Gaussian Cooling and Dikin Walks: The Interior-Point Method for Logconcave Sampling

Yunbum Kook YB.KOOK@GATECH.EDU
Georgia Institute of Technology

Santosh S. Vempala VEMPALA@GATECH.EDU
Georgia Institute of Technology

Editors: Shipra Agrawal and Aaron Roth

Abstract

The connections between (convex) optimization and (logconcave) sampling have been considerably enriched in the past decade with many conceptual and mathematical analogies. For instance, the Langevin algorithm can be viewed as a sampling analogue of gradient descent and has condition-number-dependent guarantees on its performance. In the early 1990s, Nesterov and Nemirovski developed the Interior-Point Method (IPM) for convex optimization based on self-concordant barriers, providing efficient algorithms for structured convex optimization, often faster than the general method. This raises the following question: can we develop an analogous IPM for structured sampling problems?

In 2012, Kannan and Narayanan proposed the Dikin walk for uniformly sampling polytopes, and an improved analysis was given in 2020 by Laddha-Lee-Vempala. The Dikin walk uses a local metric defined by a self-concordant barrier for linear constraints. Here we generalize this approach by developing and adapting IPM machinery together with the Dikin walk for poly-time sampling algorithms. Our IPM-based sampling framework provides an efficient warm start and goes beyond uniform distributions and linear constraints. We illustrate the approach on important special cases, in particular giving the fastest algorithms to sample uniform, exponential, or Gaussian distributions on a truncated PSD cone. The framework is general and can be applied to other sampling algorithms.

Keywords: Dikin walk, Interior-Point method, sampling with local geometry, warm-start generation.

1. Introduction

Consider the following motivating problem: how can we efficiently sample a $d \times d$ matrix from a distribution with the following density?

$$\begin{aligned} &\text{sample } X \sim \exp\Big(-\big(\langle A, X\rangle + \|X-B\|_F + \|X-C\|_F^2 - \log\det X\big)\Big) \\ &\quad \text{s.t. } X \succeq 0,\ \langle D_i, X\rangle \geq c_i\,, \quad \forall i \in [m]\,. \end{aligned}$$

This rather complicated looking distribution recovers as special cases the problems of sampling from the Max-Cut semi-definite programming relaxation and the set of minimum (or bounded) volume ellipsoids that contain a given set of points. The above density is logconcave, so we can use Ball walk or Hit-and-Run (with rounding) to sample the distribution with $\mathcal{O}(d^8 \log d)$ membership/evaluation queries. This "general-purpose" sampler already gives a poly-time mixing algorithm. However, each term in the density and constraints is "structured", which poses the following natural question: can we leverage *structure* inherent in the problem to get more efficient algorithms?

Let us consider an 'optimization' version of the sampling question above for a moment, motivated by a conceptual analogy between sampling and optimization: sample from $\exp(-V)$ vs minimize V. Then, the optimization version turns into a *structured* convex optimization problem with structured objectives and constraints. Formally, for proper convex functions f_i and h_j, we want to minimize $\sum_i f_i(x)$ subject to $h_j(x) \le 0$, for which the *interior-point method* (IPM) is a powerful framework.

Interior-point method. Given a convex optimization problem – $\min_{x \in K} f(x)$ for a real-valued convex function f on convex $K \subset \mathbb{R}^d$ – the IPM first replaces f by a new variable t and adds the epigraph constraint $\{f(x) \le t\}$. Here, the *structured* constraints / objectives admit a self-concordant barrier ϕ for the augmented constraints, and IPM tackles a 'regularized' problem – $\min f_\lambda(x,t) := t + \frac{1}{\lambda}\phi(x,t)$ with parameter $\lambda > 0$. Then, an optimization step that is aware of the local geometry given by $\nabla^2\phi$ moves the current point closer to $\arg\min f_\lambda$, with the barrier function ϕ designed to prevent escaping from the feasible region. Increasing λ a bit, IPM repeats this procedure with the updated point used as a starting point for solving $\min f_{\lambda+\delta}$. As λ increases, the effect of $\frac{1}{\lambda}\phi$ vanishes in the regularized problem, which gradually brings us to a point sufficiently closer to the minimum.

Returning to the sampling problem with IPM in mind, we pose our main question:

Problem *Let f_i be a proper convex function and h_j a convex function on $\mathbb{R}^d$ for $i \in [I]$ and $j \in [J]$. Then the goal is to develop a sampling IPM for solving*

$$\text{sample } x \sim \pi \propto \exp\Big(-\sum_i f_i\Big) \quad \text{s.t.} \quad x \in K := \bigcap_{j \in [J]} \{x \in \mathbb{R}^d : h_j(x) \le 0\}\,, \qquad \text{(strLC)}$$

where we assume that K has non-empty interior and π has finite second moments.

In this paper, we derive an IPM framework for structured logconcave sampling, answering two important questions: (1) What is a geometry-aware algorithm sampling from a regularized distribution, (2) what should be an annealing schedule of λ, and how to control closeness of distributions?

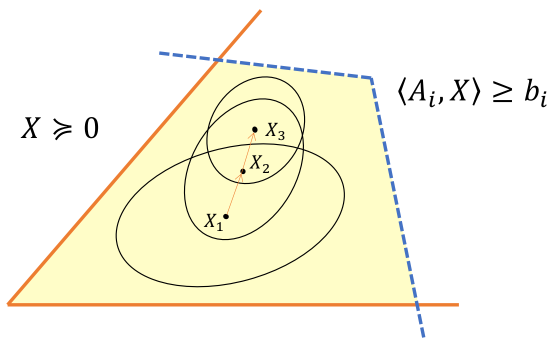
As for (1), we use Dikin walk as a sampler to implement the "inner" step of IPM, providing a gentle introduction to Dikin walk and self-concordance in §2. We then provide a mixing-time bound for Dikin walk, going beyond uniform distributions (§3.1) and recovering previous work as special cases. This generalization is necessary to be able to utilize Dikin walk within the IPM framework. As for (2), we present in §3.2 the sampling IPM with its role of warm-start generation and theoretical guarantees. Our framework is suited for breaking down complicated sampling problems into smaller structured problems (i.e., write $f = \sum_i f_i$ and $K = \cap_j K_j$). Toward this divide-and-conquer approach, we develop in §3.3 a "calculus" for combining multiple constraints and objectives, and deriving the resulting theoretical guarantees (analogous to and inspired by the work of Nesterov and Nemirovskii (1994) for optimization). To provide concrete understanding and instances, we illustrate the framework on some well-known families of constraints in §3.4, in particular obtaining faster algorithms to sample uniform, exponential, or Gaussian distributions on truncated Positive Semi-Definite (PSD) cones in §3.5. We refer readers to §4 for background and related work, and to Figure A.1 for an overall structure of the paper.

2. Warm-up: Dikin walk and self-concordance

We use the same symbol for a distribution and its density w.r.t. the Lebesgue measure. We use $\mathbb{S}_+^d$ (and $\mathbb{S}_{++}^d$) to denote the set of $d \times d$ positive semidefinite (and definite) matrices, respectively.

For two matrices A, B, we use $A \asymp B$ to mean $A \precsim B$ and $B \precsim A$. A *local metric* g defines at each point $x \in K \subset \mathbb{R}^d$ a positive-definite inner product $\langle \cdot, \cdot \rangle_{g(x)} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, which induces the local norm $\|v\|_{g(x)} := [\langle v, v \rangle_{g(x)}]^{1/2}$. We use $\|v\|_x$ to refer to $\|v\|_{g(x)}$ when the context is clear. We abuse notation and use $g(x)$ to denote the $d \times d$ positive-definite matrix represented with respect to the canonical basis $\{e_1, \dots, e_d\}$. For a function f defined on $K \subset \mathbb{R}^d$, we let $\mathrm{D}^i f(x)[h_1, \dots, h_i]$ denote the i-th directional derivative of f at x in directions $h_1, \dots, h_i \in \mathbb{R}^d$, i.e., $\mathrm{D}^i f(x)[h_1, \dots, h_i] = \frac{\mathrm{d}^i}{\mathrm{d}t_1 \cdots \mathrm{d}t_i} f(x + \sum_{j=1}^i t_j h_j)|_{t_1,\dots,t_i=0}$. We let $\mathcal{N}_g^r(x) := \mathcal{N}(x, \frac{r^2}{d} g(x)^{-1})$ be the normal distribution with mean x and covariance $\frac{r^2}{d} g(x)^{-1}$. See §A for other notation.

Dikin walks and self-concordance. Given a local metric g in $\mathbb{R}^d$, the Dikin ellipsoid of radius r at $x \in \mathbb{R}^d$ is defined as $\mathcal{D}_g^r(x) \overset{\text{def}}{=} \{y \in \mathbb{R}^d : \|y - x\|_{g(x)} \le r\}$. i.e., it is a norm ball of radius r defined by the local metric. From this perspective, Dikin walk (Algorithm 1) is a natural generalization of Ball walk to a local metric setting.



Algorithm 1: Dikin walk (π_0, π, g, r, T)

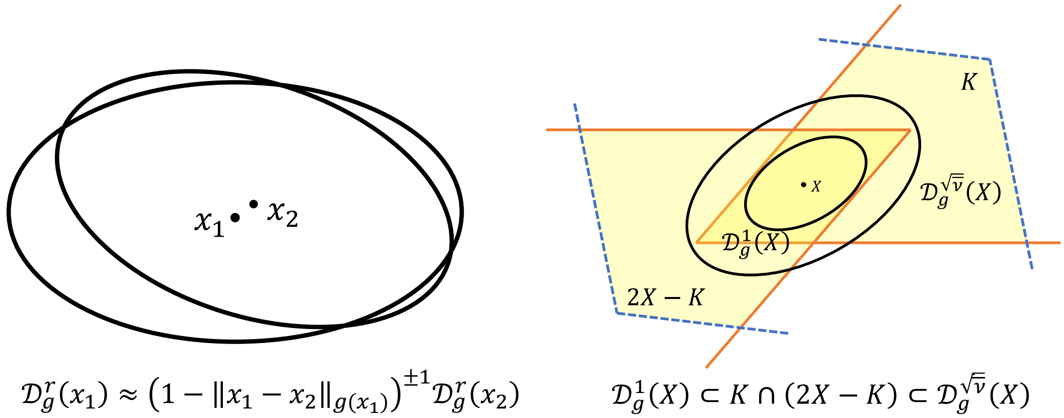
Input: Initial dist. π_0, target dist. $\pi \propto e^{-f}\, \mathbf{1}_K$, local metric g, step size r, # iterations T.
Output: x_T
Sample $x_0 \sim \pi_0$ at random.
for $t = 0, \cdots, T-1$ **do**
  Sample $z \sim \mathcal{N}\big(x_t, \frac{r^2}{d} g(x_t)^{-1}\big)$.
  $x_{t+1} \leftarrow z$ w.p. $A_{x_t}(z) := 1 \wedge \big(\frac{p_z(x_t)\pi(z)}{p_{x_t}(z)\pi(x_{t-1})}\big)$.
  Otherwise, $x_{t+1} \leftarrow x_t$.
end

Table 1: Iterates of the Dikin walk. Solid lines centered at X_i indicate Dikin ellipsoids, $\mathcal{D}_g^r(X_i)$. The pdf of $\mathcal{N}(x, \frac{r^2}{d} g(x)^{-1})$ is denoted by p_x.

The metric g used to define Dikin walk plays a crucial role in its convergence. Our metrics will be defined by Hessians of convex self-concordant barrier functions. We now collect definitions of these functions; they will be important to state our general guarantees for the mixing of Dikin walk. The concept we need is summarized by the definition of a $(\nu, \bar{\nu})$-*Dikin-amenable metric*.

Definition 2.1 (Self-concordance (brief version of Definition A.1)) *For convex $K \subset \mathbb{R}^d$, let $\phi : \mathrm{int}(K) \to \mathbb{R}$ be a smooth convex function, $g(\cdot) \asymp \nabla^2 \phi(\cdot)$, and $\mathcal{N}_g^r(x) := \mathcal{N}\big(x, \frac{r^2}{d} g(x)^{-1}\big)$.*

- ν-self-concordant barrier *(SC): (i)* $|\mathrm{D}^3 \phi(x)[h, h, h]| \le 2\|h\|^3_{\nabla^2 \phi(x)}$ *for any* $x \in \mathrm{int}(K)$ *and* $h \in \mathbb{R}^d$, *(ii)* $\lim_{x \to \partial K} \phi(x) = \infty$, *and (iii)* $\|\nabla \phi(x)\|^2_{[\nabla^2 \phi(x)]^{-1}} \le \nu$ *for any* $x \in \mathrm{int}(K)$.
- Highly SC *(HSC):* $|\mathrm{D}^4 \phi(x)[h, h, h, h]| \le 6\|h\|^4_{\nabla^2 \phi(x)}$ *for any* $x \in \mathrm{int}(K)$ *and* $h \in \mathbb{R}^d$, *and* $\lim_{x \to \partial K} \phi(x) = \infty$.
- Strong SC *(SSC):* $\|g(x)^{-\frac{1}{2}} \mathrm{D}g(x)[h]\, g(x)^{-\frac{1}{2}}\|_F \le 2\|h\|_{g(x)}$ *for any* $x \in \mathrm{int}(K)$ *and* $h \in \mathbb{R}^d$.
- Strongly lower trace SC *(SLTSC):* $\mathrm{Tr}\big(\big(\bar{g}(x) + g(x)\big)^{-1} \mathrm{D}^2 g(x)[h, h]\big) \ge -\|h\|^2_{g(x)}$ *for any* $\bar{g} : \mathrm{int}(K) \to \mathbb{S}^d_+$, $x \in \mathrm{int}(K)$, *and* $h \in \mathbb{R}^d$. *We call it lower trace self-concordant (LTSC) if it is satisfied when* $\bar{g} = 0$.



(a) Self-concordance of barrier/metric (b) $\bar{\nu}$-symmetry

Figure 2.1: (a) Self-concordance of barrier/metric (Definition 2.1) ensures that the Hessian (so Dikin ellipsoids) changes smoothly. (b) $\bar{\nu}$-symmetry (Definition 2.2) indicates how well a Dikin ellipsoid $\mathcal{D}_g^r(X)$ approximates the locally symmetrized convex body, $K \cap (2X - K)$.

- Strongly average SC *(SASC): For any* $\varepsilon > 0$ *and* $\bar{g} : \operatorname{int}(K) \to \mathbb{S}_+^d$*, there exists* $r_\varepsilon > 0$ *such that* $\mathbb{P}_{z \sim \mathcal{N}_{g+\bar{g}}^r(x)}\left(\|z - x\|_{g(z)}^2 - \|z - x\|_{g(x)}^2 \le \frac{2\varepsilon r^2}{d}\right) \ge 1 - \varepsilon$ *for* $r \le r_\varepsilon$*. We call it average self-concordant (ASC) if this is satisfied when* $\bar{g} = 0$*.*

SC imposes regularity on the eigenvalues of the directional derivative $\mathrm{D}g[h]$ through $-2\|h\|_g^2 g \preceq \mathrm{D}g[h] \preceq 2\|h\|_g^2 g$ (or equivalently the largest magnitude of eigenvalues of $g^{-1/2}\mathrm{D}g[h]g^{-1/2}$), and HSC does the same on the higher-order derivative $\mathrm{D}^2 g[h, h]$. SSC introduced by Laddha et al. (2020) imposes *stronger* regularity on the eigenvalues of $\mathrm{D}g[h]$ by definition, as SSC is stated in terms of the *Frobenius norm* of $g^{-\frac{1}{2}}\mathrm{D}g[h]g^{-\frac{1}{2}}$. LTSC relaxes 'convexity of $\log\det g$' required by Laddha et al. (2020). In particular, SSC and LTSC control the change of $\log\det g$, leading to a refined analysis of Dikin walk. Lastly, ASC is pertinent to the average of the squared local norm difference of $z - x$ computed at z and x, which controls the acceptance-probability of each iterate of Dikin walk.

These notions are sophisticated enough to carry out a tight mixing analysis of Dikin walk, but also simple enough for us to develop a "calculus" for combining metrics for multiple constraints in §3.3. Moreover, these conditions may look difficult to verify, but we show that a proper scaling of (H)SC barriers immediately makes them satisfy these properties.

Next, we recall a symmetry parameter of a self-concordant metric. We will later see that it has a natural connection to Cheeger isoperimetry.

Definition 2.2 ($\bar{\nu}$-symmetry) *For convex* $K \subset \mathbb{R}^d$*, a PSD matrix function* $g : \operatorname{int}(K) \to \mathbb{S}_+^d$ *is said to be* $\bar{\nu}$-symmetric *if* $\mathcal{D}_g^1(x) \subseteq K \cap (2x - K) \subseteq \mathcal{D}_g^{\sqrt{\bar{\nu}}}(x)$ *for any* $x \in K$*.*

We note that $K \cap (2x - K)$ is the locally symmetrized convex body with respect to x. Hence, $\bar{\nu}$-symmetry measures how accurately a Dikin ellipsoid approximates the locally symmetrized body. One can show that $\bar{\nu} = \mathcal{O}(\nu^2)$ for any metric induced by a self-concordant barrier.

Going forward, we call a PD matrix function $\bar{\nu}$-*Dikin-amenable* if it is SSC, LTSC, ASC, and $\bar{\nu}$-symmetric. We sometimes call it $(\nu, \bar{\nu})$-Dikin-amenable to reveal its self-concordance parameter ν. For example, the Hessian of the log-barrier for m linear constraints is (m, m)-Dikin-amenable. We present several concrete examples in §3.4.

3. Main results and technical overview

We provide four necessary components for a sampling IPM — Dikin walk for the inner-step of IPM (§3.1), algorithm design (§3.2), calculus for the divide-and-conquer approach (§3.3), and parameter estimations for well-known families of distributions/constraints (§3.4). Putting them together, we obtain end-to-end guarantees for sampling from structured logconcave distributions. Generating a warm start without overhead, the sampling IPM serves a faster sampling algorithm for numerous constrained distributions previously studied (§3.5). We discuss interesting future directions in §3.6.

3.1. Result 1 - Mixing of Dikin walk for general well-conditioned distributions

We begin with our general analysis of Dikin walk (see §B for the details).

Theorem 3.1 Let $K \subset \mathbb{R}^d$ be convex and $0 \le \alpha \le \beta < \infty$.

- (Local metric) Assume that a C^1-matrix function $g : \mathrm{int}(K) \to \mathbb{S}^d_{++}$ is $\bar{\nu}$-Dikin-amenable.
- (Distribution) Let π_0 and $\pi \propto \exp(-f) \cdot \mathbf{1}_K$ be an initial and target distribution respectively, where f is α-relatively strongly convex and β-smooth in g. Let $\|\pi_0/\pi\| = \mathbb{E}_{\pi_0}\big[\frac{\mathrm{d}\pi_0}{\mathrm{d}\pi}\big]$ and P be the transition kernel of Dikin walk with the local metric g and step size $r = \mathcal{O}(1 \wedge \beta^{-1/2})$.

Then for any $\varepsilon > 0$, it holds that $d_{\mathrm{TV}}(\pi_0 P^{(T)}, \pi) \le \varepsilon$ for $T \gtrsim d(1 \vee \beta)\,(\bar{\nu} \wedge 1/\alpha) \log \frac{\|\pi_0/\pi\|}{\varepsilon}$.

This result serves as a unifying framework that recovers as special cases previous works on Dikin walk for uniform sampling (Kannan and Narayanan, 2012; Narayanan, 2016; Chen et al., 2018; Laddha et al., 2020), as seen later in §3.5. Our analysis extends beyond uniform sampling, considering Dikin walk under a more general setting where the potential f satisfies $\alpha g \preceq \nabla^2 f \preceq \beta g$ on $\mathrm{int}(K)$. This setting is a generalization of $\alpha I \preceq \nabla^2 f \preceq \beta I$ to a local metric $\nabla^2 \phi \asymp g$. We also note that Dikin walk is the first *implementable* algorithm that provides a clean mixing guarantee under this general setting, which is a necessary ingredient for theory of our sampling IPM.

Challenges. Laddha et al. (2020) attempted to characterize properties of g (or ϕ) that determine mixing times of Dikin walks for uniform sampling. These necessitates that g satisfy $\bar{\nu}$-symmetry, SSC, convexity of $\log\det g(x)$, and $x \in \mathcal{D}_g^r(z)$ w.h.p. $z \sim \mathrm{Unif}\big(\mathcal{D}_g^r(x)\big)$. However, when a constraint is given as a set of convex sets, their framework encounters a challenge arising from the difficulty of verifying the convexity of $\log\det(g_1 + g_2)$ when $\log\det g_i$ is convex for each $i = 1, 2$.

To address this and succinctly characterize essential characteristics of a metric for one-step coupling, we relax the convexity of $\log\det$ to (S)LTSC and introduce the notion of ASC to account for the condition "$x \in \mathcal{D}_g^r(z)$ w.h.p.". Then under Dikin-amenability of a metric, we establish a one-step coupling, one of main proof ingredients in obtaining a mixing-time guarantee of Dikin walk.

Proof ideas. We use a conductance argument (Lovász and Simonovits, 1993), lower-bounding the conductance of a Markov chain, which follows from two ingredients: one-step coupling and isoperimetry. We focus on the first, a main difficulty when extending to general distributions.

One-step coupling requires that transition kernels at two nearby points have TV-distance bounded away from one. All previous analyses of Dikin walk do not go through for general distributions, since those techniques either have gap or yield a wrong proof.

A key distinction in extension lies in establishing a lower bound for the ratio $\frac{\exp(f(x))}{\exp(f(z))}$ to ensure a high acceptance probability. One can show $\frac{\exp(f(x))}{\exp(f(z))} \ge 1 - \varepsilon$ at the expense of $\frac{1}{2} + \varepsilon$ probability, by using the Taylor expansion of f, self-concordance of g, and symmetry of the proposal. However, this

$\frac{1}{2}+\varepsilon$ loss is incompatible with previous approach based on the triangle inequality: for a transition kernel T and proposal kernel P, we have $d_{\text{TV}}(T_x,T_y)\le d_{\text{TV}}(T_x,P_x)+d_{\text{TV}}(P_x,P_y)+d_{\text{TV}}(P_y,T_y)$, but the bound of $\frac{1}{2}+\varepsilon$ for both $d_{\text{TV}}(T_x,P_x)$ and $d_{\text{TV}}(T_y,P_y)$ makes the RHS vacuous.

We instead work with the exact formula for $d_{\text{TV}}(T_x,T_y)$: for the Gaussian $p_x=\mathcal{N}(x,\frac{r^2}{d}g(x)^{-1})$,

$$R_x(z)=\frac{p_z(x)}{p_x(z)}\frac{\pi(z)}{\pi(x)}=\frac{p_z(x)}{p_x(z)}\frac{\exp(f(x))}{\exp(f(z))},\qquad A_x(z)=\min\bigl(1,R_x(z)\,\mathbf{1}_K(z)\bigr)\,,$$

the transition kernel is $T_x(dz)=(1-\mathbb{E}_{p_x}A_x)\,\delta_x(\mathrm{d}z)+A_x(z)\,p_x(\mathrm{d}z)$. For $r_x:=1-\mathbb{E}_{p_x}A_x$,

$$d_{\text{TV}}(T_x,T_y)=\frac{r_x+r_y}{2}+\frac{1}{2}\int|A_x(z)\,p_x(z)-A_y(z)\,p_y(z)|\,\mathrm{d}z\,.$$

As for r_x and r_y, we bound below $\frac{p_z(x)}{p_x(z)}$ by $1-\varepsilon$ at the cost of ε-probability through SSC, LTSC, and ASC of g. As mentioned earlier, $\frac{\exp(f(x))}{\exp(f(z))}\ge 1-\varepsilon$ at the cost of $\frac{1}{2}+\varepsilon$ probability. Combining these results, we obtain upper bounds of $\frac{1}{2}+\varepsilon$ for small $\varepsilon>0$ on r_x and r_y.

Establishing a bound of $1/4+\varepsilon$ on the second term is a more involved task. It requires the closeness of acceptance probabilities $A_x(z)$ and $A_y(z)$ as well as that of the probability densities $p_x(z)$ and $p_y(z)$. This closeness can be achieved through sophisticated conditioning on high-probability events due to ASC, SSC, and symmetry of Gaussian proposals. We refer readers to the sketch in §B.1.

3.2. Result 2 - Sampling IPM: Gaussian cooling with the Dikin walk (GCDW)

We present Gaussian cooling (see §C for details), essentially a sampling analogue of the optimization IPM. The *function counterpart* refers to a self-concordant barrier ϕ such that $\nabla^2\phi\asymp g$ on $\operatorname{int}(K)$.

Theorem 3.2 For convex $K\subset\mathbb{R}^d$, suppose that $g:\operatorname{int}(K)\to\mathbb{S}^d_{++}$ is $(\nu,\bar{\nu})$-Dikin-amenable and ϕ is its function counterpart such that $\min_K\phi$ exists. Gaussian cooling with Dikin walk (Algorithm 3 with Dikin walk serving as a non-Euclidean sampler) generates a sample that is ε-close to $\exp(-f)\cdot\mathbf{1}_K$ in TV-distance using $\mathcal{O}\bigl(d\,(d\frac{\nu\beta+d}{\nu\alpha+d}\vee\nu\vee\bar{\nu})\log\frac{d\nu}{\varepsilon}\bigr)$ iterations of Dikin walk with g, where a C^2-function $f:\operatorname{int}(K)\to\mathbb{R}$ satisfies $\alpha\nabla^2\phi\preceq\nabla^2 f\preceq\beta\nabla^2\phi$ on K for $0\le\alpha\le\beta<\infty$. In particular, when $f(x)=\alpha^\mathsf{T}x$ or $c\phi(x)$ for $\alpha\in\mathbb{R}^d$ and $c\in\mathbb{R}_+$, the algorithm uses $\widetilde{\mathcal{O}}(d\,(d\vee\nu\vee\bar{\nu}))$ iterations of the Dikin walk.

The inner loop of Gaussian cooling runs Dikin walk. The basic GC algorithm was introduced in Cousins and Vempala (2018) for efficient sampling and volume computation. Lee and Vempala (2018) studied its extension to Hessian manifolds for uniformly sampling polytopes. Our framework is more general in that it handles more general distributions through a sophisticated annealing scheme.

This framework provides an efficient algorithm for generating a warm start for constrained log-concave distributions. If we were to apply Theorem 3.1 with initial distribution being a single point at some distance from boundary, even for the simplest case of uniform sampling, then an additional factor of d would be incurred. On the other hand, given that ν and $\bar{\nu}$ are typically $\mathcal{O}(d)$, our framework only has a logarithmic (in dimension) factor overhead for generating a warm start. An important reason why this works is the affine-invariance of Dikin walk. Samplers like Ball walk have to apply isotropic transformation to achieve a warm start efficiently, which requires a near-linear number of samples and thus have at least a linear in dimension overhead.

Derivation of the algorithm. We describe this algorithm alongside its interpretation as a 'sampling analogue of the *interior-point method*'. To this end, we revisit high-level ideas of IPM, derive its sampling version via a conceptual analogy between optimization and sampling, and then refine the derived sampling IPM by highlighting the distinctions between the two method. See §C.1 for details.

(1) Optimization IPM (Algorithm 2). In solving the optimization problem, $\min_{x\in K} f(x)$ for a real-valued convex function f on convex $K \subset \mathbb{R}^d$, IPM first replaces f by a new variable t and appends the epigraph $\{(x,t) \in \mathbb{R}^{d+1} : f(x) \leq t\}$ to the constraint in addition to $x \in K$. Then summation of self-concordant barriers for K and the epigraph results in a ν-self-concordant barrier ϕ for the augmented constraints. This barrier ϕ allows one to convert the constrained problem to a unconstrained one, $\min f_\lambda(x,t) := t + \frac{1}{\lambda}\phi(x,t)$ for a parameter $\lambda > 0$. Then an optimization step (e.g., the Newtonian gradient descent) that takes into account the local geometry given by $\nabla^2\phi$ moves a current point closer to an optimal point, with the barrier ϕ preventing escape from the constraints. Increasing $\lambda \leftarrow \lambda(1 + \frac{1}{\sqrt{\nu}})$, IPM repeats this procedure with the updated point used as a starting point. As λ increases (until $\lambda \leq \nu/\varepsilon$ for target accuracy $\varepsilon > 0$), the effect of $\frac{1}{\lambda}\phi(x,t)$ vanishes in the regularized problem, which gradually brings us to a point sufficiently closer to the minimum.

(2) Translation to sampling (Figure C.1). We recall the following conceptual match between convex optimization and logconcave sampling: for convex $K \subset \mathbb{R}^d$ and convex function $f : K \to \mathbb{R}$

$$\min f(x) \quad \text{s.t.} \quad x \in K \quad \longleftrightarrow \quad \text{sample } x \sim \pi \propto \exp(-f) \quad \text{s.t.} \quad x \in K\,.$$

With the connection in mind, we can translate IPM's machinery into the sampling context. As in IPM, we replace f by a new variable t, introduce the epigraph constraint, and attempt to sample a 'regularized' distribution $\mu_{\sigma^2}(x,t) \propto \exp\bigl(-f_{\sigma^2}(x,t)\bigr) = \exp\bigl(-\bigl(t + \frac{1}{\sigma^2}\phi(x,t)\bigr)\bigr)$, where a parameter σ^2 corresponds to λ above. This sampling step should be carried out with a sampler *aware* of the local geometry given by $\nabla^2\phi$ (call it NE-sampler, which is Dikin walk in our case). Then we increase σ^2 slightly, and using the previous regularized distribution μ_{σ^2} as a warm start, we sample a next regularized distribution $\mu_{\sigma^2+\varepsilon}$. This iterative procedure continues until σ^2 reaches ν.

(3) Refinements (Figure C.2). We now make this conceptual algorithm concrete in Algorithm 3. The finalized sampling IPM[1] consists of four phases — Phase 1 for initialization, Phase 2 and 3 for increasing σ^2 with control, and Phase 4 for high-accuracy sampling.

Phase 1 initializes the algorithm by a Gaussian truncated over a Dikin ellipsoid of radius $\mathcal{O}(d^{-\Theta(1)})$. This Gaussian serves as a good warm start for a regularized distribution with small σ^2.

The sampling IPM, in contrast to both the optimization IPM and the basic GC algorithm, proceeds with a distinct annealing scheme. Phase 2 updates $\sigma^2 \leftarrow \sigma^2(1 + 1/\sqrt{d})$ until σ^2 reaches ν/d, annealing not only ϕ but also the 'modified' potential $\nu t/d$. While $\frac{\nu}{d} \leq \sigma^2 \leq \nu$, Phase 3 updates $\sigma^2 \leftarrow \sigma^2(1 + \sigma/\sqrt{\nu})$ but only ϕ part with the potential t now fixed. We note that the basic GC anneals only regularization term throughout.

Lastly, the sampling IPM runs Dikin walk once in Phase 4. If one stopped after Phase 3 (when σ^2 reaches as the optimization version, then the total iterates of Dikin walk would be $\mathcal{O}(d\,(d \vee \nu)/\text{poly}(\varepsilon))$. This guarantee can avoid the symmetry parameter, but this comes at the cost of low-accuracy of the sampler (i.e., dependence on $\text{poly}(\varepsilon^{-1})$). Hence, we finish up the algorithm with another execution of Dikin walk, obtaining high-accuracy $\mathcal{O}(d\,(d \vee \nu \vee \bar{\nu}) \log \frac{1}{\varepsilon})$-mixing.

GCDW is exactly this refined algorithm with Dikin walk used for the NE-sampler (Algorithm 3). Specifically in the inner loop, it runs Dikin walk to sample regularized exponential distributions of

1. We focus on e^{-t} for exposition. Our algorithm can deal with more general potentials (relatively convex and smooth).

the form $\exp(-(c_1 t + c_2\phi(x,t))$ subject to $x \in K$ and $\{(x,t) \in \mathbb{R}^{d+1} : f(x) \le t\}$, where the local metric therein consists of the Hessians of self-concordant barriers for K and the level set of f. Comparing with Ball walk for a general logconcave distribution (Lovász and Vempala, 2007), incorporating the geometry of a level set of f (not $\nabla^2 f$) is a natural approach to sampling from e^{-f}.

Closeness of consecutive distributions. At the heart of the algorithm lies closeness of regularized distributions in consecutive iterations. That is, a distribution $\mu_i := \mu_{\sigma_i^2}$ serves as a good warm start for the subsequent distribution $\mu_{i+1} := \mu_{\sigma_{i+1}^2}$ (i.e., $\|\mu_i/\mu_{i+1}\|$ is small).

For the first two phases, closeness of consecutive distributions follows purely from a property of log-concave distributions, which is independent of local metrics. To sketch the idea, in Phase 2, it holds that for $\psi := \frac{\nu t}{d} + \phi$ on K and $F(\sigma^2) := \int_K \exp(-\psi/\sigma^2)$,

$$\|\mu_i/\mu_{i+1}\| = F\Big(\big(\frac{2}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2}\big)^{-1}\Big) F(\sigma_{i+1}^2)/F(\sigma_i^2)^2\,.$$

As the function $a \mapsto a^d F\big(\frac{\sigma^2}{a}\big)$ is log-concave (Lemma C.2), the update rule $\sigma_{i+1}^2 = (1 + \frac{1}{\sqrt{d}})\sigma_i^2$ and the definition of logconcavity with endpoints $\frac{2}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2}$ and $\frac{1}{\sigma_{i+1}^2}$, and the middle point $\frac{1}{\sigma_i^2}$, lead to

$$\frac{F\big((\frac{2}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2})^{-1}\big)\, F(\sigma_{i+1}^2)}{F(\sigma_i^2)^2} \le \left(\frac{\big(\frac{1}{\sigma_i^2}\big)^2}{\big(\frac{2}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2}\big)\frac{1}{\sigma_{i+1}^2}}\right)^d = \left(\frac{\big(1+\frac{1}{\sqrt{d}}\big)^2}{1+\frac{2}{\sqrt{d}}}\right)^d \le \Big(1+\frac{1}{d}\Big)^d \le e\,.$$

In Phase 3, for $r = \frac{\sigma_i}{\sqrt{\nu}}$, $s = \frac{r}{1+r}$, $\sigma = \sigma_i$, and $F(\sigma^2) = \int \exp(-t - \phi/\sigma^2)$, we have

$$\mathcal{R}_2(\mu_i \,\|\, \mu_{i+1}) = {}^{2}\log\frac{F\big(\frac{\sigma^2}{1+s}\big)\,F\big(\frac{\sigma^2}{1-s}\big)}{F(\sigma^2)^2} = \int_0^s\int_{1-l}^{1+l} g''(q)\,\mathrm{d}q\,\mathrm{d}l = \int_0^s\int_{1-l}^{1+l}\frac{1}{\sigma^4}\,\mathrm{Var}_{\nu_q}\phi\,\mathrm{d}q\,\mathrm{d}l$$

where $g(l) := \log F\big(\frac{\sigma^2}{l}\big)$ for $l > 0$ and $\nu_q \propto \exp\big(-t - \frac{q\phi}{\sigma^2}\big)$ is a probability measure. By the Brascamp-Lieb inequality with a function $V := t + \frac{q\phi}{\sigma^2}$, for a self-concordance parameter ν

$$\mathrm{Var}_{\nu_q}\phi \le \mathbb{E}_{\nu_q}\big[(\nabla\phi)^\mathsf{T}\big(\nabla^2 V\big)^{-1}\nabla\phi\big] \le \frac{\sigma^2}{q}\,\mathbb{E}_{\nu_q}[\|\nabla\phi\|^2_{(\nabla^2\phi)^{-1}}] \le \frac{\sigma^2\nu}{q}\,.$$

Putting this back to the integral, one can check $\mathcal{R}_2(\mu_i \,\|\, \mu_{i+1}) \lesssim \frac{\nu s^2}{\sigma^2}$. It follows from $s = \frac{r}{1+r}$ and $r = \frac{\sigma}{\sqrt{\nu}}$ that μ_i is an $\mathcal{O}(1)$-warm start for μ_{i+1}. As for Phase 4, the same technique along with a limiting argument shows that the final distribution μ_ν is an $\mathcal{O}(1)$-warm start for the target π.

Inexact error analysis. The law of iterate of Dikin walk slightly deviates from an actual target in each phase, so the sampling IPM uses as an initial distribution an inexact one $\bar{\mu}_{\sigma^2}$, not μ_{σ^2}. One can resolve this discrepancy through the triangle inequality.

Let $m = \mathcal{O}(\sqrt{d})$ be the number of total phases involved in the sampling IPM with initial distribution π_0, a target distribution π_i in phase $i \in [m]$, and the Markov kernel P_i defined by Dikin walk such that $d_{\mathrm{TV}}(\pi_{i-1}P_i, \pi_i) \le \varepsilon$. Then, the law of sample at the end of each phase is actually $\hat{\pi}_i := \pi_0 P_1 \cdots P_i = \hat{\pi}_{i-1}P_i$. Using the triangle inequality and data-processing inequality,

$$d_{\mathrm{TV}}(\pi_i, \hat{\pi}_i) \le d_{\mathrm{TV}}(\pi_i, \pi_{i-1}P_i) + d_{\mathrm{TV}}(\pi_{i-1}P_i, \hat{\pi}_{i-1}P_i) \le \varepsilon + d_{\mathrm{TV}}(\pi_{i-1}, \hat{\pi}_{i-1}) \le i\varepsilon\,.$$

Hence, $d_{\mathrm{TV}}(\pi_m, \hat{\pi}_m) \le m\varepsilon$, so it suffices to work with ε/m in place of ε throughout the analysis.

2. $\mathcal{R}_2 := \log(\chi^2 + 1) = \log\|\mu_i/\mu_{i+1}\|$ is the 2-Rényi divergence.

3.3. Result 3 - Self-concordance theory for combining barriers

The sampling IPM allows us to focus on the following reduced problem: Let $t_1, \dots, t_I \in \mathbb{R}$ and $y = (x, t_1, \dots, t_I) \in \mathbb{R}^d \times \mathbb{R}^I = \mathbb{R}^{d+I}$. We denote $E_i := \{(x, t_i) \in \mathbb{R}^{d+1} : f_i(x) \le y_{n+i}\}$ for $i \in [I]$ and $K_j := \{x \in \mathbb{R}^d : h_j(x) \le 0\}$ for $j \in [J]$, whose convexity follows from convexity of f_i and h_j. Denoting the embeddings of E_i and K_j onto $\mathbb{R}^{d+I}$ by $\bar{E}_i$ and $\bar{K}_j$, we can reduce (strLC) to

$$\text{sample } y \sim \tilde{\pi} \propto \exp\bigl(-(\underbrace{0, \dots, 0}_{d \text{ times}}, \underbrace{1, \dots, 1}_{I \text{ times}})^\mathsf{T} y\bigr) \quad \text{s.t.} \quad y \in K' := \bigcap_{i=1}^{I} \bar{E}_i \cap \bigcap_{j=1}^{J} \bar{K}_j\,, \qquad (\mathsf{redLC})$$

where K' is closed convex and has non-empty interior, and we are given self-concordant barriers for each E_i and K_j. As the x-marginal of $\tilde{\pi}$ is π, we just project a drawn sample from $\tilde{\pi}$ to the x-space. When $f_i(x)$ can be written as d separable terms (i.e., $f_i(x) = \sum_{l=1}^d f_{i,l}(x_l)$), it is more convenient to introduce d many variables $t_{i,1}, \dots t_{i,d}$ for $f_{i,1}(x_1), \dots, f_{i,d}(x_d)$.

In §D, we study how to combine a self-concordant metric and its parameters from each epigraph E_i and convex set K_j (for the mixing estimation of Dikin walk). As in the optimization IPM, the addition of all barriers is actually a good candidate of a barrier for K', but under an appropriate scaling. However, the sampling version requires not only self-concordance parameters but also symmetry parameters, SSC, and LTSC for final mixing time guarantees. Notably, SSC and LTSC assume invertibility of a local matrix function, but the Hessian of a barrier for a lower-dimensional space is *degenerate w.r.t.* the augmented variable $y \subset \mathbb{R}^{d+I}$. We address this technical issue by working with Definition D.16 and several matrix lemmas to study how to maintain or update each of the main properties such as symmetry, SSC, and LTSC under addition and scaling.

Using these notions, we can state how to put together information of a barrier for each constraint / epigraph. The readers can note the analogy to Nesterov and Nemirovski's IPM theory for optimization.

Theorem 3.3 In the reduced problem (redLC), assume the following:
- For $i \in [I]$, the epigraph E_i admits a PSD matrix function $g_i^e(x, t_i)$ (or $g_i^e(x, t_{i,1}, \dots, t_{i,d})$) that is a $(\nu_i, \bar{\nu}_i)$-SC barrier, SSC along some subspace, SLTSC, and SASC.
- For $j \in [J]$, the constraint K_j admits a PSD matrix function $g_j^c(x)$ that is a $(\eta_j, \bar{\eta}_j)$-SC barrier, SSC along some subspace, SLTSC, and SASC.

For appropriate projections π_i^e and π^c, a matrix function g on $y \in \operatorname{int}(K')$ defined by

$$\langle u, v \rangle_{g(y)} := (I + J) \Bigl(\sum_{i=1}^{I} \langle \pi_i^e u, \pi_i^e v \rangle_{g_i^e(\pi_i^e(y))} + \sum_{j=1}^{J} \langle \pi^c u, \pi^c v \rangle_{g_j^c(\pi^c(y))}\Bigr) \quad \text{for } u, v \in \mathbb{R}^d$$

is $\bigl((I+J)(\sum_{i=1}^I \nu_i + \sum_{j=1}^J \eta_j),\ (I+J)(\sum_{i=1}^I \bar{\nu}_i + \sum_{j=1}^J \bar{\eta}_j)\bigr)$-Dikin-amenable on K'.

3.4. Result 4 - Metrics for well-known structured instances

In §E, we examine required parameters and properties of a barrier for a structured constraint and potential, such as linear, quadratic, entropy, ℓ_p-norm, and PSD cone. See Table 2.

(1) Linear constraints. We start with linear constraints given by $K := \{x \in \mathbb{R}^d : Ax \ge b\}$ for $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$, where A is assumed to have no all-zero rows. For $x \in \operatorname{int}(K)$ and $i \in [m]$, let a_i be the i-th row of A, and denote $S_x := \operatorname{Diag}(a_i^\mathsf{T} x - b_i) \in \mathbb{R}^{m \times m}$ and $A_x := S_x^{-1} A \in \mathbb{R}^{m \times d}$.

These linear constraints admit efficiently computable self-concordant barriers: logarithmic barrier, Vaidya metric, and Lewis-weight metric. The simplest one is the logarithmic barrier defined by

Constraints / Epigraphs	Barrier	ν	$\bar{\nu}$	SSC	LTSC	SLTSC	ASC	SASC
$Ax \geq b$	ϕ_{log}	m	m					
	g_{Vaidya}	$\sqrt{md}$	$\sqrt{md}$					
	g_{Lw}	d	d		$\sqrt{d}$	$\sqrt{d}$	$\sqrt{d}$	$\sqrt{d}$
$\lVert x-\mu\rVert_{\Sigma}^{2} \leq 1$	ϕ_{ellip}			d			d	d
$\lVert x-\mu\rVert_{\Sigma}^{2} \leq t$	ϕ_{Gauss}			d			d	d
$\lVert x-\mu\rVert_{\Sigma} \leq t$	ϕ_{SOC}			d	d	d	d	d
$X \succeq 0$	ϕ_{PSD}	d	d	d			d	d^2
$-x_i \log x_i \leq t_i \; \forall i \in [d]$	ϕ_{ent}	d	d				d	d
$\lvert x_i\rvert^p \leq t_i \; \forall i \in [d]$	ϕ_{power}	d	d				d	d

Table 2: Self-concordance and symmetry parameters, and required scaling factors for a family of barriers. Here, we assume $A \in \mathbb{R}^{m\times d}$, $x \in \mathbb{R}^d$, and $X \in \mathbb{S}_{+}^{d}$. Empty entries indicate $\mathcal{O}(1)$-scalings.

$\phi_{\text{log}}(x) := -\sum_{i=1}^{m} \log(a_i^{\mathsf{T}} x - b_i)$. When the number of constraints m is large, one can use a self-concordant metric due to Vaidya (1996). For a full-rank matrix A, the resulting Vaidya metric takes advantage of the *leverage scores* $\sigma(A_x)$ of A_x, the diagonal entries of the orthogonal projection matrix $P_x = A_x(A_x^{\mathsf{T}} A_x)^{-1} A_x \in \mathbb{R}^{m\times m}$, i.e., $[\sigma(A_x)]_i := (P_x)_{ii} > 0$ for $i \in [m]$. For $\Sigma_x = \operatorname{Diag}(\sigma(A_x)) \in \mathbb{R}^{m\times m}$, the Vaidya metric is defined by $g_{\text{Vaidya}}(x) := \mathcal{O}(1)\sqrt{\frac{m}{d}}\, A_x^{\mathsf{T}}\big(\Sigma_x + \frac{d}{m} I_m\big) A_x$, which satisfies $g_{\text{Vaidya}} \asymp \nabla^2\big(\sqrt{\frac{m}{d}}(\phi_{\text{vol}} + \frac{d}{m}\phi_{\text{log}})\big)$ for $\phi_{\text{vol}} := \frac{1}{2}\log\det(\nabla^2 \phi_{\text{log}})$.

Its self-concordance parameter is still $\operatorname{poly}(m)$. It turns out that dependence on m can be made poly-logarithmic by a Lewis-weight metric that utilizes the *Lewis weights* of A_x. The ℓ_p-Lewis weight of A_x is the vector $w_x \in \mathbb{R}^m$ satisfying the implicit equation $w_x = \sigma\big(\operatorname{Diag}(w_x)^{1/2-1/p} A_x\big)$. Note that the leverage scores can be recovered as the ℓ_2-Lewis weight of A_x. Then the Lewis-weight metric is defined by $g_{\text{Lw}}(x) := \mathcal{O}(\log^{\mathcal{O}(1)} m)\, A_x^{\mathsf{T}} W_x A_x$. With $p = \mathcal{O}(\log^{\Theta(1)} m)$, the self-concordance parameter of this metric can be made $\mathcal{O}^*(d)$.

For the sampling purpose, we need to look into other properties: SSC, SLTSC, and SASC, etc. The log-barrier and Vaidya metric fulfill these without additional scaling, while the Lewis-weight metric requires a $\sqrt{d}$-scaling for SLTSC and SASC. We summarize these results below.

Theorem 3.4 (Linear constraints (Informal)) *We assume $m \geq d$ in the cases of the Vaidya and Lewis-weight. Let w_x be the ℓ_p-Lewis weights with $p = \mathcal{O}(\log^{\Theta(1)} m)$.*
- *Log-barrier ϕ_{log}: $g = \nabla^2 \phi_{\text{log}}$ satisfies $\nu, \bar{\nu} \leq m$, SSC, SLTSC, and SASC.*
- *Vaidya metric g_{Vaidya}: $g = 44 g_{\text{Vaidya}}$ satisfies $\nu, \bar{\nu} = \mathcal{O}(\sqrt{md})$, SSC, SLTSC, and SASC.*
- *Lewis-weight metric g_{Lw}: $g = \sqrt{d} g_{\text{Lw}}$ satisfies $\nu, \bar{\nu} = \widetilde{\mathcal{O}}(d^{3/2})$, SSC, SLTSC, and SASC.*

(2) Quadratic potentials and constraints. Now consider quadratic potential (i.e., Gaussian) and constraints (i.e., ellipsoid and second-order cone). A self-concordant barrier introduced by Nesterov and Nemirovskii (1994) serves as an efficient barrier for each constraint or epigraph of a potential. We show that all barriers are HSC, so the scaling of d makes it satisfy SLTSC and SASC.

Theorem 3.5 (Quadratic) *Let* $K_1 = \{x \in \mathbb{R}^d : \frac{1}{2}x^\top Qx + p^\top x + l \le 0\}$ *with* $p \in \mathbb{R}^d$ *and* $0 \ne Q \in \mathbb{S}^d_+$. *Let* $K_2 = \{(x,t) \in \mathbb{R}^{d+1} : \frac{1}{2}\|x-\mu\|^2_\Sigma \le t\}$ *and* $K_3 = \{(x,t) \in \mathbb{R}^{d+1} : \|x-\mu\|_\Sigma \le t\}$ *with* $\mu \in \mathbb{R}^d$ *and* $\Sigma \in \mathbb{S}^d_{++}$. *Let* $x \in \operatorname{int}(K_i)$ *and* $h \in \mathbb{R}^{\dim(K_i)}$.

- *Ellipsoid* $\phi_{\text{ellip}}(x) = -\log(-l - p^\top x - \frac{1}{2}x^\top Qx)$ *for* K_1*:* $g = d\,\nabla^2\phi_{\text{ellip}}$ *satisfies* $\nu, \bar\nu = \mathcal{O}(d)$*, SSC when* $Q \in \mathbb{S}^d_{++}$*,* $\mathrm{D}^2 g(x)[h,h] \succeq 0$ *(so SLTSC), and SASC.*
- *Gaussian* $\phi_{\text{Gauss}}(x,t) = -\log(t - \frac{1}{2}\|x-\mu\|^2_\Sigma)$ *for* K_2*:* $g = d\,\nabla^2\phi_{\text{Gauss}}$ *satisfies* $\nu, \bar\nu = \mathcal{O}(d)$*, SSC, and* $\mathrm{D}^2 g(x,t)[h,h] \succeq 0$ *(so SLTSC), and SASC.*
- *Second-order cone* $\phi_{\text{SOC}}(x,t) = -\log(t^2 - \|x-\mu\|^2_\Sigma)$ *for* K_3*:* $g = d\,\nabla^2\phi_{\text{SOC}}$ *satisfies* $\nu, \bar\nu = \mathcal{O}(d)$*, SSC, SLTSC, and SASC.*

(3) PSD cone. Another fundamental constraint is the PSD cone $\mathbb{S}^d_+$. This convex region admits a d-self-concordant barrier $\phi_{\text{PSD}}(\cdot) = -\log\det(\cdot)$. We show that it satisfies SLTSC, while the d-scaling further guarantees SSC and ASC. In establishing ASC, we find an interesting connection to the *Gaussian orthogonal ensemble* (GOE), one of the main objects studied in random matrix theory. However, we cannot prove SASC, so we need the $d(d+1)/2$-scaling for SASC (due to HSC of ϕ_{PSD}).

Theorem 3.6 (PSD cone) *Let* $K = \mathbb{S}^d_+$*,* $X \in \operatorname{int}(K)$*, and* $H \in \mathbb{S}^d$*.Then,* $d\,\nabla^2\phi_{\text{PSD}}$ *satisfies* $\nu, \bar\nu = \mathcal{O}(d^2)$*, SSC,* $\mathrm{D}^2 g(X)[H,H] \succeq 0$ *(so SLTSC), and ASC.* $\frac{d(d+1)}{2}\nabla^2\phi_{\text{PSD}}$ *is SASC.*

(4) Entropy and ℓ_p-norm. It is sometime more convenient to introduce d many new variables.

Theorem 3.7 (Entropy and ℓ_p-norm) *Let* $K_1 = \prod_{i=1}^d \{(x_i,t_i) \in \mathbb{R}^2 : x_i \ge 0,\, t_i \ge x_i \log x_i\}$ *and* $K_2 = \prod_{i=1}^d \{(x_i,t_i) \in \mathbb{R}^2 : |x_i|^p \le t_i\}$.

- *Entropy* $\phi_{\text{ent}}(x,t) = -\sum_{i=1}^d \big(\log(t_i - x_i\log x_i) + 36\log x_i\big)$ *for* K_1*:* $g = d\,\nabla^2\phi_{\text{ent}}$ *satisfies* $\nu, \bar\nu = \mathcal{O}(d^2)$*, SSC, SLTSC, and SASC.*
- *The p-th power of* ℓ_p*-norm* $\phi_{\text{power}}(x,t) = -\sum_{i=1}^d \big(\log(t_i^{2/p} - x_i^2) + 72\log t_i\big)$ *for* K_2*:* $g = d\,\nabla^2\phi_{\text{power}}$ *satisfies* $\nu, \bar\nu = \mathcal{O}(d^2)$*, SSC, SLTSC, and SASC.*

3.5. Examples

Our theory (Theorem 3.2 and 3.3) with the study of barriers (Table 2) proposes local metrics for structured instances. GCDW with them mixes in poly-time faster than Ball walk. For fair comparison, the complexity of Ball walk refers to that of isotropic rounding[3] (see §F).

Motivating example: Let us introduce a variable for each of $\|X - B\|_F$ and $\|X - C\|_F^2$. Then our theory suggests the following barrier: $4(\phi_{\log} + d^2\phi_{\text{Gaussian}} + d^2\phi_{\text{SOC}} + d^2\phi_{\text{PSD}})$, which is $\mathcal{O}(1)$ $(m + d^3, m + d^3)$-self-concordant, SSC, LTSC, and ASC. By Theorem 3.2 with $\alpha = 0$ and $\beta = 1$ (due to ϕ_{PSD} in the potential), we need $\widetilde{\mathcal{O}}\big(d^2(m+d^3)\big)$ iterations of Dikin walk in total.

Uniform and exponential sampling: Let us first consider uniform sampling over linear constraints given by $Ax \ge b$ for $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$. Recall that for uniform sampling Ball walk mixes in $\widetilde{\mathcal{O}}(d^3)$ iterations (including isotropic rounding). On the other hand, $\widetilde{\mathcal{O}}(md)$ queries are enough for GCDW with the (m,m)-Dikin amenable metric induced by $\phi_{\log}$. This recovers the mixing time of Kannan and Narayanan (2012) *without* warmness. If we use the $(\sqrt{md}, \sqrt{md})$-Dikin-amenable Vaidya or $(d^{3/2}, d^{3/2})$-Dikin-amenable Lewis-weight metric instead,

3. For general logconcave sampling, Ball walk needs isotropic rounding, using $\widetilde{\mathcal{O}}(d^4)$ queries (Lovász and Vempala, 2006b, 2007), **after which** it mixes in $\widetilde{\mathcal{O}}(d^2)$ queries. Without rounding, it is not necessarily poly-time mixing. For uniform sampling *only*, the complexity of isotropic rounding was improved to $\widetilde{\mathcal{O}}(d^3)$ by Jia et al. (2021).

then GCDW with each metric recovers the $\widetilde{\mathcal{O}}(m^{1/2}d^{3/2})$ and $\widetilde{\mathcal{O}}(d^{5/2})$ mixing of the Vaidya walk and Approximate John walk (Chen et al., 2018) *without* warmness. For a second-order cone with linear constraints, we can use the Hessian of $2(\phi_{\text{log}} + d\phi_{\text{SOC}})$ that is $(m+d, m+d)$-Dikin-amenable, with which GCDW mixes in $\widetilde{\mathcal{O}}(d\,(m+d))$ iterations in total. Lastly, for the PSD cone with linear constraints, we can use the $(m+d^3, m+d^3)$-Dikin-amenable $2\nabla^2(\phi_{\text{log}} + d^2\phi_{\text{PSD}})$. GCDW with this needs $\widetilde{\mathcal{O}}(d^2(m+d^3))$ queries. For large m, we use the (d^3, d^3)-Dikin-amenable $2(dg_{\text{Lw}}+d^2\nabla^2\phi_{\text{PSD}})$, with which GCDW mixes in $\widetilde{\mathcal{O}}(d^5)$ iterations. In the same setting, Ball walk needs $\widetilde{\mathcal{O}}(d^6)$ queries.

For exponential sampling, GCDW requires the same number of iterations of Dikin walk for each case (i.e., polytope, second-order cone, PSD), while Ball walk needs $\widetilde{\mathcal{O}}(d^4)$ iterations for the polytope and second-order cone, and $\widetilde{\mathcal{O}}(d^8)$ iterations for the PSD cone. Detailed statements on the mixing times and efficient per-step implementation can be found in §F.3.

Uniform sampling over hyperbolic cones: Narayanan (2016) went beyond linear constraints and analyzed Dikin walk for uniform sampling over a convex region given as the intersection of (1) linear constraints, (2) a hyperbolic cone with a ν_h-SC hyperbolic barrier ϕ_h, and (3) a general convex set with a ν_s-SC barrier ϕ_s. Using $\nabla^2(\phi_{\text{log}} + d\phi_h + d^2\phi_s)$ as a local metric, this work shows that Dikin walk mixes in $\mathcal{O}\big(d\big(m+d\nu_h+(d\nu_s)^2\big)\big)$ steps from a warm start. The term $d(d\nu_s)^2$ induced by self-concordance alone is typically the largest one in the provable guarantee. Interesting results of this work arise when K is the intersection of (1) and (2). Since a hyperbolic barrier is HSC (Güler, 1997, Theorem 4.2), the d-scaling of a HSC barrier makes it SSC, SLTSC, and SASC. Also, as a ν_h-SC hyperbolic barrier is $\mathcal{O}(\nu_h)$-symmetric (implied in Güler (1997, §4)), it follows that $d\phi_h$ is $(d\nu_h, d\nu_h)$-Dikin-amenable. Hence, $\phi_{\text{log}} + d\phi_h$ induces an $(m+d\nu_h, m+d\nu_h)$-Dikin-amenable metric, and Dikin walk with this metric mixes in $\mathcal{O}(d\,(m+d\nu_h))$ iterations from a warm start by Theorem 3.1. Without warmness, Narayanan (2016) showed that Dikin walk started at $x \in K$, where $s \geq |p|/|q|$ for any chord $\overline{pq}$ of K passing through x, mixes in $\mathcal{O}\big(d(m+d\nu_h)\big[d\log\big(s(m+d\nu_h)\big) + \log\frac{1}{\varepsilon}\big]\big)$ steps. On the other hand, GCDW requires only $\mathcal{O}\big(d(m+d\nu_h)\log\frac{d(m+d\nu_h)}{\varepsilon}\big)$ iterations.

Gaussian sampling: Going forward, we consider only logarithmic barriers for linear constraints. Ball walk for general log-concave distributions mixes in $\widetilde{\mathcal{O}}(d^4)$ iterations. As per our reduction, we first replace a quadratic potential (coming from the Gaussian distribution) by a new variable, adding its epigraph to a constraint. For a polytope, one can use the $(m+d, m+d)$-Dikin-amenable $2\nabla^2(\phi_{\text{log}} + d\phi_{\text{Gauss}})$, so GCDW needs $\widetilde{\mathcal{O}}(d\,(m+d))$ iterations of Dikin walk. For the second-order cone with linear constraints, GCDW with the $(m+d, m+d)$-Dikin-amenable metric $3\nabla^2(\phi_{\text{log}} + d\phi_{\text{SOC}} + d\phi_{\text{Gauss}})$ requires $\widetilde{\mathcal{O}}(d\,(m+d))$ iterations. For the PSD cone with linear constraints, GCDW with the $(m+d^3, m+d^3)$-Dikin-amenable metric $3\nabla^2(\phi_{\text{log}} + d^2\phi_{\text{PSD}} + d^2\phi_{\text{Gauss}})$ mixes in $\widetilde{\mathcal{O}}(d^2(m+d^3))$ iterations. Ball walk is much slower, requiring $\widetilde{\mathcal{O}}(d^8)$ iterations in this setting.

Entropy sampling: For a polytope, we use the $(m+d^2, m+d^2)$-Dikin-amenable $2\nabla^2(\phi_{\text{log}} + d\phi_{\text{ent}})$ in $2d$-dimensional space. Thus, GCDW needs $\widetilde{\mathcal{O}}(d\,(m+d^2))$ iterations of Dikin walk. For the second-order cone with linear constraints, GCDW with the $(m+d^2, m+d^2)$-Dikin-amenable $3\nabla^2(\phi_{\text{log}} + d\phi_{\text{SOC}} + d\phi_{\text{ent}})$, requires in $\widetilde{\mathcal{O}}(d\,(m+d^2))$ iterations. Lastly, for the PSD cone with linear constraints, GCDW with the $(m+d^4, m+d^4)$-Dikin-amenable $3\nabla^2(\phi_{\text{log}} + d^2\phi_{\text{PSD}} + d^2\phi_{\text{ent}})$ mixes in $\widetilde{\mathcal{O}}(d^2(m+d^4))$ iterations. Ball walk mixes in $\widetilde{\mathcal{O}}(d^8)$ iterations in this setting.

3.6. Discussion

The inner loop of the sampling IPM samples from a distribution whose potential is of the form $c^{\mathsf{T}}x + \alpha\phi(x)$. Thus, the study of other non-Euclidean samplers for relatively convex and smooth potentials will be interesting future work. Next, one question unanswered is if the d^2-scaling of ϕ_{PSD} can be improved, which is mathematically interesting in its own right. The d-scaling for ASC is shown through the random matrix theory, which is challenging to extend to SASC (see Remark G.7).

4. Background and related work

Our problem (strLC) is a special case of *logconcave sampling*: sample from a distribution π with density proportional to $\exp(-V)$ for a convex function V on $\mathbb{R}^d$. This problem has spawned a long line of research in several communities, as it captures various important distributions, including uniform distributions over convex bodies and Gaussians.

A large body of recent work in machine learning and statistics makes the assumption of $0 \preceq \alpha I \preceq \nabla^2 V \preceq \beta I$ on $\mathbb{R}^d$ (i.e., α-strong convexity and β-smoothness of the potential V) for the logconcave sampling, where the strong-convexity assumption is sometimes relaxed to isoperimetry assumptions such as log-Sobolev (LSI), Poincaré (PI), and Cheeger isoperimetry (see Chewi (2024) for a survey on this topic). The guarantees provided on the mixing time of samplers under this assumption have polynomial dependence on the condition number defined as β/α (or α is replaced by the isoperimetric constant). These guarantees do not apply to constrained sampling. For example, in uniform sampling, the simplest constrained sampling problem, V is set to be a constant within the convex body and infinity outside the body, which leads to discontinuity of V and $\beta = \infty$. The sudden change of V around the boundary requires special consideration, such as small step size, use of a Metropolis filter, projection, etc., making it a more challenging problem.

Uniform sampling. Uniform sampling can be accomplished through Ball walk (Lovász and Simonovits, 1993; Kannan et al., 1997), Hit-and-Run (Smith, 1984), and In-and-Out (Kook et al., 2024), which only require access to a function proportional to the density. When a convex body $K \subset \mathbb{R}^d$ satisfies $B_r(x_0) \subset K \subset B_R(x_0)$ for some x_0, both Ball walk (Kannan et al., 1997) and Hit-and-Run (Lovász, 1999; Lovász and Vempala, 2006a) mix from a warm start in the total variation (TV) using $\widetilde{\mathcal{O}}(d^2(R/r)^2)$ queries. In-and-Out, the proximal sampler for a uniform distribution, mixes in the q-Rényi divergence using $\widetilde{\mathcal{O}}(qd^2(R/r)^2)$ queries from a warm start. Lovász and Vempala (2007) further extended these results to general logconcave distributions. These algorithms need to use a "step size" of $\Omega(1/\sqrt{d})$, and their mixing is affected by the skewed geometry of the convex body (i.e., when $R/r \gg 1$). The latter can be addressed by *rounding* the body, after which the three samplers mix in $\widetilde{\mathcal{O}}(d^2)$ steps from a warm start, due to bounds on the KLS constant by Chen (2021); Klartag (2023) and stochastic localization by Chen and Eldan (2022).

Sampling with local geometry. Ball walk uses the same radius ball for every point in the convex body. One might want to use a different radius depending on the distance to the boundary. This by itself does not work as it simply makes the current point converge to the boundary. However, replacing balls with ellipsoids whose shape changes based on the proximity to the boundary does work. Several sampling algorithms are motivated by the use of local metrics: Dikin walk (Kannan and Narayanan, 2012), Riemannian Hamiltonian Monte Carlo (RHMC), Riemannian Langevin algorithm (RLA) (Girolami and Calderhead, 2011), etc.

Which local metrics would be suitable candidates? It turns out that a suitable metric can be derived from self-concordant barriers, a concept dating back to the development of the interior-point method in convex-optimization literature (Nesterov and Nemirovskii, 1994). It is well-known that any convex body admits a d-self-concordant barrier such as universal barrier (Nesterov and Nemirovskii, 1994; Lee and Yue, 2021) and entropic barrier (Bubeck and Eldan, 2015; Chewi, 2023), but these are computationally expensive. Moreover, as noted in Laddha et al. (2020), the symmetry parameter of these general barriers is $\Omega(d^2)$ for d-dimensional bodies (even for second-order cones), and so the resulting complexity for Dikin walk on the PSD cone is $\Omega(d^2 \cdot d^4) = \Omega(d^6)$. Thus, there is a need to find barriers that are more closely aligned with the structure of sets we wish to sample.

Polytope sampling. Samplers such as Ball walk and Hit-and-Run can be used to sample polytopes, but they do not really use any special properties of polytopes.

For polytopes with m linear constraints in $\mathbb{R}^d$ ($m > d$), the first theoretical result via self-concordant barriers dates back to Kannan and Narayanan (2012) which proposed Dikin walk with the m-self-concordant logarithmic barrier and established the mixing rate of $\widetilde{\mathcal{O}}(md)$ for uniform sampling. Chen et al. (2018) revisited the idea of Vaidya (1996) using the $\mathcal{O}(m^{1/2}d^{1/2})$-self-concordant barrier, which is a hybrid of the volumetric barrier and log-barrier, for a faster IPM. They presented Dikin walk with the hybrid barrier giving an $\widetilde{\mathcal{O}}(m^{1/2}d^{3/2})$-mixing guarantee.

While the next point proposed by all these Markov chains is obtained by a Euclidean straight line step, Geodesic walk and RHMC use curves (geodesics and Hamiltonian-preserving curves respectively). Lee and Vempala (2017) and Lee and Vempala (2018) showed that for uniform sampling, Geodesic walk and RHMC with the log-barrier mix in $\widetilde{\mathcal{O}}(md^{3/4})$ and $\widetilde{\mathcal{O}}(md^{2/3})$ steps respectively. Kook et al. (2023) extended theoretical analysis of RHMC to truncated exponential distributions and showed that discretization of Hamilton's equations by practical numerical integrators maintains a fast mixing rate. Gatmiry et al. (2023) showed that RHMC with a hybrid barrier consisting of the Lewis weights and log-barrier mixes in $\widetilde{\mathcal{O}}(m^{1/3}d^{4/3})$ steps. Their proof is based on developing suitable properties and algorithmic bounds for Riemannian manifolds.

Generalization of approaches. Extending these non-Euclidean methods to general domains (e.g., the PSD cone) and to more general densities (e.g., Relatively strong convex and smooth, Gaussian) to potentially improve the complexity of the problem significantly beyond the bounds that follow from general convex body sampling, have been open research directions and motivate our paper.

Narayanan (2016) explored the first direction, analyzing Dikin walk for uniform sampling over the intersection of linear constraints, a hyperbolic cone with a hyperbolic barrier, and a general convex set with a self-concordant barrier. Our current understanding of the second direction is rather limited. A line of work has focused on the analysis of first-order non-Euclidean samplers, such as Mirror Langevin algorithm (MLA) or RLA but under strong assumptions. For example, Li et al. (2022) provided mixing-rate guarantees of MLA under the *modified self-concordance* (msc) of ϕ in the setting $\alpha\nabla^2\phi \preceq \nabla^2 f \preceq \beta\nabla^2\phi$. However, the msc is not affine-invariant, so it does not correctly capture affine-invariance of the algorithm. Ahn and Chewi (2021); Gatmiry and Vempala (2022) avoid the msc, analyzing MLA and RLA under an alternative discretization scheme that requires an exact simulation of the Brownian motion $\nabla^2\phi(X_t)^{-1/2}\,\mathrm{d}W_t$ which is not known to be achievable algorithmically. Gopi et al. (2023) proposed a non-Euclidean version of the proximal sampler based on the log-Laplace transformation (LLT) and analyzed its mixing when a potential is strongly convex and *Lipschitz* (not smooth) relatively in $\nabla^2\phi$. However, the LLT has no closed form in general.

Our study of Dikin walk for general cones and general densities provides a rather complete picture of zeroth-order non-Euclidean samplers. It also provides a general framework and improved bounds as well as a "handbook" for structured sampling.

Acknowledgments

This work was supported in part by NSF awards CCF-2007443 and CCF-2134105.

References

Kwangjun Ahn and Sinho Chewi. Efficient constrained sampling via the mirror-Langevin algorithm. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 28405–28418, 2021.

Kurt M Anstreicher. Volumetric path following algorithms for linear programming. *Mathematical Programming*, 76:245–263, 1997.

Sébastien Bubeck and Ronen Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. In *Conference on Learning Theory (COLT)*, volume 40 of *Proceedings of Machine Learning Research*, pages 279–279. PMLR, 2015.

Yuansi Chen. An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geometric and Functional Analysis (GAFA)*, 31:34–61, 2021.

Yuansi Chen and Ronen Eldan. Hit-and-run mixing via localization schemes. *arXiv preprint arXiv:2212.00297*, 2022.

Yuansi Chen, Raaz Dwivedi, Martin J Wainwright, and Bin Yu. Fast MCMC sampling algorithms on polytopes. *The Journal of Machine Learning Research (JMLR)*, 19(1):2146–2231, 2018.

Sinho Chewi. *The entropic barrier is n-self-concordant*, pages 209–222. Springer International Publishing, Cham, 2023.

Sinho Chewi. Log-concave sampling. *Book draft available at https://chewisinho.github.io*, 2024.

Ben Cousins and Santosh Vempala. Gaussian Cooling and $\mathcal{O}^*(n^3)$ algorithms for volume and Gaussian volume. *SIAM Journal on Computing (SICOMP)*, 47(3):1237–1273, 2018.

Khashayar Gatmiry and Santosh S Vempala. Convergence of the Riemannian Langevin algorithm. *arXiv preprint arXiv:2204.10818*, 2022.

Khashayar Gatmiry, Jonathan Kelner, and Santosh S Vempala. Sampling with barriers: Faster mixing via Lewis weights. *arXiv preprint arXiv:2303.00480*, 2023.

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. In *Conference on Learning Theory (COLT)*, volume 195 of *Proceedings of Machine Learning Research*, pages 2399–2439. PMLR, 2023.

Osman Güler. Hyperbolic polynomials and interior point methods for convex programming. *Mathematics of Operations Research*, 22(2):350–377, 1997.

He Jia, Aditi Laddha, Yin Tat Lee, and Santosh Vempala. Reducing isotropy and volume to KLS: an $\mathcal{O}^*(n^3\psi^2)$ volume algorithm. In *Symposium on Theory of Computing (STOC)*, pages 961–974, 2021.

Adam Tauman Kalai and Santosh Vempala. Simulated annealing for convex optimization. *Mathematics of Operations Research*, 31(2):253–266, 2006.

Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $\mathcal{O}^*(n^5)$ volume algorithm for convex bodies. *Random Structures & Algorithms (RS&A)*, 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.

Boáz Klartag. Logarithmic bounds for isoperimetry and slices of convex sets. *Ars Inveniendi Analytica*, 2023. doi: 10.15781/jsjy-0b06.

Yunbum Kook, Yin Tat Lee, Ruoqi Shen, and Santosh Vempala. Condition-number-independent convergence rate of Riemannian Hamiltonian Monte Carlo with numerical integrators. In *Conference on Learning Theory (COLT)*, volume 195 of *Proceedings of Machine Learning Research*, pages 4504–4569. PMLR, 2023.

Yunbum Kook, Santosh S Vempala, and Matthew S Zhang. In-and-out: Algorithmic diffusion for sampling convex bodies. *arXiv preprint arXiv:2405.01425*, 2024.

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

Robert Lang. A note on the measurability of convex sets. *Archiv der Mathematik*, 47:90–92, 1986.

François Le Gall. Powers of tensors and fast matrix multiplication. In *International Symposium on Symbolic and Algebraic Computation (ISSAC)*, pages 296–303, 2014.

Yin Tat Lee and Aaron Sidford. Solving linear programs with $\widetilde{\mathcal{O}}(\sqrt{\text{rank}})$ linear system solves. *arXiv preprint arXiv:1910.08033*, 2019.

Yin Tat Lee and Santosh S Vempala. Geodesic walks in polytopes. In *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 *Symposium on Theory of Computing (STOC)*, pages 1115–1121, 2018.

Yin Tat Lee and Man-Chung Yue. Universal barrier is n-self-concordant. *Mathematics of Operations Research*, 46(3):1129–1148, 2021.

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. Hit-and-run mixes fast. *Mathematical programming*, 86:443–461, 1999.

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

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

László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms (RS&A)*, 30(3):307–358, 2007.

László Lovász and Santosh S Vempala. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *Symposium on Foundations of Computer Science (FOCS)*, pages 57–68. IEEE, 2006b.

Jan R Magnus and Heinz Neudecker. The elimination matrix: some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods (SADM)*, 1(4):422–449, 1980.

Hariharan Narayanan. Randomized interior point methods for sampling and optimization. *The Annals of Applied Probability*, 26(1):597–641, 2016.

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

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

Yurii Nesterov et al. *Lectures on convex optimization*, volume 137. Springer, 2018.

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

R Tyrrell Rockafellar. *Convex analysis*, volume 11. Princeton university press, 1997.

Sushant Sachdeva and Nisheeth K Vishnoi. The mixing time of the Dikin walk in a polytope: a simple proof. *Operations Research Letters*, 44(5):630–634, 2016.

Robert L Smith. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32(6):1296–1308, 1984.

Pravin M Vaidya. A new algorithm for minimizing convex functions over convex sets. *Mathematical programming*, 73(3):291–341, 1996.

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

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Contents

Appendix A. Missing notation

Definition A.1 (Self-concordance) *For convex $K \subset \mathbb{R}^d$, let $\phi : \operatorname{int}(K) \to \mathbb{R}$ be a convex function, $g : \operatorname{int}(K) \to \mathbb{S}^d_+$ a PSD matrix function, and $\mathcal{N}^r_g(x) := \mathcal{N}\big(x, \frac{r^2}{d} g(x)^{-1}\big)$.*

- Self-concordance *(SC): A C^3-function ϕ is called a self-concordant barrier if $|\mathrm{D}^3\phi(x)[h,h,h]| \le 2\|h\|^3_{\nabla^2\phi(x)}$ for any $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^d$, and $\lim_{x\to\partial K}\phi(x) = \infty$. The first condition is equivalent to $-2\|h\|_{\nabla^2\phi(x)}\nabla^2\phi(x) \preceq \mathrm{D}^3\phi(x)[h] \preceq 2\|h\|_{\nabla^2\phi(x)}\nabla^2\phi(x)$. We call it a ν-self-concordant barrier for K if $\sup_{h\in\mathbb{R}^d}(2\langle\nabla\phi(x), h\rangle - \|h\|^2_{\nabla^2\phi(x)}) \le \nu$ for any $x \in \operatorname{int}(K)$ in addition to self-concordance. A C^1-PSD matrix function $g : \operatorname{int}(K) \to \mathbb{S}^d_+$ is called self-concordant if $-2\|h\|_{g(x)}g \preceq \mathrm{D}g(x)[h] \preceq 2\|h\|_{g(x)}g$ for any $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^d$, and there exists a self-concordant function $\phi : \operatorname{int}(K) \to \mathbb{R}$ such that $\nabla^2\phi \asymp g$ on $\operatorname{int}(K)$. We call it a ν-self-concordant barrier for K if its counterpart ϕ is ν-self-concordant.*
- Highly self-concordant function *(HSC): A C^4-function ϕ is called highly self-concordant if $|\mathrm{D}^4\phi(x)[h,h,h,h]| \le 6\|h\|^4_{\nabla^2\phi(x)}$ for any $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^d$, and $\lim_{x\to\partial K}\phi(x) = \infty$.*
- Strong self-concordance *(SSC): A SC matrix function g is called strongly self-concordant if g is PD on $\operatorname{int}(K)$ and $\|g(x)^{-1/2}\mathrm{D}g(x)[h]\, g(x)^{-1/2}\|_F \le 2\|h\|_{g(x)}$ for any $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^d$. We call a SC function ϕ strongly self-concordant if $\nabla^2\phi(x)$ is strongly self-concordant.*
- Lower trace self-concordant matrix *(LTSC): A SC matrix function g is called lower trace self-concordant if g is PD on $\operatorname{int}(K)$ and $\operatorname{Tr}\big(g(x)^{-1}\mathrm{D}^2 g(x)[h,h]\big) \ge -\|h\|^2_{g(x)}$ for any $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^d$. We call it strongly lower trace self-concordant (SLTSC) if for any PSD matrix function $\bar{g}$ on $\operatorname{int}(K)$ it holds that $\operatorname{Tr}\big(\big(\bar{g}(x)+g(x)\big)^{-1}\mathrm{D}^2 g(x)[h,h]\big) \ge -\|h\|^2_{g(x)}$ for any $x \in \operatorname{int}(K)$ and $h \in \mathbb{R}^d$.*
- Average self-concordance *(ASC): A matrix function g is called average self-concordant if for any $\varepsilon > 0$ there exists $r_\varepsilon > 0$ such that $\mathbb{P}_{z\sim\mathcal{N}^r_g(x)}\big(\|z-x\|^2_{g(z)} - \|z-x\|^2_{g(x)} \le \frac{2\varepsilon r^2}{d}\big) \ge 1-\varepsilon$ for $r \le r_\varepsilon$. We call it strongly average self-concordant (SASC) if for $\varepsilon > 0$ and any PSD matrix function $\bar{g}$ on $\operatorname{int}(K)$ it holds that $\mathbb{P}_{z\sim\mathcal{N}^r_{g+\bar{g}}(x)}\big(\|z-x\|^2_{g(z)} - \|z-x\|^2_{g(x)} \le \frac{2\varepsilon r^2}{d}\big) \ge 1-\varepsilon$ for $r \le r_\varepsilon$.*

Basics. For $n \in \mathbb{N}$, let $[n] := \{1, \cdots, n\}$. We use $f \lesssim g$ to denote $f \le cg$ for some universal constant $c > 0$. The $\widetilde{\mathcal{O}}$ complexity notation suppresses poly-logarithmic factors and dependence on error parameters. For $v \in \mathbb{R}^d$, the Euclidean norm (or ℓ_2-norm) is denoted by $\|v\|_2 \overset{\text{def}}{=} \sqrt{\sum_{i\in[d]} v_i^2}$, and the infinity norm is denoted by $\|v\|_\infty \stackrel{\text{def}}{=} \max_{i\in[d]} |v_i|$. A Gaussian distribution with mean $\mu \in \mathbb{R}^d$ and covariance $\Sigma \in \mathbb{R}^{d\times d}$ is denoted by $\mathcal{N}(\mu, \Sigma)$.

Matrices. We use $\mathbb{S}^d$ to denote the set of symmetric matrices of size $d \times d$. For $X \in \mathbb{S}^d$, we call it *positive semidefinite* (PSD) (resp. *positive definite* (PD)) if $h^\mathsf{T} X h \geq 0$ (> 0) for any $h \in \mathbb{R}^d$. We use $\mathbb{S}^d_+$ to denote the set of positive definite matrices of size $d \times d$. Note that their effective dimension is $d_s := d(d+1)/2$ due to symmetry. For a positive (semi) definite matrix X, its *square root* is denoted as $X^{\frac{1}{2}}$, and is the unique positive (semi) definite matrix satisfying $X^{\frac{1}{2}} X^{\frac{1}{2}} = X$. For $A, B \in \mathbb{S}^d$, we use $A \preceq B$ $(A \prec B)$ to indicate that $B - A$ is PSD (PD). We use $A \asymp B$ to indicate $A \precsim B$ and $B \precsim A$. For a matrix $A \in \mathbb{R}^{d\times d}$, its *trace* is denoted by $\operatorname{Tr}(A) = \sum_{i=1}^d A_{ii}$. The *operator norm* and *Frobenius norm* are denoted by $\|A\|_2 \stackrel{\text{def}}{=} \sup_{x\in\mathbb{R}^d} \|Ax\|_2/\|x\|_2$ and $\|A\|_F \stackrel{\text{def}}{=} \big(\sum_{i,j=1}^d A_{ij}^2\big)^{1/2} = \sqrt{\operatorname{Tr}(A^\mathsf{T} A)}$, respectively.

Basic operations. For $X \in \mathbb{S}^d$, its *vectorization* $\operatorname{vec}(X) \in \mathbb{R}^{d^2}$ is obtained by stacking each column of X vertically. Its symmetric vectorization $\operatorname{svec}(X) \in \mathbb{R}^{d_s}$ is obtained by stacking the lower triangular part in vertical direction. For a matrix $A \in \mathbb{R}^{d\times d}$ and vector $x \in \mathbb{R}^d$, we use $\operatorname{diag}(A)$ to denote the vector in $\mathbb{R}^d$ with $[\operatorname{diag}(A)]_i = A_{ii}$ for $i \in [d]$, $\operatorname{Diag}(A)$ to denote the diagonal matrix with $[\operatorname{Diag}(A)]_{ii} = A_{ii}$ for $i \in [d]$ and $\operatorname{Diag}(x)$ to denote the diagonal matrix in $\mathbb{R}^{d\times d}$ with $[\operatorname{Diag}(x)]_{ii} = x_i$ for $i \in [d]$.

Matrix operations. For matrices $A, B \in \mathbb{R}^{d\times d}$, their inner product is defined as the inner product of $\operatorname{vec}(A)$ and $\operatorname{vec}(B)$, denoted by $\langle A, B\rangle = \operatorname{Tr}(A^\mathsf{T} B)$. Their *Hadamard product* $A \circ B$ is the matrix of size $d \times d$ defined by $(A \circ B)_{ij} = A_{ij} B_{ij}$ (i.e., obtained by element-wise multiplication). For $A \in \mathbb{R}^{p\times q}$ and $B \in \mathbb{R}^{r\times s}$, their *Kronecker product* $A \otimes B$ is the $(pr \times qs)$ matrix defined by

$$A \otimes B = \begin{bmatrix} A_{11}B & \cdots & A_{1q}B \\ \vdots & & \vdots \\ A_{p1}B & \cdots & A_{pq}B \end{bmatrix},$$

where $A_{ij}B$ is a matrix of size $r \times s$ obtained by multiplying each entry of B by the scalar A_{ij}.

Projection matrix, Leverage score and Lewis weights. For a full-rank matrix $A \in \mathbb{R}^{m\times d}$ with $m \geq d$, we recall that $P(A) := A(A^\mathsf{T} A)^{-1} A^\mathsf{T}$ is the orthogonal projection matrix onto the column space of A. The leverage scores of A is denoted by $\sigma(A) := \operatorname{diag}\big(P(A)\big) \in \mathbb{R}^m$. We let $\Sigma(A) := \operatorname{Diag}\big(\sigma(A)\big) = \operatorname{Diag}\big(P(A)\big)$ and $P^{(2)}(A) := P(A) \circ P(A)$. The ℓ_p-Lewis weights of A is denoted by $w(A)$, the solution w to the equation $w(A) = \operatorname{diag}\big(W^{1/2-1/p} A (A^\mathsf{T} W^{1-2/p} A)^{-1} A^\mathsf{T} W^{1/2-1/p}\big) \in \mathbb{R}^m$ for $W = \operatorname{Diag}(w)$. When $m < d$ or A is not full rank, both leverage scores and Lewis weights can be generalized via the Moore-Penrose inverse in place of the inverse in the definitions.

Derivatives. For a function $f : \mathbb{R}^d \to \mathbb{R}$, let $\nabla f(x) \in \mathbb{R}^d$ denote the gradient of f at x (i.e., $[\nabla f(x)]_i = \frac{\partial f}{\partial x_i}(x)$) and $\nabla^2 f(x) \in \mathbb{R}^{d\times d}$ denote the Hessian of f at x (i.e., $[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$). For a matrix function $g : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ in x, we use $\mathrm{D}g$ and $\mathrm{D}^2 g$ to denote the third-order and fourth-order tensor defined by $[\mathrm{D}g(x)]_{ijk} = \frac{\partial [g(x)]_{ij}}{\partial x_k}$ and $[\mathrm{D}^2 g(x)]_{ijkl} = \frac{\partial^2 [g(x)]_{ij}}{\partial x_k \partial x_l}$. We use the following shorthand notation: $g'_{x,h} := \mathrm{D}g(x)[h]$ and $g''_{x,h} := \mathrm{D}^2 g(x)[h,h]$. We let $\mathrm{D}^i g(x)[h_1, \cdots, h_i] = \mathrm{D}^i g(x)[h_1 \otimes \cdots \otimes h_i]$ denote the i-th directional derivative of g at x in directions $h_1, \cdots, h_i \in \mathbb{R}^d$, i.e.,

$$\mathrm{D}^i g(x)[h_1, \cdots, h_i] = \frac{\mathrm{d}^i}{\mathrm{d}t_1 \cdots \mathrm{d}t_i} g\Big(x + \sum_{j=1}^{i} t_j h_j\Big)\Big|_{t_1, \cdots, t_i = 0}.$$

Local norm. At each point x in a set $K \subset \mathbb{R}^d$, a *local metric* g, denoted as g_x or $g(x)$, is a positive-definite inner product $g_x : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, which naturally induces the local norm as $\|v\|_{g(x)} := \sqrt{g_x(v, v)}$. We use $\|v\|_x$ to refer to $\|v\|_{g(x)}$ when the context is clear. When an ambient space has an orthogonal basis as in our setting (e.g., $\{e_1, \ldots, e_d\}$), the local metric g_x can be represented as a positive-definite matrix of size $d \times d$. With this perspective, the inner product can be written as $g_x(v, w) = v^\mathsf{T} g(x) w$. Going forward, we use $g_x = g(x)$ to denote a local metric (or positive definite matrix of size $\dim(x) \times \dim(x)$) at each point $x \in K$. The local metric g is assumed to be at least twice differentiable.

Markov chains. We use the same symbol for a distribution and its density with respect to the Lebesgue measure. Many sampling algorithms are based on *Markov chains*. A *transition kernel* $P : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to \mathbb{R}_{\geq 0}$ (or *one-step distribution*) for the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ quantifies the probability of the Markov chains transitioning from one point to another measurable set. The next-step distribution is defined by $P_x(A) := P(x, A)$, which is the probability of a step from x landing in the set A. The transition kernel characterizes the Markov chain in the sense that if a current distribution is μ, then the distribution after n steps can be expressed as $\mu P^{(n)}$, where $\mu P^{(i)}(x) := \int_{\mathbb{R}^d} P(\cdot, x)\, \mu P^{(i-1)}$ is defined recursively for $i \in [n]$ with the convention $\mu P^{(0)} = \mu$. We call π a *stationary distribution* of the Markov chain if $\pi = \pi P$. If the stationary distribution further satisfies $\int_A P(x, B)\, \pi(\mathrm{d}x) = \int_B P(x, A)\, \pi(\mathrm{d}x)$ for any two measurable subsets A, B, then the Markov chain is said to be *reversible* with respect to π.

It is expected that the Markov chain approaches the stationary distribution. We measure this with the *total variation distance* (TV-distance): for two distributions μ and π on $\mathbb{R}^d$, the TV-distance is defined as $d_{\mathrm{TV}}(\mu, \pi) \overset{\mathrm{def}}{=} \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\mu(A) - \pi(A)| = \frac{1}{2} \int_{\mathbb{R}^d} \big|\frac{\mathrm{d}\mu}{\mathrm{d}x} - \frac{\mathrm{d}\pi}{\mathrm{d}x}\big|\, \mathrm{d}x$, where the last equality holds when the two distributions admit densities with respect to the Lebesgue measure on $\mathbb{R}^d$. We also recall other probabilistic distances: when $\mu \ll \nu$,

$$\text{The chi-squared divergence } \chi^2(\mu \,\|\, \nu) \overset{\mathrm{def}}{=} \int \big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1\big)\, \mathrm{d}\nu\,,$$

$$L^2\text{-distance } \|\mu/\nu\| \overset{\mathrm{def}}{=} \int \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\, \mathrm{d}\mu = \chi^2(\mu \,\|\, \nu) + 1\,.$$

Moreover, the rate of convergence can be quantified by the *mixing time*: for an error parameter $\varepsilon \in (0, 1)$ and an initial distribution π_0, the mixing time is defined as the smallest $n \in \mathbb{N}$ such that $d_{\mathrm{TV}}(\pi_0 P^{(n)}, \pi) \leq \varepsilon$. In this paper, we consider a *lazy* Markov chain, which does not move with probability $\frac{1}{2}$ at each step, in order to avoid a uniqueness issue of a stationary distribution. Note that this change worsens the mixing time by at most a factor of 2. One of the standard tools to control progress made by each iterate is the *conductance* Φ of the Markov chain with its stationary distribution π, defined by

$$\Phi \overset{\mathrm{def}}{=} \inf_{\text{measurable } S} \frac{\int_S P(x, S^c)\, \pi(\mathrm{d}x)}{\pi(S) \wedge \pi(S^c)}\,.$$

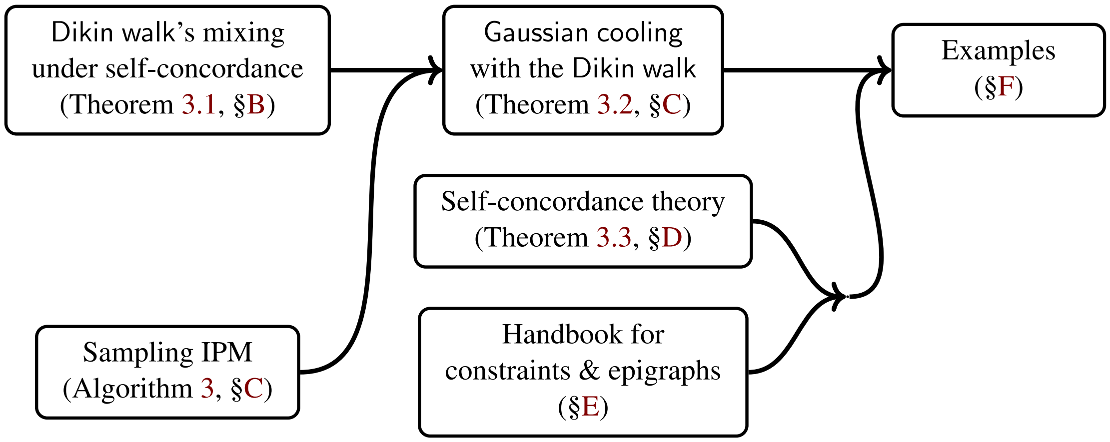


Figure A.1: Outline

Another crucial factor affecting the convergence rate is geometry of the stationary distribution π, as measured by *Cheeger isoperimetry*

$$\psi_\pi \stackrel{\text{def}}{=} \inf_{\text{measurable } S} \frac{\lim_{\delta\to 0^+} \frac{1}{\delta}\,\pi\big(\{x : 0 < d(S,x) \le \delta\}\big)}{\pi(S) \wedge \pi(S^c)}\,,$$

where $d(S,x)$ is some distance between x and the set S.

Appendix B. Mixing of the Dikin walk

We follow a standard conductance based argument (see e.g., Lovász and Simonovits (1993); Vempala (2005)). A lower bound on the conductance of a Markov chain provides an upper bound on the mixing time of the Markov chain due to the following result.

Lemma B.1 (Lovász and Simonovits (1993)) *Let π_T be the distribution obtained after T steps of a lazy reversible Markov chain of conductance at least Φ with stationary distribution π and initial distribution π_0. For $\Lambda = \mathbb{E}_{\pi_0}\big[\frac{\mathrm{d}\pi_0}{\mathrm{d}\pi}\big]$ and any $\varepsilon > 0$, we have $d_{\text{TV}}(\pi_T,\pi) \le \varepsilon + (\varepsilon^{-1}\Lambda)^{1/2}\big(1-\frac{\Phi^2}{2}\big)^T$.*

A lower bound on the conductance follows from two ingredients: **(i)** one-step coupling and **(ii)** isoperimetry. The first refers to showing that the one-step distributions of the Dikin walk from two nearby points have TV-distance bounded away from one. The second is a purely geometry property about the expansion of the target distribution. Combining these two leads to a lower bound on the conductance:

Lemma B.2 (Kook et al. (2023), Adapted from Proposition 9) *Let π be the stationary distribution of a lazy reversible Markov chain on $\mathcal{M}$ with a transition kernel P_x. Assume the isoperimetry $\psi_{\mathcal{M}}$ under the Riemannian distance $d_g(x,y)$ and the following one-step coupling: if $\|x-y\|_{g(x)} \le \Delta < 1$ for $x,y\in\mathcal{M}$, then $d_{\text{TV}}(P_x,P_y) \le 0.9$. Then the conductance Φ of the Markov chain is bounded lower by $\Omega(\psi_{\mathcal{M}}\Delta)$.*

B.1. One-step coupling and isoperimetry

Recall that a $\bar{\nu}$-Dikin-amenable metric is $\bar{\nu}$-symmetric, SSC, LTSC, and ASC. Laddha et al. (2020) was the first to attempt characterizing essential properties of g (or ϕ) that determine mixing times of Dikin walks for uniform sampling. Their framework necessitates that g satisfies $\bar{\nu}$-symmetric, SSC, convexity of $\log\det g(x)$, and $x \in \mathcal{D}_g^r(z)$ w.h.p. (where $z \sim \text{Unif}\big(\mathcal{D}_g^r(x)\big)$).

However, their framework encounters a challenge when further incorporating the work of Narayanan (2016), which analyzes the Dikin walk for uniform sampling over a convex region given as the intersection of various convex sets. The challenge arises from the difficulty of verifying the convexity of $\log\det(g_1 + g_2)$ when $\log\det g_i$ is convex for each $i = 1, 2$.

To address this challenge and succinctly characterize essential characteristics of a metric for one-step coupling, we relax the convexity of $\log\det$ to (S)LTSC and introduce the notion of ASC to account for the condition "$x \in \mathcal{D}_g^r(z)$ w.h.p.". We show that one-step coupling lemma below, one of main proof ingredients in obtaining a mixing-time guarantee of the Dikin walk, can be established under Dikin-amenability of a metric. Our characterization of a metric for achieving one-step coupling is general and unifies previous work on Dikin walks (Kannan and Narayanan, 2012; Narayanan, 2016; Chen et al., 2018; Laddha et al., 2020).

We now proceed to establish one-step coupling under the relative smoothness in ϕ.

Lemma B.3 (One-step coupling) *For convex $K \subset \mathbb{R}^d$, let $g : \text{int}(K) \to \mathbb{S}_{++}^d$ be SSC, ASC, LTSC, and $\phi : \text{int}(K) \to \mathbb{R}$ be its function counterpart. Suppose that the potential f of the target distribution π is β-relatively smooth in ϕ. Then there exist constants $s_1, s_2 > 0$ such that if $\|x-y\|_{g(x)} \le s_1 r/\sqrt{d}$ with $r = s_2\,(1 \wedge 1/\sqrt{\beta})$ for $x, y \in \text{int}(K)$, then $d_{\text{TV}}(P_x, P_y) \le \frac{3}{4} + 0.01$.*

We provide a sketch of the proof (see §G.1.1 for the full proof). A key distinction when extending beyond uniform distributions lies in establishing a lower bound for the ratio $\frac{\exp(f(x))}{\exp(f(z))}$ to ensure a high acceptance probability. To tackle this issue, we use the symmetry of the proposal distribution, claiming $\exp(f(x))/\exp(f(z)) \ge 1-\varepsilon$ at the expense of $\frac{1}{2}+\varepsilon$ probability. However, this $\frac{1}{2}+\varepsilon$ probability loss is incompatible with previous proof techniques based on the triangle inequality: for a transition kernel T and proposal kernel P, the triangle inequality leads to

$$d_{\text{TV}}(T_x, T_y) \le d_{\text{TV}}(T_x, P_x) + d_{\text{TV}}(P_x, P_y) + d_{\text{TV}}(P_y, T_y)\,,$$

and then bound the second term in the RHS by Pinsker's inequality, making it arbitrarily small by taking $r = \mathcal{O}(1)$ small enough. However, this approach yields a bound of $\frac{1}{2} + \varepsilon$ for both $d_{\text{TV}}(T_x, P_x)$ and $d_{\text{TV}}(T_y, P_y)$, making the RHS vacuous.

We instead work with the exact formula for $d_{\text{TV}}(T_x, T_y)$: for the Gaussian $p_x = \mathcal{N}(x, \frac{r^2}{d} g(x)^{-1})$,

$$R_x(z) = \frac{p_z(x)}{p_x(z)}\frac{\pi(z)}{\pi(x)} = \frac{p_z(x)}{p_x(z)}\frac{\exp(f(x))}{\exp(f(z))}, \qquad A_x(z) = \min\big(1, R_x(z)\,\mathbf{1}_K(z)\big)\,,$$

the transition kernel T_x of the Dikin walk started at x can be written as

$$T_x(dz) = \underbrace{\big(1 - \mathbb{E}_{p_x}[A_x(\cdot)]\big)}_{=:r_x}\,\delta_x(\mathrm{d}z) + A_x(z)\,p_x(\mathrm{d}z)\,.$$

Then,

$$d_{\text{TV}}(T_x, T_y) = \frac{r_x + r_y}{2} + \frac{1}{2}\int |A_x(z)\,p_x(z) - A_y(z)\,p_y(z)|\,\mathrm{d}z\,.$$

As for r_x and r_y, we bound below $\frac{p_z(x)}{p_x(z)}$ by $1-\varepsilon$ at the cost of ε-probability through SSC, LTSC, and ASC of g, following Laddha et al. (2020) with convexity of $\log\det$ replaced by LTSC. As mentioned earlier, we also deduce ${}^{\exp(f(x))}/_{\exp(f(z))} \geq 1-\varepsilon$ through the symmetry of Gaussian distributions at the cost of $\frac{1}{2}$ probability. Combining these results, we obtain upper bounds of $\frac{1}{2}+\varepsilon$ for small $\varepsilon > 0$ on r_x and r_y.

Establishing a bound of $1/4+\varepsilon$ on the second term is a more involved task. It requires the closeness of acceptance probabilities $A_x(z)$ and $A_y(z)$ as well as that of the probability densities $p_x(z)$ and $p_y(z)$. This closeness can be achieved through sophisticated conditioning on high-probability events due to ASC, SSC, and symmetry of Gaussian proposals. To be precise, define good events $G_x = \cap_{i=0,2,3} B_{x,i}^c$ and $G_y = \cap_{i=0,2,3} B_{y,i}^c$ such that $\mathbb{P}_{\mathcal{N}_g^r(x)}(G_x^c) \leq 3\varepsilon$ and $\mathbb{P}_{\mathcal{N}_g^r(y)}(G_y^c) \leq 3\varepsilon$, where

$$B_{x,0} = \{\|z-x\|_x \geq cr\} \text{ with } c \geq 1 + \frac{2}{\sqrt{d}}\log\frac{1}{\varepsilon}\,, \quad \text{(Tail bound for Gaussian)}$$

$$B_{x,1} = \{-\langle \nabla f(x), x-z\rangle \leq 0\}\,, \quad \text{(Symmetry of Gaussian)}$$

$$B_{x,2} = \{\|z-x\|_z^2 - \|z-x\|_x^2 > 2\varepsilon \frac{r^2}{d}\}\,, \quad \text{(ASC of } g\text{)}$$

$$B_{x,3} = \big\{\langle \nabla\varphi(x), z-x\rangle \leq -2\frac{r}{\sqrt{d}}\,\|g(x)^{-1/2}\nabla\varphi(x)\|_2 \log\frac{1}{\varepsilon}\big\}\,. \quad \text{(SSC \& tail bound for Gaussian)}$$

We further denote $G := G_x \cup G_y$ and a partition of G by

$$G_{x\backslash y} := G_x \backslash G_y\,, \qquad G_{x,y} := G_x \cap G_y\,, \qquad G_{y\backslash x} := G_y \backslash G_x\,.$$

Then,

$$\frac{1}{2}\int \underbrace{|A_x(z)\,p_x(z) - A_y(z)\,p_y(z)|}_{=:Q}\,\mathrm{d}z \leq 3\varepsilon + \underbrace{\frac{1}{2}\int_{G_{x\backslash y}} Q\,\mathrm{d}z}_{=:\mathcal{A}} + \underbrace{\frac{1}{2}\int_{G_{y\backslash x}} Q\,\mathrm{d}z}_{=:\mathcal{B}} + \underbrace{\frac{1}{2}\int_{G_{x,y}} Q\,\mathrm{d}z}_{=:\mathcal{C}}\,.$$

We can bound $\mathcal{A}$ and $\mathcal{B}$ by $\mathcal{O}(\varepsilon)$ by Pinsker's inequality and a well-known formula for the KL divergence between two Gaussians. As for $\mathcal{C}$, conditioning on $B_{x,1}$ and using the triangle inequality lead to

$$\mathcal{C} \leq \frac{1}{4} + 2\varepsilon + \frac{1}{2}\int_{G_x\cap G_y \cap B_{x,1}^c} \Big|\min\Big(1, \underbrace{\frac{\exp f(x)}{\exp f(z)}\,\frac{p_z(x)}{p_x(z)}}_{=:\mathsf{U}}\Big) - \min\Big(\underbrace{\frac{p_y(z)}{p_x(z)}}_{=:\mathsf{V}}, \underbrace{\frac{\exp f(y)}{\exp f(z)}\,\frac{p_z(y)}{p_x(z)}}_{=:\mathsf{W}}\Big)\Big|\,p_x(z)\,\mathrm{d}z\,.$$

The bound of $\log\mathsf{U} \geq -4\varepsilon$ was already obtained when bounding r_x. We then show that $|\log\mathsf{V}| \leq 5\varepsilon$ and $\log\mathsf{W} \geq -7\varepsilon$ conditioned on $G_x \cap G_y \cap B_{x,1}^c$ via closeness of SSC (Lemma D.6). Using these,

$$\int_{G_x\cap G_y\cap B_{x,1}^c} |1\wedge \mathsf{U} - \mathsf{V}\wedge\mathsf{W}|\,p_x(z)\,\mathrm{d}z \leq e^{5\varepsilon} - e^{4\varepsilon}\,,$$

which results in $\mathcal{C} \leq 1/4 + \mathcal{O}(\varepsilon)$. Putting the bounds on $r_x, r_y, \mathcal{A}, \mathcal{B}$, and $\mathcal{C}$ together, we conclude that the TV-distance is bounded by $3/4 + \mathcal{O}(\varepsilon)$.

Remark B.4 *We further note that $\|x-y\|_x$ can be replaced by the Riemannian distance $d_\phi(x,y)$ with the metric defined by $\nabla^2\phi$, since these two distance are within a constant factor of each other:*

Lemma B.5 (Nesterov et al. (2002), Lemma 3.1) *Let $\phi : \operatorname{int}(K) \to \mathbb{R}$ be self-concordant, and $x, y \in \operatorname{int}(K)$ with $\delta := \|x - y\|_x < 1$. Then,*

$$\delta - \frac{1}{2}\delta^2 \le d_\phi(x, y) \le -\log(1 - \delta)\,.$$

Next, we present two isoperimetric inequalities derived from distinct sources: the first comes from the symmetry of a barrier, while the second arises from strong convexity in a local metric.

Isoperimetry via barrier parameters. The first one states that isoperimetry of log-concave distributions under distance $d_g(x, y)$ (or $\|x - y\|_{g(x)}$ due to Lemma B.5) is $\Omega(1/\sqrt{\bar{\nu}})$. The following lemma is an extension of Laddha et al. (2020) from uniform distributions (over a convex body) to general log-concave distributions. We defer the proof to §G.1.2.

Lemma B.6 *For a log-concave distribution π, isoperimetry ψ_π under distance d_ϕ is $\Omega(1/\sqrt{\bar{\nu}})$.*

Isoperimetry from relative strong convexity. Another kind of isoperimetry comes from relative strong-convexity of the potential of a distribution. For a scalar $\alpha > 0$, isoperimetry of $e^{-\alpha\phi}$ on a Hessian manifold equipped with the metric $\nabla^2\phi$ is $\Omega(\sqrt{\alpha})$ if $\mathrm{D}^4\phi(x)\left[h^{\otimes 4}\right] \ge 0$ for all $x \in K$ and $h \in \mathbb{R}^d$ (Lee and Vempala, 2018, Lemma 37). Gopi et al. (2023, Lemma 9) further generalizes this to show that if ϕ is self-concordant and the potential f is α-relatively strong convex, then its isoperimetry is $\Omega(\sqrt{\alpha})$. We can adapt this lemma by restricting this to a convex set K (not necessarily bounded). See §G.1.2 for the proof.

Lemma B.7 (Gopi et al. (2023), Adapted from Lemma 9) *For a closed convex set $K \subset \mathbb{R}^d$, let a convex function $\phi : \operatorname{int}(K) \to \mathbb{R}$ be self-concordant on K, $f : \operatorname{int}(K) \to \mathbb{R}$ α-relatively strongly convex in ϕ, and π a log-concave distribution with $\pi \propto \exp(-f) \cdot \mathbf{1}_K$. For a partition $\{S_1, S_2, S_3\}$ of K and the Riemannian distance d_ϕ induced by the inner product $\langle a, b\rangle_x := a^\mathsf{T}\nabla^2\phi(x)\, b$, it holds that*

$$\pi(S_3) \gtrsim \sqrt{\alpha}\, d_\phi(S_1, S_2)\, \pi(S_1)\, \pi(S_2)\,.$$

B.2. Mixing time: Proof of Theorem 3.1

Putting all these components together, we obtain the following mixing-time bounds for the Dikin walk.

Theorem 3.1 Let $K \subset \mathbb{R}^d$ be convex and $0 \le \alpha \le \beta < \infty$.

- (Local metric) Assume that a C^1-matrix function $g : \operatorname{int}(K) \to \mathbb{S}^d_{++}$ is $\bar{\nu}$-Dikin-amenable.
- (Distribution) Let π_0 and $\pi \propto \exp(-f) \cdot \mathbf{1}_K$ be an initial and target distribution respectively, where f is α-relatively strongly convex and β-smooth in g. Let $\|\pi_0/\pi\| = \mathbb{E}_{\pi_0}\left[\frac{\mathrm{d}\pi_0}{\mathrm{d}\pi}\right]$ and P be the transition kernel of Dikin walk with the local metric g and step size $r = \mathcal{O}(1 \wedge \beta^{-1/2})$.

Then for any $\varepsilon > 0$, it holds that $d_{\mathrm{TV}}(\pi_0 P^{(T)}, \pi) \le \varepsilon$ for $T \gtrsim d(1 \vee \beta)\,(\bar{\nu} \wedge 1/\alpha) \log \frac{\|\pi_0/\pi\|}{\varepsilon}$.

Proof Lemma B.2 ensures that $\Phi \gtrsim \frac{r}{\sqrt{d}}\psi$ due to the one-step coupling in Lemma B.3. Lemma B.6 leads to $\psi \gtrsim \frac{1}{\sqrt{\bar{\nu}}}$, while Lemma B.7 implies $\psi \gtrsim \sqrt{\alpha}$ due to $\nabla^2\phi \asymp g$. Thus,

$$\Phi \gtrsim \frac{1}{\sqrt{d}}\left(\sqrt{\alpha} \vee \frac{1}{\sqrt{\bar{\nu}}}\right)\left(1 \vee \frac{1}{\sqrt{\beta}}\right),$$

and using Lemma B.1, we can enforce $d_{\mathrm{TV}}(\pi_T, \pi) \le \varepsilon$ by solving $\frac{\varepsilon}{2} + \sqrt{\frac{\Lambda}{\varepsilon/2}} e^{-T\Phi^2/2} \le \varepsilon$ for T, which results in $T \gtrsim d\,(1 \vee \beta)\,\left(\bar{\nu} \wedge \frac{1}{\alpha}\right) \log \frac{\Lambda}{\varepsilon}$. ∎

Algorithm 2: Interior-Point Method

Input: A ν-self-concordant barrier ϕ for a constraint
Output: y_λ
Denote $f_\lambda(y) := c^\mathsf{T} y + \frac{1}{\lambda}\,\phi(y)$.
`// Phase 1: Starting feasible point`
Find $y_0 = \arg\min \phi(y)$, set $\lambda = \frac{1}{6}\,\|c\|^{-1}_{[\nabla^2\phi(y_0)]^{-1}}$, and $\bar{y}_\lambda \leftarrow y_0$.
`// Phase 2: Increasing` λ `until` $\lambda \le \frac{\nu+1}{\varepsilon}$
while $\lambda \le \frac{\nu+1}{\varepsilon}$ **do**
  $\bar{y}_\lambda \leftarrow \bar{y}_\lambda - [\nabla^2 f_\lambda(\bar{y}_\lambda)]^{-1}\nabla f_\lambda(\bar{y}_\lambda)$ `// ``Opt. step'' (e.g., the Newton step)`
  $\lambda \leftarrow (1+r)\,\lambda$ with $r = \frac{1}{9\sqrt{\nu}}$. `// Increase` λ
end

Appendix C. Gaussian cooling on manifolds revisited: IPM framework for sampling

We derive a sampling analogue of the Interior-Point Method through comparison with IPM in optimization, by extending *Gaussian cooling on manifolds* introduced in Cousins and Vempala (2018); Lee and Vempala (2018). Combining the sampling IPM framework with the Dikin walk efficiently generates a warm start for a target distribution $\pi \propto e^{-f} \cdot \mathbf{1}_K$ with finite second moment.

C.1. Derivation of sampling IPM

Let us recall our setup. Let $K \subset \mathbb{R}^d$ be a closed convex set, $g : \text{int}(K) \to \mathbb{S}^d_{++}$ a $(\nu, \bar{\nu})$-SC matrix function, and $\phi : \text{int}(K) \to \mathbb{R}$ its (strictly convex) SC counterpart. We assume $\min_x \phi(x) = 0$ by considering $\phi - \min_x \phi(x)$ (here, $\arg\min \phi(x)$ can be efficiently found by the optimization IPM). We assume that f is α-relatively strongly convex and β-relatively smooth in ϕ for $0 \le \alpha \le \beta < \infty$, i.e., $0 \preceq \alpha\nabla^2\phi \preceq \nabla^2 f \preceq \beta\nabla^2\phi$ on $\text{int}(K)$. We define $\bar{f}(\cdot) := \frac{\nu}{d}\, f(\cdot)$ and $g_\phi(\cdot) := \nabla^2\phi(\cdot)$.

Interior-point method for optimization. A structural convex optimization problem is formulated as $\min_{x\in K} f(x)$, where $f : \mathbb{R}^d \to \mathbb{R}$ is a convex function, and $K \subset \mathbb{R}^d$ is a closed convex set. Also, both K and $\{(x,t) : f(x) \le t\}$ admit efficiently computable self-concordant barriers denoted by ϕ_1 and ϕ_2, respectively. We can simplify the problem by equivalently solving $\min_{x\in K,\,\{(x,t):f(x)\le t\}} t$ and in general focus on $\min_{x\in K,\,\{(x,t):f(x)\le t\}} c^\mathsf{T}(x,t)$ for a constant $c \in \mathbb{R}^{d+1}$.

IPM then regularizes $c^\mathsf{T}(x,t)$ by adding $\frac{1}{\lambda}\,\phi(x,t) = \frac{1}{\lambda}\big(\phi_1(x) + \phi_2(x,t)\big)$ for $\lambda > 0$. This regularization removes the hard constraint of $K \cap \{f(x) \le t\}$, and the resulting formulation becomes

$$\min_{y=(x,t)\in\mathbb{R}^{d+1}} f_\lambda(y) := c^\mathsf{T} y + \frac{1}{\lambda}\,\phi(y)\,,$$

where $\phi(y)$ blows up as y approaches the boundary of the constraint. For each fixed $\lambda > 0$, there exists a minimum y_λ of the convex function $f_\lambda(y)$. Intuitively, as $\lambda \to \infty$ the regularization term $\frac{1}{\lambda}\,\phi(y)$ vanishes, so y_λ converges to $\arg\min_{y\in K\cap\{f(x)\le t\}} c^\mathsf{T} y$. The path followed by $\{y_\lambda\}_{\lambda>0}$ is called the *central path*, and IPM aims to approximately follow this central path as λ increases.

To be precise, suppose that for $\lambda_1 > 0$, an approximation solution $\bar{y}_{\lambda_1}$ maintained by IPM is close enough to y_{λ_1}. Then IPM takes an optimization step (e.g., a Newton step), which takes into account the local geometry induced by the Hessian of the barrier ϕ, to find an approximate solution

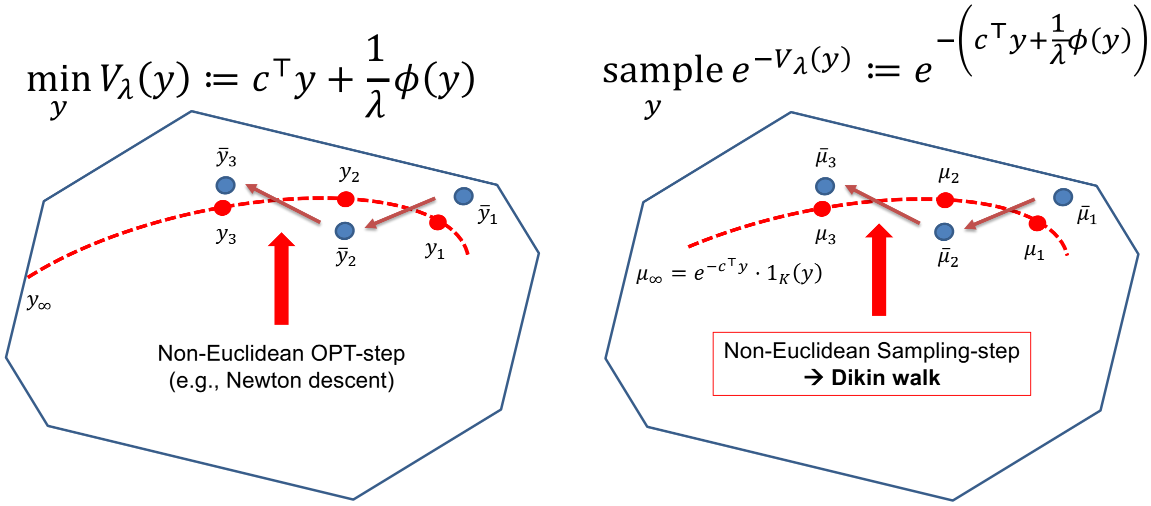


Figure C.1: Comparison between the optimization IPM and the sampling IPM.

$\bar{y}_{\lambda_2}$ when $\lambda_2 > \lambda_1$. As long as $\bar{y}_{\lambda_1}$ is sufficiently close to y_{λ_1}, this approximate solution $\bar{y}_{\lambda_1}$ serves a good starting point for the non-Euclidean optimizer, which takes $\bar{y}_{\lambda_1}$ to $\bar{y}_{\lambda_2}$. IPM alternates between increasing λ and updating $\bar{y}_\lambda$, until λ reaches ν/ε. This is described formally as Algorithm 2.

The ideas behind IPM are justified by the following theoretical guarantee: Algorithm 2 returns y in $\mathcal{O}\big(\sqrt{\nu}\,\log\big(\frac{\nu}{\varepsilon}\|c\|_{[\nabla^2\phi(y_0)]^{-1}}\big)\big)$ iterations such that $c^\top y \le c^\top y^* + \varepsilon$ for $y^* = \arg\min_{y\in K\cap\{f(x)\le t\}} c^\top y$.

Translation to sampling. Now let us adapt each step of IPM into the sampling context with the conceptual analogy between convex optimization and logconcave sampling in mind: For convex $K \subset \mathbb{R}^d$ and convex function $f : K \to \mathbb{R}$

$$\begin{aligned} &\min f(x) && \longleftrightarrow && \text{sample } x \sim \exp(-f) \\ &\text{s.t. } x \in K && && \text{s.t. } x \in K\,. \end{aligned}$$

Similar to the optimization IPM, we first replace $f(x)$ by a new variable t and add the constraint $\{f(x) \le t\}$ (which is convex due to convexity of f), resulting in the following sampling problem: sample (x,t) from a distribution with density proportional to e^{-t} subject to $x \in K$ and $\{(x,t) \in \mathbb{R}^{d+1} : f(x) \le t\}$. We note that this is indeed an equivalent sampling problem, since the x-marginal of the distribution is $\exp(-f) \cdot \mathbf{1}_K$:

$$\int_{\{(x,t)\in\mathbb{R}^{d+1}:f(x)\le t\}} \exp(-t)\cdot \mathbf{1}_K(x)\,\mathrm{d}t = \int_{f(x)}^{\infty} \exp(-t)\cdot \mathbf{1}_K(x)\,\mathrm{d}t = \exp(-f)\cdot \mathbf{1}_K\,.$$

Now assume that $K \cap \{f(x) \le t\}$ admits a barrier ϕ. Thus, this motivates our focus on sampling from distributions of the form $\exp(-c^\top y)$ subject to a convex region K with a barrier ϕ, where $y := (x,t) \in \mathbb{R}^{d+1}$ is a variable in the augmented space and $c \in \mathbb{R}^{d+1}$ is a vector.

Regularizing the potential $c^\top y$ of the distribution by adding $\frac{1}{\sigma^2}\,\phi(y)$ for some $\sigma^2 > 0$, we can ignore the hard constraint K and obtain the following formulation: for $f_{\sigma^2} := \langle c,\cdot\rangle + \frac{1}{\sigma^2}\,\phi$,

$$\text{sample } y \sim \mu_{\sigma^2} \propto \exp(-f_{\sigma^2}(y)) = \exp\Big(-\big(c^\top y + \frac{1}{\sigma^2}\,\phi(y)\big)\Big)\,,$$

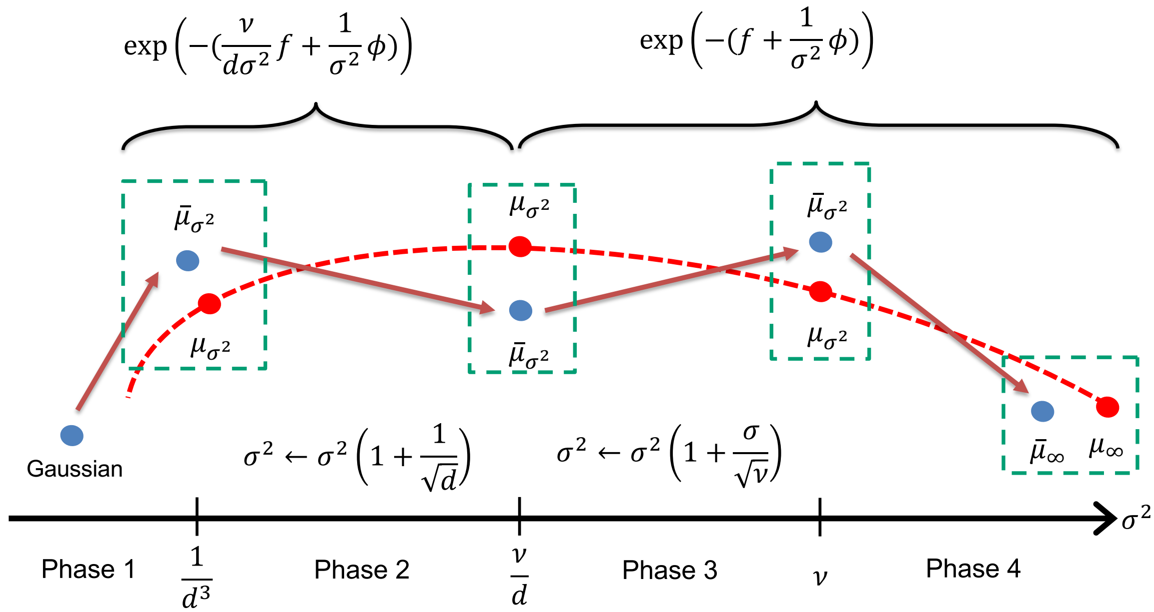


Figure C.2: We refine the derived sampling IPM to obtain the Gaussian cooling on manifolds. The red dashed line indicates a centra path of measures. The red dots are target probability measures appearing in the sampling IPM, while blue dots are probability measures given by a non-Euclidean sampler, which are approximately close to those target measures (red dots). Closeness of two dots (bounded by the green dashed boxes) is quantified by the TV-distance.

where $\phi(y)$ goes to infinity as it approaches the boundary of K. The regularization $\frac{1}{\sigma^2}\phi$ vanishes as $\sigma^2 \to \infty$, so we can expect $\mu_{\sigma^2} \to \pi \propto \exp(-\langle c, \cdot \rangle) \cdot \mathbf{1}_K$. Comparing this with the optimization IPM, the path of measures $\{\mu_{\sigma^2}\}_{\sigma^2>0}$ can be viewed as the central path in the space of measures. In an ideal scenario, a sampling IPM should closely follow this central path while increasing σ^2 along the path. To this end, we update the current distribution $\bar{\mu}_{\sigma^2}$, which is already close to μ_{σ^2} on the central path. This update should leverage a *sampling step* that is aware of the local geometry induced by $\nabla^2\phi$, which may involve running a non-Euclidean sampler such as the Dikin walk. This update brings $\bar{\mu}_{\sigma^2}$ to a new distribution $\bar{\mu}_{\sigma^2+\delta}$ that should be close to $\mu_{\sigma^2+\delta}$ for small $\delta > 0$, while $\bar{\mu}_{\sigma^2}$ serves a good starting point for this sampling step to find $\bar{\mu}_{\sigma^2+\delta}$. This procedure is repeated until σ^2 becomes large enough.

To use this sampling IPM, we further refine the framework via *Gaussian cooling on manifolds*.

Comparison with the Gaussian cooling on manifolds (GCM). Gaussian Cooling introduced in Cousins and Vempala (2018) was extended to manifolds by Lee and Vempala (2018). It was initially proposed for volume computation but shares remarkable similarities with our sampling IPM. In fact, GCM can be identified with the sampling IPM with $c = 0$ (i.e., uniform sampling) and the Riemannian Hamiltonian Monte Carlo employed for the non-Euclidean sampling step.

Returning to the comparison with the optimization IPM, we note that two algorithms use different rules for updating σ^2. While the optimization IPM updates $\sigma^2 \leftarrow \left(1 + \frac{1}{\sqrt{\nu}}\right)\sigma^2$, GCM utilizes two distinct annealing schemes:

$$\sigma^2 \leftarrow \begin{cases} \sigma^2\left(1+\frac{1}{\sqrt{d}}\right) & \text{if } \sigma^2 \le \frac{\nu}{d} \\ \sigma^2\left(1+\frac{\sigma}{\sqrt{\nu}}\right) & \text{o.w.} \end{cases}$$

While the first type of update in the small regime of σ^2 relies on a property of logconcavity of regularized distributions $\mu_{\sigma^2} \propto \exp\bigl(-\bigl(s\phi(y)+c^\mathsf{T}y\bigr)\bigr)$, the second type of update in the large regime of σ^2 is justified by concentration of measure $e^{-s\phi}$ in a thin shell for $s>0$. We note that the second type in fact accelerates the annealing process.

However, significant challenges remain for the sampling IPM. First, we need to extend this annealing scheme to exponential distributions (recall that GCM was proposed for uniform sampling). To be precise, we must account for the linear term $c^\mathsf{T}y$ (in addition to the ϕ term) when designing the annealing scheme. Unfortunately, the previous update scheme (which is applied only to ϕ part) with its analysis do not go through for this purpose.

To address this issue, we introduce a further generalization of the GCM annealing scheme in the small regime of σ^2, enabling us to leverage logconcavity of μ_{σ^2}. In the large regime of σ^2, we use the same annealing scheme but employ a different analytical approach, utilizing a functional inequality with no need to quantify the thin-shell phenomenon of μ_{σ^2}.

To discuss another remaining issue, we note that a non-Euclidean sampler used in the sampling step must have a provable mixing-time guarantee for μ_{σ^2}. We already provided this through Theorem 3.1 in §B for the Dikin walk, since the target potential is s-relatively strongly convex and s-relatively smooth in ϕ!

C.2. IPM algorithm for sampling

Our algorithm consists of four phases, where each phase updates a current distribution in a different way. For generality, we present this annealing process for a general potential f instead of linear functions, where $\alpha\nabla^2\phi \preceq \nabla^2 f \preceq \beta\nabla^2\phi$.

Going forward, we use the following notation: for $\bar{f}(x) := \frac{\nu}{d}\, f(x)$,

$$F(\sigma^2) := \begin{cases} \int_K \exp\bigl(-\frac{\bar{f}(x)+\phi(x)}{\sigma^2}\bigr)\,\mathrm{d}x & \text{if } \sigma^2 \le \frac{\nu}{d}\,, \\ \int_K \exp\bigl(-f(x) - \frac{\phi(x)}{\sigma^2}\bigr)\,\mathrm{d}x & \text{if } \frac{\nu}{d} \le \sigma^2 \le \nu\,. \end{cases}$$

We can show that $x^* = \arg\min_K(\bar{f}+\phi)$ exists in Line 3 of Algorithm 3 and that all distributions involved in the algorithm are indeed integrable. We defer the proof to §G.2.1.

Proposition C.1 *Each probability density involved in the algorithm is integrable.*

C.2.1. CLOSENESS OF DISTRIBUTIONS IN SAMPLING IPM

In this section, we demonstrate that within each phase a probability distribution $\mu_{\sigma_i^2}$ serves as a good warm start for sampling the subsequent distribution $\mu_{\sigma_{i+1}^2}$.

For the first two phases, closeness of consecutive distributions follow purely from a property of log-concave distributions, which is independent of local metrics.

Lemma C.2 (Kalai and Vempala (2006), Lemma 3.2) *For a log-concave function $g:\mathbb{R}^d\to\mathbb{R}$, the function $a \mapsto a^d \int g(x)^a\,\mathrm{d}x$ is log-concave in a.*

Algorithm 3: Interior-Point Method for sampling

Input: Target accuracy ε, local metric g, its counterpart ϕ, non-Euclidean sampler $\mathsf{NE\text{-}Sampler}(g, \varepsilon)$, target distribution $\pi \propto \exp(-f)$.

Output: x'

Let $\bar{f} = \frac{\nu}{d} f$ and $\mu_{\sigma^2} \propto \exp(-V_{\sigma^2})$, where

$$V_{\sigma^2} := \begin{cases} \frac{\bar{f}+\phi}{\sigma^2} & \text{if } \sigma^2 \le \frac{\nu}{d}\,, \\ f + \frac{1}{\sigma^2}\,\phi & \text{o.w.} \end{cases}$$

`// Phase 1: Initial distribution`

Find $x^* = \arg\min_{x \in K}(\bar{f} + \phi)$ and let $D := \mathcal{D}_g^{3\sigma_0\sqrt{d}}(x^*)$ for $\sigma_0^2 := 10^{-5}/d^3$.

Draw $x_0 \sim \mathsf{NE\text{-}Sampler}\big(g, \frac{\varepsilon}{\sqrt{d}}\big)$ with initial dist. $\mathcal{N}\big(x^*, \frac{\sigma_0^2}{1+\nu\beta/d}\, g(x^*)^{-1}\big) \cdot \mathbf{1}_D$ and target $\mu_{\sigma_0^2}$.

`// Phase 2 & 3: Annealing until` $\sigma^2 \le \nu$

while $\sigma^2 \le \nu$ **do**

Update σ^2 by

$$\sigma^2 \leftarrow \begin{cases} \sigma^2\left(1 + \frac{1}{\sqrt{d}}\right) & \text{if } \sigma^2 \le \frac{\nu}{d} \text{ (Phase 2)} \\ \sigma^2\left(1 + \frac{\sigma}{\sqrt{\nu}}\right) & \text{if } \frac{\nu}{d} \le \sigma^2 \le \nu \text{ (Phase 3)}, \end{cases}$$

Draw $x_{i+1} \sim \mathsf{NE\text{-}Sampler}\big(g, \frac{\varepsilon}{\sqrt{d}}\big)$ started at x_i with target dist. μ_{σ^2}, and increment i.

end

`// Phase 4: Sampling from` e^{-f}

Draw $x' \sim \mathsf{NE\text{-}Sampler}\big(g, \frac{\varepsilon}{\sqrt{d}}\big)$ started at x_i with target dist. π.

In Phase 1, we leverage another fundamental property of log-concave distributions. It allows us to establish that the Gaussian distribution truncated over a small Dikin ellipsoid in Phase 1 provides an $\mathcal{O}\big((\frac{\nu\beta+d}{\nu\alpha+d})^d\big)$-warm start for $\mu_{\sigma_0^2}$. Thus, the Dikin walk which has a log-dependency on the warmness parameter introduces an additional factor of d.

Lemma C.3 (Lovász and Vempala (2007), Lemma 5.16) *Let X be a random point drawn from a log-concave distribution with a density $g : \mathbb{R}^d \to \mathbb{R}$. If $\gamma \ge 2$, then*

$$\mathbb{P}\big(g(X) \le e^{-\gamma(d-1)} \max g\big) \le (\gamma\, e^{1-\gamma})^{d-1}\,.$$

Remark C.4 *If we can show that the* Dikin walk *has a* $\log\log$*-dependency through the* blocking conductance *or* Gaussian isoperimetry*, or if we utilize a non-Euclidean sampler with a double-log dependency, we can avoid the additional factor of* d*.*

We defer the proofs for closeness to §G.2.2.

Lemma C.5 (Phase 1) *Let $x^* = \arg\min_K(\bar{f} + \phi)$. For $\sigma^2 = 10^{-5}/d^3$ and $g = \nabla^2\phi$, let μ be the Gaussian distribution $\mathcal{N}\big(x^*, \frac{\sigma^2}{1+\nu\beta/d}\, g(x^*)^{-1}\big)$ truncated over $\mathcal{D}_g^{3\sigma\sqrt{d}}(x^*)$, and μ_0 the initial distribution used in Phase 2 such that $\mu_0 \propto \exp\big(-\frac{\bar{f}+\phi}{\sigma^2}\big) \cdot \mathbf{1}_K$. Then $\|\mu/\mu_0\| \lesssim \big(\frac{\nu\beta+d}{\nu\alpha+d}\big)^d$.*

In the following lemmas, we show that within each phase of our algorithm $\mu_{\sigma_i^2}$ serves as an $\mathcal{O}(1)$-warm start for the following distribution $\mu_{\sigma_{i+1}^2}$. In Phase 2, for $1/d^3 \lesssim \sigma^2 \le \nu/d$ the multiplicative update of $(1+1/\sqrt{d})$ allows us to achieve an $\mathcal{O}(1)$-warm start.

Lemma C.6 (Phase 2) *In Phase 2 (i.e., $\sigma_i^2 \le \nu/d$ with the update $\sigma_{i+1}^2 = (1+{}^1\!/\!\sqrt{d})\,\sigma_i^2$), a previous distribution μ_i serves as an $\mathcal{O}(1)$-warm start for the next distribution μ_{i+1}, i.e., $\|\mu_i/\mu_{i+1}\| = \mathcal{O}(1)$.*

In the large regime of $\nu/d \le \sigma^2 \le \nu$ during Phase 3, we leverage the Brascamp-Lieb inequality to show that the accelerated update of $(1+\sigma/\sqrt{\nu})$ ensures an $\mathcal{O}(1)$-warm start. Moreover, we employ the same technique along with a limiting argument to show that in Phase 4 the final distribution of μ_ν is an $\mathcal{O}(1)$-warm start for the target distribution π.

Lemma C.7 (Phase 3 and 4) *In Phase 3 (i.e., $\nu/d \le \sigma_i^2 \le \nu$ with the update $\sigma_{i+1}^2 = \sigma_i^2(1+\sigma_i/\sqrt{\nu})$, a previous distribution μ_i serves as an $\mathcal{O}(1)$-warm start for the next distribution μ_{i+1}, i.e., $\|\mu_i/\mu_{i+1}\| = \mathcal{O}(1)$. In Phase 4, the distribution $\mu \propto \exp\big(-(f+\phi/\nu)\big)\cdot \mathbf{1}_K$ is an $\mathcal{O}(1)$-warm start for the target distribution $\pi \propto \exp(-f)\cdot \mathbf{1}_K$.*

C.2.2. PROOF OF THEOREM 3.2

We now prove Theorem 3.2, Algorithm 3 with the Dikin walk employed for the non-Euclidean sampler.

Theorem 3.2 For convex $K \subset \mathbb{R}^d$, suppose that $g : \mathrm{int}(K) \to \mathbb{S}_{++}^d$ is $(\nu,\bar{\nu})$-Dikin-amenable and ϕ is its function counterpart such that $\min_K \phi$ exists. Gaussian cooling with Dikin walk (Algorithm 3 with Dikin walk serving as a non-Euclidean sampler) generates a sample that is ε-close to $\exp(-f)\cdot \mathbf{1}_K$ in TV-distance using $\mathcal{O}\big(d\,(d\frac{\nu\beta+d}{\nu\alpha+d} \vee \nu \vee \bar{\nu})\log\frac{d\nu}{\varepsilon}\big)$ iterations of Dikin walk with g, where a C^2-function $f : \mathrm{int}(K) \to \mathbb{R}$ satisfies $\alpha\nabla^2\phi \preceq \nabla^2 f \preceq \beta\nabla^2\phi$ on K for $0 \le \alpha \le \beta < \infty$. In particular, when $f(x) = \alpha^\mathsf{T} x$ or $c\phi(x)$ for $\alpha \in \mathbb{R}^d$ and $c \in \mathbb{R}_+$, the algorithm uses $\widetilde{\mathcal{O}}(d\,(d\vee\nu\vee\bar{\nu}))$ iterations of the Dikin walk.

Proof By Theorem 3.1, if the potential V of a target distribution satisfies $\alpha\nabla^2\phi \preceq \nabla^2 V \preceq \beta\nabla^2\phi$, the mixing time of the Dikin walk is $d\,(1\vee\beta)\,(\bar{\nu}\wedge {}^1\!/\!\alpha)\,\log\frac{\Lambda}{\varepsilon}$. Let $\bar{\kappa} = \frac{\nu\beta+d}{\nu\alpha+d}$.

- Phase 1: When a target distribution is $\exp\big(-\frac{\bar{f}+\phi}{\sigma^2}\big)$ with $\sigma^2 = 10^{-5}/d^3$,
$$d^2\big(1+\frac{\nu\beta d^{-1}+1}{\sigma^2}\big)\,\min\big(\bar{\nu}, \frac{\sigma^2}{1+\nu\alpha d^{-1}}\big)\,\log\big(\frac{\nu\beta+d}{\nu\alpha+d}\big) \le d^2\bar{\kappa}\log\bar{\kappa}\,.$$
- Phase 2 ($1/d^3 \lesssim \sigma^2 \le \nu/d$): Note that we need $\mathcal{O}^*(\sqrt{d})$-many iterations to double σ^2. Hence, in this phase the number of iterations of the Dikin walk with a target $\exp\big(-\frac{\bar{f}+\phi}{\sigma^2}\big)$ adds up to
$$d\,\big(1+\frac{\nu\beta d^{-1}+1}{\sigma^2}\big)\,\min\big(\bar{\nu}, \frac{\sigma^2}{1+\nu\alpha d^{-1}}\big)\cdot\sqrt{d} \le d^{1.5}\bar{\kappa}+\sqrt{d}\nu\,.$$
- Phase 3 ($\nu/d \le \sigma^2 \le \nu$): We need $\mathcal{O}^*\big(\frac{\sqrt{\nu}}{\sigma}\big)$-many iterations to double σ^2. Hence, in this phase the total number of iterations of the Dikin walk with a target $\exp\big(-\big(f+\frac{\phi}{\sigma^2}\big)\big)$ is
$$d\,\big(1+\beta+\frac{1}{\sigma^2}\big)\,\min\big(\bar{\nu}, \frac{1}{\alpha+\sigma^{-2}}\big)\cdot\frac{\sqrt{\nu}}{\sigma} \le \frac{d\sqrt{\nu}}{\sigma}\big(\bar{\kappa}+\sigma^2\big) \le (d^{1.5}\bar{\kappa}+\sqrt{d}\nu)\vee(d\bar{\kappa}+d\nu)$$
- Phase 4: The Dikin walk takes $\mathcal{O}(d\bar{\nu})$ iterations.

Adding up all iterations, we need $\widetilde{\mathcal{O}}(d\,(d\bar{\kappa}\vee\nu\vee\bar{\nu}))$ iterations of the Dikin walk in total. ■

Appendix D. Self-concordance theory for sampling IPM

Theorem 3.2 shows that GCDW running with a $(\nu, \bar{\nu})$-Dikin-amenable metric for exponential distributions mixes in $\widetilde{\mathcal{O}}(d \max(d, \nu, \bar{\nu}))$ iterations. Since every log-concave sampling problem can be reduced to an exponential sampling problem (as shown in (redLC)), Theorem 3.2 ensures a poly-time mixing algorithm that utilizes local geometry if we have a $(\nu, \bar{\nu})$-Dikin-amenable metric for the reduced sampling problem.

This poses a natural question of how to construct such an efficiently computable Dikin-amenable metric for structured sampling problems. Suppose that the structured sampling problems assume a Dikin-amenable metric for each constraint and epigraph of potentials. Motivated by self-concordance theory of the optimization IPM, we consider the sum of each barrier (and thus, the sum of metrics) as a candidate for the metric of the reduced sampling problem. In fact, this choice aligns seamlessly with the Dikin walk. However, obtaining a provable guarantee of the sampling IPM with the Dikin walk necessitates a comprehensive understanding not only of self-concordance but also of SSC, SLTSC, SASC, and $\bar{\nu}$-symmetry under the addition of barriers (or metrics).

In this section, we develop a "calculus" for combining metrics for multiple constraints and epigraphs, deriving the resulting theoretical guarantees (Theorem 3.3). This leads to a consistent analogy with the work of Nesterov and Nemirovskii (1994) for the optimization IPM.

D.1. Basic properties: Scaling, addition and closeness

Self-concordance is a central notion in the theory of interior-point methods for optimization (we refer interested readers to Nesterov and Nemirovskii (1994); Nesterov et al. (2018)). We first recall basic properties of self-concordance and then investigate those of strong self-concordance and lower trace self-concordance, which are crucial to our analysis.

Self-concordance.

Lemma D.1 (Nesterov (2003)) *Let f_i be a ν_i-self-concordant function on a convex set $K_i \subset \mathbb{R}^d$ for $i \in [2]$, and $\alpha > 0$ be a scalar.*
- *(Theorem 4.1.1 and 4.2.2) $f_1 + f_2$ is $(\nu_1 + \nu_2)$-self-concordant on $K_1 \cap K_2$.*
- *(Corollary 4.1.2) $g = \nabla^2(\alpha f_1)$ satisfies $\|g(x)^{-1/2} \mathrm{D}g(x)[h]\, g(x)^{-1/2}\|_2 \leq \frac{2}{\sqrt{\alpha}} \|h\|_{g(x)}$ for $x \in \mathrm{int}(K_1 \cap K_2)$ and $h \in \mathbb{R}^d$.*
- *If f_1 is a ν-self-concordant, then cf_1 is $(c\nu)$-self-concordant for $c > 1$.*

We can extend this to self-concordant matrices as well.

Lemma D.2 *Let $g_i : \mathrm{int}(K_i) \to \mathbb{S}_+^d$ be a PSD matrix function on a convex set $K_i \subset \mathbb{R}^d$ for $i \in [2]$, and $\alpha > 0$ be a scalar.*
- *$g_1 + g_2$ is $(\nu_1 + \nu_2)$-self-concordant on $K_1 \cap K_2$.*
- *If g_1 is self-concordant, then αg_1 satisfies $\mathrm{D}(\alpha g_1)(x)[h] \preceq \frac{2}{\sqrt{\alpha}} \|h\|_{\alpha g_1} (\alpha g_1)$ for $x \in \mathrm{int}(K_1 \cap K_2)$ and $h \in \mathbb{R}^d$.*
- *If g_1 is ν-self-concordant, then cg_1 is $(c\nu)$-self-concordant for $c > 1$.*

Proof Let ϕ_i be a ν_i-self-concordant function counterpart of g_i on K_i for $i \in [2]$. Then for $x \in \mathrm{int}(K_1 \cap K_2)$ and $h \in \mathbb{R}^d$

$$\mathrm{D}(g_1 + g_2)(x)[h] \preceq 2\left(\|h\|_{g_1} g_1 + \|h\|_{g_2} g_2\right) \preceq 2\left(\|h\|_{g_1+g_2} g_1 + \|h\|_{g_1+g_2} g_2\right) = 2\, \|h\|_{g_1+g_2} (g_1 + g_2)\,.$$

Clearly, $\phi_1 + \phi_2$ is a function counterpart of $g_1 + g_2$. Thus, $g_1 + g_2$ is a $(\nu_1 + \nu_2)$-self-concordant matrix function on $K_1 \cap K_2$.

For $c > 1$, if g_1 is self-concordant, then $\mathrm{D}(cg_1)(x)[h] \preceq \frac{2}{\sqrt{c}} \|h\|_{cg_1}(cg_1) \preceq 2\|h\|_{cg_1}(cg_1)$, and its function counterpart $c\phi_1$ is $(c\nu)$-self-concordant by Lemma D.2. Hence, cg_1 is $(c\nu)$-self-concordant. ■

The following lemma ensures that the Dikin walk stays inside the convex body. This lemma was proven only for self-concordant function in Nesterov et al. (2018, Theorem 5.1.5), but it can be straightforwardly extended to self-concordant matrices as well.

Lemma D.3 *$\mathcal{D}_g^1(x) \subset K$ for a convex set K and self-concordant matrix function g on K.*

Proof Consider a matrix function g_ε from $\mathrm{int}(K)$ to $\mathbb{S}_{++}^d$ defined by $g_\varepsilon(x) := g(x) + \varepsilon I$. It is self-concordant with a function counterpart $\phi(x) + \frac{\varepsilon}{2}\|x\|^2$, where $\phi : \mathrm{int}(K) \to \mathbb{R}$ is a function counterpart of g. For fixed $x \in \mathrm{int}(K)$ and $h \in \mathbb{R}^d$, let us define a function defined by $\psi(t) := \bigl(h^\mathsf{T} g_\varepsilon(x+th)\, h\bigr)^{-1/2}$ for any feasible t. Then,

$$\psi'(t) = -\frac{\mathrm{D}g_\varepsilon(x+th)[h^{\otimes 3}]}{2\|h\|^3_{g_\varepsilon(x+th)}}\,,$$

and the definition of self-concordance leads to $|\psi'(t)| \le 1$. This function can be defined on the interval $\bigl(-\psi(0), \psi(0)\bigr)$ due to $\psi(t) \ge \psi(0) - |t|$ (see Nesterov et al. (2018, Corollary 5.14)). This implies that K contains the set

$$\bigl\{x + th : |t| \le \psi(0) = \|h\|^{-1}_{g_\varepsilon(x)}\bigr\} = \{x + th : \|th\|_{g_\varepsilon(x)} \le 1\}\,.$$

By sending $\varepsilon \to 0$, the claim follows. ■

The following lemma states that self-concordant metrics are similar for nearby points.

Lemma D.4 (Nesterov (2003), Theorem 4.1.6) *Given any self-concordant matrix function g on $K \subset \mathbb{R}^d$ and $x, y \in K$ with $\|x - y\|_{g(x)} < 1$, we have*

$$(1 - \|x-y\|_{g(x)})^2 g(x) \preceq g(y) \preceq (1 - \|x-y\|_{g(x)})^{-2} g(x)\,.$$

Strong self-concordance. Strong self-concordance is additive up to a constant scaling. See §G.3.1 for the proof.

Lemma D.5 *If g_i is a SSC matrix function on K_i for $i \in [2]$, then $2\,(g_1 + g_2)$ is strongly self-concordant on $K_1 \cap K_2$.*

Note that if we add k-many strongly self-concordant metrics, then we need the scaling of $2^{\log_2 k} = k$. We remark that the factor of 2 above might be redundant. Next, we recall an analogue of Lemma D.4 for strong self-concordance.

Lemma D.6 (Laddha et al. (2020), Lemma 1.2) *Given a strongly self-concordant matrix function g on K, and any $x, y \in K$ with $\|x - y\|_{g(x)} < 1$,*

$$\|g(x)^{-1/2}\bigl(g(y) - g(x)\bigr)\, g(x)^{-1/2}\|_F \le (1 - \|x-y\|_{g(x)})^{-2}\|x-y\|_{g(x)}\,.$$

Symmetry. Recall that $\bar{\nu}$-symmetry requires two-sided inclusion: the first part is $\mathcal{D}_g^1(x) \subset K \cap (2x - K)$, and the second part is $K \cap (2x - K) \subset \mathcal{D}_g^{\sqrt{\bar{\nu}}}(x)$. The first part immediately follows when a metric is induced by a self-concordant function.

Lemma D.7 *If ϕ is a self-concordant function on K, then $\mathcal{D}_g^1(x) \subset K \cap (2x - K)$ for $g = \nabla^2 \phi$ and $x \in K$.*

Proof Lemma D.3 ensures that $y \in K$ whenever $y \in \mathcal{D}_g^1(x)$. Then $2x - y \in \mathcal{D}_g^1(x)$ and thus $2x - y \in K$. It implies that $y \in 2x - K$. ■

When a metric is induced by a self-concordant barrier with a barrier parameter ν, it holds that $\bar{\nu} = \mathcal{O}(\nu^2)$.

Lemma D.8 *For a self-concordant barrier ϕ with a barrier parameter ν on K and $g = \nabla^2 \phi$, it follows that $\bar{\nu} = \mathcal{O}(\nu^2)$.*

Proof By Nesterov (2003, Theorem 4.2.5), for any $x, y \in K$ with $\nabla \phi(x) \cdot (y - x) \geq 0$ it follows that $\|y - x\|_{g(x)} \leq \nu + 2\sqrt{\nu}$. Now, let $x \in K$ and $y \in K \cap (2x - K)$. The latter implies that $y - x = x - z$ for some $z \in K$.

If $\nabla \phi(x) \cdot (y - x) \geq 0$, then $\|y - x\|_{g(x)} \leq \nu + 2\sqrt{\nu}$. If $\nabla \phi(x) \cdot (y - x) < 0$, then $\nabla \phi(x) \cdot (z - x) > 0$ and thus $\|y - x\|_{g(x)} = \|z - x\|_{g(x)} \leq \nu + 2\sqrt{\nu}$. From these two cases, it holds in general that $\|y - x\|_{g(x)} \leq \nu + 2\sqrt{\nu}$ and thus $K \cap (2x - K) \subset \mathcal{D}_g^{\nu + 2\sqrt{\nu}}(x)$. By Lemma D.7, $\mathcal{D}_g^1(x) \subset K \cap (2x - K)$ and thus $\bar{\nu} = \mathcal{O}(\nu^2)$. ■

For affine constraints $Ax \geq b$, the first inclusion above has a useful equivalent description as follows:

Lemma D.9 *Let $x \in K = \{Ax > b\}$. It holds that $y \in K \cap (2x - K)$ if and only if $\|A_x(y - x)\|_\infty \leq 1$.*

Proof For $y \in K$, we have $Ay > b$ and thus $s_x = Ax - b > A(x - y)$ (elementwise inequality). As $s_x > 0$, we have $A_x(x - y) \leq 1$. When $y \in (2x - K)$, we can write $y = 2x - z$ for some $z \in K$. Note that

$$A(x - y) = A(z - x) > b - Ax = -s_x ,$$

and thus $A_x(x - y) \geq -1$. Therefore, $\|A_x(y - x)\|_\infty \leq 1$. ■

Lemma D.10 *For $\alpha \geq 1$, if g is $\bar{\nu}$-symmetric, then αg is $\alpha \bar{\nu}$-symmetric.*

Symmetry parameters and self-concordance parameters are additive.

Lemma D.11 *If a PSD matrix function g_i is $\bar{\nu}_i$-symmetric on K_i for $i \in [2]$, then $g_1 + g_2$ is $(\bar{\nu}_1 + \bar{\nu}_2)$-symmetric on $K_1 \cap K_2$.*

Proof For $g := g_1 + g_2$, let $y \in \mathcal{D}_g^1(x)$. It implies $y \in \mathcal{D}_{g_1}^1(x) \cap \mathcal{D}_{g_2}^1(x)$ and so $y \in K_i \cap (2x - K_i)$. Due to $\cap_i \big(K_i \cap (2x - K_i)\big) = K \cap (2x - K)$, we have $y \in K \cap (2x - K)$ and so $\mathcal{D}_g^1(x) \subset K \cap (2x - K)$.

Now let $y \in K \cap (2x - K)$. It is obvious that $y \in K_i \cap (2x - K_i)$ for $i = 1, 2$, and thus

$$(y - x)^\mathsf{T} g_1(x)(y - x) \leq \nu_1 , \qquad \text{and} \qquad (y - x)^\mathsf{T} g_2(x)(y - x) \leq \nu_2 .$$

By adding up these two, it follows that $\|y - x\|_{g(x)}^2 \leq \nu_1 + \nu_2$. ■

Lower trace self-concordance. It readily follows that (strongly) LTSC holds under scaling by a scalar greater than or equal to 1.

We provide a useful sufficient condition under which the sum of PSD matrix functions is LTSC.

Lemma D.12 *For a PSD matrix function g_i on K_i, let $g := \sum_i g_i$ be PD on $\bigcap_i K_i$. If g_i is SLTSC on K_i, then g is LTSC on $\bigcap_i K_i$.*

We note that $\mathrm{D}^2 g_i(x)[h,h] \succeq 0$ is a stronger condition than $\mathrm{Tr}\big(g(x)^{-1}\mathrm{D}^2 g_i(x)[h,h]\big) \geq -\|h\|_{g_i(x)}^2$. Thus, a special case of the lemma is that if $\mathrm{D}^2 g_1[h,h] \succeq 0$ and $\mathrm{D}^2 g_2[h,h] \succeq 0$, then $g_1 + g_2$ is LTSC. Note that this condition is *additive*.

We also find that highly self-concordance is a handy sufficient condition by which one can establish strongly lower trace self-concordance, whose proof is deferred to §G.3.2.

Lemma D.13 *For $K \subset \mathbb{R}^d$, let $\bar{g} : \mathrm{int}(K) \to \mathbb{S}_+^d$ be a HSC matrix function, and define another matrix function by $g := d\bar{g}$ on K. Then g is SLTSC.*

Average self-concordance. Just as (S)LTSC, (S)ASC still holds under scaling by a scalar greater than or equal to 1. Also, the definition of SASC immediately leads to the following additive condition:

Lemma D.14 *For a PSD matrix function g_i on K_i for $i \in [m]$, let $m = \mathcal{O}(1)$ and $g := \sum_{i=1}^m g_i$ be PD on $\bigcap_i K_i$. If g_i is SASC on K_i, then g is ASC on $\bigcap_i K_i$.*

Proof Fix $\varepsilon > 0$. Each g_i invokes $r_i(\varepsilon)$ such that if $r \leq r_i(\varepsilon/m)$, then

$$\mathbb{P}_z\Big(\|z-x\|_{g_i(x)}^2 - \|z-x\|_{g_i(x)}^2 \leq \frac{2\varepsilon}{m}\frac{r^2}{d}\Big) \geq 1 - \frac{\varepsilon}{m}\,.$$

If $r \leq \bar{r}(\varepsilon) := \min_i r_i(\varepsilon/m)$, then the union bound leads to ASC of $\sum g_i$ on $\bigcap_i K_i$. ■

When does SASC hold? It is implied in Narayanan (2016) that HSC implies SASC. For completeness, we provide the proof in §G.3.3.

Lemma D.15 (HSC to SASC) *If $\phi : \mathrm{int}(K) \to \mathbb{R}$ is HSC, then $d\phi$ is SASC.*

D.2. Collapse and embedding: Lifting up SSC, SLTSC, and SASC

SSC, (S)LTSC, (S)ASC of a local metric do not carry over into an extended space in the reduced sampling problem. For instance, SSC assumes the invertibility of the local metric, which may become singular in the extended space. To address this challenge, we introduce the notions of *collapse* and *embedding*, based on which we can pass those properties from the original sampling problem to the reduced problem.

Definition D.16 *Let K and K' be convex sets in $\mathbb{R}^d$ and in $\mathbb{R}^m$ with $d \leq m$, respectively. Let $g : \mathrm{int}(K) \to \mathbb{S}_+^d$ be a PSD matrix function.*

- *We say g is* collapsed *onto a linear subspace $W \subset \mathbb{R}^d$ if $\langle u, v\rangle_{g(x)} = \langle P_W u, P_W v\rangle_{g(x)}$ for any $x \in \mathrm{int}(K)$ and $u, v \in \mathbb{R}^d$ where P_W is the orthogonal projection onto W.*
  - *In other words, for an orthonormal basis $\{u_1, \dots, u_k\}$ of W there exists the PSD matrix function $g_W : \mathrm{int}(K) \to \mathbb{S}_+^k$ such that $\langle e_i, e_j\rangle_{g_W(x)} = \langle u_i, u_j\rangle_{g(x)}$ for $i, j \in [k]$ (i.e., $g_W(x) = U^\mathsf{T} g(x) U$ where the columns of $U \in \mathbb{R}^{d\times k}$ are $\{u_1, \dots, u_k\}$).*

- *For g collapsed onto W, we say*
  - *g is PD along W if g_W is PD. In other words, $\|h\|_{g(x)} = 0$ implies $h \perp W$.*
  - *g is SSC along W if g is a self-concordant matrix function and $g_W \succ 0$ satisfies*

$$\|g_W(x)^{-1/2} \mathrm{D} g_W(x)[h] \, g_W(x)^{-1/2}\|_F \leq 2\|h\|_g \quad \textit{for any } x \in \mathrm{int}(K) \textit{ and } h \in \mathbb{R}^d\,.$$

- Embedding *$\bar{g}$ of g into K'*
  - *Let $P : \mathbb{R}^m \to \mathbb{R}^d$ be the projection onto the set of coordinates appearing in the variable x of g. The embedding of g onto K' is a PSD matrix function $\bar{g}(y) : \mathrm{int}(K') \to \mathbb{S}_+^m$ such that $\langle u, v \rangle_{\bar{g}(y)} = \langle Pu, Pv \rangle_{g(P(y))}$.*

We note that these notions are well-defined independently of the choice of an orthonormal basis of W. The proof can be found in §G.3.4.

Proposition D.17 *Let $K \subset \mathbb{R}^d$ be convex and $g : \mathrm{int}(K) \to \mathbb{S}_+^d$ a PSD matrix function collapsed onto a subspace $W \subset \mathbb{R}^d$. Then PD and SSC along W are well-defined (i.e., the condition for each property holds for any orthonormal basis of W).*

Affine transformation. Using these notions, we can make it precise that an inverse mapping of affine transformations preserves SSC. We begin with a barrier version and subsequently extend it to a matrix-function version. The detailed proofs are deferred to §G.3.5.

Lemma D.18 *Let $T : \mathbb{R}^d \to \mathbb{R}^m$ be a linear operator defined by $T(x) = Ax + b$ for $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$. Let $\phi(y) : \mathrm{int}(K) \subset \mathbb{R}^m \to \mathbb{R}$ be a self-concordant barrier for K and define $\psi(x) := \phi(T(x)) = \phi(y)$ on $\bar{K} := T^{-1}K \subset \mathbb{R}^d$.*
- *If ϕ is a $(\nu, \bar{\nu})$-self-concordant barrier for K, so is ψ for $\bar{K}$.*
- *If $\mathrm{D}^4\phi(y)[v, v] \succeq 0$ for $y \in \mathrm{int}(K)$ and $v \in \mathbb{R}^m$, then $\mathrm{D}^4\psi(x)[u, u] \succeq 0$ for $x \in \mathrm{int}(\bar{K})$ and $u \in \mathbb{R}^d$.*
- *If ϕ is HSC, so is ψ.*

Lemma D.19 *Let $g : \mathrm{int}(K) \subset \mathbb{R}^m \to \mathbb{S}_+^m$ be a self-concordant matrix function and $T(x) = Ax+b$ with $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$ be a linear operator. Let $\bar{g}(x) := A^\mathsf{T} g(Tx)A$ be a PSD matrix function from $\bar{K} := T^{-1}K \subset \mathbb{R}^d$ to $\mathbb{S}_+^d$.*
- *If g is $(\nu, \bar{\nu})$-self-concordant barrier, so is $\bar{g}$ for $\bar{K}$.*
- *If g is SSC, then $\bar{g}$ is SSC along $W = \mathrm{row}(A)$.*
- *If $\mathrm{D}^2 g(y)[h, h] \succeq 0$ for $y \in \mathrm{int}(K)$ and $h \in \mathbb{R}^m$, then $\mathrm{D}^2\bar{g}(x)[\bar{h}, \bar{h}] \succeq 0$ for $x \in \mathrm{int}(\bar{K})$ and $\bar{h} \in \mathbb{R}^d$.*
- *If A is invertible and g is SLTSC, then $\bar{g}$ is SLTSC.*
- *If A is invertible and g is SASC, then $\bar{g}$ is SASC.*

Intuitively, embedding should not affect self-concordance and symmetry parameter, which is indeed the case.

Corollary D.20 *Assume $K \subset \mathbb{R}^d$ is embeddable into $K' \subset \mathbb{R}^m$. If $g : \mathrm{int}(K) \to \mathbb{S}_+^d$ is a $(\nu, \bar{\nu})$-self-concordant matrix function, then its embedding $\bar{g} : \mathrm{int}(K') \to \mathbb{S}_+^m$ is a $(\nu, \bar{\nu})$-self-concordant matrix function.*

Proof Since K can be embedded into K', there exists a projection matrix $P \in \{0,1\}^{d\times m}$ such that $\bar{g}(y) = P^{\mathsf{T}} g(Py)P$ with $x = Py \in \operatorname{int}(K)$ and $y \in \operatorname{int}(K')$. As we can view $\bar{g}$ as a matrix function induced by the inverse of the linear map $x = Py$, Lemma D.19 shows that $\bar{g}$ is a $(\nu,\bar{\nu})$-self-concordant matrix function for $K' = P^{-1}K$. ■

Lifting up SSC, SLTSC, and SASC via embedding. In reduction to the exponential sampling problem, passing essential properties (e.g., SSC, SLTSC, and SASC) of metrics from the original space to the extended space poses technical issues. We address these issues in the following two lemmas, whose proofs are deferred to §G.3.6.

As mentioned earlier, SSC in the original space does not automatically imply SSC for its embedding $\bar{g}$, as SSC assumes invertibility. However, there is a useful method for extending SSC from the original space to the extended space.

Lemma D.21 *For convex $K \subset \mathbb{R}^d$, let $g : \operatorname{int}(K) \to \mathbb{S}^d_+$ be SSC along a subspace $W \subset \mathbb{R}^d$, and assume K is embeddable into convex $K' \subset \mathbb{R}^m$ with $m \geq d$. For the embedding $\bar{g} : \operatorname{int}(K') \to \mathbb{S}^m_+$ of g into K', it holds that $\bar{g} + \varepsilon I_m$ is SSC on K' for any $\varepsilon > 0$.*

When extending SLTSC and SASC to the embedding space, we encounter a different subtlety. The conditions in SLTSC and SASC of $\bar{g}$ consider every PSD matrix functions g' such that $\bar{g} + g'$ is invertible in the extended space $\bar{K}$. However, the embedding $\bar{g}$ of g is collapsed onto the subspace corresponding to the original space K. As SLTSC and SASC convolve $\bar{g}$ and g' by considering $(\bar{g} + g')^{-1}$ in their formulations, it is not evident whether SLTSC and SASC can be transferred to the extended space $\bar{K}$ from the original space K. However, by employing with Schur complements we can show that these properties can indeed carry over into the extended space.

Lemma D.22 *For convex $K \subset \mathbb{R}^d$, let $g : \operatorname{int}(K) \to \mathbb{S}^d_+$ is SLTSC, and assume K is embeddable into convex $K' \subset \mathbb{R}^m$ with $m \geq d$. Then its embedding $\bar{g} : \operatorname{int}(K') \to \mathbb{S}^m_+$ is also SLTSC. The same is true for SASC.*

D.3. Proof of Theorem 3.3

With our understanding of how to combine properties of barriers for constraints and epigraphs, we are prepared to prove Theorem 3.3. Let us revisit the reduced sampling problem in (redLC):

$$\text{sample } y \sim \tilde{\pi} \propto \exp\Big(-\langle(\underbrace{0,\dots,0}_{d \text{ times}}, \underbrace{1,\dots,1}_{I \text{ times}}), \cdot\rangle\Big)$$

$$\text{s.t. } y \in \bigcap_{i=1}^{I} E_i \cap \underbrace{\bigcap_{j=1}^{J} K_j}_{=::K} =: K' \,,$$

where $E_i := \big\{y = (x, t_1, \dots, t_I) \in \mathbb{R}^{d+I} : f_i(x) \leq y_{d+i}\big\}$ for a proper closed convex function f_i and $i \in [I]$, and $K_j := \big\{y = (x, t_1, \dots, t_I) \in \mathbb{R}^{d+I} : h_j(x) \leq 0\big\}$ for a closed convex function h_j and $j \in [J]$, and K has non-empty interior.

We begin with a useful geometric property of K'.

Lemma D.23 *If the original sampling problem (strLC) is well-defined, then the extended convex region K' in the reduced sampling problem (redLC) has non-empty interior and no straight line.*

Proof Since f_i and h_j are closed and convex, K' is convex and closed. Since f_i is continuous on $\text{int}(K)$ due to convexity (see Rockafellar (1997, Theorem 10.1)), its epigraph has non-empty interior. Thus, K' has non-empty interior.

Since K' is closed and convex, it can be written as $K' = \bigcap_i H_i$ where $H_i = \{x : a_i^\mathsf{T} x \geq b_i\}$ is any halfspace containing K'. Suppose K' contains a straight line $\ell := \{p + th : t \in \mathbb{R}\}$ for some $p, h \in \mathbb{R}^d$. Then $\ell \subset H_i$ for any i, and thus ℓ must be parallel to any halfspace H_i (i.e., $h \perp a_i$).

Fix $y \in \text{int}(K')$. The translated line ℓ_y of ℓ containing y is still included in H_i for all i. As $y \in \text{int}(K')$, the distance from y to ∂H_i is bounded lower by $\delta > 0$ for all i. Hence, $\ell_y + B_\delta$ is fully contained in H_i and thus in K'.

Clearly, integration of the exponential distribution along the fiber ℓ_y is infinite. Since K' contains the cylinder $\ell_y + B_\delta$, integration of the exponential distribution over K' must be infinite, leading to contradiction. ∎

The following is the extension of Nesterov et al. (2018, Theorem 5.1.6) to self-concordant matrix functions, which implies invertibility of Dikin-amenable metrics in the reduced problem.

Lemma D.24 *For convex $K \subset \mathbb{R}^d$ containing no straight line, a self-concordant matrix function $g : \text{int}(K) \to \mathbb{S}_+^d$ is non-degenerate on K.*

Proof Suppose $\|h\|_{g(x)} = 0$ for some $0 \neq h \in \mathbb{R}^d$ and $x \in \text{int}(K)$. Clearly, the line $x + th$ for $t \in \mathbb{R}$ is contained in $\mathcal{D}_g^1(x)$. As $\mathcal{D}_g^1(x) \subset K$ due to Lemma D.3, it implies that K contains a straight line $x + th$, which leads to contradiction. ∎

Theorem 3.3 In the reduced problem (redLC), assume the following:

- For $i \in [I]$, the epigraph E_i admits a PSD matrix function $g_i^e(x, t_i)$ (or $g_i^e(x, t_{i,1}, \ldots, t_{i,d})$) that is a $(\nu_i, \bar{\nu}_i)$-SC barrier, SSC along some subspace, SLTSC, and SASC.
- For $j \in [J]$, the constraint K_j admits a PSD matrix function $g_j^c(x)$ that is a $(\eta_j, \bar{\eta}_j)$-SC barrier, SSC along some subspace, SLTSC, and SASC.

For appropriate projections π_i^e and π^c, a matrix function g on $y \in \text{int}(K')$ defined by

$$\langle u, v \rangle_{g(y)} := (I + J) \Big(\sum_{i=1}^{I} \langle \pi_i^e u, \pi_i^e v \rangle_{g_i^e(\pi_i^e(y))} + \sum_{j=1}^{J} \langle \pi^c u, \pi^c v \rangle_{g_j^c(\pi^c(y))} \Big) \quad \text{for } u, v \in \mathbb{R}^d$$

is $\big((I + J)(\sum_{i=1}^I \nu_i + \sum_{j=1}^J \eta_j),\ (I + J)(\sum_{i=1}^I \bar{\nu}_i + \sum_{j=1}^J \bar{\eta}_j)\big)$-Dikin-amenable on K'.

Proof First of all, $\bar{g}_i^e$ is $(\nu_i, \bar{\nu}_i)$-self-concordant (Corollary D.20), and SLTSC and SASC on K' (Lemma D.22). For fixed $\varepsilon > 0$, $\bar{g}_i^e + \varepsilon I$ is SSC by Lemma D.21. We can make similar arguments for $\bar{g}_j^c$ regarding self-concordance, symmetry, SLTSC, SASC, and SSC. Hence, $g + (I + J)\varepsilon I$ is SSC by Lemma D.5. Since g is self-concordant on K' by Lemma D.2 and K' contains no straight line, g is PD by Lemma D.24. Sending ε to 0, we can obtain SSC of g. LTSC and ASC of g follows from Lemma D.12 and D.14. The symmetry parameter of g follows from Lemma D.11. ∎

D.4. Direct product

For $i \in [m]$ and domain $E_i \subset \mathbb{R}^{d_i}$, let $g_i(x_i) : \operatorname{int}(E_i) \to \mathbb{S}^{d_i}_{++}$ be a self-concordant matrix. For $l := \sum_i d_i$ and $E := \prod_i E_i$, we define a self-concordant matrix g on $E \subset \mathbb{R}^l$ with block diagonals being g_i. To be precise, we can write

$$g(x) = g(x_1, \ldots, x_m) := \sum_i \bar{g}_i(x)\,,$$

where $\bar{g}_i : \mathbb{R}^l \to \mathbb{S}^l_+$ is a matrix function whose entry is all zero but the i-th block diagonal being g_i.

When handling the direct product of domains, it is common for each domain to have an $\mathcal{O}(1)$-dimension. In such cases, scaling the barriers by dimension worsens mixing time at most constant factors while making the barriers SSC and SLTSC. We defer the proofs to §G.3.7.

Lemma D.25 (SSC under direct product) *For open $E_i \subset \mathbb{R}^{d_i}$, let $g_i : E_i \to \mathbb{S}^{d_i}_{++}$ be SC. Then $g := \sum d_i \bar{g}_i$ defined on $\prod E_i$ is SSC.*

Lemma D.26 (SLTSC under direct product) *For open $E_i \subset \mathbb{R}^{d_i}$, let $g_i : E_i \to \mathbb{S}^{d_i}_{++}$ be HSC. Then $g := \sum d_i \bar{g}_i$ defined on $\prod E_i$ is SLTSC.*

D.5. Inverse images under non-linear mappings

Nesterov and Nemirovskii (1994) introduced the notion of *compatibility* with a convex domain while constructing a self-concordant barrier for a wider class of structured constraints. We generalize this notion to the fourth order, by which we can easily construct a SSC, SLTSC, and SASC barrier. For a convex cone K, we use $a \le_K b$ to denote $b - a \in K$.

Definition D.27 (Compatibility) *Let $\beta, \gamma \ge 0$. Let K be a convex cone in $\mathbb{R}^m$ and Γ be a closed convex domain in $\mathbb{R}^d$. A mapping $\mathcal{A} : \operatorname{int}(\Gamma) \to \mathbb{R}^m$ of class C^4 is called (K, β, γ)-compatible with the domain Γ if*

- *$\mathcal{A}$ is concave with respect to K. That is, $t\mathcal{A}(x) + (1-t)\,\mathcal{A}(y) \le_K \mathcal{A}(tx + (1-t)\,y)$ for all $t \in [0,1]$ and $x, y \in \operatorname{int}(\Gamma)$. Equivalently, $-\mathrm{D}^2\mathcal{A}(x)[h,h] \in K$ for any $x \in \operatorname{int}(\Gamma)$ and $h \in \mathbb{R}^m$.*
- *For any $x \in \operatorname{int}(\Gamma)$, $y \in \Gamma \cap (2x - \Gamma)$, and $h = y - x$, it holds that*

$$\beta \mathrm{D}^2\mathcal{A}(x)[h,h] \le_K \mathrm{D}^3\mathcal{A}(x)[h,h,h] \le_K -\beta \mathrm{D}^2\mathcal{A}(x)[h,h]\,,$$
$$\gamma \mathrm{D}^2\mathcal{A}(x)[h,h] \le_K \mathrm{D}^4\mathcal{A}(x)[h,h,h,h] \le_K -\gamma \mathrm{D}^2\mathcal{A}(x)[h,h]\,.$$

Example 1 *An affine mapping is $(\{0\}, 0, 0)$-compatible with any closed convex domain. We note that a function that is $(\mathbb{R}_+, \beta, \gamma)$-compatible with $\mathbb{R}_+$ is a C^4-smooth concave real-valued function $f : (0, \infty) \to \mathbb{R}$ such that for any $t > 0$,*

$$|f'''(t)| \le -\frac{\beta}{t}\, f''(t) \quad \textit{and} \quad |f^{(4)}(t)| \le -\frac{\gamma}{t^2}\, f''(t)\,.$$

- *Let $0 < p \le 1$. Then the function of $f(t) = t^p$ is $(\mathbb{R}_+, 2-p, (2-p)\,(3-p))$-compatible with $\mathbb{R}_+$.*
- *$f(t) = \log t$ is $(\mathbb{R}_+, 2, 6)$-compatible with $\mathbb{R}_+$.*

The following lemma is an extension of Nesterov and Nemirovskii (1994, Lemma 5.1.3) to our fourth-order compatibility.

Lemma D.28 *Let K, K_1, K_2 be convex cones in $\mathbb{R}^m, \mathbb{R}^{m_1}, \mathbb{R}^{m_2}$ respectively.*

- *If $\mathcal{A} : \mathrm{int}(\Gamma) \to \mathbb{R}^m$ is (K, β, γ)-compatible with Γ and $K \subset K'$ is a closed convex cone in $\mathbb{R}^m$, then $\mathcal{A}$ is (K', β, γ)-compatible with Γ.*
- *If $\mathcal{A}_i : \mathrm{int}(\Gamma_i) \to \mathbb{R}^{m_i}$ is (K_i, β_i, γ_i)-compatible with Γ_i for $i = 1, 2$, then $\mathcal{A} : \mathrm{int}(\Gamma_1 \times \Gamma_2) \to \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ mapping $(x, y) \to (\mathcal{A}_1(x), \mathcal{A}_2(y))$ is $(K_1 \times K_2, \max(\beta_1, \beta_2), \max(\gamma_1, \gamma_2))$-compatible with $\Gamma_1 \times \Gamma_2$.*

We now introduce a main result in this section (see §G.3.8). To begin with, we recall that for a closed convex domain $G \subset \mathbb{R}^d$ the *recessive cone* $R(G)$ of G is $\{h \in \mathbb{R}^d : x + th \in G \text{ for all } x \in G \text{ and } t > 0\}$.

Lemma D.29 *Let G be a closed convex domain in $\mathbb{R}^m$, F be a highly θ-self-concordant barrier for G, Γ be a closed convex domain in $\mathbb{R}^d$, and Π be a highly ν-self-concordant barrier for Γ. Let $\mathcal{A}$ be a (K, β, γ)-compatible with Γ, where K is a ray contained in the recessive cone $R(G)$. Assume that $\mathcal{A}(\mathrm{int}(\Gamma)) \cap G \neq \emptyset$.*

- *The set $G^+ = \overline{\mathrm{int}(\Gamma) \cap \mathcal{A}^{-1}\big(\mathrm{int}(G)\big)}$ is a closed convex domain in $\mathbb{R}^d$.*
- *For $\delta = \max(\beta, \gamma, 2)$, the function $\Psi(x) = F(\mathcal{A}(x)) + \delta^2\, \Pi(x)$ is a $(\theta + \delta^2\nu)$-self-concordant barrier for G^+.*
- *Ψ is highly self-concordant.*

Using this result, we can obtain a useful tool in establishing lower trace self-concordance of a barrier for the direct product of structured sets.

Lemma D.30 *Let f be a C^4 concave function on $\{t > 0\}$ such that $|f'''(t)| \leq \frac{\beta}{t}\,|f''(t)|$ and $|f^{(4)}(t)| \leq \frac{\gamma}{t^2}\,|f''(t)|$ for $t > 0$. Then the function*

$$F(t, x) = -\log\big(f(t) - x\big) - \max(4, \beta^2, \gamma^2)\,\log t$$

is a highly $(1 + \max(4, \beta^2, \gamma^2))$-self-concordant barrier for the two dimensional convex domain

$$G_f = \overline{\{(t, x) \in \mathbb{R}^2 : t > 0,\, x \leq f(t)\}}\,.$$

Proof From the discussion in Example 1, the map $f(t) : (0, \infty) \to \mathbb{R}$ is $(\mathbb{R}_+, \beta, \gamma)$-compatible with $\mathbb{R}_+$. Clearly, the identity map from $\mathbb{R}$ to $\mathbb{R}$ is $(\{0\}, 0, 0)$-compatible with $\mathbb{R}$. Hence by Lemma D.28-(2) implies that the map $\mathcal{A} : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}^2$ defined by $\mathcal{A}(t, x) = (f(t), x)$ is $(\{0\} \times \mathbb{R}_+, \beta, \gamma)$-compatible with $\mathbb{R}_+ \times \mathbb{R}$.

Now observe that G_f can be written as $\mathcal{A}^{-1}\big(\{(t, x) : x \leq t\}\big)$ and that $K = \{0\} \times \mathbb{R}_+$ is a ray contained in the recessive cone $R(G)$ for $G := \{(t, x) : x \leq t\}$. By applying Lemma D.29 to the highly 1-self-concordant barriers $F(t, x) = -\log(t - x)$ for G and $\Phi(t, x) = -\log t$ for $\mathbb{R}_+ \times \mathbb{R}$, it follows that F is is a highly $(1 + \max(4, \beta^2, \gamma^2))$-self-concordant barrier for G_f. ■

We can prove a similar result for a convex f as follows:

Lemma D.31 *Let f be a C^4 convex function on $\{x > 0\}$ such that $|f'''(x)| \leq \frac{\beta}{x}\, f''(x)$ and $|f^{(4)}(x)| \leq \frac{\gamma}{x^2}\, f''(x)$ for $x > 0$. Then the function*

$$F(t, x) = -\log\big(t - f(x)\big) - \max(4, \beta^2, \gamma^2)\,\log x$$

is a highly $(1 + \max(4, \beta^2, \gamma^2))$-self-concordant barrier for the two dimensional convex domain

$$G_f = \overline{\{(t, x) \in \mathbb{R}^2 : x > 0,\, t \geq f(x)\}}\,.$$

Its proof follows from applying Lemma D.30 to the image of G_f under the map $(t, x) \to (-x, t)$.

Appendix E. Structured densities and constraint families

In order to obtain a mixing-time bound of the Dikin walk for the reduced problem, a concrete understanding of properties and parameters of barriers for K_i and K_j is essential. To this end, we revisit self-concordant barriers for structured convex constraints and level sets, examining the required scaling factors which ensure those properties.

E.1. Linear constraints

Consider a set of linear constraints: $K = \{x \in \mathbb{R}^d : Ax \geq b\}$ for $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$, where A has no all-zero rows. We use $s_x := Ax - b$ to denote the slack at x, and $A_x := S_x^{-1}A$ to denote the constraints normalized by the slack, where $S_x := \mathrm{Diag}(s_x)$ is the diagonalization of the slack.

We now introduce three barriers (and metrics) for handling the linear constraints.

Logarithmic barrier. The logarithmic barrier $\phi_{\log}(x) := -\sum_{i=1}^m \log(a_i^\mathsf{T} x - b_i)$ is the simplest self-concordant barrier for linear constraints. We refer readers to §I.1 for gentle introduction to the log-barriers. As seen below, we demonstrate that the metric induced by the logarithmic barrier has $\nu, \bar{\nu} = m$ and requires no scaling to achieve SSC, SLTSC, and SASC.

Lemma E.1 (Logarithmic barrier) *For a closed convex $K = \{x \in \mathbb{R}^d : Ax \geq b\}$ with $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$, let $\phi_{\log}(x) = -\sum_{i=1}^m \log(a_i^\mathsf{T} x - b_i)$ and define $g(x) := \nabla^2 \phi_{\log}(x) = A_x^\mathsf{T} A_x$.*
- $\nu = m$ (Nesterov and Nemirovskii, 1994).
- *SSC along* $\mathrm{row}(A)$ *and* $\bar{\nu} = m$ (Lemma E.5).
- $\mathrm{D}^2 g(x)[h,h] \succeq 0$ *for any* $h \in \mathbb{R}^d$ *(so SLTSC)* (Claim I.1).
- *SASC* (Lemma E.10).

Vaidya metric. In sampling over a polytope K, the number m of constraints is assumed to be greater than the ambient dimension d. Given that the mixing time of the Dikin walk for uniform sampling is $\widetilde{\mathcal{O}}(d\bar{\nu}) = \widetilde{\mathcal{O}}(dm)$, a larger m leads to a worse mixing time. Is there a self-concordant barrier that has a better dependence on m for its self-concordance and symmetry parameters, without compromising SSC, SLTSC, and SASC?

Let us recall the *leverage score* first and move onto such improved self-concordant barriers. For a full-rank matrix $A \in \mathbb{R}^{m\times d}$ with $m \geq d$, we recall that $P(A) = A(A^\mathsf{T} A)^{-1}A^\mathsf{T}$ is the orthogonal projection matrix onto the column space of A, and the leverage scores of A is $\sigma(A) = \mathrm{diag}(P(A)) \in \mathbb{R}^m$. We let $\Sigma(A) := \mathrm{Diag}(\sigma(A)) = \mathrm{Diag}(P(A))$ and $P^{(2)}(A) = P(A) \circ P(A)$, where $P(A) \circ P(A)$ is the Hadamard product of size $d \times d$ defined by $(P(A) \circ P(A))_{ij} = [P(A)]_{ij}^2$.

Vaidya (1996) introduced the *volumetric barrier* for K defined by

$$\phi_{\mathrm{vol}} = \frac{1}{2}\log\det(\nabla^2 \phi_{\log}) = \frac{1}{2}\log\det(A_x^\mathsf{T} A_x)\,.$$

Then the Hessian of ϕ_{vol} can be written as

$$\nabla^2 \phi_{\mathrm{vol}} = A_x^\mathsf{T}(3\Sigma_x - 2P_x^{(2)})A_x\,,$$

where $\Sigma_x = \mathrm{Diag}(\sigma(A_x))$ is the diagonalized leverage scores, and this Hessian satisfies

$$A_x^\mathsf{T} \Sigma_x A_x \preceq \nabla^2 \phi_{\mathrm{vol}}(x) \preceq 3A_x^\mathsf{T} \Sigma_x A_x\,.$$

We refer readers to §I.2 for details. In other words, the *approximate* volumetric metric $A_x^\mathsf{T}\Sigma_x A_x$ serves as an $\mathcal{O}(1)$-approximation of the local metric $\nabla^2\phi_{\text{vol}}$ (i.e., $A_x^\mathsf{T}\Sigma_x A_x \asymp \nabla^2\phi_{\text{vol}}(x)$). We find in Lemma E.5 that the local metric $40\sqrt{m}A_x^\mathsf{T}\Sigma_x A_x$ is SSC with $\nu, \bar\nu = \mathcal{O}(\sqrt{md})$, but in some regime of d this parameter leads to worse mixing of the Dikin walk. In the same paper, Vaidya (1996) introduced a *regularized* volumetric metric by adding $\mathcal{O}(\nabla^2\phi_{\log})$, which we call the *Vaidya metric*:

$$g(x) := \sqrt{\frac{m}{d}}\, A_x^\mathsf{T}\big(\Sigma_x + \frac{d}{m}I_m\big)A_x\,.$$

Note that $g(x) \asymp \nabla^2\big(\sqrt{\frac{m}{d}}\big(\phi_{\text{vol}} + \frac{d}{m}\phi_{\log}\big)\big)$. We show that the Vaidya metric is also SSC, SLTSC, and SASC without additional scaling, while it has a better ν and $\bar\nu$ than the logarithmic barrier.

Lemma E.2 (Vaidya metric) *For a closed convex $K = \{x \in \mathbb{R}^d : Ax \geq b\}$ with $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$, let $g(x) = \sqrt{\frac{m}{d}}A_x^\mathsf{T}\big(\Sigma_x + \frac{d}{m}I_m\big)A_x$.*
- $\nu = \mathcal{O}(\sqrt{md})$ (Anstreicher, 1997, Theorem 5.2).
- *SSC and* $\bar\nu = \mathcal{O}(\sqrt{md})$ (Lemma E.5).
- *SLTSC (Lemma E.6) and SASC* (Lemma E.11).

Lewis weights metric. Self-concordance and symmetry parameters of $\mathcal{O}(\sqrt{md})$ is certainly better than $\mathcal{O}(m)$, but can we even achieve an $\mathcal{O}(d\log^{\mathcal{O}(1)} m)$ bound on those parameters?

Let us recall the ℓ_p*-Lewis weights*. The ℓ_p-Lewis weight of A is denoted by $w(A)$, the solution w to the equation $w(A) = \operatorname{diag}\big(W^{1/2-1/p}A(A^\mathsf{T}W^{1-2/p}A)^{-1}A^\mathsf{T}W^{1/2-1/p}\big) \in \mathbb{R}^m$ for $W := \operatorname{Diag}(w)$. For $W_x = \operatorname{Diag}(w(A_x))$ and $p \geq 2$, the Lewis weight barrier function is defined by

$$\phi_{\mathsf{Lw}}(x) := \log\det(A_x^\mathsf{T}W_x^{1-2/p}A_x)\,.$$

Note that the leverage score and volumetric barrier can be recovered as a special case of the Lewis weight and barrier by setting $p = 2$. As done for the Vaidya metric, it is natural to consider the Lewis weight metric with $p = \Theta(\log^{\mathcal{O}(1)} m)$, defined as

$$g(x) := \mathcal{O}(\log^{\mathcal{O}(1)} m)\, A_x^\mathsf{T}W_xA_x\,.$$

In fact, this metric serves as an $\mathcal{O}(\log^{\mathcal{O}(1)} m)$-approximation of $\nabla^2\phi_{\mathsf{Lw}}$, as demonstrated in the following relation proven in Lee and Sidford (2019, Lemma 31):

$$A_x^\mathsf{T}\Sigma_xA_x \preceq \nabla^2\phi_{\mathsf{Lw}} \preceq (1+p)\,A_x^\mathsf{T}\Sigma_xA_x\,.$$

Ignoring the logarithmic factors we have $\nabla^2\phi_{\mathsf{Lw}} \asymp g$. Notably, the Lewis-weight metric needs an additional $\sqrt{d}$-scaling for SLTSC and SASC, unlike the logarithmic barrier and Vaidya metric. Hence, when combining this with other metrics, one should use $\sqrt{d}g$, which leads to $\nu, \bar\nu = \mathcal{O}(d^{3/2}\log^{\mathcal{O}(1)} m)$.

Lemma E.3 (Lewis weight metric) *For a closed convex $K = \{x \in \mathbb{R}^d : Ax \geq b\}$ with $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$, let $g(x) = \mathcal{O}(\log^{\mathcal{O}(1)} m)\,A_x^\mathsf{T}W_xA_x$.*
- $\nu = \mathcal{O}(d\log^5 m)$ (Lee and Sidford, 2019, Theorem 30).
- *SSC and* $\bar\nu = \mathcal{O}(d\log^{\mathcal{O}(1)} m)$ (Lemma E.5).
- $\sqrt{d}g$ *is SLTSC* (Lemma E.7) *and SASC* (Lemma E.12).

E.1.1. ANALYSIS OF SELF-CONCORDANT METRICS FOR LINEAR CONSTRAINTS

Strong self-concordance and symmetry. We defer the proofs of two lemmas below to §G.4.1. We study SSC and symmetry of the metrics of the form $A_x^\top D_x A_x$ in Lemma E.4, where $D_x \in \mathbb{R}^{m\times m}$ is a diagonal matrix used to address the constraints of the form $Ax \geq b$ for $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$. Specifically, we relate the notions of SSC and symmetry to well-studied terms in the field of optimization, namely $\max_i [\sigma(\sqrt{D_x}A_x)]_i/[D_x]_{ii}$ and $\|\mathrm{D}D_x[h]\|^2_{D_x^{-1}}$.

Lemma E.4 *For a diagonal $D_x \in \mathbb{S}^m_+$, let $g(x) = A_x^\top D_x A_x \in \mathbb{R}^{d\times d}$ on $\mathrm{int}(K)$.*
- *For any PSD matrix function g' such that $g' + g$ is invertible on the domain,*

$$\|(g'(x)+g(x))^{-1/2}\mathrm{D}g(x)[h]\,(g'(x)+g(x))^{-1/2}\|_F^2 \leq 4\max_i \frac{[\sigma(\sqrt{D_x}A_x)]_i}{[D_x]_{ii}} \cdot \Big(\|h\|^2_{g(x)} + \sum_{i=1}^m \frac{(\mathrm{D}D_x[h])^2_{ii}}{[D_x]_{ii}}\Big)\,.$$

- $\max_{h:\|h\|_{g(x)}=1} \|A_x h\|_\infty = \big(\max_{i\in[m]} \frac{[\sigma(\sqrt{D_x}A_x)]_i}{[D_x]_{ii}}\big)^{1/2}$.
- $K \cap (2x - K) \subset \mathcal{D}_g^{\sqrt{\mathrm{Tr}(D_x)}}(x)$.

Then for each metric we refer to existing bounds on these terms, estimating the smallest possible scaling required for SSC and symmetry.

Lemma E.5 (Strong self-concordance and symmetry) *Let $A \in \mathbb{R}^{m\times d}$, $\Sigma_x = \mathrm{Diag}(\sigma(A_x)) \in \mathbb{R}^{m\times m}$, and $W_x = \mathrm{Diag}(w_x) \in \mathbb{R}^{m\times m}$ for the ℓ_p-Lewis weight w_x with $p = \mathcal{O}(\log m)$.*
- *Logarithmic metric: $g(x) = A_x^\top A_x$ with $D_x = I_m$ is SSC along $\mathrm{row}(A)$ with $\bar\nu = m$.*
- *Approximate volumetric metric: $g(x) = 40\sqrt{m}A_x^\top \Sigma_x A_x$ with $D_x = 40\sqrt{m}\Sigma_x$ is SSC with $\bar\nu = \mathcal{O}(\sqrt{m}d)$.*
- *Vaidya metric: $g(x) = 22\sqrt{\frac{m}{d}}A_x^\top\big(\Sigma_x + \frac{d}{m}I_m\big)A_x$ with $D_x = 22\sqrt{\frac{m}{d}}\big(\Sigma_x + \frac{d}{m}I_m\big)$ is SSC with $\bar\nu = \mathcal{O}(\sqrt{md})$.*
- *Lewis-weight metric: $\exists$ positive constants c_1 and c_2 such that $g(x) = c_1(\log m)^{c_2}A_x^\top W_x A_x$ is SSC and $\bar\nu$-symmetric with $\bar\nu = \mathcal{O}^*(d)$.*

Strongly lower trace self-concordance We show SLTSC of the Vaidya and Lewis-weight metric. Let g_2 be either Vaidya or Lewis-weight metric, and g_1 be an arbitrary PSD matrix function on K such that $g = g_1 + g_2$ is PD on $\mathrm{int}(K)$. Ensuring (S)LTSC of the Vaidya or Lewis-weight metrics is challenging, as $\mathrm{D}^2 g_2[h,h] \succeq 0$ is difficult to verify due to complicated expressions for $\mathrm{D}^2\Sigma_x[h,h]$ and $\mathrm{D}^2 W_x[h,h]$. As for the Vaidya metric, we compute higher-order derivatives of leverage scores and other pertinent matrices in Lemma I.4, finding succinct formulas by using algebraic properties of the Hadamard product. We then show SLTSC of g_2 using these results (see §G.4.2 for the proof):

Lemma E.6 (SLTSC of Vaidya) $\mathrm{Tr}\big(g^{-1}\mathrm{D}^2 g_2(x)[h,h]\big) \geq -\|h\|^2_{g_2(x)}/2$ *for the Vaidya metric g_2.*

For the Lewis-weights metric, analysis is more involved due to numerous terms appearing in $\mathrm{D}^2 W_x[h,h]$. In order to avoid dealing with each of the terms, we employ existing bounds on derivatives of W_x and other relevant matrices in §I.3. This approach significantly simplifies the computation but comes at the cost of an additional scaling of $\sqrt{d}$, which as far as we can tell might be unavoidable. We refer readers to §G.4.3 for the proof.

Lemma E.7 (SLTSC of Lewis-weight) $\mathrm{Tr}\big(g(x)^{-1}\mathrm{D}^2 g_2(x)[h,h]\big) \geq -\|h\|^2_{g_2(x)}$, *where $g_2(x) = cA_x^\top W_x A_x$ with $c = c_1(\log m)^{c_2}\sqrt{d}$ for some constants $c_1, c_2 > 0$.*

Strongly average self-concordance. Typically, (S)ASC is the most challenging property to verify, often requiring involved analysis in order to establish it *without* additional scalings. Since the three metrics are HSC (e.g., see Lemma I.10 for Lewis-weight metrics), scaling by d leads to SASC by Lemma D.15. However, for linear constraints one can still achieve SASC without scaling (or with a smaller scaling) through more sophisticated concentration techniques.

To sketch this idea, we recall that SASC requires showing that for small enough r

$$\|z-x\|_{g(z)}^2 - \|z-x\|_{g(x)}^2 \le 2\varepsilon\frac{r^2}{d}\,.$$

Taylor's expansion of $\|z-x\|_{g(z)}^2$ at $z=x$ up to second-order necessitates bounds on

$$\mathrm{D}g(x)[(z-x)^{\otimes 3}] = \frac{r^3}{d^{3/2}}\mathrm{D}g(x)[h^{\otimes 3}] \qquad \text{and} \qquad \mathrm{D}g(x')[(z-x)^{\otimes 4}] = \frac{r^4}{d^2}\mathrm{D}^2 g(x')[h^{\otimes 4}]\,,$$

for some $x' \in [x,z]$ and $h\sim\mathcal{N}(0,I_d)$. Observe that the first-order term $P(h) := \frac{r^3}{d^{3/2}}\mathrm{D}g(x)[h^{\otimes 3}]$ is a Gaussian polynomial in h, and this is where we can invoke the following concentration phenomenon:

Lemma E.8 (Concentration of Gaussian polynomials) *For $d\ge 1$, let $P:\mathbb{R}^d\to\mathbb{R}$ be a polynomial of degree n. For any $t\ge (2e)^{n/2}$,*

$$\mathbb{P}_{h\sim\mathcal{N}(0,I_d)}\Big[|P(h)| \ge t\sqrt{\mathbb{E}[P(h)^2]}\Big] \le \exp\big(-\frac{n}{2e}\,t^{2/n}\big)\,.$$

This concentration inequality necessitates bounding $\mathbb{E}[P(h)^2]$, and this is where Stein's lemma comes into play:

Lemma E.9 *For $h=(h_1,\dots,h_d)\sim\mathcal{N}(0,I_d)$, it holds that $\mathbb{E}[h_i f(h)] = \mathbb{E}[\partial_i f(h)]$.*

Unlike the first-order term, the second-order term is *not* a Gaussian polynomial due to x' depending on z. To address this issue, we derive an upper bound (in absolute value) of the quadratic form. Using coordinate-wise closeness of slacks, leverage scores, and Lewis weights at two nearby points, we replace every value estimated at z by those at x, removing dependence on z in the quadratic bound. The resulting quadratic bound is now a Gaussian polynomial, so we follow the same proof approach as with the first-order term.

This approach was used by Sachdeva and Vishnoi (2016) for ASC of log-barriers and by Chen et al. (2018) for that of Vaidya and Lewis-weight metrics. We further extend this approach to achieve SASC of those metrics, going beyond ASC.

Lemma E.10 (SASC of logarithmic barrier) *$g(x) = \nabla^2\phi_{\log}(x) = A_x^\top A_x$ is SASC.*

See §G.4.4 for the proof.

Lemma E.11 (SASC of Vaidya metric) *$g(x) = \mathcal{O}\big(\sqrt{\frac{m}{d}}\big)\,A_x^\top(\Sigma_x + \frac{d}{m}I_m)A_x$ is SASC.*

See §G.4.4 for the proof.

Lemma E.12 (SASC of Lewis-weight metric) *There exists constants c_1 and c_2 such that $g(x) = c_1\sqrt{d}\log^{c_2} m\, A_x^\top W_x A_x = \mathcal{O}^*(\sqrt{d})\,A_x^\top W_x A_x$ is SASC.*

See §G.4.4 for the proof.

E.2. Quadratic potentials and constraints

Suppose that in (redLC) we have either $f_i(x),\ h_j(x) = \|x-\mu\|_\Sigma^2$ or $\frac{1}{2}x^\mathsf{T}Qx+p^\mathsf{T}x+l$ for $\mu, p \in \mathbb{R}^d$, $\Sigma \in \mathbb{S}_{++}^d$, and $0 \neq Q \in \mathbb{S}_+^d$.

Quadratic constraint. Consider a second-order region given by $K = \{x \in \mathbb{R}^d : \frac{1}{2}x^\mathsf{T}Qx + p^\mathsf{T}x + l \leq 0\}$. Nesterov and Nemirovskii (1994) shows that $\phi := -\log f$ is an 1-self-concordant barrier for K, when $f(x) = -\frac{1}{2}\|x-\mu\|_\Sigma^2$ or $-(\frac{1}{2}x^\mathsf{T}Qx + p^\mathsf{T}x + l)$. Since $\bar{\nu} = \mathcal{O}(\nu^2)$ for a self-concordant barrier due to Lemma D.8, ϕ is $\mathcal{O}(1)$-symmetric. In case we consider $\|x-\mu\|_\Sigma^2$, the trivial scaling by dimension d implies that $d\phi$ is SSC and $\mathcal{O}(d)$-symmetric.

Moreover, $d\phi$ is SASC by Lemma D.15 by HSC of ϕ. For HSC of ϕ, we develop a handy tool for checking HSC. See §G.4.5 for the proof.

Lemma E.13 *For a real-valued function f on $K \subset \mathbb{R}^d$, let $\psi = -\log f$ be a ν-self-concordant barrier for K. Then,*

$$|\mathrm{D}^4\psi(x)[h^{\otimes 4}]| \lesssim \nu^2 \|h\|_{\nabla^2\psi(x)}^2 + \Big|\frac{\mathrm{D}^4 f(x)[h^{\otimes 4}]}{f(x)}\Big| .$$

Using this tool, we can study properties of the barrier for the quadratic constraints. We provide the proof in §G.4.5.

Lemma E.14 (Quadratic constraint) *For a closed convex $K = \{x \in \mathbb{R}^d : \frac{1}{2}x^\mathsf{T}Qx+p^\mathsf{T}x+l \leq 0\}$ with $p \in \mathbb{R}^d$ and $0 \neq Q \in \mathbb{S}_+^d$, let $\phi(x) = -\log(-l - p^\mathsf{T}x - \frac{1}{2}x^\mathsf{T}Qx)$ and $g = d\,\nabla^2\phi$.*

- *ν, $\bar{\nu} = \mathcal{O}(d)$.*
- *SSC when $Q \succ 0$, and SASC.*
- *$\mathrm{D}^2 g(x)[h,h] \succeq 0$ for any $x \in \mathrm{int}(K)$ and $h \in \mathbb{R}^d$ (so SLTSC).*

Gaussian distribution ($f(x) = \frac{1}{2}\|x-\mu\|_\Sigma^2$). Suppose the quadratic term $f(x) = \frac{1}{2}\|x-\mu\|_\Sigma^2$ appears in a potential of a target distribution. Then its epigraph is

$$\{(x,t) \in \mathbb{R}^{d+1} : \frac{1}{2}\|x-\mu\|_\Sigma^2 - t \leq 0\} ,$$

and clearly $q(x,t) = \frac{1}{2}\|x-\mu\|_\Sigma^2 - t$ is a quadratic function in (x,t). Hence, this level set admits an 1-self-concordant barrier

$$\phi(x,t) = -\log(t - \frac{1}{2}\|x-\mu\|_\Sigma^2) .$$

Our earlier discussion immediately leads to the following result:

Lemma E.15 (Quadratic potential) *Consider a closed convex $K = \{(x,t) : \frac{1}{2}\|x-\mu\|_\Sigma^2 \leq t\}$ with $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{S}_{++}^d$, and let $\phi(x) = -\log(t - \frac{1}{2}\|x-\mu\|_\Sigma^2)$ and $g = d\,\nabla^2\phi$.*

- *ν_g, $\bar{\nu}_g = \mathcal{O}(d)$.*
- *SSC and SASC.*
- *$\mathrm{D}^2 g(x,t)[h,h] \succeq 0$ for any $(x,t) \in \mathrm{int}(K)$ and $h \in \mathbb{R}^{d+1}$.*

Second-order cone ($f(x) = \frac{1}{2}\|x-\mu\|_\Sigma$). It is common that a potential includes a non-smooth term like $\|Ax-b\|_2$ in many applications, and we can handle such potentials via our framework. Nesterov and Nemirovskii (1994, Lemma 4.3.3) shows that

$$\phi(x,t) = -\log(t^2 - \|x\|^2)$$

is a 2-self-concordant for a level set $K = \{(x,t) \in \mathbb{R}^d \times \mathbb{R} : \|x\|_2 \le t\}$ (here we may assume that $\mu = 0$ and $\Sigma = I$ due to Lemma D.18). This level set is called a *second-order cone* or Lorentz cone.

Applying Lemma E.13 to $f(x,t) = t^2 - \|x\|^2$ with $\nu = 2$, we immediately show HSC of ϕ. Thus, $d\phi$ satisfies SLTSC and SASC by Lemma D.13 and Lemma D.15, respectively.

Lemma E.16 (Second-order cone) *Consider a closed convex $K = \{(x,t) : \|x-\mu\|_\Sigma \le t\}$ with $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{S}^d_{++}$, and let $\phi(x,t) = -\log(t^2 - \|x-\mu\|^2_\Sigma)$ and $g = d\,\nabla^2\phi$.*

- *ν_g, $\bar{\nu}_g = \mathcal{O}(d)$.*
- *SSC, SASC, and SLTSC.*

E.3. PSD cone

The function $\phi(X) = -\log\det X$ serves as an d-self-concordant barrier for the PSD cone $\mathbb{S}^d_+$. While achieving self-concordance does not require additional scaling, it turns out that SSC requires a scaling of $\Theta(d)$. Notably, this scaling is less than the trivial dimension-based scaling of $d_s := d(d+1)/2$. Also, direct computation leads to $\mathrm{D}^4\phi(X)[H,H] \succeq 0$ (so SLTSC).

As ϕ is HSC, scaling by d_s ensures SASC. However, we can achieve ASC with a smaller scaling by $\mathcal{O}(d)$ via the random matrix theory.

Lemma E.17 (PSD cone) *On a closed convex $K = \mathbb{S}^d_+$, let $\phi(X) = -\log\det X$ and define $g = d\,\nabla^2\phi$.*

- *$\nu = d^2$* (Nesterov and Nemirovskii, 1994) *and $\bar{\nu} = d^2$* (Lemma E.21).
- *SSC* (Corollary E.24).
- $\mathrm{D}^2 g(X)[H,H] \succeq 0$ *for any $X \in \mathrm{int}(K)$ and $H \in \mathbb{S}^d$* (Lemma E.25).
- *ASC* (Lemma E.27), *and $d_s\,\nabla^2\phi$ is SASC.*

E.3.1. FORMALISM VIA MATRIX-VECTOR TRANSFORMATIONS

In analyzing ϕ, we work in $\mathbb{R}^{d_s} = \mathbb{R}^{d(d+1)/2}$ and $\mathbb{S}^d$ simultaneously in the sequel, moving back and forth between them implicitly. We justify this identification as follows.

Measure on $\mathbb{S}^d$. We can define and work with the Lebesgue measure on $\mathbb{S}^d$ by identifying it with the Lebesgue measure on $\mathbb{R}^{d_s}$, where each component in the Lebesgue measure on $\mathbb{S}^d$ corresponds to each entry in the upper triangular part. Hence, with the Lebesgue measure $\mathrm{d}X$ on $\mathbb{S}^d$ it is straightforward to define a probability distribution on $\mathbb{S}^d$ whose probability density function with respect to $\mathrm{d}X$ is proportional to $\exp(-f)$ for a function $f : \mathbb{S}^d \to \mathbb{R}$. For instance, the uniform distribution over a region corresponds to f being constant in the region and infinity outside of the region, and an exponential distribution to $f(X) = \langle C, X\rangle = \mathrm{Tr}(C^\mathsf{T} X)$ for $C \in \mathbb{S}^d$.

Directional derivatives. A function $\phi : \mathbb{S}^d \to \mathbb{R}$ induces its counterpart $\psi : \mathbb{R}^{d_s} \to \mathbb{R}$ defined by $\psi(x) = \phi(X)$ for $x := \operatorname{svec}(X)$. For symmetric matrices $\{H_i\}_{i\le k}$, the k-th directional derivative of ϕ in directions $H_1, \dots, H_k$ is

$$\mathrm{D}^k\phi(X)[H_1, \cdots, H_k] \overset{\text{def}}{=} \frac{\mathrm{d}^k}{\mathrm{d}t_k \cdots \mathrm{d}t_1}\phi\Big(X + \sum_{i=1}^k t_i H_i\Big)\Big|_{t_1,\dots,t_k=0}\,.$$

For $h_i := \operatorname{svec}(H_i)$, it follows that $\phi(X + \sum_{i=1}^k t_i H_i) = \psi(x + \sum_{i=1}^k t_i h_i)$ and thus

$$\mathrm{D}^k\phi(X)[H_1, \cdots, H_k] = \mathrm{D}^k\psi(x)[h_1, \cdots, h_k]\,.$$

With this identification in hand, since the notion of (symmetric or strong) self-concordance is formulated in terms of directional derivatives, we can deal with both representations without having to specify one of them.

Important operators. We introduce three linear operators that enable us to make smooth transitions between $\mathbb{S}^d$ and $\mathbb{R}^{d_s}$.

Definition E.18 (Magnus and Neudecker (1980)) *Let $E_{ij} = e_i e_j^\mathsf{T} \in \mathbb{R}^{d\times d}$ be the matrix with a single 1 in the (i, j) position and zeros elsewhere.*

- *$M : \mathbb{R}^{d_s} \to \mathbb{R}^{d^2}$ is the linear operator that maps $\operatorname{svec}(\cdot)$ to $\operatorname{vec}(\cdot)$ (i.e., $M \circ \operatorname{svec} = \operatorname{vec}$). It can be written as $M = \sum_{i\ge j} \operatorname{vec}(T_{ij}) u_{ij}^\mathsf{T}$, where $T_{ij} \in \mathbb{R}^{d\times d}$ has all zero entries except for 1 at (i, j) and (j, i) positions (i.e., $T_{ij} = E_{ij} + E_{ji}$ if $i \ne j$ and E_{ij} if $i = j$), and $u_{ij} = \operatorname{svec}(E_{ij})$.*
- *$N : \mathbb{R}^{d^2} \to \mathbb{R}^{d^2}$ is the linear operator that maps $\operatorname{vec}(A)$ to $\operatorname{vec}\big(\frac{1}{2}(A + A^\mathsf{T})\big)$ for a matrix $A \in \mathbb{R}^{d\times d}$.*
- *$L : \mathbb{R}^{d_s} \to \mathbb{R}^{d^2}$ is the linear operator that maps $\operatorname{vec}(A)$ to $\operatorname{svec}(A)$ for a matrix $A \in \mathbb{R}^{d\times d}$. It can be written as $L = \sum_{i\ge j} u_{ij} \operatorname{vec}(E_{ij})^\mathsf{T}$.*

Lemma E.19 (Magnus and Neudecker (1980)) *Let M, N, L be matrices in Definition E.18.*

- *(Lemma 2.1) $N = N^\mathsf{T} = N^2$ and $N(A \otimes A) = (A \otimes A)N$ for any $d \times d$ matrix A.*
- *(Lemma 3.5) $MLN = N$.*

E.3.2. ANALYSIS OF A SELF-CONCORDANT METRIC FOR THE PSD CONE

We first examine properties of the metric defined by the Hessian of self-concordant barrier $\phi(X) = -\log\det X$ (see Nesterov (2003, Theorem 4.3.3) for self-concordance). In this case, its Hessian and inverse have clean formulas.

Proposition E.20 *Let $\nabla_X^2\phi(X) = -\nabla_x^2 \log\det(\operatorname{svec}^{-1}(x)) \in \mathbb{R}^{d_s\times d_s}$ for $X \in \mathbb{S}_+^d$. Then,*

$$\nabla^2\phi(X) = M^\mathsf{T}(X^{-1}\otimes X^{-1})M = M^\mathsf{T}(X\otimes X)^{-1}M\,,$$

$$\big(\nabla^2\phi(X)\big)^{-1} = M^\dagger(X\otimes X)\big(M^\dagger\big)^\mathsf{T} = LN(X\otimes X)NL^\mathsf{T}\,,$$

where $M^\dagger = (M^\mathsf{T}M)^{-1}M^\mathsf{T} \in \mathbb{R}^{d_s\times d^2}$ is the Moore-Penrose inverse of $M \in \mathbb{R}^{d^2\times d_s}$.

We defer the proof to Appendix H.2. We remark that as an immediate corollary to this, the local norm of $h \in \mathbb{R}^{d_s}$ with metric $\nabla^2\phi(X)$ is

$$\|h\|_X^2 = \operatorname{svec}(H)^\mathsf{T} M^\mathsf{T}(X^{-1}\otimes X^{-1})M\operatorname{svec}(H) \underset{\text{(i)}}{=} \operatorname{Tr}(HX^{-1}HX^{-1}) =: \|H\|_X^2\,,$$

where (i) follows from $\operatorname{vec} = M \circ \operatorname{svec}$ (Definition E.18) and $\operatorname{Tr}(DB^\mathsf{T}A^\mathsf{T}C) = \operatorname{vec}(A)^\mathsf{T}(B\otimes C)\operatorname{vec}(D)$ (Lemma H.1).

Symmetry.

Lemma E.21 (**$\bar{\nu}$-symmetry**) *For $X \in K = \mathbb{S}^d_+$, the barrier $\phi(X) = -\log\det X$ is d-symmetric.*

Proof For $X \in K$, pick any $Y \in K \cap (2X - K)$, and define a symmetric matrix $H := Y - X$. Since $Y \in K$ and $2X - Y \in K$, we have $X + H \in K$ and $X - H \in K$. Thus,

$$-I \preceq X^{-1/2} H X^{-1/2} \preceq I\,,$$

and the magnitude of each eigenvalue $\{\lambda_i\}_{i=1}^d$ of $X^{-1/2} H X^{-1/2}$ is bounded by 1. Hence,

$$\|H\|_X^2 = \mathrm{Tr}(X^{-1} H X^{-1} H) = \|X^{-1/2} H X^{-1/2}\|_F^2 \le \sum_{i=1}^d \lambda_i^2 \le d\,.$$

■

Convexity of log-determinant of Hessian and SSC. Next, the convexity of the log-determinant of $\nabla^2 \phi$ can be checked via properties of Kronecker products. See §G.4.6 for the proof.

Proposition E.22 (Convexity of log-determinant of Hessian) $\log\det(\nabla^2\phi(\cdot))$ *is convex.*

We move onto SSC of $d\phi(X)$.

Lemma E.23 *For $\psi_X := \sup_{H \in \mathbb{S}^d} \|(\nabla^2\phi(X))^{-1/2} \mathrm{D}^3\phi(X)[H]\,(\nabla^2\phi(X))^{-1/2}\|_F / \|H\|_X$, we have*

$$\sqrt{2(d+1)} \le \psi_X \le 2\sqrt{d}\,.$$

We present the proof in §G.4.6. This result informs us of the best possible scaling of ϕ that ensures SSC. Recall that if g satisfies $\|g^{-1/2}\mathrm{D}g[h]g^{-1/2}\|_F \le 2\alpha\|h\|_g$ for $\alpha > 0$, then $\alpha^2 g$ is SSC. We remark that the scaling of d is obviously better than the trivial scaling of $d_s = \Theta(d^2)$.

Corollary E.24 (Strong self-concordance) *A function $d\phi$ is a strongly self-concordant barrier for $\mathbb{S}^d_+$. Moreover, the scaling factor of d cannot be further improved.*

Strongly lower trace self-concordance. SLTSC of ϕ can be easily checked by noting $g(X)[H, H] = \mathrm{Tr}(X^{-1} H X^{-1} H)$ and using the chain rule. See the details in §G.4.7.

Lemma E.25 (SLTSC) $\mathrm{D}^2 g(X)[H, H] \succeq 0$ *for any $X \in \mathrm{int}(K)$ and $H \in \mathbb{S}^d$.*

Average self-concordance. In establishing ASC, we find an interesting connection to a *Gaussian orthogonal ensemble* (GOE), one of the main objects studied in the random matrix theory. We prove the following lemmas and explain challenges when extending our arguments to SASC in §G.4.8.

Lemma E.26 *For $d_s = \frac{d(d+1)}{2}$ and $\mathrm{svec}(H) \sim \mathcal{N}\big(0, \frac{r^2}{d_s} g(X)^{-1}\big)$, $\frac{\sqrt{d_s d}}{r} X^{-1/2} H X^{-1/2}$ is a GOE.*

Lemma E.27 (ASC) $-d\,\log\det X$ *is ASC.*

E.4. Logarithm, exponential, entropy, and ℓ_p-norm (power function)

Logarithm in potentials. Consider $Q_1 = \{(x,t) \in \mathbb{R}^2 : -\log x \le t, x > 0\}$. As $f(\cdot) = -\log(\cdot)$ is convex on $\mathbb{R}_+$ and satisfies the condition in Lemma D.31 with $\beta = 2$ and $\gamma = 6$,

$$F(x,t) = -\log(t + \log x) - 36 \log x$$

is a highly 37-self concordant barrier for Q_1. Therefore, $2F$ is SSC and SLTSC with $\bar{\nu} = \mathcal{O}(1)$.

Lemma E.28 (Logarithm) *Consider the direct product of level sets*

$$K = \prod_{i=1}^{d} \{(x_i, t_i) \in \mathbb{R}^2 : -\log x_i \le t_i,\, x_i > 0\} ,$$

and let $\phi(x,t) = -\sum_{i=1}^d \big(\log(t_i + \log x_i) + 36 \log x_i\big)$ *and* $g = 2\nabla^2 \phi$.
- ν, $\bar{\nu} = \mathcal{O}(d)$.
- *SSC and SLTSC.*
- $d\, \nabla^2 \phi$ *is SASC.*

Proof For $i \in [d]$, let $Q_i = \{(x_i, t_i) \in \mathbb{R}^2 : -\log x_i \le t_i,\, y_i > 0\}$ and $F_i(x_i, t_i)$ be the self-concordant barrier above. Note that $2F_i$ is SSC and SLTSC. By Lemma D.25 and D.26, the Hessian of $F(x,t) := 2\sum_{i=1}^d F_i(x_i, t_i)$ is SSC and SLTSC. The last item on SASC follows from Lemma D.15. ■

Exponent in potentials. Consider $Q_2 = \{(x,t) \in \mathbb{R}^2 : e^x \le t\} = \{(x,t) \in \mathbb{R}^2 : t > 0,\, x \le \log t\}$. As $f(t) = \log t$ is concave and satisfies the condition in Lemma D.30 with $\beta = 2$ and $\gamma = 6$,

$$F(x,t) = -\log(\log t - x) - 36 \log t$$

is a highly 37-self concordant barrier for Q_2. Therefore, $2F$ is SSC and SLTSC with $\bar{\nu} = \mathcal{O}(1)$.

Lemma E.29 (Exponential) *Consider the direct product of level sets*

$$K = \prod_{i=1}^{d} \{x_i, t_i) \in \mathbb{R}^2 : \exp(x_i) \le t_i\} ,$$

and let $\phi(x,t) = -\sum_{i=1}^d (\log(\log t_i - x_i) + 36 \log t_i)$ *and* $g = 2\nabla^2 \phi$.
- ν, $\bar{\nu} = \mathcal{O}(d)$.
- *SSC and SLTSC.*
- $d\, \nabla^2 \phi$ *is SASC.*

Proof For $i \in [d]$, let $Q_i = \{(x_i, t_i) \in \mathbb{R}^2 : e^{x_i} \le t_i\}$ and $F_i(x_i, t_i)$ be the self-concordant barrier above. Note that $2F_i$ is SSC and SLTSC. By Lemma D.25 and D.26, the Hessian of $F(x,t) := 2\sum_{i=1}^d F_i(x_i, t_i)$ is SSC and SLTSC. The last item on SASC follows from Lemma D.15. ■

Entropy in potentials. Consider $Q_3 = \{(x,t) \in \mathbb{R}^2 : x \geq 0,\, t \geq x \log x\}$. Note that $f(x) = x \log x$ is convex on $\{x > 0\}$ and satisfies the condition in Lemma D.31 with $\beta = 1$ and $\gamma = 2$. Hence,

$$F(x,t) = -\log(t - x\log x) - 36\log x$$

is a highly 5-self concordant barrier for Q_3. Therefore, $2F$ is SSC and SLTSC with $\bar{\nu} = \mathcal{O}(1)$.

Lemma E.30 (Entropy) *Consider the direct product of level sets*

$$K = \prod_{i=1}^{d} \{(x_i, t_i) \in \mathbb{R}^2 : x_i \geq 0,\, t_i \geq x_i \log x_i\}\,,$$

and let $\phi(x,t) = -\sum_{i=1}^d \bigl(\log(t_i - x_i \log x_i) + 36 \log x_i\bigr)$ *and* $g = 2\nabla^2 \phi$.
- $\nu,\, \bar{\nu} = \mathcal{O}(d)$.
- *SSC and SLTSC.*
- $d\,\nabla^2\phi$ *is SASC.*

Proof For $i \in [d]$, let $Q_i = \{(x_i,t_i) \in \mathbb{R}^2 : x_i \geq 0,\, t_i \geq x_i \log x_i\}$ and $F_i(x_i,t_i)$ be the self-concordant barrier above. Note that $2F_i$ is SSC and SLTSC. By Lemma D.25 and D.26, the Hessian of $F(x,t) := 2\sum_{i=1}^d F_i(x_i,t_i)$ is SSC and SLTSC. The last item on SASC follows from Lemma D.15. ∎

ℓ_p-norm (power function). We start with the power functions. For $p \geq 1$, consider $Q_4 = \{(x,t) \in \mathbb{R}^2 : t \geq \max(0,x)^p\} = \{(x,t) \in \mathbb{R}^2 : t \geq 0,\, x \leq t^{1/p}\}$. Note that $f(t) = t^{1/p}$ is concave on $t > 0$ and satisfies the condition in Lemma D.30 with $\beta = 2$ and $\gamma = 6$. Hence,

$$F_4(x,t) = -\log(t^{1/p} - x) - 36\log t$$

is a highly 37-self-concordant barrier for Q_4. Similarly, $F_5(t,x) = -\log(t^{1/p} + x) - 36\log t$ is a highly 37-self concordant barrier for the convex set $Q_5 = \{(x,t) \in \mathbb{R}^2 : t \geq \max(0,-x)^p\}$. Since the convex set $Q_6 = \{(x,t) \in \mathbb{R}^2 : t \geq |x|^p\}$ is equal to $Q_4 \cap Q_5$, the sum of $F_4 + F_5$, which is

$$F_6(x,t) = -\log(t^{2/p} - x^2) - 72\log t$$

is a highly 72-self-concordant barrier for Q_6. Hence, $2F$ is SSC and SLTSC with $\bar{\nu} = \mathcal{O}(1)$.

Lemma E.31 (ℓ_p-norm) *Consider the direct product of level sets* $K = \prod_{i=1}^d \{(x_i,t_i) \in \mathbb{R}^2 : |x_i|^p \leq t_i\}$*, and let* $\phi(x,t) = -\sum_{i=1}^d \bigl(\log(t_i^{2/p} - x_i^2) + 72\log t_i\bigr)$ *and* $g = 2\nabla^2\phi$.
- $\nu,\, \bar{\nu} = \mathcal{O}(d)$.
- *SSC and SLTSC.*
- $d\,\nabla^2\phi$ *is SASC.*

Proof Consider a highly 72-self-concordant barrier F_i above for $\{(x_i,t_i) : |x_i|^p \leq t_i\}$ for $i \in [d]$. Note that $2F_i$ is SSC and SLTSC. By Lemma D.25 and D.26, the Hessian of $F(x,t) := 2\sum_{i=1}^d F_i(x_i,t_i)$ is SSC and SLTSC. The last item on SASC follows from Lemma D.15. ∎

Appendix F. Examples

For given constraints and epigraphs, combining metrics for them (according to the self-concordance theory for sampling developed in §D) and employing GCDW with the combined metric lead to a poly-time mixing sampling algorithm. Compared to the state-of-the-art poly-time mixing algorithm, the Ball walk, GCDW offers several advantages. First, it does not require any preprocessing (e.g., rounding) due to affine invariance. Also, it achieves faster mixing by leveraging inherent geometric information in sampling problems.

The per-step complexity of Dikin walks, however, is in general higher than that of the Ball walk. The primary computational bottleneck lies in computing the inverse of a local metric. Nevertheless, efficient implementation of inverse maintenance can significantly reduce the per-step complexity, improving the total complexity (# iterations needed for mixing times the per-step complexity).

In this section, we illustrate how our framework recovers theoretical guarantees of previous work on Dikin walks for uniform sampling and extends beyond uniform sampling. In particular, we show that GCDW is a poly-time mixing algorithm capable of sampling uniform, exponential, or Gaussian distributions on second-order cones or truncated PSD cones. Additionally, we illustrate an efficient per-step implementation that yields a faster total complexity when compared to general-purpose samplers such as the Ball walk.

F.1. Polytope sampling

Consider a set of linear constraints given by $K = \{x \in \mathbb{R}^d : Ax \geq b\}$ with $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$.

Uniform sampling. Kannan and Narayanan (2012) first studied the Dikin walk for uniformly sampling a polytope, where a local metric is set to be the Hessian of the logarithmic barrier, $g = \nabla^2\phi_{\log} = A_{(\cdot)}^\top A_{(\cdot)}$. They showed that the Dikin walk with the log-barrier mixes in $\mathcal{O}\big(md\log\frac{M}{\varepsilon}\big)$ iterations with a warmness parameter M. An immediate consequence of our work is that GCDW achieves the mixing time of $\widetilde{\mathcal{O}}(md)$ *without a warmness assumption*, as $\bar{\nu}, \nu = m$ and g is SSC, LTSC, and ASC by Lemma E.1.

Chen et al. (2018) introduced the Vaidya walk and the Approximate John walk, which are essentially Dikin walks with the Vaidya metric $\nabla^2\phi_{\text{Vaidya}}$ and a version of the Lewis-weight metric $\sqrt{d}\,\nabla^2\phi_{\text{Lw}}$. Their work showed that both walks achieves mixing times of $\mathcal{O}\big(\sqrt{m}d^{3/2}\log\frac{M}{\varepsilon}\big)$ and $\mathcal{O}\big(d^{5/2}\log^{\mathcal{O}(1)} m\,\log\frac{M}{\varepsilon}\big)$, respectively. Building upon our analysis of the Vaidya metric and Lewis-weight metric in Lemma E.2 and E.3, we find that GCDW with these metrics achieves the same mixing but without any warmness assumption.

We note that for the same task the Ball walk without a warm start requires $\widetilde{\mathcal{O}}(d^3)$ membership queries due to Kannan et al. (1997); Jia et al. (2021). Given that a membership query involves $\mathcal{O}(md)$ arithmetic operations, the total complexity of the Ball walk is $\widetilde{\mathcal{O}}(md^4)$. In contrast, the per-step of the Dikin walk with the log-barrier can be run in $\mathcal{O}(md^{\omega-1})$ operations through the fast matrix multiplication, so the total number of arithmetic operations is $\widetilde{\mathcal{O}}(m^2d^\omega)$. Thus, for m close to d GCDW is provably faster than the Ball walk. When an efficient inverse maintenance proposed in Laddha et al. (2020) is employed, the per-step complexity can be improved to $\mathcal{O}(d^2 + \mathsf{nnz}(A)) = \mathcal{O}(md)$. In such cases GCDW is faster in a broader range of m. In particular, if A is as sparse as $\mathsf{nnz}(A) = \mathcal{O}(d^2)$, then GCDW is always faster than the Ball walk. Moreover, GCDW with the Lewis-weight metric mixes in $\widetilde{\mathcal{O}}(d^{2.5})$ steps with the per-step complexity of $\widetilde{\mathcal{O}}(md^{\omega-1})$, so it is always faster than the Ball walk for any m.

Exponential and Gaussian sampling. The current mixing bound of the Ball walk for general log-concave sampling is $\widetilde{\mathcal{O}}(d^4)$ due to Lovász and Vempala (2007). On the other hand, the Dikin walk employed with any metric above for exponential sampling converges in the same iterations as the Dikin walk for uniform sampling. Since only difference between two sapling is the additional term of $\exp\bigl(-(f(z)-f(x))\bigr)$ in the Metropolis filter, the fast implementation techniques mentioned earlier can be applied to the context of exponential sampling. As a result, for the exponential sampling each of the Dikin walks described above surpasses the Ball walk by a larger margin.

For Gaussian sampling over a polytope, we first reduce it to the exponential sampling as in (redLC): for $y=(x,t)\in\mathbb{R}^{d+1}$

$$\begin{aligned}\text{sample } & y\sim\tilde{\pi}\propto\exp(-t)\\ \text{s.t. } & Ax\geq b,\ \frac{1}{2}\|x-\mu\|_\Sigma^2\leq t\,.\end{aligned}$$

According to our theory, it is natural to use the metric given by

$$g(x,t)=2\begin{bmatrix}\nabla_x^2\phi_{\text{log}}(x) & \\ & 0\end{bmatrix}+2(d+1)\,\nabla_{(x,t)}^2\phi_{\text{Gauss}}(x,t)\,,$$

which is $\bigl(\mathcal{O}(m+d),\mathcal{O}(m+d)\bigr)$-Dikin-amenable due to Lemma E.15. Thus, GCDW needs $\widetilde{\mathcal{O}}(d(m+d))$ iterations of the Dikin walk. We note that the log-barrier can be replaced by the Vaidya or Lewis-weight metrics, and in such cases one can obtain provable guarantees on the mixing time by computing ν and $\bar{\nu}$, referring to §E or Table 2.

F.2. Second-order cone sampling

We consider a region given by $\|x-\mu\|_\Sigma\leq t$ and $A\begin{bmatrix}x & t\end{bmatrix}^\mathsf{T}\leq b$ for $A\in\mathbb{R}^{m\times(d+1)}$, $b\in\mathbb{R}^m$, $\mu\in\mathbb{R}^d$, and $\Sigma\in\mathbb{S}_{++}^d$.

Uniform and exponential sampling. In this case, our self-concordance theory suggests using

$$\nabla^2(2\sqrt{d+1}\phi_{\text{Lw}}+2(d+1)\phi_{\text{SOC}})\quad\text{or}\quad\nabla^2(2\phi_*+2(d+1)\,\phi_{\text{SOC}})\text{ for }*=\text{log, Vaidya}\,,$$

to deal with the truncated SOC constraint. For the log-barrier case, this yields an $\bigl(\mathcal{O}(m+d),\mathcal{O}(m+d)\bigr)$-Dikin-amenable metric due to Lemma E.16, with which GCDW requires $\widetilde{\mathcal{O}}(d(m+d))$ iterations of the Dikin walk.

Gaussian sampling. Following the reduction as in the polytope sampling, we should use

$$g(x,t,t')=3\begin{bmatrix}\nabla_{(x,t)}^2\phi_{\text{log}}(x,t)+(d+1)\,\nabla_{(x,t)}^2\phi_{\text{SOC}}(x,t) & \\ & 0\end{bmatrix}+3(d+2)\,\nabla_{(x,t,t')}^2\phi_{\text{Gauss}}(x,t,t'),$$

which is $\bigl(\mathcal{O}(m+d),\mathcal{O}(m+d)\bigr)$-Dikin-amenable, and thus GCDW needs $\widetilde{\mathcal{O}}(d(m+d))$ iterations of the Dikin walk.

F.3. PSD cone sampling

For a matrix $X\in\mathbb{R}^{d\times d}$, recall that $\text{vec}(X)\in\mathbb{R}^{d^2}$ denotes the vector obtained by stacking columns of X vertically. Additionally, we define $A\in\mathbb{R}^{m\times d^2}$, $S_X\in\mathbb{R}^{m\times m}$, and $A_X\in\mathbb{R}^{m\times d^2}$ by

$$A:=\begin{bmatrix}\text{vec}(A_1) & \cdots & \text{vec}(A_m)\end{bmatrix}^\mathsf{T},\quad S_X:=\text{Diag}(\langle A_i,X\rangle-b_i),\quad A_X:=S_X^{-1}A\,,$$

where we assume A has no all-zero rows and $(S_X)_{ii}>0$ for $i\in[m]$.

Uniform and exponential sampling. The metric below comes from the Hessian of

$$-2d^2 \log\det X - 2\sum_{i=1}^m \log(\langle A_i, X\rangle - b_i)\,.$$

Here the first term, the log-determinant, serves as a barrier for the PSD cone while the second term is the standard logarithmic barrier for linear constraints. We note that the $-\log\det X$ is strictly convex on $x \in \operatorname{int}(K)$ for K the truncated PSD cone, so all metrics g introduced in our main results are positive definite. Thus, the Dikin walk with those g is well-defined.

Proposition F.1 *Let K be the truncated PSD cone and g be the local metric such that at each $X \in \operatorname{int}(K)$, for symmetric matrices H_1, H_2,*

$$g_X(H_1, H_2) = 2d^2 \operatorname{Tr}(X^{-1}H_1X^{-1}H_2) + 2\ \operatorname{vec}(H_1)^\mathsf{T} A_X^\mathsf{T} A_X \operatorname{vec}(H_2)\,.$$

Then GCDW *needs $\widetilde{\mathcal{O}}((d^3+m)d^2)$ steps of the* Dikin walk *with the local metric g, where each step runs in $\mathcal{O}\big((md^\omega + m^2d^2) \wedge (d^{2\omega} + md^{2(\omega-1)})\big)$ time*[4].

Since g_X is $\big(\mathcal{O}(m+d^3), \mathcal{O}(m+d^3)\big)$-Dikin-amenable by Lemma E.17, GCDW requires $\widetilde{\mathcal{O}}(d^2\,(d^3+m))$ iterations of the Dikin walk. As mentioned earlier, efficient maintenance of the inverse of a metric function could lead to a faster per-step complexity. As an example, we provide such an implementation of Proposition F.1 in §F.3.1. Putting these together, for an interesting regime of $m = \mathcal{O}(1)$, GCDW is faster than the Ball walk by a factor of d in terms of the total complexity.

If we replace the log-barrier by the Vaidya metric, then the dependence on m is improved to $\sqrt{m}$ as in the polytope sampling. See §G.5.1 for the proofs of the two claims below.

Proposition F.2 *Let K be the truncated PSD cone and g be the local metric such that at each $X \in \operatorname{int}(K)$, for symmetric matrices H_1, H_2,*

$$g_X(H_1, H_2) = 2d^2 \operatorname{Tr}(X^{-1}H_1X^{-1}H_2) + 44\sqrt{\frac{m}{d}}\ \operatorname{vec}(H_1)^\mathsf{T} A_X^\mathsf{T}\big(\Sigma_X + \frac{d}{m}I_m\big)A_X \operatorname{vec}(H_2)\,.$$

Then GCDW *needs $\widetilde{\mathcal{O}}((d^2+\sqrt{m})d^3)$ steps of the* Dikin walk *with the local metric g, with each step running in $\widetilde{\mathcal{O}}(md^{2(\omega-1)})$ amortized time.*

Lastly, the dependence on m can be made poly-logarithmic by working with the Lewis-weight metric. We remark that for uniform sampling the total complexity of GCDW is less than that of the Ball walk by the order of $d^{5-2\omega}$.

Proposition F.3 *Let K be the truncated PSD cone and g be the local metric such that at each $X \in \operatorname{int}(K)$, for symmetric matrices H_1, H_2,*

$$g_X(H_1, H_2) = 2d^2 \operatorname{Tr}(X^{-1}H_1X^{-1}H_2) + dc_1(\log m)^{c_2}\ \operatorname{vec}(H_1)^\mathsf{T} A_X^\mathsf{T} W_X A_X \operatorname{vec}(H_2)\,,$$

where W_X is the diagonalized ℓ_p-Lewis weight of A_X with $p = \mathcal{O}(\log m)$, and $c_1, c_2 > 0$ are universal constants. Then GCDW *requires $\widetilde{\mathcal{O}}(d^5)$ steps of the* Dikin walk*, with each step running in $\widetilde{\mathcal{O}}(md^{2(\omega-1)})$ amortized time.*

4. Here $\omega < 2.373$ is the current matrix multiplication complexity exponent (Le Gall, 2014)).

Algorithm 4: Computation of $g(X)^{-1}v$

Input: $X \in \mathbb{S}_+^d$, vector $v \in \mathbb{R}^{d_s}$, local metric g.
Output: $g(X)^{-1}v$
Prepare the column vectors u_i of $U = M^\mathsf{T} A^\mathsf{T} S_X^{-1}$.
For $\bar{g}_0 := g_1(X)$, compute $\bar{g}_0^{-1}v$ and $\bar{g}_0^{-1}u_i$ for $i \in [m]$.
for $i = 1, \cdots, m$ **do**
  Compute $\bar{g}_i^{-1}v$ and $\bar{g}_i^{-1}u_j$ for $j \in [m]$, according to

$$\bar{g}_i^{-1}w = \bar{g}_{i-1}^{-1}w - \frac{\bar{g}_{i-1}^{-1}u_i \cdot u_i^\mathsf{T} \bar{g}_{i-1}^{-1}w}{1 + u_i^\mathsf{T} \bar{g}_{i-1}^{-1}u_i}\,.$$

end
Output $\bar{g}_m^{-1}v$.

Gaussian sampling. Just as in polytope or second-order cone sampling, we introduce a new variable t by replacing a quadratic term in the potential. This reduces the Gaussian sampling problem to an exponential sampling problem. We then work with a local metric

$$g(X,t) = 3\left(d\begin{bmatrix} \nabla_X^2 \phi_{\mathrm{Lw}}(X) & \\ & 0 \end{bmatrix} + d^2 \begin{bmatrix} \nabla_X^2 \phi_{\mathrm{PSD}}(X) & \\ & 0 \end{bmatrix} + d^2 \nabla_{(X,t)}^2 \phi_{\mathrm{Gauss}}(X,t)\right),$$

which is $\big(\mathcal{O}^*(d^3), \mathcal{O}^*(d^3)\big)$-Dikin-amenable. Thus, GCDW needs $\widetilde{\mathcal{O}}(d^5)$ iterations of the $\mathsf{Dikin\ walk}$ with the local metric g, and the per-step complexity remains $\widetilde{\mathcal{O}}(md^{2(\omega-1)})$ in amortized time.

F.3.1. PER-STEP IMPLEMENTATION

Now we design an oracle that implements each iteration of the $\mathsf{Dikin\ walk}$ (Algorithm 1). This can be implemented as follows: when the current point is x,

- Sample $z \sim \mathcal{N}\big(0, \frac{r^2}{d} g(x)^{-1}\big)$.
- Compute $y = x + g(x)^{-1/2}z$ and propose it.
- Accept y with probability $1 \wedge \Big(\sqrt{\frac{\det g(y)}{\det g(x)}}\, \frac{\exp f(x)}{\exp f(y)}\Big)$.

We provide two algorithms with the complexity of $\mathcal{O}(md^\omega + m^2d^2)$ and $\mathcal{O}(d^{2\omega} + md^{2(\omega-1)})$. We can implement each iteration in $\mathcal{O}\big((md^\omega + m^2d^2) \wedge (d^{2\omega} + md^{2(\omega-1)})\big)$ time by using the former for small m and the latter for large m. This completes the second half of Theorem F.1.

Algorithm for small m. For simplicity here, we ignore the constant factors of $g = g_1 + g_2$, where

$$g_1(X) = M^\mathsf{T}(X \otimes X)^{-1}M =: BB^\mathsf{T} \qquad \text{and} \qquad g_2(X) = M^\mathsf{T}A^\mathsf{T}S_X^{-2}AM =: UU^\mathsf{T}\,.$$

where $B := M^\mathsf{T}(X \otimes X)^{-1/2} \in \mathbb{R}^{d_s \times d^2}$ and $U := M^\mathsf{T}A^\mathsf{T}S_X^{-1} \in \mathbb{R}^{d_s \times m}$. Letting u_i be the i-th column of U for $i \in [m]$, we note that $g_2 = \sum_{i=1}^m u_iu_i^\mathsf{T}$.

We start with a subroutine for computing $g(X)^{-1}v$ for given $v \in \mathbb{R}^{d_s}$ in $\mathcal{O}(md^\omega + m^2d^2)$ time.

Proposition F.4 *Algorithm 4 computes $g(X)^{-1}v$ in $\mathcal{O}(md^\omega + m^2d^2)$ time for a query $v \in \mathbb{R}^{d_s}$.*

Algorithm 5: Implementation of the Dikin walk

Input: current point $X \in \mathbb{S}_+^d$, local metric g

```
// Step 1:  Sampling from
```
$\mathcal{N}\big(0, \frac{r^2}{d} g(X)^{-1}\big)$

Draw $w \sim \mathcal{N}(0, I_{d^2+m})$ and $v \leftarrow g(X)^{-1} \begin{bmatrix} B & U \end{bmatrix} w$ by Algorithm 4.

Propose $y \leftarrow \text{svec}(X) + \frac{r}{\sqrt{d}} v$.

```
// Step 2:  Computation of acceptance probability
```

Use Algorithm 4 to prepare $\{\bar{g}_i^{-1} u_1, \dots, \bar{g}_i^{-1} u_m\}_{i=0}^m$ at X and $Y := \text{svec}^{-1}(y)$.

$\det \bar{g}_0(\cdot) \leftarrow 2^{d(d-1)/2} (\det(\cdot))^{-(d+1)}$ ($\because$ Lemma H.1-7)

for $i = 1, \cdots, m$ **do**

  $\det(\bar{g}_{i+1}) \leftarrow \det \bar{g}_i \cdot (1 + u_{i+1}^\mathsf{T} \bar{g}_i^{-1} u_{i+1})$.

end

Accept Y with probability $1 \wedge \Big(\sqrt{\frac{\det \bar{g}_m(Y)}{\det \bar{g}_m(X)}} \frac{\exp f(X)}{\exp f(Y)}\Big)$.

See §G.5.2 for the proof. With this subroutine in hand, we proceed to an efficient implementation of two tasks – computation of (1) $g(x)^{-\frac{1}{2}} z$ for a given vector $z \in \mathbb{R}^{d_s}$ and (2) $\sqrt{\frac{\det g(y)}{\det g(x)}} \frac{\exp f(x)}{\exp f(y)}$.

Lemma F.5 *Algorithm 5 implements the* Dikin walk *with per-step complexity of* $\mathcal{O}(md^\omega + m^2 d^2)$.

Algorithm for large m. The algorithm right above has quadratic dependence on the number m of constraints, which could become expensive for large m. In this regime, we just fully compute the whole matrix function of size $\mathbb{R}^{d_s \times d_s}$, which takes $\mathcal{O}(d^{2\omega} + md^{2(\omega-1)})$ time, and computing its inverse, square-root, and determinant takes $\mathcal{O}(d^{2\omega})$ time.

F.3.2. HANDLING APPROXIMATE LEWIS WEIGHTS

When implementing the Dikin walk with the Lewis-weights metric, we use an approximation algorithm presented in Lee and Sidford (2019) for computing and updating the Lewis weight, which ensures

$$(1-\delta)\widetilde{W}_X \preceq W_X \preceq (1+\delta)\widetilde{W}_X$$

for the approximate Lewis weights $\widetilde{W}_X$ and a target accuracy parameter δ (note that the initialization and update times of the Lewis weight above hide poly-logarithmic dependence on $\log(1/\delta)$). Strictly speaking, we should check that these approximate Lewis weights do not affect the theoretical guarantees above.

To see this, let us define $\widetilde{g} = 2(dg_1 + \widetilde{g}_2)$, where for some constants $c_1, c_2 > 0$

$$g_1(X) = d^2 M^\mathsf{T} (X \otimes X)^{-1} M \qquad \text{and} \qquad \widetilde{g}_2 = dc_1 (\log m)^{c_2} M^\mathsf{T} A_X^\mathsf{T} \widetilde{W}_X A_X M\,.$$

First of all, the Dikin walk with $\widetilde{g}$ still converges to a target distribution, since the approximation algorithm in Lee and Sidford (2019) is deterministic and thus the condition of detailed balance still holds under the acceptance probability of $1 \wedge \Big(\sqrt{\frac{\det \widetilde{g}(Y)}{\det \widetilde{g}(X)}} \frac{\exp f(X)}{\exp f(Y)}\Big)$. For $\widetilde{P}_X$ the one-step distribution of the Dikin walk started at X with $\widetilde{g}$, we can show one-step coupling similar to Lemma B.3, following the overall proof therein and taking $\delta = 1/\text{poly}(d)$ small enough. See §G.5.3 for the proof.

Lemma F.6 (One-step coupling) *For convex $K \subset \mathbb{R}^d$, let $g : \operatorname{int}(K) \to \mathbb{S}^d_{++}$ be SSC, ASC, LTSC, and $\phi : \operatorname{int}(K) \to \mathbb{R}$ be its function counterpart. Suppose that the potential f of the target distribution π is β-relatively smooth in ϕ. Then there exist constants $s_1, s_2 > 0$ such that if $\|x-y\|_{g(x)} \le s_1 r/\sqrt{d}$ with $r = s_2\,(1 \wedge 1/\sqrt{\beta})$ for $x, y \in \operatorname{int}(K)$, then $d_{\mathrm{TV}}(\widetilde{P}_x, \widetilde{P}_y) \le \frac{3}{4} + 0.01$.*

Appendix G. Proofs

We collect deferred proofs in this section.

G.1. Mixing of the Dikin walk (§B)

G.1.1. ONE-STEP COUPLING

We start with the one-step coupling of the Dikin walk under the setting $\alpha\nabla^2\phi \preceq \nabla^2 f \preceq \beta\nabla^2\phi$ on $\operatorname{int}(K)$. Roughly speaking, if $\|x-y\|_x \le r/\sqrt{d}$ with $r \lesssim 1 \wedge 1/\sqrt{\beta}$, then $d_{\mathrm{TV}}(P_x, P_y) \le 0.99$.

Proof of Lemma B.3. For $\pi \propto \exp(-f)\cdot \mathbf{1}_K$ and $z \sim \mathcal{N}(x, \frac{r^2}{d} g(x)^{-1})$, let us denote

$$p_x = \mathcal{N}\Big(x, \frac{r^2}{d} g(x)^{-1}\Big), \qquad R(x,z) = \frac{p_z(x)}{p_x(z)}\frac{\pi(z)}{\pi(x)}, \qquad A(x,z) = \min\big(1, R(x,z)\,\mathbf{1}_K(z)\big)\,.$$

The transition kernel $P(x,\cdot)$ of the Dikin walk started at x can be written as

$$P(x,dz) = \underbrace{\big(1 - \mathbb{E}_{p_x}[A(x,\cdot)]\big)}_{=:r_x}\,\delta_x(\mathrm{d}z) + A(x,z)\,p_x(\mathrm{d}z)\,.$$

Thus, for $x, y \in \operatorname{int}(K)$,

$$d_{\mathrm{TV}}(P_x, P_y) = \underbrace{\frac{1}{2}(r_x + r_y)}_{\mathsf{I}} + \underbrace{\frac{1}{2}\int |A(x,z)\,p_x(z) - A(y,z)\,p_y(z)|\,\mathrm{d}z}_{\mathsf{II}}\,. \qquad \text{(TV-decomposition)}$$

Let $h \sim \mathcal{N}(0, I_d)$ and denote a bad event $B_0 = \{z \in \mathbb{R}^d : \|z-x\|_x \ge cr\}$ with c determined later. Due to $\|z-x\|_x = \frac{r}{\sqrt{d}}\|h\|$ (in law) and concentration of the standard Gaussian in a thin shell of radius $\sqrt{d}$ with annulus $\mathcal{O}(1)$[5], we have $\mathbb{P}_z(B_0) = \mathbb{P}_h(\|h\| \ge c\sqrt{d}) \le \exp\big(-(c-1)\sqrt{d}/2\big)$. Hence, $\mathbb{P}(B_0) \le \varepsilon$ for $c \ge 1 + \sqrt{\frac{2}{d}\log\frac{1}{\varepsilon}}$.

Rejection probability r_x and r_y (Term I). Note that

$$r_x = 1 - \mathbb{E}_{p_x}[A(x,z)] = 1 - \int \min\Big(1, \underbrace{\mathbf{1}_K(z)\frac{\exp f(x)}{\exp f(z)}}_{=:\mathsf{A}}\underbrace{\frac{p_z(x)}{p_x(z)}}_{=:\mathsf{B}}\Big)p_x(\mathrm{d}z)\,.$$

5. A standard Gaussian $h \sim \mathcal{N}(0, I_d)$ is concentrated around a thin sell of radius $\sqrt{d}$ with annulus $\mathcal{O}(1)$: For $t > 0$,

$$\mathbb{P}_h(\|h\|_2 \ge \sqrt{d} + t) \le \exp(-t^2/2)\,.$$

As for A, we let $\nabla^2\phi \preceq c_\phi g$ for some $c_\phi > 0$ and use Taylor's expansion at $x \in K \cap B_0^c$ to show that for some $x^* \in [x,z]$,

$$\begin{aligned} f(x) - f(z) + \nabla f(x)^{\mathsf{T}}(z-x) &= -\|z-x\|^2_{\nabla^2 f(x^*)} \geq -c_\phi\beta\,\|z-x\|^2_{g(x^*)} \\ &\underset{\text{(i)}}{\geq} -c_\phi\beta\,\|z-x\|_x^2 \cdot (1+2\|x-z\|_x)^2 \geq -c_\phi\beta c^2 r^2 (1+2cr)^2 \underset{\text{(ii)}}{\geq} -\varepsilon\,, \end{aligned}$$

where we used Lemma D.4 in (i) and took $r \leq r_1(\varepsilon)$ in (ii), which is defined so that $\beta c_\phi c^2 r^2 (1+cr)^2 \leq \varepsilon$ for any $r \leq r_1(\varepsilon)$. It follows from $\mathcal{D}_g^1(x) \subset K$ and symmetry of $\mathcal{N}_g^r(x)$ that there exists a half-ellipsoid $G \subset \mathcal{D}_g^1(x)$ in which $\langle \nabla f(x), z-x\rangle \leq 0$. Thus, $f(x) - f(z) \geq -\varepsilon$ holds on $z \in G$.

For a bad event $B_1 := G^c$, it holds that

$$\mathbb{P}_z(B_1) \leq \frac{1}{2} + \mathbb{P}_z\big(\mathcal{D}_g^1(x)^c\big) = \frac{1}{2} + \mathbb{P}_z(\|z-x\|_x \geq 1) = \frac{1}{2} + \mathbb{P}_h\Big(\|h\| \geq \frac{\sqrt{d}}{r}\Big) \leq \frac{1}{2} + \varepsilon\,,$$

where the last inequality follows from concentration of h for any $r \leq r_2(\varepsilon) := \big(1 + \frac{2}{\sqrt{d}}\,\log\frac{1}{\varepsilon}\big)^{-1}$.

As for B, for $\varphi(x) := \frac{1}{2}\log\det g(x)$ we have

$$\log \mathsf{B} = -\frac{d}{2r^2}\big(\|z-x\|_z^2 - \|z-x\|_x^2\big) + \big(\varphi(z) - \varphi(x)\big)\,.$$

Invoking ASC of ϕ, we can take $r_3(\varepsilon)$ so that $\mathbb{P}_z\big(\|z-x\|_z^2 - \|z-x\|_x^2 \leq 2\varepsilon r^2/d\big) \geq 1-\varepsilon$ for any $r \leq r_3(\varepsilon)$ and control the first term. Let the complement of this event be our second bad event B_2.

For $\varphi(z) - \varphi(x)$, Taylor's expansion of φ at x leads to

$$\varphi(z) - \varphi(x) = \underbrace{\langle \nabla\varphi(x), z-x\rangle}_{=:\mathsf{A}'} + \underbrace{\frac{1}{2}\langle \nabla^2\varphi(x^*)(z-x), z-x\rangle}_{=:\mathsf{B}'} \text{ for some } x^* \in [x,z]\,.$$

As for A', we have $\langle \nabla\varphi(x), z-x\rangle = \frac{r}{\sqrt{d}}\langle g(x)^{-1/2}\nabla\varphi(x), h\rangle$, and a standard tail bound for h leads to

$$\mathbb{P}_z\Big(\langle \nabla\varphi(x), z-x\rangle \leq -\frac{r}{\sqrt{d}}\,\|g(x)^{-1/2}\nabla\varphi(x)\|_2 \cdot 2\log\frac{1}{\varepsilon}\Big) \leq \varepsilon\,.$$

We call this event B_3 and bound $\|g(x)^{-1/2}\nabla\varphi(x)\|_2$ via SSC of g as follows: omitting x for simplicity,

$$\begin{aligned} \|g^{-\frac{1}{2}}\nabla\varphi\|_2 &= \sup_{v:\|v\|_2=1}\langle \nabla\varphi, g^{-\frac{1}{2}}v\rangle \underset{\text{(i)}}{=} \sup_v \operatorname{Tr}(g^{-1}\mathrm{D}g[g^{-\frac{1}{2}}v]) = \sup_v \operatorname{Tr}(g^{-\frac{1}{2}}\mathrm{D}g[g^{-\frac{1}{2}}v]\,g^{-\frac{1}{2}}) \\ &\underset{\text{(ii)}}{\leq} \sup_v \sqrt{d}\,\|g^{-\frac{1}{2}}\mathrm{D}g[g^{-\frac{1}{2}}v]\,g^{-\frac{1}{2}}\|_F \underset{\text{(iii)}}{\leq} \sup_v 2\sqrt{d}\|g^{-\frac{1}{2}}v\|_x = 2\sqrt{d}\,, \end{aligned}$$

where (i) follows from (H.1), (ii) is due to $\operatorname{Tr}(A) \leq \sqrt{d}\|A\|_F$ for $A \in \mathbb{R}^{d\times d}$, and (iii) is due to SSC. Conditioned on B_3^c, taking $r \leq r_4(\varepsilon) := \varepsilon(4\log\frac{1}{\varepsilon})^{-1}$, we have

$$\mathsf{A}' = \langle \nabla\varphi(x), z-x\rangle \geq -4r\,\log\frac{1}{\varepsilon} \geq -\varepsilon\,.$$

As for B', denoting $u = z - x$ for $z \in B_0^c$

$$\mathrm{D}^2\varphi(x^*)[u,u] \underset{\text{(H.3)}}{=} \operatorname{Tr}\big(g(x^*)^{-1}\mathrm{D}^2 g(x^*)[u,u]\big) - \|g(x^*)^{-\frac{1}{2}}\mathrm{D}g(x^*)[u]\,g(x^*)^{-\frac{1}{2}}\|_F^2$$

$$\begin{aligned}
&\underset{\text{(i)}}{\geq} -\|u\|_{x^*}^2 - \|g(x^*)^{-\frac{1}{2}}\mathrm{D}g(x^*)[u]\,g(x^*)^{-\frac{1}{2}}\|_F^2 \geq -\|u\|_{x^*}^2 - 4\|u\|_{x^*}^2 \\
&\underset{\text{(ii)}}{\geq} -5(1-\|x-x^*\|_x)^{-2}\|u\|_x^2 \qquad \text{(G.1)}\\
&\geq -5(1+2cr)^2c^2r^2\,,
\end{aligned}$$

where (i) follows from LTSC and (ii) follows from Lemma D.4. Hence, $\mathsf{B}' \geq -\varepsilon/2$ by taking $r \leq r_5(\varepsilon)$ so that $5(1+2cr_5)^2c^2r_5^2 = \varepsilon$.

In summary, conditioned on $G := \bigcap_{i=0}^3 B_i^c$ with $\mathbb{P}_z(G) \geq \frac{1}{2} - 4\varepsilon$ due to the union bound, we have

$$\mathsf{A}: \frac{\exp f(x)}{\exp f(z)} \geq \exp(-\varepsilon)\,, \qquad \text{(G.2)}$$

$$\mathsf{B}: \frac{p_z(x)}{p_x(z)} \geq \exp(-3\varepsilon)\,, \qquad \text{(G.3)}$$

$$\varphi(z) - \varphi(x) \geq -2\varepsilon\,. \qquad \text{(G.4)}$$

Combining these together,

$$r_x = 1 - \int \min\Bigl(1, \mathbf{1}_K(z)\frac{\exp f(x)}{\exp f(z)}\frac{p_z(x)}{p_x(z)}\Bigr)p_x(\mathrm{d}z) \leq 1 - \int_G (1 \wedge e^{-\varepsilon}e^{-3\varepsilon})\,\mathbb{P}_z(G) \leq \frac{1}{2} + 5\varepsilon\,.$$

Bounding r_y in the same way, we conclude that $\mathsf{I} \leq \frac{1}{2} + 5\varepsilon$ in (TV-decomposition).

Overlapping part (Term II). WLOG, assume $f(y) \geq f(x)$. We denote good events by $G_x = \cap_{i=0,2,3}B_{x,i}^c$ and $G_y = \cap_{i=0,2,3}B_{y,i}^c$ such that $\mathbb{P}_{p_x}(G_x^c) \leq 3\varepsilon$ and $\mathbb{P}_{p_y}(G_y^c) \leq 3\varepsilon$, where

$$B_{x,0} = \{\|z-x\|_x \geq cr\} \text{ with } c \geq 1 + \frac{2}{\sqrt{d}}\log\frac{1}{\varepsilon}, \text{ and } B_{x,2} = \Bigl\{\|z-x\|_z^2 - \|z-x\|_x^2 > \frac{2\varepsilon r^2}{d}\Bigr\}$$

$$B_{x,3} = \Bigl\{\nabla\varphi(x)^\mathsf{T}(z-x) \leq -\frac{2r\log\frac{1}{\varepsilon}}{\sqrt{d}}\,\|g(x)^{-\frac{1}{2}}\nabla\varphi(x)\|_2\Bigr\}\,.$$

Let $G := G_x \cup G_y$, and define a partition of G by

$$G_{x\backslash y} := G_x \backslash G_y\,, \qquad G_{x,y} := G_x \cap G_y\,, \qquad G_{y\backslash x} := G_y \backslash G_x\,.$$

Now we decompose the term II as follows: for $Q := |A(x,z)p_x(z) - A(y,z)p_y(z)|$,

$$\begin{aligned}
\mathsf{II} &= \frac{1}{2}\int_{K\backslash G} Q\,\mathrm{d}z + \underbrace{\frac{1}{2}\int_{G_{x\backslash y}} Q\,\mathrm{d}z}_{=:\mathcal{A}} + \underbrace{\frac{1}{2}\int_{G_{y\backslash x}} Q\,\mathrm{d}z}_{=:\mathcal{B}} + \underbrace{\frac{1}{2}\int_{G_{x,y}} Q\,\mathrm{d}z}_{=:\mathcal{C}} \\
&\leq \frac{1}{2}\bigl(\mathbb{P}_{p_x}(K\backslash G) + \mathbb{P}_{p_y}(K\backslash G)\bigr) + \mathcal{A} + \mathcal{B} + \mathcal{C} \leq \frac{1}{2}\bigl(\mathbb{P}_{p_x}(G_x^c) + \mathbb{P}_{p_y}(G_y^c)\bigr) + \mathcal{A} + \mathcal{B} + \mathcal{C} \\
&\leq 3\varepsilon + \mathcal{A} + \mathcal{B} + \mathcal{C}\,.
\end{aligned}$$

The term $\mathcal{A}$ can be further decomposed by

$$2\mathcal{A} \leq \int_{G_{x\backslash y}} A(x,z)\,|p_x(z) - p_y(z)|\,\mathrm{d}z + \int_{G_{x\backslash y}} |A(x,z) - A(y,z)|\,p_y(\mathrm{d}z)$$

$$\leq \int_{G_{x\setminus y}} |p_x(z) - p_y(z)|\,\mathrm{d}z + \mathbb{P}_{p_y}(G_{x\setminus y}) \leq \int_{G_{x\setminus y}} |p_x(z) - p_y(z)|\,\mathrm{d}z + \underbrace{\mathbb{P}_{p_y}(G_y^c)}_{\leq 3\varepsilon},$$

and in a similar way $\mathcal{B} \leq \frac{1}{2}\int_{G_{y\setminus x}} |p_x(z) - p_y(z)|\,\mathrm{d}z + 3\varepsilon/2$. Combining these together,

$$\mathcal{A} + \mathcal{B} \leq 3\varepsilon + \frac{1}{2}\int_{G_{x\setminus y}\cup G_{y\setminus x}} |p_x(z) - p_y(z)|\,\mathrm{d}z \leq 3\varepsilon + d_{\mathsf{TV}}(p_x, p_y) \leq 4\varepsilon\,,$$

where we used $d_{\mathsf{TV}}(p_x, p_y) \leq \varepsilon$; to see this, recall Pinsker's inequality and a formula for the KL divergences between two Gaussians:

$$2[d_{\mathsf{TV}}(p_x, p_y)]^2 \leq \mathsf{KL}(p_y \,\|\, p_x) = \frac{1}{2}\Bigl(\operatorname{Tr}\bigl(g(y)^{-1}g(x)\bigr) - d + \log\det\bigl(g(y)g(x)^{-1}\bigr) + \frac{d}{r^2}\,\|y-x\|_x^2\Bigr)\,.$$

Let $\{\lambda_i\}_{i\in[d]}$ be the eigenvalues of $g(x)^{-\frac{1}{2}}g(y)g(x)^{-\frac{1}{2}}$ and $\|x-y\|_x \leq \frac{sr}{\sqrt{d}}$ with $s > 0$ to be determined. Then, $\frac{1}{2} \leq \lambda_i \leq 1 + 8\|x-y\|_x$ by Lemma D.4. Using this and $\log x \leq x - 1$ for $x > 0$,

$$2\,\mathsf{KL}(p_y \,\|\, p_x) = \sum_{i=1}^d \Bigl(\lambda_i - 1 + \log\frac{1}{\lambda_i}\Bigr) + \frac{d}{r^2}\,\|y-x\|_x^2 \leq \sum_{i=1}^d \frac{(\lambda_i - 1)^2}{\lambda_i} + s^2 \leq s^2\,(128r^2 + 1)\,,$$

Taking $s \leq s_1(\varepsilon) := \varepsilon$ and $r \leq r_6(\varepsilon)$ so that $\sqrt{128r_6^2 + 1} \leq 2$, we obtain

$$d_{\mathsf{TV}}(p_x, p_y) \leq \sqrt{\frac{1}{2}\,\mathsf{KL}(p_y \,\|\, p_x)} \leq \frac{s}{2}\sqrt{128r^2 + 1} \leq \varepsilon\,, \tag{G.5}$$

We now bound $\mathcal{C}$. Recall $B_{x,1} = \{\langle \nabla f(x), z - x\rangle \geq 0\}$ and $\mathbb{P}_{p_x}(B_{x,1}) \leq \frac{1}{2} + \mathcal{O}(\varepsilon)$. Then,

$$\begin{aligned}
2\mathcal{C} &= \int_{(G_x\cap G_y)\setminus B_{x,1}^c} Q\,\mathrm{d}z + \int_{G_x\cap G_y\cap B_{x,1}^c} Q\,\mathrm{d}z \leq \int_{B_{x,1}} \underbrace{Q}_{\text{The traingle inequality}} \mathrm{d}z + \int_{G_x\cap G_y\cap B_{x,1}^c} Q\,\mathrm{d}z \\
&\leq \int_{B_{x,1}} |A(x,z) - A(y,z)|\,p_x(\mathrm{d}z) + \int_{B_{x,1}} A(y,z)\,|p_x(z) - p_y(z)|\,\mathrm{d}z + \int_{G_x\cap G_y\cap B_{x,1}^c} Q\,\mathrm{d}z \\
&\leq \underbrace{\mathbb{P}_{p_z}(B_{x,1})}_{\leq\frac{1}{2}+\varepsilon} + 2\underbrace{d_{\mathsf{TV}}(p_x, p_y)}_{\leq\varepsilon\ (\text{G.5})} + \int_{G_x\cap G_y\cap B_{x,1}^c} |A(x,z)\,p_x(z) - A(y,z)\,p_y(z)|\,\mathrm{d}z \\
&\leq \frac{1}{2} + 2\varepsilon + \int_{G_x\cap G_y\cap B_{x,1}^c} |A(x,z)\,p_x(z) - A(y,z)\,p_y(z)|\,\mathrm{d}z\,.
\end{aligned}$$

One can check that

$$|A(x,z)\,p_x(z) - A(y,z)\,p_y(z)|\,\mathrm{d}z = \Bigl|\min\Bigl(1, \underbrace{\frac{\exp f(x)}{\exp f(z)}\,\frac{p_z(x)}{p_x(z)}}_{=:\mathsf{U}}\Bigr) - \min\Bigl(\underbrace{\frac{p_y(z)}{p_x(z)}}_{=:\mathsf{V}}, \underbrace{\frac{\exp f(y)}{\exp f(z)}\,\frac{p_z(y)}{p_x(z)}}_{=:\mathsf{W}}\Bigr)\Bigr|\,p_x(\mathrm{d}z)\,.$$

Here we note that $\mathsf{U} \geq e^{-4\varepsilon}$ due to $\frac{\exp f(x)}{\exp f(z)} \geq e^{-\varepsilon}$ and $\frac{p_z(x)}{p_x(z)} \geq e^{-3\varepsilon}$ from (G.2) and (G.3).

We now show that under additional conditioning, $|\log\mathsf{V}|\lesssim\varepsilon$ and $\log\mathsf{W}\gtrsim-\varepsilon$ on $z\in G_x\cap G_y\cap B^c_{x,1}$. For $\varphi(\cdot)=\frac{1}{2}\log\det g(\cdot)$ and $\mathsf{L}:=-\frac{d}{2r^2}(\|z-y\|_y^2-\|z-x\|_x^2)$,

$$\begin{aligned}\log\mathsf{V}&=-\frac{d}{2r^2}(\|z-y\|_y^2-\|z-x\|_x^2)+\varphi(y)-\varphi(x)\\&=\mathsf{L}+\langle\nabla\varphi(x),y-x\rangle+\underbrace{\frac{1}{2}\langle\nabla^2\varphi(x^*)(y-x),y-x\rangle}_{\text{Use (G.1)}}\quad\text{for some }x^*\in[x,y]\qquad\text{(G.6)}\\&\geq\mathsf{L}-\|g(x)^{-1/2}\nabla\varphi(x)\|_2\|y-x\|_x-5\underbrace{(1+2\|x-y\|_x)^2}_{\leq 2}\|y-x\|_x^2\\&\geq\mathsf{L}-2\sqrt{d}\cdot s\frac{r}{\sqrt{d}}-10s^2\frac{r^2}{d}\geq\mathsf{L}-\varepsilon\,,\qquad\text{(G.7)}\end{aligned}$$

where the inequality follows from $s\leq\frac{\varepsilon}{10}$ and $r\leq r_7(\varepsilon):=1$.

As for W, due to $f(y)\geq f(x)$ and $\exp(f(x)-f(z))\geq\exp(-\varepsilon)$,

$$\begin{aligned}\log\mathsf{W}&\geq\log\Big(\frac{\exp f(x)}{\exp f(z)}\frac{p_z(y)}{p_x(z)}\Big)\geq-\varepsilon-\frac{d}{2r^2}(\|z-y\|_z^2-\|z-x\|_x^2)+\varphi(z)-\varphi(x)\\&\underset{\text{(i)}}{\geq}-\varepsilon-\frac{d}{2r^2}\Big(\|z-y\|_y^2+2\varepsilon\frac{r^2}{d}-\|z-x\|_x^2\Big)-2\varepsilon=\mathsf{L}-4\varepsilon\,,\qquad\text{(G.8)}\end{aligned}$$

where (i) follows from $\|z-y\|_z^2-\|z-y\|_y^2\leq 2\varepsilon r^2/d$ on $z\in B^c_{y,2}$, and $\varphi(z)-\varphi(x)\geq-2\varepsilon$ on $z\in B^c_{x,3}$ from (G.4).

Lastly, we show that $|\mathsf{L}|$ is bounded by $\mathcal{O}(\varepsilon)$ with high probability (w.r.t. p_x). Due to affine invariance of the algorithm, we may assume that $x=0$ and $g(x)=I_d$ (so $p_x=\mathcal{N}(0,I_d)$). Therefore,

$$\|z-y\|_y^2-\|z-x\|_x^2=\|z-y\|_y^2-\|z\|^2=\|z\|^2_{g(y)-I_d}-2\langle z,y\rangle_y+\|y\|_y^2\,.$$

The last term is bounded by $2\|y\|^2$ due to SC of g. Using a tail bound for Gaussians, we have $\mathbb{P}_{p_x}\big(|\langle z,y\rangle_y|\geq\frac{r}{\sqrt{d}}\|g(y)y\|_2\cdot 2\log\frac{1}{\varepsilon}\big)\leq\varepsilon$ and call this event C_1. In addition, SC of g leads to $g(y)\preceq 2I_d$, so $\|g(y)y\|\leq 2\|y\|$.

To bound $\|z\|^2_{g(y)-I_d}$, we note that $\|y\|=\|y-x\|_x\leq 1/\sqrt{2}$ and so

$$\begin{aligned}\|g(y)-I_d\|_F&\leq(1+2\|y\|)^2\|y\|\leq 2s\frac{r}{\sqrt{d}}\,,\quad\text{(Lemma D.6)}\\\mathbb{E}[\|z\|^2_{g(y)-I_d}]&=\frac{r^2}{d}\operatorname{Tr}(g(y)-I_d)\leq\frac{r^2}{d}\sqrt{d}\,\|g(y)-I_d\|_F\leq\frac{r^2}{d}\cdot 2rs\,.\end{aligned}$$

By the Hanson-Wright inequality[6], for universal constants $K_1,K_2>0$ and $t\geq 0$ it holds that

$$\mathbb{P}_{z\sim\mathcal{N}(0,I_d)}\big(\big|\|z\|^2_{g(y)-I_d}-\mathbb{E}[\|z\|^2_{g(y)-I_d}]\big|\geq t\big)\leq 2\exp\Big(-K_1\Big(\frac{t^2}{K_2^4\frac{r^4}{d^2}\|g(y)-I_d\|_F^2}\wedge\frac{t}{K_2^2\frac{r^2}{d}\|g(y)\|_2}\Big)\Big)\,.$$

6.

Lemma (Hanson-Wright; Adapted to Gaussian). Let $h\sim\mathcal{N}(0,\sigma^2 I_d)$ and $M\in\mathbb{R}^{d\times d}$. Then there exists universal constants $c,K>0$ such that for $t\geq 0$

$$\mathbb{P}\big(\big|\|h\|_A^2-\mathbb{E}[\|h\|_A^2]\big|>t\big)\leq 2\exp\Big(-c\min\Big(\frac{t^2}{K^4\sigma^4\|M\|_F^2},\frac{t}{K^2\sigma^2\|M\|_2}\Big)\Big)\,.$$

By taking $r \le r_8(\varepsilon) := \frac{\sqrt{K_1}}{2K_2^2}$ and $s \le s_2(\varepsilon) := \varepsilon(1+\sqrt{\log\frac{2}{\varepsilon}})^{-1}$, it follows that $\|z\|_{g(y)-I_d}^2 \le \frac{2\varepsilon r^2}{d}$ with probability at least $1-\varepsilon$. Denote the complement of this event by C_2.

Conditioned on $z \in C_1^c \cap C_2^c$, we conclude that

$$\left|\|z-y\|_y^2 - \|z-x\|_x^2\right| \le \|z\|_{g(y)-I_d}^2 + 2|\langle z,y\rangle_y| + 2\|y\|^2 \le \frac{2r^2\varepsilon}{d} + \frac{8r\|y\|}{\sqrt{d}}\log\frac{1}{\varepsilon} + 2\|y\|^2 \le \frac{2r^2}{d}\cdot 3\varepsilon\,,$$

where the last inequality follows from $\|y\| \le \frac{sr}{\sqrt{d}}$ when $s \le s_3(\varepsilon) := \varepsilon\,(4\log\frac{1}{\varepsilon})^{-1}$. Hence, $|\mathsf{L}| \le 3\varepsilon$ on $C_1^c \cap C_2^c$. Putting this into (G.7) and (G.8),

$$\log \mathsf{V} \ge \exp(-4\varepsilon) \qquad \text{and} \qquad \log \mathsf{W} \ge \exp(-7\varepsilon)\,.$$

We can also show $\log \mathsf{V} \le 5\varepsilon$. Conditioned on $z \in C_1^c \cap C_2^c$,

$$-\log \mathsf{V} = -\mathsf{L} + \varphi(x) - \varphi(y) \ge -3\varepsilon + \varphi(x) - \varphi(y) \ge -5\varepsilon\,,$$

since $\varphi(x)-\varphi(y)$ can be lowered bounded by -2ε as in (G.6). Hence, $\log \mathsf{V} \le 5\varepsilon$.

For $F := G_x \cap G_y \cap B_{x,1}^c$ and $C := (C_1 \cup C_2)^c$, since $e^{-4\varepsilon} \le \mathsf{V} \le e^{5\varepsilon}$, $e^{-7\varepsilon} \le \mathsf{W}$, and $\mathsf{U} \ge e^{-4\varepsilon}$,

$$\begin{aligned}
&\int_F |A(x,z)\,p_x(z) - A(y,z)\,p_y(z)|\,\mathrm{d}z \le \int_{C^c} (\cdot)\,\mathrm{d}z + \int_{F\cap C} (\cdot)\,\mathrm{d}z \\
&\le \underbrace{\mathbb{P}_{p_x}(C^c)}_{\le 2\varepsilon} + 2\underbrace{d_{\mathrm{TV}}(p_x,p_y)}_{\le\varepsilon} + \int_{F\cap C} (\cdot)\,\mathrm{d}z \le 4\varepsilon + \int_{F\cap C} |1\wedge \mathsf{U} - \mathsf{V}\wedge\mathsf{W}|\,p_x(\mathrm{d}z) \le 4\varepsilon + (e^{5\varepsilon} - e^{-4\varepsilon}) \\
&\le 18\varepsilon\,.
\end{aligned}$$

Using this, we can bound $\mathcal{C}$ by

$$\mathcal{C} \le \frac{1}{4} + \varepsilon + \frac{1}{2}\int_F |A(x,z)\,p_x(z) - A(y,z)\,p_y(z)|\,\mathrm{d}z \le \frac{1}{4} + 10\varepsilon\,.$$

Therefore, $\mathsf{II} \le 3\varepsilon + \mathcal{A} + \mathcal{B} + \mathcal{C} \le 3\varepsilon + 4\varepsilon + \frac{1}{4} + 10\varepsilon \le \frac{1}{4} + 17\varepsilon$. Along with $\mathsf{I} \le \frac{1}{2} + 5\varepsilon$, we can conclude that if $r \le \min_i r_i(\varepsilon)$ and $s \le \min_i s_i(\varepsilon)$, then $d_{\mathrm{TV}}(P_x,P_y) \le \frac{3}{4} + 23\varepsilon$. ■

G.1.2. ISOPERIMETRIC INEQUALITY

We now prove an isoperimetric inequality arising from the a SC barrier. Recall the *cross-ratio distance* d_K defined on a convex body K: for $x, y \in \mathrm{int}(K)$, suppose that the chord passing through x, y has endpoints p and q in the boundary ∂K (so the order of points is p, x, y, q), then the cross-ratio distance between x and y is defined by

$$d_K(x,y) \overset{\text{def}}{=} \frac{\|x-y\|_2\|p-q\|_2}{\|p-x\|_2\|y-q\|_2}\,.$$

The first type of isoperimetric inequalities says $\psi_\pi \gtrsim 1/\sqrt{\bar\nu}$.

Proof of Lemma B.6. For a ball $B_r(0)$ of radius $r>0$ centered at the origin, we define a convex body $K_r := K \cap B_r(0)$ and use π_r to denote the truncated distribution of π over K_r. Let $\{S_1, S_2, S_3\}$

be a partition of K and define $S_i^r := S_i \cap K_r$ for $i \in [3]$. By Lovász and Vempala (2007, Theorem 2.5), we have

$$\pi_r(S_3^r) \geq d_{K_r}(S_1^r, S_2^r)\, \pi_r(S_1^r)\, \pi_r(S_2^r)\,,$$

where $d_{K_r}(S_1^r, S_2^r) = \inf_{x\in S_1^r, y\in S_2^r} d_{K_r}(x,y)$. Due to $d_{K_r}(x,y) \geq \|x-y\|_x/\sqrt{\bar\nu}$ for any $x, y \in K_r$ (see Laddha et al. (2020, Lemma 2.3)),

$$\pi_r(S_3^r) \geq \inf_{x\in S_1^r,\, y\in S_2^r} \frac{\|x-y\|_x}{\sqrt{\bar\nu}}\, \pi_r(S_1^r)\, \pi_r(S_2^r) \geq \frac{1}{\sqrt{\bar\nu}} \inf_{x\in S_1,\, y\in S_2} \|x-y\|_x\, \pi_r(S_1^r)\, \pi_r(S_2^r)\,.$$

As $r \to \infty$, the bounded convergence theorem implies $\pi_r(S_i^r) \to \pi(S_i)$ for $i \in [3]$, completing the proof. ■

We provide the deferred proof for another isoperimetric inequality, $\psi_\pi \gtrsim \sqrt{\alpha}$, originating from α-relatively strong-convexity of the potential with respect to $\nabla^2\phi$.

Proof of Lemma B.7. The proof essentially follows Gopi et al. (2023). Their first proof ingredient is a modified localization lemma (Gopi et al., 2023, Lemma 8); let f_1, f_2, f_3, f_4 be non-negative functions on $\mathbb{R}^d$ such that f_1 and f_2 are upper semicontinuous, and f_3 and f_4 are lower semicontinuous, and $\phi : \mathbb{R}^d \to \mathbb{R}$ be convex. Then the following are equivalent:

- For any density $\pi : \mathbb{R}^d \to \mathbb{R}$ which is 1-relatively strongly logconcave in ϕ,

$$\int f_1\, \mathrm{d}\pi \cdot \int f_2\, \mathrm{d}\pi \leq \int f_3\, \mathrm{d}\pi \cdot \int f_4\, \mathrm{d}\pi\,.$$

- Let $\int_E h := \int_0^1 h((1-t)\,a + tb) e^{-\gamma t}\, \mathrm{d}t$. Then $\int_E f_1 e^{-\phi} \cdot \int_E f_2 e^{-\phi} \leq \int_E f_3 e^{-\phi} \cdot \int_E f_4 e^{-\phi}$ for any $a, b \in \mathbb{R}^d$ and $\gamma \in \mathbb{R}$.

First of all, this can be generalized to an extended convex function f and ϕ, whose values outside of $\operatorname{int}(K)$ are set to ∞. Since the density π and a needle $\exp(\gamma t - \phi((1-t)a + tb))$ for $\gamma \in \mathbb{R}$ and $a, b \in \mathbb{R}^d$ (induced by the extended f and ϕ) vanish outside of $\operatorname{int}(K)$, integrands above become zero on $\operatorname{int}(K)^c$, and thus the integrals above remain the same.

As in Gopi et al. (2023, Lemma 9), the proof boils down to the case of $\alpha = 1$, and it suffices to show that there exists a constant $C > 0$ such that

$$C \cdot d_\phi(S_1, S_2) \int_{S_1} e^{-f} \cdot \int_{S_2} e^{-f} \leq \int e^{-f} \int_{S_3} e^{-f}\,.$$

We can replace $S_i \leftarrow$ its closure $\bar{S}_i$ for $i \in [2]$, which only increases the LHS. Also, we can replace $S_3 \leftarrow$ an open set $\operatorname{int}(K)\backslash \bar{S}_1 \backslash \bar{S}_2$, which does not change the RHS since the boundary of a convex set is a null set (Lang, 1986, Theorem 1). By taking $f_i = \mathbf{1}_{S_i}$ for $i \in [3]$ and $f_4 = (C\, d_\phi(S_1, S_2))^{-1}$, we only need to show that for some $0 \leq c < d \leq 1$,

$$C \cdot d_\phi(S_1, S_2) \int_c^d e^{\gamma t - \phi((1-t)\,a + tb)} \mathbf{1}_{S_1}((1-t)\,a + tb)\, \mathrm{d}t \cdot \int_c^d e^{\gamma t - \phi((1-t)\,a + tb)} \mathbf{1}_{S_2}((1-t)\,a + tb)\, \mathrm{d}t$$
$$\leq \int_c^d e^{\gamma t - \phi((1-t)\,a + tb)}\, \mathrm{d}t \cdot \int_c^d e^{\gamma t - \phi((1-t)\,a + tb)} \mathbf{1}_{S_3}((1-t)\,a + tb)\, \mathrm{d}t\,,$$

where $\phi((1-t)\,a + b) < \infty$ for $t \in (c, d)$. The rest of the proof is similar to Gopi et al. (2023, Lemma 9). ■

G.2. Sampling IPM (§C)

G.2.1. WELL-DEFINEDNESS OF SAMPLING IPM

Proposition G.1 *Let $p:\mathbb{R}^d\to\mathbb{R}$ be a log-concave density with finite second moment. Then p is bounded on $\mathbb{R}^d$.*

Proof Let $X\sim p$ and denote the mean and covariance of the distribution p by $\mu:=\mathbb{E}[X]$ and $\Sigma:=\mathbb{E}[(X-\mu)(X-\mu)^\mathsf{T}]$. Then the pushforward $T_\# p$ of p via the map $T:x\mapsto y:=\Sigma^{-1/2}(x-\mu)$ is an isotropic log-concave, and satisfy $(T_\# p)(y)=\frac{p(x)}{|\det T|}$. Since $T_\# p$ is bounded on $\mathbb{R}^d$ (Lovász and Vempala, 2007, Theorem 5.14 (e)), p is bounded as well. ■

Next, we show that every measure appearing within the sampling IPM is integrable.

Proof of Proposition C.1. Recall that we may assume $\phi\ge 0$. Hence, all μ_i's in Phase 3 and 4 are well-defined

$$\int_K \exp\Bigl(-\bigl(f(x)+\frac{\phi(x)}{\sigma_i^2}\bigr)\Bigr)\,\mathrm{d}x\le\int_K\exp(-f(x))\,\mathrm{d}x<\infty\,.$$

In particular, $\exp\bigl(-(f+\frac{\phi}{\nu/d})\bigr)$ is integrable with finite second moment. By Proposition G.1, $f(x)+\frac{\phi(x)}{\nu/d}$ achieves a global minimum m in K. As $\sigma_i^2\le\sigma_{i_0}^2=\nu/d$ in Phase 2, we have

$$\begin{aligned}\int_K\exp\Bigl(-\frac{\sigma_{i_0}^2 f+\phi}{\sigma_{i_0}^2}\Bigr)&=\int_K\exp\Bigl(-\frac{\sigma_{i_0}^2 f+\phi-\min(\sigma_{i_0}^2 f+\phi)}{\sigma_{i_0}^2}-\frac{\min(\sigma_{i_0}^2 f+\phi)}{\sigma_{i_0}^2}\Bigr)\\ &\ge\int_K\exp\Bigl(-\frac{\bar f+\phi-\sigma_{i_0}^2 m}{\sigma_i^2}-m\Bigr)=\exp\Bigl(m\bigl(\frac{\sigma_{i_0}^2}{\sigma_i^2}-1\bigr)\Bigr)\int_K\exp\Bigl(-\frac{\bar f+\phi}{\sigma_i^2}\Bigr)\,,\end{aligned}$$

where the inequality holds due to $\min(\sigma_{i_0}^2 f+\phi)=\sigma_{i_0}^2 m$ and $\bar f=\sigma_{i_0}^2 f$. Therefore, μ_i's in Phase 2 are also well-defined. ■

G.2.2. CLOSENESS OF DISTRIBUTIONS IN SAMPLING IPM

We begin with closeness between $\mathcal{N}\bigl(x^*,\frac{\sigma_0^2}{1+\nu\beta d^{-1}}g(x^*)^{-1}\bigr)\cdot\mathbf{1}_{\mathcal{D}_g^{3\sigma_0\sqrt d}(x^*)}$ and $\exp\bigl(-\frac{\bar f+\phi}{\sigma_0^2}\bigr)$ in Phase 1.

Proof of Lemma C.5. Let $\gamma=9$, $r=(\gamma\sigma_0^2 d)^{1/2}<0.01$, $\psi:=\bar f+\phi$, and $S=\{x\in K:\psi(x)\le\psi(x^*)+r^2/4\}$. For $\widetilde\mu_0=\exp(-\psi/\sigma_0^2)\cdot\mathbf{1}_K\propto\mu_0$ and $x\in S$, we have $\mu_0(x)\ge e^{-\gamma d}\mu_0(x^*)$. Due to $\mu_0(S^c)\le\exp(-\gamma d/3)$ (Lemma C.3), $1=\mu_0(S)+\mu_0(S^c)\le\mu_0(S)+\exp(-\gamma d/3)$ and

$$1\le\bigl(1+2\exp(-\gamma d/3)\bigr)\,\mu_0(S)=\bigl(1+2\exp(-\gamma d/3)\bigr)\,\widetilde\mu_0(S)/\widetilde\mu_0(\mathbb{R}^d)\,. \tag{G.9}$$

We show $S\subset D=\mathcal{D}_g^{3\sigma_0\sqrt d}(x^*)$. For $x\in S$, use Taylor's expansion of ψ at x^*: for some $\bar x\in[x^*,x]$

$$\psi(x)-\psi(x^*)=\frac12(x-x^*)^\mathsf{T}\nabla^2\psi(\bar x)(x-x^*)\ge\frac12(x-x^*)^\mathsf{T}\nabla^2\phi(\bar x)(x-x^*)\,. \tag{G.10}$$

As $\psi(x)-\psi(x^*)\le r^2/4$ on $x\in S$, we have $\|\bar{x}-x^*\|_{\bar{x}}^2\le\|x-x^*\|_{\bar{x}}^2\le 2(\psi(x)-\psi(x^*))\le r^2/2$. Thus, by self-concordance of ϕ

$$\exp(-3r)\,\|x-x^*\|_{x^*}^2\le\|x-x^*\|_{\bar{x}}^2\le\exp(3r)\,\|x-x^*\|_{x^*}^2\,, \tag{G.11}$$

and it follows that $\|x-x^*\|_{x^*}^2\le r^2$, showing $S\subset D$.

Combining (G.10), (G.11), and $(1+\nu\alpha d^{-1})\,\nabla^2\phi\preceq\nabla^2\psi\preceq(1+\nu\beta d^{-1})\,\nabla^2\phi$, we have

$$\frac{\exp(-3r)}{2}\Big(1+\frac{\nu\alpha}{d}\Big)\,\|x-x^*\|_{x^*}^2\underset{(*)}{\le}\psi(x)-\psi(x^*)\underset{(\#)}{\le}\frac{\exp(3r)}{2}\Big(1+\frac{\nu\beta}{d}\Big)\,\|x-x^*\|_{x^*}^2\,, \tag{G.12}$$

and thus for a constant $c:=1+\nu\beta d^{-1}$ and function $h(x):=-(2\sigma_0^2)^{-1}\|x-x^*\|_{x^*}^2$,

$$\begin{aligned}
\|\mu/\mu_0\| &= \mathbb{E}_\mu\Big[\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}\Big]=\frac{\int_D\exp\Big(-\frac{c}{\sigma_0^2}\|x-x^*\|_{x^*}^2+\frac{\psi}{\sigma_0^2}\Big)\cdot\widetilde{\mu}_0(\mathbb{R}^d)}{\Big[\int_D\exp\Big(-\frac{c}{2\sigma_0^2}\,\|x-x^*\|_{x^*}^2\Big)\Big]^2}\\
&\underset{(\text{G.9})}{\le}\frac{1}{\Big[\int_D\exp(c\cdot h)\Big]^2}\int_D\exp\Big(-\frac{c}{\sigma_0^2}\|x-x^*\|_{x^*}^2+\underbrace{\frac{\psi}{\sigma_0^2}}_{\text{Use }(\#)\text{ in (G.12)}}\Big)\big(1+2\exp(-\gamma n/3)\big)\underbrace{\widetilde{\mu}_0(S)}_{\text{Use }(*)}\\
&\lesssim\frac{\int_D\exp\Big(-\frac{1}{2\sigma_0^2}\big(2c-e^{3r}(1+\nu\beta d^{-1})\big)\,\|x-x^*\|_{x^*}^2\Big)\int_D\exp\big(-\frac{1}{2\sigma_0^2}e^{-3r}(1+\nu\alpha d^{-1})\,\|x-x^*\|_{x^*}^2\big)}{\Big[\int_D\exp(c\cdot h)\Big]^2}\\
&=\underbrace{\frac{\int_D\exp\Big(\big(2c-c\,e^{3r}\big)\,h(x)\Big)\cdot\int_D\exp\big(c\,e^{3r}h(x)\big)}{\Big[\int_D\exp(c\cdot h)\Big]^2}}_{=:\mathsf{A}}\;\underbrace{\frac{\int_D\exp\big(e^{-3r}(1+\nu\alpha d^{-1})\,h(x)\big)}{\int_D\exp\big(c\,e^{3r}h(x)\big)}}_{=:\mathsf{B}}\,.
\end{aligned}$$

As for A, Lemma C.2 leads to

$$\mathsf{A}\le\Big(\frac{c^2}{(2c-c\,e^{3r})\,ce^{3r}}\Big)^d=\Big(\frac{1}{(2-e^{3r})e^{3r}}\Big)^d=(1+\mathcal{O}(r^2))^d=\mathcal{O}(1)\,.$$

As for B, let $c_1=e^{-3r}\,(1+\nu\alpha d^{-1})$ and $c_2=e^{3r}\,(1+\nu\beta d^{-1})$. With the change of variable $y=\sigma_0^{-1}\sqrt{c_i}g(x^*)^{1/2}(x-x^*)$ for $i\in[2]$, it follows that for $r_i:=r\sigma_0^{-1}\sqrt{c_i}(\ge 3\sqrt{d})$

$$\mathsf{B}=\Big(\frac{c_2}{c_1}\Big)^{d/2}\frac{\int_{B_{r_1}}\exp(-\frac{1}{2}\|y\|^2)\,\mathrm{d}y}{\int_{B_{r_2}}\exp(-\frac{1}{2}\|y\|^2)\,\mathrm{d}y}\le\Big(\frac{c_2}{c_1}\Big)^{d/2}\lesssim\Big(\frac{\nu\beta+d}{\nu\alpha+d}\Big)^d e^{3rd}\lesssim\Big(\frac{\nu\beta+d}{\nu\alpha+d}\Big)^d\,.$$

■

Now we show closeness of two consecutive distributions in Phase 2, i.e., $\sigma_{i+1}^2=\sigma_i^2\big(1+\frac{1}{\sqrt{d}}\big)$.

Proof of Lemma C.6. Observe that for $\psi=\bar{f}+\phi=\frac{\nu}{d}f+\phi$ on K and $F(\sigma^2)=\int_K\exp(-\psi/\sigma^2)$,

$$\|\mu_i/\mu_{i+1}\|=\mathbb{E}_{\mu_i}\Big[\frac{\mathrm{d}\mu_i}{\mathrm{d}\mu_{i+1}}\Big]=\frac{\int_K\exp\big(-2\frac{\psi}{\sigma_i^2}+\frac{\psi}{\sigma_{i+1}^2}\big)\cdot\int_K\exp\big(-\frac{\psi}{\sigma_{i+1}^2}\big)}{\Big(\int_K\exp\big(-\frac{\psi}{\sigma_i^2}\big)\Big)^2}=\frac{F\big(\big(\frac{2}{\sigma_i^2}-\frac{1}{\sigma_{i+1}^2}\big)^{-1}\big)\,F(\sigma_{i+1}^2)}{F(\sigma_i^2)^2}\,.$$

By Lemma C.2, the function $a^d F\big(\frac{\sigma^2}{a}\big)$ is log-concave in a. Using the definition with endpoints $\frac{2}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2}$ and $\frac{1}{\sigma_{i+1}^2}$, and the middle point $\frac{1}{\sigma_i^2}$, we obtain

$$\frac{F\big(\big(\frac{2}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2}\big)^{-1}\big)\, F(\sigma_{i+1}^2)}{F(\sigma_i^2)^2} \le \left(\frac{\big(\frac{1}{\sigma_i^2}\big)^2}{\big(\frac{2}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2}\big)\,\frac{1}{\sigma_{i+1}^2}}\right)^d = \left(\frac{\big(1+\frac{1}{\sqrt{d}}\big)^2}{1+\frac{2}{\sqrt{d}}}\right)^d \le \Big(1+\frac{1}{d}\Big)^d \le e\,.$$

■

We now establish closeness in Phase 3, during which we use the update of $\sigma_{i+1}^2 = \sigma_i^2\big(1+\frac{\sigma_i}{\sqrt{\nu}}\big)$.

Proof of Lemma C.7. The update is $\sigma_{i+1}^2 = \sigma_i^2\,(1+r)$ for $r = \frac{\sigma_i}{\sqrt{\nu}}$. For $s := \frac{r}{1+r}$, $\sigma := \sigma_i$, and $F(\sigma^2) = \int \exp(-f - \phi/\sigma^2)$, we have

$$\|\mu_i/\mu_{i+1}\| = \frac{F\big(\big(\frac{2}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2}\big)^{-1}\big)\, F(\sigma_{i+1}^2)}{F(\sigma_i^2)^2} = \frac{F\big(\frac{\sigma^2}{1+s}\big)\, F\big(\frac{\sigma^2}{1-s}\big)}{F(\sigma^2)^2}\,.$$

Let $g(t) := \log F\big(\frac{\sigma^2}{t}\big)$ for $t > 0$. Then,

$$\log \|\mu_i/\mu_{i+1}\| = g(1+s) + g(1-s) - 2g(1) = \int_0^s \big(g'(1+t) - g'(1-t)\big)\,\mathrm{d}t = \int_0^s \int_{1-t}^{1+t} g''(q)\,\mathrm{d}q\,\mathrm{d}t \tag{G.13}$$

and for a probability measure $\nu_q \propto \exp\big(-f - \frac{q\phi}{\sigma^2}\big)$,

$$\begin{aligned} g''(q) &= \frac{\mathrm{d}^2}{\mathrm{d}q^2}\Big[\log \int_K \exp\Big(-f - \frac{q\phi}{\sigma^2}\Big)\Big] = -\frac{1}{\sigma^2}\,\frac{\mathrm{d}}{\mathrm{d}q}\left[\frac{\int_K \phi \cdot \exp\Big(-f - \frac{q\phi}{\sigma^2}\Big)}{\int_K \exp\Big(-f - \frac{q\phi}{\sigma^2}\Big)}\right] \\ &= -\frac{1}{\sigma^2}\left(-\frac{1}{\sigma^2}\,\frac{\int_K \phi^2 \cdot \exp\Big(-f - \frac{q\phi}{\sigma^2}\Big)}{\int_K \exp\Big(-f - \frac{q\phi}{\sigma^2}\Big)} + \frac{1}{\sigma^2}\,\frac{\Big[\int_K \phi \cdot \exp\Big(-f - \frac{q\phi}{\sigma^2}\Big)\Big]^2}{\Big[\int_K \exp\Big(-f - \frac{q\phi}{\sigma^2}\Big)\Big]^2}\right) \\ &= \frac{1}{\sigma^4}\Big(\mathbb{E}_{\nu_q}[\phi^2] - (\mathbb{E}_{\nu_q}\phi)^2\Big) = \frac{1}{\sigma^4}\,\mathrm{Var}_{\nu_q}\phi\,. \end{aligned}$$

By the Brascamp-Lieb inequality with $V(\cdot) := f(\cdot) + \frac{q\phi(\cdot)}{\sigma^2}$,

$$\mathrm{Var}_{\nu_q}\phi \le \mathbb{E}_{\nu_q}\big[(\nabla\phi)^{\mathsf{T}}\big(\nabla^2 V\big)^{-1}\nabla\phi\big] \le \frac{\sigma^2}{q}\,\mathbb{E}_{\nu_q}\|\nabla\phi\|^2_{(\nabla^2\phi)^{-1}} \le \frac{\sigma^2\nu}{q}\,,$$

and thus $g''(q) \le \frac{\nu}{q\sigma^2}$. Putting this back to (G.13), we acquire

$$\log\|\mu_i/\mu_{i+1}\| \le \frac{\nu}{\sigma^2}\int_0^s \int_{1-t}^{1+t} \frac{1}{q}\,\mathrm{d}q\,\mathrm{d}t = \frac{\nu}{\sigma^2}\int_0^s \big(\log(1+t) - \log(1-t)\big)\,\mathrm{d}t$$

$$= \frac{\nu}{\sigma^2}\big((1+s)\,\log(1+s) + (1-s)\,\log(1-s)\big) \lesssim \frac{\nu s^2}{\sigma^2}\,. \tag{G.14}$$

It follows from $s = \frac{r}{1+r}$ and $r = \frac{\sigma}{\sqrt{\nu}}$ that μ_i is an $\mathcal{O}(1)$-warm start for μ_{i+1}.

For Phase 4, observe that for $\mu \propto \exp(-f - \phi/\sigma^2)$ with $\sigma^2 = \nu$,

$$\begin{aligned}
\|\mu/\pi\| &= \frac{\int_K \exp\big(-f - \frac{\phi}{\sigma^2/2}\big)\cdot \int_K \exp(-f)}{\Big[\int_K \exp\Big(-f - \frac{\phi}{\sigma^2}\Big)\Big]^2} \underset{\text{(i)}}{=} \lim_{r\to 1} \frac{F\big(\frac{\sigma^2}{1+r}\big)\cdot F\big(\frac{\sigma^2}{1-r}\big)}{F(\sigma^2)} \\
&\underset{\text{(ii)}}{\le} \lim_{r\to 1} \exp\Big(\mathcal{O}(1)\frac{\nu}{\sigma^2}\,\big((1+r)\,\log(1+r) + (1-r)\,\log(1-r)\big)\Big) \\
&= \exp\Big(\mathcal{O}(1)\frac{\nu}{\sigma^2}\Big) = \exp(\mathcal{O}(1))\,.
\end{aligned}$$

where (i) holds due to the monotone convergence theorem, and (ii) follows from (G.14). Therefore, μ serves as an $\mathcal{O}(1)$-warm start for π. ■

G.3. Self-concordance theory (§D)

G.3.1. BASIC PROPERTIES: STRONG SELF-CONCORDANCE

We show that $2(g_1 + g_2)$ is SSC if g_1 and g_2 are SSC.

Proof of Lemma D.5. For fixed $x \in K_1 \cap K_2$ and $h \in \mathbb{R}^d$, let $\mathrm{D}g_i := \mathrm{D}g_i(x)[h]$ for $i = 1, 2$. Note that

$$\begin{aligned}
&\|(g_1+g_2)^{-\frac12}\mathrm{D}(g_1+g_2)\,(g_1+g_2)^{-\frac12}\|_F \\
&\le \sum_{i=1}^2 \|(g_1+g_2)^{-\frac12}\mathrm{D}g_i\,(g_1+g_2)^{-\frac12}\|_F = \sum_{i=1}^2 \sqrt{\mathrm{Tr}\big((g_1+g_2)^{-1}\mathrm{D}g_i\,(g_1+g_2)^{-1}\mathrm{D}g_i\big)} \\
&= \Big[\mathrm{Tr}\Big(\big(\underbrace{I + g_1^{-\frac12}g_2g_1^{-\frac12}}_{=:E_1}\big)^{-1}\underbrace{g_1^{-\frac12}\mathrm{D}g_1\,g_1^{-\frac12}}_{=:T_1}\big(I + g_1^{-\frac12}g_2g_1^{-\frac12}\big)^{-1}g_1^{-\frac12}\mathrm{D}g_1\,g_1^{-\frac12}\Big)\Big]^{1/2} \\
&\quad + \Big[\mathrm{Tr}\Big(\big(\underbrace{I + g_2^{-\frac12}g_1g_2^{-\frac12}}_{=:E_2}\big)^{-1}\underbrace{g_2^{-\frac12}\mathrm{D}g_2\,g_2^{-\frac12}}_{=:T_2}\big(I + g_2^{-\frac12}g_1g_2^{-\frac12}\big)^{-1}g_2^{-\frac12}\mathrm{D}g_2\,g_2^{-\frac12}\Big)\Big]^{1/2} \\
&= \sum_{i=1}^2 \sqrt{\mathrm{Tr}(E_i^{-1}T_iE_i^{-1}T_i)} \le \sum_{i=1}^2 \sqrt{\mathrm{Tr}(T_iE_i^{-2}T_i)}\,,
\end{aligned}$$

where we used the Cauchy-Schwarz inequality $\mathrm{Tr}(A^2) \le \mathrm{Tr}(A^\mathsf{T}A)$ in the last line. It follows from $I \preceq E_i$ that $I \preceq E_i^2$ and $I \succeq E_i^{-2} \succ 0$. Therefore,

$$\sum_{i=1}^2 \sqrt{\mathrm{Tr}(T_iE_i^{-2}T_i)} \le \sum_{i=1}^2 \|T_i\|_F \le 2\sum_{i=1}^2 \|h\|_{g_i(x)}^2 \le 2\sqrt{2}\|h\|_{(g_1+g_2)(x)}\,.$$

Putting these together completes the proof. ■

G.3.2. BASIC PROPERTIES: LOWER TRACE SELF-CONCORDANCE

We now show that if g is HSC, then dg is SLTSC.

Proof of Lemma D.13. We first consider when $\bar{g}$ is positive definite on K. By HSC of $\bar{g}$, it holds that $-\|h\|_{\bar{g}}^2\,\bar{g} \lesssim \mathrm{D}^2\bar{g}[h,h]$, and thus

$$-\frac{1}{d}\,\|h\|_g^2\,(g'+g)^{-\frac{1}{2}}g\,(g'+g)^{-\frac{1}{2}} \lesssim (g'+g)^{-\frac{1}{2}}\mathrm{D}^2 g[h,h]\,(g'+g)^{-\frac{1}{2}}\,.$$

Hence,

$$\begin{aligned}\mathrm{Tr}\big((g'+g)^{-1}\mathrm{D}^2 g[h,h]\big) &\gtrsim -\frac{1}{d}\,\|h\|_g^2\,\mathrm{Tr}\Big((g'+g)^{-\frac{1}{2}}g\,(g'+g)^{-\frac{1}{2}}\Big) = -\frac{1}{d}\,\|h\|_g^2\,\mathrm{Tr}\big(g^{\frac{1}{2}}(g'+g)^{-1}g^{\frac{1}{2}}\big)\\ &\geq -\frac{1}{d}\,\|h\|_g^2\,\mathrm{Tr}(g^{\frac{1}{2}}g^{-1}g^{\frac{1}{2}}) = -\|h\|_g^2\,.\end{aligned}$$

When g is singular, we consider $\bar{g}_\varepsilon = \bar{g} + \frac{\varepsilon}{d}I \in \mathbb{S}_{++}^d$ for $\varepsilon > 0$. Then $\bar{g}_\varepsilon$ is HSC, so for $g_\varepsilon = d\bar{g}_\varepsilon$

$$\mathrm{Tr}\big((g'+g_\varepsilon)^{-1}\mathrm{D}^2 g[h,h]\big) \gtrsim -\|h\|_{g_\varepsilon}^2\,.$$

From $(g'+g_\varepsilon)^{-1} = \frac{1}{\det(g'+g_\varepsilon)}\,\mathrm{adj}(g'+g_\varepsilon)$, the LHS is continuous in ε, and the RHS is too clearly. Sending $\varepsilon \to 0$ completes the proof. ■

G.3.3. BASIC PROPERTIES: STRONGLY AVERAGE SELF-CONCORDANCE

To prove Lemma D.15, we first recall a concentration bound.

Lemma G.2 (Narayanan (2016), Lemma 4) *Let h be drawn from $\mathbb{S}^{d-1}$ uniformly at random. For any odd k, C^k-smooth $F:\mathbb{R}^d\to\mathbb{R}$, and $\varepsilon > 0$,*

$$\mathbb{P}_h\Big(|\mathrm{D}^k F(x)[h^{\otimes k}]| > k\varepsilon\cdot\sup_{\|v\|\le 1}\mathrm{D}^k F(x)[v^{\otimes k}]\Big) \le \exp\Big(-\frac{d\varepsilon^2}{2}\Big)\,.$$

We show that if g is HSC, then dg is SASC, using this lemma and following Narayanan (2016).

Proof of Lemma D.15. Let $g = d\,\nabla^2\phi$ and consider $g' : \mathrm{int}(K) \to \mathbb{S}_+^d$ such that $\bar{g} = g + g'$ is PD. For fixed $w \in \mathbb{R}^d$, apply Taylor's expansion to $\varphi(z) := \|w\|_{g(z)}^2$ at $z = x$, so there exists $p_w \in [x,z]$ such that $w^\mathsf{T} g(z) w = w^\mathsf{T} g(x) w + \mathrm{D}g(x)[z,w,w] + \frac{1}{2}\,\mathrm{D}^2 g(p_w)[z,z,w,w]$. Putting $z = w$ here,

$$\big|\|z\|_{g(z)}^2 - \|z\|_{g(x)}^2\big| \le |\mathrm{D}^3 g(x)[z^{\otimes 3}]| + \frac{1}{2}|\mathrm{D}^2 g(p_z)[z^{\otimes 4}]|\,.$$

Going forward, we can assume that $x = 0$ and $\bar{g}(x) = I$ due to affine invariance, and then z equals $rh/\sqrt{d}$ for $h \sim \mathcal{N}(0, I_d)$ in law. Using a standard tail bound on the standard Gaussian, we have $\mathbb{P}_h(\|h\| \ge -\sqrt{d}\cdot 2\log\varepsilon) \le \varepsilon$. Call this event B_1. In addition, Lemma G.2 implies that

$$\mathbb{P}\Big(\Big|\mathrm{D}^3\phi(x)\Big[\frac{h^{\otimes 3}}{\|h\|^3}\Big]\Big| \ge 3\frac{\varepsilon}{\sqrt{d}}\cdot\sup_{\|v\|\le 1}\mathrm{D}^3\phi(x)[v^{\otimes 3}]\Big) \le \varepsilon\,,$$

and call this event B_2. Conditioned on B_2^c,

$$\Big|\mathrm{D}^3\phi(x)\Big[\frac{h^{\otimes 3}}{\|h\|^3}\Big]\Big| \le \frac{3\varepsilon}{\sqrt{d}}\sup_{\|v\|\le 1}\mathrm{D}^3\phi(x)[v^{\otimes 3}] \le \frac{6\varepsilon}{\sqrt{d}}\sup_{\|v\|\le 1}\|v\|_{g(x)/d}^3 \le \frac{6\varepsilon}{d^2}\sup_{\|v\|\le 1}\|v\|_{g(x)}^3 \underbrace{\le}_{g(x)\preceq I_d} \frac{6\varepsilon}{d^2}\,.$$

Hence, conditioned on $z \in B_1^c \cap B_2^c$

$$|\mathrm{D}^3 g(x)[z^{\otimes 3}]| = \frac{r^3}{\sqrt{d}}\,\mathrm{D}^3\phi(x)[h^{\otimes 3}] \le \frac{r^3}{\sqrt{d}}\,\frac{6\varepsilon}{d^2}\,\|h\|^3 \le \frac{r^2}{d}\cdot 48r\varepsilon\Big(\log\frac{1}{\varepsilon}\Big)^3\,.$$

By taking $r_1(\varepsilon)$ so that $-48r_1\varepsilon\,(\log\varepsilon)^3 \le \varepsilon$, we can ensure $|\mathrm{D}^3 g(x)[z^{\otimes 3}]| \le \varepsilon r^2/d$ for any $r \le r_1(\varepsilon)$.

As for $|\mathrm{D}^2 g(p_z)[z^{\otimes 4}]|$, HSC of ϕ and Lemma D.4 lead to

$$\begin{aligned}
\frac{1}{2}\,|\mathrm{D}^2 g(p_z)[z^{\otimes 4}]| &\le 3d\,\|z\|_{\nabla^2\phi(p_z)}^4 \le \frac{3}{d}\|z\|_{\nabla^2\phi(x)}^4\,(1+2\,\|z\|_{\nabla^2\phi(x)}^2)^2 = \frac{3}{d}\,\|z\|_{g(x)}^4\,\big(1+\frac{2}{d}\,\|z\|_{g(x)}^2\big)^2 \\
&\underset{g\preceq I_d}{\le} \frac{3}{d}\|z\|^4\big(1+\frac{2}{d}\,\|z\|^2\big)^2 = \frac{3}{d}\,\frac{r^4}{d^2}\,\|h\|^4\Big(1+\frac{2r^2}{d^2}\|h\|^2\Big)^2 \\
&\le \frac{r^2}{d}\cdot 3r^2\,\big(2\log\frac{1}{\varepsilon}\big)^4\Big(1+2r^2\big(2\log\frac{1}{\varepsilon}\big)^4\Big)^2\,.
\end{aligned}$$

By taking $r_2(\varepsilon)$ and $r_3(\varepsilon)$ so that $\Big(1+2r_2^2\big(2\log\frac{1}{\varepsilon}\big)^4\Big)^2 \le 2$ and $2^2\cdot 3r_3^2\big(2\log\frac{1}{\varepsilon}\big)^4 \le \varepsilon$ respectively, it holds that on $B_1^c \cap B_2^c$

$$\frac{1}{2}\,|\mathrm{D}^2 g(p_z)[z^{\otimes 4}]| \le \varepsilon\frac{r^2}{d} \text{ for any } r \le \min r_i(\varepsilon).$$

Putting all these together, it follows that $|\|z\|_{g(z)}^2 - \|z\|_{g(x)}^2| \le 2\varepsilon r^2/d$ with probability at least $1-2\varepsilon$. By replacing $2\varepsilon \leftarrow \varepsilon$, the claim follows. ■

G.3.4. COLLAPSE AND EMBEDDING: WELL-DEFINEDNESS

We start with well-definedness of the notions of collapse and embedding (Definition D.16).

Proof of Proposition D.17. Let $k := \dim(W)$, and U and V be matrices in $\mathbb{R}^{d\times k}$, where the columns of each matrix form an orthonormal basis of W. Let us denote by $g_1 := U^\mathsf{T} gU$ and $g_2 := V^\mathsf{T} gV$ matrices represented with respect to U and V, and define the invertible matrix $M = V^{-1}U \in \mathbb{R}^{k\times k}$. Since U and V are full-column rank, if g_1 is PD, so is g_2.

Suppose g is SSC along W. Then,

$$\begin{aligned}
4\|h\|_g^2 &\ge \mathrm{Tr}(g_1^{-1}\mathrm{D}g_1[h]\,g_1^{-1}\mathrm{D}g_1) = \mathrm{Tr}\big((U^\mathsf{T} gU)^{-1}\cdot U^\mathsf{T}\mathrm{D}g[h]\,U\cdot(U^\mathsf{T} gU)^{-1}\cdot U^\mathsf{T}\mathrm{D}g[h]\,U\big) \\
&= \mathrm{Tr}\Big((M^\mathsf{T} V^\mathsf{T} gVM)^{-1}\cdot M^\mathsf{T} V^\mathsf{T}\mathrm{D}g[h]\,VM\cdot(M^\mathsf{T} V^\mathsf{T} gVM)^{-1}\cdot M^\mathsf{T} V^\mathsf{T}\mathrm{D}g[h]\,VM\Big) \\
&= \mathrm{Tr}\Big((V^\mathsf{T} gV)^{-1}V^\mathsf{T}\mathrm{D}g[h]\,V\,(V^\mathsf{T} gV)^{-1}V^\mathsf{T}\mathrm{D}g[h]\,V\Big) = \|g_2^{-\frac{1}{2}}\mathrm{D}g_2[h]\,g_2^{-\frac{1}{2}}\|_F^2\,,
\end{aligned}$$

and thus g_2 also satisfies the definition. ■

G.3.5. COLLAPSE AND EMBEDDING: AFFINE TRANSFORMATION

We begin with a barrier version.

Proof of Lemma D.18. For the first part, ψ is a ν-self-concordant barrier for $\bar{K}$ by Nesterov (2003, Theorem 4.2.3), so $\mathcal{D}^1_{\bar{g}}(x) \subset \bar{K} \cap (2x - \bar{K})$ for $\bar{g}(\cdot) := \nabla^2\psi(\cdot)$ by Lemma D.7. Now let $z \in \bar{K} \cap (2x - \bar{K})$. Then $Tz \in K$ and $T(2x - z) \in K$, and the latter implies $2y - Tz \in K$. Thus $Tz \in K \cap (2y - K)$ and $Tz \in \mathcal{D}^{\sqrt{\bar{\nu}}}_g(y)$. Due to

$$\mathrm{D}^2\psi(x)[(z-x)^{\otimes 2}] = \mathrm{D}^2\phi(y)[\big(A(z-x)\big)^{\otimes 2}] = \mathrm{D}^2\phi(y)[(Tz-y)^{\otimes 2}] \le \bar{\nu}\,,$$

it follows that ψ is also $\bar{\nu}$-symmetric.

For the second part, observe that $\mathrm{D}^4\psi(x)[v,v,h,h] = \mathrm{D}^4\phi(y)[Av,Av,Ah,Ah] \ge 0$ for any $v, h \in \mathbb{R}^d$. The third part can be proven similarly. ■

Next is a matrix version.

Proof of Lemma D.19. Let ϕ be a ν-self-concordant function counterpart of g. Then $\psi(x) := \phi(Tx)$ defined on $\mathrm{int}(\bar{K})$ is ν-self-concordant by Lemma D.18. For any $h \in \mathbb{R}^d$ and $y := Tx$, we have

$$\mathrm{D}\bar{g}(x)[h] = A^{\mathsf{T}}\mathrm{D}g(y)[Ah]\,A \preceq 2\|Ah\|_{g(y)}\,A^{\mathsf{T}}g(y)A = 2\|h\|_{\bar{g}(x)}\,\bar{g}(x)\,.$$

Consider a sequence $\{x_n\} \subset \bar{K}$ converging to a boundary point $x \in \partial\bar{K}$. If $Tx \notin \partial K$, then $Tx \in \mathrm{int}(K)$, and the continuity of T implies x is also in $\mathrm{int}(\bar{K})$. Thus, $Tx \in \partial K$ and $\psi(x_n) = \phi(Tx_n) \to \phi(Tx) = \infty$. Lastly, $\nabla^2\phi \asymp g$ leads to $\nabla^2\psi = A^{\mathsf{T}}\nabla^2\phi\,A \asymp A^{\mathsf{T}}gA = \bar{g}$, and $\bar{g}$ is ν-self-concordant for $\bar{K}$.

As for symmetry, since $\bar{g}$ is self-concordant, $\mathcal{D}^1_{\bar{g}}(x) \subset \bar{K} \cap (2x - \bar{K})$ for $x \in \mathrm{int}(\bar{K})$ by Lemma D.3. For $z \in \bar{K} \cap (2x - \bar{K})$, as $Tz \in K \cap (2Tx - K)$ holds, it follows that

$$\bar{\nu} \ge \|Tz - Tx\|^2_{g(y)} = \|z-y\|^2_{A^{\mathsf{T}}g(y)A} = \|z-y\|^2_{\bar{g}(x)}\,,$$

and thus $\bar{g}$ is $\bar{\nu}$-symmetric.

As for the second item, we first show that $\bar{g}$ is collapsed onto $W = \mathrm{row}(A)$ (i.e., $\bar{g} = P_W\bar{g}P_W$ for the orthogonal projection P_W onto W). To see this, observe that

$$P_W\bar{g}P_W = P_WA^{\mathsf{T}}gAP_W = A^{\mathsf{T}}(AA^{\mathsf{T}})^{\dagger}A \cdot A^{\mathsf{T}}gA \cdot A^{\mathsf{T}}(AA^{\mathsf{T}})^{\dagger}A\,,$$

and due to $AA^{\mathsf{T}}(AA^{\mathsf{T}})^{\dagger}A = AA^{\mathsf{T}}(A^{\mathsf{T}})^{\dagger}A^{\dagger}A = AA^{\dagger}A = A$, we have $P_W\bar{g}P_W = A^{\mathsf{T}}gA = \bar{g}$.

We now show that $\bar{g}$ is SSC along W. For $k := \dim(W)$, take $U \in \mathbb{R}^{d\times k}$ whose columns form an orthonormal basis of W. It suffices to show that $g_W := U^{\mathsf{T}}\bar{g}U = U^{\mathsf{T}}A^{\mathsf{T}}gAU = M^{\mathsf{T}}gM$ for $M := AU \in \mathbb{R}^{m\times k}$ is SSC. First of all, we can check PDness of g_W as follows: Suppose $g_Wv = 0$ for some $v \in \mathbb{R}^k$. Then $0 = \|v\|_{g_W} = \|g^{1/2}Mv\|_2$ and $AUv = Mv = 0$. Since $Uv \in \mathrm{row}(A) \cap \ker(A)$ and U is full-rank, we have $v = 0$. Next, for $h \in \mathbb{R}^k$ and $x \in \mathrm{int}(\bar{K})$

$$\mathrm{Tr}\big(g_W(x)^{-1}\mathrm{D}g_W(x)[h]\,g_W(x)^{-1}\mathrm{D}g_W(x)[h]\big) = \mathrm{Tr}\Big(\big(g^{\frac{1}{2}}M(M^{\mathsf{T}}gM)^{-1}M^{\mathsf{T}}g^{\frac{1}{2}}\cdot g^{-\frac{1}{2}}\mathrm{D}g(Tx)[Ah]\,g^{-\frac{1}{2}}\big)^2\Big)$$

$$\underset{\text{(i)}}{\le} \mathrm{Tr}\Big(\big(g^{-\frac{1}{2}}\mathrm{D}g(Tx)[Ah]\,g^{-\frac{1}{2}}\big)^2\Big) \le \|g^{-\frac{1}{2}}\mathrm{D}g(Tx)[Ah]\,g^{-\frac{1}{2}}\|_F^2 \le 4\|Ah\|^2_{g(Tx)} = 4\|h\|^2_{\bar{g}(x)}\,,$$

where in (i) we used $P(g^{\frac{1}{2}}M) = g^{\frac{1}{2}}M(M^\mathsf{T}gM)^{-1}M^\mathsf{T}g^{\frac{1}{2}} \preceq I$. Thus, $\bar{g}$ is SSC along $W = \text{row}(A)$.

The third item immediately follows from $\mathrm{D}^2\bar{g}(x)[h,h] = A^\mathsf{T}\mathrm{D}^2 g(y)[Ah, Ah]\,A \succeq 0$ for any $h \in \mathbb{R}^d$.

As for the fourth item, for any PSD matrix function g' on $\bar{K}$ we have

$$\operatorname{Tr}\big((g'+\bar{g})^{-1}\mathrm{D}^2\bar{g}[h,h]\big) = \operatorname{Tr}\Big((g'+A^\mathsf{T}gA)^{-1}A^\mathsf{T}\mathrm{D}^2 g[Ah,Ah]\,A\Big)$$
$$= \operatorname{Tr}\Big((A^{-\mathsf{T}}g'A^{-1}+g)^{-1}\mathrm{D}^2 g[Ah,Ah]\Big) \geq -\|Ah\|_g^2 = -\|h\|_{\bar{g}}^2\,.$$

The last item is straightforward to check by the change of variable. ■

G.3.6. COLLAPSE AND EMBEDDING: LIFTING UP SSC, SLTSC, AND SASC

In passing SSC to an augmented space, the Woodbury matrix identity is a main technical tool used: for matrices with compatible sizes

$$(I+UV)^{-1} = I - U\,(I+VU)^{-1}V\,.$$

Using this, we show that if $g \in \mathbb{S}_{++}^d$ is SSC, then $\bar{g} + \varepsilon I_m$ is SSC.

Proof of Lemma D.21. Fix $\varepsilon > 0, y \in \text{int}(K')$, and $h \in \mathbb{R}^m$. Take a projection matrix $P \in \{0,1\}^{d\times m}$ such that $PP^\mathsf{T} = I_d$ and $\bar{g}(y) = P^\mathsf{T}g(Py)P$ for $x = Py \in \text{int}(K)$. Also for $k := \dim(W)$, take a matrix $U \in \mathbb{R}^{d\times k}$ whose columns form an orthonormal basis of W. Then $\bar{g}(y) = P^\mathsf{T}g(Py)P$ and $g(x) = Ug_W(x)U$, so for $M := U^\mathsf{T}P \in \mathbb{R}^{k\times m}$,

$$\bar{g}(y) = P^\mathsf{T}Ug_W(Py)U^\mathsf{T}P = M^\mathsf{T}g_W(Py)M\,.$$

Note that $MM^\mathsf{T} = I_k$. Thus,

$$\|(\bar{g}(y)+\varepsilon I)^{-\frac{1}{2}}\mathrm{D}(\bar{g}+\varepsilon I)(y)[h]\,(\bar{g}(y)+\varepsilon I)^{-\frac{1}{2}}\|_F^2 = \operatorname{Tr}\Big(\big((\bar{g}(y)+\varepsilon I)^{-1}\mathrm{D}\bar{g}(y)[h]\big)^2\Big)$$
$$= \operatorname{Tr}\Big(\big(M(M^\mathsf{T}g_W(x)\,M+\varepsilon I)^{-1}M^\mathsf{T}\cdot \mathrm{D}g_W(x)[Ph]\big)^2\Big) \underset{\text{(i)}}{=} \operatorname{Tr}\Big(\big((g_W(x)+\varepsilon I_k)^{-1}\mathrm{D}g_W(x)[Ph]\big)^2\Big)$$
$$\leq \|g_W(x)^{-\frac{1}{2}}\mathrm{D}g_W(x)[Ph]\,g_W(x)^{-\frac{1}{2}}\|_F^2 \leq 4\|Ph\|_{g(x)}^2 = 4\|h\|_{\bar{g}(y)}^2\,,$$

where in (i) we used the identity $M\big(M^\mathsf{T}g_W(x)\,M+\varepsilon I\big)^{-1}M^\mathsf{T} = (g_W(x)+\varepsilon I_k)^{-1}$. To see this, we use the Woodbury matrix identity to get

$$(\varepsilon I_m + M^\mathsf{T}g_W M)^{-1} = \frac{1}{\varepsilon}I_m - \frac{1}{\varepsilon^2}M^\mathsf{T}g_W^{\frac{1}{2}}\big(I_k+\frac{1}{\varepsilon}g_W\big)^{-1}g_W^{\frac{1}{2}}M\,,$$

and thus conjugating both sides by M results in

$$M\big(M^\mathsf{T}g_W M+\varepsilon I_m\big)^{-1}M^\mathsf{T} = \frac{1}{\varepsilon}I_k - \frac{1}{\varepsilon}g_W^{\frac{1}{2}}(g_W+\varepsilon I_k)^{-1}g_W^{\frac{1}{2}} = \frac{1}{\varepsilon}I_k - \frac{1}{\varepsilon}(g_W+\varepsilon I_k)^{-1}g_W\,.$$

Then, the identity follows from

$$(g_W+\varepsilon I_k)\cdot\big(\frac{1}{\varepsilon}I_k - \frac{1}{\varepsilon}\,(g_W+\varepsilon I_k)^{-1}g_W\big) = \frac{1}{\varepsilon}(g_W+\varepsilon I_k) - \frac{1}{\varepsilon}g_W = I_k\,.$$

■

In extending SLTSC and SASC, we need two technical lemmas: the inverse of a block matrix and connection between P(S)Dness and Schur complements.

Lemma G.3 *If D and its Schur complement $A - BD^{-1}C$ are invertible, then*

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & * \\ * & * \end{bmatrix}.$$

Lemma G.4 (Schur complement) *Let $A \in \mathbb{R}^{d\times d}, B \in \mathbb{R}^{d\times m}, C \in \mathbb{R}^{m\times m}$ and define a matrix $M \in \mathbb{R}^{(m+d)\times(m+d)}$ by*

$$M = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix}.$$

Then $M \succ 0$ if and only if $A \succ 0$ and $C - BA^{-1}B^\top \succ 0$ if and only $C \succ 0$ and $A - B^\top C^{-1} B \succ 0$.

Using these, we show that if g is SLTSC and SASC, then $\bar{g}$ is SLTSC and SASC.

Proof of Lemma D.22. Take a full row-rank projection matrix $P \in \{0,1\}^{d\times m}$ such that $\bar{g}(y) = P^\top g(Py)P$, where the rows of P forms a subset of the canonical basis $\{e_1, \ldots, e_m\}$. We can augment the rows of P with the rest of the canonical basis so that the augmented matrix $\bar{P} \in \mathbb{R}^{m\times m}$ is an orthonormal matrix. Then we can represent $\bar{g}$ by

$$\bar{g}(y) = \bar{P}^\top \begin{bmatrix} g(Py) & 0 \\ 0 & 0 \end{bmatrix} \bar{P}.$$

Consider a PSD matrix function $g' : \mathrm{int}(K') \to \mathbb{S}_+^m$ such that $g' + \bar{g}$ is PD on K'. Representing them in the block form with $g_A \in \mathbb{R}^{d\times d}, g_B \in \mathbb{R}^{d\times(m-d)}$, and $g_C \in \mathbb{R}^{(m-d)\times(m-d)}$

$$\bar{g} + g' = \bar{P}^\top \left(\begin{bmatrix} g & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} g_A & g_B \\ g_B^\top & g_C \end{bmatrix} \right) \bar{P} = \bar{P}^\top \underbrace{\begin{bmatrix} g + g_A & g_B \\ g_B^\top & g_C \end{bmatrix}}_{=:g^*} \bar{P}.$$

Since g^* is PD, g_C and its Schur complement $(g + g_A) - g_B g_C^{-1} g_B^\top$ are PD. Thus by Lemma G.3,

$$\begin{bmatrix} g + g_A & g_B \\ g_B^\top & g_C \end{bmatrix}^{-1} = \begin{bmatrix} (g + g_A - g_B g_C^{-1} g_B^\top)^{-1} & * \\ * & * \end{bmatrix}.$$

Hence,

$$\mathrm{Tr}\big((\bar{g} + g')^{-1} \mathrm{D}^2 \bar{g}(y)[h,h]\big) = \mathrm{Tr}\left(\bar{P}^\top \begin{bmatrix} g + g_A & g_B \\ g_B^\top & g_C \end{bmatrix}^{-1} \bar{P}\bar{P}^\top \begin{bmatrix} \mathrm{D}^2 g(Py)[Ph,Ph] & 0 \\ 0 & 0 \end{bmatrix} \bar{P} \right)$$

$$= \mathrm{Tr}\left(\begin{bmatrix} g + g_A & g_B \\ g_B^\top & g_C \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{D}^2 g(Py)[Ph,Ph] & 0 \\ 0 & 0 \end{bmatrix} \right) = \mathrm{Tr}\big((g + \underbrace{g_A - g_B g_C^{-1} g_B^\top}_{\succeq 0})^{-1} \mathrm{D}^2 g(Py)[Ph,Ph]\big)$$

$$\geq -\|Ph\|_{g(Py)}^2 = -\|h\|_{\bar{g}(y)}^2,$$

where in the last inequality we used STLSC of g, since $g' \succeq 0$ ensures that its Schur complement satisfies $g_A - g_B g_C^{-1} g_B^\top \succeq 0$ by Lemma G.4.

For SASC, consider any PSD matrix function $g' : \mathrm{int}(K') \to \mathbb{S}_+^m$. For $x = Py$ and $z_x = Pz_y \in \mathbb{R}^d$ with $z_y \sim \mathcal{N}\big(y, \frac{r^2}{m}(\bar{g} + g)(y)^{-1}\big)$, we have

$$\|z_y - y\|_{\bar{g}(z_y)}^2 - \|z_y - y\|_{\bar{g}(y)}^2 = \|z_x - x\|_{g(z_x)}^2 - \|z_x - x\|_{g(x)}^2.$$

Also, $z_x - x = P\,(z_y - y)$ is a Gaussian with zero mean and covariance

$$\frac{r^2}{m}\,P\,(\bar g + g')(y)^{-1}P^\mathsf{T} = \frac{r^2}{m}\,P\bar P^\mathsf{T}\left(\begin{bmatrix} g & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix} g_A & g_B\\ g_B^\mathsf{T} & g_C\end{bmatrix}\right)^{-1}\bar P\bar P^\mathsf{T}$$
$$=\frac{r^2}{m}\,\begin{bmatrix} I_d & 0_{d\times(m-d)}\end{bmatrix}\left(\begin{bmatrix} g & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix} g_A & g_B\\ g_B^\mathsf{T} & g_C\end{bmatrix}\right)^{-1}\begin{bmatrix} I_d\\ 0_{d\times(m-d)}\end{bmatrix} = \frac{r^2}{m}\,(g + g_A - g_B g_C^{-1} g_B^\mathsf{T})^{-1}\,.$$

Since $g_A - g_B g_C^{-1} g_B^\mathsf{T} \succeq 0$ due to $g' \succeq 0$, it holds that $g_0 := \frac{m-d}{d}g + \frac{m}{d}(g_A - g_B g_C^{-1} g_B^\mathsf{T})$ on $\mathrm{int}(K)$ is PSD. Now, it suffices to check that the covariance matrix above is equal to $\frac{r^2}{d}(g+g_0)^{-1}$:

$$\frac{d}{r^2}\,(g+g_0) = \frac{d}{r^2}\Big(g + \frac{m-d}{d}\,g + \frac{m}{d}\,(g_A - g_B g_C^{-1} g_B^\mathsf{T})\Big)\frac{m}{r^2}\,(g + g_A - g_B g_C^{-1} g_B^\mathsf{T})\,.$$

■

G.3.7. DIRECT PRODUCT: SSC AND SLTSC

We show that if $g_i \in \mathbb{S}^{d_i}_{++}$ is SC, then $g = \sum d_i \bar g_i$ is SSC.

Proof of Lemma D.25. Note that $d_i g_i$ is SSC for $i = 1, \dots, m$. For $x \in \prod E_i$ and $h = (h_1, \dots, h_m) \in \mathbb{R}^l$ with $h_i \in \mathbb{R}^{d_i}$, we have

$$\|g(x)^{-\frac12}\mathrm{D}g(x)[h]\,g(x)^{-\frac12}\|_F^2$$
$$= \left\|\begin{bmatrix} g_1(x_1)^{-\frac12}\mathrm{D}g_1(x_1)[h_1]\,g_1(x_1)^{-\frac12} & & \\ & \ddots & \\ & & g_m(x_m)^{-\frac12}\mathrm{D}g_m(x_m)[h_m]\,g_m(x_m)^{-\frac12}\end{bmatrix}\right\|_F^2$$
$$= \sum_i \|g_i(x_i)^{-\frac12}\mathrm{D}g_i(x_i)[h_i]\,g_i(x_i)^{-\frac12}\|_F^2 \le 4\sum_i \|h_i\|^2_{d_i g_i(x_i)} = 4\|h\|^2_{g(x)}\,.$$

■

Next, we show that if $g_i \in \mathbb{S}^{d_i}_{++}$ is HSC, then $g = \sum d_i \bar g_i$ is SLTSC.

Proof of Lemma D.26. For $h = (h_1, \dots, h_m)$ and any PSD matrix function g', we have

$$\begin{aligned}\mathrm{Tr}\big((g'+g)^{-1}\mathrm{D}^2 g[h^{\otimes 2}]\big) &= \sum_i \mathrm{Tr}\big((g' + (g - d_i\bar g_i) + d_i \bar g_i)^{-1}\mathrm{D}^2(d_i\bar g_i)[h^{\otimes 2}]\big)\\ &\gtrsim -\sum_i \|h\|^2_{d_i\bar g_i} = -\|h\|_g^2\,,\end{aligned}$$

where we used Lemma D.13 in the inequality. ■

G.3.8. INVERSE IMAGES UNDER NON-LINEAR MAPPINGS

Proof of Lemma D.29. Since $\mathcal{A}$ is $(R(G),\beta),\gamma)$-compatible with Γ, the first two claims immediately follow from Nesterov and Nemirovskii (1994, Proposition 5.1.7). Let $x\in G^+$ and $h\in\mathbb{R}^d$. Define the following notations:

$$u=\mathrm{D}\mathcal{A}(x)[h],\quad v=\mathrm{D}^2\mathcal{A}(x)[h^{\otimes 2}],\quad w=\mathrm{D}^3\mathcal{A}(x)[h^{\otimes 3}],\quad z=\mathrm{D}^4\mathcal{A}(x)[h^{\otimes 4}],$$
$$s=\sqrt{\mathrm{D}F(y)[v]},\quad \rho=\sqrt{\mathrm{D}^2\Pi(x)[h^{\otimes 2}]},\quad r=\sqrt{\mathrm{D}^2F(y)[u^{\otimes 2}]}\,.$$

From direct computations, we have

$$\begin{aligned}
\mathrm{D}^2\Psi(x)[h^{\otimes 2}]&=\mathrm{D}F(y)[v]+\mathrm{D}^2F(y)[u^{\otimes 2}]+\delta^2\mathrm{D}^2\Pi(x)[h^{\otimes 2}]=s^2+r^2+\delta^2\rho^2\,,\\
\mathrm{D}^3\Psi(x)[h^{\otimes 3}]&=\mathrm{D}F(y)[w]+3\mathrm{D}^2F(y)[u,v]+\mathrm{D}^3F(y)[u^{\otimes 3}]+\delta^2\mathrm{D}^3\Pi(x)[h^{\otimes 3}]\,,\\
\mathrm{D}^4\Psi(x)[h^{\otimes 4}]&=\mathrm{D}^2F(y)[w,u]+\mathrm{D}F(y)[z]+3\mathrm{D}^3F(y)[u,u,v]+3\mathrm{D}^2F(y)[v^{\otimes 2}]\\
&\qquad+3\mathrm{D}^2F(y)[u,w]+\mathrm{D}^4F(y)[u^{\otimes 4}]+3\mathrm{D}^3F(y)[u,u,v]+\delta^2\mathrm{D}^4\Pi(x)[h^{\otimes 4}]\\
&=\mathrm{D}F(y)[z]+3\mathrm{D}^2F(y)[v^{\otimes 2}]+4\mathrm{D}^2F(y)[u,w]\\
&\qquad+6\mathrm{D}^3F(y)[u,u,v]+\mathrm{D}^4F(y)[u^{\otimes 4}]+\delta^2\mathrm{D}^4\Pi(x)[h^{\otimes 4}]\,.
\end{aligned}$$

HSC of F and Π implies that

$$|\mathrm{D}^4\Pi(x)[h^{\otimes 4}]|\le 6\rho^4\,,\qquad\text{and}\qquad |\mathrm{D}^4F(y)[u^{\otimes 4}]|\le 6r^4\,.$$

Since $\mathcal{A}$ is (K,β,γ)-compatible and $K\subset R(G)$, Lemma D.28-1 implies concavity of $\mathcal{A}$ with respect to $R(G)$, which means $-v\ge_{R(G)}0$. Then, Nesterov and Nemirovskii (1994, Corollary 2.3.1) ensures

$$\sqrt{\mathrm{D}^2F(y)[v^{\otimes 2}]}\le \mathrm{D}F(y)[v]=s^2\,.$$

Hence, $|3\mathrm{D}^2F(y)[v,v]|\le 3(\mathrm{D}F(y)[v])^2=3s^4$, and self-concordance of F results in

$$|6\mathrm{D}^3F(y)[u,u,v]|\le 12r^2\sqrt{\mathrm{D}^2F(y)[v,v]}\le 12r^2s^2\,.$$

Since $\{h:h^{\mathsf{T}}\Pi(x)h\le 1\}$ is contained in $\Gamma\cap(2x-\Gamma)$, compatibility of $\mathcal{A}$ leads to

$$\beta\mathrm{D}^2\mathcal{A}(x)\Big[\Big(\frac{h}{\|h\|_{\Pi(x)}}\Big)^{\otimes 2}\Big]\le_K \mathrm{D}^3\mathcal{A}(x)\Big[\Big(\frac{h}{\|h\|_{\Pi(x)}}\Big)^{\otimes 3}\Big]\le_K -\beta\mathrm{D}^2\mathcal{A}(x)\Big[\Big(\frac{h}{\|h\|_{\Pi(x)}}\Big)^{\otimes 2}\Big]\,,$$

and thus $\beta\rho v\le_K w\le_K -\beta\rho v$. As K is a ray, $\mathrm{D}^2F(y)[w,w]\le\beta^2\rho^2\mathrm{D}^2F(y)[v,v]\le\beta^2\rho^2s^4$. Thus,

$$|4\mathrm{D}^2F(y)[u,w]|\le 4\sqrt{\mathrm{D}^2F(y)[u,u]}\sqrt{\mathrm{D}^2F(y)[w,w]}\le 4r\beta\rho s^2\,.$$

Lastly, since $\gamma v\rho^2\le_K z\le_K -\gamma v\rho^2$ and K is a ray, we have

$$|\mathrm{D}F(y)[z]|\le 3\gamma\rho^2|\mathrm{D}F(y)[v]|=3\gamma\rho^2s^2\,.$$

Putting these together,

$$\begin{aligned}
|\mathrm{D}^4\Psi(x)[h^{\otimes 4}]|&\le 3\gamma\rho^2s^2+4r\beta\rho s^2+12r^2s^2+3s^4+6\delta^2\rho^4+6r^4\\
&\le 6(\delta^2\rho^4+r^4+s^4+r^2s^2+\delta\rho^2s^2+\delta r\rho s^2)\\
&\le 6\big((\delta\rho)^4+r^4+s^4+r^2s^2+(\delta\rho)^2s^2+r^2s^2+(\delta\rho)^2s^2\big)\\
&\le 6\big((\delta\rho)^2+r^2+s^2\big)^2=6\big(\mathrm{D}^2\Psi(x)[h,h]\big)^2\,.
\end{aligned}$$

■

G.4. Main constraints and epigraphs (§E)

G.4.1. LINEAR CONSTRAINTS: STRONG SELF-CONCORDANCE AND SYMMETRY

We relate SSC and symmetry to well-studied terms in the field of optimization, such as $\max_i \frac{[\sigma(\sqrt{D_x}A_x)]_i}{[D_x]_{ii}}$ and $\|D'_{x,h}\|^2_{D_x^{-1}}$.

Proof of Lemma E.4. Let us write $g(x) = A_x^\mathsf{T} D_x A_x = A^\mathsf{T} V_x A$ for $V_x := S_x^{-1} D_x S_x^{-1}$. By Claim I.1,

$$\mathrm{D}g(x)[h] = A^\mathsf{T}(-2S_x^{-1}S_{x,h}S_x^{-1}D_x + S_x^{-1}\mathrm{D}D_x[h]\,S_x^{-1})A = A^\mathsf{T}V_x^{1/2}\overline{D}_x V_x^{1/2}A\,, \tag{G.15}$$

where $\overline{D}_x := -2S_{x,h} + D_x^{-1}\mathrm{D}D_x[h]$. Using this,

$$\begin{aligned}\|(g'+g)^{-\frac{1}{2}}\mathrm{D}g[h]\,(g'+g)^{-\frac{1}{2}}\|_F^2 &= \mathrm{Tr}\big((g'+g)^{-1}A^\mathsf{T}V_x^{1/2}\overline{D}_x \underbrace{V_x^{1/2}A(g'+g)^{-1}A^\mathsf{T}V_x^{1/2}}_{=:P'_x}\overline{D}_x V_x^{1/2}A\big)\\ &= \mathrm{Tr}(P'_x\overline{D}_x P'_x\overline{D}_x)\,.\end{aligned}$$

By Lemma J.1, we have $P'_x \preceq P_x = P(V_x^{1/2}A) = P(D_x^{1/2}A_x)$, and thus

$$\begin{aligned}\mathrm{Tr}(P'_x\overline{D}_xP'_x\overline{D}_x) \le \mathrm{Tr}(P_x\overline{D}_xP_x\overline{D}_x) &\underset{\text{(i)}}{=} \mathrm{diag}(\overline{D}_x)^\mathsf{T}P_x^{(2)}\,\mathrm{diag}(\overline{D}_x) \underset{\text{(ii)}}{\le} \mathrm{diag}(\overline{D}_x)^\mathsf{T}\Sigma_x\,\mathrm{diag}(\overline{D}_x)\\ &\underset{\text{(iii)}}{\le} 4\sum_{i=1}^m[\sigma(D_x^{1/2}A_x)]_i\,\big((A_xh)_i^2 + (D_x^{-1}\mathrm{D}D_x[h])_i^2\big)\\ &\le 4\max_i\frac{[\sigma(D_x^{1/2}A_x)]_i}{[D_x]_{ii}}\cdot\sum_{i=1}^m[D_x]_{ii}\,\big((A_xh)_i^2 + (D_x^{-1}\mathrm{D}D_x[h])_i^2\big)\\ &\underset{\text{(iv)}}{=} 4\max_i\frac{[\sigma(D_x^{1/2}A_x)]_i}{[D_x]_{ii}}\cdot\Big(\|h\|^2_{g(x)} + \sum_{i=1}^m[D_x^{-1}]_{ii}(\mathrm{D}D_x[h])_i^2\Big)\,,\end{aligned}$$

where (i) holds due to $x^\mathsf{T}(A\circ B)y = \mathrm{Tr}\big(\mathrm{Diag}(x)A\,\mathrm{Diag}(y)B^\mathsf{T}\big)$ (Lemma H.2), (ii) follows from $P_x^{(2)} \preceq \Sigma_x$ (Claim I.3)[7], (iii) uses $(a+b)^2 \le 2\,(a^2+b^2)$ for $a,b\in\mathbb{R}$ and $\Sigma_x = \mathrm{Diag}(P_x) = \sigma(D_x^{1/2}A_x)$, and (iv) holds due to $\sum_{i=1}^m[D_x]_{ii}\,(A_xh)_i^2 = h^\mathsf{T}A_x^\mathsf{T}D_xA_xh = h^\mathsf{T}g(x)h$.

As for the second claim,

$$\max_{h:\|h\|_{g(x)}=1}\|A_xh\|_\infty = \max_h\max_{i\in[m]}\left|\frac{a_i^\mathsf{T}h}{s_i}\right| = \max_{i\in[m]}\max_{u:\|u\|_2=1}\left|\frac{a_i^\mathsf{T}g(x)^{-1/2}u}{s_i}\right|$$

$$= \max_{i\in[m]}\left\|g(x)^{-1/2}\frac{a_i}{s_i}\right\|_2 = \max_{i\in[m]}\sqrt{\frac{1}{s_i^2}a_i^\mathsf{T}g(x)^{-1}a_i} = \sqrt{\max_{i\in[m]}e_i^\mathsf{T}A_xg(x)^{-1}A_x^\mathsf{T}e_i} = \sqrt{\max_{i\in[m]}\frac{[\sigma(D_x^{1/2}A_x)]_i}{[D_x]_{ii}}}\,.$$

As for the last claim, for $h\in\mathbb{R}^d$ such that $\|A_xh\|_\infty \le 1$ (i.e., $h\in K\cap(2x-K)$ for $K=\{Ax\ge b\}$ due to Lemma D.9) we have

$$h^\mathsf{T}g(x)h = h^\mathsf{T}A_x^\mathsf{T}D_xA_xh = \sum_{i=1}^m(D_x)_{ii}(A_xh)_i^2 \le \|A_xh\|_\infty^2\sum_{i=1}^m(D_x)_{ii} \le \mathrm{Tr}(D_x)\,.$$

7. Even though this lemma is proven for leverage scores, the proof there can be extended to any orthogonal projection matrices.

■

Now we establish SSC and compute the symmetry parameters of metrics of the form $A_x^\mathsf{T} D_x A_x$:

Proof of Lemma E.5. Logarithmic barrier: To show that g is SSC along $\text{row}(A)$, consider a self-concordant matrix $g(y) = S_y^{-2} = -\nabla_y^2(\sum_{i=1}^m \log y_i)$ defined on $\{y \in \mathbb{R}^m : y \geq 0\}$. By putting $D_x = I_m$ and $A_x = S_x^{-1}$ into Lemma E.4-1, since $\sigma(A_x) \leq 1$

$$\|g(x)^{-\frac{1}{2}} \mathrm{D}g(x)[h]\, g(x)^{-\frac{1}{2}}\|_F \leq 2\Big(\max_{i\in[m]} \sigma(A_x)_i\Big)^{1/2} \|h\|_{g(x)} \leq 2\|h\|_{g(x)}\,.$$

Through the linear map $Tx = Ax - b = y$, we recover $g(x) = \nabla^2\phi_{\log}(x) = A^\mathsf{T} S_y^{-2} A = A_x^\mathsf{T} A_x$, which is SSC along $\text{row}(A)$ by Lemma D.19. For the $\bar\nu$-symmetry, the first part (i.e., $\mathcal{D}_g^1(x) \subset K \cap (2x - K)$) follows from Lemma D.7. The second part is immediate from $\bar\nu = \text{Tr}(I_m) = m$ and Lemma E.4-3.

Approximate volumetric barrier: For $D_x = \Sigma_x = \Sigma(A_x)$, by Lemma I.5-1 and 3 with $p = 2$,

$$\max_i \frac{[\sigma(D_x^{1/2} A_x)]_i}{[D_x]_{ii}} \leq 2\sqrt{m}\,, \quad \text{and} \quad \sum_{i=1}^m [D_x^{-1}]_{ii}\, (\mathrm{D}D_x[h])_i^2 = \|\Sigma_x^{-1}\operatorname{diag}(\mathrm{D}\Sigma_x[h])\|_{\Sigma_x}^2 \leq 4\|h\|_{g(x)}^2\,.$$

Using Lemma E.4-1,

$$\|g(x)^{-\frac{1}{2}} \mathrm{D}g(x)[h]\, g(x)^{-\frac{1}{2}}\|_F^2 \leq 4\max_i \frac{[\sigma(D_x^{1/2} A_x)]_i}{[D_x]_{ii}} \big(\|h\|_{g(x)}^2 + \sum_{i=1}^m [D_x^{-1}]_{ii} (\mathrm{D}D_x[h])_i^2\big) \leq 40\sqrt{m}\|h\|_{g(x)}^2\,.$$

For the $\bar\nu$-symmetry, $\|A_x(y-x)\|_\infty^2 \leq \max_{i\in[m]} \frac{[\sigma(D_x^{1/2} A_x)]_i}{[D_x]_{ii}} \leq 2m^{1/2}$ for $y \in \mathcal{D}_g^1(x)$ by Lemma E.4-2. Also, Lemma E.4-3 implies that y with $\|A_x(y-x)\|_\infty \leq 1$ is contained in $\mathcal{D}_g^{\sqrt{\text{Tr}(D_x)}}(x)$, where $\text{Tr}(D_x) = \text{Tr}(P_x) \leq d$. Therefore, $\tilde g(x) := 40\sqrt{m} g(x) = 40\sqrt{m} A_x^\mathsf{T} \Sigma_x A_x$ is SSC with the symmetry parameter $\bar\nu = \mathcal{O}(\sqrt{m}d)$.

Vaidya metric: Consider the metric without scaling: $g(x) := A_x^\mathsf{T} D_x A_x$ with $D_x = \Sigma_x + \frac{d}{m} I_m$. Then, using Anstreicher (1997, (4.5)) in (i) below

$$\max_i \frac{[\sigma(D_x^{1/2} A_x)]_i}{[D_x]_{ii}} \underset{\text{Lemma E.4-2}}{=} \Big(\max_{h\in\mathbb{R}^d} \frac{\|A_x h\|_\infty}{\|h\|_{g(x)}}\Big)^2 \underset{\text{(i)}}{\leq} \sqrt{\frac{m}{d}}\,, \tag{G.16}$$
$$\sum_{i=1}^m [D_x^{-1}]_{ii}\, (\mathrm{D}D_x[h])_i^2 \underset{\text{(ii)}}{\leq} \sum_{i=1}^m [\Sigma_x^{-1}]_{ii} (\mathrm{D}\Sigma_x[h])_i^2 \underset{\text{Lemma I.5-3}}{\leq} 4h^\mathsf{T} A_x^\mathsf{T} \Sigma_x A_x h \leq 4\|h\|_{g(x)}^2\,.$$

Putting these back to Lemma E.4-1,

$$\|g(x)^{-\frac{1}{2}} \mathrm{D}g(x)[h]\, g(x)^{-\frac{1}{2}}\|_F^2 \leq 4\max_i \frac{[\sigma(D_x^{1/2} A_x)]_i}{[D_x]_{ii}} \big(\|h\|_{g(x)}^2 + \sum_{i=1}^m [D_x^{-1}]_{ii} (\mathrm{D}D_x[h])_i^2\big) \leq 20\sqrt{\frac{m}{d}}\|h\|_{g(x)}^2\,.$$

Thus, $\tilde g(x) := 22\sqrt{\frac{m}{d}} g(x) = 22\sqrt{\frac{m}{d}} A_x^\mathsf{T}\big(\Sigma_x + \frac{d}{m} I_m\big) A_x$ is SSC. For the $\bar\nu$-symmetry, Lemma E.4-2 implies that for $y \in \mathcal{D}_g^1(x)$,

$$\|A_x(y-x)\|_\infty^2 \leq \max_i \frac{[\sigma(D_x^{1/2} A_x)]_i}{[D_x]_{ii}} \underset{\text{(G.16)}}{\leq} \sqrt{\frac{m}{d}}\,.$$

Also, Lemma E.4-3 implies that y with $\|A_x(y-x)\|_\infty \le 1$ is contained in $\mathcal{D}_g^{\sqrt{\mathrm{Tr}(D_x)}}(x)$, where

$$\mathrm{Tr}(D_x) = \mathrm{Tr}\big(\Sigma_x + \frac{d}{m} I_m\big) = \mathrm{Tr}(\Sigma_x) + d \le 2d\,.$$

Therefore, $\tilde{g}(x)$ satisfies $\mathcal{D}_{\tilde{g}}^1(x) \subset K \cap (2x - K) \subset \mathcal{D}_{\tilde{g}}^{\sqrt{44(md)^{1/2}}}(x)$, so $\tilde{g}$ is $\mathcal{O}(\sqrt{md})$-symmetric.

Lewis-weight metric: Consider the unscaled version first: $g(x) = A_x^\intercal W_x A_x$. By Lemma E.4-1

$$\begin{aligned}
\|g(x)^{-\frac{1}{2}} \mathrm{D}g(x)[h]\, g(x)^{-\frac{1}{2}}\|_F^2 &\le 4 \max_i \frac{[\sigma(W_x^{1/2} A_x)]_i}{[W_x]_{ii}} \Big(\|h\|_{g(x)}^2 + \sum_{i=1}^m [W_x^{-1}]_{ii} (\mathrm{D}W_x[h])_i^2\Big) \\
&\underset{\text{(i)}}{\le} 8 m^{\frac{2}{p+2}} \big(\|h\|_{g(x)}^2 + p^2 \|h\|_{g(x)}^2\big) \le \big(8 m^{\frac{2}{p+2}} (1+p^2)\big) \|h\|_{g(x)}^2\,,
\end{aligned}$$

where in (i) we used Lemma I.5-1 and 3.

For the first part of the $\bar{\nu}$-symmetry, Lemma E.4-2 implies that

$$\max_{h:\|h\|_{g(x)}=1} \|A_x h\|_\infty = \sqrt{\max_i \frac{[\sigma(W_x^{1/2} A_x)]_i}{[W_x]_{ii}}} \le \sqrt{2 m^{\frac{2}{p+2}}}\,,$$

and Lemma E.4-3 leads to $K \cap (2x - K) \subset \mathcal{D}_g^{\sqrt{d}}(x)$ due to

$$\mathrm{Tr}(W_x) = \mathrm{Tr}\big(W_x^{\frac{1}{2}-\frac{1}{p}} A_x (A_x^\intercal W_x^{1-\frac{2}{p}} A_x)^{-1} A_x^\intercal W_x^{\frac{1}{2}-\frac{1}{p}}\big) = \mathrm{Tr}\big(A_x^\intercal W_x^{1-\frac{2}{p}} A_x (A_x^\intercal W_x^{1-\frac{2}{p}} A_x)^{-1}\big) = d\,.$$

Therefore, $16p^2 m^{\frac{2}{p+2}} A_x^\intercal W_x A_x$ is SSC with $\mathcal{O}\big(dm^{\frac{2}{p+2}}\big)$-symmetry by Lemma D.9. By setting $p = \mathcal{O}(\log m)$, the claim follows. ■

G.4.2. LINEAR CONSTRAINTS: STRONGLY LOWER TRACE SELF-CONCORDANCE OF VAIDYA

Let $\theta_1(x) := A_x^\intercal \Sigma_x A_x$, $\theta_2(x) := A_x^\intercal A_x$, and $\Gamma_x := \mathrm{Diag}\big(A_x g(x)^{-1} A_x^\intercal\big)$. Recall $g = g_1 + g_2$ for a PSD matrix function g_1 and the Vaidya metric g_2.

Lemma G.5 $\|\Gamma_x\|_\infty \le \frac{1}{44}$.

Proof For $\bar{g}_2 := \theta_1 + \frac{d}{m}\theta_2 = \frac{1}{44}\sqrt{\frac{d}{m}} g_2$, it follows from $g^{-1} \preceq g_2^{-1} = \frac{1}{44}\sqrt{\frac{d}{m}} \bar{g}_2^{-1}$ that

$$44 \|\Gamma_x\|_\infty \le 4 \sqrt{\frac{d}{m}} \|\,\mathrm{Diag}(A_x \bar{g}_2^{-1} A_x^\intercal)\|_\infty = \sqrt{\frac{d}{m}} \max_{i\in[m]} \frac{\big[\sigma\big(\sqrt{\Sigma_x + \frac{d}{m} I_m} A_x\big)\big]_i}{\big[\Sigma_x + \frac{d}{m} I_m\big]_{ii}} \underset{\text{(G.16)}}{\le} 1\,.$$

■

Now we show SLTSC of the Vaidya metric:

Proof of Lemma E.6. As $\mathrm{D}^2\theta_2(x)[h,h] \succeq 0$ by Claim I.1, we have

$$\mathrm{Tr}\big(g^{-1} \mathrm{D}^2\theta_2(x)[h,h]\big) = \mathrm{Tr}\big(g^{-\frac{1}{2}} \mathrm{D}^2\theta_2(x)[h,h] g^{-\frac{1}{2}}\big) \ge 0\,.$$

As for θ_1, by Lemma I.4-6, $\mathrm{D}^2\theta_1[h,h] \succeq -16A_x^\top \operatorname{Diag}(S_{x,h}P_xS_{x,h}P_x)A_x - 6A_x^\top \operatorname{Diag}(P_xS_{x,h}^2P_x)A_x$, so

$$\operatorname{Tr}\big(g^{-1}\mathrm{D}^2\theta_1(x)[h,h]\big) \ge -16\operatorname{Tr}(\Gamma_xS_{x,h}P_xS_{x,h}P_x) - 6\operatorname{Tr}(\Gamma_xP_xS_{x,h}^2P_x)\,.$$

We first note that $\operatorname{Tr}(S_{x,h}P_xS_{x,h}) = s_{x,h}^\top(P_x \circ I)s_{x,h} = s_{x,h}^\top\Sigma_x s_{x,h} = \|h\|_{\theta_1}^2$. Using this,

$$\begin{aligned}\operatorname{Tr}(\Gamma_xS_{x,h}P_xS_{x,h}P_x) &= \operatorname{Tr}(\Gamma_x^{1/2}S_{x,h}P_x\cdot S_{x,h}P_x\Gamma_x^{1/2}) \le \sqrt{\operatorname{Tr}(\Gamma_x^{\frac12}S_{x,h}P_x^2S_{x,h}\Gamma_x^{\frac12})\operatorname{Tr}(\Gamma_x^{\frac12}P_xS_{x,h}^2P_x\Gamma_x^{\frac12})}\\ &= \sqrt{\operatorname{Tr}(P_xS_{x,h}\Gamma_xS_{x,h}P_x)}\sqrt{\operatorname{Tr}(S_{x,h}P_x\Gamma_xP_xS_{x,h})} = \|\Gamma_x\|_\infty\|h\|_{\theta_1}^2\,,\\ \operatorname{Tr}(\Gamma_xP_xS_{x,h}^2P_x) &= \operatorname{Tr}(S_{x,h}P_x\Gamma_xP_xS_{x,h}) \le \|\Gamma_x\|_\infty\operatorname{Tr}(S_{x,h}P_xS_{x,h}) \underset{\text{(i)}}{=} \|\Gamma_x\|_\infty\|h\|_{\theta_1}^2\,.\end{aligned}$$

Putting these together and using Lemma G.5,

$$\operatorname{Tr}\big(g^{-1}\mathrm{D}^2\theta_1(x)[h,h]\big) \ge -22\|\Gamma_x\|_\infty\|h\|_{\theta_1}^2 \ge -\frac12\|h\|_{\theta_1}^2\,,$$

and it follows from $g_2 = 44\sqrt{\frac{m}{d}}\,\big(\theta_1 + \frac{d}{m}\theta_2\big)$ that $\operatorname{Tr}\big(g^{-1}\mathrm{D}^2g_2(x)[h,h]\big) \ge -\frac12\|h\|_{g_2}^2$. ■

G.4.3. LINEAR CONSTRAINTS: STRONGLY LOWER TRACE SELF-CONCORDANCE OF LEWIS-WEIGHT

For $\theta(x) := A_x^\top W_xA_x$ (i.e., the unscaled version of g_2), we write $g_2 = c\cdot\theta$ for a constant c, which will be set to $c_1(\log m)^{c_2}\sqrt{d}$ for some constants $c_1, c_2 > 0$ later. Going forward, P_x indicates the projection matrix of $W_x^{1/2-1/p}A_x$ (i.e., $P_x = P(W_x^{1/2-1/p}A_x)$).

Lemma G.6 $\|\Gamma_x\|_\infty \le 2c^{-1}m^{\frac{2}{p+2}}$.

Proof Note that $0 \preceq \Gamma_x = \operatorname{Diag}(A_xg^{-1}A_x^\top) \preceq c^{-1}\operatorname{Diag}(A_x\theta^{-1}A_x^\top)$. By Lemma I.5-1,

$$\|\operatorname{Diag}(A_x\theta^{-1}A_x^\top)\|_\infty = \max_{i\in[m]}\frac{\big[\sigma\big(W_x^{1/2}A_x\big)\big]_i}{\big[W_x\big]_{ii}} \le 2m^{\frac{2}{p+2}}\,.$$

■

Now we show SLTSC of the Lewis-weight metric:

Proof of Lemma E.7. From (I.5), $\mathrm{D}^2\theta[h,h] \succeq -4A_x^\top W'_{x,h}S_{x,h}A_x + A_x^\top W''_{x,h}A_x$. Thus,

$$\operatorname{Tr}(g^{-1}\mathrm{D}^2\theta[h,h]) \ge \operatorname{Tr}\big(\Gamma_x(W''_{x,h} - 4W'_{x,h}S_{x,h})\big) = -4\operatorname{Tr}(\Gamma_xW'_{x,h}S_{x,h}) + \operatorname{Tr}(\Gamma_xW''_{x,h})\,.$$

As for the first term, $\operatorname{Tr}(\Gamma_xW'_{x,h}S_{x,h}) \le p\,\|\Gamma_x\|_\infty\|h\|_\theta^2$ follows from (I.7) with Γ_x replacing $s_{x,h}^2$.

As for the second term $\operatorname{Tr}(\Gamma_xW''_{x,h})$ (i.e., (I.4) with $\Gamma = \Gamma_x$), each term there is of the form $\operatorname{Tr}(\Gamma_x\operatorname{Diag}(v))$ for $v \in \mathbb{R}^m$, which can be bounded as follows:

$$\begin{aligned}\big|\operatorname{Tr}\big(\Gamma_x\operatorname{Diag}(v)\big)\big| &= |\operatorname{Tr}(\Gamma_xW_x^{\frac12}W_x^{-\frac12}\operatorname{Diag}(v))| \le \sqrt{\operatorname{Tr}(W_x^{\frac12}\Gamma_x^2W_x^{\frac12})}\sqrt{\operatorname{Tr}\big(\operatorname{Diag}(v)W_x^{-1}\operatorname{Diag}(v)\big)}\\ &\le \|\Gamma_x\|_\infty\sqrt{\operatorname{Tr}(W_x)}\|v\|_{W_x^{-1}} = \sqrt{d}\|\Gamma_x\|_\infty\|v\|_{W_x^{-1}}\,.\end{aligned}$$

Then, we obtain $|\operatorname{Tr}(\Gamma_x W''_{x,h})| \lesssim \sqrt{d}\|\Gamma_x\|_\infty \|h\|_\theta^2$ for $p = \mathcal{O}(\log m)$ by using this inequality together with the norm bounds in Lemma I.7.

Putting things together, we conclude that

$$\operatorname{Tr}(g^{-1}\mathrm{D}^2\theta[h,h]) \gtrsim -p\|\Gamma_x\|_\infty\|h\|_\theta^2 - \sqrt{d}\|\Gamma_x\|_\infty\|h\|_\theta^2 \gtrsim -c^{-1}\sqrt{d}\|h\|_\theta^2\,,$$

where the last line follows from Lemma G.6. Therefore, there exists positive constants d_1 and d_2 such that $\operatorname{Tr}(g^{-1}\mathrm{D}^2\theta[h,h]) \geq -c^{-1}d_1(\log m)^{d_2}\sqrt{d}\|h\|_\theta^2$, which implies

$$\operatorname{Tr}(g^{-1}\mathrm{D}^2 g_2[h,h]) \geq -c^{-1}d_1(\log m)^{d_2}\sqrt{d}\|h\|_{g_2}^2\,.$$

By taking $c = d_1(\log m)^{d_2}\sqrt{d}$, the metric $g_2 = c\theta = d_1(\log m)^{d_2}\sqrt{d}A_x^\mathsf{T} W_x A_x$ is SLTSC. ■

G.4.4. LINEAR CONSTRAINTS: STRONGLY AVERAGE SELF-CONCORDANCE

We proceed with a general form of the metric $g(x) = A_x^\mathsf{T} D_x A_x$ with a diagonal matrix $0 \prec D_x \in \mathbb{R}^m$. Then we provide computational lemmas used when proving SASC of barriers for the linear constraints.

We pick any $g' : \operatorname{int}(K) \to \mathbb{S}_+^d$ such that $\bar{g} := g + g' \succ 0$. By affine invariance, we may assume $\bar{g}(x) = I$ and $x = 0$. Note that $g(x) \preceq I_d$, and z equals $rh/\sqrt{d}$ for $h \sim \mathcal{N}(0, I_d)$ in law. Applying Taylor's expansion to $\|z-x\|_{g(z)}^2$ at $z = x$ (as in the proof of Lemma D.15), for some $p_z \in [x,z]$

$$\left|\|z-x\|_{g(z)}^2 - \|z-x\|_{g(x)}^2\right| \leq \frac{r^2}{d}\left(\frac{r}{\sqrt{d}}\underbrace{|\mathrm{D}g(x)[h^{\otimes 3}]|}_{=:\mathsf{A}} + \frac{r^2}{2d}\underbrace{|\mathrm{D}^2 g(p_z)[h^{\otimes 4}]|}_{=:\mathsf{B}}\right).$$

It suffices to show that $|\mathrm{D}g(x)[h^{\otimes 3}]| = \mathcal{O}(d^{1/2})$ and $|\mathrm{D}^2 g(p_z)[h^{\otimes 4}]| = \mathcal{O}(d)$ with high probability.

Term A. By (G.15), we have $\mathrm{D}g(x)[h^{\otimes 3}] = -2s_{x,h}^\mathsf{T} D_x S_{x,h} s_{x,h} + s_{x,h}^\mathsf{T} D'_{x,h} s_{x,h}$. Let a_i denote the i-th row of A_x for $i \in [m]$, and define two polynomials in h as follows:

$$P_1(h) := s_{x,h}^\mathsf{T} D_x S_{x,h} s_{x,h} = \operatorname{Tr}(D_x S_{x,h}^3) = \sum_{i=1}^m d_i\,(a_i^\mathsf{T} h)^3\,, \quad \text{and} \quad P_2(h) := s_{x,h}^\mathsf{T} D'_{x,h} s_{x,h}\,. \tag{G.17}$$

By Lemma J.1, $D_x^{1/2} A_x A_x^\mathsf{T} D_x^{1/2} \preceq P(D_x^{1/2} A_x)$ and thus

$$\max_{i\in[m]} \|a_i\|^2 = \|\operatorname{Diag}(A_x A_x^\mathsf{T})\|_\infty \leq \max_i \frac{[\sigma(D_x^{1/2} A_x)]_i}{[D_x]_{ii}}\,. \tag{G.18}$$

By Lemma J.3,

$$\begin{aligned}\mathbb{E}[P_1(h)^2] &= \mathbb{E}\Big[\Big\{\sum_{i=1}^m d_i (a_i\cdot h)^3\Big\}^2\Big] = 9\sum_{i,j=1}^m \|d_i^{1/3} a_i\|^2 \|d_j^{1/3} a_j\|^2 \langle d_i^{1/3} a_i, d_j^{1/3} a_j\rangle + 6\sum_{i,j}\langle d_i^{1/3} a_i, d_j^{1/3} a_j\rangle^3 \\ &= 9\cdot 1^\mathsf{T} \operatorname{Diag}(A_x A_x^\mathsf{T})\, D_x^{1/2} \underbrace{D_x^{1/2} A_x A_x^\mathsf{T} D_x^{1/2}}_{\preceq P(D_x^{1/2} A_x) \preceq I_m} D_x^{1/2} \operatorname{Diag}(A_x A_x^\mathsf{T})\, 1 + 6\sum_{i,j} d_i d_j (a_i\cdot a_j)^3\end{aligned}$$

$$
\begin{aligned}
&\lesssim \|\operatorname{Diag}(A_x A_x^\mathsf{T})\|_\infty \operatorname{Tr}\bigl(\operatorname{Diag}(A_x A_x^\mathsf{T})\, D_x\bigr) + \max_i \|a_i\|^2 \cdot \sum_{i,j} d_i d_j (a_i \cdot a_j)^2 \\
&= \max_i \|a_i\|^2 \operatorname{Tr}(A_x^\mathsf{T} D_x A_x) + \max_i \|a_i\|^2 \cdot \sum_j \operatorname{Tr}(d_j a_j^\mathsf{T} A_x^\mathsf{T} D_x A_x a_j) \\
&\underset{\text{(i)}}{\le} 2 \max_i \|a_i\|^2 \operatorname{Tr}(A_x^\mathsf{T} D_x A_x) \le 2d \max_i \|a_i\|^2 \,, \qquad \text{(G.19)}
\end{aligned}
$$

where (i) follows from $A_x^\mathsf{T} D_x A_x \preceq I_d$ and $\sum_j \operatorname{Tr}(d_j a_j^\mathsf{T} A_x^\mathsf{T} D_x A_x a_j) \le \sum_j \operatorname{Tr}(d_j a_j^\mathsf{T} a_j) = \operatorname{Tr}(A_x^\mathsf{T} D_x A_x)$.

Another polynomial $P_2(h)$ requires a different strategy for bounding $\mathbb{E}[P_2(h)^2]$ for each barrier. This polynomial vanishes for the log-barrier, while the Vaidya and Lewis-weight metrics requires rather involved tasks for bounding $\mathbb{E}[P_2(h)^2]$.

Term B. Due to (I.6) (with W_x replaced by D_x), $|\mathrm{D}^2 g(p_z)[h^{\otimes 4}]|$ consists of three polynomials:

$$
\bar{P}_3(h) := \operatorname{Tr}(D_{p_z} S_{p_z,h}^4)\,, \quad \bar{P}_4(h) = \operatorname{Tr}(D'_{p_z,h} S_{p_z,h}^2)\,, \quad \bar{P}_5(h) = \operatorname{Tr}(D''_{p_z,h} S_{p_z,h}^2)\,. \qquad \text{(G.20)}
$$

For each $i = 3,4,5$, we define $P_i(h)$ by $\bar{P}_i(h)$ with p_z replaced by x. For the log-barrier, $\bar{P}_3(h)$ only matters since $D_{(\cdot)} = I_m$. For the Vaidya metric, $\bar{P}_4(h)$ and $\bar{P}_5(h)$ can be bounded by multiples of $\bar{P}_3(h)$. For the Lewis-weight metric, each $\bar{P}_i$ requires a different procedure for bounding $\mathbb{E}[\bar{P}_i(h)^2]$. Moreover, we can show $\bar{P}_i(h) \lesssim P_i(h)$ and

$$
\begin{aligned}
\mathbb{E}[P_3(h)^2] &= \sum_{i,j \in [m]} \mathbb{E}[d_i d_j\, (a_i \cdot h)^4 (a_j \cdot h)^4] \underset{\text{CS}}{\le} \sum_{i,j} d_i d_j \sqrt{\mathbb{E}[(a_i \cdot h)^8]} \sqrt{\mathbb{E}[(a_j \cdot h)^8]} \\
&\underset{\text{(i)}}{\lesssim} \Bigl(\sum_i d_i \|a_i\|^4\Bigr)^2 \le \max_i \|a_i\|^4 \Bigl(\sum_i d_i \|a_i\|^2\Bigr)^2 \underset{\text{(ii)}}{\le} d^2 \max_i \|a_i\|^4 \,, \qquad \text{(G.21)}
\end{aligned}
$$

where we used $a_i \cdot h \sim \mathcal{N}(0, \|a_i\|^2)$ in (i), and $\sum_i d_i \|a_i\|^2 = \operatorname{Tr}(A_x^\mathsf{T} D_x A_x) \le \operatorname{Tr}(I_d)$ in (ii).

We now show SASC of the three barriers for linear constraints, using this proof outline.

SASC of log-barriers (Lemma E.10). Set $g(x) = A_x^\mathsf{T} A_x$ (with $D_x = I_m$). By (G.18),

$$
\max_{i \in [m]} \|a_i\|^2 \le \max[\sigma(A_x^{1/2})]_i \le 1\,.
$$

As for the term A, it suffices to bound $P_1(h) = \operatorname{Tr}(S_{x,h}^3)$. Since $\mathbb{E}[P_1(h)^2] \lesssim d$ by (G.19), by Lemma E.8 with $t = (2e)^{3/2} \vee \bigl(\frac{2e}{3} \log \frac{2}{\varepsilon}\bigr)^{3/2}$ and $r_1(\varepsilon) := \varepsilon(2\sqrt{60}t)^{-1}$, we have that for any $r \le r_1(\varepsilon)$,

$$
\text{Event } B_1 : \quad \mathbb{P}_h\Bigl(\frac{r}{\sqrt{d}}\, |P_1(h)| \ge \varepsilon\Bigr) \le \varepsilon\,.
$$

As for the term B, recall $\mathbb{P}_z\bigl(\|z\| \ge -r \cdot 2\log\varepsilon\bigr) \le \varepsilon$ and call this event B_2. We take $r_2(\varepsilon)$ so that $1 - 2r_2 \log \varepsilon \le 1.1$, which ensures $\|z\| \le 2r$ conditioned on B_2^c for $r \le r_2$. Next, we establish coordinate-wise closeness of s_x at close-by points. Let $x_t = x + \frac{tr}{\sqrt{d}} h$, and $s_t = Ax_t - b$. For $t \in [0,1]$,

$$
\Bigl\| S_0^{-1} \frac{\mathrm{d}s_t}{\mathrm{d}t} \Bigr\|_\infty = \frac{r}{\sqrt{d}} \|A_x h\|_\infty \le \frac{r}{\sqrt{d}} \|h\|_{g(x)} \le \frac{r}{\sqrt{d}} \|h\| = \|z\|\,,
$$

and conditioned on $z \in B_2^c$ we know $\|z\| \le 2r \log \frac{1}{\varepsilon} \le 0.1$ for $r \le r_2$. Hence,

$$\max_{i\in[m]} \Big|\frac{s_{p,i} - s_{x,i}}{s_{x,i}}\Big| \le \int_0^1 \Big\| S_0^{-1} \frac{\mathrm{d}s_t}{\mathrm{d}t}\Big\|_\infty \, \mathrm{d}t \le 0.1\,,$$

and thus $1.2 \ge s_{x,i}/s_{p,i} \ge 0.9$ for all $i \in [m]$ (i.e., $S_p^{-1} \preceq 1.2 S_x^{-1}$).

Using this, we bound $\bar{P}_3(h) = \operatorname{Tr}(S_{p,h}^4)$ by a multiple of $P_3(h) = \operatorname{Tr}(S_{x,h}^4)$ as follows:

$$\operatorname{Tr}(S_{p,h}^4) = \operatorname{Tr}(h^\mathsf{T} A^\mathsf{T} S_{p,h} S_p^{-2} S_{p,h} A h) \le 2 \operatorname{Tr}(h^\mathsf{T} A^\mathsf{T} S_{p,h} S_x^{-2} S_{p,h} A h) = 2 \operatorname{Tr}(S_{x,h}^2 S_{p,h}^2) \le 4 \operatorname{Tr}(S_{x,h}^4)\,.$$

Hence, $\mathbb{E}[\bar{P}_3(h)^2] \lesssim \mathbb{E}[P_3(h)^2] \lesssim d^2$ by (G.21). Using Lemma E.8 with $t = (2e)^2 \vee \big(\frac{2e}{4} \log \frac{2}{\varepsilon}\big)^{3/2}$ and taking $r_3(\varepsilon) := (\varepsilon/c_1 t)^{1/2}$, we obtain

$$\text{Event } B_3: \quad \mathbb{P}\Big(\frac{r^2}{2d} \cdot 16 \bar{P}_3(h) \ge \varepsilon\Big) \ge \varepsilon\,,$$

Combining bounds on A and B conditioned on $\cap_i B_i^c$, we have with probability at least $1 - 3\varepsilon$

$$\big|\|z - x\|_{g(z)}^2 - \|z - x\|_{g(x)}^2\big| \le 2\varepsilon \frac{r^2}{d} \quad \text{for any } r \le \min_i r_i(\varepsilon)\,.$$

By replacing $3\varepsilon \leftarrow \varepsilon$, the claim follows.

SASC of Vaidya metric (Lemma E.11). Set $g(x) = A_x^\mathsf{T} D_x A_x$ with $D_x = \sqrt{\frac{m}{d}}(\Sigma_x + \frac{d}{m} I_m)$. By (G.18) and (G.16),

$$\max_{i\in[m]} \|a_i\|^2 \le \max_i \frac{[\sigma(D_x^{1/2} A_x)]_i}{[D_x]_{ii}} \le 1\,.$$

Term A. As A consists of P_1 and P_2 (see (G.17)), we show $\mathbb{E}[P_i(h)^2] \lesssim d$ for $i \in [2]$, which by Lemma E.8 implies $|\mathsf{A}| \le \sqrt{d}$ w.h.p. As for $P_1(h) = \operatorname{Tr}(D_x S_{x,h}^3)$, we have $\mathbb{E}[P_1(h)]^2 \lesssim d$ from (G.19).

As for $P_2(h) = \operatorname{Tr}(D'_{x,h} S_{x,h}^2)$, our approach is similar to Chen et al. (2018). By Lemma I.4,

$$\begin{aligned}|P_2(h)| &= \Big|\sqrt{\frac{m}{d}} \operatorname{Tr}\Big(\operatorname{Diag}\big((\Sigma_x - P_x^{(2)})\, s_{x,h}\big)\, S_{x,h}^2\Big)\Big| \\ &\le |P_1(h)| + |\operatorname{Tr}(S_{x,h}^3)| + \sqrt{\frac{m}{d}}\, \big|\operatorname{Tr}\big(\operatorname{Diag}(P_x^{(2)} s_{x,h})\, S_{x,h}^2\big)\big|\,.\end{aligned}$$

Since we already established a high-probability bound for both $|P_1(h)|$ and $|\operatorname{Tr}(S_{x,h}^3)|$ (which is $P_1(h)$ for the log-barrier), we focus on the third term in the RHS.

For $\sigma_x := \operatorname{diag}(P_x)$ and $\sigma_{x,i,j} := (P_x)_{ij}$, it follows from $P_x^2 = P_x$ that $\sigma_{x,i} = \sum_j \sigma_{x,i,j}^2$. Hence,

$$\operatorname{Tr}(\Sigma_x S_{x,h}^3) = 1^\mathsf{T} \Sigma_x s_{x,h}^3 = \sum_i (s_{x,h})_i^3 \sigma_{x,i} = \sum_{i,j=1}^m \sigma_{x,i,j}^2 (s_{x,h})_i^3\,,$$

$$\operatorname{Tr}\big(\operatorname{Diag}(P_x^{(2)} s_{x,h})\, S_{x,h}^2\big) = \sum_{i,j=1}^m \sigma_{x,i,j}^2 (s_{x,h})_i^2 (s_{x,h})_j \underset{\text{symmetry}}{=} \sum_{i,j=1}^m \sigma_{x,i,j}^2 (s_{x,h})_j^2 (s_{x,h})_i\,.$$

Combining these leads to

$$\begin{aligned}
&2\,\mathrm{Tr}(\Sigma_x S_{x,h}^3) + 6\,\mathrm{Tr}\big(\mathrm{Diag}(P_x^{(2)} s_{x,h}) S_{x,h}^2\big) \\
&= \sum_{i,j=1}^m \sigma_{x,i,j}^2 \big((s_{x,h})_i^3 + 3(s_{x,h})_i^2 (s_{x,h})_j + 3(s_{x,h})_i (s_{x,h})_j^2 + (s_{x,h})_j^3\big) = \sum_{i,j=1}^m \sigma_{x,i,j}^2 \big((s_{x,h})_i + (s_{x,h})_j\big)^3\,,
\end{aligned}$$

so we handle $\sum_{i,j} \sigma_{x,i,j}^2 \big((s_{x,h})_i + (s_{x,h})_j\big)^3$ instead of $\mathrm{Tr}\big(\mathrm{Diag}(P_x^{(2)} s_{x,h}) S_{x,h}^2\big)$, as we already bounded $\sqrt{\frac{m}{d}}\,\mathrm{Tr}(\Sigma_x S_{x,h}^3) = P_1(h) - \sqrt{\frac{d}{m}}\,\mathrm{Tr}(S_{x,h}^3)$. Due to $(s_{x,h})_i + (s_{x,h})_j = (a_i + a_j)^\mathsf{T} h$, for $c_{ij} := a_i + a_j$

$$\begin{aligned}
\mathbb{E}\Big[\Big\{\sum_{i,j\in[m]} \sigma_{x,i,j}^2 \big((s_{x,h})_i + (s_{x,h})_j\big)^3\Big\}^2\Big] &= \sum_{i,j,k,l} \sigma_{x,i,j}^2 \sigma_{x,k,l}^2 \mathbb{E}[(c_{ij}\cdot h)^3 (c_{kl}\cdot h)^3] \\
&\underset{\text{Lemma J.3}}{=} 9 \sum_{i,j,k,l} \sigma_{x,i,j}^2 \sigma_{x,k,l}^2 \|c_{ij}\|^2 \|c_{kl}\|^2 (c_{ij}\cdot c_{kl}) + 6 \sum_{i,j,k,l} \sigma_{x,i,j}^2 \sigma_{x,k,l}^2 (c_{ij}\cdot c_{kl})^3\,. \qquad \text{(G.22)}
\end{aligned}$$

As for the first term in (G.22), we denote $z_i := \sum_j \sigma_{x,i,j}^2 \|c_{ij}\|^2$ and $Z := \mathrm{Diag}\big((z_i)_{i\in[m]}\big)$. Then,

$$\begin{aligned}
&\sum_{i,j,k,l} \sigma_{x,i,j}^2 \sigma_{x,k,l}^2 \|c_{ij}\|^2 \|c_{kl}\|^2 (c_{ij}\cdot c_{kl}) = \Big\|\sum_{ij} \sigma_{x,i,j}^2 \|c_{ij}\|^2 c_{ij}\Big\|^2 \\
&\le 2\Big\|\sum_{ij} \sigma_{x,i,j}^2 \|c_{ij}\|^2 a_i\Big\|^2 + 2\Big\|\sum_{ij} \sigma_{x,i,j}^2 \|c_{ij}\|^2 a_j\Big\|^2 = 4\Big\|\sum_{ij} \sigma_{x,i,j}^2 \|c_{ij}\|^2 a_i\Big\|^2 = \Big\|\sum_i z_i a_i\Big\|^2 \\
&= 1^\mathsf{T} Z A_x A_x^\mathsf{T} Z\, 1 \le 1^\mathsf{T} Z D_x^{-1/2} P(D_x^{1/2} A_x)\, D_x^{-1/2} Z\, 1 \le 1^\mathsf{T} Z D_x^{-1} Z\, 1 \lesssim \sqrt{\frac{d}{m}}\,\mathrm{Tr}(Z)\,, \qquad \text{(G.23)}
\end{aligned}$$

where the last inequality follows from $Z \precsim \Sigma_x \preceq \sqrt{\frac{d}{m}} D_x$ due to

$$z_i \le 2\sum_j \sigma_{x,i,j}^2 (\|a_i\|^2 + \|a_j\|^2) \lesssim \underbrace{\sigma_{x,i}\|a_i\|^2 + \sum_j \sigma_{x,i,j}^2 \|a_j\|^2}_{=:\mathsf{K}_i} \le \sigma_{x,i}\|a_i\|^2 + \sigma_{x,i} \lesssim \sigma_{x,i}\,.$$

Moreover, using the bound in K_i and $\sum_{i,j} \sigma_{x,i,j}^2 \|a_j\|^2 = \sum_j \sigma_{x,i}\|a_i\|^2$

$$\mathrm{Tr}(Z) \lesssim \sum_i (\sigma_{x,i}\|a_i\|^2 + \sum_j \sigma_{x,i,j}^2 \|a_j\|^2) = 2\,\mathrm{Tr}(A_x^\mathsf{T} \Sigma_x A_x) \lesssim \sqrt{\frac{d}{m}}\,\mathrm{Tr}(A_x^\mathsf{T} D_x A_x) \le d\sqrt{\frac{d}{m}}\,.$$

Putting this into (G.23), we obtain $\sum_{i,j,k,l} \sigma_{x,i,j}^2 \sigma_{x,k,l}^2 \|c_{ij}\|^2 \|c_{kl}\|^2 (c_{ij}\cdot c_{kl}) \lesssim d^2/m$.

As for the second term in (G.22),

$$\begin{aligned}
&\sum_{i,j,k,l} \sigma_{x,i,j}^2 \sigma_{x,k,l}^2\, (c_{ij}\cdot c_{kl})^3 \lesssim \sum_{i,j,k,l} \sigma_{x,i,j}^2 \sigma_{x,k,l}^2\, |c_{ij}\cdot c_{kl}|^2 \\
&\le \sum_{i,j,k,l} \sigma_{x,i,j}^2 \sigma_{x,k,l}^2\, (a_i\cdot a_k + a_i\cdot a_l + a_j\cdot a_k + a_j\cdot a_l)^2 \lesssim \sum_{i,j,k,l} \sigma_{x,i,j}^2 \sigma_{x,k,l}^2\, (a_i\cdot a_k)^2
\end{aligned}$$

$$= \sum_{ik} \sigma_i \sigma_k \,(a_i \cdot a_k)^2 = \sum_k \mathrm{Tr}(\sigma_k a_k^\mathsf{T} A_x^\mathsf{T} \Sigma_x A_x a_k) \le \sqrt{\frac{d}{m}} \sum_k \mathrm{Tr}(\sigma_k a_k^\mathsf{T} a_k)$$
$$= \sqrt{\frac{d}{m}}\, \mathrm{Tr}(A_x^\mathsf{T} \Sigma_x A_x) \le \frac{d^2}{m}\,.$$

This establish a high-probability bound of $\mathcal{O}(d^2/m)$ on (G.22), implying an $\mathcal{O}(\sqrt{d})$-high-probability bound on $\sqrt{\frac{m}{d}} \big| \mathrm{Tr}\big(\mathrm{Diag}(P_x^{(2)} s_{x,h})\, S_{x,h}^2\big) \big|$.

Term B. We show that s_x and s_{p_z} are close, and the same holds for σ_x and σ_{p_z}. For s_x, following the argument for the log-barrier, we let $x_t := x + th\frac{r}{\sqrt{d}}$ and $s_t := Ax_t - b$. For $0 \le t \le 1$,

$$\Big\| S_0^{-1} \frac{\mathrm{d}s_t}{\mathrm{d}t} \Big\|_\infty = \frac{r}{\sqrt{d}} \, \|A_x h\|_\infty \underset{\text{(G.16)}}{\le} \frac{r}{\sqrt{d}} \, \|h\|_{A_x^\mathsf{T} D_x A_x} \le \frac{r}{\sqrt{d}} \, \|h\| = \|z\|\,.$$

Conditioned on the high-probability bound of $\|z\| \le 2r \log \frac{1}{\varepsilon} \le 0.1$ for any r less than some $r(\varepsilon)$,

$$\max_{i \in [m]} \Big| \frac{s_{p,i} - s_{x,i}}{s_{x,i}} \Big| \le \int_0^1 \Big\| S_0^{-1} \frac{\mathrm{d}s_t}{\mathrm{d}t} \Big\|_\infty \mathrm{d}t \le 0.1\,,$$

and thus $1.2 \ge s_{x,i}/s_{p,i} \ge 0.9$ for all $i \in [m]$ (i.e., $S_p^{-1} \preceq 1.2 S_x^{-1}$). For σ_x, as we have $\Sigma_x = \mathrm{Diag}(A_x(A_x^\mathsf{T} A_x)^{-1} A_x^\mathsf{T})$, we have the same closeness between $\sigma_{x,i}$ and $\sigma_{p,i}$ for each $i \in [m]$.

Using the formulas in Lemma I.4,

$$\begin{aligned}
|\mathrm{D}^2 g(p)[h^{\otimes 4}]| &\lesssim \sqrt{\frac{m}{d}} \Big(\mathrm{Tr}\big((\Sigma_p + \frac{d}{m} I_m) S_{p,h}^4\big) + \underbrace{\mathrm{Tr}(S_{p,h}^2 P_p S_{p,h} P_p S_{p,h})}_{(*)} \\
&\qquad + \mathrm{Tr}(S_{p,h}^2 P_p S_{p,h}^2 P_p) + \underbrace{\mathrm{Tr}(S_{p,h} P_p S_{p,h} P_p S_{p,h} P_p S_{p,h})}_{\le \mathrm{Tr}(S_{p,h}^2 P_p S_{p,h}^2 P_p)} \Big) \\
&\underset{\text{(i)}}{\lesssim} \sqrt{\frac{m}{d}} \Big(\mathrm{Tr}\Big((\Sigma_p + \frac{d}{m} I_m) S_{p,h}^4\Big) + \mathrm{Tr}(S_{p,h}^2 \Sigma_p S_{p,h}^2) + \underbrace{\mathrm{Tr}(S_{p,h}^2 P_p S_{p,h}^2 P_p)}_{\text{Use Lemma H.1}} \Big) \\
&\underset{\text{(ii)}}{\lesssim} \sqrt{\frac{m}{d}}\, \mathrm{Tr}\Big((\Sigma_p + \frac{d}{m} I_m) S_{p,h}^4\Big) \underset{\text{(iii)}}{\lesssim} \sqrt{\frac{m}{d}}\, \mathrm{Tr}\Big((\Sigma_x + \frac{d}{m} I_m) S_{x,h}^4\Big) = P_3(h)\,,
\end{aligned}$$

where in (i) we used the Cauchy-Schwarz inequality on $(*)$:

$$\begin{aligned}
&\mathrm{Tr}(S_{p,h}^2 P_p S_{p,h} P_p S_{p,h}) \le \sqrt{\mathrm{Tr}(S_{p,h}^2 P_p^2 S_{p,h}^2)} \sqrt{\mathrm{Tr}(S_{p,h} P_p S_{p,h}^2 P_p S_{p,h})} \\
&\underset{\text{AM-GM}}{\le} \frac{1}{2} \big(\mathrm{Tr}(S_{p,h}^2 P_p^2 S_{p,h}^2) + \mathrm{Tr}(S_{p,h} P_p S_{p,h}^2 P_p S_{p,h}) \big) \le \frac{1}{2} \big(\mathrm{Tr}(S_{p,h}^2 \Sigma_p S_{p,h}^2) + \mathrm{Tr}(S_{p,h}^2 P_p S_{p,h}^2 P_p) \big)\,,
\end{aligned}$$

(ii) follows from $\mathrm{Tr}(S_{p,h}^2 P_p S_{p,h}^2 P_p) = s_{p,h}^2 \cdot P_p^{(2)} s_{p,h}^2 \preceq s_{p,h}^2 \cdot \Sigma_p s_{p,h}^2 \preceq s_{p,h}^2 \cdot (\Sigma_p + \frac{d}{m} I_m) s_{p,h}^2$, and in (iii) we used coordinate-wise closeness of $s_x \leftrightarrow s_p$ and $\sigma_x \leftrightarrow \sigma_p$. By (G.21), $\mathbb{E}[P_3(h)^2] \lesssim d^2$, and an $\mathcal{O}(d)$-high-probability bound on $|P_3(h)|$ (so on B) follows from Lemma E.8.

SASC of Lewis-weight (Lemma E.12). Set $g(x) = \sqrt{d}A_x^\top W_x A_x$ (with $D_x = \sqrt{d}\, W_x$). By (G.18) and Lemma I.5-1,

$$\max_{i\in[m]} \|a_i\|^2 \le \max_i \frac{[\sigma(D_x^{1/2}A_x)]_i}{[D_x]_{ii}} \le \frac{2m^{\frac{2}{p+2}}}{\sqrt{d}} \lesssim \frac{1}{\sqrt{d}}\,.$$

Term A. As done for the Vaidya metric, a high-probability bound on A requires $\mathbb{E}[P_i(h)^2] \lesssim d$ for $i = 1, 2$ (see (G.17)). Note that $\mathbb{E}[P_1(h)^2] \lesssim \sqrt{d}$ by (G.19).

As for $P_2(h) = \sqrt{d}\, s_{x,h}^\top W'_{x,h} s_{x,h}$, we show $\mathbb{E}[P_2(h)^2] \lesssim \sqrt{d}$. Due to $W'_{x,h} = -\operatorname{Diag}(W_x^{\frac12} N_x W_x^{\frac12} s_{x,h})$ (Lemma I.6), $P_2(h) = -\sqrt{d} s_{x,h}^\top \operatorname{Diag}(W_x^{\frac12} N_x W_x^{\frac12} s_{x,h}) s_{x,h} = -\sqrt{d}\operatorname{Tr}\big(\operatorname{Diag}(W_x^{\frac12} N_x W_x^{\frac12} s_{x,h}) S_{x,h}^2\big)$. Thus,

$$P_2(h) = \sqrt{d}\ \operatorname{Tr}\big(\operatorname{Diag}(N_x W_x^{\frac12} s_{x,h}) W_x^{\frac12} S_{x,h}^2\big) = \sqrt{d}\sum_{i=1}^m w_i^{1/2}(a_i\cdot h)^2(b_i\cdot h)\,,$$

where b_i is the i-th row of $B := N_x W_x^{\frac12} A_x$ for $i = 1, \dots, m$. By Lemma J.4,

$$\begin{aligned}
&\mathbb{E}\Big[\Big\{\sum_{i=1}^m w_i^{1/2}(a_i\cdot h)^2(b_i\cdot h)\Big\}^2\Big] \\
&= \sum_{i,j\in[m]} w_i^{1/2}w_j^{1/2}\|a_i\|^2\|a_j\|^2(b_i\cdot b_j) \\
&\quad + 4\sum_{i,j} w_i^{1/2}w_j^{1/2}(a_i\cdot a_j)(a_i\cdot b_i)(a_j\cdot b_j) + 4\sum_{i,j} w_i^{1/2}w_j^{1/2}\|a_i\|^2(b_i\cdot a_j)(a_j\cdot b_j) \\
&\quad + 2\underbrace{\sum_{i,j} w_i^{1/2}w_j^{1/2}(a_i\cdot a_j)^2(b_i\cdot b_j)}_{=:T_1} + 4\underbrace{\sum_{i,j} w_i^{1/2}w_j^{1/2}(a_i\cdot a_j)(a_i\cdot b_j)(a_j\cdot b_i)}_{=:T_2} \\
&= \underbrace{1^\top \operatorname{Diag}(A_xA_x^\top)\, W^{\frac12} BB^\top W^{\frac12}\ \operatorname{Diag}(A_xA_x^\top)\, 1}_{=:N_1} + 4\cdot \underbrace{1^\top \operatorname{Diag}(A_xB^\top)\, W^{\frac12} A_xA_x^\top W^{\frac12}\ \operatorname{Diag}(A_xB^\top)\, 1}_{=:N_2} \\
&\quad + 4\cdot \underbrace{[1^\top \operatorname{Diag}(A_xA_x^\top)\, W^{\frac12} B]\cdot [A_x^\top W^{\frac12}\ \operatorname{Diag}(A_xB^\top)\, 1]}_{\le N_1+N_2 \text{ by Young's inequality}} + 2T_1 + 4T_2\,.
\end{aligned}$$

As for N_1, since $B^\top B = A_x^\top W_x^{\frac12} N_x^2 W_x^{\frac12} A_x \le p^2 A_x^\top W_x A_x$ by Lemma I.8-1 and thus $B^\top B \precsim (d)^{-1/2} I_d$, Lemma J.1 ensures $BB^\top \precsim \frac{1}{\sqrt{d}} P(B) \preceq \frac{1}{\sqrt{d}} I_m$. Hence,

$$N_1 \lesssim \frac{1}{\sqrt{d}}\ \operatorname{Tr}\big(\operatorname{Diag}(A_xA_x^\top)\, W\ \operatorname{Diag}(A_xA_x^\top)\big) \le \frac{1}{\sqrt{d}}\ \operatorname{Tr}(A_x^\top W A_x)\,\|\operatorname{Diag}(A_xA_x^\top)\|_\infty \lesssim \frac{1}{\sqrt{d}}\,.$$

As for N_2, due to $A_x^\top W_x A_x \preceq \frac{1}{\sqrt{d}} I_d$ we have $W^{\frac12} A_xA_x^\top W^{\frac12} \preceq \frac{1}{\sqrt{d}} I_m$ by Lemma J.1. Thus,

$$N_2 \lesssim \frac{1}{\sqrt{d}}\ \operatorname{Tr}\big(\{\operatorname{Diag}(A_xB^\top)\}^2\big) = \frac{1}{\sqrt{d}}\sum_{i\in[m]}(a_i\cdot b_i)^2 \le \frac{1}{\sqrt{d}}\sum_i \|a_i\|^2\|b_i\|^2$$

$$\le \frac{1}{d}\,\mathrm{Tr}(BB^\mathsf{T}) \lesssim \frac{1}{d^{3/2}}\,\mathrm{Tr}\big(P(B)\big) \le \frac{1}{\sqrt{d}}\,.$$

As for T_1, by Young's inequality (i.e., $2(a\cdot b)\le\|a\|^2+\|b\|^2$)

$$\begin{aligned}
T_1 &= \sum_{i,j\in[m]} (a_i\cdot a_j)^2\,\big((w_j^{1/2}b_i)\cdot(w_i^{1/2}b_j)\big) \lesssim \sum_{i,j}(a_i\cdot a_j)^2\,(w_j\|b_i\|^2+w_i\|b_j\|^2)\\
&= 2\sum_{i,j} w_j(a_i\cdot a_j)^2\|b_i\|^2 = \sum_i \|b_i\|^2\cdot\mathrm{Tr}\Big(a_i^\mathsf{T}\Big(\sum_j a_j w_j a_j^\mathsf{T}\Big)a_i\Big)\\
&= \sum_i \|b_i\|^2\,\mathrm{Tr}(a_i^\mathsf{T}A_x^\mathsf{T}WA_x a_i) \le \frac{1}{\sqrt{d}}\sum_i \|b_i\|^2\|a_i\|^2 \le \frac{1}{d}\,\mathrm{Tr}(BB^\mathsf{T}) \le \frac{1}{\sqrt{d}}\,.
\end{aligned}$$

As for T_2, using $(a_i\cdot a_j)\le\|a_i\|\|a_j\|\lesssim\frac{1}{\sqrt{d}}$

$$\begin{aligned}
T_2 &= \sum_{i,j\in[m]} w_i^{1/2}w_j^{1/2}(a_i\cdot a_j)(a_i\cdot b_j)(a_j\cdot b_i) \lesssim \frac{1}{\sqrt{d}}\sum_{i,j\in[m]} w_i^{1/2}w_j^{1/2}(a_i\cdot b_j)(a_j\cdot b_i)\\
&= \frac{1}{\sqrt{d}}\sum_i w_i^{1/2}b_i^\mathsf{T}\sum_j a_j w_j^{1/2}b_j^\mathsf{T}a_i = \frac{1}{\sqrt{d}}\sum_i \mathrm{Tr}(a_i w_i^{1/2}b_i^\mathsf{T}A_x^\mathsf{T}W^{1/2}B)\\
&= \frac{1}{\sqrt{d}}\,\mathrm{Tr}\big((A_x^\mathsf{T}W^{1/2}B)^2\big) \underset{\mathrm{CS}}{\le} \frac{1}{\sqrt{d}}\,\mathrm{Tr}(B^\mathsf{T}W^{1/2}A_xA_x^\mathsf{T}W^{1/2}B) \le \frac{1}{d}\,\mathrm{Tr}(B^\mathsf{T}B) \le \frac{1}{\sqrt{d}}\,.
\end{aligned}$$

Putting all the bounds together, we have $\mathbb{E}[P_2(h)^2]\lesssim d\cdot\frac{1}{\sqrt{d}}=\sqrt{d}$.

Term B. We show that for any given $\alpha=\Theta(1)$, each coordinate of w_x/s_x^α and $w_{p_z}/s_{p_z}^\alpha$ is close. For $0\le t\le 1$, we define $x_t := x+\frac{r}{\sqrt{d}}th$, and s_t, w_t in the same fashion. Then for $p=\mathcal{O}(\log m)$,

$$\max_{i\in[m]}\Big|\log\frac{(w_{p_z,i})^\alpha}{s_{p_z,i}}-\log\frac{(w_{x,i})^\alpha}{s_{x,i}}\Big| \le \int_0^1\Big|\frac{\mathrm{d}}{\mathrm{d}t}\log\frac{[w_{t,i}]^\alpha}{s_{t,i}}\Big|\,\mathrm{d}t \lesssim \frac{r}{\sqrt{d}}\,\|h\|_{A_x^\mathsf{T}W_xA_x} \le \frac{1}{d^{1/4}}\|z\|\,.$$

Just as in showing SASC of the Vaidya metric, we can make this bound arbitrarily small (say $\delta\approx 0$) by conditioning on the high-probability region where $\|z\|\le r\log\frac{1}{\varepsilon}\le 0.01$. Hence,

$$e^{-\delta}\frac{(w_{x,i})^\alpha}{s_{x,i}} \le \frac{(w_{p_z,i})^\alpha}{s_{p_z,i}} \le e^{\delta}\frac{(w_{x,i})^\alpha}{s_{x,i}}\,. \tag{G.24}$$

We remark that this $\Theta(1)$-multiplicative closeness is still valid without the $\sqrt{d}$-scaling of $A_x^\mathsf{T}W_xA_x$.

Using the formula for $\mathrm{D}^2(A_x^\mathsf{T}W_xA_x)[h^{\otimes 4}]$ in (I.6),

$$\begin{aligned}
&|\mathrm{D}^2 g(p)[h^{\otimes 4}]| \lesssim \big(\bar{P}_3(h)+|\bar{P}_4(h)|+|\bar{P}_5(h)|\big) = \bar{P}_3(h)+\sqrt{d}\,\big(|\,\mathrm{Tr}(W'_{p,h}S^3_{p,h})|+|\,\mathrm{Tr}(W''_{p,h}S^2_{p,h})|\big)\\
&= \bar{P}_3(h)+\sqrt{d}\underbrace{\big|\,\mathrm{Tr}\big(S^3_{p,h}\,\mathrm{Diag}(W_p^{\frac12}N_pW_p^{\frac12}s_{p,h})\big)\big|}_{=:T_1}+\sqrt{d}\underbrace{|\,\mathrm{Tr}(S^2_{p,h}W''_{p,h})|}_{=:T_2}\,,
\end{aligned}$$

where in the last line we used the formula for $W'_{p,h}$ (Lemma I.6).

Now we show $\mathbb{E}[\bar{P}_3(h)^2]\lesssim d^2$ and $T_i\lesssim\sqrt{d}$ w.h.p. for $i=4,5$. As for $\bar{P}_3$, we have $\bar{P}_3(h)\lesssim P_3(h)$ from the closeness (G.24) of w_i/s_i^4 for each $i\in[m]$, so $\mathbb{E}[P_3(h)^2]\lesssim d^2\cdot d^{-1}=d$ from (G.21).

As for T_1, using the Cauchy-Schwarz

$$\begin{aligned}T_1 &= \big|\operatorname{Tr}\big(S_{p,h}^3 W_p^{\frac{1}{2}}\ \operatorname{Diag}(N_p W_p^{\frac{1}{2}} s_{p,h})\big)\big| \le \sqrt{\operatorname{Tr}(S_{p,h}^3 W_p S_{p,h}^3)}\sqrt{s_{p,h}^\mathsf{T} W_p^{1/2} N_p^2 W_p^{1/2} s_{p,h}} \\ &\underset{\text{(i)}}{\lesssim} \sqrt{s_{p,h}^3 W_p s_{p,h}^3}\sqrt{s_{p,h}^\mathsf{T} W_p s_{p,h}} \underset{\text{(ii)}}{\lesssim} \sqrt{s_{x,h}^3 W_x s_{x,h}^3}\sqrt{s_{x,h}^\mathsf{T} W_x s_{x,h}} = \sqrt{s_{x,h}^3 W_x s_{x,h}^3} \cdot d^{-1/4}\|h\|_{g(x)}\,,\end{aligned}$$

where in (i) we used $N_x \preceq p^2 I$ (Lemma I.8), and in (ii) the closeness of w_i/s_i^6 and w_i/s_i^2 established in (G.24). As for the first term in the RHS,

$$\begin{aligned}\mathbb{E}[(s_{x,h}^3 W_x s_{x,h}^3)^2] &\underset{\text{CS}}{\lesssim} \sum_{i,j\in[m]} w_i w_j \sqrt{\mathbb{E}[(a_i\cdot h)^{12}]}\sqrt{\mathbb{E}[(a_j\cdot h)^{12}]} = \Big(\sum_i w_i \left(\mathbb{E}[(a_i\cdot h)^{12}]\right)^2\Big)^2 \\ &\lesssim \Big(\sum_i w_i \|a_i\|^6\Big)^2 \le \Big(\frac{1}{d^{3/2}}\sum_i w_i\Big)^2 = \frac{1}{d}\,.\end{aligned}$$

As for the second term, the concentration of the standard Gaussian guarantees $\|h\|_{g(x)} \le \|h\| \lesssim \sqrt{d}$ w.h.p. Therefore, $T_1 \lesssim \sqrt{d}$ w.h.p.

As for T_2, (I.4) with $\Gamma_p = S_{p,h}^2$ equals T_2. Following (I.8) with I, II, III, IV defined in (I.5),

$$\begin{aligned}T_2 &\lesssim \sum_{v=\text{I,II,III,IV}} \sqrt{\operatorname{Tr}(W_p S_{p,h}^4)}\|v\|_{W_p^{-1}} \underset{\text{(i)}}{\lesssim} \sqrt{\operatorname{Tr}(W_p S_{p,h}^4)}\left(\operatorname{Tr}(S_{p,h}^2 W_p) + \operatorname{Tr}(S_{p,h}^4 W_p)\right) \\ &\underset{\text{(ii)}}{\lesssim} \sqrt{\operatorname{Tr}(W_x S_{x,h}^4)}\left(\operatorname{Tr}(S_{x,h}^2 W_x) + \operatorname{Tr}(S_{x,h}^4 W_x)\right),\end{aligned}$$

where (i) follows from Lemma I.7 (i.e., $\|v\|_{W_p^{-1}} \lesssim \|h\|^2_{A_p^\mathsf{T} W_p A_p} = \operatorname{Tr}(S_{p,h}^2 W_p)$ for $v = $ I, II, III, and $\|\text{IV}\|_{W_p^{-1}} \lesssim \operatorname{Tr}(S_{p,h}^4 W_p)$), and (ii) follows from the conditioned event where the closeness of w_i/s_i^2 at x and z holds. Since we already established the high-probability bounds of $d^{-1/2}P_3(h) = \operatorname{Tr}(S_{x,h}^4 W_x) \lesssim 1$ and $\operatorname{Tr}(S_{x,h}^2 W_x) \lesssim \sqrt{d}$, combining these yield $T_2 \lesssim \sqrt{d}$ w.h.p.

G.4.5. QUADRATIC CONSTRAINTS

We show that a ν-SC barrier $\psi(\cdot) = -\log f(\cdot)$ satisfies

$$|\mathrm{D}^4\psi(x)[h^{\otimes 4}]| \lesssim \nu^2 \|h\|^2_{\nabla^2\psi(x)} + \Big|\frac{\mathrm{D}^4 f(x)[h^{\otimes 4}]}{f(x)}\Big|\,.$$

Proof of Lemma E.13. Fix $h \in \mathbb{R}^d$ and $x \in \operatorname{int}(K)$, define $\phi(t) := \psi(x+th)$. Then,

$$\begin{aligned}\phi' &= -\frac{f'}{f}\,,\\ \phi'' &= \left(\frac{f'}{f}\right)^2 - \frac{f''}{f} = (\phi')^2 - \frac{f''}{f}\,,\\ \phi''' &= 2\phi'\phi'' - \frac{f'''f - f''f'}{f^2} = 2\phi'\phi'' - \frac{f'''}{f} + \frac{f''f'}{f^2} = 2\phi'\phi'' + \phi'(\phi'' - (\phi')^2) - \frac{f'''}{f} \\ &= 3\phi'\phi'' - (\phi')^3 - \frac{f'''}{f}\,,\end{aligned}$$

$$
\begin{aligned}
\phi^{(4)} &= 3(\phi'')^2 + 3\phi'\phi''' - 3(\phi')^2\phi'' - \frac{f^{(4)}f - f'''f'}{f^2} \\
&= 3(\phi'')^2 + 3\phi'\phi''' - 3(\phi')^2\phi'' + \phi'\left(\phi''' - 3\phi'\phi'' + (\phi')^3\right) - \frac{f^{(4)}}{f} \\
&= 3(\phi'')^2 + 4\phi'\phi''' - 6(\phi')^2\phi'' + (\phi')^4 - \frac{f^{(4)}}{f}\,.
\end{aligned}
$$

Since $|\phi'''| \le 2(\phi'')^{3/2}$ (SC of ϕ) and $\phi'' \ge \frac{1}{\nu}(\phi')^2$ (the definition of the barrier parameter), which is equivalent to $|\phi'| \le \sqrt{\nu}(\phi'')^{1/2}$, we can directly compute as follows:

$$
\begin{aligned}
|\phi^{(4)}| &\le 4\,|\phi'\phi'''| + 3\,|(\phi'')^2| + 6|\,(\phi')^2\phi''| + |(\phi')^4| + \Big|\frac{f^{(4)}}{f}\Big| \\
&\le 8\sqrt{\nu}\,|\phi''|^2 + 3\,|\phi''|^2 + 6\nu\,|\phi''|^2 + \nu^2\,|\phi''|^2 + \Big|\frac{f^{(4)}}{f}\Big| \lesssim \nu^2|\phi''|^2 + \Big|\frac{f^{(4)}}{f}\Big|\,.
\end{aligned}
$$

■

Using this tool, we study Dikin-amenability of barriers for quadratic constraints.

Proof of Lemma E.14. Let us check the last claim first. By Lemma D.18, we may assume that

$$
\phi(x,y) = -\log(l + q^{\mathsf T}y - \frac{1}{2}\|x\|^2)\,,
$$

and let $f(x,y) = l + q^{\mathsf T}y - \frac{1}{2}\|x\|^2$. For $z = (x,y) \in \operatorname{int}(K)$ and $u = (u_x, u_y) \in \mathbb{R}^d$, we have

$$
\begin{aligned}
\mathrm{D}\phi(z)[u] &= -\frac{1}{f}\,(q\cdot u_y - x\cdot u_x) = \frac{x\cdot u_x - q\cdot u_y}{f}\,, \\
\mathrm{D}^2\phi(z)[u,u] &= \frac{1}{f^2}\,(x\cdot u_x - q\cdot u_y)^2 + \frac{1}{f}\,\|u_x\|^2\,. \qquad \text{(G.25)}
\end{aligned}
$$

As for the first term in the RHS of (G.25), it holds that for $v = (v_x, v_y) \in \mathbb{R}^d$

$$
\begin{aligned}
\mathrm{D}\Big(\frac{(x\cdot u_x - q\cdot u_y)^2}{f^2}\Big)[v] &= \frac{2\,(x\cdot u_x - q\cdot u_y)(v_x\cdot u_x)}{f^2} + 2\,(x\cdot u_x - q\cdot u_y)^2\cdot\frac{x\cdot v_x - q\cdot v_y}{f^3}\,, \\
\mathrm{D}^2\Big(\frac{(x\cdot u_x - q\cdot u_y)^2}{f^2}\Big)[v,v] &= \frac{2\,(v_x\cdot u_x)^2}{f^2} + 4\frac{(x\cdot u_x - q\cdot u_y)(v_x\cdot u_x)(x\cdot v_x - q\cdot v_y)}{f^3} \\
&\quad + \frac{4\,(x\cdot u_x - q\cdot u_y)(v_x\cdot u_x)(x\cdot v_x - q\cdot v_y) + 2\,(x\cdot u_x - q\cdot u_y)^2\|v_x\|^2}{f^3} \\
&\quad + \frac{6\,(x\cdot u_x - q\cdot u_y)^2(x\cdot v_x - q\cdot v_y)^2}{f^4} \\
&= \frac{2\,(v_x\cdot u_x)^2}{f^2} + \frac{4\,(x_q\cdot u)(v_x\cdot u_x)(x_q\cdot v)}{f^3} \\
&\quad + \frac{4\,(x_q\cdot u)(v_x\cdot u_x)(x_q\cdot v) + 2(x_q\cdot u)^2\|v_x\|^2}{f^3} + \frac{6\,(x_q\cdot u)^2(x_q\cdot v)^2}{f^4}\,,
\end{aligned}
$$

where $x_q := (x, -q) \in \mathbb{R}^d$.

As for the second term, direct computations lead to

$$\begin{aligned}
\mathrm{D}\Bigl(\frac{\|u_x\|^2}{f}\Bigr)[v] &= \frac{1}{f^2}\,\|u_x\|^2(x\cdot v_x - q\cdot v_y)\,,\\
\mathrm{D}^2\Bigl(\frac{\|u_x\|^2}{f}\Bigr)[v,v] &= \frac{2}{f^3}\,\|u_x\|^2(x\cdot v_x - q\cdot v_y)^2 + \frac{1}{f^2}\,\|u_x\|^2\|v_x\|^2\\
&= \frac{2}{f^3}\,\|u_x\|^2(x_q\cdot v)^2 + \frac{1}{f^2}\,\|u_x\|^2\|v_x\|^2\,.
\end{aligned}$$

Putting these together, for $u,v\in\mathbb{R}^d$

$$\begin{aligned}
&\mathrm{D}^4\phi[u,u,v,v]\\
&= \frac{1}{f^2}\,\|u_x\|^2\|v_x\|^2 + \underbrace{\frac{2}{f^2}\,(v_x\cdot u_x)^2}_{\geq 0} + \frac{4}{f^3}\,\Bigl(\frac{1}{2}\,\|u_x\|^2(x_q\cdot v)^2 + 2\,(x_q\cdot u)(v_x\cdot u_x)(x_q\cdot v) + \frac{(x_q\cdot u)^2}{2}\,\|v_x\|^2\Bigr)\\
&\quad + \frac{6}{f^4}\,(x_q\cdot u)^2(x_q\cdot v)^2\\
&\geq \frac{4}{f^3}\,\Bigl(\underbrace{\frac{1}{2}\|u_x\|^2(x_q\cdot v)^2 + \frac{1}{2}\|v_x\|^2(x_q\cdot u)^2}_{\text{Use AM-GM}} + 2(x_q\cdot u)(v_x\cdot u_x)(x_q\cdot v)\Bigr)\\
&\quad + \underbrace{\frac{1}{f^2}\,\|u_x\|^2\|v_x\|^2 + \frac{6}{f^4}\,(x_q\cdot u)^2(x_q\cdot v)^2}_{\text{Use AM-GM}}\\
&\geq \frac{4}{f^3}\,\bigl(\|u_x\|\,\|v_x\|\,|x_q\cdot v|\,|x_q\cdot u| - 2|x_q\cdot u|\,|x_q\cdot v|\,\|u_x\|\,\|v_x\|\bigr) + \frac{2\sqrt{6}}{f^3}\,|x_q\cdot u|\,|x_q\cdot v|\,\|u_x\|\,\|v_x\|\\
&= \frac{4}{f^3}\,\|u_x\|\,\|v_x\|\,|x_q\cdot v|\,|x_q\cdot u|\,\Bigl(\frac{\sqrt{6}}{2} - 1\Bigr) \geq 0\,.
\end{aligned}$$

■

G.4.6. PSD: CONVEXITY AND STRONGLY SELF-CONCORDANCE

We start with convexity of $\log\det(\nabla^2\phi)$ for $\phi(X) = -\log\det X$.

Proof of Proposition E.22. Using Lemma E.20 and $\det\bigl(M^{\mathsf{T}}(A\otimes A)M\bigr) = 2^{d(d-1)/2}\,(\det A)^{d+1}$ (Lemma H.1) in the first and second equality below,

$$\log\det\bigl(\nabla^2\phi(X)\bigr) = \log\det\bigl(M^{\mathsf{T}}(X^{-1}\otimes X^{-1})M\bigr) = \frac{d(d-1)}{2}\,\log 2 - (d+1)\,\log\det X\,.$$

Since $-\log\det X$ is convex in X (H.4), the convexity of $\log\det\bigl(\nabla^2\phi(X)\bigr)$ also follows. ■

Observe from the proof that $\log\det\bigl(\nabla^2\phi(X)\bigr) = \text{const.} + (d+1)\,\phi(X)$. Differentiating both sides in direction H, by (H.1) $\mathrm{Tr}\bigl([\nabla^2\phi(X)]^{-1}\mathrm{D}^3\phi(X)[H]\bigr) = (d+1)\,\mathrm{D}\phi(X)[H]$. Hence,

$$\mathrm{Tr}\bigl([\nabla^2\phi(X)]^{-\frac{1}{2}}\mathrm{D}^3\phi(X)[H]\,[\nabla^2\phi(X)]^{-\frac{1}{2}}\bigr) = -(d+1)\;\mathrm{Tr}(X^{-1}H)\,. \tag{G.26}$$

We are ready to show SSC of ϕ.

Proof of Lemma E.23. For $H \in \mathbb{S}^d$ and $t \in \mathbb{R}$, denote $X_t := X + tH$ and $g_t := M^\mathsf{T}(X_t \otimes X_t)^{-1}M$. Note that

$$\big\|[\nabla^2\phi(X)]^{-\frac{1}{2}}\mathrm{D}^3\phi(X)[H]\,[\nabla^2\phi(X)]^{-\frac{1}{2}}\big\|_F^2 = \mathrm{Tr}(g^{-1}\partial_t g_t|_{t=0}\,g^{-1}\partial_t g_t|_{t=0})\,,$$

and

$$\begin{aligned}\partial_t g_t|_{t=0} &\underset{\text{(i)}}{=} \partial_t\big(M^\mathsf{T}(X_t\otimes X_t)^{-1}M\big)\Big|_{t=0} \underset{\text{(ii)}}{=} -M^\mathsf{T}(X\otimes X)^{-1}\,\partial_t(X_t\otimes X_t)|_{t=0}\,(X\otimes X)^{-1}M \\ &= -M^\mathsf{T}(X^{-1}\otimes X^{-1})(H\otimes X + X\otimes H)(X^{-1}\otimes X^{-1})M \\ &\underset{\text{(iii)}}{=} -M^\mathsf{T}(X^{-1}HX^{-1}\otimes X^{-1} + X^{-1}\otimes X^{-1}HX^{-1})M\,,\end{aligned}\tag{G.27}$$

where (i) follows from Lemma E.20, (ii) is due to (H.2), and (iii) follows from $(A\otimes B)(C\otimes D) = (AC)\otimes(BD)$ (Lemma H.1-3).

Recall that positive semidefinite matrices have unique positive semidefinite square roots, so $(X\otimes X)^{\frac{1}{2}} = X^{\frac{1}{2}}\otimes X^{\frac{1}{2}}$ (due to $(X^{1/2}\otimes X^{1/2})\cdot(X^{1/2}\otimes X^{1/2}) = X\otimes X$). Since $g_t = M^\mathsf{T}(X_t\otimes X_t)^{-1/2}(X_t\otimes X_t)^{-1/2}M$, the corresponding orthogonal projection is

$$P_t := P\big((X_t\otimes X_t)^{-\frac{1}{2}}M\big) = (X_t\otimes X_t)^{-\frac{1}{2}}Mg_t^{-1}M^\mathsf{T}(X_t\otimes X_t)^{-\frac{1}{2}}\,.$$

By substituting $\partial_t g_t\big|_{t=0}$ with (G.27),

$$\begin{aligned}&\mathrm{Tr}(g^{-1}\partial_t g_t|_{t=0}\,g^{-1}\partial_t g_t|_{t=0}) \\ &= \mathrm{Tr}\big(g^{-1}M^\mathsf{T}(X^{-1}HX^{-1}\otimes X^{-1} + X^{-1}\otimes X^{-1}HX^{-1})M \\ &\qquad\cdot g^{-1}M^\mathsf{T}(X^{-1}HX^{-1}\otimes X^{-1} + X^{-1}\otimes X^{-1}HX^{-1})M\big) \\ &= \mathrm{Tr}\big(Mg^{-1}M^\mathsf{T}(X^{-1}HX^{-1}\otimes X^{-1} + X^{-1}\otimes X^{-1}HX^{-1})M \\ &\qquad\cdot g^{-1}M^\mathsf{T}(X^{-1}HX^{-1}\otimes X^{-1} + X^{-1}\otimes X^{-1}HX^{-1})\big) \\ &= \mathrm{Tr}\Big(\big[Mg^{-1}M^\mathsf{T}(X^{-1}HX^{-1}\otimes X^{-1} + X^{-1}\otimes X^{-1}HX^{-1})\big]^2\Big) \\ &= \mathrm{Tr}\Big(\big[(X\otimes X)^{\frac{1}{2}}P(X\otimes X)^{\frac{1}{2}}(X^{-1}HX^{-1}\otimes X^{-1} + X^{-1}\otimes X^{-1}HX^{-1})\big]^2\Big) \\ &= \mathrm{Tr}\Big(\big[P\underbrace{(X\otimes X)^{\frac{1}{2}}(X^{-1}HX^{-1}\otimes X^{-1} + X^{-1}\otimes X^{-1}HX^{-1})(X\otimes X)^{\frac{1}{2}}}_{=:S}\big]^2\Big) \\ &= \mathrm{Tr}(PSPS)\,.\end{aligned}$$

Using Lemma H.1-3,

$$S = \underbrace{X^{-\frac{1}{2}}HX^{-\frac{1}{2}}\otimes I_d}_{=:A} + \underbrace{I_d\otimes X^{-\frac{1}{2}}HX^{-\frac{1}{2}}}_{=:B}\,.$$

By the Cauchy-Schwarz inequality along with $P^\mathsf{T}P = P^2 = P$ and $P \preceq I_d$,

$$\mathrm{Tr}(PSPS) \le \mathrm{Tr}((PS)^\mathsf{T}PS) \le \mathrm{Tr}(S^\mathsf{T}S) = \|S\|_F^2 \le (\|A\|_F + \|B\|_F)^2\,.$$

Using Lemma H.1-3,

$$\begin{aligned}\|A\|_F^2 &= \operatorname{Tr}\big((X^{-\frac{1}{2}}HX^{-\frac{1}{2}}\otimes I_d)\cdot(X^{-\frac{1}{2}}HX^{-\frac{1}{2}}\otimes I_d)\big)\\ &= \operatorname{Tr}(X^{-\frac{1}{2}}HX^{-1}HX^{-\frac{1}{2}}\otimes I_d) = \operatorname{Tr}(X^{-\frac{1}{2}}HX^{-1}HX^{-\frac{1}{2}})\operatorname{Tr}(I_d) = d\,\|H\|_X^2\,,\end{aligned}$$

and similarly $\|B\|_F^2 = d\,\|H\|_X^2$. Therefore, $\psi_X \le 2\sqrt{d}$ follows from

$$\big\|[\nabla^2\phi(X)]^{-\frac{1}{2}}\mathrm{D}^3\phi(X)[H]\,[\nabla^2\phi(X)]^{-\frac{1}{2}}\big\|_F \le \sqrt{\operatorname{Tr}(PSPS)} \le 2\sqrt{d}\,\|H\|_X\,.$$

To see the optimality of $\mathcal{O}(d^{1/2})$, we recall (G.26):

$$\operatorname{Tr}\big([\nabla^2\phi(X)]^{-\frac{1}{2}}\mathrm{D}^3\phi(X)[H]\,[\nabla^2\phi(X)]^{-\frac{1}{2}}\big) = -(d+1)\operatorname{Tr}(X^{-1}H)\,.$$

Taking supremum on both sides,

$$\begin{aligned}&\sup_{H:\|H\|_X=1}\operatorname{Tr}\big([\nabla^2\phi(X)]^{-\frac{1}{2}}\mathrm{D}^3\phi(X)[H]\,[\nabla^2\phi(X)]^{-\frac{1}{2}}\big)\\ &= \sup_{H\in\mathbb{S}^d:\|X^{-1/2}HX^{-1/2}\|_F=1} -(d+1)\operatorname{Tr}(X^{-\frac{1}{2}}HX^{-\frac{1}{2}}) = \sup_{S\in\mathbb{S}^d:\|S\|_F=1}(d+1)\operatorname{Tr}(S)\,,\end{aligned}$$

and this objective achieves the maximum at $H = -d^{-1/2}X$, with the supremum being $(d+1)\sqrt{d}$. On the other hand, due to $\operatorname{Tr}(A)\le d^{1/2}\,\|A\|_F$ for $A\in\mathbb{R}^{d\times d}$,

$$\begin{aligned}&\operatorname{Tr}\big([\nabla^2\phi(X)]^{-\frac{1}{2}}\mathrm{D}^3\phi(X)[H]\,[\nabla^2\phi(X)]^{-\frac{1}{2}}\big)\\ &\le\sqrt{\frac{d(d+1)}{2}}\cdot\big\|[\nabla^2\phi(X)]^{-\frac{1}{2}}\mathrm{D}^3\phi(X)[H]\,[\nabla^2\phi(X)]^{-\frac{1}{2}}\big\|_F \le \sqrt{\frac{d(d+1)}{2}}\cdot\psi_X\|H\|_X\,,\end{aligned}$$

and thus by taking supremum on both sides over a symmetric matrix H with $\|H\|_X = 1$, it follows that $(d+1)\sqrt{d}\le\sqrt{\frac{d(d+1)}{2}}\,\psi_X$ and thus $\sqrt{2(d+1)}\le\psi_X$. ∎

G.4.7. PSD: STRONGLY LOWER TRACE SELF-CONCORDANCE

Direct computation leads to $\mathrm{D}^2 g(X)[H,H]\succeq 0$ (so SLTSC).

Proof of Lemma E.25. For $g(X) = -\nabla^2\log\det X$, recall that $g(X)[H,H] = \operatorname{Tr}(X^{-1}HX^{-1}H)$. Thus for any $V\in\mathbb{S}^d$,

$$\begin{aligned}\mathrm{D}g(X)[H,H,V] &= -\operatorname{Tr}(X^{-1}VX^{-1}\cdot HX^{-1}H) - \operatorname{Tr}(X^{-1}H\cdot X^{-1}VX^{-1}\cdot H)\\ &= -2\operatorname{Tr}(X^{-1}VX^{-1}HX^{-1}H)\,,\end{aligned}$$

and differentiating again,

$$\begin{aligned}&\mathrm{D}^2 g(X)[H,H,V,V]\\ &= 4\operatorname{Tr}(X^{-1}VX^{-1}VX^{-1}HX^{-1}H) + 2\operatorname{Tr}(X^{-1}VX^{-1}HX^{-1}VX^{-1}H)\\ &= 4\operatorname{Tr}(X^{-\frac{1}{2}}HX^{-1}VX^{-1}VX^{-1}HX^{-\frac{1}{2}}) + 2\operatorname{Tr}(X^{-\frac{1}{2}}VX^{-1}HX^{-\frac{1}{2}}\cdot X^{-\frac{1}{2}}VX^{-1}HX^{-\frac{1}{2}})\\ &\underset{\text{(i)}}{\ge} 4\operatorname{Tr}(X^{-\frac{1}{2}}HX^{-1}VX^{-1}VX^{-1}HX^{-\frac{1}{2}}) - 2\operatorname{Tr}(X^{-\frac{1}{2}}HX^{-1}VX^{-\frac{1}{2}}\cdot X^{-\frac{1}{2}}VX^{-1}HX^{-\frac{1}{2}})\\ &= 2\operatorname{Tr}(X^{-\frac{1}{2}}HX^{-1}VX^{-1}VX^{-1}HX^{-\frac{1}{2}})\ge 0\,,\end{aligned}\tag{G.28}$$

where in (i) we used the Cauchy-Schwarz inequality. Therefore, $\mathrm{D}^2 g(X)[H,H]\succeq 0$. ∎

G.4.8. PSD: AVERAGE SELF-CONCORDANCE

We establish a connection to the Gaussian orthogonal ensemble (GOE): for $d_s = d(d+1)/2$ and $\mathrm{svec}(H) \sim \mathcal{N}\big(0, \frac{r^2}{d_s}\, g(X)^{-1}\big)$, we have $\frac{\sqrt{d_s d}}{r} X^{-\frac{1}{2}} H X^{-\frac{1}{2}}$ is the GOE.

Proof of Lemma E.26. Let $h_X := \mathrm{svec}(X^{-1/2} H X^{-1/2})$ and $h := \mathrm{svec}(H)$. It holds that

$$h_X = L(X \otimes X)^{-\frac{1}{2}} M h$$

due to $h_X = \mathrm{svec}(X^{-\frac{1}{2}} H X^{-\frac{1}{2}}) = L\, \mathrm{vec}(X^{-\frac{1}{2}} H X^{-\frac{1}{2}}) = L(X \otimes X)^{-\frac{1}{2}}\, \mathrm{vec}(H) = L(X \otimes X)^{-\frac{1}{2}} M h$. As $h \sim \mathcal{N}\big(0, \frac{r^2}{d_s}\, g(X)^{-1}\big)$, h_X is a Gaussian with zero mean and covariance

$$\begin{aligned}
&\frac{r^2}{d_s} L(X \otimes X)^{-\frac{1}{2}} M g(X)^{-1} M^\mathsf{T} (X \otimes X)^{-\frac{1}{2}} L^\mathsf{T} \\
&\underset{\text{(i)}}{=} \frac{r^2}{d_s d} L(X \otimes X)^{-\frac{1}{2}} M L N (X \otimes X) N^\mathsf{T} L^\mathsf{T} M^\mathsf{T} (X \otimes X)^{-\frac{1}{2}} L^\mathsf{T} \\
&\underset{(*)}{=} \frac{r^2}{d_s d} L(X \otimes X)^{-\frac{1}{2}} N (X \otimes X) N^\mathsf{T} (X \otimes X)^{-\frac{1}{2}} L^\mathsf{T} \\
&\underset{(*)}{=} \frac{r^2}{d_s d} L(X \otimes X)^{-\frac{1}{2}} (X \otimes X) N (X \otimes X)^{-\frac{1}{2}} L^\mathsf{T} \underset{(*)}{=} \frac{r^2}{d_s d} L N L^\mathsf{T} \\
&\underset{\text{(ii)}}{=} \frac{r^2}{d_s d} \begin{bmatrix} I_d & \\ & \frac{1}{2} I_{d(d-1)/2} \end{bmatrix},
\end{aligned}$$

where (i) follows from Proposition E.20, $(*)$ follows from Lemma E.19, and (ii) follows from Magnus and Neudecker (1980, Page 427) that LNL^T is a $d_s \times d_s$ diagonal matrix with d times 1 and $\frac{1}{2} d(d-1)$ times $1/2$. Precisely, the entries of $h_X \in \mathbb{R}^{d_s}$ corresponding to the diagonals of $X^{-1/2} H X^{-1/2}$ are 1, and its entries corresponding to off-diagonals is $1/2$. This is exactly the covariance matrix of a d_s-dimensional GOE, so $X^{-\frac{1}{2}} H X^{-\frac{1}{2}} \sim \frac{r}{\sqrt{d_s d}} G$ for the GOE G. ■

Now we show ASC of $d\phi$.

Proof of Lemma E.27. Expand $\|Z - X\|_Z^2 := \|Z - X\|_{g(Z)}^2$ at X for $Z = X + H$:

$$\|Z - X\|_Z^2 - \|Z - X\|_X^2 = \sum_{k=1}^{\infty} \frac{1}{k!} \mathrm{D}^k g(X)[H^{\otimes k+2}]\,.$$

It follows from induction that for $H_X := X^{-\frac{1}{2}} H X^{-\frac{1}{2}}$

$$\begin{aligned}
\mathrm{D} g(X)[H^{\otimes 3}] &= -2d\ \mathrm{Tr}(X^{-1} H X^{-1} H X^{-1} H) = -2\,\mathrm{Tr}(H_X^3)\,, \\
\mathrm{D}^2 g(X)[H^{\otimes 4}] &= 3!\, d\ \mathrm{Tr}(H_X^4)\,, \\
\mathrm{D}^k g(X)[H^{\otimes (k+2)}] &= (-1)^k (k+1)!\, d\ \mathrm{Tr}(H_X^{k+2})\,.
\end{aligned}$$

Putting these back into the series expansion, for H the GOE (see Lemma E.26)

$$\|Z - X\|_Z^2 - \|Z - X\|_X^2 = \sum_{k=1}^{\infty} (-1)^k (k+1) d\ \mathrm{Tr}(H_X^{k+2})$$

$$=\sum_{k=1}^{\infty}(-1)^k(k+1)d\cdot\Big(\frac{r}{\sqrt{d_sd}}\Big)^{k+2}\operatorname{Tr}(H^{k+2})=\frac{r^2}{d_s}\sum_{k=1}^{\infty}(-1)^k(k+1)\Big(\frac{r}{\sqrt{d_sd}}\Big)^{k}\operatorname{Tr}(H^{k+2})\,.$$

As for ASC, it suffices to show that $\sum_{k=1}^{\infty}(-1)^k(k+1)\big(\frac{r}{\sqrt{d_sd}}\big)^k\operatorname{Tr}(H^{k+2})$ can be made arbitrarily small. We first control $\sum_{k\geq 2}$:

$$\Big|\sum_{k\geq 2}(-1)^k(k+1)\Big(\frac{r}{\sqrt{d_sd}}\Big)^{k}\operatorname{Tr}(H^{k+2})\Big|\leq\sum_{k\geq 2}(k+1)\Big(\frac{r}{\sqrt{d_sd}}\Big)^{k}d\cdot\|H\|_{\text{op}}^{k+2}\,.$$

By Vershynin (2018, Corollary 4.4.8), $\|H\|_{\text{op}}\lesssim\sqrt{d}$ holds with high probability, and thus

$$\sum_{k\geq 2}(k+1)\Big(\frac{r}{\sqrt{d_sd}}\Big)^{k}d\cdot\|H\|_{\text{op}}^{k+2}\leq\sum_{k\geq 2}(k+1)r^k\frac{1}{d^{3k/2}}d\cdot d^{\frac{k+2}{2}}\leq\sum_{k\geq 2}(k+1)r^kd^{2-k}\,.$$

By taking $r=\Omega(1)$ small enough, we can make this series arbitrarily small.

Now we bound $\frac{r}{d^{3/2}}\operatorname{Tr}(H^3)$ ($k=1$ case). This is a Gaussian polynomial in $\operatorname{svec}(H)$, so it suffices to show $\mathbb{E}[(\operatorname{Tr}(H^3))^2]=\mathcal{O}(d^3)$; we then use Lemma E.8 to obtain a high-probability bound on the Gaussian polynomial $\frac{r}{d^{3/2}}\operatorname{Tr}(H^3)$. For $H=(H_{ab})\in\mathbb{S}^d$,

$$\big(\operatorname{Tr}(H^3)\big)^2=\sum_{ipq}H_{ip}H_{pq}H_{qi}\cdot\sum_{jrs}H_{jr}H_{rs}H_{sj}=\sum_{ipqjrs}H_{ip}H_{pq}H_{qi}H_{jr}H_{rs}H_{sj}\,,$$

where each H_{**} in the summand is an independent Gaussian with zero mean and variance 1 or $1/2$ (as H is the GOE). We can classify the indices $\{i,p,q,j,r,s\}$ into the following types:

6 distinct indices $\{a,b,c,d,e,f\}$,
5 distinct indices $\{a,b,c,d,(e,e)\}$,
4 distinct indices $\{a,b,c,(d,d,d)\},\{a,b,(c,c),(d,d)\}$,
Others $\ldots$,

where for example $\{a,b,c,d,e,f\}$ means all indices are different, and $\{a,b,c,d,(e,e)\}$ means that there appear 5 different indices $\{a,b,c,d,e\}$ but exists one pair (e,e) of the same index. Note that $\mathbb{E}H_{ip}H_{pq}H_{qi}H_{jr}H_{rs}H_{sj}=\mathcal{O}(1)$ is at most the sixth moment of a standard Gaussian. It implies that toward our goal of showing $\mathcal{O}(d^3)$-bound on $\big(\operatorname{Tr}(H^3)\big)^2$, it suffices to look into only three types of indices above. This is because the terms from other types contribute at most $\mathcal{O}(d^3)$ to $\big(\operatorname{Tr}(H^3)\big)^2$.

For any term with 6 distinct indices, we can always find an 'uncoupled' H_{**} (for example H_{ab}) in the summand that is independent of all the others, so its expectation of the summand is 0.

For the terms with 5-distinct indices $\{a,b,c,d,(e,e)\}$, due to symmetry (see Figure G.1) we can further classify the index (i,p,q,j,r,s) into either (a,b,c,d,e,e) or (a,b,e,c,d,e). In both cases , H_{ab} has no coupled Gaussian, so the expectations of the summand are also 0.

For 4-distinct indices, let us first consider $\{a,b,c,(d,d,d)\}$-type indices. In this case (i,p,q,j,r,s) is of the form either (a,a,a,b,c,d) or (a,a,b,a,c,d) due to symmetry. In both cases, H_{cd} has no coupled Gaussian. Now consider $\{a,b,(c,c),(d,d)\}$-type indices. Then (i,p,q,j,r,s) is of the form either (a,b,c,c,d,d) or (a,c,c,b,d,d) or (a,c,d,b,c,d). For each case, H_{ab},H_{cc},H_{ac} are uncoupled ones. Therefore, $\mathbb{E}[H_{ip}H_{pq}H_{qi}H_{jr}H_{rs}H_{sj}]=0$ whenever there are at least 4 distinct indices. ■

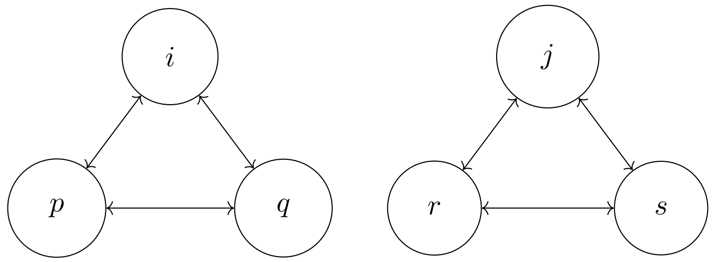


Figure G.1: A structure of indices of $H_{ip}H_{pq}H_{qi} \cdot H_{jr}H_{rs}H_{sj}$

Remark G.7 *It seems challenging to show that $d\phi$ is SASC using the same technique. When g is*

$$g = d\,\nabla^2(-\log\det X) + g'$$

for other PSD matrix function g', we know that $\operatorname{svec}(H_X) = \operatorname{svec}(X^{-\frac{1}{2}}HX^{-\frac{1}{2}})$ follows a Gaussian distribution with zero mean and covariance matrix M satisfying

$$M \preceq \begin{bmatrix} I_d & \\ & \frac{1}{2}I_{d(d-1)/2} \end{bmatrix}.$$

A main difference in the SASC setting is that the entries of $h = \operatorname{svec}(H_X)$ might exhibit dependencies, making the previous approach infeasible. This arises because many fundamental results in the random matrix theory often presume independence of the entries of a random matrix. Moreover, our combinatorial argument for the $k = 1$ case is not feasible in the presence of such dependencies.

G.5. Examples (§F)

G.5.1. ALGORITHMS FOR PSD SAMPLING

Proof of Proposition F.2. We define $g_X = g = 2(d^2g_1 + g_2)$, where

$$g_1(X) = M^{\mathsf{T}}(X \otimes X)^{-1}M \qquad \text{and} \qquad g_2(X) = 22\sqrt{\frac{m}{d}}\,M^{\mathsf{T}}A_X^{\mathsf{T}}\big(\Sigma_X + \frac{d}{m}I_m\big)A_XM\,.$$

Since d^2g_1 and g_2 are SSC, g is also SSC due to Lemma D.5 and $\mathcal{O}(d^3 + \sqrt{md^2})$-symmetric[8] due to Lemma D.11. As d^2g_1 and g_2 is SLTSC and SASC, g is LTSC and ASC. Putting these together, it follows that g is $\big(\mathcal{O}(d^3 + \sqrt{md^2}), \mathcal{O}(d^3 + \sqrt{md^2})\big)$-Dikin-amenable. Therefore, Theorem 3.2 implies that GCDW incurs $\widetilde{\mathcal{O}}(d^2(d^3 + \sqrt{md^2})) = \widetilde{\mathcal{O}}(d^3(d^2 + \sqrt{m}))$ total iterations of the Dikin walk with g.

Now we bound the per-step complexity of the Dikin walk (Algorithm 1). Recall that it requires (1) the update of the leverage scores, (2) computation of the matrix function induced by the local metric g, (3) the inverse of the matrix function and (4) its determinant. By Lee and Sidford (2019, Theorem 46) (with $p = 2$ and $d \leftarrow d_s$ therein), the initialization of the leverage scores at the beginning takes $\widetilde{\mathcal{O}}(md^{2\omega})$ and their updates takes $\widetilde{\mathcal{O}}(md^{2(\omega-1)})$ time. Since (1) takes $\widetilde{\mathcal{O}}(md^{2(\omega-1)})$, (2) takes $\widetilde{\mathcal{O}}(d^4 + md^{2(\omega-1)})$, and (3) and (4) take $\mathcal{O}\left(d^{2\omega}\right)$, each iteration runs in $\widetilde{\mathcal{O}}(d^{2\omega} + md^{2(\omega-1)})$

8. Since the dimension is d_s in the PSD setting, we should replace d by $d_s = \mathcal{O}(d^2)$ when applying Lemma E.5.

time. Even though the initialization of leverage scores takes $\widetilde{\mathcal{O}}(md^{2\omega})$ time, the amortized per-step time complexity becomes $\widetilde{\mathcal{O}}(d^{2\omega} + md^{2(\omega-1)}) = \widetilde{\mathcal{O}}(md^{2(\omega-1)})$ time, as the mixing rate is $\widetilde{\mathcal{O}}(d^3(d^2 + \sqrt{m}))$. ■

Proof of Proposition F.3. We define $g_X = g = 2(d^2 g_1 + g_2)$, where for some constants $c_1, c_2 > 0$,

$$g_1(X) = M^\top (X \otimes X)^{-1} M \qquad \text{and} \qquad g_2(X) = dc_1 (\log m)^{c_2} M^\top A_X^\top W_X A_X M\,.$$

Since $d^2 g_1$ and g_2 are SSC, g is also SSC due to Lemma D.5 and $\mathcal{O}^*(d^3)$-symmetric due to Lemma D.11. As $d^2 g_1$ and g_2 is SLTSC and SASC, g is LTSC and ASC. Putting these together, it follows that g is $\big(\mathcal{O}^*(d^3), \mathcal{O}^*(d^3)\big)$-Dikin-amenable. Therefore, Theorem 3.2 implies that GCDW requires $\widetilde{\mathcal{O}}(d^5)$ iterations of the Dikin walk with g. Since the initialization and update of the Lewis weight takes $\widetilde{\mathcal{O}}(md^{2\omega})$ and $\widetilde{\mathcal{O}}(md^{2(\omega-1)})$ time (Lee and Sidford, 2019, Theorem 46), the same implementation with Theorem F.2 also has the time complexity of $\widetilde{\mathcal{O}}(md^{2(\omega-1)})$. ■

G.5.2. EFFICIENT IMPLEMENTATION

Proof of Proposition F.4. Let $v \in \mathbb{R}^{d_s}$ be a given vector, and denote $\bar{g}_0 := g_1$ and $\bar{g}_i := \bar{g}_{i-1} + u_i u_i^\top$ for $i \in [m]$. We first prepare the column vectors u_i's of $U = M^\top A^\top S_X^{-1}$ in $\mathcal{O}(md^2)$ time and then initialize $\bar{g}_0^{-1} v$ and $\bar{g}_0^{-1} u_i$ for $i \in [m]$ in $\mathcal{O}(md^\omega)$ time. For u_i's, note that S_X can be prepared in $\mathcal{O}(md^2)$ time, and thus $A^\top S_X^{-1}$ takes $\mathcal{O}(md^2)$ time due to $A \in \mathbb{R}^{d^2 \times m}$. Since each row of $M^\top \in \mathbb{R}^{d_s \times d^2}$ has at most two non-zero entries, we can obtain u_i's in $\mathcal{O}(md^2)$ time.

For $\bar{g}_0^{-1} v$ and $\bar{g}_0^{-1} u_i$, we recall from Lemma E.20 that for a vector $z \in \mathbb{R}^{d_s}$

$$g_1^{-1} z = M^\dagger (X \otimes X)(M^\dagger)^\top z = LN(X \otimes X) N L^\top z\,.$$

Since each row of $L^\top \in \mathbb{R}^{d^2 \times d_s}$ has at most two non-zero entries, $w := L^\top z \in \mathbb{R}^{d^2}$ can be computed in $\mathcal{O}(d^2)$ time. From the definition of N, it follows that $Nw = \operatorname{vec}\big(\frac{1}{2}(W + W^\top)\big)$ for $W := \operatorname{vec}^{-1}(w) \in \mathbb{R}^{d \times d}$, which also can be computed in $\mathcal{O}(d^2)$ time. For $\overline{W} := \frac{1}{2}(W + W^\top)$, it follows that

$$(X \otimes X) N w = (X \otimes X) \operatorname{vec}(\overline{W}) \underset{\text{Lemma H.1-1}}{=} \operatorname{vec}(X \overline{W} X)\,,$$

which can be computed in $\mathcal{O}(d^\omega)$ time by the fast matrix multiplication, and in a similar way we can compute $LN \operatorname{vec}(X\overline{W}X)$ in $\mathcal{O}(d^2)$ time. Putting all these together, $\bar{g}_0^{-1} v$ can be computed in $\mathcal{O}(d^\omega)$ time, and repeating this for u_j's yields $\{\bar{g}_0^{-1} v, \bar{g}_0^{-1} u_1, \dots, \bar{g}_0^{-1} u_m\}$ in $\mathcal{O}(md^\omega)$ time.

Starting with these initializations, we recursively use the Sherman–Morrison formula: for $z \in \mathbb{R}^{d_s}$,

$$\bar{g}_i^{-1} z = \bar{g}_{i-1}^{-1} z - \frac{\bar{g}_{i-1}^{-1} u_i u_i^\top \bar{g}_{i-1}^{-1} z}{1 + u_i^\top \bar{g}_{i-1}^{-1} u_i}\,. \tag{G.29}$$

Using $\bar{g}_{i-1}^{-1} u_j$ and $\bar{g}_{i-1}^{-1} v$ from a previous iteration, we can compute each of $\bar{g}_i^{-1} u_j$ and $\bar{g}_i^{-1} v$ in the current iteration in $\mathcal{O}(d^2)$ time, and thus each round for update takes $\mathcal{O}(md^2)$ time in total. Since we iterate for m rounds, Algorithm 4 outputs $\bar{g}_m^{-1} v = g(X)^{-1} v$ in $\mathcal{O}(md^\omega + m^2 d^2)$ time. ■

Proof of Lemma F.5. Here we provide details of Algorithm 5 in two stages – (1) sampling from $\mathcal{N}\big(0, \frac{r^2}{d} g(x)^{-1}\big)$ and (2) computation of acceptance probability.

(1) Gaussian sampling: For simplicity, we ignore r^2/d and illustrate how to draw $v \sim \mathcal{N}(0, g(X)^{-1})$ without full computation of $g(X)^{-1}$ in $\mathcal{O}(md^\omega + m^2d^2)$ time.

Our approach is to compute $v := g(X)^{-1} \begin{bmatrix} B & U \end{bmatrix} w$ for $w \sim \mathcal{N}(0, I_{d^2+m})$, which follows the Gaussian distribution with covariance

$$g(X)^{-1} \begin{bmatrix} B & U \end{bmatrix} \Big(g(X)^{-1} \begin{bmatrix} B & U \end{bmatrix}\Big)^\mathsf{T} = g(X)^{-1}(BB^\mathsf{T} + CC^\mathsf{T})g(X)^{-1}g(X)^{-1} ,$$

since v is a linear transformation of the Gaussian random variable w, and $BB^\mathsf{T} + CC^\mathsf{T} = g(X)$.

Denoting $w = (w_b, w_u)$ for $w_b \sim \mathcal{N}(0, I_{d^2})$ and $w_u \sim \mathcal{N}(0, I_m)$, we can show that $\begin{bmatrix} B & U \end{bmatrix} w$ can be computed in $\mathcal{O}(d^\omega + md^2)$ time as follows:

$$\begin{aligned} \begin{bmatrix} B & U \end{bmatrix} w = Bw_b + Uw_c &= M^\mathsf{T} \underbrace{(X \otimes X)^{-1/2} w_b}_{\text{Use Lemma G.29}} + M^\mathsf{T} A^\mathsf{T} S_X^{-1} w_c \\ &= M^\mathsf{T} \Big(\operatorname{vec}\big(X^{-1/2} \operatorname{vec}^{-1}(w_b)\, X^{-1/2}\big) + A^\mathsf{T} S_X^{-1} w_c\Big) , \end{aligned}$$

where $\operatorname{vec}\big(X^{-1/2} \operatorname{vec}^{-1}(w_b)\, X^{-1/2}\big)$ and $A^\mathsf{T} S_X^{-1} w_u$ can be computed in $\mathcal{O}(d^\omega)$ and $\mathcal{O}(md^2)$ time, respectively. Since each row of $M^\mathsf{T} \in \mathbb{R}^{d_s \times d^2}$ has at most two non-zero entries, $\begin{bmatrix} B & U \end{bmatrix} w$ can be computed in $\mathcal{O}(d^\omega + md^2)$ time. Using Algorithm 4, we obtain $v = g(X)^{-1} \begin{bmatrix} B & U \end{bmatrix} w$ in $\mathcal{O}(md^\omega + m^2d^2)$ time.

(2) Computation of acceptance probability. We show that this step also takes $\mathcal{O}(md^\omega + m^2d^2)$ time. To compute $\det g(X)$, we use Algorithm 4 to prepare $\{\bar{g}_i^{-1} u_1, \dots, \bar{g}_i^{-1} u_m\}_{i=0}^m$ at X and $Y = \operatorname{svec}^{-1}(y)$ in $\mathcal{O}(md^\omega + m^2d^2)$ time. Recall the matrix determinant lemma:

$$\det(A + uu^\mathsf{T}) = (1 + u^\mathsf{T} A^{-1} u)\, \det A .$$

Using the following recursive formula

$$\det(\bar{g}_{i+1}) = \det(\bar{g}_i + u_{i+1} u_{i+1}^\mathsf{T}) = (1 + u_{i+1}^\mathsf{T} \bar{g}_i^{-1} u_{i+1})\, \det \bar{g}_i ,$$

we start with $\det \bar{g}_0 = \det g_1 = 2^{d(d-1)/2} (\det X)^{-(d+1)}$ (see Lemma H.1-7), which can be computed in $\mathcal{O}(d^\omega)$ time, and compute $\det g(X)$ (and $\det g(Y)$ in the same way) in $\mathcal{O}(md^\omega + m^2d^2)$ time. ■

G.5.3. HANDLING APPROXIMATE LEWIS WEIGHTS

Proof of Lemma F.6. We just reproduce the proof of Lemma B.3. For $\pi \propto \exp(-f) \cdot \mathbf{1}_K$, we denote

$$p_x = \mathcal{N}\Big(x, \frac{r^2}{d} g(x)^{-1}\Big), \qquad R_x(z) = \frac{p_z(x)}{p_x(z)} \frac{\pi(z)}{\pi(x)}, \qquad A_x(z) = \min\big(1, R_x(z)\, \mathbf{1}_K(z)\big) .$$

Then the transition kernel of the Dikin walk started at x can be written as

$$\widetilde{P}(x, dz) = \underbrace{(1 - \mathbb{E}_{p_x}[A_x(\cdot)])}_{=: r_x}\, \delta_x(\mathrm{d}z) + A_x(z)\, p_x(z)\, \mathrm{d}z .$$

Thus, for $x, y \in \text{int}(K)$

$$d_{\text{TV}}(P_x, P_y) = \underbrace{\frac{r_x + r_y}{2}}_{\text{I}} + \underbrace{\frac{1}{2}\int |A_x(z)\, p_x(z) - A_y(z)\, p_y(z)|\, \mathrm{d}z}_{\text{II}} \,.$$

We note that $(1-\delta)\, \widetilde{g}_2 \preceq g_2 \preceq (1+\delta)\, \widetilde{g}_2$ and thus

$$(1-\delta)\, \widetilde{g} \preceq g \preceq (1+\delta)\, \widetilde{g}\,, \tag{G.30}$$

and this implies $(1-\delta)\, I \preceq \widetilde{g}^{-1/2} g \widetilde{g}^{-1/2} \preceq (1+\delta)\, I$. Hence, $(1-\delta)^{d^2/2} \leq \sqrt{\frac{\det g}{\det \widetilde{g}}} \leq (1+\delta)^{d^2/2}$ and

$$(1-\delta)^{d^2} \sqrt{\frac{\det \widetilde{g}(z)}{\det \widetilde{g}(x)}} \leq \sqrt{\frac{\det g(z)}{\det g(x)}} \leq (1+\delta)^{d^2} \sqrt{\frac{\det \widetilde{g}(z)}{\det \widetilde{g}(x)}}\,. \tag{G.31}$$

With this in mind, recall that

$$r_x = 1 - \mathbb{E}_{p_x}[A_x(\cdot)] = 1 - \int \min\Bigl(1,\, \underbrace{\mathbf{1}_K(z) \frac{\exp(-f(z))}{\exp(-f(x))}}_{=:\mathsf{A}} \underbrace{\frac{p_z(x)}{p_x(z)}}_{=:\mathsf{B}}\Bigr)\, p_x(z)\, \mathrm{d}z.$$

We can bound A in a similar way by using (G.30). As for B,

$$\log \mathsf{B} = -\frac{d}{2r^2}(\|z-x\|_z^2 - \|z-x\|_x^2) + \frac{1}{2}(\log\det \widetilde{g}(z) - \log\det \widetilde{g}(x))\,.$$

As in Lemma B.3, the second term can be bounded lower by $\exp(-3\varepsilon)$ using (G.31). The first term can be lower-bounded by invoking ASC of g. To see this, ignoring the normalization constant of g_x

$$\begin{aligned}
(*) &= \int \mathbf{1}\Bigl(\|z-x\|_{\widetilde{g}(z)}^2 - \|z-x\|_{\widetilde{g}(x)}^2 \leq 2\varepsilon \frac{r^2}{d}\Bigr) \sqrt{|\widetilde{g}(x)|}\exp\bigl(-\frac{1}{2}\|z-x\|_{\widetilde{g}(x)}^2\bigr)\, \mathrm{d}z \\
&= \int \mathbf{1}\Bigl(\|z-x\|_{\widetilde{g}(z)}^2 - \|z-x\|_{\widetilde{g}(x)}^2 \leq 2\varepsilon \frac{r^2}{d}\Bigr) \sqrt{|g(x)|}\exp\bigl(-\frac{1}{2}\|z-x\|_{g(x)}^2\bigr) \\
&\qquad \cdot \sqrt{\left|\frac{\widetilde{g}(x)}{g(x)}\right|}\exp\bigl(-\frac{1}{2}(\|z-x\|_{\widetilde{g}(x)}^2 - \|z-x\|_{g(x)}^2)\bigr)\, \mathrm{d}z \\
&\leq \int \mathbf{1}\Bigl(\|z-x\|_{\widetilde{g}(z)}^2 - \|z-x\|_{\widetilde{g}(x)}^2 \leq 2\varepsilon \frac{r^2}{d}\Bigr) \sqrt{|g(x)|}\exp\bigl(-\frac{1}{2}\|z-x\|_{g(x)}^2\bigr) \\
&\qquad \cdot (1+\delta)^{d^2/2}\exp\bigl(\frac{\delta}{2}\|z-x\|_{g(x)}^2\bigr)\, \mathrm{d}z\,.
\end{aligned}$$

Due to $\|z-x\|_{g(x)}^2 \lesssim r^2$ w.h.p., taking $\delta = \varepsilon/d^{10}$ leads to

$$(*) \leq 2\int \mathbf{1}\Bigl(\|z-x\|_{\widetilde{g}(z)}^2 - \|z-x\|_{\widetilde{g}(x)}^2 \leq 2\varepsilon \frac{r^2}{d}\Bigr) \sqrt{|g(x)|}\exp\bigl(-\frac{1}{2}\|z-x\|_{g(x)}^2\bigr)\, \mathrm{d}z.$$

Also, due to

$$\|z-x\|_{\widetilde{g}(z)}^2 - \|z-x\|_{\widetilde{g}(x)}^2 \geq (1-\delta)\, \|z-x\|_{g(z)}^2 - (1+\delta)\, \|z-x\|_{g(x)}^2$$

$$= (1-\delta)\left(\|z-x\|_{g(z)}^2 - \|z-x\|_{g(x)}^2\right) - 2\delta\,\|z-x\|_{g(x)}^2\,,$$

we have

$$(*) \le 2\int \mathbf{1}\Big(\|z-x\|_{g(z)}^2 - \|z-x\|_{g(x)}^2 \le (2\varepsilon(1-\delta)^{-1}+\varepsilon)\,\frac{r^2}{d}\Big)\sqrt{|g(x)|}e^{-\frac{1}{2}\|z-x\|_{g(x)}^2}\,\mathrm{d}z \le 6\varepsilon$$

by invoking ASC of g in the last inequality. Putting these together, $\mathsf{I} \le \frac{1}{2}+\mathcal{O}(\varepsilon)$. For II, we can follow the proof of Lemma B.3 to show $\mathsf{II} \le \frac{1}{4}+\mathcal{O}(\varepsilon)$, and every technical issue can be resolved by repeating the same techniques above. ■

Appendix H. Backgrounds on matrix algebra

H.1. Matrix identities

We collect algebraic identities related to trace, vectorization, Kronecker and Hadamard product.

Lemma H.1 (Kronecker product) *For $A,B,C,D \in \mathbb{R}^{d\times d}$ and M in Definition E.18,*

1. $(A\otimes B)\,\mathrm{vec}(C) = \mathrm{Tr}(BCA^{\mathsf{T}})$.
2. $\mathrm{vec}(A)^{\mathsf{T}}(B\otimes C)\,\mathrm{vec}(D) = \mathrm{Tr}(DB^{\mathsf{T}}A^{\mathsf{T}}C)$.
3. $(A\otimes B)(C\otimes D) = AC\otimes BD$.
4. $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$.
5. $(A\otimes B)^{\mathsf{T}} = A^{\mathsf{T}}\otimes B^{\mathsf{T}}$.
6. $\mathrm{Tr}(A\otimes B) = \mathrm{Tr}(A)\,\mathrm{Tr}(B)$.
7. $\det\big(M^{\mathsf{T}}(A\otimes A)M\big) = 2^{d(d-1)/2}(\det A)^{d+1}$.

Lemma H.2 (Hadamard product) *Let $A,B,C,D \in \mathbb{R}^{d\times d}$, $x,y\in\mathbb{R}^d$, and $D_1,D_2\in\mathbb{R}^{d\times d}$ be diagonal matrices.*

1. $(A\circ B)y = \mathrm{diag}(A\;\mathrm{Diag}(y)B^{\mathsf{T}})$.
2. $x^{\mathsf{T}}(A\circ B)y = \mathrm{Tr}(\mathrm{Diag}(x)A\;\mathrm{Diag}(y)B^{\mathsf{T}})$.
3. $D_1(A\circ B) = (D_1A)\circ B = A\circ(D_1B)$.
4. $(A\circ B)D_2 = (AD_2)\circ B = A\circ(BD_2)$.
5. $(A\otimes B)\circ(C\otimes D) = (A\circ C)\otimes(B\circ D)$.

H.2. Matrix calculus

Let $g(x):\mathbb{R}^d\to\mathbb{R}^{d\times d}$ be a matrix function. Its gradient at x, denoted by $\mathrm{D}g(x)$, is the third-order tensor defined by $(\mathrm{D}g(x))_{ijk} = \frac{\partial g_{ij}(x)}{\partial x_k}$. Unless specified otherwise, the multiplication between higher-order tensors and a matrix of size $d\times d$ is running over (i,j)-entries. For instance, for a matrix $M\in\mathbb{R}^{d\times d}$ the product $\mathrm{D}g(x)\cdot M$ indicates the third-order tensor defined by

$$(\mathrm{D}g(x)\,M)_{\cdot,\cdot,k} = (\mathrm{D}g(x))_{\cdot,\cdot,k}M \text{ for each } k\in[d]\,.$$

In the same way, the trace is applied to a matrix spanned by (i,j)-entries, i.e.,

$$\big(\mathrm{Tr}(\mathrm{D}g(x))\big)_k = \mathrm{Tr}\Big(\big(\mathrm{D}g(x)\big)_{\cdot,\cdot,k}\Big)\,.$$

For $\varphi : \mathbb{R}^d \to \mathbb{R}$ with $\varphi(\cdot) := \log\det g(\cdot)$, its gradient and the directional derivative in $h \in \mathbb{R}^d$ are

$$\nabla\varphi(x) = \operatorname{Tr}\big(g(x)^{-1}\mathrm{D}g(x)\big)\,, \qquad \text{and} \qquad \nabla\varphi(x)\cdot h = \operatorname{Tr}\big(g(x)^{-1}\mathrm{D}g(x)[h]\big)\,. \tag{H.1}$$

For the Hessian of φ, using the product rule and

$$\mathrm{D}(g^{-1})(x) = -g(x)^{-1}\mathrm{D}g(x)\,g(x)^{-1}\,, \tag{H.2}$$

we obtain

$$\begin{aligned}\nabla^2\varphi(x) &= \mathrm{D}\operatorname{Tr}\big(g(x)^{-1}\mathrm{D}g(x)\big) = -\operatorname{Tr}\big(g(x)^{-1}\mathrm{D}g(x)\,g(x)^{-1}\mathrm{D}g(x)\big) + \operatorname{Tr}\big(g(x)^{-1}\mathrm{D}^2g(x)\big)\\ &= \operatorname{Tr}\big(g(x)^{-1}\mathrm{D}^2g(x)\big) - \|g(x)^{-\frac12}\mathrm{D}g(x)\,g(x)^{-\frac12}\|_F^2\,,\end{aligned} \tag{H.3}$$

where $\mathrm{D}^2g(x)$ is the fourth-order tensor defined by $(\mathrm{D}^2g(x))_{ijkl} = \frac{\partial(g(x))_{ij}}{\partial x_k \partial x_l}$.

We now present formulas for the Hessian and its inverse of $\phi(\cdot) = -\log\det(\cdot)$ on $\mathbb{S}^d_{++}$.

Proof of Proof of Proposition E.20. By setting $g(X) = X$ and $\phi(X) = -\varphi(X)$ above, (H.3) implies that for a symmetric matrix $H \in \mathbb{S}^d$

$$\begin{aligned}\nabla^2\phi(X)[H,H] &= \|X^{-\frac12}HX^{-\frac12}\|_F^2 = \operatorname{Tr}\big(X^{-1}HX^{-1}H\big)\\ &= \operatorname{vec}(H)^\mathsf{T}(X^{-1}\otimes X^{-1})\operatorname{vec}(H) = \operatorname{vec}(H)^\mathsf{T}(X\otimes X)^{-1}\operatorname{vec}(H)\,,\end{aligned} \tag{H.4}$$

where the last equality follows from Lemma H.1. When representing X and H in $\mathbb{R}^{d_s}$ space with notations $x := \operatorname{svec}(X)$ and $h := \operatorname{svec}(H)$, the definition of M (see Definition E.18) turns (H.4) into

$$\nabla^2\phi(x)[h,h] = h^\mathsf{T}M^\mathsf{T}(X\otimes X)^{-1}Mh\,,$$

so $g_X := \nabla_x^2\phi(x) = \nabla_X^2\phi(X)$ equals $M^\mathsf{T}(X\otimes X)^{-1}M$. The formula for the inverse, $g_X^{-1} = M^\dagger(X\otimes X)(M^\dagger)^\mathsf{T}$, is immediate from Magnus and Neudecker (1980), and another part follows from $M^\dagger = LN$ and $N^\mathsf{T} = N$ (Magnus and Neudecker, 1980, Lemma 3.6 and Lemma 2.1). ■

Appendix I. Self-concordant barriers for linear constraints

We collect details on self-concordant barriers for linear constraints, $P = \{x \in \mathbb{R}^d : Ax \geq b\}$ with $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$: the logarithmic, volumetric, and Lewis-weight barrier/metric. Recall the notations used in the paper: $s_x = \operatorname{diag}(Ax - b) \in \mathbb{R}^m$, $S_x = \operatorname{Diag}(s_x) \in \mathbb{R}^{m\times m}$, and $A_x = S_x^{-1}A \in \mathbb{R}^{m\times d}$. Also, $s_{x,h} = A_x h \in \mathbb{R}^m$ and $S_{x,h} = \operatorname{Diag}(s_{x,h}) \in \mathbb{R}^{m\times m}$. Let $h \in \mathbb{R}^d$.

I.1. Logarithmic barriers

For $x \in P$, the logarithmic barrier (or log-barrier) and the Hessian metric are given by

$$\phi_{\log}(x) := -\sum_{i=1}^m \log(a_i^\mathsf{T}x - b)\,, \qquad \text{and} \qquad g(x) = \nabla^2\phi(x) = A_x^T A_x\,.$$

Claim I.1 $\mathrm{D}S_x[h] = \operatorname{Diag}(Ah)$ *and* $\mathrm{D}S_x^{-1}[h] = -S_x^{-1}S_{x,h}$. *Also,* $\mathrm{D}g(x)[h] = -2A_x^\mathsf{T}S_{x,h}A_x$ *and* $\mathrm{D}^2g(x)[h,h] = 6A_x^\mathsf{T}S_{x,h}^2A_x \succeq 0$.

Proof The first is obvious from differentiation of $S_x = \mathrm{Diag}(Ax-b)$ w.r.t. x. As for the second,

$$\mathrm{D}S_x^{-1}[h] = -S_x^{-1}\mathrm{D}S_x[h]\,S_x^{-1} = -S_x^{-1}\,\mathrm{Diag}(Ah)S_x^{-1} = -S_x^{-1}\,\mathrm{Diag}(A_xh) = -S_x^{-1}S_{x,h}\,.$$

As for the third and fourth, as $g(x) = A^\mathsf{T} S_x^{-2}A$,

$$\mathrm{D}g(x)[h] = A^\mathsf{T}\mathrm{D}S_x^{-2}[h]\,A = -2A^\mathsf{T}S_x^{-3}\mathrm{D}S_x[h]A = -2A_x^\mathsf{T}S_x^{-1}\,\mathrm{Diag}(Ah)A_x = -2A_x^\mathsf{T}S_{x,h}A_x\,.$$
$$\mathrm{D}^2g(x)[h,h] = -2A^\mathsf{T}\mathrm{D}S_x^{-3}[h]\;\mathrm{Diag}(Ah)A = 6A^\mathsf{T}S_x^{-4}\mathrm{D}S_x[h]\;\mathrm{Diag}(Ah)A = 6A_x^\mathsf{T}S_{x,h}^2A_x\,.$$

■

I.2. Volumetric barriers

Vaidya (1996) introduced the *volumetric barrier* for P, defined by

$$\phi_{\text{vol}}(x) = \frac{1}{2}\log\det\big(\nabla^2\phi_{\log}(x)\big) = \frac{1}{2}\log\det(A_x^\mathsf{T}A_x)\,.$$

Claim I.2 $\nabla\phi_{\text{vol}}(x) = -A_x^\mathsf{T}\sigma_x$ *and* $\nabla^2\phi_{\text{vol}}(x) = A_x^\mathsf{T}(3\Sigma_x - 2P_x^{(2)})A_x$.

Proof For $P_x := P(A_x)$, using (H.1) with Claim I.1 and apply Lemma H.2 in (i),

$$\nabla\phi_{\text{vol}}(x)[h] = -\operatorname{Tr}\big((A_x^\mathsf{T}A_x)^{-1}A_x^\mathsf{T}S_{x,h}A_x\big) = -\operatorname{Tr}(P_xS_{x,h}) \underset{\text{(i)}}{=} -1^\mathsf{T}(P_x\circ I_m)s_{x,h} = -h^\mathsf{T}A_x^\mathsf{T}\sigma_x\,,$$

For the Hessian of ϕ_{vol}, let $g(x) = A_x^\mathsf{T}A_x$ and then by (H.3),

$$\nabla^2\phi_{\text{vol}}(x)[h,h] = \frac{1}{2}\left(\operatorname{Tr}(g^{-1}\mathrm{D}^2g[h,h]) - \operatorname{Tr}(g^{-1}\mathrm{D}g[h]\,g^{-1}\mathrm{D}g[h])\right).$$

As for the first term, Claim I.1 leads to

$$\operatorname{Tr}(g^{-1}\mathrm{D}^2g[h,h]) = 6\operatorname{Tr}(g^{-1}A_x^\mathsf{T}S_{x,h}^2A_x) = 6\operatorname{Tr}(P_xS_{x,h}IS_{x,h}) = 6h^\mathsf{T}A_x^\mathsf{T}(P_x\circ I)A_xh = 6h^\mathsf{T}A_x^\mathsf{T}\Sigma_xA_xh\,.$$

As for the second term,

$$\operatorname{Tr}(g^{-1}\mathrm{D}g[h]\,g^{-1}\mathrm{D}g[h]) = 4\operatorname{Tr}(P_xS_{x,h}P_xS_{x,h}) = 4(A_xh)^\mathsf{T}(P_x\circ P_x)(A_xh) = 4h^\mathsf{T}A_x^\mathsf{T}P_x^{(2)}A_xh\,.$$

Hence, $\mathrm{D}^2\phi_{\text{vol}}(x)[h,h] = h^\mathsf{T}A_x^\mathsf{T}(3\Sigma_x - 2P_x^{(2)})A_xh$, which completes the proof. ■

Claim I.3 $P_x^{(2)} \preceq \Sigma_x$, *so* $A_x^\mathsf{T}\Sigma_xA_x \preceq \nabla^2\phi_{\text{vol}}(x) \preceq 3A_x^\mathsf{T}\Sigma_xA_x$.

Proof Due to $\Sigma_x = P_x\circ I$, it suffices to show $h^\mathsf{T}P_x\circ(I-P_x)\,h \geq 0$ for any $h\in\mathbb{R}^d$. Since P_x and $I-P_x$ are orthogonal projections, for $H = \mathrm{Diag}(h)$ and $C := P_xH(I-P_x)$,

$$h^\mathsf{T}P_x\circ(I-P_x)\,h = \operatorname{Tr}\big(HP_xH(I-P_x)\big) = \operatorname{Tr}\big((I-P_x)HP_xP_xH(I-P_x)\big) = \operatorname{Tr}(C^\mathsf{T}C) \geq 0\,.$$

■

I.2.1. DERIVATIVES OF LEVERAGE SCORES AND PROJECTION MATRICES

We derive formulas for derivatives of leverage scores, orthogonal projections, and so on.

Lemma I.4 *For $x,h\in\mathbb{R}^d$, let $P_x=A_x(A_x^\top A_x)^{-1}A_x^\top$, $\Sigma_x=\operatorname{Diag}(P_x)$, and $\Lambda_x=\Sigma_x-P_x^{(2)}$. Denote $\theta(x):=A_x^\top\Sigma_x A_x$.*

- (Lee and Sidford, 2019, Lemma 24) $\Sigma'_{x,h}=-2\operatorname{Diag}(\Lambda_x s_{x,h})=2\big(\operatorname{Diag}(P_xS_{x,h}P_x)-\Sigma_xS_{x,h}\big)$.
- (Lee and Sidford, 2019, Lemma 49) $P'_{x,h}=-P_xS_{x,h}-S_{x,h}P_x+2P_xS_{x,h}P_x$.
- $\Lambda'_{x,h}=-2\operatorname{Diag}(\Lambda_xs_{x,h})+2P_x\circ P_xS_{x,h}+2S_{x,h}P_x\circ P_x-2(P_xS_{x,h}P_x)\circ P_x-2P_x\circ(P_xS_{x,h}P_x)$.
- $\Sigma''_{x,h}=6S_{x,h}\Sigma_xS_{x,h}+8\operatorname{Diag}(P_xS_{x,h}P_xS_{x,h}P_x)-6\operatorname{Diag}(P_xS_{x,h}^2P_x)-8\operatorname{Diag}(S_{x,h}P_xS_{x,h}P_x)$.
- $\mathrm{D}\theta(x)[h]=-2A_x^\top\Sigma_xS_{x,h}A_x+A_x^\top\Sigma'_{x,h}A_x$.
- $\mathrm{D}^2\theta(x)[h,h]=6A_x^\top S_{x,h}\Sigma_xS_{x,h}A_x-4A_x^\top\Sigma'_{x,h}S_{x,h}A_x+A_x^\top\Sigma''_{x,h}A_x$. *Equivalently,*

$$\begin{aligned}\mathrm{D}^2\theta(x)[h,h]&=20A_x^\top S_{x,h}\Sigma_xS_{x,h}A_x-16A_x^\top\operatorname{Diag}(S_{x,h}P_xS_{x,h}P_x)A_x\\&\quad-6A_x^\top\operatorname{Diag}(P_xS_{x,h}^2P_x)A_x+8A_x^\top\operatorname{Diag}(P_xS_{x,h}P_xS_{x,h}P_x)A_x.\end{aligned}$$

Proof As for the third item,

$$\begin{aligned}&\Lambda'_{x,h}=\Sigma'_{x,h}-P'_{x,h}\circ P_x-P_x\circ P'_{x,h}\\&=-2\operatorname{Diag}(\Lambda_xs_{x,h})-(-P_xS_{x,h}-S_{x,h}P_x+2P_xS_{x,h}P_x)\circ P_x-P_x\circ(-P_xS_{x,h}-S_{x,h}P_x+2P_xS_{x,h}P_x)\\&\underset{\text{(i)}}{=}-2\operatorname{Diag}(\Lambda_xs_{x,h})+2P_x\circ P_xS_{x,h}+2S_{x,h}P_x\circ P_x-2(P_xS_{x,h}P_x)\circ P_x-2P_x\circ(P_xS_{x,h}P_x)\,,\end{aligned}$$

where in (i) we used $D(A\circ B)=(DA)\circ B=A\circ(DB)$ and $(A\circ B)D=(AD)\circ B=A\circ(BD)$[9] for a diagonal matrix $D\in\mathbb{R}^{d\times d}$ (Lemma H.2).

As for the fourth item,

$$\begin{aligned}&\Sigma''_{x,h}=-2\mathrm{D}\big(\operatorname{Diag}(\Lambda_xs_{x,h})\big)[h]=-2\operatorname{Diag}(\Lambda'_{x,h}s_{x,h})+2\operatorname{Diag}(\Lambda_xS_{x,h}s_{x,h})\\&=-2\operatorname{Diag}\big(\big[-2\operatorname{Diag}(\Lambda_xs_{x,h})+2P_x\circ P_xS_{x,h}+2S_{x,h}P_x\circ P_x-2(P_xS_{x,h}P_x)\circ P_x-2P_x\circ(P_xS_{x,h}P_x)\big]s_{x,h}\big)\\&\qquad+2\operatorname{Diag}(\Lambda_xS_{x,h}s_{x,h})\\&=4\operatorname{Diag}(\Lambda_xs_{x,h})S_{x,h}-4\operatorname{Diag}(P_x\circ P_xS_{x,h}s_{x,h})-4\operatorname{Diag}(S_{x,h}P_x\circ P_xs_{x,h})\\&\qquad+4\operatorname{Diag}\big((P_xS_{x,h}P_x)\circ P_xs_{x,h}\big)+4\operatorname{Diag}\big(P_x\circ(P_xS_{x,h}P_x)s_{x,h}\big)+2\operatorname{Diag}(\Lambda_xS_{x,h}s_{x,h})\\&=4\operatorname{Diag}\big(S_{x,h}(\Sigma_x-P_x\circ P_x)s_{x,h}\big)-4\operatorname{Diag}(P_x\circ P_xS_{x,h}s_{x,h})-4\operatorname{Diag}(S_{x,h}P_x\circ P_xs_{x,h})\\&\qquad+4\operatorname{Diag}\big((P_xS_{x,h}P_x)\circ P_xs_{x,h}\big)+4\operatorname{Diag}\big(P_x\circ(P_xS_{x,h}P_x)s_{x,h}\big)+2\operatorname{Diag}\big((\Sigma_x-P_x\circ P_x)S_{x,h}s_{x,h}\big)\\&=4\operatorname{Diag}(S_{x,h}\Sigma_xs_{x,h})-6\operatorname{Diag}(P_x\circ P_xS_{x,h}s_{x,h})-8\operatorname{Diag}(S_{x,h}P_x\circ P_xs_{x,h})\\&\qquad+4\operatorname{Diag}\big((P_xS_{x,h}P_x)\circ P_xs_{x,h}\big)+4\operatorname{Diag}\big(P_x\circ(P_xS_{x,h}P_x)s_{x,h}\big)+2\operatorname{Diag}(\Sigma_xS_{x,h}s_{x,h})\\&=6\operatorname{Diag}(S_{x,h}\Sigma_xs_{x,h})-6\operatorname{Diag}(P_x\circ P_xS_{x,h}s_{x,h})-8\operatorname{Diag}(S_{x,h}P_x\circ P_xs_{x,h})\\&\qquad+4\operatorname{Diag}\big((P_xS_{x,h}P_x)\circ P_xs_{x,h}\big)+4\operatorname{Diag}\big(P_x\circ(P_xS_{x,h}P_x)s_{x,h}\big)\\&\underset{\text{(i)}}{=}6S_{x,h}\Sigma_x\operatorname{Diag}(s_{x,h})-6\operatorname{Diag}\Big(\operatorname{diag}\big(P_xS_{x,h}(P_xS_{x,h})^\top\big)\Big)-8\operatorname{Diag}\Big(\operatorname{diag}(S_{x,h}P_xS_{x,h}P_x^\top)\Big)\\&\qquad+4\operatorname{Diag}(P_xS_{x,h}P_xS_{x,h}P_x)+4\operatorname{Diag}\big(P_xS_{x,h}(P_xS_{x,h}P_x)^\top\big)\\&=6S_{x,h}\Sigma_xS_{x,h}-6\operatorname{Diag}(P_xS_{x,h}^2P_x)-8\operatorname{Diag}(S_{x,h}P_xS_{x,h}P_x)+8\operatorname{Diag}(P_xS_{x,h}P_xS_{x,h}P_x)\,,\end{aligned}$$

9. This property allows us to write $DA\circ B$ without parenthesis.

where in (i) we applied Lemma H.2-1 to the terms with blue.

Applying the product rule to $\theta(x) = A_x^\mathsf{T}\Sigma_x A_x = A^\mathsf{T} S_x^{-2}\Sigma_x A$,

$$\begin{aligned}
\mathrm{D}\theta[h] &= -2A^\mathsf{T} S_x^{-3}\Sigma_x \operatorname{Diag}(Ah)A + A^\mathsf{T} S_x^{-2}\Sigma'_{x,h}A = -2A_x^\mathsf{T}\Sigma_x S_{x,h}A_x + A_x^\mathsf{T}\Sigma'_{x,h}A_x\,,\\
\mathrm{D}^2\theta[h,h] &= 6A_x^\mathsf{T} S_{x,h}\Sigma_x S_{x,h}A_x - 2A_x^\mathsf{T}\Sigma'_{x,h}S_{x,h}A_x - 2A_x^\mathsf{T} S_{x,h}\Sigma'_{x,h}A_x + A_x^\mathsf{T}\Sigma''_{x,h}A_x\\
&= 6A_x^\mathsf{T} S_{x,h}\Sigma_x S_{x,h}A_x - 4A_x^\mathsf{T}\Sigma'_{x,h}S_{x,h}A_x + A_x^\mathsf{T}\Sigma''_{x,h}A_x\,.
\end{aligned}$$

By substituting $\Sigma'_{x,h}$ and $\Sigma''_{x,h}$ with our formulas above,

$$\begin{aligned}
&\mathrm{D}^2\theta[h,h] = 6A_x^\mathsf{T} S_{x,h}\Sigma_x S_{x,h}A_x - 4A_x^\mathsf{T}\Sigma'_{x,h}S_{x,h}A_x + A_x^\mathsf{T}\Sigma''_{x,h}A_x\\
&= 6A_x^\mathsf{T} S_{x,h}\Sigma_x S_{x,h}A_x + 8A_x^\mathsf{T}\big(\Sigma_x S_{x,h} - \operatorname{Diag}(P_x S_{x,h}P_x)\big)S_{x,h}A_x\\
&\quad + A_x^\mathsf{T}\Big(6S_{x,h}\Sigma_x S_{x,h} - 6\operatorname{Diag}(P_x S_{x,h}^2 P_x) - 8\operatorname{Diag}(S_{x,h}P_x S_{x,h}P_x) + 8\operatorname{Diag}(P_x S_{x,h}P_x S_{x,h}P_x)\Big)A_x\\
&= 20A_x^\mathsf{T} S_{x,h}\Sigma_x S_{x,h}A_x - 16A_x^\mathsf{T}\operatorname{Diag}(S_{x,h}P_x S_{x,h}P_x)A_x - 6A_x^\mathsf{T}\operatorname{Diag}(P_x S_{x,h}^2 P_x)A_x\\
&\quad + 8A_x^\mathsf{T}\operatorname{Diag}(P_x S_{x,h}P_x S_{x,h}P_x)A_x\,.
\end{aligned}$$

■

I.3. Lewis-weight metric

We recall preliminaries on the Lewis weights. Particularly, the leverage scores are simply the ℓ_2-Lewis weights.

Lemma I.5 (Lee and Sidford (2019)) *Let $W_x = \operatorname{Diag}(w_x(A_x)) \in \mathbb{S}_{++}^d$ be the ℓ_p-Lewis weights and $g(x) = A_x^\mathsf{T} W_x A_x$ the Lewis-weights metric, and $h \in \mathbb{R}^d$.*

- (Lemma 26) $\max_{i\in[m]} \frac{[\sigma(W_x^{1/2}A_x)]_i}{(w_x)_i} \le 2m^{\frac{2}{p+2}}$.
- (Lemma 33) $\|A_x h\|_{W_x} = \|h\|_{g(x)}$ *and* $\|A_x h\|_\infty \le \sqrt{2}m^{\frac{1}{p+2}}\|h\|_{g(x)}$.
- (Lemma 34) $\|W_x^{-1}w'_{x,h}\|_{W_x} \le p\,\|h\|_{g(x)}$.

Next is a directional derivative of the ℓ_p-Lewis weight of A_x.

Lemma I.6 (Lee and Sidford (2019), Lemma 24) *The directional derivative of the ℓ_p-Lewis weight W_x in direction $h \in \mathbb{R}^d$ is*

$$W'_{x,h} := \mathrm{D}W_x[h] = -2\operatorname{Diag}(\Lambda_x G_x^{-1}W_x s_{x,h}) = -\operatorname{Diag}(W_x^{\frac{1}{2}}N_x W_x^{\frac{1}{2}}s_{x,h})\,,$$

where $\Lambda_x \overset{\text{def}}{=} W_x - P_x^{(2)}$, $\bar\Lambda_x \overset{\text{def}}{=} W_x^{-\frac{1}{2}}\Lambda_x W_x^{-\frac{1}{2}}$, $G_x \overset{\text{def}}{=} W_x - \big(1-\frac{2}{p}\big)\Lambda_x$, *and* $N_x \overset{\text{def}}{=} 2\bar\Lambda_x(I - c_p\bar\Lambda_x)^{-1}$.

It is known that these matrices satisfy

$$P_x^{(2)} \preceq W_x \preceq I\,, \tag{I.1}$$

$$\Lambda_x \preceq W_x\,, \tag{I.2}$$

$$\frac{2}{p}W_x \preceq G_x \preceq W_x\,,\ \text{which implies } W_x^{-1} \preceq G_x^{-1} \preceq \frac{p}{2}W_x^{-1} \text{ and } I \preceq W_x^{\frac{1}{2}}G_x^{-1}W_x^{\frac{1}{2}} \preceq \frac{p}{2}I\,. \tag{I.3}$$

We can also compute the second-order directional derivative of W_x in direction $h \in \mathbb{R}^d$.

Lemma I.7 (Second-order derivative of W_x) *Let $w_x \in \mathbb{R}^m$ be the ℓ_p-Lewis weight, $\Gamma \in \mathbb{R}^{m\times m}_{\geq 0}$ a diagonal matrix, and $h \in \mathbb{R}^d$. Then,*

$$W''_{x,h} = -\operatorname{Diag}\Big(\frac{1}{2}W_x^{-\frac{1}{2}}W'_{x,h}N_xW_x^{\frac{1}{2}}s_{x,h} + W_x^{\frac{1}{2}}N'_{x,h}W_x^{\frac{1}{2}}s_{x,h} + \frac{1}{2}W_x^{\frac{1}{2}}N_xW_x^{-\frac{1}{2}}W'_{x,h}s_{x,h} + 2\Lambda_xG_x^{-1}W_xs_{x,h}^2\Big)\,,$$

$$\begin{aligned}\operatorname{Tr}(\Gamma W''_{x,h}) = &-\frac{1}{2}\operatorname{Tr}\big(\Gamma\operatorname{Diag}(\underbrace{W_x^{-\frac{1}{2}}W'_{x,h}N_xW_x^{\frac{1}{2}}s_{x,h}}_{\text{I}})\big) - \operatorname{Tr}\big(\Gamma\operatorname{Diag}(\underbrace{W_x^{\frac{1}{2}}N'_{x,h}W_x^{\frac{1}{2}}s_{x,h}}_{\text{II}})\big)\\ &-\frac{1}{2}\operatorname{Tr}\big(\Gamma\operatorname{Diag}(\underbrace{W_x^{\frac{1}{2}}N_xW_x^{-\frac{1}{2}}W'_{x,h}s_{x,h}}_{\text{III}})\big) - 2\operatorname{Tr}\big(\Gamma\operatorname{Diag}(\underbrace{\Lambda_xG_x^{-1}W_xS_{x,h}s_{x,h}}_{\text{IV}})\big)\,,\end{aligned} \tag{I.4}$$

$$\mathrm{D}^2(A_x^\top W_xA_x)[h,h] = 6A_x^\top S_{x,h}W_xS_{x,h}A_x - 4A_x^\top W'_{x,h}S_{x,h}A_x + A_x^\top W''_{x,h}A_x \tag{I.5}$$

where $\|\text{I}\|_{W_x^{-1}} \lesssim p^3m^{\frac{1}{p+2}}\|h\|_\theta^2$, $\|\text{II}\|_{W_x^{-1}} \lesssim p^{3.5}\|h\|_\theta^2$, $\|\text{III}\|_{W_x^{-1}} \lesssim p^3m^{\frac{1}{p+2}}\|h\|_\theta^2$, and $\|\text{IV}\|_{W_x^{-1}} \lesssim pm^{\frac{1}{p+2}}\|h\|_\theta^2$. Here, $\lesssim$ hides universal constants and poly-logarithmic factors in m.

Proof The formula for $W''_{x,h}$ follows from differentiating the formula for $W'_{x,h}$ (Lemma I.6). The dual local norms of I~IV can be bounded as follows:

$$\|\text{I}\|_{W_x^{-1}} = \|W_x^{-1}W'_{x,h}N_xW_x^{\frac{1}{2}}s_{x,h}\|_2 \leq \underbrace{\|W_x^{-1}W'_{x,h}\|_2}_{\text{Lemma I.8-2}}\ \underbrace{\|N_x\|_2}_{\text{Lemma I.8-1}}\ \|W_x^{\frac{1}{2}}s_{x,h}\|_2 \lesssim p^3m^{\frac{1}{p+2}}\|h\|_\theta^2\,,$$

$$\|\text{II}\|_{W_x^{-1}} = \|N'_{x,h}W_x^{\frac{1}{2}}s_{x,h}\|_2 \leq \underbrace{\|I+N_x\|_2}_{\text{Lemma I.8-1}}\underbrace{\|(I+N_x)^{-\frac{1}{2}}N'_{x,h}(I+N_x)^{-\frac{1}{2}}\|_2}_{\text{Lemma I.8-3}}\|W_x^{\frac{1}{2}}s_{x,h}\|_2 \lesssim p^{3.5}\|h\|_\theta^2\,,$$

$$\|\text{III}\|_{W_x^{-1}} = \|N_xW_x^{-\frac{1}{2}}W'_{x,h}s_{x,h}\|_2 \leq \underbrace{\|N_x\|_2}_{\text{Lemma I.8-1}}\ \underbrace{\|W_x^{-1}W'_{x,h}\|_2}_{\text{Lemma I.8-2}}\|W_xs_{x,h}\|_2 \lesssim p^3m^{\frac{1}{p+2}}\|h\|_\theta^2\,,$$

$$\begin{aligned}\|\text{IV}\|^2_{W_x^{-1}} &= s_{x,h}^\top S_{x,h}W_xG_x^{-1}\underbrace{\Lambda_xW_x^{-1}\Lambda_x}_{\preceq W_x\ (\text{I.2})}G_x^{-1}W_xS_{x,h}s_{x,h} \leq s_{x,h}^\top S_{x,h}W_x\underbrace{G_x^{-1}W_xG_x^{-1}}_{\preceq\frac{p^2}{4}W_x^{-1}\ (\text{I.3})}W_xS_{x,h}s_{x,h}\\ &\leq p^2s_{x,h}^\top W_x^{\frac{1}{2}}S_{x,h}^2W_x^{\frac{1}{2}}s_{x,h} \leq p^2\|s_{x,h}\|_\infty^2\|h\|_\theta^2 \leq p^2m^{\frac{2}{p+2}}\|h\|_\theta^4\,,\end{aligned}$$

where we used Lemma I.5-2 in the last inequality. ■

Next, we recall bounds on the derivatives of matrices relevant to Lewis weights.

Lemma I.8 (Lee and Sidford (2019)) *Let $Ax \geq b$ and $h \in \mathbb{R}^d$. For $c_p = 1 - 2/p$ with $p > 2$, let $\bar{\Lambda}_x := W_x^{-\frac{1}{2}}\Lambda_xW_x^{-\frac{1}{2}} = I - W_x^{-\frac{1}{2}}P_x^{(2)}W_x^{-\frac{1}{2}}$, $N_x \overset{\text{def}}{=} 2\bar{\Lambda}_x(I - c_p\bar{\Lambda}_x)^{-1}$ and $\theta_x = A_x^\top W_xA_x$.*

- (Lemma 31) *N_x is symmetric and $0 \preceq N_x \preceq pI$.*
- (Lemma 34) $\|W_x^{-1}w_{x,h}\|_\infty \leq p(\sqrt{2}m^{\frac{1}{p+2}} + p/2)\,\|h\|_{\theta_x}$.
- (Lemma 37) $\|(I+N_x)^{-\frac{1}{2}}\mathrm{D}N_x[h]\,(I+N_x)^{-\frac{1}{2}}\|_2 \leq 4p^{5/2}\|h\|_{\theta_x}$.

Lastly, we remind a result about closeness of the Lewis weights at close-by points.

Lemma I.9 (Lee and Sidford (2019)) *In the same setting above, let* $x_t = x + th$, $s_t = s_{x_t}$, $w_t = w_{x_t}$, *and* $z_{t,\alpha} \in \mathbb{R}^m$ *be a vector defined by* $[z_{t,\alpha}]_i := \frac{\mathrm{d}}{\mathrm{d}t} \log\Bigl(\frac{[w_{t,i}]^\alpha}{s_{t,i}}\Bigr)$. *Then,*

$$\|z_t\|_\infty \le \bigl(\sqrt{2}(1+|\alpha|p)m^{\frac{1}{p+2}} + p\,|\alpha|\,\max(1,p/2)\bigr)\,\|h\|_{A_t^\top W_t A_t}\,.$$

Now we present an auxiliary result showing HSC of the Lewis-weight metric.

Lemma I.10 *The metric* $g(x) = cA_x^\top W_x A_x$ *is HSC for* $c = c_1(\log m)^{c_2} d^{1/2}$ *with some constants* $c_1, c_2 > 0$,

Proof Let $\theta(x) = A_x^\top W_x A_x$ and $h \in \mathbb{R}^d$. From (I.5),

$$\begin{aligned}\mathrm{D}^2\theta[h,h,h,h] &= 6s_{x,h}^\top S_{x,h} W_x S_{x,h} s_{x,h} - 4s_{x,h}^\top W_{x,h}' S_{x,h} s_{x,h} + s_{x,h}^\top W_{x,h}'' s_{x,h}\\ &= \mathrm{Tr}(6S_{x,h}^4 W_x - 4S_{x,h}^3 W_{x,h}' + S_{x,h}^2 W_{x,h}'')\,. \qquad \text{(I.6)}\end{aligned}$$

As for the first term, $|\mathrm{Tr}(S_{x,h}^4 W_x)| \le \|s_{x,h}\|_\infty^2 \|h\|_\theta^2$. As for the second term,

$$\begin{aligned}|\mathrm{Tr}(S_{x,h}^3 W_{x,h}')| &\le \|s_{x,h}\|_\infty^2 \,\mathrm{Tr}\bigl(\sqrt{S_{x,h} W_{x,h}'^2 S_{x,h}}\bigr) = \|s_{x,h}\|_\infty^2 \,\mathrm{Tr}\bigl(\sqrt{W_{x,h}' W_x^{-1} W_{x,h}'}\sqrt{S_{x,h} W_x S_{x,h}}\bigr)\\ &\underset{\text{(i)}}{\le} \|s_{x,h}\|_\infty^2 \sqrt{\mathrm{Tr}(W_{x,h}' W_x^{-1} W_{x,h}')}\sqrt{\mathrm{Tr}(S_{x,h} W_x S_{x,h})} = \|s_{x,h}\|_\infty^2 \|W_x^{-1} w_{x,h}'\|_{W_x} \|h\|_\theta\\ &\underset{\text{(ii)}}{\le} p\|s_{x,h}\|_\infty^2 \|h\|_\theta^2 \qquad \text{(I.7)}\end{aligned}$$

where we used the Cauchy-Schwarz in (i) and Lemma I.5-3 in (ii).

As for the last term, we first use the formula for $\mathrm{Tr}(S_{x,h}^2 W_{x,h}'')$ with $\Gamma = S_{x,h}^2$ in Lemma I.7. Each term there is of the form $\mathrm{Tr}(S_{x,h}^2 \,\mathrm{Diag}(v))$ for $v =$ I ˜ IV, which can be bounded as follows:

$$\begin{aligned}\bigl|\mathrm{Tr}\bigl(S_{x,h}^2 \,\mathrm{Diag}(v)\bigr)\bigr| &= \bigl|\mathrm{Tr}\bigl(S_{x,h}^2 W_x^{\frac12} W_x^{-\frac12}\,\mathrm{Diag}(v)\bigr)\bigr| \le \sqrt{\mathrm{Tr}(W_x^{\frac12} S_{x,h}^4 W_x^{\frac12})}\sqrt{\mathrm{Tr}\bigl(\mathrm{Diag}(v) W_x^{-1}\,\mathrm{Diag}(v)\bigr)} \qquad \text{(I.8)}\\ &\le \|s_{x,h}\|_\infty \,\|h\|_\theta\, \|v\|_{W_x^{-1}}\,.\end{aligned}$$

Using the norm bounds in Lemma I.7, it follows that $|\mathrm{Tr}(S_{x,h}^2 W_{x,h}'')| \lesssim \|h\|_\theta^4$ for $p = \mathcal{O}(\log m)$. Putting everything together with $\|s_{x,h}\|_\infty \le \sqrt{2} m^{\frac{1}{p+2}} \|h\|_\theta \lesssim \|h\|_\theta$ (Lemma I.5-2),

$$|\mathrm{D}^2\theta[h,h,h,h]| \lesssim \|s_{x,h}\|_\infty^2 \|h\|_\theta^2 + \|s_{x,h}\|_\infty \|h\|_\theta^3 \lesssim \|h\|_\theta^4\,.$$

■

Appendix J. Technical lemmas

Lemma J.1 *For a matrix* $M \in \mathbb{R}^{m\times d}$ *and* $E \in \mathbb{R}^{d\times d}$ *such that* $E + M^\top M \succ 0$, *it holds that*

$$M(E + M^\top M)^{-1} M^\top \preceq P(M) = M(M^\top M)^\dagger M^\top\,.$$

Proof Let us denote the LHS by P' and the RHS by P. We show $I - P' \succeq I - P$ instead. First, $(P')^2 \preceq P'$ and $(I-P')^2 \preceq I - P'$ follow from

$$P'P' = M(E + M^\top M)^{-1} \underbrace{M^\top M}_{\preceq E + M^\top M} (E + M^\top M)^{-1} M^\top \preceq M(E + M^\top M)^{-1} M^\top = P' ,$$

$$(I - P')^2 = I + P'P' - 2P' \preceq I - P' .$$

It follows from $(I - P')^2 \preceq I - P'$ that for any $v \in \mathbb{R}^m$

$$v^\top (I - P')v \geq \|(I - P')v\|^2 \geq \|(I - P)v\|_2^2 = v^\top (I - P)v ,$$

where the inequality holds due to $P'v, Pv \in \mathrm{range}(M)$ and $Pv = \arg\min_{w \in \mathrm{range}(M)} \|v - w\|_2^2$. ∎

Proposition J.2 *Let $v, w, p, q, r, s \in \mathbb{R}^d$ and $h \sim \mathcal{N}(0, I_d)$.*
- $\mathbb{E}[(v\cdot h)(w\cdot h)^3] = 3\|w\|^2(v\cdot w)$.
- $\mathbb{E}[(v\cdot h)^2(w\cdot h)^2] = \|v\|^2\|w\|^2 + 2(v\cdot w)^2$.
- $\mathbb{E}[(p\cdot h)^2(r\cdot h)(s\cdot h)] = \|p\|^2(r\cdot s) + 2(p\cdot s)(p\cdot r)$.

Proof Using Stein's lemma (Lemma E.9),

$$\begin{aligned}
\mathbb{E}[(v\cdot h)(w\cdot h)^3] &\underset{\text{Stein}}{=} \sum_i w_i \mathbb{E}[h_i (v\cdot h)(w\cdot h)^2] = \sum_i w_i \big(v_i \mathbb{E}[(w\cdot h)^2] + 2w_i \mathbb{E}[(v\cdot h)(w\cdot h)]\big) \\
&= (v\cdot w)\|w\|^2 + 2\|w\|^2(v\cdot w) = 3\|w\|^2(v\cdot w) , \\
\mathbb{E}[(v\cdot h)^2(w\cdot h)^2] &= \sum_i v_i \mathbb{E}[h_i (v\cdot h)(w\cdot h)^2] \underset{\text{Stein}}{=} \sum_i v_i \left(v_i \mathbb{E}[(w\cdot h)^2] + 2w_i \mathbb{E}[(v\cdot h)(w\cdot h)]\right) \\
&= \|v\|^2\|w\|^2 + 2(v\cdot w)^2 , \\
\mathbb{E}[(p\cdot h)^2(r\cdot h)(s\cdot h)] &= \sum_i p_i \mathbb{E}[h_i (p\cdot h)(r\cdot h)(s\cdot h)] \\
&\underset{\text{Stein}}{=} \sum p_i \left(p_i \mathbb{E}[(r\cdot h)(s\cdot h)] + r_i \mathbb{E}[(p\cdot h)(s\cdot h)] + s_i \mathbb{E}[(p\cdot h)(r\cdot h)]\right) \\
&= \|p\|^2(r\cdot s) + (p\cdot r)(p\cdot s) + (p\cdot s)(p\cdot r) = \|p\|^2(r\cdot s) + 2(p\cdot s)(p\cdot r) .
\end{aligned}$$

∎

These estimations result in a useful lemma for establishing SASC of barriers for linear constraints.

Lemma J.3 *For $v, w \in \mathbb{R}^d$ and $h \sim \mathcal{N}(0, I_d)$, $\mathbb{E}[(v\cdot h)^3(w\cdot h)^3] = 9\|v\|^2\|w\|^2(v\cdot w) + 6(v\cdot w)^3$.*

Proof Using Stein's lemma,

$$\begin{aligned}
\mathbb{E}[(v\cdot h)^3(w\cdot h)^3] &= \sum_i v_i \mathbb{E}[h_i (v\cdot h)^2(w\cdot h)^3] = \sum v_i \big(2v_i \mathbb{E}[(v\cdot h)(w\cdot h)^3] + 3w_i \mathbb{E}[(v\cdot h)^2(w\cdot h)^2]\big) \\
&\underset{\text{(i)}}{=} 2\|v\|^2 \cdot 3\|w\|^2(v\cdot w) + 3(v\cdot w)\big(\|v\|^2\|w\|^2 + 2(v\cdot w)^2\big) = 9\|v\|^2\|w\|^2 + 6(v\cdot w)^3 ,
\end{aligned}$$

where in (i) we used Proposition J.2-1 and 2. ∎

Lemma J.4 *For $p, q, r, s \in \mathbb{R}^d$ and $h \sim \mathcal{N}(0, I_d)$,*

$$\mathbb{E}[(p \cdot h)^2 (q \cdot h)(r \cdot h)^2 (s \cdot h)] = (q \cdot s)\|p\|^2 \|r\|^2 + 4(p \cdot r)(p \cdot q)(r \cdot s)$$
$$+2\|p\|^2 (r \cdot q)(r \cdot s) + 2\|r\|^2 (p \cdot q)(p \cdot s) + 2(p \cdot r)^2 (q \cdot s) + 4(p \cdot s)(p \cdot r)(r \cdot q)\,.$$

Proof Using Stein's lemma,

$$\begin{aligned}
&\mathbb{E}[(p \cdot h)^2 (q \cdot h)(r \cdot h)^2 (s \cdot h)] = \sum_i q_i \mathbb{E}[h_i (p \cdot h)^2 (r \cdot h)^2 (s \cdot h)] \\
&= \sum q_i \big(2p_i \mathbb{E}[(p \cdot h)(r \cdot h)^2 (s \cdot h)] + 2r_i \mathbb{E}[(p \cdot h)^2 (r \cdot h)(s \cdot h)] + 2s_i \mathbb{E}[(p \cdot h)^2 (r \cdot h)^2]\big) \\
&\underset{\text{(i)}}{=} 2(p \cdot q)\big(\|r\|^2 (p \cdot s) + 2(p \cdot r)(r \cdot s)\big) + 2(r \cdot q)\big(\|p\|^2 (r \cdot s) + 2(p \cdot s)(p \cdot r)\big) \\
&\qquad + (q \cdot s)\big(\|p\|^2 \|r\|^2 + 2(p \cdot r)^2\big) \\
&= (q \cdot s)\|p\|^2 \|r\|^2 + 4(p \cdot r)(p \cdot q)(r \cdot s) + 2\|p\|^2 (r \cdot q)(r \cdot s) + 2\|r\|^2 (p \cdot q)(p \cdot s) \\
&\qquad + 2(p \cdot r)^2 (q \cdot s) + 4(p \cdot s)(p \cdot r)(r \cdot q)\,.
\end{aligned}$$

In (i), we used Proposition J.2-3 to the first two terms and Proposition J.2-2 to the third term. ■