Approximating Permanent of a Matrix

Rishav Gupta

May 2024

Contents

1 Definitions

For our case we assume the graph $\mathcal{G}$ to be bipartite in all definitions and theorems.

- **Perfect Matching-** For a graph $\mathcal{G}$, a perfect matching M is a subset of edges of the graph $\mathcal{G}$ in which each vertex has exactly one edge incident onto it.
- **Near Perfect Matching-** For a graph $\mathcal{G}$ a near perfect matching M is a set of edges in which all vertices except exactly two vertices say u, v have exactly one edge incident onto it in M . For our case of bipartite graphs it is easy to see that the vertices u and v are in different bipartitions.
  Also for a near perfect matching M in which u, v are not matched we say u and v the markov are the holes of the near perfect matching M.
- **Permanent-** The permanent of a $n \times n$ matrix A is defined as follows

$$per(A) = \sum_{\sigma \in S_n} \prod_i a_{i,\sigma(i)}$$

- **Mixing time of a Markov Chain -** The mixing time $\tau_x(\delta)$ of a markov chain $\mathcal{M}$ is defined as follows

$$\tau_x(\delta) = \min\{t | \forall s \geq t, d(P^s(x), \pi_{\mathcal{M}}) < \delta\}$$

  . Where $P^s(x)$ is the distribution on the states obtained after walking on the markov chain for s steps starting from x and $\pi_{\mathcal{M}}$ is the stationary distribution of $\mathcal{M}$.
  Let the spectral gap of the markov chain that is difference between the norm of the largest(norm) two eigenvalues (the largest being 1) of the transition probability matrix of the markov chain be denoted by γ.
  Then we have the following bounds on the mixing time

$$\tau_x(\delta) \leq \frac{1}{\gamma}(\log \frac{1}{\pi_{\mathcal{M}}(x)} + \log \frac{1}{\delta})$$

2 Classical Approximation Algorithm

2.1 Getting uniform sampler from a certain Markov Chain

So our objective is to generate an $FPRAS$ for approximating the permanent of matrices with non negative entries. For now we will focus on the case of $0-1$ matrices we will show how to generalize for any non-negative matrix at last. Now notice computing the permanent of a $0-1$ matrix is same as counting the number of perfect matchings in the bipartite graph obtained by considering the matrix A as the bipartite adjeceny matrix whose permanent we want to compute. So we will focus on generating a uniform sampler for the perfect matchings. While doing so we will get a sequence of parameters and markov chains which are parametrized by these parameters , which we will use to approximate the number of matchings and hence the permanent.

So given a bipartite graph $\mathcal{G}$ on $2n$ vertices we need to get an almost uniform sampler on the perfect matchings of the graph $\mathcal{G}$.
So for making such a sampler we will make a markov chain on the space $\Omega = \mathcal{M} \cup \mathcal{N}$ that is union of perfect matchings and near perfect matchings. (We denote $\mathcal{M}$ to be set of all perfect matchings possible on $K_{n,n}$ and $\mathcal{N}$ to be the set of all near perfect matchings possible on $K_{n,n}$) where the probability to land on a valid matching M that is $M \in \mathcal{M}_{\mathcal{G}}$,($\mathcal{M}_{\mathcal{G}}$ is the set of all valid perfect matchings on $\mathcal{G}$) in the stationary probability distribution of the markov chain is quite high like $\Omega(\frac{1}{n^2})$ and is uniform on the set $\mathcal{M}_{\mathcal{G}}$.In other words, we will get a markov chain MC with state space being Ω as defined above such that

$$\pi_{MC}(\mathcal{M}_{\mathcal{G}}) \geq \frac{1}{(4n^2+2)} = \Omega(1/n^2)$$

After getting such a markov chain we use the following algorithm to get an almost uniform sampler for the perfect matchings of $\mathcal{G}$.

Set $\hat{\delta} \leftarrow \delta/\left(12n^2+6\right)$.
Repeat $T = \left\lceil \left(6n^2+2\right)\ln(3/\delta) \right\rceil$ times:
    Simulate the Markov chain for $\tau(\hat{\delta})$ steps;
    output the final state if it belongs to $\mathcal{M}_{\mathcal{G}}$ and halt.
Output an arbitrary perfect matching if all trials fail.

Now consider a subset $S \subseteq \mathcal{M}_{\mathcal{G}}$ now note the following, for the final output M

$$\begin{aligned}
\Pr(M \in \mathcal{S}) &\geq \frac{\hat{\pi}(\mathcal{S})}{\hat{\pi}(\mathcal{M}_G)} - (1-\hat{\pi}(\mathcal{M}_G))^T \\
&\geq \frac{\pi(\mathcal{S}) - \hat{\delta}}{\pi(\mathcal{M}_G) + \hat{\delta}} - \exp(-\hat{\pi}(\mathcal{M}_G)T) \\
&\geq \frac{\pi(\mathcal{S})}{\pi(\mathcal{M}_G)} - \frac{2\hat{\delta}}{\pi(\mathcal{M}_G)} - \exp(-(\pi(\mathcal{M}_G) - \hat{\delta})T) \\
&\geq \frac{\pi(\mathcal{S})}{\pi(\mathcal{M}_G)} - \frac{2\delta}{3} - \frac{\delta}{3}
\end{aligned}$$

Hence we get an almost uniform sampler on the perfect matchings of $\mathcal{G}$ using the promised Markov Chain.

2.2 Designing the desired Markov Chain

Now we wish to design the markov chain on the set $\Omega = \mathcal{M} \cup \mathcal{N}$ whose stationary probability distribution π has the property that when restricted to elements of $\mathcal{M}_G$ it is uniform and $\pi(\mathcal{M}_G)$ (which is $\sum_{M \in \mathcal{M}_G} \pi(M)) \geq \frac{1}{(4n^2+2)}$.

For designing this Markov chain we will define a sequence of parameters each parameter will describe a markov chain which will be used to determine the parameters for the next iteration.
So we have parameters $\lambda(u, v)$ and $w(u, v)$ for each pair (u, v) where $u \in V_1(G)$ and $v \in V_2(G)$.
So initially $\lambda(u, v) = 1, w(u, v) = n, \forall (u, v) \in V_1 \times V_2$, then progressively we would make $\lambda(u, v) = 1/n!$ if $(u, v) \notin E(G)$ and $\lambda(u, v) = 1$ if $(u, v) \in E(G)$. We will have the stationary distribution of markov chains propotional to the product of λ's of edges (at that stage) contained in the matching. So the weight of invalid perfect matchings and invalid near perfect matchings in the stationary distribution of the final markov chain would be very low. We will give details of how we iterate and how the parameters are modified in the following paragraph.

- So again we denote the value of $\lambda(u, v)$ at iteration i to be $\lambda_i(u, v)$. Also we define $\lambda_i(M) = \prod_{e \in M} \lambda_i(e)$ for $M \in \Omega$ and $\lambda_i(\mathcal{S}) = \sum_{M \in \mathcal{S}} \lambda_i(M)$ for $\mathcal{S} \subset \Omega$.

- At each iteration for some vertex v we change the λ of all non edges connected to that vertex by a factor of $e^{-1/2}$. That is for all non edges connected to v we have $\lambda(e) \to e^{-1/2}\lambda(e)$.

- Since we have n vertices and for non-edges we go from $1 \to 1/n!$ and $\log n! \leq n \log n$ so we have total of $r = n^2 \log n$ iterations.

- Now we will define the random walk which describes the Markov Chain at each iteration.
  So we want the stationary probability π_i distribution of the markov chain at iteration i to be propotional to the following quantity Λ_i.(Basically $\pi_i(M) \propto \Lambda_i(M)$) Where Λ_i is defined as follows.

$$\Lambda_i(M) = \begin{cases} \lambda_i(M) w_i(u, v) & \text{if } M \in \mathcal{M}(u, v) \text{ for some } u, v; \\ \lambda_i(M), & \text{if } M \in \mathcal{M}. \end{cases}$$

- The transition from a matching M, for a random walk at iteration i (MC_i) is described as follows for the random walk-

- If $M \in \mathcal{M}$, choose an edge $e = (u, v)$ uniformly at random from M; set $M' = M \backslash e$.
- If $M \in \mathcal{M}(u, v)$, choose z uniformly at random from $V_1 \cup V_2$.
  - if $z \in \{u, v\}$ and $(u, v) \in E$, let $M' = M \cup (u, v)$;
  - if $z \in V_2, (u, z) \in E$ and $(x, z) \in M$, let $M' = M \cup (u, z) \backslash (x, z)$;
  - if $z \in V_1, (z, v) \in E$ and $(z, y) \in M$, let $M' = M \cup (z, v) \backslash (z, y)$;
  - otherwise, let $M' = M$.
- With probability $\min\{1, \Lambda_i(M')/\Lambda_i(M)\}$ go to M'; otherwise, stay at M.
- Now with probability $1/2$ we stay at M and with $1/2$ probability we move using above transition.

- Now for the above markov chain we can see that the stationary probability distribution is propotional to Λ_i. Now we will explain how we update w's.
  We want $w_i(u, v)$ to be close to $w_i^*(u, v)$, where $w_i^*(u, v) = \frac{\lambda_i(\mathcal{M})}{\lambda_i(\mathcal{M}(u,v))}$.
  The intuition behind choosing w^* is that if $w = w^*$ then each pattern (pattern is defined as the sets $\mathcal{M} \cup \{\mathcal{M}(u, v) | (u, v) \in V_1 \times V_2\}$, there are $(n^2 + 1)$ patterns) would be equally likely in the stationary probability distribution.

- We want that $w^*(u, v)/2 \leq w(u, v) \leq 2w^*(u, v)$ because of the following theorem.

  Theorem 2.1. *Assuming the weight function w_i satisfies inequality $w_i^*(u, v)/2 \leq w_i(u, v) \leq 2w_i^*(u, v)$ for all $(u, v) \in V_1 \times V_2$ with $\mathcal{M}(u, v) \neq \emptyset$, then the mixing time of the Markov chain MC_i is bounded above by $\tau_M(\delta) = O\left(n^6 g \left(\log\left(\pi(M)^{-1}\right) + \log \delta^{-1}\right)\right)$.*

- This theorem is the heart of the algorithm it is proved by using a multi commodity flow and bounding the congestion(ϱ). I will use the bound on the congestion and the following theorem to give bounds for the running time of the sampler and hence the approximation algorithm.

  Theorem 2.2. *For an ergodic, reversible Markov chain with self-loop probabilities $P(M, M) \geq 1/2$ for all states M, and any initial state $M_0 \in \Omega$,*

$$\tau_{M_0}(\delta) \leq \varrho \left(\ln \pi \left(M_0\right)^{-1} + \ln \delta^{-1}\right).$$

- We will analyze the complexities at last for now we will just describe how we update $w's$. We will show the following relation holds

$$\frac{\Lambda_i(\mathcal{M}(u, v))}{\Lambda_i(\mathcal{M})} = \frac{\pi_i(\mathcal{M}(u, v))}{\pi_i(\mathcal{M})} = \frac{w_i(u, v)}{w_i^*(u, v)}$$

$$\begin{aligned}
\Lambda_i(\mathcal{M}(u,v)) &= \sum_{M \in \mathcal{M}(u,v)} \lambda_i(M) w_i(u,v) \\
&= w_i(u,v) \sum_{M \in \mathcal{M}(u,v)} \lambda_i(M) \\
&= w_i(u,v) \lambda_i(\mathcal{M}(u,v)) \\
&= w_i(u,v) \frac{\lambda_i(\mathcal{M})}{w_i^*(u,v)} \\
\implies \frac{\Lambda_i(\mathcal{M}(u,v))}{\Lambda_i(\mathcal{M})} &= \frac{\pi_i(\mathcal{M}(u,v))}{\pi_i(\mathcal{M})} = \frac{w_i(u,v)}{w_i^*(u,v)}
\end{aligned}$$

Now we want to use an unbiased estimator for estimating $\pi_i(\mathcal{M})$ and $\pi_i(\mathcal{M}(u,v))$ in order to get a 5/6 factor approximation to $w^*(u,v)$.We will use this approximation to get w_{i+1}.We will use the condition on w and Theorem 2.1 to put bound on mixing time to get get δ close sampler for π_i.

Although we do not know how to sample from π_i exactly, Theorem 2.1 does allow us to sample, in polynomial time, from a distribution $\hat{\pi_i}$ that is within variation distance δ of π_i. We shall see presently that setting $\delta = O\left(n^{-2}\right)$ suffices in the current situation; certainly, the exact value of δ clearly does not affect the leading term in the mixing time promised by Theorem 2.1. So suppose we generate S samples from $\hat{\pi_i}$, and for each pair $(u,v) \in V_1 \times V_2$ we consider the proportion $\alpha(u,v)$ of samples with hole pair u, v, together with the proportion α of samples that are perfect matchings. (For the rest of the paragraph I will drop the i subscript as it is same for all i) Clearly,

$$\mathbb{E}\alpha(u,v) = \hat{\pi}(\mathcal{M}(u,v)) \quad \text{and} \quad \mathbb{E}\alpha = \hat{\pi}(\mathcal{M}).$$

I will also prove one fact that assuming conditions on w

$$\pi(\mathcal{M}(u,v)) \geq \frac{1}{4(n^2+1)}$$

Since $\Omega = \mathcal{M} \cup \mathcal{N}$ we have-

$$\begin{aligned}
&\implies \pi(\mathcal{M}) + \sum_{(u,v)} \pi(\mathcal{M}(u,v)) = 1 \\
&\implies \pi(\mathcal{M}) + \sum_{(u,v)} \frac{\lambda(\mathcal{M}(u,v))w(u,v)}{\Lambda(\Omega)} = 1 \\
&\implies \pi(\mathcal{M}) + \sum_{(u,v)} \frac{\lambda(\mathcal{M})w(u,v)}{\Lambda(\Omega)w^*(u,v)} = 1 \\
&\implies \pi(\mathcal{M}) + \sum_{(u,v)} \frac{2\lambda(\mathcal{M})}{\Lambda(\Omega)} \geq 1 \\
&\implies \pi(\mathcal{M})(1 + 2n^2) \geq 1 \\
&\implies \pi(\mathcal{M}) \geq \frac{1}{(2n^2+1)} \\
&\implies \pi(\mathcal{M}(u,v)) \geq \frac{1}{4(n^2+1)}
\end{aligned}$$

Naturally, it is always possible that some sample average $\alpha(u,v)$ will be far from its expectation, so we have to allow for the possibility of failure. We denote by $\hat{\eta}$ the (small) failure probability we are prepared to tolerate. Provided the sample size S is large enough, $\alpha(u,v)$ (respectively, α) approximates $\hat{\pi}(\mathcal{M}(u,v))$ (respectively, $\hat{\pi}(\mathcal{M})$) within ratio $c^{1/4}$, with probability at least $1-\hat{\eta}$. Furthermore, if δ is small enough, $\hat{\pi}(\mathcal{M}(u,v))$ (respectively, $\hat{\pi}(\mathcal{M})$) approximates $\pi(\mathcal{M}(u,v))$ (respectively, $\pi(\mathcal{M})$) within ratio $c^{1/4}$. Then, we have, with probability at least $1-\left(n^2+1\right)\hat{\eta}$, approximations within ratio c to all of the target weights $w^*(u,v)$.

It remains to determine bounds on the sample size S and sampling tolerance δ that make this all work. Condition on w entails

$$\mathbb{E}\alpha(u,v) = \hat{\pi}(\mathcal{M}(u,v)) \geq \pi(\mathcal{M}(u,v)) - \delta \geq \frac{1}{4\left(n^2+1\right)} - \delta.$$

Assuming $\delta \leq 1/8\left(n^2+1\right)$, it follows from any of the standard Chernoff bounds , that $O\left(n^2 \log(1/\hat{\eta})\right)$ samples from $\hat{\pi}$ suffice to estimate $\mathbb{E}\alpha(u,v) = \hat{\pi}(\mathcal{M}(u,v))$ within ratio $c^{1/4}$ with probability at least $1-\hat{\eta}$. Again using the fact that $\pi(\mathcal{M}(u,v)) \geq 1/4\left(n^2+1\right)$, we see that $\hat{\pi}(\mathcal{M}(u,v))$ will approximate $\pi(\mathcal{M}(u,v))$ within ratio $c^{1/4}$ provided $\delta \leq c_1/n^2$ where $c_1 > 0$ is a sufficiently small constant. (Note that we also satisfy the earlier constraint on δ by this setting.) Hence we take $c = 5/6$ and $S = O\left(n^2 \log(1/\hat{\eta})\right)$ to get a refined estimate to w^*. For i th iteration we label this approximation to w_i^* as w_i'.

- We return w_{i+1} as w_i'. Because w_i' is within 5/6 factor of w_i^* and w_{i+1}^* is within a factor of $e^{-1/2}$ of w_i^*. Since $5/6 \cdot e^{-1/2} = 0.505 \geq 0.5$ w_{i+1}

satisfies the condition that is within a factor of 2 of w^*_{i+1}. Hence we can repeat this step for $r = O(n^2 \log n)$ iterations to get our desired markov chain and MC_r will be our desired markov chain.

2.3 Approximating the number of perfect matchings

Now to get the number of perfect matchings we will use the following annealing scheme-

$$\Lambda_0(\Omega) \cdot \frac{\Lambda_1(\Omega)}{\Lambda_0(\Omega)} \cdot \frac{\Lambda_1(\Omega)}{\Lambda_2(\Omega)} \cdots \frac{\Lambda_r(\Omega)}{\Lambda_{r-1}(\Omega)} \cdot \frac{\Lambda_r(\mathcal{M}_G)}{\Lambda_r(\Omega)}$$

We know that $\Lambda_0(\Omega) = (n^2+1)n!$. We will now use MC_i to estimate $\frac{\Lambda_{i+1}(\Omega)}{\Lambda_i(\Omega)}$. and MC_r to estimate $\frac{\Lambda_r(\mathcal{M}_G)}{\Lambda_r(\Omega)}$.
Now we will explain how we estimate $\frac{\Lambda_{i+1}(\Omega)}{\Lambda_i(\Omega)}$.
First we will prove one bound that

$$\frac{1}{4e} \leq \frac{\Lambda_{i+1}(M)}{\Lambda_i(M)} \leq 4e, \forall M \in \Omega$$

For $M \in \mathcal{M}$ it is easy to the bound holds since for $\frac{\Lambda_{i+1}(M)}{\Lambda_i(M)}$, $\Lambda(M)$ as we go from $i \to i+1$ either remains same or gets reduced by a factor of $e^{-1/2}$.
For $M \in \mathcal{N}$ note that since, $5/12 \leq w_i/w_{i+1} \leq 12/5$ we get that

$$\frac{5}{12e^{1/2}} \leq \frac{\Lambda_{i+1}(M)}{\Lambda_i(M)} \leq \frac{12e^{1/2}}{5}, \forall M \in \mathcal{N}$$

Hence now it is obvious that the above claim holds.

Let π_i denote the stationary distribution of the Markov chain used in phase i, so that $\pi_i(M) = \Lambda_i(M)/\Lambda_i(\Omega)$. Let Z_i denote the random variable that is the outcome of the following experiment: By running the desired Markov chain MC with parameters $\Lambda = \Lambda_i$ and $\delta = \varepsilon/80e^2r$, obtain a sample matching M from a distribution that is within variation distance $\varepsilon/80e^2r$ of π_i. Return $\Lambda_{i+1}(M)/\Lambda_i(M)$.
If we had sampled M from the exact stationary distribution π_i instead of an approximation, then the resulting modified random variable Z'_i would have satisfied

$$\mathbb{E}Z'_i = \sum_{M \in \Omega} \frac{\Lambda_i(M)}{\Lambda_i(\Omega)} \frac{\Lambda_{i+1}(M)}{\Lambda_i(M)} = \frac{\Lambda_{i+1}(\Omega)}{\Lambda_i(\Omega)}.$$

As it is, noting the particular choice for δ and bounds on the ratios, and using the fact that $\exp(-x/4) \leq 1 - \frac{1}{5}x \leq 1 + \frac{1}{5}x \leq \exp(x/4)$ for $0 \leq x \leq 1$, we must settle for

$$\exp\left(-\frac{\varepsilon}{4r}\right) \frac{\Lambda_{i+i}(\Omega)}{\Lambda_i(\Omega)} \leq \mathbb{E}Z_i \leq \exp\left(\frac{\varepsilon}{4r}\right) \frac{\Lambda_{i+i}(\Omega)}{\Lambda_i(\Omega)}.$$

Now suppose s independent trials are conducted for each i using the above experiment, and denote by $\bar{Z}_i$ the sample mean of the results. Then $\mathbb{E}\bar{Z}_i = \mathbb{E}Z_i$ and then we have.

$$\exp\left(-\frac{\varepsilon}{4}\right)\frac{\Lambda_r(\Omega)}{\Lambda_0(\Omega)} \leq \mathbb{E}\left(\bar{Z}_0\bar{Z}_1\ldots\bar{Z}_{r-1}\right) \leq \exp\left(\frac{\varepsilon}{4}\right)\frac{\Lambda_r(\Omega)}{\Lambda_0(\Omega)}$$

For s sufficiently large, $\prod_i \bar{Z}_i$ will be close to $\prod_i \mathbb{E}\bar{Z}_i$ with high probability. With a view to quantifying "sufficiently large," observe that in the light of bounds of the ratio,

$$\frac{\mathrm{Var}\left[\bar{Z}_i\right]}{\left(\mathbb{E}\bar{Z}_i\right)^2} \leq \frac{64e^3}{s}$$

Thus, taking $s = \Theta\left(r\varepsilon^{-2}\right)$ we get

$$\begin{aligned}
\frac{\mathrm{Var}\left[\bar{Z}_0\cdots\bar{Z}_{r-1}\right]}{\left(\mathbb{E}\left[\bar{Z}_0\cdots\bar{Z}_{r-1}\right]\right)^2} &= \prod_{i=0}^{r-1}\frac{\mathbb{E}\bar{Z}_i^2}{\left(\mathbb{E}\bar{Z}_i\right)^2} - 1 \\
&= \prod_{i=0}^{r-1}\left(1 + \frac{\mathrm{Var}\,\bar{Z}_i}{\left(\mathbb{E}\bar{Z}_i\right)^2}\right) - 1 \\
&\leq \left(1 + \frac{64e^3}{s}\right)^r - 1 \\
&= O\left(\varepsilon^2\right).
\end{aligned}$$

So, by Chebyshev's inequality,

$$\Pr\left[\exp(-\varepsilon/4)\mathbb{E}\left(\bar{Z}_0\cdots\bar{Z}_{r-1}\right) \leq \bar{Z}_0\cdots\bar{Z}_{r-1} \leq \exp(\varepsilon/4)\mathbb{E}\left(\bar{Z}_0\cdots\bar{Z}_{r-1}\right)\right] \geq \frac{11}{12},$$

assuming the constant implicit in the setting $s = \Theta\left(r\varepsilon^{-2}\right)$ is chosen appropriately. Combining above with the fact that $\Lambda_0(\Omega) = \left(n^2+1\right)n!$, we obtain

$$\Pr\left[\exp(-\varepsilon/2)\Lambda_r(\Omega) \leq \left(n^2+1\right)n!\bar{Z}_0\cdots\bar{Z}_{r-1} \leq \exp(\varepsilon/2)\Lambda_r(\Omega)\right] \geq \frac{11}{12}$$

Now we wish to estimate the ratio $\frac{\Lambda_r(\mathcal{M}_G)}{\Lambda_r(\Omega)}$ which is basically the probability of hitting a valid matching in the stationary distribution of MC_r . And we know that $\frac{\Lambda_r(\mathcal{M}_G)}{\Lambda_r(\Omega)} \geq \frac{1}{(4n^2+2)}$. So now we will use a unbiased estimator to estimate this probability.

For this we will carry out the following experiment -

- By running the Markov chain MC_r and $\delta = \varepsilon/80e^2$, obtain a sample matching M from a distribution that is within variation distance $\varepsilon/80e^2$ of π_r.
- Return 1 if $M \in \mathcal{M}_G$, and 0 otherwise.

The outcome of this experiment is a random variable that we denote by Y. If M had been sampled from the exact stationary distribution π_r then its expectation would have been $\Lambda_r(\mathcal{M}_G)/\Lambda(\Omega)$; as it is, we have

$$\exp\left(-\frac{\varepsilon}{4}\right)\frac{\Lambda_r(\mathcal{M}_G)}{\Lambda_r(\Omega)} \leq \mathbb{E}Y \leq \exp\left(\frac{\varepsilon}{4}\right)\frac{\Lambda_r(\mathcal{M}_G)}{\Lambda_r(\Omega)}.$$

Let $\bar{Y}$ denote the sample mean of $s' = \Theta\left(n^2\varepsilon^{-2}\right)$ independent trials of the above experiment. Since $\mathbb{E}\bar{Y} = \mathbb{E}Y = \Omega\left(n^{-2}\right)$, Chebyshev's inequality gives

$$\Pr[\exp(-\varepsilon/4)\mathbb{E}\bar{Y} \leq \bar{Y} \leq \exp(\varepsilon/4)\mathbb{E}\bar{Y}] \geq \frac{11}{12},$$

as before. Combining this with the probability for the previous ratios , we get

$$\Pr\left[\exp(-\varepsilon)\,|\mathcal{M}_G| \leq \left(n^2+1\right)n!\bar{Y}\bar{Z}_0\bar{Z}_1\cdots\bar{Z}_{r-1} \leq \exp(\varepsilon)|\mathcal{M}_G|\right] \geq \frac{5}{6}.$$

All this was under the assumption that the initialization procedure succeeded. But provided we arrange for the failure probability of initialization to be at most 1/12, it will be seen that $\left(n^2+1\right)n!\bar{Y}\bar{Z}_0\bar{Z}_1\cdots\bar{Z}_{r-1}$ is an estimator for the permanent that meets the specification of an FPRAS.

2.4 Complexity Analysis

In total, the above procedure requires $rs + s' = O\left(\varepsilon^{-2}n^4(\log n)^2\right)$ samples; by Theorem 2.1, $O\left(n^7\log n\right)$ time is sufficient to generate each sample. (Since there is no point in setting $\varepsilon = o(1/n!)$, the $\log\delta^{-1}$ term in Theorem 2.1 can never dominate.) The running time is thus $O\left(\varepsilon^{-2}n^{11}(\log n)^3\right)$. Note that this is sufficient to absorb the cost of the initialization procedure as well, which we will show is $O\left(n^{11}(\log n)^3\right)$.
Now for the intialization step that is setting up $w_i(u,v)$ for all i,u,v. Since markov chain took $O(n^7\log n)$ steps and for each i we generated $O(n^2\log(1/\hat{\eta}))$ and there are total $r = n^2\log n$ iterations of i so we have total $O(n^{11}(\log n)^2\cdot\log(1/\hat{\eta}))$ steps in the intialization process. Now we need to just set $\hat{\eta}$ hence η.
Recall that $\hat{\eta}$ is the failure probability on each occasion that we use a sample mean to estimate an expectation. If we are to achieve overall failure probability η then we must set $\hat{\eta} = O\left(\eta/\left(n^4\log n\right)\right)$, since there are

$O\left(n^4 \log n\right)$ individual estimates to make in total. And we set η to $\delta/2$ or some lesser factor of δ. Now we get total steps equals $O(n^{11}(\log n)^2 \cdot (\log(1/\eta) + \log(n)))$ since $\log(n)$ dominates over $\log(1/\delta) = \log(1/\eta)$ hence we get that the initialization process takes total $O\left(n^{11}(\log n)^3\right)$ time hence we are done.

2.5 General Weights

The algorithm is exactly the same just updating of the w_i's and λ_i's and their initial values is a little bit tweaked . The following is the changed pseudocode for the updates.

Initialize $\lambda(u, v) \leftarrow a_{\max}$ for all $(u, v) \in V_1 \times V_2$. Initialize $w(u, v) \leftarrow n a_{\max}$ for all $(u, v) \in V_1 \times V_2$.
While there exists a pair y, z with $\lambda(y, z) > a(y, z)$ do:
    Take a sample of size S from MC with parameters λ, w, using a simulation of T steps in each case.
    Use the sample to obtain estimates $w'(u, v)$ satisfying condition (8), for all u, v, with high probability.
    Set $\lambda(y, v) \leftarrow \max\{\lambda(y, v) \exp(-1/2), a(y, v)\}$, for all $v \in V_2$, and $w(u, v) \leftarrow w'(u, v)$ for all u, v.
Output the final weights $w(u, v)$.

Where a_{max} and a_{min} are the maximum and minimum non-zero entries of the matrix M whose permanent we need to compute. We actually tweak the error ratios a bit for this and get an approximation algorithm polynomial in n , $\log\left(a_{max}/a_{min}\right)$ and obviously $1/\varepsilon$.

2.6 Improving the complexity

We would now present the enhanced complexity achieved by the paper (). The enhanced complexity is a result of better analysis of the mixing time and a different way to decrease the λ's. In this method the λ's for all non edges get reduced in each iteration (unlike former where we decreased all the λ's for non edges connected to one vertex at a time in one iteration). We have the following algorithm to get the λ's.

Algorithm for computing the cooling schedule λ, given parameters s, c, γ, and D :

Set $\hat{\lambda}_0 = 1, i = s$ and $j = 0$.
While $\hat{\lambda}_j > 1/\gamma$ do
    Set $\hat{\lambda}_{j+1} = c^{-1/i} \hat{\lambda}_j$.
    If $i > D + 1$ and $\hat{\lambda}_{j+1} < (s/\gamma)^{1/(i-D)}$,
        Set $\hat{\lambda}_{j+1} = (s/\gamma)^{1/(i-D)}$ and decrement $i = i - 1$.
    Increment $j = j + 1$.
Set $\ell = j$.

The following lemma estimates the number of intermediate temperatures in the above cooling schedule, i.e., the length ℓ of the $\hat{\lambda}$ sequence.(Previously $\ell = r = n^2 \log(n)$)

Lemma 2.3. *Let $c, \gamma > 0, D \geq 0$ and let $\hat{\lambda}_0, \ldots, \hat{\lambda}_\ell$ be the sequence computed by the above algorithm. Then $\ell = O\left([(D+1)\log(s-D) + s/(s-D)]\log_c \gamma\right)$. If c and D are constants independent of s, then $\ell = O(\log s \log \gamma)$.*

The following lemma shows that for the sequence of the λ_i the value of $Z(\lambda)$ changes by a factor $\leq c$ for consecutive λ_i and λ_{i+1}.

Lemma 2.4. *Let $c, \gamma, D \geq 0$ and let $Z_1, \ldots, Z_q$ be a collection of polynomials of degree s. Suppose that for every $i \in [q]$, the polynomial Z_i satisfies the following conditions: i) Z_i has non-negative coefficients, ii) there exists $d \leq D$ such that the coefficient of x^d in Z_i is at least $Z_i(1)/\gamma$.*
Let $\hat{\lambda}_0, \hat{\lambda}_1, \ldots, \hat{\lambda}_\ell$ be the sequence constructed by the above algorithm. Then

$$Z_i\left(\hat{\lambda}_j\right) \leq cZ_i\left(\hat{\lambda}_{j+1}\right) \quad \text{for every } i \in [q] \text{ and } j \in [\ell].$$

For the above algorithm we feed $s = n, c = \sqrt{2}, \gamma = n!, D = 1$. To get the λ_i's for our implementation. This improves the number of iterations from $n^2 \log(n)$ to $O(n \log^2(n))$. Also from better analysis of mixing time we get the complexity to be $O(n^7 \log^4(n))$, which is best known classically.

3 Quantum Walks

So what we know is that the mixing time of a random walk classically is propotional to $\frac{1}{\delta}$ but we will introduce the notion of quantum walks in this section which gives us a quadratic speedup and we get the mixing time propotional to $\frac{1}{\sqrt{\delta}}$. We would want to use this speedup to get an algorithm to compute the approximate number of perfect matchings in a bipartite graph which has better complexity than the best known classical bound. For that we want to simulate the above markov chain using quantum walks.

3.1 Syzedegy's Walk and Spectral Analysis

We will first present the notion of Syzedgy's quantum walk and do a spectral analysis for it.
We start with the introduction of bipartite walks in the classical framework remarking that every walk can be made bipartite by a simple "duplicating" operation. Let X and Y be two finite sets and $P = (p_{x,y}), Q = (q_{y,x})$ be matrices describing probabilistic maps from X to Y and Y to X, respectively. Since P and Q are stochastic, we have $\sum_{y \in Y} p_{x,y} = 1$ for every $x \in X$ and $\sum_{x \in X} q_{y,x} = 1$ for every $y \in Y$, and all $p_{x,y}, q_{y,x}$ are non-negative. If we have a single probabilistic function P from X to X (i.e. a Markov chain), in order

to create a bipartite walk we can set $q_{y,x} = p_{y,x}$ for every $x, y \in X$, i.e. we set $Q = P$.

We quantize walk (P, Q) by defining two operators on the Hilbert space with basis states

$$\{|x\rangle|y\rangle \mid x \in X, y \in Y\}$$

Define states

$$\phi_x = \sum_{y \in Y} \sqrt{p_{x,y}}|x\rangle|y\rangle$$

for every $x \in X$ and

$$\psi_y = \sum_{x \in X} \sqrt{q_{y,x}}|x\rangle|y\rangle$$

for every $y \in Y$. Let $A = (\phi_x)_x$ be the matrix composed of column vectors $\phi_x (x \in X)$ and $B = (\psi_y)_y$ be the matrix composed of column vectors $\psi_y (y \in Y)$. Our walk operator, W will be the product of

$$\begin{aligned} \mathrm{ref}_1 &= 2AA^* - I; \\ \mathrm{ref}_2 &= 2BB^* - I. \end{aligned}$$

In expression:

$$W = \mathrm{ref}_2 \, \mathrm{ref}_1.$$

Let $\mathcal{C}(A)$ is the column space of A and $\mathcal{C}(B)$ be the column space of B. Observe that $A^*A = I_X$, therefore $(2AA^* - I)\, A = A$. Also, for any $\phi \in \mathcal{C}(A)^\perp$ we have $(2AA^* - I)\, \phi = -\phi$. Therefore ref_1 is a reflection on the subspace $\mathcal{C}(A)$. Similarly, ref $_2$ is a reflection on the subspace $\mathcal{C}(B)$. In the study of W a central role will be played by the matrix $D = A^*B$. We call D the discriminant of matrix of the quantized walk operator W. It follows from the definitions that

$$D_{x,y} = \sqrt{p_{x,y} q_{y,x}} \quad \text{for every } x \in X, y \in Y.$$

Definition(Quantization) Let P, Q, A, B, D, W as above. Then the quantization of bipartite walk (P, Q) is the unitary operator $W_{P,Q} = W$. We also write $A_{P,Q}$, $B_{P,Q}$ and $D_{P,Q}$ for A, B and D, respectively, if (P, Q) is not clear from the context. If $P = Q$ then we use the shorthand W_P for $W_{P,Q}$, etc. Let the columns of matrices A and B be elements of a Hilbert space H and let $A^*A = I_n, B^*B = I_m$. Then operators $ref_A = 2AA^* - I, ref_B = 2BB^* - I$ are reflections on subspaces $\mathcal{C}(A)$ and $\mathcal{C}(B)$, respectively.

In this section we compute spectra of $W = ref_A ref_B$. Our formula turns out to be very useful in the study of quantum walks. Our expressions will use the singular values and vectors of the following matrix:

(Discriminant Matrix) The discriminant matrix for $\mathrm{ref_B}\, \mathrm{ref_A}$ is

$$D(A, B) = A^*B.$$

If we allow the subspaces on which we reflect be represented by arbitrary orthonormed bases (rather than by A and B in particular) then the discriminant matrix is determined only up to unitary multiplicators from left and right.

We can naturally interpret $v \to D(A,B)v$ as a map from $\mathcal{C}(B)$ to $\mathcal{C}(A)$ if we view v as a vector expressed in the $\{b_j\}_{j=1}^m$ bases and the result vector as a one expressed in the $\{a_i\}_{i=1}^n$ bases. To put this into formulas, we can define a map, where Bv is mapped into $(AA^*)\,Bv = AD(A,B)v$. This map is an orthogonal projection of $\mathcal{C}(B)$ to $\mathcal{C}(A)$, since AA^* is an orthogonal projector to the space $\mathcal{C}(A)$. Similarly $D(A,B)^*$ can be interpreted as an orthogonal projection from $\mathcal{C}(A)$ to the space $\mathcal{C}(B)$. Let λ be a singular value of $D(A,B)$ with associated right singular unit vector v and left singular unit vector w. Then we have

$$D(A,B)v = \lambda w$$
$$D(A,B)^* w = \lambda v.$$

Since $|Bv| = |v|, |Aw| = |w|$, and since projections do not increase length we get:

Lemma 3.1. *All singular values of $D(A,B)$ are at most one.*

Using the above observation, for a left-right singular vector pair v, w of $D(A,B)$ we can write the associated singular value as $\cos\theta$ for some $0 \le \theta \le \frac{\pi}{2}$. In the linear algebra literature θ is called a canonical angle between subspaces $\mathcal{C}(A)$ and $\mathcal{C}(B)$. We have that $\theta = 0$ iff the corresponding singular vectors are in $\mathcal{C}(A) \cap \mathcal{C}(B)$. The angle θ has a geometric meaning: it is the angle between Aw and Bv (indeed, $w^*A^*Bv = \cos\theta$).

Theorem 3.2. *Theorem 1 (Spectral Lemma) Let H be a Hilbert space, $A, B \le H$ subspaces, and let $\mathrm{ref_A}$ and $\mathrm{ref_B}$ be reflections on $\mathcal{C}(A)$ and $\mathcal{C}(B)$. Let $\cos\theta_1, \ldots, \cos\theta_l$ be those singular values of $D(A,B) = A^*B$ that lie in the $(0,1)$ open interval, and let the associated singular vector pairs be v_k, w_k for $(1 \le k \le l)$. Then those eigenvalues of the unitary operator $W = \mathrm{ref_B}\,\mathrm{ref_A}$ that have non-zero imaginary part are exactly*

$$e^{-2i\theta_1}, e^{2i\theta_1}, \ldots, e^{-2i\theta_l}, e^{2i\theta_l}.$$

The (un-normalized) eigenvectors associated with these eigenvalues (in order) are

$$Aw_1 - e^{-i\theta_1}Bv_1, Aw_1 - e^{i\theta_1}Bv_1, \ldots, Aw_l - e^{-i\theta_l}Bv_l, Aw_l - e^{i\theta_l}Bv_l.$$

In (3) the singular values are listed with multiplicity. Furthermore,

1. *On $\mathcal{C}(A) \cap \mathcal{C}(B)$ operator W acts as the identity. $\mathcal{C}(A) \cap \mathcal{C}(B)$ coincides with the set of left (right) singular vectors of $D(A,B)$ with singular value 1 .*

2. *On $\mathcal{C}(A) \cap \mathcal{C}(B)^\perp$ operator W acts as reflection on the center (i.e. as $-I$).$\mathcal{C}(A) \cap \mathcal{C}(B)^\perp$ coincides with the set of left singular vectors of $D(A,B)$ with singular value 0 .*

3. *On $\mathcal{C}(B)\cap\mathcal{C}(A)^\perp$ operator W acts as reflection on the center. $\mathcal{C}(B)\cap\mathcal{C}(A)^\perp$ coincides with the set of right singular vectors of $D(A,B)$ with singular value 0.*

4. *On $\mathcal{C}(A)^\perp \cap \mathcal{C}(B)^\perp$ operator W acts as the identity.*

The above is a complete description of the eigenvalues and eigenvectors of operator W acting on H.

Proof. Let $\pi_A = AA^*$ and $\pi_B = BB^*$ be the orthogonal projector operators (in H) to $\mathcal{C}(A)$ and $\mathcal{C}(B)$, respectively. Then $W = (2\pi_B - I)(2\pi_A - I)$. Since for every v, w singular vector pair with singular value $\cos\theta$ we have $\pi_A Bv = (\cos\theta)Aw$ and $\pi_B Aw = (\cos\theta)Bv$ we can conclude that the (at most) two-dimensional subspace $\langle Bv, Aw\rangle$ is invariant under the action W. Moreover, this action is a composition of two reflections, the first on axis Aw and then on axis Bv. The case when $\langle Bv, Aw\rangle$ is one dimensional gives 1-3, and we leave this easy case to the reader. The subspace $\langle Bv, Aw\rangle$ is two-dimensional if and only if $\cos\theta \in (0,1)$. Considering that the angle between Bv and Aw is θ, the formulas for the spectrum and eigenvalues of W on $\langle Bv, Aw\rangle$ follow from the familiar expressions for the two dimensional case. In particular, it is well known that if we reflect on two axes, we obtain a rotation with an angle that is twice of the one between the axes. Since the set of left and right singular vectors form complete systems, we are done with the description of W on $\langle \mathcal{C}(A), \mathcal{C}B\rangle$. We need to describe W on $\langle \mathcal{C}(A), \mathcal{C}(B)\rangle^\perp = \mathcal{C}(A)^\perp \cap \mathcal{C}(B)^\perp$. On this subspace W is simply a product of two reflections on the center, therefore it is the identity. □

3.2 Application of Syzedegy's Walk for sampling Markov Chains

Let $\mathcal{H} = \mathbb{C}^N \otimes \mathbb{C}^N$. The basis states of $\mathcal{H}$ are denoted by $|xy\rangle$ for $x, y \in \Omega$. For $x \in \Omega$, define the normalized vectors

$$|p_x\rangle = \sum_{y\in\Omega} \sqrt{p_{xy}}|y\rangle$$

where p_{xy} denotes the transition probability from x to y. A quantum update is any unitary U that satisfies

$$U|x\rangle|0\rangle = |x\rangle\,|p_x\rangle$$

for some fixed state $0 \in \Omega$ and all $x \in \Omega$. We refer to the cost to realize U and its inverse $U^\dagger$ as the quantum update cost.

To construct the quantum walk, we define the subspaces

$$\mathcal{A} = \text{span}\{|x\rangle|0\rangle : x \in \Omega\}$$
$$\mathcal{B} = U^\dagger S U \mathcal{A}$$

where S denotes that the unitary operator swapping the two tensor components of $\mathcal{H}$. For $\mathcal{K} = \mathcal{A}, \mathcal{B}$, denote by $\Pi_{\mathcal{K}}$ the orthogonal projection onto $\mathcal{K}$ and by

$$R_{\mathcal{K}} = 2\Pi_{\mathcal{K}} - I.$$

the reflection around $\mathcal{K}$.

Definition 3.1 (Quantum Walk)**.** *The quantum walk $W(P)$ based on the classical reversible Markov chain P is defined to be the unitary operation (rotation)*

$$W(P) = R_{\mathcal{B}} \cdot R_{\mathcal{A}}$$

The quantum walk $W(P)$ can be realized by applying both U and $U^\dagger$ twice:

$$W(P) = U^\dagger \cdot S \cdot U \cdot R_{\mathcal{A}} \cdot U^\dagger \cdot S \cdot U \cdot R_{\mathcal{A}}.$$

This is different from the definition used in the previous subsection which is.

$$\tilde{W}(P) = R_{\tilde{\mathcal{B}}} \cdot R_{\tilde{\mathcal{A}}}$$

where

$$\tilde{\mathcal{A}} = \text{span}\left\{|x\rangle\,|p_x\rangle : x \in \Omega\right\} = U\mathcal{A}$$
$$\tilde{\mathcal{B}} = \text{span}\left\{|p_x\rangle\,|x\rangle : x \in \Omega\right\} = U\mathcal{B}.$$

Since $W(P)$ and $\tilde{W}(P)$ are equal up to conjugation by U, we can apply the spectral analysis from previous section to determine the spectrum of $W(P)$. We refer to the subspace $\mathcal{A} + \mathcal{B}$ as the busy subspace and to its orthogonal complement, i.e., $\mathcal{A}^\perp \cap \mathcal{B}^\perp$, as the idle subspace. Clearly, the operator $W(P)$ acts as identity on the idle subspace. On the busy subspace, the spectrum of $W(P)$ is as follows.

Theorem 3.3. *Let P be a time-reversible Markov chain. Let $\theta_1, \ldots, \theta_M \in \left(0, \frac{\pi}{2}\right)$ be such that $|\lambda_1| = \cos(\theta_1), \ldots, |\lambda_M| = \cos\theta_M$ where $M \leq N-1$ and the remaining eigenvalues are equal to 0 , i.e., $\lambda_{M+1}, \ldots, \lambda_{N-1} = 0$.*

1. *On $\mathcal{A} \cap \mathcal{B}$ the operator $W(P)$ acts as the identity I. This subspace is one dimensional and is spanned by the eigenvector $|\pi\rangle|0\rangle$ where*

$$|\pi\rangle = \sum_x \sqrt{\pi_x}|x\rangle$$

   is the quantum sample of the stationary distribution π of P.

2. *On $\mathcal{A} \cap \mathcal{B}^\perp$ and $\mathcal{A}^\perp \cap \mathcal{B}$ the operator $W(P)$ acts as $-I$. The dimensions of $\mathcal{A} \cap \mathcal{B}^\perp$ and $\mathcal{A}^\perp \cap \mathcal{B}$ are equal to $N - 1 - M$, i.e., the dimension of the kernel of P.*

3. *On $\mathcal{A} + \mathcal{B}$ those eigenvalues of $W(P)$ that have nonzero imaginary part are exactly $e^{\pm 2i\theta_1}, \ldots, e^{\pm 2i\theta_M}$ with the same multiplicity.*

4. $W(P)$ has no other eigenvalues on $\mathcal{A}+\mathcal{B}$.

Now we have the following lemmas which lead us to the main result of how to sample using the markov chain.

Lemma 3.4. *Let $|t_0\rangle, \ldots, |t_r\rangle$ be arbitrary quantum states in $\mathbb{C}^d$ with $|\langle t_i \mid t_{i+1}\rangle|^2 \geq p$ for $i = 0, \ldots, r-1$. Given the state $|t_0\rangle$, we can prepare a state $|\tilde{t}_r\rangle$ such that*

$$\| \, |\tilde{t}_r\rangle - |t_r\rangle \, \| \leq \epsilon_1,$$

for any $\epsilon_1 > 0$, by invoking the unitaries from $\left\{R_i, R_i^\dagger \ : \ i = 0, \ldots, r\right\}$ no more than

$$L = \frac{12r \log\left(2r/\epsilon_1\right)}{\log(1/(1-p))}$$

times. Where R_i is defined as the $(2\Pi_i - 1)$, and $\Pi_i = |t_i\rangle\langle t_i|$

Lemma 3.5. *Let W be a unitary acting on $\mathbb{C}^d$ with unique eigenvector $|\psi_0\rangle$ with eigenvalue $\lambda_0 = 1$. Denote the remaining eigenvectors and eigenvalues of W by $|\psi_j\rangle$ and $\lambda_j = e^{2\pi i \varphi_j}$ for $j = 1, \ldots, d-1$, respectively. Let*

$$\Delta = \min_{j=1,\ldots,d-1} |\varphi_j|$$

be the phase gap of W. Let Π be the projector onto the space spanned by $|\psi_0\rangle$ and $\Pi^\perp$ the projector onto the orthogonal complement. Let R be the unitary that acts on $\mathbb{C}^d$ as follows

$$R = \omega\Pi + \Pi^\perp.$$

Let

$$a = \lceil \log(1/\Delta) \rceil$$
$$c = \lceil \log\left(1/\sqrt{\epsilon_2}\right) \rceil$$

for some $\epsilon_2 > 0$. Then, there is a quantum circuit $\tilde{R}$ acting on $\mathbb{C}^d \otimes \left(\mathbb{C}^2\right)^{\otimes ac}$ that invokes the controlled-W gate $2^{a+1} \cdot c$ times and has the following properties: for $j = 0$,

$$\tilde{R} \, |\psi_0\rangle \, |0\rangle^{\otimes ac} = \left(R \, |\psi_0\rangle\right) |0\rangle^{\otimes ac}$$

and for $j \neq 0$,

$$\tilde{R} \, |\psi_j\rangle \, |0\rangle^{\otimes ac} = \left(R \, |\psi_j\rangle\right) |0\rangle^{\otimes ac} + |\xi\rangle,$$

where $|\xi\rangle$ is some error vector in $\mathbb{C}^d \otimes \left(\mathbb{C}^2\right)^{\otimes ac}$ with

$$\| |\xi\rangle \| \leq 2\sqrt{\epsilon_2}.$$

The above two lemmas give us the following theorem . We will quote the following paper () for the proves of the above lemmas. Lemma 3.4 is proved using Grover's Amplitude Amplification for whose proof we quote the following paper ().

Theorem 3.6. *Let $P_0, P_1, \ldots, P_r$ be classical Markov chains with stationary distributions $\pi_0, \pi_1, \ldots, \pi_r$ and spectral gaps $\delta_0, \delta_1, \ldots, \delta_r$, respectively. Assume the stationary distributions of adjacent Markov chains are close to each other in the sense that their stationary distributions π_i and π_{i+1} are close with respect to fidelity, i.e.,*

$$\left(\sum_{x\in\Omega}\sqrt{\pi_i(x)}\sqrt{\pi_{i+1}(x)}\right)^2 = |\langle \pi_i \mid \pi_{i+1}\rangle|^2 \geq p$$

for $i = 0, \ldots, r-1$,

$$\min\{\delta_i : i = 0, \ldots, r\} \geq \delta,$$

and we can prepare the quantum sample $|\pi_0\rangle$. Then, for any $\epsilon > 0$, there is a quantum sampling algorithm, making it possible to sample according to a probability distribution $\tilde{\pi}_r$ that is close to π_r with respect to the total variation distance, i.e., $D(\tilde{\pi}_r, \pi_r) \leq \epsilon$.

The algorithm invokes the controlled- W_i operators at most $2^{a+1}\cdot c\cdot L$ times where

$$\begin{aligned} L &= \frac{12r\log(8r/\epsilon)}{\log(1/(1-p))} \\ a &= \lceil \log(1/\Delta)\rceil \\ c &= \left\lceil \log\left(\frac{96r\log(8r/\epsilon)}{\epsilon\log(1/(1-p))}\right)\right\rceil. \end{aligned}$$

Proof. Lemma 3.4 shows that, given the initial state $|\pi_0\rangle$, we can prepare a state $|\tilde{\pi}_r\rangle$ with $\| |\pi_r\rangle - |\tilde{\pi}_r\rangle \| \leq \epsilon_1$ by invoking the unitaries from the set $\left\{R_i, R_i^\dagger : i = 0, \ldots, r-1\right\}$ no more than

$$L = \frac{12r\log(2r/\epsilon_1)}{\log(1/(1-p))}$$

times. In Lemma 3.4 we assumed that we can implement these exactly. However, in reality we can only implement the operators $\tilde{R}_i$ and their inverses $\tilde{R}_i^\dagger$ as described in Lemma 3.5. This approximation adds an error vector $|\xi\rangle$ every time an operator $\tilde{R}_i$ or $\tilde{R}_i^\dagger$ is applied, where $\||\xi\rangle\| \leq 2\sqrt{\epsilon_2}$ for some $\epsilon_2 > 0$.

Let $|\tilde{\psi}\rangle$ be the state obtained by implementing Lemma 3.5 using $\tilde{R}_i$ to approximate R_i. Then, since these operators or their inverses are invoked no more than L times we have

$$|\tilde{\psi}\rangle = |\tilde{\pi}_r\rangle_{\mathcal{H}} |0\rangle_{\mathcal{A}}^{\otimes ac'} + |\xi\rangle$$

where $c' = \left\lceil \log\left(1/\sqrt{\epsilon_2}\right) \right\rceil$ and $|\xi\rangle$ is some vector with

$$\||\xi\rangle\| \leq 2L\sqrt{\epsilon_2}.$$

$\mathcal{H}$ is the Hilbert space that our quantum samples live in and $\mathcal{A}$ is the Hilbert space of the ancilla qubits that are required to implement the approximate phase gates $\tilde{R}_i$ and their inverses.

Let

$$|\psi\rangle = |\pi_r\rangle\,|0\rangle^{\otimes ac}$$

be the ideal state. We choose $\epsilon_1 = \epsilon/4$ and $\epsilon_2 = \epsilon^2/\left(64L^2\right)$ so that

$$\begin{aligned}
\||\psi\rangle - |\tilde{\psi}\rangle\| &\leq \||\psi\rangle - |\tilde{\pi}_r\rangle\,|0\rangle^{\otimes ac}\| + \|\,|\tilde{\pi}_r\rangle\,|0\rangle^{\otimes ac} - |\tilde{\psi}\rangle\| \\
&\leq \epsilon_1 + 2L\sqrt{\epsilon_2} \\
&= \epsilon/4 + \epsilon/4 \\
&= \epsilon/2.
\end{aligned}$$

For each $x \in \Omega$, we define the projector

$$\Lambda_x = |x\rangle\langle x| \otimes |0\rangle\langle 0|^{\otimes ac}$$

acting on $\mathcal{H} \otimes \mathcal{A}$. Let

$$\Lambda_0 = I_{\mathcal{HA}} - I_{\mathcal{H}} \otimes |0\rangle\langle 0|^{\otimes ac}.$$

Let $\Omega' = \Omega \cup \{0\}$. Observe that the desired distribution π is equal to the probability distribution given by

$$\pi(x) = \|\Lambda_x|\psi\rangle\|^2.$$

Our protocol yields the probability distribution

$$\tilde{\pi}(x) = \|\Lambda_x|\tilde{\psi}\rangle\|^2.$$

We now bound the total variation distance between π and $\tilde{\pi}$ from above. For a subset $S \subseteq \Omega'$, let

$$\Lambda_S = \sum_{x \in S} \Lambda_x$$

We have

$$\begin{aligned}
D(\pi, \tilde{\pi}) &= \max_{S \subseteq \Omega'} |\pi(S) - \pi(\tilde{S})| \\
&. = \max_{S \subseteq \Omega'} |\|\Lambda_S|\psi\rangle\|^2 - \|\Lambda_S|\tilde{\psi}\rangle\|^2\,| \\
&\leq 2 \max_{S \subseteq \Omega'} |\|\Lambda_S|\psi\rangle\| - \|\Lambda_S|\tilde{\psi}\rangle\|\,| \\
&\leq 2\||\psi\rangle - |\tilde{\psi}\rangle\| \\
&\leq \epsilon
\end{aligned}$$

It follows from above lemmas that we invoke the controlled- W_i operators or their inverses at most $2^{a+1} \cdot c \cdot L$ times. □

We would like to use the above theorem to implement our intermediate Markov Chains as they also don't change much in subsequent iterations we would like to use the above algorithm to perform our desired sampling. For that we need to implement the walk operators W_i's. As we can see we use a metropolis filter in the classical algorithm to get a certain probability distribution. So we need to find a method which implements the walk operators W_i given a set of parameters for the metropolis rule.

3.3 Implementing the walk operators

We will try to use the ideas of the following paper () to implement a walk operator. A brief of the ideas of the paper is written below.
Here we will want to prepare a distribution over eigen states of a general hamiltonian. We want a state with the following density operator

$$\rho = \sum_{i \in \Omega} \pi_i |\varphi_i\rangle\langle\varphi_i|$$

, Where φ_i are the eigenstates of a general hamiltonian and π is the stationary probability distribution of a Markov chain over Ω whose transition rule is guided by a metropolis filter. We have

$$\pi_i m_{ij} = \pi_j m_{ji}$$

$$m_{ij} = s_{ij} \cdot z_{ij}$$

$$s_{ij} = s_{ji}$$

$$z_{ij} = \min\{1, \frac{\pi_j}{\pi_i}\}$$

We will denote π_i by $e^{-\beta E_i}$, just to follow the notation of the paper. We will design an unitary for which we get the above state by tracing out from an unique eigenvector with eigenvalue 1 on a certain subspace.
Now the implementation of the paper is as follows.

Now, let us include into our Hilbert space an extra qubit initialized in $|0\rangle$ and define a more compact notation,

$$|i\rangle \equiv |\varphi_i\rangle\,|\tilde{\varphi}_i\rangle\,|0\rangle$$

The information of a Markov chain can be encoded in a pair of unitary operators U_X and U_Y (We will define the construction of U_X in the following part),

$$U_X|i\rangle = \sum_k \left(\sigma_{ik}\,|\varphi_i\rangle\,|\varphi_k\rangle\,|0\rangle + \gamma_{ik}\,|\varphi_i\rangle\,|\varphi_k\rangle\,|1\rangle\right)$$

$$U_Y|j\rangle = \sum_m \left(\sigma_{jm}\,|\varphi_m\rangle\,|\varphi_j\rangle\,|0\rangle + \gamma_{jm}\,|\varphi_j\rangle\,|\varphi_m\rangle\,|1\rangle\right)$$

where $\sigma_{ik} \equiv \alpha_{k\tilde{i}}\sqrt{z_{ik}}, \gamma_{ik} \equiv \alpha_{k\tilde{i}}\sqrt{1-z_{ik}}$, and $\alpha_{k\tilde{i}} \equiv \langle\varphi_k|K|\tilde{\varphi}_i\rangle$. Here K is an unitary operator which encodes the transition without the metropolis filter(we will go in detail in the construction part). Note that U_Y can be obtained from U_X by controlled-SWAP operation.
From the definitions of U_X and U_Y we get that -

$$\left\langle j \left| U_X^\dagger U_Y \right| i \right\rangle = \left| \alpha_{j\tilde{i}} \right|^2 (z_{ij} z_{ji})^{1/2},$$

where we used $\langle\varphi_i \mid \tilde{\varphi}_j\rangle = \langle\varphi_j \mid \tilde{\varphi}_i\rangle$. On the other hand,

$$\left\langle i \left| U_X^\dagger U_Y \right| i \right\rangle = |\alpha_{i\tilde{i}}|^2 + \sum_k |\alpha_{k\tilde{i}}|^2 (1 - z_{ik}),$$

which, as we shall see, can be interpreted as the probability of not undergoing a transition.
Construction of the operator W - Now, using above equation, we obtain the following decomposition:

$$\left\langle j \left| U_X^\dagger U_Y \right| i \right\rangle = \left\langle j \left| D_\pi^{1/2} \right| j \right\rangle \langle j|M|i\rangle \left\langle i \left| D_\pi^{-1/2} \right| i \right\rangle$$

where $D_\pi \equiv \sum_{j=0}^{N-1} \pi_j |j\rangle\langle j|$ is a diagonal matrix, and $M \equiv \sum_{i,j} m_{ij} |j\rangle\langle i|$, with $m_{ii} \equiv \left\langle i \left| U_X^\dagger U_Y \right| i \right\rangle$ and $m_{ij} \equiv \left| \alpha_{j\tilde{i}} \right|^2 z_{ij}$ for $j \neq i$ is the transition matrix of the Markov chain. Within the subspace $\{|i\rangle\}$, the above equation implies that $U_X^\dagger U_Y$ and M are similar matrices, which means that they have the same set of eigenvalues λ_k . This property allows us to construct an operator

$$W \equiv (2\Lambda_2 - I)(2\Lambda_1 - I) \quad ,$$

where

$$\Lambda_1 \equiv \sum_{i=0}^{N-1} |i\rangle\langle i| \quad \text{and} \quad \Lambda_2 \equiv U_X^\dagger U_Y \Lambda_1 U_Y^\dagger U_X.$$

The spectral properties of W can be seen in the following way: define $|\alpha_k\rangle \equiv \sum_{i=0}^{N-1} a_{ki} |i\rangle$ to be the eigenvectors of $\Lambda_1 U_X^\dagger U_Y \Lambda_1$, the eigenvalue equation can be written as

$$\Lambda_1 U_X^\dagger U_Y |\alpha_k\rangle = \lambda_k |\alpha_k\rangle.$$

On the other hand, using the fact that $\Lambda_1 U_X^\dagger U_Y \Lambda_1 = \Lambda_1 U_Y^\dagger U_X \Lambda_1$, we have,

$$\Lambda_2 |\alpha_k\rangle = \lambda_k U_X^\dagger U_Y |\alpha_k\rangle.$$

Now if we start with vectors within Λ_1, W can be block-diagonalized into subspace of 2×2 matrices w_k spanned by the basis $\left\{ |\alpha_k\rangle, U_X^\dagger U_Y |\alpha_k\rangle \right\}$. Explicitly,

$$w_k = \begin{bmatrix} \cos(2\theta_k) & -\sin(2\theta_k) \\ \sin(2\theta_k) & \cos(2\theta_k) \end{bmatrix},$$

where $\cos\theta_k \equiv \lambda_k$. Note that the eigenvalues of w_k is $e^{\pm i\theta_k}$. In the case of $k = 0$ where $\lambda_0 = 1$ (or $\theta_0 = 0$), $w_0 = I$ is simply an identity. Now we get that ,

$$|\alpha_0\rangle = \sum_{i=0}^{N-1} \sqrt{\pi_i}|i\rangle.$$

Recall that $|i\rangle \equiv |\varphi_i\rangle\,|\tilde{\varphi}_i\rangle\,|0\rangle$, the state $|\alpha_0\rangle$ becomes the Gibbs thermal state $\rho_{th} = e^{-\beta H}/\operatorname{Tr}\left[e^{-\beta H}\right]$ when the other qubits are traced out.
One of the most important features about W is that the minimum eigenvalue gap $\Delta_{\min} \equiv |2\theta_1|$ of W is less than two times the square root of the gap $\delta \equiv 1 - \lambda_1$ of the transition matrix M (using $2\theta \geq \left|1 - e^{2i\theta}\right| = 2\sqrt{1 - \cos^2\theta}$)

$$\Delta_{\min} \geq 2\sqrt{\delta}$$

Construction of U_X - Here we show how one may construct the unitary operator U_X defined as

$$U_X|i\rangle = \sum_k \left(\sigma_{ik}\,|\varphi_i\rangle\,|\varphi_k\rangle\,|0\rangle + \gamma_{ik}\,|\varphi_i\rangle\,|\varphi_k\rangle\,|1\rangle\right),$$

where

$$\sigma_{ik} \equiv \alpha_{k\tilde{i}}\sqrt{z_{ik}} \quad , \quad \gamma_{ik} \equiv \alpha_{k\tilde{i}}\sqrt{1 - z_{ik}}$$

and

$$\alpha_{k\tilde{i}} \equiv \langle\varphi_k|K|\tilde{\varphi}_i\rangle\,.$$

Here K is an unitary operator which plays the same role as the spin-flip in the classical Metropolis method(K in the basis $\{|\varphi_i\rangle\}_i$ is kind of equivalent to transition without the metropolis filter), and z_{ik} is the Metropolis filter . Note that U_Y can be obtained from U_X by a controlled-SWAP. We assume that K is symmetrical in the computational basis:

$$\langle x'|K|x\rangle = \langle x|K|x'\rangle\,.$$

We are now ready to consider the explicit procedure for constructing U_X . Starting with the paired state

$$|\varphi_i\rangle\,|\tilde{\varphi}_i\rangle$$

we apply the "kick" operator K to $|\tilde{\varphi}_i\rangle$, and write

$$K\,|\tilde{\varphi}_i\rangle = \sum_k \alpha_{k\tilde{i}}\,|\varphi_k\rangle\,.$$

Next, we implement the Metropolis filter by performing a controlled-rotation (based on the difference of the eigenvalues):

$$|\varphi_i\rangle\,|\varphi_k\rangle\,|0\rangle \to |\varphi_i\rangle\,|\varphi_k\rangle\left(\sqrt{z_{ik}}|0\rangle + \sqrt{1 - z_{ik}}|1\rangle\right),$$

where

$$z_{ij} \equiv \left\{1, e^{-\beta(E_j - E_i)}\right\}$$

This creates a state as described by definition of U_X.