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 $w^*(u, v)/2 \le w(u, v) \le 2w^*(u, v)$ because of the following theorem.

  Theorem 2.1. *Assuming the weight function w_i satisfies inequality $w_i^*(u, v)/2 \le w_i(u, v) \le 2w_i^*(u, v)$ for all $(u, v) \in V_1 \times V_2$ with $\mathcal{M}(u, v) \ne \emptyset$, 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 $P(M, M) \ge 1/2$ for all states M, and any initial state $M_0 \in \Omega$,*
  $$\tau_{M_0}(\delta) \le \varrho \left(\ln \pi (M_0)^{-1} + \ln \delta^{-1}\right).$$

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

$$\Lambda_i(\mathcal{M}(u,v)) = \sum_{M \in \mathcal{M}(u,v)} \lambda_i(M) w_i(u,v)$$

$$= w_i(u,v) \sum_{M \in \mathcal{M}(u,v)} \lambda_i(M)$$

$$= w_i(u,v) \lambda_i(\mathcal{M}(u,v))$$

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

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

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 $\delta = O(n^{-2})$ 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-

$$\implies \pi(\mathcal{M}) + \sum_{(u,v)} \pi(\mathcal{M}(u,v)) = 1$$

$$\implies \pi(\mathcal{M}) + \sum_{(u,v)} \frac{\lambda(\mathcal{M}(u,v))w(u,v)}{\Lambda(\Omega)} = 1$$

$$\implies \pi(\mathcal{M}) + \sum_{(u,v)} \frac{\lambda(\mathcal{M})w(u,v)}{\Lambda(\Omega)w^*(u,v)} = 1$$

$$\implies \pi(\mathcal{M}) + \sum_{(u,v)} \frac{2\lambda(\mathcal{M})}{\Lambda(\Omega)} \geq 1$$

$$\implies \pi(\mathcal{M})(1 + 2n^2) \geq 1$$

$$\implies \pi(\mathcal{M}) \geq \frac{1}{(2n^2 + 1)}$$

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

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 $1 - (n^2 + 1) \hat{\eta}$, 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

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

Assuming $\delta \leq 1/8 (n^2 + 1)$, it follows from any of the standard Chernoff bounds , that $O(n^2 \log(1/\hat{\eta}))$ 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 $\pi(\mathcal{M}(u,v)) \geq 1/4 (n^2 + 1)$, 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 $S = O(n^2 \log(1/\hat{\eta}))$ 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 $\frac{\Lambda_{i+1}(\Omega)}{\Lambda_i(\Omega)}$ 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,

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

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

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

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

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

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 $\log(1/\delta) = \log(\eta)$ 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 $w(u,v) \leftarrow na_{\max}$ 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 $1/\varepsilon$.