Error bound for dual-process matrix multiplication difference

Problem: Let $A \in \mathbb{R}^{M \times K}, B \in \mathbb{R}^{K \times N}$ be random uniform matrices ($A_{ij}, B_{ij} \sim \mathcal{U}[-a, a]$). Consider two processes that calculate $C = AB$, giving results C^1, C^2. Assuming the processes are computer-based (representing numbers as `float32`, subjected to floating-point errors, etc.), find a upper bound for $E = \max \left|C^1 - C^2\right|_{ij}$.

Folded Gaussian Lemma

Lemma: The absolute value of a Gaussian $X \sim \mathcal{N}(0, \sigma)$ has mean $\sigma\sqrt{\frac{2}{\pi}}$.

Proof

Integrating:

$$
\begin{aligned}
\mathbb{E}[|X|] &= \int_{\mathbb{R}} |x| \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathrm{d}x \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathrm{d}(x^2) \\
&= \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_0^\infty \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathrm{d}\left(\frac{x^2}{2\sigma^2}\right) \\
&= \sigma\sqrt{\frac{2}{\pi}}.
\end{aligned}
$$

Derivation of error upper bound

Let $\Delta C = C^1 - C^2$, assuming ΔC are independent, then: $\max|\Delta C_{ij}|$ is the maximum of $M \times N$ i.i.d. random variables.

We look at each $\left|\Delta C_{ij}\right|$. The `f32` error of one floating-point operation can be estimated as a uniform distribution $\mathcal{U}[-t, t]$, for $t = \varepsilon\left|C_{ij}\right|$, where ε is the machine epsilon. The accumulation of such error can be modelled as the sum of i.i.d. random variables, which is the Gaussian $\mathcal{N}(0, \sigma)$, where σ is:

$$
\sigma_1 = \sqrt{K\frac{t^2}{3}} = \varepsilon\left|C_{ij}\right|\sqrt{\frac{K}{3}}.
$$

To estimate $\left|C_{ij}\right|$, we use the Central Limit Theorem again: it is the sum of $A_{ik}B_{kj}$, which are independent products of two uniformly distributed variables. We can calculate the mean (which is 0) and the standard deviation of each product as

$$
\begin{aligned}
\sigma_2 = \sqrt{\mathbb{V}[C_{ij}]} &= \sqrt{K\mathbb{V}[A_{ik}B_{kj}]} \\
&= \sqrt{K\left(\mathbb{E}[A_{ik}]^2\mathbb{V}[B_{kj}] + \mathbb{E}[B_{kj}]^2\mathbb{V}[A_{ik}] + \mathbb{V}[A_{ik}]\mathbb{V}[B_{kj}]\right)} \\
&= \sqrt{K}\mathbb{V}[A_{ik}] = \frac{a^2}{3}\sqrt{K}.
\end{aligned}
$$

Then, $\left|C_{ij}\right|$ has mean $\sigma'\sqrt{\frac{2}{\pi}}$, using the lemma above. Substituting everything in:

$$
\sigma_1 = \varepsilon\frac{a^2}{3}\sqrt{\frac{2K}{\pi}}\sqrt{\frac{K}{3}} = \sqrt{\frac{2}{27\pi}}\varepsilon a^2 K.
$$

Once we have the distribution of the error of one floating-point process, we can simply subtract them to find the distribution of the error between the two processes. Assuming the two process is independent, which means the two errors aer also independent, so the difference is yet another Gaussian $\mathcal{N}\left(0, \sigma\sqrt{2}\right)$. Then, the absolute value of that difference is just the absolute value of a Gaussian. Denote $\sigma = \sigma_1\sqrt{2} = \sqrt{\frac{4}{27\pi}}\varepsilon a^2 K$.

Now, onto the maximum part. We simply evaluate the CDF:

$$F_E(x) = \mathbb{P}(E \leq x) = \mathbb{P}\left(\left|(\Delta C)_{ij}\right| \leq x, \forall i, j\right) = \operatorname{erf}\left(\frac{x}{\sigma\sqrt{2}}\right)^{MN}.$$

To find an upper bound U_α that works $1 - \alpha$ of the time, we need:

$$F_E(U_\alpha) = 1 - \alpha \Rightarrow \operatorname{erf}\left(\frac{U_\alpha}{\sigma\sqrt{2}}\right) = \sqrt[MN]{1-\alpha} \Rightarrow U_\alpha = \sigma\sqrt{2}\operatorname{erf}^{-1}\left(\sqrt[MN]{1-\alpha}\right).$$

For the full formula:

$$U_\alpha = \sqrt{\frac{8}{27\pi}}\varepsilon a^2 K \operatorname{erf}^{-1}\left(\sqrt[MN]{1-\alpha}\right).$$

Asymptotically, the erf term can be reduced:

$$\operatorname{erf}^{-1}\left(\sqrt[MN]{1-\alpha}\right) \approx \operatorname{erf}^{-1}\left(1 - \frac{\alpha}{MN}\right) \approx \frac{1}{\sqrt{2}}\sqrt{\log\frac{2}{\pi b^2} - \log\log\frac{2}{\pi b^2}},$$

where $b = \frac{\alpha}{MN}$. Ignoring the log-log term:

$$\operatorname{erf}^{-1}\left(\sqrt[MN]{1-\alpha}\right) \approx \sqrt{\frac{1}{2}\log\frac{2(MN)^2}{\pi\alpha^2}} = \sqrt{\log\frac{MN}{\alpha}\sqrt{\frac{2}{\pi}} - \log\log\frac{MN}{\alpha}\sqrt{\frac{2}{\pi}}},$$

so in conclusion,

$$U_\alpha = \sqrt{\frac{8}{27\pi}\left(\log\frac{MN}{\alpha}\sqrt{\frac{2}{\pi}} - \log\log\frac{MN}{\alpha}\sqrt{\frac{2}{\pi}}\right)\varepsilon a^2 K}.$$