Error bounds for SVD verification in Halo2

Ashutosh Marwah, Guillaume Remy, Zhengxun Wu

November 5, 2023

There are two types of errors associated with SVD verification in Halo2- the first error due to the numpy computation algorithm (or any other algorithm for computing SVD; see [Bla99a, Bla99b]) and the second due to the quantization of the fixed point chip. In this note, we will relate these errors to the tolerance parameter for the zk circuit.

We will use the following norms for a $n \times m$ matrix X,

$$\|X\|_2 := \max_{\|v\|_2 \leq 1} \|Xv\|_2 \tag{1}$$

$$\|X\|_m := \max_{i,j} |X_{ij}|. \tag{2}$$

It should be noted that $\|X\|_2$ is the operator norm (also called the spectral norm)- one of the most commonly used matrix norm. It is in particular a unitarily invariant and submultiplicative norm [Bha97, Section 4.2]. $\|X\|_m$ is the element wise maximum norm, which is neither unitarily invariant nor submultiplicative but turns out to be easy to implement in Halo2, and nicely captures the error due to quantization. Also, it should be noted that for all matrices X,

$$\|X\|_m \leq \|X\|_2 \leq \sqrt{nm}\, \|X\|_m . \tag{3}$$

Let the $n \times m$ matrix M be provided in the fixed point quantization. Suppose, the mathematical (perfect) SVD for this matrix is given by

$$M = U\Sigma V \tag{4}$$

where matrices U and V are orthogonal and Σ is diagonal. Suppose further that numpy (or any other library) outputs the matrices $\hat{U}$, $\hat{\Sigma}$ and $\hat{V}$ as the SVD of M, which satisfy the guarantees

$$\left\|M - \hat{U}\hat{\Sigma}\hat{V}\right\|_2 \leq \epsilon_{svd} \tag{5}$$

$$\left\|\Sigma - \hat{\Sigma}\right\|_2 \leq \epsilon_{svd} \tag{6}$$

$$\left\|\mathbb{1} - \hat{U}\hat{U}^T\right\|_2 \leq \epsilon_U \tag{7}$$

$$\left\|\mathbb{1} - \hat{V}\hat{V}^T\right\|_2 \leq \epsilon_U. \tag{8}$$

and that $\hat{U}$ and $\hat{V}$ are such that $\left\|\hat{U}\right\|_m \leq 1$ and $\left\|\hat{V}\right\|_m \leq 1$.
The error bounds given in [Bla99b] appear in a similar form. We do not necessarily need the second condition above, a similar bound is implied by the other bounds, but we include it to avoid unnecessary calculations. From [Bla99b], we get the following "rough" values for the values of $(\epsilon_{svd}, \epsilon_U)$ for the SVD function of numpy

$$\epsilon_{svd} = \|M\|_2 \, p(m,n)\epsilon \tag{9}$$
$$\epsilon_U = \max\{m,n\}\epsilon \tag{10}$$

where $\epsilon = 2^{-53} \approx 1.1 \times 10^{-16}$ is the precision of double arithmetic and $p(m,n)$ is "modestly growing function" of m and n[(1)].

We will write the fixed-point quantized versions of these using a q subscript, as $\hat{U}_q$, $\hat{\Sigma}_q$ and $\hat{V}_q$. For a precision of P bits for the fixed-point chip, the quantized version of a matrix X is given by

$$(X_q)_{ij} = 2^{-P}\lfloor 2^P X_{ij} + 0.5 \rfloor \tag{11}$$

so that,

$$\|X_q - X\|_m \leq 2^{-(P+1)}. \tag{12}$$

In our zk-circuits, we test the following three conditions for the quantizations of the SVD matrices

$$\left\|\mathbb{1} - \hat{U}_q \hat{U}_q^T\right\|_m \leq \text{err}_U \tag{13}$$
$$\left\|\mathbb{1} - \hat{V}_q \hat{V}_q^T\right\|_m \leq \text{err}_U \tag{14}$$
$$\left\|\hat{U}_q \hat{\Sigma}_q - M\hat{V}_q^T\right\|_m \leq \text{err}_{svd}. \tag{15}$$

In this note, we will relate the above errors with ϵ_{svd} and ϵ_U.

First, let's see how we can bound the corresponding errors for the numpy produced matrices. The conditions on the unitaries $\hat{U}$ and $\hat{V}$ in Eq. 7 and 8 are already of the above form. In order to bound the the third condition we will need bounds for $\left\|\hat{U}\right\|_2$, $\left\|\hat{V}\right\|_2$ and $\left\|\hat{\Sigma}\right\|_2$. For $\left\|\hat{U}\right\|_2$ we have

$$\begin{aligned} \left|1 - \left\|\hat{U}\right\|_2^2\right| &\leq \left\|\mathbb{1} - \hat{U}\hat{U}^T\right\|_2 \leq \epsilon_U \\ \Rightarrow \sqrt{1-\epsilon_U} &\leq \left\|\hat{U}\right\|_2 \leq \sqrt{1+\epsilon_U}. \end{aligned} \tag{16}$$

[(1)] [Bla99c] states that in practice $p(m,n)$ is linear in m and/ or n, but one can only prove bounds, which are cubic $O(n^3)$. Since, in a cryptographic protocol, we are in an adversarial setting, one needs to be very careful of what an honest algorithm for the prover can prove. We leave this for future work and simply use the average observed error during numpy's SVD for our demo.

Similarly,

$$\sqrt{1-\epsilon_U} \leq \left\|\hat{V}\right\|_2 \leq \sqrt{1+\epsilon_U}. \tag{17}$$

For $\left\|\hat{\Sigma}\right\|_2$, we have

$$\left\|\hat{\Sigma}\right\|_2 \leq \|M\|_2 + \epsilon_{svd}. \tag{18}$$

For the third condition, we have

$$\begin{aligned}
\left\|\hat{U}\hat{\Sigma} - M\hat{V}^T\right\|_m &\leq \left\|\hat{U}\hat{\Sigma} - M\hat{V}^T\right\|_2 \\
&\leq \left\|\hat{U}\hat{\Sigma} - \hat{U}\hat{\Sigma}\hat{V}\hat{V}^T\right\|_2 + \left\|\hat{U}\hat{\Sigma}\hat{V}\hat{V}^T - M\hat{V}^T\right\|_2 \\
&\leq \left\|\hat{U}\right\|_2 \left\|\hat{\Sigma}\right\|_2 \left\|\mathbb{1} - \hat{V}\hat{V}^T\right\|_2 + \left\|\hat{V}\right\|_2 \left\|\hat{U}\hat{\Sigma}\hat{V} - M\right\|_2 \\
&\leq \sqrt{1+\epsilon_U}(\|M\|_2 + \epsilon_{svd})\epsilon_U + \epsilon_{svd}\sqrt{1+\epsilon_U}
\end{aligned} \tag{19}$$

Now, we will simply use the following Lemma along with the triangle inequality for norms to bound errors above for the quantized matrices.

Lemma 0.1. *For an $n \times k$ matrix A and a $k \times m$ matrix B and their quantizations A_q and B_q, we have*

$$\|AB - A_qB_q\|_m \leq 2^{-(P+1)}k\left(\|A\|_m + \|B\|_m + 2^{-(P+1)}\right) \tag{20}$$

Proof. We have

$$\begin{aligned}
\|AB - A_qB_q\|_m &= \max_{ij}\left|\sum_{l=1}^{k}\left(A_{il}B_{lj} - (A_q)_{il}(B_q)_{lj}\right)\right| \\
&\leq \max_{ij}\left\{\sum_{l=1}^{k}\left|A_{il}B_{lj} - (A_q)_{il}B_{lj}\right| + \left|(A_q)_{il}B_{lj} - (A_q)_{il}(B_q)_{lj}\right|\right\} \\
&\leq \sum_{l=1}^{k}\max_{ij}|A_{il} - (A_q)_{il}|\,|B_{lj}| + \max_{ij}\left|B_{lj} - (B_q)_{lj}\right|\,|(A_q)_{il}| \\
&\leq 2^{-(P+1)}k\left(\|A\|_m + \|B\|_m + 2^{-(P+1)}\right)
\end{aligned}$$

□

For the quantized matrices, we have

$$\begin{aligned}
\left\|\mathbb{1} - \hat{U}_q\hat{U}_q^T\right\|_m &\leq \left\|\mathbb{1} - \hat{U}\hat{U}^T\right\|_m + \left\|\hat{U}_q\hat{U}_q^T - \hat{U}\hat{U}^T\right\|_m \\
&\leq \epsilon_U + 2^{-(P+1)}n\left(2(1+\epsilon_U) + 2^{-(P+1)}\right)
\end{aligned} \tag{21}$$

where we use the fact that $\left\|\hat{U}\right\|_m \leq 1 + \epsilon_U$. Similarly, for the unitary $\hat{V}_q$, we have

$$\left\|\mathbb{1} - \hat{V}_q \hat{V}_q^T\right\|_m \leq \epsilon_U + 2^{-(P+1)} m \left(2(1+\epsilon_U) + 2^{-(P+1)}\right). \tag{22}$$

Finally, for the bound

$$\begin{aligned}
\left\|\hat{U}_q\hat{\Sigma}_q - M\hat{V}_q^T\right\|_m &\leq \left\|\hat{U}_q\hat{\Sigma}_q - \hat{U}\hat{\Sigma}\right\|_m + \left\|\hat{U}\hat{\Sigma} - M\hat{V}^T\right\|_m + \left\|M\hat{V}^T - M\hat{V}_q^T\right\|_m \\
&\leq 2^{-(P+1)} n \left(1 + \|M\|_2 + \epsilon_{svd} + 2^{-(P+1)}\right) + \left\|M\hat{V}^T - M\hat{V}_q^T\right\|_2 + \left\|\hat{U}\hat{\Sigma} - M\hat{V}^T\right\|_m \\
&\leq 2^{-(P+1)} n \left(1 + \|M\|_2 + \epsilon_{svd} + 2^{-(P+1)}\right) + \|M\|_2 \left\|\hat{V}^T - \hat{V}_q^T\right\|_2 + \left\|\hat{U}\hat{\Sigma} - M\hat{V}^T\right\|_m \\
&\leq 2^{-(P+1)} n \left(1 + \|M\|_2 + \epsilon_{svd} + 2^{-(P+1)}\right) + \|M\|_2 m 2^{-(P+1)} + \left\|\hat{U}\hat{\Sigma} - M\hat{V}^T\right\|_m \\
&\leq 2^{-(P+1)} n \left(1 + \|M\|_2 + \epsilon_{svd} + 2^{-(P+1)}\right) + \|M\|_2 m 2^{-(P+1)} \\
&\qquad + \sqrt{1+\epsilon_U}(\|M\|_2 + \epsilon_{svd})\epsilon_U + \sqrt{1+\epsilon_U}\epsilon_{svd}.
\end{aligned}$$

References

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169. Springer, 1997.

[Bla99a] Susan Blackford. Error bounds for the singular value decomposition (lapack). https://www.netlib.org/lapack/lug/node96.html, 1999. Accessed: 2023-10-30.

[Bla99b] Susan Blackford. Further details: Error bounds for the singular value decomposition (lapack). https://www.netlib.org/lapack/lug/node97.html, 1999. Accessed: 2023-10-30.

[Bla99c] Susan Blackford. Further details: How to measure errors (lapack). https://www.netlib.org/lapack/lug/node76.html#secbackgroundnormnotation, 1999. Accessed: 2023-10-30.