

Approximate Cholesky Factorization

Lecturer: Sushant Sachdeva

Scribe: Maxim Goukhshtein

1 Notation

We introduce the following notation:

- $S^{(k)}$: Schur complement at k i.e., $S^{(k)} = S^{(k-1)} + c_k c_k^T$ where $c_k = \frac{1}{\sqrt{S_{\pi(k),\pi(k)}^{(k-1)}}} S_{:, \pi(k)}^{(k-1)}$.
- Cl_k : Clique at k .
- $(L)_v$: Laplacian of a star graph incident on v with all edges belonging to the graph whose Laplacian is L . Therefore, we can write $S^{(k)} = S^{(k-1)} + \left(S^{(k-1)}\right)_{\pi(k)} + Cl_k$.
- For a matrix M , $\overline{M} = L^{-\frac{1}{2}} M L^{-\frac{1}{2}}$, where L is the Laplacian of the original graph.
- We use $\mathbb{E}_{|k-1}[\cdot]$ to denote the expectation conditioned on everything that happened in the first $k-1$ iterations.

2 Clique Sampling

In the previous lecture, the following clique sampling algorithm was introduced.

Algorithm 1 CliqueSample(G, v)

- 1: **for** every edge $e = (v, u_1) \in E$ **do**
 - 2: Sample another edge (v, u_2) with probability $\frac{w(v, u_2)}{\deg(v)}$
 - 3: **if** $u_1 \neq u_2$ **then**
 - 4: $Y_e \leftarrow \frac{w(v, u_1)w(v, u_2)}{w(v, u_1) + w(v, u_2)} L_{(u_1, u_2)}$
 - 5: **end if**
 - 6: **end for**
 - 7: **Return** $\widehat{Cl} = \sum_e Y_e$.
-

3 Approximate Cholesky Factorization Algorithm

Algorithm 2 is used for obtaining the approximate Cholesky factorization.

Algorithm 2 ApproximateCholesky(G)

- 1: Replace each edge e with ρ parallel copies, each with weight $\frac{w(e)}{\rho}$.
 - 2: Pick a random permutation π on V .
 - 3: **for** $k = 1$ **to** n **do**
 - 4: “Record column”: $c_k \leftarrow \frac{1}{\sqrt{S_{\pi(k), \pi(k)}^{(k-1)}}} S_{:, \pi(k)}^{(k-1)}$.
 - 5: “Eliminate and Sample”: $\widehat{Cl}_k \leftarrow \text{CliqueSample}\left(S^{(k-1)}, \pi(k)\right)$,
 - 6: $\widehat{S}^{(k)} \leftarrow \widehat{S}^{(k-1)} + (\widehat{S}^{(k-1)})_{\pi(k)} + \widehat{Cl}_k$.
 - 7: **end for**
 - 8: **Return** $[c_1, c_2, \dots, c_n]$.
-

4 Proof of Algorithm Approximation Result

Our goal is to show that the returned approximate normalized Laplacian satisfies the target approximation $\widehat{L}^n \approx_{\frac{1}{3}} L$, or equivalently

$$\begin{aligned}
& \frac{2}{3}L \preceq \widehat{L}^n \preceq \frac{4}{3}L \\
& \Leftrightarrow -\frac{1}{3}L \preceq \widehat{L}^n - L \preceq \frac{1}{3}L \\
& \Leftrightarrow -\frac{1}{3}\Pi \preceq \widehat{L}^n - \bar{L} \preceq \frac{1}{3}\Pi \\
& \Leftrightarrow \|\widehat{L}^n - \bar{L}\| \leq \frac{1}{3}.
\end{aligned}$$

In order to show this result, we first introduce a number of useful lemmas.

Lemma 4.1. $\mathbb{E}_{|k-1}[\widehat{Cl}_k] = Cl_k$

Proof.

$$\begin{aligned}
\mathbb{E}_{|k-1}[\widehat{Cl}_k] &= \mathbb{E}_{|k-1} \left[\sum_e Y_e \right] \\
&= \sum_e \mathbb{E}_{|k-1}[Y_e] \\
&= \sum_{u_1} \sum_{u_2} \frac{w(v, u_2)}{\deg(v)} \frac{w(v, u_1) w(v, u_2)}{w(v, u_1) + w(v, u_2)} L_{(u_1, u_2)} \\
&= \sum_{u_1 < u_2} \frac{w(v, u_1) w(v, u_2)}{\deg(v)} \frac{w(v, u_1) + w(v, u_2)}{w(v, u_1) + w(v, u_2)} L_{(u_1, u_2)} \\
&= Cl_k
\end{aligned}$$

□

Consider the difference of the resulting Laplacians in two consecutive iterations:

$$\begin{aligned}
\widehat{L}^{(k)} - \widehat{L}^{(k-1)} &= \widehat{S}^{(k)} - \widehat{S}^{(k-1)} + c_k c_k^T \\
&= - \left(\widehat{S}^{(k-1)} \right)_{\pi(k)} + \widehat{C} l_k + c_k c_k^T \\
&= -c_k c_k^T - C l_k + \widehat{C} l_k + c_k c_k^T \\
&= \widehat{C} l_k - C l_k.
\end{aligned}$$

Therefore, it follows from Lemma 4.1, that $\mathbb{E}_{|k-1} [\widehat{L}^{(k)}] = \widehat{L}^{(k-1)}$ (hence, $\{\widehat{L}^{(k)}\}_{k \geq 0}$ is a matrix martingale).

The following Lemma concerning a distance property of the effective resistance will prove useful.

Lemma 4.2 (Triangle Inequality for Effective Resistance). $\forall a, b, c \in V$, $\text{Reff}(a, b) \leq \text{Reff}(a, c) + \text{Reff}(c, b)$, or equivalently, $\|\bar{L}_{a,b}\| \leq \|\bar{L}_{a,c}\| + \|\bar{L}_{c,b}\|$.

Proof.

$$\begin{aligned}
\text{Reff}(a, b) &= (\mathbb{1}_a - \mathbb{1}_b)^T L^\dagger (\mathbb{1}_a - \mathbb{1}_b) \\
&= (\mathbb{1}_a - \mathbb{1}_c + \mathbb{1}_c - \mathbb{1}_b)^T L^\dagger (\mathbb{1}_a - \mathbb{1}_c + \mathbb{1}_c - \mathbb{1}_b) \\
&= (\mathbb{1}_a - \mathbb{1}_c)^T L^\dagger (\mathbb{1}_a - \mathbb{1}_c) + (\mathbb{1}_c - \mathbb{1}_b)^T L^\dagger (\mathbb{1}_c - \mathbb{1}_b) + 2 (\mathbb{1}_a - \mathbb{1}_c)^T L^\dagger (\mathbb{1}_c - \mathbb{1}_b) \\
&= \text{Reff}(a, c) + \text{Reff}(c, b) + 2 (\mathbb{1}_a - \mathbb{1}_c)^T L^\dagger (\mathbb{1}_c - \mathbb{1}_b) \\
&\leq \text{Reff}(a, c) + \text{Reff}(c, b),
\end{aligned}$$

where the inequality holds since $(\mathbb{1}_a - \mathbb{1}_c)^T L^\dagger (\mathbb{1}_c - \mathbb{1}_b) \leq 0$. This is true since $L^\dagger (\mathbb{1}_c - \mathbb{1}_b)$ corresponds to the voltage drops in a resistor network where an external current of one unit is sent from c to b , hence $(\mathbb{1}_a - \mathbb{1}_c)^T L^\dagger (\mathbb{1}_c - \mathbb{1}_b)$ is the voltage drop from a to c , that is $v(a) - v(c)$. Since c is the source in this network, it has the highest potential in the network, therefore $v(c) \geq v(a)$. \square

Lemma 4.3. Any edge e added in Algorithm 2 to \widehat{S}^k , satisfies $\|Y_e\| = \|w(e)\bar{L}_e\| = w(e)\|L^{-\frac{1}{2}}L_eL^{-\frac{1}{2}}\| \leq \frac{1}{\rho}$.

Proof. We prove the lemma by induction. At $k = 0$, before “splitting” $w(e)\bar{L}_e \leq \Pi$, therefore after “splitting” $\|\frac{w(e)}{\rho}\bar{L}_e\| \leq \frac{1}{\rho}$. Now, suppose that the lemma holds for step $k - 1$. Suppose edge (u_1, u_2) is added at step k , then

$$\begin{aligned}
\frac{w(v, u_1) w(v, u_2)}{w(v, u_1) + w(v, u_2)} \|\bar{L}_{u_1, u_2}\| &\leq \frac{w(v, u_1) w(v, u_2)}{w(v, u_1) + w(v, u_2)} (\|\bar{L}_{v, u_1}\| + \|\bar{L}_{v, u_2}\|) \\
&= \frac{w(v, u_2)}{w(v, u_1) + w(v, u_2)} (w(v, u_1) \|\bar{L}_{v, u_1}\|) \\
&\quad + \frac{w(v, u_1)}{w(v, u_1) + w(v, u_2)} (w(v, u_2) \|\bar{L}_{v, u_2}\|) \\
&\leq \frac{w(v, u_2)}{w(v, u_1) + w(v, u_2)} \frac{1}{\rho} + \frac{w(v, u_1)}{w(v, u_1) + w(v, u_2)} \frac{1}{\rho}
\end{aligned}$$

$$= \frac{1}{\rho},$$

where the first inequality follows from Lemma 4.2 and the second from the inductive hypothesis. \square

Finally, in order to prove the approximation result, we will make use of the following matrix concentration result.

Theorem 4.4 (Matrix Freedman Inequality). *Consider a matrix martingale $\{Y_k \in \mathbb{R}^{n \times n}\}_{k \geq 0}$ with $Y_0 = 0$ (i.e., a random process satisfying $\mathbb{E}_{|k-1}[Y_k] = Y_{k-1}$). Define $X_k = Y_k - Y_{k-1}$ such that $\lambda_{\max}(X_k) \leq R$ for all k . Further, define the “predictable quadratic variation” $W_k = \sum_{i=1}^k \mathbb{E}_{|i-1}[X_i^2]$. Then, $\forall t, \sigma^2 \geq 0$,*

$$\Pr \left(\exists j : \lambda_{\max}(Y_j) \geq t \text{ and } \lambda_{\max}(W_k) \leq \sigma^2 \right) \leq n \exp \left\{ -\frac{\frac{t^2}{2}}{\sigma^2 + \frac{tR}{3}} \right\}.$$

Proof of Algorithm Approximation Result. In the following we will consider each edge sampling individually. We introduce a notation to indicate the call index to CliqueSample and sampled edge index. For example, $Y_e^{(k)}$ is the Laplacian of the edge e sampled at the k^{th} call to CliqueSample. We will use $e-1$ to indicate the edge sampled before e (i.e., $Y_{e-1}^{(k)}$ is the Laplacian of the edge $e-1$ sampled just before edge e , whose Laplacian is $Y_e^{(k)}$).

We had previously shown that

$$\begin{aligned} \widehat{\bar{L}}^{(k)} - \widehat{\bar{L}}^{(k-1)} &= \widehat{\bar{C}l}_k - \bar{C}l_k \\ &= \sum_e \bar{Y}_e + \sum_e \mathbb{E}_{|k-1} [\bar{Y}_e] \\ &= \sum_e \bar{Y}_e + \mathbb{E}_{|k-1} [\bar{Y}_e], \end{aligned}$$

from which it follows that

$$\widehat{\bar{L}}^{(k)} - \widehat{\bar{L}}^{(0)} = \sum_{i=1}^k \sum_e \bar{Y}_e^{(i)} + \mathbb{E}_{|i-1} [\bar{Y}_e^{(i)}], \quad (1)$$

where the summation of the edges is over **all** the edges in the each i^{th} call of CliqueSample (i.e., in the j^{th} call to CliqueSample we consider all the edges incident on vertex v in Algorithm 1).

Consider the random matrix $T_{k,e'}$ sequence defined as

$$T_{k,e'} = \left(\sum_{i=1}^{k-1} \sum_e \bar{Y}_e^{(i)} + \mathbb{E}_{|i-1} [\bar{Y}_e^{(i)}] \right) + \left(\sum_{e \leq e'} Y_e^{(k)} - \mathbb{E}_{|k-1} [\bar{Y}_e^{(k)}] \right), \quad (2)$$

where the summation over edges in the k^{th} call to CliqueSample is over all the edges up to and including edge e' . To reiterate, the difference between equations 1 and 2 is that in equation 1 we consider the summation over all the edges in each of the calls of CliqueSample, whereas in equation 2 we consider all the sampled edges in the first $k-1$ calls and the sum of all the edges sampled “so far” in the k^{th} call to CliqueSample.

To simplify notation, we let $X_e^{(k)} = \bar{Y}_e^{(k)} - \mathbb{E}_{|i-1} [\bar{Y}_e^{(k)}]$ and write $T_{k,e'} = \sum_{(i,e) \leq (k,e')} X_e^{(i)}$. Furthermore, we use $\mathbb{E}_{|<(k,e')}[\cdot]$ to denote the expectation conditioned on all edge samplings until (and including) the sampling of edge e' in the k^{th} call to CliqueSample.

Note that $\|T_{k,e'}\| \leq \frac{1}{3}$ for all k and e' implies $\|\bar{L}^k - \bar{L}\| \leq \frac{1}{3}$. Therefore, we will use $T_{k,e'}$ to prove the algorithm approximation result. First note that $T_{k,e'}$ is a martingale of the form described in Theorem 4.4, since trivially $T_{0,0} = 0$ and

$$\begin{aligned}
\mathbb{E}_{|<(k,e')} [T_{k,e'}] &= \mathbb{E}_{|<(k,e')} \left[\sum_{(i,e) \leq (k,e')} X_e^{(i)} \right] \\
&= \mathbb{E}_{|<(k,e')} \left[\sum_{(i,e) < (k,e')} X_e^{(i)} \right] - \mathbb{E}_{|<(k,e')} [X_{e'}^{(k)}] \\
&= \sum_{(i,e) < (k,e')} X_e^{(i)} + \mathbb{E}_{|<(k,e')} [\bar{Y}_{e'}^{(k)} - \mathbb{E}_{|k-1} [\bar{Y}_{e'}^{(k)}]] \\
&\stackrel{(*)}{=} \sum_{(i,e) < (k,e')} X_e^{(i)} + \mathbb{E}_{|k-1} [\bar{Y}_{e'}^{(k)}] - \mathbb{E}_{|k-1} [\bar{Y}_{e'}^{(k)}] \\
&= \sum_{(i,e) < (k,e')} X_e^{(i)} \\
&= T_{k,e'-1},
\end{aligned}$$

where $\mathbb{E}_{|<(k,e')} [\bar{Y}_{e'}^{(k)}] = \mathbb{E}_{|k-1} [\bar{Y}_{e'}^{(k)}]$ in $(*)$ follows from the fact that sampling of each edge within a call to CliqueSample is independent of any other edge sampling in the same call.

Now, note that $T_{k,e'} - T_{k,e'-1} = X_{e'}^{(k)}$, then

$$\begin{aligned}
\|X_{e'}^{(k)}\| &= \|Y_{e'}^{(k)} - \mathbb{E}_{|k-1} [Y_{e'}^{(k)}]\| \\
&\leq \|Y_{e'}^{(k)}\| \\
&\leq \frac{1}{\rho},
\end{aligned}$$

where the last inequality follows from Lemma 4.3. Therefore, when can bound the maximum eigenvalue of $X_{k,e'}$ as per Theorem 4.4, $\lambda_{\max}(X_{e'}^{(k)}) \leq R = \frac{1}{\rho}$.

Next, we wish to bound $W_{k,e'} = \sum_{(i,e) \leq (k,e')} \mathbb{E}_{|<(i,e)} \left[\left(X_{e'}^{(k)} \right)^2 \right]$,

$$\begin{aligned}
W_{k,e'} &\preceq \sum_{i=1}^n \sum_e \mathbb{E}_{|i-1} \left[\left(X_e^{(k)} \right)^2 \right] \\
&= \sum_{i=1}^n \sum_e \mathbb{E}_{|i-1} \left[\overline{(Y_e^{(i)})^2} - \left(\mathbb{E}_{|i-1} [Y_e^{(i)}] \right)^2 \right]
\end{aligned}$$

$$\begin{aligned}
& \preccurlyeq \sum_{i=1}^n \sum_e \mathbb{E}_{|i-1} \left[(\overline{Y_e^{(i)}})^2 \right] \\
& \preccurlyeq \frac{1}{\rho} \sum_{i=1}^n \sum_e \mathbb{E}_{|i-1} \left[\overline{Y_e^{(i)}} \right] \\
& = \frac{1}{\rho} \sum_{i=1}^n \overline{Cl_i}.
\end{aligned}$$

We know construct a new martingale to bound $W_{k,e'}$ by applying Theorem 4.4 a second time. Let $Z_j = \frac{1}{\rho} \sum_{i=1}^j \overline{Cl_i}$ and

$$\begin{aligned}
A_j &= \left(Z_j - \mathbb{E}_{|j-1} [Z_j] \right) + A_{j-1} \\
&= \frac{1}{\rho} \sum_{i=1}^j \overline{Cl_i} - \frac{1}{\rho} \sum_{i=1}^{j-1} \overline{Cl_i} - \frac{1}{\rho} \mathbb{E}_{|j-1} [\overline{Cl_j}] + A_{j-1} \\
&= \frac{1}{\rho} \left(\overline{Cl_j} - \mathbb{E}_{|j-1} [\overline{Cl_j}] \right) + A_{j-1}.
\end{aligned}$$

A_j is a martingale of the type used in Theorem 4.4, since $A_0 = 0$ and

$$\mathbb{E}_{|j-1} [A_j] = \frac{1}{\rho} \left(\mathbb{E}_{|j-1} [\overline{Cl_j}] - \mathbb{E}_{|j-1} [\overline{Cl_j}] \right) + \mathbb{E}_{|j-1} [A_{j-1}] = A_{j-1}.$$

Since $\overline{Cl_j} \preccurlyeq (\overline{L})_j \preccurlyeq \widehat{\overline{L^{j-1}}} \preccurlyeq \frac{4}{3}\Pi$, we have $0 \preccurlyeq \frac{1}{\rho} \overline{Cl_j} \preccurlyeq \frac{4}{3\rho}\Pi$ and similarly $0 \preccurlyeq \mathbb{E}_{|j-1} \left[\frac{1}{\rho} \overline{Cl_j} \right] \preccurlyeq \frac{4}{3\rho}\Pi$. Therefore, $-\frac{4}{3\rho}\Pi \preccurlyeq \frac{1}{\rho} \overline{Cl_j} - \mathbb{E}_{|j-1} \left[\frac{1}{\rho} \overline{Cl_j} \right] \preccurlyeq \frac{4}{3\rho}\Pi$. Hence,

$$\|A_j - A_{j-1}\| = \frac{1}{\rho} \|\overline{Cl_j} - \mathbb{E}_{|j-1} [\overline{Cl_j}]\| \leq \frac{4}{3\rho},$$

so we choose $\lambda_{\max}(A_j - A_{j-1}) \leq \frac{4}{3\rho} = R$.

Now we find a value for the bound σ^2 ,

$$\begin{aligned}
\sum_{j=1}^n \mathbb{E}_{|j-1} \left[(A_j - A_{j-1})^2 \right] &= \sum_{j=1}^n \mathbb{E}_{|j-1} \left[\left(\frac{1}{\rho} \left(\overline{Cl_j} - \mathbb{E}_{|j-1} [\overline{Cl_j}] \right) \right)^2 \right] \\
&\preccurlyeq \sum_{j=1}^n \mathbb{E}_{|j-1} \left[\frac{4}{3\rho^2} \overline{Cl_j} \right] \\
&\preccurlyeq \frac{4}{3\rho^2} \sum_{j=1}^n \mathbb{E}_{|j-1} (\overline{L})_j \\
&= \frac{4}{3\rho^2} \sum_{j=1}^n \frac{1}{n-j+1} \widehat{\overline{L^{(j-1)}}} \\
&\preccurlyeq \frac{4}{3\rho^2} \frac{4}{3} \sum_{j=1}^n \frac{2}{n-j+1} \Pi
\end{aligned}$$

$$\preccurlyeq \frac{64}{9\rho^2} \log n \Pi.$$

Hence, we can choose $\sigma^2 = \frac{64}{9\rho^2} \log n$ (and, as stated above, $R = \frac{4}{3\rho}$) giving us according to Theorem 4.4 that $\lambda_{\max}(A_j) \geq t$ with probability $\leq n \exp \left\{ -\frac{t^2}{\sigma^2 + \frac{tR}{3}} \right\}$. We let $n \exp \left\{ -\frac{t^2}{\sigma^2 + \frac{tR}{3}} \right\} \leq \frac{1}{n^3}$, which implies that we need $\frac{t^2}{\sigma^2 + \frac{tR}{3}} \geq 8 \log n$. This in turn implies that we need to choose $t \geq \max\{\sqrt{\sigma^2 8 \log n}, \frac{R}{3} 8 \log n\} = \max\{\frac{16}{3\rho} \sqrt{2 \log n}, \frac{32}{9\rho} \log n\}$. By choosing $t = \frac{32}{3\rho} \log n$, we get that $\lambda_{\max}(A_j) \leq t$ with high probability. Therefore, by induction $\lambda_{\max}(A_n) \leq \frac{32}{3\rho} \log n$.

Furthermore,

$$\begin{aligned} A_n &= \frac{1}{\rho} \left(\sum_{j=1}^n \overline{Cl}_j - \sum_{j=1}^n \mathbb{E}_{|j-1} [\overline{Cl}_j] \right) \\ &\succcurlyeq \frac{1}{\rho} \left(\sum_{j=1}^n \overline{Cl}_j - \sum_{j=1}^n \mathbb{E}_{|j-1} \left[\left(\overline{L} \right)_j \right] \right) \\ &= \frac{1}{\rho} \left(\sum_{j=1}^n \overline{Cl}_j - \sum_{j=1}^n \frac{1}{n-j+1} \overline{2L^{(j-1)}} \right) \\ &\succcurlyeq \frac{1}{\rho} \sum_{j=1}^n \overline{Cl}_j - \frac{1}{\rho} \frac{4}{3} \log n \Pi \\ &= Z_n - \frac{16}{3\rho} \log n \Pi, \end{aligned}$$

Going back to the first application of Theorem 4.4 we find

$$W_{k,e'} \preccurlyeq Z_n \preccurlyeq A_n + \frac{16}{3\rho} \log n \Pi \preccurlyeq \frac{16}{\rho} \log n \Pi.$$

Therefore, to apply Theorem 4.4 on the original martingale we have (from before) $R = \frac{1}{\rho}$ and $\sigma^2 = \frac{16}{\rho} \log n$. As before, we need to choose $t \geq \max\{\sqrt{\sigma^2 8 \log n}, \frac{R}{3} 8 \log n\} = \max\{\sqrt{\frac{2}{\rho}} 8 \log n, \frac{1}{3\rho} 8 \log n\}$.

By choosing $t = \frac{16}{3\rho} \log n$, $\lambda_{\max}(T_{k,e'}) \leq t$. To get our result, we want to choose $t \leq \frac{1}{3}$. And so by choosing $\rho = \frac{16}{9} \log n$, we get that $\|T_{k,e'}\| \leq \frac{1}{3}$ implying $\|\widehat{\overline{L}}^k - \overline{L}\| \leq \frac{1}{3}$. Using induction, we find $\|\widehat{\overline{L}}^n - \overline{L}\| \leq \frac{1}{3}$, that is $\frac{2}{3}L \preccurlyeq \widehat{L}^n \preccurlyeq \frac{4}{3}L$. \square