A. Proofs

Here, we provide proofs, and other theoretical details, left out in the the main text.

A.1. Spaces of point clouds

In the main paper, we have, in the interest of readability, intentionally refrained from being too formal. In particular, we have equated point clouds with vectors in $\mathbb{C}^m$ in a quite streamlined fashion. As we want to present formal proofs here, this will no longer suffice. In particular, since we are aiming to apply the Stone-Weierstrass theorem, we will need to consider the point clouds as points in a metric space. We will therefore consider the following well-known approach (the same ideas were applied in e.g. [20, 30].)

Definition 5. *For a subgroup $G \subseteq S_m$, let $\sim_G$ denote the equivalence relation*

$$Z \sim_G W \Leftrightarrow \exists \pi \in G, Z = \pi^* W$$

on $\mathbb{C}^m$. We can equip the set of equivalence classes $\mathbb{C}^m / \sim_G$ with the metric

$$d_G(Z, W) \ = \ \inf_{\pi \in G} ||Z - \pi^* W||.$$

For $G = S_m$, we denote the resulting metric space $\mathcal{P}^m$. For $G = \mathrm{Stab}(0)$, we denote it $\mathcal{P}^m_0$.

On $\mathcal{P}^m$ and $\mathcal{P}^m_0$, we may define a further equivalence relation via $Z \sim_{\mathbb{S}} W \Leftrightarrow Z = \theta W$ for some $\theta \in \mathbb{S}$. We can again define a metric on the set of equivalence classes under this relation via

$$d_{\mathbb{S}}(Z, W) = \inf_{\theta \in \mathbb{S}} d_G(Z, \theta W),$$

where d_G is the metric from above. We call the resulting metric spaces $\mathcal{RP}^m$ and $\mathcal{RP}^m_0$

In the following, we will without comment equip all spaces of continuous functions with the topology induced by the supremum norm on compact sets. If M is a metric space, we let $\mathcal{C}(M)$ denote the space of complex-valued continuous functions on M.

Remark 6. *(i) It is clear that permutation invariant functions $F \in \mathcal{C}(\mathbb{C}^m)$ can be identified with functions in $\mathcal{C}(\mathcal{P}^m)$. If they are additionally rotation invariant, we can even identify them with functions on $\mathcal{C}(\mathcal{RP}^m)$. Similar statements hold for $\mathcal{C}(\mathcal{P}^m_0)$ and $\mathcal{C}(\mathcal{RP}^m_0)$.*

(ii) In the following, we will sometimes consider expressions in which functions defined on $\mathcal{P}^m_0$, or $\mathbb{C}^m$, are applied to members in $Z \in \mathcal{P}^m$. This is clearly in general not formally well-defined. However, in each such expression, there are other operations present which makes the object per se well defined again. For instance, $\nu_i(Z) = |z_i|$ is not well defined on $\mathcal{P}^m$, but $\nu(Z) = \sup_{i \in [m]} |z_i|$ is. In the interest of readability, we will not comment on this in detail every time.

A.2. Proof of Proposition 1

Let us begin by proving the no-go result of Proposition 1, stating that the most straighforward way of making the pointnet architecture rotation equivariant will not yield a universal architecture.

Proof of Proposition 1. Let us call a cloud Z for which all points have the same norm and obey $\sum_{i \in [m]} z_i = 0$ *balanced*. We claim that every function of the form $\chi(\sum_{i \in [m]} \varrho(z_i))$ is constant on the set of balanced clouds.

To see this, let us first notice that if $\varrho : \mathbb{C} \to \mathbb{C}^K$ is rotation equivariant, it must be possible to write it on the form $\varrho(z) = \nu(|z|)z$ for some function $\nu : \mathbb{R}_+ \to \mathbb{C}^K$. A formal way to prove this is to notice that the function $z \mapsto \overline{z}\varrho(z)$ is rotation equivariant, and hence can only depend on the modulus of z.

Now, if r is the common value for the norms in a balanced cloud Z, we have

$$\chi\big(\sum_{i \in [m]} \varrho(z_i)\big) = \chi\big(\nu(r) \sum_{i \in [m]} z_i\big) = \chi(0).$$

Hence $\chi(Z) = \chi(0)$ for all such clouds. To finish the proof, it is therefore enough to prove the existence function $f \in \mathcal{R}(m)$ that is not constant on the set of balanced clouds.

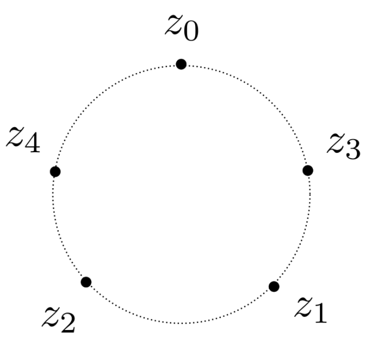

Figure 6. The balanced cloud Z used in the proof of Prop 1. Note that among the pairwise distances $|z_i - z_j|$, only $|z_1 - z_2|$ is equal to $2\sqrt{5}/3$.

Towards this endeavour, let $a : \mathbb{R} \to \mathbb{C}$ be a function and consider

$$f(Z) = \sum_{i<j\in[m]} a\left(|z_i - z_j|\right) \cdot \sum_{k\notin\{i,j\}} z_k$$

That is, in words; First, for each pair z_i, z_j of points, calculate $a(|z_i - z_j|)$ and multiply that with the sum of the rest of the points. Then sum over the set of such pairs. It is not hard to realize that such functions are members of $\mathcal{R}(m)$.

Now let a, for some $\epsilon > 0$, be equal to 1 in $2\sqrt{5}/3$ and zero outside $[2\sqrt{5}/3 - \epsilon, 2\sqrt{5}/3 + \epsilon]$. Then, if Z is a cloud with all pairwise distances smaller than $2\sqrt{5}/3 - \epsilon$, $f(Z) = 0$. There exists balanced clouds with that property for all m. Therefore, if f is constant on the set of balanced clouds, we must have $f(Z) = 0$ for all such. We can however construct a balanced cloud for which $f(Z) \neq 0$ as follows:

Let us first assume that $m = 2k + 3$ is odd. We define

$$z_1 = i, \quad z_{2,3} = -\frac{2i}{3} \pm \frac{\sqrt{5}}{3}, \quad z_{2\ell,2\ell+1} = \frac{i}{6k} \pm \frac{\sqrt{36k^2 - 1}}{6k}.$$

All points in these clouds have the norm 1, and

$$\sum_{k=1}^{m} z_k = i - \frac{4i}{3} + 2k \cdot \frac{i}{6k} = 0.$$

Note that we used that the real parts of the points cancel each other. Thus, the set is balanced. (See also Fig. 6).

Now, by calculating all distances between points, we see that $|z_1 - z_2| = 2\sqrt{5}/3$, and that all other pairwise distances $|z_i - z_j|$ for $(i, j) \neq (1, 2)$ are unequal to $2\sqrt{5}/3$. Therefore, if we choose the parameter ϵ from above small enough, we get

$$a(|z_i - z_j|) = \begin{cases} 1 & \text{if } i = 1, j = 2 \\ 0 & \text{else,} \end{cases}$$

and

$$f(Z) = \sum_{k\notin\{1,2\}} z_k = i + 2k \cdot \frac{i}{6k} = \frac{4i}{3} \neq 0.$$

Hence, f is not constant on the set of balanced clouds, and the argument is finished.

In the case of even m, we proceed as above, but interchange $z_0 = i$ with the two points $z_{-1,0} = 0.5i \pm \frac{\sqrt{3}}{2}$. The argument then proceeds just as above. □

A.3. Proof of Theorem 2

Here, we prove that functions of the form $\sum_{i\in[m]} \gamma(\tau_i^* Z) z_i$ are dense in $\mathcal{R}(m)$. Before starting the actual proof, let us agree on a simplifying notational convention. For a complex polynomial q, we will refer to the function

$$p(Z) = q(Z, \overline{Z}).$$

as a *real polynomial* in Z. Note that the set of these functions is dense in $\mathcal{C}(\mathbb{C}^m)$ with respect to supremum norm *compact sets*, n.b.. To see this, note that the classical Stone-Weierstrass theorem states that for any $N \in \mathbb{N}$ the set of real polynomials is dense in $\mathcal{C}(\mathbb{R}^n)$. By equating $\mathbb{C}^m$ with $\mathbb{R}^{2m}$, we see that the space of real polynomials *in the real and imaginary parts* of $Z \in \mathbb{C}^m$,

$$r(\mathrm{re}(Z), \mathrm{im}(Z))$$

is dense $\mathcal{C}(\mathbb{C}^m)$. Since we however for each such polynomial r can find a complex q with $r(\mathrm{re}(Z), \mathrm{im}(Z)) = q(Z, \overline{Z})$ for all Z, the claim follows.

Having established that density result we now move on to prove that in order to approximate functions in $\mathcal{R}(m)$, it is enough to consider polynomials with the same equivariance properties. Similar statements have been proven in e.g. [8, 50].

Lemma 1. *The set of real polynomials p that are permutation invariant and rotation equivariant is dense in $\mathcal{R}(m)$.*

Proof. Let us first prove that it suffices to consider rotation equivariant polynomials, we argue as follows. For some multi-indices $\alpha, \beta \in \mathbb{N}^m$, consider the 'real monomial'

$$\mu_{\alpha\beta}(Z) = Z^\alpha \overline{Z}^\beta.$$

It is clear that $\mu_{\alpha\beta}$ is rotationally equivariant if and only if $|\alpha| = |\beta| + 1$. This together with the fact that $\frac{1}{2\pi}\int_{\mathbb{S}} \theta^k \mathrm{d}\theta = \delta_{k,0}$ implies that

$$\int_{\mathbb{S}} \overline{\theta} \mu_{\alpha\beta}(\theta Z)\, \mathrm{d}\theta \neq 0 \quad \Longleftrightarrow \quad \mu_{\alpha\beta} \text{ rotationally equivariant.} \tag{3}$$

Also notice that if f is rotationally equivariant,

$$\tfrac{1}{2\pi}\int_{\mathbb{S}} \overline{\theta} f(\theta Z)\, \mathrm{d}\theta = \tfrac{1}{2\pi}\int_{\mathbb{S}} f(Z)\, \mathrm{d}\theta = f(Z).$$

Now fix a compact set $K \subseteq \mathbb{C}^m$, which without loss of generality has the property $Z \in K \Leftrightarrow \theta Z \in K$, $\theta \in \mathbb{S}$. For every $f \in \mathcal{R}(m)$, there exists a real polynomial p with $\sup_{Z\in K} |p(Z) - f(Z)| \leq \epsilon$. We now split the monomial terms in p according to whether they are rotationally equivariant or not. This defines two polynomials p_0 and p_1. Now notice that for each $Z \in K$,

$$|f(Z) - p_0(Z)| = \left| \tfrac{1}{2\pi}\int_{\mathbb{S}} \overline{\theta}(f(\theta Z) - p_0(\theta Z) - p_1(\theta Z))\, \mathrm{d}\theta \right| \leq \sup_{Z\in K} |f(Z) - p(Z)|$$

We used that p_0 and f are rotationally equivariant, and also (3) together with the fact that p_1 only consists of monomial terms that are not rotationally equviarant. This means that the rotationally equivariant real polynomial $p_0 \in \mathcal{R}(m)$ has a supremum distance at most ϵ to f on K, and we hence we might as well use q to approximate f.

The permutation invariance part is now easily handeled by symmetrization. That is if, p is a non-symmetric polynomial approximating f well, the symmetric polynomial

$$\widehat{p}(Z) = \tfrac{1}{|S_m|} \sum_{\pi\in S_m} p(\pi^* Z)$$

will approximate f just as good – see for instance [30]. □

With the previous lemma in our toolbox, the proof of Theorem 2 is relatively simple.

Proof of Theorem (2). Fix a compact set and a function f. By Lemma 1, there exists a real, symmetric and rotation equivariant polynomial

$$p(Z) = \sum_{\alpha,\beta} c_{\alpha,\beta} Z^{\alpha} \overline{Z}^{\beta},$$

which is close to f. Since p is rotation invariant, it must be $c_{\alpha,\beta} = 0$ for all (α,β) with $|\alpha| \neq |\beta| + 1$. Due to its permutation invariance, we furthermore have $c_{\alpha,\beta} = c_{\pi^*\alpha,\pi^*\beta}$ for all $\pi \in S_m$ and multiindices α, β. Hence, p consists of terms of the form

$$q_{\alpha,\beta}(Z) = \sum_{\pi \in S_m} Z^{\pi^*\alpha} \overline{Z}^{\pi^*\beta}, \ |\alpha| = |\beta| + 1., \tag{4}$$

and it is therefore enough to approximate such terms. Here, by the permutation equivariance, we can WLOG assume that the indices α_i are in ascending order. Consequently, we can write $\alpha = \hat{\alpha} + e_0$ for some $\hat{\alpha}$ with $|\hat{\alpha}| = |\beta|$.

Now let us split the sum in (4) over S_m in accordance to the value of $\pi(0)$

$$q_{\alpha,\beta}(Z) = \sum_{i \in [m]} \sum_{\pi(0)=i} Z^{\pi^*(e_0+\hat{\alpha})} \overline{Z}^{\pi^*\beta} = \sum_{i \in [m]} \sum_{\pi(i)=0} Z^{\pi^*\hat{\alpha}} \overline{Z}^{\pi^*\beta} z_i, \tag{5}$$

where we in the last step used that $\pi^* e_0 = e_{\pi(0)} = e_i$. It is clear that we can write each π with $\pi(0) = i$ as $\tau_i \circ \sigma$ for a unique $\sigma \in \mathrm{Stab}(0)$. We have

$$Z^{\pi^*\hat{\alpha}} \overline{Z}^{\pi^*\beta} = Z^{\tau_i^*\sigma^*\hat{\alpha}} \overline{Z}^{\tau_i^*\sigma^*\hat{\beta}} z_i$$

Since $(\tau_i^* Z)^{\alpha} = Z^{\tau_i^*\alpha}$, we see that our sum turns into

$$\sum_{i \in [m]} \sum_{\sigma \in \mathrm{Stab}(0)} (\tau_i^* Z)^{\sigma^*\hat{\alpha}} (\tau_i^* \overline{Z})^{\sigma^*\beta} z_i = \sum_{i \in [m]} \gamma(\tau_i^* Z) z_i,$$

where we defined

$$\gamma(Z) = \sum_{\sigma \in \mathrm{Stab}(0)} Z^{\sigma^*\hat{\alpha}} \overline{Z}^{\sigma^*\beta}$$

The function γ is clearly $\mathrm{Stab}(0)$-invariant, and also rotation invariant due to $|\hat{\alpha}| = |\beta|$. The proof is finished. □

A.4. $\mathrm{Stab}(0)$-equi- and invariant linear maps

Our architectures make heavy use of linear layers which are equi- and invariant to the action of the $\mathrm{Stab}(0)$ group. It is a priori not clear how to construct such, and in particular parametrize all of them. In this section, we provide such a description.

We let $\mathbb{K}$ denote either of the fields $\mathbb{R}$ or $\mathbb{C}$. For a tensor $T \in (\mathbb{K}^m)^{\otimes k}$, i.e. of order k, we define the action of a permutation $\pi \in S_m$ on T through

$$(\pi^* T)_{i_0,\dots,i_{k-1}} = T_{\pi^{-1}(i_0),\dots,\pi^{-1}(i_{k-1})}.$$

This is exactly as in [29]. Let us begin by introducing some notation for the spaces we are interested in.

Definition 7. *For $k, \ell \in \mathbb{N}$, we let $\mathcal{L}(k,\ell)$ denote the space of linear operators $L : (\mathbb{C}^m)^{\otimes k} \to (\mathbb{C}^m)^{\otimes \ell}$ which are S_m-equivariant. The space of operators of the same kind which are* $\mathrm{Stab}(0)$*-equivariant is denoted $\mathcal{L}_0(k,\ell)$.*

Let us briefly comment on two special cases. First, if $\ell = 0$, the spaces $\mathcal{L}(k,0)$ and $\mathcal{L}_0(k,0)$ can be identified with the space of invariant functionals of the respective kind. This is because of the fact that the action of S_m on scalars $v \in \mathbb{K}$ is trivial. In the same manner, the spaces $\mathcal{L}(0,k)$ and $\mathcal{L}_0(0,k)$ denote constant k-tensors which are invariant to the action of the respective groups. Such elements can be used as biases in our architecture.

Remark 8. *In our architecture, we are actually dealing with linear layers mapping multi-tensors to multi-tensors. It is however clear that such a mapping can be seen as a matrix of linear maps L_{ij}, where each L_{ij} corresponds to one input-output-channel pair. As such, it is enough to characterize the spaces $\mathcal{L}(k,\ell)$ and $\mathcal{L}_0(k,\ell)$ to obtain a way to parametrize the linear layers of our architecture.*

Let us reiterate that the results of [29] give a complete characterization of the spaces $\mathcal{L}(k,\ell)$[7]. In brief, they identify such maps as fixed points of a certain linear equation, which they then explicitly calculate. We refer to [29] for details.

In particular, the results in the mentioned paper prove that $\dim \mathcal{L}(k,\ell) \leq B_{k+\ell}$, where B_n denotes the n:th *Bell number*. As noted in [11], the dimension of the space cannot get larger than the dimension of the space of all linear maps from $(\mathbb{K}^m)^{\otimes k}$ to $(\mathbb{K}^m)^{\otimes \ell}$, which is $m^{k+\ell}$. In all cases, the number of scalars needed to describe a map in $\mathcal{L}(k,\ell)$ can be bounded independent of m.

Our idea here is to link the spaces $\mathcal{L}_0(k,\ell)$ with spaces $\mathcal{L}(k',\ell')$. In doing so, the following simple Lemma will be convenient . For completeness. we include a proof.

Lemma 2. *For k, ℓ in N, consider the map*

$$\Phi_{k,\ell} : L \mapsto \lambda, \quad \lambda(\overline{S} \otimes T) = \langle S, L(T) \rangle, \; T \in (\mathbb{K}^m)^{\otimes k}, S \in (\mathbb{K}^m) \otimes \ell.$$

Hereby, $\langle \cdot, \cdot \rangle$ denotes the canonical scalar product on $(\mathbb{K}^m)^{\otimes \ell}$, i.e.

$$\langle M, N \rangle = \sum_{i_0,\ldots,i_{\ell-1}} \overline{M_{i_0,\ldots,i_{\ell-1}}} N_{i_0,\ldots,i_{\ell-1}}$$

(i) $\Phi_{k,\ell}$ is an isomorphism between the spaces of linear maps $(\mathbb{K}^m)^{\otimes k} \to (\mathbb{K}^m)^{\otimes \ell}$ and functionals on $(\mathbb{K}^m)^{\otimes (k+\ell)}$.

(ii) $\Phi_{k\ell}$ maps $\mathcal{L}(k,\ell)$ to $\mathcal{L}(k+\ell,0)$ and $\mathcal{L}_0(k,\ell)$ to $\mathcal{L}_0(k+\ell,0)$. In particular, the respective pairs of spaces are isomorphic.

Proof. To not overload the notation, we fix k and ℓ and drop the index on Φ.

Ad (i): The linearity is evident. For proving the injectivity, suppose that $\lambda = \Phi(L)$ is the zero functional. That means per definition that $\langle S, L(T) \rangle = 0$ for all $S \in (\mathbb{K}^m)^{\otimes \ell}$, which implies that $L(T) = 0$ for all T in $(\mathbb{K}^m)^{\otimes k}$, i.e. that $L = 0$. The surjectivity now follows from dimensionality considerations.

Ad (ii) We concentrate on the case of S_m-equivariant maps, since the $\mathrm{Stab}(0)$-case is proven in exactly the same way. We need to prove two things: First, we need to show that $\Phi(L) \in \mathcal{L}(k+\ell, 0)$ for all $L \in \mathcal{L}(k,\ell)$. Secondly, we need to show that for every $\lambda \in \mathcal{L}(k+\ell,0)$, the (unique) L with $\Phi(L) = \lambda$ is in $\mathcal{L}(k,\ell)$.

To prove the former, let $L \in \mathcal{L}(k,\ell)$ and $\pi \in S_m$ be arbitrary. Writing $\lambda = \Phi(L)$, we have

$$\lambda(\pi^*(\overline{S} \otimes T)) = \langle \pi^* S, L(\pi^* T) \rangle = \langle \pi^* S, \pi^* L(T) \rangle = \langle S, L(T) \rangle = \lambda(\overline{S} \otimes T),$$

for each S and T. Note that we used the equivariance of L in the second step, and the (obvious) invariance of the scalar product under permutations in the third. This exactly means that $\lambda \in \mathcal{L}(k+\ell, 0)$.

To prove the latter, let $\lambda \in \mathcal{L}(k+\ell,0)$, $L = \Phi^{-1}(\lambda)$ and $\pi \in S_m$ arbitrary. For S and T arbitrary, defining $R = (\pi^{-1})^* S$, we then get

$$\langle S, L(\pi^* T) \rangle = \langle \pi^* R, L(\pi^* T) \rangle = \lambda(\pi^*(\overline{R} \otimes T)) = \lambda(\overline{R} \otimes T) = \langle R, L(T) \rangle = \langle \pi^* R, \pi^* L(T) \rangle = \langle S, \pi^* L(T) \rangle.$$

We used the invariance of λ in the third step, and the invariance of the scalar product in the fifth. Since S is arbitrary, this proves that $L(\pi^* T) = \pi^* L(T)$ for all T, i.e., $L \in \mathcal{L}(k,\ell)$ □

The above lemma links spaces of equivariant linear maps to spaces of invariant functionals, in an isomorphic fashion. This means that in order to link the spaces $\mathcal{L}(k,\ell)$ to the spaces $\mathcal{L}_0(k,\ell)$, it suffices to provide a link between one space of functionals of the one kind to a space of equivariant maps of the other. This is the purpose of the following theorem.

Theorem 9. *The map*

$$\Psi : \mathcal{L}(k,1) \to \mathcal{L}_0(k,0), L \mapsto \lambda, \quad \lambda(T) = \langle e_0, L(T) \rangle$$

is an isomorphism. In particular, $\mathcal{L}_0(k,0) \simeq \mathcal{L}(k,1)$ and $\dim(\mathcal{L}_0(k,0)) = B_{k+1}$.

[7] Technically, they only state their theorems in the case $\mathbb{K} = \mathbb{R}$, but their proofs go through also for $\mathbb{K} = \mathbb{C}$

Proof. Let us begin by proving that Ψ is well-defined, i.e. that $\Psi(L) \in \mathcal{L}_0(k,0)$ for each $L \in \mathcal{L}(k,1)$. Let $\sigma \in \mathrm{Stab}(0)$ be arbitrary. Due to the equivariance of L and invariance of the scalar product, we then get

$$\lambda(\sigma^* T) = \langle e_0, L(\sigma^* T)\rangle = \langle e_0, \sigma^* L(T)\rangle = \langle (\sigma^{-1})^* e_0, L(T)\rangle = \langle e_0, L(T)\rangle = \lambda(T).$$

In the penultimate step, we used that $(\sigma^{-1})^* e_0 = e_{\sigma^{-1}(0)} = e_0$ for $\sigma \in \mathrm{Stab}(0)$. This means that λ is invariant, and that Ψ indeed is well defined.

Now for the isomorphy. It is clear that Ψ is linear. To prove injectivity, assume that $\lambda = \Psi(L) = 0$. Due to the equivariance of L, we then get for every $i \in [m]$ and $T \in (\mathbb{K}^m)^k$

$$0 = \lambda(\tau_i^* T) = \langle e_0, L(\tau_i^* T)\rangle = \langle \tau_i^* e_0, L(T)\rangle = \langle e_i, L(T)\rangle.$$

i.e. $L = 0$. To show surjectivity, let $\lambda \in \mathcal{L}_0(k,0)$ be arbitrary. Define a map $L : (\mathbb{K}^m)^{\otimes k} \to \mathbb{K}$ through

$$\langle e_i, L(T)\rangle = \lambda(\tau_i^* T), \quad i \in [m]$$

We then have $\Psi(L)(T) = \langle e_0, L(T)\rangle = \lambda(\tau_0^* T) = \lambda(T)$, i.e., $\lambda = \Psi(L)$. If we can prove that L is equivariant, we are done. So let $\pi \in S_m$ and $i \in [m]$ be arbitrary. A direct computation shows that $\tau_i \circ \pi \circ \tau_{\pi^{-1}(i)} \in \mathrm{Stab}(0)$. This, together with the assumed invariance of λ, shows that

$$\begin{aligned}\langle e_i, L(\pi^* T)\rangle &= \lambda(\tau_i^* \pi^* T) = \lambda(\tau_i^* \pi^* \tau_{\pi^{-1}(i)}^* \tau_{\pi^{-1}(i)}^* T) = \lambda(\tau_{\pi^{-1}(i)}^* T) = \langle e_{\pi^{-1}(i)}, L(T)\rangle \\ &= \langle (\pi^{-1})^* e_i, L(T)\rangle = \langle e_i, \pi^* L(T)\rangle.\end{aligned}$$

Since i is arbitrary, this means that $L(\pi^* T) = \pi^* L(T)$, i.e., that L is equivariant. The proof is finished. □

We can now use Lemma 2 and Theorem 9 to construct an isomorphism between $\mathcal{L}_0(k,\ell)$ and $\mathcal{L}(k,\ell+1)$

Corollary 1. $\mathcal{L}(k,\ell+1) \simeq \mathcal{L}_0(k,\ell)$. *An isomorphism is given by*

$$\Xi : \mathcal{L}_0(k,\ell) \to \mathcal{L}(k,\ell+1), L_0 \mapsto K, \quad K(T) = \sum_{i\in[m]} e_i \otimes \tau_i^* L_0(\tau_i^* T).$$

Proof. If $\Phi_{k,\ell}$ and Ψ are as in Lemma 2 and Theorem 9, respectively, we define the isomorphism Ξ through the following chain

$$\begin{tikzcd} \mathcal{L}_0(k,\ell) \arrow[r, "\Phi_{k,\ell}"] & \mathcal{L}_0(k+\ell,0) \arrow[r, "\Psi^{-1}"] & \mathcal{L}(k+\ell,1) \arrow[r, "\Phi_{k+\ell,1}"] & \mathcal{L}(k+\ell+1,0) \arrow[r, "\Phi_{k,\ell+1}^{-1}"] & \mathcal{L}(k,\ell+1) \\ L_0 & \lambda_0 & L & \lambda & K \end{tikzcd}.$$

It now only is left to prove that Ξ has the claimed form. For convenience, we named all of the intermediate objects above. For $u \in \mathbb{K}^m$, $S \in (\mathbb{K}^m)^{\otimes \ell}$ and $T \in (\mathbb{K}^m)^{\otimes k}$, we calculate

$$\langle u \otimes S, K(T)\rangle = \lambda(\overline{u} \otimes \overline{S} \otimes T) = \langle u, L(\overline{S} \otimes T)\rangle = \sum_{i\in[m]} \overline{u}_i \langle e_i, L(\overline{S} \otimes T)\rangle.$$

Now, notice that since $L \in \mathcal{L}(k+\ell,1)$ and the scalar product is S_m-invariant, we have

$$\langle e_i, L(\overline{S} \otimes T)\rangle = \langle \tau_i^* e_0, L(\overline{S} \otimes T)\rangle = \langle e_0, \tau_i^* L(\overline{S} \otimes T)\rangle = \langle e_0, L(\tau_i^*(\overline{S} \otimes T))\rangle.$$

Consequently ,

$$\begin{aligned}\sum_{i\in[m]} \overline{u}_i \langle e_i, L(\overline{S} \otimes T)\rangle &= \sum_{i\in[m]} \overline{u_i} \langle e_0, L(\tau_i^*(\overline{S} \otimes T))\rangle = \sum_{i\in[m]} \overline{u_i} \lambda_0(\tau_i^* \overline{S} \otimes \tau_i^* T) = \sum_{i\in[m]} \overline{u_i} \langle \tau_i^* S, L_0(\tau_i^* T)\rangle \\ &= \sum_{i\in[m]} \overline{u_i} \langle S, \tau_i^* L_0(\tau_i^* T)\rangle = \sum_{i\in[m]} \langle u, e_i\rangle \langle S, \tau_i^* L_0(\tau_i^* T)\rangle = \langle u \otimes S, \sum_{i\in[m]} e_i \otimes \tau_i^* L_0(\tau_i^* T)\rangle.\end{aligned}$$

Since u and S are arbitrary, we obtain the claim. □

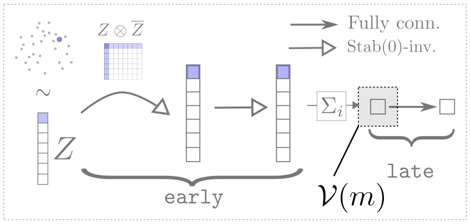

Figure 7. Definition of the space $\mathcal{V}(m)$.

We may now easily construct spanning systems of $\mathcal{L}_0(k,\ell)$ by, using the last corollary, transforming the spanning sets of $\mathcal{L}(k,\ell+1)$ from [29]. In Section C, we carry this out and write down explicit spanning sets for the spaces $\mathcal{L}_0(k,\ell)$ for $0 \leq k,\ell \leq 2$.

Let us here only comment that the above Corollary in particular proves that $\dim \mathcal{L}_0(k,\ell) = \dim \mathcal{L}(k,\ell+1) \leq B_{k+\ell+1}$. In particular, we may describe each linear input-output channel pair of the first layer of our weight units (which is an element of $\mathcal{L}_0(2,1)$) with $B_{2+1+1} = 15$ parameters, and each channel of the bias (which is an element of $\mathcal{L}_0(0,1)$) with $B_{1+1} = 2$ parameters. For the later early layers, we need $B_{1+1+1} = 5$ parameters per input-output-channel pair for the linear part (which is then an element of $\mathcal{L}_0(1,1)$) , and $B_{1+1} = 2$ parameter per output channel bias (which is still an element of $\mathcal{L}_0(0,1)$) .

A.5. Proof of Theorem 3

We now prove the main result. Note that we have to assume that the activation function in the weight units is not a polynomial (in order to be able to apply the classical universality result for neural networks [6].) The first step is to prove that the $\mathcal{NS}(m)$-architecture, i.e. the ones for the weight units is universal for functions restricted to a subset of $\mathcal{RP}_0^m$.

Lemma 3. *For $\epsilon > 0$, define the set*

$$C_\epsilon^m = \{Z \in \mathcal{RP}_0^m \,|\, |z_0| > \epsilon\} \,.$$

Then, $\mathcal{NS}(m)$ is dense in $\mathcal{C}(C_\epsilon^m)$.

Proof. We aim to apply the the Stone-Weierstrass theorem [35, Th.7.32]. This theorem says that if a set S of continuous functions defined on a compact metric space M

- separates points, e.g. if there for each $x \neq y \in M$ exists an $f \in S$ with $f(x) \neq f(y)$,
- vanishes nowhere, e.g. that there for $x \in M$ exists a $f \in S$ such that $f(x) \neq 0$,

the algebra generated by S is dense in $\mathcal{C}(M)$. Note that we may use the real version of the theorem, since we are applying real-linear layers.

In our setting, we want to apply the theorem with M equal to an arbitrary compact subset of C_ϵ^m, and S equal to the functions v defined by the averaging the output of the last early layer of the networks in $\mathcal{NS}(m)$. For convenience, let us call this set $\mathcal{V}(m)$. (See also Figure 7.) Due to the classical universality result of neural networks [6], the final fully connected layers can namely generate the algebra of those functions.

That $\mathcal{V}(m)$ is nowhere vanishing is imminent, simply due to the fact that the linear layers have biases. Thus, we can concentrate on proving that it separates points. So let $Z \neq W \in C_\epsilon^m$. We aim to show that if $v(Z) = v(W)$ for all functions in $\mathcal{V}(m)$, Z must be equal to W as points in $\mathcal{RP}_0^m$, i.e., up to a $\mathrm{Stab}(0)$-permutation and global rotation. For convenience, let us introduce the notations $Z_\wedge = (0, z_1, \ldots, z_{m-1})$ and $Z_\vee = (z_1, \ldots, z_{m-1})$

Claim 1: $|z_0| = |w_0|$. The map $T \mapsto e_0 T_{00}$ is a member of $\mathcal{L}_0(2,1)$. This can be seen through a direct calculation (see also Section C.) Therefore, channels of the first layer of α can be chosen to output $|z_0|^2 e_0$. By choosing the subsequent input-output-channel pairs as multiples of the identity, it can therefore be achieved that the output of the L:th layer can be made equal to $\psi(|z_0|^2)e_0$ for some neural network, which surely can be designed to be arbitrarily close to the identity (we hereby again appeal to the classical universality result). This vector is of course summed to (something arbitrarily close to) $|z_0|^2$. Hence, $|z_0|^2$ can be approximated arbitrarily well with functions in $\mathcal{V}(m)$, and consequently, $|z_0| = |w_0|$.

Claim 2: $z_0\overline{Z_\vee} = w_0\overline{W_\vee}$ up to a permutation. Now we use that the maps $T \mapsto Te_0$ and $T \mapsto T^Te_0$ are members of $\mathcal{L}_0(2,1)$ (This can again be realized through a direct calculation, or a consultation of Section C). Since we apply such functions on $Z \otimes \overline{Z}$ in the very first layer of α, channels of its output can be chosen equal to output $z_0\overline{Z}$ and $\overline{z_0}Z$. By subtracting the map $|z_0|^2e_0$ from above, we may even make them equal to $z_0\overline{Z_\wedge}$ and $\overline{z_0}Z_\wedge$. By taking linear combinations of those two, we may hence make the very first layer equal

$$Y_\lambda = \mathrm{re}(z_0\overline{Z_\wedge}) + \lambda\mathrm{im}(\overline{z_0}Z_\wedge).$$

Now, by letting each input-output-channel of the subsequent layers be a multiple of the identity, we can see to it that the output of the L:th layer is equal to $\psi(Y_\lambda)$, where $\psi : \mathbb{C} \to \mathbb{C}$ is any neural network applied pointwise. By the classical universality result, we can in particular make it arbitrarily close to $(Y_\lambda)^k$ for any $k \in \mathbb{N}$. These vectors are averaged to the so called *powersum polynomials* in Y_λ, i.e.

$$ps_k(Y_\lambda) = \sum_{i\geq 1}(Y_\lambda)_i^k$$

These polynomials are, of course, exactly equal to the powersum polynomials in

$$Y_\lambda^\vee = \mathrm{re}(z_0\overline{Z_\vee}) + \lambda\mathrm{im}(\overline{z_0}Z_\vee).$$

Let us correspondingly write $X_\lambda^\vee = \mathrm{re}(w_0\overline{W_\vee}) + \lambda\mathrm{im}(\overline{w_0}W_\vee)$ Since the set of powersum polynomials generate the algebra of symmetrical polynomials [9], which in turn are dense in $\mathcal{C}(\mathcal{P}^m)$, we conclude (due to Urysohn's separation lemma) that if $v(Z) = v(W)$ for all $v \in \mathcal{V}(m)$, there must for every lambda be $Y_\lambda^\vee = X_\lambda^\vee$ as points in $\mathcal{P}^{m-1}$, i.e. up to a permutation π_λ

$$Y_\lambda^\vee = \pi_\lambda^* X_\lambda^\vee. \tag{6}$$

Now, simply because S_m is finite, there must exist a π_0 and a sequence $\lambda_n \to 0$ with $\pi_{\lambda_n} = \pi_0$ for all 0. Inserting $\lambda = \lambda_n$ into equation (6) and letting $\lambda \to \infty$ we get, since π_0^* is continuous, that

$$\mathrm{re}(z_0\overline{Z_\vee}) = \pi_0^*\mathrm{re}(w_0\overline{W_\vee}).$$

By subsequently inserting a small but non-zero λ_n into (6) and subtracting $\mathrm{re}(z_0\overline{Z_\vee}) = \pi_0^*\mathrm{re}(w_0\overline{W_\vee})$ from both sides, we obtain

$$\lambda_n\mathrm{im}(z_0\overline{Z}_\vee) = \lambda_n\pi_0^*\mathrm{im}(w_0\overline{W_\vee}) \Rightarrow \mathrm{im}(z_0\overline{Z_\vee}) = \pi_0^*\mathrm{im}(w_0\overline{W_\vee}).$$

Hence, $z_0\overline{Z_\vee} = w_0\overline{W_\vee}$ up to a permutation, as claimed.

Claim 3: $Z = W$. Since $|z_0| = |w_0|$, we must have $z_0 = \theta w_0$ for some $\theta \in \mathbb{S}$. By inserting this into Claim 2 and dividing by $w_0 \neq 0$ (which is true due to $W \in C_\epsilon$), we get that $\theta\overline{Z_\vee}$ equals $\overline{W_\vee}$ up to a permutation. By conjugating that equality, and using that $\theta^{-1} = \overline{\theta}$, we get $Z_\vee = \theta W_\vee$ up to a permutation. This together with $z_0 = \theta w_0$ however exactly means that $Z = W$ as points in $\mathcal{RP}_0(m)$.

The claim now follows from Stone-Weierstrass.

□

The previous lemma shows that $\mathcal{NS}(m)$ is capable of approximating the function γ in Theorem 2 to arbitrary precision, as long as cases where z_0 is close to the origin is ignored. In order to handle also cases in which z_0 is zero, we need to choose the vector unit ψ in a certain manner. This is what the following, simple, lemma is for.

Lemma 4. *Let $\epsilon > 0$. There exists a function $s \in \mathcal{NC}$ which vanishes for $|z| < \epsilon$, equals z for $|z| > 2\epsilon$, and satisfies $|s(z)| \leq |z|$ everywhere.*

Proof. One easily realizes that

$$n(t) = \tfrac{1}{\epsilon}\left(\mathrm{ReLU}(t-\epsilon) - \mathrm{ReLU}(t-2\epsilon)\right) \begin{cases} 0 & \text{if } |z| < \epsilon \\ \frac{t-\epsilon}{\epsilon} & \text{if } \epsilon \leq t < 2\epsilon \\ 1 & \text{else.} \end{cases}$$

$$m(t) = \tfrac{1}{2}\left(\mathrm{ReLU}(t-\epsilon) + \mathrm{ReLU}(t-2\epsilon)\right) \begin{cases} 0 & \text{if } |z| < \epsilon \\ \frac{t-\epsilon}{2} & \text{if } \epsilon \leq t < 2\epsilon \\ t - \frac{3}{2}\epsilon & \text{else.} \end{cases}$$

If follows that $m(t) + \frac{3}{2}\epsilon n(t)$ equals zero for $t < \epsilon$, equals t for $t > 2\epsilon$, and is smaller than t for all $t \geq 0$. Consequently,

$$s(z) = \left(m(z) + \tfrac{3}{2}\epsilon n(z)\right)\tfrac{z}{|z|}$$

fulfills the requirements of the lemma and is, due to the definition of $\rho_{\mathbb{C}}$, in $\mathcal{NC}$. □

We can now prove the universality of our architecture.

Proof of Theorem 3. Fix a compact, arbitrary set $K \subseteq \mathcal{P}^m$, $\delta > 0$, and $f \in \mathcal{C}(\mathcal{RP}^m)$ arbitrary. Our goal is to show that there exists a $\Psi \in \mathcal{NR}(m)$ with $\sup_{Z \in K} |\Psi(Z) - f(Z)| \leq \delta$. For future reference, set $\omega = \sup_{Z\in K} \sup_{i \in [m]} |z_i|$.

By Theorem 2, there exists a function of the form

$$g(Z) = \sum_{i\in[m]} \gamma(\tau_i^* Z) z_i \tag{7}$$

with $\sup_{Z\in K} |f(Z) - g(Z)| < \frac{\delta}{2}$ and $\gamma \in \mathcal{C}(\mathcal{RP}_0^m)$. Write $\omega' = \sup_{Z \in K} \sup_{i \in \tau_i^*} |\gamma(\tau_i^* Z)|$, and define

$$\epsilon = \tfrac{\delta}{4(5m\omega' + 2m)}$$

Lemma 3 proves that there exists an $\alpha \in \mathcal{NS}(m)$ with

$$\sup_{Z \in C_\epsilon^m \cap K} |\alpha(Z) - \gamma(Z)| \leq \delta' := \min(\tfrac{\delta}{4m\omega}, 1).$$

Concretely, this means that

$$|\alpha(\tau_i^* Z) - \gamma(\tau_i^* Z)| \leq \delta' \text{ if } |z_i| \geq \epsilon. \tag{8}$$

Applying Lemma 4, we may further choose ψ equal to s as defined in that Lemma. Then, by definition,

$$\Psi(Z) = \sum_{i \in [m]} \alpha(\tau_i^* Z) s(z_i) \in \mathcal{NR}(m).$$

We now have

$$|\Psi(Z) - g(Z)| \leq \underbrace{\sum_{i:|z_i|<\epsilon} |\alpha(\tau_i^* Z)s(z_i) - \gamma(\tau_i^* Z) z_i|}_{(I)} + \underbrace{\sum_{i:\epsilon \leq |z_i| < 2\epsilon} |\alpha(\tau_i^* Z)s(z_i) - \gamma(\tau_i^* Z) z_i|}_{(II)}$$
$$+ \underbrace{\sum_{i:2\epsilon\leq|z_i|} |\alpha(\tau_i^* Z)s(z_i) - \gamma(\tau_i^* Z) z_i|}_{(III)}.$$

Let us discuss each of these terms these terms separately.

(I) For this terms, we have $s(z_i) = 0$, and z_i is small. Therefore,

$$(I) = \sum_{i:|z_i|<\epsilon} |\gamma(\tau_i^* Z) z_i| \leq m\omega'\epsilon.$$

$\underline{(III)}$ On this set, $s(z_i) = z_i$. Therefore

$$(III) = \sum_{i:2\epsilon \leq |z_i|} |\alpha(\tau_i^* Z) - \gamma(\tau_i^* Z)||z_i| \leq m\delta'\omega,$$

due to (8).

$\underline{(II)}$ For these i, we have $|s(z_i) - z_i| \leq |s(z_i)| + |z_i| \leq 4\epsilon$, and $|s(z_i)| \leq |z_i| \leq 2\epsilon$. Again using (8), we consequently obtain

$$(II) \leq \sum_{i:\epsilon \leq |z_i| < 2\epsilon} |\alpha(\tau_i^* Z) - \gamma(\tau_i^* Z)||s(z_i)| + |\gamma(\tau_i^* Z)||s(z_i) - z_i| \leq 2m\delta'\epsilon + 4m\omega'\epsilon \leq m(2 + 4\omega')\epsilon$$

Using the above three estimates, and our definition of δ' and ϵ, we obtain

$$|\Psi(Z) - g(Z)| \leq \epsilon(5m\omega' + 2m) + \delta' m\omega$$

The proof is finished. □

A.6. Proof of Proposition 4

Here, we prove that the networks in $\mathcal{NR}^+(m)$ are rotation equivariant and permutation invariant, and that the set of them includes the networks in $\mathcal{NR}(m)$.

Proof of Proposition 4. (i). It is clear that each $\alpha^+ \in \mathcal{NS}^+(m)$ still is rotation invariant(this follows from the transition to $Z \otimes \overline{Z}$ in the very first step) and that each $\psi^+ \in \mathcal{NC}^+(m)$ still is rotation equivariant (this follows from the fact that $\mathbb{C}$-linear maps and $\rho_\mathbb{C}$ both are). Since all of the linear layers are permutation equivariant, and all nonlinearities are applied pointwise, it also obvious that they are both permutation equivariant. Because of this,

$$\begin{aligned}\Psi^+(\theta\pi^* Z) &= \sum_{i\in[m]} \alpha^+(\theta\pi^* Z)_i \cdot \psi^*(\theta\pi^* Z)_i = \sum_{i\in[m]} \alpha^+(Z)_{\pi^{-1}(i)} \cdot \theta\psi^+(\pi^* Z)_{\pi^{-1}(i)} = \left\lceil k = \pi^{-1}(i) \right\rceil \\ &= \theta \cdot \sum_{k\in[m]} \alpha^+(Z)_k \cdot \psi^+(\pi^* Z)_k = \theta \cdot \Psi^+(Z),\end{aligned}$$

i.e. $\Psi^+ \in \mathcal{C}(\mathcal{PR}(m))$.

(ii) First, by choosing all input-output-channel pairs in the linear layers of ψ^+ as multiples of the identity, we can for any $\psi \in \mathcal{NC}(m)$ achieve $\psi^+(Z)_i = \psi(z_i)$, $i \in [m]$. We may hence concentrate our efforts of proving that for any $\alpha \in \mathcal{NS}(m)$, it is possible to choose the S_m-invariant layers of an $\alpha^+ \in \mathcal{NS}^+(m)$ such that $\alpha^+(Z) = \alpha(\tau_i^* Z)_i$, $i \in [m]$. We do this in three steps.

Step 1: We claim that there for each first linear layer B_0 of an $\alpha \in \mathcal{NS}(m)$ exists a first linear layer B_0^+ of an $\alpha^+ \in \mathcal{NS}^+(m)$ with

$$B_0^+(T) = \sum_{i\in[m]} e_i \otimes \tau_i^* B_0(\tau_i^* T),$$

where B_0 is a linear layer of an α-unit. It is enough to prove that this is true for each input-output-channel pair of the linear layer. However, this is exactly the statement of Corollary (1).

Step 2: Now we claim that for each subsequent linear layer B_0 of an α, there exists a corresponding linear layer B_0^+ of an α^+ so that

$$B_0^+\left(\sum_{i\in[m]} e_i \otimes v_i\right) = \sum_{i\in[m]} e_i \otimes \tau_i^* B_0(\tau_i^* v_i)$$

It is again enough to prove this for each input-output-channel pair. Each such in B_0 is a map $L_0 \in \mathcal{L}_0(1,1)$. Hence, it suffices to show that the the map defined by

$$\mathcal{K}\left(\sum_{i\in[m]} e_i \otimes v_i\right) = \sum_{i\in[m]} e_i \otimes \tau_i^* L_0(\tau_i^* v_i)$$

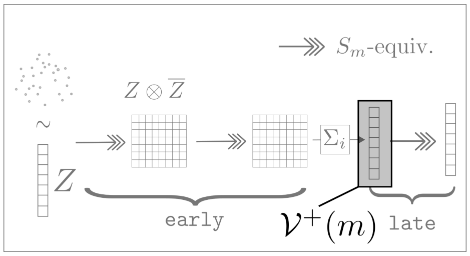

Figure 8. Definition of the space $\mathcal{V}^+(m)$.

is in $\mathcal{L}(2,2)$. To this end, let $\pi \in S_m$ be arbitrary. We have

$$\mathcal{K}(\pi^*(\sum_{i\in[m]} e_i \otimes v_i))) = \mathcal{K}(\sum_{i\in[m]} e_{\pi(i)} \otimes \pi^* v_i)) = \mathcal{K}(\sum_{i\in[m]} e_i \otimes \pi^* v_{\pi^{-1}(i)})) = \sum_{i\in[m]} e_i \otimes \tau_i^* L_0(\tau_i^* \pi^* v_{\pi^{-1}(i)}). \tag{9}$$

We performed an index shift in the second step,. Now we utilize that $\tau_i \circ \pi \circ \tau_{\pi^{-1}(i)} \in \mathrm{Stab}(0)$ to see that

$$L_0(\tau_i^* \pi^* v_{\pi^{-1}(i)}) = L_0(\tau_i^* \pi^* \tau_{\pi^{-1}(i)}^* \tau_{\pi^{-1}(i)}^* v_{\pi^{-1}(i)}) = \tau_i^* \pi^* \tau_{\pi^{-1}(i)}^* L_0(\tau_{\pi^{-1}(i)}^* v_{\pi^{-1}(i)}),$$

since L_0 is $\mathrm{Stab}(0)$-equivariant. Consequently, (9) is equal to

$$\sum_{i\in[m]} e_i \otimes \pi^* \tau_{\pi^{-1}(i)}^* L_0(\tau_{\pi^{-1}(i)}^* v_{\pi^{-1}(i)}) = \sum_{i\in[m]} e_{\pi(i)} \otimes \pi^* \tau_i^* L_0(\tau_i^* v_i) = \pi^*(\sum_{i\in[m]} e_i \otimes \tau_i^* L_0(\tau_i^* v_i)) = \pi^* \mathcal{K}(\sum_{i\in[m]} e_i \otimes v_i).$$

We again performed index shifts. Thus, $\mathcal{K}$ is S_m-equivariant, which was to be proven.

Step 3: By inductively applying Step 1 and 2, we obtain that there for every function f corresponding to the early layers of a network in $\mathcal{NS}(m)$, there exists a network in $\mathcal{NS}(m)^+$ whose first early layers have an output

$$f^+(Z) = \sum_{i\in[m]} e_i \otimes \tau_i^*(f(\tau_i^* Z)).$$

We now carry out the summation over one of the tensor dimensions of this to obtain

$$\sum_{j\in[m]} f^+(Z)_{ji} = \sum_{j\in[m]} [\tau_i^*(f(\tau_i^* Z))]_j = \lceil k = \tau_i(j) \rceil = \sum_{k\in[m]} (f(\tau_i^* Z))_k$$

Remember the definition of the space $\mathcal{V}(m)$ in the proof of Lemma 3. If we correspondingly define $\mathcal{V}^+(m)$ as the set of functions defined by summing the output of the early layers of members of $\mathcal{NS}^+(m)$-networks (see Figure 8), the above shows there for every $v \in \mathcal{V}(m)$ exists a $v^+ \in \mathcal{V}^+(m)$ with

$$v^+(Z)_i = v(\tau_i^* Z), \quad i \in [m].$$

By subsequently choosing all channels in the final layers as appropriate multiples of the identity, we can therefore achieve that $\alpha^+(Z)_i = \alpha(\tau_i^* Z)$ for all i, which was to be proven.

□

A.7. The two-cloud architecture

Here, we provide a discussion on the architectures for handling pairs of point clouds. Similarly as in the proof of the main result, we first need to equip the space of clouds of point pairs with a metric structure.

Definition 10. *For a subgroup of $G \subseteq S_m$, we let $\sim_G$ denote the equivalence relation*

$$(Z, X) \sim (W, Y) \Leftrightarrow \exists \pi \in G : (Z, X) = (\pi^* W, \pi^* Y)$$

on $\mathbb{C}^m \times \mathbb{C}^m$. We equip the set of such equivalence classes with the metric

$$d_G((Z,X),(W,Y)) = \inf_{\pi \in G} \left(||Z - \pi^* W||^2 + ||W - \pi^* Y||^2\right)^{1/2}$$

We denote the space that emerges for $G = S_m$ with $\mathcal{PP}^m$, and for $G = \mathrm{Stab}(0)$ with $\mathcal{PP}^m_0$.

On $\mathcal{PP}^m$ and $\mathcal{PP}^m_0$ we define a further equivalence relation via

$$(Z,X) \sim (W,Y) \Leftrightarrow \exists\, \theta, \omega \in \mathbb{S} : \ Z = \theta W, X = \omega Y.$$

On the resulting spaces of equivalence classes, which we denote $\mathcal{RPP}^m$ and $\mathcal{RPP}^m_0$, we define a metric through

$$d_{\mathbb{S}^2}((Z,X),(W,Y)) = \inf_{\theta,\omega \in \mathbb{S}} d((Z,X),(\theta W, \omega Y)). \tag{10}$$

Recall that $\mathcal{R}_2(m)$ was the space of functions in $\mathcal{C}(\mathcal{PP}^m)$ which were rotation equivariant with respect to the first cloud, and rotation invariant to the second, and the neural network architectures $\mathcal{NR}_2(m)$ and $\mathcal{NR}_2^+(m)$ proposed in Section 4 of the main paper.

The first result we wish to present for $\mathcal{NR}_2(m)$ is a negative one. Its proof explicitly utilizes the basis for $\mathcal{L}_2(2,1)$ provided in Section C. Hence, it might be wise to familiarize oneself with that basis before reading the proof.

Proposition 11. *$\mathcal{NR}^2(m)$ is not dense in $\mathcal{R}^2(m)$ for any $m \geq 5$.*

Proof. First, let us notice that since we only modify the architectures for calculating the weight units compared to the one-cloud case, the networks in $\mathcal{NR}_2(m)$ all have the form

$$\Psi(Z,X) = \sum_{i \in [m]} \alpha(\tau_i^* Z, \tau_i^* X)\psi(z_i).$$

with α $\mathrm{Stab}(0)$-invariant and invariant to rotations of either cloud.

Let us call clouds X with $\sum_{i \in [m]} x_i = 0$ and $x_0 = 0$ *centered*. Consider the basis $(K_i)_{i \in [15]}$ of $\mathcal{L}_0(2,1)$ described in Section C. All of their action on elements of the form $X \otimes \overline{X}$ (see in particular the final paragraph of the mentioned section) are identically zero, except for

$$K_1(X \otimes \overline{X}) = e_0 \sum_{i \in [m]} |x_i|^2, K_5(X \otimes \overline{X}) = \mathbb{1} \sum_{i \in [m]} |x_i|^2 \text{ and } K_{14}(X \otimes \overline{X}) = (|x_i|^2)_{i \in [m]}.$$

Consequently, when X is centered, the very first layer of the network, and therefore the entire value $\alpha(Z,X)$, can only depend on the norms $(|x_i|)_{i \in [m]}$ (and Z). Hence, if $X, \widetilde{X}$ are centered clouds with $|x_i| = |\tilde{x}_i|$ for all i, there must be

$$\alpha(Z,X) = \alpha(Z,\widetilde{X}) \tag{11}$$

To increase readability, let us refer to such pairs of centered clouds as *norm-equal*.

We now show that (11) leads to a contradiction. Consider functions of the form

$$f(Z,X) = \sum_i \sup_{j \neq i} \inf_{k \neq j,i} a(|x_j - x_k|) \cdot b(|z_i|)\, z_i., \tag{12}$$

where a and b are monotone functions. That is, in words: for each i, we go over all of the points x_j, $j \neq i$, and calculate the distance to nearest neighbor which is not equal to x_i. We then insert those distances into a, choose the biggest of the resulting values, and multiply it with $b(|z_i|)$ to obtain a weight for z_i to use in a weighted average. It is not hard to realize that these are in $\mathcal{R}_2(m)$.

Let us be a bit more concrete and choose b to be equal to 0 on $[0, 1/2]$ and equal to 1 on $[1, \infty[$ and a in a similar fashion be equal to 0 on $[0, 1/4]$ and equal to 1 on $[1/2, \infty]$. Now, let Z be a cloud with all points equal to 0 except for z_0, which has norm 1. We then have

$$f(Z,X) = \sup_{j \neq 0} \inf_{k \neq j,0} a(|x_j - x_k|) z_0.$$

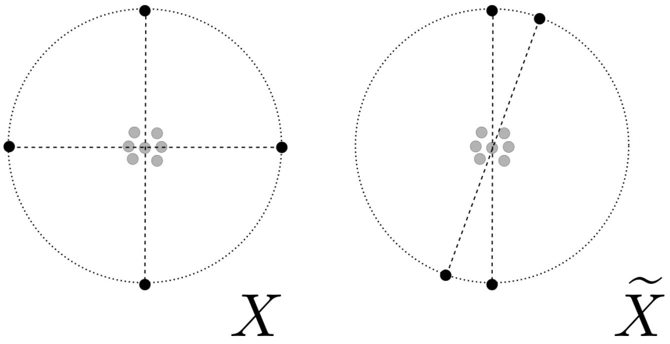

Figure 9. The norm-equal pair of centered clouds $X, \widetilde{X}$ used in the proof of Proposition 11.

Note that since both $\rho_{\mathbb{C}}$ for all $\theta > 0$ and all linear layers map 0 to 0, we must have $\psi(z_i) = 0$ for all $i \neq 0$ and $\psi \in \mathcal{NC}$. Consequently, for all $\Psi \in \mathcal{NR}_2(m)$ and Z as above, we have

$$\Psi(Z, X) = \alpha(Z, X)\psi(z_0). \tag{13}$$

Now suppose that we can construct an norm-equal pair of balanced clouds $X, \widetilde{X}$ with

(i) $|x_i| = |\tilde{x}_i| \leq \frac{1}{2}$ for all i

(ii) $\sup_{j \neq 0} \inf_{k \neq j, 0} a(|x_j - x_k|) = 1$, but $\sup_{j \neq 0} \inf_{k \neq j} a(|\tilde{x}_j - \tilde{x}_k|) = 0$,

then $f(Z, X) = z_0$, but $f(Z, \widetilde{X}) = 0$. Consquently, (11) would then imply that (13) cannot approximate (12) for both (Z, X) and $(Z, \widetilde{X})$. To see that this is possible, consider a cloud X with $x_0 = 0$, $x_{1,2} = {}^1\!/_2$, $x_{3,4} = \pm{}^i\!/_2$ and, if needed, the rest of the points arranged in a balanced fashion close to the origin. Then, X is balanced, and surely fulfills (i). We would furthermore have

$$\sup_{j \neq 0} \inf_{k \neq j, 0} a(|x_j - x_k|) \geq \inf_{k \neq 1, 0} a(|x_1 - x_k|) = 1,$$

since all points in the cloud not equal to 1 are at a distance further than ${}^1\!/_4$ from x_1. Now define $\widetilde{X}$ by letting all points in X be fixed, but rotating x_3 and x_4 using the same rotation θ (see Fig. (9)). Then, $(X, \tilde{X})$ surely is a norm-equal pair. However, we can rotate x_3 and x_4 in a fashion so that each point in $\tilde{X}$ has a nearest neighbor at a distance smaller than ${}^1\!/_4$. Consequently,

$$\sup_{j \neq 0} \inf_{k \neq j, 0} a(|\tilde{x}_j - \tilde{x}_k|) = 0.$$

This proves the proposition.

□

The last proposition shows that in order to prove a universality result, we need to restrict the set of functions we want to approximate. The following theorem describes one such possible restriction: If we are only concerned with pairs (Z, X) for which $|z_i| \lesssim |x_i|$, i.e. cloud pairs for which points close to the origin in X correspond to points close to the origin in Z, we again obtain universality

Theorem 12. *For $a > 0$, define the set*

$$D_a = \{(Z, X) \in \mathcal{PP}^m \,|\, a|z|_i \leq |x|_i \,, i \in [m]\}\,.$$

Then, both $\mathcal{NR}_2(m)$ and $\mathcal{NR}_2^+(m)$ are dense in the space of $\mathcal{C}(D_a)$-functions which are rotation-equivariant with respect to the first cloud.

Proof. The proof follows the beats of Theorem 3 very closely. We will therefore only provide a sketch, concentrating on the parts of the argument which are significantly different.

One proves $\mathcal{NR}_2(m) \subseteq \mathcal{NR}_2^+(m)$ just as the corresponding result for single cloud networks. Hence, it is enough to prove universality for $\mathcal{NR}_2(m)$. To do that, on first generalizes Theorem 2 by proving that the set of functions

$$g(Z, X) = \sum_{i \in [m]} \gamma(\tau_i^* Z, \tau_i^* X) z_i,$$

where γ is arbitrary in the space of $\mathcal{C}(\mathcal{RPP}_0^m)$, is dense in $\mathcal{R}_2(m)$ The proof is more or less verbatim equal to the proof of the $\mathcal{R}(m)$-result : One first proves that we can approximate the function using a polynomial in $\mathcal{R}^2(m)$, similarly as in Lemma 1. The proof then boils down to rewriting polynomials of the form

$$\sum_{\pi \in S_m} Z^{\pi^* \alpha_0} \overline{Z}^{\pi^* \beta_0} Z^{\pi^* \alpha_1} \overline{X}^{\pi^* \beta_1}$$

with $|\alpha_0| = |\beta_0| + 1$ and $|\alpha_1| = |\beta_1|$. It should be stressed that the last equalities are consequences of the 'separate equivariance' property.

Next, one moves on to generalizing Lemma 3. One proves that the space $\mathcal{NS}^2(m)$ of two-cloud α-units is dense in $\mathcal{C}(C_{a,\epsilon})$, where

$$C_{a,\epsilon} = \{(Z, X) \in D_a \,|\, |z_0| \geq \epsilon\}\,.$$

Note that if $(Z, X) \in C_{a,\epsilon}$, we also have $|x_0| \geq a|z_0| > 0$.

The idea of the proof is again to apply the Stone-Weierstrass theorem, with the functions $\mathcal{V}_2(m)$ that are given by outputs of α-units after the invarization step as the function set S (see the proofs of Lemma 3 and Proposition 4, as well as Figures 7 and 8). To do this, let us first note that by letting the very first layer of α only depend on either cloud, and applying the same steps as before, we get that if $v(Z, X) = v(W, Y)$ for all $v \in \mathcal{V}_2(m)$, we must have $|z_0| = |w_0|$ and $|x_0| = |y_0|$. Now notice that for every $\lambda > 0$, we can also choose the output of the very first linear layer of α equal to

$$z_0 \overline{Z_\wedge} + \lambda x_0 \overline{X_\wedge}, \quad \overline{z_0} Z_\wedge + \lambda \overline{x_0} X_\wedge,$$

using the same notation as in the previous proof. By subsequently following the same arguments as in the one-cloud proof, we see that there must be

$$z_0 \overline{Z_\vee} + \lambda x_0 \overline{X_\vee} = \pi_\lambda^* (w_0 \overline{W_\vee} + \lambda y_0 \overline{Y_\vee}) \tag{14}$$

for some permutation π_λ, possibly dependent on λ. By applying the same trick as we did to the real and imaginary parts of $z_0 \overline{Z_\vee}$ and $w_0 \overline{W_\vee}$ to conclude that they were equal to each other up to a permutation, we conclude that there exists a *common* $\pi_0 \in S_m$ with

$$z_0 \overline{Z_\vee} = \pi_0^* w_0 \overline{W_\vee}, \quad x_0 \overline{X_\vee} = \pi_0^* y_0 \overline{Y_\vee}.$$

We may now proceed as before – notice that we can divide by both z_0 and x_0, since they are both unequal to 0.

Now, the final argumentation proceeds just as in the proof of Theorem 3. □

B. Experiments

We implemented ZZ-net in PyTorch [32] using PyTorch Lightning [10]. For the essential matrix problem we performed hyper parameter tuning using Ray Tune [26].

B.1. Estimating rotations between noisy point clouds

Here, we provide some additional information on experiments on the toy problem.

Data generation A cloud Z is formed of $m = 100$ points distributed on a random triangle. These are subsequently rotated to a cloud X by a random rotation $\theta \in \mathbb{S}$, and low-level inlier noise is added to both clouds. We subsequently, with a probability r, exchange each correspondence with an outlier $(\hat{z}_i, \hat{x}_i)$ chosen completely at random. An example of a resulting pair for $r = 0.4$ is shown in Figure 4. We generate 2000, 500 and 300 cloud pairs for training, validation and testing, respectively. Step by step, the generation procedure is as follows:

- To generate the original cloud, without outliers, we first choose three points uniformly randomly on the unit disk - these are the corners of the triangle.
- Next, we choose $m = 100$ new points uniformly randomly on the unit disk. For each of the points, we choose one of the three sides of the triangle, and orthogonally project the point onto that side. This leaves us with an inlier cloud Z_{in}.
- Next, a rotation $\theta \in \mathbb{S}$ is chosen uniformly at random, and we define the other cloud as $X_{\text{in}} = \theta Z_{\text{in}}$. We add independent Gaussian noise to each of the points in either cloud, with a standard deviation of $\sigma = 0.03$.
- Then, we go through the point pairs, throwing each one out with a probability r. The ones that are thrown out are replaced with a pair of points (z_i, x_i) independently chosen uniformly on the unit disk.

Comparison models Here we outline the two comparative methods for the experiments on rotation estimation. The first one is a PointNet with 5 equivariant layers and a head with 5 fully connected layers, with additional learnable batch normalization layers. The model as a whole has around $34K$ parameters. We also consider a model better adapted to handle outliers, incorporating an attentive context normalization [36] with 7 layers, for a total of around $11K$ parameters. We refer to the latter as 'ACNe–', since it lacks a lot of mechanisms (such as group normalization, skip connections, and other things) compared to the actual ACNe model. To reiterate, we think it would be dishonest to claim that we in this experiment compare our method with [36]. Our aim is rather to show that our approach can compete also with networks tailor-made for outlier-heavy scenarios. Both of these models take in the correspondences as vectors in $\mathbb{R}^4$, used as the channels in the first layer, and outputs two real scalars, which we reinterpret as a complex outputs. They are in particular not rotation equivariant.

The 'ACNe–'-model Let us discuss our implementation of an 'ACNe-architecture' *inspired* by [36]. The ACNe–model consists of so called ACNe-units. In each such, each point in the input is first fed through one linear layer and an activation function to produce a cloud of features $F \in (\mathbb{R}^C)^m$. These weights are then fed through two different linear layers to produce two vectors $v_1, v_2 \in \mathbb{R}^m$. A sigmoid is applied pointwise to v_1 to produce the *local weight vector* w_1. $\mathrm{SoftMax}$ is applied to v_2 to produce a *global* weight vector w_2. These are then multiplied pointwise, and normalize to sum to one, to produce the final weight vector w.

This vector is subsequently used to *context normalize* the feature cloud F. That is, each channel is normalized to have zero mean and unit variance, with respect to the probability distribution defined by w. That is, with $\hat{F} = \sum_{j \in [m]} w_j F_j$, the k:th channel of the output of the ACNe unit is equal to

$$G_i^k = \frac{F_i^k - \hat{F}^k}{\left(\sum_{i \in [m]} w_i (F_i^k - \hat{F}^k)^2 \right)^{1/2}}.$$

The entire 'ACNe–'-net has two additional steps: First, the initial input is fed through one perceptron layer before being fed to the first ACNe-unit. The actual output of the net is formed by the weighted average $\hat{F}$ of the final ACNe unit. This is different from [36], where the output of the final layer is processed further in a problem-dependent manner.

Model sizes For the broad model, the number of channel in the early layers are both equal to 4, the late layers have 4, 16, 4 and 1 channels, respectively. The vector unit layers have 32 and 1 channel, respectively.

For the deep model, each $\mathcal{R}^2(m)$-unit has 4 channels in the early layer. The late layers in the two earlier units have 4, 8 and 4 units each – the final unit instead as late layers with 4, 8 and 1 channels, respectively. The first two vector layers have 4 channels, whereas the last has 1.

The permutation equivariant layers of the PointNet have 32, 64, 128, 64, 64 and 64 layers. The layers of the fully connected head have 64, 32, 16 and 2 channels. We use max-pooling in between the permutation-equivariant layers and the fully connected head.

The layers of the 'ACNe–' model have 4, 32, 32, 64, 64, 32, 32 and 2 layers, respectively.

Max. test rot. $a =$	0°	30°	60°	180°
ZZ-net (Ours)	0.15	0.15	**0.16**	**0.15**
ACNe	**0.58**	**0.16**	0.087	0.0096
CNe	0.30	0.077	0.058	0.0
OANet	0.30	0.14	0.038	0.0

Table 3. Essential matrix estimation. mAP at $w = 10°$ error in the estimated translation and rotation vectors for different values of image plane rotations a at test time.

Max. test rot. $a =$	0°	30°	60°	180°
ZZ-net (Ours)	0.33	**0.33**	**0.33**	**0.33**
ACNe	**0.72**	0.32	0.20	0.054
CNe	0.50	0.21	0.15	0.022
OANet	0.50	0.30	0.12	0.026

Table 4. Essential matrix estimation. mAP at $w = 30°$ error in the estimated translation and rotation vectors for different values of image plane rotations a at test time.

Nonlinearities We use the ReLU as a non-linearity for the PointNet, and leaky ReLUs (where the slope parameter is set to the PyTorch standard of .01) for our models and the perceptrons in the 'ACNe–'-model.

In addition to the mechanisms described in the main paper, we choose, for the deep and broad model, to normalize each channel of the weight unit, which is a vector in $\mathbb{C}^m$, to have ℓ_2-norm 1. We found this useful to prohibit the model to not get stuck at outputs of very small magnitudes. The learnable θ-parameters in the complex ReLUs are initalized to 0.1.

Training details For the training of the PointNet, we use a stochastic gradient descent with a momentum of 0.9. The learning rate is set to 10^{-3} and we train it for 400 epochs.

For the training of the ACNe model, we use Adam [21]. The learning is initially set to 10^{-3}, and halved after 200 and 300 epochs. We train it for 400 epochs.

The broad and deep models are trained using Adam. We set the initial learning rate to $5 \cdot 10^{-3}$, and half it after 70 and 150 epochs. We train it for 300 epochs.

All models are evaluated at the final epoch, with the exception of the experiment of the broad model for $r = 0.8$, which severly overfitted the data (the final model had scores 0, 0 and .02 on the three metrics). Therefore, we (manually) stopped it early after 120 epochs, when the validation loss still was low.

B.2. Essential Matrix Estimation

In this section we present more information on the experiment on essential matrix estimation from Section 5.2

Loss function Let $\{(\xi_1, \xi_2)\}$ denote a set of virtual matches (generated as the authors of OANet do by using the OpenCV `correctMatches` function), where ξ_1 and ξ_2 are in $\mathbb{R}^2$ and $\tilde{\xi}_1$ and $\tilde{\xi}_2$ are the homogeneous representations. Then the symmetric squared epipolar loss of an estimated essential matrix E is

$$\frac{(\tilde{\xi}_2^T E \tilde{\xi}_1)^2}{(E\tilde{\xi}_1)^2_{[0]} + (E\tilde{\xi}_1)^2_{[1]}} + \frac{(\tilde{\xi}_2^T E \tilde{\xi}_1)^2}{(E^T\tilde{\xi}_2)^2_{[0]} + (E^T\tilde{\xi}_2)^2_{[1]}},$$

which we average over the set of virtual matches.

Evaluation metric The mAP score proposed by [51] is obtained by first, for equispaced angle values $v = 5°, 10°, \dots, 30°$, calculating the proportion of estimated E-matrices that have an error in angle of both the translation vector and the rotation axis vector below v. The obtained proportion can be called the precision at v. The mAP at an angle w is then obtained by averaging the precision at all $v \leq w$.

Further results We present mAP scores at 10° and 30° in Tables 3 and 4. Once again our results are averaged over two runs. The maximum difference between the scores in these two runs for mAP at 10° was 0.03 and at 30° it was 0.02.

Model details The layer structures are as follows. The backbone $\mathcal{B}$ has three ZZ-units. The first has two early layers which both have 8 output channels, two late layers which have 8 and 3 output channels, and two vector layers which have 8 and 3 output channels. The second ZZ-unit has two early layers again both with 8 output channels, two late layers with 8 and 3 output channels, and two vector layers with 8 and 3 output channels. The last ZZ-unit has one early layer with 8 output channels, one late layer with 8 output channels and one vector layer with 8 output channels. We add skip connections so that the input to each ZZ-unit is both the input to the previous unit as well as the previous unit's output.

The equivariant angle predictor $\mathcal{E}$ consist of one ZZ-unit. It has one early layer with 8 output channels, one late layer with 1 output channel and two vector layers with 8 and 1 output output channels. The output of $\mathcal{E}$ is averaged over the point cloud to predict one complex number, interpreted as one angle.

The invariant angle predictor $\mathcal{I}$ takes the outputted α^+-weights of the backbone (which are rotation invariant) as input and applies a PointNet/Deepset to it. Here the real and imaginary channels are treated like any other channel, i.e. the number of input channels to $\mathcal{I}$ is twice the number of (complex) output channels of $\mathcal{B}$. $\mathcal{I}$ consists of three layers, with 32, 64 and 4 output channels respectively. The output of $\mathcal{I}$ is averaged over the point cloud to get permutation invariance and the 4 outputted real numbers are then reinterpreted as 2 complex numbers or angles.

We add context normalization (CN) [51] between the early and late layers as well as after the vector layers in each ZZ-unit. CN normalizes the features within a point cloud to mean 0 and variance 1.

Training details We implemented our model in Pytorch using Pytorch Lightning. We used Ray Tune to find reasonable hyperparameters and then retrained the method with those.

We train the model for 30 epochs using early stopping on the validation loss. We use a learning rate of 0.01 and train using Adam. We use a batch size of 1 due to the heavy memory need.

For all comparisons we use the settings supplied by the respective authors, except for the number of training iterations which we change to 100000 to compare with our method (30 epochs corresponds to $30 \cdot 3302 = 99060$ iterations).

C. Spanning sets for spaces of $\mathrm{Stab}(0)$-equivariant linear maps

Here we present explicit spanning sets for the spaces $\mathcal{L}_0(k, \ell)$ from Section C. They are obtained via applying the isomorphism given in 1 to the spanning sets of $\mathcal{L}(k, \ell + 1)$ described in [29].

$\mathcal{L}_0(0, 0)$ This is simply the space scalars, i.e. $\mathbb{K}$.

$\mathcal{L}_0(1, 0)$ The space has dimension $B_2 \leq 2$. A basis is given by

$$\mu_0(v) = v_0,\ \mu_1(v) = \langle \mathbb{1}, v\rangle.$$

$\mathcal{L}_0(0, 1)$ The space has dimension $B_2 \leq 2$. A basis is given by

$$w_0 = e_0,\ w_1 = \mathbb{1}.$$

$\mathcal{L}_0(2, 0)$ The space has dimension $B_3 \leq 5$. A basis is given by

$$\begin{aligned}&\lambda_0(T) = \langle \mathbb{1}, T\mathbb{1}\rangle,\ \lambda_1(T) = \langle \mathbb{1}, \mathrm{diag}(T)\rangle,\ \lambda_2(T) = T_{00}\\ &\lambda_3(T) = \langle e_0, T\mathbb{1}\rangle,\ \lambda_4(T) = \langle e_0, T^T\mathbb{1}\rangle.\end{aligned}$$

$\mathcal{L}_0(1, 1)$ The space has dimension $B_3 \leq 5$. A basis is given by

$$\begin{aligned}&L_0(v) = \langle \mathbb{1}, v\rangle\mathbb{1},\ L_1(v) = v,\ L_2(v) = v_0 e_0\\ &L_3(v) = \langle \mathbb{1}, v\rangle e_0,\ L_4(T) = v_0\mathbb{1}.\end{aligned}$$

$\mathcal{L}_0(0, 2)$ This space has dimension $B_3 \leq 5$. A basis is given by

$$\begin{aligned}&T_0 = \mathbb{1} \otimes \mathbb{1},\ T_1 = \mathrm{diag}^*(\mathbb{1}), T_2 = e_0 \otimes e_0\\ &T_3 = e_0 \otimes \mathbb{1}, T_3 = \mathbb{1} \otimes e_0\end{aligned}$$

where $\mathrm{diag}^* : \mathbb{K}^m \to \mathbb{K}^m \otimes \mathbb{K}^m$ is the dual operator of diag. Concretely, $\mathrm{diag}^*(v)$ is the tensor with diagonal v.

$\mathcal{L}_0(2,1)$ The space has dimension $B_4 \leq 15$. If we let λ_i denote the basis of $\mathcal{L}_0(2,0)$ from above, the first 10 basis elements are given by

$$K_i(T) = \lambda_i(T)e_0,\ K_{4+i}(T) = \lambda_i(T)\mathbb{1},\ i = 0, \dots, 4.$$

The final five are given by

$$\begin{aligned} &K_{10}(T) = Te_0,\ K_{11}(T) = T^T e_0,\ K_{12}(T) = T\mathbb{1} \\ &K_{13}(T) = T^T\mathbb{1},\ K_{14}(T) = \operatorname{diag}(T) \end{aligned}$$

$\mathcal{L}_0(1,2)$ The space has dimension $B_4 \leq 15$. If we let T_i denote the basis of $\mathcal{L}_0(0,2)$ from above, the first 10 basis elements are given by

$$L_i(v) = v_0 T_i,\ L_{4+i}(v) = \langle \mathbb{1}, v\rangle T_i,\ i = 0, \dots, 4.$$

The final five are given by

$$\begin{aligned} L_{10}(v) &= e_0 \otimes v,\ L_{11}(T) = v \otimes e_0\ L_{12}(T) = \mathbb{1} \otimes v \\ L_{13}(T) &= v \otimes \mathbb{1},\ L_{14}(T) = \operatorname{diag}^*(v) \end{aligned}$$

$\mathcal{L}_0(2,2)$ The space has dimension $B_5 \leq 52$. If we let T_i denote the basis of $\mathcal{L}_0(0,2)$ and λ_i the one of $\mathcal{L}_0(2,0)$, from above, the first 25 basis elements are given by

$$\mathcal{K}_{5i+j}(T) = \lambda_j(T)T_i,\ i,j = 0, \dots, 4.$$

Letting K_i denote the basis of $\mathcal{L}_0(2,1)$ and L_i the one of $\mathcal{L}_0(1,2)$, the next 25 are given by

$$\mathcal{K}_{25+5i+j}(T) = L_{10+i}(K_{10+j}(T)), i,j = 0, \dots, 4$$

The final two are given by

$$\mathcal{K}_{50}(T) = T, \mathcal{K}_{51} = T^T.$$

Applying $\mathcal{L}(2,1)$-maps to $Z \otimes \overline{Z}$. When describing the $\mathcal{NS}(m)$-architecture, we argued that the very first layer of an $\mathcal{NS}(m)$-unit can be applied without calculating $Z \otimes \overline{Z}$. Let us show this. We have

$$\begin{aligned} &\lambda_0(Z \otimes \overline{Z}) = \big|\sum_{i \in [m]} z_i\big|^2, \quad \lambda_1(Z \otimes \overline{Z}) = \sum_{i \in [m]} |z_i|^2, \quad \lambda_2(Z \otimes \overline{Z}) = |z_0|^2 \\ &\lambda_3(Z \otimes \overline{Z}) = z_0 \cdot \overline{\sum_{i \in [m]} z_i}, \quad \lambda_4(Z \otimes \overline{Z}) = \overline{z_0} \cdot \sum_{i \in [m]} z_i. \end{aligned}$$

Clearly, all of these expressions can be calculated directly from $Z \in \mathbb{C}^m$, which implies that the same is true for K_i, $i = 0, \dots, 9$. As for the last five maps, we have

$$\begin{aligned} &K_{10}(Z \otimes \overline{Z}) = \overline{z_0} Z, \quad K_{11}(Z \otimes \overline{Z}) = z_0 \overline{Z}, \quad K_{12}(Z \otimes \overline{Z}) = \overline{\sum_{i \in [m]} z_i} \cdot Z \\ &K_{13}(Z \otimes \overline{Z}) = \big(\sum_{i \in [m]} z_i\big) \cdot \overline{Z}, \quad K_{14}(Z) = (|z_i|^2)_{i \in [m]} \end{aligned}$$

These expressions can clearly also be calculated without actually accessing $Z \otimes \overline{Z}$ as a tensor.