Appendix

A. Notation

For $n \in \mathbb{N}$, we use $[n]$ to denote the set $\{0, \ldots, n\}$. For a vector $\mathbf{v}$, we use $\mathbf{v}_j$ to denote the element in the j^{th} position of the vector. We use $\mathbf{A}_{j,:}$ and $\mathbf{A}_{:,k}$ to denote the j^{th} row and k^{th} column of the matrix $\mathbf{A}$ respectively. We assume both $\mathbf{A}_{j,:}$, $\mathbf{A}_{:,k}$ to be column vectors (thus $\mathbf{A}_{j,:}$ is the transpose of j^{th} row of $\mathbf{A}$). $\mathbf{A}_{j,k}$ denotes the element in j^{th} row and k^{th} column of $\mathbf{A}$. $\mathbf{A}_{j,:k}$ and $\mathbf{A}_{:j,k}$ denote the vectors containing the first k elements of the j^{th} row and first j elements of k^{th} column, respectively. $\mathbf{A}_{:j,:k}$ denotes the matrix containing the first j rows and k columns of $\mathbf{A}$. The same rules can be directly extended to higher order tensors. We use bold zero i.e $\mathbf{0}$ to denote the matrix (or tensor) consisting of zero at all elements, $\mathbf{I}_n$ to denote the identity matrix of size $n \times n$. We use $\mathbb{C}$ to denote the field of complex numbers and $\mathbb{R}$ for real numbers. For a scalar $a \in \mathbb{C}$, $\overline{a}$ denotes its complex conjugate. For a vector $\mathbf{v}$ or matrix (or tensor) $\mathbf{A}$, $\overline{\mathbf{v}}$ or $\overline{\mathbf{A}}$ denotes the element-wise complex conjugate. For $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{A}^H$ denotes the hermitian transpose i.e $\mathbf{A}^H = \overline{\mathbf{A}^T}$. For a scalar $a \in \mathbb{C}$, $\mathrm{Re}(a)$, $\mathrm{Im}(a)$ and $|a|$ denote the real part, imaginary part and modulus of a respectively. We use $[a, b)$ where $a, b \in \mathbb{C}$ to denote the set consisting of complex scalars on the line connecting a and b (including a, but excluding b). $\mathbf{A} \otimes \mathbf{B}$ denotes the kronecker product between matrices $\mathbf{A}$ and $\mathbf{B}$. We use ι to denote *iota* (i.e $\iota^2 = -1$).

For a matrix $\mathbf{A} \in \mathbb{C}^{q \times r}$ and a tensor $\mathbf{B} \in \mathbb{C}^{p \times q \times r}$, $\overrightarrow{\mathbf{A}}$ denotes the vector constructed by stacking the rows of $\mathbf{A}$ and $\overrightarrow{\mathbf{B}}$ by stacking the vectors $\overrightarrow{\mathbf{B}_{j,:,:}}$, $j \in [p-1]$ so that:

$$\left(\overrightarrow{\mathbf{A}}\right)^T = \left[\mathbf{A}_{0,:}^T \,,\, \mathbf{A}_{1,:}^T \,,\, \ldots \,,\, \mathbf{A}_{q-1,:}^T\right]$$
$$\left(\overrightarrow{\mathbf{B}}\right)^T = \left[\left(\overrightarrow{\mathbf{B}_{0,:,:}}\right)^T \,,\, \left(\overrightarrow{\mathbf{B}_{1,:,:}}\right)^T \,,\, \ldots \,,\, \left(\overrightarrow{\mathbf{B}_{p-1,:,:}}\right)^T\right]$$

For a 2D convolution filter, $\mathbf{L} \in \mathbb{C}^{p \times q \times r \times s}$, we define the tensor $\mathrm{conv\_transpose}(\mathbf{L}) \in \mathbb{C}^{q \times p \times r \times s}$ as follows:

$$[\mathrm{conv\_transpose}(\mathbf{L})]_{i,j,k,l} = [\overline{\mathbf{L}}]_{j,i,r-1-k,s-1-l} \tag{7}$$

Note that this is very different from the usual matrix transpose. Given an input $\mathbf{X} \in \mathbb{C}^{q \times n \times n}$, we use $\mathbf{L} \star \mathbf{X} \in \mathbb{C}^{p \times n \times n}$ to denote the convolution of filter $\mathbf{L}$ with $\mathbf{X}$. The notation $\mathbf{L} \star^i \mathbf{X} \triangleq \mathbf{L} \star^{i-1} (\mathbf{L} \star \mathbf{X})$. Unless specified otherwise, we assume zero padding and stride 1 in each direction.

B. Proofs

B.1. Proof of Theorem 1

Theorem. *Consider a convolution filter* $\mathbf{L} \in \mathbb{C}^{m \times m \times (2p+1) \times (2q+1)}$ *applied to an input* $\mathbf{X} \in \mathbb{C}^{m \times n \times n}$ *that results in output* $\mathbf{Y} = \mathbf{L} \star \mathbf{X} \in \mathbb{C}^{m \times n \times n}$. *Let* $\mathbf{J}$ *be the jacobian of* $\overrightarrow{\mathbf{Y}}$ *with respect to* $\overrightarrow{\mathbf{X}}$ *, then the jacobian for convolution with the filter* $\mathrm{conv\_transpose}(\mathbf{L})$ *is equal to* $\mathbf{J}^H$.

Proof. We first prove the above result assuming $m = 1$.
Assuming $\mathbf{m = 1}$:
We know that $\mathbf{J}$ is a doubly toeplitz matrix of size $n^2 \times n^2$:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}^{(0)} & \mathbf{J}^{(-1)} & \cdots & \mathbf{J}^{(-p)} & 0 \\ \mathbf{J}^{(1)} & \mathbf{J}^{(0)} & \mathbf{J}^{(-1)} & \ddots & \ddots \\ \vdots & \mathbf{J}^{(1)} & \mathbf{J}^{(0)} & \ddots & \ddots \\ \mathbf{J}^{(p)} & \ddots & \ddots & \ddots & \mathbf{J}^{(-1)} \\ 0 & \ddots & \ddots & \mathbf{J}^{(1)} & \mathbf{J}^{(0)} \end{bmatrix}$$

In the above equation, each $\mathbf{J}^{(i)}$ is a toeplitz matrix of size $n \times n$. Define $\mathbf{P}^{(k)}$ as a $n \times n$ matrix with $\mathbf{P}^{(k)}_{i,j} = 1$ if $i - j = k$ and 0 otherwise. Thus $\mathbf{J}$ can be written as:

$$\mathbf{J} = \sum_{i=-p}^{p} \mathbf{P}^{(i)} \otimes \mathbf{J}^{(i)}$$

Since each matrix $\mathbf{J}^{(i)}$ is a toeplitz matrix, it can be written as follows. Because the first two dimensions of filter $\mathbf{L}$ are of size 1, we index $\mathbf{L}$ using only the last two indices:

$$\mathbf{J}^{(i)} = \sum_{j=-q}^{q} \mathbf{L}_{p+i,q+j} \mathbf{P}^{(j)}$$

Thus, $\mathbf{J}$ can be written as:

$$\mathbf{J} = \sum_{i=-p}^{p} \sum_{j=-q}^{q} \mathbf{L}_{p+i,q+j} \left(\mathbf{P}^{(i)} \otimes \mathbf{P}^{(j)}\right)$$

Thus, $\mathbf{J}^H$ can be written as:

$$\mathbf{J}^H = \sum_{i=-p}^{p} \sum_{j=-q}^{q} \overline{\mathbf{L}_{p+i,q+j}} \left(\mathbf{P}^{(i)} \otimes \mathbf{P}^{(j)}\right)^T$$
$$\mathbf{J}^H = \sum_{i=-p}^{p} \sum_{j=-q}^{q} \overline{\mathbf{L}_{p+i,q+j}} \left(\mathbf{P}^{(i)}\right)^T \otimes \left(\mathbf{P}^{(j)}\right)^T$$
$$\mathbf{J}^H = \sum_{i=-p}^{p} \sum_{j=-q}^{q} \overline{\mathbf{L}_{p+i,q+j}} \left(\mathbf{P}^{(-i)} \otimes \mathbf{P}^{(-j)}\right)$$
$$\mathbf{J}^H = \sum_{i=-p}^{p} \sum_{j=-q}^{q} \overline{\mathbf{L}_{p-i,q-j}} \left(\mathbf{P}^{(i)} \otimes \mathbf{P}^{(j)}\right)$$

Thus $\mathbf{J}^H$ corresponds to the jacobian of the convolution filter flipped along the third, fourth axis and each individual element conjugated.

Next, we prove the result when $m > 1$.
Assuming $\mathbf{m} > \mathbf{1}$:
We know that $\mathbf{J}$ is a matrix of size $mn^2 \times mn^2$. Let $\mathbf{J}^{(i,j)}$ denote the block of size $n^2 \times n^2$ as follows:

$$\mathbf{J}^{(i,j)} = \mathbf{J}_{in^2:(i+1)n^2, jn^2:(j+1)n^2}$$

Note that $\mathbf{J}^{(i,j)}$ is the jacobian of convolution with 1×1 filter $\mathbf{L}_{i:i+1,j:j+1,:,:}$. Now consider the $(i,j)^{th}$ block of $\mathbf{J}^H$. Using definition of conjugate transpose (i.e H operator):

$$\left(\mathbf{J}^H\right)^{(i,j)} = \left(\mathbf{J}^{(j,i)}\right)^H \tag{8}$$

Consider the 1×1 filter at the $(i,j)^{th}$ index in $\mathrm{conv\_transpose}(\mathbf{L})$. By the definition of $\mathrm{conv\_transpose}$ operator, we have:

$$\begin{aligned}&[\mathrm{conv\_transpose}(\mathbf{L})]_{i:i+1,j:j+1,:,:}\\ &= \mathrm{conv\_transpose}(\mathbf{L}_{j:j+1,i:i+1,:,:})\end{aligned} \tag{9}$$

Using equations (8) and (9) and the proof for the case $\mathbf{m} = 1$, we have the desired proof.

□

B.2. Proof of Theorem 2

Theorem. *Consider a convolution filter* $\mathbf{L} \in \mathbb{C}^{m\times m\times(2p+1)\times(2q+1)}$. *Given an input* $\mathbf{X} \in \mathbb{C}^{m\times n\times n}$, *output* $\mathbf{Y} = \mathbf{L} \star \mathbf{X} \in \mathbb{C}^{m\times n\times n}$. *The jacobian of* $\overrightarrow{\mathbf{Y}}$ *with respect to* $\overrightarrow{\mathbf{X}}$ *(call it* $\mathbf{J}$*) will be a matrix of size* $n^2m \times n^2m$. $\mathbf{J}$ *is a skew hermitian matrix if and only if:*

$$\mathbf{L} = \mathbf{M} - \mathrm{conv\_transpose}(\mathbf{M})$$

for some filter $\mathbf{M} \in \mathbb{C}^{m\times m\times(2p+1)\times(2q+1)}$.

Proof. We first prove that if $\mathbf{J}$ is a skew-hermitian matrix, then:

$$\mathbf{L} = \mathbf{M} - \mathrm{conv\_transpose}(\mathbf{M})$$

Let $\mathbf{J}^{(i,j)}$ denote the block of size $n^2 \times n^2$ as follows:

$$\mathbf{J}^{(i,j)} = \mathbf{J}_{in^2:(i+1)n^2, jn^2:(j+1)n^2}$$

so that $\mathbf{J}$ can be written in terms of the blocks $\mathbf{J}^{(i,j)}$:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}^{(0,0)} & \mathbf{J}^{(0,1)} & \dots & \mathbf{J}^{(0,m-1)} \\ \mathbf{J}^{(1,0)} & \mathbf{J}^{(1,1)} & \dots & \mathbf{J}^{(1,m-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{J}^{(m-1,0)} & \mathbf{J}^{(m-1,1)} & \dots & \mathbf{J}^{(m-1,m-1)} \end{bmatrix}$$

Since $\mathbf{J}$ is skew-hermitian, we have:

$$\mathbf{J}^{(i,j)} = -\left(\mathbf{J}^{(j,i)}\right)^H, \quad \forall\, i,\ j \in [m-1]$$

It is readily observed that $\mathbf{J}^{(i,j)}$ corresponds to the jacobian of convolution with 1×1 filter $\mathbf{L}_{i:i+1,j:j+1,:,:}$. For some given filter $\mathbf{A}$, we use $\mathbf{A}^{(i,j)}$ to denote the 1×1 filter $\mathbf{A}_{i:i+1,j:j+1,:,:}$ for simplicity. Thus, the above equation can be rewritten as:

$$\mathbf{L}^{(i,j)} = -\mathrm{conv\_transpose}\left(\mathbf{L}^{(j,i)}\right), \quad \forall\, i,\ j \in [m-1] \tag{10}$$

Now construct a filter $\mathbf{M}$ such that for $i \neq j$:

$$\mathbf{M}^{(i,j)} = \begin{cases} \mathbf{L}^{(i,j)}, & i < j \\ \mathbf{0}, & i > j \end{cases} \tag{11}$$

For $i = j$, $\mathbf{M}$ is given as follows:

$$\mathbf{M}^{(i,i)}_{r,s} = \begin{cases} \mathbf{L}^{(i,i)}_{r,s}, & r \leq p-1 \\ \mathbf{L}^{(i,i)}_{r,s}, & r = p,\ s \leq q-1 \\ 0.5 \times \mathbf{L}^{(i,i)}_{r,s}, & r = p,\ s = q \\ 0, & otherwise \end{cases} \tag{12}$$

Next, our goal is to show that:

$$\mathbf{L} = \mathbf{M} - \mathrm{conv\_transpose}(\mathbf{M})$$

Now by the definition of $\mathrm{conv\_transpose}$, we have:

$$\begin{aligned}&[\mathbf{M} - \mathrm{conv\_transpose}(\mathbf{M})]^{(i,j)}\\ &= \mathbf{M}^{(i,j)} - [\mathrm{conv\_transpose}(\mathbf{M})]^{(i,j)}\\ &= \mathbf{M}^{(i,j)} - \mathrm{conv\_transpose}\left(\mathbf{M}^{(j,i)}\right)\end{aligned} \tag{13}$$

Case 1: For $i < j$, using equations (10) and (11):

$$\mathbf{M}^{(i,j)} - \mathrm{conv\_transpose}\left(\mathbf{M}^{(j,i)}\right) = \mathbf{M}^{(i,j)} = \mathbf{L}^{(i,j)}$$

Case 2: For $i > j$, using equations (10) and (11):

$$\begin{aligned}&\mathbf{M}^{(i,j)} - \mathrm{conv\_transpose}\left(\mathbf{M}^{(j,i)}\right)\\ &= -\mathrm{conv\_transpose}\left(\mathbf{M}^{(j,i)}\right)\\ &= -\mathrm{conv\_transpose}\left(\mathbf{L}^{(j,i)}\right) = \mathbf{L}^{(i,j)}\end{aligned}$$

Case 3: For $i = j$, we further simplify equation (13):

$$\begin{aligned}&\mathbf{M}^{(i,i)}_{r,s} - \left[\mathrm{conv\_transpose}\left(\mathbf{M}^{(i,i)}\right)\right]_{r,s}\\ &= \mathbf{M}^{(i,i)}_{r,s} - \overline{\mathbf{M}^{(i,i)}_{2p-r,2q-s}}\end{aligned} \tag{14}$$

Subcase 3(a): For $(r \leq p-1)$ or $(r = p,\ s \leq q-1)$, we have:

$$\mathbf{M}^{(i,i)}_{2p-r,2q-s} = 0$$

Thus for $(r \leq p-1)$ or $(r = p,\ s \leq q-1)$: equation (14) simplifies to $\mathbf{M}^{(i,i)}_{r,s}$. The result follows trivially from the very definition of $\mathbf{M}^{(i,i)}_{r,s}$, i.e equation (12).

Subcase 3(b): For $(r \geq p+1)$ or $(r = p,\ s \geq q+1)$, we have:

$$\mathbf{M}^{(i,i)}_{r,s} = 0$$

Thus, equation (14) simplifies to:

$$\mathbf{M}^{(i,i)}_{r,s} - \overline{\mathbf{M}^{(i,i)}_{2p-r,2q-s}} = -\overline{\mathbf{M}^{(i,i)}_{2p-r,2q-s}}$$

Since $(r \geq p+1)$ or $(r = p,\ s \geq q+1)$, we have: $(2p-r \leq p-1)$ or $(2p-r = p,\ 2q-s \leq q-1)$ respectively. Thus using equation (12), we have:

$$-\overline{\mathbf{M}^{(i,i)}_{2p-r,2q-s}} = -\overline{\mathbf{L}^{(i,i)}_{2p-r,2q-s}}$$

Since $\mathbf{L}^{(i,i)}$ is a skew-hermitian filter, we have from Theorem 1:

$$\mathbf{L}^{(i,i)}_{r,s} = -\overline{\mathbf{L}^{(i,i)}_{2p-r,2q-s}}$$

Thus in this subcase, equation (14) simplifies to $\mathbf{L}^{(i,i)}_{r,s}$ again.

Subcase 3(c): For $r = p,\ s = q$, since $\mathbf{L}^{(i,i)}$ is a skew-hermitian filter, we have:

$$\mathbf{L}^{(i,i)}_{p,q} = -\overline{\mathbf{L}^{(i,i)}_{p,q}}$$
$$\mathbf{L}^{(i,i)}_{p,q} + \overline{\mathbf{L}^{(i,i)}_{p,q}} = 0$$

Thus, $\mathbf{L}^{(i,i)}_{p,q}$ is a purely imaginary number. In this subcase

$$\mathbf{M}^{(i,i)}_{r,s} - \overline{\mathbf{M}^{(i,i)}_{2p-r,2q-s}}$$
$$= \mathbf{M}^{(i,i)}_{p,q} - \overline{\mathbf{M}^{(i,i)}_{p,q}} = 2\mathbf{M}^{(i,i)}_{p,q}$$

Using equation (12) , we have:

$$2\mathbf{M}^{(i,i)}_{p,q} = \mathbf{L}^{(i,i)}_{p,q}$$

Thus, we get:

$$\mathbf{M}^{(i,i)}_{r,s} - \left[\text{conv\_transpose}\left(\mathbf{M}^{(i,i)}\right)\right]_{r,s} = \mathbf{L}^{(i,i)}_{r,s}$$

Thus we have established: $\mathbf{L} = \mathbf{M} - \text{conv\_transpose}(\mathbf{M})$. Note that the opposite direction of the if and only if statement follows trivially from the above proof. □

B.3. Proof of Theorem 3

Theorem. *(a) For a scalar $\lambda \in \mathbb{C}$ with $\text{Re}(\lambda) = 0$, the error between $\exp(\lambda)$ and approximation $p_k(\lambda)$ given below can be bounded as follows:*

$$\exp(\lambda) = \sum_{i=0}^{\infty} \frac{\lambda^k}{i!}, \qquad p_k(\lambda) = \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} \tag{15}$$
$$\left|\exp(\lambda) - p_k(\lambda)\right| \leq \frac{|\lambda|^k}{k!}, \qquad \forall\, \lambda : \text{Re}(\lambda) = 0$$

(b) For a skew-hermitian matrix $\mathbf{A}$, the error between $\exp(\mathbf{A})$ and the series approximation $\mathbf{S}_k(\mathbf{A})$ can be bounded as follows:

$$\exp(\mathbf{A}) = \sum_{i=0}^{\infty} \frac{\mathbf{A}^i}{i!}, \qquad \mathbf{S}_k(\mathbf{A}) = \sum_{i=0}^{k-1} \frac{\mathbf{A}^i}{i!}$$
$$\|\exp(\mathbf{A}) - \mathbf{S}_k(\mathbf{A})\|_2 \leq \frac{\|\mathbf{A}\|_2^k}{k!}$$

Proof. Since $\mathbf{A}$ is skew-hermitian, it is a normal matrix and eigenvectors for distinct eigenvalues must be orthogonal. Let the eigenvalue decomposition of $\mathbf{A}$ be given as follows:

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^H$$

Note that Λ is a diagonal matrix, and each element along the diagonal is purely imaginary (since $\mathbf{A}$ is skew-hermitian). Exponentiating both sides, we get:

$$\exp(\mathbf{A}) = \mathbf{U}\exp(\Lambda)\mathbf{U}^H$$

Thus the error $\mathbf{E}_k(\mathbf{A})$ is given by:

$$\mathbf{E}_k(\mathbf{A}) = \exp(\mathbf{A}) - \mathbf{S}_k(\mathbf{A}) \tag{16}$$
$$\mathbf{E}_k(\mathbf{A}) = \mathbf{U}\left(\exp(\Lambda) - \mathbf{S}_k(\Lambda)\right)\mathbf{U}^H$$
$$\|\mathbf{E}_k(\mathbf{A})\|_2 = \|\mathbf{U}\left(\exp(\Lambda) - \mathbf{S}_k(\Lambda)\right)\mathbf{U}^H\|_2$$
$$\|\mathbf{E}_k(\mathbf{A})\|_2 = \|\left(\exp(\Lambda) - \mathbf{S}_k(\Lambda)\right)\|_2$$

Since $(\exp(\Lambda) - \mathbf{S}_k(\Lambda))$ is a diagonal matrix, we have:

$$\|\mathbf{E}_k(\mathbf{A})\|_2 = \max_i \left|\left(\exp(\Lambda_{i,i}) - p_k(\Lambda_{i,i})\right)\right| \tag{17}$$

Let λ be an arbitrary element along the diagonal of Λ i.e $\lambda = \Lambda_{i,i}$ for some i. First note that:

$$\lambda \int_0^1 \{\exp(t\lambda) - p_k(t\lambda)\}dt$$
$$= \int_0^1 \{\exp(t\lambda) - p_k(t\lambda)\}\lambda dt$$

Substituting $u = \lambda t$, we have:

$$= \int_0^\lambda \{\exp(u) - p_k(u)\}du$$
$$= \int_0^\lambda \exp(u)du - \int_0^\lambda p_k(u)du$$
$$= \exp(\lambda) - 1 - \int_0^\lambda p_k(u)du$$

Substituting $p_k(u)$ using equation (15):

$$
\begin{aligned}
&= \exp(\lambda) - 1 - \int_0^{\lambda} \sum_{i=0}^{k-1} \frac{u^i}{i!} du \\
&= \exp(\lambda) - 1 - \sum_{i=0}^{k-1} \int_0^{\lambda} \frac{u^i}{i!} du \\
&= \exp(\lambda) - 1 - \sum_{i=0}^{k-1} \frac{u^{i+1}}{(i+1)!} \Big|_0^{\lambda} \\
&= \exp(\lambda) - 1 - \sum_{i=0}^{k-1} \frac{\lambda^{i+1}}{(i+1)!} \\
&= \exp(\lambda) - 1 - \sum_{i=1}^{k} \frac{\lambda^{i}}{i!} \\
&= \exp(\lambda) - \sum_{i=0}^{k+1} \frac{\lambda^{i}}{i!} \\
&= \exp(\lambda) - p_{k+1}(\lambda)
\end{aligned}
$$

This gives the following result:

$$
\exp(\lambda) - p_{k+1}(\lambda) = \lambda \int_0^1 \Big(\exp(t\lambda) - p_k(t\lambda) \Big) dt \tag{18}
$$

We shall now prove the main result using induction and equation (18):

Base case:

Use $k = 0$ and the convention that $p_0(\lambda) = 0$. We know that $p_1(\lambda) = 1$.

$$
\begin{aligned}
\exp(\lambda) - p_1(\lambda) &= \lambda \int_0^1 \Big(\exp(t\lambda) - p_0(t\lambda) \Big) dt \\
\big|\exp(\lambda) - 1\big| &= \big|\lambda \int_0^1 \exp(t\lambda) dt\big|
\end{aligned}
$$

Since λ is purely imaginary and t is purely real, we have $\big|\exp(t\lambda)\big| = 1$:

$$
\begin{aligned}
\big|\exp(\lambda) - 1\big| &\leq \big|\lambda\big| \int_0^1 \big|\exp(t\lambda)\big| dt = \big|\lambda\big| \int_0^1 1 dt \\
\big|\exp(\lambda) - 1\big| &\leq \big|\lambda\big|
\end{aligned}
$$

Induction step:

Assuming this holds for all k i.e:

$$
\big|\exp(\lambda) - p_k(\lambda)\big| \leq \frac{|\lambda|^k}{k!} \tag{19}
$$

Now let us consider $\big|\exp(\lambda) - p_{k+1}(\lambda)\big|$:

$$
\begin{aligned}
\big|\exp(\lambda) - p_{k+1}(\lambda)\big| &\leq \left|\lambda \int_0^1 \Big(\exp(t\lambda) - p_k(t\lambda) \Big) dt\right| \\
\big|\exp(\lambda) - p_{k+1}(\lambda)\big| &\leq \big|\lambda\big| \int_0^1 \left|\Big(\exp(t\lambda) - p_k(t\lambda) \Big)\right| dt
\end{aligned}
$$

Using equation (19), we have:

$$
\begin{aligned}
\big|\exp(\lambda) - p_{k+1}(\lambda)\big| &\leq \big|\lambda\big| \int_0^1 \frac{|t\lambda|^k}{k!} dt \\
\big|\exp(\lambda) - p_{k+1}(\lambda)\big| &\leq \big|\lambda\big|^{k+1} \int_0^1 \frac{|t|^k}{k!} dt \\
\big|\exp(\lambda) - p_{k+1}(\lambda)\big| &\leq \frac{\big|\lambda\big|^{k+1}}{(k+1)!}
\end{aligned}
$$

This proves (a).

Since λ is an arbitrary element along the diagonal of eigenvalue matrix Λ, using equations (16) and (17) we have:

$$
\begin{aligned}
\|\exp(\mathbf{A}) - \mathbf{S}_k(\mathbf{A})\|_2 &= \max_i \big|\exp(\Lambda_{i,i}) - p_k(\Lambda_{i,i})\big| \\
\|\exp(\mathbf{A}) - \mathbf{S}_k(\mathbf{A})\|_2 &\leq \max_i \frac{\big|\Lambda_{i,i}\big|^k}{k!} \\
\|\exp(\mathbf{A}) - \mathbf{S}_k(\mathbf{A})\|_2 &\leq \frac{1}{k!} \max_i \big|\Lambda_{i,i}\big|^k
\end{aligned} \tag{20}
$$

Since $\mathbf{A}$ is skew-hermitian, it is a normal matrix and singular values are equal to the magnitude of eigenvalues. Thus we have from equation (20):

$$
\begin{aligned}
\max_i \big|\Lambda_{i,i}\big| &= \|\Lambda\|_2 = \|\mathbf{A}\|_2 \\
\|\exp(\mathbf{A}) - \mathbf{S}_k(\mathbf{A})\|_2 &\leq \frac{\|\mathbf{A}\|_2^k}{k!}
\end{aligned}
$$

This proves (b). □

B.4. Proof of Theorem 4

Theorem. *Given a real skew-symmetric matrix* $\mathbf{A} \in \mathbb{R}^{n \times n}$*, we can construct a real skew-symmetric matrix* $\mathbf{B} \in \mathbb{R}^{n \times n}$ *such that* $\mathbf{B}$ *satisfies: (a)* $\exp(\mathbf{A}) = \exp(\mathbf{B})$ *and (b)* $\|\mathbf{B}\|_2 \leq \pi$.

Proof. We know that for eigenvalues of real symmetric matrices are purely imaginary and come in pairs: $\lambda_1\iota,\ -\lambda_1\iota, \lambda_2\iota, -\lambda_2\iota$ where each λ_i is real. When n is an odd integer, 0 is an eigenvalue. Additionally, we know that a real skew symmetric matrix can be expressed in a block diagonal form as follows:

$$
\mathbf{A} = \mathbf{Q}\Sigma\mathbf{Q}^T \tag{21}
$$

Here $\mathbf{Q}$ is a real orthogonal matrix and Σ is a block diagonal matrix defined as follows:

$$
\Sigma_{2i:2i+2, 2i:2i+2} = \begin{bmatrix} 0 & \lambda_i \\ -\lambda_i & 0 \end{bmatrix}, \qquad 0 \leq i < \left\lfloor \frac{n}{2} \right\rfloor \tag{22}
$$

In the above equation, $\lambda_i \in \mathbb{R}$ and $\pm\lambda_i\iota$ are the eigenvalues of $\mathbf{A}$. When n is odd, we additionally have:

$$
\Sigma_{n-1,n-1} = 0
$$

Taking the exponential of both sides of equation (21):

$$\exp\left(\mathbf{A}\right) = \mathbf{Q}\exp\left(\Sigma\right)\mathbf{Q}^T \tag{23}$$

We can compute $\exp\left(\Sigma\right)$ by computing the exponential of each 2×2 block defined in equation (22):

$$\exp\left(\begin{bmatrix} 0 & \lambda_i \\ -\lambda_i & 0 \end{bmatrix}\right) = \exp\left(\lambda_i \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos(\lambda_i) & -\sin(\lambda_i) \\ \sin(\lambda_i) & \cos(\lambda_i) \end{bmatrix} \tag{24}$$

From equation (24), we observe each λ_i can be shifted by integer multiples of $2\pi\iota$ without changing the exponential. For each $\lambda_i, i \in [\lfloor n/2 \rfloor - 1]$, we define a scalar μ_i:

$$\mu_i = \lambda_i + 2\pi k_i \iota, \qquad k_i \in \mathbb{Z} \tag{25}$$
$$\mu_i \in [-\pi\iota, \pi\iota) \tag{26}$$

Construct a new matrix $\mathbf{B}$ defined as follows:

$$\mathbf{B} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T \tag{27}$$

The matrix $\mathbf{D}$ in equation (27) is defined as follows:

$$\mathbf{D}_{2i:2i+2, 2i:2i+2} = \begin{bmatrix} 0 & \mu_i \\ -\mu_i & 0 \end{bmatrix}, \qquad 0 \leq i < \left\lfloor \frac{n}{2} \right\rfloor \tag{28}$$

Let us verify that $\mathbf{B}$ satisfies the following properties. Using equations (24), (25) and (28), we know that:

$$\exp(\mathbf{D}) = \exp(\Lambda)$$

This results in the following set of equations:

$$\exp(\mathbf{B}) = \mathbf{Q}\exp(\mathbf{D})\mathbf{Q}^T$$
$$\exp(\mathbf{B}) = \mathbf{Q}\exp(\Lambda)\mathbf{Q}^T = \exp(\mathbf{A})$$

Using equations (26) and (28), we have:

$$\|\mathbf{B}\|_2 = \|\mathbf{Q}\mathbf{D}\mathbf{Q}^T\|_2 = \|\mathbf{D}\|_2$$
$$\|\mathbf{D}\|_2 \leq \pi$$

Note that $\mathbf{B}$ is a product of 3 real matrices $\mathbf{Q}$, $\mathbf{D}$ and $\mathbf{Q}^T$ and hence $\mathbf{B}$ is real. Moreover, since $\mathbf{D}$ is skew symmetric, $\mathbf{B}$ is skew symmetric. □

B.5. Proof of Theorem 5

Theorem 5. *Given a skew-hermitian matrix $\mathbf{A}$, we can construct a skew-hermitian matrix $\mathbf{B}$ by adding integer multiples of $2\pi\iota$ to eigenvalues of $\mathbf{A}$ such that $\mathbf{B}$ satisfies: (a) $\exp(\mathbf{A}) = \exp(\mathbf{B})$ and (b) $\|\mathbf{B}\|_2 \leq \pi$.*

Proof. Let the eigenvalue decomposition of $\mathbf{A}$ be given:

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^H$$

Let λ_j be some eigenvalue of $\mathbf{A}$ such that:

$$\lambda_j = \Lambda_{j,j}$$

Construct a new diagonal matrix $\mathbf{D}$ of eigenvalues such that:

$$\mathbf{D}_{j,j} = \lambda_j + 2\pi k_j \iota, \qquad k_j \in \mathbb{Z} \tag{29}$$
$$\mathbf{D}_{j,j} \in [-\pi\iota, \pi\iota) \tag{30}$$

Construct a new matrix $\mathbf{B}$ defined as follows:

$$\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{U}^H$$

Let us verify that $\mathbf{B}$ satisfies the following properties. Using equation (29), we have:

$$\exp(\mathbf{B}) = \mathbf{U}\exp(\mathbf{D})\mathbf{U}^H$$
$$\exp(\mathbf{B}) = \mathbf{U}\exp(\Lambda)\mathbf{U}^H = \exp(\mathbf{A})$$

Using equation (30), we have:

$$\|\mathbf{B}\|_2 = \|\mathbf{U}\mathbf{D}\mathbf{U}^H\|_2 = \|\mathbf{D}\|_2$$
$$\|\mathbf{D}\|_2 = \max_j |D_{j,j}| \leq \pi$$

□

B.6. Proof of Theorem 6

Theorem 6. *Consider a convolution filter $\mathbf{L} \in \mathbb{C}^{m \times m \times (2p+1) \times (2q+1) \times (2r+1)}$ applied to an input $\mathbf{X} \in \mathbb{C}^{m \times n \times n \times n}$ that results in output $\mathbf{Y} = \mathbf{L} \star \mathbf{X} \in \mathbb{C}^{m \times n \times n \times n}$. Let $\mathbf{J}$ be the jacobian of $\overrightarrow{\mathbf{Y}}$ with respect to $\overrightarrow{\mathbf{X}}$, then the jacobian for convolution with the filter* $\text{conv3d\_transpose}(\mathbf{L})$ *is equal to $\mathbf{J}^H$.*

Proof. We first prove the above result assuming $m = 1$.

Assuming $\mathbf{m} = \mathbf{1}$:

Because the first two dimensions of filter $\mathbf{L}$ are of size 1, we index $\mathbf{L}$ using only the last two indices. Define $\mathbf{P}^{(k)}$ as a $n \times n$ matrix with $\mathbf{P}^{(k)}_{i,j} = 1$ if $i - j = k$ and 0 otherwise. We know that $\mathbf{J}$ is a triply toeplitz matrix of size $n^3 \times n^3$ given as follows:

$$\mathbf{J} = \sum_{i=-p}^{p} \sum_{j=-q}^{q} \sum_{k=-r}^{r} \mathbf{L}_{p+i,q+j,r+k} \left(\mathbf{P}^{(i)} \otimes \mathbf{P}^{(j)} \otimes \mathbf{P}^{(k)}\right)$$

Thus, $\mathbf{J}^H$ can be written as:

$$\begin{aligned}
&\mathbf{J}^H \\
&= \sum_{i=-p}^{p} \sum_{j=-q}^{q} \sum_{k=-r}^{r} \overline{\mathbf{L}_{p+i,q+j,r+k}} \left(\mathbf{P}^{(i)} \otimes \mathbf{P}^{(j)} \otimes \mathbf{P}^{(k)}\right)^T \\
&= \sum_{i=-p}^{p} \sum_{j=-q}^{q} \sum_{k=-r}^{r} \overline{\mathbf{L}_{p+i,q+j,r+k}} \mathbf{P}^{(-i)} \otimes \mathbf{P}^{(-j)} \otimes \mathbf{P}^{(-k)} \\
&= \sum_{i=-p}^{p} \sum_{j=-q}^{q} \sum_{k=-r}^{r} \overline{\mathbf{L}_{p-i,q-j,r-k}} \mathbf{P}^{(i)} \otimes \mathbf{P}^{(j)} \otimes \mathbf{P}^{(k)}
\end{aligned}$$

Thus $\mathbf{J}^H$ corresponds to the jacobian of the convolution filter flipped along the third, fourth, fifth axis and each individual element conjugated.

Next, we prove the above result when $m > 1$.
Assuming $\mathbf{m} > \mathbf{1}$:
We know that $\mathbf{J}$ is a matrix of size $mn^3 \times mn^3$. Let $\mathbf{J}^{(i,j)}$ denote the block of size $n^3 \times n^3$ as follows:

$$\mathbf{J}^{(i,j)} = \mathbf{J}_{in^3:(i+1)n^3, jn^3:(j+1)n^3}$$

Note that $\mathbf{J}^{(i,j)}$ is the jacobian of convolution with 1×1 filter $\mathbf{L}_{i:i+1,j:j+1,:,:}$. Now consider the $(i,j)^{th}$ block of $\mathbf{J}^H$. Using definition of conjugate transpose (i.e H operator):

$$\left(\mathbf{J}^H\right)^{(i,j)} = \left(\mathbf{J}^{(j,i)}\right)^H \tag{31}$$

Consider the 1×1 filter at the $(i,j)^{th}$ index in $\text{conv3d\_transpose}(\mathbf{L})$. By the definition of conv3d_transpose operator, we have:

$$\begin{aligned}&[\text{conv3d\_transpose}(\mathbf{L})]_{i:i+1,j:j+1,:,:}\\ &= \text{conv3d\_transpose}(\mathbf{L}_{j:j+1,i:i+1,:,:})\end{aligned} \tag{32}$$

Using equations (31) and (32) and the proof for the case $\mathbf{m} = 1$, we have the desired proof.

□

B.7. Proof of Theorem 7

Theorem 7. *Consider a convolution filter* $\mathbf{L} \in \mathbb{C}^{m\times m\times(2p+1)\times(2q+1)\times(2r+1)}$. *Given an input* $\mathbf{X} \in \mathbb{C}^{m\times n\times n\times n}$*, output* $\mathbf{Y} = \mathbf{L} \star \mathbf{X} \in \mathbb{C}^{m\times n\times n\times n}$. *The jacobian of* $\overrightarrow{\mathbf{Y}}$ *with respect to* $\overrightarrow{\mathbf{X}}$ *(call it* $\mathbf{J}$*) will be a matrix of size* $n^3m \times n^3m$. $\mathbf{J}$ *is a skew hermitian matrix if and only if:*

$$\mathbf{L} = \mathbf{M} - \text{conv3d\_transpose}(\mathbf{M})$$

for some filter $\mathbf{M} \in \mathbb{C}^{m\times m\times(2p+1)\times(2q+1)\times(2r+1)}$.

Proof. We first prove that if $\mathbf{J}$ is a skew-hermitian matrix, then:

$$\mathbf{L} = \mathbf{M} - \text{conv3d\_transpose}(\mathbf{M})$$

Let $\mathbf{J}^{(i,j)}$ denote the block of size $n^3 \times n^3$ as follows:

$$\mathbf{J}^{(i,j)} = \mathbf{J}_{in^3:(i+1)n^3, jn^3:(j+1)n^3}$$

Since $\mathbf{J}$ is skew-hermitian, we have:

$$\mathbf{J}^{(i,j)} = -\left(\mathbf{J}^{(j,i)}\right)^H, \quad \forall\, i,\ j \in [m-1]$$

It is readily observed that $\mathbf{J}^{(i,j)}$ corresponds to the jacobian of convolution with 1×1 filter $\mathbf{L}_{i:i+1,j:j+1,:,:,:}$. For some given filter $\mathbf{A}$, we use $\mathbf{A}^{(i,j)}$ to denote the 1×1 filter $\mathbf{A}_{i:i+1,j:j+1,:,:,:}$ for simplicity. Thus, the above equation can be rewritten as:

$$\mathbf{L}^{(i,j)} = -\text{conv3d\_transpose}\left(\mathbf{L}^{(j,i)}\right), \quad \forall\, i,\ j \in [m-1] \tag{33}$$

Now construct a filter $\mathbf{M}$ such that for $i \neq j$:

$$\mathbf{M}^{(i,j)} = \begin{cases}\mathbf{L}^{(i,j)}, & i < j\\ \mathbf{0}, & i > j\end{cases} \tag{34}$$

For $i = j$, $\mathbf{M}$ is given as follows:

$$\mathbf{M}^{(i,i)}_{s,t,u} = \begin{cases}\mathbf{L}^{(i,i)}_{s,t,u}, & s \leq p-1\\ \mathbf{L}^{(i,i)}_{s,t,u}, & s = p,\ t \leq q-1\\ \mathbf{L}^{(i,i)}_{s,t,u}, & s = p,\ t = q,\ u \leq r-1\\ 0.5 \times \mathbf{L}^{(i,i)}_{s,t,u}, & s = p,\ t = q,\ u = r\\ 0, & otherwise\end{cases} \tag{35}$$

Next, our goal is to show that:

$$\mathbf{L} = \mathbf{M} - \text{conv3d\_transpose}(\mathbf{M})$$

Now by the definition of conv3d_transpose, we have:

$$\begin{aligned}&[\mathbf{M} - \text{conv3d\_transpose}(\mathbf{M})]^{(i,j)}\\ &= \mathbf{M}^{(i,j)} - [\text{conv3d\_transpose}(\mathbf{M})]^{(i,j)}\\ &= \mathbf{M}^{(i,j)} - \text{conv3d\_transpose}\left(\mathbf{M}^{(j,i)}\right)\end{aligned} \tag{36}$$

Case 1: For $i < j$, using equations (33) and (34):

$$\mathbf{M}^{(i,j)} - \text{conv3d\_transpose}\left(\mathbf{M}^{(j,i)}\right) = \mathbf{M}^{(i,j)} = \mathbf{L}^{(i,j)}$$

Case 2: For $i > j$, using equations (33) and (34):

$$\begin{aligned}&\mathbf{M}^{(i,j)} - \text{conv3d\_transpose}\left(\mathbf{M}^{(j,i)}\right)\\ &= -\text{conv3d\_transpose}\left(\mathbf{M}^{(j,i)}\right)\\ &= -\text{conv3d\_transpose}\left(\mathbf{L}^{(j,i)}\right) = \mathbf{L}^{(i,j)}\end{aligned}$$

Case 3: For $i = j$, we further simplify equation (36):

$$\begin{aligned}&\mathbf{M}^{(i,i)}_{s,t,u} - \left[\text{conv3d\_transpose}\left(\mathbf{M}^{(i,i)}\right)\right]_{s,t,u}\\ &= \mathbf{M}^{(i,i)}_{s,t,u} - \overline{\mathbf{M}^{(i,i)}_{2p-s,2q-t,2r-u}}\end{aligned} \tag{37}$$

Subcase 3(a): For $(s \leq p-1)$ or $(s = p,\ t \leq q-1)$ or $(s = p,\ t = q,\ u \leq r-1)$, we have:

$$\mathbf{M}^{(i,i)}_{2p-s,2q-t,2r-u} = 0$$

Thus for ($s \leq p-1$) or ($s = p,\ t \leq q-1$) or ($s = p,\ t = q,\ u \leq r-1$): equation (37) simplifies to $\mathbf{M}^{(i,i)}_{s,t,u}$. The result follows trivially from the very definition of $\mathbf{L}^{(i,i)}_{s,t,u}$, i.e equation (35).

Subcase 3(b): For ($s \geq p+1$) or ($s = p,\ t \geq q+1$) or ($s = p,\ t = q,\ u \geq r+1$), we have:

$$\mathbf{M}^{(i,i)}_{s,t,u} = 0$$

Thus, equation (14) simplifies to:

$$\mathbf{M}^{(i,i)}_{s,t,u} - \overline{\mathbf{M}^{(i,i)}_{2p-s,2q-t,2r-u}} = -\overline{\mathbf{M}^{(i,i)}_{2p-s,2q-t,2r-u}}$$

Since ($s \geq p+1$) or ($s = p,\ t \geq q+1$) or ($s = p,\ t = q,\ u \geq r+1$), we have: ($2p-s \leq p-1$) or ($2p-s = p,\ 2q-t \leq q-1$) or ($2p-s = p,\ 2q-t = q,\ 2u-r \leq r-1$) respectively. Thus using equation (35), we have:

$$-\overline{\mathbf{M}^{(i,i)}_{2p-s,2q-t,2r-u}} = -\overline{\mathbf{L}^{(i,i)}_{2p-s,2q-t,2r-u}}$$

Since $\mathbf{L}^{(i,i)}$ is a skew-hermitian filter, we have from Theorem 6:

$$\mathbf{L}^{(i,i)}_{s,t,u} = -\overline{\mathbf{L}^{(i,i)}_{2p-s,2q-t,2r-u}}$$

Thus in this subcase, equation (37) simplifies to $\mathbf{L}^{(i,i)}_{s,t,u}$ again.

Subcase 3(c): For $s = p,\ t = q,\ u = r$, since $\mathbf{L}^{(i,i)}$ is a skew-hermitian filter, we have:

$$\mathbf{L}^{(i,i)}_{p,q,r} = -\overline{\mathbf{L}^{(i,i)}_{p,q,r}}$$
$$\mathbf{L}^{(i,i)}_{p,q,r} + \overline{\mathbf{L}^{(i,i)}_{p,q,r}} = 0$$

Thus, $\mathbf{L}^{(i,i)}_{p,q,r}$ is a purely imaginary number. In this subcase

$$\mathbf{M}^{(i,i)}_{s,t,u} - \overline{\mathbf{M}^{(i,i)}_{2p-s,2q-t,2r-u}}$$
$$= \mathbf{M}^{(i,i)}_{p,q,r} - \overline{\mathbf{M}^{(i,i)}_{p,q,r}} = 2\mathbf{M}^{(i,i)}_{p,q,r}$$

Using equation (35) , we have:

$$2\mathbf{M}^{(i,i)}_{p,q,r} = \mathbf{L}^{(i,i)}_{p,q,r}$$

Thus, we get:

$$\mathbf{M}^{(i,i)}_{p,q,r} - \left[\text{conv3d\_transpose}\left(\mathbf{M}^{(i,i)}\right)\right]_{p,q,r} = \mathbf{L}^{(i,i)}_{p,q,r}$$

Thus we have established: $\mathbf{L} = \mathbf{M} - \text{conv3d\_transpose}(\mathbf{M})$. Note that the opposite direction of the if and only if statement follows trivially from the above proof. □

C. MaxMin Activation function

Given a feature map $\mathbf{X} \in \mathbb{R}^{2m \times n \times n}$ (we assume the number of channels in $\mathbf{X}$ is a multiple of 2), to apply the MaxMin activation function, we first divide the input into two chunks of equal size: $\mathbf{A}$ and $\mathbf{B}$ such that:

$$\mathbf{A} = \mathbf{X}_{:m,:,:}$$
$$\mathbf{B} = \mathbf{X}_{m:,:,:}$$

Then the MaxMin activation function is given as follows:

$$\text{MaxMin}(\mathbf{X})_{:m,:,:} = \max(\mathbf{A}, \mathbf{B})$$
$$\text{MaxMin}(\mathbf{X})_{m:,:,:} = \min(\mathbf{A}, \mathbf{B})$$

D. Additional Experiments

Model	Standard Accuracy		Provably Robust Accuracy	
	BCOP-20	BCOP-30	BCOP-20	BCOP-30
LipConvnet-5	74.35%	74.93%	58.01%	58.97%
LipConvnet-10	74.47%	74.63%	58.48%	58.23%
LipConvnet-15	73.86%	74.09%	57.39%	57.42%
LipConvnet-20	69.84%	70.01%	52.10%	52.59%
LipConvnet-25	68.26%	66.66%	49.92%	47.63%
LipConvnet-30	64.11%	65.77%	43.39%	45.10%
LipConvnet-35	63.05%	63.45%	41.72%	42.41%
LipConvnet-40	60.17%	59.60%	38.87%	37.75%

Table 5. Comparing between results using BCOP with 20 (BCOP-20) and 30 (BCOP-30) Bjorck iterations for provable robustness against adversarial examples (l_2 perturbation radius of $36/255$ and CIFAR-10 dataset).

Model	BCOP	SOC
LipConvnet-5	40.34%	**42.01%**
LipConvnet-10	40.77%	**44.13%**
LipConvnet-15	39.33%	**44.24%**
LipConvnet-20	34.75%	**45.18%**
LipConvnet-25	31.99%	**43.50%**
LipConvnet-30	25.02%	**42.39%**
LipConvnet-35	23.30%	**41.75%**
LipConvnet-40	21.20%	**37.88%**

Table 6. Comparing between BCOP and SOC for provably robust accuracy using l_2 perturbation radius of $72/255$.