Supplementary Material for A Polynomial Time MCMC Method for Sampling from Continuous DPPs

May 11, 2019

1 Proof of the Lowerbound on the Conductance in the Discrete case

We prove the following.

Theorem 1.1. *Let $\mathcal{M}$ be the Gibbs sampler chain for an arbitrary discrete k-DPP, then we have $\phi(\mathcal{M}) \geq \frac{1}{Ck^2}$, for a constant $C > 0$.*

Fix $\mathcal{M} = (\Omega, P, \pi)$ to be the Gibbs-sampler chain on a k-DPP defined on domain $[n]$, that is $\Omega = \binom{[n]}{k}$. Before discussing the details of the proof let us first fix some notation and recall fundamental properties of k-DPPs. For any element $1 \leq i \leq n$, define $\Omega_i, \Omega_{\bar{i}}$ be the set of all states in Ω that contain, do not contain i, respectively. Also define

$$\pi_i := \{\pi \mid i \text{ is chosen }\}, \text{ i.e. } \pi_i(\boldsymbol{x}) = \frac{\pi(\boldsymbol{x})}{\pi(\Omega_i)}, \forall x \in \Omega_i$$

$$\pi_{\bar{i}} := \{\pi \mid i \text{ is not chosen }\}, \text{ i.e. } \pi_{\bar{i}}(\boldsymbol{x}) = \frac{\pi(\boldsymbol{x})}{\pi(\Omega_{\bar{i}})}, \forall x \in \Omega_{\bar{i}}$$

It follows from [AGR16] that $\pi_i, \pi_{\bar{i}}$ can be identified with a $(k-1)$-DPP, k-DPP supported on $\Omega_i, \Omega_{\bar{i}}$, respectively. We define $\mathcal{M}_i = (\Omega_i, P_i, \pi_i), \mathcal{M}_{\bar{i}} = (\Omega_{\bar{i}}, P_{\bar{i}}, \pi_{\bar{i}})$ to be the *restricted* Gibbs samplers. So, it is straightforward to see that for any $x, y \in \Omega_i$ we get $P_i(\boldsymbol{x}, \boldsymbol{y}) = \frac{k}{k-1} \cdot P(\boldsymbol{x}, \boldsymbol{y})$. and consequently for Q_i defined as Q for $\mathcal{M}_i$, we get

$$Q_i(\boldsymbol{x}, \boldsymbol{y}) = \frac{Q(\boldsymbol{x}, \boldsymbol{y})}{\pi(\Omega_i)}. \tag{1.1}$$

Unlike P_i, $P_{\bar{i}}$ is not obtained from scaling a restriction of P. In particular, Let $\boldsymbol{x}, \boldsymbol{y} \in \Omega_{\bar{i}}$ so that $P_{\bar{i}}(\boldsymbol{x}, \boldsymbol{y}) > 0$ (which implies $|\boldsymbol{x} \cap \boldsymbol{y}| = k - 1$). Then, setting $I = \boldsymbol{x} \cap \boldsymbol{y}$ and with a bit abuse of notation $\pi(I) = \sum_{j \in [n] \setminus I} \pi(I + j)$, i.e. $\pi(I) = \mathbb{P}_{\boldsymbol{z} \sim \pi}[I \subset \boldsymbol{z}]$, we have

$$P_{\bar{i}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{k} \cdot \frac{\pi(\boldsymbol{y})}{\pi(I) - \pi(i + I)} \tag{1.2}$$

whereas $P(\boldsymbol{x}, \boldsymbol{y}) = \frac{\pi(\boldsymbol{y})}{k \cdot \pi(I)}$. For any $\boldsymbol{x} \in \Omega_i$, define $N_{\bar{i}}(\boldsymbol{x})$ be the set of its neighbours in $\Omega_{\bar{i}}$, i.e.

$$N_{\bar{i}}(\boldsymbol{x}) = \{\boldsymbol{y} \in \Omega_{\bar{i}} | P(\boldsymbol{x}, \boldsymbol{y}) > 0\}.$$

We use the following lemma to relate $Q_{\bar{i}}$ to Q.

Lemma 1.2. *Let $A \subset \Omega_{\bar{i}}$ be an arbitrary subset. For a state $\boldsymbol{x} \in \Omega_i$, consider the following partitioning of $N_{\bar{i}}(\boldsymbol{x})$: $N_A = N_{\bar{i}}(\boldsymbol{x}) \cap A$ and $N_{\overline{A}} = N_{\bar{i}}(\boldsymbol{x}) \cap (\Omega_{\bar{i}} \setminus A)$. Then we have*

$$Q(\boldsymbol{x}, N_A) + Q(N_A, N_{\overline{A}}) \geq \pi(\Omega_{\bar{i}}) \cdot Q_{\bar{i}}(N_A, N_{\overline{A}}). \tag{1.3}$$

Proof. Note that $\boldsymbol{x} \cup N_A \cup N_{\overline{A}}$ is the set of all states containing elements in $\boldsymbol{x} - i$. So by definition of Q and $Q_{\bar{i}}$, we have

$$Q(\boldsymbol{x}, N_A) + Q(N_A, N_{\overline{A}}) = \frac{1}{k} \cdot \frac{\pi(\boldsymbol{x})\pi(N_A)}{\pi(\boldsymbol{x}) + \pi(N_A) + \pi(N_{\overline{A}})} + \frac{1}{k} \cdot \frac{\pi(N_{\overline{A}})\pi(N_A)}{\pi(\boldsymbol{x}) + \pi(N_A) + \pi(N_{\overline{A}})} \tag{1.4}$$

$$= \frac{\pi(N_A)}{k} \cdot \frac{\pi(\boldsymbol{x}) + \pi(N_{\overline{A}})}{\pi(\boldsymbol{x}) + \pi(N_A) + \pi(N_{\overline{A}})} \geq \frac{\pi(N_A)}{k} \cdot \frac{\pi(N_{\overline{A}})}{\pi(N_A) + \pi(N_{\overline{A}})} = \pi(\Omega_{\bar{i}}) \cdot Q_{\bar{i}}(N_A, N_{\overline{A}}) \tag{1.5}$$

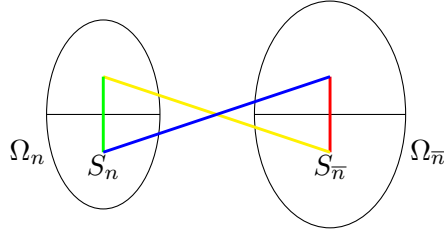where the inequality follows simply because $\pi(N_A) \geq 0$. $\qquad\square$

Figure 1: A schematic view of the restriction chains.
yellow, red, blue, and green edges correspond to $Q(S_n, \Omega_n \setminus S_n), Q(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}), Q(S_n, \Omega_{\overline{n}} \setminus S_{\overline{n}})$,
and $Q(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}})$, respectively

We use an inductive argument to prove $Q(S, \overline{S}) \geq \frac{\pi(S)}{Ck^2}$ for a subset $S \in \Omega$ with $\pi(S) \leq \frac{1}{2}$. Letting $S_n = S \cap \Omega_n$ and $S_{\overline{n}} = S \cap \Omega_{\overline{n}}$, we have

$$Q(S, \overline{S}) = Q(S_n, \Omega_n \setminus S_n) + Q(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) + Q(S_n, \Omega_{\overline{n}} \setminus S_{\overline{n}}) + Q(S_{\overline{n}}, \Omega_n \setminus S_n). \tag{1.6}$$

We carry out the induction step by lowerbounding the RHS of the above term by term. In order to bound $Q(S_n, \Omega_n \setminus S_n)$ we use induction hypothesis on $\mathcal{M}_n$. To bound $Q(S_{\overline{n}}, \Omega_n \setminus S_{\overline{n}})$, we combine the induction hypothesis on $\mathcal{M}_{\overline{n}}$ with Lemma 1.2. It remains to bound the other two terms which correspond to the contribution of the edge across $(\Omega_n, \Omega_{\overline{n}})$. To do that, we crucially use negative association of π. In particular, we use the following lemma (appeared before in [Mih92] in the unweighted case). For any set $A \in \Omega_n$, let $N_{\overline{n}}(A) = \{\boldsymbol{y} \in \Omega_{\overline{n}} : \exists \boldsymbol{x} \in A, \ P(\boldsymbol{x}, \boldsymbol{y}) > 0\}$ denote the set of neighbors of A in $\Omega_{\overline{n}}$.

Lemma 1.3 ([AGR16]). *For any subset $A \subseteq \Omega_n$,*

$$\pi_{\overline{n}}(N_{\overline{n}}(A)) \geq \pi_n(A).$$

The lemma lower bounds the vertex expansion of S_n in Ω_n and similarly vertex expansion of $S_{\overline{n}}$ in Ω_n. Later we show how to use it to bound the edge expansion which is our quantity of interest.

Proof of Theorem 1.1. We induct on $k + n$. So, assume, the conductance of the Gibbs sampler for any $(k-1)$-DPP over $n-1$ elements is at most $\frac{1}{C(k-1)^2}$ and the conductance is at most $\frac{1}{Ck^2}$ for any k-DPP over any $n-1$ elements.

Fix a set $S \subset \Omega$ where $\pi(S) \leq \frac{1}{2}$. We need to show $Q(S, \overline{S}) \geq \frac{\pi(S)}{Ck^2}$. First, consider a simple case where $\pi_n(S) \leq \frac{1}{2}$ and $\pi_{\overline{n}}(S) \leq \frac{1}{2}$. By induction hypothesis we have $Q_n(S_n, \Omega_n \setminus S_n) \geq \frac{\pi_n(S_n)}{c(k-1)^2}$. Moreover, by adding up (1.1) for the edges across the cut $(S_n, \Omega_n \setminus S_n)$, we get $Q(S_n, \Omega_n \setminus S_n) = \frac{(k-1)\pi(\Omega_n)}{k} \cdot Q_n(S_n, \Omega_n \setminus S_n)$. So combining them we have

$$Q(S_n, \Omega_n \setminus S_n) \geq \frac{\pi(S_n)}{Ck^2}. \tag{1.7}$$

Now, we use induction on $\mathcal{M}_{\overline{n}}$ along with Lemma 1.2. The induction hypothesis implies

$$Q_{\overline{n}}(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) \geq \frac{\pi_{\overline{n}}(S_{\overline{n}})}{ck^2} = \frac{\pi(S_{\overline{n}})}{\pi(\Omega_{\overline{n}}) \cdot ck^2}$$

So to prove the theorem in this case, it is enough to show the following and add it up with (1.7).

$$Q(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) + Q(S_{\overline{n}}, \Omega_n \setminus S_n) + Q(S_n, \Omega_{\overline{n}} \setminus S_{\overline{n}}) \geq \pi(\Omega_{\overline{n}}) \cdot Q_{\overline{n}}(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}). \tag{1.8}$$

To see that, it is enough to apply Lemma 1.2 and add up (1.3) for all $\boldsymbol{x} \in \Omega_n$, where subset $A \subset \Omega_{\overline{n}}$ in the lemma is determined as follows: if $\boldsymbol{x} \in S_n$ then set $A = S_{\overline{n}}$, otherwise set $A = \Omega_{\overline{n}} \setminus S_n$. Note that, doing that the RHS of the result will be exactly $\pi(\Omega_{\overline{n}}) \cdot Q_{\overline{n}}(S_{\overline{n}}, \Omega_{\overline{n}} \setminus \Omega_{\overline{n}})$, because any edge $\boldsymbol{yz}$ of that will only show up in (1.3) by having $\boldsymbol{x} = \boldsymbol{y} \cap \boldsymbol{z} + n$.

So we focus on the case $\max\{\pi_n(S_n), \pi_{\overline{n}}(S_{\overline{n}})\} > \frac{1}{2}$. Since $\pi(S) \leq \frac{1}{2}$, we have $\min\{\pi_n(S_n), \pi_{\overline{n}}(S_{\overline{n}})\} \leq \frac{1}{2}$.so So, without loss of generality, perhaps by considering $\overline{S}$ instead of S, we may assume $\pi_n(S_n) > \frac{1}{2}$ and $\pi_{\overline{n}}(S) \leq \frac{1}{2}$. Our goal is to prove

$$Q(S, \overline{S}) \geq \frac{1}{Ck^2} \cdot \min\{1 - \pi(S), \pi(S)\} \tag{1.9}$$

For every $\boldsymbol{x} \in \Omega_n$, let $N_{\overline{n}, S}(\boldsymbol{x}) := N_{\overline{n}}(\boldsymbol{x}) \cap S_{\overline{n}}$, and $N_{\overline{n}, \overline{S}}(\boldsymbol{x}) := N_{\overline{n}}(\boldsymbol{x}) \cap (\Omega_{\overline{n}} \setminus S_{\overline{n}})$ be a partitioning of $N_{\overline{n}}(\boldsymbol{x})$, so for every subset $T \in N_{\overline{n}}(\boldsymbol{x})$ we have

$$Q(\boldsymbol{x}, T) = \frac{1}{2k} \cdot \frac{\pi(\boldsymbol{x})\pi(T)}{\pi(\boldsymbol{x}) + \pi(N_{\overline{n}, S}(\boldsymbol{x})) + \pi(N_{\overline{n}, \overline{S}}(\boldsymbol{x}))} \tag{1.10}$$

Now, define $S_{\text{leave}} \subset S_n$ to be

$$S_{\text{leave}} = \{\boldsymbol{x} \in S_n : \pi(\boldsymbol{x}) + \pi(N_{\overline{n}, S}(\boldsymbol{x})) < \pi(N_{\overline{n}, \overline{S}}(\boldsymbol{x}))\},$$

in other words, $S_{\text{leave}} \in S_n$ is the subset of states so that, if the chain takes one step from S_{leave} by removing and resampling element n, then with probability at least $\frac{1}{2}$ it leaves S and enters $N_{\overline{n}, \overline{S}}(\boldsymbol{x})$. We also let $S_{\text{stay}} = S_n \setminus S_{\text{leave}}$. On the other hand, starting from S_{stay} and by resampling n, the chain with probability at least half stays in S. It is straight-forward to see

$$Q(S_{\text{leave}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) \geq \frac{\pi(S_{\text{leave}})}{4k} \tag{1.11}$$

3

To see that, note that definition of S_{leave} and setting $T = \Omega_{\overline{n}} \setminus S_{\overline{n}}$ in (1.10) implies that for any $\boldsymbol{x} \in S_{\text{leave}}$, we have $Q(\boldsymbol{x}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) \geq \frac{\pi(\boldsymbol{x})}{4k}$. To get (1.11), it suffices to sum up this over all states of S_{leave}. The bound (1.11) shows that $Q(S_{\text{leave}}, \overline{S}) \gg \frac{\pi(S_{\text{leave}})}{k^2}$. So roughly speaking, to prove the theorem, it suffices to show $\phi(S_{\text{stay}}) \cup S_{\overline{n}}) \geq \frac{1}{Ck^2}$. consider two cases: if $\pi_n(S_{\text{stay}}) \lesssim \frac{1}{2}$, we essentially use the same argument as in the case $\pi_n(S_n), \pi_{\overline{n}}(S_{\overline{n}}) \leq \frac{1}{2}$. Otherwise we combine the induction with Lemma 1.3 to bound the expansion.

- **Case 1:** $\pi_n(S_{\textbf{stay}}) \leq \frac{1}{2} + \frac{1}{4k}$. We show $Q(S, \overline{S}) \geq \frac{\pi(S)}{Ck^2}$. To do that, we use the induction hypothesis on $\mathcal{M}_n$, and the following claim which is the stronger version of (1.8).

    Claim 1.4.

    $$Q(S_{\overline{n}}, \overline{S}) + Q(S_n, \Omega_{\overline{n}} \setminus S_{\overline{n}}) - \frac{1}{2}Q(S_{leave}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) \geq \pi(\Omega_{\overline{n}}) \cdot Q_{\overline{n}}(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) \qquad (1.12)$$

    Proof. The claim is implied by combining the summation of (1.13),(1.14), and (1.15) over $\Omega_n \setminus S_n$, S_{stay} and S_{leave}, respectively. Let $\boldsymbol{x} \in \Omega_n \setminus S_n$. Then by applying Lemma 1.2 for $\boldsymbol{x}$ and $A = S_{\overline{n}}$, we get

    $$Q(N_{\overline{n},S}(\boldsymbol{x}), \{\boldsymbol{x}\} \cup N_{\overline{n},\overline{S}}(\boldsymbol{x})) \geq \pi(\Omega_{\overline{n}}) \cdot Q_{\overline{n}}(N_{\overline{n},S}(\boldsymbol{x}), N_{\overline{n},\overline{S}}(\boldsymbol{x})) \qquad (1.13)$$

    Similarly if $\boldsymbol{x} \in S_n$, by applying Lemma 1.2 for $\boldsymbol{x}$ and $A = \Omega_{\overline{n}} \setminus S_{\overline{n}}$, we have

    $$Q(\boldsymbol{x} \cup N_{\overline{n},S}(\boldsymbol{x}, N_{\overline{n},\overline{S}}(\boldsymbol{x})) \geq \pi(\Omega_{\overline{n}}) \cdot Q_{\overline{n}}(N_{\overline{n},S}(\boldsymbol{x}), N_{\overline{n},\overline{S}}(\boldsymbol{x})). \qquad (1.14)$$

    Finally, for $\boldsymbol{x} \in S_{\text{leave}}$, we have

    $$\begin{aligned} Q(N_{\overline{n},S}(\boldsymbol{x}), N_{\overline{n},\overline{S}}(\boldsymbol{x})) + \frac{1}{2}Q(\boldsymbol{x}, N_{\overline{n},\overline{S}}(\boldsymbol{x})) &= \frac{\pi(N_{\overline{n},\overline{S}}(\boldsymbol{x}))}{2k \cdot (\pi(\boldsymbol{x}) + \pi(N_{\overline{n},S}(\boldsymbol{x})) + \pi(N_{\overline{n},\overline{S}}(\boldsymbol{x})))} \cdot \left(\pi(N_{\overline{n},S}(\boldsymbol{x})) + \frac{\pi(\boldsymbol{x})}{2}\right) \\ &\geq \frac{1}{2k} \cdot \frac{\pi(N_{\overline{n},S}(\boldsymbol{x}))\pi(N_{\overline{n},\overline{S}}(\boldsymbol{x}))}{\pi(N_{\overline{n},S}(\boldsymbol{x})) + \pi(N_{\overline{n},\overline{S}}(\boldsymbol{x}))} \\ &= \pi(\Omega_{\overline{n}}) \cdot Q_{\overline{n}}(N_{\overline{n},S}(\boldsymbol{x}), N_{\overline{n},\overline{S}}(\boldsymbol{x})) \end{aligned}$$

    $$(1.15)$$

    where the inequality follows since $\pi(\boldsymbol{x}) + \pi(N_{\overline{n},S}(\boldsymbol{x})) < \pi(N_{\overline{n},\overline{S}}(\boldsymbol{x}))$ for $\boldsymbol{x} \in S_{\text{leave}}$. $\square$

    In particular, we use the above claim to get

    $$\begin{aligned} Q(S, \overline{S}) &= Q(S_n, \Omega_n \setminus S_n) + Q(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) + Q(S_n, \Omega_{\overline{n}} \setminus S_{\overline{n}}) + Q(S_{\overline{n}}, \Omega_n \setminus S_n) \\ &\geq Q(S_n, \Omega_n \setminus S_n) + \frac{1}{2}Q(S_{\text{leave}} + \Omega_{\overline{n}} \setminus S_{\overline{n}}) + \pi(\Omega_{\overline{n}})Q_{\overline{n}}(S_{\overline{n}}, \overline{S_{\overline{n}}}) \quad \text{By Claim 1.4} \\ &\geq \frac{\pi(\Omega_n) - \pi(S_n)}{Ck(k-1)} + \frac{1}{2}Q(S_{\text{leave}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) + \frac{\pi(S_{\overline{n}})}{Ck^2} \quad \text{induction Hyp. on } \mathcal{M}_n \text{ and } \mathcal{M}_{\overline{n}} \\ &\geq \frac{\pi(\Omega_n) - \pi(S_{\text{leave}}) - \pi(S_{\text{stay}})}{Ck(k-1)} + \frac{\pi(S_{\text{leave}})}{8k} + \frac{\pi(S_{\overline{n}})}{Ck^2} \quad \text{By (1.11) and } S_n = S_{\text{leave}} \cup S_{\text{stay}} \end{aligned}$$

    $$(1.16)$$

    To finish the proof, we need to show the RHS of the above is at least $\frac{\pi(S)}{Ck^2}$. To see that note that since $\frac{\pi(S_{\text{leave}})}{8k} \geq \pi(S_{\text{leave}}) \cdot \left(\frac{1}{Ck^2} + \frac{1}{Ck(k-1)}\right)$ for sufficiently large k, it suffices to show $\frac{\pi(\Omega_n) - \pi(S_{\text{stay}})}{Ck(k-1)} \geq \frac{\pi(S_{\text{stay}})}{Ck^2}$, which can be directly verified for $\pi_n(S_{\text{stay}}) \leq \frac{1}{2} + \frac{1}{4k}$.

4

- **Case 2:** $\pi_n(S_{\mathbf{stay}}) > \frac{1}{2} + \frac{1}{4k}$. We prove

$$Q(S, \bar{S}) \geq \frac{1 - \pi(S)}{Ck^2}.$$

Lemma 1.3 states that the vertex expansion of S_{stay} is proportional to $\pi_n(S_{\text{stay}}) - \pi_{\overline{n}}(S_{\overline{n}})($ which is positive in this case by the assumption). We use it to bound $Q(S, \overline{S})$ by relating vertex expansion of S_{stay} to $Q(S, \overline{S})$. In particular, we show the following claim.

Claim 1.5.

$$Q(S_{stay}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) + Q(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) \geq \frac{\pi(\Omega_{\overline{n}})}{2k} \cdot (\pi_n(S_{stay}) - \pi_{\overline{n}}(S_{\overline{n}}))$$

Proof. Note that for any $\boldsymbol{x} \in S_{\text{stay}}$, since $\pi(N_{\overline{n}, \overline{S}}(\boldsymbol{x})) \leq \pi(\boldsymbol{x}) + \pi(N_{\overline{n}, S}(\boldsymbol{x}))$, we have

$$Q(\boldsymbol{x}, N_{\overline{n}, \overline{S}}(\boldsymbol{x})) + Q(N_{\overline{n}, S}(\boldsymbol{x}), N_{\overline{n}, \overline{S}}(\boldsymbol{x})) = \frac{1}{2k} \cdot \frac{\pi(N_{\overline{n}, \overline{S}}(\boldsymbol{x})) \cdot [\pi(\boldsymbol{x}) + \pi(N_{\overline{n}, S}(\boldsymbol{x}))]}{\pi(\boldsymbol{x}) + \pi(N_{\overline{n}, S}(\boldsymbol{x})) + \pi(N_{\overline{n}, \overline{S}}(\boldsymbol{x}))} \geq \frac{1}{2k} \cdot \frac{\pi(N_{\overline{n}, \overline{S}}(\boldsymbol{x}))}{2},$$

To complete the proof, it is enough to sum up the above over S_{stay} to get the following

$$
\begin{aligned}
Q(S_{\text{stay}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) + Q(S_{\overline{n}}, \Omega_{\overline{n}} \setminus S_{\overline{n}}) &\geq \sum_{x \in S_{\text{stay}}} \frac{\pi(N_{\overline{n}, \overline{S}}(\boldsymbol{x}))}{4k} \geq \pi \left(\bigcup_{\boldsymbol{x} \in S_{\text{stay}}} N_{\overline{n}, \overline{S}}(\boldsymbol{x}) \right). \\
&\geq \pi(N_{\overline{n}}(S_{\text{stay}})) - \pi(S_{\overline{n}}) \\
&\geq \pi(\Omega_{\overline{n}}) \cdot (\pi_n(S_{\text{stay}}) - \pi_n(S_{\overline{n}})) \quad \text{By Lemma 1.3}
\end{aligned}
$$

$\square$

Claim 1.5 and (1.16) implies $Q(S, \overline{S}) \geq \max\{L_1, L_2\}$ defined as above

$$L_1 := \frac{\pi(S_{\text{leave}})}{8k} + \frac{\pi(\Omega_n) - \pi(S_{\text{leave}}) - \pi(S_{\text{stay}})}{Ck(k-1)} + \frac{\pi(S_{\overline{n}})}{Ck^2} \qquad \text{By (1.16)}$$

$$L_2 := \frac{\pi(\Omega_{\overline{n}})}{4k} \cdot (\pi_n(S_{\text{stay}}) - \pi_{\overline{n}}(S_{\overline{n}})) \qquad \text{By Claim 1.5.}$$

So we need to prove $\max\{L_1, L_2\} \geq \frac{1 - \pi(S)}{Ck^2}$. To prove that, we show that $L_1 + \frac{L_2}{k-1} \geq (1 + \frac{1}{k-1}) \cdot \frac{1 - \pi(S)}{Ck^2}$. Replacing values of L_1 and L_2 in the above and simplifying the resulting inequality, we need to show

$$\frac{\pi(S_{\text{leave}})}{8k} + \frac{\pi(S_{\overline{n}})}{Ck^2} + \frac{\pi(\Omega_{\overline{n}})}{4k(k-1)} \cdot (\pi_n(S_{\text{stay}}) - \pi_{\overline{n}}(S_{\overline{n}})) \geq \frac{\pi(\Omega_{\overline{n}}) - \pi(S_{\overline{n}})}{Ck(k-1)}.$$

Ignoring the $\frac{\pi(S_{\text{leave}})}{8k}$ term and rearranging the other terms, it is enough to show

$$\frac{\pi(\Omega_{\overline{n}})}{4k(k-1)} \cdot (\pi_{\overline{n}}(S_{\text{stay}}) - \pi_{\overline{n}}(S_{\overline{n}})) \geq \frac{\pi(\Omega_{\overline{n}})}{Ck(k-1)} \cdot (1 - \frac{2k-1}{k} \cdot \pi_{\overline{n}}(S_{\overline{n}})).$$

The above can be verified for $C > 16$, by noting that by assumption $\pi_n(S_{\text{stay}}) \geq \frac{1}{2} + \frac{1}{4k}$ and $\pi_{\overline{n}}(S_{\overline{n}}) \leq \frac{1}{2}$.

$\square$

2 Markov Chains with Measurable State Space

Here, we provide a more formal overview of Markov chains defined on measurable state spaces. More details, can also be found at [LS93]. Let $(\Omega, \mathcal{B})$ be a measurable space. In the most general setting, a Markov chain is defined by the triple $(\Omega, \mathcal{B}, \{P_x\}_{x \in \Omega})$, where for every $x \in \Omega$, $P_x : \mathcal{B} \to \mathbb{R}_+$ is a probability measure on $(\Omega, \mathcal{B})$. Also, for every fixed $B \in \mathcal{B}$, $P_x(B)$ is a measurable function in terms of x. In this setting starting from a distribution μ_0, after one step the distribution μ_1 would be given by

$$\mu_1(B) = \int_\Omega P_x(B) d\mu_0(x), \ \forall B \in \mathcal{B}.$$

From now on, assume $\Omega \subset \mathbb{R}^k$ and $\mathcal{B}$ is the standard Borel σ-algebra. In our setting, we can assume the transition probabilities are given by a kernel *transition kernel* $P : \Omega \times \Omega \to \mathbb{R}_+$ where for any measurable $A \subset \Omega$, we can write

$$P_x(A) = \int_A P(x,y) dy.$$

In this notation, we use $P(x, B)$ and $P_x(B)$ interchangeably. $P^n(x, .)$ would also denote the probability distribution of the states after n steps of the chain started at x. Similar to the discrete setting, we can define the *stationary measure* for the chain. A probability distribution π on Ω is stationary if and only if for every measurable set B, we have

$$\pi(B) = \int_\Omega \int_B P(x,y) dy d\pi(x).$$

We call $\mathcal{M}$ *ϕ-irreducible* for a probability measure ϕ if for any set $B \in \mathcal{B}$ with $\phi(B) > 0$, and any state x, there is $t \in \mathbb{N}$ such that $P^t(x, B) > 0$. It is called *strongly ϕ-irreducible* if for any $B \subseteq \Omega$ with non-zero measure and $x \in \Omega$, there exists $t \in \mathbb{N}$ such that for any $m \geq t$, $P^m(x, B) > 0$. We say $\mathcal{M}$ is *reversible* with respect to a measure π if for any two sets A and B we have

$$\int_B \int_A P(y,x) dx d\pi(y) = \int_A \int_B P(x,y) dy d\pi(x).$$

In particular, reversibility with respect to a measure, implies it is a stationary measure. Is is immediate from this to verify that for a Gibbs sampler of a k-DPPs π, the π itself is the stationary measure. Moreover, if the kernel of the k-DPP is continuous, it is straight-forward to see that it is π-strongly irreducible. The following lemma also shows π is the unique stationary measure, and as the number of steps increases, the chain approaches to the unique stationary measure.

Lemma 2.1 ([DF97])**.** *If π is a stationary measure of $\mathcal{M}$, and $\mathcal{M}$ is strongly π-irreducible. Then for any other distribution μ which is absolutely continuous with respect to π, $\lim_{n\to\infty} |P^n(\mu, .) - \pi|_{TV} = 0$.*

From now on, consider $\mathcal{M} = (\Omega, P, \pi)$ is chain with state space Ω, probability transition function P, and a unique stationary measure π. Let us describe some results about mixing time in the Markov chains defined on continuous spaces. But before that we need to setup some notation. Consider a Hilbert space $L^2(\Omega, \pi)$ equipped with the following inner product.

$$\langle f, g \rangle_\pi = \int_\Omega f(x) g(x) d\pi(x).$$

P defines an operator in this space where for any function $f \in \ell^2(\Omega, \pi)$ and $x \in \Omega$,

$$(Pf)(x) = \int_\Omega P(x, y) f(y) dy.$$

In particular $\mathcal{M}$ being reversible is equivalent to P being self-adjoint. For a reversible chain $\mathcal{M}$ and a function $f \in L^2(\Omega, \pi)$, the Dirichlet form $\mathcal{E}_P(f, f)$ is defined as

$$\mathcal{E}_P(f, f) = \frac{1}{2} \int_\Omega \int_\Omega (f(x) - f(y))^2 P(x, y) d\pi(x) dy.$$

We also define the *Variance* of f with respect to π as

$$\mathrm{var}_\pi(f) := \int_\Omega (f(x) - \mathbb{E}_\pi(f))^2 d\pi(x).$$

We may drop the subscript if the underlying stationary distribution is clear in the context. One way for upperbounding the mixing time of a chain is to use is to its spectral gap which is also known as *Poincaré Constant*.

Definition 2.2 (Poincaré Constant)**.** *. The Poincaré constant of the chain is defined as follows,*

$$\lambda := \inf_{f : \pi \to \mathbb{R}} \frac{\mathcal{E}_P(f, f)}{\mathrm{var}_\pi(f)},$$

where the infimum is only taken over all functions in $L^2(\Omega, \pi)$ with non-zero variance.

In this paper, we use the following theorem to upperbound the mixing time of the chain relevant to us.

Theorem 2.3 ([KM12])**.** *For any reversible, lazy, and π-irreducible Markov chain $M = (\Omega, P, \pi)$, if $\lambda > 0$, then the distribution of the chain started from μ (which is absolute continuous with respect to π) is*

$$\|P^t(\mu, .) - \pi\|_{TV} \leq \frac{1}{2}(1 - \lambda)^t \sqrt{\mathrm{var}_\pi\left(\frac{f_\mu}{f_\pi}\right)}.$$

For the sake of completeness, we include a proof of the above theorem which is an extension of the proof of the analogous discrete result in [Fil91]. We need the following simple lemma known as Mihail's identity.

Lemma 2.4 (Mihail's identity, [Fil91])**.** *For any reversible irreducible Markov chain $\mathcal{M} = (\Omega, P, \pi)$, and any function f in $L^2(\pi)$,*

$$\mathrm{var}_\pi(f) = \mathrm{var}_\pi(Pf) + \mathcal{E}_{P^2}(f, f).$$

Proof of Theorem 2.3. First of all, one can easily verify that if a chain is lazy and irreducible, then it is strongly-irreducible. Combining it with Lemma 2.1 would guarantee the uniqueness of the stationary measure. Let $\mu_0 = \mu$ be the starting distribution and define $\mu_t = P^t(\mu, .)$ be the distribution at time t. Set $f_t := \frac{f_{\mu_t}}{f_\pi}$, by reversibility of the chain we have

$$(Pf_t)(x) = \int_\Omega P(x, y) \frac{f_{\mu_t}(y)}{f_\pi(y)} dy = \int_\Omega \frac{P(y, x) f_{\mu_t}(y)}{f_\pi(x)} dy = \frac{f_{\mu_{t+1}}}{f_\pi}(x) = f_{t+1}(x)$$

7

which implies
$$\text{var}_\pi(Pf_t) = \text{var}_\pi(f_{t+1}) \qquad (2.1)$$

So applying Mihail's identity on $\frac{f_{\mu_n}}{f_\pi}$ and using (2.1) , we conclude
$$\text{var}_\pi(f_t) = \text{var}_\pi(f_{t+1}) + \mathcal{E}_{P^2}(f_t, f_t). \qquad (2.2)$$

Now, note that P^2 has the same stationary distribution π, so its Poincaré constant is at most
$$\lambda(P^2) \le \frac{\mathcal{E}_{P^2}(f_t, f_t)}{\text{var}_\pi(f_t)}.$$

Combining this with (2.2), and using induction we can deduce
$$\text{var}_\pi(f_t) \le (1 - \lambda(P^2))^t \, \text{var}_\pi(f_0).$$

Note that, since P is the kernel for a lazy chain, it has no negative values in its spectrum, implying $1 - \lambda(P^2) = (1 - \lambda(P))^2$. So in order to complete the proof it is enough show
$$4\|\mu_t - \pi\|_{TV}^2 \le \text{var}(f_t).$$

This can be seen using an application of Cauchy-Schwarz's inequality. We have
$$\begin{aligned}
4\|\mu_t - \pi\|_{TV}^2 &= \left(\int_\Omega |f_{\mu_t}(x) - f_\pi(x)| dx \right)^2 \\
&= \left(\int_\Omega f_\pi(x) \left| \frac{f_{\mu_t}(x)}{f_\pi(x)} - 1 \right| dx \right)^2 \\
&\le \int_\Omega f_\pi(x) \left| \frac{f_{\mu_t}(x)}{f_\pi(x)} - 1 \right|^2 dx = \text{var}\left(\frac{f_{\mu_t}}{f_\pi}\right)
\end{aligned}$$

The last identity uses that $\mathbb{E}_\pi \frac{f_{\mu_t}}{f_\pi} = 1$. This completes the proof. $\qquad\square$

In order to take advantage of Theorem 2.3, we need to lowerbound the Poicaré constant of our chain. This can be done by lowerbounding the *Ergodic Flow* of the chain.

Definition 2.5 (Ergodic Flow). *For a chain $\mathcal{M} = (\Omega, P, \pi)$, the ergodic flow $Q : \mathcal{B} \to [0, 1]$ is defined by*
$$Q(B) = \int_B \int_{\Omega \setminus B} P(u, v) dv f_\pi(u) du.$$

The *conductance* of a set B is defined by, $\phi(B) := \frac{Q(B)}{\pi(B)}$, and the conductance of the chain is
$$\phi(\mathcal{M}) = \min_{0 < \pi(B) \le \frac{1}{2}} \phi(B).$$

The following theorem which is an extension of the Cheeger's inequality for the Markov chains on a continuous space, relates the spectral gap to conductance.

Theorem 2.6 ([LS88])**.** *For a chain $\mathcal{M}$ defined on a general state space with spectral gap λ we have*
$$\frac{\phi(\mathcal{M})^2}{8} \le \lambda \le 2\phi(\mathcal{M}).$$

8

3   Conductance of the Continuous Case

Let $\mathcal{M}$ be the Gibbs sampler for a k-DPP defined by a continuous kernel L. As the rest of the paper, we assume L is a continuous, symmetric, and PSD kernel where $\int_C |L(x,x)|dx < \infty$. We prove the following.

Theorem 3.1. *Let $\mathcal{M}$ be the Gibbs sampler for a k-DPP defined by $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, then*

$$\phi(\mathcal{M}) \gtrsim \frac{1}{k^2}. \tag{3.1}$$

Proof. Recall that by Theorem 1.1 the conductance of a Gibbs sampler for any discrete k-DPP is at least $\Omega(\frac{1}{k^2})$. The key observation is that this bound is independent of the number of states. Therefore, we can obtain this bound for arbitrarily fine discretizations of $\mathcal{M}$, and with a limiting argument extend it to $\mathcal{M}$.

For simplicity, we assume $d = 1$. It is straight-forward to extend the argument to higher dimensions. Let us denote the state space by Ω. Fix a measurable subset $S \subset \Omega$ with $\pi(S) \leq \frac{1}{2}$. Our goal is to prove $\phi(S) = \frac{Q(S,\overline{S})}{\pi(S)} \geq \Omega(\frac{1}{k^2})$. Without loss of generality, we can only consider restriction of Ω and S to a bounded set. To see that, note that if we set $\Omega_n = \binom{[-n,n]}{k}$, then clearly, $\lim_{n\to\infty} \frac{Q(S\cap\Omega_n,\overline{S}\cap\Omega_n)}{\pi(S\cap\Omega_n)} = \phi(S)$, and so for large values of n, $\frac{Q(S\cap\Omega_n,\overline{S})}{\pi(S\cap\Omega_n)} = \Theta(\phi(S))$. So suppose that $\Omega = \binom{[0,1]}{k}$. For an integer n, we consider a discretization $\mathcal{M}_n$ of $\mathcal{M}$ defined as follows. We use n in subscript to denote quantities related to $\mathcal{M}_n$. We partition $[0,1]$ into intervals of length $\frac{1}{n}$, and identify each interval with an element in the ground set of $\mathcal{M}_n$, so $\Omega_n = \binom{[n]}{k}$. $\mathcal{M}_n$ is defined by a kernel L_n characterized below. For $i \in [n]$ let $I_i = [\frac{i-1}{n}, \frac{i}{n}]$. For any $i, j \in [n]$, we define $L_n(i,j) = \int_{I_i} \int_{I_j} L(u,v)dudv$, be the accumulative value of L over $I_i \times I_j$. One can easily see L_n is a PSD matrix, as L is a PSD operator. Moreover, L and consequently $\det_L$ is a continuous function on a closed domain, so it is uniform continuous, implying for any $\epsilon > 0$, there exists an integer $n(\epsilon)$ so that for all $n > n(\epsilon)$ and any two states $\{x_1, \ldots, x_k\}$ and $\{y_1, \ldots, y_k\}$ with $|y_i - x_i| \leq \frac{1}{n}$, we have $|\det_L(x_1, \ldots, x_k) - \det_L(y_1, \ldots, y_k)| \leq \epsilon$. Now, note that $f_\pi(y_1, \ldots, y_k) = \frac{\det_L(y_1,\ldots,y_k)}{\frac{1}{k!}\int \det_L(x_1,\ldots,x_k)dx_1\ldots dx_k}$. So, using the simple fact that for any two sequences of numbers $\{a_n\}$ and $\{b_n\}$,

$$\left(\lim_{n\to\infty} a_n = a\right) \wedge \left(\lim_{n\to\infty} b_n = b \neq 0\right) \implies \lim_{n\to\infty} \frac{a_n}{b_n} = \frac{a}{b} \tag{3.2}$$

we get that for any $\epsilon > 0$, there exists an integer $m(\epsilon)$, where $m(\epsilon)$ depends on $n(\epsilon)$, such that

$$\forall n \geq m(\epsilon), \forall \{t_1, \ldots, t_k\} \in \binom{[n]}{k} : \left|\pi_n(t_1, \ldots, t_k) - \pi(\prod_{i=1}^k I_{t_i})\right| \leq \frac{\epsilon}{n^k} \tag{3.3}$$

We define a set $S_n \subset \Omega_n$ corresponding to S for any n, so that

$$\lim_{n\to\infty} \phi_n(S_n) = \phi(S). \tag{3.4}$$

Clearly, the above proves the theorem as by Theorem 1.1, we know that $\phi_n(S_n) \gtrsim \frac{1}{k^2}$ for any n. In what follows, we use $A \subset B$ to denote both of $A - B$ and $B - A$ have Lebesgue measure zero. Also, define

$$S_n = \left\{ \{t_1, \ldots, t_k\} \in \binom{[n]}{k} \,\middle|\, I_{t_1} \times \cdots \times I_{t_k} \subset S \right\}.$$

9

Following (3.2), to prove (3.4), it is enough to argue that $\lim_{n\to\infty} Q_n(S_n, \overline{S_n}) = Q(S, \overline{S})$, and $\lim_{n\to\infty} \pi_n(S_n) = \pi(S)$. We first show the latter. This follows by (3.3) and that

$$\lim_{n\to\infty} \mu \left(\cup_{\{t_1,\dots,t_k\} \in S_n} \prod_{i=1}^{k} I_{t_i} \right) = \mu(S) \tag{3.5}$$

for μ being the Lebesgue measure.

It remains to see $\lim_{n\to\infty} Q_n(S_n, \overline{S_n}) = Q(S, \overline{S})$. First, note that $[0,1]^{k-1}$ is a closed set, so for any $\delta > 0$ and $\epsilon > 0$, there exists an integer $n(\delta, \epsilon)$ so that for any $n > n(\delta, \epsilon)$, and points $x_1, \dots, x_k, x_{k+1}$ and $y_1, \dots, y_k, y_{k+1}$ with $|x_i - y_i| \leq \frac{1}{n}$, and $\int_0^1 \det_L(x_1, \dots, x_{k-1}, \tau) d\tau \geq \delta$, we have

$$\left| \frac{\det_L(x_1, \dots, x_k) \det_L(x_1, \dots, x_{k-1}, x_{k+1})}{\int_0^1 \det_L(x_1, \dots, x_{k-1}, \tau) d\tau} - \frac{\det_L(y_1, \dots, y_k) \det_L(y_1, \dots, y_{k-1}, y_{k+1})}{\int_0^1 \det_L(y_1, \dots, y_{k-1}, \tau) d\tau} \right| \leq \epsilon.$$

Therefore, similar to the case for π_n, it follows that for any $\epsilon, \delta > 0$, there exists integer $m(\delta, \epsilon)$ depending on $n(\delta, \epsilon)$ so that for any $n \geq m(\delta, \epsilon)$ and for all $t_1, \dots, t_{k-1}, s, t \in \binom{[n]}{k+1}$ with $\sum_{i=1}^{n} \pi_n(t_1, \dots, i) \geq \frac{\delta}{n^{k-1}}$

$$\left| Q_n(\{t_1, \dots, t_{k-1}, t\}, \{t_1, \dots, t_{k-1}, s\}) - Q(I_t \times \prod_{i=1}^{k-1} I_{t_i}, I_s \times \prod_{i=1}^{k-1} I_{t_i}) \right| \leq \frac{\epsilon}{n^{k+1}}. \tag{3.6}$$

Now, combining the above equation with (3.5), and noting ϵ and δ can be chosen arbitrary close to zero, we obtain $\lim_{n\to\infty} Q_n(S_n, \overline{S_n}) = Q(S, \overline{S})$, which completes the proof. $\qquad\square$

Remark 3.2. *It is straight-forward to use a similar discretization argument to prove Theorem 3.1, when the domain of the kernel is restricted to a closed subset $C \subset \mathbb{R}^d$ which can be nicely discretized as in Theorem 3.1. In particular, we assume C is an sphere in the next section. More precisely, C could be any closed subset which has the same measure as its interior.*

We also point out that, irreducibly of the chain can also be deduced in the same way; More precisely, it is shown [Brä07] that the support of the chain corresponding to a discrete k-DPP is the set of bases of a matroid [Brä07, Cor 3.4]; so the chain is irreducible in the discrete case. It can be extended to the continuous case with the same limiting argument.

4 Proof of the Lemma 4.3 on Starting Distribution

Recall that we use a randomized greedy algorithm to find a starting state and we state the following lemma in the analysis of our algorithm in the paper.

Lemma 4.1. *Let ν be the probability distribution of the output of our to find the starting state. Also let f_ν and f_π denote the p.d.f. for ν and π which is the k-DPP defined by L. Then, for any $\{x_1, \dots, x_k\} \subset C$,*

$$f_\nu(\{x_1, \dots, x_k\}) \leq (k!)^2 f_\pi(\{x_1, \dots, x_k\}).$$

As stated the proof essentially follows from a similar argument in [DV07] which appears in a discrete setting. However, for the sake of completeness we provide a full proof here.

Proof. For any $x \in \mathbb{R}^d$, let f_x be the corresponding feature map, i.e. $f_x : \mathcal{H} \to \mathbb{R}$ for some Hilbert space $\mathcal{H}$ and for any $x, y \in \mathbb{R}^d$, $L(x, y) = \langle f_x, f_y \rangle$. Fix $\boldsymbol{x} = \{x_1, \ldots, x_k\}$, and let S_k be the set of all permutations of $\{x_1, \ldots, x_k\}$. Also, for any $\sigma \in S_k$ and for any $1 \leq i \leq k - 1$, define $H_\sigma^i = \langle f_{\sigma(1)}, \ldots, f_{\sigma(i)} \rangle$. In the above the range of all integrals is $\mathbb{R}^d$. We have

$$f_\nu(\boldsymbol{x}) = \sum_{\sigma \in S_k} \left[\frac{\|f_{\sigma(1)}\|^2}{\int \|f_y\|^2 \, dy} \cdot \frac{d(f_{\sigma(2)}, H_\sigma^1)^2}{\int d(f_y, H_\sigma^1)^2 dx} \cdots \frac{d(f_{\sigma(k)}, H_\sigma^{k-1})^2}{\int d(f_y, H_\sigma^{k-1})^2 dy} \right].$$

Note that the above integrals are well-defined since our kernel is continuous. For any $1 \leq i \leq k-1$, let $H_*^i = \arg\min_{H = \langle f_{y_1}, \ldots, f_{y_i} \rangle} \int d(f_y, H)^2 dy$, where $y_1 \ldots, y_i$ range over $\mathbb{R}^d$. Note that, the minimum of the quantity is defined since L is continuous on a closed set. Combining with the above, and noting that for any σ, $\det(x_1, \ldots, x_k) = \|f_{\sigma(1)}\|^2 \cdot d(f_{\sigma(2)}, H_\sigma^1)^2 \ldots d(f_{\sigma(k)}, H_\sigma^{k-1})^2$, we obtain

$$f_\nu(\boldsymbol{x}) \leq k! \cdot \frac{\det(x_1, \ldots, x_k)}{\int \|f_y\|^2 \, dy \cdot \int d(f_y, H_*^1)^2 dy \cdot \int d(f_y, H_*^{k-1} dy}$$
$$\leq k! \cdot \frac{f_\pi(\boldsymbol{x}) \cdot \int \cdots \int_C \det(y_1, \ldots, y_k) dy_k \ldots dy_1}{k! \cdot \int \|f_y\|^2 \, dy \cdot \int d(f_y, H_*^1)^2 dx \cdots \int d(f_y, H_*^{k-1})^2 dy}.$$

So, rearranging the above to show $\frac{f_\nu(\boldsymbol{x})}{f_\pi(\boldsymbol{x})} \leq (k!)^2$, it suffices to show

$$\frac{\int \cdots \int \det(y_1, \ldots, y_k) dy_k \ldots dy_1}{\int \|f_y\|^2 \, dy \cdot \int d(f_y, H_*^1)^2 dx \cdots \int d(f_y, H_*^{k-1})^2 dy} \leq (k!)^2. \tag{4.1}$$

To proof the above, we use induction on k. For $k = 1$, the statement is obvious as for any $y \in \mathbb{R}^d$, $\det(y) = L(y, y) = \|f_y\|^2$. It is straight-forward to see, applying the above claim will prove the induction step, and completes the proof.

Claim 4.2.

$$\int \cdots \int \det(y_1, \ldots, y_k) dy_k \ldots dy_1 \leq k^2 \left(\int d(f_y, H_*^{k-1})^2 dy \right) \left(\int \cdots \int \det(y_1, \ldots, y_{k-1}) dy_{k-1} \ldots dy_1 \right) \tag{4.2}$$

Proof of Claim 4.2. For any $\boldsymbol{y} = \{y_1, \ldots, y_k\} \subset \mathbb{R}^d$, let $G_{\boldsymbol{y}}$ be a $(k-1)$-dimensional linear subspace of $\langle f_{y_1} \ldots, f_{y_k} \rangle$ which contains the projection of $H_*^{(k-1)}$ onto $\langle f_{y_1} \ldots, f_{y_k} \rangle$. Now, for any $\boldsymbol{y}$, using Lemma 4.3, we get

$$\det(\boldsymbol{y}) \leq \left(\sum_{i=1}^k d(f_{y_i}, G_{\boldsymbol{y}}) \sqrt{\det(\boldsymbol{y} - y_i)} \right)^2$$
$$\leq k \left(\sum_{i=1}^k d(f_{y_i}, G_{\boldsymbol{y}})^2 \det(\boldsymbol{y} - y_i) \right) \quad \text{Cauchy-Schwarz Inequality.}$$

By integerating the above, we get

$$\int \cdots \int \det(\boldsymbol{y})d\boldsymbol{y} \leq k \int_{\mathbb{R}^d} \cdots \int \sum_{i=1}^{k} d(f_{y_i}, G_{\boldsymbol{y}})^2 \det(\boldsymbol{y} - y_i)d\boldsymbol{y}$$

$$\leq k^2 \int_{y \in \mathbb{R}^d} \int_{z_1 \in \mathbb{R}^d} \cdots \int_{z_{k-1} \in \mathbb{R}^d} d(f_y, G_{\boldsymbol{z}+y})^2 \det(\boldsymbol{z})dzdy \ \ (\text{setting } \boldsymbol{z} = \{z_1, \ldots, z_{k-1}\})$$

$$\leq \int_{y \in \mathbb{R}^d} \int_{z_1 \in \mathbb{R}^d} \cdots \int_{z_{k-1} \in \mathbb{R}^d} d(f_y, H_*^{k-1})^2 \det(\boldsymbol{z})dzdy$$

$$= \left(\int d(f_y, H_*^{k-1})^2 dy \right) \left(\int_{z_1 \in \mathbb{R}^d} \cdots \int_{z_{k-1} \in \mathbb{R}^d} \det(\boldsymbol{z})d\boldsymbol{z} \right),$$

where in the third inequality, the fact $d(f_y, G_{\boldsymbol{z}+y}) \leq d(f_y, H_*^{k-1})$ holds because $f_y \in \langle f_{z_1}, \ldots, f_{z_{k-1}}, f_y \rangle$, and $G_{\boldsymbol{z}+y}$ contains the projection of H_*^{k-1} onto this space. Thus, the proof of the claim and the theorem is complete. □

□

Lemma 4.3 (Lemma 2 of [DV07]). *Let S be a set of k vectors, and H be any $(k-1)$-dimesnsional subspace of $\langle S \rangle$. Then*

$$\text{vol}(S) \leq \sum_{v \in S} d(v, H) \text{vol}(S - v),$$

where volume of a set of vectors, refer to the volume of the parallelopiped spanned by them.

5   Missing Proofs for the Conditional Sampling

Here, we include the analysis of our proposed rejection sampler for polynomial and spherical Gaussian kernels and provide the proof for Lemma 5.3 of the paper. The eigenvalues for Spherical Gaussian kernels defined on a unit sphere is explicitly given in [MNY06].

Lemma 5.1 ([MNY06]). *Let $\mathcal{G}_\sigma$ be the Gaussian kernel with variance σ^2 restricted to the unit sphere with the uniform measure, i.e. for any $x, y \in \mathbb{S}^{d-1}$: we have $\mathcal{G}_\sigma(x, y) = \frac{\exp(-\|x-y\|^2/\sigma^2)}{\text{vol}(S^{d-1})}$. For any integer $k \geq 0$, $\mathcal{G}_\sigma$ has an eigenvalue μ_k with multiplicity $N(d, k) = \frac{(2k+d-2)(k+d-3)!}{k!(d-2)!}$ where*

$$\left(\frac{2e}{\sigma^2} \right)^k \cdot \frac{A_1}{(2k+d-2)^{k+\frac{d-1}{2}}} \leq \mu_k \leq \left(\frac{2e}{\sigma^2} \right)^k \cdot \frac{A_2}{(2k+d-2)^{k+\frac{d-1}{2}}}, \tag{5.1}$$

for $A_1 = e^{-\frac{2}{\sigma^2} - \frac{1}{12}} \frac{1}{\sqrt{\pi}} (2e)^{\frac{d}{2}-1} \Gamma \left(\frac{d}{2} \right)$ and $A_2 = A_1 \cdot e^{\frac{1}{12} + \frac{1}{\sigma^4}}$.

Proof of Lemma 5.3 of the paper. Let $\lambda_0 \geq \lambda_1 \ldots$ be eigenvalues of $\mathcal{G}_\sigma$. Note that $\mathcal{G}_\sigma(x, x) = 1$ for all x. Combining that with Lemma 5.1 and the fact that we are considering the kernel with respect to the uniform measure [1] we get

$$\mathbb{E}[T] \leq \frac{1}{\sum_{j=k}^{\infty} \lambda_j}.$$

---

[1] The kernel, we are considering in the paper is not normalized by the uniform measure. Note that, after normalizing the volume term cancels out.

We first prove the second part by showing if $\sigma \leq \frac{1}{2\sqrt{\log k}}$, then $\sum_{j=k}^{\infty} \lambda_j \geq \Omega(1)$. Using the Cauchy-Schwarz inequality we have $k \cdot \sum_{i=0}^{k-1} \lambda_i^2 \geq \left(\sum_{i=0}^{k-1} \lambda_i\right)^2 = \left(1 - \sum_{j=k}^{\infty} \lambda_j\right)^2$. We show $\sum_{i=0}^{k-1} \lambda_i^2 \leq \frac{1}{k^2}$ which implies $\sum_{j=k}^{\infty} \lambda_j \geq (1 - 1/\sqrt{k})$ which completes the proof. To see that, note that

$$\sum_{i=0}^{k-1} \lambda_i^2 \leq \text{tr}(\mathcal{G}_\sigma^2) = \mathbb{E}_{x,y\sim\mu} e^{-\|x-y\|^2/2\sigma^2},$$

where μ is the uniform measure on the sphere. Fix $x \in \mathbb{S}^{d-1}$. It follows from basic concentration inequalities for Gaussian measures that $\mathbb{E}_{y\sim\mu} e^{-\|x-y\|^2/2\sigma^2} \leq e^{-1/2\sigma^2}$ which implies the bound on the trace and finishes the proof of this case.

So from now on, we only need to prove for any σ

$$\sum_{j=k}^{\infty} \lambda_j \gtrsim \frac{e^{\frac{-2}{\sigma^2}}}{t! \cdot \sigma^{2t}}. \tag{5.2}$$

Let $\mu_0 > \mu_1 > \ldots$ be distinct eigenvalues of the kernel given by Lemma 5.1 where for any j, the multiplicity of μ_j is $n_j = N(d, j)$. It suffices to show $\frac{n_t \mu_t}{2} \geq \frac{e^{\frac{-2}{\sigma^2}}}{t! \cdot \sigma^{2t}}$ where we are using the fact that for any j, $n_j \geq \frac{d^j}{j!}$, and so $n_t \geq 2k$. Now using $n_t \geq \frac{d^t}{t!}$, and the bound on μ_t by Lemma 5.1, we get

$$
\begin{aligned}
n_t \mu_t &\gtrsim \frac{d^t}{t!} \cdot \frac{e^{\frac{-2}{\sigma^2}} (2e)^{t+\frac{d}{2}} \Gamma(\frac{d}{2})}{\sigma^{2t} \cdot (2t+d)^{t+\frac{d+1}{2}}} \\
&\gtrsim \frac{d^t}{t!} \cdot \frac{e^{\frac{-2}{\sigma^2}} (2e)^t \cdot d^{\frac{d+1}{2}}}{\sigma^{2t} (2t+d)^{t+\frac{d+1}{2}}} && \text{Sterling's approximation} \\
&\geq \frac{e^{\frac{-2}{\sigma^2}} (2e)^t}{\sigma^{2t} \cdot t! \cdot (1+\frac{2t}{d})^{t+\frac{d+1}{2}}} \gtrsim \frac{e^{\frac{-2}{\sigma^2}} 2^t}{\sigma^{2t} \cdot t! \cdot e^{\frac{2t^2}{d}}} && \text{by } (1+2t/d) \leq e^{2t/d}.
\end{aligned}
$$

Noting that $k \leq \exp(d/4)$ implies $t \leq \frac{d}{4}$ and $\exp(2t/d) \leq 2$, completes the proof of (5.2). $\qquad \square$

References

[AGR16] Nima Anari, Shayan Oveis Gharan, and Alireza Rezaei. Monte carlo markov chain algorithms for sampling strongly rayleigh distributions and determinantal point processes. In *Conference on Learning Theory*, pages 103–115, 2016. 1, 2

[Brä07] Petter Brändén. Polynomials with the half-plane property and matroid theory. *Advances in Mathematics*, 216(1):302–320, 2007. 10

[DF97] Persi Diaconis and David Freedman. On markov chains with continuous state space. Technical report, Technical Report, 1997. 6

[DV07] Amit Deshpande and Kasturi Varadarajan. Sampling-based dimension reduction for subspace approximation. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 641–650. ACM, 2007. 10, 12

[Fil91]   James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process. *The annals of applied probability*, pages 62–87, 1991. 7

[KM12]   Ioannis Kontoyiannis and Sean P Meyn. Geometric ergodicity and the spectral gap of non-reversible markov chains. *Probability Theory and Related Fields*, pages 1–13, 2012. 7

[LS88]   Gregory F Lawler and Alan D Sokal. Bounds on the l^2 spectrum for markov chains and markov processes: a generalization of cheeger's inequality. *Transactions of the American mathematical society*, 309(2):557–580, 1988. 8

[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. 6

[Mih92]   Milena Mihail. On the expansion of combinatorial polytopes. In *International Symposium on Mathematical Foundations of Computer Science*, pages 37–49. Springer, 1992. 2

[MNY06] Ha Quang Minh, Partha Niyogi, and Yuan Yao. Mercer's theorem, feature maps, and smoothing. In *International Conference on Computational Learning Theory*, pages 154–168. Springer, 2006. 12