let $D \subseteq M$ be open and convex and let $f\colon D \to \mathbb{R}$ be 1-self-concordant. Then $g(p,t) = -\log(t - f(p)) + f(p)$ is a convex and 1-self-concordant function on the open epigraph E_f° of f, see Eq. (4.7). It satisfies $\lambda_g(p,t) \leq \sqrt{1 + \lambda_f(p)^2}$ for all $(p,t) \in E_f^\circ$. If f is strongly self-concordant, so is g. In particular, if f is a θ-barrier, g is a $(1+\theta)$-barrier for E_f°.*

To end this section, we provide a variant of the above barrier for level sets of a convex function which does not use the notion of compatibility, but has a parameter that depends on the variation of the function. For a convex function $f\colon M \to \mathbb{R}$ and $\eta \in \mathbb{R}$ for which there is $p \in M$ with $f(p) < \eta$, the open level set $\mathcal{L}_{f,\eta}^\circ \subseteq M$ is defined by

$$\mathcal{L}_{f,\eta}^\circ = \{p \in M \mid f(p) < \eta\}. \tag{4.12}$$

Define the logarithmic barrier $F_\eta\colon \mathcal{L}_{f,\eta}^\circ \to \mathbb{R}$ by

$$F_\eta(p) = -\log(\eta - f(p)) \quad (p \in \mathcal{L}_{f,\eta}^\circ). \tag{4.13}$$

The logarithmic barrier is convex and has bounded Newton decrements as follows.

Lemma 4.13. *The function $F = F_\eta$ defined in Eq. (4.13) is smooth, closed convex, and satisfies*

$$dF_p(u)^2 \le (\nabla^2 F)_p(u, u) \quad (u \in T_pM, \ p \in \mathcal{L}^\circ_{f,\eta}). \tag{4.14}$$

Proof. Let $\omega(p) := \eta - f(p) > 0$. Then we have

$$dF_p(u) = \frac{df_p(u)}{\omega(p)}, \quad (\nabla^2 F)_p(u, u) = \frac{(\nabla^2 f)_p(u, u)}{\omega(p)} + \frac{df_p(u)^2}{\omega(p)^2}. \tag{4.15}$$

Then by convexity of f, $(\nabla^2 F)_p(u, u) \ge 0$ and hence F is convex, and satisfies $(\nabla^2 F)_p(u, u) \ge dF_p(u)^2$.

The closedness of F is seen as follows: Consider a sequence (p_k, z_k) in the epigraph of F, that converges to $(p_\infty, z_\infty) \in M \times \mathbb{R}$. Note that f is smooth on M, and hence so is F on $\mathcal{L}^\circ_{f,\eta}$. By continuity of f, $\mathcal{L}^\circ_{f,\eta}$ is open, hence disjoint from its boundary in M. Therefore any boundary point q of $\mathcal{L}^\circ_{f,\eta}$ satisfies $f(q) \ge \eta$. Therefore, it is impossible for p_∞ to belong to the boundary of $\mathcal{L}^\circ_{f,\eta}$: that would imply $f(p_\infty) \ge \eta$, which would imply $z_\infty \ge \infty$. Hence $p_\infty \in \mathcal{L}^\circ_{f,\eta}$, and $F(z_\infty) = \lim_{k\to\infty} F(p_k) \le \lim_{k\to\infty} z_k = z_\infty$. □

If an α-self-concordant function F satisfies Eq. (4.14), then F/α is an α-barrier. The following is an extension of [Nes18, Thm. 5.1.4] to our setting:

Theorem 4.14 (Barriers for level sets). *Suppose that $f\colon M \to \mathbb{R}$ is α-self-concordant. Then $F_\eta\colon \mathcal{L}^\circ_{f,\eta} \to \mathbb{R}$ is α'-self-concordant for*

$$\alpha' = \frac{4(\eta - f^*)/\alpha + 1}{(2(\eta - f^*)/\alpha + 1)^2} \tag{4.16}$$

where $f^* := \inf_{x \in M} f(x)$. In particular, F_η/α' is an $O((\eta - f^*)/\alpha)$-barrier for $\mathcal{L}^\circ_{f,\eta}$.

When only considering self-concordance along geodesics, the constant α' can be taken as $\alpha/((\eta - f^*) + \alpha)$, which is exactly what is proven in [Nes18, Thm. 5.1.4]. For self-concordance, however, a little modification is required, which leads to a weaker constant.

Proof. Our starting point is Eq. (4.15), where we recall that $\omega(p) = \eta - f(p)$. Since $df_p(u)^2 = (df_p \otimes df_p)(u, u)$, and $(\nabla_v(df \otimes df))_p(u, u) = ((\nabla_v df)_p \otimes df_p + df_p \otimes (\nabla_v df)_p)(u, u) = 2df_p(u)(\nabla^2 f)_p(u, v)$, the covariant derivative of $\nabla^2 F$ is given by (suppressing p's for convenience)

$$\nabla^3 F(v, u, u) = \frac{\nabla^3 f(v, u, u)}{\omega} + \frac{df(v)\nabla^2 f(u, u)}{\omega^2} + \frac{2df(u)\nabla^2 f(u, v)}{\omega^2} + \frac{2df(v)df(u)^2}{\omega^3}. \tag{4.17}$$

Hence we have

$$|\nabla^3 F(v, u, u)| \le \frac{2\sqrt{\nabla^2 f(v, v)}\nabla^2 f(u, u)}{\sqrt{\alpha}\omega} + \frac{|df(v)|\nabla^2 f(u, u)}{\omega^2}$$
$$+ \frac{2|df(u)|\sqrt{\nabla^2 f(v, v)}\sqrt{\nabla^2 f(u, u)}}{\omega^2} + \frac{2|df(v)|df(u)^2}{\omega^3}.$$

Define τ_1, τ, ξ_1, ξ by

$$\tau_1 := \sqrt{\nabla^2 f(v, v)/\omega}, \ \tau := \sqrt{\nabla^2 f(u, u)/\omega}, \ \xi_1 := |df(v)|/\omega, \ \xi := |df(u)|/\omega.$$

Then we have

$$\frac{|\nabla^3 F(v, u, u)|}{2\sqrt{\nabla^2 F(v, v)}\nabla^2 F(u, u)} \le \frac{(1/\sqrt{\alpha})\omega^{1/2}\tau_1\tau^2 + (1/2)\xi_1\tau^2 + \xi\tau_1\tau + \xi_1\xi^2}{(\tau_1^2 + \xi_1^2)^{1/2}(\tau^2 + \xi^2)}. \tag{4.18}$$

We bound the right-hand side as follows. By homogeneity, we may consider the optimization problem:

$$\text{maximize } (1/\sqrt{\alpha})\omega^{1/2}\tau_1\tau^2 + (1/2)\xi_1\tau^2 + \xi\tau_1\tau + \xi_1\xi^2 \ \text{ s.t. } \ \tau_1^2 + \xi_1^2 = 1, \ \tau^2 + \xi^2 = 1.$$

For fixed (τ, ξ), optimizing with respect to (τ_1, ξ_1) is a linear optimization over the unit circle. The optimum is $((1/\sqrt{\alpha})\omega^{1/2}\tau^2 + \xi\tau, (1/2)\tau^2 + \xi^2)/\sqrt{((1/\sqrt{\alpha})\omega^{1/2}\tau^2 + \xi\tau)^2 + ((1/2)\tau^2 + \xi^2)^2}$. Then the problem reduces to

$$\text{maximize } \sqrt{((1/\sqrt{\alpha})\omega^{1/2}\tau^2 + \xi\tau)^2 + ((1/2)\tau^2 + \xi^2)^2} \ \text{ s.t. } \ \tau^2 + \xi^2 = 1.$$

This optimization problem can be solved using the method of Lagrange multipliers. For convenience set $c = \sqrt{\omega/\alpha}$, and define $q(\tau, \xi) = (\sqrt{\omega/\alpha}\tau^2 + \xi\tau)^2 + (\tau^2/2 + \xi^2)^2$. The system of equations

$$\partial_\tau q(\tau, \xi) = \mu\tau, \quad \partial_\xi q(\tau, \xi) = \mu\xi, \quad \tau^2 + \xi^2 = 1, \quad \mu \in \mathbb{R}$$

has six solutions (τ, ξ, μ), given by

$$(0, \pm 1, 4), \ \frac{1}{\sqrt{4c^2 + 1}}(2c, 1, 16c^4 + 16c^2 + 4), \ \frac{1}{\sqrt{4c^2 + 1}}(-2c, -1, 16c^4 + 16c^2 + 4),$$

$$\frac{1}{\sqrt{4c^2 + 9}}(3, -2c, 16c^2 + 9), \ \frac{1}{\sqrt{4c^2 + 9}}(-3, 2c, 16c^2 + 9)$$

and the largest value attained of $q(\tau, \xi)$ attained at any of these points is $(2c^2 + 1)^2/(4c^2 + 1)$. Therefore, the right-hand side of Eq. (4.18) is at most

$$\sqrt{\frac{(2(\omega/\alpha) + 1)^2}{4(\omega/\alpha) + 1}}.$$

In other words, this gives that $\alpha' = (4(\omega/\alpha) + 1)/(2(\omega/\alpha) + 1)^2$ is a suitable self-concordance constant at p. Taking the maximum over $p \in \mathcal{L}_{f,\eta}^\circ$ yields the choice of α' in Eq. (4.16). $\qquad\square$

4.3 Path-following method

We now discuss a path-following method for objectives which are compatible with a barrier. To this end, we consider the approach of [NN94, Ch. 3]. Their Euclidean framework is rather general, and deals with *self-concordant families*. We specialize to self-concordant families generated by a barrier, and generalize the corresponding path-following method to the Riemannian setting. The goal is to minimize a convex objective function f over an open convex domain D, that is, to find $p \in D$ such that $f(p) \approx \inf_{q \in D} f(q)$. The running assumption we shall make is that we have a barrier F for the domain D such that the function

$$F_t := tf + F \colon D \to \mathbb{R}$$

is α-self-concordant for all $t \geq 0$, with a parameter α that is independent of t. One way to guarantee this is to assume that f is compatible with F, as shown before in Proposition 4.10.

The basic idea of the path-following method is as follows. The algorithm keeps track of two pieces of data, a point p in the domain D and a time parameter t. The initial data to the algorithm is specified by a point $p_{-1} \in D$ such that $\lambda_{F,\alpha}(p_{-1})$ is small. We then choose a time parameter $t_0 > 0$ such that we are in the quadratic convergence regime for Newton's method for F_{t_0} as determined by Theorem 3.17, say $\lambda_{F_{t_0},\alpha}(p_{-1}) < \lambda_* = 1 - 1/\sqrt{2}$. Such initial data can be obtained for instance by using the damped Newton method of Theorem 3.18, or in the Euclidean setting by a similar (reverse) path-following method. We then iterate the following procedure for $k = 0, 1, 2, \ldots$:

(i) Update p_{k-1} to $p_k \in D$ by taking one Newton step with respect to F_{t_k}, so that $\lambda_{F_{t_k},\alpha}(p_{k+1})$ becomes smaller.

(ii) Increase t_k to some t_{k+1} by a constant factor such that one still has $\lambda_{F_{t_{k+1}},\alpha}(p_k) < \lambda_*$.

Throughout the algorithm, p_k will be an approximate minimizer of F_{t_k}. One can also show that if t_k is large enough, approximate minimizers of F_{t_k} are approximate minimizers of f.

We first determine by what factor one can increase t while keeping the Newton decrement below some threshold. The following result is a translation of [NN94, Thm. 3.1.1] to our setting. Note that here, we do not assume that $tf + F$ is self-concordant.

Lemma 4.15. *Let $D \subseteq M$ be open and convex, let $F \colon D \to \mathbb{R}$ be a θ-barrier along geodesics, and let $f \colon D \to \mathbb{R}$ be a convex function. Furthermore, let $t, t', \alpha, c > 0$ and $p \in D$ be such that*

$$\left(1 + \frac{\sqrt{\theta}}{c\sqrt{\alpha}}\right)\left|\log \frac{t'}{t}\right| \leq 1 - \frac{\lambda_{F_t,\alpha}(p)}{c}.$$

Then $\lambda_{F_t,\alpha}(p) \leq c$ implies that $\lambda_{F_{t'},\alpha}(p) \leq c$.

Proof. Let $p \in D$. Throughout the proof, all derivatives of functions defined on M will be taken at the point p, hence we shall omit the subscript. We will assume that $t' \geq t$, but the proof for $t' \leq t$ is analogous. For every $0 \neq u \in T_pM$, define a function $\phi_u \colon [t, t'] \to \mathbb{R}$ by

$$\phi_u(s) = \frac{dF_s(u)}{\sqrt{\nabla^2 F_s(u, u)}}.$$

To prove the lemma, it suffices to show that $|\phi_u(t')| \leq c\sqrt{\alpha}$ for all $u \neq 0$. Since $\phi_{-u} = -\phi_u$, we may assume without loss of generality that $\phi_u(t') \geq 0$. We first compute the derivative of ϕ_u:

$$\begin{aligned}
\partial_s \phi_u(s) &= \frac{df(u)}{\sqrt{\nabla^2 F_s(u, u)}} - \frac{1}{2}\frac{dF_s(u) \cdot \nabla^2 f(u, u)}{(\nabla^2 F_s(u, u))^{3/2}} \\
&= \frac{1}{s}\phi_u(s) - \frac{1}{s}\frac{dF(u)}{\sqrt{\nabla^2 F_s(u, u)}} - \frac{1}{2}\frac{dF_s(u) \cdot \nabla^2 f(u, u)}{(\nabla^2 F_s(u, u))^{3/2}} \\
&= \frac{1}{2s}\phi_u(s) - \frac{1}{s}\frac{dF(u)}{\sqrt{\nabla^2 F_s(u, u)}} + \frac{1}{2s}\frac{dF_s(u) \cdot \nabla^2 F(u, u)}{(\nabla^2 F_s(u, u))^{3/2}} \\
&= \frac{1}{2s}\phi_u(s)\left(1 + \frac{\nabla^2 F(u, u)}{\nabla^2 F_s(u, u)}\right) - \frac{1}{s}\frac{dF(u)}{\sqrt{\nabla^2 F_s(u, u)}}.
\end{aligned}$$

Let t_0 be the largest $s \in [t, t']$ such that $\phi_u(t_0) = 0$; if such an s does not exist, then set $t_0 = t$. Let $t^* \in [t_0, t']$ be such that $\phi_u(t^*)$ is maximal over this interval, and set $\phi_u^* = \phi_u(t^*)$. Then,

$$\begin{aligned}
\phi_u^* &= \phi_u(t_0) + \int_{t_0}^{t^*} \partial_s \phi(s)\, ds \\
&\leq \phi_u(t_0) + \int_{t_0}^{t^*}\left[\frac{1}{2s}\phi_u(s)\left(1 + \frac{\nabla^2 F(u, u)}{\nabla^2 F_s(u, u)}\right) + \frac{1}{s}\frac{|dF(u)|}{\sqrt{\nabla^2 F_s(u, u)}}\right] ds \\
&\leq |\phi_u(t_0)| + \int_{t_0}^{t^*}\left[\frac{1}{s}\phi_u(s) + \frac{1}{s}\sqrt{\theta}\right] ds \\
&\leq |\phi_u(t)| + (\phi_u^* + \sqrt{\theta})\log \frac{t^*}{t_0};
\end{aligned}$$

41

the second inequality follows since $\nabla^2 F_s \geq \nabla^2 F$ as f is convex and using that F is a θ-barrier; the last inequality is ensured by our choice of t_0. Using $|\phi_u(t)| \leq \sqrt{\alpha}\lambda_{F_t,\alpha}(p)$, we obtain

$$\phi_u^*\left(1 - \log\frac{t^*}{t_0}\right) \leq \sqrt{\alpha}\,\lambda_{F_t,\alpha}(p) + \sqrt{\theta}\log\frac{t^*}{t_0}, \tag{4.19}$$

On the other hand, since $t \leq t_0 \leq t^* \leq t'$, our assumption implies that

$$\left(1 + \frac{\sqrt{\theta}}{c\sqrt{\alpha}}\right)\log\frac{t^*}{t_0} \leq \left(1 + \frac{\sqrt{\theta}}{c\sqrt{\alpha}}\right)\left|\log\frac{t'}{t}\right| \leq 1 - \frac{\lambda_{F_t,\alpha}(p)}{c},$$

or equivalently

$$\sqrt{\alpha}\,\lambda_{F_t,\alpha}(p) + \sqrt{\theta}\log\frac{t^*}{t_0} \leq c\sqrt{\alpha}\left(1 - \log\frac{t^*}{t_0}\right). \tag{4.20}$$

Combining Eqs. (4.19) and (4.20) gives $\phi_u^* \leq c\sqrt{\alpha}$, implying that $|\phi_u(t')| \leq c\sqrt{\alpha}$ as desired. $\qquad\square$

We now show that for large $t > 0$, approximate minimizers of F_t correspond to approximate minimizers of f. The proposition and proof we give below are adapted from [NN94, Prop. 3.2.4].

Proposition 4.16. *Let $D \subseteq M$ be open and convex, let $F\colon D \to \mathbb{R}$ be a θ-barrier along geodesics for D, and let $f\colon D \to \mathbb{R}$ be a smooth convex function which has a closed convex extension. For some fixed $t > 0$, suppose that $F_t := tf + F$ is α-self-concordant along geodesics for some $\alpha > 0$ and that it is bounded from below. Then for every $p \in D$ such that $\lambda_{F_t,\alpha}(p) < \frac{1}{3}$, we have*

$$f(p) - \inf_{q \in D} f(q) \leq \frac{2\theta + \alpha\rho(\lambda_{F_t,\alpha}(p))}{t},$$

where we recall from Eq. (3.17) that $\rho(r) = -r - \log(1 - r)$.

Proof. By Lemma 2.3, F_t is closed convex and hence strongly α-self-concordant along geodesics. Because it also has positive definite Hessians and we have $\lambda_{F_t,\alpha}(p) < 1$, Proposition 3.14 implies that F_t attains its minimum at a unique minimizer $p_{t,*} \in D$ and moreover

$$F_t(p) - F_t(p_{t,*}) \leq \alpha\rho(\lambda_{F_t,\alpha}(p)). \tag{4.21}$$

Furthermore, Lemma 3.15 shows that if $u \in T_pM$ is such that $\mathrm{Exp}_p(u) = p_{t,*}$, then

$$\|u\|_{F_t,p,\alpha} \leq \frac{\lambda_{F_t,\alpha}(p)}{1 - \lambda_{F_t,\alpha}(p)} < \frac{1}{2}$$

where the last inequality follows from $\lambda_{F_t,\alpha}(p) < \frac{1}{3}$. Using Corollary 3.11, we obtain that

$$\mathrm{Exp}_{p_{t,*}}(v) = \mathrm{Exp}_p(2u) \in D,$$

where $v = \tau_{\gamma,1}u$ is the parallel transport of u from p to $p_{t,*}$ along the geodesic $\gamma(t) := \mathrm{Exp}_p(tu)$. By Proposition 4.4, it follows that

$$dF_{p_{t,*}}(v) \leq \theta$$

and hence, using convexity of F and $\mathrm{Exp}_{p_{t,*}}(v) = p$,

$$F(p_{t,*}) - F(p) \leq -dF_{p_{t,*}}(-v) = dF_{p_{t,*}}(v) \leq \theta. \tag{4.22}$$

Together, Eqs. (4.21) and (4.22) then show that

$$
\begin{aligned}
f(p) &= \frac{F_t(p) - F(p)}{t} \\
&\leq \frac{F_t(p_{t,*}) + \alpha\rho(\lambda_{F_t,\alpha}(p)) - F(p)}{t} \\
&= f(p_{t,*}) + \frac{F(p_{t,*}) - F(p) + \alpha\rho(\lambda_{F_t,\alpha}(p))}{t} \\
&\leq f(p_{t,*}) + \frac{\theta + \alpha\rho(\lambda_{F_t,\alpha}(p))}{t}.
\end{aligned}
\tag{4.23}
$$

We will now give an upper bound on $f(p_{t,*}) - f(q)$ for every $q \in D$. Let $v \in T_{p_{t,*}}M$ be such that $\operatorname{Exp}_{p_{t,*}}(v) = q$. Using the convexity of f, the fact that $p_{t,*}$ is a minimizer of F_t, and Proposition 4.4 (in this order) gives

$$
f(p_{t,*}) - f(q) \leq -df_{p_{t,*}}(v) = \frac{dF_{p_{t,*}}(v)}{t} \leq \frac{\theta}{t}.
$$

Combining this with Eq. (4.23) and optimizing over $q \in D$ gives the desired bound. $\qquad\square$

We now come to the main result of this section, giving a path-following method which converges to a minimizer of the objective, generalizing [NN94, Prop. 3.2.4] to our setting.

Theorem 4.17. *Let $D \subseteq M$ be an open, convex, and bounded domain. Let $F\colon D \to \mathbb{R}$ be a θ-barrier for D, and let $f\colon D \to \mathbb{R}$ be a smooth convex function with a closed convex extension. Let $\alpha > 0$ be such that $F_t := tf + F$ is α-self-concordant for all $t \geq 0$. Choose $1 > \lambda^{(1)} > \lambda^{(2)} > 0$ such that $\left(\frac{\lambda^{(1)}}{1-\lambda^{(1)}}\right)^2 \leq \lambda^{(2)} < \frac{1}{3}$; a suitable choice is given by $\lambda^{(1)} = \frac{1}{4}$, $\lambda^{(2)} = \frac{1}{9}$. Finally, let $p \in D$ be given such that $\lambda_{F,\alpha}(p) < \lambda^{(1)}$, and assume that p is not a minimizer of f. Define a sequence of time parameters*

$$
t_0 = \frac{\sqrt{\alpha}\lambda^{(1)} - \lambda_F(p)}{\|df_p\|^*_{F,p}}, \quad t_\ell = t_0 \cdot \exp\left(\ell\frac{\lambda^{(1)} - \lambda^{(2)}}{\lambda^{(1)} + \sqrt{\theta/\alpha}}\right) \text{ for } \ell = 0, 1, 2, \ldots,
$$

and a sequence of points

$$
p_{-1} = p, \quad p_\ell = (p_{\ell-1})_{F_{t_\ell},+} \text{ for } \ell = 0, 1, 2, \ldots.
$$

i.e., p_ℓ is the Newton iterate of $p_{\ell-1}$ with respect to F_{t_ℓ}. Then this sequence is well-defined, in the sense that $p_\ell \in D$ for all $\ell \geq 0$, and it satisfies

$$
f(p_\ell) - \inf_{q \in D} f(q) \leq \frac{2(\theta + \alpha)}{t_\ell} = \frac{2(\theta + \alpha)\|df_p\|^*_{F,p}}{\sqrt{\alpha}\lambda^{(1)} - \lambda_F(p)} \exp\left(-\ell \cdot \frac{\lambda^{(1)} - \lambda^{(2)}}{\lambda^{(1)} + \sqrt{\theta/\alpha}}\right).
$$

Proof. By the assumptions on f and strong self-concordance of F, we see from Lemma 2.3 that F_t is strongly α-self-concordant on D for all $t \geq 0$. We shall prove by induction on ℓ that for every $\ell \geq 0$, we have $p_\ell \in D$ and

$$
\lambda_{F_{t_\ell},\alpha}(p_{\ell-1}) \leq \lambda^{(1)}, \quad \lambda_{F_{t_\ell},\alpha}(p_\ell) \leq \lambda^{(2)}.
$$

Let us first check that $\lambda_{F_{t_0},\alpha}(p_{-1}) = \lambda_{F_{t_0},\alpha}(p) \leq \lambda^{(1)}$. For every $u \neq 0$, we have

$$
|d(F_{t_0})_p(u)| \leq t_0|df_p(u)| + |dF_p(u)|
$$

43

$$= (\sqrt{\alpha}\lambda^{(1)} - \lambda_F(p))\frac{|df_p(u)|}{\|df_p\|_{F,p}^*} + |dF_p(u)|$$

$$\leq (\sqrt{\alpha}\lambda^{(1)} - \lambda_F(p))\|u\|_{F,p} + \|dF_p\|_{F,p}^*\|u\|_{F,p}$$

$$= \sqrt{\alpha}\lambda^{(1)}\|u\|_{F,p}$$

$$\leq \sqrt{\alpha}\lambda^{(1)}\|u\|_{F_{t_0},p},$$

hence $\|d(F_{t_0})_p\|_{F_{t_0},p}^* \leq \sqrt{\alpha}\lambda^{(1)}$, which is equivalent to $\lambda_{F_{t_0},\alpha}(p) \leq \lambda^{(1)}$. Next, if $\lambda_{F_{t_\ell},\alpha}(p_{\ell-1}) \leq \lambda^{(1)}$ for some $\ell \geq 0$, then by applying Theorem 3.16, we find that the Newton iterate p_ℓ is in D satisfies

$$\lambda_{F_{t_\ell},\alpha}(p_\ell) \leq \left(\frac{\lambda^{(1)}}{1 - \lambda^{(1)}}\right)^2 \leq \lambda^{(2)}.$$

Lastly, it remains to verify that if $\lambda_{F_{t_\ell},\alpha}(p_\ell) \leq \lambda^{(2)}$ for some $\ell \geq 0$, then $\lambda_{F_{t_{\ell+1}},\alpha}(p_\ell) \leq \lambda^{(1)}$. The t_ℓ are chosen exactly so that

$$\left(1 + \frac{\sqrt{\theta}}{\lambda^{(1)}\sqrt{\alpha}}\right)\left|\log\frac{t_\ell}{t_{\ell+1}}\right| = \left(1 + \frac{\sqrt{\theta}}{\lambda^{(1)}\sqrt{\alpha}}\right)\left(\frac{\lambda^{(1)} - \lambda^{(2)}}{\lambda^{(1)} + \sqrt{\theta/\alpha}}\right) = 1 - \frac{\lambda^{(2)}}{\lambda^{(1)}}.$$

We conclude that $\lambda_{F_{t_{\ell+1}},\alpha}(p) \leq \lambda^{(1)}$ by Lemma 4.15. Lastly, the bound on $f(p_l) - \inf_{q \in D} f(q)$ follows from Proposition 4.16, where we use that $\lambda^{(2)} < \frac{1}{3}$ and $\rho(\frac{1}{3}) \approx 0.072 \leq 2$. $\qquad\square$

We end with a simple but useful lemma to upper bound the quantity $\|df_p\|_{F,p}^*$.

Lemma 4.18. *Let $p \in D$, and $f, F\colon D \to \mathbb{R}$ be such that f is convex and F is strongly 1-self-concordant on D. Then*

$$\|df_p\|_{F,p}^* \leq \sup_{q \in D} f(q) - f(p) \leq \sup_{q \in D} f(q) - \inf_{q \in D} f(q).$$

Proof. By Corollary 3.8, the Dikin ellipsoid $B := B_{F,p}^\circ(1)$ of radius 1 is contained in D. Then the convexity of f gives

$$\|df_p\|_{F,p}^* = \sup_{\substack{u \in T_pM \\ \|u\|_{F,p} < 1}} |df_p(u)| = \sup_{\substack{u \in T_pM \\ \|u\|_{F,p} < 1}} df_p(u) \leq \sup_{\substack{u \in T_pM \\ \|u\|_{F,p} < 1}} f(\mathrm{Exp}_p(u)) - f(p) = \sup_{q \in B} f(q) - f(p),$$

which is at most $\sup_{q \in D} f(q) - f(p)$ as $B \subseteq D$. $\qquad\square$

5 The squared distance function

In this section we discuss self-concordance of the squared distance function to a point. In Section 5.1 we recall some useful formulas that apply for arbitrary Hadamard manifolds. In Section 5.2 we focus on the space $\mathrm{PD}(n)$ of positive-definite complex $n \times n$ matrices and prove that the distance squared to any point is self-concordant. This relies on explicit computations of higher covariant derivatives. Next, in Section 5.3 we use these same formulas to deduce stronger self-concordance estimates in the case of hyperbolic space $\mathbb{H}^n$, and use these to construct a barrier for the distance function rather than its square; all this generalizes readily to the model spaces of arbitrary constant negative curvature. The results of this section are applied in Section 6.

5.1 Hadamard manifolds

Let M be a Hadamard manifold, i.e., a simply-connected geodesically-complete Riemannian manifold with non-positive sectional curvature (cf. Section 2.3). Fix $p_0 \in M$ and consider the function that computes the *squared distance* to the point p_0, that is,

$$f\colon M \to \mathbb{R}, \quad f(p) = d(p, p_0)^2.$$

Then it is known that f is 2-strongly convex (which follows from variational principles for the energy of a curve, cf. [Lee18, Thm. 10.22]). In fact, this is a defining property of the more general class of CAT(0)-spaces [BH99]. It will also be useful to consider the distance to p_0,

$$g\colon M \to \mathbb{R}, \quad g(p) = d(p, p_0),$$

which is still convex. The following lemma summarizes well-known properties of these functions.

Lemma 5.1. *Let M be a Hadamard manifold, let $p_0 \in M$, and define $f, g\colon M \to \mathbb{R}$ by $f(p) = d(p, p_0)^2$ and $g(p) = d(p, p_0)$. Then f is 2-strongly convex and g is convex. For every $p \neq p_0$, g is smooth at p, and the differentials and Hessians satisfy*

$$df_p = 2g(p)dg_p = -2 \left\langle \mathrm{Exp}_p^{-1}(p_0), \cdot \right\rangle_p, \tag{5.1}$$

$$\nabla^2 f = 2g\nabla^2 g + 2\, dg \otimes dg \succeq 2\, dg \otimes dg = \frac{df \otimes df}{2f}. \tag{5.2}$$

Proof. The strong convexity of f and convexity of g hold on any CAT(0)-space [BH99, Cor. II.2.5]. Whenever $p \neq p_0$, $f(p) \neq 0$ and hence $g = \sqrt{f}$ is smooth at p. By the chain rule, $df = 2g\, dg$. To compute these, note that g is 1-Lipschitz by the triangle inequality, so $|dg_p(u)| \leq \|u\|_p$ for all $u \in T_p M$. But since the geodesic from p in the direction $\mathrm{Exp}_p^{-1}(p_0)$ has constant speed and reaches p_0 at time 1, it follows that

$$dg_p(\mathrm{Exp}_p^{-1}(p_0)) = -g(p).$$

As $\|\mathrm{Exp}_p^{-1}(p_0)\|_p = g(p)$, this means that the Cauchy–Schwarz inequality applied to

$$g(p) = |dg_p(\mathrm{Exp}_p^{-1}(p_0))| = |\langle (\mathrm{grad}\, g)_p, \mathrm{Exp}_p^{-1}(p_0)\rangle| \leq \|(\mathrm{grad}\, g)_p\|_p \|\mathrm{Exp}_p^{-1}(p_0)\|_p \leq \|\mathrm{Exp}_p^{-1}(p_0)\|_p$$

holds with equality, hence $(\mathrm{grad}\, g)_p = -g(p)^{-1}\mathrm{Exp}_p^{-1}(p_0)$ and $dg_p = -g(p)^{-1}\langle \mathrm{Exp}_p^{-1}(p_0), \cdot\rangle$, and $df_p = -2\langle \mathrm{Exp}_p^{-1}(p_0), \cdot\rangle_p$ follows. We finally derive the formulas for the Hessians. Applying the product rule to $df = 2g\, dg$ yields

$$(\nabla^2 f)_p = 2g(p)(\nabla^2 g)_p + 2\, dg_p \otimes dg_p,$$

The lower bound in Eq. (5.2) follows since $(\nabla^2 g)_p \succeq 0$, as a consequence of the convexity of g. $\square$

Corollary 5.2. *The Newton decrement of $f(p) = d(p, p_0)^2$ is given by $\lambda_f(p) = \sqrt{2}\, d(p, p_0)$.*

Proof. Recall the variational characterization of the Newton decrement in Eq. (3.20):

$$\lambda_f(p) = \min\left\{\lambda \geq 0 : df_p \otimes df_p \preceq \lambda^2 (\nabla^2 f)_p\right\}.$$

Thus, $\lambda_f \leq \sqrt{2f}$ by Eq. (5.2). As g is linear in the direction $\mathrm{Exp}_p^{-1}(p_0)$, its Hessian vanishes in this direction and so we in fact have equality, by the first equality in Eq. (5.2). $\square$

We use Lemma 5.1 to prove the following result, which is used later to prove Theorem 5.22.

Lemma 5.3. *Let* $\Psi\colon M \times \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}$ *be the function defined by*

$$\Psi(p, R, S) = R - S^{-1}d(p, p_0)^2.$$

Then Ψ *is concave, with Hessian given by*

$$\nabla^2 \Psi = -\frac{2\left(S^{-1}g\,dS - dg\right)^{\otimes 2} + \left(\nabla^2 f - 2dg \otimes dg\right)}{S} \preceq 0,$$

where f, g *are as in Lemma 5.1, dS is the differential of the projection* $(p, R, S) \mapsto S$, *and we write* dg *for the differential of* $(p, R, S) \mapsto g(p)$ *by a slight abuse of notation. Moreover, for* $u = (u_p, u_R, u_S)$ *and* $w = (w_p, w_R, w_S)$ *tangent vectors at* (p, R, S), *one has*

$$\nabla^3 \Psi(w, u, u) = -2\frac{u_S}{S}\nabla^2 \Psi(w, u) - \frac{w_S}{S}\nabla^2 \Psi(u, u) - \frac{1}{S}\nabla^3 f(w_p, u_p, u_p).$$

Proof. Clearly,

$$d\Psi = dR + S^{-2}f\,dS - S^{-1}\,df.$$

Since $\nabla dR \equiv 0 \equiv \nabla dS$, this yields

$$\nabla^2 \Psi = -2S^{-3}f\,dS \otimes dS + S^{-2}\,df \otimes dS + S^{-2}\,dS \otimes df - S^{-1}\nabla^2 f. \tag{5.3}$$

We now use Eqs. (5.1) and (5.2) to rewrite the above as

$$\begin{aligned}
\nabla^2 \Psi &= -2S^{-3}g^2\,dS \otimes dS + 2S^{-2}g\,dg \otimes dS + 2S^{-2}g\,dS \otimes dg - S^{-1}(2g\nabla^2 g + 2\,dg \otimes dg) \\
&= -2S^{-1}(S^{-1}g\,dS - dg)^{\otimes 2} - 2S^{-1}g\nabla^2 g.
\end{aligned}$$

Taking one more derivative in Eq. (5.3), we obtain

$$\begin{aligned}
\nabla^3 \Psi(w, u, u) &= 6S^{-4}f\,dS(w)\,dS(u)^2 - 2S^{-3}df(w)\,dS(u)^2 - 4S^{-3}dS(w)\,df(u)\,dS(u) \\
&\quad + 2S^{-2}\nabla^2 f(w, u)\,dS(u) + S^{-2}dS(w)\nabla^2 f(u, u) - S^{-1}\nabla^3 f(w, u, u) \\
&= -2S^{-1}dS(u)\,\nabla^2 \Psi(w, u) - S^{-1}dS(w)\,\nabla^2 \Psi(u, u) - S^{-1}\nabla^3 f(w, u, u). \qquad \square
\end{aligned}$$

Corollary 5.4. *Let* $D = \{(p, R, S) \in M \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} : RS - f(p) > 0\}$. *Then the function* $F\colon D \to \mathbb{R}$ *defined by* $F(p, R, S) = -\log(R - S^{-1}d(p, p_0)^2)$ *is convex.*

5.2 Positive definite matrices

In this subsection, we specialize to the space $\mathrm{PD}(n) = \mathrm{PD}(n, \mathbb{C})$ of positive definite Hermitian $n \times n$ matrices, which is a Hadamard manifold when endowed with a well-known Riemannian metric. We collect a number of well-known results from the literature and then derive explicit formulas for the higher derivatives of the squared distance on this space by using techniques from matrix analysis. The main result of this section is Theorem 5.15, where we show that the squared distance is self-concordant on $\mathrm{PD}(n)$. As explained in the introduction, this implies that the squared distance is self-concordant on arbitrary Hadamard symmetric spaces.

We will often use notation of the form $h(P)$ where $h\colon \mathbb{R}_{>0} \to \mathbb{R}$ is some scalar-valued function, which refers to the Hermitian matrix obtained by expanding P in an eigenbasis and applying h to its eigenvalues. Examples include but are not limited to expressions of the form P^t with $t \in \mathbb{R}$, $P + \lambda = P + \lambda I$ where $\lambda \in \mathbb{R}$, $\log(P)$, et cetera.

We think of $\mathrm{PD}(n)$ as an open submanifold of the $n \times n$ Hermitian matrices $\mathrm{Herm}(n) \subseteq \mathbb{C}^{n \times n}$, so that we can identify $T_P\mathrm{PD}(n) \cong \mathrm{Herm}(n)$ at any $P \in \mathrm{PD}(n)$. Concretely, $X \in \mathrm{Herm}(n)$ corresponds to the tangent vector of the curve $t \mapsto P + Xt = P^{1/2}(I + tP^{-1/2}XP^{-1/2})P^{1/2}$ at $t = 0$. These curves would be geodesics if we equipped $\mathrm{PD}(n)$ with the Euclidean metric inherited from $\mathrm{Herm}(n)$. Instead, we introduce the following Riemannian metric on $\mathrm{PD}(n)$:

$$\langle X, Y \rangle_P := \mathrm{Tr}\left[(P^{-1/2}XP^{-1/2})(P^{-1/2}YP^{-1/2})\right] = \mathrm{Tr}\left[P^{-1}XP^{-1}Y\right] \tag{5.4}$$

for $X, Y \in T_P\mathrm{PD}(n)$. This is real-valued as the Hilbert-Schmidt inner product of two Hermitian matrices. Interstingly, $\langle \cdot, \cdot \rangle_P$ is also the *Euclidean* Hessian of the function $P \mapsto -\log\det(P)$, which is a Euclidean self-concordant barrier for $\mathrm{PD}(n)$.

It is immediate from the definition that for every $P \in \mathrm{PD}(n)$, the bijection $Q \mapsto P^{1/2}QP^{1/2}$ is a Riemannian isometry of $\mathrm{PD}(n)$, meaning it preserves inner products between tangent vectors. Then it also preserves the distance between any two points: for any $P, Q, Q' \in \mathrm{PD}(n)$, we have

$$d(Q, Q') = d(P^{1/2}QP^{1/2}, P^{1/2}Q'P^{1/2}).$$

Therefore, if one is interested in properties of squared distance $f(P) = d(P, P_0)^2$, one may choose $P_0 = I$ without loss of generality. This will be convenient for our purposes.

We now give explicit formulas for the geodesics on $\mathrm{PD}(n)$. For any $P \in \mathrm{PD}(n)$, the exponential map at P reads

$$\mathrm{Exp}_P(X) = P^{1/2}e^{P^{-1/2}XP^{-1/2}}P^{1/2} \tag{5.5}$$

and hence the geodesics through P take the form

$$P(t) = \mathrm{Exp}_P(tX) = P^{1/2}e^{tP^{-1/2}XP^{-1/2}}P^{1/2}.$$

In particular, the geodesics through $P = I$ are of the form $\mathrm{Exp}_I(tX) = e^{tX}$. From the description of the exponential map above it follows that $\mathrm{Exp}_P : T_P\mathrm{PD}(n) \to \mathrm{PD}(n)$ is a smooth bijection for all P, with smooth inverse given by

$$\mathrm{Exp}_P^{-1}(Q) = P^{1/2}\log(P^{-1/2}QP^{-1/2})P^{1/2}.$$

By the Hopf–Rinow theorem, there exists a length-minimizing geodesic, which is unique by the bijectivity of the exponential map; hence the distance induced by the Riemannian metric is

$$d(P, Q) = \|\log(P^{-1/2}QP^{-1/2})\|_{\mathrm{HS}} = \|\log(Q^{-1/2}PQ^{-1/2})\|_{\mathrm{HS}}.$$

where $\|\cdot\|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt (Frobenius) norm, because $d(P, Q) = \|\mathrm{Exp}_P^{-1}(Q)\|_P$.

The geodesics on $\mathrm{PD}(n)$ can be naturally described using the operator geometric mean, which is defined for $P, Q = \mathrm{PD}(n)$ and $t \in [0, 1]$ to be

$$P \#_t Q := P^{1/2}(P^{-1/2}QP^{-1/2})^t P^{1/2}.$$

The above formula for the geodesics through P shows that this is equal to $\mathrm{Exp}_P(t\mathrm{Exp}_P^{-1}(Q))$, and so it is the "time-t"-geodesic-midpoint between P and Q.

One can also explicitly describe the parallel transport along geodesics. For $P, Q \in \mathrm{PD}(n)$, the parallel transport of $X \in T_P \mathrm{PD}(n)$ along the unique geodesic from P to Q is given by[9]

$$\tau_{P \to Q}(X) = P^{1/2}(P^{-1/2}QP^{-1/2})^{1/2}P^{-1/2}XP^{-1/2}(P^{-1/2}QP^{-1/2})^{1/2}P^{1/2}. \tag{5.6}$$

This may be conveniently restated as

$$\tau_{P \to \mathrm{Exp}_P(tY)}(X) = P^{1/2}(e^{\frac{t}{2}P^{-1/2}YP^{-1/2}})P^{-1/2}XP^{-1/2}(e^{\frac{t}{2}P^{-1/2}YP^{-1/2}})P^{1/2} \tag{5.7}$$

which for the geodesics emanating from the identity specializes to

$$\tau_{I \to e^{tZ}}(X) = e^{\frac{t}{2}Z}Xe^{\frac{t}{2}Z},$$

i.e.,

$$\tau_{I \to Q}(X) = Q^{1/2}XQ^{1/2}.$$

Now consider a function $f \colon \mathrm{PD}(n) \to \mathbb{R}$. It follows from the previous considerations and the discussion in Section 2.3 that the third derivative at $I \in \mathrm{PD}(n)$ can be computed as follows for $X, Z \in T_I \mathrm{PD}(n)$:

$$(\nabla^3 f)_I(Z, X, X) = \partial_{t=0}(\nabla^2 f)_I(\tau_{I \to \exp_I(tZ)}(X), \tau_{I \to \exp_I(tZ)}(X))$$
$$= \partial_{t=0}(\nabla^2 f)_I(e^{\frac{t}{2}Z}Xe^{\frac{t}{2}Z}, e^{\frac{t}{2}Z}Xe^{\frac{t}{2}Z}).$$

Although we will not need it explicitly, one can also use the above to determine the covariant derivative of a general vector field. More precisely, the covariant derivative $\nabla_X Y$, where $X \in T_P \mathrm{PD}(n)$ and $Y(t)$ is a vector field defined along the curve $P(t) = \mathrm{Exp}_P(tX)$, is given by

$$\nabla_X Y = \partial_{t=0}\tau_{P(t) \to P}(Y(t)).$$

For $P = I$, we have

$$\nabla_X Y = \partial_{t=0}\tau_{e^{tX} \to I}(Y(t)) = \partial_{t=0}e^{-\frac{t}{2}X}Y(t)e^{-\frac{t}{2}X} = \dot{Y}(0) - \frac{1}{2}\{X, Y(0)\}$$

where we write $\{X, Y\} = XY + YX$ for the anticommutator of X and Y.

Lastly, we have an explicit expression for the Riemann curvature tensor on $\mathrm{PD}(n)$. The fact that the curvature tensor is of this form follows from [Hel79, Thm. IV.4.2], and the prefactor of $\frac{1}{4}$ can be deduced from the fact that $\mathrm{SPD}(2, \mathbb{C})$ is a model space for constant curvature $-\frac{1}{2}$ (the prefactor appears because we work directly with positive-definite matrices, rather than the quotient $\mathrm{GL}_n(\mathbb{C})/\mathrm{U}(n)$). Alternatively, one may consult the self-contained explicit proof available in [DP14]:

---

[9]One way of proving Eq. (5.6) is as follows [Sak96, Lem. IV.6.2]: for every $P \in \mathrm{PD}(n)$, the *geodesic inversion* map $s_P \colon \mathrm{PD}(n) \to \mathrm{PD}(n)$ given by $s_P(Q) = \mathrm{Exp}_P(-\mathrm{Exp}_P^{-1}(Q)) = PQ^{-1}P$ is an isometry (more generally, the maps $Q \mapsto Q^{-1}$ and $Q \mapsto AQA^*$ are isometries for every $A \in \mathrm{GL}(n, \mathbb{C})$). Let $P_0, P_1 \in \mathrm{PD}(n)$, and let $\gamma \colon \mathbb{R} \to M$ be the unique geodesic such that $\gamma(0) = P_0$ and $\gamma(1) = P_1$. Then $s_{P_0}(\gamma(t)) = \gamma(-t)$ and $s_{P_1}(\gamma(t)) = \gamma(1 - t)$. If X_t is a parallel vector field along γ, then so is $d(s_{P_0})(X_{-t})$, as s_{P_0} is an isometry; but $d(s_{P_0})_{P_0} = -I_{T_{P_0}\mathrm{PD}(n)}$, and so $d(s_{P_0})(X_{-t}) = -X_t$ by the uniqueness of parallel vector fields. Similarly, $d(s_{\gamma(1/2)})(X_{1/2-t}) = -X_{1/2+t}$, and so $d(s_{\gamma(1/2)} \circ s_{P_0})(X_0) = X_1 = \tau_{P_0 \to P_1}(X_0)$. Expanding the definition of $s_{\gamma(1/2)} \circ s_{P_0}$ (also called a *transvection*), it is easy to see that its derivative is exactly the right-hand side in Eq. (5.6).

Lemma 5.5. *The Riemann curvature $(1,3)$-tensor at $P \in \mathrm{PD}(n)$ is given by*

$$R(X,Y)Z = -\frac{1}{4}[[P^{-1/2}XP^{-1/2}, P^{-1/2}YP^{-1/2}], P^{-1/2}ZP^{-1/2}]$$

for every $X, Y, Z \in T_P\mathrm{PD}(n)$. In particular, the curvature tensor is parallel along any geodesic.

This last property may be more succinctly stated as follows: if one thinks of R as a $(0,4)$-tensor, then $\nabla R \equiv 0$. Therefore $\mathrm{PD}(n)$ is a *locally symmetric space*, see [Lee18, Thm. 10.19], and because it is simply connected, it is also a globally symmetric space. A simple computation using the above lemma shows that $\mathrm{PD}(n)$ has sectional curvatures bounded by an n-independent constant with our normalization of the metric:

Lemma 5.6. *The space $\mathrm{PD}(n)$ has all sectional curvatures in $[-\frac{1}{2}, 0]$.*

Proof. Let $X, Y \in T_I\mathrm{PD}(n) = \mathrm{Herm}(n)$ have $\|X\|_I = \|Y\|_I = 1$ and $\langle X, Y \rangle_I = \mathrm{Tr}[XY] = 0$. Assume without loss of generality that Y is diagonal. Then

$$\langle R(X,Y)Y, X \rangle = -\frac{1}{4}\sum_{i,j=1}^{n}|X_{ij}|^2(Y_{jj} - Y_{ii})^2.$$

This is clearly at most 0, and

$$\sum_{i,j=1}^{n}|X_{ij}|^2(Y_{jj} - Y_{ii})^2 \leq 2\sum_{i,j=1, i\neq j}^{n}|X_{ij}|^2(Y_{jj}^2 + Y_{ii}^2) \leq 2\sum_{i,j=1}^{n}|X_{ij}|^2\|Y\|_I^2 = 2\|X\|_I^2\|Y\|_I^2 = 2,$$

so $K(X,Y) \geq -\frac{1}{2}$. $\qquad\square$

We now turn to the task of computing higher derivatives of the squared distance on $\mathrm{PD}(n)$. Recall from Section 5.2 that the distance between $P, Q \in \mathrm{PD}(n)$ is given by $d(P,Q)^2 = \|\log(P^{-1/2}QP^{-1/2})\|_{\mathrm{HS}}^2$. To differentiate this, we use the following integral expression for the operator logarithm: for $Q \in \mathrm{PD}(n)$, one has

$$\log(Q) = \int_0^\infty \left(\frac{1}{I+\lambda} - \frac{1}{Q+\lambda}\right)d\lambda, \tag{5.8}$$

where $Q + \lambda$ is shorthand for $Q + \lambda I$, and $\frac{1}{Q+\lambda} = (Q+\lambda)^{-1}$. The advantage of this expression is that it is an integral of *rational* functions of Q, which is straightforward to differentiate using the Leibniz integral rule and the following rule for differentiating matrix inverses: if $t \mapsto Q_t \in \mathrm{PD}(n)$ is a smooth curve defined on an open interval containing 0, then

$$\partial_{t=0}(Q_t^{-1}) = -Q_0^{-1}(\partial_{t=0}Q_t)Q_0^{-1}, \tag{5.9}$$

as can be seen from differentiating the identity $Q_tQ_t^{-1} = I$.

We now use this integral representation to compute derivatives of the squared distance. For convenience, we consider only the squared distance to the identity $I \in \mathrm{PD}(n)$, but this is without loss of generality; to compute the derivatives of $d(\cdot, P)^2$ for $P \in \mathrm{PD}(n)$, one may use the fact that $Q \mapsto P^{1/2}QP^{1/2}$ is an isometry sending I to P. First, we record the formula for the first derivative.

Proposition 5.7. *Let $f(Q) = d(Q,I)^2 = \|\log(Q)\|_{\mathrm{HS}}^2$. Then for $U \in T_Q\mathrm{PD}(n)$,*

$$df_Q(U) = 2\,\mathrm{Tr}[Q^{-1}\log(Q)U] = 2\,\langle Q^{1/2}\log(Q)Q^{1/2}, U\rangle_Q,$$

where $\langle \cdot, \cdot \rangle_Q$ is the Riemannian metric in $\mathrm{PD}(n)$ defined in Eq. (5.4).

Proof. Let $Q_t = \mathrm{Exp}_Q(tU)$ be the geodesic through Q in the direction U. Then by Eq. (5.5), we have

$$Q_t = Q^{1/2} e^{tQ^{-1/2}UQ^{-1/2}} Q^{1/2},$$

and so

$$\partial_{t=0} f(Q_t) = \partial_{t=0} \|\log(Q_t)\|_{\mathrm{HS}}^2 = 2\,\mathrm{Tr}[\log(Q) \cdot \partial_{t=0} \log(Q_t)].$$

To evaluate $\partial_{t=0} \log(Q_t)$, we use Eq. (5.8) and Eq. (5.9) to obtain

$$\partial_{t=0} \log(Q_t) = \partial_{t=0} \int_0^\infty \left(\frac{1}{I+\lambda} - \frac{1}{Q_t + \lambda} \right) d\lambda = \int_0^\infty \frac{1}{Q+\lambda} U \frac{1}{Q+\lambda} \, d\lambda.$$

Therefore

$$\partial_{t=0} f(Q_t) = 2\,\mathrm{Tr}\left[\log(Q) \cdot \int_0^\infty \frac{1}{Q+\lambda} U \frac{1}{Q+\lambda} \, d\lambda \right] = 2\,\mathrm{Tr}\left[Q^{-1} \log(Q) \cdot U \right],$$

where we used cyclicity of the trace and $\int_0^\infty \frac{1}{(q+\lambda)^2} \, d\lambda = q^{-1}$. $\qquad\square$

Remark 5.8. *In the above proof, one may also use the curve $t \mapsto Q + tU$ instead of the geodesic, because they agree in first order: it holds that $\partial_{t=0}(Q + tU) = U = \partial_{t=0}\mathrm{Exp}_Q(tU)$, and hence first derivatives of functions are not affected. However, for the second derivative, $(\nabla^2 f)_P(U,U) = \partial_{t=0}^2 f(\mathrm{Exp}_Q(tU))$ and $\partial_{t=0}^2 f(Q + tU)$ are generally distinct; a simple example is given by the function $f(P) = \mathrm{Tr}[P]$, differentiating at $Q = I$.*

Remark 5.9. *One may observe that*

$$-Q^{1/2} \log(Q) Q^{1/2} = \mathrm{Exp}_Q^{-1}(I)$$

so that $df_Q(U) = -2\left\langle \mathrm{Exp}_Q^{-1}(I), U \right\rangle_Q$, which also follows from Lemma 5.1.

In the next theorem, we compute the higher covariant derivatives of the squared distance. We write $\{A, B\} := AB + BA$ for the anticommutator of two matrices.

Theorem 5.10. *Let $f(Q) = d(Q, I)^2$, and $U, W \in T_Q\mathrm{PD}(n)$. Set $\tilde{U} = Q^{-1/2}UQ^{-1/2}$ and $\tilde{W} = Q^{-1/2}WQ^{-1/2}$. Then the second derivative of f satisfies*

$$(\nabla^2 f)_Q(U,U) = \int_0^\infty d\lambda\,\mathrm{Tr}\left[\frac{1}{Q+\lambda} U \frac{1}{Q+\lambda} \{Q^{-1}, U\} \right] = \int_0^\infty d\lambda\,\mathrm{Tr}\left[\frac{1}{Q+\lambda} \tilde{U} \frac{1}{Q+\lambda} \{Q, \tilde{U}\} \right],$$

and the third derivative is given by

$$(\nabla^3 f)_Q(W,U,U)$$
$$= \int_0^\infty d\lambda\,\mathrm{Tr}\left[\frac{1}{Q+\lambda} \tilde{U} \frac{1}{Q+\lambda} (\tilde{U}\tilde{W}Q + Q\tilde{W}\tilde{U}) - \frac{1}{Q+\lambda}(\tilde{U}\frac{Q}{Q+\lambda}\tilde{W} + \tilde{W}\frac{Q}{Q+\lambda}\tilde{U})\frac{1}{Q+\lambda}\{\tilde{U}, Q\} \right].$$

Proof. For the second derivative, we use the identity $(\nabla^2 f_Q)(U,U) = \partial_{t=0}^2 f(Q_t)$ where $Q_t = \mathrm{Exp}_Q(tU)$. From Proposition 5.7 it follows that

$$\partial_t f(Q_t) = 2\,\mathrm{Tr}\left[Q_t^{-1} \log(Q_t)(\partial_t Q_t) \right].$$

50

As $Q_t = \mathrm{Exp}_Q(tU) = Q^{1/2}e^{tQ^{-1/2}UQ^{-1/2}}Q^{1/2}$, we have

$$\partial_t Q_t = UQ^{-1/2}e^{tQ^{-1/2}UQ^{-1/2}}Q^{1/2}, \quad \partial_{t=0}^2 Q_t = UQ^{-1}U,$$

which together with Eq. (5.9) leads to

$$\frac{1}{2}\partial_{t=0}^2 f(Q_t) = \mathrm{Tr}\left[(-Q^{-1}UQ^{-1}\log(Q)U) + Q^{-1}(\partial_{t=0}\log(Q_t))U + Q^{-1}\log(Q)(\partial_{t=0}^2 Q_t)\right]$$

$$= \mathrm{Tr}\left[Q^{-1}\int_0^\infty \frac{1}{Q+\lambda}U\frac{1}{Q+\lambda}\,d\lambda\, U\right]$$

$$= \int_0^\infty \mathrm{Tr}\left[\frac{1}{Q+\lambda}U\frac{1}{Q+\lambda}UQ^{-1}\right]d\lambda.$$

To replace the last UQ^{-1} by $\frac{1}{2}\{U, Q^{-1}\}$, note that

$$\mathrm{Tr}[(Q+\lambda)^{-1}U(Q+\lambda)^{-1}UQ^{-1}] = \mathrm{Tr}[(Q+\lambda)^{-1}UQ^{-1}(Q+\lambda)^{-1}U] = \mathrm{Tr}[(Q+\lambda)^{-1}U(Q+\lambda)^{-1}Q^{-1}U]$$

where we first used cyclicity and next that Q^{-1} and $(Q+\lambda)^{-1}$ commute. Using the definition $\tilde{U} = Q^{-1/2}UQ^{-1/2}$ yields the statement in the lemma.

We now turn to the third derivative. Let $U, W \in T_Q\mathrm{PD}(n)$, set $Q_t = \mathrm{Exp}_Q(tW)$ and let $U_t = \tau_{Q\to Q_t}(U)$, explicitly given in Eq. (5.7):

$$U_t = \tau_{Q\to Q_t}(U) = Q^{1/2}(e^{\frac{t}{2}Q^{-1/2}WQ^{-1/2}})Q^{-1/2}UQ^{-1/2}(e^{\frac{t}{2}Q^{-1/2}WQ^{-1/2}})Q^{1/2}.$$

Then

$$(\nabla^3 f)_Q(W, U, U) = \partial_{t=0}(\nabla^2 f)_{Q_t}(U_t, U_t).$$

The two basic derivatives that we need are

$$\partial_{t=0}U_t = \frac{1}{2}(WQ^{-1}U + UQ^{-1}W), \quad \partial_{t=0}Q_t = W.$$

This yields, again using Eq. (5.9),

$$\partial_{t=0}(\nabla^2 f)_{Q_t}(U_t, U_t) = \partial_{t=0}\int_0^\infty \mathrm{Tr}\left[\frac{1}{Q_t+\lambda}U_t\frac{1}{Q_t+\lambda}\{Q_t^{-1}, U_t\}\right]d\lambda$$

$$= \int_0^\infty \mathrm{Tr}\left[-\frac{1}{Q+\lambda}W\frac{1}{Q+\lambda}U\frac{1}{Q+\lambda}\{Q^{-1}, U\}\right]$$

$$+ \frac{1}{2}\mathrm{Tr}\left[\frac{1}{Q+\lambda}(WQ^{-1}U + UQ^{-1}W)\frac{1}{Q+\lambda}\{Q^{-1}, U\}\right]$$

$$+ \mathrm{Tr}\left[-\frac{1}{Q+\lambda}U\frac{1}{Q+\lambda}W\frac{1}{Q+\lambda}\{Q^{-1}, U\}\right]$$

$$+ \mathrm{Tr}\left[\frac{1}{Q+\lambda}U\frac{1}{Q+\lambda}\{-Q^{-1}WQ^{-1}, U\}\right]$$

$$+ \frac{1}{2}\mathrm{Tr}\left[\frac{1}{Q+\lambda}U\frac{1}{Q+\lambda}\{Q^{-1}, WQ^{-1}U + UQ^{-1}W\}\right]d\lambda$$

$$= \int_0^\infty \mathrm{Tr}\left[\frac{1}{Q+\lambda}U\frac{1}{Q+\lambda}WQ^{-1}UQ^{-1} + \frac{1}{Q+\lambda}UQ^{-1}W\frac{1}{Q+\lambda}UQ^{-1}\right]$$

$$- \frac{1}{Q+\lambda}(W\frac{1}{Q+\lambda}U + U\frac{1}{Q+\lambda}W)\frac{1}{Q+\lambda}\{Q^{-1}, U\}\Big]d\lambda.$$

Substituting $W = Q^{1/2}\tilde{W}Q^{1/2}$ and $U = Q^{1/2}\tilde{U}Q^{1/2}$ yields the desired expression. $\qquad\square$

We now explicitly compute the integral expressions in Theorem 5.10 in terms of the entries of the matrices $\tilde{U}$ and $\tilde{W}$. We assume without loss of generality that $Q = \mathrm{diag}(q_1, \ldots, q_n)$ by considering the expression in an eigenbasis of Q. Furthermore, we shall assume that all q_i are distinct; expressions at general Q may be obtained by taking limits, but the inequalities we will derive automatically hold for all Q by continuity. Let us start with the second derivative. Take $U \in \mathrm{Herm}(n)$. Then for $\tilde{U} = Q^{-1/2}UQ^{-1/2}$ we have

$$
\begin{aligned}
(\nabla^2 f)_Q(U, U) &= \int_0^\infty d\lambda \, \mathrm{Tr}\left[\frac{1}{Q+\lambda}\tilde{U}\frac{1}{Q+\lambda}\{\tilde{U}, Q\}\right] \\
&= \sum_{k,l}\int_0^\infty d\lambda \frac{1}{q_k+\lambda}\tilde{U}_{kl}\frac{1}{q_l+\lambda}\tilde{U}_{lk}(q_k+q_l) \\
&= 2\sum_k |\tilde{U}_{kk}|^2 + \sum_{k\neq l}|\tilde{U}_{kl}|^2 \frac{(q_k+q_l)\log(q_k/q_l)}{q_k-q_l}.
\end{aligned}
\tag{5.10}
$$

where we evaluated the integral using the identities

$$
\int_0^\infty \frac{1}{(x+\lambda)^2}\,d\lambda = \frac{1}{x}, \quad \int_0^\infty \frac{1}{(x+\lambda)(y+\lambda)}\,d\lambda = \frac{\log(x/y)}{x-y}
\tag{5.11}
$$

for distinct $x, y > 0$. We now evaluate the third derivative in a similar manner. The only new difficulty is in performing the integration with respect to λ, for which we record the following lemma.

Lemma 5.11. *For distinct $x, y, z > 0$, one has*

$$
\int_0^\infty \frac{1}{(x+\lambda)(y+\lambda)(z+\lambda)}d\lambda = \frac{z(\log(x)-\log(y)) + y(\log(z)-\log(x)) + x(\log(y)-\log(z))}{(x-y)(y-z)(x-z)}.
$$

Proof. One can deduce from a partial fraction decomposition that

$$
\frac{(x-y)(y-z)(x-z)}{(x+\lambda)(y+\lambda)(z+\lambda)} = \frac{y-z}{x+\lambda} + \frac{z-x}{y+\lambda} + \frac{x-y}{z+\lambda},
$$

and the latter integrates to

$$
\begin{aligned}
-\int_0^\infty \frac{y-z}{x+\lambda} + \frac{z-x}{y+\lambda} + \frac{x-y}{z+\lambda}\,d\lambda &= \int_0^\infty (y-z)\left(\frac{1}{1+\lambda} - \frac{1}{x+\lambda}\right)d\lambda \\
&\quad + \int_0^\infty (z-x)\left(\frac{1}{1+\lambda} - \frac{1}{y+\lambda}\right)d\lambda \\
&\quad + \int_0^\infty (x-y)\left(\frac{1}{1+\lambda} - \frac{1}{z+\lambda}\right)d\lambda \\
&= (y-z)\log(x) + (z-x)\log(y) + (x-y)\log(z). \qquad \square
\end{aligned}
$$

For convenience we will use the following notation. Define $H\colon \mathbb{R}_{>0}^2 \to \mathbb{R}$ by

$$
H(x, y) = \frac{(x+y)\log(x/y)}{x-y},
\tag{5.12}
$$

if $x, y > 0$ are distinct, and

$$
H(x, x) = 2.
\tag{5.13}
$$

Next, we define $T \colon \mathbb{R}^3_{>0} \to \mathbb{R}$ by

$$T(x,y,z) = \frac{x+y}{x-y}\left(\frac{x+z}{x-z}\log(x/z) - \frac{y+z}{y-z}\log(y/z)\right), \tag{5.14}$$

for distinct $x,y,z > 0$. Then T extends to a continuous function on $\mathbb{R}^3_{>0}$, such that

$$T(x,x,z) = \frac{2x^2 - 2z^2 - 4xz\log(x/z)}{(x-z)^2},$$

$$T(x,y,x) = \frac{2x^2 - 2y^2 - (x+y)^2\log(x/y)}{(x-y)^2}, \tag{5.15}$$

$$T(x,x,x) = 0.$$

Furthermore, $T(x,y,z)$ is symmetric in x and y, for every $c > 0$ satisfies $T(cx,cy,cz) = T(x,y,z)$, and $T(x^{-1}, y^{-1}, z^{-1}) = -T(x,y,z)$. Then we have the following proposition.

Proposition 5.12. *Let $f(Q) = d(Q,I)^2$ and $U, W \in T_Q\mathrm{PD}(n)$. Then for $Q = \mathrm{diag}(q_1, \ldots, q_n)$, and $\tilde{U} = Q^{-1/2}UQ^{-1/2}$, $\tilde{W} = Q^{-1/2}WQ^{-1/2}$, one has*

$$(\nabla^2 f)_Q(U,U) = \sum_{k,l=1}^{n} |\tilde{U}_{kl}|^2 H(q_k, q_l), \tag{5.16}$$

$$(\nabla^3 f)_Q(W,U,U) = \sum_{k,l,m=1}^{n} \tilde{W}_{kl}\tilde{U}_{lm}\tilde{U}_{mk}T(q_k, q_l, q_m)$$

where $H \colon \mathbb{R}^2_{>0} \to \mathbb{R}$ and $T \colon \mathbb{R}^3_{>0} \to \mathbb{R}$ are defined in Eqs. (5.12) to (5.15), and the subscripts refer to the respective matrix entries.

Proof. The formula for the Hessian of f was already derived in Eq. (5.10). For the third derivative, one can evaluate the trace in Theorem 5.10 as

$$\mathrm{Tr}\left[\tilde{W}Q(Q+\lambda)^{-1}\tilde{U}(Q+\lambda)^{-1}\tilde{U}\right] = \sum_{k,l,m} \tilde{W}_{kl}\frac{q_l}{q_l+\lambda}\tilde{U}_{lm}\frac{1}{q_m+\lambda}\tilde{U}_{mk},$$

$$\mathrm{Tr}\left[\tilde{W}\tilde{U}(Q+\lambda)^{-1}\tilde{U}(Q+\lambda)^{-1}Q\right] = \sum_{k,l,m} \tilde{W}_{kl}\tilde{U}_{lm}\frac{1}{q_m+\lambda}\tilde{U}_{mk}\frac{q_k}{q_k+\lambda},$$

$$\mathrm{Tr}\left[\tilde{W}(Q+\lambda)^{-1}\{\tilde{U},Q\}(Q+\lambda)^{-1}\tilde{U}(Q+\lambda)^{-1}Q\right] = \sum_{k,l,m} \tilde{W}_{kl}\tilde{U}_{lm}\frac{q_l+q_m}{(q_l+\lambda)(q_m+\lambda)}\tilde{U}_{mk}\frac{q_k}{q_k+\lambda},$$

$$\mathrm{Tr}\left[\tilde{W}Q(Q+\lambda)^{-1}\tilde{U}(Q+\lambda)^{-1}\{\tilde{U},Q\}(Q+\lambda)^{-1}\right] = \sum_{k,l,m} \tilde{W}_{kl}\frac{q_l}{q_l+\lambda}\tilde{U}_{lm}\frac{q_m+q_k}{(q_m+\lambda)(q_k+\lambda)}\tilde{U}_{mk},$$

so that the third derivative satisfies

$(\nabla^3 f)_Q(W,U,U)$

$$= \int_0^\infty d\lambda \sum_{k,l,m} \tilde{W}_{kl}\tilde{U}_{lm}\tilde{U}_{mk}\left(\frac{q_k}{(q_k+\lambda)(q_m+\lambda)}\left(1 - \frac{q_l+q_m}{q_l+\lambda}\right) + \frac{q_l}{(q_l+\lambda)(q_m+\lambda)}\left(1 - \frac{q_k+q_m}{q_k+\lambda}\right)\right).$$

Using Eq. (5.11) and Lemma 5.11, this integrates to (interpreting expressions as limits whenever not all q_k, q_l, q_m are distinct)

$$\sum_{k,l,m} \tilde{W}_{kl}\tilde{U}_{lm}\tilde{U}_{mk}\left(\frac{q_k\log(q_k/q_m)}{q_k-q_m} + \frac{q_l\log(q_l/q_m)}{q_l-q_m}\right)$$

$$-\sum_{k,l,m} \tilde{W}_{kl}\tilde{U}_{lm}\tilde{U}_{mk}\frac{(q_k(q_l+q_m)+q_l(q_k+q_m))(q_m\log(q_k/q_l)+q_l\log(q_m/q_k)+q_k\log(q_l/q_m))}{(q_k-q_l)(q_l-q_m)(q_k-q_m)}$$

$$=\sum_{k,l,m} \tilde{W}_{kl}\tilde{U}_{lm}\tilde{U}_{mk}\frac{q_k+q_l}{q_k-q_l}\left(\frac{(q_k+q_m)(q_l-q_m)\log(q_k/q_m)-(q_l+q_m)(q_k-q_m)\log(q_l/q_m)}{(q_l-q_m)(q_k-q_m)}\right)$$

$$=\sum_{k,l,m} \tilde{W}_{kl}\tilde{U}_{lm}\tilde{U}_{mk}\frac{q_k+q_l}{q_k-q_l}\left(\frac{q_k+q_m}{q_k-q_m}\log(q_k/q_m)-\frac{q_l+q_m}{q_l-q_m}\log(q_l/q_m)\right)$$

$$=\sum_{k,l,m} \tilde{W}_{kl}\tilde{U}_{lm}\tilde{U}_{mk}T(q_k,q_l,q_m),$$

which is exactly the desired expression for the third derivative. $\square$

We note here that Proposition 5.12 can be used to verify that the squared distance is 2-strongly convex, which is a general property of Hadamard manifolds as mentioned before. Indeed, $\|U\|_Q = \|\tilde{U}\|_{\mathrm{HS}}$ by definition of the Riemannian metric, so one has to show that $(\nabla^2 f)_Q(U,U) \geq 2\|\tilde{U}\|_{\mathrm{HS}}^2$. In view of Eq. (5.16), it suffices to prove that $H(x,y) \geq 2$. This follows directly from the logarithmic-arithmetic mean inequality: for every $x, y > 0$, one has

$$\frac{x-y}{\log(x)-\log(y)} \leq \frac{x+y}{2}, \tag{5.17}$$

where the quantity $(x-y)/(\log(x)-\log(y))$ is known as the *logarithmic mean* of x and y (it is defined as x when $x = y$). It is known to be inbetween the geometric and arithmetic mean of x and y [Car72]. A short proof of Eq. (5.17) is as follows. Assume without loss of generality that $x < y$; then the lower bound of the Hermite–Hadamard inequality applied to the function $z \mapsto 1/z$ yields

$$\frac{\log(y)-\log(x)}{y-x} = \frac{1}{y-x}\int_x^y \frac{1}{z}\,dz \geq \left(\frac{x+y}{2}\right)^{-1}.$$

One can also reverse this strategy: $\mathrm{PD}(n)$ is a Hadamard manifold, hence the squared distance is 2-strongly convex, which in turn implies the logarithmic-arithmetic mean inequality. It would be interesting to understand whether there is a more direct relation between the logarithmic-arithmetic mean inequality and the 2-strong-convexity of the squared distance, for instance via midpoint-strong-convexity considerations.

We now study the coefficients appearing in Proposition 5.12 to show that the squared distance is self-concordant on $\mathrm{PD}(n)$. Let $a = \log(q_k/q_m)$ and $b = \log(q_l/q_m)$. Then

$$T(q_k,q_l,q_m) = \coth((a-b)/2)\left(a\coth(a/2) - b\coth(b/2)\right),$$

whereas the square root of the product of the coefficients of $|\tilde{W}_{kl}|^2$, $|\tilde{U}_{lm}|^2$, and $|\tilde{U}_{mk}|^2$ in $\nabla^2 f$ is

$$\sqrt{H(q_k,q_l)H(q_l,q_m)H(q_k,q_m)} = \sqrt{ab(a-b)\coth(a/2)\coth(b/2)\coth((a-b)/2)}.$$

Lemma 5.13. *The constant $C = \sqrt{2}$ is such that for all $a, b \in \mathbb{R}$, one has*

$$|\coth((a-b)/2)\left(a\coth(a/2) - b\coth(b/2)\right)| \leq C\sqrt{ab(a-b)\coth(a/2)\coth(b/2)\coth((a-b)/2)}.$$

As a consequence, for all $x, y, z > 0$, we have

$$|T(x,y,z)| \leq C\sqrt{H(x,y)H(y,z)H(x,z)}. \tag{5.18}$$

Remark 5.14. *We conjecture, based on numerical evidence, that the optimal constant in the above inequality is $C = 1/\sqrt{2}$. Let $A(x,y) = (x+y)/2$ and $G(x,y) = \sqrt{xy}$ be the arithmetic and geometric mean, respectively. The inequality for $C = 1/\sqrt{2}$ is equivalent to the following "reverse arithmetic-geometric mean inequality": for all $a, b \in \mathbb{R}$,*

$$\frac{A(a^2 \coth(a)^2, b^2 \coth(b)^2)}{G(a^2 \coth(a)^2, b^2 \coth(b)^2)} \leq 1 + \frac{(a-b)\tanh(a-b)}{2}.$$

Proof of Lemma 5.13. Consider $h(x) = x \coth(x/2)$. Then h is 1-Lipschitz: its derivative is given by

$$\partial_x h(x) = \frac{\sinh(x) - x}{\cosh(x) - 1}.$$

It is clear that $|\sinh(x) - x| \leq \cosh(x) - 1$: for $x \geq 0$, the difference is $\cosh(x) - 1 - (\sinh(x) - x) = x + e^{-x} - 1$, which is convex and has zero derivative at $x = 0$, where it evaluates to 0. For $x \leq 0$, the difference is $\cosh(x) - 1 + \sinh(x) - x = e^x - x - 1 \geq 0$.

We rewrite the left- and right-hand sides of the inequality:

$$\coth((a-b)/2)\,(a \coth(a/2) - b \coth(b/2)) = \frac{h(a-b)(h(a) - h(b))}{a - b}$$

and

$$\sqrt{ab(a-b) \coth(a/2) \coth(b/2) \coth((a-b)/2)} = \sqrt{h(a)h(b)h(a-b)}.$$

Therefore it suffices to prove that

$$\left| \frac{h(a) - h(b)}{a - b} \right| \leq C \sqrt{\frac{h(a)h(b)}{h(a-b)}}.$$

Because h is 1-Lipschitz, the left-hand side is at most 1.

We now claim that the following lower- and upper bounds on h hold: $h(x) \geq 1 + \frac{|x|}{2}$, and $h(x) \leq 2 + |x|$. The upper bound follows from h being 1-Lipschitz and $h(0) = 2$. For the lower bound, we restrict to $x \geq 0$, in which case it suffices to prove $x \cosh(x/2) \geq (1 + x/2) \sinh(x/2)$. This is simple: we have $x \cosh(x/2) \geq 2 \sinh(x/2)$ (by a power series comparison for $x \cosh(x)$ and $\sinh(x)$), and $x \cosh(x/2) \geq x \sinh(x/2)$ since $\cosh(x/2) \geq \sinh(x/2)$. Therefore $x \cosh(x/2)$ is greater than their average.

We now finish up the argument: we have

$$\frac{h(a)h(b)}{h(a-b)} \geq \frac{1 + \frac{|a|+|b|}{2} + \frac{|ab|}{4}}{2 + |a| + |b|} \geq \frac{1}{2},$$

so we conclude that

$$C\sqrt{\frac{h(a)h(b)}{h(a-b)}} \geq \frac{C}{\sqrt{2}} \geq 1 \geq \frac{h(a) - h(b)}{a - b}.$$

holds for $C = \sqrt{2}$. $\square$

This directly implies that the squared distance is self-concordant (with an n-independent constant), hence also proving Theorem 1.4.

Theorem 5.15. *Let $C \geq 0$ be such that the inequality in [Lemma 5.13](#) holds. Then the function $f\colon \mathrm{PD}(n) \to \mathbb{R}$ defined by $f(Q) = d(Q,I)^2$ satisfies for $Q \in \mathrm{PD}(n)$ and $U, W \in T_Q\mathrm{PD}(n)$ the inequality*

$$\left|(\nabla^3 f)_Q(W,U,U)\right| \leq C\sqrt{(\nabla^2 f)_Q(W,W)\,(\nabla^2 f)_Q(U,U)}$$

In particular, from the choice $C = \sqrt{2}$ it follows that f is 2-self-concordant.

Proof. By [Eq. (5.18)](#) and consecutive applications of Cauchy–Schwarz, we have

$$
\begin{aligned}
&\left|(\nabla^3 f)_Q(W,U,U)\right| \\
&\leq \sum_{k,l,m} |\tilde{W}_{kl}\tilde{U}_{lm}\tilde{U}_{mk}||T(q_k,q_l,q_m)| \\
&\leq C \sum_{k,l,m} |\tilde{W}_{kl}\tilde{U}_{lm}\tilde{U}_{mk}|\sqrt{H(q_k,q_l)H(q_l,q_m)H(q_k,q_m)} \\
&\leq C \sqrt{\sum_{k,l}|\tilde{W}_{kl}|^2 H(q_k,q_l)} \sqrt{\sum_{k,l}\left(\sum_m |\tilde{U}_{lm}\tilde{U}_{mk}|\sqrt{H(q_l,q_m)H(q_k,q_m)}\right)^2} \\
&\leq C \sqrt{\sum_{k,l}|\tilde{W}_{kl}|^2 H(q_k,q_l)} \sqrt{\sum_{k,l}\left(\sum_m |\tilde{U}_{lm}|^2 H(q_l,q_m)\right)\left(\sum_m |\tilde{U}_{mk}|^2 H(q_k,q_m)\right)} \\
&= C \sqrt{\sum_{k,l}|\tilde{W}_{kl}|^2 H(q_k,q_l)} \sqrt{\left(\sum_{l,m}|\tilde{U}_{lm}|^2 H(q_l,q_m)\right)^2} \\
&= C\sqrt{(\nabla^2 f)_Q(W,W)(\nabla^2 f)_Q(U,U)}. \qquad\qquad \square
\end{aligned}
$$

One can use this to construct a strongly self-concordant function on the open epigraph of the squared distance using [Theorem 4.11](#), hence also proving [Theorem 1.6](#). By imposing an additional upper bound on the value of the squared distance one can use this to construct a barrier for the epigraph, albeit with a distance-dependent barrier parameter; see [Section 6](#) for similar constructions.

5.3 Constant negative curvature

In this subsection, we prove that the squared distance on n-dimensional hyperbolic space $\mathbb{H}^n$ is self-concordant with a larger self-concordance parameter, and other refinements of the self-concordance estimate. We use this to construct a barrier for the epigraph of the (squared) distance in [Theorem 5.22](#), which is useful for our applications in [Section 6](#). Instead of dealing just with $\mathbb{H}^n$, we consider, more generally, the *model spaces* $M^n_{-\kappa}$ with constant sectional curvature $-\kappa < 0$ (we recall that $\mathbb{H}^n$ is M^n_{-1}). The main result of this subsection is the following.

Theorem 5.16. *Let $n \geq 2$, $\kappa > 0$, set $M = M^n_{-\kappa}$, let $p_0 \in M$, and consider $f, g\colon M \to \mathbb{R}$ defined by $f(p) = d(p,p_0)^2$ and $g(p) = d(p,p_0)$. One has the following estimates:*

   (i) $|(\nabla^3 f)_p(w,u,u)| \leq \sqrt{\dfrac{\kappa}{2}}\sqrt{(\nabla^2 f)_p(w,w)(\nabla^2 f)_p(u,u)}$, so f is $\frac{8}{\kappa}$-self-concordant, and this constant cannot be improved.

(ii) $|(\nabla^3 f)_p(u,u,u)| \leq \sqrt{\dfrac{8\kappa}{27}}((\nabla^2 f)_p(u,u))^{3/2}$, so f is $\frac{27}{2\kappa}$-self-concordant along geodesics, and this constant cannot be improved.

(iii) $|(\nabla^3 f)_p(w,u,u)| \leq 2\zeta\sqrt{\kappa}|dg_p(w)|((\nabla^2 f)_p(u,u) - 2dg_p(u)^2)$

$$+ 2\sqrt{\kappa}|dg_p(u)|\sqrt{(\nabla^2 f)_p(u,u) - 2dg_p(u)^2}\sqrt{(\nabla^2 f)_p(w,w) - 2dg_p(w)^2}$$

$$\leq 2\zeta\sqrt{\kappa}\|w\|_p(\nabla^2 f)_p(u,u) + 2\sqrt{\kappa}\|u\|_p\sqrt{(\nabla^2 f)_p(u,u)}\sqrt{(\nabla^2 f)_p(w,w)},$$

where $\zeta = \sup_{x\in\mathbb{R}}|\sinh(x)^{-1} - x^{-1}| \leq \frac{1}{2}$.

By Lemmas 2.1 and 3.2 it suffices to prove the above estimates for $M = M^n_{-1}$ and then to appropriately rescale the estimate when the curvature changes. The estimate in (iii) is a refinement of self-concordance for f (albeit with different constants), because $2\|W\|^2_Q \leq \|W\|^2_{f,Q}$ by the 2-strong-convexity of f (and in the presence of curvature, these norms can differ by a factor that scales with the distance to the base point and the curvature). The estimate also implies that, in the terminology of Section 4.2, the squared distance is compatible with *every* strongly convex function, which is relevant for computing geometric means on $M^n_{-\kappa}$ as discussed in Section 6.4. The presence of the "correction terms" $-2dg_p(u)^2$ and similar for w will also be useful for proving Theorem 5.22, which we use later for the purpose of computing geometric medians.

Before starting with the proof of Theorem 5.16, we provide estimates on some single-variable functions which we use.

Lemma 5.17. (i) Define $\Phi\colon \mathbb{R} \to \mathbb{R}$ by

$$\Phi(x) := \partial_x(x\coth(x)) = \coth(x) + x - x\coth(x)^2, \quad x \neq 0, \tag{5.19}$$

and $\Phi(0) = 0$. Then Φ is smooth, and for $x \in \mathbb{R}_{\geq 0}$, it holds that

$$0 \leq \Phi(x) \leq \min(x,1), \tag{5.20}$$

and $\lim_{x\to\infty}\Phi(x) = 1$.

(ii) It holds that

$$\zeta := \sup_{x\in\mathbb{R}_{\geq 0}}\frac{\Phi(x)}{2x\coth(x)} = \sup_{x\in\mathbb{R}}\left|\frac{1}{\sinh(x)} - \frac{1}{x}\right| < \frac{1}{2}. \tag{5.21}$$

We note here that numerical evaluation suggests the value of ζ is approximately 0.23536, which is slightly smaller than $\frac{1}{3\sqrt{2}} \approx 0.23570$.

Proof. We first prove (i). By $\sinh(x) = x + (1/3!)x^3 + \cdots$ and $\cosh(x) = 1 + (1/2!)x^2 + \cdots$, and by the identities $\cosh(x)^2 - \sinh(x)^2 = 1$, $2\cosh(x)\sinh(x) = \sinh(2x)$, and $2\sinh(x)^2 = \cosh(2x) - 1$, it holds that

$$\Phi(x) = \frac{\cosh(x)}{\sinh(x)} + x(1 - \coth(x)^2) = \frac{\sinh(x)\cosh(x) - x}{\sinh(x)^2} = \frac{\sinh(2x) - 2x}{\cosh(2x) - 1} = \frac{(2x)^3/3! + \cdots}{(2x)^2/2! + \cdots}.$$

From this, we deduce that $\Phi(x) \geq 0$ for $x \geq 0$, and

$$\lim_{x\to 0}\Phi(x) = 0 = \Phi(0).$$

Therefore Φ is continuous at 0. The above argument shows that Φ is a ratio of the analytic functions $\sinh(2x) - 2x$ and $\cosh(2x) - 1$, and the continuity at 0 shows that Φ has no singularity at 0, which is the only zero of $\cosh(2x) - 1$; hence Φ must in fact be smooth on $\mathbb{R}$.

We now show that $\Phi(x) \le \min(x, 1)$ for $x \ge 0$. We have

$$\lim_{x \to 0} x \coth(x) = \lim_{x \to 0} \frac{x(1 + x^2/2! + \cdots)}{x + x^3/3! + \cdots} = 1.$$

By $\partial_x(x \coth x) = \Phi(x) \ge 0$ for $x \ge 0$, we have

$$x \coth(x) \ge 1.$$

This implies that Φ is nondecreasing, since

$$\partial_x \Phi(x) = \frac{2(x \coth(x) - 1)}{\sinh(x)^2} \ge 0.$$

Thus we have

$$\sup_{x \in [0, \infty)} \Phi(x) = \lim_{x \to \infty} \Phi(x) = \lim_{x \to \infty} \coth x - x/\sinh^2 x = 1.$$

Lastly, $\Phi(x) \le x$ follows from

$$x - \Phi(x) = \coth(x) (x \coth(x) - 1) \ge 0.$$

We now prove (ii). Observe that $\lim_{x \to 0} \sinh(x)^{-1} - x^{-1} = 0$ by two applications of L'Hôpital's rule, so $\sinh(x)^{-1} - x^{-1}$ has a continuous extension to all of $\mathbb{R}$. A similar argument shows that $\coth(x) - x^{-1}$ can be continuously extended to $x = 0$ with value 0. For both inequalities it suffices to treat the case $x > 0$. The inequality $|\sinh(x)^{-1} - x^{-1}| \le \frac{1}{2}$ is equivalent to

$$|x - \sinh(x)| = \sinh(x) - x \le \frac{x \sinh(x)}{2}.$$

We have equality for $x = 0$, and

$$\partial_x(\sinh(x) - x) = \cosh(x) - 1, \quad \partial_x \sinh(x) = \sinh(x) + x \cosh(x)$$

agree for $x = 0$ as well. Differentiating once more yields

$$\partial_x^2(\sinh(x) - x) = \sinh(x), \quad \partial_x^2(x \sinh(x)) = 2 \cosh(x) + x \sinh(x).$$

Clearly, $\frac{1}{2}(2 \cosh(x) + x \sinh(x)) \ge \cosh(x) \ge \sinh(x)$, and so we have proven $\zeta \le \frac{1}{2}$. $\qquad \square$

Although there are several models of $M^n_{-\kappa}$ in which explicit computations can be performed (such as $\mathrm{SPD}(2, \mathbb{C})$, which is $M^3_{-1/2}$), for proving Theorem 5.16, we take a "model-free" approach based on Jacobi fields. For a geodesic $\gamma \colon [0, l] \to M$, a *Jacobi field* along γ is a vector field $X = (X(t))_{t \in [0, l]}$ along γ, where $X(t) \in T_{\gamma(t)} M$ satisfies the Jacobi equation:[10]

$$\nabla_{\dot\gamma(t)} \nabla_{\dot\gamma(t)} X(t) + R(X(t), \dot\gamma(t)) \dot\gamma(t) = 0, \quad t \in [0, l]. \tag{5.22}$$

This is a linear differential equation. Therefore, the solution $X(t)$ is uniquely determined by the initial values $X(0), \nabla_{\dot\gamma(0)} X(0)$, or by its boundary values $X(0), X(l)$. Jacobi fields are relevant to the task of differentiating the squared distance because they arise variation fields of geodesics: the distance $d(p_0, p)$ is the minimal length of a geodesic between p_0 and p, and varying p leads to a *family* of geodesics. More precisely, one has the following classical result:

---

[10]The meaning of $\nabla_{\dot\gamma(t)}$ here is slightly different from its previous meaning: instead of acting on tensor fields on an open subset of M, it acts on tensor fields along the curve γ. The two notions they agree whenever $X(t)$ is locally the restriction of a vector field on M, see [Lee18, Ch. 4] for more information.

Lemma 5.18 (see [Sak96, p.35, 36]). *Let $\alpha\colon [0,l] \times (-\epsilon,\epsilon) \to M$ be a smooth map such that the curve $t \mapsto \alpha(t,s)$ is a geodesic for each $s \in (-\epsilon,\epsilon)$. Then $d\alpha(t,0)(\frac{\partial}{\partial s})$ is a Jacobi field along geodesic $t \mapsto \alpha(t,0)$.*

It can also be shown that every Jacobi field (along a geodesic on a compact interval) arises in this way [Lee18, Prop. 10.4], but we will not need this fact. The derivative and the Hessian of $p \mapsto f(p) = d(p,p_0)^2$ can be determined using Jacobi fields as follows.

Lemma 5.19 (see [Sak96, p.108–110]). *Let $p, p_0 \in M$ be distinct points, let $\gamma\colon [0,l] \to M$ be the unique unit-speed geodesic with $\gamma(0) = p_0$, $\gamma(l) = p$, and $l := g(p) = d(p,p_0)$. For $u \in T_pM$, it holds that:*

(i) $dg_p(u) = \langle \dot{\gamma}(l), u \rangle_p,$

(ii) $df_p(u) = 2l\langle \dot{\gamma}(l), u \rangle_p,$ *and*

(iii) $(\nabla^2 f)_p(u,u) = 2l \langle \nabla_{\dot{\gamma}(l)} X(l), u \rangle_p,$ *where X is the Jacobi field along γ under the boundary condition*

$$X(0) = 0, \quad X(l) = u.$$

Note that (i) and (ii) are reformulations of Eq. (5.1), and in light of Eq. (5.2), (iii) is essentially a claim about $(\nabla^2 g)_p$.

We shall use the following fact about spaces of constant curvature $-\kappa$ [Sak96, Lem. II.3.3]: their Riemann curvature tensor R satisfies

$$R(X,Y)Z = -\kappa(\langle Y,Z \rangle X - \langle X,Z \rangle Y), \tag{5.23}$$

where we recall that $\langle \cdot, \cdot \rangle$ is the Riemannian metric. This allows one to explicitly write down the solutions of the Jacobi equation, as given in the following lemma. While this, and explicit expressions for the Hessian of the (squared) distance are well-known (see e.g. [Sak96, p. 136, p. 154] or [Lee18, Prop. 10.12, Prop. 11.3]), we provide a proof for completeness.

Lemma 5.20. *Let $p, p_0 \in M = \mathbb{H}^n$ with $p \neq p_0$, and let $\gamma\colon [0,l] \to M$ be the unit-speed geodesic from p_0 to p with $l := g(p) = d(p,p_0)$. Let $u \in T_pM$ and decompose $u = u^\top + u^\perp$ such that $u^\top = \langle u, \dot{\gamma}(l) \rangle_p \dot{\gamma}(l)$ is the part of u parallel to $\dot{\gamma}(l)$, and $u^\perp$ orthogonal to $\dot{\gamma}$, i.e., $\langle u^\perp, \dot{\gamma}(l) \rangle_p = 0$. Then the unique Jacobi field $X(t)$ along γ with $X(0) = 0$ and $X(l) = u$ satisfies*

$$X(t) = \frac{t}{l}\tau_{\gamma,t-l}u^\top + \frac{\sinh(t)}{\sinh(l)}\tau_{\gamma,t-l}u^\perp,$$

where $\tau_{\gamma,t-l}\colon T_{\gamma(l)}M \to T_{\gamma(t)}M$ is the parallel transport along γ.

Proof. It is clear that $X(l) = u$ and $X(0) = 0$. Therefore it remains to check that X is a Jacobi field: we have

$$\nabla_{\dot{\gamma}(t)} X(t) = \frac{1}{l}\tau_{\gamma,t-l}u^\top + \frac{\cosh(t)}{\sinh(l)}\tau_{\gamma,t-l}u^\perp$$

and

$$\nabla_{\dot{\gamma}(t)}\nabla_{\dot{\gamma}(t)} X(t) = \frac{\sinh(t)}{\sinh(l)}\tau_{\gamma,t-l}u^\perp.$$

From Eq. (5.23) it follows that

$$R(X(t), \dot{\gamma}(t))\dot{\gamma}(t) = -[X(t) - \langle X(t), \dot{\gamma}(t) \rangle_{\gamma(t)} \dot{\gamma}(t)].$$

59

Therefore

$$\nabla_{\dot\gamma(t)}\nabla_{\dot\gamma(t)}X(t) + R(X(t),\dot\gamma(t))\dot\gamma(t) = \frac{\sinh(t)}{\sinh(l)}\tau_{\gamma,t-l}u^\perp - X(t) + \langle X(t),\dot\gamma(t)\rangle_{\gamma(t)}\dot\gamma(t)$$

$$= -\frac{t}{l}\tau_{\gamma,t-l}u^\top + \langle X(t),\dot\gamma(t)\rangle_{\gamma(t)}\dot\gamma(t)$$

$$= -\frac{t}{l}\langle u,\dot\gamma(l)\rangle_p\,\dot\gamma(t) + \langle X(t),\dot\gamma(t)\rangle_{\gamma(t)}\dot\gamma(t)$$

$$= 0,$$

where the penultimate equality follows from $u^\top = \langle u,\dot\gamma(l)\rangle_{\gamma(l)}\dot\gamma(l)$ and $\tau_{\gamma,t-l}\dot\gamma(l) = \dot\gamma(t)$, and the last equality follows from $\tau_{\gamma,t-l}$ being an isometry and $\langle u,\dot\gamma(l)\rangle = \langle u^\top,\dot\gamma(l)\rangle$. $\qquad\square$

Using this description of the Jacobi fields leads to the following description of the Hessian, and the third covariant derivative of the squared distance.

Proposition 5.21. *Let $p,p_0 \in M = \mathbb{H}^n$ with $p \neq p_0$ and let $\gamma\colon [0,l] \to M$ be the unique geodesic from p_0 to p with $l := g(p) = d(p,p_0)$. Then $f(p) = d(p,p_0)^2$ satisfies*

$$(\nabla^2 f)_p(u,u) = 2(l\coth l)\left(\langle u,u\rangle_p - \langle u,\dot\gamma(l)\rangle_p^2\right) + \langle u,\dot\gamma(l)\rangle_p^2, \tag{5.24}$$

$$(\nabla^3 f)_p(w,u,u) = 2\Phi(l)\langle w,\dot\gamma(l)\rangle_p\left(\langle u,u\rangle_p - \langle u,\dot\gamma(l)\rangle_p^2\right)$$
$$\qquad + 4\left(l - \Phi(l)\right)\langle u,\dot\gamma(l)\rangle_p\left(\langle w,\dot\gamma(l)\rangle_p\langle u,\dot\gamma(l)\rangle_p - \langle u,w\rangle_p\right). \tag{5.25}$$

Proof. By Lemma 5.20, the Jacobi field $X(t)$ along γ with $X(0) = 0$ and $X(l) = u$ satisfies

$$X(t) = \frac{t}{l}\tau_{\gamma,t-l}u^\top + \frac{\sinh(t)}{\sinh(l)}\tau_{\gamma,t-l}u^\perp$$

where $u = u^\top + u^\perp$ is a decomposition with $u^\top = \langle u,\dot\gamma(l)\rangle_p\,\dot\gamma(l)$ parallel and $u^\perp = u - u^\top$ orthogonal to $\dot\gamma(l)$, respectively. Therefore

$$\nabla_{\dot\gamma(l)}X(l) = \frac{1}{l}u^\top + \frac{\cosh(l)}{\sinh(l)}u^\perp = \frac{1}{l}\langle u,\dot\gamma(l)\rangle_p\,\dot\gamma(l) + \frac{\cosh(l)}{\sinh(l)}(u - u^\top) \tag{5.26}$$

Now apply Lemma 5.19(iii) to obtain Eq. (5.24).

Consider the geodesic $s \mapsto c(s) := \mathrm{Exp}_p(sw)$. Let $\gamma_s\colon [0,l] \to M$ be the geodesic from p to $c(s)$ (not necessarily parametrized by the arc-length). For $s \in (-\epsilon,\epsilon)$, let $l_s := d(c(s),p_0)$ and $u_s := \tau_{c,s}u$. Applying Eq. (5.24) to the reparametrized geodesic $t \mapsto \gamma_s((l/l_s)t)$ $(t \in [0,l_s])$, we obtain

$$(\nabla^2 f)_{c(s)}(u_s,u_s) = 2(l_s\coth(l_s))\langle u_s,u_s\rangle + 2\left(1 - l_s\coth(l_s)\right)(l/l_s)^2\langle u_s,\dot\gamma_s(l)\rangle^2. \tag{5.27}$$

By Eq. (2.3), the covariant derivative $(\nabla^3 f)_p(w,u,u)$ is obtained by computing the s-derivative of Eq. (5.27) at $s = 0$. We use that

$$\partial_{s=0}l_s = \langle\dot\gamma(l),w\rangle, \quad \partial_{s=0}\langle u_s,u_s\rangle = 0, \quad \partial_{s=0}\langle u_s,\dot\gamma_s(l)\rangle = \langle u,\nabla_{\dot c(s)}\dot\gamma_s(l)\big|_{s=0}\rangle,$$

where the first equality follows from Lemma 5.19(i), and the other two follow from $X\langle Y,Z\rangle = \langle\nabla_X Y,Z\rangle + \langle Y,\nabla_X Z\rangle$ and $\nabla_{\dot c(s)}u_s = 0$. Hence we have

$$(\nabla^3 f)_p(w,u,u) = 2\Phi(l)\langle\dot\gamma(l),w\rangle\langle u,u\rangle + 2\left(-\Phi(l) - 2/l + 2\coth l\right)\langle\dot\gamma(l),w\rangle\langle u,\dot\gamma(l)\rangle^2$$
$$\qquad + 4\left(1 - l\coth l\right)\langle u,\dot\gamma(l)\rangle\langle u,\nabla_{\dot c(s)}\dot\gamma_s(l)\big|_{s=0}\rangle$$

60

$$
\begin{aligned}
&= 2\Phi(l)\langle\dot{\gamma}(l),w\rangle(\langle u,u\rangle - \langle u,\dot{\gamma}(l)\rangle^2) \\
&\quad + 4\left(1 - l\coth l\right)\langle u,\dot{\gamma}(l)\rangle\left[\langle u, \nabla_{\dot{c}(s)}\dot{\gamma}_s(l)\big|_{s=0}\rangle - \langle w,\dot{\gamma}(l)\rangle\langle u,\dot{\gamma}(l)\rangle/l\right].
\end{aligned}
\tag{5.28}
$$

To compute $\nabla_{\dot{c}(s)}\dot{\gamma}_s(l)|_{s=0}$, consider the (smooth) map $\alpha\colon [0,l]\times(-\epsilon,\epsilon) = M$ given by $(t,s)\mapsto\gamma_s(t)$. Let $\frac{\partial\alpha}{\partial s}(t,s) := d\alpha_{(t,s)}(\frac{\partial}{\partial t})$ and $\frac{\partial\alpha}{\partial t}(t,s) := d\alpha_{(t,s)}(\frac{\partial}{\partial s})$. Then $\nabla_{\dot{c}(s)}\dot{\gamma}_s(l)\big|_{s=0} = \nabla_{\frac{\partial\alpha}{\partial s}}\frac{\partial\alpha}{\partial t}(l,0) = \nabla_{\frac{\partial\alpha}{\partial t}}\frac{\partial\alpha}{\partial s}(l,0)$, since $\nabla_{\frac{\partial\alpha}{\partial s}}\frac{\partial\alpha}{\partial t} = \nabla_{\frac{\partial\alpha}{\partial t}}\frac{\partial\alpha}{\partial s} + [\frac{\partial\alpha}{\partial s},\frac{\partial\alpha}{\partial t}]$ and $[\frac{\partial\alpha}{\partial s},\frac{\partial\alpha}{\partial t}] = d\alpha([\frac{\partial}{\partial s},\frac{\partial}{\partial t}]) = 0$; see [Sak96, Lem. II.2.2] or [Lee18, Lem. 6.2]. By Lemma 5.18, $Y(t) := \frac{\partial\alpha}{\partial s}(t,0)$ is a Jacobi field along the geodesic γ, and satisfies $Y(0) = 0$ and $Y(l) = w$. Therefore Eq. (5.26) yields

$$
\nabla_{\dot{c}(s)}\dot{\gamma}_s(l)\big|_{s=0} = \nabla_{\dot{\gamma}(l)}Y(l) = \frac{1}{l}\dot{\gamma}(l)\langle w,\dot{\gamma}(l)\rangle + \coth(l)(w - \dot{\gamma}(l)\langle w,\dot{\gamma}(l)\rangle).
$$

By substituting this into Eq. (5.28), we obtain Eq. (5.31). $\qquad\square$

We are now ready to prove Theorem 5.16.

Proof of Theorem 5.16. We first restrict to the case $\kappa = -1$. We are going to bound

$$
\sigma_p(u,w) := \frac{|(\nabla^3 f)_p(w,u,u)|}{\sqrt{(\nabla^2 f)_p(w,w)(\nabla^2 f)_p(u,u)}}, \quad u,v\in T_pM\setminus\{0\}.
$$

From $d(p,p_0) = l$, it holds that $\|\dot{\gamma}(l)\| = 1$. We can also assume that $\|u\|_p = \|w\|_p = 1$. Therefore, $u,v,\dot{\gamma}(l)$ can be assumed to be unit vectors in $\mathbb{R}^3$, and represented in the spherical coordinate system as $\dot{\gamma}(l) = (0,0,1)$, $u = (\sin\theta,0,\cos\theta)$, $w = (\sin\varphi\cos\alpha,\sin\varphi\sin\alpha,\cos\varphi)$ for $\theta,\varphi\in[0,\pi]$ and $\alpha\in[0,2\pi]$. By Proposition 5.21, we have

$$
(\nabla^2 f)_p(w,w) = 2\cos^2\varphi + 2l\coth l\sin^2\varphi, \tag{5.29}
$$
$$
(\nabla^2 f)_p(u,u) = 2\cos^2\theta + 2l\coth l\sin^2\theta, \tag{5.30}
$$
$$
(\nabla^3 f)_p(w,u,u) = 2\Phi(l)\cos\varphi\sin^2\theta + 4\left(l - \Phi(l)\right)\cos\theta\sin\varphi\sin\theta(-\cos\alpha). \tag{5.31}
$$

By Lemma 5.17(i) the quantities $\Phi(l)$, $l - \Phi(l)$, $\sin(\theta)$, and $\sin(\varphi)$ in Eq. (5.31) are all non-negative. Thus

$$
|(\nabla^3 f)_p(w,u,u)| \leq 2\Phi(l)|\cos(\varphi)|\sin(\theta)^2 + 4\left(l - \Phi(l)\right)\sin(\varphi)\sin(\theta)|\cos(\theta)|. \tag{5.32}
$$

For $C := l\coth(l)\geq 1$, observe that

$$
\max_{\phi\in[0,\pi]}\frac{|\cos\phi|}{\sqrt{\cos^2\phi + C\sin^2\phi}} = 1, \quad \max_{\phi\in[0,\pi]}\frac{\sin\phi}{\sqrt{\cos^2\phi + C\sin^2\phi}} = \frac{1}{\sqrt{C}},
$$
$$
\max_{\theta\in[0,\pi]}\frac{\sin\theta|\cos\theta|}{\cos^2\theta + C\sin^2\theta} = \max_{\theta\in[0,\pi]}\frac{|\tan\theta|}{1 + C\tan^2\theta} = \max_{z\in[0,\infty)}\frac{z}{1 + Cz^2} = \frac{1}{2\sqrt{C}}.
$$

Therefore

$$
\begin{aligned}
\sigma_p(u,w) &\leq \max_{\varphi,\theta\in[0,\pi]}\frac{2\Phi(l)|\cos\varphi|\sin^2\theta + 4(l - \Phi(l))\sin\varphi\sin\theta|\cos\theta|}{\sqrt{2\cos^2\varphi + 2C\sin^2\varphi}\left(2\cos^2\theta + 2C\sin^2\theta\right)} \\
&\leq \frac{\Phi(l)}{\sqrt{2C}} + \frac{l - \Phi(l)}{\sqrt{2C}} = \frac{\tanh(l)}{\sqrt{2}} \leq \frac{1}{\sqrt{2}}.
\end{aligned}
$$

This shows the 8-self-concordance of f on M^n_{-1}.

We now show that this estimate is tight. Choose $\varphi = \pi/2$, $\tan^2 \theta = 1/C$, and $\alpha \in \{0, \pi\}$. From Eq. (5.31) we have

$$\sigma_p(u, w) = \frac{2(l - \Phi(l))|\cos(\theta)|\sin(\theta)}{\sqrt{2C}(\cos(\theta)^2 + C\sin(\theta)^2)} = \frac{l - \Phi(l)}{\sqrt{2C}} = \frac{(l - \Phi(l))\tanh(l)}{\sqrt{2l}} = \frac{l\coth(l) - 1}{\sqrt{2l}}.$$

For $l \to \infty$, it holds that $\sigma_p(u, w) \to 1/\sqrt{2}$, and so the estimate $\sigma_p(u, w) \leq 1/\sqrt{2}$ is tight. This completes the proof of (i). Note the choice of α guarantees that we are essentially working with $u, w, \dot{\gamma}(l) \in \mathbb{R}^2$, so the argument is still valid for $n = 2$.

For (ii), we consider the case of $u = w$; then $\varphi = \theta$ and $\alpha = 0$. From Eq. (5.31), we have

$$(\nabla^3 f)_p(u, u, u) = 2(-2l + 3\Phi(l))\cos\theta\sin^2\theta. \tag{5.33}$$

Then we have

$$\sup_{u \in T_pM} \sigma_p(u, u) = \max_{\theta \in [0, \pi/2]} \frac{|2l - 3\Phi(l)|\tan^2\theta}{\sqrt{2}(1 + C\tan^2\theta)^{3/2}} = \max_{z \in [0, \infty)} \frac{|2l - 3\Phi(l)|z}{\sqrt{2}(1 + Cz)^{3/2}}$$

$$= \sqrt{\frac{2}{27}}\frac{|2l - 3(\coth l + l - l\coth^2 l)|}{C} = \sqrt{\frac{2}{27}}|-3/l - \tanh l + 3\coth l|,$$

where the maximum of $z/(1 + Cz)^{3/2}$ is attained at $z = 2/C = 2(\tanh l)/l$. The supremum of the last quantity is attained at $l \to \infty$, and equals $\sqrt{2/27}$. This implies (ii), i.e., that f is $27/2$-self-concordant along geodesics, and that this bound is tight.

Finally we show (iii). Again, we may assume $\|w\|_p = \|u\|_p = 1$, and we use the above spherical coordinates. By Eqs. (5.29) and (5.30) and Lemma 5.19(i), we have

$$|\sin\theta| = \sqrt{\frac{(\nabla^2 f)_p(u, u) - 2dg_p(u)^2}{2l\coth l}}, \quad |\sin\varphi| = \sqrt{\frac{(\nabla^2 f)_p(w, w) - 2dg_p(w)^2}{2l\coth l}}.$$

By substituting these into Eq. (5.31) and using $dg_p(u) = \cos\theta$ and $dg_p(w) = \cos\varphi$ we obtain

$$\begin{aligned}
(\nabla^3 f)_p(w, u, u) &\leq \frac{\Phi(l)}{l\coth l}|dg_p(w)|((\nabla^2 f)_p(u, u) - 2dg_p(u)^2) \\
&+ \frac{2(l - \Phi(l))}{l\coth l}|dg_p(u)|\sqrt{(\nabla^2 f)_p(w, w) - 2dg_p(w)^2}\sqrt{(\nabla^2 f)_p(u, u) - 2dg_p(u)^2} \\
&\leq 2\zeta((\nabla^2 f)_p(u, u) - 2dg_p(u)^2) \\
&+ 2\sqrt{(\nabla^2 f)_p(w, w) - 2dg_p(w)^2}\sqrt{(\nabla^2 f)_p(u, u) - 2dg_p(u)^2},
\end{aligned}$$

where we used Lemma 5.17 for the second inequality. This implies (iii) for $\kappa = 1$.

Finally, the statements for $M^n_{-\kappa}$ follow from Lemmas 2.1 and 3.2. Note for part (iii) that rescaling the Riemannian metric on M^n_{-1} by a factor $1/\kappa$ yields sectional curvature κ, and rescales the distance g by a factor $1/\sqrt{\kappa}$, so to compensate one must use the prefactors $2\zeta\sqrt{\kappa}$ and $2\sqrt{\kappa}$. $\quad\square$

We now use Theorem 5.16 to prove the following theorem, which for $\kappa = 1$ yields Theorem 1.7:

Theorem 5.22. *Let $\kappa > 0$, $M = M^n_{-\kappa}$, $p_0 \in M$, and define $f: M \to \mathbb{R}$ by $f(p) = d(p, p_0)^2$. Define an open convex set $D \subseteq M \times \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ by*

$$D = \{(p, R, S) \in M \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} : RS - f(p) > 0\},$$

and define a function $F: D \to \mathbb{R}$ by

$$F(p, R, S) = -\log(RS - f(p)) + \kappa f(p)$$

Then F is convex and strongly $\frac{1}{2}$-self-concordant. Furthermore, $\lambda_{F, \frac{1}{2}}(p, R, S)^2 \leq 4 + 4\kappa f(p)$.

Proof. Recall from Corollary 5.4 that F is convex. Let $u = (u_p, u_R, u_S)$ and $w = (w_p, w_R, w_S)$ be tangent vectors at $(p, R, S) \in D$. Throughout the rest of this proof, we suppress the base point (p, R, S) for derivatives. Set

$$\Psi(p, R, S) = R - S^{-1}d(p, p_0)^2.$$

Instead of immediately taking F as stated, we leave the prefactor of f as a quantity $\xi > 0$ to be chosen later. The derivative of $F = -\log \Psi - \log S + \xi f$ is given by

$$dF(u) = -\frac{1}{\Psi}d\Psi(u) - \frac{u_S}{S} + \xi\, df(u_p).$$

Define $A_u = d\Psi(u)/\Psi$, $B_u = \sqrt{-\nabla^2\Psi(u, u)/\Psi}$, $C_u = S^{-1}u_S$ and $D_u = \sqrt{\xi\,\nabla^2 f(u_p, u_p)}$. We recall from Lemma 5.3 that Ψ is concave, so that B_u is well-defined. The Hessian of F is then given by

$$\nabla^2 F(u, u) = \underbrace{\frac{1}{\Psi^2}(d\Psi(u))^2}_{=A_u^2} - \underbrace{\frac{1}{\Psi}\nabla^2\Psi(u, u)}_{=B_u^2} + \underbrace{\frac{1}{S^2}u_S^2}_{=C_u^2} + \underbrace{\xi\,\nabla^2 f(u_p, u_p)}_{=D_u^2}. \tag{5.34}$$

For convenience we also write $B_{uw} = -\nabla^2\Psi(u, w)$. The third derivative of F is given by

$$\nabla^3 F(w, u, u) = -2\frac{1}{\Psi^3}(d\Psi(w))(d\Psi(u))^2 + 2\frac{1}{\Psi^2}(d\Psi(u))(\nabla^2\Psi(w, u)) + \frac{1}{\Psi^2}d\Psi(w)(\nabla^2\Psi(u, u))$$

$$- \frac{1}{\Psi}\nabla^3\Psi(w, u, u) - 2\frac{1}{S^3}w_S u_S^2 + \xi\,\nabla^3 f(w_p, u_p, u_p)$$

$$= -2A_w A_u^2 - 2A_u B_{uw} - A_w B_u^2 - 2C_w C_u^2 - \frac{1}{\Psi}\nabla^3\Psi(w, u, u) + \xi\,\nabla^3 f(w_p, u_p, u_p). \tag{5.35}$$

It is easy to see that the first four terms in Eq. (5.35) are bounded by a constant multiple of $\sqrt{\nabla^2 F(w, w)}\nabla^2 F(u, u)$, and similar for the last term (by α-self-concordance of f). The term $\nabla^3\Psi(w, u, u)/\Psi$ requires more effort. Recall from Lemma 5.3, if $g = d(p, p_0) = \sqrt{f}$, then

$$\nabla^2\Psi = -S^{-1}(2(S^{-1}g\,dS - dg)^{\otimes 2} + (\nabla^2 f - 2\,dg \otimes dg))),$$

and the third derivative satisfies

$$\nabla^3\Psi(w, u, u) = -2S^{-1}u_S\nabla^2\Psi(w, u) - S^{-1}w_S\nabla^2\Psi(u, u) - S^{-1}\nabla^3 f(w_p, u_p, u_p).$$

Therefore

$$\nabla^3 F(w, u, u) = = -2A_w A_u^2 - 2B_{wu}(A_u + C_u) - B_u^2(A_w + C_w) - 2C_w C_u^2 + \left(\frac{1}{\Psi S} + \xi\right)\nabla^3 f$$

$$= -2A_w\left(A_u^2 - \frac{1}{2}B_u^2\right) - 2B_{wu}(A_u + C_u) - 2C_w\left(\frac{1}{2}B_u^2 + C_u^2\right) + \left(\frac{1}{\Psi S} + \xi\right)\nabla^3 f.$$

We now use the bound from Theorem 5.16(iii) and the 2-strong-convexity of f:

$$\left|\frac{\nabla^3 f(w_p, u_p, u_p)}{S\Psi}\right| \leq \frac{|dg(w_p)| \cdot |\nabla^2 f(u_p, u_p) - 2(dg(u_p))^2|}{S\Psi} \cdot C_1$$

$$+ \frac{|dg(u_p)| \cdot \sqrt{\nabla^2 f(u_p, u_p) - 2(dg(u_p))^2} \cdot \sqrt{\nabla^2 f(w_p, w_p) - 2(dg(w_p))^2}}{S\Psi} \cdot C_2$$

$$\leq \frac{1}{\sqrt{2}}\sqrt{\nabla^2 f(w_p, w_p)}B_u^2 \cdot C_1 + \frac{1}{\sqrt{2}}\sqrt{\nabla^2 f(u_p, u_p)}B_u B_w \cdot C_2$$

where $C_1 = 2\zeta\sqrt{\kappa}$ and $C_2 = 2\sqrt{\kappa}$, and $\zeta \leq \frac{1}{2}$ is defined in Lemma 5.17. Furthermore, f is α-self-concordant with $\alpha = 8/\kappa$ (cf. Theorem 5.16(i)). The triangle inequality gives

$$|\nabla^3 F(w, u, u)|$$
$$\leq 2|A_w(A_u^2 - \frac{1}{2}B_u^2)| + 2|B_{wu}||A_u + C_u| + 2|C_w(\frac{1}{2}B_u^2 + C_u^2)|$$
$$+ \sqrt{\nabla^2 f(w_p, w_p)}(\frac{C_1}{\sqrt{2}}B_u^2 + \frac{2\xi}{\sqrt{\alpha}}\nabla^2 f(u_p, u_p)) + \frac{C_2}{\sqrt{2}}|B_w||B_u|\sqrt{\nabla^2 f(u_p, u_p)}$$
$$= 2|A_w(A_u^2 - \frac{1}{2}B_u^2)| + 2|B_{wu}||A_u + C_u| + 2|C_w(\frac{1}{2}B_u^2 + C_u^2)|$$
$$+ D_w\left|\frac{C_1}{\sqrt{2\xi}}B_u^2 + \frac{2}{\sqrt{\alpha\xi}}D_u^2\right| + |B_w B_u|\frac{C_2}{\sqrt{2\xi}}D_u$$
$$\leq 2|A_w(A_u^2 - \frac{1}{2}B_u^2)| + 2|B_w||B_u|(|A_u + C_u| + \frac{C_2}{2\sqrt{2\xi}}D_u) + 2|C_w(\frac{1}{2}B_u^2 + C_u^2)|$$
$$+ D_w\left|\frac{C_1}{\sqrt{2\xi}}B_u^2 + \frac{2}{\sqrt{\alpha\xi}}D_u^2\right|$$
$$\leq 2\sqrt{A_w^2 + B_w^2 + C_w^2 + D_w^2}\sqrt{L},$$

where we applied $|B_{uw}| \leq |B_u||B_w|$ to get the penultimate inequality, Cauchy–Schwarz to get the last inequality, and L is defined as

$$L = (A_u^2 - \frac{1}{2}B_u^2)^2 + |B_u|^2(|A_u + C_u| + \frac{C_2}{2\sqrt{2\xi}}D_u)^2 + (\frac{1}{2}B_u^2 + C_u^2)^2 + \left|\frac{C_1}{2\sqrt{2\xi}}B_u^2 + \frac{1}{\sqrt{\alpha\xi}}D_u^2\right|^2.$$

We now show that $L \leq 2(\nabla^2 F(u, u))^2$ for the choice $\xi = \kappa$. First, we use that $C_1 = 2\zeta\sqrt{\kappa}$, $C_2 = 2\sqrt{\kappa}$ and $\alpha = 8/\kappa$. Therefore L is

$$L = (A_u^2 - \frac{1}{2}B_u^2)^2 + |B_u|^2(|A_u + C_u| + \sqrt{\frac{\kappa}{2\xi}}D_u)^2 + (\frac{1}{2}B_u^2 + C_u^2)^2 + \left|\frac{\zeta\sqrt{\kappa}}{\sqrt{2\xi}}B_u^2 + \sqrt{\frac{\kappa}{8\xi}}D_u^2\right|^2$$
$$= A_u^4 - A_u^2 B_u^2 + \frac{1}{4}B_u^4 + |B_u|^2(|A_u + C_u|^2 + \sqrt{\frac{2\kappa}{\xi}}|A_u + C_u|D_u + \frac{\kappa}{2\xi}D_u^2)$$
$$+ \frac{1}{4}B_u^4 + B_u^2 C_u^2 + C_u^4 + \frac{\zeta^2\kappa}{2\xi}B_u^4 + \frac{\zeta\kappa}{2\xi}B_u^2 D_u^2 + \frac{\kappa}{8\xi}D_u^4$$
$$= A_u^4 + B_u^4\left(\frac{1}{2} + \frac{\zeta^2\kappa}{2\xi}\right) + C_u^4 + \frac{\kappa}{8\xi}D_u^4$$
$$+ 2B_u^2|A_u||C_u| + 2B_u^2 C_u^2 + \sqrt{\frac{2\kappa}{\xi}}B_u^2|A_u + C_u|D_u + \frac{\kappa}{2\xi}(1 + \zeta)B_u^2 D_u^2.$$

As $\zeta \leq \frac{1}{2}$, we have $\zeta^2 \leq \frac{1}{4}$. Therefore the choice $\xi = \kappa$ ensures that

$$L \leq A_u^4 + \frac{5}{8}B_u^4 + C_u^4 + \frac{1}{8}D_u^4$$
$$+ 2B_u^2|A_u||C_u| + 2B_u^2 C_u^2 + \sqrt{2}B_u^2|A_u + C_u|D_u + \frac{3}{4}B_u^2 D_u^2$$
$$\leq A_u^4 + \frac{5}{8}B_u^4 + C_u^4 + \frac{1}{8}D_u^4$$

$$+ \frac{1}{2}B_u^4 + 2A_u^2 C_u^2 + \frac{\sqrt{2}}{2}B_u^2(A_u^2 + C_u^2 + 2D_u^2) + \frac{3}{4}B_u^2 D_u^2$$
$$\leq \frac{9}{8}(\nabla^2 F(u,u))^2 \leq 2(\nabla^2 F(u,u))^2$$

as $\nabla^2 F(u,u) = A_u^2 + B_u^2 + C_u^2 + D_u^2$. To conclude, we have shown that

$$|\nabla^3 F(w,u,u)| \leq 2\sqrt{2}\sqrt{\nabla^2 F(w,w)}\,\nabla^2 F(u,u)$$

and F is $\frac{1}{2}$-self-concordant.[11]

We now verify the bound on the Newton decrement. For $u = (u_p, u_R, u_S) \in T_{(p,R,S)}D$ such that $\nabla^2 F(u,u) \neq 0$ and $u_p \neq 0$, we have

$$|dF(u)| = |-A_u - C_u + \xi df(u_p)| \leq \sqrt{A_u^2 + C_u^2 + D_u^2}\sqrt{1 + 1 + \frac{\xi^2 |df(u_p)|^2}{D_u^2}},$$

and

$$\frac{\xi^2 |df(u_p)|^2}{D_u^2} = \frac{\xi^2 |df(u_p)|^2}{\xi \nabla^2 f(u_p, u_p)} \leq \xi \lambda_f(p)^2 = 2\xi f(p)$$

by Corollary 5.2. Since we chose $\xi = \kappa$, this shows that $\lambda_{F,1/2}(p,R,S)^2 \leq 2(2 + 2\kappa f(p))$. $\qquad\square$

6 Applications

In this section, we discuss applications of our interior-point method framework. In Section 6.1 we show that the framework can be used to solve non-commutative optimization and scaling problems. In Sections 6.2 to 6.4, we use the previously constructed barriers for the epigraph of the squared distance on Hadamard symmetric spaces and the epigraph of the distance on the model spaces for constant negative sectional curvature to the natural geometric problems of computing minimum enclosing balls, geometric medians, and Riemannian barycenters. To achieve the above, we build on the results of Sections 4 and 5.

6.1 Non-commutative optimization and scaling problems

In this subsection we show that the problem of minimizing log-norm or Kempf–Ness functions, as discussed in Section 1.4, can be solved using our interior-point methods. This leads to also naturally leads to algorithms for *scaling problems*.

We briefly recap the general setup for the norm minimization problem and refer to [BFG$^+$19] for more detail. Throughout this section we let $G \subseteq \mathrm{GL}(n,\mathbb{C})$ be a connected algebraic Lie group such that $g^* \in G$ for every $g \in G$. (In mathematics, such groups are known as reductive and they are particularly well-behaved [Wal17].) We also fix $\pi\colon G \to \mathrm{GL}(V)$ to be a finite-dimensional rational complex representation of G. Let $K = G \cap \mathrm{U}(n)$, which is a maximal compact subgroup of G, and assume that V is endowed with a K-invariant inner product $\langle\cdot|\cdot\rangle$.[12] For a non-zero vector $0 \neq v \in V$, the goal is to minimize $\|\pi(g)v\|^2 = \langle v|\pi(g)^*\pi(g)|v\rangle = \langle v|\pi(g^*g)|v\rangle$ over $g \in G$, where we used that $\pi(g)^* = \pi(g^*)$.[13] Therefore, this is equivalent to minimizing $\langle v|\pi(p)|v\rangle$

---

[11]Bounding L by $(\nabla^2 F(u,u))^2$ would lead to 1-self-concordance of F, but it is not clear whether there is a choice of $\xi > 0$ such that F is 1-self-concordant and its Newton decrement is not too adversely affected.

[12]Following Dirac notation, we will also write $\langle v|A|w\rangle := \langle v|Aw\rangle$ for vectors $v, w \in v$ and operators A on V.

[13]Because K acts unitarily and the Lie algebra representation $\Pi = d\pi_I$ is complex linear, one has $\Pi(X^*) = \Pi(X)^*$ for $X \in \mathrm{Lie}(G)$. By the Cartan decomposition every $g \in G$ is a product $g = k\exp(H)$ with $k \in K$ and $H \in i\mathrm{Lie}(K)$, so $\pi(g)^* = (\pi(k)\exp(\Pi(H)))^* = \exp(\Pi(H))\pi(k)^{-1} = \exp(\Pi(H))\pi(k^{-1}) = \pi(g^*)$ (cf. [BFG$^+$19, Hir22]).

over $p \in M = \{g^*g : g \in G\} = G \cap \mathrm{PD}(n) \subseteq \mathrm{PD}(n)$. The log-norm or Kempf–Ness function computes the logarithm of this quantity:[14]

Definition 6.1 (Kempf–Ness function). *Let $M = \{g^*g : g \in G\} = G \cap \mathrm{PD}(n) \subseteq \mathrm{PD}(n)$. For $0 \neq v \in V$, the* Kempf–Ness function ϕ_v *is defined by*

$$\phi_v \colon M \to \mathbb{R}, \quad \phi_v(p) = \log \langle v | \pi(p) | v \rangle. \tag{6.1}$$

For the special case where $G = \mathrm{GL}(n, \mathbb{C})$, $V = \mathbb{C}^n$ and π is the identity map, we write

$$f_v \colon \mathrm{PD}(n) \to \mathbb{R}, \quad f_v(P) = \log \langle v | P | v \rangle. \tag{6.2}$$

We note that M is a convex subset of $\mathrm{PD}(n)$ [BH99, Thm. 10.58, Lem. 10.59], so the geodesics in M are precisely the geodesics in $\mathrm{PD}(n)$ which lie completely in G. Thus the tangent space $T_I M$ consists of those Hermitian matrices $H \in \mathrm{Herm}(n) = T_I \mathrm{PD}(n)$ which also are in $\mathrm{Lie}(G) := T_I G$, the Lie algebra of G. For $G = \mathrm{GL}(n, \mathbb{C})$, we simply have that $T_I M = \mathrm{Herm}(n)$.

Because K acts unitarily, π restricts to a map $M \to \mathrm{PD}(V)$, and one can verify that it sends geodesics to geodesics (i.e., it is *geodesically affine*). At the identity, we have the explicit description

$$\pi\left(\mathrm{Exp}_I(tH)\right) = \mathrm{Exp}_I(t\Pi(H))$$

for $H \in T_I M$ and $\Pi \colon \mathrm{Lie}(G) \to \mathrm{End}(V) = \mathrm{Lie}(\mathrm{GL}(V))$ is given by the derivative of π, i.e., $\Pi = d\pi_I$. The linear map Π is also known as the Lie algebra homomorphism induced by π. Therefore, the Kempf–Ness function is the composition of the geodesically affine map $M \to \mathrm{PD}(V)$, $p \mapsto \pi(p)$, and the map $\mathrm{PD}(V) \to \mathbb{R}$ given by $P \mapsto \log \langle v | P | v \rangle$, i.e., the Kempf–Ness function for the definining representation of $\mathrm{GL}(V)$. To establish bounds on the derivatives of ϕ_v, it therefore suffices to prove bounds on the derivatives of f_v, and to translate the results via Π.

Below, we prove the well-known fact that the Kempf–Ness functions are convex on M (see, e.g., [BFG$^+$19]). As explained above it suffices to prove this for the special case where $G = \mathrm{GL}(n, \mathbb{C})$ and $V = \mathbb{C}^n$, with $\pi \colon G \to \mathrm{GL}(V)$ given by the identity map.

Proposition 6.2. *For $0 \neq v \in \mathbb{C}^n$, the Hessian of the function $f_v \colon \mathrm{PD}(n) \to \mathbb{R}$ defined in Eq. (6.2) satisfies for every $P \in \mathrm{PD}(n)$ and $U \in T_P \mathrm{PD}(n)$ the identity*

$$(\nabla^2 f_v)_P(U, U) = \frac{\langle \tilde{v} | (\tilde{U} - \frac{\langle \tilde{v} | \tilde{U} | \tilde{v} \rangle}{\langle \tilde{v} | \tilde{v} \rangle} I)^2 | \tilde{v} \rangle}{\langle \tilde{v} | \tilde{v} \rangle},$$

where we use the notation $\tilde{v} = P^{1/2} v$ and $\tilde{U} = P^{-1/2} U P^{-1/2}$. As a consequence, all Kempf–Ness functions are convex.

Proof. We compute the Hessian of $f := f_v$. First off, we have

$$\partial_t f(\mathrm{Exp}_P(tU)) = \partial_t \log \langle v | \mathrm{Exp}_P(tU) | v \rangle = \frac{\langle \tilde{v} | \tilde{U} e^{t\tilde{U}} | \tilde{v} \rangle}{\langle \tilde{v} | e^{t\tilde{U}} | \tilde{v} \rangle}.$$

The second derivative is given by

$$\partial_{t=0}^2 f(\mathrm{Exp}_P(tU)) = \frac{\langle \tilde{v} | \tilde{U}^2 | \tilde{v} \rangle \langle \tilde{v} | \tilde{v} \rangle - \langle \tilde{v} | \tilde{U} | \tilde{v} \rangle^2}{\langle \tilde{v} | \tilde{v} \rangle^2} = \frac{\langle \tilde{v} | (\tilde{U} - \frac{\langle \tilde{v} | \tilde{U} | \tilde{v} \rangle}{\langle \tilde{v} | \tilde{v} \rangle} I)^2 | \tilde{v} \rangle}{\langle \tilde{v} | \tilde{v} \rangle},$$

hence is non-negative. $\qquad \square$

---

[14]Alternatively, because of the K-invariance, $g \mapsto \|\pi(g)v\|^2$ descends to a map on the quotient $K \backslash G$. This space is naturally isometric to M via the map $Kg \mapsto g^*g$: for $G = \mathrm{GL}(n, \mathbb{C})$ one can prove this using the polar decomposition, which generalizes to the Cartan decomposition for reductive G. As such, this is the same as Definition 6.1.

The expression for the first- and second derivatives can be understood in terms of the expectation and variance of corresponding random variables, as pointed out in [BFG$^+$19].[15] This will be useful for bounding the third derivative. Define a linear map $\Phi_v \colon \mathbb{C}^{n \times n} \to \mathbb{C}$ by

$$\Phi_v(A) = \frac{\langle v|A|v \rangle}{\langle v|v \rangle}. \tag{6.3}$$

Then Φ_v is what is known as a *completely positive* and *unital* map.[16] Such a map is to be interpreted as taking the expectation with respect to a random variable, where the random variable is now specified by a complex matrix. One can define the *covariance* between two matrices $A, B \in \mathbb{C}^{n \times n}$ as

$$\mathrm{Cov}_v(A, B) = \Phi_v(A^*B) - \Phi_v(A)^*\Phi_v(B). \tag{6.4}$$

The variance of A is defined accordingly as $\mathrm{Var}_v(A) = \mathrm{Cov}_v(A, A)$. With this notation, we can more succinctly write

$$(\nabla^2 f_v)_P(U, U) = \mathrm{Var}_{\tilde{v}}(\tilde{U}),$$

where $\tilde{v} = P^{1/2}v$ and $\tilde{U} = P^{-1/2}UP^{-1/2}$ as before. Then the third derivative can be computed as follows.

Proposition 6.3. *Let $0 \neq v \in \mathbb{C}^n$ and let $f_v \colon \mathrm{PD}(n) \to \mathbb{R}$ be as defined in Eq. (6.2). Then for every $U, W \in T_I\mathrm{PD}(n) = \mathrm{Herm}(n)$, its third derivative satisfies*

$$(\nabla^3 f_v)_I(W, U, U)$$
$$= \frac{1}{2}\frac{\langle v|\{W, U^2\}|v \rangle}{\langle v|v \rangle} - \frac{\langle v|U^2|v \rangle \langle v|W|v \rangle}{\langle v|v \rangle^2} - \frac{\langle v|U|v \rangle \langle v|\{W, U\}|v \rangle}{\langle v|v \rangle^2} + 2\frac{\langle v|U|v \rangle^2 \langle v|W|v \rangle}{\langle v|v \rangle^3}$$
$$= \mathrm{Re}\left(\mathrm{Cov}(W, U^2 - 2\Phi(U)U)\right).$$

Proof. To compute the third derivative of $f := f_v$ at $I \in \mathrm{PD}(n)$, note that

$$\partial_{t=0}(\nabla^2 f)_{\mathrm{Exp}_I(tW)}(\tau_{I \to \mathrm{Exp}_I(tW)}U, \tau_{I \to \mathrm{Exp}_I(tW)}U)$$
$$= \partial_{t=0}\left(\frac{\langle v|e^{\frac{t}{2}W}U^2 e^{\frac{t}{2}W}|v \rangle}{\langle v|e^{tW}|v \rangle} - \frac{\langle v|e^{\frac{t}{2}W}Ue^{\frac{t}{2}W}|v \rangle^2}{\langle v|e^{tW}|v \rangle^2}\right)$$
$$= \frac{1}{2}\frac{\langle v|\{W, U^2\}|v \rangle}{\langle v|v \rangle} - \frac{\langle v|U^2|v \rangle \langle v|W|v \rangle}{\langle v|v \rangle^2} - \frac{\langle v|U|v \rangle \langle v|\{W, U\}|v \rangle}{\langle v|v \rangle^2} + 2\frac{\langle v|U|v \rangle^2 \langle v|W|v \rangle}{\langle v|v \rangle^3}.$$

Using the map Φ and the associated covariance defined in Eqs. (6.3) and (6.4) , we may rewrite the above more succinctly as

$$(\nabla^3 f)_I(W, U, U) = \frac{1}{2}\Phi(\{W, U^2\}) - \Phi(U^2)\Phi(W) - \Phi(U)\Phi(\{W, U\}) + 2\Phi(U)^2\Phi(W)$$
$$= \frac{1}{2}(\mathrm{Cov}(W, U^2) + \mathrm{Cov}(U^2, W)) - \Phi(U)(\mathrm{Cov}(U, W) + \mathrm{Cov}(W, U))$$
$$= \mathrm{Re}\left(\mathrm{Cov}(W, U^2 - 2\Phi(U)U)\right). \qquad \square$$

---

[15]Similarly, the higher derivatives *along geodesics* can be related to higher cumulants, see [BFG$^+$19, Rem. 3.16].

[16]This means that $\Phi_v(I) = 1$, and the complete positivity refers to the fact that for every $n' \geq 1$, the map $\Phi_v \otimes I_{\mathbb{C}^{n' \times n'}} \colon \mathbb{C}^{n \times n} \otimes \mathbb{C}^{n' \times n'} \to \mathbb{C}^{n' \times n'}$ sends positive-semidefinite operators to positive-semidefinite operators.

Remark 6.4. *The Kempf–Ness functions are not necessarily self-concordant, even along geodesics. To see this, consider $v = \frac{1}{\sqrt{2}}(e_1 - e_2)$ and for $z \in \mathbb{R}$ the matrix $U_z \in \mathrm{Herm}(2)$ given by*

$$U_z = \begin{bmatrix} 1 & z \\ z & 0 \end{bmatrix}.$$

Then

$$(\nabla^2 f_v)_I(U_z, U_z) = \frac{1}{4}, \quad (\nabla^3 f_v)_I(U_z, U_z, U_z) = \frac{z}{2},$$

so $|(\nabla^3 f_v)_I(U_z, U_z, U_z)|$ can be arbitrarily large compared to $(\nabla^2 f_v)_I(U_z, U_z)^{3/2}$.

Although self-concordance does not hold, we do have the following bound on its third derivative, which implies that it is compatible (in the sense of Section 4.2) with any strongly convex function. This generalizes [BFG$^+$19, Prop. 3.15] beyond the case $W = U$.

Theorem 6.5. *Let $0 \neq v \in \mathbb{C}^n$ and let $f_v \colon \mathrm{PD}(n) \to \mathbb{R}$ be as defined in Eq. (6.2). For every $P \in \mathrm{PD}(n)$ and $U, W \in T_P\mathrm{PD}(n) = \mathrm{Herm}(n)$, one has the estimate*

$$\begin{aligned} \left| (\nabla^3 f_v)_P(W, U, U) \right| &\leq 4\|\tilde{U}\|_\infty \sqrt{(\nabla^2 f_v)_P(W, W)}\sqrt{(\nabla^2 f_v)_P(U, U)} \\ &\leq 4\|U\|_P \sqrt{(\nabla^2 f_v)_P(W, W)}\sqrt{(\nabla^2 f_v)_P(U, U)} \\ &= 4\|U\|_P \|W\|_{f_v, P}\|U\|_{f_v, P}. \end{aligned}$$

where $\tilde{U} = P^{-1/2}UP^{-1/2}$, and $\|\cdot\|_\infty$ is the spectral norm.

Proof. We prove the statement for $P = I$, and set $f := f_v$. Writing $\mathrm{Var}(A) = \mathrm{Cov}(A, A)$, an operator version of the Cauchy–Schwarz inequality [BD00] yields

$$\left| (\nabla^3 f)_I(W, U, U) \right|^2 \leq \left| \mathrm{Cov}(W, U^2 - 2\Phi(U)U) \right|^2 \leq \mathrm{Var}(W)\,\mathrm{Var}(U^2 - 2\Phi(U)U).$$

Using that for every $A, B \in \mathbb{C}^{n \times n}$,

$$\mathrm{Var}(A + B) = \mathrm{Var}(A) + \mathrm{Var}(B) + \mathrm{Cov}(A, B) + \mathrm{Cov}(B, A) \leq 2(\mathrm{Var}(A) + \mathrm{Var}(B)),$$

one can deduce for Hermitian A that

$$\begin{aligned} \mathrm{Var}(U^2 - 2\Phi(U)U) &\leq 2\,\mathrm{Var}(U(U - \Phi(U))) + 2\,\mathrm{Var}(\Phi(U)U) \\ &\leq 2\|U^2\|_\infty \mathrm{Var}(U - \Phi(U)) + 2\Phi(U)^2\,\mathrm{Var}(U) \\ &\leq 4\|U\|_\infty^2\,\mathrm{Var}(U) \end{aligned}$$

where the second inequality follows from

$$\begin{aligned} \mathrm{Var}(U(U - \Phi(U))) &\leq \Phi((U - \Phi(U))UU(U - \Phi(U))) \\ &= \frac{\langle v|(U - \Phi(U))UU(U - \Phi(U))|v\rangle}{\langle v|v\rangle} \\ &\leq 2\|U^2\|_\infty \frac{\langle v|(U - \Phi(U))^2|v\rangle}{\langle v|v\rangle} \\ &= 2\|U^2\|_\infty \mathrm{Var}(U - \Phi(U)). \end{aligned}$$

The theorem now follows from the observation that $(\nabla^2 f)_I(U, U) = \mathrm{Var}(U - \Phi(U)) = \mathrm{Var}(U)$. $\quad\square$

Corollary 6.6. *For $0 \neq v \in V$, the Kempf–Ness function ϕ_v defined in Eq. (6.1) satisfies for all $p \in M$ and $u, w \in T_pM$ the inequality*

$$\left|(\nabla^3 \phi_v)_p(w, u, u)\right| \leq 4\|d\pi_p(u)\|_{\pi(p)}\sqrt{(\nabla^2 \phi_v)_p(w, w)}\sqrt{(\nabla^2 \phi_v)_p(u, u)}.$$

The quantity $\|d\pi_p(u)\|_{\pi(p)}$ can be understood by observing that

$$d\pi_p(u) = \partial_{t=0}\pi(\mathrm{Exp}_p(tu)) = \partial_{t=0}\pi\left(p^{1/2}e^{tp^{-1/2}up^{-1/2}}p^{1/2}\right) = \pi(p^{1/2})\Pi(p^{-1/2}up^{-1/2})\pi(p^{1/2}).$$

Therefore $\|d\pi_p(u)\|_{\pi(p)} = \|\Pi(p^{-1/2}up^{-1/2})\|_I$. For convenience, we write $N(\pi) = \|\Pi\|$ for the operator norm of $\Pi\colon \mathrm{Lie}(G) \to \mathrm{End}(V)$. This quantity is known as the *weight norm* of π in [BFG+19], as it is determined as the largest norm of any highest weight appearing in the decomposition of the representation π into irreducible components. Then the above computation shows that the operator norm of $d_p\pi$ with respect to $\|\cdot\|_p$ and $\|\cdot\|_{\pi(p)}$ is exactly the weight norm $N(\pi)$.

Corollary 6.7. *Let $N(\pi) = \|\Pi\|$ the weight norm of π. Then for $0 \neq v \in V$, the Kempf–Ness function ϕ_v defined in Eq. (6.1) satisfies for all $p \in M$ and $u, w \in T_pM$ the inequality*

$$\left|(\nabla^3 \phi_v)_p(w, u, u)\right| \leq 4N(\pi)\,\|u\|_p\sqrt{(\nabla^2 \phi_v)_p(w, w)}\sqrt{(\nabla^2 \phi_v)_p(u, u)}.$$

We now apply the above to obtain an algorithmic result for optimizing Kempf–Ness functions over balls of fixed radius. Recall from Theorem 5.15 that $h(p) = \frac{1}{2}d(p, I)^2$ is 1-self-concordant on $\mathrm{PD}(n)$. Therefore, the same holds on M. It directly follows from Theorem 4.11 that one can construct a strongly self-concordant function on its open epigraph, as h is $(0, 1)$-compatible with itself:

Proposition 6.8. *Let $h\colon M \to \mathbb{R}$ be defined by $h(p) = \frac{1}{2}d(p, I)^2$. Let $S_0 > 0$ and consider $D = \{p \in M : h(p) < S_0\}$. Then the function $F\colon D \to \mathbb{R}$ defined by*

$$F(p) = -\log(S_0 - h(p)) + h(p)$$

is a self-concordant barrier for D with barrier parameter $\theta = 1 + S_0$.

The claim that it has barrier parameter at most $1 + S_0$ follows from $\lambda_F(p)^2 \leq 1 + \lambda_h(p)^2$ and Corollary 5.2. Since h is 1-strongly convex, we see that the Kempf–Ness function ϕ_v is $(0, 2N(\pi))$-compatible with F in the sense of Definition 4.8. Therefore by Proposition 4.10, for every $t \geq 0$, the function $F_t := t\phi_v + F$ is α-self-concordant, where α is given by

$$\alpha = \begin{cases} \frac{4N(\pi)^2 - 1}{4N(\pi)^4} & \text{if } 2N(\pi)^2 > 1, \\ 1 & \text{otherwise.} \end{cases} \tag{6.5}$$

Lastly, we can exactly give the analytic center of F: one easily verifies that it is given by $p = I$. We obtain the following algorithmic result.

Theorem 6.9. *For $0 \neq v \in V$, let $\phi_v\colon M \to \mathbb{R}$ be the Kempf–Ness function defined in Eq. (6.1). Let $\alpha \geq 0$ be as in Eq. (6.5). Then for every $S_0 > 0$, using*

$$\left(\frac{9}{5} + \frac{36}{5}\sqrt{\frac{1 + S_0}{\alpha}}\right)\log\left(\frac{8(1 + S_0 + \alpha)}{\sqrt{\alpha}\varepsilon}\right)$$

iterations of the path-following method, one can compute a point $P_\varepsilon \in M$ such that

$$\phi_v(p_\varepsilon) - \inf_{p \in D}\phi_v(p) \leq \varepsilon.$$

Proof. Set $\lambda^{(1)} = \frac{1}{4}$ and $\lambda^{(2)} = \frac{1}{9}$. Let p_i be the sequence of points defined in Theorem 4.17 with these choices of $\lambda^{(i)}$. These satisfy

$$\phi_v(p_i) - \inf_{p \in D} \phi_v(p) \leq \frac{2(1 + S_0 + \alpha)\|d\phi_v\|_{F,p}^*}{\sqrt{\alpha}\lambda^{(1)}} \exp\left(-i\frac{\lambda^{(1)} - \lambda^{(2)}}{\lambda^{(1)} + \sqrt{(1 + S_0)/\alpha}}\right)$$

where we used that the barrier parameter θ of F is $1 + S_0$. The norm $\|d(\phi_v)_p\|_{F,p}^*$ is at most $N(\pi)$, because F is strongly 1-convex and $d(\phi_v)_p$ is $N(\pi)$-Lipschitz: f_v is easily checked to be 1-Lipschitz, and π is $N(\pi)$-Lipschitz. Therefore we just need to ensure that

$$i\frac{\lambda^{(1)} - \lambda^{(2)}}{\lambda^{(1)} + \sqrt{(1 + S_0)/\alpha}} \geq \log\left(\frac{2(1 + S_0 + \alpha)}{\sqrt{\alpha}\lambda^{(1)}\varepsilon}\right),$$

which amounts to

$$i \geq \frac{\lambda^{(1)} + \sqrt{(1 + S_0)/\alpha}}{\lambda^{(1)} - \lambda^{(2)}} \log\left(\frac{2(1 + S_0 + \alpha)}{\sqrt{\alpha}\lambda^{(1)}\varepsilon}\right) = \left(\frac{9}{5} + \frac{36}{5}\sqrt{\frac{1 + S_0}{\alpha}}\right) \log\left(\frac{8(1 + S_0 + \alpha)}{\sqrt{\alpha}\varepsilon}\right). \quad \square$$

Corollary 6.10. *For $0 \neq v \in V$, let ϕ_v be the Kempf–Ness function defined in Eq. (6.1). Then for every $\varepsilon > 0$ and $R_0 > 0$, an ε-approximate minimizer of ϕ_v over a ball of radius R_0 around $I \in M \subseteq \mathrm{PD}(n)$ can be found using*

$$\mathcal{O}\left((1 + R_0)(1 + N(\pi)) \log\left(\frac{R_0\, N(\pi)}{\varepsilon}\right)\right)$$

iterations of the path-following method.

We shall not explicitly relate the norm minimization problem to the scaling problem here, but note that approximate minimizers of the Kempf–Ness function necessarily have small gradient (hence their *moment map* image is close to zero), and determining whether the gradient can become arbitrarily close to zero is the *null-cone problem*, to which the general scaling problem can be reduced. See [BFG+19] for details.

We briefly comment on the geometric meaning of the Kempf–Ness functions. For the purpose of optimization, it is natural to consider whether there exists an analogue of (non-constant) linear functions on $\mathbb{R}^n$. This is generally not the case; in fact, if M is a complete Riemannian manifold with a non-constant smooth function $h\colon M \to \mathbb{R}$ such that $\nabla^2 h = 0$, then M is isometric to a product $M' \times \mathbb{R}$, such that after this identification, h is some multiple of the projection onto the second coordinate [Inn82].[17] There does exist another useful generalization, namely the class of Busemann functions; see [BH99, II.8] for general background. These may be defined on any Hadamard manifold M (and also more generally) as follows [Hir22]: for a (not necessarily unit-speed) geodesic $\gamma\colon \mathbb{R} \to M$ with $\dot{\gamma} \neq 0$, define $b_\gamma\colon M \to \mathbb{R}$ by

$$b_\gamma(p) := \|\dot{\gamma}(0)\|\left(\lim_{t \to \infty} d(p, \gamma(t/\|\dot{\gamma}(0)\|)) - t\right). \tag{6.6}$$

This limit is well-defined and the resulting function turns out to be convex, and in the specific case of $M = \mathbb{R}^n$, reduces to an arbitrary (suitably normalized) affine function. For $M = \mathrm{PD}(n)$,

---

[17]For Hadamard M, this may be deduced as follows: $\nabla^2 h = 0$ implies that $\|dh\|$ is a constant function on M. Since h is non-constant, $\|dh\|$ is nonzero. The gradient flow of h is by isometries, without fixed points. If $z\colon M \to M$ denotes the map given by following the gradient flow for time 1, then $d(z(p), p)$ is also constant as a function of $p \in M$, and the subgroup of the isometries of M generated by z acts properly by semi-simple isometries on M, in the sense of [BH99, Def. I.8.2, Def. II.6.1]. Hence by [BH99, Thm. 7.1], M splits as a product $M' \times \mathbb{R}$.

whenever γ converges to a *rational* point at infinity, the Busemann function is a multiple of the Kempf–Ness function associated with a highest weight vector for an irreducible representation of $\mathrm{GL}(n)$. This follows, e.g., by comparing [Hir22, Lem. 2.34] and [FW22, Thm. 5.7]. The Kempf–Ness functions for $v \in \mathbb{C}^n$ with $\|v\| = 1$, considered as a vector in the defining representation of $\mathrm{GL}(n)$, correspond to those γ for which $\dot{\gamma}(0)$ is $-vv^*$, which may also be deduced from e.g. [BH99, Prop. 10.69].

6.2 The minimum enclosing ball problem

In the remainder of this section we show to apply the results of Sections 4 and 5 to various geometric problems, all of which involve the distance function or its square.

We first study the *minimum enclosing ball problem (MEB)* on a manifold M: given $m \geq 3$ distinct points $p_1, p_2, \ldots, p_m$ in M, find the smallest ball containing all of them. More formally, finding the MEB amounts to solving the following nonsmooth optimization problem:

$$\text{minimize } R \text{ s.t. } (p, R) \in M \times \mathbb{R}, \; d(p, p_i) \leq R \; (i = 1, 2, \ldots, m). \tag{6.7}$$

In the case of Euclidean space $M = \mathbb{R}^n$, MEB is a well-studied problem in computational geometry, and can be formulated as a second-order cone program to which an interior-point method is applicable; see e.g. [KMY04].

Nielsen and Hadjeres [NH15] addressed this problem for a hyperbolic space M. We shall assume that M is a complete convex submanifold of $\mathrm{PD}(n)$, but we note that similar results may be obtained for products of (rescalings of) these spaces, hence for all Hadamard symmetric spaces as explained in Section 1.3. To apply our framework, we reformulate Eq. (6.7) as a convex optimization problem over the following bounded domain.

Lemma 6.11. *Set* $S_0 = \max_{i \neq j} d(p_i, p_j)^2$. *Let* $D \subseteq M \times \mathbb{R}$ *be defined by*

$$D = \{(p, S) \in M \times \mathbb{R} \mid d(p, p_i)^2 < S < 2S_0 \quad (i = 1, 2, \ldots, m)\}. \tag{6.8}$$

Then D *is convex, open, bounded and non-empty, as* $(p_j, \frac{3}{2}S_0) \in D$ *for every* $j = 1, \ldots, m$.

Proof. Since D is the intersection of open epigraphs of squared distance functions and an open halfspace defined by $S < 2S_0$, it is open and convex. The boundedness of D is clear, as is the containment $(p_j, \frac{3}{2}S_0) \in D$ for every $j = 1, \ldots, m$. $\square$

Clearly, the optimal radius of a MEB is at most $R_0 := \max_{i \neq j} d(p_i, p_j) = \sqrt{S_0}$. It is also at least half of that:

Lemma 6.12. *Let* R_* *be the optimum of* Eq. (6.7) *and* $R_0 = \max_{i \neq j} d(p_i, p_j)$. *Then* $2R_* \geq R_0$.

Proof. For every $p \in M$, we have

$$d(p_i, p_j) \leq d(p_i, p) + d(p, p_j) \leq 2 \max_k d(p_k, p).$$

Minimizing the right-hand side with respect to $p \in M$ yields $d(p_i, p_j) \leq 2R_*$ for every i, j; maximizing over $i \neq j$ gives the desired bound. $\square$

Replacing the objective function R by $R^2 = S$, finding the MEB is equivalent to solving

$$\text{minimize } S \text{ s.t. } (p, S) \in D. \tag{6.9}$$

As a natural application of our results, we obtain a self-concordant barrier for D.

Proposition 6.13. *Let D be as in Lemma 6.11. Define $G \colon D \to \mathbb{R}$ by*

$$G(p, S) = -\log(2S_0 - S) + \sum_{i=1}^{m} \left(-\log(S - d(p, p_i)^2) + \frac{1}{2}d(p, p_i)^2 \right).$$

Then G is a self-concordant barrier for D, with barrier parameter $\theta = 1 + m(1 + 2S_0)$.

Proof. Let $F_i(p, S) := -\log(S - d(p, p_i)^2) + \frac{1}{2}d(p, p_i)^2$. By Corollary 1.5 and Theorem 1.6, F_i is is 1-self-concordant. Furthermore, it satisfies $\lambda_{F_i}(p, S)^2 \le 1 + d(p, p_i)^2 \le 1 + 2S_0$. As $-\log(2S_0 - S)$ is 1-self-concordant, so is G. The Newton decrement of G then satisfies $\lambda_G(p, S)^2 \le 1 + m(1 + 2S_0)$. Hence G is a self-concordant barrier with the claimed parameter. $\qquad \square$

To initialize the path-following method, we use the damped Newton method from Theorem 3.18. To estimate its iteration complexity, we need a lower bound on G.

Lemma 6.14. *For every $(p, S) \in D$, we have*

$$G(p, S) \ge -(1 + m)\log(2S_0).$$

Proof. Since $x \mapsto -\log(x)$ is decreasing, $d(p, p_i)^2 \ge 0$ and $S > 0$, we have $G(p, S) \ge -\log(2S_0) - m\log(2S_0) = -(1 + m)\log(2S_0)$. $\qquad \square$

The main result of this subsection is then the following.

Theorem 6.15. *Let $p_1, p_2, \ldots, p_m \in M$, and let R_* denote the radius of the minimum enclosing ball for these points. Set $R_0 = \max_{i \ne j} d(p_i, p_j)$. For $\varepsilon > 0$, with $\mathcal{O}(mR_0^2)$ iterations of a damped Newton method and*

$$\mathcal{O}\left(\sqrt{1 + m(R_0^2 + 1)} \log\left(\frac{m(R_0^2 + 1)}{\varepsilon} \right) \right)$$

iterations of the path following method, one can find $(p_\varepsilon, R_\varepsilon) \in M \times \mathbb{R}$ such that $R_\varepsilon \le R_* + \varepsilon$, and the ball with center p_ε and radius R_ε includes $p_1, p_2, \ldots, p_m$.

Proof. Set $\lambda^{(1)} = \frac{1}{4}$, $\lambda^{(2)} = \frac{1}{9}$. The damped Newton method of Theorem 3.18 with starting point $(p_j, \frac{3}{2}S_0)$ yields a point (q, S) with $\lambda_G(q, S) \le \frac{1}{2}\lambda^{(1)}$ within the order of

$$\frac{G(p_j, \frac{3}{2}S_0) - \inf_{(p, S) \in D} G(p, S)}{\frac{1}{2}\lambda^{(1)}}$$

$$\le \frac{-\log(S_0/2) + \sum_{i=1}^{m}(-\log((3/2)S_0 - d(p_j, p_i)^2) + (1/2)d(p_j, p_i)^2) + (1 + m)\log(2S_0)}{\frac{1}{2}\lambda^{(1)}}$$

$$\le \frac{-\log(S_0/2) - m\log(S_0/2) + (m/2)S_0 + (1 + m)\log(2S_0)}{\frac{1}{2}\lambda^{(1)}}$$

$$= \frac{(1 + m)\log 4 + (m/2)S_0}{\frac{1}{2}\lambda^{(1)}}$$

iterations. Consider the path-following method in Theorem 4.17 from the initial point (q, S), with objective $s \colon D \to \mathbb{R}$ defined by $(p, S) \mapsto S$. Since this is a linear map, $ts + G$ is 1-self-concordant for all $t > 0$. The starting time t_0 is given by

$$t_0 = \frac{\lambda^{(1)} - \lambda_G(q, S)}{\|ds_{(q,S)}\|_{G,(q,S)}^*} \ge \frac{\lambda^{(1)} - \lambda_G(q, S)}{2S_0},$$

where $\|ds_{(q,S)}\|^*_{G,(q,S)}$ is bounded by $2S_0$ by Lemma 4.18. Thus the path-following method yields a sequence of points (q_l, S_l) such that

$$S_l - R_*^2 \leq \frac{8S_0(\theta+1)}{\lambda^{(1)}} \exp\left(-l\frac{\lambda^{(1)} - \lambda^{(2)}}{\lambda^{(1)} + \sqrt{\theta}}\right),$$

where $\theta = 1 + m(1 + 2S_0)$ is the barrier parameter of G and we used $\lambda(q,S) \leq \lambda^{(1)}/2$. For $\varepsilon' > 0$, after

$$l \geq \frac{\frac{1}{4} + \sqrt{\theta}}{\frac{1}{4} - \frac{1}{9}} \log\left(\frac{32(\theta+1)}{\varepsilon'}\right)$$

iterations, we have

$$S_l - R_*^2 \leq \varepsilon' S_0.$$

For $R_l = \sqrt{S_l}$, it holds that

$$R_l - R_* \leq \varepsilon' S_0/(R_l + R_*) \leq \varepsilon' S_0/2R_* \leq \varepsilon' S_0/R_0,$$

where the last inequality follows from Lemma 6.12. Therefore, choosing $\varepsilon' = \varepsilon R_0/S_0 = \varepsilon/R_0$ yields the desired estimate. $\qquad\square$

6.3 The geometric median on model spaces

In this subsection we show how to apply the methods from Section 4 to compute geometric medians on the model spaces $M^n_{-\kappa}$ for constant sectional curvature $-\kappa$, where $\kappa > 0$. For now, we shall work with general M; later, we restrict to the model spaces because it is there that we have a barrier for the epigraph of the distance function (cf. Theorem 5.22). Recall from the introduction that the geometric median problem is as follows: given $m \geq 3$ points $p_1, \ldots, p_m \in M$, not all contained in a single geodesic, find $p_0 \in M$ such that

$$p_0 \in \operatorname*{argmin}_{p \in M} s(p) := \sum_{i=1}^m d(p, p_i). \tag{6.10}$$

This is a convex optimization objective, as the distance to a point is convex by Lemma 5.1. Let us first construct define a suitable domain to optimize over.

Lemma 6.16. *Set $R_0 = \max_{i \neq j} d(p_i, p_j)$. Let $D \subseteq M \times \mathbb{R}$ be defined by*

$$D = \{(p, R) \in M \times \mathbb{R}^m : R_i^2 > d(p, p_i)^2, \, 2R_0 > R_i > 0\}.$$

Then D is convex, open, and non-empty: for every $j \in [m]$, we have $(p_j, \frac{3}{2}R_0\mathbf{1}) \in D$, where $\mathbf{1} \in \mathbb{R}^m$ is the all-ones vector.

Proof. The convexity of D follows from the convexity of the distance function, see Lemma 5.1. The fact that D is open is obvious. Lastly, the given points are in D because

$$d(p_j, p_i) \leq R_0 < \frac{3}{2}R_0. \qquad\square$$

Lemma 6.17. *Define $c\colon M \times \mathbb{R}^m \to \mathbb{R}$ by $c(p, R) = \sum_{i=1}^m R_i$, and let $s\colon M \to \mathbb{R}$ be as in Eq. (6.10). Then*

$$\inf_{(p,R)\in D} c(p, R) = \inf_{p \in M} s(p)$$

The proof relies on the fact that the geometric median of $p_1, \ldots, p_m$ is contained in the convex hull of these points, for which we essentially follow the argument given in [Yan10, Prop. 2.4], where this fact is proven for more general distributions (rather than just discrete distributions).

Proof. First, we observe that for fixed $(p, R) \in D$,

$$\inf_{R' : (p, R') \in D} c(p, R') = \sum_{i=1}^{m} d(p, p_i) = s(p).$$

Thus it suffices to prove that if $p_0 \in \operatorname{argmin}_{p \in M} s(p)$, then there exists some $R \in \mathbb{R}^m$ such that $(p_0, R) \in D$. We claim that any such p_0 is in the convex hull of the p_j. From this claim one immediately deduces that $(p_0, R) \in D$ for $R = \frac{3}{2} R_0 \mathbf{1}$, since by Lemma 6.16, D is convex and $(p_j, \frac{3}{2} R_0 \mathbf{1}) \in D$ for every $j \in [m]$.

We now establish the claim by proving its contrapositive. Suppose p is not in the convex hull C of the points $p_1, \ldots, p_m$, and let q be the projection of p onto C, which is automatically distinct from p. We use the notion of Alexandrov angle, which for three points $a, b, c \in M$ with $a \neq b, c$ is defined as the unique $\angle_a(b, c) \in [0, \pi]$ such that

$$\cos \angle_a(b, c) = \frac{\langle \operatorname{Exp}_a^{-1}(b), \operatorname{Exp}_a^{-1}(c) \rangle_a}{d(a, b) \, d(a, c)}.$$

Suppose first that $q = p_j$ for some $j \in [m]$. Then $\angle_p(q, p_j) = 0$. On the other hand, if $q \neq p_j$, by [BH99, Prop. II.2.4], we have $\angle_q(p, p_j) \geq \pi/2$. On a Hadamard manifold, the angles of a triangle add to at most π, hence $\angle_p(q, p_j) \leq \pi/2$. Since we have $m \geq 3$, there must exist at least two j such that $q \neq p_j$. Furthermore, for at least one such j, the inequality must be strict: if the inequality is not strict then we must have $\angle_{p_j}(q, p) = 0$, so p, q, p_j all lie on a single geodesic. Since p is distinct from all p_j, s is differentiable at p, and it follows from Lemma 5.1 that

$$\operatorname{grad}(s)_p = -\sum_{j=1}^{m} \frac{\operatorname{Exp}_p^{-1}(p_j)}{d(p, p_j)}.$$

Since we have shown that $\angle_p(q, p_j) \leq \pi/2$ for every $j \in [m]$, with strict inequality for at least one j, we have

$$\langle \operatorname{grad}(s)_p, \operatorname{Exp}_p^{-1}(q) \rangle_p = -d(p, q) \sum_{j=1}^{m} \cos \angle_p(q, p_j) < 0$$

because $d(p, q) \neq 0$. In particular, $\operatorname{grad}(s)_p \neq 0$ and p is not a minimizer of s. $\qquad \square$

We now construct a barrier for the domain D. From here onwards, we assume that $M = M_{-\kappa}^n$ with $\kappa > 0$.

Proposition 6.18. *Let D be as in Lemma 6.16. Define $G \colon D \to \mathbb{R}$ by*

$$G(p, R) = \sum_{i=1}^{m} \left(-\log(2R_0 - R_i) - 2\log(R_i^2 - d(p, p_i)^2) + 2\kappa \, d(p, p_i)^2 \right).$$

Then G is a self-concordant barrier for D, with barrier parameter $\theta = 5m + 16m\kappa R_0^2$.

Proof. Let $\Psi(r) = -\log(2R_0 - r)$ and recall from Theorem 5.22 that $F_i(p, R, S) = -\log(RS - d(p, p_i)^2) + \kappa\, d(p, p_i)^2$ is strongly $\frac{1}{2}$-self-concordant. Using Lemma 3.2 and the strong 1-self-concordance of $-\log(2R_0 - R)$, we deduce that G is strongly 1-self-concordant. Then for every $(p, R) \in D$, we have

$$d((p, R), (p_1, R_0))^2 = d(p, p_1)^2 + |R - R_0|^2 \le 2R^2 - 2R_0 R + R_0^2 \le 8R_0^2 - 4R_0^2 + R_0^2 = 5R_0^2.$$

where d on the left-hand side refers to the distance on $M \times \mathbb{R}$. Furthermore, for every $(p, R) \in D$, the bound on $\lambda_{F_i, 1/2}(p, R, S) = \lambda_{2F_i, 1}(p, R, S)$ from Theorem 5.22 implies that

$$\lambda_G(p, R)^2 \le \sum_{i=1}^m \lambda_\Psi(R_i)^2 + \lambda_{2F_i, 1}^2(p, R, R) \le m + \sum_{i=1}^m (4 + 4\kappa\, d(p, p_i)^2) \le 5m + 16m\kappa R_0^2.$$

Therefore G is a self-concordant barrier with barrier parameter $\theta = 5m + 16m\kappa R_0^2$. $\qquad\square$

We now consider how to initialize the path-following method for the objective

$$c(p, R) = \sum_{i=1}^m R_i,$$

which is such that $tc + G$ is 1-self-concordant for every $t \ge 0$, because c is linear. To apply Theorem 4.17, we need to find a point $(q, S) \in D$ such that $\lambda_G(q, S) < \lambda^{(1)}$.[18] We can do this using the damped Newton method from Theorem 3.18. To bound the number of iterations, we must bound the potential gap of G.

Lemma 6.19. *For every $(p, R) \in D$, we have*

$$G(p, R) \ge -m \log(32 R_0^5).$$

Proof. The function $x \mapsto -\log(x)$ is decreasing. Because $R_i > 0$ for every $i \in [m]$, we have $-\log(2R_0 - R_i) \ge -\log(2R_0)$. Similarly, because $R_i < 2R_0$ and $d(p, p_i) \ge 0$ for every $i \in [m]$, each $-\log(R_i^2 - d(p, p_i)^2)$ term is at least $-\log(4R_0^2)$. Hence $G(p, R) \ge -m \log(2R_0) - 2m \log(4R_0^2) = -m \log(32 R_0^5)$, concluding the proof. $\qquad\square$

We now prove the main result of this subsection.

Theorem 6.20. *Let $p_1, \ldots, p_m \in M_{-\kappa}^n$ with $\kappa > 0$ be $m \ge 3$ points, not all on one geodesic, and set $R_0 = \max_{i \ne j} d(p_i, p_j)$. Define $s(p) = \sum_{j=1}^m d(p, p_j)$, and let $\varepsilon > 0$. Then with $\mathcal{O}((m+1)\kappa R_0^2)$ iterations of a damped Newton method and*

$$\mathcal{O}\left(\sqrt{m(\kappa R_0^2 + 1)} \log\left(\frac{mR_0(\kappa R_0^2 + 1)}{\varepsilon}\right)\right)$$

iterations of the path following method, one can find $p_\varepsilon \in M_{-\kappa}^n$ such that

$$s(p_\varepsilon) - \inf_{q \in M} s(q) \le \varepsilon.$$

---

[18]For *fixed* q, it is easy to determine the optimal S, by explicitly solving the first-order optimality conditions.

Proof. Set $\lambda^{(1)} = \frac{1}{4}$, $\lambda^{(2)} = \frac{1}{9}$. The damped Newton method of Theorem 3.18 with starting point $(p_j, \frac{3}{2}R_0\mathbf{1})$ yields a point (q, S) with $\lambda_G(q, S) \leq \frac{1}{2}\lambda^{(1)}$ within the order of

$$\frac{G(p_j, \frac{3}{2}R_0\mathbf{1}) - \inf_{(p,R)\in D} G(p, R)}{\frac{1}{2}\lambda^{(1)}}$$

$$\leq \frac{G(p_j, \frac{3}{2}R_0\mathbf{1}) + m\log(32R_0^5)}{\frac{1}{2}\lambda^{(1)}}$$

$$= \frac{-m\log(\frac{R_0}{2}) - 2\sum_{i=1}^m \log(\frac{9}{4}R_0^2 - d(p_j, p_i)^2) + m\log(32R_0^5) + 2\kappa\sum_{i=1}^m d(p_j, p_i)^2}{\frac{1}{2}\lambda^{(1)}}$$

$$\leq \frac{-m\log(\frac{R_0}{2}) - 2\sum_{i=1}^m \log(\frac{5}{4}R_0^2) + m\log(32R_0^5) + 8\kappa m R_0^2}{\frac{1}{2}\lambda^{(1)}}$$

$$= \frac{m\log(\frac{1024}{25}) + 8m\kappa R_0^2}{\frac{1}{2}\lambda^{(1)}}$$

iterations. A suitable choice of starting time is for the path-following method from Theorem 4.17 is then

$$t_0 = \frac{\lambda^{(1)} - \lambda_G(q, S)}{\|dc_{(q,S)}\|^*_{G,(q,S)}}.$$

It remains to be shown that this is not too small. We give an upper bound on $\|dc_{(q,S)}\|^*_{G,(q,S)}$. The domain D is constructed so that $c(p, R) \leq 2mR_0$ for every $(p, R) \in D$, and $c(q, S) \geq 0$. It follows by Lemma 4.18 that

$$\|dc_{(q,S)}\|^*_{G,(q,S)} \leq 2mR_0,$$

and so $t_0 \geq (\lambda^{(1)} - \lambda_G(q, S))/(2mR_0)$. Therefore, initializing the algorithm from Theorem 4.17 with initial point (q, S) and the above t_0 yields a sequence of points (q_l, S_l) such that

$$c(q_l, S_l) - \inf_{(p,R)\in D} c(p, R) \leq \frac{4mR_0(\theta + 1)}{\lambda^{(1)}} \exp\left(-l\frac{\lambda^{(1)} - \lambda^{(2)}}{\lambda^{(1)} + \sqrt{\theta}}\right)$$

where θ is the barrier parameter of G, and we used that $\lambda^{(1)} - \lambda_G(q, S)$ is at least $\frac{1}{2}\lambda^{(1)}$. Rewriting the above and using Lemma 6.17 shows that

$$s(q_l) - \inf_{q\in M} s(q) \leq c(q_l, S_l) - \inf_{(q,R)\in D} c(q, R) \leq \varepsilon$$

whenever

$$l \geq \frac{\frac{1}{4} + \sqrt{\theta}}{\frac{1}{4} - \frac{1}{9}} \log\left(\frac{4mR_0(\theta + 1)}{\varepsilon}\right).$$

The theorem now follows from filling in $\theta = 5m + 16m\kappa R_0^2$. $\qquad\square$

6.4 The Riemannian barycenter

We end this section by briefly commenting on the problem of finding the Riemannian barycenter, first introduced by Cartan, and sometimes also called the Fréchet or Karcher mean, see e.g. [Afs11] for some historical context on this topic. It is defined as follows: given points $p_1, \ldots, p_m \in M$, find $p_0 \in M$

$$p_0 \in \operatorname*{argmin}_{p\in M} f(p) := \sum_{i=1}^m d(p, p_i)^2.$$

The point p_0 is known as the barycenter of $p_1, \ldots, p_m$, and is unique on Hadamard manifolds by strong convexity of f. It is trivial to find p_0 when $M = \mathbb{R}^n$ is Euclidean space, as it is given by $p_0 = \frac{1}{m} \sum_{i=1}^m p_i$. Furthermore, the solution is unique on any Hadamard manifold, as the squared distance is 2-strongly convex, and hence f is $2m$-strongly convex. Even for hyperbolic space it is not clear whether one can give a closed-form solution to the above problem. However, if M has sectional curvatures in $[-\kappa, 0]$, then f is $\mathcal{O}(m\sqrt{\kappa}R/\tanh(R\sqrt{\kappa}))$-smooth at p with $R = \max_j d(p, p_j)$, which follows from standard variational arguments [Lee18, Prop. 10.12, Thm. 10.22], hence the function f is well-conditioned. Therefore a standard gradient descent method gives an algorithm which converges relatively quickly; one can find an ε-approximate minimizer of f in $\mathcal{O}(\sqrt{\kappa}R_0 \log([f(p) - \inf_q f(q)]/\varepsilon)/\tanh(\sqrt{\kappa}R_0))$ iterations, where R_0 is some a priori bound on size of the domain one restricts to, and p is the starting point. This can be deduced from a simple adaptation of the Euclidean argument in [BG19, Thm. 3.8]). We note that one could also apply more sophisticated first-order methods such as accelerated gradient descent to this problem, see [AS20].

It is natural to determine what complexity our interior-point methods give for this problem. In the setting of $M = M_{-\kappa}^n$, we can (up to logarithmic factors) recover the above iteration complexity. We restrict the above optimization problem to a ball of radius $R_1 = \max_{j \neq 1} d(p_1, p_j)$ around the point p_1, and use the barrier $F(p) = -\log(R_1^2 - d(p, p_1)^2) + \kappa d(p, p_1)^2$, which has barrier parameter $1 + \mathcal{O}(\kappa R_1^2)$. Then, observe that by Theorem 5.16(iii) and Lemma 4.9, f is $(\sqrt{2}\zeta\sqrt{\kappa}, \sqrt{2}\kappa)$-compatible with any squared distance function, as each of the $d(p, p_i)^2$'s is. As a consequence, f is $(\sqrt{2}\zeta, \sqrt{2})$-compatible with F, and $tf + F$ is $\mathcal{O}(1)$-self-concordant for every $t \geq 1$ by Proposition 4.10. The path-following method, initialized with starting point p_1 (which is the analytic center of F), then yields an ε-approximate minimizer of f within $\mathcal{O}((1 + \sqrt{\kappa}R_1) \log(m\kappa R_1/\varepsilon))$ iterations. While this specific choice of barrier may seem odd, it has the advantage that we know its analytic center to be p_1, so it is easy to initialize the path-following method. This shows again that it is useful to have a general path-following method capable of dealing with compatible objectives, rather than just linear ones: if one included a barrier term for the epigraph of every $d(p, p_i)^2$, then it would both be harder to find the analytic center (for initialization), and the barrier parameter would scale with m. We note that a similar approach works on $\mathrm{PD}(n)$ if one suitably generalized Theorem 5.16(iii).

7 Outlook

In this work, we extend the basic theory of interior-point methods to manifolds, and show that the developed framework is capable of capturing interesting geodesically convex optimization problems. In particular, we define a suitable version of self-concordance on Riemannian manifolds, and show that it gives the same guarantees for Newton's method as in the Euclidean setting. This is used to analyze a path-following method for the optimization of compatible objectives over domains for which one has a self-concordant barrier. We exhibit non-trivial examples of self-concordant functions, namely squared distance functions on $\mathrm{PD}(n)$, and more generally symmetric spaces with non-positive curvature, and construct related self-concordant barriers. The framework is able to capture the optimization of Kempf–Ness functions, a problem which has connections to many areas of mathematics and computer science, leading to algorithms with state-of-the-art complexity guarantees. It also applies to computing the geometric median on hyperbolic space, for which we give an algorithm capable of finding high-precision solutions. This demonstrates the power of the framework, and we believe that it encompasses many more problems. Our work suggests several directions for further investigation:

- It is natural to search for self-concordant barriers for the aforementioned applications which have better barrier parameters. Alternatively, is it possible to prove lower bounds that show

that the constructions given in our work are essentially optimal?

- In Euclidean convex optimization, there are universal constructions of self-concordant barriers, cf. [NN94, Hil14, Fox15, BE19, Che21]. Can one find such a construction for manifolds? We describe a concrete proposal. Let $D \subseteq M$ be a compact convex subset of a Hadamard manifold M, with non-empty interior. Denote by CM^∞ the cone over the boundary at infinity of M [Hir22]. Its elements can be identified with the geodesic rays γ emanating from a fixed base point and hence determine Busemann functions b_γ as in Eq. (6.6). Define $F^*\colon CM^\infty \to \mathbb{R}$ by $F^*(\gamma) = \log \int_D \exp(-b_\gamma(q)) \, d\mathrm{vol}(q)$. Then the inverse Legendre–Fenchel conjugate $F\colon D \to \mathbb{R}$ of F^*, given by $F(p) = \sup_{\gamma \in CM^\infty} -b_\gamma(p) - F^*(\gamma)$, is a natural candidate for a barrier for D. Indeed, for Euclidean space $M = \mathbb{R}^n$ it reduces precisely to the *entropic barrier* of Bubeck and Eldan [BE19].

- From the perspective of interior-point methods, we currently only treat the *main stage*, which minimizes an objective given a starting point that is well-centered with respect to the barrier F. Can one give a general procedure for finding such a starting point from an arbitrary feasible point $p \in D$? In the Euclidean setting, this is achieved by applying the path-following method with the linear objective $f := -\langle \mathrm{grad}(F)_p, \cdot \rangle$ *in reverse*, starting at $t = 1$. This is sensible as p is exactly a minimizer of $F_t = tf + F$ at $t = 1$. Busemann functions generalize linear functions to Hadamard manifolds, hence is natural to instead use $f = b_\gamma$ with γ the geodesic ray starting at $p \in M$ with direction $\mathrm{grad}(F)_p$. When f is compatible with F (as we show in Section 6.1 for specific f and F), then one can use the same time steps as for the main stage, and switch to the main stage as soon as $\lambda_{F,\alpha} \leq \frac{1}{3}$. One method for lower bounding the t for which this happens is as follows: if F is μ-strongly convex and f is ν-smooth, then $\lambda_{F,\alpha}(q)$ is at most $\lambda_{F_t,\alpha}(q)\sqrt{1 + t\nu/\mu} + t\|df_q\|^*_{F,q,\alpha}/\alpha$, and $\|df_q\|^*_{F,q,\alpha}$ can be bounded (for instance) using Lipschitzness of f and strong convexity of F. We leave a more careful analysis of this idea to future work. We note that in the Euclidean setting, the complexity is often bounded in terms of the *asymmetry* of domain D with respect to the point p, see [NN94, Eq. (3.2.24)] for details, but such a bound does not seem to generalize to the Riemannian setting.

- It would be interesting to understand whether there is a suitable notion of primal-dual methods in the Riemannian setting, or a notion of duality which interacts well with self-concordance. While there exists a version of Legendre–Fenchel duality for Hadamard manifolds M, where the dual space is CM^∞, the cone over the boundary at infinity of M discussed above, the conjugate of a convex function need not be convex [Hir22]. Other proposals such as [BHSL$^+$21] require a stronger notion of convexity.

Acknowledgements

References

[ABM08] F. Alvarez, J. Bolte, and J. Munier. A Unifying Local Convergence Result for Newton's Method in Riemannian Manifolds. *Foundations of Computational Mathematics*, 8(2):197–226, April 2008. `doi:10.1007/s10208-006-0221-6`.

[Afs11] Bijan Afsari. Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011. `doi:10.1090/S0002-9939-2010-10541-5`.

[AGL$^+$18] Zeyuan Allen-Zhu, Ankit Garg, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson. Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing*, pages 172–181, 2018. `doi:10.1145/3188745.3188942`.

[AKRS21a] Carlos Améndola, Kathlén Kohn, Philipp Reichenbach, and Anna Seigal. Invariant Theory and Scaling Algorithms for Maximum Likelihood Estimation. *SIAM Journal on Applied Algebra and Geometry*, 5(2):304–337, 2021. `doi:10.1137/20M1328932`.

[AKRS21b] Carlos Améndola, Kathlén Kohn, Philipp Reichenbach, and Anna Seigal. Toric invariant theory for maximum likelihood estimation in log-linear models. *Algebraic Statistics*, 12(2):187–211, 2021. `doi:10.2140/astat.2021.12.187`.

[ALOW17] Zeyuan Allen-Zhu, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson. Much Faster Algorithms for Matrix Scaling. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 890–901, 2017. `doi:10.1109/FOCS.2017.87`.

[AMN$^+$22] Arturo Acuaviva, Visu Makam, Harold Nieuwboer, David Pérez-García, Friedrich Sittner, Michael Walter, and Freek Witteveen. The minimal canonical form of a tensor network, 2022. `arXiv:2209.14358`.

[AMS09] P.-A. Absil, R. Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009. `doi:10.1515/9781400830244`.

[AN13] Marc Arnaudon and Frank Nielsen. On approximating the Riemannian 1-center. *Computational Geometry*, 46(1):93–104, 2013. `doi:10.1016/j.comgeo.2012.04.007`.

[AS20] Kwangjun Ahn and Suvrit Sra. From Nesterov's Estimate Sequence to Riemannian Acceleration. In *Proceedings of Thirty Third Conference on Learning Theory*, pages 84–118. PMLR, 2020.

[Ban38] Stefan Banach. Über homogene Polynome in (L^2). *Studia Mathematica*, 7(1):36–44, 1938.

[BCMW17] Peter Bürgisser, Matthias Christandl, Ketan D. Mulmuley, and Michael Walter. Membership in Moment Polytopes is in NP and coNP. *SIAM Journal on Computing*, 46(3):972–991, January 2017. `doi:10.1137/15M1048859`.

[BD00] Rajendra Bhatia and Chandler Davis. More Operator Versions of the Schwarz Inequality. *Commun. Math. Phys*, 215:239–244, 2000.

[BDWY12] Boaz Barak, Zeev Dvir, Avi Wigderson, and Amir Yehudayoff. Fractional Sylvester–Gallai theorems. *Proceedings of the National Academy of Sciences*, 2012. `doi:10.1073/pnas.1203737109`.

[BE19]   Sébastien Bubeck and Ronen Eldan. The entropic barrier: Exponential families, log-concave geometry, and self-concordance. *Mathematics of Operations Research*, 44(1):264–276, 2019. `doi:10.1287/moor.2017.0923`.

[BFG⁺18] Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Oliveira, Michael Walter, and Avi Wigderson. Efficient algorithms for tensor scaling, quantum marginals and moment polytopes. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 883–897, 2018. `arXiv:1804.04739`, `doi:10.1109/FOCS.2018.00088`.

[BFG⁺19] Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Oliveira, Michael Walter, and Avi Wigderson. Towards a theory of non-commutative optimization: Geodesic first and second order methods for moment maps and polytopes. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 845–861, 2019. Citations refer to the long version on arXiv. `arXiv:1910.12375`, `doi:10.1109/FOCS.2019.00055`.

[BG19]   Nikhil Bansal and Anupam Gupta. Potential-Function Proofs for Gradient Methods. *Theory of Computing*, 15(4):1–32, 2019. `doi:10.4086/toc.2019.v015a004`.

[BGO⁺18] Peter Bürgisser, Ankit Garg, Rafael Oliveira, Michael Walter, and Avi Wigderson. Alternating Minimization, Scaling Algorithms, and the Null-Cone Problem from Invariant Theory. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, volume 94 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 24:1–24:20, 2018. `doi:10.4230/LIPIcs.ITCS.2018.24`.

[BH99]   Martin R. Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1999. `doi:10.1007/978-3-662-12494-9`.

[BHSL⁺21] Ronny Bergmann, Roland Herzog, Maurício Silva Louzeiro, Daniel Tenbrinck, and José Vidal-Núñez. Fenchel Duality Theory and a Primal-Dual Algorithm on Riemannian Manifolds. *Foundations of Computational Mathematics*, 21(6):1465–1504, 2021. `doi:10.1007/s10208-020-09486-5`.

[BLNW20] Peter Bürgisser, Yinan Li, Harold Nieuwboer, and Michael Walter. Interior-point methods for unconstrained geometric programming and scaling problems, 2020. `arXiv:2008.12110`.

[Bou23]   Nicolas Boumal. *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023. URL: https://www.nicolasboumal.net/book, `doi:10.1017/9781009166164`.

[vdB19]   Jan van den Brand. A Deterministic Linear Program Solver in Current Matrix Multiplication Time. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 259–278. Society for Industrial and Applied Mathematics, 2019. `doi:10.1137/1.9781611975994.16`.

[Car72]   B. C. Carlson. The Logarithmic Mean. *The American Mathematical Monthly*, 79(6):615–618, 1972. `doi:10.1080/00029890.1972.11993095`.

[CB22]   Christopher Criscitiello and Nicolas Boumal. Negative curvature obstructs acceleration for geodesically convex optimization, even with exact first-order oracles, 2022. `arXiv:2111.13263`.

[Che21]   Sinho Chewi. The entropic barrier is n-self-concordant, 2021. `arXiv:2112.10947`.

[CLM+16] Michael B. Cohen, Yin Tat Lee, Gary Miller, Jakub Pachocki, and Aaron Sidford. Geometric median in nearly linear time. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC)*, pages 9–21. Association for Computing Machinery, 2016. `doi:10.1145/2897518.2897647`.

[CMTV17] Michael B. Cohen, Aleksander Madry, Dimitris Tsipras, and Adrian Vladu. Matrix Scaling and Balancing via Box Constrained Newton's Method and Interior Point Methods. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 902–913, 2017. `doi:10.1109/FOCS.2017.88`.

[Cut13] Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

[Dan63] George Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, 1963. `doi:10.1515/9781400884179`.

[DGOS18] Zeev Dvir, Ankit Garg, Rafael Oliveira, and József Solymosi. Rank bounds for design matrices with block entries and geometric applications. *Discrete Analysis*, 5(2018):1–24, 2018. `doi:10.19086/da.3118`.

[DP14] Alberto Dolcetti and Donato Pertici. Some differential properties of $\mathrm{GL}_n(\mathbb{R})$ with the trace metric, 2014. `arXiv:1412.4565`.

[DPM03] Jean-Pierre Dedieu, Pierre Priouret, and Gregorio Malajovich. Newton's method on Riemannian manifolds: Covariant alpha theory. *IMA Journal of Numerical Analysis*, 23(3):395–419, July 2003. `doi:10.1093/imanum/23.3.395`.

[DSW14] Zeev Dvir, Shubhangi Saraf, and Avi Wigderson. Improved rank bounds for design matrices and a new proof of Kelly's theorem. *Forum of Mathematics, Sigma*, 2:e4, March 2014. `doi:10.1017/fms.2014.2`.

[Dui99] J. J. Duistermaat. On the boundary behaviour of the Riemannian structure of a self-concordant barrier function, 1999.

[Ebe97] Patrick Eberlein. *Geometry of nonpositively curved manifolds*. University of Chicago Press, 1997.

[FORW21] Cole Franks, Rafael Oliveira, Akshay Ramachandran, and Michael Walter. Near optimal sample complexity for matrix and tensor normal models via geodesic convexity, 2021. `arXiv:2110.07583`.

[Fox15] Daniel JF Fox. A Schwarz lemma for Kähler affine metrics and the canonical potential of a proper convex cone. *Annali di Matematica Pura ed Applicata (1923-)*, 194(1):1–42, 2015. `doi:10.1007/s10231-013-0362-6`.

[FR21] W. Cole Franks and Philipp Reichenbach. Barriers for Recent Methods in Geodesic Optimization. In *36th Computational Complexity Conference (CCC 2021)*, volume 200 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 13:1–13:54, 2021. `doi:10.4230/LIPIcs.CCC.2021.13`.

[Fra18]   Cole Franks. Operator scaling with specified marginals. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC)*, pages 190–203. Association for Computing Machinery, 2018. `doi:10.1145/3188745.3188932`.

[Fri55]   KR Frisch. The logarithmic potential method of convex programming. *Memorandum, University Institute of Economics, Oslo*, 5(6), 1955.

[FS02]   O. P. Ferreira and B. F. Svaiter. Kantorovich's Theorem on Newton's Method in Riemannian Manifolds. *Journal of Complexity*, 18(1):304–329, March 2002. `doi:10.1006/jcom.2001.0582`.

[FVJ09]   P. Thomas Fletcher, Suresh Venkatasubramanian, and Sarang Joshi. The Geometric Median on Riemannian Manifolds with Application to Robust Atlas Estimation. *NeuroImage*, 45:S143–S152, 2009. `doi:10.1016/j.neuroimage.2008.10.052`.

[FW22]   Cole Franks and Michael Walter. Minimal length in an orbit closure as a semiclassical limit, 2022. `arXiv:2004.14872`.

[GGOW16] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. A Deterministic Polynomial Time Algorithm for Non-commutative Rational Identity Testing. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 109–117. IEEE, 2016. `doi:10.1109/FOCS.2016.95`.

[GGOW17] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. Algorithmic and optimization aspects of Brascamp-Lieb inequalities, via operator scaling. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC)*, pages 397–409. Association for Computing Machinery, 2017. `doi:10.1145/3055399.3055458`.

[GGOW20] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. Operator Scaling: Theory and Applications. *Foundations of Computational Mathematics*, 20(2):223–290, 2020. `doi:10.1007/s10208-019-09417-z`.

[GMS⁺86] Philip E. Gill, Walter Murray, Michael A. Saunders, J. A. Tomlin, and Margaret H. Wright. On projected newton barrier methods for linear programming and an equivalence to Karmarkar's projective method. *Mathematical Programming*, 36(2):183–209, 1986. `doi:10.1007/BF02592025`.

[Gur04]   Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448–484, 2004. `doi:10.1016/j.jcss.2004.06.003`.

[Hel79]   Sigurdur Helgason. *Differential geometry, Lie groups, and symmetric spaces*. Academic Press, 1979.

[Hil14]   Roland Hildebrand. Canonical barriers on convex cones. *Mathematics of operations research*, 39(3):841–850, 2014. `doi:10.1287/moor.2013.0640`.

[Hir22]   Hiroshi Hirai. Convex analysis on Hadamard spaces and scaling problems, 2022. `arXiv:2203.03193`.

[HM21a]   Linus Hamilton and Ankur Moitra. No-go theorem for acceleration in the hyperbolic plane, 2021. `arXiv:2101.05657`.

[HM21b] Linus Hamilton and Ankur Moitra. The Paulsen problem made simple. *Israel Journal of Mathematics*, 246(1):299–313, December 2021. `doi:10.1007/s11856-021-2245-7`.

[IMW17] Christian Ikenmeyer, Ketan D. Mulmuley, and Michael Walter. On vanishing of Kronecker coefficients. *computational complexity*, 26(4):949–992, 2017. `arXiv:1507.02955`, `doi:10.1007/s00037-017-0158-y`.

[Inn82] Nobuhiro Innami. Splitting theorems of riemannian manifolds. *Compositio Mathematica*, 47(3):237–247, 1982.

[Ji07] Huibo Ji. *Optimization approaches on smooth manifolds*. PhD thesis, Australian National University, 2007.

[JMJ07] Danchi Jiang, John B Moore, and Huibo Ji. Self-concordant functions for optimization on smooth manifolds. *Journal of Global Optimization*, 38(3):437–457, 2007. `doi:10.1007/s10898-006-9095-z`.

[Kar84a] N. Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the 16th Annual ACM Symposium on Theory of Computing (STOC)*, pages 302–311. Association for Computing Machinery, 1984. `doi:10.1145/800057.808695`.

[Kar84b] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984. `doi:10.1007/BF02579150`.

[Kha80] L. G. Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72, 1980. `doi:10.1016/0041-5553(80)90061-0`.

[KLLR18] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, and Akshay Ramachandran. The Paulsen problem, continuous operator scaling, and smoothed analysis. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC)*, pages 182–189. Association for Computing Machinery, June 2018. `doi:10.1145/3188745.3188794`.

[KMY04] Piyush Kumar, Joseph S. B. Mitchell, and E. Alper Yildirim. Approximate minimum enclosing balls in high dimensions using core-sets. *ACM Journal of Experimental Algorithmics*, 8:1.1–es, December 2004. `doi:10.1145/996546.996548`.

[KN79] George Kempf and Linda Ness. The length of vectors in representation spaces. In Knud Lønsted, editor, *Algebraic Geometry*, volume 732, pages 233–243. Springer Berlin Heidelberg, 1979. `doi:10.1007/BFb0066647`.

[Lee18] John M. Lee. *Introduction to Riemannian manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, 2018. `doi:10.1007/978-3-319-91755-9`.

[LS20] Yin Tat Lee and Aaron Sidford. Solving Linear Programs with Sqrt(rank) Linear System Solves, 2020. `arXiv:1910.08033`.

[LSW00] Nathan Linial, Alex Samorodnitsky, and Avi Wigderson. A Deterministic Strongly Polynomial Algorithm for Matrix Scaling and Approximate Permanents. *Combinatorica*, 20(4):545–568, 2000. `doi:10.1007/s004930070007`.

[LY22] Zhijian Lai and Akiko Yoshise. Riemannian Interior Point Methods for Constrained Optimization on Manifolds, 2022. `arXiv:2203.09762`.

[MFK94] David Mumford, John Fogarty, and Frances Kirwan. *Geometric Invariant Theory*. Springer Science & Business Media, 1994.

[Nes18] Yurii Nesterov. *Lectures on convex optimization*, volume 137. Springer, 2018. `doi: 10.1007/978-3-319-91578-4`.

[NH15] Frank Nielsen and Gaëtan Hadjeres. Approximating Covering and Minimum Enclosing Balls in Hyperbolic Geometry. In *Geometric Science of Information*, Lecture Notes in Computer Science, pages 586–594, Cham, 2015. Springer International Publishing. `doi: 10.1007/978-3-319-25040-3_63`.

[NM84] Linda Ness and David Mumford. A Stratification of the Null Cone Via the Moment Map. *American Journal of Mathematics*, 106(6):1281, 1984. `doi:10.2307/2374395`.

[NN94] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. SIAM, 1994. `doi: 10.1137/1.9781611970791`.

[NN08] Y. Nesterov and A. Nemirovski. Primal Central Paths and Riemannian Distances for Convex Sets. *Foundations of Computational Mathematics*, 8(5):533–560, 2008. `doi:10. 1007/s10208-007-9019-4`.

[NT02] Yu. E. Nesterov and M. J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Foundations of Computational Mathematics*, 2(4):333–361, 2002. `doi:10.1007/s102080010032`.

[Ost78] Lawrence M. Ostresh. On the Convergence of a Class of Iterative Methods for Solving the Weber Location Problem. *Operations Research*, 26(4):597–609, 1978. `doi:10.1287/opre. 26.4.597`.

[Per20] Frank Permenter. A geodesic interior-point method for linear optimization over symmetric cones, 2020. `arXiv:2008.08047`.

[Per22] Frank Permenter. Log-domain interior-point methods for convex quadratic programming, 2022. `arXiv:2212.02294`.

[Ren01] James Renegar. *A mathematical view of interior-point methods in convex optimization*. MPS/SIAM Series on Optimization. SIAM, 2001. `doi:10.1137/1.9780898718812`.

[Rus19] Alexander Rusciano. A Riemannian Corollary of Helly's Theorem, 2019. `arXiv:1804.10738`.

[Sak96] Takashi Sakai. *Riemannian Geometry*. American Mathematical Society, 1996.

[Sat21] Hiroyuki Sato. *Riemannian Optimization and Its Applications*. SpringerBriefs in Electrical and Computer Engineering. Springer International Publishing, 2021. `doi:10.1007/ 978-3-030-62391-3`.

[SH15] Suvrit Sra and Reshad Hosseini. Conic geometric optimisation on the manifold of positive definite matrices. *SIAM Journal on Optimization*, 25(1):713–739, 2015. `arXiv:1312.1039`, `doi:10.1137/140978168`.

[SV14] Mohit Singh and Nisheeth K. Vishnoi. Entropy, optimization and counting. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 50–59. Association for Computing Machinery, 2014. `doi:10.1145/2591796.2591803`.

[SW22] Vishwak Srinivasan and Ashia Wilson. Sufficient conditions for non-asymptotic convergence of Riemannian optimisation methods, 2022. `arXiv:2212.05972`.

[Udr94] Constantin Udrişte. *Convex functions and optimization methods on Riemannian manifolds*. Springer Netherlands, 1994. `doi:10.1007/978-94-015-8390-9`.

[Udr97] Constantin Udrişte. Optimization Methods on Riemannian Manifolds. *Algebras, Groups and Geometries*, 14:339–359, 1997.

[Wal14] Michael Walter. *Multipartite quantum states and their marginals*. PhD thesis, ETH Zurich, 2014. `arXiv:1410.6820`, `doi:10.3929/ethz-a-010250985`.

[Wal17] Nolan R. Wallach. *Geometric Invariant Theory*. Universitext. Springer International Publishing, Cham, 2017. `doi:10.1007/978-3-319-65907-7`.

[WDGC13] Michael Walter, Brent Doran, David Gross, and Matthias Christandl. Entanglement Polytopes: Multiparticle Entanglement from Single-Particle Information. *Science*, 340(6137):1205–1208, 2013. `arXiv:1208.0365`, `doi:10.1126/science.1232957`.

[Wei37] E. Weiszfeld. Sur le point pour lequel la Somme des distances de n points donnés est minimum. *Tohoku Mathematical Journal, First Series*, 43:355–386, 1937.

[WS22] Melanie Weber and Suvrit Sra. Riemannian Optimization via Frank-Wolfe Methods. *Mathematical Programming*, 2022. `doi:10.1007/s10107-022-01840-5`.

[XY97] Guoliang Xue and Yinyu Ye. An Efficient Algorithm for Minimizing a Sum of Euclidean Norms with Applications. *SIAM Journal on Optimization*, 7(4):1017–1036, 1997. `doi:10.1137/S1052623495288362`.

[Yan10] Le Yang. Riemannian median and its estimation. *LMS Journal of Computation and Mathematics*, 13:461–479, 2010. `doi:10.1112/S1461157020090531`.

[ZS16] Hongyi Zhang and Suvrit Sra. First-order Methods for Geodesically Convex Optimization. In *Conference on Learning Theory*, pages 1617–1638. PMLR, 2016.