Tensor learning with orthogonal, Lorentz, and symplectic symmetries

Wilson G. Gregory
Dept. Applied Math. & Stats.
Math. Institute for Data Science
Johns Hopkins University
Baltimore, MD, USA
wgregor4@jhu.edu

Josué Tonelli-Cueto
Dept. of Quantitative Methods
CUNEF Universidad
Madrid, SPAIN
josue.tonelli.cueto@bizkaia.eu

Nicholas F. Marshall
Dept. Mathematics
Oregon State University
Corvallis, OR, USA
marsnich@oregonstate.edu

Andrew S. Lee
Dept. Mathematics & Computer Science
Adelphi University
Garden City, NY, USA
alee2@adelphi.edu

Soledad Villar
Dept. Applied Math. & Stats.
Math. Institute for Data Science
Johns Hopkins University
Baltimore, MD, USA
Flatiron Institute
New York, NY, USA
svillar3@jhu.edu

Abstract

Tensors are a fundamental data structure for many scientific contexts, such as time series analysis, materials science, and physics, among many others. Improving our ability to produce and handle tensors is essential to efficiently address problems in these domains. In this poster, we show how to exploit the underlying symmetries of functions that map tensors to tensors. More concretely, we develop universally expressive equivariant machine learning architectures on tensors that exploit that, in many cases, these tensor functions are equivariant with respect to the diagonal action of the orthogonal, Lorentz, and/or symplectic groups. We showcase our results on three problems coming from material science, theoretical computer science, and time series analysis. Our numerical experiments show that our equivariant models perform better than corresponding non-equivariant baselines.

This is work accepted and presented at ICLR'26 [1], which is of interest to ISSAC.

1 Introduction

Tensors are fundamental mathematical objects that appear in a broad spectrum of domains. In many problems in the sciences and engineering, we find tensor-valued data at their center. In these settings, we are interested in learning functions of the form $(a_1, \ldots, a_n) \mapsto f(a_1, \ldots, a_n)$ that map a tuple of tensors to another tuple of tensors. Moreover, in many applications, such as physics (see [1] for references, we expect such tensor-maps to be equivariant to some group action, such as the natural action of the orthogonal group on tensors. Using these symmetries, we can reduce the number of parameters to be learned in our training model, but, more importantly, we can guarantee better generability of the learned function outside the training data.

In this short communication, based on [1], we showcase universally expressive equivariant machine learning architectures for equivariant tensor-maps with respect not only the natural actions of the usual orthogonal, Lorentz and symplectic groups but a wider class of real linear algebraic groups. We refer to [1] for details regarding numerical experiments and applications, which showcase the advantage of our arquitectures.

2 Orthogonal Case

The orthogonal group $O(d)$ acts on vectors of $\mathbb{R}^d$ through two possible actions:

$$g \cdot v = \det(g)^{\frac{1-p}{2}} gv$$

where $p \in \{-1, 1\}$. This gives us two $O(d)$-modules: $\mathcal{T}_1(\mathbb{R}^d, -1)$ and $\mathcal{T}_1(\mathbb{R}^d, +1)$, called, respectively, *pseudovectors* and *vectors* in physics to keep track of which $O(d)$-action we are considering. This action extends naturally to tensors through the tensor product as follows:

$$\mathcal{T}_k(\mathbb{R}^d, p) := \mathcal{T}_1(\mathbb{R}^d, p_1) \otimes \cdots \otimes \mathcal{T}_1(\mathbb{R}^d, p_k)$$

where $p = p_1 \cdots p_k$. Such an $O(d)$-module depends only on p and not the particular p_i and, in physics, we call the objects in it $k_{(p)}$*-tensors*, which allows us to keep track of the kind of action.

In this setting, we are interested in parameterizing the $O(d)$-equivariant maps of the form $f : \prod_{i=1}^n \mathcal{T}_{k_i}(\mathbb{R}^d, p_i) \to \mathcal{T}_{k'}(\mathbb{R}^d, p')$, which are the maps satisfying $f(g \cdot a_1, \ldots, g \cdot a_n) = g \cdot f(a_1, \ldots, a_n)$ for all $g \in O(d)$ and $(a_1, \ldots, a_n) \in \prod_{i=1}^n \mathcal{T}_{k_i}(\mathbb{R}^d, p_i)$. To do so, we need to recall some further notions.

First, the $O(d)$-module of $k_{(p)}$-tensors, $\mathcal{T}_k(\mathbb{R}^d, p)$, admits an action by the symmetric group of degree n, Σ_n, by just permuting the indices of the tensor. We denote this action by the right-action a^σ to stress the fact that this action commutes with the $O(d)$-action. Note that this action satisfies $(v_1 \otimes \cdots \otimes v_k)^\sigma = v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(k)}$.

Second, we have a *k-contraction map* $\iota_k : \mathcal{T}_{2k+k'}(\mathbb{R}^d, p) \to \mathcal{T}_{k'}(\mathbb{R}^d, p)$ that contracts together the first $2k$ indices of the tensor using the usual inner product. This map is determined by the identity

$$\iota_k(u_1 \otimes \cdots \otimes u_k \otimes v_1 \otimes \cdots \otimes v_k \otimes w_1 \otimes \cdots \otimes w_{k'}) = \langle u_1, v_1 \rangle \cdots \langle u_k, v_k \rangle \, w_1 \otimes \cdots \otimes w_{k'}$$

where $\langle \cdot \, , \, \cdot \rangle$ is the bilinear form preserved by $O(d)$ in $\mathbb{R}^d$, i.e., the usual inner product.

Third, an *isotropic* $k_{(p)}$-tensor is just a tensor in $\mathcal{T}_k(\mathbb{R}^d, p)$ that is fixed under the $O(d)$-action. In this setting, we have two special isotropic $k_{(p)}$-tensors. On the one hand, we have the *Kronecker delta*, δ, which is the isotropic $2_{(+1)}$-tensor given by $\delta_{i,j} = 1$ whenever $i = j$ and $\delta_{i,j} = 0$ otherwise. On the other hand, the *Levi-Civita symbol*, ϵ, which is the isotropic $d_{(-1)}$-tensor given by $\epsilon_{i_1,\ldots,i_d} = 1$

if $(i_1, \ldots, i_d)$ is an even permutation of $(1, \ldots, d)$, $\epsilon_{i_1,\ldots,i_d} = -1$ if it is an odd one, and $\epsilon_{i_1,\ldots,i_d} = 0$ otherwise.

We can now state the main results regarding $O(d)$-equivariant functions we prove in [1].

Theorem 1. *[1, Theorem 1] Let $f : \prod_{i=1}^n \mathcal{T}_{k_i}(\mathbb{R}^d, p_i) \to \mathcal{T}_{k'}(\mathbb{R}^d, p')$ be an $O(d)$-equivariant polynomial function of degree at most R. Then we may write f as follows:*

$$f(a_1, \ldots, a_n) = \sum_{r=0}^{R} \sum_{1 \leq \ell_1 \leq \cdots \leq \ell_r \leq n} \iota_{k_{\ell_1,\ldots,\ell_r}}(a_{\ell_1} \otimes \ldots \otimes a_{\ell_r} \otimes c_{\ell_1,\ldots,\ell_r})$$

where $c_{\ell_1,\ldots,\ell_r}$ is an $O(d)$-isotropic $(k_{\ell_1,\ldots,\ell_r} + k')_{(p_{\ell_1,\ldots,\ell_r}\, p')}$-tensor with order and parity chosen to be consistent with the output's ($k_{\ell_1,\ldots,\ell_r} = \sum_{q=1}^r k_{\ell_q}$ and $p_{\ell_1,\ldots,\ell_r} = \prod_{q=1}^r p_{\ell_q}$).

The above theorem gives an effective parameterization, because we can parameterize the $O(d)$-isotropic $k_{(p)}$-tensors using the Kronecker delta and Levi-Civita symbol. See [1, Appendix C] and references therein for more details.

Lemma 1 (Characterization of $O(d)$-isotropic $k_{(p)}$-tensors)**.** *Suppose $c \in \mathcal{T}_k(\mathbb{R}^d, p)$ is $O(d)$-isotropic. Then the following holds:*

Case $p = +1$: *Assume $p = +1$. If k is even, then c can be written in the form $c = \sum_{\sigma \in \Sigma_k} \alpha_\sigma (\delta^{\otimes \frac{k}{2}})^\sigma$ where $\alpha_\sigma \in \mathbb{R}$. Otherwise, if k is odd, then $c = 0$ is the zero tensor.*

Case $p = -1$: *Assume $p = -1$. If $k - d$ is even and $k \geq d$, then c can be written in the form $c = \sum_{\sigma \in S_k} \beta_\sigma (\delta^{\otimes \frac{k-d}{2}} \otimes \epsilon)^\sigma$ where $\beta_\sigma \in \mathbb{R}$. Otherwise, if $k - d$ is odd or $k < d$, then $c = 0$ is the zero tensor.*

We note that not all permutations are needed in the above expressions, we refer to [1, Appendix D] for further details. In the following special case, we get a simpler form:

Corollary 1. *[1, Corollary 1] Let $f : \prod_{i=1}^n \mathcal{T}_1(\mathbb{R}^d, +) \to \mathcal{T}_{k'}(\mathbb{R}^d, +)$ be an $O(d)$-equivariant polynomial function. Then, we may write it as*

$$f(v_1, \ldots, v_n) = \sum_{t=0}^{\lfloor \frac{k'}{2} \rfloor} \sum_{\sigma \in S_{k'}} \sum_{1 \leq J_1 \leq \ldots \leq J_{k'-2t} \leq n} q_{t,\sigma,J}((\langle v_i, v_j \rangle)_{i,j=1}^n)\, (v_{J_1} \otimes \ldots \otimes v_{J_{k'-2t}} \otimes \delta^{\otimes t})^\sigma,$$

where $J = (J_1, \ldots, J_{k'-2t})$ are indices of the input tensors, and the function $q_{t,\sigma,J}$ which depends on the tuple (t, σ, J) is a polynomial of the inner products of the input vectors.

3 General Case

The above tensor setting for the orthogonal group can be generalized easily for groups beyond the orthogonal group. We refer to [1, Section 4 and Appendix G] for the proofs and all the details. Now, consider a self-paired real vector space $(V, \langle \cdot\,, \cdot \rangle)$, where $\langle \cdot\,, \cdot \rangle$ is a non-degenerate (not necessarily symmetric) bilinear form. Through the tensor product and its universal property we get too self-paired real vector spaces of tensors $(V^{\otimes k}, \langle \cdot\,, \cdot \rangle)$. Now, if we have a group G acting (rationally) on $(V, \langle \cdot\,, \cdot \rangle)$ in a structure-preserving way (i.e., linearly, preserving $\langle \cdot\,, \cdot \rangle$), we have, through the universal property of tensor product, an extension of this action to the $(V^{\otimes k}, \langle \cdot\,, \cdot \rangle)$. In this setting, we get the family of (rational) G-modules $\mathcal{T}_k(V,\chi) := (V^{\otimes k}, \langle\,\cdot\,,\,\cdot\,\rangle)$ where $\chi : G \to \mathbb{R}^*$ is a one-dimensional (rational) group-homomorphism of G where the action of G satisfies $g\cdot(v_1\otimes\cdots\otimes v_k) = \chi(g)(g\cdot v_1)\otimes\cdots\otimes(g\cdot v_k)$ for all $g\in G$ and $v_1,\dots,v_n\in V$. In this setting, we still have G-isotropic tensors and a G-equivariant *contraction map*

$$\iota_k^G : \mathcal{T}_{2k+k'}(V,\chi) \to \mathcal{T}_{k'}(V,\chi)$$

satisfying $\iota_k^G(a\otimes b\otimes c) = \langle a,b\rangle c$ for $a,b\in\mathcal{T}_k(V,\chi_0)$, being χ_0 the trivial character, and $c\in\mathcal{T}_{k'}(V,\chi)$.

Our result then extends from orthogonal groups to the so-called class of complexly averageable real linear algebraic groups, which are not necessarily compact.

Definition 1. A real linear algebraic group G is *complexly averageable* if it's Zariski-dense in its complexification and its complexification admits a Zariski-dense compact subgroup closed under complex conjugation.

Theorem 2. *Let $G\subset GL(V)$ be either a compact or a complexly averageable real linear algebraic group acting rationally in a structure-preserving way on a self-paired vector space $(V,\langle\,\cdot\,,\,\cdot\,\rangle)$ and $f:\prod_{i=1}^n\mathcal{T}_{k_i}(V,\chi_i)\to\mathcal{T}_{k'}(V,\chi')$ a G-equivariant entire function. Then we may write f as follows:*

$$f(a_1,\dots,a_n) = \sum_{r=0}^{\infty}\sum_{1\le\ell_1\le\cdots\le\ell_r\le n}\iota^G_{k_{\ell_1,\dots,\ell_r}}(a_{\ell_1}\otimes\ldots\otimes a_{\ell_r}\otimes c_{\ell_1,\dots,\ell_r})$$

where $c_{\ell_1,\dots,\ell_r}\in\mathcal{T}_{k_{\ell_1,\dots,\ell_r}+k'}(\mathbb{R}^d,\chi_{\ell_1,\dots,\ell_r}\chi')$ is a G-isotropic tensor for $k_{\ell_1,\dots,\ell_r} := \sum_{q=1}^r k_{\ell_q}$ and $\chi_{\ell_1,\dots,\ell_r} = \prod_{q=1}^r\chi_{\ell_q}$.

In the special case of the *indefinite orthogonal group* (which is the linear part of the *Lorentz group* when $d=4$ and $s\in\{1,3\}$) and the *symplectic group*, given respectively by

$$O(s,d-s) := \{g\in\mathrm{GL}(\mathbb{R}^d)\mid g^\top\mathbb{I}_{s,d-s}g=\mathbb{I}_{s,d-s}\}\quad\text{and}\quad Sp(d) := \{g\in\mathrm{GL}(\mathbb{R}^d)\mid g^\top\mathbb{J}_d g=\mathbb{J}_d\}$$

where $\mathbb{I}_{s,d-s} := \begin{pmatrix}\mathbb{I}_s & \\ & -\mathbb{I}_{d-s}\end{pmatrix}$ and $\mathbb{J}_d := \begin{pmatrix} & \mathbb{I}_{d/2}\\ -\mathbb{I}_{d/2} & \end{pmatrix}$, we get the following improvement using the permutation action. Below, $[A]_{i,j}$ denotes the 2-tensor in $(\mathbb{R}^d)^{\otimes 2}$ induced by the matrix $A\in\mathbb{R}^{d\times d}$.

Corollary 2. *Let G be either $O(s,d-s)$ or $Sp(d)$ and $f:\prod_{i=1}^n\mathcal{T}_1(\mathbb{R}^d,\chi_0)\to\mathcal{T}_k(\mathbb{R}^d,\chi_0)$, with $\chi_0(g)=1$ for all g, be a G-equivariant entire function. Then we may write f as follows:*

$$f(v_1,\dots,v_n) = \sum_{t=0}^{\lfloor\frac{k}{2}\rfloor}\sum_{\sigma\in S_k}\sum_{1\le J_1\le\cdots\le J_{k-2t}\le n} q_{t,\sigma,J}\big((\langle v_i,v_j\rangle_G)_{i,j=1}^n\big)\,(v_{J_1}\otimes\ldots\otimes v_{J_{k-2t}}\otimes\theta_G^{\otimes t})^\sigma$$

where $\langle v,w\rangle_G = v^T\mathbb{I}_{s,d-s}w$ and $\theta_G = [\mathbb{I}_{s,d-s}]_{i,j}$ if $G=O(s,d-s)$, and $\langle v,w\rangle_G = v^T\mathbb{J}_d w$ and $\theta_G=[\mathbb{J}_d]_{i,j}$ if $G=Sp(d)$, and $q_{t,\sigma,J}$ is an entire function that depends on the tuple (t,σ,J) and whose inputs are all possible inner products between the input vectors and whose output is a scalar.

References

[1] W.G. Gregory, J. Tonelli-Cueto, N.F. Marshall, A.S. Lee, and S. Villar. Tensor learning with orthogonal, Lorentz, and symplectic symmetries. In *The Fourteenth International Conference on Learning Representations (ICLR2026)*, 2026. https://openreview.net/forum?id=1FCZ4f8dAY.