

Contents

1	General Sketching method	2
2	Block Coordinate Descent Method	4
2.1	Algorithm	4
2.2	Convergence rate	4
3	Randomized orthonormal systems	6
3.1	Algorithm	6
3.2	Convergence rate	6
4	Count-min Sketches	8
4.1	Algorithm	8
4.2	Convergence rate	8
4.3	Sparse Shuffling (Spashu)	10
5	Conclusion	14

1. General Sketching method

A is a $n \times n$ positive definite matrix representing our problem.

s is the sketch size.

$\{S_i\}_{i=1,\dots,r}$ is the set of r realizations of our $s \times n$ sketch matrix.

We denote by S the $s \times n$ random sketch matrix, which is such that $S = S_i$ with probability p_i .

Throughout the computations, we denote by $Z = AS^T(SAS^T)^{-1}SA$. That is a quantity that intervenes in the computation of the convergence rate¹.

The convergence rate is defined by $\rho = 1 - \lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}})$.

By definition, $A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = \sum_i p_i A^{\frac{1}{2}} S_i^T (S_i A S_i^T)^{-1} S_i A^{\frac{1}{2}}$

for any $i \in \{1, \dots, r\}$, $A^{\frac{1}{2}} S_i^T (S_i A S_i^T)^{-1} S_i A^{\frac{1}{2}}$ is a projection matrix (a matrix such that $M^2 = M$) and then its eigenvalues are a nonempty subset of $\{0, 1\}$.

Since λ_{\max} is a convex function, we obtain that :

$$0 \leq \lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \leq \lambda_{\max}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \leq \sum_i p_i \lambda_{\max}(A^{\frac{1}{2}} S_i^T (S_i A S_i^T)^{-1} S_i A^{\frac{1}{2}}) \leq 1.$$

Denote by $C = (S_1^T, \dots, S_r^T)$ which is of size $n \times rs$.

Lemma 1.0.1 $A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = (A^{\frac{1}{2}}CD)(DC^T A^{\frac{1}{2}})$ where $D = \text{diag}(\sqrt{p_1}(S_1 A S_1^T)^{-\frac{1}{2}}, \dots, \sqrt{p_r}(S_r A S_r^T)^{-\frac{1}{2}}) \in \mathcal{M}_{rs}(\mathbb{R})$. Plus :

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \frac{\lambda_{\min}(A)\lambda_{\min}(CC^T)}{\lambda_{\max}(A)} \min_i \frac{p_i}{\lambda_{\max}(S_i^T S_i)}$$

Proof :

$$A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = \sum_i p_i A^{\frac{1}{2}} S_i^T (S_i A S_i^T)^{-1} S_i A^{\frac{1}{2}}$$

Then we straightforwardly obtain that : $A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = A^{\frac{1}{2}}CD^2C^T A^{\frac{1}{2}}$.

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \lambda_{\min}(C^T AC)\lambda_{\min}(D^2)$$

¹will put before the intervention of the convergence rate in the convergence of our sequence to the optimal solution

$$\lambda_{\min}(D^2) = \min_i \frac{p_i}{\lambda_{\max}(S_i A S_i^T)} \geq \min_i \frac{p_i}{\lambda_{\max}(S_i^T S_i) \lambda_{\max}(A)}.$$

Therefore, $\lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}}) \geq \min_i \frac{p_i \lambda_{\min}(\mathbf{C}^T \mathbf{A} \mathbf{C})}{\lambda_{\max}(S_i^T S_i) \lambda_{\max}(A)} = \frac{\lambda_{\min}(A) \lambda_{\min}(\mathbf{C} \mathbf{C}^T)}{\lambda_{\max}(A)} \min_i \frac{p_i}{\lambda_{\max}(S_i^T S_i)}$

□

2. Block Coordinate Descent Method

Robert: Explain why is this called block coordinate descent? Write little intro.

2.1 Algorithm

2.2 Convergence rate

A is a $n \times n$ positive definite matrix representing our problem.

For any subset C of $\{1, \dots, n\}$ of length s , we denote by I_C the $s \times n$ matrix which rows are $\{e_i^T\}_{i \in C}$ up to a permutation, where $\{e_i\}_{i=1, \dots, n}$ is a canonical basis of \mathbb{R}^n .

Denote by $\{C_i\}_{i=1, \dots, r}$ the subsets of $\{1, \dots, n\}$ of size s : that implies that $r \stackrel{\text{def}}{=} \binom{n}{s}$.

Throughout the computations, we denote by $Z = AI_C^T(I_C AI_C^T)^{-1}I_C A$.

The convergence rate is defined by $\rho = 1 - \lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}})$.

Denote by $C = (I_{C_1}^T, \dots, I_{C_r}^T)$ which is of size $n \times rs$.

By **lemma 1.0.1**, we have that : $\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \frac{\lambda_{\min}(A)\lambda_{\min}(CC^T)}{\lambda_{\max}(A)} \min_i \frac{p_i}{\lambda_{\max}(I_{C_i}^T I_{C_i})}$

For any $i \in \{1, \dots, n\}$, for any x in \mathbb{R}^n , $\langle I_{C_i}^T I_{C_i} x | x \rangle = \|I_{C_i} x\|^2 \leq \|x\|^2$, then $\lambda_{\max}(I_{C_i}^T I_{C_i}) \leq 1$.

Therefore, $\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \frac{\lambda_{\min}(A)\lambda_{\min}(CC^T)}{\lambda_{\max}(A)} \min_i p_i$.

$CC^T = \sum_{i=1}^r I_{C_i}^T I_{C_i} = \binom{n-1}{s-1} I_n$ and then we obtain that corollary :

Corollary 2.2.1

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \binom{n-1}{s-1} \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)} \min_i p_i.$$

If we choose $\{p_i\}_{i=1}^r$ as the uniform probability of choosing s rows uniformly on $\{1, \dots, n\}$, i.e. for any i , $p_i = \frac{1}{\binom{n}{s}}$, then :

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \frac{s}{n} \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$$

Robert: This is already pretty interesting! It shows an improvement for using bigger batchsize! We should try to push this further, for instance, when $s = n$ we know the method converges in one step. It would be great if we have a convergence rate that shows this phenomena. In other words, when $s = n$ we have $\lambda_{\min}(A^{-1/2}E[Z]A^{-1/2}) = 1$! Also, please have a look at the paper “paving_kaczmarz.pdf” which I’ve just added to our repo.

3. Randomized orthonormal systems

This type of randomized sketch is well-suited for big data regression, thanks to the efficiency of matrix multiplication used in this method.

When the dimension of our matrix A is n , we denote by H_n the Hadamard matrix (well defined if the dimension of the problem n is a power of 2) defined recursively as :

$$H_{2^p} = \begin{pmatrix} H_{2^{p-1}} & -H_{2^{p-1}} \\ H_{2^{p-1}} & H_{2^{p-1}} \end{pmatrix} \text{ for } p = 1, 2, \dots \text{ and } H_1 = 1.$$

The Hadamard sketch consists of choosing a random sketch matrix $S \in \mathcal{M}_{s,n}$ where s is the sketch size of the problem, as follows :

we sample s *i.i.d.* rows of the form $s^T = e_j^T H_n D$ with probability $\frac{1}{n}$ for $j = 1, \dots, n$, where $(e_j)_j$ forms a canonical basis of \mathbb{R}^n , and $D = \text{diag}(\nu)$ is a diagonal matrix of *i.i.d.* Rademacher variables $\nu \in \{-1, 1\}^n$.

3.1 Algorithm

3.2 Convergence rate

Now we denote by $Z = AS^T(SAS^T)^{-1}SA$, where S is our Hadamard random matrix.

For any subset C of $\{1, \dots, n\}$ of length s , we denote by I_C the $s \times n$ matrix which rows are $\{e_i^T\}_{i \in C}$ up to a permutation, where $\{e_i\}_{i=1, \dots, n}$ is a canonical basis of \mathbb{R}^n .

By construction, $S = I_C H D$ where C is a uniform random subset of $\{1, \dots, n\}$ of size s , H is the Hadamard matrix ($HH^T = nI_n$) and $D = \text{diag}(\nu)$ is a diagonal matrix of *i.i.d.* Rademacher variables $\nu \in \{-1, 1\}^n$.

Recall that the convergence rate is $\rho = 1 - \lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}})$. From **lemma 1.0.1**, we have that :

Corollary 3.2.1

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}}) \geq \frac{s}{n} \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$$

Robert: In the statement of these corollary's and lemmas you should state the resulting convergence rate of the algorithm. The quantity $\lambda_{\min}(A^{-1/2} E[Z] A^{-1/2})$ is not an intuitive/interpretable quantity.

Proof :

Let's condition on the Rademacher diagonal matrix D .

Define by $\tilde{A}_D = \frac{H}{\sqrt{n}} D A D \frac{H^T}{\sqrt{n}}$. We obtain that :

$$\begin{aligned}
 A^{-\frac{1}{2}} E[Z|D] A^{-\frac{1}{2}} &= E[A^{\frac{1}{2}} S^T (S A S^T)^{-1} S A^{\frac{1}{2}} | D] \\
 &= \sum_i p_i A^{\frac{1}{2}} D H^T I_{C_i}^T (I_{C_i} H D A D H^T I_{C_i}^T)^{-1} I_{C_i} H D A^{\frac{1}{2}} \\
 &= \frac{1}{n} A^{\frac{1}{2}} D H^T E[I_C^T (I_C \tilde{A}_D I_C^T)^{-1} I_C] H D A^{\frac{1}{2}} \\
 &= D H^{-1} H D \frac{1}{n} A^{\frac{1}{2}} D H^T E[I_C^T (I_C \tilde{A}_D I_C^T)^{-1} I_C] H D A^{\frac{1}{2}} D H^T (H^T)^{-1} D \\
 &= D H^{-1} \tilde{A}_D^{\frac{1}{2}} E[I_C^T (I_C \tilde{A}_D I_C^T)^{-1} I_C] \tilde{A}_D^{\frac{1}{2}} n (H^T)^{-1} D \\
 &= D H^{-1} \tilde{A}_D^{\frac{1}{2}} E[I_C^T (I_C \tilde{A}_D I_C^T)^{-1} I_C] \tilde{A}_D^{\frac{1}{2}} H D
 \end{aligned}$$

Hence :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}}) = \lambda_{\min} \left(E_D \left[D H^{-1} \tilde{A}_D^{\frac{1}{2}} E[I_C^T (I_C \tilde{A}_D I_C^T)^{-1} I_C] \tilde{A}_D^{\frac{1}{2}} H D \right] \right).$$

Denote by $(D_i)_{i=1, \dots, 2^n}$ the 2^n possible values of the random matrix D .

We obtain that :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}}) = \lambda_{\min} \left(\sum_{i=1}^{2^n} \frac{1}{2^n} D_i H^{-1} \tilde{A}_{D_i}^{\frac{1}{2}} E[I_C^T (I_C \tilde{A}_{D_i} I_C^T)^{-1} I_C] \tilde{A}_{D_i}^{\frac{1}{2}} H D_i \right).$$

And thanks to the concavity of λ_{\min} , we obtain that :

$$\begin{aligned}
 \lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}}) &\geq \sum_{i=1}^{2^n} \frac{1}{2^n} \lambda_{\min} \left(D_i H^{-1} \tilde{A}_{D_i}^{\frac{1}{2}} E[I_C^T (I_C \tilde{A}_{D_i} I_C^T)^{-1} I_C] \tilde{A}_{D_i}^{\frac{1}{2}} H D_i \right) \\
 &= \sum_{i=1}^{2^n} \frac{1}{2^n} \lambda_{\min} \left(\tilde{A}_{D_i}^{\frac{1}{2}} E[I_C^T (I_C \tilde{A}_{D_i} I_C^T)^{-1} I_C] \tilde{A}_{D_i}^{\frac{1}{2}} \right)
 \end{aligned}$$

We then straightforwardly use the uniform case in **Corollary 2.2.1** to obtain that :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}}) \geq \sum_{i=1}^{2^n} \frac{1}{2^n} \frac{s}{n} \frac{\lambda_{\min}(\tilde{A}_{D_i})}{\lambda_{\max}(\tilde{A}_{D_i})}.$$

For all $i = 1, \dots, 2^n$, \tilde{A}_{D_i} is similar to A , and then finally :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}}) \geq \frac{s}{n} \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)} \quad \square$$

4. Count-min Sketches

4.1 Algorithm

4.2 Convergence rate

Denote by $(e_i)_{i=1,\dots,n}$ a canonical basis of \mathbb{R}^n and $(f_i)_{i=1,\dots,s}$ a canonical basis of \mathbb{R}^s .

Then we obtain that every count-min random matrix is of the form :

$$S = \sum_{i=1}^n \epsilon(i) f_{\pi(i)} e_i^T \in \mathcal{M}_{s,n}(\mathbb{R}), \text{ where } \epsilon : \{1, \dots, n\} \rightarrow \{1, -1\} \text{ and } \pi : \{1, \dots, n\} \rightarrow \{1, \dots, s\}.$$

We therefore can rewrite S as :

$$S = \begin{pmatrix} \epsilon(1)f_{\pi(1)}, \epsilon(2)f_{\pi(2)}, \dots, \epsilon(n)f_{\pi(n)} \end{pmatrix} \begin{pmatrix} e_1^T \\ \vdots \\ e_n^T \end{pmatrix} = \begin{pmatrix} f_{\pi(1)}, f_{\pi(2)}, \dots, f_{\pi(n)} \end{pmatrix} \text{diag}(\epsilon(1), \dots, \epsilon(n)).$$

For any $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, s\}$, define by f_π the $s \times n$ matrix $\begin{pmatrix} f_{\pi(1)}, f_{\pi(2)}, \dots, f_{\pi(n)} \end{pmatrix}$.

Let S be a random count-min sketch matrix.

$S = f_\pi D$ where π is a uniform random element of $\{1, \dots, s\}^{\{1,\dots,n\}}$ and $D = \text{diag}(\nu)$ is a diagonal matrix of *i.i.d.* Rademacher variables $\nu \in \{-1, 1\}^n$.

Denote again by $Z = AS^T(SAS^T)^{-1}SA$, where S is our count-min random matrix.

Recall that the convergence rate is $\rho = 1 - \lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}})$.

Denote $r \stackrel{\text{def}}{=} s^n$ and $\{\pi_1, \dots, \pi_r\}$ the elements of $\{1, \dots, s\}^{\{1,\dots,n\}}$ which is of size $r = s^n$.

Then, $\pi = \pi_k$ with probability $p_k \stackrel{\text{def}}{=} s^{-n}$.

Denote by $C = (f_{\pi_1}^T, \dots, f_{\pi_r}^T)$ which is a $n \times rs$ matrix.

Lemma 4.2.1

Let M be a $n \times n$ matrix of the form $aI_n + bJ_n$ where a and b are two real numbers. Then :
 $a - b$ and $a + (n - 1)b$ are the eigenvalues of M with $a - b$ of multiplicity at least $n - 1$.

Proof :

Let's observe that by subtracting $(a-b)I_n$ from M , we obtain : $M - (a-b)I_n =$

$$\begin{pmatrix} b & & & \\ & b & & \mathbf{b} \\ & & \ddots & \\ \mathbf{b} & & & b \\ & & & & b \end{pmatrix}$$

which is a matrix of range at most 1 (of range 0 if $b = 0$). Therefore $a - b$ is an eigenvalue of M of multiplicity at least $n - 1$.

Then, $\text{Tr}(M) - (n - 1)(a - b)$ is the remaining eigenvalue. And since $\text{Tr}(M) = na$, then $a + (n - 1)b$ is the remaining eigenvalue. \square

Corollary 4.2.2

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \frac{(s - 1) \lambda_{\min}(A)}{n s \lambda_{\max}(A)}$$

Proof :

Denote by $\tilde{A} = DAD$.

$$\begin{aligned} A^{-\frac{1}{2}}E[Z|D]A^{-\frac{1}{2}} &= E[A^{\frac{1}{2}}S^T(SAS^T)^{-1}SA^{\frac{1}{2}}|D] \\ &= \sum_i p_i A^{\frac{1}{2}}Df_{\pi_i}^T(f_{\pi_i}DADf_{\pi_i}^T)^{-1}f_{\pi_i}DA^{\frac{1}{2}} \\ &= A^{\frac{1}{2}}DE[f_{\pi}^T(f_{\pi}\tilde{A}_Df_{\pi}^T)^{-1}f_{\pi}]DA^{\frac{1}{2}} \\ &= D\tilde{A}_D^{\frac{1}{2}}E[f_{\pi}^T(f_{\pi}\tilde{A}_Df_{\pi}^T)^{-1}f_{\pi}]\tilde{A}_D^{\frac{1}{2}}D \end{aligned}$$

Then :

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) = \lambda_{\min}\left(E_D\left[D\tilde{A}_D^{\frac{1}{2}}E[f_{\pi}^T(f_{\pi}\tilde{A}_Df_{\pi}^T)^{-1}f_{\pi}]\tilde{A}_D^{\frac{1}{2}}D\right]\right).$$

Denote again by $(D_i)_{i=1,\dots,2^n}$ the 2^n possible values of the random matrix D . We obtain that :

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) = \lambda_{\min}\left(\sum_{i=1}^{2^n} \frac{1}{2^n} D_i \tilde{A}_{D_i}^{\frac{1}{2}} E[f_{\pi}^T(f_{\pi}\tilde{A}_{D_i}f_{\pi}^T)^{-1}f_{\pi}] \tilde{A}_{D_i}^{\frac{1}{2}} D_i\right).$$

And thanks to the concavity of λ_{\min} , we obtain that :

$$\begin{aligned} \lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) &\geq \sum_{i=1}^{2^n} \frac{1}{2^n} \lambda_{\min}\left(D_i \tilde{A}_{D_i}^{\frac{1}{2}} E[f_{\pi}^T(f_{\pi}\tilde{A}_{D_i}f_{\pi}^T)^{-1}f_{\pi}] \tilde{A}_{D_i}^{\frac{1}{2}} D_i\right) \\ &= \sum_{i=1}^{2^n} \frac{1}{2^n} \lambda_{\min}\left(\tilde{A}_{D_i}^{\frac{1}{2}} E[f_{\pi}^T(f_{\pi}\tilde{A}_{D_i}f_{\pi}^T)^{-1}f_{\pi}] \tilde{A}_{D_i}^{\frac{1}{2}}\right) \end{aligned}$$

Then by **lemma 1.0.1** :

$$\begin{aligned}\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) &\geq \sum_{i=1}^{2^n} \frac{1}{2^n} \frac{\lambda_{\min}(\tilde{A}_{D_i})\lambda_{\min}(\mathbf{C}\mathbf{C}^T)}{\lambda_{\max}(\tilde{A}_{D_i})} \min_k \frac{p_k}{\lambda_{\max}(f_{\pi_k}^T f_{\pi_k})} \\ &= \frac{\lambda_{\min}(A)\lambda_{\min}(\mathbf{C}\mathbf{C}^T)}{\lambda_{\max}(A)} \min_k \frac{p_k}{\lambda_{\max}(f_{\pi_k}^T f_{\pi_k})}\end{aligned}$$

Recall that $p_k = s^{-n}$ for any $k \in \{1, \dots, r\}$.

For any x in \mathbb{R}^n , for any $k \in \{1, \dots, r\}$,

$$\langle f_{\pi_k}^T f_{\pi_k} x | x \rangle = \|f_{\pi_k} x\|^2 = \left\| \sum_{i=1}^n x_i f_{\pi_k(i)} \right\|^2 \leq \left(\sum_{i=1}^n |x_i| \right)^2 \leq n \|x\|^2 \text{ and then } \lambda_{\max}(f_{\pi_k}^T f_{\pi_k}) \leq n.$$

$$\mathbf{C}\mathbf{C}^T = \sum_{k=1}^r f_{\pi_k}^T f_{\pi_k} = s^{n-1} \begin{pmatrix} s & & & \\ & s & & \mathbf{1} \\ & & \ddots & \\ \mathbf{1} & & & s & \\ & & & & s \end{pmatrix}, \text{ thanks to the facts that :}$$

$$\text{For all } i \neq j, \sum_{k=1}^r f_{\pi_k(i)}^T f_{\pi_k(j)} = r = s^n \text{ and } \sum_{k=1}^r f_{\pi_k(i)}^T f_{\pi_k(j)} = \sum_{k=1}^r 1_{\{\pi_k(i)=\pi_k(j)\}} = s \times s^{n-2} = s^{n-1}.$$

$$\text{Denote by } M = \frac{1}{s^{n-1}} \mathbf{C}\mathbf{C}^T.$$

By using the **lemma 4.2.1**, $(s-1)s^{n-1}$ and $s^n + (n-1)s^{n-1} = (n+s-1)s^{n-1}$ are the eigenvalues of $\mathbf{C}\mathbf{C}^T$.

Hence, $\lambda_{\min}(\mathbf{C}\mathbf{C}^T) = (s-1)s^{n-1}$.

Thereby we obtain that :

$$\lambda_{\min}(A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}}) \geq \frac{\lambda_{\min}(A)(s-1)s^{n-1}}{\lambda_{\max}(A)} \frac{s^{-n}}{n} = \frac{(s-1)\lambda_{\min}(A)}{n s \lambda_{\max}(A)} \quad \square$$

4.3 Sparse Shuffling (Spashu)

Let $\phi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ be a permutation, selected uniformly at random from all the $n!$ possible permutations. Let $s \in \mathbb{N}$ be an integer that divides n , that is, there exists $m \in \mathbb{N}$ such that $n = ms$. We define the Spashu (Sparse Shuffling) sketch $S \in \mathbb{R}^{n \times s}$ as a $s \times n$ when

$$S := \sum_{i=1}^s f_i \sum_{j=1+m(i-1)}^{mi} \epsilon(j) e_{\phi(j)}^\top.$$

Note that there are n non-zeros elements in S , with exactly m non-zero elements in each row of S .

Robert: To include subsampling, we will either have to re-define $\phi : C \rightarrow \{1, \dots, ms\}$ where $C \subset \{1, \dots, n\}$ and $|C| = ms$, or re-define the ϵ function to return $\{-1, 1, 0\}$.

We can also define a subsampled Spashu by considering $m \in \mathbb{N}$ as a free parameter such that $m \leq \lfloor \frac{n}{s} \rfloor$.

Notice that S can be rewritten as : $S = \sum_{j=1}^n \epsilon_j f_{\pi(j)} e_{\phi(j)}^T$, where π is the function $\left\{ \begin{array}{l} \{1, \dots, n\} \longrightarrow \{1, \dots, s\} \\ j \longmapsto \lceil \frac{j}{m} \rceil \end{array} \right.$

π verifies that for all $i \in \{1, \dots, s\}$, for all $j \in \{1 + m(i-1), \dots, mi\}$, $\pi(j) = i$.

For any permutation ϕ on $\{1, \dots, n\}$, denote by P_ϕ the $n \times n$ matrix $\begin{pmatrix} e_{\phi(1)}^T \\ \vdots \\ e_{\phi(n)}^T \end{pmatrix}$.

Denote by $\phi_1, \dots, \phi_{n!}$ the different permutations over $\{1, \dots, n\}$ and define $(p_k)_{k=1, \dots, n!}$ such that $p_k = \frac{1}{n!}$ for all k , which is the uniform probability distribution over all permutations. Then $\phi = \phi_k$ with probability $\frac{1}{n!}$.

Let ϵ be a uniform random vector of $\{-1, 1\}^n$ and ϕ a uniform random permutation of \mathfrak{S}_n .

Let S be a random shuffling sketch such that : $S = \sum_{j=1}^n \epsilon_j f_{\pi(j)} e_{\phi(j)}^T$.

Denote by $f_\pi = (f_{\pi(1)}, f_{\pi(2)}, \dots, f_{\pi(n)})$ and $D = \text{diag}(\epsilon(1), \dots, \epsilon(n))$.

We have that :

$$S = (\epsilon(1)f_{\pi(1)}, \epsilon(2)f_{\pi(2)}, \dots, \epsilon(n)f_{\pi(n)}) \begin{pmatrix} e_{\phi(1)}^T \\ \vdots \\ e_{\phi(n)}^T \end{pmatrix} = (f_{\pi(1)}, f_{\pi(2)}, \dots, f_{\pi(n)}) \text{diag}(\epsilon(1), \dots, \epsilon(n)) P_\phi.$$

Then : $S = f_\pi D P_\phi$.

Denote by $\mathbf{C}_D = ((P_{\phi_1}^T D f_\pi^T, \dots, P_{\phi_{n!}}^T D f_\pi^T)$ which is a $n \times n! n$ matrix.

Recall that $Z = A S^T (S A S^T)^{-1} S A$, where S is our sparse shuffling random matrix, and that the convergence rate is $\rho = 1 - \lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}})$.

Corollary 4.3.1

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}}) \geq \frac{s}{n} \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)} \left(1 - \sqrt{\frac{n}{s(n-1)}}\right)$$

Proof :

The **lemma1.0.1** gives us that :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z|D] A^{-\frac{1}{2}}) \geq \frac{\lambda_{\min}(A) \lambda_{\min}(\mathbf{C}_D \mathbf{C}_D^T)}{\lambda_{\max}(A)} \min_k \frac{p_k}{\lambda_{\max}(P_{\phi_k}^T D f_\pi^T f_\pi D P_{\phi_k})}.$$

For all $k = 1, \dots, n!$, $p_k = \frac{1}{n!}$ and P_{ϕ_k} is an orthogonal matrix (i.e. $P_{\phi_k} P_{\phi_k}^T = I_n$). Therefore one obtains that :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z|D] A^{-\frac{1}{2}}) \geq \frac{\lambda_{\min}(A) \lambda_{\min}(\mathbf{C}_D \mathbf{C}_D^T)}{n! \lambda_{\max}(A) \lambda_{\max}(f_\pi^T f_\pi)}.$$

For any positive integer k , denote by $J_k \in \mathcal{M}_k(\mathbb{R})$ the all-ones matrix of size k , i.e. $J_k(i, j) = 1$ for all $i, j = 1, \dots, k$.

$$\begin{aligned} \mathbf{C}_D \mathbf{C}_D^T &= \sum_{k=1}^n P_{\phi_k}^T D f_{\pi}^T f_{\pi} D P_{\phi_k} \\ &= (n-1)! \begin{pmatrix} \text{Tr}(f_{\pi}^T f_{\pi}) & & & & \\ & \text{Tr}(f_{\pi}^T f_{\pi}) & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \text{Tr}(f_{\pi}^T f_{\pi}) \end{pmatrix} \\ &\quad + \begin{pmatrix} & & & \frac{\text{Tr}(D f_{\pi}^T f_{\pi} D (J - I_n))}{n-2} & \\ & & & & \\ & & & & \\ & \frac{\text{Tr}(D f_{\pi}^T f_{\pi} D (J - I_n))}{n-2} & & & \\ & & & & \text{Tr}(f_{\pi}^T f_{\pi}) \end{pmatrix} \end{aligned}$$

By using the **lemma 4.2.1**, $\lambda_1 = (n-1)! \text{Tr}(f_{\pi}^T f_{\pi}) - \frac{(n-1)!}{n-2} \text{Tr}(D f_{\pi}^T f_{\pi} D (J - I_n))$ and $\lambda_2 = (n-1)! \text{Tr}(f_{\pi}^T f_{\pi}) + (n-1)! \frac{n-1}{n-2} \text{Tr}(D f_{\pi}^T f_{\pi} D (J - I_n))$ are the eigenvalues of $\mathbf{C}_D \mathbf{C}_D^T$. Hence, $\lambda_{\min}(\mathbf{C}_D \mathbf{C}_D^T) = (n-1)! \text{Tr}(f_{\pi}^T f_{\pi}) - \frac{(n-1)!}{n-2} \text{Tr}(D f_{\pi}^T f_{\pi} D (J - I_n))$.

Now denote by $\mathbf{1}_m = \underbrace{(1, \dots, 1)}_{m \text{ times } 1}$.

One observes that $f_{\pi} = (f_1 \mathbf{1}_m, f_2 \mathbf{1}_m, \dots, f_s \mathbf{1}_m)$.

Then :

$$f_{\pi}^T f_{\pi} = (1_m^T f_i^T f_j 1_m)_{i,j=1,\dots,s} = \begin{pmatrix} 1_m^T 1_m & & & & \\ & 1_m^T 1_m & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1_m^T 1_m \end{pmatrix} = \begin{pmatrix} J_m & & & & \\ & J_m & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & J_m \end{pmatrix}.$$

When we use the **lemma 4.2.1** on J_m , we obtain that 0 and m are the eigenvalues of J_m , with 0 of multiplicity $m-1$.

Then :

$$\lambda_{\max}(f_{\pi}^T f_{\pi}) = m \text{ and } \text{Tr}(f_{\pi}^T f_{\pi}) = s \text{Tr}(J_m) = s m = n.$$

A modifier à partir d'ici

Right now we have that :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z|D] A^{-\frac{1}{2}}) \geq \frac{\lambda_{\min}(A) (n! - (n-2)! \text{Tr}(D f_{\pi}^T f_{\pi} D (J - I_n)))}{n! m \lambda_{\max}(A)}.$$

By Cauchy-Schwarz inequality, $\text{Tr}(D f_{\pi}^T f_{\pi} D (J - I_n)) \leq \sqrt{\text{Tr}(D f_{\pi}^T f_{\pi} D^2 f_{\pi}^T f_{\pi} D)} \sqrt{\text{Tr}(J - I_n)^2}$.

Then : $\text{Tr}(D f_{\pi}^T f_{\pi} D (J - I_n)) \leq \sqrt{\text{Tr}(f_{\pi}^T f_{\pi} f_{\pi}^T f_{\pi})} \sqrt{n^2 - n} \leq \sqrt{s m^2} \sqrt{n^2 - n}$.

Therefore :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z|D] A^{-\frac{1}{2}}) \geq \frac{\lambda_{\min}(A) (n! - (n-2)! m \sqrt{s n (n-1)})}{n! m \lambda_{\max}(A)} = \frac{s \lambda_{\min}(A)}{n \lambda_{\max}(A)} \left(1 - \frac{m \sqrt{s n (n-1)}}{n(n-1)} \right).$$

Then :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z|D] A^{-\frac{1}{2}}) \geq \frac{s}{n} \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)} \left(1 - \frac{\sqrt{sn(n-1)}}{s(n-1)}\right) = \frac{s}{n} \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)} \left(1 - \sqrt{\frac{n}{s(n-1)}}\right).$$

We finally finish the proof thanks to the concavity of the function λ_{\min} :

$$\lambda_{\min}(A^{-\frac{1}{2}} E[Z] A^{-\frac{1}{2}}) \geq E_D \left[\lambda_{\min}(A^{-\frac{1}{2}} E[Z|D] A^{-\frac{1}{2}}) \right] \geq \frac{s}{n} \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)} \left(1 - \sqrt{\frac{n}{s(n-1)}}\right) \quad \square$$

5. *Conclusion*

References

- [1] ROBERT GOWER AND PETER RICHTARIK, Randomized iterative methods for linear systems, SIAM, (2015).