Trace Estimation I
– Sampling –
Lecture 5

F. M. Faulstich

01/23/2024

Trace of a matrix

- Given $\mathbf{A} \in \mathbb{F}^{n \times n}$, we want to compute

$$\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{n} (\mathbf{A})_{ii}$$

- Not a problem if we have inexpensive access to matrix elements
- But what if we do not have that access?
  $\rightarrow$ We might only have $\mathbf{u} \mapsto \mathbf{A}\mathbf{u}$ implicitly

Idea:

Construct an unbiased estimator for the trace and then average independent copies to reduce the variance of the estimator.

Trace of a matrix

- Given $\mathbf{A} \in \mathbb{F}^{n \times n}$, we want to compute

$$\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{n} (\mathbf{A})_{ii}$$

- Not a problem if we have inexpensive access to matrix elements
- But what if we do not have that access?
  $\rightarrow$ We might only have $\mathbf{u} \mapsto \mathbf{Au}$ implicitly

Idea:

Construct an unbiased estimator for the trace and then average independent copies to reduce the variance of the estimator.
(... Monte-Carlo)

Girard–Hutchinson Estimator

We assume $\mathbf{A} \in \mathbb{H}_n$, $0 \preccurlyeq \mathbf{A}$, and $\operatorname{Tr}(\mathbf{A}) \neq 0$.

- Consider a random test vector $\boldsymbol{\omega} \in \mathbb{F}^n$ with $\mathbb{E}(\boldsymbol{\omega}\boldsymbol{\omega}^*) = \mathbf{I}$ we say $\boldsymbol{\omega}$ is isotropic.
- Then $X = \boldsymbol{\omega}^*(\mathbf{A}\boldsymbol{\omega})$ satisfies

$$\mathbb{E}(X) = \operatorname{Tr}(\mathbf{A})$$

$\Rightarrow X$ is an unbiased estimator of the trace.

Girard–Hutchinson Estimator

We assume $\mathbf{A} \in \mathbb{H}_n$, $0 \preccurlyeq \mathbf{A}$, and $\mathrm{Tr}(\mathbf{A}) \neq 0$.

- Consider a random test vector $\boldsymbol{\omega} \in \mathbb{F}^n$ with $\mathbb{E}(\boldsymbol{\omega}\boldsymbol{\omega}^*) = \mathbf{I}$ we say $\boldsymbol{\omega}$ is isotropic.
- Then $X = \boldsymbol{\omega}^*(\mathbf{A}\boldsymbol{\omega})$ satisfies

$$\mathbb{E}(X) = \mathrm{Tr}(\mathbf{A})$$

  $\Rightarrow$ X is an unbiased estimator of the trace.
- [Girard–Hutchinson Estimator] Reduce the variance by taking k copies

$$\bar{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i$$

  where $X_i \sim X$ are i.i.d.
  $\rightarrow$ By linearity $\mathbb{E}(\bar{X}_k) = \mathrm{Tr}(\mathbf{A})$
  $\rightarrow$ But the variance decreases: $\mathbb{V}(\bar{X}_k) = \frac{1}{k}\mathbb{V}(X)$

Algorithm I (naïve Monte-Carlo Estimator)

- Input: $\mathbf{A} \in \mathbb{H}_n$, $k \in \mathbb{N}$
- for i=1:k

  Draw isotropic $\boldsymbol{\omega}_i \in \mathbb{F}^n$

  Compute $X_i = \boldsymbol{\omega}_i^* \mathbf{A} \boldsymbol{\omega}_i$
- Compute trace estimator $\bar{X}_k = k^{-1} \sum_{i=1}^{k} X_i$
- Compute sample variance $S_k = (k-1)^{-1} \sum_{i=1}^{k} (X_i - \bar{X}_k)^2$

Cost and Variance bounds

Cost:

- Simulate k independent copies $\boldsymbol{\omega}_i$
- Perform k matrix vector products
- Perform $\mathcal{O}(kn)$ additional operations

Note that with n vector multiplications we can compute $\mathrm{Tr}(\mathbf{A})$ exactly!

- [Girard 1989] Consider $\boldsymbol{\omega} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ then

$$\mathbb{V}(\bar{X}_k) = \frac{2}{k} \sum_{i,j=1}^{n} |\mathbf{A}_{ij}|^2 = \frac{2}{k} \|\mathbf{A}\|_F^2 \le \frac{2}{k} \|\mathbf{A}\| \mathrm{Tr}(\mathbf{A})$$

- [Hutchinson 1990] Consider $\boldsymbol{\omega} \sim \mathcal{U}\{\pm 1\}^n$ (Rademacher r.v.) then

$$\mathbb{V}(\bar{X}_k) = \frac{4}{k} \sum_{1 \le i < j \le n} |\mathbf{A}_{ij}|^2 < \frac{2}{k} \|\mathbf{A}\|_F^2 \le \frac{2}{k} \|\mathbf{A}\| \mathrm{Tr}(\mathbf{A})$$

a priori Error Estimates

Chebychev's inequality:

$$\mathbb{P}(|\bar{X}_k - \mathrm{Tr}(\mathbf{A})| \geq t) \leq \frac{\mathbb{V}(X)}{kt^2} \qquad \forall t > 0$$

The specific trace estimator will determine $\mathbb{V}(X)$:

- [Girard 1989] $\boldsymbol{\omega} \sim \mathcal{N}(0, \mathbf{I})$

  $$\mathbb{P}(|\bar{X}_k - \mathrm{Tr}(\mathbf{A})| \geq t\mathrm{Tr}(\mathbf{A})) \leq \frac{2}{kt^2 \mathrm{intdim}(\mathbf{A})}$$

  The bound improves as the intrinsic dimension of **A** increases.
- Much stronger bounds can be obtained when exponential concentration inequalities are used!
  (Cramér–Chernoff method)