CPSC 661: Sampling Algorithms in Machine Learning	1 March, 2021

Lecture 8

Lecturer: Andre Wibisono — *Scribe: Anay Mehrotra*

1 Outline

Today's lecture covers two main topics:

1. Effective diameter and concentration inequalities for strongly log-concave distributions, and
2. lower bounds on the s-conductance of Markov chains.

2 Preliminaries

Notation. Given $x \in \mathbb{R}^n$ and $r \geq 0$, let $\mathbb{B}(x, r) \subseteq \mathbb{R}^n$ denote the ℓ_2-ball centered at x with radius r.[1] Given two distributions ν and ρ on $\mathbb{R}^n$, define the total variation distance between them as

$$\mathrm{TV}\left(\nu, \rho\right) \coloneqq \sup_{A \subseteq \mathbb{R}^n} \left|\nu(A) - \rho(A)\right|.$$

Given two sets $S, T \subseteq \mathbb{R}^n$, define the distance between them to be

$$d(S, T) \coloneqq \inf_{x \in S, y \in T} \left\|x - y\right\|_2.$$

2.1 Markov chains

Let P be a Markov chain. Suppose that the current point of P is $x \in \mathbb{R}^n$. Let P_x denote the density of the next point of P. For a Markov chain P, define the ergodic flow from a set $A \subseteq \mathbb{R}^n$ to A^c as

$$\Phi(A) \coloneqq \int_A P_x(A^c)\ \mathrm{d}\nu(x). \tag{Ergodic flow}$$

In other words, $\Phi(A)$ is the probability that P moves from a point in A (sampled from ν) to a point in A^c.

Definition 1 (**s-conductance**)**.** *For $s \in [0, 1/2)$ define the s-conductance of P as*

$$\Phi_s(P) \coloneqq \inf_{A \subseteq \mathbb{R}^n :\ s < \nu(A) < 1-s} \frac{\Phi(A)}{\min\left\{\nu(A) - s, \nu(A^c) - s\right\}}. \tag{s-conductance}$$

[1] In other words, $B(x, r) \coloneqq \left\{y \in \mathbb{R}^n :\ \|y - x\|_2 \leq r\right\}$.

2.2 Strong log-concavity and strong convexity

From the last lecture, recall that for any $\alpha > 0$ a twice differentiable function, f, is said to be α-strongly convex if $\nabla^2 f(x) \succeq \alpha I$. Equivalently, f is said to be α-strongly convex if the smallest eigenvalue of $\nabla^2 f(x)$ is at least α. Further, a distribution ν is said to be α-strongly logconcave if the function $f \colon \mathbb{R}^n \to \mathbb{R}$ defined as $f \coloneqq -\log(\nu)$ is α-strongly convex. In particular, this implies that a α-strongly convex distribution has a unique mode. For further discussion of log-concave distributions, we refer the reader to [Vem05, Section 2] and [LV07, Section 5].

Example 1. *Consider the Gaussian distribution $\rho = \mathcal{N}(\nu, \Sigma)$ with mean $\nu \in \mathbb{R}^n$ covariance $\Sigma \in \mathbb{R}^{n\times n}$:*

$$\rho(x) \coloneqq \frac{1}{\sqrt{\det(2\pi\Sigma)}} \cdot \exp\left(-\frac{1}{2} \cdot (x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

In this example, $f(x) \coloneqq -\log(\rho) \circ (x) = \frac{1}{2} \cdot (x-\mu)^\top \Sigma^{-1}(x-\mu) + \frac{1}{2} \cdot \log\det(2\pi\Sigma)$. The hessian of f is $\nabla^2 f(x) = \Sigma^{-1}$. Therefore, f is $\frac{1}{\lambda_{\max}(\Sigma)}$-strongly convex and ρ is $\frac{1}{\lambda_{\max}(\Sigma)}$-strongly log-concave.

2.3 Isoperimetry for strongly logconcave distributions

Recall the following isoperimetry-inequality for strongly logconcave distributions presented in Lecture 7.

Theorem 1 (Isoperimetry constant for strongly logconcave distributions). *All α-strongly log-concave distribution ν on $\mathbb{R}^n$ satisfy isoperimetry with constant*

$$\Psi \geq \log 2 \cdot \sqrt{\alpha}.$$

That is, for any partition $S \cup T \cup (\mathbb{R}^n \backslash S \backslash T)$, it holds that

$$\nu(\mathbb{R}^n \backslash S \backslash T) \geq \log 2 \cdot \sqrt{\alpha} \cdot d(S,T) \cdot \min\left\{\nu(S), \nu(T)\right\}. \tag{1}$$

Interestingly, the isoperimetry constant Ψ is independent of the dimension n for strongly logconcave distributions. Kannan, Lovász and Simonovits conjectured that this is true for all log-concave distributions (not just strongly log-concave ones); this is known as the KLS conjecture [KLS95]. Since its proposal in 1995, it has been a subject of intense study–both in attempts to resolve the conjecture [Kla07b, FGP07, Kla07a, Fle10, GM11, LV17a] and in establishing its relation to other well-known conjectures [Bal06, Eld13]. We refer the reader to the excellent survey [LV18] for an in-depth discussion and several more references.

Typically proofs of lower bounds on isoperimetry constants proceed through the idea of "Localization," which reduces inequalities in measures of sets (i.e., integrals in $\mathbb{R}^n$) to one-dimensional integrals. At a high-level, the reduction proceeds by starting with $\mathbb{R}^n$ and repeatedly bisecting the space so that the original inequality holds in the remaining space after the bisection. In the limit, this process reduces the dimension of the space from n to 1. We refer the reader to [YTL19, Section 9.5] for an overview of localization; and [Eld13, EL14, Fle10, LV17b] for a discussion of state-of-the-art localization techniques.

3 Effective diameter and concentration

In this section, we motivate the notion of effect diameter and present concentration result for strongly log-concave distributions. Consider an α-strongly log-concave distribution, $\nu \propto e^{-f}$ on $\mathbb{R}^n$. Let $\overline{x}$ be the mean of ν, i.e., $\overline{x} := \frac{\int_{\mathbb{R}^n} x e^{-f(x)} \mathrm{d}x}{\int_{\mathbb{R}^n} e^{-f(x)} \mathrm{d}x}$. Let $x^\star$ be the mode of ν, i.e., $x^\star := \mathrm{argmax}_{x \in \mathbb{R}^n} \nu(x)$. Then one can show that $\overline{x}$ and $x^\star$ are not too far from each other. Further, on average, a sample from ν is close to both $\overline{x}$ and $x^\star$. Formally, we have the following result.

Lemma 1. *Suppose ν is α-strongly log-concave. Then it holds that*

1. $\mathrm{Var}_\nu(x) := \mathbb{E}_\nu[\|x - \overline{x}\|_2^2] \leq \frac{n}{\alpha}$,
2. $\mathbb{E}_\nu[\|x - x^\star\|_2^2] \leq \frac{n}{\alpha}$,
3. $\|\overline{x} - x^\star\|_2^2 \leq \frac{n}{\alpha}$.

Example 2. *Consider the Gaussian distribution $\rho = \mathcal{N}(\nu, \Sigma)$ with mean $\nu \in \mathbb{R}^n$ covariance $\Sigma \in \mathbb{R}^{n \times n}$. From Example 1, we know that ρ is $\frac{1}{\lambda_{\max}(\Sigma)}$-strongly log-concave. In this case,* $\mathrm{Var}_\nu(x) = \mathrm{Tr}(\Sigma) = \sum_{i=1}^n \lambda_i(\Sigma) \leq n \cdot \lambda_{\max}(\Sigma) = \frac{n}{\alpha}$.

See [DCWY19, Section 5.3] for an outline of the proof of Lemma 1. Lemma 1 suggests that "effectively" samples from ν lie at a distance of at most $\approx \sqrt{n/\alpha}$ from each other. $\sqrt{n/\alpha}$ is popularly known as the effective diameter of the distribution. Lemma 2 formalizes this idea.

For all $s > 0$, define $r(s) := 2 + 2 \cdot \max\{\left(n^{-1} \log s^{-1}\right)^{\frac{1}{4}}, \left(n^{-1} \log s^{-1}\right)^{\frac{1}{2}}\}$. Define $\mathcal{R}_s \subseteq \mathbb{R}^n$ as

$$\mathcal{R}_s := \mathbb{B}(x^\star, r(s) \cdot \sqrt{n/\alpha}) \tag{2}$$

Lemma 2 ([DCWY19, Lemma 1])**.** *If ν is α-strongly log-concave on $\mathbb{R}^n$, then $\nu(\mathcal{R}_s) \geq 1 - s$.*

Note that, if $s = e^{-cn}$ then $r(s) \approx \sqrt{c}$. Therefore, by Lemma 2, at most $e^{-\Omega(n)}$ fraction of the samples from ν lie farther than $O(\sqrt{n/\alpha})$ distance from its $x^\star$. Using this, one can show that the restriction of ν to $\mathcal{R}_s$ is $(\frac{s}{1-s})$-close in total variation distance to ν.[2] Thus, for a sufficiently small s, it suffices to restrict ν to the ball $\mathcal{R}_s$.

Interlude to optimization. Equivalently, one can define $x^\star$, the mode of ν, as the minimizer of f (where $\nu \propto e^{-f}$), i.e.,

$$x^\star = \underset{x \in \mathbb{R}^n}{\mathrm{argmin}}\, f(x).$$

Since ν is α-strongly log-concave, f is α-strongly convex. Assume that f is L-smooth. Therefore, we have $\alpha I \preceq \nabla^2 f(x) \preceq LI$. Define the *condition number* of f as $\kappa := L/\alpha$. As we will see in later

[2]To see this, note that $\nu_s(A) - \nu(A) = \frac{\nu(A \cap \mathcal{R}_s)}{\nu(\mathcal{R}_s)} - \nu(A) \geq \frac{\nu(A) - \nu(\mathcal{R}_s^c)}{\nu(\mathcal{R}_s)} - \nu(A) \geq \nu(A) - s - \nu(A) = s$ and $\nu_s(A) - \nu(A) = \frac{\nu(A \cap \mathcal{R}_s)}{\nu(\mathcal{R}_s)} - \nu(A) \leq \frac{\nu(A)}{\nu(\mathcal{R}_s)} - \nu(A) \leq \frac{\nu(A)}{1-s} - \nu(A) \leq \frac{s}{1-s}$.

lectures, it is a theorem that gradient descent finds $x^\star$ using $\widetilde{O}(\kappa)$ calls to first-order oracle for f [Vis21, Chapter 6]. Further, it is also a theorem that, accelerated gradient descent finds $x^\star$ using $\widetilde{O}(\sqrt{\kappa})$ calls to first-order oracle for f [Vis21, Chapter 8] and this is optimal: any gradient-based algorithm requires at least $\Omega(\sqrt{\kappa})$ oracle-calls [Vis21, Exercise 8.4].

4 Isoperimetry with one-step overlap implies s-conductance

In this section, we show that if a Markov chain satisfies a "one-step overlap" property and its stationary distribution's isoperimetry constant is not too large, then its s-conductance is not too small. Formally, we prove Theorem 2.

Theorem 2 (Isoperimetry with one-step overlap implies s-conductance [DCWY19, Lemma 2]). *Suppose ν is a α-strongly log-concave distribution on $\mathbb{R}^n$. Let P be a Markov chain with stationary distribution is ν. For all $s \in (0, 1/2]$ and $\Delta_s \geq 0$, if P satisfies the following* one-step overlap *property (in the euclidean norm):*

$$\textit{For all } x, y \in \mathcal{R}_s, \quad \textit{if } \|x-y\|_2 \leq \Delta_s, \quad \textit{then } \mathrm{TV}\left(P_x, P_y\right) \leq \frac{3}{4}, \tag{3}$$

then P has s-conductance at least $\min\{1/16, \sqrt{\alpha}\cdot\Delta_s/128\}$, i.e.,

$$\phi_s(P) \geq \min\left\{\frac{1}{16}, \frac{\sqrt{\alpha}\cdot\Delta_s}{128}\right\}. \tag{4}$$

Recall that, from Theorem 1 we know that any strongly log-concave distribution has a small isoperimetry constant. Thus, any Markov chain P with the one-step overlap property and a α-strongly log-concave distribution, has s-conductance at least $\sqrt{\alpha}\cdot\Delta_s/128$. Intuitively, the property requires that if two points are close together, then the next-step distributions on starting P from either point are also close to each other. It turns out several fundamental Markov chains have this property, including the Metropolis random walk (e.g., [LS93]) and MALA (e.g., [MV19]). We will prove this property for the Metropolis random walk in the next lecture. (Somewhat interestingly, it also holds for Markov chains like hit-and-run [Lov99] and Gibbs-sampler [LV20, NS20] which take large steps in the domain.)

Remark 1 (The constant in the one-step property). *The constant $3/4$ in Equation (3) is not crucial. In fact, for any constant $c > 0$, one can replace $3/4$ with $1 - c$ in Equation (3). This would only change other constants in Theorem 2.*

Combining Theorem 2 with a bound on the mixing-time in terms of s-conductance (from Lecture 6), we get the following corollary.

Corollary 1 (**Bound on mixing time**). *Let P be a Markov chain with stationary distribution is ν. For all $s \in (0, 1/2]$ and $\Delta_s \geq 0$, if P satisfies the following* one-step overlap *property (in the euclidean norm):*

$$\text{For all } x, y \in \mathcal{R}_s, \quad \text{if } \|x-y\|_2 \leq \Delta_s, \quad \text{then } \mathrm{TV}\left(P_x, P_y\right) \leq \frac{3}{4}, \tag{5}$$

then starting from an M-warm start P converges to a distribution ε-close in total variation distance to ν in $O(\phi_t^{-2} \cdot \log(M/\varepsilon)) = O(\alpha^{-1} \cdot \Delta_t^{-2} \cdot \log(M/\varepsilon))$ steps, where $t = \varepsilon/2M$.

4.1 Proof of Theorem 2

Fix any set $A \subseteq \mathbb{R}^n$ with $s < \nu(A) < 1-s$. We will show that $C(A) \geq \min\{1/16, \sqrt{\alpha} \cdot \Delta_s/128\}$, where

$$C(A) \coloneqq \frac{\Phi(A)}{\min\left\{\nu(A)-s, \nu(A^c)-s\right\}}.$$

Then, Theorem 2 follows by using Definition 1.

Given $A \subseteq \mathbb{R}^n$, define sets $A_1 \subseteq A$ and $A_2 \subseteq A^c$, containing points "deep inside" A and A^c respectively, as

$$A_1 \coloneqq \left\{x \in A \cap \mathcal{R}_s \mid P_x(A^c) < 1/8\right\}, \tag{6}$$

$$A_2 \coloneqq \left\{y \in A^c \cap \mathcal{R}_s \mid P_y(A) < 1/8\right\}. \tag{7}$$

Further, define $A_3 \coloneqq \mathcal{R}_s \backslash A_1 \backslash A_2$ as the remaining portion of $\mathcal{R}_s$. Thus, $\{A_1, A_2, A_3\}$ partitions $\mathcal{R}_s$.

Intuitively, points in A_1 (resp. A_2) are unlikely to cross over to A^c (resp. A) in the next step. Therefore, if the measure of A_1 or A_2 is small (relative to $\nu(A)$ and $\nu(A^c)$), then the ergodic flow $\Phi(A)$ should be large. Lemma 3 formalizes this idea.

Lemma 3. *If $\nu(A_1) \leq \frac{1}{2} \cdot \nu(A \cap \mathcal{R}_s)$ or $\nu(A_2) \leq \frac{1}{2} \cdot \nu(A^c \cap \mathcal{R}_s)$, then $C(A) \geq \frac{1}{16}$.*

Proof. First, suppose that $\nu(A_1) \leq \frac{1}{2} \cdot \nu(A \cap \mathcal{R}_s)$. In this case we have

$$\begin{aligned}
\Phi(A) \quad &= \quad \int_{x \in A} P_x(A^c) \mathrm{d}\nu(x) \\
&\geq \quad \int_{x \in (A \cap \mathcal{R}_s) \backslash A_1} P_x(A^c) \mathrm{d}\nu(x) \\
&\overset{(6)}{\geq} \quad \frac{1}{8} \cdot \nu((A \cap \mathcal{R}_s) \backslash A_1) \\
&\geq \quad \frac{1}{16} \cdot \nu(A \cap \mathcal{R}_s) \qquad\qquad (\text{Using that } \nu(A_1) \leq \tfrac{1}{2} \cdot \nu(A \cap \mathcal{R}_s)) \\
&= \quad \frac{1}{16} \cdot (\nu(A) - \nu(A \cap \mathcal{R}_s^c)) \\
&\geq \quad \frac{1}{16} \cdot (\nu(A) - \nu(\mathcal{R}_s^c)) \\
&\overset{\text{Lemma 2}}{\geq} \quad \frac{1}{16} \cdot (\nu(A) - s)
\end{aligned}$$

$$\geq \quad \frac{1}{16} \cdot \min\left\{\nu(A) - s, \nu(A^c) - s\right\}.$$

Thus, it follows that

$$C(A) = \frac{\Phi(A)}{\min\left\{\nu(A) - s, \nu(A^c) - s\right\}} \geq \frac{1}{16}.$$

By symmetry, if $\nu(A_2) \leq \frac{1}{2} \cdot \nu(A^c \cap \mathcal{R}_s)$, then we have $C(A^c) \geq \frac{1}{16}$. Since ν is the stationary distribution of P, we have $C(A) = C(A^c)$. Thus, $C(A) \geq \frac{1}{16}$. □

It remains to prove Equation (4), when $\nu(A_1) \geq \frac{1}{2} \cdot \nu(A \cap \mathcal{R}_s)$ and $\nu(A_2) \geq \frac{1}{2} \cdot \nu(A^c \cap \mathcal{R}_s)$.

Lemma 4. *If $\nu(A_1) \geq \frac{1}{2} \cdot \nu(A \cap \mathcal{R}_s)$ and $\nu(A_2) \geq \frac{1}{2} \cdot \nu(A^c \cap \mathcal{R}_s)$, then $\Phi(A) \geq \frac{1}{16} \cdot \nu(A_3)$.*

Proof.

$$\begin{aligned}
\Phi(A) &= \frac{1}{2} \cdot (\Phi(A) + \Phi(A^c)) \qquad \text{(Using that } \nu \text{ is the stationary distribution of } P\text{)} \\
&= \frac{1}{2} \cdot \left(\int_{x \in A} P_x(A^c)\mathrm{d}\nu(v) + \int_{x \in A^c} P_x(A)\mathrm{d}\nu(v)\right) \\
&\geq \frac{1}{2} \cdot \left(\int_{x \in (A \cap \mathcal{R}_s) \setminus A_1} P_x(A^c)\mathrm{d}\nu(v) + \int_{x \in (A^c \cap \mathcal{R}_s) \setminus A_2} P_x(A)\mathrm{d}\nu(v)\right) \\
&\overset{(6),(7)}{\geq} \frac{1}{16} \cdot \nu(A_3). \qquad (8)
\end{aligned}$$

□

Towards proving Equation (4), we would like to lower bound $\nu(A_3)$ when $\nu(A_1)$ and $\nu(A_2)$ are not too small–they satisfy

$$\nu(A_1) \geq \frac{1}{2} \cdot \nu(A \cap \mathcal{R}_s) \quad \text{and} \quad \nu(A_2) \geq \frac{1}{2} \cdot \nu(A^c \cap \mathcal{R}_s). \qquad (9)$$

This is where isoperimetry comes in. However, we cannot use Equation (1) with the measure ν and sets $A_1, A_2, A_3 \subseteq \mathbb{R}^n$, because A_1, A_2, and A_3 do not partition $\mathbb{R}^n$. We bypass this by considering the restriction, ν_s, of ν to $\mathcal{R}_s$. Formally, let ν_s be the distribution satisfying that for all sets $B \subseteq \mathbb{R}^n$ that

$$\nu_s(B) \coloneqq \frac{\nu(B \cap \mathcal{R}_s)}{\nu(\mathcal{R}_s)}. \qquad (10)$$

Fact 1. *If ν is α-strongly log-concave and $\mathcal{R}_s$ is a convex set, then ν_s is also α-strongly log-concave.*

Fact 1 allows us to use Equation (1) for ν_s with the partition $\mathcal{R}_s = A_1 \cup A_2 \cup A_3$.

$$\nu_s(A_3) \geq \frac{\sqrt{\alpha}}{4} \cdot d(A_1, A_2) \cdot \min\left\{\nu_s(A_1), \nu_s(A_2)\right\}$$

We can transfer this lower bound on $\nu_s(A_3)$ to a lower bound on $\nu(A_3)$ by scaling the above inequality by $\nu(\mathcal{R}_s) > 0$ and using Equation (10).

$$\begin{aligned}
\nu(A_3) &\geq \frac{\sqrt{\alpha}}{4} \cdot d(A_1, A_2) \cdot \min\left\{\nu(A_1), \nu(A_2)\right\} \\
&\overset{(9)}{\geq} \frac{\sqrt{\alpha}}{8} \cdot d(A_1, A_2) \cdot \min\left\{\nu(A) - s, \nu(A^c) - s\right\}.
\end{aligned} \tag{11}$$

Finally, we require a lower bound $d(A_1, A_2)$. This follows from the one-step overlap property.

Lemma 5. *It holds that* $d(A_1, A_2) \geq \Delta_s$.

Proof. For any $x \in A_1$ and $y \in A_2$, we have that

$$\begin{aligned}
\mathrm{TV}\left(P_x, P_y\right) &= \sup_{B \subseteq \mathbb{R}^n} \left|P_x(B) - P_y(B)\right| \\
&\geq P_x(A) - P_y(A) \\
&= 1 - P_x(A^c) - P_y(A) \\
&\geq 1 - \frac{1}{8} - \frac{1}{8} \\
&= \frac{3}{4}.
\end{aligned}$$

Therefore, by the (contrapositive of) one-step overlap property (Equation (5)), it must hold that

$$\left\|x - y\right\|_2 > \Delta_s. \tag{12}$$

Since this holds for all $x \in A_1$ and $y \in A_2$, we have that

$$d(A_1, A_2) = \inf_{x \in A_1,\ y \in A_2} \left\|x - y\right\|_2 \overset{(12)}{\geq} \Delta_s.$$

(Notice that we loose the strict inequality, why?) □

Combining Lemma 5 with Equation (11), we get that, if Equation (9) holds, then

$$\nu(A_3) \geq \frac{\sqrt{\alpha} \cdot \Delta_s}{8} \cdot \min\left\{\nu(A) - s, \nu(A^c) - s\right\}.$$

Substituting this into Lemma 4, we get

$$\Phi(A) \geq \frac{\sqrt{\alpha} \cdot \Delta_s}{128} \cdot \min\left\{\nu(A) - s, \nu(A^c) - s\right\}.$$

Thus, $C(A) \geq \frac{1}{128} \cdot \sqrt{\alpha} \cdot \Delta_s$. Combining this with Lemma 3 completes the proof.

5 An aside on the Kullback-Leibler divergence

Let ρ and ν be probability distributions on some domain $\mathcal{X}$. The Kullback-Leibler (KL) divergence of ρ with respect to ν is defined as

$$H_\nu(\rho) \coloneqq \int_{x\in\mathcal{X}} \rho(x)\cdot\log\frac{\rho(x)}{\nu(x)}\mathrm{d}x.$$

KL divergence is also known as relative-entropy, due to its close connection to the Shanon entropy [Sha48]. KL divergence has several nice properties–Lemma 6 summarizes some of them.

Lemma 6. *The Kullback-Leibler (KL) divergence satisfies the following properties:*

1. *For all distributions ρ and ν, $H_\nu(\rho)\geq 0$.*
2. *For all distributions ρ and ν, $H_\nu(\rho)\geq 0$ if and only if $\rho=\nu$.*
3. *For all distributions ρ and ν, if ρ has density $h=\frac{\mathrm{d}\rho}{\mathrm{d}\nu}\colon\mathcal{X}\to\mathbb{R}$ w.r.t. ν, then*

$$H_\nu(\rho)=\mathop{\mathbb{E}}_{\nu}[h(x)\log h(x)].$$

Note that KL divergence is *not* a metric, in particular, it is not symmetric and does not atisfy the triangle inequality.

We are interested in KL divergence because of its relation to the total variation distance: For any distributions ρ and ν, it holds that

$$\mathrm{TV}\left(\rho,\nu\right)\leq\sqrt{\frac{1}{2}\cdot H_\nu(\rho)}. \qquad \text{(Pinsker's inequality)}$$

In the next lecture, we will use KL divergence and Pinsker's inequality to show that Metropolis Random Walk (MRW) satisfies the one-step overlap property (in total variation distance).

References

[Bal06] K. Ball. *Lecture.* Institut Henri Poincare, Paris, 2006.

[DCWY19] Raaz Dwivedi, Yuansi Chen, Martin J. Wainwright, and Bin Yu. Log-concave sampling: Metropolis-hastings algorithms are fast. *J. Mach. Learn. Res.*, 20:183:1–183:42, 2019.

[EL14] Ronen Eldan and Joseph Lehec. Bounding the norm of a log-concave vector via thin-shell estimates. In *Geometric Aspects of Functional Analysis*, pages 107–122. Springer, 2014.

[Eld13] Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013.

[FGP07] Bruno Fleury, Olivier Guédon, and Grigoris Paouris. A stability result for mean width of lp-centroid bodies. *Advances in Mathematics*, 214(2):865–877, 2007.

[Fle10] Bruno Fleury. Concentration in a thin euclidean shell for log-concave measures. *Journal of Functional Analysis*, 259(4):832–841, 2010.

[GM11] Olivier Guédon and Emanuel Milman. Interpolating thin-shell and sharp large-deviation estimates for lsotropic log-concave measures. *Geometric and Functional Analysis*, 21(5):1043, 2011.

[Kla07a] B Klartag. Power-law estimates for the central limit theorem for convex sets. *Journal of Functional Analysis*, 245(1):284–310, 2007.

[Kla07b] Bo'az Klartag. A central limit theorem for convex sets. *Inventiones mathematicae*, 168(1):91–131, 2007.

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

[Lov99] László Lovász. Hit-and-run mixes fast. *Mathematical Programming*, 86(3):443–461, 1999.

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

[LV07] László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & 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 *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 998–1007. IEEE, 2017.

[LV17b] Yin Tat Lee and Santosh Srinivas Vempala. Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In *FOCS*, pages 998–1007. IEEE Computer Society, 2017.

[LV18] Yin Tat Lee and Santosh S Vempala. The kannan-lov\'asz-simonovits conjecture. *arXiv preprint arXiv:1807.03465*, 2018.

[LV20] Aditi Laddha and Santosh Vempala. Convergence of gibbs sampling: Coordinate hit-and-run mixes fast. *arXiv preprint arXiv:2009.11338*, 2020.

[MV19] Oren Mangoubi and Nisheeth K. Vishnoi. Nonconvex sampling with the metropolis-adjusted langevin algorithm. In *COLT*, volume 99 of *Proceedings of Machine Learning Research*, pages 2259–2293. PMLR, 2019.

[NS20] Hariharan Narayanan and Piyush Srivastava. On the mixing time of coordinate hit-and-run. *CoRR*, abs/2009.14004, 2020.

[Sha48] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[Vem05] Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52(573-612):2, 2005.

[Vis21] Nisheeth K. Vishnoi. *Algorithms for Convex Optimization*. Cambridge University Pres, 2021.

[YTL19] Santosh Vempala Yin Tat Lee. *Techniques in Optimization and Sampling*. 2019.