Contents

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 rate[1].

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

By defiition, $A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = \sum_i p_i A^{\frac{1}{2}} S_i^T (S_i A S_i^T)^{-1} S_i A^{\frac{1}{2}}$
for any $i \in \{1, \dots, n\}$, $A^{\frac{1}{2}} S_i^T (S_i A S_i^T)^{-1} S_i A^{\frac{1}{2}}$ 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 :

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

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

Lemma 1.0.1 $A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = (A^{\frac{1}{2}}\mathbf{C}D)(D\mathbf{C}^T A^{\frac{1}{2}})$ 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 :

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

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 : $A^{-\frac{1}{2}}E[Z]A^{-\frac{1}{2}} = A^{\frac{1}{2}}\mathbf{C}D^2\mathbf{C}^T A^{\frac{1}{2}}$.

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

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

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

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

Block Coordinate Descent Method

Algorithm

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 $\left\{e_i^T\right\}_{i \in C}$ 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 $r \stackrel{def}{=} \binom{n}{s}$.

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

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

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

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

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

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

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

Corollary 2.2.1

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

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, $p_i = \dfrac{1}{\binom{n}{s}}$, then :

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

Robert: This is already pretty interesting! It shows an improvement for using bigger bachsize! 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.

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 $D = diag(\nu)$ is a diagonal matrix of $i.i.d.$ Rademacher variables $\nu \in \{-1, 1\}^n$.

Algorithm

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 $\left\{e_i^T\right\}_{i \in C}$ 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 $D = diag(\nu)$ 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

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

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

Proof :

Let's condition on the Rademacher diagonal matrix D.

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

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

Hence :

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

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

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

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

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

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

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

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

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

Count-min Sketches

Algorithm

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

We therefore can rewrite S as :

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

For any $\pi : \{1, \dots, n\} \to \{1, \dots, s\}$, define by f_π the $s \times n$ matrix $\left(f_{\pi(1)}, f_{\pi(2)}, \dots, f_{\pi(n)}\right)$.

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 $D = diag(\nu)$ 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 $r \overset{def}{=} s^n$ 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 $p_k \overset{def}{=} s^{-n}$.
Denote by $\mathbf{C} = (f_{\pi_1}^T, \dots, f_{\pi_r}^T)$ which is a $n \times rs$ matrix.

Corollary 4.2.1

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

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 :

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

Then by **lemma** 1.0.1 :

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

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

For any x in $\mathbb{R}^n$, for any $k \in \{1,\dots,r\}$,
$\left\langle f_{\pi_k}^Tf_{\pi_k}x \,|\, x\right\rangle = \|f_{\pi_k}x\|^2 = \|\sum_{i=1}^n x_if_{\pi_k(i)}\|^2 \leqslant \left(\sum_{i=1}^n|x_i|\right)^2 \leqslant n\|x\|^2$ and then $\lambda_{max}(f_{\pi_k}^Tf_{\pi_k}) \leqslant n$.

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

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

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

By subtracting $(s-1)I_n$ from M, we recognize that $s-1$ is an eigenvalue of M with multiplicity $n-1$. Then the trace of M gives us that $n+s-1$ is the other eigenvalue of M.
Hence, $\lambda_{min}(\mathbf{C}\mathbf{C}^T) = (s-1)s^{n-1}$.

Thereby we obtain that :

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

Sparse Shuffling (Spashu)

Robert: I was calling this Radamacher sketch before, but in truth it is not the Radamacher sketch. So we need to give this a new name. How about Sparse Shuffling Sketch? Or a Spashu sketch for short :)

Let $\phi : \{1,\ldots,n\} \to \{1,\ldots,n\}$ 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.

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

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 $\begin{cases} \{1,\ldots,n\} \longrightarrow \{1,\ldots,s\} \\ j \longmapsto -\lfloor -\frac{j}{m} \rfloor \end{cases}$

Robert: I think it should be $\pi : j \to \lceil \frac{j}{m} \rceil$

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

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

Denote by $\phi_1,\ldots,\phi_{n!}$ the different permutations of $\mathfrak{S}_n$ and define $(p_k)_{k=1,\ldots,n!}$ such that $p_k = \frac{1}{n!}$ for all k.
Let's consider that uniform probability on $\mathfrak{S}_n$.
Then $\phi = \phi_k$ with probability $\dfrac{1}{n!}$.

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 $f_\pi = \left(f_{\pi(1)}, f_{\pi(2)}, \ldots, f_{\pi(n)}\right)$ and $D = \mathrm{diag}\,(\epsilon(1),\ldots,\epsilon(n))$.
We have that :

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

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

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

Recall that $Z = AS^T(SAS^T)^{-1}SA$, 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

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

Proof :

The **lemma**1.0.1 gives us that :
$$\lambda_{min}(A^{-\frac{1}{2}} E\left[Z|D\right] A^{-\frac{1}{2}}) \geqslant \frac{\lambda_{min}(A)\lambda_{min}(\mathbf{C}_D \mathbf{C}_D^T)}{\lambda_{max}(A)} \min_k \frac{p_k}{\lambda_{max}(P_{\phi_k}^T D f_\pi^T f_\pi D P_{\phi_k})}.$$
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 :

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

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 $i, j = 1, \dots, k$.

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

Denote by $\lambda_1 = (n-1)!\,\mathrm{Tr}(f_\pi^T f_\pi) - (n-2)!\,\mathrm{Tr}\left(D f_\pi^T f_\pi D (J - I_n)\right)$ and
$\lambda_2 = (n-1)!(n-1)\,\mathrm{Tr}(f_\pi^T f_\pi) + (n-2)!\,\mathrm{Tr}\left(D f_\pi^T f_\pi D (J - I_n)\right)$.
By subtracting $\lambda_1 I_n$ from $\mathbf{C}_D \mathbf{C}_D^T$, we straightforwardly observe that λ_1 is an eigenvalue of $\mathbf{C}_D \mathbf{C}_D^T$ of multiplicity $n-1$. And then taking the trace shows that λ_2 is the remaining eigenvalue.
Hence, $\lambda_{min}(\mathbf{C}_D \mathbf{C}_D^T) = (n-1)!\,\mathrm{Tr}(f_\pi^T f_\pi) - (n-2)!\,\mathrm{Tr}\left(D f_\pi^T f_\pi D (J - I_n)\right)$.

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

One observes that $f_\pi = \left(f_1 1_m, f_2 1_m, \dots, f_s 1_m\right)$.

Then :

$$f_\pi^T f_\pi = \left(1_m^T f_i^T f_j 1_m\right)_{i,j=1,\dots,s} = \begin{pmatrix} 1_m^T 1_m & & & \\ & 1_m^T 1_m & \mathbf{0} & \\ & & \ddots & \\ & \mathbf{0} & 1_m^T 1_m & \\ & & & 1_m^T 1_m \end{pmatrix} = \begin{pmatrix} J_m & & & \\ & J_m & \mathbf{0} & \\ & & \ddots & \\ & \mathbf{0} & J_m & \\ & & & J_m \end{pmatrix}.$$

Then :
$\lambda_{max}(f_\pi^T f_\pi) = m$ and $\mathrm{Tr}(f_\pi^T f_\pi) = n$.

Right now we have that :

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

By Cauchy-Schwarz inequality, $\mathrm{Tr}\left(D f_\pi^T f_\pi D(J - I_n)\right) \leqslant \sqrt{\mathrm{Tr}\left(D f_\pi^T f_\pi D^2 f_\pi^T f_\pi D\right)}\sqrt{\mathrm{Tr}\left(J - I_n\right)^2}$.
Then : $\mathrm{Tr}\left(D f_\pi^T f_\pi D(J - I_n)\right) \leqslant \sqrt{\mathrm{Tr}\left(f_\pi^T f_\pi f_\pi^T f_\pi\right)}\sqrt{n^2 - n} \leqslant \sqrt{sm^2}\sqrt{n^2 - n}$.
Therefore :

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

Then :

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

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

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

Conclusion

References

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