Dual Quaternion Rotational and Translational Equivariance in 3D Rigid Motion Modelling

1. Guilherme Vieira – Universidade Estadual de Campinas (Unicamp)
2. Eleonora Grassucci – Sapienza Università di Roma
3. Marcos Eduardo Valle – Universidade Estadual de Campinas (Unicamp)
4. Danilo Comminiello – Sapienza Università di Roma

33rd IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2023)

Motivation

Human pose forecasting aims to predict the position of human keypoints.

Applications include home healthcare, security surveillance, and self-driving cars.

Forecasting human pose should be equivariant to rigid motion, which means the movement does not depend on the observer.

Usually, the data comprise a sequence of highly correlated points in the $3D$ space. Formally, the pose at instant t is represented by

$$\boldsymbol{P}(t) = \begin{bmatrix} x_1(t) & y_1(t) & z_1(t) \\ \vdots & \vdots & \vdots \\ x_n(t) & y_n(t) & z_n(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1(t) \\ \vdots \\ \boldsymbol{p}_n(t) \end{bmatrix}. \tag{1}$$

Vector and Hypercomplex-Valued Networks

Vector-valued neural networks (V-nets) are designed to process arrays of vectors (Valle, 2023).

They are obtained by replacing real numbers with vectors. Formally, the field $(\mathbb{R}, +, \cdot)$ is substituted by an algebra $\mathbb{V}$.

Definition 1 (Algebra – Schafer (1961))

An algebra $\mathbb{V}$ over a field $\mathbb{F}$ is a vector space over $\mathbb{F}$ with an additional bilinear operation called multiplication.

A hypercomplex algebra is an algebra with additional algebraic or geometric properties (Catoni et al., 2008; Kantor and Solodovnikov, 1989).

Quaternions

Quaternions, denoted by $\mathbb{H}$, are well-known hypercomplex algebras.

A quaternion is a number of the form

$$q = q_W + q_X\hat{\imath} + q_Y\hat{\jmath} + q_Z\hat{\kappa} \in \mathbb{H}, \tag{2}$$

where q_W, q_X, q_Y, q_Z are real numbers.

The hyperimaginary units satisfy the Hamilton rules:

$$\hat{\imath}^2 = \hat{\jmath}^2 = \hat{\kappa}^2 = \hat{\imath}\hat{\jmath}\hat{\kappa} = -1 \tag{3}$$

In polar form, we have

$$q = \|q\| \left(\cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\theta}{2}\right) \bar{\mathbf{u}} \right), \tag{4}$$

where $\theta \in [0, 2\pi)$ and $\bar{\mathbf{u}} = u_X\hat{\imath} + u_Y\hat{\jmath} + u_Z\hat{\kappa}$ is a pure quaternion.

Dual Numbers

A dual number has the form

$$\widehat{a} = a_0 + \varepsilon a_\varepsilon, \tag{5}$$

where $a_0, a_\varepsilon \in \mathbb{R}$ and ε, the dual unit, satisfies $\varepsilon^2 = 0$.

For example, the following holds

$$(a_0 + \varepsilon a_\varepsilon)(b_0 + \varepsilon b_\varepsilon) = a_0 b_0 + \varepsilon(a_0 b_\varepsilon + a_\varepsilon b_0). \tag{6}$$

Dual Quaternions

Dual quaternions, denoted by $\mathbb{D}$, are quaternions whose components are dual numbers.

Formally, a dual quaternion is given by $\widehat{\mathbf{q}} = \widehat{q}_W + \widehat{q}_X \hat{\imath} + \widehat{q}_Y \hat{\jmath} + \widehat{q}_Z \hat{\kappa}$, where $\widehat{q}_W, \widehat{q}_X, \widehat{q}_Y, \widehat{q}_Z$ are dual numbers.

Equivalently, a dual quaternion is a dual numbers in which each part is a quaternion.

Thus, a dual quaternion can be represented by $\widehat{\mathbf{q}} = (q_0 + \varepsilon q_\varepsilon)$, where $q_0, q_\varepsilon \in \mathbb{H}$.

We say that $\widehat{\mathbf{q}}$ is a unit dual quaternion if

$$\|\widehat{\mathbf{q}}\| = \sqrt{\|q_0\|^2 + \|q_\varepsilon\|^2} = 1. \tag{7}$$

Dual Quaternion Representation of Rigid Motions

A rigid motion in 3D space is a rotation and a translation.

The rotation of $u = u_x\hat{\imath} + u_y\hat{\jmath} + u_z\hat{\kappa}$ by an angle θ followed by a translation by $d = d_x\hat{\imath} + d_y\hat{\jmath} + d_z\hat{\kappa}$ yields the dual quaternion

$$\widehat{\mathbf{q}} = \widehat{q}_W + \widehat{q}_X\hat{\imath} + \widehat{q}_Y\hat{\jmath} + \widehat{q}_Z\hat{\kappa}, \tag{8}$$

where

$$\begin{aligned}
\widehat{q}_W &= \cos\frac{\theta}{2} - \frac{\varepsilon}{2}\vec{u}\cdot\vec{d}\sin\frac{\theta}{2},\\
\widehat{q}_X &= u_x\sin\frac{\theta}{2} + \frac{\varepsilon}{2}\left[d_x\cos\frac{\theta}{2} - \sin\frac{\theta}{2}(u_y d_z - u_z d_y)\right],\\
\widehat{q}_Y &= u_y\sin\frac{\theta}{2} + \frac{\varepsilon}{2}\left[d_y\cos\frac{\theta}{2} - \sin\frac{\theta}{2}(u_z d_x - u_x d_z)\right],\\
\widehat{q}_Z &= u_z\sin\frac{\theta}{2} + \frac{\varepsilon}{2}\left[d_z\cos\frac{\theta}{2} - \sin\frac{\theta}{2}(u_x d_y - u_y d_x)\right].
\end{aligned} \tag{9}$$

If there is no translation ($\overrightarrow{d} = 0\hat{\imath} + 0\hat{\jmath} + 0\hat{\kappa}$), then the dual part becomes null, and we have a quaternion in the polar form.

If there is no rotation ($\theta = 0$), we obtain $\widehat{\mathbf{q}} = 1 + \frac{\varepsilon}{2}\overrightarrow{d}$. Thus, the dual part is responsible for the translation.

Dual quaternion contain the information necessary do describe rigid motions.

Question:

Do neural networks based on dual quaternions exhibit rigid motion equivariance?

Experiment – Lorenz System

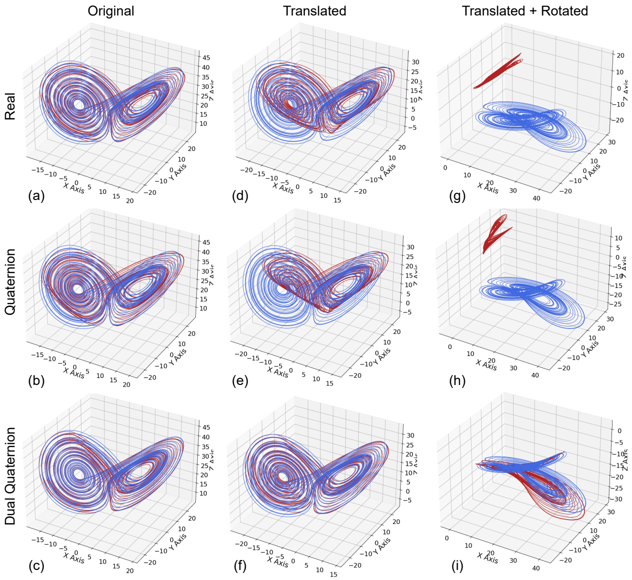
The Lorenz system describes the movement of a free particle in atmospheric domain effects.

We used a time series of 10.000 consecutive positions, 10% of which are used for training, and the remaining 90% are used for testing.

Single hidden-layer MLPs with ReLU activation are used as follows:

- **Real-Valued Model (6-128-3 architecture, 1280 parameters):** $(x_{t-1}, y_{t-1}, z_{t-1}, x_t, y_t, z_t) \mapsto (x_{t+1}, y_{t+1}, z_{t+1})$.
- **Quaternion-Valued Model (2-80-1 architecture, 1280 params):** $(0 + x_{t-1}\hat{\imath} + y_{t-1}\hat{\jmath} + z_{t-1}\hat{\kappa}, 0 + x_t\hat{\imath} + y_t\hat{\jmath} + z_t\hat{\kappa}) \mapsto$ $(0 + x_{t+1}\hat{\imath} + y_{t+1}\hat{\jmath} + z_{t+1}\hat{\kappa})$.
- **Dual Quaternion-Valued Model (1-53-1 arch., 1280 params):** $(0 + x_{t-1}\hat{\imath} + y_{t-1}\hat{\jmath} + z_{t-1}\hat{\kappa}) + \varepsilon(0 + x_t\hat{\imath} + y_t\hat{\jmath} + z_t\hat{\kappa}) \mapsto$ $(0 + x_{t+1}\hat{\imath} + y_{t+1}\hat{\jmath} + z_{t+1}\hat{\kappa}) + \varepsilon(0 + 0\hat{\imath} + 0\hat{\jmath} + 0\hat{\kappa})$.

Predicted trajectories are in red and expected in blue.



Model	Test MSE↓			
	Original	Translated	Rotated	T+R
Real	0.433	4.648	118.719	192.766
Quaternion	0.756	7.952	178.102	263.051
Dual Quat.	**0.272**	**0.183**	**2.140**	**3.617**

Model	Test Prediction Gain↑			
	Original	Translated	Rotated	T+R
Real	57.187	31.050	14.024	10.360
Quaternion	51.918	24.841	6.847	3.729
Dual Quat.	**63.606**	**52.722**	**43.797**	**40.108**

Experiment – Human Pose Forecasting

Human pose forecast using variational autoencoder endowed with dual quaternion numbers based on the decoupled representations for pose forecasting (DeRPoF) (Parsaeifard et al., 2021).

The position (x, y, z) of a joint is encoded by

$$\widehat{\mathbf{q}} = x_c\hat{\imath} + y_c\hat{\jmath} + z_c\hat{\kappa} + \varepsilon\left((x - x_c)\,\hat{\imath} + (y - y_c)\,\hat{\jmath} + (z - z_c)\,\hat{\kappa}\right), \quad (10)$$

where (x_c, y_c, z_c) represents the center of mass.

We used the 3D Poses in the Wild (3DPW) structured dataset (13 key joints and the center of mass), which contains over 51.000 registered frames from 60 short videos with hugging, arguing, playing basketball, and dancing, among others.

Model	VIM↓	FDE↓	Val Loss↓
SC-MPF (Adeli et al., 2020)	46.28	-	-
Nearest Neighbour (Zhang et al., 2019)	27.34	-	-
Zero Velocity (Martinez et al., 2017)	29.35	-	-
DeRPoF (Parsaeifard et al., 2021)	19.07 $\pm$ .005	0.360 $\pm$ .007	-
CoRPoF (Parsaeifard et al., 2021)	16.76 $\pm$ .003	0.317 $\pm$ .001	0.118 $\pm$ .004
Quaternion CoRPoF	16.35 $\pm$ .009	0.271 $\pm$ .002	0.105 $\pm$ .010
Dual Quaternion CoRPoF	**15.23** $\pm$ .002	**0.266** $\pm$ .001	**0.103** $\pm$ .006

Metrics:

- Visibility ignored metric (VIM): the average of the distances between each predicted joint and the ground truth, in centimeters Parsaeifard et al. (2021).
- Final displacement error (FDE): An L2 distance.

Concluding Remarks

Vector-valued networks (V-nets) are designed to process vector information. In particular, hypercomplex-valued neural networks are particular V-nets enriched with algebraic or geometric properties.

Dual quaternions contain information on rigid motions in 3D.

We provided a practical example of the translation and rotation equivariance properties using the Lorenz system.

We proceed to show how models endowed with this formulation outperform other approaches for human pose forecasting.

Our results show that models utilizing dual quaternions are able to maintain their performance even when data are translated and rotated.

Acknowledge

I would like to thank the financial support by FAPESP, CNPq, and FAEPEX-Unicamp:

FAPESP CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico FAEPEX

Thanks for your attention!

References (1)

V. Adeli, E. Adeli, I. D. Reid, J. C. Niebles, and H. Rezatofighi. Socially and contextually aware human motion and pose forecasting. *IEEE Robotics and Automation Letters*, 5:6033–6040, 2020.

F. Catoni, D. Boccaletti, R. Cannata, V. Catoni, E. Nichelatti, and P. Zampetti. *The Mathematics of Minkowski Space-Time*. Birkhäuser Basel, 2008. doi: 10.1007/978-3-7643-8614-6.

I. L. Kantor and A. S. Solodovnikov. *Hypercomplex Numbers: An Elementary Introduction to Algebras*. Springer New York, 1989. doi: 10.1007/978-1-4612-3650-4.

J. Martinez, M. J. Black, and J. Romero. On human motion prediction using recurrent neural networks. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 4674–4683, 2017.

References (2)

B. Parsaeifard, S. Saadatnejad, Y. Liu, T. Mordan, and A. Alahi. Learning decoupled representations for human pose forecasting. *IEEE/CVF Int. Conf. on Computer Vision Workshops (ICCVW)*, pages 2294–2303, 2021.

R. Schafer. *An Introduction to Nonassociative Algebras*. Project Gutenberg, 1961. URL `https://www.gutenberg.org/ebooks/25156`.

M. E. Valle. Understanding vector-valued neural networks and their relationship with real and hypercomplex-valued neural networks, 2023.

J. Y. Zhang, P. Felsen, A. Kanazawa, and J. Malik. Predicting 3d human dynamics from video. *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, pages 7113–7122, 2019.