Prof. Dr. D. Kressner
H. Y. Lam

Instructions

- For every question, provide a detailed and concise justification for your answer.
- You are required to typeset a report and submit scripts that reproduce each plot, figure, or table included in your report.
- Your scripts should be written in MATLAB, Python, or Julia, and they must be able to reproduce the plots, figures, or tables without requiring additional input.
- Unless otherwise mentioned in the question, only results covered in the lecture/exercise may be used without proof. Please provide exact citations for the corresponding lecture slides/exercise.
- Deadline for project submission: 20 Dec at midnight.
- Submit the project by email to: hysan.lam@epfl.ch

Prof. Dr. D. Kressner
H. Y. Lam

1. Randomized Nyström Preconditioning with RPCholesky

▶ Setting

This project is based on the paper [2], with an extended version available at [3]. Read the introduction of [2] for the setting and motivation of this project.

▶ Tasks

1. Using the result from L6S76, show that the approximation $\hat{A}^{(k)}$ returned after k steps of RPCholesky satisfies

$$\mathbb{E}\|A - \hat{A}^{(k)}\|_2 \leq 3 \cdot \mathrm{sr}_p(A) \cdot \lambda_p,$$

for $k \geq (p-1)\big(1/2 + \log(\eta^{-1}/2)\big)$ and with $\mathrm{sr}_p(A)$ defined in [2].

2. By mimicking the proof of Theorem 5.1 in [2], derive a sensible upper bound on

$$\mathbb{E}\big[\kappa_2\big(P^{-1/2} A_\mu P^{-1/2}\big)\big],$$

where P is constructed as described in equations (1.3) of [2], with $\hat{A}_{\mathrm{nys}}$ replaced by $\hat{A}^{(k)}$ for a suitable value for k. Explain what this bound means in terms of the quality of the preconditioner.

3. The proof of Proposition 2.2 from [2] on the quality of the Nyström approximation (with Gaussian random sketches) uses a squared Chevet bound. Provide a *detailed* proof of this bound (see Section B.2 in [3]) in your own words. Include all missing details (such as verifying the conditions of Slepian's lemma).

4. Perform numerical experiments that illustrate the effectiveness of the preconditioners constructed by RPCholesky and the Nyström approximation (with Gaussian random sketches) when using preconditioned CG for solving $(A + \mu I)x = b$. For this purpose, generate A as a kernel matrix from sampling the Gaussian kernel with scale parameter $\sigma > 0$ on the interval $[0, 1]$. Experiment with different scale parameters σ and regularization parameters μ.

5. **Bonus:** Use the preconditioner constructed by RPCholesky to solve a kernel ridge regression problem for a real data set discussed in the literature, for example in [1] or [2].

▶ References

[1] A. Alaoui and M. W. Mahoney. Fast randomized kernel ridge regression with statistical guarantees. NIPS, Vol. 28, 2015, pp. 775–783.

[2] Zachary Frangella, Joel A. Tropp, and Madeleine Udell. Randomized Nyström preconditioning. SIAM Journal on Matrix Analysis and Applications, 2024. `https://doi.org/10.1137/21M1466244`

[3] Zachary Frangella, Joel A. Tropp, and Madeleine Udell. Randomized Nyström preconditioning. arXiv preprint arXiv:2110.02820 (2021). `https://arxiv.org/abs/2110.02820`

Prof. Dr. D. Kressner
H. Y. Lam

2. Randomized spectral density approximation

▶ Setting

Given a symmetric matrix $B \in \mathbb{R}^{n\times n}$, the distribution of its eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ can be represented by the spectral density

$$\phi(t) = n^{-1}\sum_{i=1}^{n} \delta(t-\lambda_i), \quad \text{for } t \in \mathbb{R},$$

where δ denotes the Dirac delta function.

The delta function makes the problem of spectral density approximation complicated. In many application it suffices to approximate a smoothed spectral density instead:

$$\phi(t) \approx \phi_\sigma(t) := n^{-1}\sum_{i=1}^{n} g_\sigma(t-\lambda_i),$$

where

$$g_\sigma(s) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{s^2}{2\sigma^2}}$$

for some fixed scalar parameter $\sigma > 0$.

For a review of spectral density approximation, see [1,2].

▶ Tasks

1. Prove that

$$\phi_\sigma(t) = \frac{1}{n}\mathrm{trace}\big(g_\sigma(tI_n - B)\big),$$

where the right-hand side denotes a matrix function g_σ evaluated at $tI_n - B$ as defined in L6S3.

Note that $g_\sigma(tI_n - B)$ depends on the parameter t and one is often interested in evaluating $\phi_\sigma(t)$ for many values of t. More generally, we now consider the trace estimation for a parameter-dependent matrix $A(t)$ that depends continuously on $t \in [a,b] \subset \mathbb{R}$. A straightforward idea to approximate $\mathrm{trace}(A(t))$ is to draw standard random Gaussian vector $\omega^{(t)}$ and sample $(\omega^{(t)})^\top A(t)\omega^{(t)}$ at every $t \in \mathbb{R}$. However, this is wasteful if $A(t)$ does not change rapidly. Instead, one could try to use a fixed standard random Gaussian vector ω, which does not depends on t, and estimate the trace

$$\mathrm{trace}(A(t)) \approx \mathrm{trace}_\ell(A(t)) = \frac{1}{\ell}\sum_{i=1}^{\ell} \omega_i^\top A(t)\omega_i, \quad \text{i.i.d. } \omega_i \sim N(0, I_n).$$

2. For a fixed value of $t \in [a,b]$,

   a.) Using the linearity properties of sub-exponential random variables, show that $\mathrm{trace}(A(t)) - \mathrm{trace}_\ell(A(t))$ is a mean-zero, sub-exponential random variable. Provide the corresponding sub-exponential parameters.

   b.) Instead of directly using the linearity properties of sub-exponential random variables. Show that $\mathrm{trace}(A(t)) - \mathrm{trace}_\ell(A(t))$ is a mean-zero, sub-exponential random variable with parameters

$$\big(\sqrt{2\|A(t)\|_F^2(1-1/\beta)^{-1}/\ell}, 2\|A(t)\|_2\beta/\ell\big),$$

   for any $\beta > 1$.

   Hint: Directly obtain expresson of $\mathbb{E}(e^{\lambda(\mathrm{trace}(A(t)) - \mathrm{trace}_\ell(A(t)))})$ and find the upper bound of the function.

3. For fixed $t \in [a,b]$ and integer $p \geq 1$, using the result of 2b), provide an upper bound for

$$\left(\mathbb{E}|\text{trace}(A(t)) - \text{trace}_\ell(A(t))|^p\right)^{\frac{1}{p}},$$

using properties of sub-exponential random variables.

4. Show that for any integer $p \geq 1$ and $\gamma \geq 1$,

$$\int_a^b |\text{trace}(A(t)) - \text{trace}_\ell(A(t))| \mathrm{d}t \leq C\left(\frac{\gamma p}{\ell}\right) \int_a^b \|A(t)\|_F \mathrm{d}t$$

holds with probability at least $1 - \gamma^{-p}$. The C denotes a constant.

Hint: You may use Minkowski's integral inequality, but you need to verify that it can be applied to this situation.

5. Implement the estimator for $g_\sigma(tI_n - B)$, where $B = Q\Lambda Q^\top \in \mathbb{R}^{500\times 500}$, $\Lambda_{ii} = 1/i$, and Q is an orthogonal matrix obtained by orthogonalizing a random Gaussian matrix. Here, your function should receive an input of M values of t, g_σ, and B. The function should then output an array containing the trace approximation of $g_\sigma(tI_n - B)$ at the given input values of t. Compare the error

$$\int_a^b |\text{trace}(g_\sigma(tI_n - B)) - \text{trace}_\ell(g_\sigma(tI_n - B))| \mathrm{d}t$$

vs ℓ. Here take a, b such that it contains the spectrum of B (you may use an existing implementation of numerical quadrature rule to approximate the integral).

► References

[1] Lin Lin Randomized estimation of spectral densities of large matrices made accurate. Numer. Math. 136.1 (2017)

[2] Lin Lin, Yousef Saad, and Chao Yang Approximating Spectral Densities of Large Matrices. SIAM review, 58(1), 34-65 (2016)

Prof. Dr. D. Kressner
H. Y. Lam

3. Estimating the 2-norm of the columns of a matrix.

▶ Setting

In this project, we will derive a random estimator to estimate the (squared) 2-norm of each column of a matrix $A \in \mathbb{R}^{m\times n}$, which is a useful quantity in numerical linear algebra e.g. [1]. Let $A = [a_1, \dots, a_n] \in \mathbb{R}^{m\times n}$, our goal is to estimate the vector

$$a = [\|a_1\|_2^2, \|a_2\|_2^2, \dots, \|a_n\|_2^2]^\top \in \mathbb{R}^n,$$

where A can only be accessed through matrix-vector products with A and A^T.

▶ Tasks

1. Let $B = A^\top A \in \mathbb{R}^{n\times n}$ and $\omega \in \mathbb{R}^n$ be a Rademacher (random± 1) vector, show that
$$a = \bar{b} \quad \text{where} \quad \bar{b} := \mathbb{E}[\omega \odot B\omega],$$
the $\odot$ denotes the Hadamard product.

Now we have a column (squared) 2-norm estimator by taking

$$a \approx \frac{1}{\ell}\sum_{i=1}^{\ell} \omega_i \odot B\omega_i := \bar{b}_\ell, \tag{1}$$

where ω_i are i.i.d. Rademacher vectors.

2. Define $e_i = (\omega_i \odot B\omega_i) - a$, show that
$$\mathbb{E}[\|\bar{b}_\ell - a\|_2^2] = \mathbb{E}\Big[\Big\|\frac{1}{\ell}\sum_{i=1}^{\ell} e_i\Big\|_2^2\Big] = \frac{1}{\ell}(\|B\|_F^2 - \|a\|_2^2).$$

Let $e = \sum_{i=1}^{\ell} e_i$,the idea we use to provide an error estimate is to bound the k-th moment of $\|e\|_2^2$ by using a scalar random variable.

3. By Jensen's inequaility, show that for a fixed integer $k > 0$,
$$\mathbb{E}[\|e\|_2^{2k}] \le \mathbb{E}[\|\hat{e} - \tilde{e}\|_2^{2k}] = \mathbb{E}\Big[\|\sum_{i=1}^{\ell}(\hat{e_i} - \tilde{e_i})\|_2^{2k}\Big], \tag{2}$$
here $\hat{e}$ and $\tilde{e}$ are i.i.d copies of e and $\hat{e_i}$ and $\tilde{e_i}$ are i.i.d copies of e_i.

Now, our aim is to provide an upper bound for the right-hand side of (2).

4. Denote a scalar random variable $W_i = r_i\|\hat{e_i} - \tilde{e_i}\|_2$, where r_i is another independent Rademacher random variable. Show that
$$\mathbb{E}\Big[\|\sum_{i=1}^{\ell}(\hat{e_i} - \tilde{e_i})\|_2^{2k}\Big] \le \mathbb{E}\Big[(\sum_{i=1}^{\ell} W_i)^{2k}\Big], \tag{3}$$
and conclude that
$$\mathbb{E}[e^{\lambda\|e\|_2^2}] \le \mathbb{E}[e^{\lambda(\sum_{i=1}^{\ell} W_i)^2}].$$

5. Show that
$$\mathbb{E}[e^{\lambda\|W_i\|_2^2}] \le \mathbb{E}[e^{4\lambda\|e_i\|_2^2}] \le [e^{c\lambda \mathrm{trace}(D)}],$$
for $D := (B - diag(a))^\top(B - diag(a))$ and for some constant c with $diag(a)$ denotes the diagonal matrix whose diagonal entries are the entries of a. Conclude that $\sum_{i=1}^{\ell} W_i$ is sub-Gaussian random variable with parameter $C\sqrt{\ell}\sqrt{\mathrm{trace}(D)}$, where $C > 0$ is some (absolute) constant.

   Hint: Relate $\|e_i\|_2^2$ to Huchinson'estimator applied to D and use the relation between sub-expontential random variable and sub-Gaussian random variable.

6. Using part 4 and 5, prove that for any $\delta \in (0, 1]$ and $\ell \geq 1$

$$\|a - \bar{b}_\ell\|_2 \leq c\sqrt{\frac{\log(2/\delta)}{\ell}}\|B - diag(a)\|_F,$$

holds with probability $1 - \delta$, where c is some (absolute) constant.

7. (Bonus) Implement the estimator (1), check the error vs ℓ, and investigate the tightness of probability bound in 6). With matrix $A = U\Sigma V^\top$ where $U, V \in \mathbb{R}^{1000\times 1000}$ independent randomly generated orthogonal matrices and $\Sigma_{ii} = 1/i$.

► References

[1] Stanislav Budzinskiy When big data actually are low-rank, or entry-wise approximation of certain function-generated matrices. arXiv preprint arXiv:2407.03250, 2024

Prof. Dr. D. Kressner
H. Y. Lam

4. Low-rank approximation of matrix square roots

▶ Setting

Given a symmetric positive semi-definite (SPSD) matrix $A \in \mathbb{R}^{n\times n}$, the (matrix) square root of A is defined through its spectral decomposition:

$$A = Q \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} Q^T \quad \Rightarrow \quad A^{1/2} = Q \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix} Q^T.$$

In the following, we will sort eigenvalues decreasingly: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Square roots can be used, for example, to sample from non-standard multivariate normal distribution. If $\omega \sim N(0, I_n)$ then

$$\tilde{\omega} = A^{1/2}\omega \sim N(0, A), \tag{1}$$

which allows us to sample from $N(0, A)$ by sampling from $N(0, I_n)$ and multiplying with $A^{1/2}$.

We consider three different Nyström-inspired approximations of $A^{1/2}$:

$$\begin{aligned} N_1(A) &= A^{1/2}\Omega(\Omega^T A^{1/2}\Omega)^{-1}\Omega^T A^{1/2}, \\ N_2(A) &= A\Omega(\Omega^T A^{3/2}\Omega)^{-1}\Omega^T A, \\ N_3(A) &= (A\Omega(\Omega^T A\Omega)^{-1}\Omega^T A)^{1/2}, \end{aligned}$$

where Ω is an $n \times (r + p)$ Gaussian random matrix with $r + p < n$.

▶ Tasks

1. Assume that A has rank $r \geq 1$ and $p = 0$. Show for each of the approximations $N_i(A)$, $i = 1, 2, 3$, that $N_i(A) = A$ holds almost surely.

2. Given an orthonormal basis $Q \in \mathbb{R}^{n\times k}$ and an SPSD matrix $B \in \mathbb{R}^{k\times k}$, show that $(QBQ^T)^{1/2} = QB^{1/2}Q^T$. Use this result to develop an algorithm of complexity $O(n^2(r + p))$ for computing $N_3(A)$ when A is an explicitly given, dense SPSD matrix.

   In the following, we will assume that $p \geq 2$ and rank$(A) \geq r + p$.

3. Which of the three approximations satisfy $N_i(A) \leq A^{1/2}$? Which of the three approximations satisfy $N_i(A)^2 \leq A$? (A brief justification of your answer is sufficient.)

4. Use a result from L6 to state an upper bound on $\mathbb{E}\,\mathrm{trace}(A^{1/2} - N_1(A))$ in terms of the best rank-r approximation error trace(Λ_2) with

$$\Lambda_2 = \mathrm{diag}(\lambda_{r+1}, \cdots, \lambda_n).$$

5. We now aim at deriving an error bound for $N_2(A)$. First, show that

$$\mathrm{trace}(A^{1/2} - N_2(A)) = \|(I - \Pi_{A^{3/4}\Omega})A^{1/4}\|_F^2,$$

   where Π_Y denotes the orthogonal projector onto the range of Y. Then, adapt the arguments from L6S39–S42 to show that

$$\mathrm{trace}(A^{1/2} - N_2(A)) \leq \|\Lambda_2^{1/4}\|_F^2 + \frac{\sqrt{\lambda_{r+1}}}{\sqrt{\lambda_r}}\|\Lambda_2^{1/4}\Omega_2\Omega_1^{\dagger}\|_F^2,$$

   for independent Gaussian random matrices Ω_1, Ω_2 of suitable size. Use this result to state an upper bound on $\mathbb{E}\,\mathrm{trace}(A^{1/2} - N_2(A))$ in terms of the best rank-r approximation error.

6. Apply Theorem 1.1 from [1] to derive an an upper bound on $\mathbb{E}\,\mathrm{trace}(A^{1/2} - N_3(A))$ in terms of the best rank-r approximation error.

7. Compare the bounds derived in Points 4–6. From these bounds, which approximation do you expect to deliver the lowest error? Verify the error numerically by implementing all three approximations and performing experiments with several SPSD matrices A of different eigenvalue decay. Which method do you observe to perform best?

8. **Bonus:** Consider the distributions $\rho = N(0, A)$ for SPD A and $\hat{\rho} = N(0, \hat{A})$ for $\hat{A} = A\Omega(\Omega^T A\Omega)^{-1}\Omega^T A$ (in turn, ρ can be cheaply and approximately sampled from $\hat{\rho}$ using $N_3(A)$). Then the following bound on the Wasserstein 2-distance between the two distributions holds:

$$W_2(\rho, \hat{\rho}) \leq \sqrt{\mathrm{trace}(A - \hat{A})}.$$

Do you find this result in the literature? If yes, provide the precise reference. If not, try to prove it yourself.

► References

[1] David Persson, Raphael A. Meyer, and Christopher Musco Algorithm-agnostic low-rank approximation of operator monotone matrix functions. arXiv preprint arXiv:2311.14023 (2023). https://arxiv.org/abs/2311.14023

5. Johnson-Lindenstrauss moment property

Prof. Dr. D. Kressner
H. Y. Lam

▶ Setting

In this project, we will explore an important technique used in random embedding called the Johnson-Lindenstrauss moment property. The Johnson-Lindenstrauss moment (JL-moment) property is defined as follow:

Definition. For $\delta, \epsilon > 0$ and a fixed positive integer p, a random $n \times \ell$ matrix Ω satisfies the (ϵ, δ, p)-JL moment property if

$$\left(\mathbb{E}\|\Omega^T x\|_2^2 - 1]^p\right)^{\frac{1}{p}} \leq \epsilon\delta^{\frac{1}{p}} \text{ and } \mathbb{E}\|\Omega^T x\|_2^2 = 1.$$

hold for any $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$.

▶ Tasks

1. Prove that if Ω satisfies the (ϵ, δ, p)-JL moment property then Ω satisfies the Johnson-Lindenstrauss property, i.e.

$$\mathbb{P}\{|\|\Omega^T x\|_2^2 - 1| > \epsilon\} \leq \delta.$$

2. Let $U \in \mathbb{R}^{n\times d}$ with orthonormal columns and Ω satisfies the $(\epsilon/2, \delta/9^d, p)$-JL moment property. Prove that

$$\mathbb{E}\|(\Omega^\top U)^\top(\Omega^\top U) - I\|_2^p \leq \epsilon^p \cdot \delta.$$

   Using above result, prove that for any matrices A and B with n rows and the sum of ranks is at most d, then

$$\mathbb{E}\|(\Omega^\top A)^\top(\Omega^\top B) - A^\top B\|_2^p \leq \epsilon^p \|A\|_2^p \|B\|_2^p \cdot \delta.$$

   Hint: Use lecture 5 to show that there exists $X \subset \mathbb{R}^d, |X| \leq 9^d$ such that

$$\mathbb{E}[\|(\Omega^\top U)^T(\Omega^\top U) - I\|_2^p] \leq 2^p \cdot \mathbb{E}\sup_{x\in X} |\|\Omega^\top x\|_2^2 - 1|^p \leq 2^p \cdot \sum_{x\in X} \mathbb{E}|\|\Omega^\top x\|_2^2 - 1|^p.$$

3. Let $\Omega = \frac{1}{\sqrt{\ell}}\Psi \in \mathbb{R}^{n\times\ell}$, where Ψ is a standard Gaussian random matrix. By the sub-exponential property provide a lower bound on ℓ such that Ω satisfies the (ϵ, δ, p)-JL moment property.

From 2.), we see that if Ω satisfies the JL moment property, such sketches can be used to approximate matrix multiplication. However, 2.) only applies to low-rank matrices A and B. It is difficult to generalize this to the general case; therefore, we will simply state the theorem here, and you may use it without proof.

Theorem. Given an integer $d \geq 1$ and real numbers $\epsilon \in (0, 1]$, $\delta \in (0, 1/2)$, let $\Omega \in \mathbb{R}^{n\times d}$ be a random matrix satisfying the $(\epsilon/2, \delta/9^{2d}, p)$-JL moment property for some $p \geq 2$. Then, for any matrices A and B with n rows, the following bound holds:

$$\mathbb{P}\left\{ \left\|(\Omega^T A)^T(\Omega^T B) - A^T B\right\|_2 > \varepsilon\sqrt{\left(\|A\|_2 + \frac{\|A\|_F}{d}\right)\left(\|B\|_2 + \frac{\|B\|_F}{d}\right)} \right\} < \delta \tag{1}$$

4. Assume we use a Gaussian random matrix as definied in part 3.). Perform a numerical experiment for part 3.) to explore the tightness of the bound you provided. Also, study the tightness of the above theorem by exploring the parameter d while keeping all other parameters fixed.

5. In the rest of the project, we study a structured sketch $\Omega = \frac{1}{\sqrt{\ell}}[\omega_1, \dots \omega_\ell]$, with $\omega_i = \hat{r}_i \otimes \tilde{r}_i$ where $\tilde{r}_1, \dots, \tilde{r}_\ell, \hat{r}_1, \dots, \hat{r}_\ell \in \mathbb{R}^{\sqrt{n}}$ are i.i.d. Rademacher random vectors, and $\otimes$ denotes the Kronecker product. Using the above theorem and the fact that (you do not need to prove):

   If $\delta \le e^{-8}$, then Ω satisfies the (ε, δ, p)-JL moment property with $p = \lceil \frac{1}{2} \log(\frac{1}{\delta}) \rceil$, provided that

   $$d \ge C(\log(1/\delta)\varepsilon^{-2} + \log(1/\delta)\varepsilon^{-1}), \tag{2}$$

   for some constant $C > 0$.

   Prove that Ω has (k, ϵ, δ)-OSE property holds provided that

   $$\ell \ge C' \cdot (k^{3/2}\epsilon^{-2} + k\log(1/\delta)\epsilon^{-2} + k^{1/2}\log^2(1/\delta)\epsilon^{-1}), \tag{3}$$

   where $C' > 0$ is some constant.

6. (Bonus) Another way to obtain OSE is by obtaining the JL property via (2) and using the union bound. Compare the bound obtained by this approach with (3).

6. Sketching twice

Prof. Dr. D. Kressner
H. Y. Lam

▶ Setting

The goal of this project is to work out the idea mentioned on L5S18 to combine two sketching operations. More specifically, we consider two OSEs: (1) A random matrix $S_1 \in \mathbb{R}^{d_1 \times d}$ that is an (k, ϵ, δ)-OSE for k-dimensional subspaces of $\mathbb{R}^d$; and (2) a random matrix $S_2 \in \mathbb{R}^{d_2 \times d_1}$ that is an (k, ϵ, δ)-OSE for k-dimensional subspaces of $\mathbb{R}^{d_1}$. It is assumed that S_1, S_2 are independent and $0 < \delta < 1/2$.

▶ Tasks

1. Given a fixed matrix $V \in \mathbb{R}^{d_1 \times k}$ of rank k, show that

$$\sigma_{\max}(S_2V)^2 \leq (1+\epsilon)\|V\|_2^2, \quad \sigma_{\min}(S_2V)^2 \geq (1-\epsilon)\sigma_{\min}(V)^2$$

   hold with probability at least $1 - \delta$.

2. Using Point 1, show that S_2S_1 is an $(k, (2+1/3)\epsilon, 2\delta)$-OSE for k-dimensional subspaces of $\mathbb{R}^d$ if $0 < \epsilon \leq 1/3$.

3. Explain why the assumption on the independence of S_1, S_2 is needed in your proof of Point 2. Give an example of (k, ϵ, δ)-OSEs S_1, S_2 such that S_2S_1 is *not* an $(k, \tilde{\epsilon}, 2\delta)$-OSE for any $\tilde{\epsilon}$. Explain in detail why your construction is not an OSE.

4. Discuss the complexity of sketch-and-solve (L5S4) for computing an approximate solution of $\tilde{x}$ solving the linear least-squares matrix $\min \|Ax - b\|_2$ with a dense matrix $A \in \mathbb{R}^{d \times k}$, where $d > k$, for fixed δ when (a) using a Gaussian random matrix for sketching, (b) using a sparse sign matrix for sketching, and (c) using S_2S_1, where S_1 is a sparse sign matrix and S_2 is a Gaussian random matrix. Assume that the sketching matrices are chosen such that $\|A\tilde{x} - b\|_2^2 \leq 3 \min \|Ax - b\|_2^2$ is satisfied (using the estimates from L5S18).

5. Do the complexities derived in Point 3 (a), (b), (c), change when using CG preconditioned with the sketched matrix (instead of sketch-and-solve), as described in Exercise 6, Problem 2? Implement this preconditioned CG method with the sketching matrix S_2S_1 from Point 3 (choosing reasonably small constants when fixing the sketching sizes) and test it with the data from Exercise 6, Problem 2c). Note: You may use an existing implementation of the preconditioned CG method, but you need to implement the sparse sign sketch by yourself such that it it attains its (minimal) complexity.