Equivariant Learning of Stochastic Fields

Peter Holderrieth*†, Michael Hutchinson*†, Yee Whye Teh†‡

† Department of Statistics, University of Oxford, ‡ Deepind, London

Motivation

Scalar valued stochastic processes are well studied, i.e. random functions

$$f : \mathcal{X} \to \mathbb{R} \text{ or even } f : \mathcal{X} \to \mathbb{R}^n$$

If we want to model functions whose outputs are **vector** valued, or higher-order tensors, we can still model these as functions

$$f : \mathcal{X} \to \mathbb{R}^n$$

but we have the additional baggage of choosing a **basis** to be able to represent the vector values as a series of scalars.

- Choice of basis is arbitrary

$\implies$ Shouldn't affect our inference

- Choice of coordinate frame is arbitrary

$\implies$ Shouldn't affect our inference

The is strongly related to the physical principle of the **homogeneity of space** and the symmetry of natural laws.

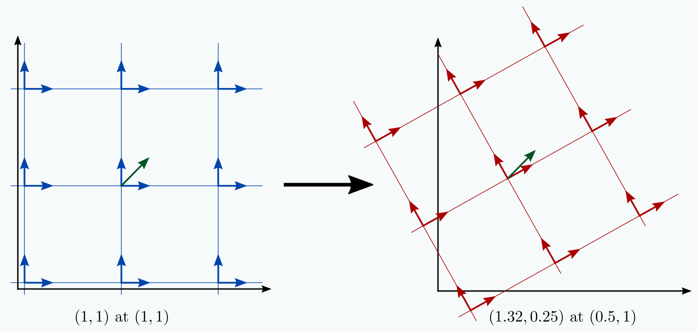


Figure 1: A change of basis and coordinate system changes how we represent the location and value of a vector.

Changing Coordinate System

Changes of coordinate system in $\mathbb{R}^n$ form a group, the Euclidean group $E(n)$. Every element can be written as

$$g = tr, \quad g \in E(n),\ t \in T(n)\ r \in SO(n)$$

Changes of basis (for orthogonal, same-handed bases) on $\mathbb{R}^n$ also form a group, $SO(n)$.

- Scalar functions, $f : \mathbb{R}^n \to \mathbb{R}$, transform like

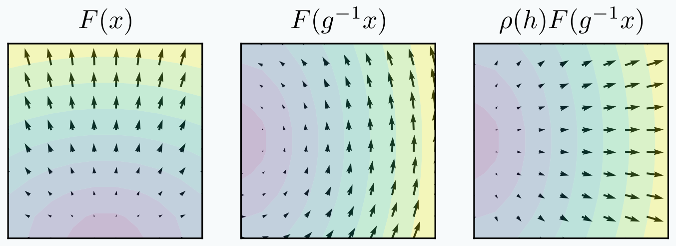
$$g \cdot f(x) = f(g^{-1} \cdot x)$$

- Vector valued functions transform like

$$g \cdot f(x) = \rho(r) f(g^{-1} \cdot x) \quad \text{for } g = tr$$

with

$$\rho : G \to \mathbb{R}^{d \times d} \quad \rho(g_1)\rho(g_2) = \rho(g_1 \cdot g_2)$$



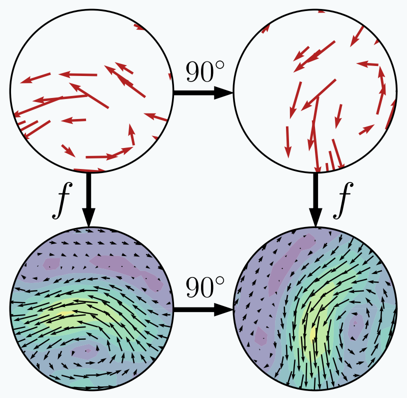
Stochastic Process Conditions

If we have a stochastic process $F \sim P$, the we can define the transform stochastic process $g \cdot P$ as the distribution of $g \cdot F$.

Write the posterior map conditioned on some observations $Z = \{(x_i, y_i)\}_{i=1}^n$ as $Z \mapsto P_Z$

Two natural implications of the independance from coordinate systems and bases:

- The *prior* should be *invariant*, $g \cdot P = P\ \forall g \in G$
- The *posterior map* should be *equivariant*, $P_{g \cdot Z} = g \cdot P_Z\ \forall g \in G$



Proposition 1

Invariant Prior $\iff$ Equivariant Posterior Map (1)

$$g \cdot P = P\ \forall g \in G \iff P_{g.Z} = g.P_Z\ \forall g \in G \quad (2)$$

Equivariant Gaussian Processes

Theorem 1

A Gaussian process $\mathcal{GP}(\mathbf{m}, K)$ is G-invariant, equivalently the posterior G-equivariant, if and only if

1. For all $\mathbf{x} \in \mathbb{R}^n$, $g \in G$,
   (a) $\mathbf{m}(\mathbf{x}) = \mathbf{m}$
   (b) $\rho(r)\mathbf{m} = \mathbf{m}$
2. For all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$, $g \in G$
   (a) $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x} - \mathbf{x}', \mathbf{0}) := \hat{K}(\mathbf{x} - \mathbf{x}')$
   (b) $K(r \cdot \mathbf{x}, r \cdot \mathbf{x}') = \rho(h) K(\mathbf{x}, \mathbf{x}') \rho(r)^{-1}$

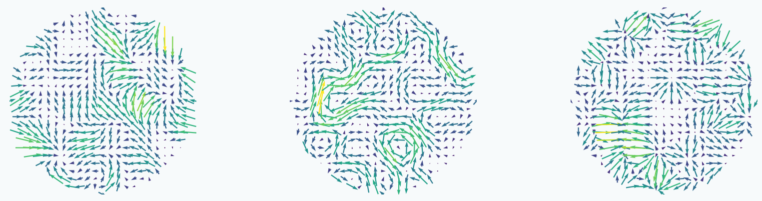


Figure 4: Example RBF, divergence free, and curl free equivariant Guassian Processes

Equivariant DeepSets

We can characterise all functions of the form

$$Z \mapsto f_Z \quad \text{subject to} \quad f_{g \cdot Z}(x) = \rho(r) f_Z(g^{-1} \cdot x)$$

These are functions that take some set of points, and produce a *field*, i.e. a function that can be evaluated anywhere. Any such function can be expressed as

$$f_Z = D\left(\sum_{i=1}^{n} K(\cdot, x_i) \begin{bmatrix} y_i \\ 1 \end{bmatrix}\right) \quad \text{subject to} \quad \begin{cases} K \text{ is an equivariant kernel} \\ D \text{ is an equivariant map between fields} \end{cases}$$

Equivariant Conditional Neural Processes

Neural Processes are *approximations* to stochastic processes. They directly learn an approximate posterior map $Q_Z \approx P_Z$, implicitly learning a prior from data

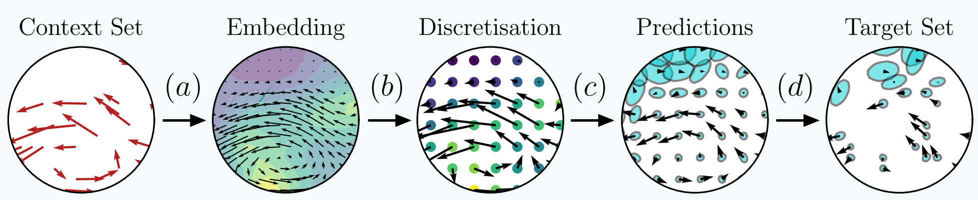
Building on proposition 1, we should ensure this approximation is *equivariant*.

Conditional Neural Process models learn the marginals of each location, in this case a Gaussian marginal, parametrised by mean and covariance functions.

$$Q_Z = \mathcal{N}(\mu_Z(x), \Sigma_Z(x))$$

To ensure this approximation is equivariant, we need the mean and covariance maps to be equivariant

$$\mu_{g \cdot Z}(x) = \rho(r)\mu_Z(g \cdot x) \quad \Sigma_{g \cdot Z}(x) = \rho(r)\Sigma_Z(g \cdot x)\rho(r)^{-1}$$



MNIST Image Experiments

Train dataset	MNIST	rotMNIST	MNIST
Test dataset	MNIST	MNIST	extMNIST
Model			
GP	0.39±0.30	0.39±0.30	0.72±0.17
CNP	0.76±0.05	0.66±0.06	-1.11±0.06
ConvCNP	1.01±0.01	0.95±0.01	1.08±0.02
SteerCNP(C_4)	1.05±0.02	**1.02**±**0.03**	1.14±0.02
SteerCNP(C_8)	**1.07**±**0.03**	**1.05**±**0.04**	**1.16**±**0.03**
SteerCNP(C_{16})	**1.08**±**0.03**	**1.04**±**0.03**	**1.17**±**0.05**
SteerCNP(D_4)	**1.08**±**0.03**	**1.05**±**0.03**	1.14±0.03
SteerCNP(D_8)	**1.08**±**0.03**	**1.04**±**0.04**	**1.17**±**0.02**

Table 1: MNIST completion results

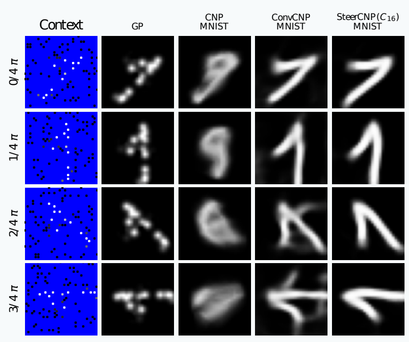


Figure 2: Example inferences

ERA5 Weather Experiments

Model	US	China
GP	0.386±0.005	-0.755±0.001
CNP	0.001±0.017	-2.456±0.365
ConvCNP	0.898±0.045	-0.890±0.059
SteerCNP (C_4)	**1.255**±0.019	-0.578±0.173
SteerCNP (C_8)	1.038±0.026	-0.582±0.104
SteerCNP (C_{16})	1.094±0.015	-0.550±0.073
SteerCNP (D_4)	1.037±0.037	**-0.429**±0.067
SteerCNP (D_8)	1.032±0.011	-0.539±0.129

Table 2: Weather completion results

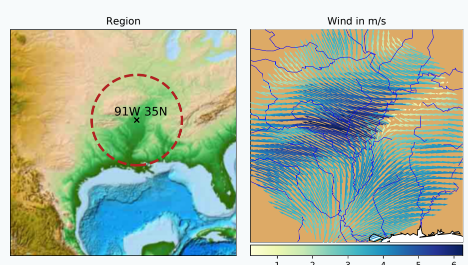


Figure 3: Example region and wind map