EQNIO: SUBEQUIVARIANT NEURAL INERTIAL ODOMETRY

Royina Karegoudra Jayanth[*], **Yinshuang Xu**[*], **Ziyun Wang**, **Evangelos Chatzipantazis**, **Kostas Daniilidis**[†], **Daniel Gehrig**[†]
University of Pennsylvania
{royinakj,xuyin,ziyunw,vaghat,dgehrig}@seas.upenn.edu
kostas@cis.upenn.edu

ABSTRACT

Neural network-based odometry using accelerometer and gyroscope readings from a single IMU can achieve robust, and low-drift localization capabilities, through the use of *neural displacement priors (NDPs)*. These priors learn to produce denoised displacement measurements but need to ignore data variations due to specific IMU mount orientation and motion directions, hindering generalization. This work introduces EqNIO, which addresses this challenge with *canonical displacement priors*, i.e., priors that are invariant to the orientation of the gravity-aligned frame in which the IMU data is expressed. We train such priors on IMU measurements, that are mapped into a learnable canonical frame, which is uniquely defined via three axes: the first is gravity, making the frame gravity aligned, while the second and third are predicted from IMU data. The outputs (displacement and covariance) are mapped back to the original gravity-aligned frame. To maximize generalization, we find that these learnable frames must transform equivariantly with global gravity-preserving roto-reflections from the subgroup $O_g(3) \subset O(3)$, acting on the trajectory, rendering the NDP $O(3)$-*subequivariant*. We tailor specific linear, convolutional, and non-linear layers that commute with the actions of the group. Moreover, we introduce a bijective decomposition of angular rates into vectors that transform similarly to accelerations, allowing us to leverage both measurement types. Natively, angular rates would need to be inverted upon reflection, unlike acceleration, which hinders their joint processing. We highlight EqNIO's flexibility and generalization capabilities by applying it to both filter-based (TLIO), and end-to-end (RONIN) architectures, and outperforming existing methods that use *soft* equivariance from auxiliary losses or data augmentation on various datasets. We believe this work paves the way for low-drift and generalizable neural inertial odometry on edge devices. The project details and code can be found at https://github.com/RoyinaJayanth/EqNIO.

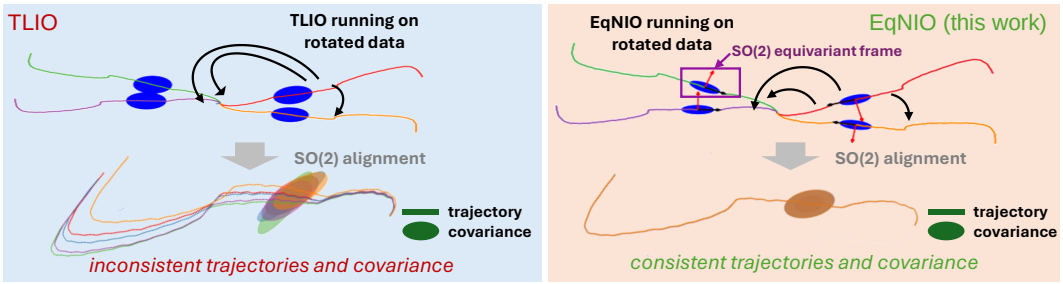

Figure 1: Neural Displacement Priors (NDPs) that rely on data augmentation (TLIO, left) produce a different trajectory for different reference frames while strictly equivariant approaches (EqNIO, right) yield one trajectory independent of the reference frame. EqNIO achieves this by learning an equivariant canonical frame aligned, per definition, with the predicted covariance (ellipsoids,right). In test time, the gravity-aligned IMU orientation estimate of the Kalman filter is the reference frame.

[*] denotes equal contribution, † denotes equal advising

1 INTRODUCTION

Inertial Measurement Units (IMUs) are commodity sensors that measure the accelerations and angular velocities of a body. Due to their low cost, they are widely used in diverse applications such as robot navigation and Mixed Reality for precise and rapid tracking of body frames. However, since they are differential sensors, relying only on IMUs invariably results in drift. Traditional Visual Inertial Odometry (VIO) approaches can effectively mitigate this drift by combining IMU measurements with features extracted from camera images. Still, these images are of limited use in high-speed scenarios with challenging lighting conditions, since they can suffer from saturation, and blurring artifacts.

Recently, a novel class of methods has emerged that instead mitigates this drift with neural displacement priors (NDPs), that are directly learned from IMU data alone (Liu et al., 2020; Herath et al., 2020). These methods perform competitively with VIO methods despite using only a single IMU sensor. However, learning generalizable priors proves challenging: while identifying specific motion patterns, they must learn to ignore data differences due to particular IMU mount orientations, and the direction of the motion patterns. However, they often fail to do so in practice (see Fig. 1, left state-of-the-art method TLIO (Liu et al., 2020)), yielding large trajectory variations when observing simply rotated input data. These variations persist, in spite of data augmentation strategies (TLIO Liu et al. (2020), RONIN Herath et al. (2020)) that include random data rotations or the enforcement of equivariance constraints via auxiliary consistency losses during training (RIO Cao et al. (2022)).

In this work, we simplify this task by introducing **EqNIO**, which leverages *canonical displacement priors (CDPs)*. CDPs are invariant to IMU orientation and, as a result, easier to learn, more robust, and more generalizable than regular NDPs. EqNIO first maps the IMU data into a learnable, equivariant, and gravity-aligned frame F, before passing them to the CDP, and mapping the outputs back to the original frame. Our design can flexibly integrate arbitrary off-the-shelf methods such as TLIO (Liu et al., 2020) and RONIN (Herath et al., 2020), provided that suitable change of basis maps are defined for the network outputs (displacement and covariance of TLIO, and only linear velocity for RONIN). The frame F is composed of two transformations: The first, gravity-alignment, aligns the z-axis with the gravity direction, estimated from an off-the-shelf filter (Liu et al., 2020), and is traditionally used in inertial odometry. The second is a learnable, gravity-preserving roto-reflection.

To be maximally generalizable, we show that this roto-reflection must generalize to arbitrary gravity-preserving roto-reflections of the input data. We design a specific model that achieves this by processing IMU data *equivariantly*, *i.e.*, in a way that commutes with the action of gravity-preserving roto-reflections on the input data. We identify these roto-reflections as elements of the group $O_g(3)$, a subgroup of the orthogonal group $O(3)$, which is isomorphic to $O(2)$. Equivariance is ensured by a preliminary, unique preprocessing step that maps accelerometer and gyroscope measurements into a space that transforms consistently under the group action, and subsequent processing with equivariant MLPs, convolutions, and non-linear layers. Due to the isomorphism with $O(2)$ these layers are $O(2)$ equivariant, and, due to the subgroup property, also called $O(3)$ *subequivariant*.

Contributions: *(i)* We apply a canonicalization scheme that maps IMU measurement, and NDP outputs to and from a gravity-aligned, subequivariant, canonical frame. This procedure can be flexibly applied to arbitrary off-the-shelf network architectures. NDPs trained on such canonical data produce inherently more robust, and generalizable results than previous work. *(ii)* We formalize the group actions of gravity-preserving roto-reflection from $O_g(3)$ on IMU measurements and derive unique preprocessing steps that map both accelerometer and gyroscope measurements into a space in which these actions are consistent. *(iii)* We tailor an $O(2)$ equivariant network that regresses canonical roto-reflections from 2D vector features, and 1D scalar derived from IMU data. It leverages specialized $O(2)$ equivariant MLPs and convolution, to process vector features, conventional layers to process scalar features, and equivariant non-linearities to mix vector and scalar features.

We demonstrate the generality of our framework by applying it to two neural inertial odometry methods, TLIO (Liu et al., 2020), and RONIN (Herath et al., 2020). Extensive qualitative and quantitative results comparing EqNIO against previous works across diverse benchmarks establish a new state-of-the-art in inertial-only odometry. EqNIO significantly enhances the accuracy, reliability, and generalization of existing methods.

2 RELATED WORK

Neural inertial odometry is highly related to inertial navigation systems (INS) which can be broad classified into estimation-based, multi-sensor data fusion-based (MSDF) techniques, and learning-

based techniques. Recent works in inertial attitude estimation (Asgharpoor Golroudbari & Sabour, 2023; Ge et al., 2024), VIO (van Goor & Mahony, 2023), LIDAR odometry (Zheng et al., 2024) and SLAM (Kim & Sukkarieh, 2003) are highly related and are hence discussed in App. A.1.

Model-based Inertial Odometry: Purely Inertial Odometry can be broadly classified into two categories: kinematics-based and learning-based approaches. Kinematics-based approaches (Leishman et al., 2014; Titterton et al., 2004; Bortz, 1971) leverage analytical solutions based on double integration that suffers from drift accumulation over time when applied to consumer-grade IMUs. To mitigate this drift, loop closures (Solin et al., 2018), and other pseudo measurements derived from IMU data that are drift-free have been explored (Groves, 2015; Hartley et al., 2020; Brajdic & Harle, 2013). In the context of Pedestrian Dead Reckoning (Jimenez et al., 2009), these include step counting (Ho et al., 2016; Brajdic & Harle, 2013), detection of the system being static (Foxlin, 2005; Rajagopal, 2008) and gait estimation (Beaufils et al., 2019).

Learning-based Inertial Odometry: Recently, RIDI (Yan et al., 2018), PDRNet (Asraf et al., 2022) and RONIN (Herath et al., 2020) proposed CNN, RNN, and TCN-based velocity regression. RIDI uses these velocities to correct the IMU measurements, while RONIN directly integrates them while assuming given orientation information. Denoising networks either regress IMU biases (Brossard et al., 2020b; Buchanan et al., 2023; Brossard et al., 2020a) denoised IMU measurements directly (Steinbrener et al., 2022). While Buchanan et al. (2023) uses constant covariance, AI-IMU (Brossard et al., 2020a) estimates the covariance for automotive applications. Displacement-based methods like IONet (Chen et al., 2018a), TLIO (Liu et al., 2020), RNIN-VIO (Chen et al., 2021a), and IDOL (Sun et al., 2021) directly estimate 2D/3D displacement. Unlike TLIO which regresses a diagonal covariance matrix, Russell & Reale (2021) estimates the full covariance matrix parameterized via Pearson correlations. RNIN-VIO extends TLIO to continuous human motion adding a loss function for long-term accuracy. Unlike these methods, EqNIO learns canonical displacement priors and thus generalizes better to arbitrary IMU orientation and motion directions.

Equivariant Inertial Odometry: The previous learning-based approaches (Liu et al., 2020; Chen et al., 2021a) use $SO(2)$ augmentation strategies to achieve approximate $SO(2)$ equivariance. MotionTransformer (Chen et al., 2019) used GAN-based RNN encoder to transfer IMU data into domain-invariant space by separating the domain-related constant. Recently, RIO (Cao et al., 2022) demonstrated the benefits of approximate $SO(2)$ equivariance with an auxiliary loss, introduced Adaptive Test Time Training (TTT), and uncertainty estimation via ensemble of models (See Appendix A.4 for more details). We propose integrating strict equivariance by design directly into the framework. Additionally, no prior work has addressed reflection equivariance, which requires specific preprocessing of gyroscope data for it to adhere to the right-hand rule. Our novel $O(2)$ equivariant framework can be seamlessly integrated with existing learning-based inertial navigation systems.

Equivariant Networks: Group equivariant networks (Cohen & Welling, 2016) commute by design with group actions on the input, and have been tailored to a variety of inputs and architecture designs. These include point clouds (Thomas et al., 2018; Chen et al., 2021b; Deng et al., 2021), 2D (Worrall et al., 2017; Weiler & Cesa, 2019), 3D (Weiler et al., 2018; Esteves et al., 2019), and spherical images (Cohen et al., 2018; Esteves et al., 2018; 2020; 2023), graphs (Satorras et al., 2021) and general manifolds (Cohen et al., 2019b;a; Weiler et al., 2021; Xu et al., 2024). Yet, equivariant networks tailored to IMU data and their symmetries have not been studied, and are introduced in the current work. Our method draws on general theories and methods developed for equivariance to $E(n)$ its subgroups. Cesa et al. (2021); Xu et al. (2022) use Fourier analysis to design steerable CNN kernels on homogeneous space, while Finzi et al. (2021) proposed an algorithm for finding a kernel by solving a linear equivariant map constraint. Villar et al. (2021) demonstrated that any $O(n)$ equivariant function can be represented using a set of scalars and vectors. However, applying these to the neural integration of IMUs is not straightforward as gravity's presence introduces subequivariance, angular velocity in the input data follows the right-hand rule, and the input is a sequence with a time dimension. Related approaches (Han et al., 2022; Chen et al., 2023) tackle subequivariance using equivariant graph networks and calculating gram matrices achieving simple $O(2)$ equivariance. However, dealing with data that obey the right-hand rule (*e.g.* angular rates), has been underexplored, and is addressed in the current work. Canonicalization is a prevalent approach to equivariance. Kaba et al. (2023) first introduced learned canonicalization functions for equivariant networks, which we adopt in our design. We discuss additional works on equivariant canonicalization in Appendix A.1.

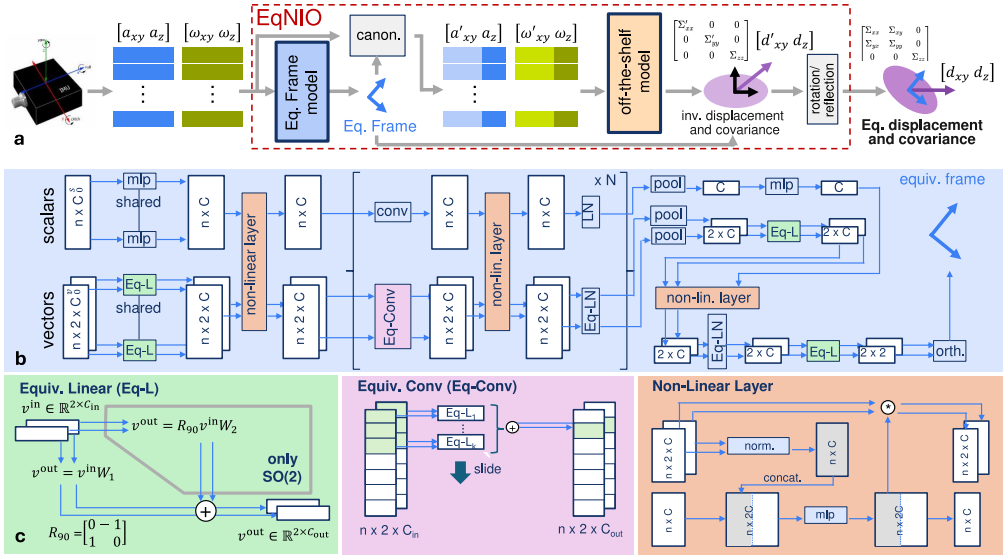

Figure 2: EqNIO (a) processes gravity-aligned IMU measurements, $\{(a_i, \omega_i)\}_{i=1}^n$. An equivariant network (blue) predicts a canonical equivariant frame F into which IMU measurements are mapped, *i.e. canonicalized*, yielding invariant inputs $\{(a'_i, \omega'_i)\}_{i=1}^n$. A conventional neural network then predicts invariant displacement (d') and covariance (Σ') which are mapped back yielding equivariant displacement (d) and covariance (Σ). The equivariant network (b) takes as input $n \times C_0^s$ scalars, and $n \times C_0^v$ vectors: Vectors are processed by equivariant layers (Eq-L, Eq-Conv, Eq-LN), while scalars are separately processed with conventional layers. Eq-L (green) uses two weights W_1, W_2 for SO(2) equivariance, and only W_1 for O(2) equivariance. Eq-Conv (pink) uses Eq-L to perform 1-D convolutions over time. The equivariant non-linear layer (orange) mixes vector and scalar features.

3 Problem Setup

This paper targets neural inertial odometry using data from a single IMU, comprised of an accelerometer (giving linear accelerations $a_i \in \mathbb{R}^3$) and gyroscope (giving angular velocity $\omega_i \in \mathbb{R}^3$). IMU's measure sequences of data $\{(a_i, \omega_i)\}_{i=1}^n$, each expressed in the local IMU body frame, b, at time t_i. These are related to the true IMU acceleration $\bar{a}_i$ and angular rates $\bar{\omega}_i$ via

$$\tilde{\omega}_i = \bar{\omega}_i + b_i^g + \eta_i^g \qquad \tilde{a}_i = \bar{a}_i - {}^w_b R_i{}^T g + b_i^a + \eta_i^a \tag{1}$$

where g is gravity pointing downward in world frame w, ${}^w_b R_i$ is the transformation between b and w at time t_i, and b_i^g, b_i^a and η_i^g, η_i^a are IMU biases and noises respectively. Naively integrating angular rates and accelerations to get positions p_i and orientations ${}^w_b R_i$ leads to significant drift due to sensor noise and unknown biases. We thus turn our attention to neural displacement priors Φ, which regress accurate 2D linear velocities (Herath et al., 2020) or 3D displacement $d \in \mathbb{R}^3$ and covariances $\Sigma \in \mathbb{R}^{3\times3}$ (Liu et al., 2020) from sequences of bias-corrected and gravity aligned IMU measurements

$$\omega_i = {}^g_b R_i(\tilde{\omega}_i - b^g) \qquad a_i = {}^g_b R_i(\tilde{a}_i - b^a) \tag{2}$$

where ${}^g_b R_i$ aligns the z-axis with gravity, and is defined as ${}^g_b R_i = R_\gamma \, {}^w_b R_i$ with some unobservable yaw rotation R_γ. The neural displacement prior has the form

$$d, \Sigma = \Phi\left(\{(a_i, \omega_i)\}_{i=1}^n\right) \tag{3}$$

where $d \in \mathbb{R}^3$ denotes displacement on the time interval $[t_1, t_n]$, and $\Sigma \in \mathbb{R}^{3\times3}$ denotes associated covariance prediction. For instance, Liu et al. (2020), uses these network predictions as measurements and fuses them in an EKF estimating the IMU state in w, *i.e.* orientation, position, velocity, and IMU biases. Preliminaries on terms used in inertial odometry are included in App. A.2.2 and details of EKF and IMU measurement model are included in App. A.6.

We simplify the learning of informative priors by suitably canonicalizing the IMU measurements in two steps: First, we gravity-align IMU measurements by rotating them into the frame ${}^g_b R$, such that the z-axis of the IMU frame and world frame coincide. We use the EKF orientation state from (Liu et al., 2020) to find the gravity direction (see App. A.6). Later we empirically show the robustness of our method to noise originating from this estimation. In what follows we thus

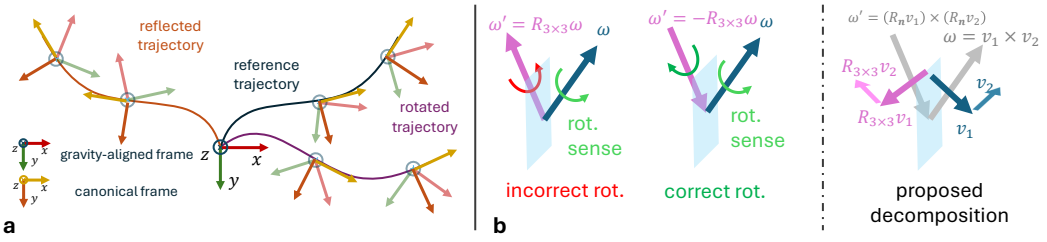

Figure 3: Symmetries in neural inertial odometry. (a) An IMU undergoes three trajectories in xy-plane, each related to a reference (blue) via rotation (purple) and/or reflection (orange) around gravity (parallel to the z-axis). At a fixed time, IMU measurements on different trajectories, expressed in the corresponding local gravity-aligned frame (red-green) differ only by an unknown yaw roto-reflection $R_{3\times3}$. Mapping these measurements to a canonical frame (yellow-red) that transforms equivariantly under roto-reflections of the trajectory eliminates this ambiguity enhancing the sample efficiency of downstream neural networks. (b) Expressed in alternative roto-reflected frames, acceleration, and angular rates transform as $a' = R_{3\times3}a$ and $\omega' = \det(R_{3\times3})R_{3\times3}\omega$. Angular rate must follow the right-hand-rule, and thus be also inverted when reflected. To ensure a similar transformation rule as a, we decompose $\omega = v_1 \times v_2$ and process v_1, v_2 instead, which transform as $v'_{1/2} = R_{3\times3}v_{1/2}$

assume accelerations and angular rates to be expressed in the gravity-aligned frame and illustrate these frames in Fig. 3 (a) for three rotated trajectories. Gravity alignment reduces data variability by two degrees of freedom. However, this frame is not unique, since simply rotating it around z or reflecting it across planes parallel to z (applications of rotations or roto-reflections from the groups $SO_g(3) = \{R \in SO(3) | Rg = g\}$ and $O_g(3) = \{R \in O(3) | Rg = g\}$) result in new valid gravity-aligned frames. Hence, secondly, we predict a canonical frame and map the IMU data into this frame, which we later show to be subequivariant to roto-reflections. In what follows we will restrict our discussion to the $O_g(3)$ case but note that, where not explicitly stated, this discussion carries over to $SO_g(3)$ as well. Next, we will introduce our canonicalization procedure to ensure better network generalization.

4 Methodology

Our goal is to predict a canonical yaw frame $F = \Psi(\{(a_i, \omega_i)\}_{i=1}^n) \in O_g(3)$ from data, which generalizes across arbitrary yaw orientations. We use this frame to map IMU data into a canonical frame before giving as input to the NN, and mapping the outputs back, (see Fig. 2 (a)),*i.e.*

$$d', \Sigma' = \Phi(\{(a'_i, \omega'_i)\}_{i=1}^n) \qquad \text{with } a'_i = \rho_a(F^{-1})a_i \text{ and } \omega'_i = \rho_\omega(F^{-1})\omega_i, \tag{4}$$

where a', ω' are expressed in the canonical frame. Finally, we map d', Σ' back to the original frame via $d = \rho_d(F)d'$ and $\text{vec}(\Sigma) = \rho_\Sigma(F)\text{vec}(\Sigma')$. Here vec(.) stacks the columns of Σ into a single vector, and ρ is a homomorphism that maps group elements F to corresponding matrices, called *matrix representations*. These capture the transformation of a, ω, d and Σ under the action of F.

While $\rho_a(F) = \rho_d(F) = F_{3\times3}$, with $F_{3\times3} \in \mathbb{R}^{3\times3}$ being the rotation matrix corresponding to element F, covariances transform as $\rho_\Sigma(F) = F_{3\times3} \otimes F_{3\times3}$, where $\otimes$ is the Kronecker product. Unfortunately, reflections ($\det(F_{3\times3}) = -1$) induce a reflection *and inversion* of angular rates to preserve the right-hand-rule (see Fig. 3 b), *i.e.* $\rho_\omega(F) = \det(F_{3\times3})F_{3\times3}$, and for reflections. As discussed later $\rho_\omega(F) \neq \rho_a(F)$ hinders joint processing of accelerations and angular rates. Next, we will discuss the design of Ψ which ensures generalization across arbitrary yaw rotations.

4.1 Equivariant Frame

Here we derive a property of the frame network Ψ such that it can generalize to arbitrary roto-reflections of the IMU body frame. To generalize, canonical IMU measurement inputs a'_i, ω'_i to the network must look identical under arbitrary roto-reflections $R \in O_g(3)$. Let a_i, ω_i, and a^*_i, ω^*_i denote quantities before and after application of R. Then $a^*_i = \rho_a(R)a_i$ and $\omega^*_i = \rho_\omega(R)\omega_i$. Enforcing identical inputs under both rotations, *i.e.* ${a^*}'_i = a'_i$ we have

$${a^*}'_i = \rho_a({F^*}^{-1})a^*_i \quad {\omega^*}'_i = \rho_\omega({F^*}^{-1})\omega^*_i \tag{5}$$

We see that choosing $F^* = RF$, *i.e.* that F transforms equivariantly leads to

$${a^*}'_i = \rho_a({F^*}^{-1})a^*_i = \rho_a(F^{-1}R^{-1})\rho_a(R)a_i = \rho_a(F^{-1}R^{-1}R)a_i = a^*_i \tag{6}$$

where we have used the fact that ρ_a is a homomorphism. This shows the invariance of a'_i and a similar proof can be done for ω'_i. This equality puts a constraint on the NN that estimates F, namely

$$RF = \Psi(\{(\rho_a(R)a_i, \rho_\omega(R)\omega_i)\}_{i=1}^n) \tag{7}$$

i.e. Ψ must be a function that is *equivariant* with respect to group actions by elements from $O_g(3)$. Since this is a subgroup of $O(3)$ we also say that Ψ must be *subequivariant* with respect to $O(3)$.

In addition, this equivariance property of Ψ induces end-to-end equivariance to predicted displacements $d = \rho_d(F)d'$ and covariances $\text{vec}(\Sigma) = \rho_\Sigma(F)\text{vec}(\Sigma')$. This is because

$$d^* = \rho_d(F^*)d^{*\prime} = \rho_d(RF)d' = \rho_d(R)\rho_d(F)d' = \rho_d(R)d \tag{8}$$

$$\text{vec}(\Sigma^*) = \rho_\Sigma(F^*)\text{vec}(\Sigma'^*) = \rho_\Sigma(RF)\text{vec}(\Sigma') = \rho_\Sigma(R)\rho_\Sigma(F)\text{vec}(\Sigma') = \rho_\Sigma(R)\text{vec}(\Sigma) \tag{9}$$

using, again, the homomorphism of ρ and the fact that d', $\text{vec}(\Sigma')$ are, by construction of equation 6, invariant to rotations by R.

Diagonal Covariance: We show empirically in Sec. 6 that the diagonal parameterization of Σ' aids in stabilization and convergence of the network. Therefore, we assume the displacement uncertainties $\Sigma_{d,xz} = \Sigma_{d,yz} = 0$ and without loss in generality choose $\Sigma' = \text{diag}(e^{2u_x}, e^{2u_y}, e^{2u_z})$, where u_x, u_y, u_z are learnable, as in TLIO (Liu et al., 2020). Since our network predicts $\text{vec}(\Sigma) = \rho_\Sigma(F)\text{vec}(\Sigma')$, where both Σ' and F are learned, the resulting covariance is $\Sigma = F\Sigma' F^T$ (in matrix format). Via the transformation F we can learn arbitrarily rotated Σ in the xy-plane. We posit that this forces the frame network Ψ to learn F that aligns with the principle axes of the statistical uncertainty in displacement Σ_d. See App. A.2.3 for details on covariance parameterizations. Writing the singular value decomposition (SVD) we see that $\Sigma_d = U\text{diag}(\Sigma_{xx}, \Sigma_{yy}, \Sigma_{zz})U^T$. By inspection, this uncertainty is matched when F aligns with principle directions U and Σ' aligns with the true uncertainties in those directions.

In the next section, let us now discuss the specific issue that arises when designing an equivariant frame network to process both a_i and ω_i, and the specific preprocessing step to remedy it.

4.2 Decomposition of Angular Rates

As previously discussed a and ω transform under different representations $\rho_a \neq \rho_\omega$. This hinders joint feature learning since this would entail forming linear combinations of a and ω, and these linear combinations will not transform under ρ_a or ρ_ω. We propose a preprocessing step that decomposes ω_i into perpendicular vectors $v_{1,i}, v_{2,i}$ via a bijection $\mathcal{F} : \mathbb{R}^3 \to \mathbb{R}^3 \times \mathbb{R}^3$:

$$\mathcal{F}(\omega) = (v_1, v_2) = \left(\sqrt{\|\omega\|}\frac{w_1}{\|w_1\|}, \sqrt{\|\omega\|}\frac{w_2}{\|w_2\|}\right) \qquad \mathcal{F}^{-1}(v_1, v_2) = \omega = v_1 \times v_2 \tag{10}$$

We define $w_1 = [-\omega_y\, \omega_x\, 0]^T$ and $w_2 = \omega \times w_1$. If $\omega_x = \omega_y = 0$, we use $w_1 = a \times \omega$ and if both $\omega_x = \omega_y = 0$ and $a \times \omega = \mathbf{0}$, we use $w_1 = \omega \times [1\, 0\, 0]^T$.

Fig. 3 (b) shows that v_1 and v_2 transform with representation $\rho_{v_1}(F) = \rho_{v_2}(F) = F_{3\times3}$. Let variables with $*$ denote transformed vectors according to rotation R. Their cross product has the desirable property $\omega^* = v_1^* \times v_2^* = (R_{3\times3}v_1) \times (R_{3\times3}v_2) = \det(R_{3\times3})R_{3\times3}(v_1 \times v_2) = \det(R_{3\times3})R_{3\times3}\omega = \rho_\omega(R)\omega$, using the standard cross-product property $(Ax) \times (Ay) = \det(A)A(x \times y)$, and recalling that $R_{3\times3}$ is the matrix representation of R. The group action on ω exactly coincides with what was derived in Sec. 3. We only use this decomposition when with $O_g(3)$, where we process a, v_1, v_2 in a unified way. For $SO_g(3)$, we process a, ω which transform similarly since $\det(R_{3\times3}) = 1$.

4.3 Transition to $O(2)$ Equivariance and Basic Network Layers

Expressed in the gravity-aligned frame, representations $R_{3\times3}$ of $R \in O_g$ leave the z-axis unchanged, and can thus be decomposed into $R_{3\times3} = R_{2\times2} \oplus 1$, where the direct sum $\oplus$ constructs a block-diagonal matrix of its arguments. This decomposition motivates the decomposition $a_i = a_{i,xy} \oplus a_{i,z}$ and $v_{1/2,i} = v_{1/2,i,xy} \oplus v_{1/2,i,z}$, where $\oplus$ concatenates the xy-coordinates of each vector which transform with representation $R_{2\times2}$ and its z component which transforms with representation 1, *i.e.* z is invariant. This means that the xy-components transform according to representations of group $O(2)$ and implies that $O(2) \cong O_g(3)$. Inspired by Villar et al. (2021), we design our frame network to learn universally $O(2)$ equivariant outputs from invariant features alongside 2D vector features.

We convert the sequence of n IMU measurements into $n \times C_0^s$ rotation invariant scalar features and $n \times 2 \times C_0^v$ vector features. As vector features we select the xy-components of each input

vector ($C_0^v = 2$ for $SO(2)$ corresponding with $a_{i,xy}, \omega_{i,xy}$ and $C_0^v = 3$ for $O(2)$ corresponding with $a_{i,xy}, v_{1,i,xy}, v_{2,i,xy}$). Instead, as scalar features we select *(i)* the z-components of each vector, *(ii)* the norm of the xy-components of each vector, and *(iii)* the pairwise dot-product of the xy-components of each vector. For $SO(2)$ we have $C_0^s = 2+2+1 = 5$, while for $O(2)$ we have $C_0^s = 3+3+3 = 9$. We process scalar features with MLPs and standard 1-D convolutions, vector features with specific linear and convolution layers, and combine both with specialized non-linear layers.

Equivariant Linear Layer Following Villar et al. (2021), we design a $2D$ version of vector neuron (Deng et al., 2021) to process the vector features, enhancing efficiency. Following Finzi et al. (2021), we consider learnable linear mappings $v^{\text{out}} = Wv^{\text{in}}$, with input and output vector features $v^{\text{in}}, v^{\text{out}} \in \mathbb{R}^2$ and seek a basis of weights $W \in \mathbb{R}^{2\times 2}$, which satisfy $R_{2\times 2}Wv = WR_{2\times 2}v$, i.e., equivariantly transform vector features $v \in \mathbb{R}^2$. This relation yields the constraint

$$(R_{2\times 2} \otimes R_{2\times 2})\text{vec}(W) = \text{vec}(W), \tag{11}$$

Solving the above equation amounts to finding the eigenspace of the left-most matrix with eigenvalue 1. Such analysis for $R_{2\times 2} \in SO(2)$ yields $W_{SO(2)} = w_1 I_{2\times 2} + w_2 R_{90}$, where R_{90} denotes a 90 degree counter-clockwise rotation in 2D, and $w_1, w_2 \in \mathbb{R}$ are learnable weights. Similarly, for $O(2)$ we find $W_{O(2)} = w_1 I_{2\times 2}$. Vectorizing this linear mapping to multiple input and output vector features we have the following $SO(2)$ and $O(2)$ equivariant linear layers:

$$SO(2): \quad v^{\text{out}} = v^{\text{in}}W_1 + R_{90}v^{\text{in}}W_2 \qquad O(2): \quad v^{\text{out}} = v^{\text{in}}W_1 \tag{12}$$

with $v^{\text{in}} \in \mathbb{R}^{2\times C_{\text{in}}}$, $v^{\text{out}} \in \mathbb{R}^{2\times C_{\text{out}}}$ and $W_1, W_2 \in \mathbb{R}^{C_{\text{in}}\times C_{\text{out}}}$. Note that the $SO(2)$ layer has twice as many parameters as the $O(2)$ layer. We stack the above linear components into a kernel to design equivariant 1-D convolution layers. Since the IMU data forms a time sequence, we implement convolutions across time. We visualize our Linear and Convolutional Layers in Fig. 2.

Nonlinear Layer Previous works (Weiler et al., 2018; Weiler & Cesa, 2019) propose various nonlinearities such as norm-nonlinearity, tensor-product nonlinearity, and gated nonlinearity for $SO(3)$ and $O(2)$ equivariance in an equivariant convolutional way; while Deng et al. (2021) applies per-point nonlinearity for vector features only. Since we already apply convolutions over time we simply apply a non-linearity pointwise. Unlike Deng et al. (2021), we need to mix scalar and vector features and thus adapt the gated nonlinearity (Weiler et al., 2018) to pointwise nonlinearity. Specifically, for n vector and scalar features $v^{\text{in}} \in \mathbb{R}^{n\times 2\times C}$, $s^{\text{in}} \in \mathbb{R}^{n\times C}$, we concatenate the norm features $\|v^{\text{in}}\| \in \mathbb{R}^{n\times C}$ with s^{in}. We run a single MLP with an output of size $n \times 2C$, and split it into new norm features $\gamma \in \mathbb{R}^{n\times C}$ and new activations $\beta \in \mathbb{R}^{n\times C}$ which we modulate with a non-linearity $s^{\text{out}} = \sigma(\beta)$. Finally, we rescale the original vector features according to the new norm:

$$\gamma, \beta = \text{mlp}(\|v^{\text{in}}\| \oplus_c s^{\text{in}}) \qquad v^{\text{out}} = \gamma v^{\text{in}} \qquad s^{\text{out}} = \sigma(\beta) \tag{13}$$

where $\oplus_c$ concatenates along the feature dimension. See Fig. 2b for more details.

5 Experiments

We apply our framework to two types of neural inertial navigation systems: *(i)* an end-to-end deep learning approach (RONIN), and *(ii)* a filter-based approach with a learned prior (TLIO). Both networks process IMU samples in a gravity-aligned frame without gravity compensation, i.e., removing the gravity vector from the accelerometer reading. While RONIN regresses only a 2D velocity, TLIO estimates the orientation, position, velocity, and IMU biases using an EKF in 3D, which propagates states using raw IMU measurements and applies measurement updates with predicted displacement and uncertainty from a NN. Sec. 6 presents extensive ablations.

Datasets: Our TLIO variant is trained on the TLIO Dataset (Liu et al., 2020) and tested on TLIO and Aria Everyday Activities (Aria) Datasets (Lv et al., 2024). Our RONIN variant is trained on RONIN Dataset (Herath et al., 2020). We train on the 50% open-sourced data. We test our RONIN variant on three popular pedestrian datasets RONIN (Herath et al., 2020), RIDI (Yan et al., 2018) and OxIOD (Chen et al., 2018b), which specifically target 2D trajectory tracking. RONIN-U (Unseen) contains IMU measurements from people who did not participate in the training and validation data collection. The people used to record RONIN-S (Seen) overlap with those from the training and validation set, but their data is disjoint from these sets. See App.A.3 for more dataset details, and App.A.5 and Fig. 2 for more details and visualizations of the equivariant network.

Baselines: We compare EqNIO with TLIO with yaw augmentation (Liu et al., 2020), on the 3D benchmarks, and RONIN and RIO (Cao et al., 2022) on the 2D benchmarks. We also report the naive

Model	TLIO Dataset						Aria Dataset					
	MSE* ($10^{-2}m^2$)	ATE (m)	ATE* (m)	RTE (m)	RTE* (m)	AYE (deg)	MSE* ($10^{-2}m^2$)	ATE (m)	ATE* (m)	RTE (m)	RTE* (m)	AYE (deg)
TLIO	3.333	1.722	3.079	0.521	0.542	2.366	15.248	1.969	4.560	0.834	0.977	2.309
+ rot. aug.	3.242	1.812	3.722	0.500	0.551	2.376	5.322	1.285	2.103	0.464	0.521	2.073
+ **SO(2) Eq. Frame**	3.194	1.480	2.401	0.490	0.501	2.428	2.457	1.178	1.864	0.449	0.484	2.084
+ **O(2) Eq. Frame**	2.982	1.433	2.382	0.458	0.479	2.389	2.304	1.118	1.850	0.416	0.465	2.059

Table 1: Trajectory errors, lower being better. + rot. aug. is trained with yaw augmentations. Lowest, and second lowest values are marked in red and orange. * no EKF.

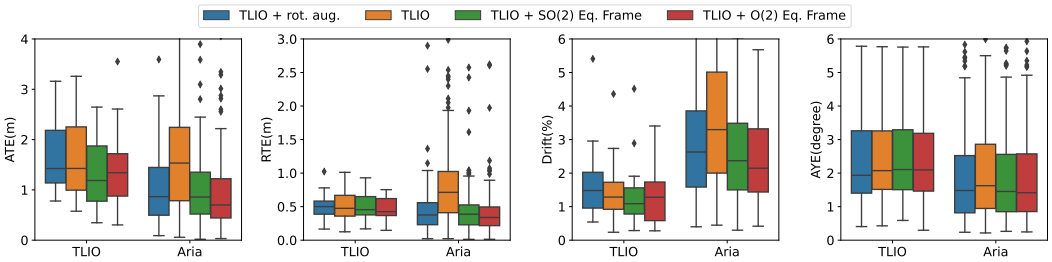

Figure 4: Trajectory errors for EqNIO applied to TLIO compared to vanilla TLIO trained with and without yaw augmentations on TLIO and Aria Datasets visualized with a box plot.

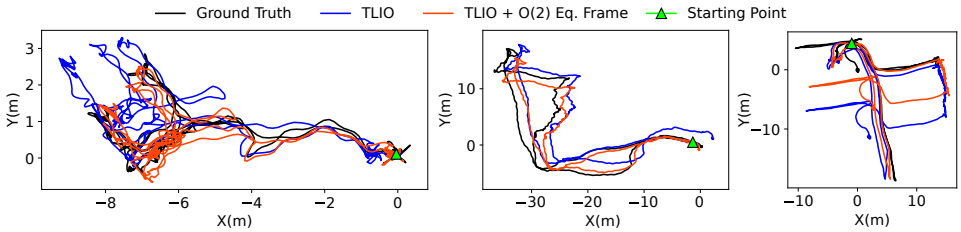

Figure 5: Groundtruth (black), and predicted trajectories on the TLIO Dataset by TLIO (Blue), and EqNIO (Red). Left and right have high, and the middle has medium difficulty.

IMU double integration (NDI) from Herath et al. (2020), and RONIN results trained on 100% of the RONIN training dataset. All other methods use only the public 50% of this dataset. RIO extends RONIN with two features: Joint optimization of an MSE loss on velocity predictions and cosine similarities with an equivariance constraint modeled using an auxiliary loss ($+J$) and an Adaptive Test-Time-Training strategy ($+TTT$). Finally, $+J+TTT$ combines both.

Metrics: The NN performance (indicated with $*$) is reported in terms of Mean Squared Error (MSE) in $10^{-2}m^2$, Absolute Translation Error (ATE) in m, and Relative Translation Error (RTE) in m, on the cumulative sum of the displacements predicted by the network as done in prior work. For TLIO, we also report the ATE in m, RTE in m, and Absolute Yaw Error (AYE) in degrees when running the EKF. See App. A.7 for more metric details. In what follows, we abbreviate EqNIO with +Eq. Frame.

5.1 Results using the TLIO Architecture

Tab. 1 compares baseline TLIO, trained with yaw augmentations as in (Liu et al., 2020) (+ rot.aug.), TLIO without augmentation (termed TLIO), and our two methods applied to TLIO without yaw augmentations (termed +$SO(2)$ and +$O(2)$ Eq. Frame). As seen in Tab. 1, +$O(2)$ Eq. Frame outperforms TLIO on metrics that ignore the EKF (with *) by a large margin of 57%, 12%, and 11% on MSE*, ATE*, and RTE* respectively. The +$SO(2)$ Eq. Frame model follows closely with 54%, 11%, and 7% respectively on the Aria Dataset. The performance of our methods is consistent across TLIO and Aria Datasets illustrating our generalization ability. Tab. 1 and Fig. 4 show our method surpasses the baseline on most metrics while remaining comparable in AYE. The superior performance of our model as compared to baseline TLIO when the NN is combined with EKF (*i.e.*, performance on ATE, RTE, and AYE metrics) is attributed to its generalization ability when the orientation estimate is not very accurate as well as the equivariant covariance predicted by the network. See Fig. 5 and App. A.8 for trajectory plots from the TLIO test dataset.

5.2 Results using the RONIN Architecture

On the RONIN dataset, we compare against RONIN and RIO† (indicated with +J, +TTT, and +J+TTT). RIO does not provide results for the RONIN Seen Dataset (RONIN-S) or RIDI Cross

Model (RONIN)	RONIN-U ATE* (m)	RONIN-U RTE* (m)	RONIN-S ATE* (m)	RONIN-S RTE* (m)	RIDI-T ATE* (m)	RIDI-T RTE* (m)	RIDI-C ATE* (m)	RIDI-C RTE* (m)	OxIOD ATE* (m)	OxIOD RTE* (m)
+ 100% data	5.14	4.37	3.54	2.67	1.63	1.91	1.67	1.62	3.46	4.39
+ 50% data †	5.57	4.38	-	-	1.19	1.75	-	-	3.52	4.42
+ 50% data + J †	5.02	4.23	-	-	1.13	1.65	-	-	3.59	4.43
+ 50% data + TTT †	5.05	4.14	-	-	1.04	1.53	-	-	2.92	3.67
+ 50% data + J +TTT †	5.07	4.17	-	-	1.03	1.51	-	-	2.96	3.74
+ 50% data + **SO(2) Eq. Frame**	5.18	4.35	3.67	2.72	0.86	1.59	0.63	1.39	1.22	2.39
+ 50% data + **O(2) Eq. Frame**	4.42	3.95	3.32	2.66	0.82	1.52	0.70	1.41	1.28	2.10
Naive Double Integration (NDI)	458.06	117.06	675.21	1.6948	31.06	37.53	32.01	38.04	1941.41	848.55

Table 2: Trajectory errors with the RONIN architecture (lower is better), on the RONIN Unseen (-U), Seen (-S) (people used to record the test dataset are the same as train and validation), RIDI Test (-T), Cross Subject (-C), and the Oxford Inertial Odometry Datasets (OxIOD). Red (lowest), orange (second lowest). † results from Cao et al. (2022). * indicates no-EKF.

Model	TLIO Dataset MSE* ($10^{-2}m^2$)	TLIO ATE (m)	TLIO ATE* (m)	TLIO RTE (m)	TLIO RTE* (m)	TLIO AYE (deg)	Aria Dataset MSE* ($10^{-2}m^2$)	Aria ATE (m)	Aria ATE* (m)	Aria RTE (m)	Aria RTE* (m)	Aria AYE (deg)
TLIO	3.333	1.722	3.079	0.521	0.542	2.366	15.248	1.969	4.560	0.834	0.977	2.309
+ rot. aug.	3.242	1.812	3.722	0.500	0.551	2.376	5.322	1.285	2.102	0.464	0.521	2.073
+ rot. aug. + more layers	3.047	1.613	2.766	0.524	0.519	2.397	2.403	1.189	2.541	0.472	0.540	2.081
+ rot. aug. + Non Eq. Frame	3.008	1.429	2.443	0.495	0.496	2.411	2.437	1.213	2.071	0.458	0.508	2.096
+ rot. aug. + PCA Frame	3.473	1.506	2.709	0.523	0.535	2.459	6.558	1.717	4.635	0.771	0.976	2.232
+ $SO(2)$ Eq. Frame + S	3.331	1.626	2.796	0.524	0.536	2.440	2.591	1.146	2.067	0.466	0.517	2.089
+ $SO(2)$ Eq. Frame + P	3.298	1.842	2.652	0.588	0.523	2.537	2.635	1.592	2.303	0.585	0.539	2.232
+ **SO(2) Eq. Frame**	3.194	1.480	2.401	0.490	0.501	2.428	2.457	1.178	1.864	0.449	0.484	2.084
+ $O(2)$ Eq. Frame + S	3.061	1.484	2.474	0.462	0.481	2.390	2.421	1.175	1.804	0.421	0.458	2.043
+ $O(2)$ Eq. Frame + P	2.990	1.827	2.316	0.578	0.478	2.534	2.373	1.755	1.859	0.564	0.468	2.223
+ **O(2) Eq. Frame**	2.982	1.433	2.382	0.458	0.479	2.389	2.304	1.118	1.849	0.416	0.465	2.059
Eq CNN	3.194	1.580	3.385	0.564	0.610	2.394	8.946	3.223	6.916	1.091	1.251	2.299

Table 3: Ablations with the TLIO architecture, lower is better. We test non-equivariant (+Non Eq. Frame), PCA-based (+PCA Frame), $SO(2)$ equivariant (+$SO(2)$ Eq. Frame), and $O(2)$ equivariant (+$O(2)$ Eq. Frame) frames, and yaw augmentation (+ rot. aug.). We also test xy-isotropic (+S) and Pearson-based (+P) covariance parameterizations. Eq CNN is a fully equivariant CNN. Red (lowest), orange (second lowest), and yellow (third lowest). * indicates no-EKF.

Subject Dataset (RIDI-C). As the RONIN base model does not use an EKF, we only report metrics with $*$. As Seen in Tab. 2, our methods significantly outperform the original RONIN by a large margin of 14% and 9% on ATE* and RTE* respectively even on the RONIN-U dataset. Our methods have better generalization as seen on RIDI-T and OxIOD, outperforming even +J+TTT† by a margin of 56% and 43% on ATE* and RTE* respectively on OxIOD Dataset. The +$O(2)$ Eq. Frame model converges at 38 epochs compared to over 100 in RONIN implying faster network convergence with our framework as compared to data augmentation. This demonstrates superior generalization of our strictly equivariant architecture. RIO's approach, involving multiple data rotations, test optimization, and deep ensemble at test time, would result in higher computational and memory costs as compared to our method. Finally, the comparison with NDI highlights the need for a neural displacement prior.

6 Ablation Study

Here, we show the necessity of incorporating equivariance in inertial odometry, the choice of equivariant architecture, and covariance parameterization. We present all the ablations using the TLIO base model in Tab. 3, both with and without integrating the EKF. App. A.9 further contains the performance of all models above on a test dataset which is augmented with rotations and/or reflections. App. A.12, A.13, A.15, A.14 present sensitivity studies on the input sequence length, estimated gravity direction, IMU biases and IMU sampling rate.

Baseline Ablation: **Is yaw augmentation needed when the input is in a local gravity-aligned frame?** We trained TLIO both with and without yaw augmentation using identical hyperparameters and the results in Tab. 3 (rows 1 and 2) reveal that augmentation enhances the network's generalization, improving all metrics for the Aria dataset with the lowest margin of 10% on AYE and highest margin of 65% for MSE*. This underscores the importance of equivariance for network generalization. **Does a Deeper TLIO with a comparable number of parameters match the performance of equivariant methods?** We enhanced the residual depth of the original TLIO architecture from 4

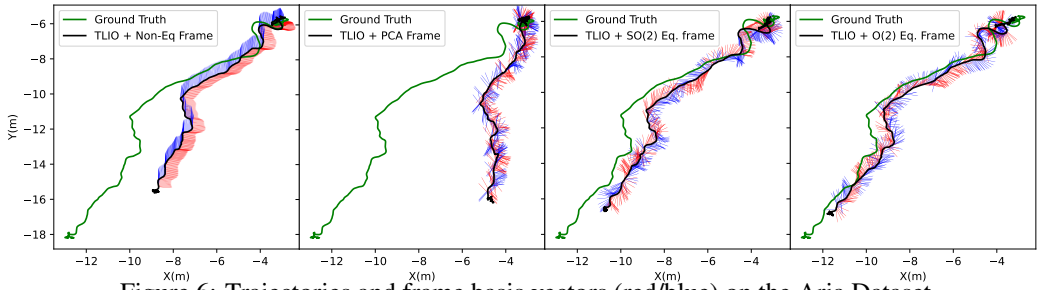

Figure 6: Trajectories and frame basis vectors (red/blue) on the Aria Dataset.

residual blocks of depth 2 each to depth 3 each (row 3) to match the number of parameters with our $+SO(2)$ Eq. Frame model (row 8). Despite having fewer parameters due to the removal of the orthogonal basis in $SO(2)$ vector neuron-based architecture, $+O(2)$ Eq. Frame model (row 11) still outperformed the deeper TLIO. The data from Tab. 3 show that merely increasing the network's size, without integrating true equivariance, is insufficient for achieving precise inertial odometry.

Frame Ablation: **Can a non-equivariant MLP predict meaningful frames?** We trained TLIO with yaw augmentation and identical hyperparameters alongside an additional MLP mirroring the architecture of our method to predict a frame and term this baseline +Non Eq. Frame (row 4). We observed that +Non Eq. Frame tends to overfit to the TLIO dataset, and thus produce worse results on the Aria dataset. The predicted frames also poorly correlate with the underlying trajectory, as illustrated in Fig. 6. **Can frames predicted using PCA (handcrafted equivariant frame) achieve the same performance?** Using PCA to generate frames leads to underperformance on the Aria dataset, and worse results than the original TLIO which is likely due to PCA's noise sensitivity as shown in Fig. 6. Additionally, PCA cannot distinguish between $SO(2)$ and $O(2)$ transformations. Fig. 6 also shows that $O(2)$ does not have frames as smooth as $SO(2)$ as the reflected bends have reflected frames. See App. A.16 for an additional ablation using frame-averaging for canonicalization.

Architecture Ablation: **Are fully equivariant architectures better ?** We trained a fully equivariant 1-D CNN using the layers in Sec. 4.3. Tab. 3 (row 12) shows our frame-based methods outperforming the equivariant CNNs, likely by leveraging the power of scalars and conventional backbones. We believe the fully equivariant architecture is overly restrictive, while also requiring a full network redesign. By contrast, our method can flexibly integrate existing state-of-the-art displacement priors.

Covariance Ablation: **Do we need equivariant covariance?** We investigated the importance of equivariant covariance for both $SO(2)$ and $O(2)$ groups, as described in Sec. 4.1(See App. A.2.3 for covariance parameterizations). In Tab. 3, the models +S (rows 6 and 9) are trained with invariant covariance parameterized as $\Sigma = \text{diag}(e^{2u_x}, e^{2u_y}, e^{2u_z})$, that is unaffected by application of F. The results show that the equivariant covariance yields better performance, especially when combined with EKF, as it provides a more accurate estimate of the prediction covariance. **Can a full covariance matrix with Pearson parameterization improve performance?** In Tab. 3, our model outperforms +P (rows 7 and 10) in most cases indicating that by aligning the principal covariance axis into the basis of the equivariant frame, we intrinsically force covariance in the equivariant frame to be diagonal, which reduces ambiguity while training. Diagonal covariances improve convergence stability during optimization as stated in Liu et al. (2020). App. A.11 visualizes the covariance consistency of EqNIO and we conduct an analysis on the covariance parameterization in App. A.17.

7 Conclusion

We introduce a robust and generalizable neural displacement prior that combats drift in IMU-only neural inertial odometry via equivariant canonicalization. Our canonicalization scheme is generally applicable and eliminates the underlying yaw ambiguity in gravity-aligned frames which arise from roto-reflections in the plane around gravity. We fully characterize actions from this group on all relevant inputs and outputs of the prior and leverage this insight to design a network that produces learned frames that are $O_g(3)$ equivariant to these actions. By reducing the data variability seen by neural networks these frames boost the generalization of existing networks and enforce exact equivariance, unlike existing strategies that use data augmentation or equivariant consistency losses to enforce approximate equivariance. We demonstrate the generality of our framework through extensive validation on various datasets and applications to two base architectures (TLIO and RONIN). We believe this work paves the way for robust, and low-drift odometry running on edge devices.

ACKNOWLEDGEMENTS

We gratefully acknowledge support by the following grants: NSF FRR 2220868, NSF IIS-RI 2212433, and ONR N00014-22-1-2677.

REFERENCES

Arman Asgharpoor Golroudbari and Mohammad Hossein Sabour. Generalizable end-to-end deep learning frameworks for real-time attitude estimation using 6dof inertial measurement units. *Measurement*, 217:113105, 2023. ISSN 0263-2241. doi: https://doi.org/10.1016/j.measurement.2023.113105. URL https://www.sciencedirect.com/science/article/pii/S0263224123006693.

Omri Asraf, Firas Shama, and Itzik Klein. Pdrnet: A deep-learning pedestrian dead reckoning framework. *IEEE Sensors Journal*, 22(6):4932–4939, 2022. doi: 10.1109/JSEN.2021.3066840.

Justin Baker, Shih-Hsin Wang, Tommaso de Fernex, and Bao Wang. An explicit frame construction for normalizing 3d point clouds. In *Forty-first International Conference on Machine Learning*, 2024.

Bertrand Beaufils, Frédéric Chazal, Marc Grelet, and Bertrand Michel. Robust stride detector from ankle-mounted inertial sensors for pedestrian navigation and activity recognition with machine learning approaches. *Sensors*, 19(20):4491, 2019.

John E. Bortz. A new mathematical formulation for strapdown inertial navigation. *IEEE Transactions on Aerospace and Electronic Systems*, AES-7(1):61–66, 1971. doi: 10.1109/TAES.1971.310252.

Tarek Bouazza, Robert Mahony, and Tarek Hamel. Exploiting polar symmetry in designing equivariant observers for vision-based motion estimation. *IEEE Control Systems Letters*, 8:1192–1197, 2024. doi: 10.1109/LCSYS.2024.3406227.

Agata Brajdic and Robert Harle. Walk detection and step counting on unconstrained smartphones. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pp. 225–234, 2013.

Martin Brossard, Axel Barrau, and Silvère Bonnabel. Ai-imu dead-reckoning. *IEEE Transactions on Intelligent Vehicles*, 5(4):585–595, 2020a. doi: 10.1109/TIV.2020.2980758.

Martin Brossard, Silvère Bonnabel, and Axel Barrau. Denoising imu gyroscopes with deep learning for open-loop attitude estimation. *IEEE Robotics and Automation Letters*, 5(3):4796–4803, 2020b. doi: 10.1109/LRA.2020.3003256.

James Brotchie, Wei Shao, Wenchao Li, and Allison Kealy. Leveraging self-attention mechanism for attitude estimation in smartphones. *Sensors*, 22(22), 2022. ISSN 1424-8220. doi: 10.3390/s22229011. URL https://www.mdpi.com/1424-8220/22/22/9011.

Russell Buchanan, Varun Agrawal, Marco Camurri, Frank Dellaert, and Maurice Fallon. Deep imu bias inference for robust visual-inertial odometry with factor graphs. *IEEE Robotics and Automation Letters*, 8(1):41–48, 2023. doi: 10.1109/LRA.2022.3222956.

Xiya Cao, Caifa Zhou, Dandan Zeng, and Yongliang Wang. Rio: Rotation-equivariance supervised learning of robust inertial odometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6614–6623, 2022.

Gabriele Cesa, Leon Lang, and Maurice Weiler. A program to build e (n)-equivariant steerable cnns. In *International conference on learning representations*, 2021.

Changhao Chen, Xiaoxuan Lu, Andrew Markham, and Niki Trigoni. Ionet: Learning to cure the curse of drift in inertial odometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018a.

Changhao Chen, Peijun Zhao, Chris Xiaoxuan Lu, Wei Wang, Andrew Markham, and Niki Trigoni. Oxiod: The dataset for deep inertial odometry. *CoRR*, abs/1809.07491, 2018b. URL http://arxiv.org/abs/1809.07491.

Changhao Chen, Yishu Miao, Chris Xiaoxuan Lu, Linhai Xie, Phil Blunsom, Andrew Markham, and Niki Trigoni. Motiontransformer: Transferring neural inertial tracking between domains. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 8009–8016, 2019.

Chuchu Chen, Yulin Yang, Patrick Geneva, and Guoquan Huang. Fej2: A consistent visual-inertial state estimator design. In *2022 International Conference on Robotics and Automation (ICRA)*, pp. 9506–9512, 2022. doi: 10.1109/ICRA46639.2022.9811831.

Danpeng Chen, Nan Wang, Runsen Xu, Weijian Xie, Hujun Bao, and Guofeng Zhang. Rnin-vio: Robust neural inertial navigation aided visual-inertial odometry in challenging scenes. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 275–283. IEEE, 2021a.

Haiwei Chen, Shichen Liu, Weikai Chen, Hao Li, and Randall Hill. Equivariant point network for 3d point cloud analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14514–14523, 2021b.

Runfa Chen, Jiaqi Han, Fuchun Sun, and Wenbing Huang. Subequivariant graph reinforcement learning in 3d environments. In *International Conference on Machine Learning*, pp. 4545–4565. PMLR, 2023.

Antônio C. B. Chiella, Bruno O. S. Teixeira, and Guilherme A. S. Pereira. Quaternion-based robust attitude estimation using an adaptive unscented kalman filter. *Sensors*, 19(10), 2019. ISSN 1424-8220. doi: 10.3390/s19102372. URL https://www.mdpi.com/1424-8220/19/10/2372.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral cnn. In *International conference on Machine learning*, pp. 1321–1330. PMLR, 2019a.

Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. *arXiv preprint arXiv:1801.10130*, 2018.

Taco S Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant cnns on homogeneous spaces. *Advances in neural information processing systems*, 32, 2019b.

Giulio Delama, Alessandro Fornasier, Robert Mahony, and Stephan Weiss. Equivariant imu preintegration with biases: an inhomogeneous galilean group approach, 2024. URL https://arxiv.org/abs/2411.05548.

Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas J Guibas. Vector neurons: A general framework for so (3)-equivariant networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12200–12209, 2021.

Zihao Dong, Jeff Pflueger, Leonard Jung, David Thorne, Philip R. Osteen, Christa S. Robison, Brett T. Lopez, and Michael Everett. Lidar inertial odometry and mapping using learned registration-relevant features, 2024. URL https://arxiv.org/abs/2410.02961.

Weitao Du, He Zhang, Yuanqi Du, Qi Meng, Wei Chen, Nanning Zheng, Bin Shao, and Tie-Yan Liu. Se (3) equivariant graph neural networks with complete local frames. In *International Conference on Machine Learning*, pp. 5583–5608. PMLR, 2022.

Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning so (3) equivariant representations with spherical cnns. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 52–68, 2018.

Carlos Esteves, Yinshuang Xu, Christine Allen-Blanchette, and Kostas Daniilidis. Equivariant multi-view networks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1568–1577, 2019.

Carlos Esteves, Ameesh Makadia, and Kostas Daniilidis. Spin-weighted spherical cnns. *Advances in Neural Information Processing Systems*, 33:8614–8625, 2020.

Carlos Esteves, Jean-Jacques Slotine, and Ameesh Makadia. Scaling spherical cnns. *arXiv preprint arXiv:2306.05420*, 2023.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *International conference on machine learning*, pp. 3318–3328. PMLR, 2021.

Alessandro Fornasier, Pieter van Goor, Eren Allak, Robert Mahony, and Stephan Weiss. Msceqf: A multi state constraint equivariant filter for vision-aided inertial navigation. *IEEE Robotics and Automation Letters*, 9(1):731–738, 2024. doi: 10.1109/LRA.2023.3335775.

Eric Foxlin. Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Computer graphics and applications*, 25(6):38–46, 2005.

Yixiao Ge, Behzad Zamani, Pieter van Goor, Jochen Trumpf, and Robert Mahony. Geometric data fusion for collaborative attitude estimation. *IFAC-PapersOnLine*, 58(17):392–397, 2024. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2024.10.201. URL https://www.sciencedirect.com/science/article/pii/S2405896324019566. 26th International Symposium on Mathematical Theory of Networks and Systems MTNS 2024.

Paul D Groves. Principles of gnss, inertial, and multisensor integrated navigation systems, [book review]. *IEEE Aerospace and Electronic Systems Magazine*, 30(2):26–27, 2015.

James K. Hall, Nathan B. Knoebel, and Timothy W. McLain. Quaternion attitude estimation for miniature air vehicles using a multiplicative extended kalman filter. In *2008 IEEE/ION Position, Location and Navigation Symposium*, pp. 1230–1237, 2008. doi: 10.1109/PLANS.2008.4570043.

Jiaqi Han, Wenbing Huang, Hengbo Ma, Jiachen Li, Josh Tenenbaum, and Chuang Gan. Learning physical dynamics with subequivariant graph neural networks. *Advances in Neural Information Processing Systems*, 35:26256–26268, 2022.

Ross Hartley, Maani Ghaffari, Ryan M Eustice, and Jessy W Grizzle. Contact-aided invariant extended kalman filtering for robot state estimation. *The International Journal of Robotics Research*, 39(4): 402–430, 2020.

Sachini Herath, Hang Yan, and Yasutaka Furukawa. Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, and new methods. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3146–3152. IEEE, 2020.

Joel A. Hesch, Dimitrios G. Kottas, Sean L. Bowman, and Stergios I. Roumeliotis. Consistency analysis and improvement of vision-aided inertial navigation. *IEEE Transactions on Robotics*, 30 (1):158–176, 2014. doi: 10.1109/TRO.2013.2277549.

Ngoc-Huynh Ho, Phuc Huu Truong, and Gu-Min Jeong. Step-detection and adaptive step-length estimation for pedestrian dead-reckoning at various walking speeds using a smartphone. *Sensors*, 16(9):1423, 2016.

Guoquan P. Huang, Anastasios I. Mourikis, and Stergios I. Roumeliotis. Analysis and improvement of the consistency of extended kalman filter based slam. In *2008 IEEE International Conference on Robotics and Automation*, pp. 473–479, 2008. doi: 10.1109/ROBOT.2008.4543252.

Guoquan P. Huang, Anastasios I. Mourikis, and Stergios I. Roumeliotis. A first-estimates jacobian ekf for improving slam consistency. In Oussama Khatib, Vijay Kumar, and George J. Pappas (eds.), *Experimental Robotics*, pp. 373–382, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-00196-3.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015.

A. Javanmard-Gh., Dorota Iwaszczuk, and Stefan Roth. Deeplio: Deep lidar inertial sensor fusion for odometry estimation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2021. URL https://api.semanticscholar.org/CorpusID:237383396.

Antonio R Jimenez, Fernando Seco, Carlos Prieto, and Jorge Guevara. A comparison of pedestrian dead-reckoning algorithms using a low-cost mems imu. In *2009 IEEE International Symposium on Intelligent Signal Processing*, pp. 37–42. IEEE, 2009.

Xiaofei Jing, Jiarui Cui, Hongtai He, Bo Zhang, Dawei Ding, and Yue Yang. Attitude estimation for uav using extended kalman filter. In *2017 29th Chinese Control And Decision Conference (CCDC)*, pp. 3307–3312, 2017. doi: 10.1109/CCDC.2017.7979077.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, pp. 15546–15566. PMLR, 2023.

Rudolph Emil Kalman and Others. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

Jong-Hyuk Kim and Salah Sukkarieh. Airborne simultaneous localisation and map building. volume 1, pp. 406 – 411 vol.1, 10 2003. ISBN 0-7803-7736-2. doi: 10.1109/ROBOT.2003.1241629.

Robert C. Leishman, John C. Macdonald, Randal W. Beard, and Timothy W. McLain. Quadrotors and accelerometers: State estimation with an improved dynamic model. *IEEE Control Systems Magazine*, 34(1):28–41, 2014. doi: 10.1109/MCS.2013.2287362.

Feiran Li, Kent Fujiwara, Fumio Okura, and Yasuyuki Matsushita. A closer look at rotation-invariant deep point cloud analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16218–16227, 2021.

Wenxin Liu, David Caruso, Eddy Ilg, Jing Dong, Anastasios I Mourikis, Kostas Daniilidis, Vijay Kumar, and Jakob Engel. Tlio: Tight learned inertial odometry. *IEEE Robotics and Automation Letters*, 5(4):5653–5660, 2020.

David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004.

Shitong Luo, Jiahan Li, Jiaqi Guan, Yufeng Su, Chaoran Cheng, Jian Peng, and Jianzhu Ma. Equivariant point cloud analysis via learning orientations for message passing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18932–18941, 2022.

Zhaoyang Lv, Nickolas Charron, Pierre Moulon, Alexander Gamino, Cheng Peng, Chris Sweeney, Edward Miller, Huixuan Tang, Jeff Meissner, Jing Dong, et al. Aria everyday activities dataset. *arXiv preprint arXiv:2402.13349*, 2024.

George Ma, Yifei Wang, Derek Lim, Stefanie Jegelka, and Yisen Wang. A canonicalization perspective on invariant and equivariant learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Robert Mahony, Pieter van Goor, and Tarek Hamel. Observer design for nonlinear systems with equivariance. *Annual Review of Control, Robotics, and Autonomous Systems*, 5(Volume 5, 2022):221–252, 2022. ISSN 2573-5144. doi: https://doi.org/10.1146/annurev-control-061520-010324. URL https://www.annualreviews.org/content/journals/10.1146/annurev-control-061520-010324.

Arnab Kumar Mondal, Siba Smarak Panigrahi, Oumar Kaba, Sai Rajeswar Mudumba, and Siamak Ravanbakhsh. Equivariant adaptation of large pretrained models. *Advances in Neural Information Processing Systems*, 36:50293–50309, 2023.

Anastasios I. Mourikis and Stergios I. Roumeliotis. A multi-state constraint kalman filter for vision-aided inertial navigation. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pp. 3565–3572, 2007. doi: 10.1109/ROBOT.2007.364024.

Paul Newman. On the structure and solution of the simultaneous localisation and map building problem. 1999. URL https://api.semanticscholar.org/CorpusID:86793851.

Yonhon Ng, Pieter van Goor, Tarek Hamel, and Robert Mahony. Equivariant observers for second-order systems on matrix lie groups. *IEEE Transactions on Automatic Control*, 68(4):2468–2474, 2023. doi: 10.1109/TAC.2022.3173926.

Taku Okawara, Kenji Koide, Shuji Oishi, Masashi Yokozuka, Atsuhiko Banno, Kentaro Uno, and Kazuya Yoshida. Tightly-coupled lidar-imu-wheel odometry with online calibration of a kinematic model for skid-steering robots. *IEEE Access*, PP:1–1, 01 2024. doi: 10.1109/ACCESS.2024.3461655.

Youqi Pan, Wugen Zhou, Yingdian Cao, and Hongbin Zha. Adaptive VIO: Deep Visual-Inertial Odometry with Online Continual Learning . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18019–18028, Los Alamitos, CA, USA, June 2024. IEEE Computer Society. doi: 10.1109/CVPR52733.2024.01706. URL https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.01706.

Xiongfeng Peng, Zhihua Liu, Weiming Li, Ping Tan, Soon Yong Cho, and Qiang Wang. Dvi-slam: A dual visual inertial slam network. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 12020–12026, 2024. doi: 10.1109/ICRA57147.2024.10610042.

Omri Puny, Matan Atzmon, Heli Ben-Hamu, Ishan Misra, Aditya Grover, Edward J Smith, and Yaron Lipman. Frame averaging for invariant and equivariant network design. *arXiv preprint arXiv:2110.03336*, 2021.

Yuheng Qiu, Chen Wang, Can Xu, Yutian Chen, Xunfei Zhou, Youjie Xia, and Sebastian Scherer. Airimu: Learning uncertainty propagation for inertial odometry, 2024. URL https://arxiv.org/abs/2310.04874.

Sujatha Rajagopal. Personal dead reckoning system with shoe mounted inertial sensors. *Master's Degree Project, Stockholm, Sweden*, 2008.

Rebecca L Russell and Christopher Reale. Multivariate uncertainty in deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):7937–7943, 2021.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Arno Solin, Santiago Cortes, Esa Rahtu, and Juho Kannala. Inertial odometry on handheld smartphones. In *2018 21st International Conference on Information Fusion (FUSION)*, pp. 1–5. IEEE, 2018.

Jan Steinbrener, Christian Brommer, Thomas Jantos, Alessandro Fornasier, and Stephan Weiss. Improved state propagation through ai-based pre-processing and down-sampling of high-speed inertial data. In *2022 International Conference on Robotics and Automation (ICRA)*, pp. 6084–6090, 2022. doi: 10.1109/ICRA46639.2022.9811989.

Scott Sun, Dennis Melamed, and Kris Kitani. Idol: Inertial deep orientation-estimation and localization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 6128–6137, 2021.

Yong Tang, Jianhua Gong, Yi Li, Guoyong Zhang, Banghui Yang, and Zhiyuan Yang. Wavelet transform-based inertial neural network for spatial positioning using inertial measurement units. *Remote Sensing*, 16(13), 2024. ISSN 2072-4292. doi: 10.3390/rs16132326. URL https://www.mdpi.com/2072-4292/16/13/2326.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

D. Titterton, J.L. Weston, Institution of Electrical Engineers, American Institute of Aeronautics, and Astronautics. *Strapdown Inertial Navigation Technology*. IEE Radar Series. Institution of Engineering and Technology, 2004. ISBN 9780863413582. URL https://books.google.com/books?id=WwrCrn54n5cC.

S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, 1991. doi: 10.1109/34.88573.

Pieter van Goor and Robert Mahony. Eqvio: An equivariant filter for visual-inertial odometry. *IEEE Transactions on Robotics*, 39(5):3567–3585, 2023. doi: 10.1109/TRO.2023.3289587.

Pieter van Goor, Tarek Hamel, and Robert Mahony. Equivariant filter (eqf). *IEEE Transactions on Automatic Control*, 68(6):3501–3512, 2023. doi: 10.1109/TAC.2022.3194094.

Soledad Villar, David W Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. *Advances in Neural Information Processing Systems*, 34:28848–28863, 2021.

Dian Wang, Jung Yeon Park, Neel Sortur, Lawson L. S. Wong, Robin Walters, and Robert Platt. The surprising effectiveness of equivariant models in domains with latent symmetry, 2023. URL https://arxiv.org/abs/2211.09231.

Yifu Wang, Yonhon Ng, Inkyu Sa, Álvaro Parra, Cristian Rodriguez-Opazo, Taojun Lin, and Hongdong Li. Mavis: Multi-camera augmented visual-inertial slam using se2(3) based exact imu pre-integration. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1694–1700, 2024. doi: 10.1109/ICRA57147.2024.10609982.

Daniel Weber, Clemens Gühmann, and Thomas Seel. Riann—a robust neural network outperforms attitude estimation filters. *AI*, 2(3):444–463, 2021. ISSN 2673-2688. doi: 10.3390/ai2030028. URL https://www.mdpi.com/2673-2688/2/3/28.

Maurice Weiler and Gabriele Cesa. General e (2)-equivariant steerable cnns. *Advances in neural information processing systems*, 32, 2019.

Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco S Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. *Advances in Neural Information Processing Systems*, 31, 2018.

Maurice Weiler, Patrick Forré, Erik Verlinde, and Max Welling. Coordinate independent convolutional networks–isometry and gauge equivariant convolutions on riemannian manifolds. *arXiv preprint arXiv:2106.06020*, 2021.

S.B. Williams, P. Newman, G. Dissanayake, and H. Durrant-Whyte. Autonomous underwater simultaneous localisation and map building. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, volume 2, pp. 1793–1798 vol.2, 2000. doi: 10.1109/ROBOT.2000.844855.

Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5028–5037, 2017.

Yinshuang Xu, Jiahui Lei, Edgar Dobriban, and Kostas Daniilidis. Unified fourier-based kernel and nonlinearity design for equivariant networks on homogeneous spaces. In *International Conference on Machine Learning*, pp. 24596–24614. PMLR, 2022.

Yinshuang Xu, Jiahui Lei, and Kostas Daniilidis. $se(3)$ equivariant convolution and transformer in ray space. *Advances in Neural Information Processing Systems*, 36, 2024.

Hang Yan, Qi Shan, and Yasutaka Furukawa. Ridi: Robust imu double integration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

Cem Yüceer and Kemal Oflazer. A rotation, scaling, and translation invariant pattern classification system. *Pattern recognition*, 26(5):687–710, 1993.

Ji Zhang and Sanjiv Singh. Loam: Lidar odometry and mapping in real-time. In *Robotics: Science and Systems*, 2014. URL https://api.semanticscholar.org/CorpusID:18612391.

Chunran Zheng, Wei Xu, Zuhao Zou, Tong Hua, Chongjian Yuan, Dongjiao He, Bingyang Zhou, Zheng Liu, Jiarong Lin, Fangcheng Zhu, Yunfan Ren, Rong Wang, Fanle Meng, and Fu Zhang. Fast-livo2: Fast, direct lidar-inertial-visual odometry, 2024. URL https://arxiv.org/abs/2408.14035.