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 $M$ be the Gibbs sampler chain for an arbitrary discrete $k$-DPP, then we have

$$\phi(M) \geq \frac{1}{Ck^2},$$

for a constant $C > 0$.

Fix $M = (\Omega, P, \pi)$ 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 $\Omega_i$, $\Omega^i$ be the set of all states in $\Omega$ that contain, do not contain $i$, respectively. Also define $\pi_i := \{\pi | i \text{ is chosen}\}$, i.e. $\pi_i(x) = \frac{\pi(x)}{\pi(\Omega_i)}$, $\forall x \in \Omega_i$. Similarly, $\pi^i := \{\pi | i \text{ is not chosen}\}$, i.e. $\pi^i(x) = \frac{\pi(x)}{\pi(\Omega^i)}$, $\forall x \in \Omega^i$.

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

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

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

$$P^i(x, y) = \frac{1}{k} \cdot \frac{\pi(y)}{\pi(I) - \pi(i + I)} \quad (1.2)$$

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

$$N^i(x) = \{y \in \Omega^i | P(x, y) > 0\}.$$

We use the following lemma to relate $Q_i$ to $Q$. 

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

\[
Q(x, N_A) + Q(N_A, N_{\pi^c}) \geq \pi(\Omega_\pi) \cdot Q_{\pi}(N_A, N_{\pi^c}).
\]

Proof. Note that \( x \cup N_A \cup N_{\pi^c} \) is the set of all states containing elements in \( x \). So by definition of \( Q \) and \( Q_{\pi} \), we have

\[
Q(x, N_A) + Q(N_A, N_{\pi^c}) = \frac{1}{k} \cdot \frac{\pi(x) \pi(N_A)}{\pi(x) + \pi(N_A) + \pi(N_{\pi^c})} + \frac{1}{k} \cdot \frac{\pi(N_{\pi^c}) \pi(N_A)}{\pi(x) + \pi(N_A) + \pi(N_{\pi^c})} \geq \frac{\pi(N_A)}{k} \cdot \frac{\pi(N_{\pi^c})}{\pi(x) + \pi(N_A) + \pi(N_{\pi^c})} = \pi(\Omega_\pi) \cdot Q_{\pi}(N_A, N_{\pi^c})
\]

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

\[\Box\]

Figure 1: A schematic view of the restriction chains.

yellow, red, blue, and green edges correspond to \( Q(S_n, \Omega_n \setminus S_n), Q(S_{\pi}, \Omega_{\pi} \setminus S_{\pi}), Q(S_n, \Omega_{\pi} \setminus S_{\pi}), \) and \( Q(S_{\pi}, \Omega_{\pi} \setminus S_{\pi}) \), respectively.

We use an inductive argument to prove \( Q(S, \Omega) \geq \frac{\pi(S)}{C_{360}} \) for a subset \( S \in \Omega \) with \( \pi(S) \leq \frac{1}{2} \).

Letting \( S_n = S \cap \Omega_n \) and \( S_{\pi} = S \cap \Omega_{\pi} \), we have

\[
Q(S, \Omega) = Q(S_n, \Omega_n \setminus S_n) + Q(S_{\pi} \setminus S_{\pi}, \Omega_{\pi} \setminus S_{\pi}) + Q(S_n, \Omega_{\pi} \setminus S_{\pi}) + Q(S_{\pi}, \Omega_n \setminus S_n).
\]

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 \( Q(S_{\pi}, \Omega_n \setminus S_{\pi}) \), we combine the induction hypothesis on \( \mathcal{M}_{\pi} \) with Lemma 1.2. It remains to bound the other two terms which correspond to the contribution of the edge across \( (\Omega_n, \Omega_{\pi}) \). To do that, we crucially use negative association of \( \pi \). In particular, we use the following lemma (appeared before in [Mih92] in the unweighted case). For any set \( A \in \Omega_n \), let \( \pi_{\pi}(A) = \{ y \in \Omega_{\pi} : \exists x \in A, P(x, y) > 0 \} \) denote the set of neighbors of \( A \) in \( \Omega_{\pi} \).

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

\[
\pi_{\pi}(N_{\pi}(A)) \geq \pi_n(A).
\]

The lemma lower bounds the vertex expansion of \( S_n \) in \( \Omega_n \) and similarly vertex expansion of \( S_{\pi} \) in \( \Omega_{\pi} \). 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 \( \pi_n(S) \leq \frac{1}{2} \). 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 (1.7) for the edges across the cut \( (S_n, \Omega_n \setminus S_n) \), we get \( Q(S_n, \Omega_n \setminus S_n) = \frac{(k-1)\pi(O_n)}{Ck^2} \cdot Q_n(S_n, \Omega_n \setminus S_n) \). 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 \( M_\pi \) along with Lemma 1.2. The induction hypothesis implies

\[
Q(\pi(S_n, \Omega_n \setminus S_n)) \geq \frac{\pi(\pi(S_n))}{ck^2} = \frac{\pi(S_n)}{\pi(\pi(S))} \cdot \frac{\pi(S_n)}{ck^2}
\]

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

\[
Q(S_n, \Omega_n \setminus S_n) + Q(\pi(S_n, \Omega_n \setminus S_n)) + Q(S_n, \Omega_n \setminus S_n) \geq \pi(\Omega_n) \cdot Q_n(S_n, \Omega_n \setminus S_n). \tag{1.8}
\]

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

So we focus on the case \( \max\{\pi_n(S_n), \pi_n(S_n)\} > \frac{1}{2} \). Since \( \pi(S) \leq \frac{1}{2} \), we have \( \min\{\pi_n(S_n), \pi_n(S_n)\} \leq \frac{1}{2} \) 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 \( \pi_n(S_n) \leq \frac{1}{2} \). 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 \( x \in \Omega_n \), let \( N_{\pi,S}(x) := N_{\pi}(x) \setminus S_n \), and \( N_{\pi,\overline{S}}(x) := N_{\pi}(x) \setminus S_n \) be a partitioning of \( N_{\pi}(x) \), so for every subset \( T \in N_{\pi}(x) \) we have

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

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

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

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

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

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

Claim 1.4.

\[ Q(S_{\pi}, \overline{S}) + Q(S_n, \Omega_{\pi} \setminus S_{\pi}) - \frac{1}{2} Q(S_{\text{leave}}, \Omega_{\pi} \setminus S_{\pi}) \geq \pi(\Omega_{\pi}) \cdot Q(\pi(S_{\pi}, \Omega_{\pi} \setminus S_{\pi})) \]  
(1.12)

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

\[ Q(N_{\pi,S}(x), \{x\} \cup N_{\pi,S}(x)) \geq \pi(\Omega_{\pi}) \cdot Q(\Omega_{\pi}, S_{\pi}, N_{\pi,S}(x)) \]  
(1.13)

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

\[ Q(x \cup N_{\pi,S}(x), N_{\pi,S}(x)) \geq \pi(\Omega_{\pi}) \cdot Q(\Omega_{\pi}, S_{\pi}, N_{\pi,S}(x)) \]  
(1.14)

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

\[ Q(N_{\pi,S}(x), N_{\pi,S}(x)) + \frac{1}{2} Q(x, N_{\pi,S}(x)) = \frac{\pi(N_{\pi,S}(x))}{2k \cdot (\pi(x) + \pi(N_{\pi,S}(x)) + \pi(N_{\pi,S}(x)))} \cdot \left( \pi(N_{\pi,S}(x)) + \frac{\pi(x)}{2} \right) \]

\[ \geq \frac{1}{2k} \cdot \frac{\pi(N_{\pi,S}(x)) \pi(N_{\pi,S}(x))}{\pi(N_{\pi,S}(x)) + \pi(N_{\pi,S}(x))} \]

\[ = \pi(\Omega_{\pi}) \cdot Q(\pi(N_{\pi,S}(x), N_{\pi,S}(x)) \]  
(1.15)

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

In particular, we claim the above claim to get

\[ Q(S, \overline{S}) = Q(S_n, \Omega_{\pi} \setminus S_{\pi}) + Q(S_{\pi}, \Omega_{\pi} \setminus S_{\pi}) + Q(S_{\pi}, \Omega_{\pi} \setminus S_{\pi}) + Q(S_n, \Omega_{\pi} \setminus S_{\pi}) \]

\[ \geq Q(S_n, \Omega_{\pi} \setminus S_{\pi}) + \frac{1}{2} Q(S_{\text{leave}}, \Omega_{\pi} \setminus S_{\pi}) + \pi(\Omega_{\pi}) Q(S_{\pi}, \overline{S}) \]  
By Claim 1.4

\[ \geq \frac{\pi(\Omega_{\pi}) - \pi(S_{\text{leave}})}{Ck(k-1)} + \frac{1}{2} Q(S_{\text{leave}}, \Omega_{\pi} \setminus S_{\pi}) + \frac{\pi(S_{\text{stay}})}{Ck^2} \]  
induction Hyp. on $M_n$ and $M_{\pi}$

\[ \geq \frac{\pi(\Omega_{\pi}) - \pi(S_{\text{leave}}) - \pi(S_{\text{stay}})}{Ck(k-1)} + \frac{\pi(S_{\text{leave}})}{8k} + \frac{\pi(S_{\text{stay}})}{Ck^2} \]  
By (1.11) and $S_n = S_{\text{leave}} \cup S_{\text{stay}}$

(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 $\frac{\pi(S_{\text{leave}})}{8k} \geq \pi(S_{\text{leave}}) \cdot \left( \frac{1}{Ck^2} + \frac{1}{k(k-1)} \right)$ for sufficiently large $k$, it suffices to show

\[ \frac{\pi(\Omega_{\pi}) - \pi(S_{\text{leave}})}{Ck(k-1)} \geq \pi(S_{\text{stay}}) \cdot \frac{\pi(S_{\text{stay}})}{Ck^2}, \]

which can be directly verified for $\pi_n(S_{\text{stay}}) \leq \frac{1}{2} + \frac{1}{4k}$. 

4
\textbf{Case 2: } \( \pi_n(S_{\text{stay}}) > \frac{1}{2} + \frac{1}{4k} \). We prove
\[ Q(S, \bar{S}) \geq \frac{1 - \pi(S)}{Ck^2}. \]

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

\textbf{Claim 1.5.}
\[ Q(S_{\text{stay}}, \Omega_{\bar{S}} \setminus S) + Q(S, \Omega_{\bar{S}} \setminus S) \geq \frac{\pi(\Omega_{\bar{S}})}{2k} \cdot (\pi_n(S_{\text{stay}}) - \pi_n(S)) \]

\textbf{Proof.} Note that for any \( x \in S_{\text{stay}} \), since \( \pi(N_{\bar{S},\bar{S}}(x)) \leq \pi(x) + \pi(N_{\bar{S},S}(x)) \), we have
\[ Q(x, N_{\bar{S},\bar{S}}(x)) + Q(N_{\bar{S},S}(x), N_{\bar{S},\bar{S}}(x)) = \frac{1}{2k} \cdot \pi(N_{\bar{S},\bar{S}}(x)) \cdot \left[ \pi(x) + \pi(N_{\bar{S},S}(x)) \right] \geq \frac{1}{2k} \pi(N_{\bar{S},S}(x)). \]

To complete the proof, it is enough to sum up the above over \( S_{\text{stay}} \) to get the following
\[ Q(S_{\text{stay}}, \Omega_{\bar{S}} \setminus S) + Q(S, \Omega_{\bar{S}} \setminus S) \geq \sum_{x \in S_{\text{stay}}} \frac{\pi(N_{\bar{S},\bar{S}}(x))}{4k} \geq \pi \left( \bigcup_{x \in S_{\text{stay}}} N_{\bar{S},\bar{S}}(x) \right). \]

\[ \geq \pi(N_{\bar{S}}(S_{\text{stay}})) - \pi(S) \]
\[ \geq \pi(\Omega_{\bar{S}}) \cdot (\pi_n(S_{\text{stay}}) - \pi_n(S)) \quad \text{By Lemma 1.3} \]

\textbf{Claim 1.5 and (1.16)} implies \( Q(S, \bar{S}) \geq \max\{L_1, L_2\} \) defined as above
\[ L_1 := \frac{\pi(S_{\text{leave}})}{8k} + \frac{\pi(\Omega_{\bar{S}}) - \pi(S_{\text{leave}}) - \pi(S_{\text{stay}})}{Ck(k-1)} + \frac{\pi(S)}{Ck^2} \quad \text{By (1.16)} \]
\[ L_2 := \frac{\pi(\Omega_{\bar{S}})}{4k} \cdot (\pi_n(S_{\text{stay}}) - \pi_n(S)) \quad \text{By Claim 1.5.} \]

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
\[ \frac{\pi(S_{\text{leave}})}{8k} + \frac{\pi(S)}{Ck^2} + \frac{\pi(\Omega_{\bar{S}})}{4k(k-1)} \cdot (\pi_n(S_{\text{stay}}) - \pi_n(S)) \geq \frac{\pi(\Omega_{\bar{S}}) - \pi(S)}{Ck(k-1)}. \]

Ignoring the \( \frac{\pi(S_{\text{leave}})}{8k} \) term and rearranging the other terms, it is enough to show
\[ \frac{\pi(\Omega_{\bar{S}})}{4k(k-1)} \cdot (\pi_n(S_{\text{stay}}) - \pi_n(S)) \geq \frac{\pi(\Omega_{\bar{S}})}{Ck(k-1)} \cdot \left( 1 - \frac{2k-1}{k} \cdot \pi_n(S) \right). \]

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 \( \pi_n(S) \leq \frac{1}{2} \).
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 $\mu_0$, after one step the distribution $\mu_1$ 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 $\sigma$-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 $\pi$ on $\Omega$ 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 $\mathcal{M}$ $\phi$-irreducible for a probability measure $\phi$ 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 $\phi$-irreducible if for any $B \subset \Omega$ 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 $\pi$ 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 $\pi$, the $\pi$ itself is the stationary measure. Moreover, if the kernel of the $k$-DPP is continuous, it is straight-forward to see that it is $\pi$-strongly irreducible. The following lemma also shows $\pi$ 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 $\pi$ is a stationary measure of $\mathcal{M}$, and $\mathcal{M}$ is strongly $\pi$-irreducible. Then for any other distribution $\mu$ which is absolutely continuous with respect to $\pi$, $\lim_{n \to \infty} |P^n(\mu, .) - \pi|_{TV} = 0$.

From now on, consider $\mathcal{M} = (\Omega, P, \pi)$ is chain with state space $\Omega$, probability transition function $P$, and a unique stationary measure $\pi$. 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 $\pi$ as

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

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 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,

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

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 $\pi$-irreducible Markov chain $M = (\Omega, P, \pi)$, if $\lambda > 0$, then the distribution of the chain started from $\mu$ (which is absolute continuous with respect to $\pi$) is

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

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 $L^2(\pi)$,

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

**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 $f_t := \frac{f_\mu}{f_\pi}$, by reversibility of the chain we have

$$(Pf_t)(x) = \int_{\Omega} P(x, y) \frac{f_\mu(y)}{f_\pi(y)} dy = \int_{\Omega} \frac{P(y, x) f_\mu(y)}{f_\pi(x)} dy = \frac{f_{\mu_{t+1}}}{f_\pi}(x) = f_{t+1}(x).$$
which implies
\[ \text{var}_\pi(Pf_t) = \text{var}_\pi(f_{t+1}) \]  
(2.1)

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

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

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

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
\[ 4\|\mu_t - \pi\|_T^2 \leq \text{var}(f_t). \]

This can be seen using an application of Cauchy-Schwarz’s inequality. We have
\[
4\|\mu_t - \pi\|_T^2 = \left( \int_\Omega |f_{\mu_t}(x) - f_\pi(x)| dx \right)^2 \\
= \left( \int_\Omega f_\pi(x) \left| \frac{f_{\mu_t}(x)}{f_\pi(x)} - 1 \right| dx \right)^2 \\
\leq \int_\Omega f_\pi(x) \left| \frac{f_{\mu_t}(x)}{f_\pi(x)} - 1 \right|^2 dx = \text{var}(\frac{f_{\mu_t}}{f_\pi})
\]

The last identity uses that \( \mathbb{E}_\pi f_{\mu_t} = 1 \). This completes the proof.

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 \( M = (\Omega, P, \pi) \), the ergodic flow \( Q : B \to [0,1] \) 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
\[ \phi(M) = \min_{0 < \pi(B) \leq \frac{1}{2}} \phi(B). \]

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 \( M \) defined on a general state space with spectral gap \( \lambda \) we have
\[ \frac{\phi(M)^2}{8} \leq \lambda \leq 2\phi(M). \]
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

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

**Proof.** Recall that by Theorem 1.1 the conductance of a Gibbs sampler for any discrete $k$-DPP is at least $\Omega\left(\frac{1}{k^2}\right)$. 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 straightforward to extend the argument to higher dimensions. Let us denote the state space by $\Omega$. Fix a measurable subset $S \subset \Omega$ with $\pi(S) \leq \frac{1}{2}$. Our goal is to prove $\phi(S) = \frac{Q(S, S)}{\pi(S)} \geq \Omega\left(\frac{1}{k^2}\right)$. Without loss of generality, we can only consider restriction of $\Omega$ and $S$ to a bounded set. To see that, note that if we set $\Omega_n = \left(1^{-n}, n \right]$, then clearly, $\lim_{n \to \infty} \frac{Q(S \cap \Omega_n, S \cap \Omega_n)}{\pi(S \cap \Omega_n)} = \phi(S)$, and so for large values of $n$, $\frac{Q(S \cap \Omega_n, S \cap \Omega_n)}{\pi(S \cap \Omega_n)} = \Theta(\phi(S))$. So suppose that $\Omega = \left(0, 1\right]$. 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 $\Omega_n = \left(0, 1\right]$. $\mathcal{M}_n$ is defined by a kernel $L_n$ characterized below. For $i \in \left[n\right]$ let $I_i = \left[\frac{i-1}{n}, \frac{i}{n}\right]$. For any $i, j \in \left[n\right]$, we define $L_n(i, j) = \int \int L(u, v) dudv$, 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 $|y_i - x_i| \leq \frac{1}{n^2}$, we have $|det_L(x_1, \ldots, x_k) - det_L(y_1, \ldots, y_k)| \leq \epsilon$. Now, note that $f_\pi(y_1, \ldots, y_k) = \frac{det_L(y_1, \ldots, y_k)}{\int_{\Omega} det_L(x_1, \ldots, x_k) dx_1 \ldots dx_k}$.

So, using the simple fact that for any two sequences of numbers $\{a_n\}$ and $\{b_n\}$,

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

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

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

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

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

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

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

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

for \( \mu \) 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 \( x_1, \ldots, x_k, x_{k+1} \) and \( y_1, \ldots, y_k, y_{k+1} \) with \( |x_i - y_i| \leq \frac{1}{n} \), and \( \int_0^1 \det_L(x_1, \ldots, x_{k-1}, \tau)d\tau \geq \delta \), we have

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

Therefore, similar to the case for \( \pi_n \), 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 \( t_1, \ldots, t_{k-1}, s, t \in \binom{[n]}{k+1} \) with \( \sum_{i=1}^{n} \pi_n(t_1, \ldots, i) \geq \frac{\delta}{n^{k-1}} \),

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

Now, combining the above equation with (3.5), and noting \( \epsilon \) and \( \delta \) 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. \( \square \)

**Remark 3.2.** It is straightforward to use a similar discretization argument to prove Theorem 3.1, when the domain of the kernel is restricted to a closed subset \( C \subseteq \mathbb{R}^d \) 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 our algorithm in the paper.

**Lemma 4.1.** Let \( \nu \) be the probability distribution of the output of our to find the starting state. Also let \( f_\nu \) and \( f_\pi \) denote the p.d.f. for \( \nu \) and \( \pi \) which is the k-DPP defined by \( L \). Then, for any \( \{x_1, \ldots, x_k\} \subseteq C \),

\[
f_\nu(\{x_1, \ldots, x_k\}) \leq (k!)^2 f_\pi(\{x_1, \ldots, x_k\}).
\]

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 \( x = \{x_1, \ldots, x_k\} \), 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 \( H_{\sigma} = \langle f_{\sigma(1)}, \ldots, f_{\sigma(i)} \rangle \). In the above the range of all integrals is \( \mathbb{R}^d \). We have

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

Note that the above integrals are well-defined since our kernel is continuous. For any \( 1 \leq i \leq k-1 \), let \( H_{\sigma} = \arg \min_{H = \langle f_{y_1}, \ldots, f_{y_i} \rangle} \int d(f_y, H)^2 \, dy \), 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 \( \sigma \), \( \det(x_1, \ldots, x_k) = \|f_{\sigma(1)}\|^2 \cdot d(f_{\sigma(2)}, H_{\sigma}^1)^2 \cdots d(f_{\sigma(k)}, H_{\sigma}^{k-1})^2 \), we obtain

\[
F_{\nu}(x) \leq k! \cdot \frac{\det(x_1, \ldots, x_k)}{\int \|f_y\|^2 \, dy \cdot \int d(f_y, H_1)^2 \, dx \cdots \int d(f_y, H_{k-1}^{k-1})^2 \, dy} \cdot \frac{\det(y_1, \ldots, y_k) \, dy_k \cdots dy_1}{k!} \cdot \frac{\int \|f_y\|^2 \, dy \cdot \int d(f_y, H_1)^2 \, dx \cdots \int d(f_y, H_{k-1}^{k-1})^2 \, dy}{(k!)^2}.
\]

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

\[
\int \int \cdots \det(y_1, \ldots, y_k) \, dy_k \cdots dy_1 \leq (k!)^2.
\]

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.

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

Proof of Claim 4.2. For any \( y = \{y_1, \ldots, y_k\} \subset \mathbb{R}^d \), let \( G_y \) be a \((k-1)-\text{dimensional linear}

subspace of \( \langle f_{y_1}, \ldots, f_{y_k} \rangle \) which contains the projection of \( H_{y}^{(k-1)} \) onto \( \langle f_{y_1}, \ldots, f_{y_k} \rangle \). Now, for any \( y \), using Lemma 4.3, we get

\[
\det(y) \leq \left( \sum_{i=1}^{k} d(f_{y_i}, G_y) \sqrt{\det(y - y_i)} \right)^2 \leq k \left( \sum_{i=1}^{k} d(f_{y_i}, G_y)^2 \det(y - y_i) \right) \text{Cauchy-Schwarz Inequality.}
\]
By integrating the above, we get
\[
\int \cdots \int \det(y)dy \leq k \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} d(f_{y_i}, G_y)^2 \det(y - y_i)dy
\]
\[
\leq k^2 \int_{y \in \mathbb{R}^d} \int_{z_1 \in \mathbb{R}^d} \cdots \int_{z_{k-1} \in \mathbb{R}^d} d(f_y, G_{z+y})^2 \det(z)dzdy \quad \text{(setting } z = \{z_1, \ldots, z_{k-1}\})
\]
\[
\leq \int_{y \in \mathbb{R}^d} \int_{z_1 \in \mathbb{R}^d} \cdots \int_{z_{k-1} \in \mathbb{R}^d} d(f_y, H_k)^2 \det(z)dzdy
\]
\[
= \left( \int d(f_y, H_k)^2dy \right) \left( \int_{z_1 \in \mathbb{R}^d} \cdots \int_{z_{k-1} \in \mathbb{R}^d} \det(z)dz \right),
\]
where in the third inequality, the fact \(d(f_y, G_{z+y}) \leq d(f_y, H_k)\) holds because \(f_y \in \langle f_{z_1}, \ldots, f_{z_{k-1}}, f_y \rangle\), and \(G_{z+y}\) contains the projection of \(H_k\) 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 \((k-1)\)-dimensional subspace of \((S)\. 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 \(G_\sigma\) be the Gaussian kernel with variance \(\sigma^2\) restricted to the unit sphere with the uniform measure, i.e. for any \(x, y \in S^{d-1}\): we have \(G_\sigma(x, y) = \frac{\exp(-\|x-y\|^2/\sigma^2)}{\text{vol}(S^{d-1})}\). For any integer \(k \geq 0\), \(G_\sigma\) has an eigenvalue \(\mu_k\) with multiplicity \(N(d, k) = \frac{(2k+d-2)(k+d-3)!}{k!(d-2)!}\) where
\[
\left( \frac{2e}{\sigma^2} \right)^k \cdot \frac{A_1}{(2k + d - 2)^{k+\frac{d-1}{2}}} \leq \mu_k \leq \left( \frac{2e}{\sigma^2} \right)^k \cdot \frac{A_2}{(2k + d - 2)^{k+\frac{d-1}{2}}},
\]
for \(A_1 = e^{-\frac{2}{\sigma^2}} \cdot \frac{1}{\sqrt{\pi}} (2e)^{\frac{d}{2}} \Gamma \left( \frac{d}{2} \right)\) and \(A_2 = A_1 \cdot e^{\frac{1}{12\sigma^2} + \frac{1}{\sigma^2}}\).

**Proof of Lemma 5.3 of the paper.** Let \(\lambda_0 \geq \lambda_1 \ldots\) be eigenvalues of \(G_\sigma\). Note that \(G_\sigma(x, x) = 1\) 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
\[ \sum_{i=0}^{k-1} \lambda_i^2 \leq \text{tr}(G_{\sigma}^2) = \mathbb{E}_{x,y \sim \mu} e^{-\|x-y\|^2/2\sigma^2}, \]
where $\mu$ is the uniform measure on the sphere. Fix $x \in S^{d-1}$. 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 $\sigma$
\begin{equation}
\sum_{j=k}^{\infty} \lambda_j \gtrsim \frac{e^{-\sigma^2/2t}}{t! \cdot \sigma^2 t}.
\end{equation}
Let $\mu_0 > \mu_1 > \ldots$ be distinct eigenvalues of the kernel given by Lemma 5.1 where for any $j$, the multiplicity of $\mu_j$ is $n_j = N(d,j)$. It suffices to show \( \frac{n_t \mu_t}{2} \geq \frac{e^{-\sigma^2/2t}}{t! \cdot \sigma^2 t} \) 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 $\mu_t$ by Lemma 5.1, we get
\begin{align*}
n_t \mu_t & \geq \frac{d^t}{t!} \cdot \frac{e^{-\sigma^2/2t}}{\sigma^2 t \cdot (2t + d)^{t+\frac{d+1}{2}}} \\
& \geq \frac{d^t}{t!} \cdot \frac{e^{-\sigma^2/2t}}{\sigma^2 t (2t + d)^{t+\frac{d+1}{2}}} \\
& \geq \frac{e^{-\sigma^2/2t}}{\sigma^2 t! \cdot (1 + \frac{2t}{d})^{t+\frac{d+1}{2}}} \geq \frac{e^{-\sigma^2/2t}}{\sigma^2 t! \cdot t! \cdot e^{2t/d}} \text{ by } (1 + 2t/d) \leq e^{2t/d}.
\end{align*}
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).

References


