

Approximate Gaussian Elimination for Laplacian Systems

MA4291 Presentation

Ang Yan Sheng
A0144836Y

6 Nov 2017

Laplacian matrices

Definition

A *Laplacian matrix* is a symmetric matrix \mathbf{L} such that:

- $L_{ii} \geq 0$, and $L_{ij} \leq 0$ for $i \neq j$; and
- the entries in each row sum to 0.

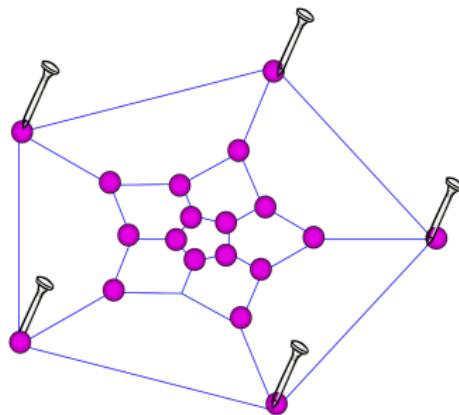
$$\begin{array}{ccc} \text{Laplacians} & \longleftrightarrow & \text{Weighted graphs} \\ \mathbf{L} \in \mathbb{R}^{n \times n} & & G = ([n], E), w : E \rightarrow \mathbb{R}_{\geq 0} \end{array}$$

$$\mathbf{L} = \sum_{i,j} w_{ij} \times \begin{matrix} i & j \\ j & i \end{matrix} \begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix}$$

$$= \sum_{i,j} w_{ij} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$$

$$\mathbf{x}^T \mathbf{L} \mathbf{x} = \sum_{i,j} w_{ij} (x_i - x_j)^2$$

Example: springy graphs



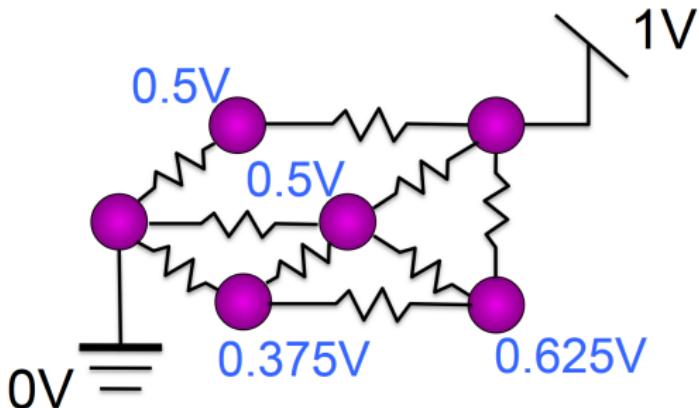
Vertex positions (x_i, y_i) , spring constants w_{ij}

Elastic potential energy: $\frac{1}{2} \sum_{i,j} w_{ij}(x_i - x_j)^2$

Energy minimised when

$$x_i = \frac{\sum_{j \neq i} w_{ij} x_j}{\sum_j w_{ij}} \implies \mathbf{Lx} = \mathbf{b}$$

Example: resistor networks

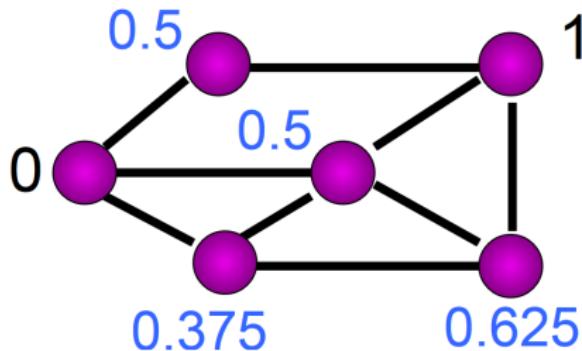


Voltages x_i , resistances w_{ij}^{-1}

Voltages minimise power dissipated $\frac{1}{2} \sum_{i,j} w_{ij}(x_i - x_j)^2$
(Dirichlet Principle)

$$x_i = \frac{\sum_{j \neq i} w_{ij} x_j}{\sum_j w_{ij}} \implies \mathbf{L}\mathbf{x} = \mathbf{b}$$

Example: learning on graphs



Function values x_i , edge weights w_{ij}

$$\text{Minimise } \frac{1}{2} \sum_{i,j} w_{ij} (x_i - x_j)^2$$

$$x_i = \frac{\sum_{j \neq i} w_{ij} x_j}{\sum_j w_{ij}} \implies \mathbf{L}\mathbf{x} = \mathbf{b}$$

The main problem

Main problem

Given Laplacian $\mathbf{L} \in \mathbb{R}^{n \times n}$ with m nonzero entries, and $\mathbf{b} \in \mathbb{R}^n$, efficiently solve $\mathbf{Lx} = \mathbf{b}$ to accuracy ε .

LU/Cholesky factorisation: $O(n^3)$, even for sparse \mathbf{L}

Spielman-Teng 2004: $O(m \ln^{50} n \ln(1/\varepsilon))$

Cohen et al. 2014: $O(m \ln^{1/2} n \ln(1/\varepsilon))$

Kyng 2017: $O(m \ln^2 n \ln(1/\varepsilon))$, simple

Linear iteration

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Iterative method:

$$\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k + \mathbf{Nb}$$

$$\mathbf{x}_* = \mathbf{M}\mathbf{x}_* + \mathbf{Nb}$$

Convergence:

$$\mathbf{e}_{k+1} = \mathbf{Me}_k$$

$$\implies \|\mathbf{M}\| < 1$$

Consistency:

$$\mathbf{x} = \mathbf{x}_* = (\mathbf{I} - \mathbf{M})^{-1}\mathbf{Nb}$$

$$\implies \mathbf{M} = \mathbf{I} - \mathbf{NA}$$

Definition

$$\mathbf{B} \approx_c \mathbf{A} \iff \|\mathbf{B}^{-1}\mathbf{A} - \mathbf{I}\| \leq c$$

Want: approximate Cholesky factorisation

$$\mathbf{U} \text{ sparse upper triangular, } \mathbf{U}^T \mathbf{U} \approx_{1/2} \mathbf{L}$$

Cholesky factorisation

$$\begin{pmatrix} 16 & -4 & -8 & -4 \\ -4 & 5 & 0 & -1 \\ -8 & 0 & 9 & -1 \\ -4 & -1 & -1 & 6 \end{pmatrix} = \mathbf{c}_1 \mathbf{c}_1^T + \begin{pmatrix} 4 & -2 & -2 \\ -2 & 5 & -3 \\ -2 & -3 & 5 \end{pmatrix} \quad \mathbf{c}_1 = \begin{pmatrix} 4 \\ -1 \\ -2 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} 4 & -2 & -2 \\ -2 & 5 & -3 \\ -2 & -3 & 5 \end{pmatrix} = \mathbf{c}_2 \mathbf{c}_2^T + \begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix} \quad \mathbf{c}_2 = \begin{pmatrix} 0 \\ 2 \\ -1 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix} = \mathbf{c}_3 \mathbf{c}_3^T \quad \mathbf{c}_3 = \begin{pmatrix} 0 \\ 0 \\ 2 \\ -2 \end{pmatrix}$$

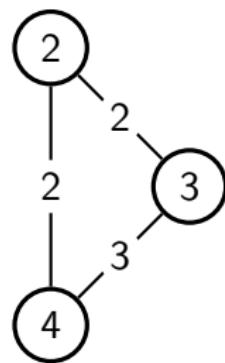
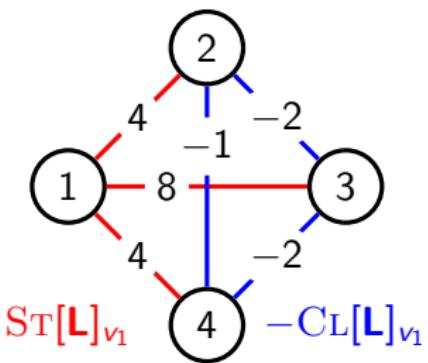
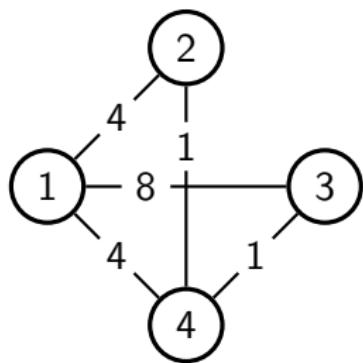
Cholesky factorisation

$$\begin{pmatrix} 16 & -4 & -8 & -4 \\ -4 & 5 & 0 & -1 \\ -8 & 0 & 9 & -1 \\ -4 & -1 & -1 & 6 \end{pmatrix} = (\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3 \quad \mathbf{0}) (\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3 \quad \mathbf{0})^T \\ = \mathbf{U}^T \mathbf{U}$$

$$\mathbf{U} = \begin{pmatrix} 4 & -1 & -2 & -1 \\ & 2 & -1 & -1 \\ & & 2 & -1 \\ & & & 0 \end{pmatrix}$$

Cholesky factorisation

$$\begin{pmatrix} 16 & -4 & -8 & -4 \\ -4 & 5 & 0 & -1 \\ -8 & 0 & 9 & -1 \\ -4 & -1 & -1 & 6 \end{pmatrix} = \begin{pmatrix} 16 & -4 & -8 & -4 \\ -4 & 1 & 2 & 1 \\ -8 & 2 & 4 & 2 \\ -4 & 1 & 2 & 1 \end{pmatrix} + \begin{pmatrix} 4 & -2 & -2 \\ -2 & 5 & -3 \\ -2 & -3 & 5 \end{pmatrix}$$



Remove one vertex, add a clique to its neighbours

The algorithm

CHOLFAC(**L**)

Split each edge into $\rho = 12 \ln^2(1/\delta)$ copies

Pick random permutation of vertices v_1, \dots, v_n

S₀ \leftarrow **L**

for $i = 1, \dots, n$ **do**

$$\mathbf{c}_i \leftarrow \begin{cases} \frac{1}{\sqrt{\mathbf{S}_{i-1}(v_i, v_i)}} \mathbf{S}_{i-1}(v_i, :) & \text{if } \mathbf{S}_{i-1}(v_i, v_i) \neq 0 \\ \mathbf{0} & \text{else} \end{cases}$$

C_i \leftarrow CL[**S**_{i-1}]_{v_i}

S_i \leftarrow **S**_{i-1} - ST[**S**_{i-1}]_{v_i} + **C**_i

end for

$$\mathbf{U} \leftarrow \begin{pmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_n \end{pmatrix}^T$$

return **U**

The algorithm

CHOLAPX'(\mathbf{L}, δ)

Split each edge into $\rho = 12 \ln^2(1/\delta)$ copies

Pick random permutation of vertices v_1, \dots, v_n

$\mathbf{S}_0 \leftarrow \mathbf{L}$

for $i = 1, \dots, n$ **do**

$$\mathbf{c}_i \leftarrow \begin{cases} \frac{1}{\sqrt{\mathbf{S}_{i-1}(v_i, v_i)}} \mathbf{S}_{i-1}(v_i, :) & \text{if } \mathbf{S}_{i-1}(v_i, v_i) \neq 0 \\ \mathbf{0} & \text{else} \end{cases}$$

$\mathbf{C}_i \leftarrow \text{CLIQUESAMPLE}(\mathbf{S}_{i-1}, v_i)$

$\mathbf{S}_i \leftarrow \mathbf{S}_{i-1} - \text{ST}[\mathbf{S}_{i-1}]_{v_i} + \mathbf{C}_i$

end for

$$\mathbf{U} \leftarrow \begin{pmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_n \end{pmatrix}^T$$

return \mathbf{U}

Idea: add $\deg(v)$ random edges instead of $\binom{\deg(v)}{2}$

The algorithm

CLIQUESAMPLE(\mathbf{S}, v)

for $e = 1, \dots, \deg_{\mathbf{S}}(v)$ **do**

 Sample (v, u_1) with probability $w(v, u_1)/w_{\mathbf{S}}(v)$

 Sample (v, u_2) uniformly

$$\mathbf{Y}_e \leftarrow \frac{w(v, u_1)w(v, u_2)}{w(v, u_1) + w(v, u_2)} (\mathbf{e}_{u_1} - \mathbf{e}_{u_2})(\mathbf{e}_{u_1} - \mathbf{e}_{u_2})^T$$

end for

return $\sum_e \mathbf{Y}_e$ ▷ expected value = CL[\mathbf{S}] $_v$

Theorem (Kyng 2017)

CHOLAPX'(\mathbf{L}, δ) ($\delta < n^{-100}$) returns \mathbf{U} with $O(m \ln^2(1/\delta) \ln n)$ nonzero entries such that $\mathbf{U}^T \mathbf{U} \approx_{1/2} \mathbf{L}$ with probability $1 - O(\delta)$. Moreover, CHOLAPX'(\mathbf{L}, δ) runs in $O(tm \ln^2(1/\delta) \ln n)$ time with probability $1 - n^{-t}$ for $t > 1$.

The matrix martingale

$$\mathbf{c}_i \leftarrow \frac{1}{\sqrt{\mathbf{S}_{i-1}(v_i, v_i)}} \mathbf{S}_{i-1}(v_i, :) \quad \text{if } \mathbf{S}_{i-1}(v_i, v_i) \neq 0$$

$\mathbf{C}_i \leftarrow \text{CLIQUESAMPLE}(\mathbf{S}_{i-1}, v_i)$

$\mathbf{S}_i \leftarrow \mathbf{S}_{i-1} - \text{ST}[\mathbf{S}_{i-1}]_{v_i} + \mathbf{C}_i$

Let $\mathbf{L}_i = \mathbf{S}_i + \sum_{j=1}^i \mathbf{c}_j \mathbf{c}_j^T$. Note that $\mathbf{L}_n = \mathbf{U}^T \mathbf{U}$.

$$\begin{aligned}\mathbf{L}_i - \mathbf{L}_{i-1} &= \mathbf{S}_i - \mathbf{S}_{i-1} + \mathbf{c}_i \mathbf{c}_i^T \\ &= \mathbf{C}_i - \text{ST}[\mathbf{S}_{i-1}]_{v_i} + \mathbf{c}_i \mathbf{c}_i^T \\ &= \mathbf{C}_i - \text{CL}[\mathbf{S}_{i-1}]_{v_i}\end{aligned}$$

$(\mathbf{L}_i - \mathbf{L})_{i=0}^n$ is a zero-mean martingale!

We want to bound

$$\mathbb{P} \left(\|\mathbf{L}^{-1} \mathbf{L}_n - \mathbf{I}\| \geq \frac{1}{2} \right) = \mathbb{P} \left(\|\overline{\mathbf{L}_n} - \mathbf{I}\| \geq \frac{1}{2} \right),$$

where $\overline{\mathbf{A}} := \mathbf{L}^{-1/2} \mathbf{A} \mathbf{L}^{-1/2}$ if \mathbf{A} symmetric with $\ker(\mathbf{L}) \subseteq \ker(\mathbf{A})$.

First try

$$\mathbf{Y}_{i,e} \leftarrow \frac{w(v,u_1)w(v,u_2)}{w(v,u_1)+w(v,u_2)}(\mathbf{e}_{u_1} - \mathbf{e}_{u_2})(\mathbf{e}_{u_1} - \mathbf{e}_{u_2})^T$$

Let $\mathbf{X}_{i,e} = \mathbf{Y}_{i,e} - \mathbb{E}_{<(i,e)} \mathbf{Y}_{i,e}$:

$$\mathbf{L}_i - \mathbf{L}_{i-1} = \mathbf{C}_i - \text{CL}[\mathbf{S}_{i-1}]_{v_i} = \sum_e \mathbf{X}_{i,e}.$$

Can show $\|\overline{\mathbf{Y}_{i,e}}\| \leq 1/\rho$, so

$$\|\overline{\mathbf{X}_{i,e}}\| \leq \max(\|\overline{\mathbf{Y}_{i,e}}\|, \|\mathbb{E}_{<(i,e)} \overline{\mathbf{Y}_{i,e}}\|) \leq 1/\rho.$$

Matrix Azuma: martingale length $\sim m\rho[\frac{1}{n} + \dots + \frac{1}{2} + 1] \sim m\rho \ln n$,

$$\begin{aligned}\mathbb{P}(\|\overline{\mathbf{L}_n} - \mathbf{I}\| \geq t) &\leq \exp\left(\frac{-t^2}{8 \sum 1/\rho^2}\right) \\ &\leq \exp\left(\frac{-Kt^2}{m\rho \ln n / \rho^2}\right) \\ &= \exp\left(\frac{-K't^2 \ln n}{m}\right)\end{aligned}$$

Second try

$$\mathbf{Y}_{i,e} \leftarrow \frac{w(v,u_1)w(v,u_2)}{w(v,u_1)+w(v,u_2)}(\mathbf{e}_{u_1} - \mathbf{e}_{u_2})(\mathbf{e}_{u_1} - \mathbf{e}_{u_2})^T$$

$$\mathbf{X}_{i,e} = \mathbf{Y}_{i,e} - \mathbb{E}_{<(i,e)} \mathbf{Y}_{i,e}$$

$$\mathbf{L}_i - \mathbf{L}_{i-1} = \sum_e \mathbf{X}_{i,e}$$

$$\|\overline{\mathbf{X}_{i,e}}\| \leq 1/\rho$$

Define a new martingale with stopping condition:

$$\mathbf{Z}_{i,e} = \begin{cases} \overline{\mathbf{X}_{i,e}} & \text{if } \|\sum_{j < i} \sum_f \mathbf{Z}_{j,f}\| \leq \frac{1}{2} \\ \mathbf{0} & \text{else} \end{cases}$$

$$\mathbf{T}_{i,e} = \sum_{(j,f) \leq (i,e)} \mathbf{Z}_{j,f}$$

Then

$$\mathbb{P} \left(\|\overline{\mathbf{L}_n} - \mathbf{I}\| \geq \frac{1}{2} \right) \leq \mathbb{P} \left(\exists (i, e) : \|\mathbf{T}_{i,e}\| \geq \frac{1}{2} \right).$$

Matrix Freedman inequality

Theorem (Freedman 1975)

Let $(A_k)_{k \geq 0}$ be a real-valued martingale with $B_k \leq R$, where $B_k = A_k - A_{k-1}$. Let $W_k = \sum_{j=1}^k \mathbb{E}_{<j}(B_j^2)$. Then for all $t \geq 0$ and $\sigma^2 > 0$,

$$\mathbb{P}(\exists k : A_k \geq t \text{ and } W_k \leq \sigma^2) \leq \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right).$$

Theorem (Tropp 2011)

Let $(\mathbf{A}_k)_{k \geq 0}$ be a symmetric $d \times d$ -matrix martingale with $\lambda_{\max}(\mathbf{B}_k) \leq R$, where $\mathbf{B}_k = \mathbf{A}_k - \mathbf{A}_{k-1}$. Let $\mathbf{W}_k = \sum_{j=1}^k \mathbb{E}_{<j}(\mathbf{B}_j^2)$. Then for all $t \geq 0$ and $\sigma^2 > 0$,

$$\mathbb{P}(\exists k : \lambda_{\max}(\mathbf{A}_k) \geq t \text{ and } \|\mathbf{W}_k\| \leq \sigma^2) \leq 2d \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right).$$

Freedman inequality: proof

The function $g(x) = \frac{e^x - 1 - x}{x^2}$ is increasing.

$$e^{\lambda x} \leq 1 + \lambda x + (e^\lambda - 1 - \lambda)x^2 \text{ for all } \lambda \geq 0, x \leq 1.$$

Let B be a r.v. with $B \leq 1$, $\mathbb{E}B = 0$. Then

$$\begin{aligned}\mathbb{E} \exp(\lambda B) &\leq 1 + (e^\lambda - 1 - \lambda)\mathbb{E}(B^2) \\ &\leq \exp((e^\lambda - 1 - \lambda)\mathbb{E}(B^2)).\end{aligned}$$

Take $B = B_n$:

$$\begin{aligned}&\exp(\lambda A_{k-1} - (e^\lambda - 1 - \lambda)W_{k-1}) \\ &\geq \mathbb{E}_{< k} \exp(\lambda(A_{k-1} + B_k) - (e^\lambda - 1 - \lambda)(W_{k-1} + \mathbb{E}_{< k}(B_k^2))) \\ &= \mathbb{E}_{< k} \exp(\lambda A_k - (e^\lambda - 1 - \lambda)W_k)\end{aligned}$$

Hence $(\exp(\lambda A_k - (e^\lambda - 1 - \lambda)W_k))_k$ is a supermartingale.

Freedman inequality: proof

Let $s = \min\{k : A_k \geq t\}$ if defined, ∞ otherwise. Then

$$\mathbb{E} \left(\mathbb{1}_{s < \infty} \exp(\lambda A_s - (e^\lambda - 1 - \lambda) W_s) \right) \leq 1.$$

Let E be the event that $A_k \geq t$ and $W_k \leq \sigma^2$ for some k .

Now $s < \infty$ on E , so

$$\begin{aligned} 1 &\geq \mathbb{E} \left(\mathbb{1}_E \exp(\lambda A_s - (e^\lambda - 1 - \lambda) W_s) \right) \\ &\geq \mathbb{P}(E) \exp(\lambda t - (e^\lambda - 1 - \lambda) \sigma^2) \\ \mathbb{P}(E) &\leq \exp(-\lambda t + (e^\lambda - 1 - \lambda) \sigma^2). \end{aligned}$$

Use $e^\lambda - 1 - \lambda \leq \frac{\lambda^2}{2!} \left(1 + \frac{t}{3} + (\frac{t}{3})^2 + \dots\right) = \frac{\lambda^2/2}{1-\lambda/3}$, $\lambda = \frac{t}{\sigma^2+t/3}$:

$$\mathbb{P}(E) \leq \exp \left(\frac{-t^2/2}{\sigma^2 + t/3} \right).$$

Matrix Freedman: proof sketch

If \mathbf{B} is a symmetric r.v. with $\mathbb{E}\mathbf{B} = \mathbf{0}$ and $\lambda_{\max}(\mathbf{B}) \leq 1$, then

$$\mathbb{E} \exp(\theta \mathbf{B}) \preccurlyeq \exp\left((e^\theta - \theta - 1)\mathbb{E}(\mathbf{B}^2)\right).$$

Let $S_k(\theta) := \text{Tr} \exp(\theta \mathbf{A}_k - (e^\theta - \theta - 1)\mathbf{W}_k)$. Then $S_0 = d$, and

$$\begin{aligned} S_{k-1} &= \text{Tr} \exp(\theta \mathbf{A}_{k-1} - (e^\theta - \theta - 1)\mathbf{W}_{k-1}) \\ &= \text{Tr} \exp(\theta \mathbf{A}_{k-1} - (e^\theta - \theta - 1)(\mathbf{W}_k - \mathbb{E}_{<k}(\mathbf{B}_j^2))) \\ &\geq \text{Tr} \exp(\theta \mathbf{A}_{k-1} - (e^\theta - \theta - 1)\mathbf{W}_k + \ln \mathbb{E}_{<k} \exp(\theta \mathbf{B}_j)) \\ &\stackrel{*}{\geq} \mathbb{E}_{<k} \text{Tr} \exp(\theta \mathbf{A}_{k-1} - (e^\theta - \theta - 1)\mathbf{W}_k + \theta \mathbf{B}_k) \\ &= \mathbb{E}_{<k} S_k. \end{aligned}$$

Theorem (Lieb 1973)

If \mathbf{H} is symmetric, then $\mathbf{A} \mapsto \text{Tr} \exp(\mathbf{H} + \ln \mathbf{A})$ is concave on the set of positive definite matrices.

Hence $(S_k)_k$ is a supermartingale, finish as before.

Main theorem: proof sketch

$$\|\mathbf{Z}_{i,e}\| \leq 1/\rho$$

$$\mathbf{T}_{i,e} = \sum_{(j,f) \leq (i,e)} \mathbf{Z}_{j,f}$$

Let $\mathbf{W}_{i,e} = \sum_{(j,f) \leq (i,e)} \mathbb{E}_{<(j,f)}(\overline{\mathbf{Z}_{j,f}}^2)$. Then

$$\begin{aligned} \mathbb{P} \left(\exists(i, e) : \|\mathbf{T}_{i,e}\| \geq \frac{1}{2} \right) &\leq \mathbb{P} \left(\exists(i, e) : \|\mathbf{T}_{i,e}\| \geq \frac{1}{2} \text{ and } \|\mathbf{W}_{i,e}\| \leq \sigma^2 \right) \\ &\quad + \mathbb{P} \left(\exists(i, e) : \|\mathbf{W}_{i,e}\| \geq \sigma^2 \right) \end{aligned}$$

First term: Matrix Freedman ($\sigma^2 = \ln(1/\delta)/\rho$, $R = 1/\rho$, $t = 1/2$)

$$\mathbb{P}_1 \leq n \exp \left(\frac{-t^2/2}{\sigma^2 + Rt/3} \right) \leq 2\delta.$$

Second term: Matrix Freedman

$$\mathbf{W}_i = \mathbf{W}_{i,e_{\text{last}}}$$

$$\mathbf{V}_i = \mathbf{W}_i - \mathbb{E}_{< i} \mathbf{W}_i$$

$$\mathbf{R}_i = \sum_{j=1}^i \mathbf{V}_j$$

$$\mathbf{M}_i = \sum_{j=1}^i \mathbb{E}_{< j} \mathbf{V}_j^2$$

CHOLAPX(\mathbf{L}, δ)

...

Pick v_i uniformly from vertices with **at most twice average degree**

...

Theorem (Kyng 2017)

CHOLAPX(\mathbf{L}, δ) ($\delta < n^{-100}$) returns \mathbf{U} with $O(m \ln^2(1/\delta) \ln n)$ nonzero entries such that $\mathbf{U}^T \mathbf{U} \approx_{1/2} \mathbf{L}$ with probability $1 - O(\delta)$. Moreover, CHOLAPX(\mathbf{L}, δ) runs in $O(m \ln^2(1/\delta) \ln n)$ time.

Random Gaussian row elimination

- Consistent runtimes for different graph families
- Open problem: why does this work?

References

- ▶ R. Kyng.
Approximate Gaussian Elimination.
Yale University, PhD dissertation, 2017.
- ▶ D. Spielman.
“The Laplacian Matrices of Graphs.”
IEEE International Symposium on Information Theory.
Barcelona, 16 Jul 2016. Plenary Lecture.
- ▶ J. Tropp.
“Freedman’s Inequality for Matrix Martingales.”
Elect. Comm. in Probab., **16**, 2011, pp. 262–270.