Sampling Polytopes with Riemannian HMC: Faster Mixing via the Lewis Weights Barrier

Khashayar Gatmiry GATMIRY@MIT.EDU
MIT

Jonathan Kelner KELNER@MIT.EDU
MIT

Santosh S. Vempala VEMPALA@GATECH.EDU
Georgia Tech

Editors: Shipra Agrawal and Aaron Roth

Abstract

We analyze Riemannian Hamiltonian Monte Carlo (RHMC) on a manifold endowed with the metric defined by the Hessian of a convex barrier function and apply it to sample a polytope defined by m inequalities in $\mathbb{R}^n$. The advantage of RHMC over Euclidean methods such as the ball walk, hit-and-run and the Dikin walk is in its ability to take longer steps. However, in all previous work, the mixing rate of RHMC has a linear dependence on the number of inequalities. We introduce a hybrid of the Lewis weight barrier and the standard logarithmic barrier and prove that the mixing rate for the corresponding RHMC is bounded by $\tilde{O}(m^{1/3}n^{4/3})$, improving on the previous best bound of $\tilde{O}(mn^{2/3})$ (based on the log barrier). This continues the general parallels between optimization and sampling, with the latter typically leading to new tools and requiring more refined analysis. To prove our main results, we overcomes several challenges relating to the smoothness of Hamiltonian curves and self-concordance properties of the barrier. In the process, we give a general framework for the analysis of Markov chains on Riemannian manifolds, derive new smoothness bounds on Hamiltonian curves, a central topic of comparison geometry, and extend self-concordance theory to the infinity norm, which gives sharper bounds; these properties all appear to be of independent interest.

Keywords: Riemannian Hamiltonian Monte Carlo, Lewis weights, Markov chains, Calabi estimates, Self-concordance, Interior point theory, Isoperimetric inequality, Geometric sampling

1. Introduction

Generating nearly uniform random samples from a high-dimensional polytope is a fundamental algorithmic problem with a rich history and powerful applications, notably including the only known fully polynomial-time approximation schemes for computing a polytope's volume. All efficient algorithms known for this problem work by designing a Markov chain whose stationary distribution is uniform over the polytope and showing that it mixes in a small number of steps.

In this paper, our main result is that we can construct such a Markov chain with an improved bound on its mixing time. For a polytope given by m linear inequalities in $\mathbb{R}^n$, we describe chain that mixes in $\tilde{O}\left(m^{1/3}n^{4/3}\right)$ steps, improving on the best previous bound of $\tilde{O}\left(mn^{2/3}\right)$. This allows us to approximate the volume within relative error ϵ using $\tilde{O}\left(m^{1/3}n^{4/3}/\epsilon^2\right)$ steps, which is a similar improvement over the best existing bound of $\tilde{O}\left(mn^{2/3}/\epsilon^2\right)$.

1.1. Background and Related Work

In their seminal work Dyer et al. (1991), Dyer, Frieze and Kannan gave the first polynomial-time algorithm for this problem, as well as for the more general problem of sampling from a convex body specified by a membership oracle. The Markov chain in their algorithm was a *grid walk*, which takes steps along the edges of the graph obtained by intersecting the convex body with a discrete grid supported on $\delta\mathbb{Z}^n$ for some $\delta = 1/\mathrm{poly}(n)$. This graph is heavily dependent on the coordinate system—its diameter is proportional to the diameter of the convex body, and its conductance can be arbitrarily small if the convex body is scaled so that is very long in some directions but short in others. However, they showed that, if one changes to a basis in which the convex body is appropriately "well-rounded," the grid walk mixes in polynomial time and that one can use a random sample from the grid to obtain a one from the convex body.

The polynomial for the mixing time in Dyer et al. (1991) was quite large, and a sequence of later papers improved this by modifying the Markov chains and refining the analysis. Because one often wants to draw many samples from the body, these papers typically provide two bounds on the number of steps required: a bound when starting from an arbitrary point and including the cost of any preprocessing; and a bound when given a *warm start*, where the preprocessing has already been performed and the starting point is drawn from a distribution that is not too far from uniform.

In Kannan et al. (1997), Kannan, Lovász, and Simonovits showed that a *ball walk* whose steps are chosen uniformly from a Euclidean ball around the current point mixes in $\tilde{O}(n^3)$ steps from a warm start and $\tilde{O}(n^5)$ steps from an arbitrary starting point and including preprocessing. Later, Lovász and Vempala Lovász and Vempala (2006) studied the "hit-and-run" walk, which chooses a line in a random direction from the current point and then picks the next point randomly from the intersection of this line with the body, and they showed it also mixed in $\tilde{O}(n^3)$ steps from a warm start but needed only $\tilde{O}(n^4)$ steps for first sample and preprocessing. These algorithms work on general convex bodies presented by oracles, but like the grid walk, they are strongly coordinate dependent, and they thus require strong additional assumptions about the coordinate system. In particular, analyses of these algorithms typically assume that the body is close to being isotropic, i.e., that the covariance matrix of a random sample from the body is approximately the identity, and applying these algorithms to more general bodies requires costly preprocessing.

The dependence on the coordinate system in the aforementioned Markov chains comes from the dependence of the transition probabilities on the extrinsic geometry of the ambient Euclidean space. The impact of this extends beyond the overhead from the isotropy requirements. The geometry of the ambient space does not incorporate any information about how close a point is to the boundary, which typically leads to difficulties making progress with steps near the boundary. For example, if one is running a ball walk with step radius δ an n-dimensional cube, and the current point is some distance $d \ll \delta$ from one of the corners, a random point from the radius δ ball will lie outside the cube with probabability exponentially close to 1, so naively trying random points until obtaining one in the cube would take a large number of tries. Moreover, even if one could sample a random point in the intersection of the ball with the cube, restricting the step to points inside the cube would distort the stationary distribution, and it would no longer be uniform. Remedying such difficulties typically involves (depending on the paper) some combination of taking smaller steps, enlarging the convex body (and failing if the walk ends up at a point outside the original body), and employing rejection sampling or a Metropolis filter to correct the stationary probabilities, all of which increase the required number of steps.

For polytopes specified by an explicit collection of linear constraints, one can use the barrier functions employed by interior point methods to design random walks whose steps depend only on the intrinsic geometry of the polytope and are independent of the basis chosen for the ambient space. The idea behind these random walks is to use the Hessian of the barrier function to define a local norm/Riemannian metric on the interior of the polytope and specify the steps in terms of the resulting geometry. This mitigates some of the problems described above and has led to Markov chains whose mixing times grow with the number of constraints but depend more mildly on the dimension.

In the first such work, Kannan and Narayanan Kannan and Narayanan (2012) introduced the *Dikin walk* and gave a mixing time bound of $O(mn)$ from a warm start for a polytope with m facets in $\mathbb{R}^n$. This walk is similar to the ball walk, but it chooses its steps from Dikin ellipsoids, which are balls with respect to the Hessian of the standard logarithmic barrier function on the polytope. In Laddha et al. (2020), Laddha, Lee, and Vempala studied the analogous walk with respect to any self-concordant barrier and showed that it mixes in $\tilde{O}(n\bar{\nu})$ steps, where $\bar{\nu}$ is a parameter they called the *barrier parameter*. By bounding this parameter for a different barrier function (a variant of a barrier due to Lee and Sidford Lee and Sidford (2014)), they obtained an improved mixing rate bound of $\tilde{O}(n^2)$.

In 2017, Lee and Vempala Lee and Vempala (2017) reduced the mixing rate to $\tilde{O}\left(mn^{3/4}\right)$ using a process they called the *geodesic walk*. Similar to the Dikin Walk, the steps are constructed using the Hessian of a barrier function. However, instead of using the Hessian to define a Euclidean ellipsoid, they use it to define a Riemannian metric, and they then solve a differential equation in each step to follow geodesics on the resulting manifold. These geodesics tend to avoid the polytope's boundary, which allows longer steps in each iteration.

In 2018, Lee and Vempala Lee and Vempala (2018) improved this to $\tilde{O}\left(mn^{2/3}\right)$ using Riemannian Hamiltonian Monte Carlo (RHMC) Girolami and Calderhead (2011), which is the class of processes we will use in this paper. While there is a large literature on using RHMC and related methods to sample *smooth* densities Dalalyan (2017); Durmus et al. (2019); Chewi et al. (2022); Vempala and Wibisono (2019); Li et al. (2022), there are relatively few provable results about applying it in constrained non-smooth settings like polytope sampling. Roughly speaking, this improvement over the geodesic walk came from RHMC's ability to avoid the use of a Metropolis filter, which the geodesic walk requires in order to obtain the correct stationary distribution (even when the target distribution is uniform). RHMC chooses its trajectories according to a differential equation that, remarkably, yields a reversible random walk with the desired stationary distribution, thus eliminating the need for a Metropolis filter and allowing greater progress in each step.

While barriers with better parameters have led to improved mixing times for the Dikin walk, obtaining similar improvements for the geodesic walk or RHMC have remained elusive, and improving upon the $\tilde{O}(mn^{2/3})$ bound attained by RHMC using the standard log barrier has been a major open problem for the past 5+ years.

The core issue that prevents the authors of Lee and Vempala (2017) and Lee and Vempala (2018) from using other barriers in place of the log barrier is that the geodesic walk and RHMC use their barrier functions in a fundamentally different way from how they are used in the Dikin walk. The Dikin walk, like the interior point methods for which self-concordant barriers were originally defined, uses the Hessian of the barrier function at each point to specify an ellipsoid centered at the point and contained in the polytope and chooses its next iterate from this ellipsoid.

The geodesic walk and RHMC, however, use the barrier function to define the local geometry of a manifold, and they take a step by simulating the trajectory of a particle according to a corresponding second-order differential equation. The solution depends on the geometry at every point of the trajectory, rather than just at the point where the particle was at the beginning of the iteration. As such, the steps of the random walk depend on the geometric structure at all of the points of the trajectory, and analyzing them requires one to understand how the locally-defined structure at each point relates to those at other nearby points. This leads to a dependence on higher-order derivatives of the barrier function than the ones that self-concordance was designed to control. As a result, self-concordance by itself does not seem to be sufficient in this setting, and analyzing these walks requires the authors of Lee and Vempala (2017) and Lee and Vempala (2018) to rely on new but ad hoc arguments tied to specific properties of the logarithmic barrier.

Year	Algorithm	Steps
1997 Kannan et al. (1997)	Ball walk#	n^3 $(+n^5)$
2003 Lovász and Vempala (2006)	Hit-and-run#	n^3 $(+n^4)$
2009 Kannan and Narayanan (2012)	Dikin walk	mn
2017 Lee and Vempala (2017)	Geodesic walk	$mn^{3/4}$
2018 Lee and Vempala (2018)	RHMC with log barrier	$mn^{2/3}$
2020 Laddha et al. (2020)	Weighted Dikin walk	n^2
2021 Jia et al. (2021)	Ball walk#	n^2 $(+n^3)$
This paper	RHMC with Hybrid barrier	$m^{1/3}n^{4/3}$

Table 1: The complexity of uniformly sampling a polytope from a warm start. All algorithms have a logarithmic dependence on the warm start parameter and each uses $\widetilde{O}(n)$ bit of randomness. The entries marked # are for general convex bodies presented by oracles, while the rest are for polytopes. The additive terms are pre-processing costs for rounding the polytope.

1.2. Background on Riemannian Hamiltonian Monte Carlo

The motivation for RHMC comes from the Hamiltonian formulation of classical Newtonian mechanics. Hamiltonian mechanics parameterizes a physical system in terms of a *position* vector x and a corresponding *momentum* vector v (which is also referred to as "velocity" in some prior work on sampling polytopes with RHMC). The physics of the system are encoded in its *Hamiltonian* $H(x, v)$, which is simply the energy of the system written as a function of x and v, and its time evolution is determined by *Hamilton's equations*:

$$\begin{aligned}\frac{dx}{dt} &= \frac{\partial H}{\partial v}(x, v)\\ \frac{dv}{dt} &= -\frac{\partial H}{\partial x}(x, v).\end{aligned}$$

With the appropriate choice of H, these reproduce Newton's laws of motion, but they also generalize quite broadly, including to Riemannian manifolds.

In RHMC, one defines a Markov chain by choosing a Hamiltonian that appropriately encodes the target distribution. At each step, the Markov chain chooses a random momentum vector and then finds the next point by numerically solving a differential equation to follow the trajectory given by Hamilton's equations.

One can show that the value of the Hamiltonian and the volume element in the space of pairs (x, v) are conserved along the trajectory, which can be used to show that the trajectories are preserved by time reversal. One can then use this to show that, if one uses the Hamiltonian defined below, the marginal distribution of x will converge to the desired target distribution without requiring a Metropolis filter. (See Girolami and Calderhead (2011) for the derivation for general RHMC and Lee and Vempala (2018) for the specific class of Hamiltonians given below.)

More precisely, let the Hamiltonian at a point $x \in \mathbb{R}^n$ for a vector $v \in \mathbb{R}^n$ be defined as

$$H(x, v) = f(x) + \frac{1}{2}v^\top g^{-1}(x)v + \frac{1}{2}\log\det g(x)$$

where $g(x)$ is a positive definite matrix defining a Riemannian metric at each point x as $\|u\|_g \triangleq \|u\|_{g(x)} \triangleq \sqrt{u^\top g(x)u}$, and the target density to be sampled is proportional to e^{-f} restricted to the support of g. One step of RHMC consists of the following: first pick v from the Gaussian $\mathcal{N}(x, g(x)^{-1})$. Then for time δ

follow the Hamiltonian curve jointly on (x, v):

$$\begin{aligned} \frac{dx}{dt} &= \frac{\partial H}{\partial v}(x,v) = g^{-1}(x)v \\ \frac{dv}{dt} &= -\frac{\partial H}{\partial x}(x,v) = -\nabla f(x) + \frac{1}{2}\mathrm{tr}(g(x)^{-1}Dg(x)) - \frac{1}{2}Dg(x)\left[\frac{dx}{dt}, \frac{dx}{dt}\right]. \end{aligned} \tag{1}$$

The final x at time δ is the sampled point from the Markov Kernel. A natural choice for the metric g turns out to be the Hessian of a self-concordant barrier function inside the polytope $\mathcal{P}$. The standard logarithmic barrier, $\phi_\ell(x) = -\sum_{i=1}^m \log(a_i^\top x - b_i)$, was used in Lee and Vempala (2018) to prove that the resulting RHMC mixes in $mn^{2/3}$ steps, where the polytope is defined by the inequalities $\{{a_i}^\top x \geq b_i\}_{i=1}^m$. Improving on this bound is our motivating open problem.

1.3. Results

Algorithmic results. Our main algorithmic result in this paper is the construction of a random walk based on RHMC for sampling and approximating the volume of polytopes using only $\tilde{O}(m^{1/3}n^{4/3})$ steps. Using a Hamiltonian based on the standard logarithmic barrier yields a mixing rate that depends linearly on m, the number of inequalities. We improve on this by developing a theoretical framework for designing and analyzing barriers for RHMC.

Our framework is motivated by the ways in which the requirements of RHMC differ from those of the Dikin walk. In the case of Dikin walk, we care about how much volume two nearby ellipsoids defined by our metric have in common, which can be controlled given that the first derivative of the metric is bounded, a property that self-concordant barriers possess Nesterov and Nemirovskii (1994). Namely, the derivative of the Hessian of a self-concordant barrier ϕ is controlled by the Hessian itself, which can be seen as a property of the metric $g = \nabla^2\phi$,

$$-\|v\|_g g \preccurlyeq \mathrm{D}g(v) \preccurlyeq \|v\|_g g,$$

where $\mathrm{D}g(v)$ is the directional derivative of g along direction v.

On the other hand, to define the Markov kernel for RHMC, one reparameterizes the open set inside the polytope by the Hamiltonian trajectories, which map the tangent space of the current point to the manifold. This means the density of the RHMC Markov kernel depends on the distortion properties of this map, such as how fast the Hamiltonian curves with different initial conditions converge or diverge. Therefore, showing that this density is Lipschitz is linked to the geometry imposed by the Hessian of the barrier, and analyzing the second-order ODE regarding RHMC demands estimates on more than just the first order-derivative of the metric. As a result, the existing notions of self-concordance are insufficient for bounding the spectral gap of RHMC.

In this paper, we define a stronger notion, *third-order ℓ_∞-self-concordance*, that is stringent enough to guarantee the stronger properties required by RHMC, but we show that it still admits a construction that improves upon the logarithmic barrier. Third-order ℓ_∞-self-concordance strengthens standard self-concordance in two ways: it controls the higher order derivatives of g up to third order; and, instead of the local ellipsoidal norm $\|.\|_g$ that is conventionally used in self-concordance, we measure the spectral change of the metric in the local infinity norm $\|.\|_{x,\infty}$, which we define next.

Definition 1 *For arbitrary vector v we define the local norm $\|.\|_{x,\infty}$ at point $x \in \mathcal{P}$ as the maximum relative change of the distance of x to an arbitrary facet of the polytope after taking step v. Formally:*

$$\|v\|_{x,\infty} \triangleq \|S_x^{-1}\mathrm{A}v\|_\infty,$$

where the polytope is defined by the inequalities $Ax \geq b$ and S_x is the diagonal matrix whose entries are the slacks of the linear constraints at point x, i.e. $(S_x)_{ii} \triangleq a_i^\top x - b_i$.

An intuitive description of $\|.\|_{x,\infty}$ is via its unit ball; namely, $\|.\|_{x,\infty}$ is the unique norm whose unit ball is the symmetrized polytope $\mathcal{P} \cap 2x - \mathcal{P}$ around x, as illustrated in Figure 2 in Appendix A. ($2x - \mathcal{P}$ is the reflection of $\mathcal{P}$ around x.) Using this generalized notion, we are not only able to control the change of density of the RHMC kernel (Section D), but also prove the stability of the Hamiltonian curves (Lemma 16), which is required for bounding the conductance.

Definition 2 (Third-order ℓ_∞-self-concordance) *We say that ϕ is c-*third-order ℓ_∞-self-concordant *if its Hessian $g = \nabla^2\phi$ obeys*

$$
\begin{aligned}
-c\|v\|_{x,\infty} g(x) \preccurlyeq \mathrm{D}g(x)[v] \preccurlyeq c\|v\|_{x,\infty} g(x), \\
-c\|v\|_{x,\infty}\|z\|_{x,\infty} g(x) \preccurlyeq \mathrm{D}^2 g(x)[v,z] \preccurlyeq c\|v\|_{x,\infty}\|z\|_{x,\infty} g(x), \\
-c\|v\|_{x,\infty}\|z\|_{x,\infty}\|u\|_{x,\infty} g(x) \preccurlyeq \mathrm{D}^3 g(x)[v,z,u] \preccurlyeq c\|v\|_{x,\infty}\|z\|_{x,\infty}\|u\|_{x,\infty} g(x),.
\end{aligned} \tag{2}
$$

*Furthermore, we say that ϕ is a (c,ν)-*third-order ℓ_∞-self-concordant barrier *if in addition to the estimates (2) the norm of its gradient in the local norm is bounded as*

$$
(\mathrm{D}\phi)^\top g^{-1} \mathrm{D}\phi \le \nu,
$$

where $\mathrm{D}\phi$ refers to the Euclidean gradient of ϕ.

Here $\preccurlyeq$ is the Löwner ordering between matrices ignoring logarithmic factors. Our second major contribution is to construct a barrier for polytopes that satisfies third-order ℓ_∞-self-concordance. Namely, we construct a hybrid barrier inside the polytope based on the *Lewis weight barrier*

$$
\phi_p(x) \triangleq \log\det\left(\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p} \mathrm{A}_x\right), \tag{3}
$$

where $\mathbf{W}_x$ is a diagonal matrix whose diagonal entries are the *Lewis weights* of the rescaled matrix $\mathrm{A}_x = {S_x}^{-1}\mathrm{A}$, which we define in Section 1.4.

The hybrid barrier ϕ for a polytope is then defined as a combination of the Lewis weight barrier and the log barrier. This combination is necessary so that the resulting manifold with metric $g = \nabla^2\phi$ satisfies a suitable isoperimetric inequality.

Definition 3 (Hybrid barrier) *We define the* hybrid *barrier ϕ inside a polytope $Ax \ge b$ as*

$$
\phi(x) \triangleq -\left(\frac{m}{n}\right)^{\frac{2}{p+2}} \left(\log\det \mathrm{A}_x^\top \mathbf{W}_x^{1-2/p} \mathrm{A}_x + \frac{n}{m}\sum_i \log(s_i)\right), \tag{4}
$$

where $s_i = a_i^\top x - b_i$ are the slacks at point x. We denote the normalizing factor of ϕ by $\alpha_0 \triangleq (\frac{m}{n})^{\frac{2}{p+2}}$.

Our main theorem is a bound on the mixing rate of RHMC with this hybrid barrier.

Theorem 4 (Mixing) *Given a polytope $\mathcal{P}$, let π be the distribution with density proportional to $e^{-\alpha\phi(x)}$ over the open set inside $\mathcal{P}$. Then, RHMC with stationary distribution π on the manifold of the open set inside P equipped with metric g defined by the Hessian of the hybrid barrier ϕ with $p = 4 - (1/\log(m))$ has mixing rate bounded by*

$$
\tilde{O}\left(\min\{\alpha^{-1}n^{2/3} + \alpha^{-1/3}n^{5/9}m^{1/9} + n^{1/3}m^{1/6}, n^{4/3}m^{1/3}\}\right).
$$

In particular, for the uniform distribution over $\mathcal{P}$ (i.e. $\alpha = 0$), the mixing rate is $\tilde{O}\left(m^{1/3}n^{4/3}\right)$.

Specifically, the Markov chain starting at π_0 reaches π_t with TV-distance at most ϵ to the target after

$$
\tilde{O}\left(m^{1/3}n^{4/3}\log(M/\epsilon)^2\log(M)\log\log(M/\epsilon)^2\right) \tag{5}
$$

steps, where $M \triangleq sup_{x\in P}\frac{d\pi_0(x)}{d\pi(x)}$ and $\tilde{O}$ in Equation (5) hides $\mathrm{polylog}(m)$ *factors.*

Note that without a warm start, the $\log(M)$ dependence in Theorem 4 could be another factor of n to the mixing time. However, applying the Gaussian Cooling framework Cousins and Vempala (2018) extended to manifolds Lee and Vempala (2018) lets us sample from $e^{-\alpha\phi}$ for any α without a warm start penalty, and also allows us to compute the volume of the polytope without a significant overhead. (Recent work Kook and Vempala (2023) shows how to leverage the Gaussian Cooling method in more general metrics and for avoiding the warm start penalty for sampling also,)

Corollary 5 (Any start; Volume) *For the manifold Gaussian Cooling scheme in Lee and Vempala (2018) with the hybrid barrier (4) applied to sample from the density $e^{-\alpha\phi(x)}$ inside a given polytope starting from $\arg\min \phi(x)$, the total number of RHMC steps for any $\alpha \geq 0$ is bounded by*

$$\tilde{O}\left(m^{1/3}n^{4/3}\log(1/\epsilon)^2 \log\log(1/\epsilon)^2\right),$$

Moreover, to compute the integral of $e^{-\alpha\phi}$ in the polytope and in particular the volume of the polytope up to multiplicative error $1 \pm \epsilon'$, the total number of RHMC steps is bounded by $\tilde{O}(m^{1/3}n^{4/3}/\epsilon'^2)$.

This improves on the previous best bound of $mn^{2/3}$ due to Lee and Vempala (2018) based on the standard logarithmic barrier.

Geometric results. The proof of Theorem 4 requires the development of several technical ingredients. We summarize a few that are likely to be of independent interest.

The first is a new isoperimetric inequality for this hybrid barrier, which we prove in Section G. (See Section A.3 for the definition of the isoperimetric constant.)

Theorem 6 *[Isoperimetry of the hybrid barrier] Let g be the metric corresponding to the Hessian of the hybrid barrier, with support given by a polytope defined by m inequalities in $\mathbb{R}^n$. Then for $\alpha \geq 0$, the distribution with density proportional to $e^{-\alpha\phi}$ has isoperimetric constant at least*

$$\max\{\frac{1}{\sqrt{n}}(\frac{n}{m})^{\frac{1}{p+2}},\ poly(\frac{1}{4/p-1})\sqrt{\alpha}\}.$$

Moreover, in order to use the abstract framework that we introduce in this work to control the change of the RHMC Markov kernel, we establish the third-order ℓ_∞-self-concordance of the hybrid barrier defined in Equation (4), which we prove in Section B.

Theorem 7 (Third-order ℓ_∞-self-concordance of the hybrid barrier) *The hybrid barrier, defined in (4), is a $(c_2, \alpha_0 n)$-third-order ℓ_∞-self-concordant barrier where $c_2 = poly(\frac{1}{4/p-1})$. In particular, with our choice $p = 4 - (1/\log(m))$ in Theorem 4 we have $c_2 = polylog(m)$.*

These estimates allow us to prove important smoothness properties of certain quantities on the manifold that we are interested in. As far as we know, this is the first proof of such regularity for higher order derivatives of the Lewis weight barrier. The main challenge we face to prove the third-order ℓ_∞-self-concordance is estimating higher order derivatives of the Lewis weights in the PSD cone, which we do in Sections B and F.

We remark that our notion of third-order ℓ_∞-self-concordance is a strengthening of a well-studied notion in differential geometry. We show in Lemma 95 that $\|v\|_{x,\infty} \leq \|v\|_g$ when g is the metric derived from the hybrid barrier. This implies the following corollary, which says that we can obtain the same third-order derivative estimates of g when the norm $\|.\|_{x,\infty}$ is replaced by $\|.\|_g$ in Theorem 7.

Corollary 8 (Calabi estimates for the hybrid barrier) *The metric g of the hybrid barrier (4) satisfies Calabi estimates up to third order, namely*

$$-c_2\|v\|_g g \preccurlyeq \mathrm{D}g(v) \preccurlyeq c_2\|v\|_g g$$
$$-c_2\|v\|_g\|z\|_g g \preccurlyeq \mathrm{D}^2 g(v,z) \preccurlyeq c_2\|v\|_g\|z\|_g g$$
$$-c_2\|v\|_g\|z\|_g\|u\|_g g \preccurlyeq \mathrm{D}^3 g(v,z,u) \preccurlyeq c_2\|v\|_g\|z\|_g\|u\|_g g,$$

where $c_2 = poly(\frac{1}{4/p-1})$.

These type of estimates on the derivatives of the metric are known as the Calabi estimates in the differential geometry literature Székelyhidi (2014); Wang et al. (2006). It turns out that the Calabi-type estimates in Corollary 8 are insufficient to improve the mixing rate, which is why we develop the third-order ℓ_∞-self-concordance for our hybrid barrier to further exploit the randomness of the Hamiltonian curves.

1.4. Technical Overview

Mixing and Conductance. Our general approach to bounding the mixing rate is based on bounding the conductance Lovász and Simonovits (1993). The standard approach to bounding the conductance of geometric walks of this type is to show an isoperimetric inequality for the underlying metric space and then prove that steps of the random walk behave well with respect to the underlying metric. Formally, we show two properties for the manifold $\mathcal{M}$ obtained by equipping the interior of the polytope $\mathcal{P}$ with the metric $g = \nabla^2\phi$:

- **Isoperimetry.** The target density $e^{-\alpha\phi(x)}$ has a good isoperimetry constant on $\mathcal{M}$.
- **One-Step Coupling.** The one-step distributions of the Markov chain given two close-by points x_0, x_1 on the manifold are close in TV-distance. Namely, for some parameter $\delta > 0$, after excluding a tiny set $S^c \subseteq \mathcal{M}$, given any two points $x_0, x_1 \in \mathcal{S}$ with $d(x_0, x_1) \le \delta$ we show

$$TV(\mathcal{T}_{x_0}, \mathcal{T}_{x_1}) \le 0.01, \tag{6}$$

where $\mathcal{T}_x$ denotes the Markov kernel starting from x.

Isoperimetry. The log barrier metric gives an isoperimetric coefficient of $1/\sqrt{m}$, which leads to a factor of m in the conductance. In principle, this can be improved to $\tilde{O}(n)$ by using a barrier with barrier parameter $\nu = \tilde{O}(n)$, as the general bound on the isoperimetry is $1/\sqrt{\nu}$ for any strongly self-concordant barrier with barrier parameter ν Laddha and Vempala (2021). The barrier parameter is an indicator of how well the ellipsoids of the metric defined by the barrier approximate the symmetrized polytope $\mathcal{P} \cap 2x - \mathcal{P}$ around x. While the universal and entropic barriers have $\nu = O(n)$, they are expensive to compute. The LS barrier Lee and Sidford (2014) has $\nu = \tilde{O}(n)$ while being efficient to compute. However, as we will see in more detail, as far as we know, the derivatives of the metric of the LS barrier are not “smooth” enough in most directions, which means we would have to take rather small steps while running RHMC.

We will prove that the hybrid barrier has significantly better isoperimetry (Thm. 6) than the log barrier while maintaining sufficient smoothness.

Smoothness of Hamiltonian Curves and Comparison Geometry. The starting point of our analysis is the fact that one can look at the ordinary differential equation of RHMC in Equation (1) as a second-order ODE on the manifold $\mathcal{M}$ of the open set inside the polytope with metric g. We will introduce this alternative form shortly. Looking at the Markov Kernel $\mathcal{T}_{x_0}$ of RHMC for a fixed point x_0, the randomness to define this kernel comes from the initial velocity v_0, which can be viewed as a vector on the tangent space of x_0 on the manifold $\mathcal{M}$ distributed as a standard Gaussian with respect to the local metric, namely

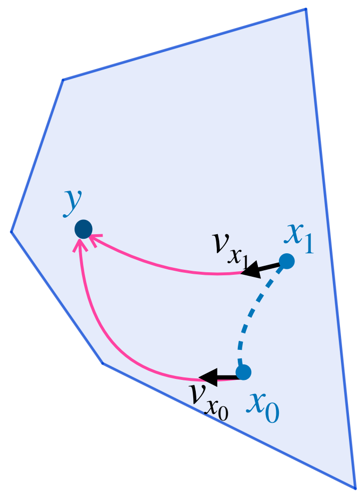

Figure 1: Family of Hamiltonian curves $\gamma_s(t)$ all ending in y with starting point varying from $x_0 = \gamma_0(0)$ to $x_1 = \gamma_{s'}(0)$, where $\gamma_s(0)$ is a geodesic in s.

$\mathcal{N}(0, g(x)^{-1})$ in the Euclidean chart. In order to show the one-step coupling (Lemma 6) for the Markov kernel of RHMC, we bound the difference between the densities $\mathcal{T}_{x_0}(y)$ and $\mathcal{T}_{x_1}(y)$ at a given point y on the manifold. These densities are the pushforwards of the Gaussian density in the tangent space of x_0 and x_1 respectively, onto the manifold through the Hamiltonian map $Ham^{\delta}(x_0, v_{x_0})$ for some fixed time δ, which maps the initial velocity v_{x_0} to the solution of the ODE $y = x(\delta)$ at time δ. The key to bound the change of density is to understand how the Hamiltonian curves vary as we change the initial point from x_0 to x_1 for a fixed destination y, given the particular geometry imposed by our hybrid barrier inside a polytope. In fact, understanding the extremal scenarios of the behavior of geometric quantities on a certain class of manifolds is the topic of Comparison Geometry Cheeger et al. (1975) Petersen (2006) Ballmann (2000). In particular, to argue that the Hamiltonian curve changes sufficiently slowly, we need the metric g of the manifold and its derivatives to be "stable". The simplest form of stability of the metric is the so-called self-concordance property, namely, g is self-concordant if the derivative of $g(x)$ in a unit direction in the tangent space is controlled by g itself. This type of self-concordance for the first derivative of the metric is already known for the Lewis weight barrier Lee and Sidford (2019). An important part of our contribution is to build an abstract framework which shows that self-concordance of the metric up to third-order derivatives is sufficient for characterizing the stability of Hamiltonian curves (see section D).

It turns out that we need to bound the rate of change of the density only for Hamiltonian curves with typical values of the initial velocity and can ignore sets with small probability when bounding the conductance. However, the typical value of $\|v\|_g$ for a Gaussian vector $v \sim \mathcal{N}(0, g^{-1})$ in the tangent space is still quite large to improve the mixing time, hence it is ideal to show self-concordance with respect to a better norm. Taking a closer look into proving self-concordance for the Lewis weight barrier, we need to control the change of the Lewis weights multiplicatively, which infinitesimally is equivalent to bounding the norm of the Jacobian of the function that maps the location $x \in \mathcal{P}$ to the logarithm of its Lewis weights. Our key observation here is that when $p < 4$, we can bound the $\left\|.\right\|_{(x,\infty)\to\infty}$ norm of this Jacobian by constant, which results in a stability argument for the Lewis weights with respect to $\|.\|_{x,\infty}$. Building upon this estimate, we show that our hybrid barrier is first-order ℓ_∞-self-concordant. This is favorable for us as the typical value of $\|v\|_{x,\infty}$ for $v \sim \mathcal{N}(0, g(x)^{-1})$ is much smaller than $\|v\|_g$. In fact, we show that the$\|.\|_{x,\infty}$ norm of the tangent vector to the RHMC curve remains small for all small enough positive times with high probability. Following this idea, to lift the ℓ_∞-self-concordance of the barrier to second and third order derivatives, we need to control the $\left\|.\right\|_{(x,\infty)\to\infty}$ norm of operators that arise from higher order derivatives of Lewis weights, and use them to estimate the derivatives of the Lewis weights barrier by analysis on the PSD cone. Our framework for obtaining these estimates is summarized in Section B.2 and elaborated upon in Section F. We use these estimates to derive self-concordance estimates for g.

Lewis weights stability. Lewis weights of A_x can be defined as the solution of the following optimization problem (for more detail, see Section A.2):

$$w_x \triangleq \mathrm{argmax}_{w\in\mathbb{R}^n_{\geq 0}} - \log\det(\mathrm{A}_x^\top \mathrm{W}^{1-2/p}\mathrm{A}_x) + (1-2/p)1^\top w.$$

In particular, in Section F we obtain the following infinity norm estimates on higher-order derivatives of the Lewis weights:

$$-\frac{1}{(4/p-1)}\|u\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}\mathbf{W}_x(z,u) \preccurlyeq \frac{1}{(4/p-1)}\|u\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x,$$
$$-\frac{1}{(4/p-1)^5}\|u\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^2\mathbf{W}_x(z,u) \preccurlyeq \frac{1}{(4/p-1)^5}\|u\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x,$$
$$-\frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^3\mathbf{W}_x(v,z,u) \preccurlyeq \frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x,$$

where we use $\preccurlyeq$ to show Löwner ordering up to universal constants. These estimates are indeed used to drive infinity norm estimates on the derivatives of the Lewis weights barrier (see Section B and F.)

Isoperimetry vs Smoothness. We show stronger stability results for the derivatives of the metric of the p-Lewis weights barrier with $p < 4$ based on the $\|.\|_{x,\infty}$ norm. However, for small p the ellipsoid of the p-Lewis weights barrier does not approximate the symmetrized polytope as well as larger p; in particular, a large subset of the ellipsoid lies outside the symmetrized polytope. This necessitates a larger barrier parameter and implies a smaller isoperimetric constant. To construct a barrier that is smooth enough along typical directions whose ellipsoids also approximate the symmetrized polytope more accurately, we go back to an idea of Vaidya from optimization and use a hybrid barrier by "regularizing" the Lewis weight barrier for $p < 4$ with the standard log barrier; we observe that penalizing the p-Lewis weights barrier with the log barrier improves its barrier parameter, while not affecting the smoothness of the barrier since the log barrier is already third-order ℓ_∞-self-concordant. Therefore, the particular choice of our barrier is essential to simultaneously guaranteeing third-order ℓ_∞-self-concordance and good isoperimetry.

Hamiltonian Curves and Variations. To see the high-level idea of how we show the one-step coupling of the Markov kernel, let the curve γ_s parameterized by $s \in [0, s']$ be a length-minimizing geodesic connecting $x_0 = \gamma_0$ to $x_1 = \gamma_{s'}$ with distance $d(x_0, x_1)$; geodesics are generalization of straight lines in the Euclidean space to arbitrary manifolds (see Section H for more background.) Suppose now that running the Hamiltonian ODE with initial location $x_0 \in \mathcal{P}$ and initial velocity v_{x_0} up to time δ takes us to a point y on the manifold. As we start moving toward x_1 on the geodesic γ_s parameterized by $s \in [0, s']$, we consider the variation of the initial Hamiltonian curve; namely a family of Hamiltonian curves parameterized by s, where the s-curve starts from point γ_s, perhaps with a different initial velocity v_{γ_s}, but ends up to the same destination y at time δ. The geodesic γ_s from x_0 to x_1 and the corresponding Hamiltonian curves are illustrated in Figure 1. Looking at the the value of the density $\mathcal{T}_{\gamma_s}(y)$ at point y after taking one step of the Markov chain starting from γ_s, we see that it depends on two major components: (1) the Gaussian density of the initial velocity v_{γ_s} which is proportional to $\exp\{-\frac{\|v_{\gamma_s}\|_g^2}{2}\}$, and (2) the determinant of the Jacobian or the differential of the map from the initial velocity v_{γ_s} to the destination point y, denoted by $J_y^{v_{\gamma_s}}$. Therefore, to study how quickly the density $\mathcal{T}_{\gamma_s}(y)$ changes from x_0 to x_1, we need to study the rate of change of the initial velocities v_{γ_s} and the Jacobians $J_y^{v_{\gamma_s}}$; the latter will depend on the rate of change of the Ricci tensor on the manifold. This necessitates studying the variations of Hamiltonian curves, which we define next.

As we mentioned earlier, one can identify the location variable x in the Hamiltonian ODE (1) as a point on the manifold $\mathcal{M}$ with metric g, and the velocity variable v as a vector in the tangent space of x, $T_x(\mathcal{M})$. Then, the Hamiltonian ODE in Equation (1) can be written as a second-order ODE on the manifold $\mathcal{M}$ using the covariant derivative, illustrated in Lemma 9. For background on Riemannian geometry and covariant differentiation, we refer the reader to Appendix H.

Lemma 9 *The Hamiltonian ODE in Equation 1 can be written using the covariant derivative of the manifold in a simplified form:*

$$\nabla_{\gamma'(t)}\gamma'(t) = \mu(\gamma(t)). \tag{7}$$

where ∇ is the covariant derivative and $\mu(x)$ is the bias (drift) vector field of the Hamiltonian curve:

$$\mu(x) \triangleq g^{-1}\mathrm{D}f(x) - \frac{1}{2}g(x)^{-1}\mathtt{tr}[g(x)^{-1}\mathrm{D}g(x)],. \tag{8}$$

In the above notation, $\mathtt{tr}[g(x)^{-1}\mathrm{D}g(x)]$ is a vector whose ith entry is $\mathtt{tr}[g(x)^{-1}\mathrm{D}_i g(x)]$. See Appendix H for a proof of Lemma 9. The above ODE (7) for Hamiltonian curves is similar to the second order ODE for geodesics; for the latter the bias vector μ is zero, i.e., the geodesic Equation is given by Do Carmo (2016)

$$\nabla_{\gamma'(t)}\gamma'(t) = 0. \tag{9}$$

In physics, the Hamiltonian ODE in Equation 7 models the motion of a particle on a manifold acting under a force field devised by μ. Next, we define the notion of a family of Hamiltonian curves and an operator $\Phi(t)$ which plays an important role in the study of variations of Hamiltonian curves.

Definition 10 (Family of Hamiltonian curves) *We say $\big(\gamma_s(t)\big)$ is a family of Hamiltonian curves ending at some fixed y whose starting point varies from $x_0 = \gamma_0(0)$ to $x_1 = \gamma_{s_1}(0)$ if for every fixed time $0 \le s \le s_1$, $\gamma_s(t)$ is a Hamiltonian curve in t, and $\gamma_s(0)$ as a function of s is a geodesic on $\mathcal{M}$ from x_0 to x_1. Unless specified otherwise, $\gamma_s(t)$ refers to a curve in t for a fixed s, and $\gamma_s'(t) = \partial_t\gamma_s(t)$ refers to its derivative.*

Definition 11 (Operators Φ and M_x) *At any point $x \in \mathcal{M}$, we define the operator M_x as*

$$\forall u \in T_x(\mathcal{M}), M_x(u) \triangleq \nabla_u\mu(x),$$

where ∇ is the covariant derivative on the manifold and μ is the Hamiltonian bias. Given the Hamiltonian curve $\gamma(t)$, we define the operator $\Phi(t) = \Phi(\gamma(t), \gamma'(t))$ on the tangent space $T_{\gamma(t)}(\mathcal{M})$ as

$$\Phi(t) \triangleq \Phi(\gamma(t)) \triangleq R(., \gamma'(t))\gamma'(t) + M_{\gamma'(t)}.$$

where R is the Riemann tensor.

Similar to Jacobi fields for geodesics (see section H.5), for a given family of Hamiltonian curves $(\gamma_s(t))$, one can write a second order ODE for the variational vector field $\tilde{J}(t) = \frac{d}{ds}\gamma_s(t)$ along the Hamiltonian curve, which depends on operator Φ (see Appendix H.4 for the proof):

Lemma 12 (ODE for Hamiltonian fields) *Given a family of Hamiltonian curves $\big(\gamma_s(t)\big)$, the vector field $\tilde{J}(t) \triangleq \partial_s\gamma_s(t)\Big|_{s=0}$ is characterized by the following second order ODE:*

$$\tilde{J}''(t) = \Phi(t)\tilde{J}(t), \tag{10}$$

where $\Phi(t)$ is defined in Definition 11. We refer to $\tilde{J}$ as a Hamiltonian field.

For variation of Hamiltonian curves, the log determinant of the Jacobian of the Hamiltonian map $J_y^{v_{\gamma_s}}$ can be characterized by a weighted integral of the trace of $\Phi(t)$ Therefore, to study the rate of change of $\det(J_y^{v_{\gamma_s}})$ as we move from x_0 to x_1, we need to study the change of $\mathtt{tr}(\Phi(t))$ along the variation of Hamiltonian curves $(\gamma_s(t))$, which in turn depends on the rate of change of the Ricci tensor and the trace of operator M_x.These ideas are formalized as the (R_1, R_2, R_3)-normality of the Hamiltonian curve in the definition below.

Definition 13 *[Normal Hamiltonian curves] We say a Hamiltonian curve $\gamma(t)$ is (R_1, R_2, R_3)-normal up to time δ if for all $0 \le t \le \delta$ we have the following:*

- *Bound on the Frobenius norm of Φ (with respect to the metric g): $\|\Phi(t)\|_F \le R_1$.*
- *For any parameterized family of curves $(\gamma_s(t))$ such that $\gamma_0(t) = \gamma(t)$ for all times $0 \le t \le \delta$, then for all such t, the derivative of $tr(\Phi(t))$ with respect to $z = \frac{d}{ds}\gamma_s(t)$ satisfies*

$$|\frac{d}{ds}tr(\Phi(t))| = |\mathrm{D}(tr(\Phi(t)))(z)| \le R_2(\|z\|_g + \delta\|\nabla_z\gamma_s'(t)\|_g).$$

- *For $\zeta(t)$ defined as the parallel transport of $\gamma'(0)$ along $\gamma(t)$: $\|\Phi(t)\zeta(t)\|_g \le R_3$.*

Parallel transport of a vector on the manifold is a generalization of shifting vectors in Euclidean space, using the covariant derivative of the manifold (see Appendix H for the rigorous definition.)

In order to show the (R_1, R_2, R_3)-normal property for the family of Hamiltonian curves, we need to define a more fundamental regularity condition for the Hamiltonian curves which states that both $\|.\|_g$ and $\|.\|_{x,\infty}$ norms remain small for the tangent vector along the Hamiltonian curve.

Definition 14 (Nice Hamiltonian curve) *We say a Hamiltonian curve $\gamma(t)$ is (c, δ)-nice if for $0 \le t \le \delta$:*

$$\|\gamma'(t)\|_g \le c\sqrt{n},$$
$$\|\gamma'(t)\|_{\gamma(t),\infty} \le c.$$

Our (c, δ)-niceness framework is a simpler and allows us to work with any third-order ℓ_∞-self-concordant barrier and avoids the technical machinery of auxiliary functions on curves used in Lee and Vempala (2018), which needs additional parameters and only works for the specialized case of log barrier.

The second major part of our contribution in this paper is that we relate this abstract notion of (R_1, R_2, R_3)-normality to the notion of third-order ℓ_∞-self-concordance. Our framework can potentially be reused on other manifolds and distributions.

Theorem 15 (From third-order ℓ_∞-self-concordance to Hamiltonian normality) *Given a Hessian manifold defined by the metric $g = \nabla^2\phi$ inside the polytope for a $(c_2, \alpha_0 n)$-third-order ℓ_∞-self-concordant barrier ϕ, define a Hamiltonian curve $\gamma(t)$ by the ODE in Equation (7) with target log density $f = \alpha\phi$. If γ is (c, δ)-nice, then it is also (R_1, R_2, R_3)-normal with parameters*

$$R_1 = c_3(c^2 + \sqrt{\alpha}_0\alpha)\sqrt{n},\ R_2 = c_3(c^2 + \sqrt{\alpha_0}\alpha)n,\ R_3 = c_3(c^2(\sqrt{n} + nc\delta) + n\delta c\alpha\sqrt{\alpha_0}),$$

where $c_3 = poly(c_2)$.

Proof The proof follows from combining Lemmas 47, 53, and 61. ∎

In order to show the closeness of one step distributions between x_0 and x_1, we need the (R_1, R_2, R_3)-normality for the family of Hamiltonian curves $(\gamma_s(t))$ for all $0 \le s \le \delta$ as we defined in 13. Therefore, we need to show that the (c, δ)-niceness property is stable for a third-order ℓ_∞-self-concordant barrier. We show this in Lemma 16, which we prove in a more technical form as Lemma 65.

Lemma 16 (Stability of norms) *In the same setting as Theorem 15, given a family of Hamiltonian curves $\gamma_s(t)$ for which $\gamma_0(t)$ is (c, δ)-nice for*

$$\delta \leq \delta' \triangleq \frac{1}{\sqrt{c^2 + \alpha\sqrt{\alpha_0} n^{1/4}}},$$

then $(\gamma_s(t))$ is a $(O(c), \delta)$-nice family of Hamiltonian curves in the interval $s \in (0, \delta)$.

Acknowledgments

We thank a bunch of people and funding agency.

References

W Ballmann. Riemannian geometry and geometric analysis by j. jost; riemannian geometry by p. petersen; riemannian geometry by t. sakai. *BULLETIN-AMERICAN MATHEMATICAL SOCIETY*, 37(4):459–466, 2000.

Jeff Cheeger, David G Ebin, and David Gregory Ebin. *Comparison theorems in Riemannian geometry*, volume 9. North-Holland Amsterdam, 1975.

Sinho Chewi, Murat A Erdogdu, Mufan Li, Ruoqi Shen, and Shunshi Zhang. Analysis of Langevin Monte Carlo from Poincaré to Log-Sobolev. In *Conference on Learning Theory (COLT)*, pages 1–2. PMLR, 2022.

Ben Cousins and Santosh Vempala. Gaussian cooling and o^*(n^3) algorithms for volume and gaussian volume. *SIAM Journal on Computing*, 47(3):1237–1273, 2018.

Arnak Dalalyan. Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent. In *Conference on Learning Theory (COLT)*, pages 678–689. PMLR, 2017.

Manfredo P Do Carmo. *Differential geometry of curves and surfaces: revised and updated second edition*. Courier Dover Publications, 2016.

Alain Durmus, Szymon Majewski, and Błażej Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *The Journal of Machine Learning Research (JMLR)*, 20(1):2666–2711, 2019.

Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991.

Mark Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

Sivakanth Gopi, Yin Tat Lee, Daogao Liu, Ruoqi Shen, and Kevin Tian. Algorithmic aspects of the log-laplace transform and a non-euclidean proximal sampler. *arXiv preprint arXiv:2302.06085*, 2023.

He Jia, Aditi Laddha, Yin Tat Lee, and Santosh Vempala. Reducing isotropy and volume to KLS: an o*(n 3 ψ 2) volume algorithm. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 961–974, 2021.

Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures & Algorithms*, 11(1):1–50, 1997.

Ravindran Kannan and Hariharan Narayanan. Random walks on polytopes and an affine interior point method for linear programming. *Mathematics of Operations Research*, 37(1):1–20, 2012.

Yunbum Kook and Santosh S. Vempala. Gaussian cooling and dikin walks: The interior-point method for logconcave sampling, 2023.

Yunbum Kook, Yin Tat Lee, Ruoqi Shen, and Santosh S Vempala. Condition-number-independent convergence rate of riemannian hamiltonian monte carlo with numerical integrators. *arXiv preprint arXiv:2210.07219*, 2022.

Aditi Laddha and Santosh Vempala. Convergence of Gibbs sampling: Coordinate Hit-and-Run mixes fast. *The 37th International Symposium on Computational Geometry (SoCG)*, 2021.

Aditi Laddha, Yin Tat Lee, and Santosh Vempala. Strong self-concordance and sampling. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1212–1222, 2020.

Yin Tat Lee and Aaron Sidford. Path finding methods for linear programming: Solving linear programs in $O(\sqrt{\text{rank}})$ iterations and faster algorithms for maximum flow. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 424–433. IEEE, 2014.

Yin Tat Lee and Aaron Sidford. Solving linear programs with sqrt (rank) linear system solves. *arXiv preprint arXiv:1910.08033*, 2019.

Yin Tat Lee and Santosh S Vempala. Geodesic walks in polytopes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on theory of Computing (STOC)*, pages 927–940, 2017.

Yin Tat Lee and Santosh S Vempala. Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1115–1121, 2018. URL https://arxiv.org/pdf/1710.06261.pdf.

Ruilin Li, Molei Tao, Santosh S Vempala, and Andre Wibisono. The mirror Langevin algorithm converges with vanishing bias. In *International Conference on Algorithmic Learning Theory (ALT)*, pages 718–742. PMLR, 2022.

László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume algorithm. *Random structures & algorithms*, 4(4):359–412, 1993.

László Lovász and Santosh Vempala. Hit-and-run from a corner. *SIAM Journal on Computing*, 35(4): 985–1005, 2006.

Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

Peter Petersen. *Riemannian geometry*, volume 171. Springer, 2006.

Gábor Székelyhidi. *An Introduction to Extremal Kahler Metrics*, volume 152. American Mathematical Soc., 2014.

Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52 (573-612):2, 2005.

Santosh Vempala and Andre Wibisono. Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. *Advances in neural information processing systems (NeurIPS)*, 32, 2019.

Xu-Jia Wang et al. Schauder estimates for elliptic and parabolic equations. 2006.

Contents

Structure of the appendices. The appendices are organized as follows:

- In Appendix A we discuss the basic tools and notation that we use throughout the paper.
- In Appendix B, we prove that the hybrid barrier is third-order ℓ_∞-self-concordant. This section relies on higher-order derivative estimates. We give an overview of these in Section B.2 and defer their proofs to Section F.
- In Appendix C, we prove one-step coupling and bound the mixing time by combining the Hamiltonian smoothness bounds for our hybrid barrier, the isoperimetry of the stationary distribution with respect to the chosen metric, and the stability of the Hamiltonian curves.
- In Appendix D, we develop our abstract framework on relating the third-order ℓ_∞-self-concordance to control the smoothness of the Hamiltonian fields on the manifold.
- In Appendix E, we prove the stability of the smoothness properties of the Hamiltonian curves as we start varying the initial location and velocity of the curve.
- In Appendix F, we prove the higher-order derivative bounds for Lewis weights and related objects that we need for Section B.
- In Appendix G, we prove an isoperimetric inequality on the Riemannian manifold $\mathcal{M}$ equipped with metric g, the Hessian of our hybrid barrier.

Appendix A. Preliminaries

A.1. Notation

We denote the target probability distribution inside the polytope by $\pi(\cdot) \sim e^{-\alpha\phi}$. Recall that for the LP polytope description $Ax \geq b$, we define the rescaling A by the slack variables, namely

$$\mathrm{A}_x = \mathtt{Diag}\big(((a_i^\top x - b_i)^{-1})_{i=1}^m\big)\mathrm{A}.$$

For a vector v in the tangent space of x, we also work with the reparameterization of v defined as

$$\begin{aligned} s_{v,x} &\triangleq \mathrm{A}_x v, \\ \mathrm{S}_{x,v} &\triangleq \mathtt{Diag}\big(s_{x,v}\big). \end{aligned} \tag{11}$$

which is the speed that v approaches the facets of the polytope normalized by the slacks. In our derivations we treat hadamard product of matrices with higher priority than matrix multiplications, namely $\mathrm{AB} \odot \mathrm{C}$ means $\mathrm{A}(\mathrm{B} \odot \mathrm{C})$. We refer to the p-Lewis weights vector of A_x by w_x and its diagonal matrix version by $\mathbf{W}_x \triangleq \mathtt{Diag}(w_x)$.

Define the log barrier by ϕ_ℓ:

$$\phi_\ell(x) \triangleq -\sum_{i=1}^{m} \log(a_i^\top x - b_i).$$

We define the metric g_2 as the Hessian of the n/m-rescaled log barrier, $g_2(x) \triangleq \frac{n}{m}\nabla^2\phi_\ell(x)$. It is easy to check that g_2-norm of v is given by the ℓ_2 norm of $s_{x,v}$:

$$\|v\|_{g_2}^2 = v^\top g_2(x) v = v^\top \mathrm{A}_x{}^\top \mathrm{A}_x v = \|s_{x,v}\|_2^2.$$

For a given point x inside polytope $\mathcal{P}$, we define the symmetrized polytope $\mathcal{P} \cap 2x - \mathcal{P}$ around x as the following: we reflect $\mathcal{P}$ around x and intersect it with the $\mathcal{P}$ namely $\mathcal{P} \cap 2x - \mathcal{P}$, as illustrated in Figure 2. The approximation of the symmetrized body by the ellipsoids corresponding to the Hessian of the barrier function plays a key role in bounding the isoperimetry constant, as we describe in Section G.

Throughout the proof, we use the notation $\lesssim$ to indicate an inequality that is true up to logarithmic factors. We use D for Euclidean derivative and ∇ and D_t for covariant differentiation with respect to the metric structure on the manifold. Moreover, we use $\preccurlyeq$ to show Löwner inequalities up to universal constants. We use $\|.\|$ with various subindices to refer to different vector norms, and $\left\|.\right\|_{\infty\to\infty}$ and $\|.\|$ to refer to the infinity to infinity operator norm and the usual operator norm of a matrix, respectively. Throughout the paper, by high probability we mean with probability $1 - 1/\mathrm{poly}(m)$.

A.2. John Ellipsoid and Lewis Weights

Proving good isoperimetry for a specific barrier can be reduced to how well the ellipsoids corresponding to the Hessian of the barrier at each point x inside the polytope approximate the symmetrized polytope around x. A natural way to approximate a symmetric polytope is via its John Ellipsoid, i.e. the ellipsoid of maximum volume contained in the polytope. Parametrizing the John ellipsoid as $\mathrm{A}_x^\top W \mathrm{A}_x$ for a positive diagonal matrix W, i.e., a weighted sum of the outer product of the rows of A_x, the weights are characterized by the following optimization problem:

$$\max_{w \in \mathbb{R}^n_{\geq 0}} \log\det(\mathrm{A}_x^\top \mathrm{W} \mathrm{A}_x) \qquad (12)$$
$$s.t.\ 1^\top w = n.$$

where $\mathrm{W} = \mathtt{Diag}\big(w\big)$ is the diagonal matrix corresponding to the vector w. The John ellipsoid approximates the symmetrized polytope in the sense that (1) it is inside the ellipsoid and (2) scaling it up by $\sqrt{n}$ will make it contains the symmetrized polytope.

On the other hand, in order to prove smoothness of the HMC curves, we need to pick a barrier whose Hessian does not change too fast as a function of x. Unfortunately the John ellipsoid is not stable. In particular, the weights W which maximize (12) are not even continuous with respect to x. An alternative is to use the p-Lewis weights to define the ellipsoid, obtained as the solution to a relaxation of the program in (12):

$$w_x \triangleq \mathrm{argmax}_{w \in \mathbb{R}^n_{\geq 0}} -\log\det(\mathrm{A}_x^\top \mathrm{W}^{1-2/p} \mathrm{A}_x) + (1-2/p) 1^\top w, \qquad (13)$$

where $\mathrm{W} = \mathtt{Diag}\big(w\big)$. Moreover, the optimal value of the program in (13) is denoted by the Lewis weight barrier at x as defined next.

Definition 17 (Lewis weight barrier) *The Lewis weight barrier can be defined as the solution of the following optimization problem:*

$$\phi_p(x) \triangleq \max_{w \in \mathbb{R}^n_{>0}} -\log\det(\mathrm{A}_x^\top \mathrm{W}^{1-2/p} \mathrm{A}_x) + (1-2/p) 1^\top w, \qquad (14)$$

Let $g_1 = \nabla^2 \phi_p$ be the metric defined by the Hessian of the Lewis weight barrier which constitute the first part of our hybrid barrier ϕ. Hence, the metric with respect to our hybrid barrier can be written as $g = \alpha_0(g_1 + g_2)$.

$$\begin{aligned} g_1 &\triangleq \nabla^2 \text{logdet}(\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p} \mathrm{A}_x), \\ g_2 &\triangleq \frac{n}{m} \mathrm{A}_x^\top \mathrm{A}_x. \\ g &\triangleq \alpha_0(g_1 + g_2). \end{aligned} \tag{15}$$

Although ϕ_p is defined as the volume of the ellipsoid when the each $a_i a_i^\top$ is reweighted by $1 - 2/p$ power of the p-Lewis weights, it is not clear if the Hessian of this barrier can be estimated explicitly by Lewis weights. It turns out that this is the case, the ellipsoid corresponding to g_1 is roughly the same as the one defined by $\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x$ (Lemma 31 in Lee and Sidford (2019)).

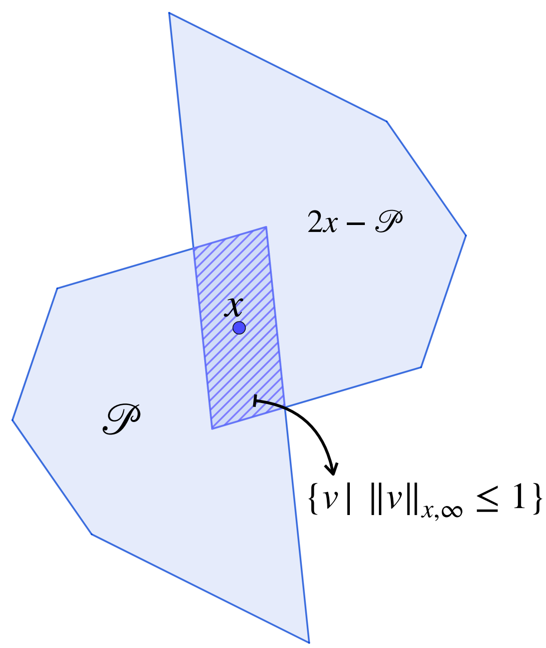

Figure 2: The unit ball of the local norm $\|.\|_{x,\infty}$ is the symmetrized polytope around $x \in \mathcal{P}$.

Lemma 18 (Restatement of Lemma 31 in Lee and Sidford (2019)) *For the Lewis weight barrier ϕ_p we can bound the local norm of its Hessian as*

$$\sum_{i=1}^m w_i(x)(s_{x,v})_i^2 \leq v^\top g_1(x) v \leq (1+p) \sum_{i=1}^m w_i(x)(s_{x,v})_i^2. \tag{16}$$

Equivalently

$$\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x \preccurlyeq g_1(x) \preccurlyeq (1+p) \mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x. \tag{17}$$

Moreover, we have the following formula for $g_1(x)$,

Lemma 19 (Equation (5.5) in Lee and Sidford (2019)) *The Lewis weight metric* $g_1(x) = \nabla^2 \log\det\left(\mathrm{A}_x^T \mathbf{W}_x^{1-2/p} \mathrm{A}_x\right)$ *can be written in the following form*

$$g_1(x) = \mathrm{A}_x^\top(\mathbf{W}_x + 2\mathbf{\Lambda}_x)\mathrm{A}_x + 2(1-2/p)\mathrm{A}_x^\top \mathbf{\Lambda}_x {\mathbf{G}_x}^{-1} \mathbf{\Lambda}_x \mathrm{A}_x, \tag{18}$$

where we define

$$\begin{aligned} \mathbf{\Lambda}_x &\triangleq \mathbf{W}_x - \mathbf{P}_x^{(2)}, \\ \mathbf{G}_x &\triangleq \mathbf{W}_x - (1-2/p)\mathbf{\Lambda}_x. \end{aligned} \tag{19}$$

In the above Lemma, $\mathbf{\Lambda}_x$, $\mathbf{G}_x$, $r_{x,v}$, and $\mathrm{R}_{x,v}$ are all functions of the location variable x. A useful fact about $\mathbf{\Lambda}_x$ and $\mathbf{G}_x$ is that they can be estimated by $\mathbf{W}_x$. It is easy to see that $\Lambda \preccurlyeq \mathbf{W}_x$ and $\frac{1}{p}\mathbf{W}_x \preccurlyeq \mathbf{G}_x \preccurlyeq \mathbf{W}_x$ (see Lemma 108 for a proof). This enables us to estimate $g_1(x)$ by the simpler form $\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x$. On the other hand, it is clear from Equation (18) that in order to estimate the first derivative of g_1 in direrction v, we need to study the derivative $\mathrm{D}\mathbf{W}_x(v)$. In Lemma 21 we illustrate the form of the Jacobian of the Lewis weights as a function of x, by taking its directional derivative in direction v based on fundamental matrices $\mathbf{\Lambda}_x$, $\mathbf{G}_x$, for any point x inside the polytope. Before that, we start by defining the projection matrix $\mathbf{P}_x$ with respect to A_x when reweighted by $\mathbf{W}_x^{1-2/p}$.

Definition 20 (Projection matrix) *we define the projection matrix* $\mathbf{P}_x$*, implicitly depending on* x*, as*

$$\mathbf{P}_x \triangleq \mathrm{P}(\mathbf{W}_x^{1/2-1/p}\mathrm{A}_x) \triangleq \mathbf{W}_x^{1/2-1/p}\mathrm{A}_x(\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p}\mathrm{A}_x)^{-1}{\mathrm{A}_x}^\top \mathbf{W}_x^{1/2-1/p},$$

where $\mathbf{W}_x$ *is the* p*-Lewis weights calculated at* x*. Moreover, we denote the Hadamard square* $\mathbf{P}_x^{\odot 2}$ *of the projection matrix by* $\mathbf{P}_x^{(2)}$*:*

$$(\mathbf{P}_x^{(2)})_{ij} \triangleq ({\mathbf{P}_\mathbf{x}}^{\odot 2})_{ij} = (\mathbf{P}_x)_{ij}^2.$$

Next, we state a formula for the derivative of the Lewis weights.

Lemma 21 (Derivative of the Lewis weights) *For arbitrary direction* $v \in \mathbb{R}^n$*, the directional derivative* $\mathrm{D}\mathbf{W}_x(v)$ *can be calculated as*

$$\mathrm{D}\mathbf{W}_x(v) = -2\mathtt{Diag}\big(\mathbf{\Lambda}_x \mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v}\big).$$

Due to the importance and repetition of the vector $\mathbf{\Lambda}_x \mathbf{G}_x^{-1}\mathbf{W}_x s_v$ in our calculations later on, we give it a separate notation

$$\begin{aligned} r_{x,v} &\triangleq \mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v}, \\ \mathrm{R}_{x,v} &\triangleq \mathtt{Diag}\big(r_{x,v}\big). \end{aligned} \tag{20}$$

Then, the derivative of $\mathbf{W}_x$ can be written as

$$\mathrm{D}\mathbf{W}_x(v) = -2\mathtt{Diag}\big(\mathbf{\Lambda}_x r_{x,v}\big).$$

Furthermore, when v is clear from the context, we denote $\mathrm{D}\mathbf{W}_x(v)$ in short by ${\mathbf{W}'}_{x,v}$.

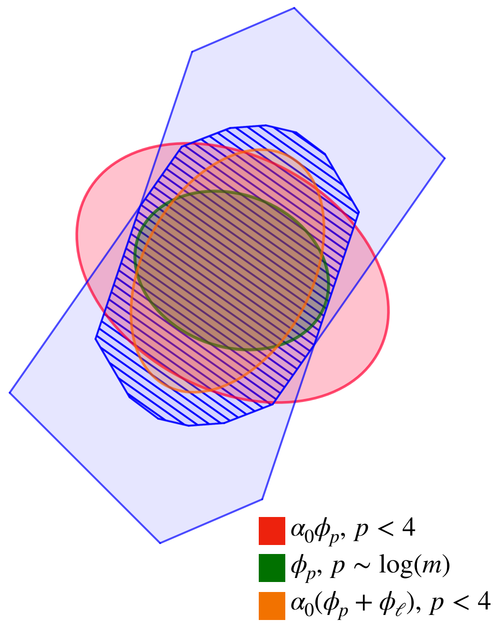

Figure 3: The Lewis weight barrier with $p \simeq \log(n)$, with $p < 4$, and our hybrid barrier ϕ which is regularized with the log barrier.

A.3. Markov Chains

For a Markov chain with state space $\mathcal{M}$, stationary distribution Q and next step distribution $p_u(\cdot)$ for any $u \in \mathcal{M}$, the conductance of the Markov chain is defined as

$$\Phi_0 \triangleq \inf_{S \subseteq \mathcal{M}} \frac{\int_S p_u(\mathcal{M} \setminus S) dQ(u)}{\min\left\{Q(S), Q(\mathcal{M} \setminus S)\right\}}.$$

The conductance of an ergodic Markov chain allows us to bound its mixing time, i.e., the rate of convergence to its stationary distribution, e.g., via the following theorem of Lovász and Simonovits. However, we will need a more refined notion of s-conductance here, to be able to ignore small subsets of small measure in bounding the conductance.

Definition 22 (s-conductance) *Consider a Markov chain with a state space $\mathcal{M}$, a transition distribution $\mathcal{T}_x$ and stationary distribution π. For any $s \in [0, 1/2)$, the s-*conductance *of the Markov chain is defined by*

$$\Phi_s \triangleq \inf_{\pi(S) \in (s, 1-s)} \frac{\int_S \mathcal{T}_x(S^c)\pi(x)dx}{\min(\pi(S) - s, \pi(S^c) - s)}.$$

A lower bound on the s-conductance of a Markov chain leads to an upper bound on its mixing rate.

Lemma 23 *Lovász and Simonovits (1993) Let π_t be the distribution of the points obtained after t steps of a lazy reversible Markov chain with the stationary distribution π. For $0 < s \leq 1/2$ and $H_s = \sup\{|\pi_0(A) - \pi(A)| : A \subset \mathcal{M},\ \pi(A) \leq s\}$, it follows that*

$$d_{TV}(\pi_t, \pi) \leq H_s + \frac{H_s}{s}\left(1 - \frac{\Phi_s^2}{2}\right)^t.$$

In order to bound the conductance or s-conductance of the Markov chain via the one-step closeness framework as in Vempala (2005), we also require an isoperimetric inequality:

Definition 24 *The isoperimetry of a metric space $\mathcal{M}$ with target distribution π is*

$$\psi_{\mathcal{M}} = \inf_{\delta \to 0} \min_{S \subseteq \mathcal{M}} \frac{\int_{\{x \mid d(S,x) \leq \delta\}} \pi(x) dx - \pi(S)}{\delta \min \{\pi(S), \pi(\mathcal{M} \setminus S)\}}$$

where d is the distance function in $\mathcal{M}$.

The following theorem (see Kook et al. (2022)) illustrates how one-step coupling with the isoperimetry leads to a lower bound on the s-conductance. Its proof is similar to that of Lemma 13 in Lee and Vempala (2018) and can be found in full detail in Appendix J.7.

Theorem 25 *For a Riemannian manifold $(\mathcal{M}, g)$, let π be the stationary distribution of a reversible Markov chain on $\mathcal{M}$ with a transition distribution $\mathcal{T}_x$. Let $\mathcal{M}' \subset \mathcal{M}$ be a subset with $\pi(\mathcal{M}') \geq 1 - \rho$ for some $\rho < \frac{1}{2}$. We assume the following one-step coupling: if $d_g(x, x') \leq \Delta \leq 1$ for $x \in \mathcal{M}'$, then $d_{TV}(\mathcal{T}_x, \mathcal{T}_{x'}) \leq 0.9$. Then for any $\rho/(\Delta \psi_{\mathcal{M}}) \leq s < \frac{1}{2}$ and given $\psi_{\mathcal{M}} \Delta \leq 1/2$, the s-conductance is bounded below by*

$$\Phi_s = \Omega(\psi_{\mathcal{M}} \Delta).$$

Appendix B. ℓ_∞-Self-Concordance of the Hybrid Barrier

In this section, we prove Theorem 7, which asserts that the hybrid barrier defined in Equation (4) is a $\left(\text{poly}(\frac{1}{4/p-1}), \alpha_0 n\right)$-third-order ℓ_∞-self-concordant barrier. We begin in Section B.1 by proving stability bounds for Lewis weights in the $\|.\|_{x,\infty}$ norm. These are sufficient to prove the first-order ℓ_∞ self-concordance, but proving second and third-order ℓ_∞ self-concordance requires bounds on higher derivatives of the Lewis weights. As these calculations are somewhat lengthy, we provide a high-level overview of them in Section B.2 and defer their full proofs to Section F. We then show in Section B.3 how to use these estimates to prove Theorem 7. As the second and third-order self-concordance proofs are similar, we derive the second-order bounds in detail in Section B.3 and prove the analogous third-order bounds in Appendix I.

B.1. Stability of Lewis Weights in the $\|.\|_{x,\infty}$ Norm

The fact that $g_1(x)$ is well-approximated by $\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x$ as stated in Lemma 18 makes it tempting to define the metric g_1 to be exactly equal to $\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x$. Note that this results in a manifold which is not Hessian anymore, i.e. its metric is not the Hessian of a convex function. Indeed, Hessian manifolds have the favorable property that the second order derivatives of the metric simplifies in the definition of the Ricci tensor as opposed to general manifolds. We use this property of Hessian manifolds in Section D.

Even though $g_1 = \nabla^2 \phi_p$ is not exactly equal to $\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x$, we still need to estimate $\mathrm{A}_x^\top \mathrm{D}\mathbf{W}_x(v) \mathrm{A}_x$ as it appears in the derivative of g_1 in direction v. To estimate $\mathrm{D}\mathbf{W}_x[v]$, While estimates of the form $\mathrm{D}\mathbf{W}_x(v) \leq O(\|v\|_{g_1}) \mathbf{W}_x$ with respect to the ellipsoidal norm $\|.\|_{g_1}$ have been derived before (Lemma 34 in Lee and Sidford (2019)), the techniques in Lee and Sidford (2019) seem insufficient to recover estimates with respect to the ℓ_∞-norm, $\|.\|_{x,\infty}$. Note that as we mentioned in Section 1.4, it is crucial in our approach to obtain estimates with respect to $\|.\|_{x,\infty}$ instead of ellipsoidal norms as for Gaussian random vectors $v \sim \mathcal{N}(0, g_1^{-1})$ in the tangent space, the typical value of $\|v\|_{x,\infty}$ is $\Theta(\sqrt{n})$ factor smaller than that of $\|v\|_{g_1}$. Unlike ellipsoidal norms, one cannot use Löwner inequalities to estimate the $\|.\|_{x,\infty}$ norm since it is not defined by a quadratic form. A key observation that we make which enables an estimate based on $\|.\|_{x,\infty}$ is that one can estimate the $\left\|.\right\|_{\infty \to \infty}$ norm of the operator $\mathbf{G}_x^{-1} \mathbf{W}_x$. We state this observation in Lemma 26 below. Note that condition $p < 4$ is vital for this norm bound.

Lemma 26 (Operator $\left\|.\right\|_{\infty\to\infty}$ norm bound on the Jacobian) *For $p < 4$, given $y = \mathbf{G}_x^{-1}\mathbf{W}_x r$ for any vector $r \in \mathbb{R}^m$, we have*

$$\|y\|_\infty \le \frac{1}{4/p-1}\|r\|_\infty.$$

Proof Set $\|r\|_\infty = \ell$. then

$$\mathbf{W}_x r = \mathbf{G}_x y = \frac{2}{p}\mathbf{W}_x y + (1-\frac{2}{p})\mathbf{P}_x^{(2)} y.$$

Now suppose $\|y\|_\infty \ge \frac{1}{4/p-1}\ell$, which implies that for the maximizing index i we have

$$|y_i| \ge \frac{1}{4/p-1}\ell.$$

But note that

$$|y^\top {\mathbf{P}_x^{(2)}}_{i,:}| \le w_i\|y\|_\infty = w_i y_i,$$

where ${\mathbf{P}_x^{(2)}}_{i,:}$ is the ith row of $\mathbf{P}_x^{(2)}$. Hence

$$y^\top \mathbf{G}_{x i,:} \ge \frac{2}{p}w_i y_i - (1-\frac{2}{p})w_i y_i = (\frac{4}{p}-1)w_i y_i > w_i\ell.$$

On the other hand

$$y^\top \mathbf{G}_{x i,:} = w_i r_i \le w_i \ell.$$

The contradiction finishes the proof. ■

Combining Lemma 26 with Lemma 21 allows us to control $\|r_{x,v}\|_{x,\infty}$, and therefore to estimate $D\mathbf{W}_x(v)$ by $\mathbf{W}_x$ and $\|v\|_{x,\infty}$:

Lemma 27 *We have*

$$-\frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathbf{W'}_{x,v} \preccurlyeq \frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x.$$

Proof Note that $\mathbf{W'}_{x,v} = -2\mathtt{Diag}\big(\mathbf{\Lambda}_x r_{x,v}\big)$ by Lemma 21. Using Lemma 26, we have $\|r_{x,v}\|_\infty \le \frac{1}{4/p-1}\|s_{x,v}\|_\infty$. Hence, for every $1 \le i \le m$:

$$|\mathbf{\Lambda}_{xi,}r_{x,v}| \le |w_i {r_{x,v}}_i| + |{\mathbf{P}_x^{(2)}}_{i,}r_{x,v}| \lesssim w_i\|r_{x,v}\|_\infty,$$

where we used the fact that the sum of the entries of the ith row of $\mathbf{P}_x^{(2)}$ is equal to w_i as it is the norm squared of the ith row of the projection matrix $\mathbf{P}_x$, which is equal to the ith diagonal entry of the $\mathbf{P}_x$, i.e. w_i. This completes the proof.

■

Next, in order to estimate the derivative of g_1, we need to estimate $\mathrm{D}\mathbf{\Lambda}_x(v)$ and $\mathrm{D}\mathbf{G}_x(v)$ based on Equation (18), for which we require the derivative of the projection matrix $\mathbf{P}_x$, since $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ are defined based on $\mathbf{W}_x$ and $\mathbf{P}_x$. We calculate the derivative of $\mathbf{P}_x$ in the following Lemma, which we prove in Appendix J.8.

Lemma 28 (Derivative of the projection matrix) *The derivative of the projection matrix* $\mathbf{P}_x = \mathbf{P}(\mathbf{W}_x^{1/2-1/p}\mathrm{A}_x)$ *in direction* v *is given by*

$$\mathrm{D}\mathbf{P}_x(v) = -\mathbf{P}_x\mathrm{R}_{x,v} - \mathrm{R}_{x,v}\mathbf{P}_x + 2\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x, \tag{21}$$

where $\mathrm{R}_{x,v}$ *is defined in Equation (20).*

Luckily, we see that the derivative of the projection matrix can be written by itself and the variable $\mathrm{R}_{x,v}$ that we defined in Equation (20). This observation completes the circle in the calculus between variables $\mathbf{W}_x, \mathbf{P}_x, \mathbf{G}_x, \mathbf{\Lambda}_x, \mathrm{R}_{x,v}$, and $\mathrm{S}_{x,v}$, i.e. the derivative of each one can be written based on others (Note that $\mathrm{D}(\mathrm{S}_{x,v})(z) = -\mathrm{S}_{x,v}\mathrm{S}_{x,z}$). Moreover, Lemma 28 enables us to use our infinity norm bound in Lemma 27 to obtain estimates on $\mathrm{D}\mathbf{G}_x(v)$ and $\mathrm{D}\mathbf{\Lambda}_x(v)$. The following Lemma is proved in Section F.

Lemma 29 *For the derivatives of* $\mathbf{G}_x$ *and* $\mathbf{\Lambda}_x$ *at some point* x *we have*

$$-\frac{1}{4/p-1}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}\mathbf{G}_x(z) \preccurlyeq \frac{1}{4/p-1}\|z\|_{x,\infty}\mathbf{W}_x,$$
$$-\frac{1}{4/p-1}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}\mathbf{\Lambda}_x(z) \preccurlyeq \frac{1}{4/p-1}\|z\|_{x,\infty}\mathbf{W}_x.$$

When v is clear from the context, we refer to $\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x$ by $\tilde{\mathbf{P}}_{x,v}$ for brevity.

B.2. Higher-Order Lewis Weight Estimates

The estimates that we built up so far are enough to show first-order ℓ_∞-self-concordance of g_1, as we show in Lemma 34. In this section, we give an overview of how we bound the higher-order derivatives, which we prove in detail in Section F.

In order to go to higher derivatives, as can be observed in Lemmas 21 and 28, we need to estimate the derivative of $r_{x,v}$ in the second direction z. In particular, note that $\mathrm{D}^2\mathbf{P}_x^{(2)}(v) = 2\mathrm{D}\mathbf{P}_x(v)\odot\mathrm{D}\mathbf{P}_x(z) + 2\mathbf{P}_x\odot \mathrm{D}^2\mathbf{P}_x(v,z)$, where the second subterm subsumes the derivative of $\mathrm{R}_{x,v}$ in direction z (recall Equation (21).) Luckily we can estimate the infinity norm of this derivative by the infinity norm of v and z.

Lemma 30 *The derivative of* $r_{x,v}$ *in direction* z *can be estimated as*

$$\|\mathrm{D}(r_{x,z})(v)\|_\infty \lesssim \frac{1}{(4/p-1)^4}\|v\|_{x,\infty}\|z\|_{x,\infty}.$$

Now using this estimate, we can derive estimates for the second derivatives of $\mathbf{W}_x$, $\mathbf{P}_x^{(2)}$, $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ by analysis on the PSD cone. We state these in the following lemma, which combines the results of Lemmas 79 and 83.

Lemma 31

$$-\frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^2\mathbf{W}_x(v,z) \preccurlyeq \frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x,$$
$$-\frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^2\mathbf{G}_x(v,z) \preccurlyeq \frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x,$$
$$-\frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^2\mathbf{\Lambda}_x(v,z) \preccurlyeq \frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x.$$

In order to go to one higher derivative, we need to control the second derivative of $r_{x,v}$. For this purpose, we need to derive the following key operator-$\left\|.\right\|_{\infty\to\infty}$ norm estimates, which we prove in Lemmas 84 and 85:

Lemma 32 *For diagonal matrices* $\mathrm{S}_1, \mathrm{S}_2$ *we have*

$$\left\|\mathbf{W}_x^{-1}\big((\mathbf{P}_x \mathrm{S}_1 \mathbf{P}_x) \odot (\mathbf{P}_x \mathrm{S}_2 \mathbf{P}_x)\big)\right\|_{\infty\to\infty} \le \left\|\mathrm{S}_1\right\|\left\|\mathrm{S}_2\right\|,$$
$$\left\|\mathbf{W}_x^{-1}(\mathbf{P}_x \odot (\mathbf{P}_x \mathrm{S}_1 \mathbf{P}_x \mathrm{S}_2 \mathbf{P}_x))\right\|_{\infty\to\infty} \le \left\|\mathrm{S}_1\right\|\left\|\mathrm{S}_2\right\|.$$

Building upon Lemma 32 and the previous estimates, we can then derive estimates on the second order derivative of $r_{x,v}$, and then third order derivatives of $\mathbf{W}_x, \mathbf{G}_x, \mathbf{\Lambda}_x$. The following estimates are proved in Lemmas 88, 89, and 91 in Section F.

Lemma 33

$$\|\mathrm{D}^2(r_{x,v})(u,z)\|_\infty \lesssim \frac{1}{(4/p-1)^6}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty},$$

and

$$-\frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^3\mathbf{W}_x(v,z,u) \preccurlyeq \frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x.$$
$$-\frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^3\mathbf{G}_x(u,v,z) \preccurlyeq \frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x,$$
$$-\frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^3\mathbf{\Lambda}_x(u,v,z) \preccurlyeq \frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x.$$

B.3. Proof of ℓ_∞-Self-Concordance of the Hybrid Barrier

In this section, we show how to use the estimates above to prove Theorem 7. To show how the proof goes, in this section we show the first and second-order ℓ_∞-self-concordance of ϕ and refer the reader to Appendix I for the proof of third-order ℓ_∞-self-concordance.

We start by the first-order ℓ_∞-self-concordance.

Lemma 34 (First-order ℓ_∞-self-concordance) *For $p < 4$ for arbitrary direction $v \in \mathbb{R}^n$,*

$$-\frac{1}{4/p-1}\|v\|_{x,\infty} g_1 \preccurlyeq \mathrm{D}g_1(v) \preccurlyeq \frac{1}{4/p-1}\|v\|_{x,\infty} g_1.$$

Proof Taking derivative from $g_1(x)$ as expanded in Lemma 19 and the fact that $\mathrm{D}(\mathrm{A}_x)(v) = \mathrm{A}_x^\top \mathrm{S}_{x,v}$:

$$\begin{aligned}\mathrm{D}g_1(x) &= \mathrm{A}_x^\top \mathrm{S}_{x,v}(\mathbf{W}_x + 2\mathbf{\Lambda}_x)\mathrm{A}_x + \mathrm{A}_x^\top(\mathbf{W}_x + 2\mathbf{\Lambda}_x)\mathrm{S}_{x,v}\mathrm{A}_x + \mathrm{A}_x^\top(\mathrm{D}\mathbf{W}_x[v] + 2\mathrm{D}\mathbf{\Lambda}_x[v])\mathrm{A}_x \\ &\quad + 2(1-2/p)\Big[\mathrm{A}_x^\top(\mathrm{D}\mathbf{\Lambda}_x[v]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x + \mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x[v]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x + \mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathrm{D}\mathbf{\Lambda}_x[v])\mathrm{A}_x\Big] \qquad (22)\end{aligned}$$

Now we bound each of the terms separately. For arbitrary vector $q \in \mathbb{R}^n$ note that $\mathrm{A}_x q = s_{x,q}$. For the first term we can write

$$\begin{aligned}q^\top \mathrm{A}_x^\top \mathrm{S}_{x,v}(\mathbf{W}_x + 2\mathbf{\Lambda}_x)\mathrm{A}_x q &= s_{x,q}^\top \mathrm{S}_{x,v}(\mathbf{W}_x + 2\mathbf{\Lambda}_x)s_{x,q} \\ &\le \sqrt{(s_{x,q}^\top \mathrm{S}_{x,v}\mathbf{W}_x \mathrm{S}_{x,v} s_{x,q})s_{x,q}^\top \mathbf{W}_x s_{x,q}} \\ &\quad + 2\sqrt{(s_{x,q}^\top \mathrm{S}_{x,v}\mathbf{\Lambda}_x \mathrm{S}_{x,v} s_{x,q})s_{x,q}^\top \mathbf{\Lambda}_x} s_{x,q} \\ &\le \sqrt{(s_{x,q}^\top \mathrm{S}_{x,v}\mathbf{W}_x \mathrm{S}_{x,v} s_{x,q})s_{x,q}^\top \mathbf{W}_x s_{x,q}} \\ &\quad + 2\sqrt{(s_{x,q}^\top \mathrm{S}_{x,v}\mathbf{W}_x \mathrm{S}_{x,v} s_{x,q})s_{x,q}^\top \mathbf{W}_x} s_{x,q} \\ &\lesssim \|s_{x,v}\|_\infty q^\top g_1(x) q, \qquad (23)\end{aligned}$$

where in the last line we used the estimate 17 and the fact that $\mathbf{W}_x$ is diagonal. Applying the same bound as Equation (23) for the second term $\mathrm{A}_x^\top(\mathbf{W}_x + 2\mathbf{\Lambda}_x)\mathrm{S}_{x,v}\mathrm{A}_x$ in Equation (22), we conclude

$$\mathrm{A}_x^\top \mathrm{S}_{x,v}(\mathbf{W}_x + 2\mathbf{\Lambda}_x)\mathrm{A}_x + \mathrm{A}_x^\top(\mathbf{W}_x + 2\mathbf{\Lambda}_x)\mathrm{S}_{x,v}\mathrm{A}_x \preccurlyeq (1+p)\|v\|_{x,\infty} g_1(x). \tag{24}$$

For the third term in Equation (23) applying the estimate in Lemmas (27) and (29) implies

$$\mathrm{A}_x^\top(\mathrm{D}\mathbf{W}_x[v] + 2\mathrm{D}\mathbf{\Lambda}_x[v])\mathrm{A}_x \preccurlyeq \frac{1}{4/p-1}\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x \preccurlyeq \frac{(1+p)\|v\|_{x,\infty}}{4/p-1} g_1(x). \tag{25}$$

For the forth term in Equation (23) using Cauchy-Schwarz inequality and then the estimate for $\mathbf{\Lambda}_x$ in Lemma 29:

$$\begin{aligned} &q^\top {\mathrm{A}_x}^\top \mathrm{D}\mathbf{\Lambda}_x[v]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x \mathrm{A}_x q \\ &\leq \sqrt{s_{x,q}^\top \mathbf{G}_x^{1/2}\left(\mathbf{G}_x^{-1/2}\mathrm{D}\mathbf{\Lambda}_x[v]\mathbf{G}_x^{-1/2}\right)^2 \mathbf{G}_x^{1/2} s_{x,q}} \sqrt{s_{x,q}^\top \mathbf{G}_x^{1/2}\left(\mathbf{G}_x^{-1/2}\mathbf{\Lambda}_x\mathbf{G}_x^{-1/2}\right)^2 \mathbf{G}_x^{1/2} s_{x,q}}. \end{aligned} \tag{26}$$

Now using the fact that $\mathrm{D}\mathbf{\Lambda}_x[v] \preccurlyeq \frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x$ and $\frac{1}{p}\mathbf{W}_x \preccurlyeq \mathbf{G}_x$, we get

$$\mathbf{G}_x^{-1/2}\mathrm{D}\mathbf{\Lambda}_x[v]\mathbf{G}_x^{-1/2} \preccurlyeq \frac{p}{4/p-1}\mathbf{W}_x^{-1/2}\mathbf{W}_x\mathbf{W}_x^{-1/2} \preccurlyeq \frac{p}{4/p-1}\mathrm{I},$$

which implies

$$\left(\mathbf{G}_x^{-1/2}\mathrm{D}\mathbf{\Lambda}_x[v]\mathbf{G}_x^{-1/2}\right)^2 \preccurlyeq \left(\frac{p}{4/p-1}\right)^2 \mathrm{I}. \tag{27}$$

Similarly, using the fact that $\mathbf{\Lambda}_x \preccurlyeq \mathbf{W}_x$,

$$\left(\mathbf{G}_x^{-1/2}\mathbf{\Lambda}_x\mathbf{G}_x^{-1/2}\right)^2 \preccurlyeq p^2 \mathrm{I}. \tag{28}$$

Plugging Equations (27) and (28) into Equation (26), we get

$$q^\top {\mathrm{A}_x}^\top \mathrm{D}\mathbf{\Lambda}_x[v]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x\mathrm{A}_x q \leq \frac{p^2}{4/p-1} s_{x,q}^\top \mathbf{G}_x s_{x,q} \leq \frac{p^2}{4/p-1} q^\top \mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x q.$$

Repeating this bound for the other symmetric quadratic form, ${\mathrm{A}_x}^\top \mathbf{\Lambda}_x \mathbf{G}_x^{-1}\mathrm{D}\mathbf{\Lambda}_x[v]\mathrm{A}_x$, implies

$${\mathrm{A}_x}^\top \mathrm{D}\mathbf{\Lambda}_x[v]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x\mathrm{A}_x + {\mathrm{A}_x}^\top \mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathrm{D}\mathbf{\Lambda}_x[v]\mathrm{A}_x \preccurlyeq \frac{p^2}{4/p-1}\mathrm{A}_x^\top \mathbf{W}_x\mathrm{A}_x \preccurlyeq \frac{(1+p)p^2\|v\|_{x,\infty}}{4/p-1} g_1(x). \tag{29}$$

Finally for the remaining quadratic form in Equation (22) we have

$$\begin{aligned} \mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x[v]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x &\preccurlyeq \frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathbf{G}_x\mathbf{G}_x^{-1}\mathbf{\Lambda}_x \\ &\preccurlyeq \frac{p}{4/p-1}\|v\|_{x,\infty}\Lambda W^{-1}\Lambda \\ &= \frac{p}{4/p-1}\|v\|_{x,\infty}\mathbf{\Lambda}_x\mathbf{\Lambda}_x^{-1}\mathbf{\Lambda}_x \preccurlyeq \frac{p}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x, \end{aligned}$$

which implies

$$\mathrm{A}_x^\top \mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x[v]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x\mathrm{A}_x \preccurlyeq \frac{(1+p)p\|v\|_{x,\infty}}{4/p-1}\mathrm{A}_x^\top \mathbf{W}_x\mathrm{A}_x. \tag{30}$$

Plugging Equations (24), (25), (29), and (30) into Equation (22) and noting the fact that $p < 4$ is constant proves the right-hand side in Equation (31). The left-hand side follows similarly. ■

Next, we move on to show the second-order ℓ_∞-self-concordance.

Lemma 35 (Second-order ℓ_∞-self-concordance) *For $p < 4$ for arbitrary direction $v \in \mathbb{R}^n$,*

$$-\frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}g_1(x) \preccurlyeq \mathrm{D}g_1(x)[v,z] \preccurlyeq \frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}g_1(x). \tag{31}$$

Proof Again we consider the formula for g_1 in Lemma 19. The first observation is that if the derivative with respect to v hits the A_x matrix (on either left or right), then we can upper bound this part of the quadratic form $q^\top \mathrm{D}g_1(v,z)q$ by $\frac{1}{4/p-1}\|\mathrm{S}_{x,v}s_{x,q}\|_{w(x)}\|s_{x,q}\|_{w(x)}$ by just reusing our estimates in the proof of Lemma 34 for showing first-order ℓ_∞-self-concordance. But $\|\mathrm{S}_{x,v}s_{x,q}\|_{w(x)}$ and $\|s_{x,q}\|_{w(x)}$ can further be upper bounded by $\|v\|_{x,\infty}\|q\|_{g_1(x)}$ and $\|q\|_{g_1(x)}$, respectively, which results in the upper and lower bounds in Equation (31). The same argument holds in the derivative with respect to z hits A_x. Hence, we only have to consider the case when both of the directional derivatives with respect to v and z do not hit any A_x. For that part of the derivative $\mathrm{D}^2g_1(v,z)$ consists of the following terms:

$$\begin{aligned}
\mathrm{D}^2g_1(x)[v,z] \to\ & \mathrm{A}_x^\top(\mathrm{D}\mathbf{W}_x[v,z] + 2\mathrm{D}\boldsymbol{\Lambda}_x[v,z])\mathrm{A}_x \\
&+ 2(1-2/p)\mathrm{A}_x^\top(\mathrm{D}\boldsymbol{\Lambda}_x[v,z]\mathbf{G}_x^{-1}\boldsymbol{\Lambda}_x + \boldsymbol{\Lambda}_x\mathbf{G}_x^{-1}\mathrm{D}\boldsymbol{\Lambda}_x[v,z])\mathrm{A}_x \\
&+ 2(1-2/p)\mathrm{A}_x^\top(\mathrm{D}\boldsymbol{\Lambda}_x[v]\mathbf{G}_x^{-1}\mathrm{D}\boldsymbol{\Lambda}_x[z] + \mathrm{D}\boldsymbol{\Lambda}_x[v]\mathbf{G}_x^{-1}\mathrm{D}\boldsymbol{\Lambda}_x[z])\mathrm{A}_x \\
&+ 2(1-2/p)\mathrm{A}_x^\top(\mathrm{D}\boldsymbol{\Lambda}_x[v]\mathrm{D}(\mathbf{G}_x^{-1})[z]\boldsymbol{\Lambda}_x + \mathrm{D}\boldsymbol{\Lambda}_x[z]\mathrm{D}(\mathbf{G}_x^{-1})[v]\boldsymbol{\Lambda}_x)\mathrm{A}_x \\
&+ 2(1-2/p)\mathrm{A}_x^\top(\boldsymbol{\Lambda}_x\mathrm{D}(\mathbf{G}_x^{-1})[z]\mathrm{D}\boldsymbol{\Lambda}_x[v] + \boldsymbol{\Lambda}_x\mathrm{D}(\mathbf{G}_x^{-1})[v]\mathrm{D}\boldsymbol{\Lambda}_x[z])\mathrm{A}_x \\
&+ 2(1-2/p)\mathrm{A}_x^\top\boldsymbol{\Lambda}_x\mathrm{D}^2(\mathbf{G}_x^{-1})[v,z]\boldsymbol{\Lambda}_x\mathrm{A}_x.
\end{aligned}$$

For the first line above, using Lemma 31,

$$\mathrm{A}_x^\top(\mathrm{D}\mathbf{W}_x[v,z] + 2\mathrm{D}\boldsymbol{\Lambda}_x[v,z])\mathrm{A}_x \preccurlyeq \frac{1}{(4/p-1)^2}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathrm{A}_x^\top(\mathbf{W}_x + 2\boldsymbol{\Lambda}_x)\mathrm{A}_x.$$

The second line of Equation (36) follows similar to the third term in Equation (22) except that instead of Lemma 29 we use Lemma 31. Regarding the third line, using the bound in Equation (27),

$$\begin{aligned}
q^\top\mathrm{A}_x^\top\mathrm{D}\boldsymbol{\Lambda}_x[z]\mathbf{G}_x^{-1}\mathrm{D}\boldsymbol{\Lambda}_x[v]\mathrm{A}_xq &\leq \sqrt{q^\top\mathrm{A}_x^\top\mathrm{D}\boldsymbol{\Lambda}_x[z]G^{-1}\mathrm{D}\boldsymbol{\Lambda}_x[z]\mathrm{A}_xq}\sqrt{q^\top\mathrm{A}_x^\top\mathrm{D}\boldsymbol{\Lambda}_x[z]G^{-1}\mathrm{D}\boldsymbol{\Lambda}_x[z]\mathrm{A}_xq} \\
&\leq \sqrt{q^\top\mathrm{A}_x^\top G^{1/2}\Big(G^{-1/2}\mathrm{D}\boldsymbol{\Lambda}_x[z]\mathbf{G}_x^{-1/2}\Big)^2G^{1/2}\mathrm{A}_xq} \\
&\times \sqrt{q^\top\mathrm{A}_x^\top G^{1/2}\Big(G^{-1/2}\mathrm{D}\boldsymbol{\Lambda}_x[v]\mathbf{G}_x^{-1/2}\Big)^2G^{1/2}\mathrm{A}_xq} \\
&\leq (\frac{p}{4/p-1})^2\|q\|_{g_1}^2,
\end{aligned}$$

which implies

$$\mathrm{A}_x^\top\mathrm{D}\boldsymbol{\Lambda}_x[z]\mathbf{G}_x^{-1}\mathrm{D}\boldsymbol{\Lambda}_x[v]\mathrm{A}_x + \mathrm{A}_x^\top\mathrm{D}\boldsymbol{\Lambda}_x[v]\mathbf{G}_x^{-1}\mathrm{D}\boldsymbol{\Lambda}_x[z]\mathrm{A}_x \preccurlyeq (\frac{p}{4/p-1})^2g_1.$$

Before going to the fourth line, we show that the derivatives of $\mathbf{G}_x^{-1}$ are controlled by $\mathbf{G}_x^{-1}$ the same way the derivatives of $\mathbf{G}_x$ are controlled by $\mathbf{G}_x$.

Lemma 36 (Control on $\mathbf{G}_x^{-1}$) *For arbitrary v, z*

$$-\frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{G}_x^{-1} \preccurlyeq \mathrm{D}\mathbf{G}_x^{-1}[v] \preccurlyeq \frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{G}_x^{-1},$$

$$-\frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{G}_x^{-1} \preccurlyeq \mathrm{D}^2\mathbf{G}_x^{-1}[v,z] \preccurlyeq \frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{G}_x^{-1}.$$

Using Lemma 36, we can bound the forth and fifth lines of Equation (36) as

$$\begin{aligned} q^\top \mathrm{A}_x \mathrm{D}\mathbf{\Lambda}_x[v]\mathrm{D}(\mathbf{G}_x^{-1})[z]\mathbf{\Lambda}_x\mathrm{A}_x q &\le \sqrt{q^\top \mathrm{A}_x\mathbf{\Lambda}_x^{1/2}\left(\mathbf{\Lambda}_x^{-1/2}\mathrm{D}\mathbf{\Lambda}_x[v]\mathbf{\Lambda}_x^{-1/2}\right)^2\mathbf{\Lambda}_x^{1/2}\mathrm{A}_x q} \\ &\times\sqrt{q^\top \mathrm{A}_x^\top\mathbf{\Lambda}_x^{1/2}\left(\mathbf{\Lambda}_x^{1/2}\mathrm{D}\mathbf{G}_x^{-1}[z]\mathbf{\Lambda}_x^{1/2}\right)^2\mathbf{\Lambda}_x^{1/2}\mathrm{A}_x q} \\ &\le (1+p)(\frac{p}{4/p-1})^2 q^\top g_1(x) q, \end{aligned}$$

which implies

$$\mathrm{A}_x^\top(\mathrm{D}\mathbf{\Lambda}_x[v]\mathrm{D}(\mathbf{G}_x^{-1})[z]\mathbf{\Lambda}_x+\mathrm{D}\mathbf{\Lambda}_x[z]\mathrm{D}(\mathbf{G}_x^{-1})[v]\mathbf{\Lambda}_x)\mathrm{A}_x \preccurlyeq (\frac{p}{4/p-1})^2 g_1(x).$$

For the last line in Equation (36) we again use Lemma 36 to obtain

$$\begin{aligned} \mathrm{A}_x^\top\mathbf{\Lambda}_x\mathrm{D}^2(\mathbf{G}_x^{-1})[v,z]\mathbf{\Lambda}_x\mathrm{A}_x &\preccurlyeq \frac{1}{(4/p-1)^2}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathrm{A}_x^\top\mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathbf{\Lambda}_x\mathrm{A}_x \\ &\preccurlyeq \frac{p}{(4/p-1)^2}\mathrm{A}_x^\top\mathbf{\Lambda}_x\mathbf{\Lambda}_x^{-1}\mathbf{\Lambda}_x\mathrm{A}_x \\ &\preccurlyeq (1+p)\frac{p}{(4/p-1)^2}g_1. \end{aligned}$$

This completes the proof for the right-hand side of Equation (31). The left-hand side follows similarly. ■

Finally, a straightforward calculation shows that the log barrier is also third-order ℓ_∞-self-concordant.

Lemma 37 (Third-order ℓ_∞-self-concordance of the log barrier) *The log barrier $\phi_\ell(x) = -\sum_{i=1}^m \log(a_i^T x - b_i)$ is third-order ℓ_∞-self-concordant.*

Proof [Proof of Lemma 37] To establish the third-order ℓ_∞-self-concordance of ϕ_ℓ, we need to show that the corresponding metrix $g_2 = \nabla^2\phi_\ell(x) = \mathrm{A}_x^\top\mathrm{A}_x$ obeys the following inequalities:

$$\begin{aligned} -\|v\|_{x,\infty}g_2 \preccurlyeq \mathrm{D}g_2(v) &\preccurlyeq \|v\|_{x,\infty}g_2, \\ -\|v\|_{x,\infty}\|z\|_{x,\infty}g_2 \preccurlyeq \mathrm{D}g_2(v,z) &\preccurlyeq \|v\|_{x,\infty}\|z\|_{x,\infty}g_2, \\ -\|v\|_{x,\infty}\|z\|_{x,\infty}\|u\|_{x,\infty}g_2 \preccurlyeq \mathrm{D}^3g_2(v,z,u) &\preccurlyeq \|v\|_{x,\infty}\|z\|_{x,\infty}\|u\|_{x,\infty}g_2. \end{aligned}$$

Its directional derivative is given by

$$\mathrm{D}g_2(v) = -2\mathrm{A}_x^\top\mathrm{S}_{x,v}\mathrm{A}_x,$$

which can be bounded as

$$-\|s_{x,v}\|_\infty\mathrm{A}_x^\top\mathrm{A}_x \preccurlyeq \mathrm{A}_x^\top\mathrm{S}_{x,v}\mathrm{A}_x \preccurlyeq \|s_{x,v}\|_\infty\mathrm{A}_x^\top\mathrm{A}_x.$$

Similarly, the second and third directional derivatives of g_2 are given by

$$\begin{aligned} \mathrm{D}^2g_2(v,z) &= 6\mathrm{A}_x^\top\mathrm{S}_{x,v}\mathrm{S}_{x,z}\mathrm{A}_x, \\ \mathrm{D}^2g_2(v,z,u) &= -24\mathrm{A}_x^\top\mathrm{S}_{x,v}\mathrm{S}_{x,z}\mathrm{S}_{x,u}\mathrm{A}_x, \end{aligned}$$

which can be bounded as

$$\begin{aligned} -\|s_{x,v}\|_\infty\|s_{x,z}\|_\infty\mathrm{A}_x^\top\mathrm{A}_x \preccurlyeq \mathrm{A}_x^\top\mathrm{S}_{x,v}\mathrm{S}_{x,z}\mathrm{A}_x &\preccurlyeq \|s_{x,v}\|_\infty\|s_{x,z}\|_\infty\mathrm{A}_x^\top\mathrm{A}_x, \\ -\|s_{x,v}\|_\infty\|s_{x,z}\|_\infty\|s_{x,u}\|_\infty\mathrm{A}_x^\top\mathrm{A}_x \preccurlyeq \mathrm{A}_x^\top\mathrm{S}_{x,v}\mathrm{S}_{x,z}\mathrm{S}_{x,u}\mathrm{A}_x &\preccurlyeq \|s_{x,v}\|_\infty\|s_{x,z}\|_\infty\|s_{x,u}\|_\infty\mathrm{A}_x^\top\mathrm{A}_x. \end{aligned}$$

This completes the proof as for arbitrary vector $v \in \mathbb{R}^n$, $\|v\|_{x,\infty} = \|s_{x,v}\|_\infty$. ■

The following lemma bounds the self-concordance parameter of the hybrid barrier.

Lemma 38 (Self-concordance parameter of ϕ) *For our hybrid barrier ϕ, the self-concordance parameter is defined as*

$$\nu = \sup_{x \in \mathcal{P}} \mathrm{D}\phi(x)^\top (\mathrm{D}^2\phi(x))^{-1} \mathrm{D}\phi(x),$$

as is bounded by $\alpha_0 n$.

Proof Note that for the Lewis weights and log barrier parts of the barrier $\phi = \alpha_0\phi_p + \alpha_0\frac{n}{m}\phi_\ell$ we can bound the self-concordance parameter separately as (different from the barrier parameter which is defined in Section G])

$$\begin{aligned}\sqrt{\mathrm{D}\phi(x)^\top(\mathrm{D}^2\phi(x))^{-1}\mathrm{D}\phi(x)} &\leq \alpha_0\sqrt{\mathrm{D}\phi_p(x)^\top(\mathrm{D}^2\phi(x))^{-1}\mathrm{D}\phi_p(x)} + \alpha_0\frac{n}{m}\sqrt{\mathrm{D}\phi_\ell(x)^\top(\mathrm{D}^2\phi(x))^{-1}\mathrm{D}\phi_\ell(x)}\\ &\leq \sqrt{\alpha_0}\sqrt{\mathrm{D}\phi_p(x)^\top(\mathrm{D}^2\phi_p(x))^{-1}\mathrm{D}\phi_p(x)} + \sqrt{\alpha_0\frac{n}{m}}\sqrt{\mathrm{D}\phi_\ell(x)^\top(\mathrm{D}^2\phi_\ell(x))^{-1}\mathrm{D}\phi_\ell(x)}.\end{aligned}$$

Now for the log barrier part, we have

$$\mathrm{D}\phi_\ell(x)^\top(\mathrm{D}^2\phi_\ell(x))^{-1}\mathrm{D}\phi_\ell(x) = 1^\top \mathrm{A}_x(\mathrm{A}_x{\top}\mathrm{A}_x)^{-1}\mathrm{A}_x^\top 1 \leq m, \tag{32}$$

and for the Lewis weight barrier part, from Lemmas 117 and 18:

$$\mathrm{D}\phi_p(x)^\top(\mathrm{D}^2\phi_p(x))^{-1}\mathrm{D}\phi_p(x) \leq w_x^\top \mathrm{A}_x(\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x)^{-1}\mathrm{A}_x^\top w_x \leq \|w_x\|_2^2 = n. \tag{33}$$

Combining Equations (32) and (33) completes the proof. ■

Our main technical theorems about the hybrid barrier follow from the lemmas above.
Proof [Proof of Theorem 7] The third-order ℓ_∞-self-concordance follows from Lemmas 34, 35, 102, and 37. The $\alpha_0 n$ self-concordance parameter follows from Lemma 38. ■

Appendix C. Bounding Conductance and Mixing Time

The goal of this section is to illustrate how we combine different pieces together to prove Theorem 4. To this end, we first prove a general purpose mixing time on a manifold in Theorem 40.

We start by defining the concept of "Nice sets," which links the initial velocity v_{x_0} to the (R_1, R_2, R_3) normality.

Definition 39 (Nice set) *Given $x_0 \in \mathcal{M}$, we say a set $Q_{x_0} \subseteq T_{x_0}(\mathcal{M})$ is $(R_1, R_2, R_3, \delta, c)$-nice if for $v_{x_0} \sim \mathcal{N}(0, g(x_0)^{-1})$, we have*

1. $\mathbb{P}(v_{x_0} \notin Q_{x_0}) \leq 0.001$.
2. *For every x_1 with $d(x_1, x_0) \leq \delta$, the Hamiltonian family of curves between x_0 and x_1 ending at $Ham^\delta(x_0, v_{x_0})$, denoted by $(\gamma_s(t))$, is (R_1, R_2, R_3)-normal. Furthermore, $\|\gamma_s'(t)\| \leq 2c\sqrt{n}$ and $\|\nabla_{\frac{d}{ds}\gamma_s(t)}\gamma_s'(t)\| \leq \frac{10}{\delta}$ for all $0 \leq s \leq d(x_1, x_0)$ and $0 \leq t \leq \delta$.*

We can now state this section's main theorem, which we prove at the end of the section.

Theorem 40 *Suppose we want to sample from some distribution π on the manifold $\mathcal{M}$, starting from distribution π_0 with $M = \sup_{x\in\mathcal{M}} \frac{d\pi_0(x)}{d\pi(x)}$. Suppose there exists a set $S \subseteq \mathcal{M}$ with $\pi(S) \geq 1 - O((\epsilon\delta\psi)/M)$, such that for every $x_0 \in S$ there exists an $(R_1, R_2, R_3, \delta, c)$-nice set $Q_{x_0} \subseteq T_{x_0}(\mathcal{M})$. Moreover, let ψ be the isoperimetric constant of the pair $(\mathcal{M}, g)$. Then, for any δ satisfying $\delta^2 R_1 \leq 1$, $\delta^2 R_3 \leq 1$, $\delta^3 R_2 \leq 1$, the mixing time to reach a distribution within TV distance ϵ of π is bounded by*

$$O(\log(M + 1/\epsilon)(\psi\delta)^{-2}(\log(c) + \log(m))^2).$$

The technical core of this section is Lemma 46, which establishes the closeness of the one-step distributions of the Markov chain. This will allow us to prove Theorem 40. We will then use Theorem 40 and the bound on the isoperimetry of the target measure from Theorem 6 to prove our main mixing time result, Theorem 4.

To prove Lemma 46, we start with some definitions. The overall plan is that we approximate the density of a Hamiltonian step as written in Equation (34) as in Equation (35) and bound its change going from x_0 to x_1 for most of the vectors v_{x_0} within a nice set in the tangent space of x_0.

Definition 41 *Consider a family of Hamiltonian curves $\gamma_s(t)$ for time interval $s, t \in [0, \delta]$ all ending at y, where $\gamma(0) = x$, and $\gamma'(0) = v_x$. Define the local push-forward density of $v_x \sim \mathcal{N}(0, g^{-1})$ onto y by*

$$P^{v_x}(y) = det(J_y^{v_x})\frac{\sqrt{|g(y)|}}{\sqrt{(2\pi)^n}}e^{-\|v_x\|_g^2/2}, \tag{34}$$

where $J_y^{v_x}$ is the inverse Jacobian of the Hamiltonian after time δ, sending v_x to y, which we denote by Ham^δ. we consider the Jacobian as an operator between the tangent spaces. The push forward density at y with respect to the manifold measure is given by

$$P(y) = \sum_{v_x:\ Ham^\delta(x, v_x) = y} P^{v_x}(y).$$

Note that $dg(y)$ refers to the manifold measure. Define the approximate local push-forward density of v_x as

$$\tilde{P}^{v_x}(y) = \exp\big(-\int_{t=0}^{\delta}\frac{t(\delta - t)}{2}tr(\Phi(t))dt\big)\sqrt{|g(y)|}/\sqrt{(2\pi)^n}e^{-\|v_x\|_g^2/2}. \tag{35}$$

Lemma 42 (Lemma 22 in Lee and Vempala (2018)) *For an R_1-normal Hamiltonian curve, for $0 \leq \delta^2 \leq \frac{1}{R_1}$ we have*

$$|\log(\tilde{P}^{v_x}(y)) - \log(P^{v_x}(y))| \leq \frac{(\delta^2 R_1)^2}{10}. \tag{36}$$

Lemma 43 (Lemma 32 in Lee and Vempala (2018)) *In the setting of Lemma 44, for an (R_1, R_3) normal γ_0, denoting $\frac{d}{ds}\gamma_s(0)$ by z, we have*

$$\frac{1}{2}\delta\frac{d}{ds}\|\gamma_s'(0)\|^2 \leq |\langle v_x, z\rangle| + 3\delta^2 R_3\|z\|.$$

Lemma 44 (Change of the pushforward density) *Consider the family of smooth Hamiltonian curves $\gamma_s(t)$ up to time δ from x_0 to x_1 pointing towards y, namely $\gamma_0(0) = x_0$, $\gamma_0(\delta) = y$, and $\gamma_s'(0) = v_x$ regarding a point $x = \gamma_s(0)$ along the geodesic between x_0 to x_1 whose tangent to the geodesic is $z \triangleq \frac{d}{ds}\gamma_s(0)$. Then, given that $\gamma_s(t)$ is (R_1, R_2, R_3) normal and $\|\nabla_{\frac{d}{ds}\gamma_s(t)}\gamma_s'(t)\| \leq 10/\delta$ for $0 \leq s, t \leq \delta$, under $\delta^2 \leq \frac{1}{R_1}$ we have*

$$\delta\frac{d}{ds}\log(\tilde{P}^{v_x}(y)) \leq |\langle v_x, z\rangle| + 2\delta^3 R_2 + 3\delta^2 R_3.$$

Proof Simply differentiating Equation (35):

$$\delta\Big|\frac{d}{ds}\log(\tilde{P}^{v_x}(y))\Big| = \Big| -\delta\frac{d}{ds}\Big(\int_{t=0}^{\delta}\frac{t(\delta-t)}{\delta}tr(\Phi(t))dt\Big) - \frac{1}{2}\delta\frac{d}{ds}\|v_x\|_g^2\Big|$$
$$\leq \delta\int_{t=0}^{\delta}\frac{t(\delta-t)}{\delta}\Big|\frac{d}{ds}tr(\Phi(t))\Big|dt + |\langle v_x, z\rangle| + 3\delta^2 R_3\|z\|.$$

where we used Lemma 43 with $v_x = \gamma_s'(0)$. Furthermore, using the R_2 normality property and noting our assumption $\|z\| = \|\frac{d}{ds}\gamma_s(0)\| = 1$, we have,

$$\Big|\frac{d}{ds}tr(\Phi(t))\Big| \leq R_2(\|\frac{d}{ds}\gamma_s(t)\| + \delta\|\nabla_{\frac{d}{ds}\gamma_s(t)}\gamma_s'(t)\|) \leq 11R_2,$$

where the last inequality follows from our assumption (which will be imposed by the definition of Nice sets in Lemma 45). Therefore,

$$LHS \leq |\langle v_x, z\rangle| + \frac{11}{6}\delta^3 R_2 + 3\delta^2 R_3.$$

∎

Lemma 45 (Change in probability of events under approximate density) *Let $Q_{x_0} \subseteq T_{x_0}(\mathcal{M})$ be a $(R_1, R_2, R_3, \delta, c)$ nice set in the tangent space of x_0 with $\delta^2 R_1 \leq 1$. Let x be an arbitrary point in the geodesic between x_0 and x_1. For vector v_x in the tangent space of x with $Ham^\delta(x, v_x) = y$ we can consider the family of hamiltonian curves $\gamma_s(t)$ between $x_0 = \gamma_0(0)$ and $x_1 = \gamma_\delta(0)$ with $\gamma_s(\delta) = y$ for all $0 \leq s \leq \delta$.Now let p_n be the finite measure obtained by restricting the normal distribution in the tangent space of x to vectors v_x for which the corresponding $v_{x_0} = \gamma_0'(0) \in Q_{x_0}$. For a point $y \in \mathcal{M}$, let $\tilde{P}_x^n(y)$ be the approximate pushforward density of p_n onto $\mathcal{M}$, defined as*

$$\tilde{P}^n(y) = \tilde{P}_x^n(y) = \Big(\sum_{v_x:\ Ham^\delta(x,v_x)=y,\ v_{x_0}\in Q_{x_0}}\tilde{P}^{v_x}(y)\Big)dg(y), \tag{37}$$

where $\tilde{P}_x^{v_x}(y)$ is defined in (35). We define $\tilde{P}^n(.)$ to be the corresponding finite measure. Now given a fixed event $Y \subset \mathcal{M}$ with probability $P^n(Y) \geq 0.001$, we have

$$\delta\Big|\frac{d}{ds}\log(\tilde{P}^n(Y))\Big| \leq 2c_1 + 3\delta^3 R_2 + 2\delta^2 R_3, \tag{38}$$

and for all Y:

$$\delta\Big|\frac{d}{ds}\log(\tilde{P}^n(Y))\Big| \leq c_1\sqrt{n} + 2\delta^3 R_2 + 3\delta^2 R_3, \tag{39}$$

where c_1 is a polylogarithmic factor in m and c. Note that $\tilde{P}_x^n$ depends on $x = \gamma_s(0)$, and we are fixing the set Q_{x_0} in the tangent space of x_0.

Proof Let $\tilde{P}_1^n$ be the density of further restricting $\tilde{P}^n$ to v_x's for which $\langle v_x, z\rangle \leq c_1$ for a factor c_1 that we pick polylogarithmically large in m, where recall $z \triangleq \frac{d}{ds}\gamma_s(0)$, and $\tilde{P}_2^n$ be such that $\tilde{P}^n(y) = \tilde{P}_1^n(y) + \tilde{P}_2^n(y)$. Note that

$$\Big|\frac{\frac{d}{ds}\tilde{P}^n(Y)}{\tilde{P}^n(Y)}\Big| = \Big(\frac{\frac{d}{ds}\tilde{P}_1^n(Y)}{\tilde{P}_1^n(Y)}\Big)\Big(\frac{\tilde{P}_1^n(Y)}{\tilde{P}^n(Y)}\Big) + \Big(\frac{\frac{d}{ds}\tilde{P}_2^n(Y)}{\tilde{P}_2^n(Y)}\Big)\Big(\frac{\tilde{P}_2^n(Y)}{\tilde{P}^n(Y)}\Big)$$
$$= \mathrm{LHS}_1\Big(\frac{\tilde{P}_1^n(Y)}{\tilde{P}^n(Y)}\Big) + \mathrm{LHS}_2\Big(\frac{\tilde{P}_2^n(Y)}{\tilde{P}^n(Y)}\Big). \tag{40}$$

But note that for the first term

$$\begin{aligned}\mathrm{LHS}_1 &\leq \int_Y \sum_{v_x:\ Ham^\delta(x,v_x)=y,\ v_{x_0}\in Q_{x_0},\ \langle v_x,z\rangle\lesssim 1} \Big|\Big(\frac{\frac{d}{ds}\tilde{P}^{v_x}(y)}{\tilde{P}^{v_x}(y)}\Big)\Big(\frac{\tilde{P}^{v_x}(y)}{\tilde{P}_1^n(y)}\Big)\Big(\frac{\tilde{P}_1^n(y)}{\tilde{P}_1^n(Y)}\Big)\Big| dg(y)\\ &\leq \int_Y \Big(\frac{\tilde{P}_1^n(y)}{\tilde{P}_1^n(Y)}\Big)\Big(\Big|\langle v_x, z\rangle\Big| + \delta^3 R_2 + \delta^2 R_3\Big)\delta^{-1} dg(y)\\ &\leq (c_1 + 2\delta^3 R_2 + 3\delta^2 R_3)/\delta.\end{aligned}$$

To see why the second line holds, note that the Hamiltonian curve from x to y is (R_1, R_2, R_3) normal from our assumption for time $t \in (0, \delta)$. The second line follows from Lemma 44 and the fact that $\frac{\tilde{P}^{v_x}(y)}{\tilde{P}_1^n(y)} \leq 1$. The third line follows simply by the choice $\langle v_x, z\rangle \leq c_1$.

Similarly for the second term

$$\mathrm{LHS}_2 \leq \Big(|\langle v_x, z\rangle| + 2\delta^3 R_2 + 3\delta^2 R_3\Big)/\delta \leq \Big(2c\sqrt{n} + 2\delta^3 R_2 + 3\delta^2 R_3\Big)/\delta,$$

where in the last inequality we are using the fact that $|\langle v_x, z\rangle| \leq 2c\sqrt{n}$. This is because $\tilde{P}_x^n$ is defined in Equation (37) as a sum on only v_{x_0}'s that are in Q_{x_0}, which means $\gamma_s(t)$ is a nice family of Hamiltonian curves. Hence, from the definition of Nice sets, we have $\|v_x\| = \|\gamma_s'(0)\| \leq 2c\sqrt{n}$. Therefore, a simple Cauchy Schwarz implies $|\langle v_x, z\rangle| \leq \|v_x\|_g \|z\|_g \leq 2c\sqrt{n}$.. Now first note that combining these and putting back in (40) implies

$$\delta\Big|\frac{d}{ds}\log(\tilde{P}^n(Y))\Big| \leq \Big(\frac{\tilde{P}_1^n(Y)}{\tilde{P}^n(Y)}\Big)\big(c_1 + 2\delta^3 R_2 + 3\delta^2 R_3\big) + \Big(\frac{\tilde{P}_2^n(Y)}{\tilde{P}^n(Y)}\Big)\big(2c\sqrt{n} + 2\delta^3 R_2 + 3\delta^2 R_3\big). \tag{41}$$

This immediately implies Equation (39) as $\tilde{P}^n(Y) = \tilde{P}_1^n(Y) + \tilde{P}_2^n(Y)$. To show case (38), first we use the fact that the densities regarding $\tilde{P}^n$ and P^n are within factor two of one another from Equation (36) and the assumption $\delta^2 R_1 \leq 1$, and combine it with the assumption $\tilde{P}^n(Y) \geq 0.001$ to get:

$$0.001 \leq P^n(Y) \leq 2\tilde{P}^n(Y), \tag{42}$$
$$\tilde{P}_2^n(Y) \leq 2P_2^n(Y). \tag{43}$$

Second, note that we can bound $\langle v_x, z\rangle$ as

$$\begin{aligned}|\langle v_x, z\rangle| = |\langle \gamma_s'(0), z\rangle| &\leq |\langle \gamma_0'(0), z\rangle| + \int_{r=0}^{s} |\langle \nabla_{\frac{d}{dr}\gamma_r(0)}\gamma_r'(0), z\rangle|\\ &\leq |\langle v_{x_0}, z\rangle| + \int_{r=0}^{s} \|\nabla_{\frac{d}{dr}\gamma_r(0)}\gamma_r'(0)\|\|z\| \leq |\langle \gamma_0'(0), z\rangle| + 10,\end{aligned} \tag{44}$$

where in the last inequality we used the property of the nice sets to bound $\|\nabla_{\frac{d}{dr}\gamma_r(0)}\gamma_r'(0)\|$. Note that the variable $\langle v_{x_0}, z\rangle$ is Gaussian with variance $\|z\|_g^2 = 1$. Therefore, using Gaussian tail bound (see 1), we can pick $c \geq 20$ large enough such that $|\langle v_x, z\rangle| \leq c_1$ happens with probability at least $1/(800c\sqrt{n})$, and $c_1 = O(\log(m) + \log(c))$. Combining this with Equation (44) implies $P_2^n(Y) \leq 1/(800c\sqrt{n})$ from the definition of P_2^n. Combining this with Equation (43), we get

$$\tilde{P}_2^n(Y) \leq 2/(800c\sqrt{n}). \tag{45}$$

Finally combining Equations (42) and (45) reveals $\tilde{P}_2^n(Y)/\tilde{P}^n(Y) \leq 1/(2c\sqrt{n})$, and plugging this into Equation (41) completes the proof of Equation (38).

■

Using the bounds on smoothness, we will show that one-step distributions of RHMC from two nearby points will have large overlap (and hence TV distance less than 1).

Lemma 46 (One-step coupling for RHMC) *Consider two points x_0 and x_1 and suppose Q_{x_0} is a $(R_1, R_2, R_3, \delta, c)$-nice set in the tangent space of x_0. Now given step size δ such that $\delta^2 \leq \frac{1}{R_1}$, $\delta^3 R_2 \leq 1$, $\delta^2 R_3 \leq 1$ and close by point x_1 such that $d(x_0, x_1) \leq \tilde{\delta}$ for $\tilde{\delta} = (0.1\delta)/(2c_1 + 5)$, where d is the distance on the manifold and $c_1 = O(\log(m) + \log 9c))$ is used in Lemma 45, then the total variation distance between P_{x_0} and P_{x_1} is upper bounded by 0.5.*

Proof Similar to (37), we define

$$P^n_x(y) = \Big(\sum_{v_x:\ Ham^\delta(x,v_x)=y,\ v_{x_0}\in Q_{x_0}} P^{v_x}(y)\Big) dg(y).$$

First, note that for any event $Z \subseteq \mathcal{M}$, from the definition of Nice sets

$$|P^n_{x_0}(Z) - P_{x_0}(Z)| \leq \mathbb{P}(v_x \notin Q_{x_0}) \leq 0.001. \tag{46}$$

Suppose $Y \subseteq \mathcal{M}$ be a set for which

$$P_{x_0}(Y) - P_{x_1}(Y) > 0.5. \tag{47}$$

This means $P_{x_0}(Y) \geq 0.5$, and moreover from (46)

$$P^n_{x_0}(Y) - P^n_{x_1}(Y) \geq P_{x_0}(Y) - P_{x_1}(Y) - \mathbb{P}(v_{x_0} \notin Q_{x_0}) \geq 0.499, \tag{48}$$

which also implies

$$P^n_{x_0}(Y) \geq 0.499. \tag{49}$$

But now using the assumptions on R_2 and R_3 and plugging Equation (49) into Equation (38) in Lemma 45 we can state

$$\tilde{\delta}\Big|\frac{d}{ds}\log(\tilde{P}^n_x(Y))\Big| \leq \tilde{\delta}(2c_1 + 5)/\delta \leq 0.1,$$

which by integrating from $s = 0$ to $s = \tilde{\delta}$ implies at time $s = \tilde{\delta}$ we have

$$\log(\tilde{P}^n_{\gamma_0(0)}(Y)) - \log(\tilde{P}^n_{\gamma_{\tilde{\delta}(0)}}(Y)) \leq 0.5,$$

or in other words

$$\tilde{P}^n_{x_0}(Y)/\tilde{P}^n_{x_1}(Y) \leq 1.11. \tag{50}$$

Now again applying the boundedness of the ratio between $\tilde{P}^n$ and P^n from Equation (36) and assumption $\delta^2 R_1 \leq 1$, we obtain

$$\tilde{P}^n_{x_1}(Y)/P^n_{x_1}(Y), P^n_{x_0}(Y)/\tilde{P}^n_{x_0}(Y) \leq 1.11, \tag{51}$$

which combined with Equation (50) means

$$P^n_{x_0}(Y)/P^n_{x_1}(Y) < 1.4.$$

This implies

$$P^n_{x_1}(Y) - P^n_{x_0}(Y) < 0.4.$$

This further implies from (46):

$$P_{x_0}(Y) - P_{x_1}(Y) \leq P^n_{x_0}(Y) - P^n_{x_1}(Y) + 0.001 \leq 0.401,$$

which contradicts Equation (47). This completes the proof. ■

Theorem 40 now follows easily from Lemma 46, Lemma 23, and Theorem 25.

Proof [Proof of Theorem 40] With the given choice of δ, Lemma 46 implies that, for every $x_0 \in S$ and every x_1 within distance $d(x_0, x_1) \leq O(\delta/c_1)$ for $c_1 = O(\log(m) + \log(c))$:

$$TV(\mathcal{T}_{x_0}, \mathcal{T}_{x_1}) \leq 0.01.$$

Using Theorem 25, for $\rho = \mathbb{P}(S^c) = O((\epsilon\delta\psi)/M)$ we get a lower bound on the s-conductance for $s = O(\epsilon/M)$:

$$\Phi_s^2 \geq \Omega(c_1^2(\psi\delta)^2).$$

Now using Lemma 23 with the same choice of s,

$$d_{TV}(\pi_t, \pi) \leq H_s + \frac{H_s}{s}\left(1 - \frac{\Phi_s^2}{2}\right)^t \leq \epsilon,$$

where we used the fact that $H_s \leq Ms = O(\epsilon)$ (recall the definition of M) and the fact that we pick t of the order $\log(M + 1/\epsilon)(\psi\delta)^2 c_1^2$ as $H_s/s \leq M$. The proof is complete. ■

Finally, Combining Theorems 40 and 15 and Lemma 16, we prove the main Theorem 4.

Proof [Proof of Theorem 4] Given a fixed parameter $c > 1$, Lemma 72 implies that there exists a high probability set $S = S_c \subseteq \mathcal{M}$, with

$$\pi(S) \geq 1 - poly(m)e^{-\Theta(c^2)}, \tag{52}$$

(Recall π is the distribution supported on the polytope with density $e^{-\alpha\phi}$) such that every $x_0 \in S$ has a corresponding set $Q_{x_0} \in T_p(\mathcal{M})$ such that

- The initial velocity $v_x \in T_p(\mathcal{M})$ which is distributed according to $\mathcal{N}(0, g(x)^{-1})$ is in Q_x with probability at least 0.999,
- For every $v_x \in Q_x$, the Hamiltonian curve with initial velocity v_x is $(c, 1)$-nice.

Now given that we pick parameter δ so that it satisfies the condition in Lemma 16, i.e.,:

$$\delta \leq \delta' = \frac{1}{\tilde{c}c_2c_3\sqrt{c^2 + \alpha\sqrt{\alpha}_0}n^{1/4}},$$

and given $c > 20$, then we conclude that for any family of Hamiltonian curves $\gamma_s(t)$ for which $\gamma_0'(0) = v_x$, γ_s is $(O(c), \delta)$-nice for every $s \in (0, \delta)$.

Furthermore, Corollary 66 also guarantees $\|\nabla_{\frac{d}{ds}\gamma_s(t)}\gamma_s'(t)\| \leq \frac{10}{\delta}$ for all $0 \leq s, t \leq \delta$. Combining these with Theorem 15 and since our hybrid barrier is $\alpha_0 n$-third-order-ℓ_∞-self-concordant according to Theorem 7, we conclude that Q_{x_0} is a (δ, R_1, R_2, R_3)-nice set with parameters:

$$\begin{aligned}
R_1 &= c_4(c^2 + \alpha\sqrt{\alpha_0})\sqrt{n}, \\
R_2 &= c_4(c^2 + \alpha\sqrt{\alpha_0} + \frac{c}{\sqrt{n}\delta})n, \\
R_3 &= c_4(c^2(\sqrt{n} + cn\delta) + n\delta c\alpha\sqrt{\alpha_0},
\end{aligned}$$

where $c_4 = O(c_3)$ and c_3 is the factor in Theorem 15 and recall $\alpha_0 = \left(\frac{m}{n}\right)^{\frac{2}{p+2}}$. As we see shortly, we pick $p = 4 - \lambda$ for $\lambda = O(1/\log(m))$, in which case c_3 and c_4 become polylogarithmic in m. Now for the same parameter c we considered above, we wish to satisfy the conditions in Theorem 40 on δ, namely $\delta^2 R_1(c) \leq 1$, $\delta^2 R_3(c) \leq 1$, $\delta^3 R_2(c) \leq 1$ (We have used this notation to emphasize that R_1, R_2, R_3 are functions of c).

Hence, the conditions on δ translate into

$$\begin{aligned}
\delta &= O(\frac{1}{c_4^{1/2} n^{1/4} c}), \\
\delta &= O(\frac{1}{c_4^{1/2} n^{1/3} c}), \\
\delta &= O(\frac{1}{c_4^{1/2} n^{1/3}}), \\
\delta &= O(\frac{1}{c_4^{1/2} n^{1/3} c^{1/3} (\alpha\sqrt{\alpha_0})^{1/3}}), \\
\delta &= O(\frac{1}{(c_4 + c_3 c_2 \tilde{c})^{1/2} \alpha^{1/2} {\alpha_0}^{1/4} n^{1/4}}).
\end{aligned}$$

Note that a sufficient condition on δ which satisfies all of the above constraints is (note that $c > 1$)

$$\delta = \frac{1}{c_5 c} \min\{\frac{1}{n^{1/3}}, \frac{1}{n^{1/3}(\alpha\sqrt{\alpha_0})^{1/3}}, \frac{1}{\alpha^{1/2}{\alpha_0}^{1/4} n^{1/4}}\}, \tag{53}$$

For a factor $c_5 = O(c_4^{1/2} + c_2 c_3)$.

Now to satisfy the condition $P(S) \geq 1 - O((\epsilon\delta\psi)/M)$ in Theorem 40, noting Equation (52), we set

$$c = \sqrt{\log\left(poly(m)M/(\epsilon\psi)\right)} = \Theta\left(\sqrt{\log(Mm/\epsilon)}\right).$$

On the other hand, from Theorem 6, we see that for the choice of $p = 4 - \lambda$ converging to 4 from below (λ is a parameter smaller than one), the square of the isoperimetry constant is $\psi^2 = \Theta(\max\{m^{-\frac{2/p}{2/p+1}} n^{-\frac{1}{2/p+1}}, \text{poly}(\frac{1}{4/p-1})\alpha\})$. Indeed, by picking $\lambda = O(1/\log(m))$ for small enough constant, we can make sure

$$\psi^2 = \Theta(\max\{m^{-1/3} n^{-2/3}, c_6\alpha\})$$

for $c_6 = \text{poly}(\frac{1}{4/p-1})$. On the other hand, note that by this choice of λ, we get $\frac{1}{4/p-1} = O(\log(m))$, hence the factors c_2, c_3, c_4, c_5, and c_6 become polylogarithmic in m. Now plugging this ψ and δ from (53) into Theorem 40 and noting the choice of c we get the following mixing bound:

$$\begin{aligned}
&\min\{\alpha^{-1}, n^{2/3}m^{1/3}\} \max\{n^{2/3}, n^{2/3}(\alpha\sqrt{\alpha_0})^{2/3}, n^{1/2}\alpha\sqrt{\alpha_0}\} \log(M + 1/\epsilon) \log(Mm/\epsilon) \textbf{polylog}(m) \log(c) \\
&= \min\{\alpha^{-1}, n^{2/3}m^{1/3}\} \max\{n^{2/3}, n^{2/3}(\alpha\sqrt{\alpha_0})^{2/3}, n^{1/2}\alpha\sqrt{\alpha_0}\} \log(M + 1/\epsilon) \log(M/\epsilon) \log\log(M/\epsilon)^2 \textbf{polylog}(m).
\end{aligned}$$

But note that if $\alpha\sqrt{\alpha_0}n^{1/2} \geq n^{2/3}$ or $n^{2/3}(\alpha\sqrt{\alpha_0})^{2/3} \geq n^{2/3}$, then $\alpha^{-1} \leq n^{2/3}m^{1/3}$. Hence, the mixing time boils down to

$$\begin{aligned}&\min\{\alpha^{-1}(n^{2/3} + n^{2/3}(\alpha\sqrt{\alpha_0})^{2/3} + \alpha\sqrt{\alpha_0}n^{1/2}), n^{4/3}m^{1/3}\}\log(M)\log(M/\epsilon)\log\log(M/\epsilon)^2\mathrm{polylog}(m),\\ \leq &\min\{\alpha^{-1}n^{2/3} + \alpha^{-1/3}n^{5/9}m^{1/9} + m^{1/6}n^{1/3}, n^{4/3}m^{1/3}\}\log(M/\epsilon)^2\log\log(M/\epsilon)^2\mathrm{polylog}(m).\end{aligned}$$

■

Appendix D. Geometry and Stability of Self-concordant Hessian Manifolds

In this section, we prove the smoothness of the operator $\Phi(t)$, namely we show that a nice Hamiltonian curve is (R_1, R_2, R_3) normal. Importantly, we do not explicitly work with the expansion of the metric g and its derivatives using our hybrid barrier. Instead we exploit the strong-self concordance property that we show in Theorem 7 to prove the desired smoothness bounds, hence our framework potentially can be applied more broadly. Interestingly, in order to bound the trace of certain operators that arise from bounding the smoothness of the Hamiltonian curves on manifold, it turns out that writing them as the average of random low rank tensors will enable us to apply our strong self-concordance estimates more efficiently and provide sufficient bounds to improve the mixing time. In this section, for sake of clarity of the presentation of the proofs, we use $\lesssim$ to indicate an inequality by ignoring the polylogarithmic factors in m.

D.1. Bounding R_1

Lemma 47 *Given a $\alpha_0 n$-third-order-ℓ_∞-self-concordant barrier ϕ, assuming $\gamma(t)$ is (δ, c)-nice, then for the parameter R_1 regarding the Frobenius norm bound of operator $\Phi(t)$, we have*

$$R_1 \leq c_3(c^2 + \alpha\sqrt{\alpha_0})\sqrt{n},$$

where $c_3 = poly(\frac{1}{4/p-1})$.

Proof Directly follows from Lemmas 52 and 48. ■

Throughout the proof of Lemmas 52 and 48 we assume $\gamma(t)$ is (δ, c)-nice and avoid repeating this condition. For fixed time t, we refer to $\gamma(t)$ and $\gamma'(t)$ by x and v respectively. Furthermore, when referring to Riemann and Ricci tensor, or operators Φ or M_x, we mean these operators on the tangent space on point x. First, recall the definition of the Frobenius norm:

$$\|\Phi(t)\|_F^2 = \mathbb{E}_{v_1,v_2\sim\mathcal{N}(0,g^{-1})}\mathbb{E}\langle v_1, \Phi(t)v_2\rangle^2.$$

To bound R_1, i.e. the Frobenius norm of $\Phi(t)$, note that

$$\Phi(t) = R(., v)v + M_x(.),$$

where R is the Riemann tensor and M_x is obtained from the bias vector μ. For brevity, sometimes we refer to M_x by M. In particular, for vector ℓ we have

$$\begin{aligned} R(\ell, v)v &= g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)\ell \\ &\quad + g^{-1}\mathrm{D}g(\ell)g^{-1}\mathrm{D}g(v)v, \\ M(\ell) &= \nabla_\ell(\nabla(\alpha\phi)) + \frac{1}{2}\nabla_\ell(g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g)), \end{aligned} \tag{54}$$

where $\mathtt{tr}(g^{-1}\mathrm{D}g)$ is a vector with its ith entry equal to $tr(g^{-1}\mathrm{D}_i g)$. We start from the Riemann tensor. The proof of this bound follows directly from the second-order ℓ_∞-self-concordance of g.

Lemma 48 (Frobenius norm of random Riemann tensor) *Assuming $\|v\|_{x,\infty} \leq c$, $\|v\|_g \leq c\sqrt{n}$, we have*

$$\|R(.,v)v\|_F \leq c^2\sqrt{n}.$$

Proof For the first term of $R(.,v)v$ as written in (54):

$$\begin{aligned}\mathbb{E}_{v_1,v_2\sim\mathcal{N}(0,g^{-1})}(v_1^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)v_2)^2 &= \mathbb{E}_{v_1,v_2} v_1^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)v_2 v_2^T \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)v_1\\ &= \mathbb{E}_{v_1,v_2} v_1^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)v_1\\ &\leq \|s_v\|_\infty^4 \mathbb{E}v_1^\top g v_1 = \|s_v\|_\infty^4 n \leq c^4 n.\end{aligned}$$

For the second term of the Riemann tensor:

$$\begin{aligned}\mathbb{E}_{v_1,v_2\sim\mathcal{N}(0,g^{-1})}(v_1^\top \mathrm{D}g(v_2)g^{-1}\mathrm{D}g(v)v)^2 &= \mathbb{E}_{v_1,v_2} v^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v_2)v_1 v_1^\top \mathrm{D}g(v_2)g^{-1}\mathrm{D}g(v)v\\ &= \mathbb{E}_{v_2} v^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v_2)g^{-1}\mathrm{D}g(v_2)g^{-1}\mathrm{D}g(v)v\\ &\leq \mathbb{E}_{v_2}\|v\|_{x,\infty}^2\|v_2\|_{x,\infty}^2 v^\top g v\\ &\lesssim \|v\|_{x,\infty}^2 v^\top g v \leq c^4 n.\end{aligned}$$

■

where we used Lemma 115 to bound $\mathbb{E}_{v_2}\|v_2\|_{x,\infty}^2$. Lemma 48 states $c^2\sqrt{n}$ as an upper bound on the Frobenius norm of $R(\ell,v)v$ given that the curve is nice.

Next, we prove a lemma regarding the expansion of the operator M_x, applying the covariant derivative.

Lemma 49 (Subterms for operator M_x) *We have the following expansion for the subterms of operator M_x:*

$$\begin{aligned}\langle\nabla_{v_1}(\nabla(\alpha\phi)), v_2\rangle &= v_2^\top \mathrm{D}g(\nabla(\alpha\phi))v_1 + v_2^\top \mathrm{D}^2(\alpha\phi)v_1,\\ \langle\nabla_{v_1}(g^{-1}tr(g^{-1}\mathrm{D}g)), v_2\rangle &= v_2^\top \mathrm{D}g(\xi)v_1 + v_2^\top \mathrm{D}(g\xi)v_1,\end{aligned} \tag{55}$$

where

$$\xi \triangleq g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g).$$

Moreover,

$$v_2^\top \mathrm{D}(g\xi)v_1 = -\mathtt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)) + \mathtt{tr}(g^{-1}\mathrm{D}^2 g(v_1.v_2)). \tag{56}$$

Proof By differentiating the first term:

$$\langle\nabla_{v_1}(\nabla(\alpha\phi)), v_2\rangle = \langle -g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}(\alpha\phi) + g^{-1}\mathrm{D}^2(\alpha\phi)[v_1] + g^{-1}\mathrm{D}g(\nabla(\alpha\phi))v_1, v_2\rangle.$$

But noting that $\nabla(\alpha\phi) = g^{-1}\mathrm{D}(\alpha\phi)$, the first and third terms are the same and we get the result. For the second term:

$$\langle\nabla_{v_1}(g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g)), v_2\rangle = v_2^\top \mathrm{D}g(\xi)v_1 + v_2^\top \mathrm{D}(g\xi)v_1. \tag{57}$$

Finally, for the second argument of the Lemma

$$\begin{aligned}v_2^\top \mathrm{D}(g\xi)v_1 &= v_2^\top \mathrm{D}(g\xi)(v_1)\\ &= -v_2^\top \mathtt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g) + v_2^\top \mathtt{tr}(g^{-1}\mathrm{D}^2 g(v_1,.))\\ &= -\mathtt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)) + \mathtt{tr}(g^{-1}\mathrm{D}^2 g(v_1,v_2)).\end{aligned}$$

■

Before we proceed, we state some useful bounds for ξ and its derivative that we use later in the proof.

Lemma 50 (Bound on $\|\xi\|$) *We have*

$$\|\xi\|_g \leq \sqrt{n}.$$

Proof We have

$$\begin{aligned}\|\xi\|_g^2 &= \mathtt{tr}(g^{-1}\mathrm{D}g)^\top g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g)\\ &= \mathbb{E}_{v,v'\sim\mathcal{N}(0,g^{-1})} v^\top \mathrm{D}g(v) g^{-1}\mathrm{D}g(v')v'\\ &\leq \mathbb{E}_{v\sim\mathcal{N}(0,g^{-1})} v^\top \mathrm{D}g(v) g^{-1}\mathrm{D}g(v)v\\ &\lesssim \mathbb{E}_v \|v\|_{x,\infty}^2 v^\top g v \lesssim n, \end{aligned} \tag{58}$$

where the first inequality above is due to Cauchy-Schwarz, the second one we used the first-order ℓ_∞-self-concordance to pull out $\|v\|_{x,\infty}$, and the third one is due to Lemma 115. ■

Next, we show the following bound on the derivative of ξ, i.e. $\|\mathrm{D}(\xi)(z)\|_g^2$:

Lemma 51 *For the derivative of ξ in direction z we have*

$$\|\mathrm{D}(\xi)(z)\|_g^2 \lesssim n\|z\|_g^2.$$

Proof Note that

$$\begin{aligned}\|\mathrm{D}(\xi)(z)\|_g^2 = &\ \mathtt{tr}(g^{-1}\mathrm{D}g)^\top g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(z)g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g) && \mathrm{LHS}_1\\ &+ \mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g)^\top g^{-1} tr(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g) && \mathrm{LHS}_2\\ &+ \mathtt{tr}(g^{-1}\mathrm{D}^2 g(z,.))^\top g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}^2 g(z,.)). && \mathrm{LHS}_3\end{aligned}$$

For the first term above,

$$\begin{aligned} tr(g^{-1}\mathrm{D}g)^\top g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(z)g^{-1}tr(g^{-1}\mathrm{D}g) &\leq \mathtt{tr}(g^{-1}\mathrm{D}g)^\top g^{-1/2}(g^{-1/2}\mathrm{D}g(z)g^{-1/2})^2 g^{-1/2}\mathtt{tr}(g^{-1}\mathrm{D}g)\\ &\leq \|z\|_g^2 \mathtt{tr}(g^{-1}\mathrm{D}g)^\top g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g). \end{aligned}$$

following the argument in (58):

$$\mathrm{LHS}_1 \leq \|s_{x,z}\|_\infty^2 n = \|z\|_{x,\infty}^2 n.$$

For the second term, we write the second g^{-1} within the tracec as an expectation $\mathbb{E}_{v'\sim\mathcal{N}(0,I)} v'v'^\top$, i.e.

$$\begin{aligned}\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g) &= \mathtt{tr}(\mathrm{D}g g^{-1}\mathrm{D}g(z)g^{-1})\\ &= \mathbb{E}_{v'}\mathtt{tr}(\mathrm{D}g g^{-1}\mathrm{D}g(z)v'v'^\top)\\ &= \mathbb{E}_{v'}\mathrm{D}g(v')g^{-1}\mathrm{D}g(z)v'.\end{aligned}$$

Therefore, using independent normal vectors $v, v' \sim \mathcal{N}(0, g^{-1})$, we can rewrite the second term as

$$\begin{aligned}\mathrm{LHS}_2 &= \mathbb{E}_{v,v'} v^\top \mathrm{D}g(z)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v')g^{-1}\mathrm{D}g(z)v'\\ &= \mathbb{E}_{v,v'} z^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v')g^{-1}\mathrm{D}g(v')z\\ &\leq \mathbb{E}_v z^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)z\\ &\lesssim \mathbb{E}_v \|z\|_g^2 \|v\|_{x,\infty}^4 \lesssim \|z\|_g^2.\end{aligned}$$

where the first inequality follows from Cauchy-Schwarz and the second one follows from first-order self-concordance, and the third one from Lemma 115. For the third term similarly

$$\begin{aligned}\mathrm{LHS}_3 &= \mathbb{E}_{v,v'\sim\mathcal{N}(0,g^{-1})} v^\top \mathrm{D}g(z,v) g^{-1} \mathrm{D}g(z,v')v' \\ &= \mathbb{E}_{v,v'} z^\top \mathrm{D}g(v,v) g^{-1} \mathrm{D}g(v',v')z \\ &\le \mathbb{E}_v z^\top \mathrm{D}g(v,v) g^{-1} \mathrm{D}g(v,v)z \\ &\lesssim \mathbb{E}_v \|z\|_g^2 \|v\|_{x,\infty}^4 \lesssim \|z\|_g^2.\end{aligned}$$

Combining all three bounds similar to our argument for $\nabla(\alpha\phi)$ we conclude

$$\|\mathrm{D}(\xi)(z)\|_g \lesssim \|z\|_{x,\infty}\sqrt{n} \le \|z\|_g \sqrt{n}. \tag{59}$$

■

Next, using Lemma 50 we bound the Frobenius norm of the M part in the following lemma, again only using second-order ℓ_∞-self-concordance of g to bound each of the four terms.

Lemma 52 (Frobenius norm of operator M) *We have*

$$\|M_x\|_F \lesssim \alpha\sqrt{\alpha_0 n}.$$

Proof To bound the Frobenius norm of the first part of the first term of operator M stated in Lemma 49:

$$\begin{aligned}\mathbb{E}_{v_1,v_2\sim\mathcal{N}(0,g^{-1})}(v_1^\top \mathrm{D}g(\nabla(\alpha\phi))v_2)^2 &= \mathbb{E}_{v_1,v_2} v_1^\top \mathrm{D}g(\nabla(\alpha\phi))v_2 v_2^\top \mathrm{D}g(\nabla(\alpha\phi))v_1 \\ &= \alpha^2 \mathbb{E}_{v_1} v_1^\top \mathrm{D}g(\nabla\phi) g^{-1} \mathrm{D}g(\nabla\phi)v_1 \\ &= \alpha^2 \mathbb{E}_{v_1} \nabla\phi^\top \mathrm{D}g(v_1) g^{-1} \mathrm{D}g(v_1)\nabla\phi \\ &\lesssim \alpha^2 \mathbb{E}_{v_1} \|v_1\|_{x,\infty}^2 \|\nabla\phi\|_g^2 \lesssim n\alpha^2\alpha_0,\end{aligned}$$

where in the second line we are switching v_1 and $\nabla\phi$ as $v_1^\top \mathrm{D}g(\nabla\phi)$ into $\nabla\phi^\top \mathrm{D}g(v_1)$, which is true due to the symmetry of the derivatives of the metric on Hessian manifolds, i.e. $\partial_k g_{ij} = \partial_i g_{jk} = \partial_j g_{ik}$. In the last inequality, we used the fact that ϕ is has $\alpha_0 n$ as its self-concordance parameter. For the second part of first term of M, note that $\mathrm{D}^2\phi = g$, so the Frobenius norm is at most n automatically. Next, for the first part of the second term of M, again based on Lemma 49

$$\begin{aligned}\mathbb{E}_{v_1,v_2\sim\mathcal{N}(0,g^{-1})}(v_1^\top \mathrm{D}g(\xi)v_2)^2 &= \mathbb{E} v_2^\top \mathrm{D}g(\xi) v_1 v_1^\top \mathrm{D}g(\xi)v_2 \\ &= \mathbb{E}\xi^\top \mathrm{D}g(v_2) g^{-1} \mathrm{D}g(v_2)\xi \\ &\le \mathbb{E}\|s_{x,v_2}\|_\infty^2 \xi^\top g\xi \le n,\end{aligned}$$

where in the last line we used Lemma 50. For the second part of the second term of M, from Lemma 49:

$$\mathbb{E}_{v_1,v_2}(v_1^\top \mathrm{D}(g\zeta)v_2)^2 \lesssim \mathbb{E}\mathtt{tr}^2(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)) + \mathbb{E}\mathtt{tr}^2(g^{-1}\mathrm{D}^2 g(v_1.v_2))$$

for the first part

$$\begin{aligned}\mathbb{E}\mathtt{tr}^2(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)) &= \mathbb{E}_{v_1,v_2,v}(\mathbb{E}_{v\sim\mathcal{N}(0,g^{-1})} v^\top \mathrm{D}g(v_1) g^{-1} \mathrm{D}g(v_2)v)^2 \\ &\le \mathbb{E}_{v_1,v_2,v}(v^\top \mathrm{D}g(v_1) g^{-1} \mathrm{D}g(v_2)v)^2 \\ &= \mathbb{E}_{v_1,v_2,v} v^\top \mathrm{D}g(v_1) g^{-1} \mathrm{D}g(v) v_2 v_2^\top \mathrm{D}g(v) g^{-1} \mathrm{D}g(v_1)v \\ &= \mathbb{E}_{v_1,v} v^\top \mathrm{D}g(v_1) g^{-1} \mathrm{D}g(v) g^{-1} \mathrm{D}g(v) g^{-1} \mathrm{D}g(v_1)v \\ &\lesssim \mathbb{E}_{v,v_1}\|s_{x,v}\|_\infty^2 \|s_{x,v_1}\|_\infty^2 \|v\|_g^2 \lesssim n.\end{aligned}$$

For the second part:

$$\begin{aligned}
\mathbb{E}_{v_1,v_2}\mathtt{tr}^2(g^{-1}\mathrm{D}^2 g(v_1,v_2)) &\leq \mathbb{E}_{v_1,v_2,v}(v^\top \mathrm{D}^2 g(v_1,v_2)v)^2\\
&= \mathbb{E}_{v_1,v_2,v}(v^\top \mathrm{D}^2 g(v_1,v)v_2)^2\\
&= \mathbb{E}_{v_1,v_2,v}v^\top \mathrm{D}^2 g(v_1,v)v_2 {v_2}^\top \mathrm{D}^2 g(v_1,v)v\\
&= \mathbb{E}_{v_1,v}v^T \mathrm{D}^2 g(v_1,v) g^{-1}\mathrm{D}^2 g(v_1,v)v\\
&\lesssim \mathbb{E}_{v_1,v}\|s_{x,v_1}\|_\infty^2\|s_{x,v}\|_\infty^2\|v\|_g^2 \lesssim n.
\end{aligned}$$

■

D.2. Bounding R_2

Here we state the bound on R_2.

Lemma 53 *Given a $(c_2,\alpha_0 n)$-third-order-ℓ_∞-self-concordant ϕ, for point $x = \gamma_s(t)$ on a (c,δ)-nice Hamiltonian curve ending at $\gamma_0(\delta)$ with $v = v_s(t) = \gamma_s'(t)$, namely that $\|\gamma_s'(t)\|_{\gamma_s(t),\infty} \leq c$ and $\|\gamma_s'(t)\|_g \leq c\sqrt{n}$ along the curve up to time $t = \delta$, suppose we wish to bound the change of the trace of the operator $\Phi(t)$ in parameter in direction $z = \frac{d}{ds}\gamma_s(t)$. Then the curve is R_2-normal with*

$$R_2 = nc_3(c^2 + \sqrt{\alpha_0}\alpha + \frac{c}{\sqrt{n}\delta}),$$

according to Definition 13, where $c_3 = poly(c_2)$. In particular,

$$|\frac{d}{ds}tr(\Phi(t))| \lesssim R_2\Big(\|\frac{d}{ds}\gamma_s(t)\|_g + \delta\|\nabla_{\frac{d}{ds}\gamma_s(t)}\gamma_s'(t)\|\Big).$$

Proof Directly from Lemmas 54 and 60. ■

In the following, we show Lemmas 54 and 60. Again we assume that γ_s is $(c,\delta)-$nice in all of the sections without repeating, and refer to $\gamma_s(t)$ by x. In sections D.2.1 and D.2.2, we bound the change in the M_x part and the Ricci part of Φ, respectively.

D.2.1. BOUNDING THE CHANGE IN M_x

Given a distribution $e^{-\alpha\phi(x)}$ that we want to sample from, we study the properties of the derivatives of the corresponding operator M which is defined as

$$M_x(v_1,v_2) = \langle \nabla_{v_1}\mu(x), v_2\rangle, \tag{60}$$

where

$$\mu(x) = \nabla(\alpha\phi)(x) + \frac{1}{2}g^{-1}tr(g^{-1}\mathrm{D}g) = g^{-1}\mathrm{D}(\alpha\phi) + \frac{1}{2}g^{-1}tr(g^{-1}\mathrm{D}g),$$

Recall from Lemma 49:

$$\text{LHS} = \langle \nabla_{v_1}(\nabla(\alpha\phi)) + \frac{1}{2}\nabla_{v_1}(g^{-1}tr(g^{-1}\mathrm{D}g)), v_2\rangle = \langle A_1(v_1), v_2\rangle + \langle A_2(v_1), v_2\rangle. \tag{61}$$

where we defined matrices $A_1(v_1)$ and $A_2(v_1)$. Here we introduce the main lemma of this section which bounds the derivative of the trace of M:

Lemma 54 (Bound on the change of operator M) *For the operator M defined in (60) and for any direction z, we have*

$$|\mathrm{D}(\mathtt{tr}M_x)(z)| \lesssim (1+\sqrt{\alpha_0}\alpha)n\|z\|_g.$$

Proof To prove Lemma 54, we bound the derivative of $tr(A_1)$ and $tr(A_2)$ in direction z separately in Lemmas 55 and 57. As a result, the proof of Lemma 54 directly follows from Lemmas 55 and 57. ■

We start from $tr(A_1)$ in the following Lemma.

Lemma 55 (Trace of A_1) *Regarding the operator $A_1(v_1) = \nabla_{v_1}(\nabla(\alpha\phi))$, we have*

$$\mathrm{D}(\mathtt{tr}(A_1))(z) \lesssim \alpha\sqrt{n\alpha_0}\|z\|_g.$$

Proof Note that from Lemma 49:

$$\begin{aligned}\mathrm{D}(\mathtt{tr}(A_1))(z) &= \langle \mathrm{D}(\nabla(\alpha\phi))[v_1] + \frac{1}{2}g^{-1}\mathrm{D}g(\nabla(\alpha\phi))v_1, v_2\rangle \\ &= v_2^\top \mathrm{D}g(\nabla(\alpha\phi))v_1 + v_2^\top \mathrm{D}^2\phi v_1.\end{aligned} \tag{62}$$

For the second part, note that $D^2(\alpha\phi) = \alpha g$. Hence

$$\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}^2\phi))(z) = 0.$$

So we only need to handle the derivative of the first part. First, note that the g-norm of the gradient $\nabla\phi$ is bounded.

Lemma 56 *For the gradient of the potential ϕ we have*

$$\|\mathrm{D}(\alpha\phi)\|_{g^{-1}} \leq \alpha\sqrt{n\alpha_0}.$$

Proof Directly from the fact that ϕ is a $\sqrt{\alpha_0 n}$-third-order self-concordant barrier. ■

Now we got back to bound the first term in (62), which we can expand as

$$\begin{aligned}\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}g(\nabla(\alpha\phi))))[z] &= \mathtt{tr}(g^{-1}\mathrm{D}^2g(z,\nabla(\alpha\phi))) + \mathtt{tr}(g^{-1}\mathrm{D}g(\mathrm{D}(\nabla(\alpha\phi))[z])) \qquad (63)\\ &\quad - \mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(\nabla(\alpha\phi))). \qquad (64)\end{aligned}$$

For the first term in (64), according to Lemma 56:

$$\begin{aligned}&\mathtt{tr}(g^{-1}\mathrm{D}^2g(z,\nabla(\alpha\phi)))\\ &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}{v'}^\top \mathrm{D}^2g(z,\nabla(\alpha\phi))v'\\ &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}{v'}^\top \mathrm{D}^2g(z,v')\nabla(\alpha\phi)\\ &\leq \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}\|s_{x,z}\|_\infty\|s_{x,v'}\|_\infty\sqrt{{v'}^\top g v'}\sqrt{\nabla(\alpha\phi)^\top g\nabla(\alpha\phi)}\\ &\lesssim \alpha\sqrt{\alpha_0}\sqrt{n}\sqrt{n}\|s_{x,z}\|_\infty\\ &\leq \alpha\sqrt{\alpha_0}n\|z\|_g,\end{aligned} \tag{65}$$

where we used Lemma 115 to bound $\mathbb{E}_{v'}\|s_{v'}\|_\infty\sqrt{v'^\top g v'}$ and used Lemma 95. For the second term in (64), we follow a similar reasoning:

$$\begin{aligned} tr(g^{-1}\mathrm{D}g(\mathrm{D}(\nabla(\alpha\phi))(z))) &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}{v'}^\top \mathrm{D}g(\mathrm{D}(\nabla(\alpha\phi))(z))v' \\ &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}{v'}^\top \mathrm{D}g(v')\mathrm{D}(\nabla(\alpha\phi))(z) \\ &\lesssim \mathbb{E}_{v'}\|v'\|_{x,\infty}\sqrt{v'^\top g v'}\sqrt{\mathrm{D}(\nabla(\alpha\phi))(z)^\top g\mathrm{D}(\nabla(\alpha\phi))(z)}. \end{aligned} \tag{66}$$

Therefore, bounding $\mathtt{tr}(g^{-1}\mathrm{D}g(\mathrm{D}(\nabla(\alpha\phi))(z)))$ boils down to bounding $\|\mathrm{D}(\nabla(\alpha\phi))(z)\|_g$. Now using the fact that $\nabla^2\phi = g$, we can write

$$\begin{aligned} \sqrt{\mathrm{D}(\nabla(\alpha\phi))(z)^\top g\mathrm{D}(\nabla(\alpha\phi))(z)} &= \alpha\sqrt{(-(\mathrm{D}\phi)^\top g^{-1}\mathrm{D}g(z)g^{-1} + z^\top)g(-g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}\phi + z)} \\ &\le \alpha\sqrt{(\mathrm{D}\phi)^\top g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}\phi} + \alpha\sqrt{z^\top g z} \\ &\lesssim \alpha\sqrt{\alpha_0 n}\|z\|_{x,\infty} + \alpha\|z\|_g. \end{aligned} \tag{67}$$

Plugging Equation (67) back into Equation (66) and using the fact that $\|z\|_{x,\infty} \le \|z\|_g$ implies the following bound on the second term in Equation (64):

$$\mathtt{tr}(g^{-1}\mathrm{D}g(\mathrm{D}(\nabla(\alpha\phi))(z))) \le n\sqrt{\alpha}_0\alpha\|z\|_g. \tag{68}$$

For the third term in (64), we reduce it to the first group of terms. Note that

$$\begin{aligned} \mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(\nabla(\alpha\phi))) &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}{v'}^\top\mathrm{D}g(z)g^{-1}\mathrm{D}g(v')\nabla(\alpha\phi) \\ &= \mathbb{E}_{v'}\sqrt{{v'}^\top\mathrm{D}g(z)g^{-1}\mathrm{D}g(z)v'}\sqrt{\nabla(\alpha\phi)^\top\mathrm{D}g(v')g^{-1}\mathrm{D}g(v')\nabla(\alpha\phi)} \\ &\lesssim \mathbb{E}_{v'}\|v'\|_{x,\infty}\|z\|_{x,\infty}\|v'\|_g\sqrt{\nabla(\alpha\phi)^\top g^{-1}\nabla(\alpha\phi)} \\ &\lesssim \alpha\sqrt{\alpha_0}n\|z\|_g, \end{aligned} \tag{69}$$

which is the same upper bound obtained in Equation (65) and (68). Note that we used Lemma 115 to bound $\mathbb{E}_{v'}\|v'\|_{x,\infty}\|v'\|_g$. Hence, combining Equations (65), (68), and (69) we conclude

$$\mathrm{D}(\mathtt{tr}(A_1))(z) \lesssim \alpha\sqrt{n\alpha_0}\|z\|_g.$$

■

Next, we focus on the second term in (61) and bound the derivative of the trace of the operator $A_2(v_1) = \nabla_{v_1}(g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g))$.

Lemma 57 (Trace of A_2) *For operator $A_2(v_1) = \nabla_{v_1}(g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g))$ as defined in Equation (61) we have*

$$|\mathrm{D}(\mathtt{tr}(A_2))(z)| \le n\|z\|_g.$$

Proof Recall the definition of ξ:

$$\xi = g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g).$$

From Lemma 49, we have

$$\langle\nabla_{v_1}(g^{-1}tr(g^{-1}\mathrm{D}g)), v_2\rangle = v_2^\top\mathrm{D}g(\xi)v_1 + v_2^\top\mathrm{D}(g\xi)v_1, \tag{70}$$

We bound the derivatives of the two terms in Equation (70) separately in Lemmas 58 and 59. Hence, the proof of Lemma 57 directly follows from these Lemmas. ■

We start from bounding the derivative of the first term in Equation (70), i.e. we wish to bound $|\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}g(\xi)))(z)|$.

Lemma 58 *Regarding the first quadratic form in Equation (70), we can bound its trace as*

$$|\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}g(\xi)))(z)| \lesssim n\|z\|_g.$$

Proof
According to Lemma 51, by substituting w with $D\xi(z)$ in Lemma 112, we get

$$|\mathtt{tr}(g^{-1}\mathrm{D}g(\mathrm{D}\xi(z)))| \leq \sqrt{n}\|\mathrm{D}\xi(z)\|_g \lesssim \|z\|_g n. \tag{71}$$

Moreover, according to Lemma 113 and Lemma 50:

$$|\mathtt{tr}(g^{-1}\mathrm{D}g(\xi, z))| \leq \sqrt{n}\|\xi\|_g\|z\|_g \leq n\|z\|_g. \tag{72}$$

Further, using Lemma 112 combined with Lemma 50:

$$|\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(\xi))| \leq \sqrt{n}\|z\|_g\|\xi\|_g \leq n\|z\|_g. \tag{73}$$

Finally, combining Equations (71), (72), and (73),

$$\begin{aligned}|\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}g(\xi)))(z)| &\lesssim |\mathtt{tr}(g^{-1}\mathrm{D}g(\mathrm{D}\xi(z)))| + |\mathtt{tr}(g^{-1}\mathrm{D}g(\xi, z))| + |\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(\xi))|\\ &\lesssim n\|z\|_g,\end{aligned}$$

completes the bound for the trace of the first part $Dg(\xi)$ of the operator in Equation (70) and the proof of Lemma 58 is complete. ∎

Finally, we move on to bound the derivative of the trace of the second operator in Equation (70), namely $\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}(g\xi)))(z)$.

Lemma 59 *We can bound the derivative of the trace of the second operator in Equation (70) as*

$$|\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}(g\xi)))(z)| \lesssim n\|z\|_g.$$

Proof Recall from Lemma 49:

$$\begin{aligned}v_2^\top \mathrm{D}(g\xi)v_1 &= -\mathtt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)) + \mathtt{tr}(g^{-1}\mathrm{D}^2g(v_1.v_2)) &(74)\\ &= -v_1^\top B_1 v_2 + v_1^\top B_2 v_2. &(75)\end{aligned}$$

Now we wish to calculate the derivative of the trace of this operator, namely

$$\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}(g\xi)))(z). \tag{76}$$

We separate the case when the derivative hits the outer g^{-1} in (76). First, we calculate the derivative with respect to the outer g^{-1} regarding the term $\mathtt{tr}(g^{-1}B_1)$:

$$|\mathtt{tr}(\mathrm{D}(g^{-1})(z)B_1)| = |\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}B_1)|. \tag{77}$$

Note that

$$v_1^\top B_1 v_2 = \mathtt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)).$$

Note that this 2-form is symmetric and PSD since

$$\mathtt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_1)) = \mathtt{tr}((g^{-1/2}\mathrm{D}g(v_1)g^{-1/2})^2) \geq 0.$$

Moreover, note that

$$g^{-1}\mathrm{D}g(z)g^{-1} \leq \|z\|_{x,\infty}g^{-1}.$$

Hence, Equation (77) can further be upper bounded as

$$\|z\|_{x,\infty}tr(g^{-1}B_1) = \|z\|_{x,\infty}\mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}\mathtt{tr}(g^{-1}\mathrm{D}g(v')g^{-1}\mathrm{D}g(v')).$$

But we have already bounded the operator norm of $\mathtt{tr}(g^{-1}\mathrm{D}g(v')g^{-1}\mathrm{D}g(v'))$ in Lemma 111 by $\tilde{O}(\|s_{x,v'}\|_\infty^2)$, which implies its trace can be at most $n\tilde{O}(\|s_{x,v'}\|^2)$. Taking expectation, we have

$$\mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}\mathtt{tr}(g^{-1}\mathrm{D}g(v')g^{-1}\mathrm{D}g(v')) \lesssim n.$$

Hence, we conclude

$$|\mathtt{tr}(\mathrm{D}(g^{-1})(z)B_1)| \lesssim n\|s_{x,z}\|_\infty. \tag{78}$$

On the other hand, note that for the second term in Equation (75), there is a symmetry between the inner and outer g^{-1}:

$$\mathtt{tr}(g^{-1}B_2) = \mathtt{tr}(g^{-1}\mathrm{D}^2g[g^{-1}]).$$

Hence, it is sufficient to bound when taking derivative with respect z hit one of them, for example the inner g^{-1}.

Therefore, we move on to taking derivative with respect to the $\mathrm{D}(g\xi)$ part of $\mathtt{tr}(g^{-1}\mathrm{D}(g\xi))$. For this, we can again use the trick of writing g^{-1} as $\mathbb{E}_{v\sim\mathcal{N}(0,g^{-1})}vv^\top$:

$$\mathtt{tr}(g^{-1}\mathrm{D}(g\xi)) = \mathbb{E}_v v^\top\mathrm{D}(g\xi)v.$$

But from Equation (75), we have

$$\mathtt{tr}(g^{-1}\mathrm{D}(g\xi)) = -\mathtt{tr}(g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)) + \mathtt{tr}(g^{-1}\mathrm{D}^2g(v,v)).$$

Now taking derivative with respect to z:

$$|\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}(g\xi)))(z)| \leq |\mathbb{E}_v\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)))(z)| + |\mathbb{E}_v\mathrm{D}(\mathtt{tr}(g^{-1}\mathrm{D}g(v,v)))(z)| \tag{79}$$
$$= \mathrm{LHS}_1 + \mathrm{LHS}_2. \tag{80}$$

But for the first term in (80), we can write:

$$\begin{aligned}\mathrm{LHS}_1 &\leq 2\mathbb{E}_v|\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v))| + 2\mathbb{E}_v|\mathtt{tr}(g^{-1}\mathrm{D}^2g(v,z)g^{-1}\mathrm{D}g(v))| \\ &\lesssim \mathbb{E}_v\|z\|_{x,\infty}\mathtt{tr}((g^{-1/2}\mathrm{D}g(v)g^{-1/2})^2) + \mathbb{E}_v\|g^{-1/2}\mathrm{D}g(v,z)g^{-1/2}\|_1\|g^{-1/2}\mathrm{D}g(v)g^{-1/2}\|_{op} \\ &\lesssim \mathbb{E}_v\|v\|_{x,\infty}^2\|z\|_{x,\infty}n \\ &\leq \mathbb{E}_v\|v\|_{x,\infty}^2\|z\|_g n \\ &\lesssim \|z\|_g n.\end{aligned} \tag{81}$$

For the second term in (80):

$$\begin{aligned}\mathrm{LSH}_2 &\leq |\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(v,v))| + |\mathtt{tr}(g^{-1}\mathrm{D}g(v,v,z))| \\ &\leq \|g^{-1/2}\mathrm{D}g(z)g^{-1/2}\|\|g^{-1/2}\mathrm{D}g(v,v)g^{-1/2}\|_1 + \|s_v\|_\infty^2\|s_z\|_\infty\mathtt{tr}(g^{-1}g) \\ &\lesssim n\|z\|_{x,\infty}\|v\|_{x,\infty}^2 \\ &\leq \mathbb{E}_v n\|z\|_g\|v\|_{x,\infty}^2 \\ &\lesssim n\|z\|_g.\end{aligned} \tag{82}$$

where we used the third order self-concordance property of g with respect to the infinity norm, as shown in section I, and also Lemma 95. Combining Equations (78), (81), and (82) completes the proof of Lemma 59. ■

D.2.2. Bounding the Change in the Ricci Tensor

First, we state the main result of this section, which is a bound on the change of the Ricci tensor.

Lemma 60 (Bound on the change of Ricci tensor) *Given the assumptions of Lemma 53, we have*

$$\Big|\frac{d}{ds}\mathtt{Ricci}(v_s(t), v_s(t))\Big| \lesssim nc^2,$$

where $v_s(t) = \gamma_s'(t)$. Note that we drop the indices s, t from $v_s(t)$ for brevity.
Proof According to Lemma 101 `Ricci` has two terms. We start analyzing the first term:

$A_1 := -\frac{1}{4}\mathtt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2))$ term Taking derivative of this subterm of Ricci tensor in direction z:

$$\mathrm{D}A_1(z) = -\frac{1}{4}\mathtt{tr}(g^{-1}\mathrm{D}g(v,z)g^{-1}\mathrm{D}g(v)) + \frac{1}{4}\mathtt{tr}(g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(v)).$$

Now we use Lemmas 35 and 34 to bound these terms:

$$\begin{aligned}\mathtt{tr}(g^{-1/2}\mathrm{D}g(v,z)g^{-1/2}g^{-1/2}\mathrm{D}g(v)g^{-1/2}) &\le \|g^{-1/2}\mathrm{D}g(v,z)g^{-1/2}\|_F\|g^{-1/2}\mathrm{D}g(v)g^{-1/2}\|_F\\ &\lesssim \|v\|_{x,\infty}^2\|z\|_{x,\infty}\|g^{-1/2}gg^{-1/2}\|_F^2\\ &\le n\|v\|_{x,\infty}^2\|z\|_{x,\infty}\\ &\le n\|v\|_{x,\infty}^2\|z\|_g \le nc^2\|z\|_g.\end{aligned}$$

Similarly

$$\begin{aligned}&\mathtt{tr}(g^{-1/2}\mathrm{D}g(z)g^{-1/2}g^{-1/2}\mathrm{D}g(v)g^{-1/2}g^{-1/2}\mathrm{D}g(v)g^{-1/2})\\ &\le \|g^{-1/2}\mathrm{D}g(v)g^{-1/2}\|\|g^{-1/2}\mathrm{D}g(z)g^{-1/2}\|_F\|g^{-1/2}\mathrm{D}g(v)g^{-1/2}\|_F\\ &\lesssim n\|v\|_{x,\infty}^2\|z\|_g \le nc^2\|z\|_g.\end{aligned}$$

Terms in the derivative of A_1 that involves the derivative of v Note that $v = v_s(t)$ is a function of s so we can take its derivative in direction $z = \frac{d}{ds}\gamma_s(t)$. Differentiating v with respect to z,

$$\begin{aligned}\mathtt{tr}(g^{-1}\mathrm{D}g(Dv(z))g^{-1}\mathrm{D}g(v)) &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}v'^\top\mathrm{D}g(\mathrm{D}v(z))g^{-1}\mathrm{D}g(v)v'\\ &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}\mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}\mathrm{D}v(z)^\top{}^\top\mathrm{D}g(v')g^{-1}\mathrm{D}g(v')v\\ &\le \mathbb{E}_{v'}\sqrt{\mathrm{D}v(z)^\top\mathrm{D}g(v')g^{-1}\mathrm{D}g(v')\mathrm{D}v(z)}\sqrt{v^\top\mathrm{D}g(v')g^{-1}\mathrm{D}g(v')v}\\ &\lesssim \mathbb{E}_{v'}\|v'\|_{x,\infty}^2\|\mathrm{D}v(z)\|_g\|v\|_g\\ &\lesssim (\|v\|_{x,\infty}\|z\|_g + \|\nabla_z v\|_g)c\sqrt{n}\\ &\le c^2\sqrt{n}\|z\|_g + \delta\frac{c\sqrt{n}}{\delta}\|\nabla_z v\|_g.\end{aligned}$$

where we used Lemma 107 to bound $\|\mathrm{D}v(z)\|_g$ by $\|z\|_g$ and $\|\nabla_z v\|_g$, and also the fact that $\mathbb{E}\|v'\|_{x,\infty} = O(\sqrt{\log(m)})$ from Lemma 115. ■

Second part of the Ricci Tensor. We should take derivative of $v^\top \mathrm{D}g(g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g))v$ in direction z, which is the second term in the Ricci tensor according to Lemma 101. As a warm up, we first bound the value of this term before taking derivative:

Taking derivative in direction z. First, we differentiate the inner g^{-1} term in $v^\top \mathrm{D}g(v)g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g)$:

$$\begin{aligned}\mathrm{D}(v^\top \mathrm{D}g(v)g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g))(z) &\to v^\top \mathrm{D}g(v)g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g)\\ &= \mathbb{E}_{v'} v^\top \mathrm{D}g(v)g^{-1}\mathtt{tr}(v'^\top \mathrm{D}gg^{-1}\mathrm{D}g(z)v')\\ &= \mathbb{E}_{v'} v^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v')g^{-1}\mathrm{D}g(v')z\\ &\lesssim \mathbb{E}_{v'}\|v'\|_{x,\infty}^2\|v\|_{x,\infty}\|v\|_g\|z\|_g \lesssim c^2\sqrt{n}\|z\|_g.\end{aligned}$$

For the remaining derivatives we can substitute the inner g^{-1} by $\mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}v'v'^\top$. Specifically, for the ones which do not involve derivative with respect to v:

$$\begin{aligned}\mathbb{E}_{v'}|\mathrm{D}(v^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v')v')(z)| &\le \mathbb{E}_{v'}|v^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v',z)v'| + \mathbb{E}_{v'}|v^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(v')v'|\\ &\lesssim \mathbb{E}_{v'}\|v\|_{x,\infty}\|v\|_g\|v'\|_g\|z\|_{x,\infty}\\ &\le c^2 n\|z\|_g.\end{aligned}$$

Finally we have to check when z hits v. Again for Gaussian variable $v'\sim\mathcal{N}(0,g^{-1})$.

$$\begin{aligned}\mathrm{D}(v^\top \mathrm{D}g(v)g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g))(z) &\to \mathrm{D}v(z)^\top \mathrm{D}g(v)g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g)\\ &= \mathbb{E}_{v'}\mathrm{D}v(z)^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v')v'\\ &\le \mathbb{E}_{v'}\sqrt{\mathrm{D}v(z)^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)\mathrm{D}v(z)}\sqrt{v'^\top \mathrm{D}g(v')g^{-1}\mathrm{D}g(v')v'}\\ &\lesssim \mathbb{E}_{v'}\|\mathrm{D}v(z)\|_g\|s_v\|_\infty\|s_{v'}\|_\infty\|v'\|_g\\ &\lesssim c\sqrt{n}(\|v\|_{x,\infty}\|z\|_g + \|\nabla_z v\|_g) \le c^2\sqrt{n}\|z\|_g + \delta\frac{c\sqrt{n}}{\delta}\|\nabla_z v\|_g,\end{aligned}$$

where we used Lemma 107 to bound $\|\mathrm{D}v(z)\|_g$ in terms of $\|z\|_g$ and $\|\nabla_z v\|_g$.

D.3. Bounding R_3

Here we bound the parameter R_3 which is defined as the maximum possible value of the norm of $\Phi(t)\zeta(t)$, where $\zeta(t)$ is the parallel transport of the initial velocity. The idea is to bound the infinity norm of $\zeta(t)$ along the Hamiltonian curve, then show a more efficient bound compared to the naive operator norm of $\Phi(t)$ which works with both of the norms $\|s_{\zeta(t)}\|_\infty$ and $\|\zeta(t)\|_g$.

Recall the definition of the parameter R_3:

$$\|\Phi(t)\zeta(t)\|_g \le R_3$$

where $\zeta(t)$ is the parallel transport of $\gamma'(0)$ along the Hamiltonian curve $\gamma(t)$.

Lemma 61 (Bound on R_3) *For a $\alpha_0 n$-third-order-self-concordant barrier and given that the curve $\gamma(t)$ is (c,δ)-nice, we have*

$$R_3 = c_3(c^2(\sqrt{n} + cn\delta) + n\delta c\alpha\sqrt{\alpha_0}),$$

up to time δ, where $c_3 = poly(c_2)$.

Proof From the definition of niceness, we have a c upper bound on the infinity norm $\|s_{\gamma'}\|_\infty$. Using that, we can apply Lemma 64 to obtain

$$\|s_\zeta\|_\infty \le \delta c\sqrt{n}.$$

Finally combining this with Lemmas 62 and 63:

$$\|\Phi(t)\zeta\|_g \le c^2\|\zeta\|_g + \|\zeta\|_\infty(c^2 + \alpha\sqrt{\alpha_0})\sqrt{n} \le c^2 n^{1/2} + c\sqrt{n}\delta(c^2 + \alpha\sqrt{\alpha_0})\sqrt{n} = c^2(\sqrt{n} + cn\delta) + n\delta c\alpha\sqrt{\alpha_0}.$$

■

Here we show a norm bound for $\Phi(t)$ which we used to bound R_3. To this end, we show bounds on the Riemann tensor $R(,v)v$ and operator M separately in Lemmas 62 and 63.

Lemma 62 (Operator norm of random Riemann tensor) *Assuming $\|s_v\|_\infty \le c,\ \|v\|_g \le c\sqrt{n}$, we have*

$$\|R(\ell, v)v\|_g \le c^2\|\ell\|_g + c^2\sqrt{n}\|s_\ell\|_\infty \le c^2\sqrt{n}\|\ell\|_g.$$

Proof Similar to Lemma 48, using the form of Riemann expansion in Equation (54):

$$\begin{aligned}
\|R(\ell, v)v\|_g &\le (\ell^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)\ell)^{1/2} \\
&\quad + (v^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(\ell)g^{-1}\mathrm{D}g(\ell)g^{-1}\mathrm{D}g(v)v)^{1/2} \\
&\le \|v\|_{x,\infty}^2(\ell^\top g\ell)^{1/2} + \|v\|_{x,\infty}\|\ell\|_{x,\infty}\|v\|_g \\
&\le c^2\|\ell\|_g + c^2\sqrt{n}\|\ell\|_{x,\infty}.
\end{aligned}$$

■

Next, we state a similar mix norm bound for operator M.

Lemma 63 (Operator norm of M) *we have*

$$\|M(x)\ell\|_g \le \|\ell\|_g + (1 + \alpha\sqrt{\alpha_0})\sqrt{n}\|s_\ell\|_\infty.$$

Proof Recall from Lemma (49):

$$\langle M(x)v_1, v_2\rangle = \langle \nabla_{v_1}(\nabla\phi) + \frac{1}{2}\nabla_{v_1}(g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g)), v_2\rangle.$$

Starting from the first part of the term $\langle \nabla_{v_1}(\nabla\phi), v_2\rangle$:

$$\begin{aligned}
\|g^{-1}\mathrm{D}g(\nabla\phi)\ell\|_g &= \mathtt{tr}^{1/2}(\ell^\top \mathrm{D}g(\nabla\phi)g^{-1}\mathrm{D}g(\nabla\phi)\ell) \\
&= \mathtt{tr}^{1/2}((\nabla\phi)^\top \mathrm{D}g(\ell)g^{-1}\mathrm{D}g(\ell)\nabla\phi) \\
&\le \|s_\ell\|_\infty\|\nabla\phi\|_g \le \alpha\sqrt{n\alpha_0}\|s_\ell\|_\infty.
\end{aligned}$$

Note that for the second part, $\mathrm{D}^2\phi = g$, hence the corresponding operator is the identity and has operator norm one.
Next, we move on to the second term of M in (55). For the first part of it from Equation (70), we have:

$$\begin{aligned}
\|g^{-1}\mathrm{D}g(\xi)\ell\|_g &= \sqrt{\ell^\top \mathrm{D}g(\xi)g^{-1}\mathrm{D}g(\xi)\ell} \\
&= \sqrt{\xi^\top \mathrm{D}g(\ell)g^{-1}Dg(\ell)\xi} \\
&\le \|s_\ell\|_\infty\sqrt{\xi^\top g\xi} = \|s_\ell\|_\infty\sqrt{n}.
\end{aligned}$$

where we used Lemma 51. For the second part, note that from Equation (56):

$$v_2^\top \mathrm{D}(g\xi)v_1 = \mathtt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)) + \mathtt{tr}(g^{-1}\mathrm{D}g(v_1, v_2)). \tag{83}$$

Starting from the first part, now we rewrite this term in a better way as

$$\mathtt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)) = \mathbb{E}_v\mathtt{tr}(vv^\top \mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)) = \mathbb{E}_v v^\top \mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)v = \mathbb{E}_v v_1^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)v_2.$$

Now due to Lemma 111 the norm of the corresponding operator is one:

$$\mathbb{E}\|g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)\ell\|_g \leq \mathbb{E}_v\|v\|_{x,\infty}^2\|\ell\|_g \lesssim \|\ell\|_g. \tag{84}$$

For the second part in (83), we write it as

$$\mathtt{tr}(g^{-1}\mathrm{D}g(v_1, v_2)) = \mathbb{E}_v tr(vv^\top \mathrm{D}g(v_1, v_2)) = \mathbb{E}_v v_1 \mathrm{D}g(v, v)v_2.$$

Hence, the operator norm is bounded as

$$\mathbb{E}_v\|g^{-1}\mathrm{D}g(v, v)\ell\|_g \leq \mathbb{E}\|v\|_{x,\infty}^2\|\ell\|_g \lesssim \|\ell\|_g.$$

■

Next, we show a bound on the derivative of the infinity norm of the parallel transported vector ζ given that we know the infinity norm of γ' is constant (randomness + stability).

Lemma 64 (Infinity-norm of parallel transport) *Given $\delta \leq \frac{1}{c}$ and a (c, δ)-nice Hamiltonian curve γ, we have for $t \leq \delta$:*

$$\|s_{\zeta(t)}\|_\infty \leq \delta c\sqrt{n},$$

where ζ is the parallel transport of $\gamma'(0)$ along the curve.

Proof As ζ is the parallel transport vector, from opening up the covariant derivative being zero:

$$\begin{aligned}\frac{d}{dt}(A\zeta) &= A\zeta' - (A\zeta) \odot (A\gamma') \\ &= -\frac{1}{2}Ag^{-1}\mathrm{D}g(\gamma')\zeta - (A\zeta) \odot (A\gamma'),\end{aligned}$$

which implies using Lemma 95:

$$\begin{aligned}\|\frac{d}{dt}(A\zeta)\|_\infty &\lesssim \|Ag^{-1}\mathrm{D}g(\gamma')\zeta\|_\infty + \|s_{\gamma'}\|_\infty\|s_\zeta\|_\infty \\ &\lesssim \|g^{-1}\mathrm{D}g(\gamma')\zeta\|_g + \|s_{\gamma'}\|_\infty\|s_\zeta\|_\infty \\ &\lesssim c\|\zeta\|_g + c\|s_\zeta\|_\infty \leq c\sqrt{n} + c\|s_\zeta\|_\infty,\end{aligned}$$

where we used $\|s_{\gamma'}\|_\infty \lesssim c$ from the definition of niceness and the fact that parallel transport preserves the norm of ζ and $\|\zeta(0)\|_g = \|\gamma'(0)\|_g \leq \sqrt{n}$. This ODE implies to avoid blow up we should pick $\delta \lesssim \frac{1}{c}$. Under this condition, we further get

$$\|s_\zeta\|_\infty \lesssim \delta c\sqrt{n},$$

which completes the proof. ■

In the next section, we show the stability of the infinity norm and the metric norm of γ' along the curve, $c_t(s)$ where $c_t(s) = \gamma_s(t)$ is defined for a fixed time t.

Appendix E. Regularity and Stability of Hamiltonian curves

In this section, we show that the niceness property holds for Hamiltonian curves with high probability, and its stability gives rise to a family of nice Hamiltonian curves.

E.1. Stability of the Niceness Property

Here, we show that niceness property of Hamiltonian curves is stable in Lemma 65. This Lemma combined with Lemma 72 shows that with high probability, all of the curves in the Hamiltonian family are nice. This result can then be used by Theorem 15 (proved in Section D) and the fact that we constructed a suitable barrier in Section F to show the existence of nice sets, which we then use in Theorem 40 along with the isoperimetry result to prove the mixing time bound.

Lemma 65 (Restatement of Lemma 16) *In the same setting as Theorem 15 for a $(c_2, \alpha_0 n)$-third-order ℓ_∞-self-concordant barrier, suppose we are given a family of Hamiltonian curves $\gamma_s(t)$ with $s, t \in (0, \delta)$ with*

$$\delta^2 \leq \frac{1}{\tilde{c} c_2 c_3 (c^2 + \alpha\sqrt{\alpha}_0)\sqrt{n}},$$

where c_3 is the factor defined in the argument of Lemma 47 (will be polylog(m) in our setting) and $\tilde{c}$ is a universal constant. Now given that γ_0 is (c, δ)-nice for $c > 20$, then $\gamma_s(t)$ is also $(O(c), \delta)$-nice for all $0 \leq s \leq \delta$. In particular, given that for all $0 \leq t \leq \delta$, $\|s_{\gamma_0(t), \gamma_0'(t)}\|_\infty \leq c$ and $\|\gamma_0'(t)\|_g \leq \sqrt{n}$, then for all $0 \leq t \leq \delta$ and $0 \leq s \leq \delta$,

$$\|\gamma_s'(t)\|_{\gamma_s(t),\infty} \leq 12c,$$
$$\|\gamma_s'(t)\|_g \leq 2c\sqrt{n}.$$

Proof Suppose we denote the time until which we run the Hamiltonian curve by δ, i.e. $0 \leq t \leq \delta$. Suppose the argument is not true, and consider the set S to be the times $0 \leq s \leq \delta$ for which $f(s) = sup_{0 \leq t \leq \delta} \|s_{\gamma'(t,s)}\|_\infty < 12c$. Since $f(s)$ is continuous, the set S is open. Hence, if we consider the infimum s_0 of times s for which $f(s) \geq 1$, then the infimum is attained, i.e. $f(s_0) = 12c$, while $f(s) < 12c$ for every time $s < s_0$. Exactly the same way we can define the function $f_2(s) = \sup_{0 \leq t \leq \delta} \|\gamma_s'(t)\|_g$ and the first time s_1 (if it exists) for which $f_2(s_1) = 2c\sqrt{n}$ while $f_2(s) < 2c\sqrt{n}$ for $s < s_1$.
First assume the case where $s_0 \leq s_1$. Now again from the continuity of f and the fact that $[0, \delta]$ is a compact set, its supremum is attained in some time t_0. This means

$$\|s_{\gamma_s'(t_0)}\|_\infty < 12c, \tag{85}$$
$$\|\gamma_s'(t_0)\|_g < 2c\sqrt{n}, \tag{86}$$

for all $s < s_0$, while $\|s_{\gamma_{s_0}'(t_0)}\|_\infty = 12c$. But now using this infinity norm bound for times $s \leq s_0$ (for the fixed time t_0), we can obtain a Frobenius norm bound for $\Phi(\gamma_s(t_0), \gamma_s'(t_0))$ from Lemma 47 as

$$\|\Phi(\gamma_s(t_0), \gamma_s'(t_0))\|_F \leq R_1 = R_1(12c) = O(c_3((12c)^2 + \alpha\sqrt{\alpha_0})\sqrt{n}),$$

Let $\varrho(s)$ be the curve $\gamma_s(t_0)$ in parameter s. Now we pick the constant $\tilde{c}$ small enough to satisfy the condition $\delta^2 R_1 \leq 1$, so can apply Lemma 23 in Lee and Vempala (2018), which implies

$$\|\nabla_{\varrho'(s)} \gamma_s'(t_0)\|_g \leq 10/\delta,$$

for every $s < s_0$, where we are using the fact that $\|\varrho'(s)\|_g = 1$. But note that for $s < s_0$ we can write

$$\begin{aligned}
\|\frac{d}{ds}(A\gamma')\|_\infty &\leq \|A\nabla_{\varrho'}\gamma' - Ag^{-1}\mathrm{D}g(c')\gamma'\|_\infty + \|s_{\gamma,\varrho'}\|_\infty\|s_{\gamma,\gamma'}\|_\infty \\
&\leq \|\nabla_{\varrho'}\gamma'\|_g + \|g^{-1}Dg(\varrho')\gamma'\|_g + \|s_{\gamma,\varrho'}\|_\infty\|s_{\gamma,\gamma'}\|_\infty \\
&= \|\nabla_{\varrho'}\gamma'\|_g + \|g^{-1}\mathrm{D}g(\gamma')\varrho'\|_g + \|s_{\gamma,\varrho'}\|_\infty\|s_{\gamma,\gamma'}\|_\infty \\
&\leq \|\nabla_{\varrho'}\gamma'\|_g + c_2\|\gamma'\|_{\gamma,\infty}\|\varrho'\|_g + \|\varrho'\|_g\|\gamma'\|_{\gamma,\infty} \qquad (87) \\
&\leq \frac{10}{\delta} + c_2 c + c = \frac{10}{\delta} + (c_2+1)c,
\end{aligned}$$

where the first line follows from opening the definition of covariant derivative and in line (87) we have used the first-order ℓ_∞-self-concordance of g. Finally, this ODE implies that $\|\gamma_s'(t_0)\|_{\gamma_s(t_0),\infty} \leq s(\frac{10}{\delta} + (c_2 + 1)c) < 12c$ for all times $s < s_0$ (with the choice of $\tilde{c}$ constant large enough), which from continuity holds also for time s_0. But this contradicts $|s_{\gamma'_{s_0}(t_0)}\|_\infty = 12c$, which completes the proof for the case $s_0 \leq s_1$.

Next, we consider the latter case $s_1 < s_0$. Similar to the above argument, until time $s \leq s_1$ we have the Frobenius bound on $\Phi(t)$ from Lemma 47, and again as we have $\delta^2 R_1 \leq 1$, from Lemma 23 in Lee and Vempala (2018),

$$\|\nabla_{\varrho'(s)}\gamma'(t_0)\| \leq 10/\delta,$$

for $s \leq s_1$. Now we write an ODE to control the norm of $\|\gamma'_{s_1}(t_0)\|_g$ where t_0 is defined in the same way as the previous case, and get a contradiction:

$$\frac{d}{ds}\|\gamma'\|_g^2 = 2\langle\nabla_{\frac{d}{ds}\gamma(t_0,s)}\gamma', \gamma'\rangle \leq 2\|\gamma'\|\|\nabla_{\frac{d}{ds}\gamma(t_0,s)}\gamma'\| \leq \frac{20}{\delta}\|\gamma'\|_g,$$

which implies

$$\frac{d}{ds}\|\gamma'\|_g \leq \frac{20}{\delta}.$$

Therefore, at time $s = \delta$ the change in $\|\gamma'\|_g$ from its initial value is less than $20 < c\sqrt{n}$, which means the value of $\|\gamma'\|_g$ should have remained less than $2c\sqrt{n}$. The contradiction completes the proof for the second case. ∎

Corollary 66 *In the same setting as Lemma 16, we have for all* $0 \leq s, t \leq \delta$,

$$\|\nabla_{\frac{d}{ds}\gamma_s(t)}\gamma_s'(t)\| \leq \frac{10}{\delta}.$$

Proof Since from Lemma 16 we have R_1 normality for all $s, t \leq \delta$ with $\delta^2 R_1 \leq 1$, then from Lemma 23 in Lee and Vempala (2018) we have the desired result. ∎

Next, we show a helper lemma regarding the derivative of $\gamma_s'(t)$ in direction $\frac{d}{ds}\gamma_s(t)$:

Lemma 67 *On a* (c,δ)*-nice Hamiltonian curve with* $\delta \leq \frac{1}{n^{1/4}c}$*, We have:*

$$\|\frac{d}{ds}\gamma'(t)(s)\| \leq 1/\delta.$$

Proof Note that from Lemma 16 we have $\|s_{\gamma'_s(t_0)}\|_\infty \le c$. Hence, from Lemma 47, we can apply Lemma 23 in Lee and Vempala (2018) to obtain

$$\|\nabla_{\frac{d}{ds}\gamma}\gamma'(t)\|_g \lesssim \frac{1}{\delta}. \tag{88}$$

But now from Lemma 107, setting $v = \gamma'(t,s)$ and $z = \frac{d}{ds}\gamma(t,s)$:

$$\|\frac{d}{ds}\gamma'(t,s)\|_g \le \|\gamma'\|_\infty \|\frac{d}{ds}\gamma\|_g + \|\nabla_{\frac{d}{ds}\gamma}\gamma'(t)\|_g.$$

From Lemma 16, we have $\|\gamma'\|_\infty \le c$ and note that from our assumption on the s parameterization, $\|\frac{d}{ds}\gamma\|_g = 1$, which combined with Equation (88) finishes the proof. ■

E.2. Niceness of Hamiltonian Curves with High Probability

The goal of this section is to prove Lemma 72. This lemma shows the niceness of the Hamiltonian curve at parameter $s = 0$, which then implies the niceness of all the curves in the family by Lemma 16. To prove Lemma 72, first we show a bound on the g-norm along the Hamiltonian curve given a bound at time zero.

Recall the ODE related to RHMC for curve γ is

$$\mathrm{D}_t^2\gamma(t) = \mu(\gamma),$$

where D_t refers to covariant differentiation along the curve $\gamma(t)$. Opening it up implies

$$\gamma''(t) + \frac{1}{2}g^{-1}Dg(\gamma')\gamma' = \mu.$$

First, we show a non-random bound on the norm $\|\gamma'\|_g$ given a bound at time zero.

Lemma 68 (Boundedness of $\|.\|_g$ along the Hamiltonian curve) *Suppose at time zero, $\|\gamma'(0)\|_g \le c\sqrt{n}$. Then for all times $t \le 2$, we have*

$$\|\gamma'(t)\|_g \le c + t(1 + \alpha\sqrt{\alpha_0})\sqrt{n}.$$

Proof Note that

$$\|g^{-1}\mathrm{D}g(\gamma'(t))\gamma'(t)\|_g \le \|\gamma'(t)\|_{\gamma(t),\infty}\|\gamma'(t)\|_g,$$

hence, taking covariant derivative

$$\begin{aligned}\frac{d}{dt}\|\gamma'(t)\|_g^2 &= 2\langle \nabla_{\gamma'(t)}\gamma'(t), \gamma'(t)\rangle \\ &\le 2\|\gamma'(t)\|\|\mu(\gamma(t))\| \le 2(1+\alpha\sqrt{\alpha_0})\sqrt{n}\|\gamma'(t)\|,\end{aligned}$$

where we used Lemma 104 to bound $\|\mu\|$. This implies

$$\frac{d}{dt}\|\gamma'(t)\| \le (1+\alpha\sqrt{\alpha_0})\sqrt{n}.$$

Solving this ODE,

$$\|\gamma'(t)\| \le (c + t(1+\alpha\sqrt{\alpha_0}))\sqrt{n}. \tag{89}$$

■

Lemma 69 (Stability bound on $\|.\|_{x,\infty}$ along the curve) *In the same setting as Theorem 15 for a $(c_2, \alpha_0 n)$-third-order ℓ_∞-self-concordant barrier, for a Hamiltonian curve with $\|\gamma'(0)\|_g \leq c\sqrt{n}$, suppose for a fixed time $t_1 \leq 1$ we know $\|s_{\gamma'(t_1)}\|_\infty \leq c$ for $c \geq 1$. Then for all times $t \in (t_1 - \frac{1}{10c_2c(1+\alpha\sqrt{\alpha_0})\sqrt{n}}, t_1 + \frac{1}{10c_2c(1+\alpha\sqrt{\alpha_0})\sqrt{n}})$ we have*

$$\|\gamma'(t)\|_{\gamma(t),\infty} \leq 2c.$$

Proof Consider the Hamiltonian ODE below:

$$\gamma''(t) + \frac{1}{2}g^{-1}\mathrm{D}g(\gamma'(t))\gamma'(t) = \mu(\gamma(t)).$$

which implies

$$\frac{d}{dt}(\mathrm{A}_\gamma\gamma') = -\frac{1}{2}\mathrm{A}_\gamma g^{-1}\mathrm{D}g(\gamma')\gamma' + \mathrm{A}_\gamma\mu - (S_{\gamma,\gamma'})^{\odot 2}.$$

Hence, using Lemma 95

$$\begin{aligned}\|\frac{d}{dt}(\mathrm{A}_\gamma\gamma')\|_\infty &\leq \|g^{-1}\mathrm{D}g(\gamma')\gamma'\|_g + \|s_{\gamma,\gamma'}\|_\infty^2 + \|\mathrm{A}_\gamma\mu\|_\infty \\ &\leq c_2\|\gamma'\|_{\gamma,\infty}\|\gamma'\|_g + \|\gamma'\|_{\gamma,\infty}^2 + \|\mu\|_g,\end{aligned}$$

where we used the c_2-first-order ℓ_∞-self-concordance of the barrier and the fact that $\|\mathrm{A}_\gamma\mu\|_\infty = \|\mu\|_{\gamma,\infty} \leq \|\mu\|_g$ from Lemma 95. But using Lemma 68, having upper bound $c\sqrt{n}$ on the g-norm of γ' at time zero implies a bound $2c\sqrt{n}$ on the tangent norm of the curve until $t \leq 2$, and in particular for all $t \in (t_1 - \frac{1}{10c_2c(1+\alpha\sqrt{\alpha_0}\sqrt{n})}, t_1 + \frac{1}{10c_2c(1+\alpha\sqrt{\alpha_0}\sqrt{n})})$. Combining this with Lemma 104:

$$\|\frac{d}{dt}(\mathrm{A}_\gamma\gamma')\|_\infty \leq 2c_2c\sqrt{n}\|\gamma'\|_{\gamma,\infty} + \|\gamma'\|_{\gamma,\infty}^2 + (1+\alpha\sqrt{\alpha_0})\sqrt{n}.$$

This ODE implies that if at time t_1 the infinity norm bound $\|\gamma'(t_1)\|_{\gamma,\infty} \leq c$ for $c \geq 1$, then for times within $t_1 \pm \frac{1}{10c_2c(1+\alpha\sqrt{\alpha_0})\sqrt{n}}$ we have an $2c$ bound on the infinity norm, which completes the proof. ■

Lemma 70 (Stability bound on $\|.\|_g$ along the curve) *For a Hamiltonian curve with $\|\gamma'(0)\|_g \leq \sqrt{n}$, suppose for a fixed time t_1 we know $\|\gamma'(t_1)\|_g \leq c$. Then for all times $t \in (t_1 - \frac{1}{(1+\alpha\sqrt{\alpha_0})\sqrt{n}}, t_1 + \frac{1}{(1+\alpha\sqrt{\alpha_0})\sqrt{n}})$ we have*

$$\|\gamma'(t)\|_g \leq 2c.$$

Proof This follows directly from Lemma 68. ■

Now combining Lemmas 69 and 70, we uniformly control the norms $\|.\|_g$ and $\|.\|_\infty$ along the Hamiltonian curve.

Lemma 71 (Uniform control on the $\|.\|_g$ and $\|.\|_{x,\infty}$ along the curve) *Suppose we pick x random from $e^{-\alpha\phi(x)}$ then run a Hamiltonian curve starting from x with initial vector $\gamma'(0)$ as $\mathcal{N}(0, g^{-1})$. Then, with probability at least $1 - poly(m)ce^{-\Theta(c^2)}$, for any time $t_1 \in (0,1)$, we have*

$$\begin{aligned}&\|s_{\gamma'(t_1)}\|_\infty \leq c, \\ &\|\gamma'(t_1)\|_g \leq c\sqrt{n}.\end{aligned}$$

Proof From the property of the Hamiltonian curve, we know the joint density of $(\gamma(t), \gamma'(t))$ is $e^{-\alpha\phi(x)} \times \mathcal{N}(0, g^{-1}(x))dxdv$. Focusing on the probability of the vector field $v_t = \gamma'(t)$, we see that for each i, $a_{\gamma,i}^\top v_t$ is a Gaussian distributed variable with variance

$$a_{\gamma,i}^\top g^{-1} a_{\gamma,i} \leq 1,$$

for $a_{\gamma,i}$ the ith row of $\mathrm{A}_{\gamma(t)}$, where the inequality follows from Lemma 94. Hence, from Gaussian tail bound 1 and a union bound on top, for a fixed time t we have:

$$\mathbb{P}(\|s_{v_t}\|_\infty \geq c) \leq 2me^{-c^2/2},$$

where note that $\|s_{v_t}\|_\infty$ is just the maximum of Gaussian random variables and we applied a union bound over the entries of s_{v_t}. Moreover, note that $\|v_t\|_g$ is a sub-Gaussian random variable with mean $O(\sqrt{n})$ and sub-Gaussian parameter $O(1)$. Hence

$$\mathbb{P}(\|v_t\|_g \geq c\sqrt{n}) \lesssim e^{-\Theta(c^2)}.$$

Next, consider a cover $\mathcal{C} = \{t_i\}_{i=1}^{c(1+\alpha\sqrt{\alpha_0})\sqrt{n}}$ of equally distant times of the Hamiltonian curve from $t = 0$ to $t = 1$. Apply the above argument for all the times in this cover with a union bound on top. This implies with probability at least $1 - \mathrm{poly}(m)ce^{-\Theta(c^2)}$, we have $\|s_{v_t}\|_\infty \leq c$ and $\|v_t\|_g \leq c\sqrt{n}$ for all times $t \in \mathcal{C}$, where we used the fact that $\alpha\sqrt{\alpha_0} = poly(m)$. Now combining this with Lemmas 69 and 70 completes the proof. ∎

Next, we prove a lemma that shows the existence of nice sets, which we use in the proof of Theorem 4.

Lemma 72 (Niceness of Hamiltonian Curves) *There is a high probability region $S \subset \mathcal{M}$ such that $\pi(S) \geq 1 - O(poly(m)e^{-\Theta(c^2)})$ (where recall $\pi(.)$ is the probability distribution of density $e^{-\alpha\phi}$ inside the polytope) and for every $x \in S$, there is a high probability region Q_x in the tangent space of x, namely $\mathbb{P}(v_x \in Q_x) \geq 0.999$ such that for all $v_x \in Q_x$, the Hamiltonian curve starting from x with initial vector v_x is $(c, 1)$-nice, namely for all $0 \leq t \leq 1$:*

$$\|\gamma'(t)\|_g \leq c\sqrt{n},$$
$$\|s_{\gamma'(t)}\|_\infty \leq c.$$

Proof For every point $x \in \mathcal{M}$, define Q_x to be the set of vectors in its tangent space such that the resulting curve is c-nice up to time 1. Define region S to be the the set of points x on $\mathcal{M}$ such that $p_{v_x}(Q_x) \geq 1 - 0.0005$, where p_{v_x} denotes the density of $\mathcal{N}(0, g^{-1})$ in the tangent space of x (The constant $1 - 0.0005$ is motivated by the definition of nice sets). Now if it was the case that $\mathbb{P}(S^c) \geq \mathrm{poly}(m)ce^{-\Theta(c^2)}$, then under the joint distribution on (x, v), there is a region with probability at least $poly(m)ce^{-\Theta(c^2)}$ such that the Hamiltonian curve starting from x with initial vector v is not c-nice. But this contradicts Lemma 71. ∎

Appendix F. Derivation of Higher-order Estimates for Lewis Weights

In this section we prove the estimates stated in Section B.2 for the derivatives of the Lewis weights and the matrices $\mathbf{P}_x$,$\mathbf{G}_x$, and $\mathbf{\Lambda}_x$ defined in Section A.2. Just for this section, the notation $\lesssim$ only hides universal constants. Namely, we drop factors of p, as it is constant for $p < 4$, and only collect factors of $\frac{1}{4/p-1}$ because it goes to infinity as $p \to 4$.

Our estimates build on Lemma 26, which we restate here for convenience:

Lemma 73 (Operator $\left\|.\right\|_{\infty\to\infty}$ norm bound on the Jacobian) *For $p < 4$, given $y = \mathbf{G}_x^{-1}\mathbf{W}_x r$ for any vector $r \in \mathbb{R}^m$, we have*

$$\|y\|_\infty \leq \frac{1}{4/p - 1}\|r\|_\infty.$$

In particular, Lemma 26 allows us to estimate the first derivative of the Lewis weights based on the $\|\|_{x,\infty}$ norm, as stated in Lemma 21. Next, we make an important observation in Lemma 74 which enables us to drive estimate the first derivative of $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ in Lemma 29.

Lemma 74 *We have*

$$-\frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathbf{P}_x \odot \tilde{\mathbf{P}}_{x,v} \preccurlyeq \frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x.$$

Proof For the matrix $\tilde{\mathbf{P}}_{x,v}$ using Schur product theorem and the fact that

$$\tilde{\mathbf{P}}_{x,v} = \mathbf{P}_x \mathrm{R}_{x,v}\mathbf{P}_x \preccurlyeq \|v\|_{x,\infty}\mathbf{P}_x\mathbf{P}_x = .\|v\|_{x,\infty}\mathbf{P}_x,$$

we have

$$\begin{aligned}
\mathbf{P}_x \odot \tilde{\mathbf{P}}_{x,v} &\preccurlyeq \|r_{x,v}\|_\infty \mathbf{P}_x \odot \mathbf{P}_x \\
&\preccurlyeq \frac{1}{4/p-1}\|s_{x,v}\|_\infty \mathbf{P}_x^{(2)} \\
&\preccurlyeq \frac{1}{4/p-1}\|s_{x,v}\|_\infty \mathbf{W}_x,
\end{aligned}$$

and similarly

$$-\frac{1}{4/p-1}\|s_{x,v}\|_\infty \mathbf{W}_x \preccurlyeq \mathbf{P}_x \odot \tilde{\mathbf{P}}_{x,v}.$$

■

Now using on Lemma 74, we can prove our estimates for the first derivatives of $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$.

Lemma 75 *For the derivatives of $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ at some point x we have*

$$\begin{aligned}
-\frac{1}{4/p-1}\|z\|_{x,\infty}\mathbf{W}_x &\preccurlyeq \mathrm{D}\mathbf{G}_x(z) \preccurlyeq \frac{1}{4/p-1}\|z\|_{x,\infty}\mathbf{W}_x, \\
-\frac{1}{4/p-1}\|z\|_{x,\infty}\mathbf{W}_x &\preccurlyeq \mathrm{D}\mathbf{\Lambda}_x(z) \preccurlyeq \frac{1}{4/p-1}\|z\|_{x,\infty}\mathbf{W}_x.
\end{aligned}$$

Proof Note that from Lemma 28 we have

$$\begin{aligned}
\mathrm{D}\mathbf{P}_x^{(2)} &= 2\mathbf{P}_x \odot \mathrm{D}\mathbf{P}_x \\
&= 2\mathbf{P}_x \odot (-\mathbf{P}_x\mathrm{R}_{x,v} - \mathrm{R}_{x,v}\mathbf{P}_x + 2\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x) \\
&= -2\mathbf{P}_x^{(2)}\mathrm{R}_{x,v} - 2\mathrm{R}_{x,v}\mathbf{P}_x^{(2)} + 4\mathbf{P}_x \odot \tilde{\mathbf{P}}_{x,v}.
\end{aligned}$$

But we have already estimated $\mathbf{P}_x \odot \tilde{\mathbf{P}}_{x,v}$ in Lemma 74. It remains to deal with the first two terms. But for arbitrary vector $r \in \mathbb{R}^m$,

$$\begin{aligned}
r^\top \mathbf{P}_x^{(2)}\mathrm{R}_{x,v} r &\leq \sqrt{r^\top \mathbf{P}_x^{(2)} r}\sqrt{(\mathrm{R}_{x,v}r)^\top \mathbf{P}_x^{(2)}\mathrm{R}_{x,v}r} \\
&\leq \sqrt{r^\top \mathbf{W}_x r}\sqrt{(\mathrm{R}_{x,v}r)^\top \mathbf{W}_x \mathrm{R}_{x,v}r}.
\end{aligned}$$

Now using $\mathbf{P}_x^{(2)} \preccurlyeq \mathbf{W}_x$ (Lemma 108),

$$r^\top \mathbf{P}_x^{(2)} \mathrm{R}_{x,v} r \leq \|r\|_{w(x)} \|\mathrm{R}_{x,v} r\|_{w(x)} \\ \leq \|r_{x,v}\|_\infty \|r\|_{w(x)}^2 \\ \leq \frac{1}{4/p-1} \|v\|_{x,\infty} \|r\|_{w(x)}^2,$$

where the last line follows from the definition of $r_{x,v}$ and Lemma 29. Since r was arbitrary, we get

$$-\frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathbf{P}_x^{(2)}\mathrm{R}_{x,v} + \mathrm{R}_{x,v}\mathbf{P}_x^{(2)} \preccurlyeq \frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x.$$

This implies

$$-\frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}\mathbf{P}_x^{(2)} \preccurlyeq \frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{W}_x. \tag{90}$$

Finally noting the fact that $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ are a linear combination of $\mathbf{P}_x^{(2)}$ and $\mathbf{W}_x$ and we have the same estimates as Equation (90) for $\mathbf{W}_x$ in Lemma 27, the proof is complete. ■

Next, we move on to bound the second order derivatives of $\mathbf{W}_x$. To this end, we first estimate the first order derivative of $r_{x,v}$, but before that, first we need to generalize our result in Lemma 26. We start by making the following observation.

Lemma 76 *For arbitrary $r \in \mathbb{R}^m$,*

$$\|\mathbf{W}_x^{-1}\mathbf{P}_x \odot \tilde{\mathbf{P}}_{x,z} r\|_\infty \leq \frac{1}{4/p-1}\|r\|_\infty \|z\|_{x,\infty}.$$

Proof For the ith entry of $\mathbf{W}_x^{-1}\mathbf{P}_x \odot \tilde{\mathbf{P}}_{x,z} r$ we can write

$$|e_i^\top (\mathbf{P}_x \odot \mathbf{P}_x \mathrm{R}_{x,z} \mathbf{P}_x) r_{x,v}| = |(r_{x,z} \odot \mathbf{P}_{xi,})^\top \mathbf{P}_x (r \odot \mathbf{P}_{xi,})| \\ \leq \|r_{x,z} \odot \mathbf{P}_{xi,}\|_2 \|r \odot \mathbf{P}_{xi,}\|_2 \\ \leq \|r_{x,z}\|_\infty \|r\|_\infty w_i \\ \leq \frac{1}{(4/p-1)^2}\|z\|_{x,\infty}\|r\|_\infty w_i.$$

■

Now based on Lemma 76, we generalize Lemma 26 below.

Lemma 77 *For arbitrary $z \in \mathbb{R}^n$,*

$$\left\|\mathbf{G}_x^{-1}\mathbf{W'}_{x,z}\right\|_{\infty\to\infty} \lesssim \frac{1}{(4/p-1)^2}\|z\|_{x,\infty}, \\ \left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\right\|_{\infty\to\infty} \lesssim \frac{1}{(4/p-1)^2}\|z\|_{x,\infty}, \\ \left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{\Lambda}_x(z)\right\|_{\infty\to\infty} \lesssim \frac{1}{(4/p-1)^2}\|z\|_{x,\infty}.$$

Proof

From Lemma 76, we know

$$\left\|\mathbf{W}_x^{-1}(\mathbf{P}_x \odot (\mathbf{P}_x \mathrm{R}_{x,z} \mathbf{P}_x))\right\|_{\infty\to\infty} \leq \frac{1}{4/p-1}\|z\|_{x,\infty}.$$

Moreover,

$$\left\|\mathbf{W}_x^{-1}(\mathbf{P}_x \odot (\mathrm{R}_{x,z} \mathbf{P}_x))\right\|_{\infty\to\infty} \leq \|s_{x,z}\|_\infty \left\|\mathbf{W}_x^{-1}\mathbf{P}_x^{(2)}\right\|_{\infty\to\infty} \leq \|z\|_{x,\infty},$$

where the inequality follows as the sum of the rows of the ith row of $\mathbf{P}_x^{(2)}$ is equal to $w(x)_i$. Similarly we have

$$\left\|\mathbf{W}_x^{-1}(\mathbf{P}_x \odot (\mathbf{P}_x \mathrm{R}_{x,z}))\right\|_{\infty\to\infty} \leq \|z\|_{x,\infty}.$$

Therefore, according to the expansion of $\mathrm{D}\mathbf{P}_x^{(2)}(z)$ as in Lemma 28, we get

$$\left\|\mathbf{W}_x^{-1}(\mathrm{D}\mathbf{P}_x^{(2)}(v))\right\|_{\infty\to\infty} \lesssim \frac{1}{4/p-1}\|z\|_{x,\infty}. \tag{91}$$

On the other hand, note that

$$\left\|\mathbf{W}_x^{-1}\mathbf{W'}_{x,z}\right\|_{\infty\to\infty} \lesssim \frac{1}{4/p-1}\|z\|_{x,\infty}, \tag{92}$$

by Lemma 27, since $\mathbf{W}_x$ and $\mathbf{W'}_{x,z}$ are diagonal. Now since $\mathbf{G}_x$ is a linear combination of matrices $\mathbf{W}_x$ and $\mathbf{P}_x^{(2)}$, by combining Equations (91) and (92):

$$\left\|\mathbf{W}_x^{-1}\mathrm{D}\mathbf{G}_x(v)\right\|_{\infty\to\infty} \lesssim \frac{1}{4/p-1}\|z\|_{x,\infty}. \tag{93}$$

Now we can write

$$\begin{aligned}\left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\right\|_{\infty\to\infty} &= \left\|\mathbf{G}_x^{-1}\mathbf{W}_x\mathbf{W}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\right\|_{\infty\to\infty}\\ &\leq \left\|\mathbf{G}_x^{-1}\mathbf{W}_x\right\|_{\infty\to\infty}\left\|\mathbf{W}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\right\|_{\infty\to\infty}\\ &\lesssim \frac{1}{(4/p-1)^2}\|z\|_{x,\infty}.\end{aligned}$$

The argument for $\mathbf{W}_x$ and $\mathbf{\Lambda}_x$ follows similarly. ∎

Next, we proceed to bound the derivative of $r_{x,v}$.

Lemma 78 *The derivative of $r_{x,v}$ in direction z can be estimated as*

$$\|\mathrm{D}(r_{x,z})(v)\|_\infty \lesssim \frac{1}{(4/p-1)^4}\|v\|_{x,\infty}\|z\|_{x,\infty}.$$

Proof We can write

$$\mathrm{D}(r_{x,z})(v) = -\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v} + \mathbf{G}_x^{-1}\mathbf{W'}_{x,v}s_{x,v} - \mathbf{G}_x^{-1}\mathbf{W}_x\mathrm{S}_{x,z}s_{x,v}. \tag{94}$$

Now from Lemmas 77 and 26:

$$\left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathbf{W}_x\right\|_{\infty\to\infty} \le \left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\right\|_{\infty\to\infty}\left\|\mathbf{G}_x^{-1}\mathbf{W}_x\right\|_{\infty\to\infty} \lesssim \frac{1}{(4/p-1)^4},$$
$$\left\|\mathbf{G}_x^{-1}\mathbf{W'}_{x,v}\right\|_{\infty\to\infty} \lesssim \frac{1}{(4/p-1)^2},$$
$$\left\|\mathbf{G}_x^{-1}\mathbf{W}_x\right\|_{\infty\to\infty} \lesssim \frac{1}{(4/p-1)^2},$$

which completes the proof. ■

Now we built upon Lemma 30 to bound the second derivative of $\mathbf{W}_x$ in arbitrary directions v, z.

Lemma 79 *We have*

$$-\frac{1}{(4/p-1)^5}\|z\|_{x,\infty}\|v\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}\mathbf{W'}_{x,v}(z) \preccurlyeq \frac{1}{(4/p-1)^5}\|z\|_{x,\infty}\|v\|_{x,\infty}\mathbf{W}_x.$$

Proof Note that

$$\mathrm{D}\mathbf{W'}_{x,v}(z) = -2\mathtt{Diag}\big(\mathrm{D}\mathbf{\Lambda}_x(z)r_{x,v}\big) + -2\mathtt{Diag}\big(\mathbf{\Lambda}_x\mathrm{D}(r_{x,v})(z)\big).$$

■

But for the first term by Lemmas 77 and 26,

$$\begin{aligned}\|\mathrm{D}\mathbf{\Lambda}_x(z)r_{x,v}\|_\infty &\le \|\mathbf{W}_x^{-1}\mathbf{G}_x\mathbf{G}_x^{-1}\mathrm{D}\mathbf{\Lambda}_x(z)r_{x,z}\|_\infty\\ &\le \left\|\mathbf{W}_x^{-1}\mathbf{G}_x\right\|_{\infty\to\infty}\left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{\Lambda}_x(z)\right\|_{\infty\to\infty}\|r_{x,v}\|_\infty\\ &\lesssim \frac{1}{(4/p-1)^5}\|v\|_{x,\infty},\end{aligned}$$

and for the second term, using Lemma 30

$$\begin{aligned}\|\mathbf{W}_x^{-1}\mathbf{\Lambda}_x\mathrm{D}(r_{x,v})(z)\|_\infty &\le \left\|\mathbf{W}_x^{-1}\mathbf{\Lambda}_x\right\|_{\infty\to\infty}\|\mathrm{D}(r_{x,v})(z)\|_\infty\\ &\lesssim \frac{1}{(4/p-1)^4}\|v\|_{x,\infty}\|z\|_{x,\infty},\end{aligned}$$

which completes the proof.

Next, we hope to develop estimates for the second derivative of $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ as well. In order to do so, we need to build some additional analytical tools. In this regard, we state two important Lemmas 80 and 81.

Lemma 80 *For a symmetric matrix* S *with* $-\mathbf{W}_x \preccurlyeq \mathrm{S} \preccurlyeq \mathbf{W}_x$ *and diagonal matrices* $\mathrm{S}_1, \mathrm{S}_2$*, we have*

$$\begin{aligned}&-\mathbf{W}_x\|\mathrm{S}_1\| \preccurlyeq \mathrm{S}_1\mathrm{S} + \mathrm{S}\mathrm{S}_1 \preccurlyeq \mathbf{W}_x\|\mathrm{S}_1\|,\\ &-\mathbf{W}_x\|\mathrm{S}_1\|\|\mathrm{S}_2\| \preccurlyeq \mathrm{S}_1\mathrm{S}\mathrm{S}_2 + \mathrm{S}_2\mathrm{S}\mathrm{S}_1 \preccurlyeq \mathbf{W}_x\|\mathrm{S}_1\|\|\mathrm{S}_2\|.\end{aligned}$$

Proof Using the inequality $\mathrm{q}_1^\top\mathrm{S}\mathrm{q}_2 \le \sqrt{\mathrm{q}_1^\top\mathbf{W}_x\mathrm{q}_1}\sqrt{\mathrm{q}_2^\top\mathbf{W}_x\mathrm{q}_2}$ for vectors $\mathrm{q}_1, \mathrm{q}_2$ set as $\mathrm{q}_1 = \mathrm{S}_1\mathrm{q}$ and $\mathrm{q}_2 = \mathrm{q}$:

$$\mathrm{q}^\top\mathrm{S}_1\mathrm{S}\mathrm{q} \le \sqrt{\mathrm{q}^\top\mathrm{S}_1\mathbf{W}_x\mathrm{S}_1\mathrm{q}}\sqrt{\mathrm{q}^\top\mathbf{W}_x\mathrm{q}} \le \|\mathrm{S}_1\|\mathrm{q}^\top\mathbf{W}_x\mathrm{q}.$$

and similarly

$$\mathrm{q}^\top \mathrm{S}_1 S \mathrm{q} \lesssim \frac{1}{4/p-1}\|\mathrm{S}_1\|\mathrm{q}^\top \mathbf{W}_x \mathrm{q},$$

which completes the proof. The second estimate follows in a similar mannner. ■

The next lemma bounds the terms that are created by differentiating Hadamard combinations of the projection matrix.

Lemma 81 *For diagonal matrices* $\mathrm{S}_1, \mathrm{S}_2, \mathrm{S}_3$*(not necessarily positive),*

$$-\|\mathrm{S}_1\|\|\mathrm{S}_2\|\mathbf{P}_x^{(2)} \preccurlyeq \mathbf{P}_x\mathrm{S}_1\mathbf{P}_x \odot \mathbf{P}_x\mathrm{S}_2\mathbf{P}_x \preccurlyeq \|\mathrm{S}_1\|\|\mathrm{S}_2\|\mathbf{P}_x^{(2)},$$
$$-\|\mathrm{S}_1\|\|\mathrm{S}_2\|\|\mathrm{S}_3\|\mathbf{P}_x^{(2)} \preccurlyeq \mathbf{P}_x(\mathrm{S}_2\mathbf{P}_x\mathrm{S}_3 + \mathrm{S}_3\mathbf{P}_x\mathrm{S}_2)\mathbf{P}_x \odot \mathbf{P}_x\mathrm{S}_1\mathbf{P}_x \preccurlyeq \|\mathrm{S}_1\|\|\mathrm{S}_2\|\|\mathrm{S}_3\|\mathbf{P}_x^{(2)}.$$

Proof Consider the Cholesky decomposition of $\mathbf{P}_x$:

$$\mathbf{P}_x = \sum_{i=1}^n u_i u_i^\top,$$

Then for the first inequality, note that we can write $\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x$ as

$$\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x = \sum_{i=1}^n (u_i^\top \mathrm{S}_1 u_i) u_i u_i^\top. \tag{95}$$

Hence, for arbitrary vector ℓ:

$$\begin{aligned}
\left|\mathrm{q}^\top(\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x \odot \mathbf{P}_x\mathrm{S}_2\mathbf{P}_x)\mathrm{q}\right| &\leq \sum_{i=1}^n |u_i^\top \mathrm{S}_1 u_i||(\ell\odot u_i)^\top(\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x)(\mathrm{q}\odot u_i)| \\
&\leq \sum_{i=1}^n \|\mathrm{S}_1\||(\ell\odot u_i)^\top(\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x)(\mathrm{q}\odot u_i)| \\
&\leq \sum_i \|\mathrm{S}_1\|\|\mathrm{S}_2\|(\mathrm{q}\odot u_i)^\top\mathbf{P}_x(\mathrm{q}\odot u_i) \\
&= \|\mathrm{S}_1\|\|\mathrm{S}_2\|\mathrm{q}^\top\mathbf{P}_x^{(2)}\mathrm{q}.
\end{aligned}$$

For the second inequality, note that

$$\mathrm{q}^\top(\mathrm{S}_2\mathbf{P}_x\mathrm{S}_3 + \mathrm{S}_3\mathbf{P}_x\mathrm{S}_2)\mathrm{q} \leq 2(\mathrm{S}_2\mathrm{q})^\top\mathbf{P}_x(\mathrm{S}_3\mathrm{q}) \leq 2\|\mathrm{S}_2\mathrm{q}\|_2\|\mathrm{S}_3\mathrm{q}\|_2 \leq 2\|\mathrm{S}_2\|\|\mathrm{S}_3\|\|\mathrm{q}\|_2^2,$$

which implies

$$-\|\mathrm{S}_2\|\|\mathrm{S}_3\|\mathrm{I} \preccurlyeq \mathrm{S}_2\mathbf{P}_x\mathrm{S}_3 + \mathrm{S}_3\mathbf{P}_x\mathrm{S}_2 \preccurlyeq \|\mathrm{S}_2\|\|\mathrm{S}_3\|\mathrm{I}.$$

Therefore

$$-\|\mathrm{S}_2\|\|\mathrm{S}_3\|\mathbf{P}_x \preccurlyeq \mathbf{P}_x(\mathrm{S}_2\mathbf{P}_x\mathrm{S}_3 + \mathrm{S}_3\mathbf{P}_x\mathrm{S}_2)\mathbf{P}_x \preccurlyeq \|\mathrm{S}_2\|\|\mathrm{S}_3\|\mathbf{P}_x.$$

Now again using Equation (95):

$$\begin{aligned}
&\left|\mathrm{q}^\top(\mathbf{P}_x(\mathrm{S}_2\mathbf{P}_x\mathrm{S}_3 + \mathrm{S}_3\mathbf{P}_x\mathrm{S}_2)\mathbf{P}_x) \odot (\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x)\mathrm{q}\right| \\
&\leq \sum_{i=1}^n \left|u_i^\top \mathrm{S}_1 u_i\right| |(\mathrm{q}\odot u_i)^\top(\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x\mathrm{S}_3\mathbf{P}_x + \mathbf{P}_x\mathrm{S}_3\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x)(\mathrm{q}\odot u_i)| \\
&\leq \|\mathrm{S}_1\|\mathrm{S}_2\|\|\mathrm{S}_3\|(\mathrm{q}\odot u_i)^\top\mathbf{P}_x(\mathrm{q}\odot u_i) \\
&= \|\mathrm{S}_1\|\|\mathrm{S}_2\|\|\mathrm{S}_3\|\mathrm{q}^\top\mathbf{P}_x^{(2)}\mathrm{q}.
\end{aligned}$$

■

Next, we move on to bound the second derivative of the Hadamard square of the projection matrix.

Lemma 82 *For arbitrary v, z.*

$$-\frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^2\mathbf{P}_x^{(2)}(v,z) \preccurlyeq \frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x. \tag{96}$$

Proof Expanding the derivative of $\mathbf{P}_x^{(2)}$ based on Lemma 28:

$$\mathrm{D}^2\mathbf{P}_x^{(2)}(v,z) = 2\mathbf{P}_x \odot \mathrm{D}^2\mathbf{P}_x(v,z) + 2\mathrm{D}\mathbf{P}_x(v)\odot\mathrm{D}\mathbf{P}_x(z). \tag{97}$$

For the second term, by Lemma 81

$$\begin{aligned}
\mathrm{D}\mathbf{P}_x(v)\odot\mathrm{D}\mathbf{P}_x(z) &= \mathrm{R}_{x,z}\mathbf{P}_x^{(2)}\mathrm{R}_{x,v} + \mathrm{R}_{x,v}\mathbf{P}_x^{(2)}\mathrm{R}_{x,z} + \mathrm{R}_{x,v}\mathrm{R}_{x,z}\mathbf{P}_x^{(2)} + \mathbf{P}_x^{(2)}\mathrm{R}_{x,v}\mathrm{R}_{x,z} \\
&- 4\mathrm{R}_{x,z}(\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\odot\mathbf{P}_x) - 4(\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\odot\mathbf{P}_x)\mathrm{R}_{x,z} \\
&- 4\mathrm{R}_{x,v}(\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x\odot\mathbf{P}_x) - 4(\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x\odot\mathbf{P}_x)\mathrm{R}_{x,v} \\
&+ 8\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\odot\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x.
\end{aligned} \tag{98}$$

For the first line, using Lemma 80, the fact that $\mathbf{P}_x^{(2)} \preccurlyeq \mathbf{W}_x$, and Lemma 26:

$$\begin{aligned}
\mathrm{R}_{x,z}\mathbf{P}_x^{(2)}\mathrm{R}_{x,v} + \mathrm{R}_{x,v}\mathbf{P}_x^{(2)}\mathrm{R}_{x,z} + \mathrm{R}_{x,v}\mathrm{R}_{x,z}\mathbf{P}_x^{(2)} + \mathbf{P}_x^{(2)}\mathrm{R}_{x,v}\mathrm{R}_{x,z} &\preccurlyeq \|r_{x,v}\|_\infty\|r_{x,z}\|_\infty\mathbf{W}_x \\
&\preccurlyeq \frac{1}{(4/p-1)^2}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x.
\end{aligned} \tag{99}$$

For the second line, Note that

$$\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x \preccurlyeq \|r_{x,v}\|_\infty\mathbf{P}_x. \tag{100}$$

Hence, by Schur product theorem

$$\mathbf{P}_x\odot(\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x) \preccurlyeq \frac{1}{4/p-1}\|v\|_{x,\infty}\mathbf{P}_x^{(2)}.$$

Therefore, by Lemma 80

$$\mathrm{R}_{x,z}(\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\odot\mathbf{P}_x) + (\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\odot\mathbf{P}_x)\mathrm{R}_{x,z} \preccurlyeq \frac{1}{(4/p-1)^2}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x.$$

The third line is symmetric to the second line. For the forth line, using Lemma 81:

$$\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\odot\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x \preccurlyeq \frac{1}{(4/p-1)^2}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x. \tag{101}$$

Combining Equations (99), (100), and (101) completes the proof for RHS of Equation (96). The left hand side follows similarly. Next, we have on to bound the first term in Equation (97):

$$\begin{aligned}\mathbf{P}_x \odot \mathrm{D}^2\mathbf{P}_x(v,z) = &+ \mathbf{P}_x \odot (\mathrm{R}_{x,v}\mathrm{R}_{x,z}\mathbf{P}_x + \mathbf{P}_x\mathrm{R}_{x,v}\mathrm{R}_{x,z} + \mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z} + \mathrm{R}_{x,z}\mathbf{P}_x\mathrm{R}_{x,v})\\ &- 2\mathbf{P}_x \odot (4\mathbf{P}_x\mathrm{R}_{x,v}\mathrm{R}_{x,z}\mathbf{P}_x\\ &+ \mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z} + \mathrm{R}_{x,z}\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x + \mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x\mathrm{R}_{x,v} + \mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x)\\ &+ 8\mathbf{P}_x \odot (\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x + \mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x)\\ &+ 2\mathbf{P}_x \odot \mathbf{P}_x\mathrm{D}\mathrm{R}_{x,v}(z)\mathbf{P}_x. \end{aligned} \tag{102}$$

■

The first line in Equation (102) is similar to the first line in Equation (98). The second and third lines in Equation (102) are similar to the second line in Equation (98). For the forth line, note that by Lemma 80

$$\mathrm{R}_{x,z}\mathbf{P}_x\mathrm{R}_{x,v} + \mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z} \preccurlyeq \|r_{x,v}\|_\infty \|r_{x,z}\|_\infty \mathrm{I},$$

which implies

$$\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x + \mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x \preccurlyeq \frac{1}{(4/p-1)^2}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{P}_x,$$

and by Schur product theorem

$$\begin{aligned}\mathbf{P}_x \odot (\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x + \mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x) &\preccurlyeq \frac{1}{(4/p-1)^2}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{P}_x^{(2)}\\ &\preccurlyeq \frac{1}{(4/p-1)^2}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x.\end{aligned}$$

Finally for the last line in Equation (102). combining Lemma 30 with Schur product theorem

$$\mathbf{P}_x \odot \mathbf{P}_x\mathrm{D}\mathrm{R}_{x,v}(z)\mathbf{P}_x \preccurlyeq \|v\|_{x,\infty}\frac{1}{(4/p-1)^5}\|z\|_{x,\infty}\mathbf{P}_x^{(2)},$$

which completes the proof.

Lemma 83 *For matrices $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ we have*

$$\begin{aligned}-\frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^2\mathbf{G}_x(v,z) \preccurlyeq \frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x,\\ -\frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^2\mathbf{\Lambda}_x(v,z) \preccurlyeq \frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x.\end{aligned}$$

Proof It directly follows from Lemmas 82 and 79 and the fact that $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ are linear combinations of $\mathbf{W}_x$ and $\mathbf{P}_x^{(2)}$. ■

Next, we move on to bound the third order derivative of $\mathbf{W}_x$. Similar to the second derivative case, we need to start by estimating the second derivative of $r_{x,v}$. To control the infinity norms, we need to build new tools for operators that result from the higher derivatives. We start by two important observations in Lemmas 84 and 85.

Lemma 84 *For diagonal matrices $\mathrm{S}_1, \mathrm{S}_2$,*

$$\left\|\mathbf{W}_x^{-1}\big((\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x) \odot (\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x)\big)s_{x,v}\right\|_{\infty\to\infty} \le \|\mathrm{S}_1\|_\infty\|\mathrm{S}_2\|_\infty.$$

Proof Observe that the 2-norm of the ith row of the matrix $\mathbf{P}_x \mathrm{S}_1 \mathbf{P}_x$ is at most $\|\mathrm{S}_1\|\sqrt{w_i}$. This is because

$$\|\mathbf{P}_x \mathrm{S}_1 \mathbf{P}_x e_i\|^2 \le \|\mathrm{S}_1 \mathbf{P}_x e_i\|^2 = \sqrt{\sum_j \mathrm{S}_{1\,j,j}^{\,2} \mathbf{P}_{x\,i,j}^{\,2}} \le \|\mathrm{S}_1\|\sqrt{w_i}.$$

Now for arbitrary vector r,

$$\begin{aligned}
e_i^\top\big((\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x) \odot (\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x)\big)r &= \big(e_i^\top(\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x) \odot e_i^\top(\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x)\big)^\top r \\
&= (e_i^\top(\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x))^\top((\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x e_i) \odot r) \\
&= (\mathrm{S}_2\mathbf{P}_{xi,})\mathbf{P}_x((\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x e_i) \odot r \\
&\le \|\mathrm{S}_2\mathbf{P}_{xi,}\|_2\|(\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x e_i) \odot r\|_2 \\
&\le \|\mathrm{S}_2\|\|r\|_\infty\|\mathbf{P}_{xi,}\|\|(\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x e_i)\| \\
&= \|\mathrm{S}_2\|\|r\|_\infty\|\mathbf{P}_{xi,}\|\|\mathbf{P}_x(\mathrm{S}_1\mathbf{P}_{xi,})\| \\
&\le \|\mathrm{S}_2\|\|r\|_\infty\|\mathbf{P}_{xi,}\|\|\mathrm{S}_1\mathbf{P}_{xi,}\| \\
&\le \|\mathrm{S}_2\|_\infty\|r\|_\infty\|\mathbf{P}_{xi,}\|\|\mathrm{S}_1\|\|\mathbf{P}_{xi,}\| \\
&= w_i\|\mathrm{S}_2\|\|\mathrm{S}_1\|\|r\|_\infty.
\end{aligned}$$

■

Lemma 85 *For diagonal matrices* $\mathrm{S}_1, \mathrm{S}_2$

$$\|\mathbf{W}_x^{-1}(\mathbf{P}_x \odot (\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x))\|_{\infty\to\infty} \le \|\mathrm{S}_1\|\|\mathrm{S}_2\|.$$

Proof Note that by Cauchy Schwarz

$$\begin{aligned}
e_i^\top(\mathbf{P}_x \odot (\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x))r &\le \|\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x e_i\|_2\|\mathbf{P}_{xi,} \odot r\|_2 \\
&\le \|\mathrm{S}_1\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x e_i\|_2\sqrt{w_i}\|r\|_\infty \\
&\le \sqrt{w_i}\|\mathrm{S}_1\|\|\mathbf{P}_x\mathrm{S}_1\mathbf{P}_x e_i\|_2\|r\|_\infty \\
&\le \sqrt{w_i}\|\mathrm{S}_1\|\|\mathrm{S}_2\mathbf{P}_x e_i\|_2\|r\|_\infty \\
&\le \sqrt{w_i}\|\mathrm{S}_1\|\|\mathbf{P}_x e_i\|_2\|\mathrm{S}_2\|\|r\|_\infty \\
&= w_i\|\mathrm{S}_1\|\|\mathrm{S}_2\|\|r\|_\infty.
\end{aligned}$$

■

Building upon Lemmas 83 and 84, we generalize Lemma 77 to handle second order derivatives in Lemmas 86 and 87.

Lemma 86 *For arbitrary* u, z*,*

$$\left\|\mathbf{W}_x^{-1}\mathrm{D}^2(\mathbf{P}_x^{(2)})(z,u)\right\|_{\infty\to\infty} \lesssim \frac{1}{(4/p-1)^2}\|z\|_{x,\infty}\|u\|_{x,\infty}.$$

Proof Based on Equations (98) and (102), we collect the terms that appear in the second derivative of $\mathbf{P}_x^{(2)}$ (ignoring the constant behind.) To summarize our presentation here, we use the $\sum$ notation below to include all other permutations within u, v, and z for a particular term.

$$\begin{aligned}
\mathrm{D}^2(\mathbf{P}_x^{(2)})(z,u) &\to \sum_{u,z} \mathbf{P}_x \odot (\mathbf{P}_x \mathrm{R}_{x,z} \mathbf{P}_x \mathrm{R}_{x,u} \mathbf{P}_x) \\
&\to (\mathbf{P}_x \mathrm{R}_{x,z} \mathbf{P}_x) \odot (\mathbf{P}_x \mathrm{R}_{x,u} \mathbf{P}_x) \\
&\to \mathrm{R}_{x,u} \mathbf{P}_x^{(2)} \mathrm{R}_{x,z} \\
&\to \sum_{u,z} \mathrm{R}_{x,u} \mathrm{R}_{x,z} \mathbf{P}_x^{(2)} + \mathbf{P}_x^{(2)} \mathrm{R}_{x,u} \mathrm{R}_{x,z} \\
&\to \sum_{u,z} \mathrm{R}_{x,u} (\mathbf{P}_x \odot \mathbf{P}_x \mathrm{R}_{x,z} \mathbf{P}_x) + (\mathbf{P}_x \odot \mathbf{P}_x \mathrm{R}_{x,u} \mathbf{P}_x) \mathrm{R}_{x,u} \\
&\to \mathbf{P}_x \odot \mathbf{P}_x \mathrm{R}_{x,u} \mathrm{R}_{x,z} \mathbf{P}_x
\end{aligned} \tag{103}$$

To upper bound $\left\|\mathbf{W}_x^{-1}(\mathbf{P}_x \odot (\mathbf{P}_x \mathrm{R}_{x,z} \mathbf{P}_x \mathrm{R}_{x,u} \mathbf{P}_x))\right\|_{\infty\to\infty}$ regarding the first line in Equation (103), we use Lemma 85. For the second line we use Lemma 84. For the third and forth lines we use the fact that $\left\|\mathbf{W}_x^{-1}\mathbf{P}_x^{(2)}\right\|_{\infty\to\infty} \leq 1$. For the fifth and sixth lines we use Lemma 76, which completes the proof. ■

Lemma 87 *For arbitrary* u, z*,*

$$\begin{aligned}
\left\|\mathbf{W}_x^{-1}\mathrm{D}^2\mathbf{G}_x(z,u)\right\|_{\infty\to\infty} &\lesssim \frac{1}{(4/p-1)^2}\|z\|_{x,\infty}\|u\|_{x,\infty}, \\
\left\|\mathbf{W}_x^{-1}\mathrm{D}^2\mathbf{\Lambda}_x(z,u)\right\|_{\infty\to\infty} &\lesssim \frac{1}{(4/p-1)^2}\|z\|_{x,\infty}\|u\|_{x,\infty}.
\end{aligned}$$

Proof The proof follows from Lemma 86, noting the fact that both G and Λ are linear combinations of W and $P^{(2)}$. ■

Finally, building upon Lemma 87 we estimate the second derivative of $r_{x,v}$.

Lemma 88 *For arbitrary* v, z, u*,*

$$\|\mathrm{D}^2(r_{x,v})(z,u)\| \lesssim \frac{1}{(4/p-1)^6}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}.$$

Proof Recall the first derivative of $r_{x,z}$ in direction v is

$$\mathrm{D}(r_{x,v})(z) = -\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v} + \mathbf{G}_x^{-1}\mathbf{W'}_{x,z} s_{x,v} - \mathbf{G}_x^{-1}\mathbf{W}_x \mathrm{S}_{x,z} s_{x,v}. \tag{104}$$

The derivative of the first term with respect to v is

$$\begin{aligned}
\mathrm{D}(\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v})(u) = &- \mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(u)\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v} \\
&+ \mathbf{G}_x^{-1}\mathrm{D}^2\mathbf{G}_x(z,u)\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v} \\
&- \mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(u)\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v} \\
&+ \mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathrm{D}\mathbf{W}_x(u) s_{x,v} \\
&- \mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathbf{W}_x \mathrm{S}_{x,u} s_{x,v}.
\end{aligned} \tag{105}$$

Now from Lemmas 26 and 77, for the first line in Equation (105)

$$\begin{aligned}
&\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(u)\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v}\|_\infty \\
&\leq \left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(u)\right\|_{\infty\to\infty}\left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\right\|_{\infty\to\infty}\left\|\mathbf{G}_x^{-1}\mathbf{W}_x\right\|_{\infty\to\infty}\|s_{x,v}\|_\infty \\
&\lesssim \cdot\frac{1}{(4/p-1)^5}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty}.
\end{aligned}$$

For the second line, using Lemma 87,

$$\begin{aligned}
\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z,u)\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v}\|_\infty &\leq \left\|\mathbf{W}_x^{-1}\mathrm{D}^2\mathbf{G}_x(z,u)\right\|_{\infty\to\infty}\left\|\mathbf{G}_x^{-1}\mathbf{W}_x\right\|_{\infty\to\infty}\|s_{x,v}\|_\infty \\
&\lesssim \frac{1}{(4/p-1)^3}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty}.
\end{aligned}$$

The third term is similar to the first line. For the forth line by Lemmas 27,77, and 26

$$\begin{aligned}
\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\mathbf{G}_x^{-1}\mathrm{D}\mathbf{W}_x(u)s_{x,v}\|_\infty &\leq \left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x(z)\right\|_{\infty\to\infty}\left\|\mathbf{G}_x^{-1}\mathbf{W}_x\right\|_{\infty\to\infty}\left\|\mathbf{W}_x^{-1}\mathbf{W'}_{x,u}\right\|_{\infty\to\infty}\|s_{x,z}\|_\infty \\
&\leq \frac{1}{(4/p-1)^4}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty}.
\end{aligned}$$

The fifth term is handled similarly noting the fact that $\|\mathrm{S}_{x,u}s_{x,z}\|_\infty \leq \|s_{x,u}\|_\infty\|s_{x,z}\|_\infty$. The derivatives of the two other terms in Equation (104) are handled similarly; the only new term is $\mathbf{G}_x^{-1}\mathrm{D}\mathbf{W'}_{x,u}(z)s_{x,v}$ when taking derivative from the second term in Equation (104), which is handled by Lemma 79,

$$\left\|\mathbf{G}_x^{-1}\mathrm{D}\mathbf{W'}_{x,u}(z)\right\|_{\infty\to\infty} \leq \left\|\mathbf{G}_x^{-1}\mathbf{W}_x\right\|_{\infty\to\infty}\left\|\mathbf{W}_x^{-1}\mathrm{D}\mathbf{W'}_{x,u}(z)\right\|_{\infty\to\infty} \lesssim \frac{1}{(4/p-1)^6}.$$

The proof is complete. ∎

Now based on Lemma 88 we are ready to estimate the third order derivative of the Lewis weights.

Lemma 89 *We have*

$$-\frac{1}{(4/p-1)^7}\|z\|_{x,\infty}\|u\|_{x,\infty}\|v\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^2(\mathbf{W'}_{x,v})(z,u) \preccurlyeq \frac{1}{(4/p-1)^7}\|z\|_{x,\infty}\|u\|_{x,\infty}\|v\|_{x,\infty}\mathbf{W}_x.$$

Proof Calculating the derivative of $\mathbf{W'}_{x,v}$:

$$\begin{aligned}
\mathrm{D}^2\mathbf{W'}_{x,v}(u,z) &= \mathtt{Diag}\big(\mathrm{D}^2(\mathbf{\Lambda}_x r_{x,v})(u,z)\big) \\
&= \mathtt{Diag}\big(\mathrm{D}^2\mathbf{\Lambda}_x(u,z)r_{x,v}\big) \\
&+ \mathtt{Diag}\big(\mathrm{D}\mathbf{\Lambda}_x(u)\mathrm{D}(r_{x,v})(z)\big) + \mathtt{Diag}\big(\mathrm{D}\mathbf{\Lambda}_x(z)\mathrm{D}(r_{x,v})(u)\big) \\
&+ \mathtt{Diag}\big(\mathbf{\Lambda}_x\mathrm{D}^2(r_{x,v})(z,u)\big).
\end{aligned}$$

For the first line, using Lemma 77

$$\|\mathbf{W}_x^{-1}\mathrm{D}^2\mathbf{\Lambda}_x(u,z)r_{x,v}\|_\infty \leq \frac{1}{(4/p-1)^3}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}.$$

For the second line, using Lemma 30

$$\begin{aligned}
\|\mathbf{W}_x^{-1}\mathrm{D}\mathbf{\Lambda}_x(u)\mathrm{D}(r_{x,v})(z)\|_\infty &\leq \left\|\mathbf{G}_x^{-1}\mathbf{W}_x\right\|_{\infty\to\infty}\left\|\mathbf{W}_x^{-1}\mathrm{D}\mathbf{\Lambda}_x(u)\right\|_{\infty\to\infty}\|\mathrm{D}r_{x,v}(z)\|_\infty \\
&\lesssim \frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}.
\end{aligned}$$

Finally for the forth line, we use Lemma 88

$$\|\mathbf{W}_x^{-1}\mathbf{\Lambda}_x \mathrm{D}^2(r_{x,v})(z,u)\|_\infty \le \left\|\mathbf{W}_x^{-1}\mathbf{\Lambda}_x\right\|_{\infty\to\infty}\|\mathrm{D}^2(r_{x,v})(z,u)\|_\infty$$
$$\lesssim \frac{1}{(4/p-1)^6}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty},$$

where we used the fact that $\mathbf{\Lambda}_x$ is a linear combination of $\mathbf{W}_x$ and $\mathbf{P}_x^{(2)}$ to bound $\left\|\mathbf{W}_x^{-1}\mathbf{\Lambda}_x\right\|_{\infty\to\infty}$. This completes the proof. ■

Finally, building upon all of the tools we developed, we drive estimates for the third derivatives of $\mathbf{\Lambda}_x$ and $\mathbf{G}_x$. The key is Lemma 89 combined with the following key Lemma in which we control the third derivative of $\mathbf{P}_x^{(2)}$.

Lemma 90 *For arbitrary v, u, z,*

$$-\frac{1}{(4/p-1)^3}\|u\|_\infty\|v\|_\infty\|z\|_\infty \mathbf{W}_x \preccurlyeq \mathrm{D}^3\mathbf{P}_x^{(2)}(u,v,z) \preccurlyeq \frac{1}{(4/p-1)^3}\|u\|_\infty\|v\|_\infty\|z\|_\infty \mathbf{W}_x.$$

Proof Similar to the proof of Lemma 86 in Equation (103), we collect all possible terms when differentiating three times from $\mathbf{P}_x^{(2)}$, ignoring the constants.

$$\begin{aligned}
&\mathrm{D}^3\mathbf{P}_x^{(2)}(u,v,z)\\
&= \mathbf{P}_x \odot \Big(\sum \mathbf{P}_x \mathrm{R}_{x,u}\mathrm{R}_{x,z}\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x + \sum \mathbf{P}_x\mathrm{R}_{x,u}\mathbf{P}_x\mathrm{R}_{x,v}\mathrm{R}_{x,z}\mathbf{P}_x\\
&\quad + \sum \mathrm{R}_{x,u}\mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}\\
&\quad + \sum \mathrm{R}_{x,u}\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x + \sum \mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x\mathrm{R}_{x,u}\mathrm{R}_{x,v}\\
&\quad + \sum \mathrm{R}_{x,u}\mathbf{P}_x\mathrm{R}_{x,v}\mathrm{R}_{x,z}\mathbf{P}_x + \sum \mathbf{P}_x\mathrm{R}_{x,u}\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}\\
&\quad + \sum \mathbf{P}_x\mathrm{R}_{x,u}\mathrm{R}_{x,v}\mathrm{R}_{x,z} + \sum \mathrm{R}_{x,u}\mathrm{R}_{x,v}\mathrm{R}_{x,z}\mathbf{P}_x\Big)\\
&+ \sum(\mathbf{P}_x\mathrm{R}_{x,u}\mathbf{P}_x + \mathrm{R}_{x,u}\mathbf{P}_x + \mathbf{P}_x\mathrm{R}_{x,u}) \odot \mathbf{P}_x\mathrm{R}_{x,v}\mathrm{R}_{x,z}\mathbf{P}_x\\
&+ \sum(\mathbf{P}_x\mathrm{R}_{x,u}\mathbf{P}_x + \mathrm{R}_{x,u}\mathbf{P}_x + \mathbf{P}_x\mathrm{R}_{x,u}) \odot \mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x\\
&+ \sum(\mathbf{P}_x\mathrm{R}_{x,u}\mathbf{P}_x + \mathrm{R}_{x,u}\mathbf{P}_x + \mathbf{P}_x\mathrm{R}_{x,u}) \odot (\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}\mathbf{P}_x + \mathbf{P}_x\mathrm{R}_{x,v}\mathbf{P}_x\mathrm{R}_{x,z}).
\end{aligned} \tag{106}$$

The first four line can be upper bounded by $\frac{1}{(4/p-1)^3}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty}\mathbf{P}_x^{(2)}$ using Lemma 81, and the fifth line has the same upper bound from Lemma 80. The rest of the lines are handled similarly by Lemma 81. The point is, all the terms in Equation (106) are of the form

$$\mathrm{S}_1(\mathbf{P}_x\mathrm{S}_2\mathbf{P}_x) \odot (\mathbf{P}_x\mathrm{S}_3\mathbf{P}_x)\mathrm{S}_4,$$

for diagonal matrices S_1 and S_4, such that

$$\|\mathrm{S}_1\|\|\mathrm{S}_2\|\|\mathrm{S}_3\|\|\mathrm{S}_4\| \le \|r_{x,v}\|_\infty\|r_{x,u}\|_\infty\|r_{x,z}\|_\infty.$$

Hence, combining Lemmas 80 and 81, we get

$$-\|r_{x,u}\|_\infty\|r_{x,v}\|_\infty\|r_{x,z}\|_\infty\mathbf{W}_x \preccurlyeq \mathrm{D}^3\mathbf{P}_x^{(2)}(u,v,z) \preccurlyeq \|r_{x,u}\|_\infty\|r_{x,v}\|_\infty\|r_{x,z}\|_\infty\mathbf{W}_x,$$

which completes the proof by the fact that $\|r_{x,v}\|_\infty \le \frac{1}{(4/p-1)}\|v\|_{x,\infty}$ by Lemma 26. ■

Finally based on Lemmas 90 and 89, the estimates on the third derivatives of $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ follows.

Lemma 91 *We have*

$$-\frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^3\mathbf{G}_x(u,v,z) \preccurlyeq \frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x,$$
$$-\frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x \preccurlyeq \mathrm{D}^3\mathbf{\Lambda}_x(u,v,z) \preccurlyeq \frac{1}{(4/p-1)^7}\|u\|_{x,\infty}\|v\|_{x,\infty}\|z\|_{x,\infty}\mathbf{W}_x.$$

Proof The proof follows from the fact that $\mathbf{G}_x$ and $\mathbf{\Lambda}_x$ are a linear combination of $\mathbf{W}_x$ and $\mathbf{P}_x^{(2)}$ and third order estimates on $\mathbf{W}_x$ and $\mathbf{P}_x^{(2)}$ in Lemmas 89 and 90, respectively. ■

Appendix G. Isoperimetry of the Hybrid Barrier

In this section, we prove the bound on the isoperimetry constant for our hybrid barrier stated in Theorem 6. We prove this isoperimetry by efficiently bounding the barrier parameter for our Hybrid Barrier. In a high level, we exploit the fact that while the log barrier and the Lewis weight barrier for $p < 4$ are both smooth, the unit ellipsoids of their local norms are complementary to each other in approximating the symmetrized polytope, in a way that adding them up improves the overall quality of this approximation. In Figure 3 we have visualized our hybrid barrier and the Lewis weights barrier in approximating the symmetrized polytope. We start from the definition of barrier parameter borrowed from Laddha and Vempala (2021).

Definition 92 (Barrier parameter) *A convex barrier function ϕ for polytope $\mathcal{P}$ has symmetric self-concordance parameter (barrier parameter) $\bar{\nu}$ if at each point $x \in \mathcal{P}$ inside the polytope the Dikin ellipsoid of ϕ at x, namely $E_x \triangleq \{v \in \mathbb{R}^n | \, v^\top \nabla^2 \phi(x) v \leq 1\}$ satisfies $E_x \subseteq \mathcal{P} \cap 2x - \mathcal{P} \subseteq \sqrt{\bar{\nu}} E_x$.*

As we mentioned in Section 1.3, it is easy to check that the symmetrized polytope $\mathcal{P} \cap 2x - \mathcal{P}$ is the set of points within radius one in the $\|.\|_{x,\infty}$ norm around x, i.e.

$$\mathcal{P} \cap 2x - \mathcal{P} = \{x + v | \, \|v\|_{x,\infty} \leq 1\}.$$

Therefore, $E_x \subseteq \mathcal{P} \cap 2x - \mathcal{P}$ is equivalent to having $\|v\|_{x,\infty} \leq \|v\|_g$ for all v. Next, we provide some intuition on how we show this inequality.

For $a_{x,i}$ the ith row of A_x, note that if we have a bound on the quantity $\|a_{x,i}\|^2_{g^{-1}} = a_{x,i}^\top g^{-1} a_{x,i}$ for our metric $g = \nabla^2 \phi$ and all i, it enables us to control the infinity norm of $s_{x,v}$, which is equal to $\|v\|_{x,\infty}$, via the following simple Cauchy Schwarz on the ith entry of $s_{x,v}$:

$$|{s_{x,v}}_i| = |a_{x,i}^\top v| \leq \|v\|_g \|a_{x,i}\|_{g^{-1}}.$$

Recalling the definition of $g = g_1 + g_2$ as the sum of the Lewis Weights metric g_1 and the log barrier metric g_2, one might hope to upper bound $\|a_{x,i}\|^2_{g^{-1}}$ as

$$a_{x,i}^\top g^{-1} a_{x,i} \leq a_{x,i}^\top g_1^{-1} a_{x,i} \leq a_{x,i}^\top (\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x)^{-1} a_{x,i}, \tag{107}$$

where the last inequality use the fact that g_1 can approximated by $\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x$ as $\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x \preccurlyeq g_1 \preccurlyeq p\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x$, (Lemma 18.) On the other hand, according to the fixed point property of Lewis weights (see Lemma 116), we have

$$a_{x,i}^\top (\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p}) \mathrm{A}_x)^{-1} a_{x,i} = {w_x}_i^{2/p} \leq 1, \tag{108}$$

Comparing the left-hand side of Equation (108) and the right-hand side in Equation (107), one might hope to estimate $\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x$ by $\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p} \mathrm{A}_x$ up to log factors in order to upper bound $a_{x,i}^\top g^{-1} a_{x,i}$. Unfortunately such estimates are possible Lee and Sidford (2019) between $\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x$ and $\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p} \mathrm{A}_x$ for poly-logarithmically large p's, but here we cannot pick $p \geq 4$ as our infinity-norm estimates break. As a result, if we only consider the Hessian of the Lewis weights barrier g_1, the quantity $a_{x,i}^\top {g_1}^{-1} a_{x,i}$ might be orders of magnitude larger than its counterpart $a_{x,i}^\top (\mathrm{A}_x \mathbf{W}_x^{1-2/p}) \mathrm{A}_x)^{-1} a_{x,i}$ in Equation (108).

Nonetheless, we show that adding the log barrier to the Lewis weight barrier and appropriately rescaling the metric g indeed enables us to bound $a_{x,i}^\top g^{-1} a_{x,i}$.

To prove the desired bound on $a_{x,i}^\top g^{-1} a_{x,i}$, we start by comparing $\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p} \mathrm{A}_x$ with the matrix $\tilde{g} \triangleq \mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x + \frac{n}{m} \mathrm{A}_x^\top \mathrm{A}_x$, which is proportional to the Hessian of the hybrid barrier before scaling by α_0. This estimate enables us to analyze the quantity $a_{x,i}^\top g^{-1} a_{x,i}$ via Equation (108). We state this estimate in Lemma 93

Lemma 93 (Löwner comparison with different weighted matrices) *For the PSD matrix* $\tilde{g} \triangleq \mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x + \frac{n}{m}\mathrm{A}_x^\top \mathrm{A}_x$*, we have*

$$\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p} \mathrm{A}_x \preccurlyeq (\frac{m}{n})^{2/p} \tilde{g}.$$

Proof First, note that that for having the inequality $\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p} \mathrm{A}_x \preccurlyeq \beta\Big(\mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x + \frac{n}{m}\mathrm{A}_x^\top \mathrm{A}_x\Big)$ for a positive real β, it is enough to guarantee

$$\mathbf{W}_x^{1-2/p} \preccurlyeq \beta(\mathbf{W}_x + \mathrm{I}).$$

which is equivalent to picking β such that for all indices $1 \leq i \leq m$,

$$w_{xi}{}^{1-2/p} \leq \beta(w_{xi} + \frac{n}{m}). \tag{109}$$

The first thing we notice is that if $w_{x_i}^{1-2/p} \leq \beta\frac{n}{m}$, then the inequality is already satisfied. Hence, without loss of generality we assume

$$w_{xi} \geq (\beta\frac{n}{m})^{\frac{1}{1-2/p}}. \tag{110}$$

in this regime of w_i to pick a β which satisfies Equation (109), we need to have

$$\beta w_{xi}{}^{2/p} \geq 1.$$

But using Equation (110), it is sufficient to have

$$(\beta\frac{n}{m})^{\frac{2/p}{1-2/p}}\beta \geq 1,$$

which is satisfied if we pick β as large as

$$\beta = (\frac{m}{n})^{2/p}.$$

This completes the proof. ∎

Now based on Lemma 93 we upper bound $a_{x,i}^\top g^{-1} a_{x,i}$ for the hybrid barrier.

Lemma 94 (Taming the hybrid metric) *For the metric of our hybrid barrier* $g = \nabla^2\phi$*, we have for every* i*:*

$$a_{x,i}^\top g^{-1} a_{x,i} \leq 1.$$

Proof Note that using Lemma 18, we have

$$\tilde{g} = \mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x + \frac{n}{m}\mathrm{A}_x^\top \mathrm{A}_x \preccurlyeq g_1 + \frac{n}{m}\mathrm{A}_x^\top \mathrm{A}_x = \frac{1}{\alpha_0} g,$$

where recall that $\alpha_0 = (\frac{m}{n})^{\frac{2}{2+p}}$ is the normalizing factor of the hybrid barrier (Equation (4).) Hence, using Lemma 93:

$$\alpha_0 a_i^\top g^{-1} a_i \leq a_i^\top \tilde{g}^{-1} a_i \leq (\frac{m}{n})^{2/p} a_i^\top (\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p} A)^{-1} a_i \leq (\frac{m}{n})^{2/p} w_{xi}{}^{2/p}. \tag{111}$$

On the other hand, by the definition of $\tilde{g}$ and the fact that the leverage scores of any matrix are less than one:

$$\alpha_0 a_i^\top g^{-1} a_i \leq a_i^\top \tilde{g}^{-1} a_i \leq a_i^\top (\mathrm{A}_x^\top \mathbf{W}_x A)^{-1} a_i = {w_x}_i^{-1} {w_x}_i^{1/2} a_i^\top (A^\top \mathbf{W}_x \mathrm{A}_x)^{-1} a_i {w_x}_i^{1/2} \leq {w_i}^{-1}. \quad (112)$$

Balancing Equations (111) and (112) implies

$$\alpha_0 a_i^\top g^{-1} a_i \leq (\frac{m}{n})^{\frac{2/p}{1+2/p}}. \quad (113)$$

which completes the proof as $\alpha_0 = (\frac{m}{n})^{\frac{2/p}{1+2/p}}$. ■

Finally, using the result of Lemma 94 we estimate $\|.\|_{x,\infty}$ by $\|.\|_g$.

Lemma 95 (Bounding infinity norm by the ellipsoidal norm) *Given an arbitrary vector $z \in \mathbb{R}^n$,*

$$\|v\|_{x,\infty} \leq \|v\|_g, \quad (114)$$
$$\|v\|_g \leq \sqrt{\alpha_0 n(p+1)}\|v\|_{x,\infty}. \quad (115)$$

Proof First, we show Equation (114). For arbitrary $1 \leq i \leq m$, we have using Lemma 94 and Cauchy Schwarz:

$$|a_i^\top v| \leq \sqrt{a_i^\top g^{-1} a_i}\sqrt{v^\top g v} \leq \|v\|_g,$$

which means

$$\|v\|_{x,\infty} = \|s_{x,v}\|_\infty \leq \|v\|_g.$$

To show Equation (115), note that using Lemma 18, we have

$$v^\top g_1 v \leq p \sum_i {w_x}_i {s_{x,v}}_i^2 \leq p \sum_i {w_x}_i \|s_{x,v}\|_\infty^2 = np\|v\|_{x,\infty}^2,$$

and

$$v^\top g_2 v = \frac{n}{m} \sum_i {s_{x,v}}_i^2 \leq n\|s_{x,v}\|_\infty^2 = n\|v\|_{x,\infty}^2.$$

Noting the definition of $g = \alpha_0(g_1 + g_2)$ in Equation (15), the proof is completes. ■

Finally using the estimates in Lemma 95 we prove Theorem 6. To this end
Proof [Proof of Theorem 6.] From the estimate $\|v\|_g \leq \sqrt{\alpha_0 n(1+p)}\|v\|_{x,\infty}$ in Lemma 95 we see that if we scale the ellipsoid $\{v|\ v^\top g(x) v \leq 1\}$ by $\sqrt{pn}(\frac{m}{n})^{\frac{1/p}{2/p+1}}$ then it includes the symmetrized polytope around x, whose unit ball is exactly $\{v|\ \|v\|_{x,\infty} \leq 1\}$, i.e.

$$\{v|\ \|v\|_{x,\infty} \leq 1\} \subseteq \{\sqrt{pn}(\frac{m}{n})^{\frac{1/p}{2/p+1}} v|\ v^\top g(x) v \leq 1\}. \quad (116)$$

On the other hand, from Lemma 95 we have

$$\|v\|_{x,\infty} \leq \|v\|_g,$$

which implies that the unit ball of the norm, or the Dikin ellipsoid, is contained in the symmetrized polytope around x, i.e.

$$\{v|\ v^\top g(x)v \leq 1\} \subseteq \{v|\ \|v\|_{x,\infty} \leq 1\}. \tag{117}$$

Combining Equations (116) and (117) implies that the symmetric self-concordance parameter $\bar{\nu}$ defined in Laddha and Vempala (2021) is at most $\bar{\nu} \leq pn(\frac{m}{n})^{\frac{2/p}{2/p+1}}$, which in turn implies that the distribution $e^{-\alpha\phi}$ has isoperimetry with constant at least

$$\frac{1}{\sqrt{\bar{\nu}}} \geq \frac{1}{\sqrt{pn}}(\frac{n}{m})^{\frac{1/p}{2/p+1}}, \tag{118}$$

with respect to metric g as desired.

Furthermore, from the normal c_2-self-concordance of ϕ by Theorem 8 (recall $c_2 = \text{poly}(\frac{1}{4/p-1})$) and the fact that $\alpha\phi$ is α-relatively strongly convex with respect to ϕ, then Lemma 8 in Gopi et al. (2023) implies an isoperimetric inequality with constant $\text{poly}(\frac{1}{4/p-1})\sqrt{\alpha}$ for the measure $e^{-\alpha\phi}dx$ on the manifold with metric of the Hessian of ϕ. Combining this with the first isoperimetric constant in Equation (118) completes the proof. ■

Appendix H. Riemannian Geometry

H.1. Basic Manifold Definitions

In this section, we go through some basic definitions in differential geometry that are essential to know in our proofs. A manifold is defined abstractly as a topological space which locally resembles $\mathbb{R}^n$.

Definition 96 *A manifold $\mathcal{M}$ is a topological space such that for each point $p \in \mathcal{M}$, there exists an open set U around p such that U is a homeomorphism to an open set of $\mathbb{R}^n$.*

Tangent Space. *For any point $p \in \mathcal{M}$, one can define the notion of tangent space for p, $T_p(\mathcal{M})$, as the equivalence class of the set of curves γ starting from p ($\gamma(0) = p$), where we define two such curves γ_0 and γ_1 to be equivalent if for any function f on the manifold:*

$$\frac{d}{dt} f(\gamma_0(t))\big|_{t=0} = \frac{d}{dt} f(\gamma_1(t))\big|_{t=0}.$$

On can define a linear structure on $T_p(\mathcal{M})$, hence it is a vector space. Now given a positive definite quadratic form $g(p)$ on the vector space $T_p(\mathcal{M})$, one can equip the manifold $\mathcal{M}$ with metric g. While the definition of a general manifold is abstract, putting a metric on it allows us to measure length, areas, volumes, etc. on the manifold, and do calculus similar to Euclidean space. Next, we define some basic notions regarding manifolds.

Differential. For a map $f : \mathcal{M} \to \mathcal{N}$ between two manifolds, the differential df_p at some point $p \in \mathcal{M}$ is a linear map from $T_p(\mathcal{M})$ to $T_{f(p)}(\mathcal{N})$ with the property that for any curve $\gamma(t)$ on $\mathcal{M}$ with $\gamma(0) = p$, we have

$$df(\frac{d}{dt}\gamma(0)) = \frac{d}{dt} f(\gamma)(0). \tag{119}$$

. As a special case, for a function f over the manifold, the differential df at some point $p \in \mathcal{M}$ is a linear functional over $T_p(\mathcal{M})$, i.e. an element of $T_p^*(\mathcal{M})$. Writing (119) for curve γ_i with $\frac{d}{dt}\gamma_i(0) = \partial x_i$, testing property (119), we see

$$df(\partial x_i) = \frac{d}{dt} f(\gamma_i(t))\Big|_{t=0} = \frac{\partial f}{\partial x_i}(\gamma_i(0)).$$

We can write $df = \sum_i \frac{\partial f}{\partial x_i} dx_i$.

Vector field. A vector field V is a smooth choice of a vector $V(p) \in T_p(\mathcal{M})$ in the tangent space for all $p \in \mathcal{M}$.

Metric and inner product. A metric is a tensor on the manifold $\mathcal{M}$ which is simply a smooth choice of a symmetric bilinear map over $\mathcal{M}$. Alternatively, the metric or dot product $\langle , \rangle$ can be seen as a bilinear map over the space of vector fields with the tensorization property, i.e. for vector fields V, W, Z and scalar functions α, β over $\mathcal{M}$:

$$\langle V + W, Z\rangle = \langle V, Z\rangle + \langle W, Z\rangle, \tag{120}$$

$$\langle \alpha V, \beta W\rangle = \alpha\beta\langle V, W\rangle. \tag{121}$$

H.2. Manifold Derivatives, Geodesics, Parallel Transport

H.2.1. COVARIANT DERIVATIVE

Given two vector fields V and W, the covariant derivative, also called the Levi-Civita connection $\nabla_V W$ is a bilinear operator with the following properties:

$$\begin{aligned}\nabla_{\alpha_1 V_1+\alpha_2 V_2} W &= \alpha_1 \nabla_{V_1} W + \alpha_2 \nabla_{V_2} W,\\ \nabla_V(W_1+W_2) &= \nabla_V(W_1) + \nabla_V(W_2),\\ \nabla_V(\alpha W_1) &= \alpha \nabla_V(W_1) + V(\alpha) W_1\end{aligned}$$

where $V(\alpha)$ is the action of vector field V on scalar function α. Importantly, the property that differentiates the covariant derivative from other kinds of derivaties over manifold is that the covariant derivative of the metric is zero, i.e., $\nabla_V g = 0$ for any vector field V. In other words, we have the following intuitive rule:

$$\nabla_V \langle W_1, W_2\rangle = \langle \nabla_V W_1, W_2\rangle + \langle W_1, \nabla_V W_2\rangle.$$

Moreover, the covariant derivative has the property of being torsion free, meaning that for vector fields W_1, W_2:

$$\nabla_{W_1} W_2 - \nabla_{W_2} W_1 = [W_1, W_2],$$

where $[W_1, W_2]$ is the Lie bracket of W_1, W_2 defined as the unique vector field that satisfies

$$[W_1, W_2] f = W_1(W_2(f)) - W_2(W_1(f))$$

for every smooth function f.

In a local chart with variable x, if one represent $V = \sum V^i \partial x_i$, where ∂x_i are the basis vector fields, and $W = \sum W^i \partial x_i$, the covariant derivative is given by

$$\begin{aligned}\nabla_V W &= \sum_i V^i \nabla_i W = \sum_i V^i \sum_j \nabla_i (W^j \partial x_j)\\ &= \sum_i V^i \sum_j \partial_i (W^j) \partial x_j) + \sum_i V^i \sum_j W^j \nabla_i \partial x_j)\\ &= \sum_j V(W^j) \partial x_j) + \sum_i \sum_j V^i W^j \sum_k \Gamma_{ij}^k \partial x_k) =\\ &= \sum_k \big(V(W^k) + \sum_i \sum_j V^i W^j \Gamma_{ij}^k\big) \partial x_k.\end{aligned}$$

The Christoffel symbols Γ_{ij}^k are the representations of the Levi-Cevita derivatives of the basis $\{\partial x_i\}$:

$$\nabla_{\partial x_j} \partial x_i = \sum_k \Gamma_{ij}^k \partial x_k$$

and are given by the following formula:

$$\Gamma_{ij}^k = \frac{1}{2} \sum_m g^{km} (\partial_j g_{mi} + \partial_i g_{mj} - \partial_m g_{ij}).$$

Above, g^{ij} refers to the (i, j) entry of the inverse of the metric. In the following Lemma, we calculate the Christoffel symbols on a Hessian manifold and $g = D^2 \phi$ is the Hessian of a convex function.

Lemma 97 *On a Hessian manifold with metric g we have*

$$\Gamma_{ij}^k = \frac{1}{2} \sum_m g^{km} D g_{mij}.$$

Proof Since the manifold is Hessian, we have

$$\partial_j g_{mi} = \partial_i g_{jm} = \partial_m g_{ij} = D g_{ijm},$$

where Dg_{ijm} is just the notation that we use for Hessian manifolds.

■

H.2.2. PARALLEL TRANSPORT

The notion of parallel transport of a vector V along a curve γ can be generalized from Euclidean space to a manifold. On a manifold, parallel transport is a vector field restricted to γ such that $\nabla_{\gamma'}(V) = 0$. By this definition, for two parallel transport vector fields $V(t), W(t)$ we have that their dot product $\langle V(t), W(t) \rangle$ is preserved, i.e., $\frac{d}{dt}\langle V(t), W(t) \rangle = 0$.

H.2.3. GEODESIC

A geodesic is a curve γ on $\mathcal{M}$ is a "locally shortest path", i.e., the tangent to the curve is parallel transported along the curve: $\nabla_{\dot\gamma}\dot\gamma = 0$ ($\dot\gamma$ denotes the time derivative of the curve γ.) Writing this in a chart, one can see it is a second order nonlinear ODE which locally has a unique solution given initial location and speed.

$$\frac{d^2\gamma_k}{dt^2}(t) = -\frac{1}{2} \sum_{i,j} \frac{d\gamma_i}{dt} \frac{d\gamma_j}{dt} \Gamma_{ij}^k, \quad \forall k. \tag{122}$$

H.2.4. RIEMANN TENSOR

The Riemann tensor is a particular tensor on the manifold which arise from the covariant derivative. Namely, it is a linear mapping from $T_p(\mathcal{M}) \times T_p(\mathcal{M}) \times T_p(\mathcal{M}) \to T_p(\mathcal{M})$ defined as

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

The Riemann tensor can be calculated in a chart given the following formula:

$$R_{jkl}^i = \frac{\partial \Gamma_{lj}^i}{\partial x_k} - \frac{\partial \Gamma_{kj}^i}{\partial x_l} + \sum_p (\Gamma_{kp}^i \Gamma_{lj}^p - \Gamma_{lp}^i \Gamma_{kj}^p). \tag{123}$$

In the following Lemma, we calculate the Riemann tensor on a Hessian manifold:

Lemma 98 *The Riemann tensor is given by*

$$R_{jkl}^i = -\frac{1}{4} g^{i,\ell} \mathrm{D} g_{\ell,k,p} g^{p,\ell_2} \mathrm{D} g_{\ell_2,l,j} + \frac{1}{4} g^{i,\ell} \mathrm{D} g_{\ell,l,p} g^{p,\ell_2} \mathrm{D} g_{\ell_2,k,j}.$$

Proof We consider the terms in Equation (123) one by one. For the first term

$$\begin{aligned}\frac{\partial \Gamma_{lj}^i}{\partial x_k} &= \partial_{x_k}(\frac{1}{2}\sum_{\ell_2} g^{i\ell_2}\mathrm{D}g_{\ell_2 lj}) \\ &= \frac{1}{2}\sum_{\ell_2}\partial_{x_k}(g^{i\ell_2})\mathrm{D}g_{\ell_2 lj} + \frac{1}{2}\sum_{\ell_2} g^{i\ell_2}\partial_{x_k}(\mathrm{D}g_{\ell_2 lj}) \\ &= -\frac{1}{2}\sum_{\ell_2} g^{i\ell}\mathrm{D}g_{\ell kp}g^{p\ell_2}\mathrm{D}g_{\ell_2 lj} + \frac{1}{2}\sum_{\ell_2} g^{i\ell_2}D^2 g_{k\ell_2 lj}.\end{aligned}$$

Similarly

$$\frac{\partial \Gamma_{kj}^i}{\partial x_l} = -\frac{1}{2}\sum_{\ell_2} g^{i\ell}\mathrm{D}g_{\ell lp}g^{p\ell_2}\mathrm{D}g_{\ell_2 kj} + \frac{1}{2}\sum_{\ell_2} g^{i\ell_2}D^2 g_{l\ell_2 kj}.$$

Hence

$$\frac{\partial \Gamma_{lj}^i}{\partial x_k} - \frac{\partial \Gamma_{kj}^i}{\partial x_l} = -\frac{1}{2}\sum_{\ell_2} g^{i\ell}\mathrm{D}g_{\ell kp}g^{p\ell_2}\mathrm{D}g_{\ell_2 lj} + \frac{1}{2}\sum_{\ell_2} g^{i\ell}\mathrm{D}g_{\ell lp}g^{p\ell_2}\mathrm{D}g_{\ell_2 kj}. \tag{124}$$

For the third and forth terms

$$\sum_p \Gamma_{kp}^i\Gamma_{lj}^p = \frac{1}{4}\sum_{\ell_2} g^{i\ell}\mathrm{D}g_{\ell kp}g^{p\ell_2}\mathrm{D}g_{\ell_2 lj}, \tag{125}$$

$$\sum_p \Gamma_{lp}^i\Gamma_{kj}^p = \frac{1}{4}\sum_{\ell_2} g^{i\ell}\mathrm{D}g_{\ell lp}g^{p\ell_2}\mathrm{D}g_{\ell_2 kj}. \tag{126}$$

Combining Equations (124) and (126) and plugging into (123) completes the proof. ■

H.2.5. RICCI TENSOR

The Ricci tensor is just the trace of the Riemann tensor with respect to the second and third components or first and forth components, i.e. the trace of the operator $R(.,X)Y$:

$$\texttt{Ricci}(X,Y) = \texttt{tr}(R(.,X)Y).$$

Equivalently, if $\{e_i\}$ is an orthogonal basis in the tangent space, we have

$$\texttt{Ricci}(X,Y) = \sum_i \langle Y, R(X,e_i)e_i\rangle. \tag{127}$$

Lemma 99 (Form of the Ricci tensor on Hessian manifolds) *On a Hessian manifold, the Ricci tensor is given by*

$$\texttt{Ricci}(v_1,v_2) = -\frac{1}{4}\texttt{tr}(g^{-1}\mathrm{D}g(v_1)g^{-1}\mathrm{D}g(v_2)) + \frac{1}{4}v_1^\top \mathrm{D}g(g^{-1}\texttt{tr}(g^{-1}\mathrm{D}g))v_2.$$

Proof Using the form of Riemann tensor in (123) and the definition of Ricci tensor in (127)

$$\begin{aligned}\texttt{Ricci}(\partial_j, \partial_k) &= \sum_{i=l=1}^{n} \big(\frac{\partial \Gamma^i_{lj}}{\partial x_k} - \frac{\partial \Gamma^i_{kj}}{\partial x_l} + \sum_p (\Gamma^i_{kp}\Gamma^p_{lj} - \Gamma^i_{lp}\Gamma^p_{kj})\big)\\ &= \sum_{i=l=1}^{n} -\frac{1}{4} g^{i,\ell} \mathrm{D} g_{\ell,k,p} g^{p,\ell_2} \mathrm{D} g_{\ell_2,l,j} + \frac{1}{4} g^{i,\ell} \mathrm{D} g_{\ell,l,p} g^{p,\ell_2} \mathrm{D} g_{\ell_2,k,j}\\ &= -\frac{1}{4}\texttt{tr}(g^{-1}\mathrm{D}g_k g^{-1}\mathrm{D}g_j) + \frac{1}{4} e_j^\top \mathrm{D}g(g^{-1}\texttt{tr}(g^{-1}\mathrm{D}g))e_k.\end{aligned}$$

Therefore, for arbitrary vector v_1 and v_2

$$\begin{aligned}\texttt{Ricci}(v_1, v_2) &= \sum_{j,k} v_{1j} v_{2k} \big(-\frac{1}{4} Tr(g^{-1}\mathrm{D}g_k g^{-1} Dg_j) + \frac{1}{4}\mathrm{D}g(g^{-1}\texttt{tr}(g^{-1}\mathrm{D}g))\big)\\ &= -\frac{1}{4}\texttt{tr}(g^{-1}\mathrm{D}g(v_1) g^{-1} Dg(v_2)) + \frac{1}{4} v_1^\top \mathrm{D}g(g^{-1}\texttt{tr}(g^{-1}\mathrm{D}g))v_2.\end{aligned}$$

■

H.2.6. EXPONENTIAL MAP

The exponential $\exp_p(v)$ at point p is a map from $T_p(\mathcal{M})$ to $\mathcal{M}$, defined as the point obtained on a geodesic starting from p with initial speed v, after time 1. We use $\gamma_t(x)$ to denote the point after going on a geodesic starting from x with initial velocity ∇F, after time t.

Lemma 100 (Commuting derivatives) *Given a family of curves $\gamma_s(t)$ for $s \in [0, s']$ and $t \in [0, t']$, we have*

$$\mathrm{D}_s \partial_t \gamma_s(t) = \mathrm{D}_t \partial_s \gamma_s(t).$$

Proof Let ∂_s and ∂_t be the standard vector fields in the two dimensional $\mathbb{R}^2$ space (t, s). Then, we know

$$\begin{aligned}\mathrm{D}_s \partial_t \gamma_s(t) - \mathrm{D}_t \partial_s \gamma_s(t) &= [\partial_s \gamma_s(t), \partial_t \gamma_s(t)]\\ &= [\partial_t, \partial_s] = 0.\end{aligned}$$

where $[.,.]$ is the Lie bracket. ■

H.3. Hessian manifolds

In this work we are working with a specific class of manifold whose metric is impoesd by the Hessian of our hybrid barrier. A nice property of Hessian manifolds is that the terms in the Riemann tensor which depends on the second derivative of the metric cancels out, and we end up just with the first derivative and the metric itself. Specifically, for a Hessian manifold recall from Lemmas 97, 98, and 101 we have the following equations for Cristoffel symbols, the Riemann tensor, and the Ricci tensor:

$$\begin{aligned}
\Gamma^k_{ij} &= \frac{1}{2}(g^{-1}\mathrm{D}_k g)_{ij},\\
\mathrm{R}^i_{jkl} &= \frac{\partial \Gamma^i_{lj}}{\partial x_k} - \frac{\partial \Gamma^i_{kj}}{\partial x_l} + \sum_p (\Gamma^i_{kp}\Gamma^p_{lj} - \Gamma^i_{lp}\Gamma^p_{kj})\\
&= \sum_{\ell,\ell_2} -\frac{1}{4} g^{i,\ell}\mathrm{D}g_{\ell,k,p} g^{p,\ell_2}\mathrm{D}g_{\ell_2,l,j} + \frac{1}{4} g^{i,\ell}\mathrm{D}g_{\ell,l,p} g^{p,\ell_2}\mathrm{D}g_{\ell_2,k,j},\\
\mathtt{Ricci}(\partial_k, \partial_j) &= \sum_{\ell,\ell_2} -\frac{1}{4} g^{i,\ell}\mathrm{D}g_{\ell,k,p} g^{p,\ell_2}\mathrm{D}g_{\ell_2,i,j} + \frac{1}{4} g^{i,\ell}\mathrm{D}g_{\ell,i,p} g^{p,\ell_2}\mathrm{D}g_{\ell_2,k,j}.
\end{aligned}$$

As we mentioned, the change of the determinant of the Jacobian matrices $J_y^{v_{\gamma_s}}$ regarding the Hamiltonian family $(\gamma_s(t))$ between x_0 and x_1 is related to the rate of change of the Ricci tensor on the manifold. In Lemma 101 below, we concretely record the Ricci tensor for a Hessian manifold in the Euclidean chart, based on the metric g and its derivatives.

Lemma 101 (Form of Ricci tensor on Hessian manifolds) *On a Hessian manifold, the Ricci tensor is given by*

$$\mathtt{Ricci}(v_1, v_2) = -\frac{1}{4} tr(g^{-1}\mathrm{D}g(v_1) g^{-1}\mathrm{D}g(v_2)) + \frac{1}{4} v_1^\top \mathrm{D}g(g^{-1} tr(g^{-1}\mathrm{D}g)) v_2. \tag{128}$$

We use the formula of Ricci tensor on manifold in section D.2 and bound its derivative to bound the rate of change of the pushforward density of RHMC going from x_0 to x_1 in section D.2.2. Note that we only need to have a multiplicative control over the change of density of a sampled Gaussian vector on the destination point on the manifold, as we move from x_0 to x_1.

H.4. Hamiltonian Curves and Fields on Manifold

Proof [Proof of Lemma 12] We start from the ODE of HMC:

$$\gamma_s''(t) = \mu(\gamma_s(t)).$$

Taking covariant derivative in direction s:

$$\begin{aligned}
\mathrm{D}_s\mu(\gamma_s(t)) = \mathrm{D}_s\gamma_s''(t) &= \mathrm{D}_s\mathrm{D}_t\gamma_s''(t)\\
&= \nabla_{\partial_s\gamma_s(t)}\nabla_{\partial_t\gamma_s(t)}\gamma_s'(t).
\end{aligned}$$

Now we apply the definition of Riemann tensor. Namely for arbitrary vector fields X, Y, Z, we have

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

Setting $X = \partial_s\gamma_s(t)$ and $Y = \partial_t\gamma_s(t)$, we first observe that $[\partial_s\gamma_s(t), \partial_t\gamma_s(t)]$ because they are just the application of the differential of γ to the standard vectors ∂_s and ∂_t in $\mathbb{R}^2$. Applying this above

$$\mathrm{D}_s\mu(\gamma_s(t)) = \nabla_{\partial_t\gamma_s(t)}\nabla_{\partial_s\gamma_s(t)}\gamma_s'(t) - R(\partial_s\gamma_s(t), \partial_t\gamma_s(t))\gamma_s'(t). \tag{129}$$

But note that because $\partial_t\gamma_s(t)$ and $\partial_s\gamma_s(t)$ are the image of the differential of $\gamma_s(t)$ applied to ∂_t and ∂_t, we have

$$\nabla_{\partial_s\gamma_s(t)}\gamma_s'(t) = \nabla_{\partial_t\gamma_s(t)}\partial_s\gamma_s(t) = J'(t). \tag{130}$$

Applying Equation (130) to Equation (129):

$$\mathrm{D}_s\mu(\gamma_s(t)) = J''(t) - R(J(t), \gamma_s'(t))\gamma_s'(t).$$

Noting the definition of the operator M completes the proof. ■

H.5. Variations of Geodesics and One-Step Closeness of the Markov Kernel

To bound the one-step closeness of the Markov operator for our algorithm, the object of study in this paper is variations of Hamiltonian curves. However, to give some intuition and provide some more background about the discussion in Section 1.4, we first overview the variations of geodesics on the manifold; suppose $\gamma_s(t)$ is a variation of geodesics, i.e. $\gamma_s(t)$ is a geodesic in t for every fixed $s \in [0, s']$, and $\gamma_s(0)$ is also a geodesic in parameter s from x_0 to x_1. For brevity, we sometimes refer to the curve $\gamma_0(t)$ by $\gamma(t)$. To see how fast the geodesics $\gamma_s(t)$ changes as a function of s at time $s = 0$, for a fixed t we take the derivative of $\gamma_s(t)$ with respect to s at time $s = 0$; this gives us a vector field $J(t)$ along $\gamma_0(t)$:

$$J(t) = \partial_s \gamma_0(t) = \partial_s \gamma_s(t)\Big|_{s=0},$$

This vector field, called a *Jacobi field*, is a fundamental object in studying the variations of geodesics. Importantly, one can write a second-order ODE to describe how $J(t)$ evolves along the geodesic given initial conditions $J(0), J'(0)$

$$D_t^2 J(t) = R(J(t), \gamma'(t))\gamma'(t). \tag{131}$$

where the second derivative $J''(t)$ is the covariant derivative on the manifold with respect to $\gamma_0'(t)$, i.e., $D_t \triangleq \nabla_{\gamma_0'(t)}$, and R is the Riemann tensor. We will provide some intuition on how the Riemann tensor effects in the behavior of geodesics presently. The point is we can study the Jacobi field ODE to estimate how fast the initial velocity is changing along the geodesic from x_0 to x_1, for this family of Hamiltonian curves with the same destination y. Now consider a direction e perpendicular to the velocity $\gamma'(t) = \gamma_0'(t)$ of the geodesic at time t, i.e., $\langle \gamma'(t), e\rangle_g = 0$. Looking at the dot product of the vector $R(e, \gamma'(t))\gamma'(t)$ on the right hand side of the Jacobi field ODE in (131) to e itself, the quantity $\langle e, R(e, \gamma'(t))\gamma'(t)\rangle$ is intuitively measuring how much the Jacobi field is growing or shrinking in direction e, meaning whether the geodesics $\gamma_s(t)$ parameterized by s are converging or diverging in direction e at time $s = 0$. This quantity is known as the sectional curvature of the plane spanned by e and $\gamma'(t)$. Now consider a unit orthonormal parallelepiped at time $t = 0$, denoted by a set of orthonormal vectors $\{e_i\}_{i=1}^n$ in the tangent space of $\gamma(0)$, where $e_1 = \gamma'(0)$, and look at the evolution of its volume along the geodesic when each e_i evolves according to the Jacobi Equation; in each directions e_i, the parallelepiped is either expanding or squeezing, depending on if the geodesics are converging or diverging in that direction, which in turn depends on the sign of the sectional curvature $\langle e_i, R(e_i, \gamma'(0))\gamma'(0)\rangle$. Indeed, one can characterize the rate of change of this parallelepiped along the geodesic by summing the sectional curvatures for all $\{e_i\}_{i=2}^n$, which gives the Ricci curvature of the manifold at $\gamma(0)$ in the direction $\gamma'(0)$: $\texttt{Ricci}(\gamma'(0), \gamma'(0)) = \sum_{i=1}^n \langle e_i, R(e_i, \gamma'(0))\gamma'(0)\rangle$.

On the other hand, the determinant of the Jacobian $J_y^{v_{\gamma_s}}$ of the Hamiltonian map, a quantity of our interest to bound the change of density from x_0 to x_1, can be characterized by the ratio of the volume of this parallelepiped at the beginning and the ending time t. which can be written as a time-weighted integral of the Ricci curvature along the geodesic.

Fortunately, one can extend these arguments to variations of Hamiltonian curves instead of geodesics. As a result, instead of the Riemann tensor in the Jacobi fields Equation (131), we end up with a slightly different operator $\Phi(t)$ (defined in 11) which can be decomposed into a "geometric part," the Riemann tensor, and a "bias part," M_x, which comes from the derivative of the Hamiltonian bias $\mu(x)$, defined in Equation (8).

Appendix I. Third-Order ℓ_∞-Self-Concordance for the Lewis Weight Barrier

Here we show the third-order ℓ_∞-self-concordance of the hybrid barrier.

Lemma 102 (Third-order ℓ_∞-self-concordance for the Lewis weights barrier) *For $p < 4$ for arbitrary direction $v \in \mathbb{R}^n$*

$$-\frac{1}{(4/p-1)^7}\|v\|_{x,\infty}\|z\|_{x,\infty}\|u\|_{x,\infty}g_1 \preccurlyeq \mathrm{D}^3 g_1(v,z,u) \preccurlyeq \frac{1}{(4/p-1)^7}\|v\|_{x,\infty}\|z\|_{x,\infty}\|u\|_{x,\infty}g_1.$$

Proof Here we first carry the same argument as in the proof of second-order ℓ_∞-self-concordance. Namely, if the derivative with respect to one of the variables contains A_x then we can similarly reduce the problem to our estimate for the second-order ℓ_∞-self-concordance, in the proof of Lemma 35. Therefore, here we only bound the terms in which all of the derivative with respect to v, z, and u do not contain A_x. Based on the formula of g_1 in Equation (19):

$$\begin{aligned}\mathrm{D}^3 g_1[v,z,u] \to\ & \mathrm{A}_x^\top(\mathrm{D}^3\mathbf{W}_x[v,z,u] + 2\mathrm{D}^3\mathbf{\Lambda}_x[v,z,u])\mathrm{A}_x \\ &+ 2(1-2/p)\sum_{u,v,z}\mathrm{A}_x^\top(\mathrm{D}\mathbf{\Lambda}_x[v,u,z]{\mathbf{G}_x}^{-1}\mathbf{\Lambda}_x + \mathbf{\Lambda}_x{\mathbf{G}_x}^{-1}\mathrm{D}\mathbf{\Lambda}_x[v,z,u])\mathrm{A}_x \\ &+ 2(1-2/p)\sum_{u,v,z}\mathrm{A}_x^\top(\mathrm{D}\mathbf{\Lambda}_x[v,u]\mathrm{D}{\mathbf{G}_x}^{-1}[z]\mathbf{\Lambda}_x + \mathbf{\Lambda}_x\mathrm{D}{\mathbf{G}_x}^{-1}[z]\mathrm{D}\mathbf{\Lambda}_x[v,u])\mathrm{A}_x \\ &+ 2(1-2/p)\sum_{u,v,z}\mathrm{A}_x^\top(\mathrm{D}\mathbf{\Lambda}_x[v]\mathrm{D}{\mathbf{G}_x}^{-1}[z,u]\mathbf{\Lambda}_x + \mathbf{\Lambda}_x\mathrm{D}{\mathbf{G}_x}^{-1}[z,u]\mathrm{D}\mathbf{\Lambda}_x[v])\mathrm{A}_x \\ &+ 2(1-2/p)\mathrm{A}_x^\top(\mathbf{\Lambda}_x\mathrm{D}{\mathbf{G}_x}^{-1}[v,z,u]\mathbf{\Lambda}_x)\mathrm{A}_x. \end{aligned} \tag{132}$$

In the above, the sums mean all possible ways to distribute derivative directions in that particular way; as an example,

$$\begin{aligned}\sum_{u,v,z}\mathrm{A}_x^\top\mathrm{D}\mathbf{\Lambda}_x[v,u]\mathrm{D}\mathbf{G}_x^{-1}[z]\mathbf{\Lambda}_x\mathrm{A}_x = \mathrm{A}_x^\top\mathrm{D}\mathbf{\Lambda}_x[v,u]\mathrm{D}\mathbf{G}_x^{-1}[z]\mathbf{\Lambda}_x\mathrm{A}_x + \mathrm{A}_x^\top\mathrm{D} \\ + \mathbf{\Lambda}_x[v,z]\mathrm{D}\mathbf{G}_x^{-1}[u]\mathbf{\Lambda}_x\mathrm{A}_x \\ + \mathrm{A}_x^\top\mathrm{D}\mathbf{\Lambda}_x[u,z]\mathrm{D}\mathbf{G}_x^{-1}[v]\mathbf{\Lambda}_x\mathrm{A}_x.\end{aligned}$$

It is clear that we only need to deal with one of these terms per sum. We start from the first line in Equation (132). Using the estimates in Lemma 33 we can write

$$\mathrm{A}_x^\top(\mathrm{D}^3\mathbf{W}_x[v,z,u] + 2\mathrm{D}^3\mathbf{\Lambda}_x)\mathrm{A}_x \preccurlyeq \frac{1}{(4/p-1)^3}\mathrm{A}_x^\top\mathbf{W}_x\mathrm{A}_x.$$

For the second line, it is enough to bound the quadratic form $q^\top\mathrm{A}_x^\top\mathrm{D}\mathbf{\Lambda}_x[v,u,z]{\mathbf{G}_x}^{-1}\mathbf{\Lambda}_x\mathrm{A}_x q$ as the other term is symmetric.

$$q^\top\mathrm{A}_x^\top\mathrm{D}\mathbf{\Lambda}_x[v,u,z]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x\mathrm{A}_x q \leq \sqrt{q^\top\mathrm{A}_x^\top\mathbf{G}_x^{1/2}(\mathbf{G}_x^{-1/2}\mathrm{D}\Lambda[v,u,z]\mathbf{G}_x^{-1/2})^2\mathbf{G}_x^{1/2}\mathrm{A}_x q}\sqrt{q^\top\mathrm{A}_x^\top\mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathbf{\Lambda}_x\mathrm{A}_x q}. \tag{133}$$

Now similar to the trick that we used in Equation (27),

$$\begin{aligned}(\mathbf{G}_x^{-1/2}\mathrm{D}\mathbf{\Lambda}_x[v,u,z]\mathbf{G}_x^{-1/2})^2 &\preccurlyeq \frac{1}{(4/p-1)^6}\|v\|_{x,\infty}^2\|u\|_{x,\infty}^2\|z\|_{x,\infty}^2(\mathbf{G}_x^{-1/2}\mathbf{W}_x\mathbf{G}_x^{-1/2})^2 \\ &\preccurlyeq \frac{1}{(4/p-1)^6}\|v\|_{x,\infty}^2\|u\|_{x,\infty}^2\|z\|_{x,\infty}^2 p^2(\mathbf{W}_x^{-1/2}\mathbf{W}_x\mathbf{W}_x^{-1/2})^2 \\ &\leq \frac{1}{(4/p-1)^6}\|v\|_{x,\infty}^2\|u\|_{x,\infty}^2\|z\|_{x,\infty}^2 p^2\mathrm{I}.\end{aligned} \tag{134}$$

Plugging back Equation (134) into Equation (133):

$$\begin{aligned}q^\top \mathrm{A}_x^\top \mathrm{D}\mathbf{\Lambda}_x[v,u,z]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x\mathrm{A}_x q &\lesssim \frac{1}{(4/p-1)^3}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty} p\sqrt{q^\top \mathrm{A}_x^\top \mathbf{G}_x \mathrm{A}_x q}\sqrt{p}\sqrt{q^\top \mathrm{A}_x^\top \mathbf{\Lambda}_x \mathbf{\Lambda}_x^{-1}\mathbf{\Lambda}_x \mathrm{A}_x q}\\ &\le \frac{p\sqrt{p}}{(4/p-1)^3}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty} q^\top g_1(x) q.\end{aligned}$$

Therefore

$$\mathrm{A}_x^\top \mathrm{D}\mathbf{\Lambda}_x[v,u,z]\mathbf{G}_x^{-1}\mathbf{\Lambda}_x\mathrm{A}_x + \mathrm{A}_x^\top \mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathrm{D}\mathbf{\Lambda}_x[v,u,z]\mathrm{A}_x \preccurlyeq \frac{1}{(4/p-1)^3}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty} g_1.$$

For the third line, using Lemmas 36 and 31

$$\begin{aligned}q^\top \mathrm{D}\mathbf{\Lambda}_x[v,u]\mathrm{D}\mathbf{G}_x^{-1}[z]\mathbf{\Lambda}_x\mathrm{A}_x q &\le \sqrt{q^\top A^\top \mathbf{\Lambda}_x \mathrm{A}_x q}\\ &\times \sqrt{q^\top \mathrm{A}_x^\top \mathrm{D}^2\mathbf{\Lambda}_x[v,u]\mathbf{W}_x^{-1/2}\left(\mathbf{W}_x^{1/2}\mathrm{D}\mathbf{G}_x^{-1}[z]\mathbf{W}_x^{1/2}\right)^2\mathbf{W}_x^{-1/2}\mathrm{D}^2\mathbf{\Lambda}_x[v,u]\mathrm{A}_x q}\\ &\le \sqrt{q^\top A^\top \mathbf{\Lambda}_x \mathrm{A}_x q}\\ &\times \frac{p}{4/p-1}\|z\|_{x,\infty}\sqrt{q^\top \mathrm{A}_x^\top \mathbf{W}_x^{1/2}\left(\mathbf{W}_x^{-1/2}\mathrm{D}^2\mathbf{\Lambda}_x[v,u]\mathbf{W}_x^{-1/2}\right)^2\mathbf{W}_x^{1/2}\mathrm{A}_x q}\\ &\le \sqrt{q^\top A^\top \mathbf{\Lambda}_x \mathrm{A}_x q}\\ &\times \frac{p}{(4/p-1)^3}\|z\|_{x,\infty}\|v\|_{x,\infty}\|u\|_{x,\infty}\sqrt{q^\top \mathrm{A}_x^\top \mathbf{W}_x \mathrm{A}_x q}\\ &\le \frac{p}{(4/p-1)^3}\|z\|_{x,\infty}\|v\|_{x,\infty}\|u\|_{x,\infty} q^\top g_1(x) q.\end{aligned}$$

The terms in the forth line in Equation (132) can be bounded exactly similar to the third line, except that we should use Lemmas 36 and **??** instead. To deal with the term on the fifth line Equation (132) we prove the following estimate on the third order derivative of $\mathbf{G}_x^{-1}$:

Lemma 103 *We have*

$$-\frac{1}{(4/p-1)^7}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty}\mathbf{G}_x^{-1} \preccurlyeq \mathrm{D}^3\mathbf{G}_x^{-1}[v,u,z] \preccurlyeq \frac{1}{(4/p-1)^7}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty}\mathbf{G}_x^{-1}.$$

Proof Note that

$$\begin{aligned}\mathrm{D}^3\mathbf{G}_x^{-1}[v,u,z] &= \mathrm{D}^2(\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x[u]\mathbf{G}_x^{-1})[v,z]\\ &= \sum_{u,v,z}\mathrm{D}^2\mathbf{G}_x^{-1}[v,z]\mathrm{D}\mathbf{G}_x[u]\mathbf{G}_x^{-1}\\ &+ \sum_{u,v,z}\mathrm{D}\mathbf{G}_x^{-1}[v]\mathrm{D}\mathbf{G}_x[u]\mathrm{D}\mathbf{G}_x^{-1}[z]\\ &+ \sum_{u,v,z}\mathrm{D}\mathbf{G}_x^{-1}[v]\mathrm{D}\mathbf{G}_x[u,z]\mathrm{D}\mathbf{G}_x^{-1}\\ &+ \sum_{u,v,z}\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x[u,v,z]\mathbf{G}_x^{-1}.\end{aligned}$$

■

As before, we use the trick of building squares in Equation (27) to bound these terms. The first line is handled by Lemmas 36 and **??**. The second line is handled by Lemmas **??** and 36. The third line is handled by Lemmas 31 and 36. Finally the forth line is handled by Lemma 33. ∎

Now using Lemma 109 the final line in Equation (132) is upper bounded as

$$\begin{aligned} \mathrm{A}_x^\top \mathbf{\Lambda}_x \mathrm{D}^3 \mathbf{G}_x^{-1}[u,v,z] \mathbf{\Lambda}_x \mathrm{A}_x &\preccurlyeq \frac{1}{(4/p-1)^7} \|u\|_{x,\infty} \|v\|_{x,\infty} \|z\|_{x,\infty} \mathrm{A}_x^\top \mathbf{\Lambda}_x \mathrm{D}^3 \mathbf{G}_x^{-1}[u,v,z] \mathbf{\Lambda}_x \mathrm{A}_x \\ &\preccurlyeq \frac{p}{(4/p-1)^7} \|u\|_{x,\infty} \|v\|_{x,\infty} \|z\|_{x,\infty} g_1(x), \end{aligned}$$

which completes the proof.

Appendix J. Remaining Proofs

J.1. Proof of Lemma 36

Proof Note that from Lemma 29

$$\mathrm{D}\mathbf{G}_x^{-1}(v) = -\mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1} \preccurlyeq \frac{1}{4/p-1} \mathbf{G}_x^{-1}$$

and similarly

$$-\frac{1}{4/p-1} \mathbf{G}_x^{-1} \preccurlyeq \mathrm{D}\mathbf{G}_x^{-1}(v).$$

For the second derivative

$$\mathrm{D}^2 \mathbf{G}_x^{-1}(v,u) = -\mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u,v) \mathbf{G}_x^{-1} + \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u) \mathbf{G}_x^{-1} + \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u) \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1}.$$

For the first term using Lemma 31

$$\mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u,v) \mathbf{G}_x^{-1} \preccurlyeq \frac{1}{(4/p-1)^5} \|u\|_{x,\infty} \|v\|_{x,\infty} \mathbf{G}_x^{-1} \mathbf{G}_x \mathbf{G}_x^{-1} = \frac{1}{(4/p-1)^5} \|u\|_{x,\infty} \|v\|_{x,\infty} \mathbf{G}_x^{-1}. \tag{135}$$

For the second term using Cauchy Schwarz for the quadratic form $q^\top \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u) \mathbf{G}_x^{-1} q$ it is enough to upper bound $\mathrm{q}^\top \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1} \mathrm{q}$ and $\mathrm{q}^\top \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u) \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u) \mathbf{G}_x^{-1} \mathrm{q}$ by $\mathbf{G}_x^{-1}$. But by Lemma **??**

$$\begin{aligned} \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1} &= \mathbf{G}_x^{-1/2} (\mathbf{G}_x^{-1/2} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1/2})^2 \mathbf{G}_x^{-1/2} \\ &\preccurlyeq \frac{1}{(4/p-1)^2} \|v\|_{x,\infty}^2 \mathbf{G}_x^{-1}. \end{aligned}$$

Writing the same bound for $\mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u) \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u) \mathbf{G}_x^{-1}$ we conclude

$$\mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u) \mathbf{G}_x^{-1} + \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(u) \mathbf{G}_x^{-1} \mathrm{D}\mathbf{G}_x(v) \mathbf{G}_x^{-1} \preccurlyeq \frac{1}{(4/p-1)^2} \|v\|_{x,\infty} \|u\|_{x,\infty} \mathbf{G}_x^{-1}. \tag{136}$$

Combining Equations (135) and (136) implies

$$\mathrm{D}^2 \mathbf{G}_x^{-1}(v,u) \preccurlyeq \frac{1}{(4/p-1)^5} \|v\|_{x,\infty} \|u\|_{x,\infty} \mathbf{G}_x^{-1}.$$

The proof of the left side $-\frac{1}{(4/p-1)^5} \|v\|_{x,\infty} \|u\|_{x,\infty} \mathbf{G}_x^{-1} \preccurlyeq \mathrm{D}^2 \mathbf{G}_x^{-1}(v,u)$ follows similarly. ∎

J.2. Norm of the Bias

Lemma 104 *We have*

$$\|\mu\|_g \leq (1 + \alpha\sqrt{\alpha_0})\sqrt{n}.$$

Proof For the first part

$$\|\nabla\phi\|_g = \|\mathrm{D}\phi\|_{g^{-1}} \leq \alpha\sqrt{n\alpha_0}$$

from Lemma 56. For the second part, writing $tr(g^{-1}\mathrm{D}g)$ as an expectation

$$\mathtt{tr}(g^{-1}\mathrm{D}g) = \mathbb{E}_{v\sim\mathcal{N}(0,g^{-1})}\mathrm{D}g(v)v,$$

we have for independent $v, v' \sim \mathcal{N}(0, g^{-1})$:

$$\begin{aligned}\|g^{-1}\mathtt{tr}(g^{-1}\mathrm{D}g)\|_g^2 &= \mathbb{E}_{v,v'}v^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v')v' \\ &\leq \mathbb{E}_v v^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)v \\ &\leq \mathbb{E}_v\|s_v\|_\infty^2 v^\top g v \lesssim n,\end{aligned}$$

where we used Lemma 115. This completes the proof. ∎

J.3. Comparison Between Leverage Scores

Lemma 105 *Let*

$$\tilde{\sigma}_i = (\mathbf{W}_x^{1/2}\mathrm{A}_x g^{-1}\mathrm{A}_x^\top \mathbf{W}_x^{1/2})_{i,i}.$$

Then

$$\tilde{\sigma}_i/w_i \leq (\frac{m}{n})^{2/p} w_i^{2/p},$$

which implies

$$\tilde{\sigma}_i/w_i \leq (\frac{m}{n})^{\frac{2/p}{1+2/p}}.$$

Proof Simply note that $g \geq (\frac{n}{m})^{2/p}\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p}\mathrm{A}_x$, which implies

$$(\mathbf{W}_x^{1/2}\mathrm{A}_x(\mathrm{A}_x^\top \mathbf{W}_x^{1-2/p}\mathrm{A}_x)^{-1}\mathrm{A}_x^\top \mathbf{W}_x^{1/2})_{i,i} \leq w_i^{2/p}$$

∎

J.4. Norm Comparison Between Covariant and Normal Derivatives

Lemma 106 *Given a family of Hamiltonian curves γ_s for $t \in (0, \delta)$ where γ_0 is $(\delta, c)-$nice and R_1 normal, defining $v = v_s(t) = \gamma_s'(t)$, then under the assumption that $\delta^2 \lesssim 1/R_1$ and $\|\frac{d}{ds}\gamma_s(0)\Big|_{s=0}\|_g = 1$ we have for all $t \in (0, \delta)$*

$$\begin{aligned}&\|\frac{d}{ds}\gamma_0(t)\|_g \leq 5, \\ &\|\nabla_{\frac{d}{ds}\gamma_0(t)}v_0(t)\|_g \leq \frac{10}{\delta}, \\ &\|\frac{d}{ds}v_s(t)\Big|_{s=0}\|_g \leq c + \frac{10}{\delta}.\end{aligned}$$

Proof Since we have $\delta^2 \leq 1/R_1$ along the curve, Lemma 23 in Lee and Vempala (2018) implies for all $t \in (0, \delta)$:

$$\|\frac{d}{ds}\gamma_s(t)\|_g \leq 5\|\frac{d}{ds}\gamma_s(0)\|_g = 5,$$
$$\|\nabla_{\frac{d}{ds}\gamma_s(t)} v_s(t)\|_g \leq \frac{10}{\delta}.$$

But now from Lemma 107

$$\|\frac{d}{ds}v_0(t)\|_g \leq \|s_{\gamma_0(t),v}\|_\infty \|\frac{d}{ds}\gamma_0(t)\|_g + \|\nabla_{\frac{d}{ds}\gamma_0(t)} v_0(t)\|_g,$$

As always, our parameterization in s is always unit norm, so $\|\frac{d}{ds}\gamma_0(t)\|_g = 1$. From niceness of the curve we have $\|s_{\gamma_0(t),v}\|_\infty \leq c$, which completes the proof. ■

Lemma 107 *For a vector field v and arbitrary vector z at point $x \in \mathcal{M}$, denoting $\mathrm{D}v(z)$ by v', we have*

$$\|v'\|_g \leq \|v\|_{x,\infty}\|z\|_g + \|\nabla_z(v)\|_g.$$

Proof We have

$$\nabla_z(v) = v' + \frac{1}{2}g^{-1}\mathrm{D}g(v)z,$$

so

$$\|v'\|_g \leq \|\nabla_z(v)\|_g + \|g^{-1}\mathrm{D}g(v)z\|_g \leq \|\nabla_z(v)\|_g + \|s_{x,v}\|_\infty \|z\|_g.$$

■

J.5. Iteration complexity of Gaussian Cooling

Proof [Proof of Corollary 5] First, note that from Lemma 38, ϕ is self-concordant with self-concordant parameter $\nu = \alpha_0 n$. The Gaussian cooling schedule introduced by authors in Lee and Vempala (2018) can be used to relax the requirement of having a warm start for our sampling algorithm, resulting in an efficient algorithm for computing the volume as well. The idea is that sampling from Gibbs distributions $e^{-\alpha\phi(x)}$ with smaller variance or larger α is easier, so one can start from sampling in a large temperature α and gradually decrease it. Lemma 45 of Lee and Vempala (2018) shows that the Gaussian Cooling algorithm accurately estimates the volume with any self-concordant barrier function. We now proceed to bound the running time, as this depends on the sampling time in each phase of the GC algorithm.

The Gaussian cooling of Lee and Vempala (2018) evolves in phases where in the ith phase it generates k_i approximate samples from the density proportional to $e^{-\phi(x)/\sigma_i^2}$ inside the polytope, where

$$
\begin{aligned}
&k_i = \Theta(\frac{\sqrt{n}}{\epsilon^2}\log(\frac{\sqrt{n}}{\epsilon})) && \text{if } \sigma_i^2 \le \frac{\nu}{n},\\
&k_i = \Theta((\frac{\sqrt{\nu}}{\sigma}+1)\epsilon^2\log(\frac{n}{\epsilon}), && \text{O.W.}
\end{aligned}
$$

and the update rule for σ_i is

$$
\begin{aligned}
&\sigma_{i+1}^2 = \sigma_i^2(1+\frac{1}{\sqrt{n}}) && \text{if } \sigma_i^2 \le \frac{\nu}{n},\\
&\sigma_i^2 = (1+\min\{\frac{\sigma_i}{\sqrt{\nu}},\frac{1}{2}\}). && \text{O.W.}
\end{aligned}
$$

starting from $\sigma_0^2 = \Theta(\epsilon^2 n^{-3}\log^{-3}(n/\epsilon))$ until σ goes above $\Theta(\frac{\nu}{\epsilon}\log(\frac{n\nu}{\epsilon}))$. Note that the temperature parameter is given by $\alpha = 1/\sigma^2$. Now at each phase i going from temperature σ_i^2 to σ_{i+1}^2 we have an approximate sample from $e^{-\phi(x)/\sigma_i^2}$ which can be used as a warm start for sampling from $e^{-\phi(x)/\sigma_{i+1}^2}$, specially as $k_{i+1} \le k_i$. Therefore, our main Theorem 4 implies that the mixing time of sampling at each phase is of order

$$
\begin{aligned}
&\tilde{O}\big(\min\{\alpha^{-1}n^{2/3}+\alpha^{-1/3}n^{5/9}m^{1/9}+n^{1/3}m^{1/6},\ m^{1/3}n^{4/3}\}\big)\\
&=\tilde{O}(\alpha^{-1}n^{2/3}+\alpha^{-1/3}n^{5/9}m^{1/9}+n^{1/3}m^{1/6}).
\end{aligned}
$$

Now in the first case when $\sigma_i^2 \le \frac{\nu}{n} = \alpha_0$, we have $\alpha \ge \frac{1}{\alpha_0}$. On the other hand, due to the update rule of σ_i in this case, it takes $\sqrt{n}$ phase to double σ and in each phase we take $k_i = \tilde{\Theta}(\frac{\sqrt{n}}{\epsilon^2})$ samples. Hence, the total number of RHMC steps to double σ in this case is bounded by

$$
\begin{aligned}
&\tilde{O}\big((\alpha^{-1}n^{2/3}+\alpha^{-1/3}n^{5/9}m^{1/9}+n^{1/3}m^{1/6})\times\frac{\sqrt{n}}{\epsilon^2}\times\sqrt{n}\big)\\
&=\tilde{O}\big((\alpha_0 n^{2/3}+\alpha_0^{1/3}n^{5/9}m^{1/9}+n^{1/3}m^{1/6})\frac{n}{\epsilon^2}\big)\\
&=\tilde{O}(\frac{n^{4/3}m^{1/3}}{\epsilon^2}).
\end{aligned}
$$

In the other case when $\sigma_i^2 \ge \frac{\nu}{n} = \alpha_0$, we have $\alpha \le \frac{1}{\alpha_0}$. Then, the total RHMC steps to double σ in this case can be upper bounded after substituting $\nu = n\alpha_0$ as

$$
\tilde{O}\big((\alpha^{-1}n^{2/3}+\alpha^{-1/3}n^{5/9}m^{1/9}+n^{1/3}m^{1/6})\times\frac{1}{\epsilon^2}(\frac{\sqrt{\nu}}{\sigma}+1)\times(\frac{\sqrt{\nu}}{\sigma}+1)\big)=\tilde{O}(\frac{n^{4/3}m^{1/3}}{\epsilon^2}).
$$

This means we can calculate the integral of $e^{-\alpha\phi(x)}$ for any α using $\tilde{O}(\frac{n^{4/3}m^{1/3}}{\epsilon^2})$ steps of RHMC up to $1 \pm \epsilon$. Moreover, if we just want to sample from $e^{-\alpha\phi(x)}$ in the polytope, we do not have to take k_i number of samples at phase i but only need one sample. As a result, the ϵ^2 in the complexity is omitted and we end up with the complexity $\tilde{O}(n^{4/3}m^{1/3})$ for sampling without a warm start. ∎

J.6. Other helper Lemmas

Lemma 108 *We have the following relations between $\mathbf{P}_x^{(2)}$, $\mathbf{\Lambda}_x$, and $\mathbf{W}_x$.*

$$\frac{2}{p}\mathbf{W}_x \preccurlyeq \mathbf{G}_x \preccurlyeq \mathbf{W}_x,$$
$$\mathbf{P}_x^{(2)} \preccurlyeq \mathbf{W}_x,$$
$$\mathbf{\Lambda}_x \preccurlyeq \mathbf{W}_x.$$

Proof For the first inequality, note that the sum of entries of the ith row of the matrix $\mathbf{P}_x^{(2)}$ is equal to $\mathbf{W}_{xi,i}$. Hence, the matrix $\mathbf{W}_x - \mathbf{P}_x^{(2)}$ is a Laplacian so it is positive semi-definite. The second inequality follows from the fact that $\mathbf{\Lambda}_x = \mathbf{W}_x - \mathbf{P}_x^{(2)}$, and that $\mathbf{P}_x^{(2)}$ is PSD. For the third inequality, using the fact that $\mathbf{P}_x^{(2)} \preccurlyeq \mathbf{W}_x$, we can write

$$\frac{2}{p}\mathbf{W}_x \preccurlyeq \frac{2}{p}\mathbf{W}_x + (1 - \frac{2}{p})\mathbf{P}_x^{(2)} \preccurlyeq \mathbf{W}_x.$$

∎

Lemma 109 *We have*

$$-\frac{1}{(4/p-1)^3}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty}\mathbf{G}_x^{-1} \preccurlyeq \mathrm{D}^3\mathbf{G}_x^{-1}[v,u,z] \preccurlyeq \frac{1}{(4/p-1)^3}\|v\|_{x,\infty}\|u\|_{x,\infty}\|z\|_{x,\infty}\mathbf{G}_x^{-1}.$$

Proof Note that

$$\begin{aligned}
\mathrm{D}^3\mathbf{G}_x^{-1}[v,u,z] &= \mathrm{D}^2(\mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x[u]\mathbf{G}_x^{-1})[v,z] \\
&= \sum_{u,v,z} \mathrm{D}^2\mathbf{G}_x^{-1}[v,z]\mathrm{D}\mathbf{G}_x[u]\mathbf{G}_x^{-1} \\
&+ \sum_{u,v,z} \mathrm{D}\mathbf{G}_x^{-1}[v]\mathrm{D}\mathbf{G}_x[u]\mathrm{D}\mathbf{G}_x^{-1}[z] \\
&+ \sum_{u,v,z} \mathrm{D}\mathbf{G}_x^{-1}[v]\mathrm{D}\mathbf{G}_x[u,z]\mathrm{D}\mathbf{G}_x^{-1} \\
&+ \sum_{u,v,z} \mathbf{G}_x^{-1}\mathrm{D}\mathbf{G}_x[u,v,z]\mathbf{G}_x^{-1}.
\end{aligned}$$

∎

Lemma 110 *For any positive integer n, vector v, and matrix $\tilde{g} \preccurlyeq g$ we have*

$$g^{1/2}(g^{-1/2}\tilde{g}g^{-1/2})^n g^{1/2} \preccurlyeq g.$$

Proof *Directly from the fact that if $A \preccurlyeq B$, then for any matrix C we have $C^\top AC \preccurlyeq C^\top BC$.* ∎

Lemma 111 *For operator* $g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)$*, we have* $\|g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)\ell\|_g \lesssim \|s_{x,v}\|_\infty^2\|\ell\|_g$.

Proof We have

$$\|g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)\ell\|_g \leq \sqrt{\ell^\top \mathrm{D}g(v)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)g^{-1}\mathrm{D}g(v)\ell} \tag{137}$$

$$\leq \|s_{x,v}\|_\infty^2\sqrt{\ell^\top g\ell} \lesssim \|s_{x,v}\|_\infty^2\|\ell\|_g. \tag{138}$$

■

Lemma 112 *For vector field* w *on manifold* $\mathcal{M}$*, we have*

$$\mathtt{tr}(g^{-1}\mathrm{D}g(w)) \lesssim \|w\|_g.$$

Proof We have

$$\begin{aligned}\mathtt{tr}(g^{-1}\mathrm{D}g(w)) &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}{v'}^\top \mathrm{D}g(w)v' \\ &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}{v'}^\top \mathrm{D}g(v')w \\ &\leq \mathbb{E}_{v'}\|v'\|_\infty\sqrt{{v'}^\top g v'}\sqrt{w^\top g w} \\ &\lesssim \sqrt{n}\|w\|_g,\end{aligned}$$

where in the last line we used Lemma 115. ■

Lemma 113 *For arbitrary vector field* w *on* $\mathcal{M}$ *we have*

$$\left|\mathtt{tr}(g^{-1}\mathrm{D}g(z,w))\right| \leq \sqrt{n}\|w\|_g\|z\|_g.$$

Proof We can write

$$\begin{aligned}&\left|\mathtt{tr}(g^{-1}\mathrm{D}g(z,w))\right| \\ &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}{v'}^\top \mathrm{D}g(z,w)v' \\ &= \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}{v'}^\top \mathrm{D}g(z,v')w \\ &\leq \mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}\|z\|_\infty\|v'\|_\infty\sqrt{{v'}^\top g v'}\sqrt{w^\top g w} \\ &\leq \|w\|_g\sqrt{n}\|z\|_\infty \\ &\leq \|w\|_g\sqrt{n}\|z\|_g.\end{aligned}$$

■

Lemma 114 *For vector field* w *we have*

$$|\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(w))| \leq \sqrt{n}\|z\|_g\|w\|_g.$$

Proof

$$\begin{aligned}|\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(w))| &= |\mathbb{E}_{v'\sim\mathcal{N}(0,g^{-1})}v'^{\top}\mathrm{D}g(z)g^{-1}\mathrm{D}g(w)v'| \\ &= \mathbb{E}_{v'}|v'^{\top}\mathrm{D}g(z)g^{-1}\mathrm{D}g(v')w| \\ &\leq \mathbb{E}_{v'}\sqrt{v'^{\top}\mathrm{D}g(z)g^{-1}\mathrm{D}g(v')g^{-1}\mathrm{D}g(v')g^{-1}\mathrm{D}g(z)v'}\|w\|_g \\ &= \mathbb{E}_{v'}\sqrt{v'^{\top}\mathrm{D}g(z)g^{-1/2}(g^{-1/2}\mathrm{D}g(v')g^{-1/2})^2g^{-1/2}\mathrm{D}g(z)v'}\|w\|_g.\end{aligned}$$

But note that

$$g^{-1/2}\mathrm{D}g(v')g^{-1/2} \leq \|v'\|_\infty I.$$

Hence

$$\begin{aligned}|\mathtt{tr}(g^{-1}\mathrm{D}g(z)g^{-1}\mathrm{D}g(w))| &\leq \mathbb{E}_{v'}\sqrt{v'^{\top}g^{1/2}(g^{-1/2}\mathrm{D}g(z)g^{-1/2})^2g^{1/2}v'}\|w\|_g \\ &\leq \mathbb{E}_{v'}\|s_z\|_\infty\|v'\|_g\|w\|_g \\ &\leq \mathbb{E}_{v'}\|z\|_g\|v'\|_g\|w\|_g \\ &\leq \sqrt{n}\|z\|_g\|w\|_g.\end{aligned}$$

where we used Lemma 115 and Lemma 95. ■

Lemma 115 *For normal vector $v\sim\mathcal{N}(0,g(x)^{-1})$, we have*

$$\begin{aligned}&\mathbb{E}\|s_{x,v}\|_\infty v^{\top}gv = O(n\sqrt{\log(m)}), \\ &\mathbb{E}\|s_{x,v}\|_\infty = O(\sqrt{\log(m)}), \\ &\mathbb{E}\|s_{x,v}\|_\infty^2 = O(\log(m)), \\ &\mathbb{E}\|s_{x,v}\|_\infty^4 = O(\log(m)^2).\end{aligned}$$

Proof These follow directly from standard Gaussian moment bounds. ■

Fact 1 (Gaussian tail bound) *For Gaussian random variable $X\sim\mathcal{N}(0,\sigma^2)$, we have the following tail bound*

$$\mathbb{P}(|X|\geq t)\leq 2e^{-t^2/2\sigma^2}.$$

J.7. Proof of Theorem 25

Proof Consider a subset $S\subseteq\mathcal{S}$ with $0.5\geq\pi(S)=s'\geq s\geq 2\rho$. Then, to show a lower bound for s-conductance, we need to lower bound

$$P(S,S^c)/P(S),$$

where $P(.,.)=\int_{x\in S}\mathcal{T}_x(S^c)\pi(x)dx$ is the probability that we are in set S and the next step of the Markov chain we escape S and P is the probability measure corresponding to π. Recall that $\mathcal{T}_x(.)$ is the Markov kernel, specifying the distribution of the next step given we are at point x. Now assume that the conductance bound does not hold, i.e. there exists such S with

$$P(S,S^c)/P(S)=O(\Delta\psi_M).$$

Note that because the chain is reversible, we have

$$P(S, S^c) = P(S^c, S),$$

and because $\pi(S) \leq 0.5$, we have

$$P(S^c, S)/P(S^c) \leq P(S, S^c)/P(S) = O(\Delta\psi_M). \tag{139}$$

Next, define the set $\tilde{S} \subseteq S$ to be the points x from which our chance of escaping S is at least 0.01. Now if $\pi(S) \geq \Delta\psi_{\mathcal{M}}\pi(S)/2$, then given that we are in S, we have at least $\Delta\psi_{\mathcal{M}}$ chance of escaping S which contradicts (139). This means

$$\pi(\tilde{S}) \leq \Delta\psi_{\mathcal{M}}/2.\pi(S). \tag{140}$$

On the other hand, note that for point x_1 with $d(x_1, x_0) \leq \Delta$ for $x_0 \in S - \tilde{S}$, we have

$$TV(P_{x_0}, P_{x_1}) \leq 0.9, \tag{141}$$

which means x_1 cannot be in $S - \tilde{S}$, hence it should be in S^c. Therefore, defining the set $S^{+\Delta}$ as the set of points outside $\tilde{S}$ which are Δ close to a point in $S - \tilde{S} - \mathcal{M}'^c$, we have

$$S^{+\Delta} \subseteq S^c \cup \mathcal{M}'^c. \tag{142}$$

On the other hand, from isoperimetry (because $\pi(S) \leq \frac{1}{2}$) and the fact that $\Delta\psi_{\mathcal{M}} \leq 1/2$ we have

$$\pi(S^{+\Delta}) \geq \Delta\psi_{\mathcal{M}}(\pi(S - \tilde{S}) - \pi(\mathcal{M}'^c)) \geq \Delta\psi_{\mathcal{M}}(\pi(S)/2 - \rho) \geq \Delta\psi_{\mathcal{M}}(\pi(S)/4).$$

Therefore, from the assumption $s \geq \rho/(8\Delta\psi_{\mathcal{M}})$:

$$\pi(S^{+\Delta} - \mathcal{M}'^c) \geq \Delta\psi_{\mathcal{M}}(\pi(S)/4 - \pi(S)/8) \geq \Delta\psi_{\mathcal{M}}\pi(S)/8,$$

which implies from Equations (141) and (142):

$$P(S, S^c) \geq P(S, S^{+\Delta} - \mathcal{M}'^c) = P(S^{+\Delta} - \mathcal{M}'^c, S) \geq \Delta\psi_{\mathcal{M}}(\pi(S)/8) \times 0.99 \geq \Delta\psi_{\mathcal{M}}\psi(S)/16,$$

which proves that the conducance is lower bounded by $\Omega(\Delta\psi_{\mathcal{M}})$. ■

J.8. Some Properties of the Lewis Weights

In this section, we recall some properties of Lewis weights which we use in the proof.

Lemma 116 (Fixed point property of Lewis weights) *The Lewis weights of the matrix* A_x *is the unique vector* w_x *in* $\mathbb{R}^m_{\geq 0}$ *with* $\mathbf{W}_x = \mathtt{Diag}(w_x)$ *such that*

$$\sigma(\mathbf{W}_x^{1/2-1/p}\mathrm{A}_x) = \mathbf{W}_x,$$

where $\sigma(.)$ *denotes the leverage scores of the matrix.*

Proof Recall the definition of Lewis weights as the optimum of the objective in Equation (14). Taking derivative with respect to W, we get

$$-(1-2/p)\sigma/w + (1-2/p)1^\top w = 0,$$

where recall $\sigma \triangleq (W^{1/2-1/p}A_x)$ is the vector of leverage scores defined as

$$\sigma(W^{1/2-1/p}\mathrm{A}_x) \triangleq \mathtt{diag}\big(W^{1/2-1/p}\mathrm{A}_x(\mathrm{A}_x^\top W^{1-2/p}\mathrm{A}_x)^{-1}{\mathrm{A}_x}^\top W^{1/2-1/p}\big).$$

■

Proof [Proof of Lemma 21] We reuse Lemma 24 in Lee and Sidford (2019) with variable V set as S_x. Then, noting the fact that $\mathrm{D}S_x[v] = -S_x\mathrm{S}_{x,v}$, we obtain the following formula for the derivative of $\mathbf{W}_x$ with respect to x in direction v:

$$\mathrm{D}\mathbf{W}_x[v] = -2\mathtt{Diag}\Big(\mathbf{W}_x\Big(\mathbf{W}_x - (1-2/p)\mathbf{\Lambda}_x\Big)^{-1}\mathbf{\Lambda}_x s_{x,v}\Big) = -2\mathtt{Diag}\big(\mathbf{W}_x\mathbf{G}_x^{-1}\mathbf{\Lambda}_x s_{x,v}\big).$$

But simple algebra reveals,

$$2\mathbf{W}_x - (1-2/p)\mathrm{D}\mathbf{W}_x[v] = 2\mathbf{W}_x\mathbf{G}_x^{-1}\mathbf{W}_x,$$

which further implies

$$2\mathbf{W}_x - (1-2/p)\mathrm{D}\mathbf{W}_x[v] - 2(1-2/p)\mathtt{Diag}\big(\mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v}\big) = 2\mathbf{W}_x.$$

Therefore

$$\mathrm{D}\mathbf{W}_x[v] = -2\mathtt{Diag}\big(\mathbf{\Lambda}_x\mathbf{G}_x^{-1}\mathbf{W}_x s_{x,v}\big).$$

■

Proof Proof of Lemma 28 Recall the definition of $\mathbf{P}_x$:

$$\mathbf{P}_x = \mathbf{W}_x^{1/2-1/p}\mathrm{A}_x(\mathrm{A}_x^\top\mathbf{W}_x^{1-2/p}\mathrm{A}_x)^{-1}{\mathrm{A}_x}^\top\mathbf{W}_x^{1/2-1/p}.$$

Using the chain rule

$$\begin{aligned}\mathrm{D}\mathbf{P}_x[v] = {} & \mathtt{Diag}\big(\mathbf{W}_x^{-1/2-1/p}(-\mathbf{G}_x - 2(1/2-1/p)\mathbf{\Lambda}_x)r_{x,v}\big)\mathrm{A}_x^\top\mathbf{W}_x^{1-2/p}\mathrm{A}_x\mathbf{W}_x^{1/2-1/p} \\ & + \mathbf{W}_x^{1/2-1/p}\mathrm{A}_x^\top\mathbf{W}_x^{1-2/p}\mathrm{A}_x\mathtt{Diag}\big(\mathbf{W}_x^{-1/2-1/p}(-\mathbf{G}_x - 2(1/2-1/p)\mathbf{\Lambda}_x)r_{x,v}\big) \\ & - 2\mathbf{W}_x^{1/2-1/p}\mathrm{A}_x(\mathrm{A}_x^\top\mathbf{W}_x^{1-2/p}\mathrm{A}_x)^{-1}\mathrm{A}_x^\top W^{-2/p} \\ & \mathtt{Diag}\big((-2\mathbf{G}_x - 2(1-2/p)\Lambda)r_{x,v}\big)\mathrm{A}_x(\mathrm{A}_x^\top\mathbf{W}_x^{1-2/p}\mathrm{A}_x)^{-1}\mathrm{A}_x^\top\mathbf{W}_x^{1/2-1/p}.\end{aligned}$$

Noting the fact that $-\mathbf{G}_x - 2(1/2-1/p)\mathbf{\Lambda}_x = \mathbf{W}_x$ finishes the proof. ■

Lemma 117 (Gradient of the Lewis weights barrier) *The gradient of the Lewis weight barrier ϕ_p is given by*

$$\mathrm{D}\phi_p(x) = \mathrm{A}_x^\top w_x.$$

Proof Taking directional derivative in direction v, using the chain rule

$$\begin{aligned}\mathrm{D}\phi_p(x)[v] = {} & 2\mathtt{tr}\big((\mathrm{A}_x^\top\mathbf{W}_x^{1-2/p}\mathrm{A}_x)^{-1}(\mathrm{A}_x^\top\mathrm{S}_{x,v}\mathbf{W}_x^{1-2/p}\mathrm{A}_x\big) \\ & + \mathrm{D}(w_x)^\top\frac{\partial\big(-\operatorname{logdet}(\mathrm{A}_x^\top\mathbf{W}_x^{1-2/p}\mathrm{A}_x) + (1-2/p)1^T w\big)}{\partial w_x},\end{aligned}$$

But because w_x is the maximizer of $\big(-\operatorname{logdet}(\mathrm{A}_x^\top\mathbf{W}_x^{1-2/p}\mathrm{A}_x) + (1-2/p)1^\top w\big)$, the second term is zero and the proof is complete. ■