Exercise 2 – Randomized matrix computations, Fall'24

Prof. D. Kressner
H. Lam

1 ▶ The moment of a Gaussian random variable

Let $X \sim N(0, \sigma^2)$,

1. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function with a bounded, continuous derivative. Prove the Gaussian integration by parts formula:

$$\mathbb{E}[Xf(X)] = \sigma^2 \cdot \mathbb{E}[f'(X)] \tag{1}$$

   Hint: use ordinary integration by parts.

2. For $p = 2, 4, 6, \dots$, prove that

$$(\mathbb{E}|X|^p)^{1/p} \le C \cdot \sigma\sqrt{p}, \tag{2}$$

   where C is some constant.

   Hint: consider applying 1) iteratively and use the Stirling's approximation.

3. By the monotonicity of L_p norm, prove that (2) also holds for $p = 1, 3, 5, 7, \dots$ (with a different constant C).

2 ▶ Frobenius norm of the pseudo-inverse of a Gaussian random matrix

In many randomized algorithms for matrix computation, we need to obtain bounds for $\|\Omega^\dagger\|_F$, where $\Omega \in \mathbb{R}^{m\times n}$, $n \ge m+4$, is a standard Gaussian random matrix. This exercise shows the effectiveness for combining the moment and the Markov's inequality to obtain a concentration probability bound.

1. Explain why $\|\Omega^\dagger\|_F^2 = \text{trace}[(\Omega\Omega^T)^{-1}]$ holds almost surely.

2. It is well-known that the entry $[(\Omega\Omega^T)^{-1}]_{ii}$ is distributed according to the inverse chi-square distribution i.e.

$$[(\Omega\Omega^T)^{-1}]_{ii} = X_i^{-1}, \quad X_i \sim \chi^2_{n-m+1}$$

   and the expectation is

$$\mathbb{E}(X_i^{-1}) = \frac{1}{n-m-1}.$$

   Using these two facts, calculate the expectation of $\|\Omega^\dagger\|_F^2$.

3. If we fix $q = \frac{n-m}{2}$, we also have the following moment bound on the q-th moment

$$(\mathbb{E}[|X_i^{-1}|^q])^{1/q} < \frac{3}{n-m+1}.$$

   Using this q-th moment bound, provide an upper bound for $(\mathbb{E}[\|\Omega^\dagger\|_F^{2q}])^{1/q}$.

4. Apply the Markov's inequality to result from 3) to obtain concentration probability bound for $\|\Omega^\dagger\|_F$. Compare your upper bound with Proposition 10.4 in the paper by Halko, Martinsson and Tropp[1].

[1] Halko/Martinsson/Tropp 2011, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*

3 ► Starting vector of the Power method

In the lecture, we have discussed the choice of the starting vector in the Power method. Assume $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix. Without loss of generality, assume $e_1 = (1, 0, \dots, 0)^T \in \mathbb{R}^n$ is the eigenvector corresponding to the largest eigenvalue. Let v_0 be the starting vector in the Power method . The quantity of interest is $|e_1^T v_0|$, and we want it to not be too small.

1. Pick $v_0 = X/\|X\|_2$ where $X \sim N(0, I)$, using the result from Lecture 1 slide 29, prove that
$$\mathbb{P}\left(|e_1^T v_0| \le \epsilon\right) \le \sqrt{\frac{2n}{\pi}}\epsilon$$
for $\epsilon > 0$.
2. Fix ϵ and then varies $n = 5, \dots, 50$ estimate the failure probability in Matlab/ Python by computing 10000 samples. Plot the failure probability vs n.
3. Fix $n = 20$ and varies the values of $\epsilon \in [0.01, 0.1]$, estimate the failure probability in Matlab/ Python by computing 10000 samples. Plot the failure probability vs ϵ.
4. Is the dependency of n and ϵ in the probability bound tight?

4 ► Norm of a Gaussian random vector

Let $A \in \mathbb{R}^{n \times m}$ and $X \sim N(0, I)$. Prove that for $\epsilon > 1$,
$$\Pr(\|A\|_2 \le \epsilon \|AX\|_2) \ge 1 - \sqrt{\frac{2}{\pi}}\frac{1}{\epsilon}.$$