Trace Estimation by Sampling

The methods outlined in this document follow sections 4.1 and 4.2 in Martinsson & Tropp (2021)[1]. Proofs and algorithms for these methods follow.

The technical idea behind randomized methods for trace estimation is to construct an unbiased estimator for the trace and then average independent copies to reduce the variance of the estimate. This algorithm falls under the umbrella of *Monte Carlo methods* or MC.

The first problem will be to estimate the trace of a nonzero positive semi-definite (psd) matrix $\mathbf{A} \in \mathbb{H}_n$, where $\mathbb{H}_n$ is the space of self-adjoint *nxn* matrices, namely square matrices that satisfy $\mathbf{A} = \mathbf{A}'$. The goal is to produce an approximation of *trace*(**A**) and some measure of the quality of this approximation. We will extract data from the input matrix by computing the product $\mathbf{Y} = \mathbf{A\Omega}$, where $\mathbf{\Omega} \in \mathbb{R}^{nxk}$ is a random test matrix. Thereafter all operations will only involve the sample matrix **Y** and the test matrix **Ω**, reducing the number of operations and subsequently the runtime of the algorithm as compared to computing *n* matrix-vector products $\mathbf{u} \mapsto \mathbf{Au}$ to achieve the same result.

Randomized trace estimation is based on the insight that it is easy to construct a random variable whose expectation equals the trace of the input matrix. Consider a random test vector $\omega \in \mathbb{R}^n$ that is isotropic, namely: $\mathbb{E}[\omega\omega'] = \mathbf{I}$.

Proposition 1: $X = \omega'(\mathbf{A}\omega)$ satisfies $\mathbb{E}[X] = trace(\mathbf{A})$.

Proof: The isotropic property of ω can be expressed as: $\mathbb{E}[\omega_i\omega_j'] = 1$. Then:

$$\mathbb{E}[X] = \mathbb{E}[\omega'(\mathbf{A}\omega)] = \mathbb{E}[\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_j' a_{ij}\omega_i] \qquad = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[\omega_j' a_{ij}\omega_i]$$ a_{ij} can be factored out by linearity:

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\mathbb{E}[\omega_j'\omega_i] \qquad = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\mathbb{E}[\omega_i\omega_j']$$ for *i*=*j*, this becomes:

$$= \sum_{i,j=1}^{n} a_{ij}\mathbb{E}[\omega_i\omega_j'] = \sum_{i,j=1}^{n} a_{ij}$$ by the isotropic property;

$$= trace(\mathbf{A})$$

Therefore the random variable *X* is an unbiased estimator of the trace.

With only one sample of *X*, the variance of the estimator will be large and therefore unreliable. Martinsson and Tropp present the following way to reduce the variance, by averaging *k* independent copies of *X* for $k \in \mathbb{N}$:

$$\overline{X}_k \overset{\text{def}}{=} \frac{1}{k}\sum_{i=1}^{k} X_i \text{ where } X_i \sim X \text{are iid} \tag{1}$$

[1] https://arxiv.org/abs/2002.01387v1, last accessed 4/30/21

Proposition 2: $\overline{X}_k$ is an unbiased estimator of *trace*($\mathbf{A}$) with variance $\frac{1}{k}\mathrm{Var}(X)$.

Proof: $\mathbb{E}[\overline{X}_k] = \mathbb{E}[\frac{1}{k}\sum_{i=1}^{k} X_i] = \frac{1}{k}\mathbb{E}[\sum_{i=1}^{k} X_i]$ — $\frac{1}{k}$ can be factored out by linearity;

$= \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}[X_i]$ — because the X_i are iid;

$= \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}[X]$ — because $X_i \sim X$;

$= \frac{1}{k} \cdot k\,\mathbb{E}[X] = \mathbb{E}[X]$

$= trace(\mathbf{A})$

$\mathrm{Var}(\overline{X}_k) = \mathrm{Var}(\frac{1}{k}\sum_{i=1}^{k} X_i) = (\frac{1}{k})^2\mathrm{Var}(\sum_{i=1}^{k} X_i)$

$= (\frac{1}{k})^2\sum_{i=1}^{k}\mathrm{Var}(X_i)$ — because the X_i are iid;

$= (\frac{1}{k})^2\sum_{i=1}^{k}\mathrm{Var}(X)$ — because $X_i \sim X$;

$= (\frac{1}{k})^2 \cdot k\,\mathrm{Var}(X) = \frac{1}{k}\mathrm{Var}(X)$

Thus the variance of $\overline{X}_k$ is less than that of X, but we still have an unbiased estimator for the trace. An algorithm to compute this estimation follows, and is demonstrated in the accompanying code for this project.[2]

Algorithm 1: Trace estimation by random sampling

Input: psd matrix $\mathbf{A} \in \mathbb{H}_n$, number k of samples to take

Output: Trace estimate $\overline{X}_k$ and sample variance S_k

1. for $i = 1, \ldots, k$ repeat the following:
2. Draw isotropic test vector $\omega_i \in \mathbb{F}^n$
3. Compute $X_i = \omega_i'(\mathbf{A}\omega_i)$
4. Compute the trace estimator: $\overline{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i$
5. Compute the sample variance: $S_k = \frac{1}{k-1}\sum_{i=1}^{k}(X_i - \overline{X}_k)^2$

[2] https://github.com/ghostpress/comp-stats-sims/tree/final-project/final-project/trace-estim

To compute $\overline{X}_k$, the algorithm simulates k independent copies of the random vector $\omega \in \mathbb{R}^n$ and performs k matrix-vector products with **A**, each of which takes n^2 steps, plus $O(kn)$ additional arithmetic. Thus the total runtime of the algorithm is $O(kn^2 + 2kn)$.

Regardless of the distribution of the isotropic test vector ω, Chebyshev's inequality demonstrates a probability bound for the trace estimator:

$$P\{|\overline{X}_k - trace(\mathbf{A})| \geq t\} \leq \frac{1}{kt^2}\mathrm{Var}(X) \quad \text{for } t \geq 0 \tag{2}$$

Furthermore, the strong law of large numbers says that:

$$\overline{X}_k \rightarrow trace(\mathbf{A}) \ \text{ almost surely as } k \rightarrow \infty \tag{3}$$

Combining Proposition 1 and (2) and (3) means that for a large sample, the performance of the trace estimator $\overline{X}_k$ depends only on the distribution of the test vector ω through the variance of the resulting sample X. The authors note, however, that the estimator cannot overcome the Central Limit Theorem, and its accuracy will always be limited by fluctuations on the scale of $\sqrt{\mathrm{Var}(X)}$.

Up to this point, we have constrained the method by assuming that **A** is psd. The authors note that trace estimation in this way can be extended to general square matrices, but with the drawback that the variance of the estimator may no longer be comparable with the trace.