# Hutch++

*Prof. Dr. D. Kressner*
*D. Persson*

## ▶ Setting

Estimating the trace of a large symmetric matrix $A$ is a ubiquitous task in numerical linear algebra, and arises in numerous applications, such as triangle counting in graphs, statistical learning, inverse problems, lattice quantum chromodynamics, and norm estimation. The difficulty in this problem arises because we do not have explicit access to the entries of $A$, but we only have implicit access to $A$ via matrix-vector products $Ax$. By using only matrix-vector products with $A$, the task is to find an estimate $t$ of $\text{tr}(A)$ such that $|t - \text{tr}(A)|$ is small.

A common method to estimate $\text{tr}(A)$ is using a Monte-Carlo estimator. If $\omega_1, \dots, \omega_m \sim N(0, I_n)$ are independent, then

$$\text{tr}_m(A) := \frac{1}{m} \sum_{i=1}^{m} \omega_i^T A \omega_i$$

is an unbiased estimator of $\text{tr}(A)$. Furthermore, one can show that

$$\sqrt{\text{Var}(\text{tr}_m(A))} = \sqrt{\frac{2}{m}} \|A\|_F$$

Hence, for a symmetric positive semi-definite matrix one has to let $m = \mathcal{O}(\frac{1}{\varepsilon^2})$ in order to guarantee

$$\sqrt{\text{Var}(\text{tr}_m(A))} \leq \varepsilon \text{tr}(A).$$

## ▶ Tasks

1. Implement the Monte-Carlo estimator on

   $$A = Q \Lambda Q^T \in \mathbb{R}^{1000 \times 1000},$$

   where

   $$\Lambda = \text{diag}(1, 2^{-c}, 3^{-c}, \dots, 1000^{-c}),$$

   for $c = 0.1, 1, 5$, and $Q$ is an arbitrary orthogonal matrix.

   Use a log-log plot and for each $m$ show the average relative mean-square error

   $$\frac{\sqrt{\mathbb{E}[|\text{tr}_m(A) - \text{tr}(A)|^2]}}{\text{tr}(A)},$$

   where $\mathbb{E}[|\text{tr}_m(A) - \text{tr}(A)|^2]$ is estimated with 100 repetitions of the experiment, on the y-axis. Let Let $m = 9, 18, \dots, 495, 498$. Confirm the $\mathcal{O}(\frac{1}{\sqrt{m}})$ Monte-Carlo convergence rate.

   Explain why the relative mean-square error is much smaller for $c = 0.1$ compared to $c = 5$.

   *Hint:* When is $\|A\|_F \approx \text{tr}(A)$ and when is $\|A\|_F \ll \text{tr}(A)$? Explain why.

2. Read and understand the proof of Theorem 1 and Theorem 10 in [1]. Assume that you can obtain a low-rank approximation $\widehat{A}$ of $A$ in $\mathcal{O}(k)$ matrix-vector products that satisfies

   $$\mathbb{E}\|A - \widehat{A}\|_F^2 \leq \frac{C_1}{k} \text{tr}(A)^2,$$

   for some universal constant $C_1$. Show that you can construct an algorithm that outputs an estimate $t$ of $\text{tr}(A)$ such that

   $$\mathbb{E}[t] = \text{trace}(A), \quad \sqrt{\text{Var}(t)} \leq \varepsilon \text{tr}(A)$$

   using only $\mathcal{O}(\frac{1}{\varepsilon})$ matrix-vector products with $A$.

3. Implement [1, Algorithm 1] but with random Gaussian vectors instead of Rademacher vectors. Test it on the matrices from exercise 1.

   Use a log-log plot and for each $m$ show the average relative error

$$\frac{\sqrt{\mathbb{E}[|t - \mathrm{tr}(A)|^2]}}{\mathrm{tr}(A)},$$

   where $\mathbb{E}[|t - \mathrm{tr}(A)|^2$ is estimated using 100 repetitions of the experiment, on the y-axis. Let $m = 9, 18, \ldots, 495, 498$.

4. From your plots you will see that in some cases the error converges faster or slower than $\mathcal{O}(\frac{1}{m})$. Explain why this does not contradict [1, Theorem 1].

# ► References

[1] Meyer, R., Musco, C., Musco, C., Woodruff, D.P. Hutch++: Optimal Stochastic Trace Estimation arXiv preprint arXiv:2010.09649 (2021). https://arxiv.org/pdf/2010.09649