Equivariant Neural Boundary Operators

Emanuel Ström

1 Introduction

There has recently been a surge in deep-learning based methods for solving partial differential equations (PDE). In this context, a neural network is used to parameterize the PDE solution. Some research like physics-informed neural nets (PINN) focuses on designing unsupervised or semisupervised approaches, in which the PDE (or its weak form) is incorporated into the loss function.

Operator learning is a more recent direction that focuses on architecture design. The main philosophy is that the neural network parameterizes not the PDE solution, but the *solution operator*. The setting is typically quite general, but for the purpose of this work we will narrow it down to a certain problem:

Setting 1. *Suppose a manifold $\mathcal{M}$ of codimension 1, embedded in $\mathbb{R}^d$. We will study a generic class of geometry-dependent operators $A[\,\cdot\,;\mathcal{M}]\colon \mathcal{H}_s(\mathcal{M})^d \to \mathcal{L}(\mathcal{M})^d$, where the equation*

$$A[x;\mathcal{M}] = y$$

has a unique solution $x \in \mathcal{H}_s(\mathcal{M})^d$ for any $y \in \mathcal{L}(\mathcal{M})^d$.

We will specifically study operators that are equivariant to rigid motion transforms $\mathrm{SE}(d)$:

Definition 1. *A family of operatos $A[\cdot;\mathcal{M}]\colon \mathcal{H}_s(\mathcal{M})^d \to \mathcal{L}(\mathcal{M})^d$ is equivariant to $\mathrm{SE}(d)$ if, for any $g = (r,h) \in \mathrm{SE}(d)$ where r is a rotation and h a translation,*

$$A[gx; g\mathcal{M}] = gA[x;\mathcal{M}],$$

where $(gx)(t) = r(x(g^{-1}t))$ and $g\mathcal{M} = \{gt\colon t \in \mathcal{M}\}$.

A^{-1} is equivariant in the same way. Equivalently, A can be seen as a map between curves $\gamma = \{(t, x(t))\}_{t\in\mathcal{M}}$ of the fiber bundle $\mathcal{M} \times \mathbb{R}^d$. The group action $g\gamma$ is then g transforming the entire fiber bundle as $g(t,x) = (gt, rx)$, restricted to γ, and is a more natural interpretation of $A[g\gamma] = gA[\gamma]$.

The goal in this project is to develop a neural architecture that respects euclidean equivariance and can evaluate on *arbitrary* geometries $\mathcal{M}$. The data input (and output) for our architecture will be a generalized curve:

$$\gamma = \{(t, x(t))\colon t \in \mathcal{M}\}.$$

we seek an operator Φ that takes any γ, so that $g\Phi[\gamma] = \Phi[g\gamma]$ for any $g \in \mathrm{SE}(d)$.

2 Gaussian Process Regression (GPR):

Suppose $\mathcal{D} = \{(\mathcal{M}_n, x_n, y_n)\}_{n=1}^N$ is a data set and $y_n = A[x_n; \mathcal{M}_n]$ for each n. Our objective is to learn the map $x_n = A^{-1}[y_n; \mathcal{M}_n]$. Gaussian Process Y over x, with mean μ and kernel q is defined through the distribution of finite samples $\boldsymbol{Y} = (Y(x_1), \dots, Y(x_N))$ evaluated at discrete points $\boldsymbol{x} = (x_1, \dots, x_N)$ by

$$\boldsymbol{Y} \sim \mathcal{N}(\mu(\boldsymbol{x}), q(\boldsymbol{x}; \boldsymbol{x})), \quad \text{where} \quad \mu(\boldsymbol{x})_n = \mu(x_n) \quad \text{and} \quad q(\boldsymbol{x}; \boldsymbol{x})_{n,m} = q(x_n, x_m).$$

The posterior distribution of a new sample $Y(x)$ conditioned on $\boldsymbol{Y}$ is also normal:

$$\begin{aligned} Y(x) \mid \boldsymbol{Y} \sim \mathcal{N}(\mu(x \mid \boldsymbol{x}), q(x \mid \boldsymbol{x})), \quad \text{where} & \\ \mu(x \mid \boldsymbol{x}) = \mu(x) + q(x; \boldsymbol{x})^T Q(\boldsymbol{x}; \boldsymbol{x})^{-1}(\boldsymbol{Y} - \mu(\boldsymbol{x})), \quad \text{and} & \\ q(x \mid \boldsymbol{x}) = q(x; \boldsymbol{x})^T Q(\boldsymbol{x}; \boldsymbol{x})^{-1} q(x; \boldsymbol{x}), & \end{aligned}$$

In the case of MAP-estimation, the posterior mean is used as the prediction, in which case it makes sense to precompute the quantity $Q(\boldsymbol{x}; \boldsymbol{x})^{-1}(\boldsymbol{Y} - \mu(\boldsymbol{x}))$.

GPR for Hilbert Spaces: In the case where Y is a Gaussian process taking values in a Hilbert Space $\mathcal{Y}$, we have for $x, z \in X$ and test functions $h, f \in Y$

$$\mathrm{Cov}(\langle f, Y(x)\rangle, \langle h, Y(z)\rangle) = q(x, z)[f, h]$$

if the spectrum of $q(x, y)$ decays fast, there is an ON-basis $\{\varphi_\ell\}_{\ell=1}^L \subset Y$ that explains most of the variance. Then we can define $\boldsymbol{Y}_L(x) := (\langle \varphi_\ell, Y(x)\rangle)_\ell$, which will be a Gaussian process with mean $\langle \varphi_\ell, \mu(x)\rangle$ and a matrix-valued covariance kernel $q^L_{\ell,\ell'}(x, z) = q(x, z)(\varphi_\ell, \varphi_{\ell'})$.

Equivariant GPR: Let G be a group and $Y(x)$ a Gaussian random variable with mean $\mu(x)$ and covariance kernel $q(x, x')$. Furthermore, denote by gy, gx the action of some $g \in G$ on elements $x \in \mathcal{X}, y \in \mathcal{Y}$, respectively.

Definition 2. *A Gaussian process Y is equivariant in distribution with respect to G, if for any points $\boldsymbol{x} = (x_1, \dots x_N)$ and group elements $\boldsymbol{g} = (g_1, \dots, g_N)$, the processes $\boldsymbol{g}\boldsymbol{Y} = (g_1 Y(x_1), \dots g_N Y(x_N))$ and $\boldsymbol{Y} \circ \boldsymbol{g} = (Y(g_1 x_1), \dots, Y(g_N x_N))$ are equal in distribution.*

Lemma 1. *A Gaussian process Y is G-equinvariant in distribution, if and only if the kernel q and mean μ satisfy*

$$\mu(gx) = g\mu(x), \quad q(gx, hy) = gq(x, y)h^*, \quad \textit{for all } g, h \in G, x, x' \in \mathcal{X}.$$

Proof. Let $\boldsymbol{x} = (x, x')$ and take $\boldsymbol{g} = (g, h)$. Suppose $\boldsymbol{g}Y(\boldsymbol{x})$ and $Y(\boldsymbol{g}\boldsymbol{x})$ are equal in distribution. By the formulas for linear transformations of Gaussian variables, we have

$$\boldsymbol{g}Y(\boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{g}\mu(\boldsymbol{x}), \boldsymbol{g}q(\boldsymbol{x}; \boldsymbol{x})\boldsymbol{g}^*), \quad Y(\boldsymbol{g}\boldsymbol{x}) \sim \mathcal{N}(\mu(\boldsymbol{g}\boldsymbol{x}), q(\boldsymbol{g}\boldsymbol{x}; \boldsymbol{g}\boldsymbol{x})).$$

By the properties of the Gaussian distribution, the two distributions are equal if and only if the mean and covariance are the same in both cases, which reproduces the result from the theorem. □

Lemma 2. *If a Gaussian Process is G-equivariant in distribution, then the conditional probability distribution of $Y(x)$ given $\boldsymbol{Y}$ is also G-equivariant.*

Proof. The result follows from the formula for conditional normal variables, combined with lemma 1. $\square$

Example 1. *Let d be a distance measure that satisfies $d(gx, z) = d(x, g^{-1}z)$ for all $g \in \mathrm{SE}(d)$ and $x, y \in \mathcal{X}$. Define the shift $g(x, y)$ that minimizes this distance:*

$$\widehat{g}(x, y) = \operatorname{argmin}\{d(gx, y) \colon g \in \mathrm{SE}(d)\}$$

and let $\widehat{d}(x, y) = d(\widehat{g}(x, y)x, y)$. Note that $\widehat{d}$ is invariant to $\mathrm{SE}(d)$, since x is in $\mathrm{Orb}_{\mathrm{SE}(d)}(gx)$. Moreover, $\widehat{g}(gx, y) = \widehat{g}(x, y)g^{-1}$, since

$$d(g(x, y)g^{-1}gx, y) = d(g(x, y)x, y) = \widehat{d}(x, y).$$

A similar argument shows that $\widehat{g}(x, hy) = h\widehat{g}(x, y)$. Hence, $\widehat{g}(gx, hy) = h\widehat{g}(x, y)g^{-1}$. Now, define $Q(x, y) := \phi(\widehat{d}(x, z))\widehat{g}(x, y)^{-1}$ for some decreasing function ϕ.

$$Q(gx, hy) = \widehat{g}(gx, hy)^{-1}\phi(\widehat{d}(gx, hy)) = g\left[\widehat{g}(x, y)\varphi(\widehat{d}(x, y))\right]h^{-1} = gQ(x, y)h^{-1}.$$

3 Steerable Guassian Operator Processes:

We identify the space of smooth, connected domains with smooth connected boundaries with the space of smooth, injective, periodic $\mathbb{C}$-valued functions $\mathcal{C}_{\mathrm{per,inj}}([0, 2\pi], \mathbb{C})$ on $[0, 2\pi]$. Furthermore, we identify the space of vector fields on such domains with smooth, periodic $\mathbb{C}$-valued functions $\mathcal{C}_{\mathrm{per}}([0, 2\pi], \mathbb{C})$.

Let $\mathcal{Y} = \mathcal{C}_{\mathrm{per}}([0, 2\pi], \mathbb{C})$ and $\mathcal{X} = \mathcal{C}_{\mathrm{per,inj}}([0, 2\pi], \mathbb{C})$, and consider a $\mathcal{Y}$-valued Gaussian Process over $\mathcal{X}$, with kernel q and mean μ. We will now construct an SE(2)-equivariant kernel, with the intent of applying the GP model to data:

$$\mathcal{D} = \{(\mathcal{M}_n, y_n)\}_{n=1}^N \quad \text{where} \quad y_n = A[\mathcal{M}_n] \quad \text{for each } n.$$

Note here, that $\mathcal{M}_n \in \mathcal{X}$ and $y_n \in \mathcal{Y}$. Suppose we have access to a poor, but fast estimate $\tilde{A}$ of A that is also equivariant wrt. SE(2). Then, we make a simple model as follows:

$$\mu(\mathcal{M}) = \tilde{A}[\mathcal{M}], \quad Q(\mathcal{M}, \mathcal{M}') = \widehat{g}(\mathcal{M}, \mathcal{M}')^{-1}e^{-\widehat{d}(\mathcal{M}, \mathcal{M}')/\nu}.$$

Here, $\nu > 0$ is a hyper parameter linked to the error of the model $\tilde{A}$, which one can often estimate. For data $\boldsymbol{\mathcal{M}} = (\mathcal{M}_n)_{n=1}^N$, the joint kernel q is an operator on $\mathcal{Y}^N$, defined

$$q(\boldsymbol{\mathcal{M}}; \boldsymbol{\mathcal{M}})[\boldsymbol{u}, \boldsymbol{v}] = \langle \boldsymbol{u}, Q(\boldsymbol{\mathcal{M}}; \boldsymbol{\mathcal{M}})\boldsymbol{v}\rangle = \sum_{i,j}\langle u_i, \widehat{g}(\mathcal{M}_i; \mathcal{M}_j)^{-1}e^{-\widehat{d}(\mathcal{M}_i; \mathcal{M}_j)/\nu}v_j\rangle.$$

Introduce notation $g_{ij} = \widehat{g}(\mathcal{M}_i; \mathcal{M}_j)^{-1}$ and $k_{ij} = \exp\{\widehat{d}(\mathcal{M}_i, \mathcal{M}_j)/\nu\}$. Furthermore, we have $g_{ij} = g_{ji}^{-1}$ and $g_{ii} = e$, and so splitting the above sum up into a sum over $i = j$ and $i < j$ and $i > j$ gives, in the setting $\boldsymbol{u} = \boldsymbol{v}$,

$$\sum_{i,j} k_{ij} \langle u_i, g_{ij} u_j \rangle = \sum_{i=j} k_{ii} \langle u_i, u_i \rangle + 2 \sum_{i>j} k_{ij} \langle u_i, g_{ij} u_j \rangle.$$

The above is a sum of diagonal and off-diagonal terms. The diagonal terms are positive, and the off-diagonal terms can be made to vanish provided $\mathcal{M}_i \notin \mathrm{Orb}_{\mathrm{SE}(2)}(\mathcal{M}_j)$, by choosing ν small enough.

Steerable Process Let $\varphi_k(t) = e^{ikt}$ be the Fourier basis. We can now define a steerable process $Y_k(\mathcal{M}) = \langle \varphi_k, Y(\mathcal{M}) \rangle$ with mean and covariance as:

$$\mu(\mathcal{M}) = \sum_{k=1}^{L} \langle \varphi_k, \tilde{A}[\mathcal{M}] \rangle \varphi_k, \quad Q(\mathcal{M}, \mathcal{M}')_{k\ell} = \langle \varphi_k, \widehat{g}(\mathcal{M}, \mathcal{M}')^{-1} \varphi_\ell \rangle e^{-\widehat{d}(\mathcal{M}, \mathcal{M}')/\nu}$$

The kernel Q is equivariant wrt. SE(2), and the mean μ is equivariant wrt. SE(2), since $\tilde{A}$ is. Let $\widehat{g}(\mathcal{M}) = (\theta, \gamma)$ be a rotation by θ radians of the output domain, combined with a perodic shift γ of the function domain $[0, 2\pi]$. We compute Q:

$$\langle \varphi_k, (\theta, \gamma) \varphi_\ell \rangle = \int_0^{2\pi} e^{ikt} \overline{e^{i\theta} e^{i\ell(t-\gamma)}} \mathrm{d}t = e^{i(k\gamma - \theta)} \delta_{k\ell}.$$

Hence, the matrix $Q_{\ell,k}^{m,n}$ is for the full data set $\mathcal{D}$:

$$Q_{\ell,k}^{m,n} = \delta_{k\ell} \exp\left\{ i(k\gamma_{m,n} - \theta_{m,n}) - \widehat{d}_{m,n}/\nu \right\}, \quad \widehat{d}_{m,n} = \widehat{d}(\mathcal{M}_m, \mathcal{M}_n)$$

Note, that $Q_{\ell,k}^{m,n}$ is complex, because it models covariance structure between vector components. Suppose $\mu = 0$ for simplicity and let $R_{\ell,k}^{m,n} = [Q^{-1}]_{\ell,k}^{m,n}$. Then, the conditional expectation of $Y(\mathcal{M})$ given $\boldsymbol{Y}$ is:

$$\overline{q(\mathcal{M}; \mathcal{M}_m)} \sum_{k-K}^{K} \sum_{n=1}^{N} R_{\ell,k}^{m,n} Y_k^m = \sum_{n=1}^{N} R_{\ell,\ell}^{m,n} Y_\ell^m e^{-i(\ell\gamma_m(\mathcal{M}) - \theta_m(\mathcal{M}))}$$

4 Data Generation:

For data generation we convert samples from random curves $\{(t_i, z_i)\}_{i=1}^N$ to canonical Fourier parameterization, by which we mean solving the following optimization problem:

$$\min_{c_k \in \mathbb{C}} \sum_{n=1}^{N} \left\| \sum_{k=-K}^{K} c_k e^{ikt_n} - z_n \right\|^2, \qquad \text{subject to} \qquad \left\| \sum_{k=-K}^{K} ikc_k e^{ikt} \right\| = 1 \quad \text{for all} \quad t$$