Extending Trace Estimation to Schatten Norm Estimation

The methods outlined in this document follow sections 4.8 through 5.2, and 5.4 and 5.6 of Martinsson & Tropp (2021)[1]. Proofs and algorithms for these methods follow. Please find the accompanying code in [2].

The Frobenius norm and the Schatten 4-norm

Randomized trace estimators of the form given in the introduction can also be used to estimate some matrix norms, notably the Frobenius or Schatten 2-norm and the Schatten 4-norm, also referred to as the ℓ_2 and ℓ_4 norms.

Consider a rectangular matrix $\mathbf{B} \in \mathbb{R}^{mxn}$, accessed by a matrix-vector product $\mathbf{u} \mapsto \mathbf{Bu}$. Suppose that we extract test vectors from the standard normal distribution. Draw a standard normal matrix $\mathbf{\Omega} \in \mathbb{R}^{nxk}$ with columns ω_i. Then construct the random variable:

$$\overline{X}_k := \frac{1}{k}\| \mathbf{B\Omega}\|_F^2 = \frac{1}{k}\sum_{i=1}^{k} \omega_i'(\mathbf{B'B})\omega_i =: \frac{1}{k}\sum_{i=1}^{k} X_i \tag{1}$$

To analyse $\overline{X}_k$, note that it is an instance of the randomized trace estimator in (1), where $\mathbf{A} = \mathbf{B'B}$.

Proposition 3: $\overline{X}_k$ is an unbiased estimator for $\| \mathbf{B}\|_F^2$, with $\mathrm{Var}(\overline{X}_k) = \frac{2}{k}\| \mathbf{B}\|_4^4$.

Proof:

$$\begin{aligned}
\mathbb{E}[\overline{X}_k] &= \mathbb{E}[\tfrac{1}{k}\textstyle\sum_{i=1}^{k} \omega_i'(\mathbf{B'B})\omega_i] = \tfrac{1}{k}\mathbb{E}[\textstyle\sum_{i=1}^{k} \omega_i'(\mathbf{B'B})\omega_i] && \tfrac{1}{k} \text{ can be factored out by linearity;}\\
&= \tfrac{1}{k}\textstyle\sum_{i=1}^{k} \mathbb{E}[\omega_i'(\mathbf{B'B})\omega_i] && \text{because } X_i = \omega_i'(\mathbf{B'B})\omega_i \text{ are iid;}\\
&= \tfrac{1}{k}\textstyle\sum_{i=1}^{k} (\mathbf{B'B})\mathbb{E}[\omega_i'\omega_i] && (\mathbf{B'B}) \text{ can be factored out by linearity;}\\
&= \tfrac{1}{k}k\,(\mathbf{B'B})\textstyle\sum_{i=1}^{k} \mathbb{E}[\omega_i'\omega_i] = (\mathbf{B'B})\textstyle\sum_{i=1}^{k} \mathbb{E}[\omega_i'\omega_i] && \text{rewrite } (\mathbf{B'B}) \text{ as:}\\
&= (\textstyle\sum_{l=1}^{n} b_{il}b_{lj})\textstyle\sum_{i=1}^{k} \mathbb{E}[\omega_i'\omega_i] = \textstyle\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}^2 \textstyle\sum_{i=1}^{k} \mathbb{E}[\omega_i'\omega_i] && \text{use the isotropic property of } \omega_i\text{:}\\
&= \| \mathbf{B}\|_F^2
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\overline{X}_k) &= \mathrm{Var}(\tfrac{1}{k}\textstyle\sum_{i=1}^{k} \omega_i'(\mathbf{B'B})\omega_i)\\
&= \tfrac{1}{k^2}\mathrm{Var}(\textstyle\sum_{i=1}^{k} \omega_i'(\mathbf{B'B})\omega_i) && \tfrac{1}{k} \text{ can be factored out by linearity;}\\
&= \tfrac{1}{k^2}\textstyle\sum_{i=1}^{k} \mathrm{Var}(\omega_i'(\mathbf{B'B})\omega_i) && \text{because } X_i = \omega_i'(\mathbf{B'B})\omega_i \text{ are iid;}
\end{aligned}$$

---

[1] https://arxiv.org/abs/2002.01387v1, last accessed 4/30/21
[2] https://github.com/ghostpress/comp-stats-sims/tree/final-project/final-project/trace-estim

$$= \frac{1}{k^2}\sum_{i=1}^{k} (\mathbf{B'B})\ \mathrm{Var}(\omega_i'\omega_i)(\mathbf{B'B})' \qquad (\mathbf{B'B}) \text{ can be factored out by linearity;}$$

$$= \frac{1}{k^2}\sum_{i=1}^{k} (\mathbf{B'B})\ \mathbb{E}[(\omega_i'\omega_i - \mathbb{E}[\omega_i'\omega_i])(\omega_i'\omega_i - \mathbb{E}[\omega_i'\omega_i])'](\mathbf{B'B})$$

$$= \frac{1}{k^2}(\mathbf{B'B})(\mathbf{B'B})\sum_{i=1}^{k} \mathbb{E}[(\omega_i'\omega_i - \mathbf{I})(\omega_i'\omega_i - \mathbf{I})']$$

$$= \frac{1}{k^2}(\mathbf{B'B})(\mathbf{B'B})\sum_{i=1}^{k} \mathbb{E}[(\omega_i'\omega_i - \mathbf{I})(\omega_i'\omega_i - \mathbf{I})]$$

$$= \frac{1}{k^2}(\mathbf{B'B})(\mathbf{B'B})\sum_{i=1}^{k} \mathbb{E}[\omega_i'\omega_i\omega_i'\omega_i - \omega_i'\omega_i - \omega_i'\omega_i - \mathbf{I}]$$

$$= \frac{1}{k^2}(\mathbf{B'B})(\mathbf{B'B})\sum_{i=1}^{k} (\mathbb{E}[\omega_i'\omega_i\omega_i'\omega_i] - 2\mathbb{E}[\omega_i'\omega_i] - \mathbf{I})$$

$$= \frac{1}{k^2}(\mathbf{B'B})(\mathbf{B'B})\sum_{i=1}^{k} (\mathbb{E}[\omega_i'\omega_i]\mathbb{E}[\omega_i'\omega_i] - 3\mathbf{I}) \qquad \text{because } \omega_i \text{ are independent;}$$

$$= \frac{1}{k^2}(\mathbf{B'B})(\mathbf{B'B})\sum_{i=1}^{k} (\mathbf{I} - 3\mathbf{I})$$

$$= \frac{1}{k^2}(\mathbf{B'B})(\mathbf{B'B})\sum_{i=1}^{k} 2\mathbf{I}$$

$$= \frac{1}{k^2}(\mathbf{B'B})(\mathbf{B'B})2k$$

$$= \frac{2}{k}(\mathbf{B'B})(\mathbf{B'B})$$

$$= \frac{2}{k}\|\mathbf{B}\|_4^4$$

From these results we can also clearly see that rescaling $\mathrm{Var}(\overline{X}_k)$ by $\frac{k}{2}$ gives an unbiased estimate for $\|\mathbf{B}\|_4^4$, the Schatten 4-norm of **B**.

$\overline{X}_k$ is computed by simulating nk standard normal variables, taking k matrix-vector products with **B**, and performing $O(kn)$ additional arithmetic. Therefore the total runtime of this method is $O(2kn + knm)$.

<u>Schatten *p*-norm Estimation by Sampling</u>

We now extend the discussion of approximating Schatten 2- and 4-norms by sampling to the $2p$-norm for each $p \in \mathbb{N}$. Consider the general matrix $\mathbf{B} \in \mathbb{R}^{mxn}$, accessed via the matrix-vector product $\mathbf{u} \mapsto \mathbf{Bu}$. For a sample size k, let $\mathbf{\Omega} \in \mathbb{R}^{nxk}$ be a (random) test matrix that does not depend on **B**. For a natural number $p \geq 3$, the problem is to estimate the Schatten $2p$-norm $// \ \mathbf{B}\|_{2p}$ from the sample matrix $\mathbf{Y} = \mathbf{B\Omega}$ such that **Y** is an unbiased estimator of the norm. Moreover, the methods used should ideally be less expensive than the $O(\min\{mn^2, nm^2\})$ cost of computing the Singular Value Decomposition (SVD) of **B**, as in the classical algorithm.

The authors remark that the sample size k needed to estimate $\| \mathbf{B} \|_{2p}$ up to a fixed constant factor with 75% probability is unfortunately $k \gtrsim \min\{m, n\}^{1-2/p}$. In other words, the sample size must grow polynomially with the dimensions of the matrix for $p > 2$. Nevertheless, as shown below, the algorithm used is still faster than the SVD method.

Assume that the random test matrix $\mathbf{\Omega} \in \mathbb{R}^{nxk}$ has isotropic columns ω_i that are iid. Form the sample matrix $\mathbf{Y} = \mathbf{B\Omega}$. Abbreviate $\mathbf{A} = \mathbf{B}'\mathbf{B}$ and $\mathbf{X} = \mathbf{Y}'\mathbf{Y}$. Then:

$$(\mathbf{X})_{ij} = (\mathbf{Y}'\mathbf{Y})_{ij} = \omega_i'\mathbf{A}\omega_j \tag{2}$$

For any natural numbers that satisfy $1 \leq i_1, \ldots, i_p \leq k$,

$$(\mathbf{X})_{i_1 i_2}(\mathbf{X})_{i_2 i_3} \cdots (\mathbf{X})_{i_p i_1} = trace(\omega_{i_1}\omega_{i_1}' \mathbf{A} \ldots \omega_{i_p}\omega_{i_p}' \mathbf{A})$$

Under the assumption that $i_1, \ldots, i_p$ are distinct, by independence and isotropy the expectation becomes:

$$\mathbb{E}[(\mathbf{X})_{i_1 i_2}(\mathbf{X})_{i_2 i_3} \cdots (\mathbf{X})_{i_p i_1}] = trace(\mathbf{A}^p) = \| \mathbf{B} \|_{2p}^{2p}$$

And now define the estimator as below, where $C(k, p)$ is the binomial coefficient with k and p:

$$V_p = C(k, p)^{-1} \sum_{1 \leq i_1 \leq i_p \leq k} (\mathbf{X})_{i_1 i_2}(\mathbf{X})_{i_2 i_3} \cdots (\mathbf{X})_{i_p i_1}$$

Finally, we can reformulate V_p . Let $T : \mathbb{H}_k \rightarrow \mathbb{R}^{kxk}$ be the linear map that reports the strict upper triangle of a symmetric matrix. Then:

$$V_p = C(k, p)^{-1}\, trace(\ T(\mathbf{X})^{p-1}\mathbf{X}\ ) \tag{3}$$

The algorithm to compute V_p follows.

Algorithm 2: Schatten 2p-norm estimation by random sampling

Input: Matrix $\mathbf{B} \in \mathbb{R}^{mxn}$, the order p of the norm to estimate, and the number k of samples to take
Output: Schatten 2p-norm estimate V_p

1. Draw the test matrix $\mathbf{\Omega} \in \mathbb{R}^{nxk}$ with iid isotropic columns
2. Compute the sample matrix $\mathbf{Y} = \mathbf{B\Omega}$
3. Form the Gram matrix $\mathbf{X} = \mathbf{Y}'\mathbf{Y} \in \mathbb{R}^{kxk}$
4. Extract the strict upper triangle $\mathbf{T} = T(\mathbf{X})$
5. Compute $\mathbf{T}^{p-1}$ by repeated squaring
6. Return $V_p = trace(\mathbf{T}^{p-1}\mathbf{X})$

The runtime of this algorithm is dominated by the $O(k^2n)$ arithmetic required to form $\mathbf{X}$ given $\mathbf{Y}$.