Tropp: Matrix Sparsification

$A \in \mathbb{R}^{d_1 \times d_2}$ ... find sparse $\hat{A}$, such that $\|A-\hat{A}\|$ is small.

- Zvolíme $(p_{ij})_{i,j=1}^{d_1,d_2}$ pravděpodobnosti: $\sum_{i,j} p_{ij} = 1$.

- $S = \frac{a_{ij}}{p_{ij}} E_{ij}$ ... E_{ij} ... matice s 1 na (i,j), jinak nula

  - náhodná matice s $d_1 \times d_2$ možnými výsledky

- $\mathbb{E} S = \sum_{i,j} \left(\frac{a_{ij}}{p_{ij}} E_{ij}\right) p_{ij} = \sum_{i,j} a_{ij} E_{ij} = A$

- $\hat{A}_r = \frac{1}{r} \sum_{k=1}^{r} S_k$ ... součet r ~~nez~~ nezávislých kopií S

  - $\hat{A}_r$ má (nejvýše) r nezávislých nenulových hodnot

  - "výběr s opakováním": je možné, že $\hat{A}_{ij} = \frac{2a_{ij}}{p_{ij}}$ ... etc.

  - oproti výběru r párů (i,j) ... NEZÁVISLOST!

    ... jednodušší analýza!

- Ověříme předpoklady v r-ty! $p_{ij} = \frac{1}{2}\left[\frac{|a_{ij}|^2}{\|A\|_F^2} + \frac{|a_{ij}|}{\|A\|_1}\right]$ ... $\sum p_{ij} = 1$.

- Spektrální norma: ~~$\|S - \mathbb{E}S\| = \|S - A\|$~~

  - $p_{ij} \geq \frac{1}{2} \frac{|a_{ij}|^2}{\|A\|_F^2}$; $p_{ij} \geq \frac{1}{2} \frac{|a_{ij}|}{\|A\|_1}$

• $Z = \hat{A}_n - A = \frac{1}{n}\sum_{k=1}^{n} S_k - A = \frac{1}{n}\sum_{k=1}^{n}(S_k - A) = \sum_{k=1}^{n} Z_k$

Předpoklady věty!:

• Z_k jsou nezávislé

• $\mathbb{E} Z_k = \mathbb{E}\left[\frac{1}{n}(S_k - A)\right] = 0$

• $\|Z_k\| = \frac{1}{n}\|S_k - A\| \le \frac{1}{n}\max_{(i,j)} \left\|\frac{a_{ij}}{p_{ij}}E_{ij} - A\right\|$

$$\le \frac{1}{n}\max_{(i,j)}\left\{\left\|\frac{a_{ij}}{p_{ij}}E_{ij}\right\| + \|A\|\right\}$$

$$\le \frac{1}{n}\max_{(i,j)} \frac{|a_{ij}|}{p_{ij}} + \frac{1}{n}\|A\|$$

• $\|A\| = \|\mathbb{E} S\| \le \mathbb{E}\|S\| \le \max_{(i,j)} \frac{|a_{ij}|}{p_{ij}}$

$$\left\|\sum_{(i,j)} \frac{a_{ij}}{p_{ij}} \cdot p_{ij} E_{ij}\right\| \le \sum_{i,j} p_{ij}\left\|\frac{a_{ij}}{p_{ij}}E_{ij}\right\| = \cancel{\sum_{i,j} p_{ij}} = \mathbb{E}\|S\|$$

$$\Rightarrow \|Z_k\| \le \underbrace{\frac{2}{n}\max_{(i,j)} \frac{|a_{ij}|}{p_{ij}}}_{L}$$

• Rozptyl

$$v := \max\left\{\left\|\sum_k \mathbb{E}(Z_k Z_k^*)\right\|, \left\|\sum_k \mathbb{E}(Z_k^* Z_k)\right\|\right\}$$

$$\sum_{k=1}^{n} \mathbb{E}\left\{\frac{1}{n^2}(S_k - A)(S_k - A)^*\right\} = \frac{n}{n^2}\mathbb{E}(S_k - A)(S_k^* - A^*)$$

$$= \frac{1}{n}\left\{\mathbb{E}(S_k S_k^*) - \mathbb{E}(A S_k^*) - \mathbb{E}(S_k A^*) + \mathbb{E} A A^*\right\}$$

$$= \frac{1}{n}\left\{\mathbb{E}(S_k S_k^*) - AA^* - AA^* + AA^*\right\} = \frac{1}{n}\left\{\mathbb{E}(SS^*) - AA^*\right\}$$

Celkem $0 \preceq n\,\mathbb{E} ZZ^* = \sum_{k=1}^{n} \mathbb{E}(Z_k Z_k^*) = \frac{1}{n}\left[\mathbb{E}(SS^*) - AA^*\right] \preceq \frac{1}{n}\mathbb{E}(SS^*)$

$$\Rightarrow v \le \frac{1}{n}\max\left\{\|\mathbb{E}(SS^*)\|, \|\mathbb{E}(S^*S)\|\right\}$$

$$\mathbb{E}(SS^*) = \,?$$

$$(SS^*)_{uv} = \sum_{j=1}^{d_2} S_{uj} (S^*)_{jv} = \sum_{j=1}^{d_2} S_{uj} S_{vj}$$

$$= 0 \text{ for } u \neq v$$

$$= \cancel{\frac{a_{uj}}{p_{uj}}} \sum_{j=1}^{d_2} (S_{uj})^2 \text{ for } u = v$$

$$\mathbb{E}(SS^*)_{uv} = \begin{cases} 0 \ \ldots \ u \neq v \\ \displaystyle\sum_{j=1}^{d_2} p_{uj} \frac{a_{uj}^2}{p_{uj}^2} = \sum_{j=1}^{d_2} \frac{a_{uj}^2}{p_{uj}} & u = v \end{cases}$$

$$\|\mathbb{E}(SS^*)\| = \max_{i=1,\ldots,d_1} \sum_{j=1}^{d_2} \frac{a_{ij}^2}{p_{ij}}$$

$$\ldots \ \nu \leq \frac{1}{r} \max \left\{ \max_{i=1,\ldots,d_1} \sum_{j=1}^{d_2} \frac{a_{ij}^2}{p_{ij}} \,;\, \max_{j=1,\ldots,d_2} \sum_{i=1}^{d_1} \frac{a_{ij}^2}{p_{ij}} \right\}.$$

$$\Longrightarrow \mathbb{E}\Big\|\sum_k Z_k\Big\| = \mathbb{E}\|\hat{A}_r - A\| \leq \sqrt{2\nu \log(d_1+d_2)} + \frac{L}{3}\log(d_1+d_2)$$

$$L = \frac{2}{r} \max_{(i,j)} \frac{|a_{ij}|}{p_{ij}} \,;\ \nu \text{ výše},\ p_{ij} = \frac{1}{2}\left[\frac{|a_{ij}|^2}{\|A\|_F^2} + \frac{|a_{ij}|}{\|A\|_1} \right].$$

$p_{ij} \geq \frac{1}{2}\frac{|a_{ij}|^2}{\|A\|_F^2}$; $p_{ij} \geq \frac{1}{2}\frac{|a_{ij}|}{\|A\|_1} \longrightarrow \ldots \quad \frac{|a_{ij}|}{p_{ij}} \leq 2\|A\|_1$

$L \leq \frac{2}{r}\cdot 2\|A\|_1$

$\downarrow$

$\frac{a_{ij}^2}{p_{ij}} \leq 2\|A\|_F^2 \quad \ldots \quad \nu \leq \frac{1}{r}\cdot \max(d_1,d_2)\cdot 2\|A\|_F^2$

$$\Rightarrow \mathbb{E}\|\hat{A}_r - A\| \leq \sqrt{4\cdot\frac{1}{r}\cdot\max(d_1,d_2)\cdot\|A\|_F^2\log(d_1+d_2)} + \frac{4}{3}\cdot\frac{1}{r}\cdot\|A\|_1\cdot\log(d_1+d_2)$$

$$\leq 2\sqrt{\frac{1}{r}\max(d_1,d_2)\log(d_1+d_2)\cdot \mathrm{srank}(A)\|A\|^2} + \frac{4}{3}\cdot\frac{1}{r}\log(d_1+d_2)\sqrt{d_1 d_2}\underbrace{\|A\|_F}_{\leq\sqrt{\mathrm{srank}(A)}\|A\|}$$

$$\Rightarrow \frac{\mathbb{E}\|\hat{A}_r - A\|}{\|A\|} \leq \underbrace{2\sqrt{\frac{1}{r}\max(d_1,d_2)\log(d_1+d_2)\,\mathrm{srank}(A)}}_{\leq 2\varepsilon} + \underbrace{\frac{4}{3}\cdot\frac{1}{r}\cdot\log(d_1+d_2)\sqrt{d_1 d_2}\sqrt{\mathrm{srank}(A)}}_{\leq 2\varepsilon}$$

$\Rightarrow r \geq \varepsilon^{-2}\max(d_1,d_2)\log(d_1+d_2)\,\mathrm{srank}(A) \longleftarrow$ wins!

$\& \; r \geq \frac{2}{3}\log(d_1+d_2)\sqrt{d_1 d_2}\sqrt{\mathrm{srank}(A)}\cdot\varepsilon^{-1}$

$\blacksquare$

$\mathrm{srank}(A) := \frac{\|A\|_F^2}{\|A\|^2} \leq \mathrm{rank}(A)$

$\|A\|_F^2 = \sum_{j=1}^{\mathrm{rank}(A)} \sigma_j^2(A) \leq \|A\|^2\cdot\mathrm{rank}(A)$