# CS395T: Continuous Algorithms, Part XV
# Continuous random walks

Kevin Tian

## 1 Sampling from a convex body

In this lecture, we begin our study of sampling from continuous distributions (i.e., those which have absolutely continuous densities with respect to the Lebesgue measure on $\mathbb{R}^d$). To introduce key analysis techniques for continuous random walks, we use the fundamental problem of producing an approximately-uniform sample from a convex body $\mathcal{K} \subseteq \mathbb{R}^d$ as a running example throughout these notes. Specifically we study uniform sampling from convex bodies under *membership oracle* access, i.e., we assume access to an oracle which takes as input $\mathbf{x} \in \mathbb{R}^d$, and tells us whether $\mathbf{x} \in \mathcal{K}$.

Sampling from a convex body is a prominent example of the more general problem of sampling from a logconcave density on $\mathbb{R}^d$ (see Section 4, Part I for the relevant definitions). When $\mu \propto \exp(-f)$ is a logconcave density on $\mathbb{R}^d$, so $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is convex, the assumption of value oracle access to $f$ grants us membership oracle access in the special case when $f = \chi_{\mathcal{K}}$ is the 0-$\infty$ characteristic function of $\mathcal{K}$. We will discuss continuous sampling settings beyond uniform densities on convex bodies in the following lectures, which build upon the tools introduced here.

One reason that we base our exploration of continuous sampling in this uniform sampling problem is historical: uniform sampling from a convex body has been a longstanding testbed for the development of new techniques for continuous Markov chains. Indeed, the fact that polynomial-time sampling was possible even in this special case was a surprising breakthrough when first established by [DFK91]. For example, previous results by [Ele86, BF87] showed that it is impossible to approximate the volume of a convex body up to an exponential factor in the dimension in deterministic polynomial time, and [DF88] proved that it is #P-hard to exactly compute the volume. By crucially leveraging randomness, [DFK91] showed how to use their convex body sampler to estimate volume up to a $1 \pm \epsilon$ multiplicative factor with high probability in time $\text{poly}(d, \epsilon^{-1})$.

To do so, one can follow a standard reduction from counting to sampling. In particular, assuming that $\mathbb{B}(1) \subseteq \mathcal{K} \subseteq \mathbb{B}(d)$, we can estimate $\text{Vol}(\mathcal{K})$ as the product of volume ratios for sets $\mathcal{K}_i := \mathcal{K} \cap \mathbb{B}(2^{\frac{i}{d}})$ over a sequence of indices $0 \le i \le d \log_2 d$, i.e.,

$$\text{Vol}(\mathcal{K}) = \text{Vol}(\mathcal{K}_0) \cdot \prod_{0 \le i \le \lfloor d \log_2 d \rfloor} \frac{\text{Vol}(\mathcal{K}_{i+1})}{\text{Vol}(\mathcal{K}_i)} = \text{Vol}(\mathbb{B}(1)) \cdot \prod_{0 \le i \le \lfloor d \log_2 d \rfloor} \frac{\text{Vol}(\mathcal{K}_{i+1})}{\text{Vol}(\mathcal{K}_i)}.$$

Here we used that our containment assumptions imply $\mathcal{K}_0 = \mathbb{B}(1)$ and $\mathcal{K}_{\lfloor d \log_2 d \rfloor + 1} = \mathcal{K}$. Moreover, each of the volume ratios we need to estimate, of the form

$$\frac{\text{Vol}(\mathcal{K}_{i+1})}{\text{Vol}(\mathcal{K}_i)},$$

are $\Theta(1)$, so they can be efficiently estimated to multiplicative error via random sampling, simply by checking what proportion of random samples from $\mathcal{K}_{i+1}$ also fall in $\mathcal{K}_i$. Finally, as we will see in a future lecture, the assumption that $\mathbb{B}(1) \subseteq \mathcal{K} \subseteq \mathbb{B}(d)$ (up to a negligible fraction of the body) can also be achieved via random sampling, by estimating an affine transformation which makes the body isotropic,[1] and applying concentration inequalities for logconcave densities.

The original [DFK91] algorithm, while polynomial-time, used $\approx d^{19}$ membership oracle queries to compute a single sample. Since this seminal work, a long line of simplifications of the [DFK91]

[1] We say that a density $\pi$ on $\mathbb{R}^d$ is *isotropic* if $\mathbb{E}_{x \sim \pi}[x] = \mathbf{0}_d$ and $\mathbb{E}_{x \sim \pi}[xx^\top] = \mathbf{I}_d$.

framework, as well as new algorithmic tools, have considerably improved the query complexity of sampling from a convex body [AK91, LS92, LS93, KLS97, LV06a, LV06b, LV06c, LV07, CV18, JLLV21] and its extension, sampling general logconcave densities [KV25]. This lecture gives the tools needed to obtain a basic sampler in a simple regime where the convex body is assumed to be near-isotropic, and we only need low-accuracy guarantees. We will develop techniques for removing both assumptions, and extending to general logconcave densities, in the following lectures.

In Sections 2 and 3, we begin by giving general-purpose tools for analyzing the convergence of random walks in continuous space, which are not specialized to any particular random walk. We specialize Sections 4 and 5 to the analysis of a particular Markov chain, the *ball walk*, a basic building block for all algorithms we develop towards sampling convex bodies.

# 2 Mixing in the continuous setting

## 2.1 Preliminaries

Throughout these notes, for any density $\pi$ over an event space $\mathcal{E}$, we overload $\pi(S)$ to mean $\Pr_{s\sim\pi}[s \in S]$ for any $S \subseteq \mathcal{E}$, e.g., if $\pi$ is a density over $\mathbb{R}^d$, then $\pi(S)$ denotes the probability a sample from $\pi$ falls in $S \subseteq \mathbb{R}^d$. We let $\pi^\star$ denote a target stationary density on $\mathbb{R}^d$, which we assume is absolutely continuous with respect to the Lebesgue measure.[2]

We consider Markov chains parameterized by *transition distributions* $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$, which are also densities on $\mathbb{R}^d$, so

$$\int \mathcal{T}_{\mathbf{x}}(\mathbf{y})\mathrm{d}\mathbf{y} = 1 \text{ for all } \mathbf{x} \in \mathbb{R}^d. \tag{1}$$

We assume each $\mathcal{T}_{\mathbf{x}}$ is *lazy*, which means that it puts a point mass of at least $\frac{1}{2}$ on $\mathbf{x}$, and that its density on $\mathbb{R}^d \setminus \{\mathbf{x}\}$ is absolutely continuous with respect to the Lebesgue measure. This is the analog of Eq. (7), Part XIV, and as was the case there, laziness facilitates simpler proofs (but does not meaningfully affect implementation). The $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$ induce a random walk via the update

$$\mathbf{x}_{k+1} \sim \mathcal{T}_{\mathbf{x}_k}, \text{ for all } k \geq 0.$$

We denote the distribution that the random walk is initialized with by $\pi_0$, and the distribution of the $k^{\text{th}}$ iterate of the walk by $\pi_k$. By definition of the random walk, for all $k \geq 0$,

$$\pi_{k+1}(\mathbf{x}) = \int \mathcal{T}_{\mathbf{y}}(\mathbf{x})\pi_k(\mathbf{y})\mathrm{d}y. \tag{2}$$

We additionally assume that the $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$ are *reversible* with respect to $\pi^\star$, analogously to Definition 3, Part XIV. In the continuous setting, this means that for all $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$\pi^\star(\mathbf{x})\mathcal{T}_{\mathbf{x}}(\mathbf{y}) = \pi^\star(\mathbf{y})\mathcal{T}_{\mathbf{y}}(\mathbf{x}). \tag{3}$$

Under the assumption (3), it is simple to verify that $\pi^\star$ is indeed a stationary distribution for the random walk with updates (2). Indeed, supposing that $\pi_k = \pi^\star$ for some $k$, we have

$$\pi_{k+1}(\mathbf{x}) = \int \mathcal{T}_{\mathbf{y}}(\mathbf{x})\pi^\star(\mathbf{y})\mathrm{d}\mathbf{y} = \int \mathcal{T}_{\mathbf{x}}(\mathbf{y})\pi^\star(\mathbf{x})\mathrm{d}\mathbf{y} = \pi^\star(\mathbf{x}), \tag{4}$$

where the second equality used (3) and the third used (1). Moreover, given a set of *proposal distributions* $\{\mathcal{P}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$ and a target stationary distribution $\pi^\star$, such that (3) is not necessarily satisfied with respect to the $\{\mathcal{P}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$ one can apply the Metropolis-Hastings correction (Lemma 3, Part XIV) to obtain transition distributions which do satisfy (3), by letting

$$\mathcal{T}_{\mathbf{x}}(\mathbf{y}) := \mathcal{P}_{\mathbf{x}}(\mathbf{y}) \min\left(1, \frac{\pi^\star(\mathbf{y})\mathcal{P}_{\mathbf{y}}(\mathbf{x})}{\pi^\star(\mathbf{x})\mathcal{P}_{\mathbf{x}}(\mathbf{y})}\right) \text{ for all } \mathbf{y} \in \mathbb{R}^d,\ \mathbf{y} \neq \mathbf{x}. \tag{5}$$

As in Part XIV, the correction (5) has a simple interpretation as applying a filter which either accepts the proposal distribution or does not move, with an acceptance probability proportional to how "reversible" the proposals are for the relevant pair of points. Finally, when the transition distributions $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$ of a Markov chain on $\mathbb{R}^d$ are clear from context, we use $\mathcal{T}\pi$ to denote the density resulting from applying one step of the Markov chain to a point drawn from $\pi$.

[2] In Sections 4 and 5, we focus on the specific case where $\pi^\star(\mathbf{x}) \propto \mathbb{I}_{\mathbf{x}\in\mathcal{K}}$, for convex $\mathcal{K} \subseteq \mathbb{R}^d$.

## 2.2 Mixing from conductance

In this section, we introduce a continuous definition of the *conductance* of a random walk, which is used to analyze mixing, and is patterned off of the discrete analog in Definition 6, Part XIV.

**Definition 1** (Conductance). *For a random walk on $\mathbb{R}^d$ with stationary distribution $\pi^\star$ and lazy, reversible transition distributions $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$, we denote the flow out of a set $S \subseteq \mathbb{R}^d$ by*

$$Q(S) := \int_{\mathbf{x}\in S} \mathcal{T}_{\mathbf{x}}(S^c)\pi^\star(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{\mathbf{x}\in S^c} \mathcal{T}_{\mathbf{x}}(S)\pi^\star(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{6}$$

*where $S^c := \mathbb{R}^d \setminus S$. We define the* conductance *of the random walk by*

$$\Phi := \inf_{0<\pi^\star(S)\le\frac{1}{2}} \frac{Q(S)}{\pi^\star(S)}. \tag{7}$$

We note that the equality (6) holds because of the calculation

$$\begin{aligned}\int_{\mathbf{x}\in S} \mathcal{T}_{\mathbf{x}}(S^c)\pi^\star(\mathbf{x})\mathrm{d}\mathbf{x} &= \int_{\mathbf{x}\in S}\int_{\mathbf{y}\in S^c} \mathcal{T}_{\mathbf{x}}(\mathbf{y})\pi^\star(\mathbf{x})\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x} \\ &= \int_{\mathbf{x}\in S}\int_{\mathbf{y}\in S^c} \mathcal{T}_{\mathbf{y}}(\mathbf{x})\pi^\star(\mathbf{y})\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x} = \int_{\mathbf{y}\in S^c} \mathcal{T}_{\mathbf{y}}(S)\pi^\star(\mathbf{y})\mathrm{d}\mathbf{y},\end{aligned}$$

where we applied reversibility and Fubini's theorem. As in Part XIV, the quantity $\frac{Q(S)}{\pi^\star(S)}$ has a natural interpretation as the probability that one step of a random walk initialized randomly in $S$ leaves the set. The intuition is again that if the conductance $\Phi$ is large, then no set $S$ is a bottleneck for the random walk, and hence the walk will rapidly mix. In order to quantify this intuition, we require a few new definitions specialized to the continuous setting.

**Definition 2** (Warmness). *For densities $\pi_0, \pi^\star$ on $\mathbb{R}^d$, we say $\pi_0$ is $\beta$*-warm *with respect to $\pi^\star$ if*

$$\frac{\pi_0(\mathbf{x})}{\pi^\star(\mathbf{x})} \le \beta \text{ for all } \mathbf{x} \in \mathbb{R}^d.$$

**Definition 3** (Lovasz-Simonovits curve). *For a random walk on $\mathbb{R}^d$ with stationary distribution $\pi^\star$ and lazy, reversible transition distributions $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$, initialized at $\pi_0$ and with density $\pi_k$ after $k$ steps for all $k \in \mathbb{N}$, we define the* Lovasz-Simonovits curve *$h_k : [0,1] \to [0,1]$ for all $k \in \mathbb{N}$ by*

$$h_k(\tau) := \sup_{g\in G_\tau} \int g(\mathbf{x})\pi_k(\mathbf{x})\mathrm{d}\mathbf{x} - \tau, \text{ where } G_\tau := \left\{g : \mathbb{R}^d \to [0,1] \mid \int g(\mathbf{x})\pi^\star(\mathbf{x})\mathrm{d}\mathbf{x} = \tau\right\}. \tag{8}$$

Definition 2 is fairly straightforward, and it implies in particular that for any $S \subseteq \mathbb{R}^d$ that $\pi_0(S) \le \beta\pi^\star(S)$. Definition 3 is somewhat more complicated, but we briefly interpret it here. First, it is straightforward to check that if $\pi_0$ is absolutely continuous with respect to $\pi^\star$, then all $\pi_k$ for $k \in \mathbb{N}$ are too, so the supremum over $G_\tau$ in (8) is achieved by greedily placing mass on points sorted by the value of $\frac{\pi_k}{\pi}$. In other words, the supremum is achieved by a 0-1 set indicator function, $g(\mathbf{x}) = \mathbb{I}_{\mathbf{x}\in S}$, for some $S \subseteq \mathbb{R}^d$. Rewriting (8) where the supremum is taken over $S \subseteq \mathbb{R}^d$,

$$h_k(\tau) = \sup_{\substack{S\subseteq\mathbb{R}^d \\ \pi^\star(S)=\tau}} \pi_k(S) - \tau. \tag{9}$$

Next, the maximum of the above expression over $\tau \in [0,1]$ is exactly the first characterization of the total variation distance from Fact 1, Part XIV. Therefore, to show that $\pi_k \to \pi^\star$ in total variation, it suffices to uniformly bound $h_k$ over $[0,1]$, which we now provide the tools for.

**Lemma 1.** *Consider a random walk on $\mathbb{R}^d$ with stationary distribution $\pi^\star$ and lazy, reversible transition distributions $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$, initialized at $\pi_0$ and with density $\pi_k$ after $k$ steps for all $k \in \mathbb{N}$, and suppose it has conductance $\Phi$. Then the Lovasz-Simonovits curve (8) satisfies:*

$$h_k(\tau) \le \frac{1}{2}h_{k-1}\left((1-2\Phi)\tau\right) + \frac{1}{2}h_{k-1}\left((1+2\Phi)\tau\right) \text{ for all } \tau \in \left[0, \frac{1}{2}\right],$$

$$h_k(\tau) \le \frac{1}{2}h_{k-1}\left(\tau - 2\Phi(1-\tau)\right) + \frac{1}{2}h_{k-1}\left(\tau + 2\Phi(1-\tau)\right) \text{ for all } \tau \in \left[\frac{1}{2}, 1\right].$$

*Proof.* We only prove the former statement, as the latter follows analogously. We observe that $h_k$ is a concave function of $\tau$ for all $k \in \mathbb{N} \cup \{0\}$, since recalling the characterization (9), $h_k$ selects points for the optimal set greedily according to $\frac{\pi_k}{\pi}$, which is a decreasing ratio as the $\pi^\star$-measure of the set $\tau$ increases. Next, let $S$ achieve the bound in $h_k(\tau)$, following (9). We define

$$g_1(\mathbf{x}) := \begin{cases} 2\mathcal{T}_{\mathbf{x}}(S) - 1 & \mathbf{x} \in S \\ 0 & \mathbf{x} \notin S \end{cases}, \; g_2(\mathbf{x}) := \begin{cases} 1 & \mathbf{x} \in S \\ 2\mathcal{T}_{\mathbf{x}}(S) & \mathbf{x} \notin S \end{cases}.$$

By laziness, $g_1(\mathbf{x}) \in [0, 1]$ pointwise (since $2\mathcal{T}_{\mathbf{x}}(S) \geq 1$ if $\mathbf{x} \in S$); we similarly have $g_2(\mathbf{x}) \in [0, 1]$ pointwise. Moreover, $\frac{1}{2}(g_1(\mathbf{x}) + g_2(\mathbf{x})) = \mathcal{T}_{\mathbf{x}}(S)$ for all $\mathbf{x} \in \mathbb{R}^d$. Thus, letting

$$\tau_1 := \int g_1(\mathbf{x}) \pi^\star(\mathbf{x}) \mathrm{d}\mathbf{x}, \; \tau_2 := \int g_2(\mathbf{x}) \pi^\star(\mathbf{x}) \mathrm{d}\mathbf{x},$$

we have $\frac{1}{2}(\tau_1 + \tau_2) = \int \mathcal{T}_{\mathbf{x}}(S) \pi^\star(\mathbf{x}) \mathrm{d}\mathbf{x} = \pi^\star(S)$ by the definition of stationary measure. Further,

$$\begin{aligned} h_k(\tau) = \pi_k(S) - \pi^\star(S) &= \int \mathcal{T}_{\mathbf{x}}(S) \pi_{k-1}(\mathbf{x}) \mathrm{d}\mathbf{x} - \tau \\ &= \frac{1}{2}\left(\int g_1(\mathbf{x}) \pi_{k-1}(\mathbf{x}) \mathrm{d}\mathbf{x} - \tau_1\right) + \frac{1}{2}\left(\int g_2(\mathbf{x}) \pi_{k-1}(\mathbf{x}) \mathrm{d}\mathbf{x} - \tau_2\right) \\ &\leq \frac{1}{2} h_{k-1}(\tau_1) + \frac{1}{2} h_{k-1}(\tau_2), \end{aligned}$$

where the first line used (2), and the last used that $g_1 \in G_{\tau_1}$, $g_2 \in G_{\tau_2}$ by definition. Finally, we use the definition of conductance to show that $\tau_1$ and $\tau_2$ are in fact separated from $\tau$:

$$\begin{aligned} \tau_1 &= \int_{\mathbf{x} \in S} (2\mathcal{T}_{\mathbf{x}}(S) - 1) \pi^\star(\mathbf{x}) \mathrm{d}\mathbf{x} = 2 \int_{\mathbf{x} \in S} \mathcal{T}_{\mathbf{x}}(S) \pi^\star(\mathbf{x}) \mathrm{d}\mathbf{x} - \tau \\ &= 2 \int_{\mathbf{x} \in S} \left(1 - \mathcal{T}_{\mathbf{x}}(S^c)\right) \pi^\star(\mathbf{x}) \mathrm{d}\mathbf{x} - \tau \\ &= \tau - 2 \int_{\mathbf{x} \in S} \mathcal{T}_{\mathbf{x}}(S^c) \pi^\star(\mathbf{x}) \mathrm{d}\mathbf{x} = \tau - 2Q(S) \leq (1 - 2\Phi)\tau. \end{aligned}$$

In the last line, we recalled the definitions (6) and (7). This also implies $\tau_2 \geq (1 + 2\Phi)\tau$, since earlier we showed $\frac{1}{2}(\tau_1 + \tau_2) = \tau$. Finally, by concavity of $h_{k-1}$, we have the conclusion

$$h_k(\tau) \leq \frac{1}{2} h_{k-1}(\tau_1) + \frac{1}{2} h_{k-1}(\tau_2) \leq \frac{1}{2} h_{k-1}\left((1 - 2\Phi)\tau\right) + \frac{1}{2} h_{k-1}\left((1 + 2\Phi)\tau\right).$$

□

As a consequence of Lemma 1, we have the following bound on the decay of the Lovasz-Simonovits curve when the random walk is initialized at a warm start.

**Corollary 1.** *Consider a random walk on $\mathbb{R}^d$ with stationary distribution $\pi^\star$ and lazy, reversible transition distributions $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x} \in \mathbb{R}^d}$, initialized at $\pi_0$ and with density $\pi_k$ after $k$ steps for all $k \in \mathbb{N}$, and suppose it has conductance $\Phi$. Let $\pi_0$ be $\beta$-warm with respect to $\pi^\star$. Then,*

$$D_{\mathrm{TV}}\left(\pi_k, \pi^\star\right) \leq \epsilon, \text{ if } k \geq \frac{2}{\Phi^2} \log\left(\frac{\beta}{\epsilon}\right).$$

*Proof.* We claim that for all $\tau \in [0, 1]$ and all $k \in \mathbb{N} \cup \{0\}$,

$$h_k(\tau) \leq \beta \min\left(\sqrt{\tau}, \sqrt{1 - \tau}\right) \left(1 - \frac{\Phi^2}{2}\right)^k. \tag{10}$$

The conclusion then follows from (10) using (9) and the first definition of total variation distance from Fact 1, Part XIV, as well as $\exp(-c) \geq 1 - c$ for $c \in [0, 1]$. For $k = 0$, (10) follows from warmness, since (following notation (9)) any $S$ with $\pi^\star(S) = \tau$ has $\pi_0(S) \le \min(\beta\tau, 1)$. We proceed inductively: for $\tau \in [0, \frac{1}{2}]$, supposing (10) holds for some $k-1$, Lemma 1 yields

$$\begin{aligned} h_k(\tau) &\le \frac{1}{2} h_{k-1}((1-2\Phi)\tau) + \frac{1}{2} h_{k-1}((1+2\Phi)\tau) \\ &\le \beta \cdot \frac{1}{2}\left(\sqrt{(1-2\Phi)\tau} + \sqrt{(1+2\Phi)\tau}\right)\left(1 - \frac{\Phi^2}{2}\right)^{k-1} \\ &\le \beta \min\left(\sqrt{\tau}, \sqrt{1-\tau}\right)\left(1 - \frac{\Phi^2}{2}\right)^k, \end{aligned}$$

where we used that $\sqrt{1-x} + \sqrt{1+x} \le 2(1 - \frac{x^2}{2})$.[3] Finally, the case $\tau \in [\frac{1}{2}, 1]$ is handled analogously using the other conclusion in Lemma 1, establishing (10) as desired. □

The most appropriate comparison to Corollary 1 is Corollary 2, Part XIV, the analogous result we showed in the discrete setting. Compared to its discrete counterpart, Corollary 1 incurs a logarithmic dependence on the warmness $\beta$ rather than $(\min_{i \in [d]} \boldsymbol{\pi}^\star_i)^{-1}$. This latter quantity is typically meaningless in the continuous setting when the target density has no atoms.

Recall that Corollary 2, Part XIV was actually established in two steps: we first assumed a spectral gap on the Markov chain transition operator, and then used Cheeger's inequality to lower bound the spectral gap as a function of the conductance. There is an analogous definition in the continuous setting, where one defines the "spectral gap" to be the decay of the variance of the relative density $\frac{\pi_k}{\pi^\star}$, which will be formalized in a following lecture. To introduce these ideas, we mention a few useful definitions here and compare them. First, the total variation convergence result in Corollary 1 can be slightly modified to directly show that

$$\chi^2(\pi_k \| \pi^\star) := \mathbb{E}_{\pi^\star}\left[\left(\frac{\pi_k}{\pi^\star} - 1\right)^2\right] = \mathrm{Var}_{\pi^\star}\left[\frac{\pi_k}{\pi^\star}\right]$$

also decays by a factor of $\Omega(\Phi^2)$ in each iteration (we state such a result shortly). Here, $\chi^2(\mu\|\pi) := \mathrm{Var}_\pi[\frac{\mu}{\pi}]$ is the *chi-squared divergence*, which satisfies the inequality $\chi^2(\mu\|\pi) \ge D_{\mathrm{KL}}(\mu\|\pi) := \mathbb{E}_\pi[\frac{\mu}{\pi} \log \frac{\mu}{\pi}]$,[4] where $D_{\mathrm{KL}}(\mu\|\pi)$ is the KL divergence. Importantly, the continuous analog of Pinsker's inequality (see Lemma 7 and Remark 5, Part III), which states that

$$D_{\mathrm{KL}}(\mu\|\pi) \ge \frac{1}{2}\left(\int |\mu(\mathbf{x}) - \pi(\mathbf{x})| \mathrm{d}\mathbf{x}\right)^2 = 2 D_{\mathrm{TV}}(\mu, \pi)^2,$$

lets us transfer $\chi^2$ bounds to $D_{\mathrm{TV}}$ bounds. Further, under a $\beta$-warm start it is simple to see that $\chi^2(\pi_0\|\pi^\star) \le \beta^2$, so the aforementioned variance decay result (i.e., $\chi^2(\pi_k\|\pi^\star)$ falls by $\Omega(\Phi^2)$) recovers Corollary 1 up to constant factors, and yields a stronger bound in general.

One shortcoming of Corollary 1 is that in continuous settings, the warmness $\beta$ is typically exponential in the dimension. As a simple example, using the uniform distribution on $\mathbb{B}(\mathbf{0}_d, 1)$ as a warm start for the uniform distribution on $\mathbb{B}(\mathbf{0}_d, 2)$ yields warmness parameter

$$\beta = \frac{\mathrm{Vol}\left(\mathbb{B}(\mathbf{0}_d, 2)\right)}{\mathrm{Vol}\left(\mathbb{B}(\mathbf{0}_d, 1)\right)} = 2^d.$$

This phenomenon is quite general when scale parameters on the target distribution cannot be estimated to high precision, and so extraneous $\log \beta$ factors in mixing times result in dimension-dependent overheads. A natural alternative approach is to directly show that the KL divergence decays at a linear rate, because the initial KL divergence for a $\beta$-warm start $\pi_0$ is bounded by

$$D_{\mathrm{KL}}\left(\pi_0 \| \pi^\star\right) = \mathbb{E}_{\pi^\star}\left[\frac{\pi_0}{\pi^\star} \log \frac{\pi_0}{\pi^\star}\right] = \mathbb{E}_{\pi_0}\left[\log \frac{\pi_0}{\pi^\star}\right] \le \log \beta.$$

[3] The choice of the potential function $\sqrt{\cdot}$ in the proof is not particularly important. The most important property is that we use a concave potential function, so that the separation of $\tau$ values in Lemma 1 induces negative drift. In particular, a first-order Taylor approximation of $\sqrt{\cdot}$ provides intuition for the calculation used here; the first-order terms cancel, and the second-order terms induce a negative quadratic in $\Phi$ by concavity.

[4] One way to see this is by combining the facts that $\chi^2(\mu\|\pi)$ dominates the Rényi divergence of order 2, $D_{\mathrm{KL}}$ is the Rényi divergence of order 1, and Rényi divergences are monotone in the order.

Therefore, establishing this type of *entropy decay* (rather than variance decay) results in mixing time overheads of $\log\log\beta$, which is $O(\log d)$ rather than $O(d)$ when $\beta = \exp(\Theta(d))$. These types of entropy decay bounds, which are called modified log-Sobolev inequalities in the literature (see [BT03]), are typically harder to come by than the simpler spectral gap bounds which imply variance decay, though we will see some examples of entropy decay later in the course.

Finally, we mention an extension of Corollary 1 which we find useful in applications. In particular, the following extension due to [CDWY20] (building upon *average conductance* frameworks from [LK99, GMT06]) handles situations where uniform conductance bounds do not hold due to small sets with poor behavior, and also sometimes removes extraneous $\log\beta$ factors.

**Proposition 1** (Lemma 3, [CDWY20])**.** *Let $\epsilon \in (0,1)$, and consider a random walk on $\mathbb{R}^d$ with stationary distribution $\pi^\star$ and lazy, reversible transition distributions $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$, initialized at $\pi_0$ and with density $\pi_k$ after $k$ steps for all $k \in \mathbb{N}$. Let $\pi_0$ be $\beta$-warm with respect to $\pi^\star$, and suppose $\Omega \subseteq \mathbb{R}^d$ has $m := \pi^\star(\Omega) \geq 1 - \frac{\epsilon^2}{3\beta^2}$. For all $\tau \in (0, \frac{m}{2}]$, define (following notation (6))*

$$\Phi_\Omega(\tau) := \inf_{\substack{S\subseteq\Omega \\ \pi^\star(S)\leq\tau}} \frac{Q(S)}{\pi^\star(S)}. \tag{11}$$

*Then,*

$$\chi^2\left(\pi_k \| \pi^\star\right) \leq \epsilon, \text{ if } k \geq \int_{\frac{4}{\beta}}^{\frac{m}{2}} \frac{16}{\tau\Phi_\Omega(\tau)^2}\mathrm{d}\tau + \frac{32}{\Phi_\Omega(\frac{m}{2})^2}\log\left(\frac{4}{\epsilon}\right).$$

We pause to interpret Proposition 1. First, because $\Phi_\Omega(\tau)$ is decreasing as $\tau$ increases because the definition (11) includes more sets, in the special case $\Omega = \mathbb{R}^d$ and $m = 1$, so $\Phi = \Phi_\Omega(\frac{m}{2})$ is the standard conductance (Definition 1), the bound in Proposition 1 is never worse than

$$32\int_{\frac{4}{\beta}}^{\frac{4}{\epsilon}} \frac{1}{\tau\Phi^2}\mathrm{d}\tau = \Theta\left(\frac{1}{\Phi^2}\log\left(\frac{\beta}{\epsilon}\right)\right).$$

This matches Corollary 1 up to a constant factor, and bounds the $\chi^2$ divergence rather than the smaller $D_{\mathrm{TV}}$, giving a stronger bound. Further, Proposition 1 allows for excluding a small set with poor behavior, which can be useful on random walks in $\mathbb{R}^d$ where "far away" points are difficult to control, but also are unlikely under $\pi^\star$. Finally, if one can prove a bound on $\Phi_\Omega(\tau)$ which improves with $\tau$, Proposition 1 can lead to significantly sharper mixing time estimates. For example, suppose

$$\Phi_\Omega(\tau) = \Omega\left(\Phi\sqrt{\log\frac{1}{\tau}}\right),$$

which is the case for Gaussian densities (and more generally, distributions with strongly convex negative log-densities, see e.g., Theorem 2.7 in [Led99]). Then, the bound in Proposition 1 reads[5]

$$k \gtrsim \int_{\frac{4}{\beta}}^{\frac{m}{2}} \frac{1}{\tau\log(\frac{1}{\tau})\cdot\Phi^2}\mathrm{d}\tau + \frac{1}{\Phi^2}\log\frac{1}{\epsilon} \gtrsim \frac{1}{\Phi^2}\log\log\beta + \log\beta + \frac{1}{\Phi^2}\log\frac{1}{\epsilon},$$

where we used that the antiderivative of $(\tau\log(\frac{1}{\tau}))^{-1}$ is $\log\log\frac{1}{\tau}$. This can improve mixing times by dimension-dependent factors, when initialized at a "cold start" of quality $\beta = \exp(\Theta(d))$.

## 2.3 Boosting

In this section, we give a generic reduction inspired by the boosting result shown in Corollary 1, Part XIV for discrete Markov chains. Specifically, we show that if a Markov chain requires $T_{\mathrm{mix}}$ iterations to achieve constant total variation distance, where $T_{\mathrm{mix}}$ has a logarithmic overhead in the initial warmness, then it mixes to $\epsilon$ total variation distance in roughly $T_{\mathrm{mix}}\log\frac{1}{\epsilon}$ iterations. As in Corollary 1, Part XIV, this section will not make use of reversibility, so it applies generically.

**kjtian:** This section is Homework IV, Problems 1 and 5. I will update when it is due.

[5]The additive $\log\beta$ term arises because $\Phi_\Omega(\tau) \leq 1$ for all $\tau$.

## 3 Conductance via isoperimetry and transition overlaps

We established in Corollary 1 and Proposition 1 that to bound the mixing time (in $D_{\text{TV}}$ or $\chi^2$, respectively), it suffices to lower bound the conductance. In this section, we give a geometric argument by [LS93] which reduces conductance bounds to bounding two quantities which are typically simpler to reason about in applications. We define below the first such quantity.

**Definition 4** (Isoperimetry)**.** *Let $\pi^\star$ be a density on $\mathbb{R}^d$, and let $\text{m} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be a metric. We say that $\pi^\star$ has* isoperimetric constant *$\psi$ if, for any partition[6] $S_1, S_2, S_3$ of $\mathbb{R}^d$,*

$$\frac{\pi^\star(S_3)}{\text{m}(S_1, S_2)\pi^\star(S_1)\pi^\star(S_2)} \geq \psi,$$

*where we let $\text{m}(S, T) := \min_{\mathbf{x}\in S, \mathbf{y}\in T} \text{m}(\mathbf{x}, \mathbf{y})$ for $S, T \subseteq \mathbb{R}^d$.*

Importantly, the isoperimetric constant in Definition 4 is a function of only the target distribution $\pi^\star$, and has no dependence on the transition distributions $\{\mathcal{T}_\mathbf{x}\}_{\mathbf{x}\in\mathbb{R}^d}$. This definition isolates a geometric property of the distribution, which can be reasoned about separately. To gain some intuition for Definition 4, note that when $S_3$ is taken to be an infinitesimally small strip on the boundary between $S_1$ and $S_2$, then $\frac{\pi^\star(S_3)}{\text{m}(S_1,S_2)}$ is essentially the surface area of the boundary. Therefore, in this case Definition 4 reduces to a lower bound on surface area-to-volume ratios.

To provide some intuition on isoperimetry, we prove a classical fact in high-dimensional geometry.

**Proposition 2.** *Among all compact $\mathcal{K} \subseteq \mathbb{R}^d$ with a fixed volume $\text{Vol}(\mathcal{K})$, the minimum surface area is achieved when $\mathcal{K}$ is a ball of radius $(\frac{\text{Vol}(\mathcal{K})}{V_d})^{\frac{1}{d}}$.*

*Proof.* It is standard that another way to define the surface area of a set $\mathcal{K} \subseteq \mathbb{R}^d$ is

$$\text{Vol}(\partial\mathcal{K}) = \lim_{\epsilon\to 0^+} \frac{\text{Vol}(\mathcal{K} \oplus \mathbb{B}(\mathbf{0}_d, \epsilon)) - \text{Vol}(\mathcal{K})}{\epsilon},$$

where $\oplus$ is the Minkowski sum. The Brunn-Minkowski inequality (Theorem 5, Part I) then shows

$$\begin{aligned} \text{Vol}(\mathcal{K} \oplus \mathbb{B}(\epsilon))^{\frac{1}{d}} &\geq \text{Vol}(\mathcal{K})^{\frac{1}{d}} + \text{Vol}(\mathbb{B}(\epsilon))^{\frac{1}{d}} \\ \implies \lim_{\epsilon\to 0^+} \frac{\text{Vol}(\mathcal{K} \oplus \mathbb{B}(\epsilon)) - \text{Vol}(\mathcal{K})}{\epsilon} &\geq \lim_{\epsilon\to 0^+} \frac{(\text{Vol}(\mathcal{K})^{\frac{1}{d}} + \text{Vol}(\mathbb{B}(\epsilon))^{\frac{1}{d}})^d - \text{Vol}(\mathcal{K})}{\epsilon} \\ &= d \cdot \text{Vol}(\mathcal{K})^{\frac{d-1}{d}} \text{Vol}(\mathbb{B}(1))^{\frac{1}{d}}. \end{aligned}$$

It is well-known that $\text{Vol}(\partial\mathbb{B}(1)) = d \cdot \text{Vol}(\mathbb{B}(1))$, which can be derived via

$$\text{Vol}(\mathbb{B}(1)) = \int_0^1 r^{d-1}\text{Vol}(\partial\mathbb{B}(1))\text{d}r = \frac{1}{d}\text{Vol}(\partial\mathbb{B}(1)). \tag{12}$$

Hence, for $r := (\frac{\text{Vol}(\mathcal{K})}{V_d})^{\frac{1}{d}}$,

$$\frac{\text{Vol}(\partial\mathcal{K})}{\text{Vol}(\mathcal{K})} \geq d \cdot \left(\frac{V_d}{\text{Vol}(\mathcal{K})}\right)^{\frac{1}{d}} = \frac{\text{Vol}\left(\partial\mathbb{B}(r)\right)}{\text{Vol}\left(\mathbb{B}(r)\right)}.$$

□

The second quantity used in our conductance lower bound framework can informally be thought of as an overlap bound between "nearby" transition distributions, where the notion of nearby is compatible with Definition 4. Specifically, we assume that

$$D_{\text{TV}}\left(\mathcal{T}_\mathbf{x}, \mathcal{T}_\mathbf{y}\right) \leq \frac{1}{2} \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d \text{ with } \text{m}(\mathbf{x}, \mathbf{y}) \leq \Delta. \tag{13}$$

The bound (13) is a local property of transition distributions, because it only needs to hold for pairs of nearby points $(\mathbf{x}, \mathbf{y})$. We next give a representative result which shows how to lift local bounds of the form (13) and establish conductance lower bounds, a global property, via isoperimetry.

[6] That is, $\bigcup_{i\in[3]} S_i = \mathbb{R}^d$ and $S_i \cap S_j = \emptyset$ for all $i \neq j$.

**Proposition 3.** *Consider a random walk on* $\mathbb{R}^d$ *with stationary distribution* $\pi^\star$ *and lazy, reversible transition distributions* $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$*. Suppose that* $\pi^\star$ *has isoperimetric constant* $\psi$*, and that* (13) *holds. Then the conductance of the random walk satisfies*

$$\Phi \geq \min\left(\frac{1}{8}, \frac{\Delta\psi}{64}\right).$$

*Proof.* Choose an arbitrary set $S \subseteq \mathbb{R}^d$ with $0 < s := \pi^\star(S) \leq \frac{1}{2}$, so our goal is to lower bound $\frac{Q(S)}{s}$ by the stated quantity. Define the following sets:

$$S_1 := \left\{\mathbf{x} \in S \mid \mathcal{T}_{\mathbf{x}}(S^c) < \frac{1}{4}\right\},\ S_2 := \left\{\mathbf{x} \in S^c \mid \mathcal{T}_{\mathbf{x}}(S) < \frac{1}{4}\right\},\ S_3 := \mathbb{R}^d \setminus (S_1 \cup S_2).$$

Intuitively, $S_1$ and $S_2$ are the points deep in $S$ and $S^c$ respectively, since they are unlikely to cross after one random walk step. Note that for $\mathbf{x} \in S_1$, $\mathbf{y} \in S_2$, we have $D_{\mathrm{TV}}(\mathbf{x}, \mathbf{y}) > \frac{1}{2}$, and therefore $\mathrm{m}(\mathbf{x}, \mathbf{y}) > \Delta$ due to the assumption (13). If $\pi^\star(S_1) < \frac{1}{2}\pi^\star(S)$, then

$$Q(S) \geq \int_{\mathbf{x}\in S_3\cap S} \mathcal{T}_{\mathbf{x}}(S^c)\pi^\star(\mathbf{x})\mathrm{d}\mathbf{x} \geq \frac{1}{4}\pi^\star(S_3 \cap S) \geq \frac{s}{8}.$$

Similarly, if $\pi^\star(S_2) < \frac{1}{2}\pi^\star(S^c)$, then

$$Q(S) \geq \int_{\mathbf{x}\in S_3\cap S^c} \mathcal{T}_{\mathbf{x}}(S)\pi^\star(\mathbf{x})\mathrm{d}\mathbf{x} \geq \frac{1}{4}\pi^\star(S_3 \cap S^c) \geq \frac{1-s}{8} \geq \frac{s}{8}.$$

Finally, if both $\pi^\star(S_1) \geq \frac{1}{2}\pi^\star(S)$ and $\pi^\star(S_2) \geq \frac{1}{2}\pi^\star(S^c)$, we have

$$\begin{aligned} Q(S) &= \frac{1}{2}\left(\int_{\mathbf{x}\in S} \mathcal{T}_{\mathbf{x}}(S^c)\pi^\star(\mathbf{x})\mathrm{d}\mathbf{x} + \int_{\mathbf{x}\in S^c} \mathcal{T}_{\mathbf{x}}(S)\pi^\star(\mathbf{x})\mathrm{d}\mathbf{x}\right) \\ &\geq \frac{1}{8}\int_{\mathbf{x}\in S_3} \pi^\star(\mathbf{x})\mathrm{d}\mathbf{x} = \frac{\pi^\star(S_3)}{8} \geq \frac{\Delta\psi\pi^\star(S_1)\pi^\star(S_2)}{8} \geq \frac{\Delta\psi s}{64}. \end{aligned}$$

In the second-to-last inequality, we used our earlier conclusion $\mathrm{m}(S_1, S_2) \geq \Delta$, and the definition of $\psi$; in the last inequality, we used our assumptions $\pi^\star(S_1) \geq \frac{1}{2}\pi^\star(S)$ and $\pi^\star(S_2) \geq \frac{1}{2}\pi^\star(S^c) \geq \frac{1}{4}$. □

There are various extensions to the proof strategy in Proposition 3, for example handling sets of smaller measure or excluding ill-behaved regions, as suggested by Proposition 1; we will require one such extension in Section 5. Nonetheless, the overall strategy of lower bounding an appropriate notion of isoperimetry (potentially restricted to specific subsets) and establishing the overlap of nearby transition distributions remains as a common theme. In the rest of this lecture, we give techniques for reasoning about these quantities, specialized to uniform sampling of convex bodies.

# 4 Localization

Up to this point, our development has not made any structural assumption about the target distribution $\pi^\star$ (or the transition distributions $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$). We now provide an isoperimetric inequality which is specific to logconcave densities $\pi^\star$, in order to introduce a powerful tool: the *localization lemma*, first shown by [LS93, KLS95]. The localization lemma states, at a high level, that the boundary points of the set of logconcave densities satisfying one linear inequality are one-dimensional and logaffine. This is extremely useful when maximizing linear functions of logconcave densities subject to one constraint, since it means we only need to establish the inequality for one-dimensional, logaffine densities. To make these statements precise, we give a convenient formulation of the localization lemma from [FG04], building upon [LS93, KLS95].[7]

**Proposition 4** (Theorem 1, [FG04]). *Let* $\mathcal{K} \subset \mathbb{R}^d$ *be compact and convex, and let* $f : \mathcal{K} \to \mathbb{R}$ *be upper semicontinuous. Let* $\mathcal{P}(f)$ *be the set of logconcave densities* $\pi : \mathcal{K} \to \mathbb{R}_{\geq 0}$ *satisfying* $\int_{\mathbf{x}\in\mathcal{K}} f(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x} \geq 0$*. All boundary points*[8] *of* $\mathrm{Conv}(\mathcal{P}(f))$ *satisfy one of the following.*

[7] The statement in [FG04] is more general than Proposition 4, as it gives a complete characterization of extreme points, allows for multiple constraints, and extends to $s$-concave distributions, which generalizes logconcavity.

[8] Recall from Lemma 3, Part I that the boundary points of a convex set $S$ are points $\mathbf{x} \in S$ such that there do not exist $\mathbf{x}', \mathbf{x}'' \in S$ and $\lambda \in (0, 1)$ such that $\mathbf{x} = (1-\lambda)\mathbf{x}' + \lambda\mathbf{x}''$.

1. *$\pi$ is a Dirac measure at a point $\mathbf{x} \in \mathcal{K}$ where $f(\mathbf{x}) \geq 0$.*

2. *$\pi$ is a logaffine distribution satisfying $\int f(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x} = 0$, and is supported on a one-dimensional subspace, i.e., for some $\gamma \in \mathbb{R}$ and $\mathbf{a}, \mathbf{b} \in \mathcal{K}$, we have*

$$\pi(\mathbf{x}) \propto \begin{cases} \exp(\gamma t) & \mathbf{x} = (1-t)\mathbf{a} + t\mathbf{b} \text{ for some } t \in [0,1] \\ 0 & \text{otherwise} \end{cases}. \tag{14}$$

Densities of the form (14) are often called exponential needles, because they are truncated exponential distributions over one-dimensional "needles" within $\mathcal{K}$. We first briefly sketch the proof of Proposition 4, deferring more details to [FG04]. Suppose for the sake of contradiction that there is a boundary point $\pi$ of $\mathrm{Conv}(\mathcal{P}(f))$ with $\dim S \geq 2$, where $S$ is the least affine subspace containing the support of $\pi$. Let $\mathbf{x}_0 \in S$ and let $E \subseteq S$ be an arbitrary two-dimensional subspace, such that $\mathbf{x}_0 \oplus E \subseteq S$. Let $C(E)$ be the unit circle in $E$, and for any $\mathbf{u} \in C(E)$ denote

$$H_{\mathbf{u}}^{+} := \{\mathbf{x} \in S \mid \langle \mathbf{x} - \mathbf{x}_0, \mathbf{u} \rangle > 0\}, \; H_{\mathbf{u}}^{-} := \{\mathbf{x} \in S \mid \langle \mathbf{x} - \mathbf{x}_0, \mathbf{u} \rangle < 0\}.$$

By choosing $\mathbf{x}_0$ in the relative interior of the support of $\pi$, we can ensure that $\pi(H_{\mathbf{u}}^{+}), \pi(H_{\mathbf{u}}^{-}) > 0$ for all $\mathbf{u} \in C(E)$. Next, define $\phi : C(E) \to \mathbb{R}$ by $\phi(\mathbf{u}) := \int_{x \in H_{\mathbf{u}}^{+}} f(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x} - \frac{1}{2}\int f(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}$. Note that $\phi$ is a continuous function of $\mathbf{u}$, and further $\phi(\mathbf{u}) = -\phi(-\mathbf{u})$ for all $\mathbf{u} \in C(E)$.[9] The intermediate value theorem then gives $\mathbf{u} \in C(E)$ with $\phi(\mathbf{u}) = 0$, so that $\pi^{+}$ and $\pi^{-}$, the restrictions of $\pi$ to $H_{\mathbf{u}}^{+}$ and $H_{\mathbf{u}}^{-}$, have $\int f(\mathbf{x})\pi^{+}(\mathbf{x})\mathrm{d}\mathbf{x} \geq 0$ and $\int f(\mathbf{x})\pi^{-}(\mathbf{x})\mathrm{d}\mathbf{x} \geq 0$. Finally, we can write $\pi$ as a nontrivial convex combination of $\pi^{+}$ and $\pi^{-}$, so $\pi$ is not a boundary point of $\mathrm{Conv}(\mathcal{P}(f))$.

Similar arguments are then used to decompose any one-dimensional elements of $\mathcal{P}(f)$ which do not satisfy $\int f\pi\mathrm{d}\mathbf{x} = 0$, or are not logaffine, so they cannot be boundary points. For the former property, one can split the measure along the line at any point giving half the value of $\int f\pi\mathrm{d}\mathbf{x}$. For the latter property, the "largest" possible extension of a logconcave function is logaffine, just as extremal examples of convex functions are affine; one can use this strategy to obtain a logaffine function which dominates any logconcave function, and split it accordingly. For more details, we refer the reader to the (fairly short) proof of Proposition 4 in [FG04].

We note that these proofs use relatively little about the structure of logconcave functions; for example, the one-dimensional argument only used absolute continuity and that restricting to convex sets preserves logconcavity. Due to this, analogous arguments can be extended to broader families of distributions; see e.g., Lemma 1 of [GLL+23] for an extension to strongly logconcave densities.

We now give an example use of Proposition 4, in proving a variant of the localization lemma which is often easier to apply. The key idea, as mentioned previously, is that if we want to prove a linear inequality for all logconcave functions in some $\mathcal{P}(f)$, it suffices to prove the inequality for point masses and one-dimensional logaffine functions, by using Lemma 3, Part I.

**Lemma 2.** *Let $f_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, $f_2 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be upper semicontinuous and let $f_3 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, $f_4 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be lower semicontinuous. Then the following two statements are equivalent.*

1. *For every logconcave density $\pi : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$,*

$$\left(\int f_1(x)\pi(x)\mathrm{d}x\right)\left(\int f_2(x)\pi(x)\mathrm{d}x\right) \leq \left(\int f_3(x)\pi(x)\mathrm{d}x\right)\left(\int f_4(x)\pi(x)\mathrm{d}x\right).$$

2. *For all $\mathbf{x} \in \mathbb{R}^d$, $f_1(\mathbf{x})f_2(\mathbf{x}) \leq f_3(\mathbf{x})f_4(\mathbf{x})$, and for every $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and $\gamma \in \mathbb{R}$,*

$$\left(\int_0^1 f_1((1-t)\mathbf{a} + t\mathbf{b})\exp(\gamma t)\mathrm{d}t\right)\left(\int_0^1 f_1((1-t)\mathbf{a} + t\mathbf{b})\exp(\gamma t)\mathrm{d}t\right)$$
$$\leq \left(\int_0^1 f_3((1-t)\mathbf{a} + t\mathbf{b})\exp(\gamma t)\mathrm{d}t\right)\left(\int_0^1 f_4((1-t)\mathbf{a} + t\mathbf{b})\exp(\gamma t)\mathrm{d}t\right).$$

[9] To see this, we used that $\pi(\mathbb{R}^d \setminus (H_{\mathbf{u}}^{+} \cup H_{\mathbf{u}}^{-})) = 0$. This is a nontrivial fact which follows from a characterization of logconcave functions in [Bor75], showing that logconcave functions are absolutely continuous with respect to the Lebesgue measure on the least affine subspace containing their support. Therefore, they place zero measure on any lower-dimensional subspace, so we can exclude the boundary $\mathbb{R}^d \setminus (H_{\mathbf{u}}^{+} \cup H_{\mathbf{u}}^{-})$ from consideration.

*Proof.* It is obvious that Item 1 implies Item 2, since truncated logaffine functions are also logconcave, so we focus on showing Item 2 implies Item 1. By a limiting argument, it suffices to consider the case where $\pi$ is compactly supported on $\mathcal{K} \subset \mathbb{R}^d$, and $\int f_3(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x} > 0$. Then, define

$$f = f_1 - \left(\frac{\int f_1(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}{\int f_3(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}\right) f_3,\ g = \left(\frac{\int f_1(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}{\int f_3(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}\right) f_2 - f_4.$$

The maximum value of $\int g(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x}$ for $\mu \in \mathrm{Conv}(\mathcal{P}(f))$ is achieved by either a Dirac measure or a one-dimensional logaffine function, by Proposition 4. The fact that $\mu \in \mathrm{Conv}(\mathcal{P}(f))$ implies that

$$\begin{aligned}\int f_1(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x} - \left(\frac{\int f_1(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}{\int f_3(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}\right)\int f_3(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x} \geq 0 \\ \implies \frac{\int f_1(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}{\int f_3(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}} \leq \frac{\int f_1(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x}}{\int f_3(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x}}.\end{aligned} \tag{15}$$

Thus, the fact that $\pi \in \mathcal{P}(f)$ by definition of $f$ shows the desired

$$\begin{aligned}&\left(\frac{\int f_1(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}{\int f_3(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}\right)\int f_2(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x} - \int f_4(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x} \\ &\leq \left(\frac{\int f_1(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}{\int f_3(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}}\right)\int f_2(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x} - \int f_4(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x} \\ &\leq \left(\frac{\int f_1(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x}}{\int f_3(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x}}\right)\int f_2(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x} - \int f_4(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x} \leq 0,\end{aligned}$$

where the second inequality used (15), and the last used the assumption in Item 2. ☐

To demonstrate the utility of Lemma 2, we use it to give a bound on the isoperimetric constant (Definition 4) for any logconcave density $\pi$, parameterized by the average distance of points to its mean, which we denote (patterning off notation in Theorem 2, Part I) by

$$\bar{\mathbf{x}}_\pi := \mathbb{E}_{\mathbf{x}\sim\pi}[\mathbf{x}] = \int \mathbf{x}\pi(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{16}$$

**Lemma 3.** *Let $\pi : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be a logconcave density, and let $S_1, S_2, S_3$ be a partition of $\mathbb{R}^d$ with $\min_{\mathbf{x}\in S_1, \mathbf{y}\in S_2} \|\mathbf{x} - \mathbf{y}\|_2 \geq \Delta$. Then,*

$$\pi(S_1)\pi(S_2) \leq \frac{\mathbb{E}_{\mathbf{x}\sim\pi}[\|\mathbf{x} - \bar{\mathbf{x}}_\pi\|_2]}{\Delta \log 2} \cdot \pi(S_3).$$

*Proof.* Let $f_i(\mathbf{x}) := \mathbb{I}_{\mathbf{x}\in S_i}$ for $i \in [3]$ be the indicator functions associated to the partition, and let $f_4(\mathbf{x}) := \frac{1}{\Delta \log 2}\|\mathbf{x} - \bar{\mathbf{x}}_\pi\|_2$. Clearly for all $\mathbf{x} \in \mathbb{R}^d$, we have $f_1(\mathbf{x})f_2(\mathbf{x}) = 0 \leq f_3(\mathbf{x})f_4(\mathbf{x})$. Thus, to obtain the conclusion, Lemma 2 states that it suffices to show that for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and $\gamma \in \mathbb{R}$,

$$\begin{aligned}&\left(\int_0^1 \mathbb{I}_{(1-t)\mathbf{a}+t\mathbf{b}\in S_1}\exp(\gamma t)\mathrm{d}t\right)\left(\int_0^1 \mathbb{I}_{(1-t)\mathbf{a}+t\mathbf{b}\in S_2}\exp(\gamma t)\mathrm{d}t\right) \\ &\leq \frac{1}{\Delta \log 2}\left(\int_0^1 \mathbb{I}_{(1-t)\mathbf{a}+t\mathbf{b}\in S_3}\exp(\gamma t)\mathrm{d}t\right)\left(\int_0^1 \|(1-t)\mathbf{a} + t\mathbf{b} - \bar{\mathbf{x}}_\pi\|_2 \exp(\gamma t)\mathrm{d}t\right).\end{aligned}$$

We next observe that changing $\bar{\mathbf{x}}_\pi$ to its projection on the line between $\mathbf{a}, \mathbf{b}$, i.e., some point $(1-u)\mathbf{a} + u\mathbf{b}$ for $u \in [0, 1]$, can only decrease the right-hand side above. Moreover, we redefine

$$I_i := \{t \in [0, 1] \mid (1-t)\mathbf{a} + t\mathbf{b} \in S_i\} \text{ for all } i \in [3],$$

so that $\min_{t\in I_1, t'\in I_2} \|((1-t)\mathbf{a} + t\mathbf{b}) - ((1-t')\mathbf{a} + t'\mathbf{b})\|_2 = \min_{t\in I_1, t'\in I_2} |t - t'| \|\mathbf{a} - \mathbf{b}\|_2 \geq \Delta$. Therefore, rescaling $\Delta$ by a factor of $\|\mathbf{a} - \mathbf{b}\|_2$, our problem reduces to showing that for any partition $I_1, I_2, I_3$ of $[0, 1]$ with $\min_{t\in I_1, t'\in I_2} |t - t'| \geq \Delta$, and for any $u \in [0, 1]$,

$$\left(\int_{t\in I_1} \exp(\gamma t)\mathrm{d}t\right)\left(\int_{t\in I_2} \exp(\gamma t)\mathrm{d}t\right) \leq \frac{1}{\Delta \log 2}\left(\int_{t\in I_3} \exp(\gamma t)\mathrm{d}t\right)\left(\int_0^1 |u - t| \exp(\gamma t)\mathrm{d}t\right). \tag{17}$$

In the case when $I_1$, $I_2$, and $I_3$ are single intervals (where the length of $I_3$ is at least $\Delta$), the proof of (17) is carried out using elementary arguments in Theorem 5.2, [KLS95], which we omit here. We instead discuss how to reduce to this particular case for general partitions, which is a core part of localization arguments. First, we may assume that $I_3$ is open by moving a measure zero set of endpoints which does not affect the inequality. Next, every interval in $I_3$ has length at least $\Delta$, since any shorter interval must have both endpoints in $I_1$ or $I_2$, and hence can be moved to the corresponding partition piece which only makes (17) tighter. Under these simplifications, we can write $I_3 = \bigcup_{j\in[k]}(t_j, u_j)$ for a finite $k$. Applying (17) to each interval $(t_j, u_j)$ and summing,

$$\sum_{j\in[k]} \left(\int_0^{t_j} \exp(\gamma t)\mathrm{d}t\right)\left(\int_{u_j}^1 \exp(\gamma t)\mathrm{d}t\right) \le \frac{1}{\Delta \log 2}\left(\int_{t\in I_3} \exp(\gamma t)\mathrm{d}t\right)\left(\int_0^1 |u-t|\exp(\gamma t)\mathrm{d}t\right).$$

The claim follows since every pair of intervals in $I_1$ and $I_2$ are separated by some $(t_j, u_j)$, so

$$\sum_{j\in[k]} \left(\int_0^{t_j} \exp(\gamma t)\mathrm{d}t\right)\left(\int_{u_j}^1 \exp(\gamma t)\mathrm{d}t\right) \ge \left(\int_{t\in I_1} \exp(\gamma t)\mathrm{d}t\right)\left(\int_{t\in I_2} \exp(\gamma t)\mathrm{d}t\right).$$

□

We mention that the proof of Lemma 3 uses nothing specific about $\bar{\mathbf{x}}_\pi$, so it can be replaced with an arbitrary point. One reason why the bound with $\bar{\mathbf{x}}_\pi$ in particular is useful is because

$$\mathbb{E}_{\mathbf{x}\sim\pi}\left[\|\mathbf{x}-\bar{\mathbf{x}}_\pi\|_2\right] \le \sqrt{\mathbb{E}_{\mathbf{x}\sim\pi}\left[\|\mathbf{x}-\bar{\mathbf{x}}_\pi\|_2^2\right]} = \sqrt{\mathrm{Tr}\Bigg(\underbrace{\mathbb{E}_{\mathbf{x}\sim\pi}\left[(\mathbf{x}-\bar{\mathbf{x}}_\pi)(\mathbf{x}-\bar{\mathbf{x}}_\pi)^\top\right]}_{:=\mathrm{Cov}_\pi}\Bigg)}. \tag{18}$$

In other words, Lemma 3 proves that the isoperimetric constant (Definition 4) of any logconcave density $\pi : \mathbb{R}^d \to \mathbb{R}_{\ge 0}$ with respect to the $\ell_2$ norm is lower bounded by

$$\frac{\log 2}{\sqrt{\mathrm{Tr}(\mathrm{Cov}_\pi)}}.$$

This result was first established by [KLS95], and shows that e.g. if $\pi$ is isotropic (meaning $\mathrm{Cov}_\pi = \mathbf{I}_d$), the isoperimetric constant is $\Omega(d^{-\frac{1}{2}})$. The authors of [KLS95] conjectured that this bound could be improved to $\Omega(\|\mathrm{Cov}_\pi\|_{\mathrm{op}}^{-1})$, which has since come to be called the KLS conjecture, a deep result in probability theory which has intimate connections with various other results in mathematics (see [LV17b] for an extended discussion of these connections).

Recently, there has been tremendous progress on the KLS conjecture. First, [Eld13] developed a technique known as *stochastic localization*, which simulates the localization process in Proposition 4 which decomposes a measure into low-dimensional pieces, not by bisection, but rather by gradually adding a Gaussian component. We cover this technique in detail in a future lecture, as it closely-related to modern techniques for training diffusion models [Mon23].

The reason this technique is useful in the context of isoperimetry is because the KLS conjecture is much simpler to establish for Gaussians, and indeed [LV17a] built upon [Eld13] to show that the isoperimetric constant of logconcave $\pi$ is bounded by $\Omega(\|\mathrm{Cov}_\pi\|_{\mathrm{F}}^{-1})$, which is $\Omega(d^{-\frac{1}{4}})$ in the isotropic case. By bootstrapping the arguments of [Eld13, LV17a], a breakthrough was made by [Che21], who finally proved that the KLS conjecture is true up to a $d^{o(1)}$ factor. We state here the current state-of-the-art bound on the isoperimetric constant of logconcave distributions.

**Proposition 5** (Theorem 1.2, [Kla23]). *There is a universal constant $C > 0$ such that for any logconcave density $\pi : \mathbb{R}^d \to \mathbb{R}_{\ge 0}$, the isoperimetric constant $\psi$ of $\pi$ with respect to $\|\cdot\|_2$ satisfies*

$$\psi \ge \frac{1}{C\sqrt{\log d}} \cdot \frac{1}{\sqrt{\|\mathrm{Cov}_\pi\|_{\mathrm{op}}}}.$$

## 5 Ball walk analysis

In this section, we finally apply the results of Sections 2, 3, and 4 to a specific Markov chain for sampling from the uniform density over a convex body $\mathcal{K} \subseteq \mathbb{R}^d$. We fix the notation

$$\pi^\star(\mathbf{x}) := \frac{1}{\mathrm{Vol}(\mathcal{K})} \cdot \mathbb{I}_{\mathbf{x} \in \mathcal{K}} \tag{19}$$

throughout the section; by observation, $\pi^\star$ is logconcave because $\mathcal{K}$ is convex.

We primarily focus on providing intuition for how to adapt our machinery to this setting, under several simplifying assumptions, rather than giving all of the technical details (which can be found in the original source material). In later lectures, we discuss how to relax our simplifying assumptions and dramatically streamline the analysis, using more sophisticated tools.

Following the notation (16), (18), we begin by making the assumption that $\pi^\star$ is isotropic:

$$\mathrm{Cov}_{\pi^\star} := \mathbb{E}_{\mathbf{x}\sim\pi^\star}\left[(\mathbf{x}-\bar{\mathbf{x}}_\pi)(\mathbf{x}-\bar{\mathbf{x}}_\pi)^\top\right] = \mathbf{I}_d, \text{ where } \bar{\mathbf{x}}_{\pi^\star} := \mathbb{E}_{\mathbf{x}\sim\pi^\star}[\mathbf{x}]. \tag{20}$$

One convenient fact about isotropic convex bodies is that they are decently-approximated by balls.

**Fact 1** (Theorem 4.1, [KLS95])**.** *For any convex body $\mathcal{K} \subseteq \mathbb{R}^d$ satisfying (20), we have*

$$\mathbb{B}(1) \subseteq \mathcal{K} \subseteq \mathbb{B}(d+1).$$

We omit the proof of Fact 1 as it is somewhat tedious, deferring details to [KLS95]. Intuitively, the lower bound in Fact 1 holds because any one-dimensional projection being contained in an interval of length $o(1)$ would contradict the covariance being lower-bounded by 1 in all directions.

As a basic building block, we define the ball walk transition operators, denoted $\{\mathcal{T}_\mathbf{x}\}_{\mathbf{x}\in\mathcal{K}}$ throughout the section. The ball walk operator $\mathcal{T}_\mathbf{x}$ first returns $\mathbf{x}$ with probability $\frac{1}{2}$ (so it is lazy). Otherwise, it uniformly samples $\mathbf{y} \sim \mathbb{B}(\mathbf{x}, \eta)$ for a step size $\eta > 0$, and then updates $\mathbf{x}$ to $\mathbf{y}$ if $\mathbf{y} \in \mathcal{K}$, else staying at $\mathbf{x}$. This filter in the latter case preserves reversibility, as it is the result of a Metropolis filter (5) applied to the lazy proposal distributions $\mathcal{P}_\mathbf{x}$ which uniformly sample a point in $\mathbb{B}(\mathbf{x}, \eta)$. To make this definition simpler to state, we define the *local conductance* at $\mathbf{x}$,

$$\ell(\mathbf{x}) := \frac{1}{\mathrm{Vol}(\mathbb{B}(\mathbf{x},\eta))}\int_{\mathbf{y}\in\mathbb{B}(\mathbf{x},\eta)} \mathbb{I}_{\mathbf{y}\in\mathcal{K}} \mathrm{d}\mathbf{y} = \frac{\mathrm{Vol}\left(\mathbb{B}(\mathbf{x},\eta)\cap\mathcal{K}\right)}{\mathrm{Vol}(\mathbb{B}(\mathbf{x},\eta))} \tag{21}$$

to be the probability of accepting a step. Drawing $\mathbf{x}' \sim \mathcal{T}_\mathbf{x}$ can then be concisely stated as

$$\mathbf{x}' \leftarrow \begin{cases} \mathbf{x} & \text{with probability } 1 - \frac{1}{2}\ell(\mathbf{x}) \\ \text{a uniform draw from } \mathbb{B}(\mathbf{x},\eta)\cap\mathcal{K} & \text{with probability } \frac{1}{2}\ell(\mathbf{x}) \end{cases}. \tag{22}$$

Due to the way we set up the Metropolis-Hastings filter, the stationary distribution of the $\{\mathcal{T}_\mathbf{x}\}_{\mathbf{x}\in\mathbb{R}^d}$ is indeed $\pi^\star$. We finally mention one useful property of $\ell(\mathbf{x})$ which is used later.

**Lemma 4.** *The local conductance $\ell$ defined in (22) is logconcave.*

*Proof.* Logconcavity is preserved by multiplication by a constant, so it suffices to prove that $\int_{\mathbf{y}\in\mathbb{B}(\mathbf{x},\eta)} \mathbb{I}_{\mathbf{y}\in\mathcal{K}} \mathrm{d}\mathbf{y}$ is logconcave. This follows because it is the convolution of the logconcave indicator functions of $\mathbb{B}(\mathbf{0}_d, \eta)$ and $\mathcal{K}$, so the claim follows from Corollary 2, Part I. □

### 5.1 Overlap bounds

In order to bound the conductance of the ball walk, Proposition 3 requires an isoperimetric constant bound, as well as a distance $\Delta$ such that (13) holds. From Proposition 5 and the assumption (20), we know that we can take $\psi^{-1} = O(\sqrt{\log d})$, so we wish to get a handle on the $\Delta$ parameter in (13) where m is the $\ell_2$ distance. To this end, we note that for two points $\mathbf{x}, \mathbf{x}'$,

$$\begin{aligned} D_{\mathrm{TV}}\left(\mathcal{T}_\mathbf{x}, \mathcal{T}_{\mathbf{x}'}\right) &\le D_{\mathrm{TV}}\left(\mathcal{P}_\mathbf{x}, \mathcal{P}_{\mathbf{x}'}\right) + D_{\mathrm{TV}}\left(\mathcal{T}_\mathbf{x}, \mathcal{P}_\mathbf{x}\right) + D_{\mathrm{TV}}\left(\mathcal{T}_{\mathbf{x}'}, \mathcal{P}_{\mathbf{x}'}\right) \\ &\le D_{\mathrm{TV}}\left(\mathcal{P}_\mathbf{x}, \mathcal{P}_{\mathbf{x}'}\right) + \frac{1}{2}\left(1-\ell(\mathbf{x})\right) + \frac{1}{2}\left(1-\ell(\mathbf{x}')\right), \end{aligned} \tag{23}$$

where $\mathcal{P}_{\mathbf{x}}$ is the aforementioned lazy proposal distribution, which with probability $\frac{1}{2}$ draws a uniform point from $\mathbb{B}(\mathbf{x}, \eta)$. Because $\mathcal{P}_{\mathbf{x}}$ and $\mathcal{P}_{\mathbf{x}'}$ are uniform distributions over nearby balls when they are not lazy, their total variation distance enjoys a simple closed-form formula:

$$D_{\mathrm{TV}}\left(\mathcal{P}_{\mathbf{x}}, \mathcal{P}_{\mathbf{x}'}\right) = \frac{\mathrm{Vol}\left(\mathbb{B}(\mathbf{x}, \eta) \setminus \mathbb{B}(\mathbf{x}', \eta)\right)}{2\mathrm{Vol}(\mathbb{B}(\mathbf{x}, \eta))} = \frac{\mathrm{Vol}\left(\mathbb{B}(\mathbf{x}', \eta) \setminus \mathbb{B}(\mathbf{x}, \eta)\right)}{2\mathrm{Vol}(\mathbb{B}(\mathbf{x}', \eta))}. \quad (24)$$

We therefore provide a simple geometric lemma bounding the overlap of nearby balls.

**Lemma 5.** *Let $\eta > 0$, and suppose $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ satisfy $\|\mathbf{x} - \mathbf{x}'\|_2 = \frac{\alpha\eta}{\sqrt{d}}$. Then,*

$$\frac{\mathrm{Vol}\left(\mathbb{B}(\mathbf{x}, \eta) \setminus \mathbb{B}(\mathbf{x}', \eta)\right)}{\mathrm{Vol}(\mathbb{B}(\mathbf{x}, \eta))} \le \alpha.$$

*Proof.* The statement is invariant to shifts, rotations, and rescalings, so we may assume $\eta = 1$, $\mathbf{x} = \mathbf{0}_d$, and $\mathbf{x}' = \beta\mathbf{e}_1$ for simplicity, where $\beta := \frac{\alpha}{\sqrt{d}}$. We recall the formula

$$V_d := \mathrm{Vol}\left(\mathbb{B}(\mathbf{0}_d, 1)\right) = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2} + 1\right)}, \quad (25)$$

where $\Gamma$ is Euler's gamma function, satisfying $\Gamma(n+1) = n!$ for any $n \in \mathbb{N}$. We can hence compute

$$\begin{aligned}\frac{\mathrm{Vol}(\mathbb{B}(\mathbf{0}_d, 1) \setminus \mathrm{Vol}(\mathbb{B}(\beta\mathbf{e}_1, 1)))}{\mathrm{Vol}(\mathbb{B}(\mathbf{0}_d, 1))} &= \frac{1}{V_d}\int_{-1}^{\frac{\beta}{2}} V_{d-1} \cdot \left(\sqrt{1 - r^2}\right)^{d-1} \mathrm{d}r \\ &\quad - \frac{1}{V_d}\int_{-1}^{-\frac{\beta}{2}} V_{d-1} \cdot \left(\sqrt{1 - r^2}\right)^{d-1} \mathrm{d}r \\ &= \frac{V_{d-1}}{V_d}\int_{-\frac{\beta}{2}}^{\frac{\beta}{2}} \left(\sqrt{1 - r^2}\right)^{d-1} \mathrm{d}r \le \frac{\beta V_{d-1}}{V_d}.\end{aligned}$$

The first equality above followed because the symmetric difference of the two balls falls left of $\frac{\beta}{2}\mathbf{e}_1$ along the first coordinate axis, so we directly computed how much volume of each ball falls left of this point and take the difference. The conclusion follows from the inequality

$$\frac{V_{d-1}}{V_d} = \frac{\Gamma(\frac{d}{2} + 1)}{\sqrt{\pi}\Gamma(\frac{d-1}{2} + 1)} \le \sqrt{d}.$$

□

In conclusion, as long as two points $\mathbf{x}, \mathbf{x}'$ have local conductances $\ell(\mathbf{x}), \ell(\mathbf{x}') \ge \frac{3}{4}$, combining (23), (24), and Lemma 5 shows that $\|\mathbf{x} - \mathbf{x}'\|_2 \le \frac{\eta}{2\sqrt{d}}$ implies $D_{\mathrm{TV}}(\mathcal{T}_{\mathbf{x}}, \mathcal{T}_{\mathbf{x}'}) \le \frac{1}{2}$, as required by (13). We will discuss the issue of requiring large local conductances in the following sections. For now, we remark that [KLS97] established a strengthening of the argument in this section by reasoning more carefully about the relative volume of balls when intersecting with $\mathcal{K}$. In particular, they showed the following multiplicative variant of Lemma 5 which holds after intersection.

**Lemma 6** (Lemma 3.5, [KLS97])**.** *If $\mathbf{x}, \mathbf{x}' \in \mathcal{K}$ for compact, convex $\mathcal{K} \subseteq \mathbb{R}^d$ have $\|\mathbf{x} - \mathbf{x}'\|_2 \le \frac{\eta}{\sqrt{d}}$,*

$$\frac{\mathrm{Vol}\left(\mathcal{K} \cap (\mathbb{B}(\mathbf{x}, \eta) \cap \mathbb{B}(\mathbf{x}', \eta))\right)}{\mathrm{Vol}(\mathbb{B}(\mathbf{x}, \eta))} \ge \frac{1}{4}\min\left\{\ell(\mathbf{x}), \ell(\mathbf{x}')\right\}.$$

As a useful corollary of Lemma 6, we show how to bound the "crossing over" probability of nearby points on different sides of a partition, in line with what was used by the proof of Proposition 3.

**Corollary 2.** *For convex $\mathcal{K} \subseteq \mathbb{R}^d$, let $S \subseteq \mathcal{K}$, and denote $S^c := \mathcal{K} \setminus S$. For any $\mathbf{x} \in S$ and $\mathbf{x}' \in S^c$ with $\|\mathbf{x} - \mathbf{x}'\|_2 \le \frac{\eta}{\sqrt{d}}$ and $\ell(\mathbf{x}) \in [\frac{2}{3}\ell(\mathbf{x}'), \frac{3}{2}\ell(\mathbf{x}')]$,*

$$\frac{\mathcal{T}_{\mathbf{x}}(S^c)}{\ell(\mathbf{x})} + \frac{\mathcal{T}_{\mathbf{x}'}(S)}{\ell(\mathbf{x}')} \ge \frac{1}{18}.$$

*Proof.* By the definition of our lazy transition operators $\{\mathcal{T}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^d}$, we have

$$\mathcal{T}_{\mathbf{x}}(S^c) = \frac{\mathrm{Vol}(S^c \cap \mathbb{B}(\mathbf{x},\eta))}{2\mathrm{Vol}(\mathbb{B}(\mathbf{x},\eta))} \geq \frac{\mathrm{Vol}(S^c \cap (\mathbb{B}(\mathbf{x},\eta) \cap \mathbb{B}(\mathbf{x}',\eta)))}{2\mathrm{Vol}(\mathbb{B}(\mathbf{x},\eta))},$$

and similarly,

$$\mathcal{T}_{\mathbf{x}'}(S) \geq \frac{\mathrm{Vol}(S \cap (\mathbb{B}(\mathbf{x},\eta) \cap \mathbb{B}(\mathbf{x}',\eta)))}{2\mathrm{Vol}(\mathbb{B}(\mathbf{x}',\eta))}.$$

By summing these inequalities and applying Lemma 6, we hence have

$$\mathcal{T}_{\mathbf{x}}(S^c) + \mathcal{T}_{\mathbf{x}'}(S) \geq \frac{1}{8}\min\{\ell(\mathbf{x}), \ell(\mathbf{x}')\}.$$

Dividing both sides by $\ell(\mathbf{x})$ and using the assumptions then gives the claim. □

## 5.2 Speedy walk

As stated, Corollary 2 is incompatible with Proposition 3, due to the normalization by $\ell(\mathbf{x})$. However, there is a related random walk called the *speedy walk* for which Corollary 2 does imply a mixing bound. We define it in this section and prove some basic facts about its convergence.

In short, the speedy walk is the ball walk, except all "wasted nonlazy steps" are skipped. In other words, with probability $\frac{1}{2}$ a transition of the speedy walk starting from $\mathbf{x}$ does not move; else, it steps to a uniformly random point in $\mathbb{B}(\mathbf{x},\eta) \cap \mathcal{K}$. The transitions of the speedy walk, denoted $\{\widetilde{\mathcal{T}}_{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{K}}$ (to contrast with $\mathcal{T}_{\mathbf{x}}$ in (22)) are thus defined as follows: $\mathbf{x}' \sim \widetilde{\mathcal{T}}_{\mathbf{x}}$ follows

$$\mathbf{x}' \leftarrow \begin{cases} \mathbf{x} & \text{with probability } \frac{1}{2} \\ \text{a uniform draw from } \mathbb{B}(\mathbf{x},\eta)\cap\mathcal{K} & \text{with probability } \frac{1}{2} \end{cases}. \tag{26}$$

**Lemma 7.** *The speedy walk (26) with transition distributions $\{\widetilde{\mathcal{T}}_{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{K}}$ is reversible, and its stationary distribution is $\tilde{\pi}^\star$, the distribution proportional to $\ell$ over $\mathcal{K}$, defined as follows:*

$$\tilde{\pi}^\star(\mathbf{x}) = \begin{cases} \frac{\ell(\mathbf{x})}{\lambda(\mathcal{K})} \cdot \frac{1}{\mathrm{Vol}(\mathcal{K})} & \mathbf{x}\in\mathcal{K} \\ 0 & \textit{else} \end{cases}, \tag{27}$$

*where $\lambda(\mathcal{K})$ is the* average local conductance *over $\mathcal{K}$,*

$$\lambda(\mathcal{K}) := \frac{1}{\mathrm{Vol}(\mathcal{K})}\int_{\mathbf{x}\in\mathcal{K}} \ell(\mathbf{x})\mathrm{d}\mathbf{x} = \mathbb{E}_{\mathbf{x}\sim\pi^\star}[\ell(\mathbf{x})]. \tag{28}$$

*Proof.* Reversibility is immediate from the fact that for all $\mathbf{x},\mathbf{x}'\in\mathcal{K}$ with $\mathbf{x}\neq\mathbf{x}'$,

$$\begin{aligned}\tilde{\pi}^\star(\mathbf{x})\widetilde{\mathcal{T}}_{\mathbf{x}}(\mathbf{x}') &= \frac{\ell(\mathbf{x})}{\lambda(\mathcal{K})\mathrm{Vol}(\mathcal{K})} \cdot \frac{1}{2\mathrm{Vol}(\mathbb{B}(\mathbf{x},\eta)\cap\mathcal{K})} = \frac{1}{\lambda(\mathcal{K})\mathrm{Vol}(\mathcal{K})} \cdot \frac{1}{2\mathrm{Vol}(\mathbb{B}(\mathbf{x},\eta))} \\ &= \frac{1}{\lambda(\mathcal{K})\mathrm{Vol}(\mathcal{K})} \cdot \frac{1}{2\mathrm{Vol}(\mathbb{B}(\mathbf{x}',\eta))} = \tilde{\pi}^\star(\mathbf{x}')\widetilde{\mathcal{T}}_{\mathbf{x}'}(\mathbf{x}),\end{aligned}$$

where the second equality used (21). Now (4) implies that $\tilde{\pi}^\star$ is the stationary distribution. □

Of course, it is not obvious how to implement a step of the speedy walk from $\mathbf{x}\in\mathcal{K}$ using only a membership oracle. A simple way is to continue drawing random points in $\mathbb{B}(\mathbf{x},\eta)$ until it falls in $\mathcal{K}$; the distribution of the resulting point is clearly uniform in $\mathbb{B}(\mathbf{x},\eta)\cap\mathcal{K}$. We bound the expected number of such draws assuming that the current distribution is warm with respect to $\pi^\star$.

**Lemma 8.** *Let $\pi$ be $\beta$-warm with respect to $\tilde{\pi}^\star$ defined in (27). Then,*

$$\mathbb{E}_{\mathbf{x}\sim\pi}\left[\frac{1}{\ell(\mathbf{x})}\right] \leq \frac{\beta}{\lambda(\mathcal{K})}.$$

*Proof.* This follows from a direct computation: recalling the definition (27),

$$\int_{\mathbf{x}\in\mathcal{K}} \frac{1}{\ell(\mathbf{x})}\pi(\mathbf{x})\mathrm{d}x \le \beta \int_{\mathbf{x}\in\mathcal{K}} \frac{1}{\ell(\mathbf{x})}\pi^\star(\mathbf{x})\mathrm{d}x = \frac{\beta}{\lambda(\mathcal{K})}\cdot\frac{1}{\mathrm{Vol}(\mathcal{K})}\int_{\mathbf{x}\in\mathcal{K}}\mathrm{d}x = \frac{\beta}{\lambda(\mathcal{K})}.$$

□

Note that $\frac{1}{\ell(\mathbf{x})}$ is the expected number of queries to a membership oracle needed before a random draw from $\mathbb{B}(\mathbf{x},\eta)$ is found to fall in $\mathcal{K}$. Thus, Lemma 8 shows that a step of the speedy walk from a $\beta$-warm distribution can be implemented in $\approx \beta$ steps, provided $\lambda(\mathcal{K})$ is at least a constant. We next bound how much our reweighting of $\pi^\star$ by $\ell(\mathbf{x})$ affects the covariance.

**Lemma 9.** *For $\tilde{\pi}^\star$ defined in (27), we have following notation (18) that*

$$\mathrm{Cov}_{\tilde{\pi}^\star} \preceq \frac{1}{\lambda(\mathcal{K})}\mathrm{Cov}_{\pi^\star}.$$

*Proof.* We recall that

$$\mathrm{Cov}_{\tilde{\pi}^\star} = \mathbb{E}_{\mathbf{x}\sim\tilde{\pi}^\star}\left[(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})^\top)\right] = \frac{1}{\lambda(\mathcal{K})\mathrm{Vol}(\mathcal{K})}\cdot\int_{\mathbf{x}\in\mathcal{K}}\ell(\mathbf{x})(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})^\top\mathrm{d}\mathbf{x},$$
$$\mathrm{Cov}_{\pi^\star} = \mathbb{E}_{\mathbf{x}\sim\pi^\star}\left[(\mathbf{x}-\bar{x}_{\pi^\star})(\mathbf{x}-\bar{\mathbf{x}}_{\pi^\star})^\top\right] = \frac{1}{\mathrm{Vol}(\mathcal{K})}\int_{\mathbf{x}\in\mathcal{K}}(\mathbf{x}-\bar{\mathbf{x}}_{\pi^\star})(\mathbf{x}-\bar{\mathbf{x}}_{\pi^\star})^\top\mathrm{d}\mathbf{x}.$$

Hence, it suffices to show that

$$\int_{\mathbf{x}\in\mathcal{K}}\ell(\mathbf{x})(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})^\top\mathrm{d}\mathbf{x} \preceq \int_{\mathbf{x}\in\mathcal{K}}(\mathbf{x}-\bar{\mathbf{x}}_{\pi^\star})(\mathbf{x}-\bar{\mathbf{x}}_{\pi^\star})^\top\mathrm{d}\mathbf{x},$$

which follows because $\ell(\mathbf{x}) \in [0,1]$ pointwise, and

$$\begin{aligned}\mathbb{E}_{\mathbf{x}\sim\tilde{\pi}^\star}\left[(\mathbf{x}-\bar{\mathbf{x}}_{\pi^\star})(\mathbf{x}-\bar{\mathbf{x}}_{\pi^\star})\right] &= \mathbb{E}_{\mathbf{x}\sim\tilde{\pi}^\star}\left[(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})\right] + (\bar{\mathbf{x}}_{\tilde{\pi}^\star}-\bar{\mathbf{x}}_{\pi^\star})(\bar{\mathbf{x}}_{\tilde{\pi}^\star}-\bar{\mathbf{x}}_{\pi^\star})^\top \\ &\succeq \mathbb{E}_{\mathbf{x}\sim\tilde{\pi}^\star}\left[(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})(\mathbf{x}-\bar{\mathbf{x}}_{\tilde{\pi}^\star})\right].\end{aligned}$$

□

At this point we can almost directly combine our relative overlap bound (Corollary 2) with our isoperimetry bound (Lemma 3 using (18) and Lemma 9) to obtain a conductance lower bound on the speedy walk, via Proposition 3. There is one complication, which is that Corollary 2 can only compare points whose local conductances are close. To bypass this, [KLS97] define a hybrid distance which both compares the Euclidean distance and the relative change in $\ell$. They show that both cases can be handled via localization, and prove the following conductance lower bound.

**Corollary 3** (Theorem 3.1, [KLS97])**.** *The conductance of the speedy walk (26) is*

$$\Omega\left(\frac{\eta\sqrt{\lambda(\mathcal{K})}}{d^{1.5}}\right).$$

Intuitively, we expect to lose one $\frac{\eta}{\sqrt{d}}$ factor from the distance required by Corollary 2, and one $\sqrt{\lambda(\mathcal{K})/d}$ factor from the isoperimetry in Lemmas 3 and 9. The remaining factor of $\sqrt{d}$ lost in Corollary 3 is because the argument for the local conductance distance depends on the *diameter* of the convex body, which scales as $\approx d$ in Fact 1. Ultimately many of these poly($d$) factors can be shaved by using the more streamlined analysis tools we develop in later lectures.

### 5.3 Average local conductance

By applying Corollary 3 within Corollary 1, we see that the speedy walk rapidly mixes to its stationary distribution, $\tilde{\pi}^\star$. Moreover, as long as each iteration of the speedy walk is $\beta$-warm with respect to $\tilde{\pi}^\star$, then steps can be implemented efficiently using Lemma 8. What is left is to control the distance between $\tilde{\pi}^\star$ and the actual uniform distribution of interest, $\pi^\star$.

We begin by characterizing $D_{\mathrm{TV}}(\pi^\star,\tilde{\pi}^\star)$ in terms of the average local conductance $\lambda(\mathcal{K})$.

**Lemma 10.** *For $\pi^\star, \tilde{\pi}^\star$ defined in (19), (27) respectively, $D_{\mathrm{TV}}(\pi^\star, \tilde{\pi}^\star) \le 1 - \lambda(\mathcal{K})$.*

*Proof.* This is a direct calculation, using the formula for $D_{\mathrm{TV}}$ in Definition 4, Part XIV:

$$\begin{aligned} D_{\mathrm{TV}}\left(\pi^\star, \tilde{\pi}^\star\right) &= \frac{1}{2}\int_{\mathbf{x}\in\mathcal{K}} \left|\frac{1}{\mathrm{Vol}(\mathcal{K})} - \frac{\ell(\mathbf{x})}{\lambda(\mathcal{K})\mathrm{Vol}(\mathcal{K})}\right| \mathrm{d}\mathbf{x} \\ &\le \frac{1}{2\mathrm{Vol}(\mathcal{K})}\int_{\mathbf{x}\in\mathcal{K}} |1-\ell(\mathbf{x})|\,\mathrm{d}\mathbf{x} + \frac{1}{2\mathrm{Vol}(\mathcal{K})}\int_{\mathbf{x}\in\mathcal{K}} \left|\frac{\ell(\mathbf{x})}{\lambda(\mathcal{K})} - \ell(\mathbf{x})\right| \mathrm{d}\mathbf{x} \\ &= \frac{1}{2}\left(1-\lambda(\mathcal{K})\right) + \frac{1}{2}\left(1-\lambda(\mathcal{K})\right) = 1-\lambda(\mathcal{K}). \end{aligned}$$

The only inequality was the triangle inequality, and the last line used $\ell(\mathbf{x}) \in [0,1]$ and $\lambda(\mathcal{K}) \le 1$. □

We are left with estimating the average local conductance $\lambda(\mathcal{K}) = \mathbb{E}_{\mathbf{x}\sim\pi^\star}[\ell(\mathbf{x})]$ as a function of the step size $\eta$. We can relate these quantities via the isoperimetry of $\mathcal{K}$ as follows.

**Fact 2** (Corollary 4.5, [KLS97])**.** *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be convex, and let $\partial\mathcal{K}$ denote the boundary of its closure. Then,*

$$\lambda(\mathcal{K}) \ge 1 - \frac{\eta}{2\sqrt{d}} \cdot \frac{\mathrm{Vol}(\partial\mathcal{K})}{\mathrm{Vol}(\mathcal{K})}.$$

Fact 2 lets us lower bound the average local conductance of an isotropic set, by using Fact 1.

**Corollary 4.** *For any convex $\mathcal{K} \subseteq \mathbb{R}^d$ satisfying (20), we have $D_{\mathrm{TV}}(\pi^\star, \tilde{\pi}^\star) \le \frac{\eta\sqrt{d}}{2}$.*

*Proof.* We first apply Fact 1, which shows $\mathcal{K}$ contains a unit ball. Along each unit vector projection of the body, the surface area-to-volume ratio is monotone decreasing in the length $r$ of the projection, because the former scales as $r^{d-1}$ and the latter as $r^d$. Therefore, by (12),

$$\frac{\mathrm{Vol}(\partial\mathcal{K})}{\mathrm{Vol}(\mathcal{K})} \le \frac{\mathrm{Vol}(\partial\mathbb{B}(\mathbf{0}_d, 1))}{\mathrm{Vol}(\mathbb{B}(\mathbf{0}_d, 1))} = d.$$

Plugging this bound into Fact 2 gives $\lambda(\mathcal{K}) \ge 1 - \frac{\eta\sqrt{d}}{2}$, and then Lemma 10 gives the claim. □

At this point, we have given all the tools needed to obtain the following sampling guarantee.

**Theorem 1.** *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be convex and satisfy (20), and define $\pi^\star$ as in (19). We can produce a sample within $\epsilon$ total variation distance from $\pi^\star$, given an initial point drawn from a $\beta$-warm distribution for $\pi^\star$, using $\mathrm{poly}(d, \beta, \frac{1}{\epsilon})$ queries to a membership oracle for $\mathcal{K}$ in expectation.*

*Proof.* First, let $\eta = \frac{\epsilon}{\sqrt{d}}$, so that Corollary 4 shows that $D_{\mathrm{TV}}(\pi^\star, \tilde{\pi}^\star) \le \frac{\epsilon}{2}$. Next, Corollaries 1 and 3 show that, letting $\pi_k$ be the distribution of the $k^{\text{th}}$ speedy walk iterate initialized at $\pi_0$, we can achieve $D_{\mathrm{TV}}(\pi_k, \tilde{\pi}^\star) \le \frac{\epsilon}{2}$ as long as

$$k = \Omega\left(\frac{d^3}{\eta^2}\log\left(\frac{\beta}{\epsilon}\right)\right) = \Omega\left(\frac{d^4}{\epsilon^2}\log\left(\frac{\beta}{\epsilon}\right)\right).$$

By the triangle inequality we also have $D_{\mathrm{TV}}(\pi_k, \pi^\star) \le \epsilon$ as desired. Finally, each iteration of the speedy walk is implementable using $O(\beta)$ expected queries to a membership oracle by Lemma 8. □

Theorem 1 gives us a basic appreciation of the ingredients needed to analyze the convergence of structured continuous random walks for challenging, but still tractable, distributions. There are several qualitative weaknesses of Theorem 1: it depends polynomially on the warmness parameter $\beta$ and the inverse accuracy $\frac{1}{\epsilon}$, and has a rather large dependence on the dimension. The tools we will soon develop are motivated by these shortcomings and facilitate much sharper bounds.

## Source material

Portions of this lecture are based on reference material in [LS93, KLS95, KLS97], as well as the author's own experience working in the field.

## References


[AK91] David L. Applegate and Ravi Kannan. Sampling and integration of near log-concave functions. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, pages 156–163. ACM, 1991.

[BF87] Imre Bárány and Zoltán Füredi. Computing the volume is difficult. *Discrete and Computational Geometry*, 2:319–326, 1987.

[Bor75] Christer Borell. Convex set functions in d-space. *Periodica Mathematica Hungarica*, 6(2):111–136, 1975.

[BT03] Sergey G. Bobkov and Prasad Tetali. Modified log-sobolev inequalities, mixing and hypercontractivity. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 287–296. ACM, 2003.

[CDWY20] Yuansi Chen, Raaz Dwivedi, Martin J. Wainwright, and Bin Yu. Fast mixing of metropolized hamiltonian monte carlo: Benefits of multi-step gradients. *J. Mach. Learn. Res.*, 21:92:1–92:72, 2020.

[Che21] Yuansi Chen. An almost constant lower bound of the isoperimetric coefficient in the kls conjecture. *Geometric and Functional Analysis*, 31:34–61, 2021.

[CV18] Ben Cousins and Santosh S. Vempala. Gaussian cooling and $o(n^3)$ algorithms for volume and gaussian volume. *SIAM J. Comput.*, 47(3):1237–1273, 2018.

[DF88] Martin E. Dyer and Alan M. Frieze. On the complexity of computing the volume of a polyhedron. *SIAM J. Comput.*, 17(5):967–974, 1988.

[DFK91] Martin E. Dyer, Alan M. Frieze, and Ravi Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, 1991.

[Eld13] Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23:532–569, 2013.

[Ele86] György Elekes. A geometric inequality and the complexity of computing volume. *Discrete and Computational Geometry*, 1:289–292, 1986.

[FG04] Matthieu Fradelizi and Olivier Guédon. The extreme points of subsets of s-concave probabilities and a geometric localization theorem. *Discrete & Computational Geometry*, 31(2):327–335, 2004.

[GLL+23] Sivakanth Gopi, Yin Tat Lee, Daogao Liu, Ruoqi Shen, and Kevin Tian. Private convex optimization in general norms. In *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023*, pages 5068–5089. SIAM, 2023.

[GMT06] Sharad Goel, Ravi Montenegro, and Prasad Tetali. Mixing time bounds via the spectral profile. *Electronic Journal of Probability*, 11:1–26, 2006.

[JLLV21] He Jia, Aditi Laddha, Yin Tat Lee, and Santosh S. Vempala. Reducing isotropy and volume to KLS: an $o(n^3\psi^2)$ volume algorithm. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, 2021*, pages 961–974. ACM, 2021.

[Kla23] Boaz Klartag. Logarithmic bounds for isoperimetry and slices of convex sets. *Ars Inveniendi Analytica*, 2023.

[KLS95] Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discret. Comput. Geom.*, 13:541–559, 1995.

[KLS97] Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $o^*(n^5)$ volume algorithm for convex bodies. *Random Struct. Algorithms*, 11(1):1–50, 1997.

[KV25] Yunbum Kook and Santosh S. Vempala. Sampling and integration of logconcave functions by algorithmic diffusion. In *STOC '25, 57th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, 2025.

[Led99] Michel Ledoux. *Concentration of measure and logarithmic Sobolev inequalities*. Seminaire de probabilities XXXIII, 1999.

[LK99] László Lovász and Ravi Kannan. Faster mixing via average conductance. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pages 282–287. ACM, 1999.

[LS92] László Lovász and Miklós Simonovits. On the randomized complexity of volume and diameter. In *33rd Annual Symposium on Foundations of Computer Science*, pages 482–491. IEEE Computer Society, 1992.

[LS93] László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Struct. Algorithms*, 4(4):359–412, 1993.

[LV06a] László Lovász and Santosh S. Vempala. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS) 2006*, pages 57–68. IEEE Computer Society, 2006.

[LV06b] László Lovász and Santosh S. Vempala. Hit-and-run from a corner. *SIAM J. Comput.*, 35(4):985–1005, 2006.

[LV06c] László Lovász and Santosh S. Vempala. Simulated annealing in convex bodies and an $o(n^4)$ volume algorithm. *J. Comput. Syst. Sci.*, 72(2):392–417, 2006.

[LV07] László Lovász and Santosh S. Vempala. The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms*, 30(3):307–358, 2007.

[LV17a] Yin Tat Lee and Santosh Srinivas Vempala. Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017*, pages 998–1007. IEEE Computer Society, 2017.

[LV17b] Yin Tat Lee and Santosh Srinivas Vempala. The kannan-lovász-simonovits conjecture. *Current Developments in Mathematics*, 2017:1–36, 2017.

[Mon23] Andrea Montanari. Sampling, diffusions, and stochastic localization. *CoRR*, abs/2305.10690, 2023.