Unifying Regression and Uncertainty Quantification with Contrastive Spectral Representation Learning

Anonymous Author(s)
Affiliation
Address
email

Abstract

In this work, we discuss a contrastive representation learning framework, called Neural Conditional Probability (NCP) [16], which introduces a new paradigm for training deep Neural Network (NN) architectures for regression. This framework enables high-quality regression estimates *and* parametric uncertainty quantification without retraining or restrictive assumptions on the uncertainty distribution. NCP learns high-dimensional data representations that are *linearly transferable* to regression *and* uncertainty quantification tasks, backed by non-asymptotic statistical learning guarantees linking representation quality to downstream performance. Crucially, in equivariant regression contexts, the NCP framework can be adapted to train any Geometric Deep Learning (GDL) architecture, resulting in a *disentangled* equivariant representation learning algorithm with first-of-its-kind statistical guarantees for equivariant regression and symmetry-aware uncertainty quantification.

Introduction

A central problem in machine learning is modeling the conditional probability $\mathbb{P}(\mathbf{y}|\mathbf{x})$ of a target variable $\mathbf{y} \in \mathcal{Y}$ given an observed variable $\mathbf{x} \in \mathcal{X}$. This underpins robust reasoning under *uncertainty* in safety-critical applications such as medicine, finance, robotics, and physics [12, 28, 35]. In these domains, the expected value of the target variable $\mathbb{E}[\mathbf{y}|\boldsymbol{x}] := \mathbb{E}_{\mathbf{y}\sim\mathbb{P}(\mathbf{y}|\boldsymbol{x})}[\mathbf{y}]$ (i.e., regression) is often insufficient due to high spread, multimodality, or asymmetry of the conditional distribution (see Fig. 1). However, estimating conditional distributions in high-dimensional settings remains challenging without resorting to ensemble learning [10, 25] or imposing strong, often restrictive, inductive biases—e.g., that $\mathbb{P}(\mathbf{y}|\boldsymbol{x})$ belongs to a parametric distribution family (e.g., Gaussian, Poisson, Mixture of Gaussians) for all $\boldsymbol{x} \in \mathcal{X}$ [1, 13, 21, 27].

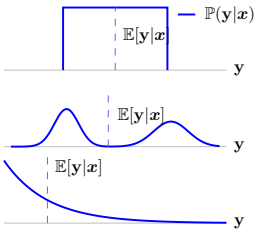

Figure 1: Conditional distributions with uninformative expected values.

To tackle conditional probability modeling—and therefore reliable uncertainty quantification—using deep NNs, several works in contrastive representation learning [16, 23, 26, 32] have shown that deep NNs can be endowed with robust conditional probability and uncertainty modeling capabilities. This is achieved when trained via a contrastive representation learning loss that aims to approximate the Pointwise Mutual Dependency (PMD) ratio, $\kappa(\boldsymbol{x},\boldsymbol{y}) = \frac{P_{\mathbf{xy}}(d\boldsymbol{x},d\boldsymbol{y})}{P_{\mathbf{x}}(d\boldsymbol{x})P_{\mathbf{y}}(d\boldsymbol{y})}$, between the joint distribution $P_{\mathbf{xy}}$ and the product of the marginals $P_{\mathbf{x}} \times P_{\mathbf{y}}$. The approach uses a bilinear model $\kappa_{\boldsymbol{\theta}} : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ (see Fig. 4), parameterized by two NNs, $\boldsymbol{\phi}_{\boldsymbol{\theta}} : \mathcal{X} \to \mathbb{R}^n$ and $\boldsymbol{\psi}_{\boldsymbol{\theta}} : \mathcal{Y} \to \mathbb{R}^n$, that learn high-dimensional representations of the input and target variables, respectively [16].

In this preliminary work, we discuss how this contrastive learning framework, referred to as NCP, could represent a paradigm shift in training deep NNs and GDL architectures for safety-critical regression applications.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

Neural Conditional Probability (NCP)

The NCP framework proposes to model conditional probability distributions by approximating the *conditional expectation operator* [2, 8, 29], $\mathsf{E}_{\mathbf{y}|\mathbf{x}}\colon \mathcal{L}^2_{\mathbf{y}} \to \mathcal{L}^2_{\mathbf{x}}$, a linear integral operator acting on the Hilbert spaces $\mathcal{L}^2_{\mathbf{x}} := \mathcal{L}^2_{P_{\mathbf{x}}}(\mathcal{X}, \mathbb{R})$ and $\mathcal{L}^2_{\mathbf{y}} := \mathcal{L}^2_{P_{\mathbf{y}}}(\mathcal{Y}, \mathbb{R})$ of square-integrable functions of the random variables $\mathbf{x}$ and $\mathbf{y}$, respectively. The action of this operator on any function $h \in \mathcal{L}^2_{\mathbf{y}}$ returns the function's *conditional expectation*:

$$[\mathsf{E}_{\mathbf{y}|\mathbf{x}} h](\boldsymbol{x}) = \mathbb{E}[h(\mathbf{y})|\mathbf{x}=\boldsymbol{x}] := \int_{\mathcal{Y}} h(\boldsymbol{y}) P_{\mathbf{y}|\mathbf{x}}(d\boldsymbol{y}|\boldsymbol{x}) = \int_{\mathcal{Y}} h(\boldsymbol{y}) \tfrac{P_{\mathbf{xy}}(d\boldsymbol{y},\boldsymbol{x})}{P_{\mathbf{x}}(d\boldsymbol{x})} = \int_{\mathcal{Y}} h(\boldsymbol{y}) \kappa(\boldsymbol{x},\boldsymbol{y}) P_{\mathbf{y}}(d\boldsymbol{y}), \quad (1)$$

where the PMD $\kappa(\boldsymbol{x},\boldsymbol{y}) := \frac{P_{\mathbf{xy}}(d\boldsymbol{x},d\boldsymbol{y})}{P_{\mathbf{x}}(d\boldsymbol{x})\, P_{\mathbf{y}}(d\boldsymbol{y})}$ becomes the kernel function defining the operator $\mathsf{E}_{\mathbf{y}|\mathbf{x}}$ (see Fig. 3) and $P_{\mathbf{y}|\mathbf{x}}$ is the conditional probability measure [30].

The practical utility of $\mathsf{E}_{\mathbf{y}|\mathbf{x}}$ lies in that it provides an infinite-dimensional linear model—in a nonlinear representation space—for computing conditional probabilities *and* expectations. To see this, note that for any $\boldsymbol{x} \in \mathcal{X}$ and any measurable set $\mathbb{B} \subset \mathcal{Y}$ we have that:

$$\mathbb{P}(\mathbf{y} \in \mathbb{B}|\mathbf{x}=\boldsymbol{x}) := \int_{\mathcal{Y}} \mathbb{1}_{\mathbb{B}}(\boldsymbol{y}) P_{\mathbf{y}|\mathbf{x}}(d\boldsymbol{y}|\boldsymbol{x}) = [\mathsf{E}_{\mathbf{y}|\mathbf{x}} \mathbb{1}_{\mathbb{B}}](\boldsymbol{x}), \quad \text{and} \quad \mathbb{E}[\mathbf{y}|\mathbf{x}=\boldsymbol{x}] := [\mathsf{E}_{\mathbf{y}|\mathbf{x}} \mathbf{y}](\boldsymbol{x}). \quad (2)$$

Therefore, to provide estimates for (2), NCP seeks the best finite-dimensional approximation of $\mathsf{E}_{\mathbf{y}|\mathbf{x}}$ as a rank-r operator $\mathsf{E}_{\boldsymbol{\theta}}$ with matrix representation $\boldsymbol{E}_{\boldsymbol{\theta}} \in \mathbb{R}^{r \times r}$ given by

$$\arg\min_{\boldsymbol{\theta}} \|\mathsf{E}_{\mathbf{y}|\mathbf{x}} - \mathsf{E}_{\boldsymbol{\theta}}\|^2_{\text{HS}} \equiv \mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}}(\kappa(\mathbf{x},\mathbf{y}) - \kappa_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}))^2, \quad \text{s.t. } \mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}}\kappa_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) = 1 \text{ and } \operatorname{rank}(\mathsf{E}_{\boldsymbol{\theta}}) \leq r. \quad (3)$$

Spectral representation learning The optimal solution to Eq. (3), denoted $\mathsf{E}_\star$, is the r-truncated Singular Value Decomposition (SVD) of $\mathsf{E}_{\mathbf{y}|\mathbf{x}}$ [6, 15, 26], namely

$$[\mathsf{E}_\star f](\boldsymbol{x}) = \textstyle\sum_{i=0}^{r} \sigma_i \langle f, v_i \rangle_{P_{\mathbf{y}}} u_i(\boldsymbol{x}), \qquad \text{with} \quad \sigma_i u_i(\boldsymbol{x}) = [\mathsf{E}_{\mathbf{y}|\mathbf{x}} v_i](\boldsymbol{x}), \ \forall i \in [1, r], \quad (4)$$

where (σ_i, u_i, v_i) denotes the ith singular value and left/right singular functions of $\mathsf{E}_{\mathbf{y}|\mathbf{x}}$ [2, 15].

Consequently, NCP parameterizes $\mathsf{E}_{\boldsymbol{\theta}}$ by a bilinear model $\kappa_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) = 1 + \boldsymbol{\phi}_{\boldsymbol{\theta}}(\boldsymbol{x})^\top \boldsymbol{E}_{\boldsymbol{\theta}} \boldsymbol{\psi}_{\boldsymbol{\theta}}(\boldsymbol{y})$, composed of two encoder NNs $\boldsymbol{\phi}_{\boldsymbol{\theta}}\colon \mathcal{X} \to \mathbb{R}^r$ and $\boldsymbol{\psi}_{\boldsymbol{\theta}}\colon \mathcal{Y} \to \mathbb{R}^r$ that aim to approximate the spectral decomposition of the operator, i.e., the *span* of the top r (non-constant) left and right singular functions of $\mathsf{E}_{\mathbf{y}|\mathbf{x}}$. See Fig. 4-left.

Since κ is generally unavailable analytically, (3) is solved via the regularized contrastive loss[1]:

$$\begin{aligned} \mathcal{L}_\gamma(\boldsymbol{\theta}) \;=\; & -2\mathbb{E}_{\mathbf{xy}}\kappa_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) + \mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}}\kappa_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})^2 + 2\gamma\big(\|\mathbb{E}_{\mathbf{x}}\boldsymbol{\phi}_{\boldsymbol{\theta}}(\mathbf{x})\|_F^2 + \|\mathbb{E}_{\mathbf{y}}\boldsymbol{\psi}_{\boldsymbol{\theta}}(\mathbf{y})\|_F^2 \\ & + \|\operatorname{Cov}(\boldsymbol{\phi}_{\boldsymbol{\theta}}) - \boldsymbol{I}_r\|_F^2 + \|\operatorname{Cov}(\boldsymbol{\psi}_{\boldsymbol{\theta}}) - \boldsymbol{I}_r\|_F^2\big), \end{aligned} \quad (5)$$

where the first two regularization terms center the representations, ensuring that $\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}}\kappa_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) \approx 1$ [15], while the last two enforce approximate orthonormality of the learned bases in $\mathcal{F}^{\boldsymbol{\theta}}_{\mathbf{x}} := \operatorname{span}(\boldsymbol{\phi}_{\boldsymbol{\theta}}) \subset \mathcal{L}^2_{\mathbf{x}}$ and $\mathcal{F}^{\boldsymbol{\theta}}_{\mathbf{y}} := \operatorname{span}(\boldsymbol{\psi}_{\boldsymbol{\theta}}) \subset \mathcal{L}^2_{\mathbf{y}}$ [13]—analogous to kernel methods [20].

Equivariant Neural Conditional Probability (eNCP)

In virtually all applications of GDL [5], symmetry priors result in the conditional $\mathbb{P}(\mathbf{y}|\mathbf{x})$ and marginal distributions $\mathbb{P}(\mathbf{x})$ and $\mathbb{P}(\mathbf{y})$ being invariant to symmetry transformations of the data (see Fig. 3), i.e:

$$\mathbb{P}(\mathbf{y}|\mathbf{x}) = \mathbb{P}(g \triangleright_{\mathcal{Y}} \mathbf{y} | g \triangleright_{\mathcal{X}} \mathbf{x}) \quad \textit{and} \quad \mathbb{P}(\mathbf{x}) = \mathbb{P}(g \triangleright_{\mathcal{X}} \mathbf{x}) \quad \Rightarrow \quad \mathbb{P}(\mathbf{y}) = \mathbb{P}(g \triangleright_{\mathcal{Y}} \mathbf{y}), \quad \forall\, g \in \mathbb{G}, \quad (6)$$

where $\mathbb{G}$ denotes a compact symmetry group acting on the data spaces $\mathcal{X}$ and $\mathcal{Y}$ via the group actions, $\triangleright_{\mathcal{X}}\colon \mathbb{G} \times \mathcal{X} \to \mathcal{X}$, and $\triangleright_{\mathcal{Y}}\colon \mathbb{G} \times \mathcal{Y} \to \mathcal{Y}$, with $g \triangleright_{\mathcal{X}} \boldsymbol{x} \in \mathcal{X}$ and $g \triangleright_{\mathcal{Y}} \boldsymbol{y} \in \mathcal{Y}$ denoting linear, invertible transformations of $\boldsymbol{x}$ and $\boldsymbol{y}$ defined by $g \in \mathbb{G}$ (see Fig. 3).

These priors imply the $\mathbb{G}$-invariance of the joint distribution $\mathbb{P}(\mathbf{x},\mathbf{y})$, as well as the $\mathbb{G}$-equivariance of conditional expectations (i.e., equivariant regression) and the $\mathbb{G}$-invariance of the PMD kernel defining the conditional expectation operator (see Fig. 3 and [23]):

$$g \triangleright_{\mathcal{Y}} \mathbb{E}[\mathbf{y}|\mathbf{x}=\boldsymbol{x}] = \mathbb{E}[\mathbf{y}|\mathbf{x}=g \triangleright_{\mathcal{X}} \boldsymbol{x}] \quad \text{and} \quad \kappa(\boldsymbol{x},\boldsymbol{y}) = \kappa(g \triangleright_{\mathcal{X}} \boldsymbol{x}, g \triangleright_{\mathcal{Y}} \boldsymbol{y}) \qquad \forall\, g \in \mathbb{G}, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}. \quad (7)$$

These symmetry priors represent powerful inductive biases for the NCP framework, summarized in the following and in [23].

[1] Used in density-ratio fitting [30], representation learning [9, 34], and mutual information estimation [31].

Task	$\boldsymbol{f}(\boldsymbol{x}) := \mathbb{E}_{\mathbf{y}}[\mathbf{y}\vert\mathbf{x}{=}\boldsymbol{x}] \approx \hat{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x})$	$\mathbb{P}[\mathbf{y}{\in}\mathbb{B}\vert\mathbf{x}\in\mathbb{A}] \approx \widehat{\mathbb{P}}_{\boldsymbol{\theta}}[\mathbf{y}\in\mathbb{B}\vert\mathbf{x}\in\mathbb{A}]$
Estimate	$\widehat{\mathbb{E}}_{\mathbf{y}}[\mathbf{y}]{+}\boldsymbol{\phi}_{\boldsymbol{\theta}}(\boldsymbol{x})^\top \boldsymbol{E}_{\boldsymbol{\theta}}\widehat{\mathbb{E}}_{\mathbf{y}}[\boldsymbol{\psi}_{\boldsymbol{\theta}}(\mathbf{y})\otimes\mathbf{y}]$	$\widehat{\mathbb{E}}_{\mathbf{y}}[\mathbb{1}_{\mathbb{B}}]{+}\frac{\widehat{\mathbb{E}}_{\mathbf{x}}[\mathbb{1}_{\mathbb{A}}(\mathbf{x})\otimes\boldsymbol{\phi}_{\boldsymbol{\theta}}(\mathbf{x})]^\top \boldsymbol{E}_{\boldsymbol{\theta}}\widehat{\mathbb{E}}_{\mathbf{y}}[\mathbb{1}_{\mathbb{B}}(\mathbf{y})\otimes\boldsymbol{\psi}_{\boldsymbol{\theta}}(\mathbf{y})]}{\widehat{\mathbb{E}}_{\mathbf{x}}[\mathbb{1}_{\mathbb{A}}(\mathbf{x})]}$
Guarantees	$\Vert\boldsymbol{f}{-}\hat{\boldsymbol{f}}_{\boldsymbol{\theta}}\Vert_{\mathcal{L}^2_{\mathbf{x}}}{\lesssim}\sqrt{\mathrm{Var}[\Vert\mathbf{y}\Vert]}\left(\mathcal{E}^r_{\boldsymbol{\theta}}{+}\frac{\ln(n_{\text{iso}}/\delta)}{(d_{\text{iso}}N)^{\frac{\alpha}{1+2\alpha}}}\right)$	$\vert\mathbb{P}{-}\widehat{\mathbb{P}}_{\boldsymbol{\theta}}\vert{\lesssim}\sqrt{\frac{\mathbb{P}[\mathbf{y}\in\mathbb{B}]}{\mathbb{P}[\mathbf{x}\in\mathbb{G}\triangleright_{\mathcal{X}}\mathbb{A}]}}\left(\mathcal{E}^r_{\boldsymbol{\theta}}{+}\frac{\ln(n_{\text{iso}}/\delta)}{(d_{\text{iso}}N)^{\frac{\alpha}{1+2\alpha}}}\right)$

Table 1: Statistical guarantees for eNCP [23]. The error bounds are shaped by (i) the structure of the symmetry group $\mathbb{G}$—the number of isotypic subspaces n_{iso} and the sum of the group's irreducible representations dimension $d_{\text{iso}} = \sum_{k=1}^{n_{\text{iso}}} d_k$, which enlarge the effective sample size—, (ii) the quality of the learned representations $\mathcal{E}^r_{\boldsymbol{\theta}} = \Vert\mathsf{E}_{\mathbf{y}|\mathbf{x}} - \mathsf{E}_{\boldsymbol{\theta}}\Vert_{\text{op}} \leq \sqrt{\mathcal{L}_\gamma(\boldsymbol{\theta}) - \mathcal{L}_\gamma(\star)}$, and (iii) the operator's singular-value decay rate $\alpha > 0$. Note that $\mathbb{G} \triangleright_{\mathcal{X}} \mathbb{A} := \cup_{g\in\mathbb{G}}\ g \triangleright_{\mathcal{X}} \mathbb{A}$ denotes the group orbit of $\mathbb{A}$. NCP's guarantees are recovered when $n_{\text{iso}} = d_{\text{iso}} = 1$.

Symmetric function spaces A significant property of the NCP and eNCP frameworks is their non-asymtotic statistical learning guarantees, summasized in Tab. 1

The $\mathbb{G}$-invariance of the marginal probabilities implies that the function spaces $\mathcal{L}^2_{\mathbf{x}}$ and $\mathcal{L}^2_{\mathbf{y}}$ are symmetric Hilbert spaces with unitary group actions $\triangleright_{\mathcal{L}^2_{\mathbf{x}}}\colon \mathbb{G}\times\mathcal{L}^2_{\mathbf{x}} \to \mathcal{L}^2_{\mathbf{x}}$ and $\triangleright_{\mathcal{L}^2_{\mathbf{y}}}\colon \mathbb{G}\times\mathcal{L}^2_{\mathbf{y}} \to \mathcal{L}^2_{\mathbf{y}}$ (see details in App. F and Fig. 11). Consequently, whenever $\mathbb{G}$ is a compact symmetry group, a standard result from harmonic analysis [18] enables the orthogonal decomposition of these spaces into $n_{\text{iso}} \leq \vert\mathbb{G}\vert$ subspaces referred to as *isotypic subspaces*: $\mathcal{L}^2_{\mathbf{x}} = \bigoplus^{\perp}_{k\in[1,n_{\text{iso}}]} \mathcal{L}^{2(k)}_{\mathbf{x}}$ and $\mathcal{L}^2_{\mathbf{y}} = \bigoplus^{\perp}_{k\in[1,n_{\text{iso}}]} \mathcal{L}^{2(k)}_{\mathbf{y}}$ (see Thm. E.8), where each $\mathcal{L}^{2(k)}_{\mathbf{x}}$ and $\mathcal{L}^{2(k)}_{\mathbf{y}}$ denotes the subspaces that transform according to a unique quotient group $\mathbb{G}^{(k)} \subseteq \mathbb{G}$.

Equivariant conditional expectation operator Due to Eq. (7), $\mathsf{E}_{\mathbf{y}|\mathbf{x}}$ becomes a $\mathbb{G}$-equivariant linear operator [23]. This means that $\mathsf{E}_{\mathbf{y}|\mathbf{x}}$ commutes with the group action on the function spaces; $g \triangleright_{\mathcal{L}^2_{\mathbf{x}}} [\mathsf{E}_{\mathbf{y}|\mathbf{x}}h](\cdot){=}\mathsf{E}_{\mathbf{y}|\mathbf{x}}[g \triangleright_{\mathcal{L}^2_{\mathbf{y}}} h](\cdot)$ for all $g \in \mathbb{G}, h \in \mathcal{L}^2_{\mathbf{y}}$, and consequently, can be decomposed (disentangled) into a direct sum of operators acting on the corresponding isotypic subspaces $[\mathsf{E}_{\mathbf{y}|\mathbf{x}}h](\cdot) = \sum_{k=1}^{n_{\text{iso}}}[\mathsf{E}^{(k)}_{\mathbf{y}|\mathbf{x}}h^{(k)}](\cdot)$, for all $h \in \mathcal{L}^2_{\mathbf{y}}$ and $h^{(k)} \in \mathcal{L}^{2(k)}_{\mathbf{y}}$, where each $\mathsf{E}^{(k)}_{\mathbf{y}|\mathbf{x}} : \mathcal{L}^{2(k)}_{\mathbf{y}} \to \mathcal{L}^{2(k)}_{\mathbf{x}}$ models $\mathbb{G}^{(k)}$-equivariant conditional expectation.

Disentangled spectral representation learning The orthogonal decomposition of the function spaces and of $\mathsf{E}_{\mathbf{y}|\mathbf{x}}$ decomposes the representation learning objective in (3) into n_{iso} independent (disentangled) representation learning subprobelms:

$$\begin{aligned} &\arg\min_{\boldsymbol{\theta}}\ \Vert\mathsf{E}_{\mathbf{y}|\mathbf{x}}{-}\mathsf{E}_{\boldsymbol{\theta}}\Vert^2_{\text{HS}} = \textstyle\sum_{k=1}^{n_{\text{iso}}}\Vert\mathsf{E}^{(k)}_{\mathbf{y}|\mathbf{x}}{-}\mathsf{E}^{(k)}_{\boldsymbol{\theta}}\Vert^2_{\text{HS}} \equiv \mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}}\sum_{k=1}^{n_{\text{iso}}}(\kappa^{(k)}(\mathbf{x},\mathbf{y}){-}\kappa^{(k)}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}))^2, \\ &\text{s.t.} \qquad \mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}}\kappa_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}){=}1, \quad \text{and} \quad \kappa_{\boldsymbol{\theta}}(g \triangleright_{\mathcal{X}} \boldsymbol{x}, g \triangleright_{\mathcal{Y}} \boldsymbol{y}){=}\kappa_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}), \quad \forall g \in \mathbb{G}, (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}\times\mathcal{Y}. \end{aligned} \tag{8}$$

Analogous to (4), the optimal truncation of each $\mathsf{E}^{(k)}_{\mathbf{y}|\mathbf{x}}$ is its truncated SVD [6]. Consequently, as explained in [23], Eq. (8) can be solved by parameterizing the approximated PMD as a bilinear model

$$\kappa_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) = \mathbb{1}_{P_{\mathbf{x}}}(\boldsymbol{x})\mathbb{1}_{P_{\mathbf{y}}}(\boldsymbol{y}) + \textstyle\sum_{k=1}^{n_{\text{iso}}}\kappa^{(k)}_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}), \qquad \kappa^{(k)}_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) := \boldsymbol{\phi}^{(k)}_{\boldsymbol{\theta}}(\boldsymbol{x})^\top \boldsymbol{E}^{(k)}_{\boldsymbol{\theta}}\boldsymbol{\psi}^{(k)}_{\boldsymbol{\theta}}(\boldsymbol{y}), \tag{9}$$

parameterized by two $\mathbb{G}$-equivariant encoder NNs $\boldsymbol{\phi}_{\boldsymbol{\theta}} : \mathcal{X} \to \mathbb{R}^r$ and $\boldsymbol{\psi}_{\boldsymbol{\theta}} : \mathcal{Y} \to \mathbb{R}^r$, defined to expose the approximated isotypic subspaces, i.e., $\boldsymbol{\phi}_{\boldsymbol{\theta}}(\cdot) = [\boldsymbol{\phi}^{(1)}_{\boldsymbol{\theta}}(\cdot)^\top, \ldots, \boldsymbol{\phi}^{(n_{\text{iso}})}_{\boldsymbol{\theta}}(\cdot)^\top]^\top$ and $\boldsymbol{\psi}_{\boldsymbol{\theta}}(\cdot) = [\boldsymbol{\psi}^{(1)}_{\boldsymbol{\theta}}(\cdot)^\top, \ldots, \boldsymbol{\psi}^{(n_{\text{iso}})}_{\boldsymbol{\theta}}(\cdot)^\top]^\top$, and a $\mathbb{G}$-equivariant truncated operator matrix parameterized in block-diagonal form $\boldsymbol{E}_{\boldsymbol{\theta}} = \bigoplus_{k=1}^{n_{\text{iso}}}\boldsymbol{E}^{(k)}_{\boldsymbol{\theta}}$ (see Fig. 4).

Statistical learning guarantees A significant property of the NCP and eNCP frameworks is their non-asymptotic statistical learning guarantees, summarized in Tab. 1, which provide first-of-its-kind learning guarantees for equivariant representations in the context of equivariant regression and conditional probability estimation. Crucially, these guarantees enable direct estimation of the benefits of leveraging symmetry priors for any finite symmetry group $\mathbb{G}$.

Experiments

We present two experiments evaluating the NCP and eNCP frameworks in $\mathbb{G}$-equivariant regression and symmetry-aware uncertainty quantification.

$\mathbb{G}$-Equivariant regression The goal is to predict a quadruped robot's Center of Mass (CoM) linear $\boldsymbol{l} \in \mathbb{R}^3$ and angular momenta $\boldsymbol{k} \in \mathbb{R}^3$ given noisy observations of the robot's generalized

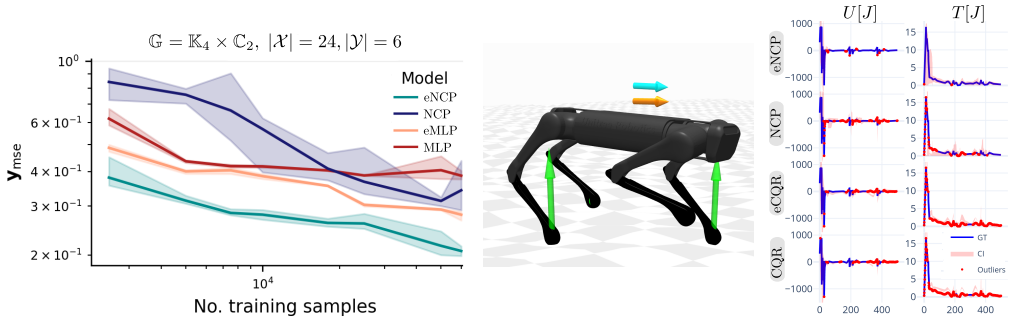

Figure 2: **Left**: Test set sample efficiency for $\mathbb{G}$-equivariant regression (MSE vs. training samples) when predicting linear and angular momentum of a quadruped robot's center of mass (CoM) from noisy joint positions and velocities. **Right**: Uncertainty quantification via prediction of 90% confidence intervals (CI, light-red area) for the robot's instantaneous work U_t and kinetic energy T_t during locomotion over rough terrain.

positions $\boldsymbol{q} \in \mathbb{R}^{12}$ and velocity coordinates $\dot{\boldsymbol{q}} \in \mathbb{R}^{12}$. We compare eNCP against NCP and two baselines—a standard Multi-Layer Perceptron (MLP) and an Equivariant MLP (eMLP)—all with equivalent architectural footprint, where MLP and eMLP are trained using Mean Squared Error (MSE) loss. The results in Fig. 2 show that NCP can beat MLP in terms of performance, while eNCP outperforms eMLP in both performance and sample complexity.

Symmetry-aware uncertainty quantification The goal is to provide uncertainty quantification for unavailable yet crucial state observables for robot control [4, 19]. Specifically, we use proprioceptive sensor readings to provide 90% confidence intervals for the robot's Ground Reaction Forces (GRF) $\boldsymbol{\tau}_{\text{grf}} \in \mathbb{R}^{12}$, instantaneous work $U(\boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{\tau}) \in \mathbb{R}$, and kinetic energy $T(\boldsymbol{q}, \dot{\boldsymbol{q}}) \in \mathbb{R}$ while the robot traverses rough terrain. Robust estimates of these quantities are crucial for optimal control [4].

For eNCP and NCP models, quantile estimation is achieved by regressing the Conditional Cumulative Distribution Function (CCDF) for each dimension of $\mathbf{y} = [\mathrm{y}_1, \dots]$ and applying linear search to extract quantiles (see Fig. 7 in the appendix). This involves discretizing each y_i range into N_b bins and estimating $\mathbb{P}(\mathrm{y}_i \in \mathbb{A}_{i,n} | \mathbf{x} = \cdot) := [\mathsf{E}_{\mathbf{y}|\mathbf{x}} \mathbb{1}_{\mathbb{A}_{i,n}}](\cdot)$ for all $n \in [N_b]$ (see Tab. 1) in a single forward pass, where $\mathbb{A}_{i,n}$ consists of the first n bins. In contrast, baseline methods Conditional Quantile Regression (CQR) [7] and its equivariant adaptation Equivariant CQR (eCQR) directly regress quantiles for fixed coverage levels and require retraining for different coverage values.

	r-Coverage ↑	Coverage ↑
eNCP	99.5±0.1%	95.0±0.4%
NCP	99.5±0.0%	56.9±0.3%
eCQR	84.2±0.7%	6.7±1.2%
CQR	80.5±3.7%	8.5±0.9%

Table 2: Relaxed coverage, see (13), and Coverage, see (12), for the test-set confidence intervals in quadruped locomotion uncertainty estimation of $\boldsymbol{\tau}_{\text{grf}}$, U and T. Target coverage is: 90%.

Results in Tab. 2, Fig. 2 show NCP and eNCP outperforming their respective baselines, with eNCP being the sole method with empirical test coverage close to the desired value, making other models unreliable for practical applications.

Conclusions

This work discusses the NCP contrastive representation learning framework for training NN architectures for regression, enabling high-quality regression estimates *and* uncertainty quantification via conditional probability estimation, without requiring retraining or assumptions on the conditional distribution shape. NCP can be extended to leverage symmetry priors, resulting in a disentangled equivariant representation learning framework capable of training any equivariant NN architecture for equivariant regression and symmetry-aware uncertainty quantification with novel non-asymptotic statistical learning guarantees. We demonstrate this framework's capabilities to replace standard training approaches on two preliminary robotics applications.

Discussion and future work From a practical perspective, the NCP framework simply requires changing the training loss from the standard MSE to the contrastive loss (5) and using an auxiliary NN model during training (see Fig. 4). Extensive experimental work—similar to that presented in [3] in the context of classification—is needed to elucidate in which applications the NCP framework can and cannot outperform standard training approaches and which engineering choices achieve optimal performance (e.g., batch size, orthonormalization technique, etc.).

References

[1] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76:243–297, 2021.

[2] Charles R Baker. Joint measures and cross-covariance operators. *Trans. Am. Math. Soc.*, 186: 273–273, 1973.

[3] Han Bao, Yoshihiro Nagano, and Kento Nozawa. On the surrogate gap between contrastive and supervised losses. In *International conference on machine learning*, pages 1585–1606. PMLR, 2022.

[4] Gerardo Bledt, Patrick M Wensing, Sam Ingersoll, and Sangbae Kim. Contact model fusion for event-based locomotion in unstructured terrains. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4399–4406. IEEE, 2018.

[5] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

[6] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[7] Shai Feldman, Stephen Bates, and Yaniv Romano. Calibrated multiple-output quantile regression with representation learning. *Journal of Machine Learning Research*, 24(24):1–48, 2023.

[8] Kenji Fukumizu, Francis R Bach, and Michael I Jordan. Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces. *Journal of Machine Learning Research*, 5(Jan): 73–99, 2004.

[9] Jeff Z. HaoChen, Colin Wei, Adrien Gaidon, and Tengyu Ma. Provable guarantees for self-supervised deep learning with spectral contrastive loss. In *Advances in Neural Information Processing Systems*, volume 34, pages 5000–5011. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/27debb435021eb68b3965290b5e24c49-Paper.pdf.

[10] Dan Hendrycks, Norman Mu, Ekin Dogus Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. Augmix: A simple data processing method to improve robustness and uncertainty. In *International Conference on Learning Representations*.

[11] Irina Higgins, David Amos, David Pfau, Sebastien Racaniere, Loic Matthey, Danilo Rezende, and Alexander Lerchner. Towards a definition of disentangled representations. *arXiv preprint arXiv:1812.02230*, 2018.

[12] Rafael Izbicki. *Machine Learning Beyond Point Predictions: Uncertainty Quantification*. imprint, 1st edition, 2025. ISBN 978-65-01-20272-3.

[13] Rafael Izbicki and Ann B. Lee. Converting high-dimensional regression to high-dimensional conditional density estimation. 2017.

[14] Anthony W. Knapp. *Representation Theory of Semisimple Groups, An Overview Based on Examples (PMS-36)*. Princeton University Press, Princeton, 1986.

[15] Vladimir R Kostic, Pietro Novelli, Riccardo Grazzi, Karim Lounici, and Massimiliano Pontil. Learning invariant representations of time-homogeneous stochastic dynamical systems. In *The Twelfth International Conference on Learning Representations*, 2024.

[16] Vladimir R Kostic, Gregoire Pacreau, Giacomo Turri, Pietro Novelli, Karim Lounici, and Massimiliano Pontil. Neural conditional probability for uncertainty quantification. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[17] Zicheng Liu, Steven J Gortler, and Michael F Cohen. Hierarchical spacetime control. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 35–42, 1994.

[18] George W. Mackey. Harmonic analysis as the exploitation of symmetry–a historical survey. *Bulletin (New Series) of the American Mathematical Society*, 3(1.P1):543 – 698, 1980.

[19] Michael Maravgakis, Despina-Ekaterini Argiropoulos, Stylianos Piperakis, and Panos Trahanias. Probabilistic contact state estimation for legged robots using inertial information. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 12163–12169. IEEE, 2023.

[20] Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, 10(1–2):1–141, 2017. URL http://dx.doi.org/10.1561/2200000060.

[21] Thomas Nagler and Claudia Czado. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, 151:69–89, October 2016. URL http://dx.doi.org/10.1016/j.jmva.2016.07.003.

[22] Ylenia Nisticò, João Carlos Virgolino Soares, Lorenzo Amatucci, Geoff Fink, and Claudio Semini. Muse: A real-time multi-sensor state estimator for quadruped robots. *IEEE Robotics and Automation Letters*, 2025.

[23] Daniel Ordoñez-Apraez, Vladimir Kostić, Alek Fröhlich, Vivien Brandt, Karim Lounici, and Massimiliano Pontil. Equivariant representation learning for symmetry-aware inference with guarantees. *arXiv preprint arXiv:2505.19809*, 2025.

[24] Daniel Ordoñez-Apraez, Giulio Turrisi, Vladimir Kostic, Mario Martin, Antonio Agudo, Francesc Moreno-Noguer, Massimiliano Pontil, Claudio Semini, and Carlos Mastalli. Morphological symmetries in robotics. *The International Journal of Robotics Research*, 0(0): 02783649241282422, 0. doi: 10.1177/02783649241282422. URL https://doi.org/10.1177/02783649241282422.

[25] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, Sebastian Nowozin, Joshua Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. *Advances in neural information processing systems*, 32, 2019.

[26] J Jon Ryu, Xiangxiang Xu, HS Erol, Yuheng Bu, Lizhong Zheng, and Gregory W Wornell. Operator svd with neural networks via nested low-rank approximation. *arXiv preprint arXiv:2402.03655*, 2024.

[27] DAVID W. Scott. Feasibility of multivariate density estimates. *Biometrika*, 78(1):197–205, 1991. URL http://dx.doi.org/10.1093/biomet/78.1.197.

[28] Ralph C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, January 2013. URL http://dx.doi.org/10.1137/1.9781611973228.

[29] Le Song, Jonathan Huang, Alex Smola, and Kenji Fukumizu. Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 961–968, 2009.

[30] Masashi Sugiyama, Taiji Suzuki, and Takafumi Kanamori. *Density ratio estimation in machine learning*. Cambridge University Press, 2012.

[31] Yao-Hung Hubert Tsai, Han Zhao, Makoto Yamada, Louis-Philippe Morency, and Russ R Salakhutdinov. Neural methods for point-wise dependency estimation. In *Advances in Neural Information Processing Systems*, volume 33, pages 62–72. Curran Associates, Inc., 2020.

[32] Giacomo Turri, Luigi Bonati, Kai Zhu, Massimiliano Pontil, and Pietro Novelli. Self-supervised evolution operator learning for high-dimensional dynamical systems. *arXiv preprint arXiv:2505.18671*, 2025.

[33] Giulio Turrisi, Valerio Modugno, Lorenzo Amatucci, Dimitrios Kanoulas, and Claudio Semini. On the benefits of gpu sample-based stochastic predictive controllers for legged locomotion. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 13757–13764, 2024. doi: 10.1109/IROS58592.2024.10801698.

[34] Ziyu Wang, Yucen Luo, Yueru Li, Jun Zhu, and Bernhard Schölkopf. Spectral representation learning for conditional moment models. *arXiv preprint arXiv:2210.16525*, 2022.

[35] Larry Wasserman. *All of Nonparametric Statistics*. Springer Texts in Statistics. Springer, New York, NY, 1 edition, May 2007.

[36] Maurice Weiler, Patrick Forré, Erik Verlinde, and Max Welling. *Equivariant and Coordinate Independent Convolutional Networks*. World Scientific, 2023. URL https://maurice-weiler.gitlab.io/cnn_book/EquivariantAndCoordinateIndependentCNNs.pdf.

Appendix

Table of Contents

A Symbols and notation **9**

B Symmetry constraints on conditional expectations **10**

C $\mathbb{G}$-Equivariant bilinear NN architecture **10**

D Experimental setup **11**

E Background on group and representation theory **18**

F Representation theory of symmetric function spaces **21**

A Symbols and notation

	Numbers and Arrays
x	A scalar, or scalar function $x(\cdot)$
$\boldsymbol{x}$	A vector, or vector-valued function $\boldsymbol{x}(\cdot)$
$\boldsymbol{x}_1 \oplus \boldsymbol{x}_2$	Direct sum (stacking) of vectors, such that $\boldsymbol{x}_1 \oplus \boldsymbol{x}_2 := \left[\begin{smallmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{smallmatrix}\right]$
$\boldsymbol{K}$	A matrix
$\boldsymbol{A} \oplus \boldsymbol{B}$	Direct sum of matrices, such that $\boldsymbol{A} \oplus \boldsymbol{B} := \left[\begin{smallmatrix}\boldsymbol{A} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{B}\end{smallmatrix}\right]$
K	A linear operator
$\boldsymbol{I}$	Identity matrix
$\delta_{i,j}$	The Kronecker function, equal to 1 when $i = j$, and 0 when $i \neq j$
	Sets, Vector Spaces, and Function Spaces
$\mathcal{X}, \mathcal{Z}, \mathcal{H}, \mathcal{F}$	A vector or Hilbert space
$\mathbb{R}, \mathbb{C}$	The set of real and complex numbers
$\mathcal{X} \oplus \mathcal{Y}$	Direct sum of vector spaces $\mathcal{X}$ and $\mathcal{Y}$, such that if $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$, then $\boldsymbol{x} \oplus \boldsymbol{y} \in \mathcal{X} \oplus \mathcal{Y}$
$\mathcal{L}^2_{\mathbf{x}} := \mathcal{L}^2_{P_{\mathbf{x}}}(\mathcal{X}, \mathbb{R})$	The Hilbert space of square-integrable functions on $\mathcal{X}$ with respect to the measure $P_{\mathbf{x}}$, defined as $\mathcal{L}^2_{P_{\mathbf{x}}}(\mathcal{X}) := \{f \mid \int_{\mathcal{X}} \lvert f(\boldsymbol{x})\rvert^2 P_{\mathbf{x}}(d\boldsymbol{x}) < \infty\}$
$\langle f, f' \rangle_{P_{\mathbf{x}}}$	Inner product between f an f' in $\mathcal{L}^2_{P_{\mathbf{x}}}\mathcal{X}$, defined as $\langle f, f' \rangle_{P_{\mathbf{x}}} := \int_{\mathcal{X}} f(\boldsymbol{x}) f'(\boldsymbol{x}) P_{\mathbf{x}}(d\boldsymbol{x})$
	Group and Representation theory
$\mathbb{G}$	A symmetry group
g, g_1, g_a	A symmetry group element
$g \triangleright \boldsymbol{x}$	The (left) group action of g on $\boldsymbol{x}$ defined by $g \triangleright \boldsymbol{x} := \boldsymbol{\rho}_{\mathcal{X}}(g)\mathcal{X}$
$\boldsymbol{\rho}_{\mathcal{X}}$	A representation of the group $\mathbb{G}$ on the vector space $\mathcal{X}$, defined for a chosen basis of $\mathcal{X}$
$\bar{\boldsymbol{\rho}}_k$	An irreducible representation Def. E.7 of the group $\mathbb{G}$
$\boldsymbol{\rho}_{\mathcal{X}}(g)$	Representation of the group element g on the vector space $\mathcal{X}$, defined for a chosen basis $\mathcal{X}$
$\boldsymbol{\rho}_{\mathcal{X}} \oplus \boldsymbol{\rho}_{\mathcal{Y}}$	Direct sum of group representations, such that $\boldsymbol{\rho}_{\mathcal{X}}(g) \oplus \boldsymbol{\rho}_{\mathcal{Y}}(g) := \begin{bmatrix}\boldsymbol{\rho}_{\mathcal{X}}(g) & \\ & \boldsymbol{\rho}_{\mathcal{Y}}(g)\end{bmatrix}$
$\mathbb{G}\boldsymbol{x}$	The group orbit of $\boldsymbol{x}$, defined as $\mathbb{G}\boldsymbol{x} := \{g \triangleright \boldsymbol{x} \mid g \in \mathbb{G}\}$
$\gamma_{\mathbb{G}'}(A)$	The symmetry index of a set $A \subseteq \mathcal{X}$ w.r.t. probability distribution on $\mathcal{X}$ and group elements $\mathbb{G}' \subseteq \mathbb{G}$
$\mathbb{G}_a \times \mathbb{G}_b$	Direct product of groups $\mathbb{G}_a$ and $\mathbb{G}_b$
$\mathbb{U}(\mathcal{X})$	Unitary group on the vector space $\mathcal{X}$
$\mathbb{GL}(\mathcal{X})$	General Linear group on the vector space $\mathcal{X}$, a.k.a the space of invertible matrices in $\mathbb{R}^{\lvert\mathcal{X}\rvert \times \lvert\mathcal{X}\rvert}$
$\mathbb{C}_n$	Cyclic group of order n
$\mathbb{K}_4$	Klein four-group
	Probability Theory
$\mathbf{x} \sim \mathbb{P}(\mathbf{x})$	Random vector $\mathbf{x} \in \mathcal{X}$ has distribution $\mathbb{P}(\mathbf{x})$
$P_{\mathbf{x}}$	A probability measure on the space $\mathcal{X}$
$\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})]$	Expectation of $f(\mathbf{x})$ with respect to $P_{\mathbf{x}}$
$\mathrm{Cov}(f(\mathbf{x}))$	Variance of $f(\mathbf{x})$ with respect to $P_{\mathbf{x}}$, define as $\mathbb{E}_{\mathbf{x}}(f(\mathbf{x}) - \mathbb{E}_{\mathbf{x}} f(\mathbf{x}))^2$
$\mathrm{Cov}(f(\mathbf{x}), h(\mathbf{y}))$	Covariance of $f(\mathbf{x})$ and $h(\mathbf{y})$ with respect to the joint distribution $P_{\mathbf{xy}}$, defined as $\mathbb{E}_{\mathbf{xy}}(f(\mathbf{x}) - \mathbb{E}_{\mathbf{x}} f(\mathbf{x}))(h(\mathbf{y}) - \mathbb{E}_{\mathbf{y}} h(\mathbf{y}))$
$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution over $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$

B Symmetry constraints on conditional expectations

Under the assumed symmetry priors in (6) the conditional expectation of $\mathbf{y}$ is a $\mathbb{G}$-equivariant function/map. This property is depicted in Fig. 3-center and proved in the following proposition.

Proposition B.1 ($\mathbb{G}$-equivariant conditional expectations)**.** *Let $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$ be two vector valued random variables satisfying the symmetry priors of Eq. (6). Then, the conditional expectation of $\mathbf{y}$ given $\mathbf{x}$ is $\mathbb{G}$-equivariant, since, for every $g \in \mathbb{G}, \boldsymbol{x} \in \mathcal{X}$,*

$$\begin{aligned}
\mathbb{E}[\mathbf{y}|\mathbf{x} = g \triangleright_{\mathcal{X}} \boldsymbol{x}] &= g \triangleright_{\mathcal{Y}} \mathbb{E}[\mathbf{y}|\mathbf{x} = \boldsymbol{x}] \\
&= \int_{\mathcal{Y}} g \triangleright_{\mathcal{Y}} \boldsymbol{y} \; P_{\mathbf{y}|\mathbf{x}}(d\boldsymbol{y}|\boldsymbol{x}) \\
&= \int_{\mathcal{Y}} \boldsymbol{y} \; P_{\mathbf{y}|\mathbf{x}}\Big(g^{-1} \triangleright_{\mathcal{Y}} d\boldsymbol{y}|\boldsymbol{x}\Big) \\
&= \int_{\mathcal{Y}} \boldsymbol{y} \; P_{\mathbf{y}|\mathbf{x}}(d\boldsymbol{y}|g \triangleright_{\mathcal{X}} \boldsymbol{x}) \quad \textit{(by Eq. (6))} \\
&= \mathbb{E}[\mathbf{y}|\mathbf{x} = g \triangleright_{\mathcal{X}} \boldsymbol{x}].
\end{aligned}$$

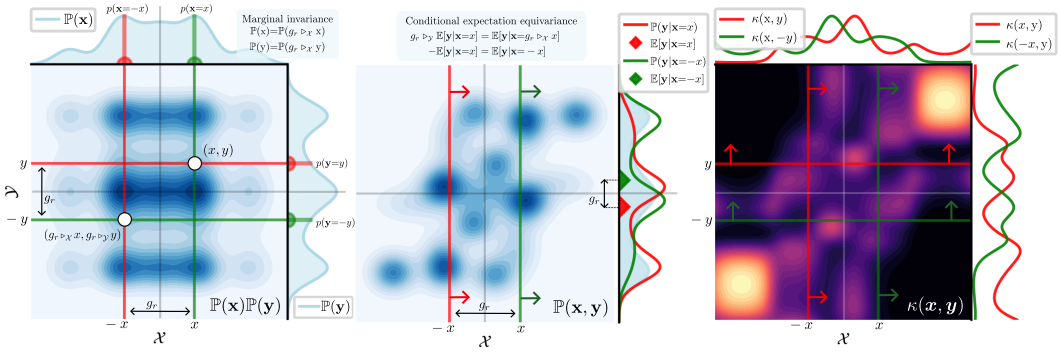

Figure 3: Example of symmetric random variables $(\mathrm{x}, \mathrm{y}) \sim \mathcal{X} \times \mathcal{Y} \subset \mathbb{R} \times \mathbb{R}$, whose marginals $\mathbb{P}(\mathrm{x})$ and $\mathbb{P}(\mathrm{y})$; joint $\mathbb{P}(\mathrm{x}, \mathrm{y})$; and conditional $\mathbb{P}(\mathrm{y}|\mathrm{x})$ distributions are invariant to reflections of the data: $g_r \triangleright_{\mathcal{X}} x = -x$ and $g_r \triangleright_{\mathcal{Y}} y = -y$, where g_r denotes the reflection element of the reflection symmetry group $\mathbb{C}_2 := \{e, g_r | g_r^2 = e\}$. Consequently, the PMD $\kappa(x, y)$ is $\mathbb{C}_2$-invariant.

C $\mathbb{G}$-Equivariant bilinear NN architecture

This section outlines how to construct a $\mathbb{G}$-equivariant disentangled representation for the random variables $\mathbf{x}$ and $\mathbf{y}$ using **any** type of $\mathbb{G}$-equivariant NN architecture backbone, such as MLP, CNNs, Transformers, and others.

Let $\boldsymbol{f_\theta} : \mathcal{X} \mapsto \mathbb{R}^r$ and $\boldsymbol{h_\theta} : \mathcal{Y} \mapsto \mathbb{R}^r$ be two $\mathbb{G}$-equivariant NNs, whose outputs will be interpreted as the basis functions of the truncated symmetric function spaces $\mathcal{F}_\mathbf{x} \subset \mathcal{L}^2_\mathbf{x}$ and $\mathcal{F}_\mathbf{y} \subset \mathcal{L}^2_\mathbf{y}$. Assume, the group representations on $\mathcal{F}_\mathbf{x}$ and $\mathcal{F}_\mathbf{y}$ are constructed from multiplicities of the group's regular representation, $\boldsymbol{\rho}_{\mathcal{F}_\mathbf{x}} = \bigoplus_{n=1}^{r/|\mathbb{G}|} \boldsymbol{\rho}_{\text{reg}}$ and $\boldsymbol{\rho}_{\mathcal{F}_\mathbf{y}} = \bigoplus_{n=1}^{r/|\mathbb{G}|} \boldsymbol{\rho}_{\text{reg}}$—as done usually in practice [36]. Since for (most) finite groups, the decomposition of $\boldsymbol{\rho}_{\text{reg}}$ into *irreps* is known or can be computed, we have access to the analytical change of basis $\boldsymbol{Q}_\mathbf{x} : \mathcal{F}_\mathbf{x} \mapsto \mathcal{F}_\mathbf{x}$ and $\boldsymbol{Q}_\mathbf{y} : \mathcal{F}_\mathbf{y} \mapsto \mathcal{F}_\mathbf{y}$ to transition to the isotypic basis. Consequently, we can directly parameterize the representations of the random variables in disentangled form as:

$$\boldsymbol{\phi_\theta}(\cdot) = \boldsymbol{Q}_\mathbf{x}^\top(\boldsymbol{f_\theta}(\cdot) - \mathbb{E}_\mathbf{x}[\boldsymbol{f_\theta}(\mathbf{x})]), \quad \boldsymbol{\psi_\theta}(\cdot) = \boldsymbol{Q}_\mathbf{y}^\top(\boldsymbol{h_\theta}(\cdot) - \mathbb{E}_\mathbf{y}[\boldsymbol{h_\theta}(\mathbf{y})]). \tag{10}$$

Given that during training these representations are not orthogonal, the truncated operator is parameterized as the trainable $\mathbb{G}$-equivariant matrix $\boldsymbol{E_\theta} = \oplus_k^{n_{\text{iso}}} \boldsymbol{E}_{\boldsymbol{\theta}}^{(k)} = \oplus_k^{n_{\text{iso}}} \boldsymbol{O}^{(k)} \otimes \boldsymbol{I}_{d_k}$ with parameters

$\{\boldsymbol{O}^{(k)} \in \mathbb{R}^{m_k \times m_k}\}_{k=1}^{n_{\text{iso}}}$. Hence, the kernel of each the truncated operator is given in terms of the model free parameters by:

$$\kappa_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{1}_{P_{\mathbf{x}}}(\boldsymbol{x})\mathbb{1}_{P_{\mathbf{y}}}(\boldsymbol{y}) + \sum_{k=1}^{n_{\text{iso}}} \sum_{s,t}^{m_k} \boldsymbol{O}_{s,t}^{(k)} \sum_{i,j}^{d_k} \phi_{s,i}^{\boldsymbol{\theta}(k)}(\boldsymbol{x}) \psi_{t,j}^{\boldsymbol{\theta}(k)}(\boldsymbol{y}). \tag{11}$$

Note that after training, the SVD of the learned operator can be computed by exploiting the constraints imposed by the operator's $\mathbb{G}$-equivariance (see Fig. 4). Importantly, once changed to the spectral basis, the group action on the approximated spectral basis matches that on the isotypic basis.

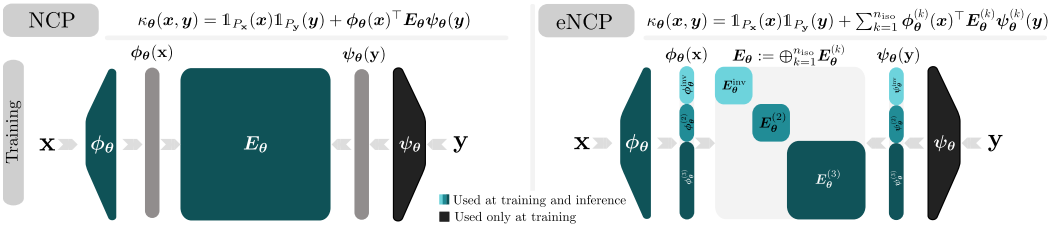

Figure 4: **Left:** NCP's bilinear NN architecture. **Right:** eNCP's $\mathbb{G}$-equivariant bilinear NN architecture, featuring $\boldsymbol{\phi}_{\boldsymbol{\theta}}$ and $\boldsymbol{\psi}_{\boldsymbol{\theta}}$ as $\mathbb{G}$-equivariant NNs and $\boldsymbol{E}_{\boldsymbol{\theta}}$ as a $\mathbb{G}$-equivariant block-diagonal matrix. Each block is equivariant to a quotient group $\mathbb{G}^{(k)} \subseteq \mathbb{G}$ and is constrained to have singular spaces of dimension at least d_k—the dimension of the irreducible representation of type k.

D Experimental setup

In this section we provide details on the experimental setup. We first describe general design choices and hyperparameters and then provide details for each experiment.

Sample efficiency experiments For both the conditional expectation operator approximation and the $\mathbb{G}$-equivariant regression experiments, we evaluate model performance by measuring sample efficiency/complexity. To do so, we partition the dataset $\mathbb{D} = \{(\boldsymbol{x}_n, \boldsymbol{y}_n)\}_{n=1}^{N}$ into training, validation, and testing splits in proportions of 70%, 15%, and 15%, respectively. With fixed validation and testing sets, we iteratively train the models on increasing portions of the training set and report the test performance for each size.

For each training set size, we select the model checkpoint with the best validation loss to compute the test performance. Thus, these experiments quantify the generalization error (or true risk) and its evolution as a function of the training set size.

NNs architectures and hyperparameters To compare our equivariant representation learning framework with other contrastive and supervised methods, all (inference) models share a similar fixed architectural footprint. For the baseline models, the only hyperparameter tuned is the learning rate, whereas for the NCP and eNCP models we additionally tune the regularization weight γ in Eq. (5). Further details for each experiment are provided in the corresponding sections below.

Code reproducibility All experiments, plots and examples are provided in the open-access repository and python package *symm_rep_learn*.

D.1 $\mathbb{G}$-equivariant regression of robot's CoM momenta

In this experiment, we evaluate the quality of the learned representations using the contrastive loss Eq. (5) alongside supervised learning baselines trained with the standard MSE loss. The task is a $\mathbb{G}$-equivariant benchmark in robotics presented in [24], with the goal of predicting a quadruped robot's CoM linear $\boldsymbol{l} \in \mathbb{R}^3$ and angular momenta $\boldsymbol{k} \in \mathbb{R}^3$ from noisy observations of the robot's generalized positions $\boldsymbol{q} \in \mathbb{R}^{12}$ and velocity coordinates $\dot{\boldsymbol{q}} \in \mathbb{R}^{12}$. Consequently, the random variables are defined as $\boldsymbol{x} = \boldsymbol{q} + \epsilon_{\boldsymbol{q}} \oplus \dot{\boldsymbol{q}} + \epsilon_{\dot{\boldsymbol{q}}}$ and $\boldsymbol{y} = \boldsymbol{l} \oplus \boldsymbol{k}$, where $\epsilon_{\boldsymbol{q}} \in \mathbb{R}^{12}$ and $\epsilon_{\dot{\boldsymbol{q}}} \in \mathbb{R}^{12}$ are independent Gaussian noise terms that model sensor noise. The function computing the CoM momenta from these proprioceptive observations is highly non-linear and $\mathbb{G}$-equivariant whenever $\mathbb{G}$ is a morphological symmetry group of the robot (see Fig. 5 and [24] for details).

The robot considered is the quadruped robot Solo (Fig. 5-right), which possesses a symmetry group of order 8: $\mathbb{G} = \mathbb{K}_4 \times \mathbb{C}_2$, as depicted in this animation showing 8 symmetric robot configurations along with their corresponding linear and angular momenta vectors.

NN architectures We configure all models under consideration (eNCP, NCP, eMLP, and MLP) to have an inference-time NN architecture with a similar footprint. In particular, the encoder network for $\mathbf{x}$ in NCP and eNCP is designed similarly to the NN used in MLP/eMLP. The idea is to test how a model with the same capacity performs on the downstream task of classification when trained using either the representation learning loss or a supervised learning loss. The backbone of all architectures is a standard multilayer perceptron consisting of three hidden layers, each with 512 units, followed by a final hidden layer containing 128 units. This final layer encodes the feature vector r for the NCP and eNCP models. Crucially, since $\mathbb{G}$-equivariance enforces weight sharing in the NN architecture, the encoder NN for eNCP and eMLP comprises $\times 8$ fewer parameters than their symmetry-agnostic counterparts.

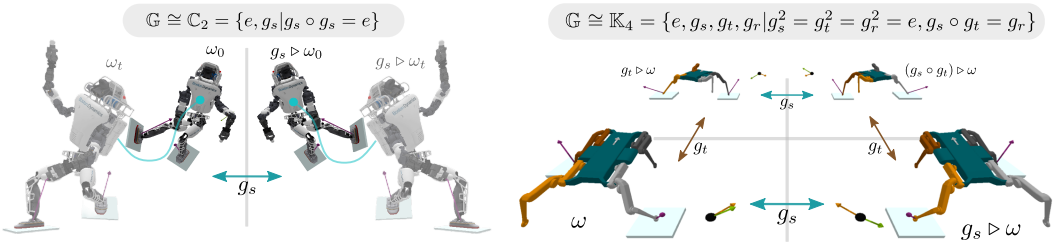

Figure 5: Example of morphological finite symmetry in robotics. **Left**: A humanoid robot with the reflectional symmetry group $\mathbb{G} \equiv \mathbb{C}_2$. **Right**: The quadruped robot Solo with the symmetry group $\mathbb{G} = \mathbb{K}_4 \times \mathbb{C}_2$ (only $\mathbb{K}_4$ is shown for clarity). The robot's center of mass linear $\boldsymbol{l} \in \mathbb{R}^3$ and angular $\boldsymbol{k} \in \mathbb{R}^3$ momentum are depicted as orange and green vectors, respectively, for each symmetric configuration. Images adapted from Ordoñez-Apraez et al. [24] with author approval.

D.2 Uncertainty quantification via conditional quantile regression

The goal of these experiments benchmark is to learn the family of conditional distributions $\mathbb{P}(\mathbf{y} \mid \mathbf{x} = \cdot)$ for a bivariate random variable $\mathbf{y} = [\mathrm{y}_0, \mathrm{y}_1] \in \mathbb{R}^2$ given a scalar covariate $\mathbf{x} \in \mathbb{R}$. Once $\mathbb{P}(\mathbf{y} \mid \mathbf{x})$ is recovered, the practitioner can estimate *conditional* $(1-\alpha)$–confidence regions by regressing the lower and upper conditional quantiles $q_{\alpha/2}(\mathbf{x})$, $q_{1-\alpha/2}(\mathbf{x})$ for any desired miscoverage level $\alpha \in (0, 1)$. In particular, a 95% confidence region corresponds to $\alpha = 0.05$, so the two quantiles of interest are $q_{0.025}(\mathbf{x})$ and $q_{0.975}(\mathbf{x})$. See Fig. 6 for a visual representation of the problem.

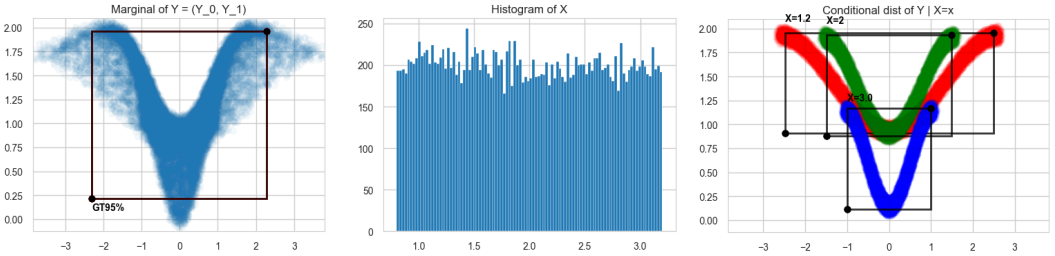

Figure 6: Synthetic experiment in uncertainty quantification, originally proposed by Feldman et al. [7]. The task is to predict the 95% confidence intervals (black bounding boxes) of a random variable $\mathbf{y} \in \mathbb{R}^2$ conditioned on a scalar random variable $\mathbf{x} \in \mathbb{R}$. **Left:** The marginal distribution $\mathbb{P}(\mathbf{y})$. **Middle:** The marginal distribution $\mathbb{P}(\mathbf{x})$. **Right:** Example conditional distributions $\mathbb{P}(\mathbf{y}|\mathbf{x} = \cdot)$ for different conditioning values.

Conditional quantile regression models We compare the NCP and proposed eNCP models to a standard baseline for parametric NN conditional quantile regression, namely CQR [7], which uses

two separate NNs to predict the lower and upper quantiles of the conditional distribution, trained with a pinball loss function (see [7] for details). Both models use MLP backbones with similar parameter counts, ensuring that improvements are solely due to the loss functions.

Furthermore, CQR can only be trained for specific confidence intervals, requiring retraining for different quantiles. In contrast, the NCP and eNCP models regress the CCDF of each dimension of $\mathbf{y}$ given $\mathbf{x}$. Thus, they can estimate conditional quantiles for any confidence interval via the quantile estimation algorithm from the CCDF described in Kostic et al. [15] without retraining. See details in Fig. 7.

Evaluation metrics: coverage and set size Let $\mathbb{C}_{1-\alpha}(\mathbf{x}) \subseteq \mathbb{R}^d$ denote a *prediction set* of nominal level $(1-\alpha)$ produced by a conditional quantile regression model for the response $\mathbf{y} \in \mathbb{R}^d$ given the covariate $\mathbf{x} \in \mathbb{R}^p$. In all experiments we assess two complementary metrics.

- **Coverage.** The conditional *coverage* of $\mathbb{C}_{1-\alpha}$ is the probability that the true response is captured by the predicted region,

$$\mathrm{c}_{1-\alpha}(\mathbf{x}) := \mathbb{P}\big(\mathbf{y} \in \mathbb{C}_{1-\alpha}(\mathbf{x}) \mid \mathbf{x}\big), \qquad \text{with the target } \ \mathrm{c}_{1-\alpha}(\mathbf{x}) \approx 1-\alpha \ \ \forall \mathbf{x}. \tag{12}$$

  In practice we report the *marginal* coverage $\widehat{\mathbb{E}}_{\mathbf{x}}[\mathrm{c}_{1-\alpha}(\mathbf{x})]$, estimated on a large held-out sample; values above (resp. below) $1-\alpha$ indicate over- (resp. under-) coverage.

- **Relaxed Coverage (r-Coverage).** The conditional *relaxed coverage* of $\mathbb{C}_{1-\alpha}$ is defined as the probability that each scalar component of the response lies within its corresponding predicted confidence interval. Formally, if $\mathbf{y} = [\mathrm{y}_1, \dots, \mathrm{y}_d]$ and $\mathbb{C}_{1-\alpha}(\mathbf{x})$ has corresponding marginal intervals $\mathbb{C}^{(i)}_{1-\alpha}(\mathbf{x})$ for $i \in \{1, \dots, d\}$, then

$$\mathrm{rc}_{1-\alpha}(\mathbf{x}) := \prod_{i=1}^{d} \mathbb{P}\Big(\mathrm{y}_i \in \mathbb{C}^{(i)}_{1-\alpha}(\mathbf{x}) \,\Big|\, \mathbf{x}\Big), \tag{13}$$

  with the target $\mathrm{rc}_{1-\alpha}(\mathbf{x}) \approx 1-\alpha$ for all $\mathbf{x}$. As with coverage, we report the *marginal* relaxed coverage $\widehat{\mathbb{E}}_{\mathbf{x}}[\mathrm{rc}_{1-\alpha}(\mathbf{x})]$.

- **Set size.** To quantify how informative the region is, we measure its *size* (volume) under the Lebesgue measure λ^d:

$$\mathrm{Size}_{1-\alpha}(\mathbf{x}) := \mathrm{vol}\big(\mathbb{C}_{1-\alpha}(\mathbf{x})\big). \tag{14}$$

  Smaller sets correspond to sharper uncertainty estimates, provided the required coverage is met. For multidimensional responses the volume is expressed in the natural units of $\mathbb{R}^d$; for $d = 1$ it reduces to the interval length. As with coverage, we report the marginal expectation $\widehat{\mathbb{E}}_{\mathbf{x}}[\mathrm{Size}_{1-\alpha}(\mathbf{x})]$ so that models can be compared fairly across the entire input distribution.

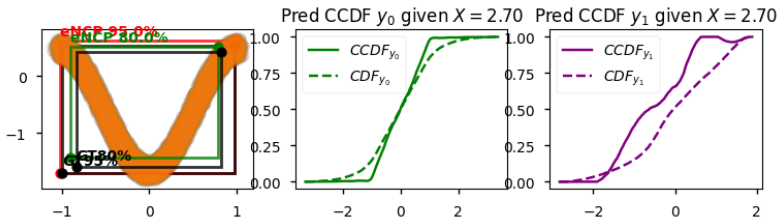

Figure 7: Prediction of the 80% and 95% confidence intervals for the random variable $\mathbf{y}$ in experiment App. D.2 using the proposed eNCP model. The model estimates the CCDF by discretizing each dimension of $\mathbf{y} = [\mathrm{y}_1, \mathrm{y}_2]$ into 100 bins and computing the conditional probabilities $\mathbb{P}(\mathrm{y}_i \in \mathbb{A}_n | \mathbf{x} = \cdot) := [\mathsf{E}_{\mathbf{y}|\mathbf{x}} \mathbb{1}_{\mathbb{A}_n}](\cdot)$ for all $n \in [100]$ based on the learned conditional expectation operator $\kappa_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y})$ (see Tab. 1). Here, $\mathbb{A}_n$ comprises the bins from the 0-th to the n-th. This yields the estimated CCDF for y_1 (center) and y_2 (right) at $\mathbf{x} = 2.7$. The CCDFs can then be used to estimate upper and lower quantiles for any confidence interval [15]. In practice, the eNCP model regresses 2×100 variables in a single forward pass. Thus, the final layer of the conditional quantile regression model is a linear layer of size $r \times (2 \times 100)$, where r is the number of features in the $\mathbf{y}$ representation.

D.2.1 Synthetic benchmark

The goal of these experiments is to learn the conditional distributions $\mathbb{P}(\mathbf{y} \mid \mathbf{x} = \cdot)$ for a bivariate random variable $\mathbf{y} = [\mathrm{y}_0, \mathrm{y}_1] \in \mathbb{R}^2$ given a scalar covariate $\mathbf{x} \in \mathbb{R}$. Following Feldman et al. [7], the covariate is sampled uniformly: $\mathbf{x} \sim \mathrm{Unif}\big(0.8,\ 3.2\big)$, and the response variable $\mathbf{y}$ is produced by a non-linear transformation of auxiliary latent variables (see Fig. 6):

$$
\begin{aligned}
\mathrm{y}_0 &= \frac{\mathrm{z}}{\beta\,\mathbf{x}} \;+\; \mathrm{r}\cos\phi, & \mathrm{z} &\sim \mathrm{Unif}(-\pi, \pi),\\
\mathrm{y}_1 &= \tfrac{1}{2}\big(-\cos\mathrm{z} + 1\big) \;+\; \mathrm{r}\sin\phi \;+\; \sin\mathbf{x}, & \phi &\sim \mathrm{Unif}(0, 2\pi),\\
& & \mathrm{r} &\sim \mathrm{Unif}(-0.1, 0.1).
\end{aligned}
$$

Here, $\beta > 0$ is a scaling constant.

The additive perturbation $r(\cos\phi, \sin\phi)$ yields heteroskedastic, anisotropic noise, whereas the $\frac{1}{2}(-\cos z + 1)$ and $\sin \mathbf{x}$ terms introduce strong non-monotonicity and interaction effects between $\mathbf{x}$ and $\mathbf{y}$. As a result, the conditional quantile functions $\mathbf{x} \mapsto q_\tau(\mathbf{x})$ are highly non-linear, making this dataset an ideal low-dimensional experiment for conditional quantile regression methods.

Results The experiment results are depicted in Fig. 9. Where the NCP and eNCP models outperform the baseline CQR model in terms of both coverage and set size. Furthermore, Fig. 8 illustrates the basis functions learned by the NCP and eNCP models for the random variable $\mathbf{y} = [\mathrm{y}_0, \mathrm{y}_1]$. In contrast to the standard NCP model, the eNCP model incorporates symmetry priors, enabling a clean separation of its latent representation into two orthogonal subspaces: one corresponding to $\mathbb{C}_2$-invariant functions and the other to functions that change sign under reflection.

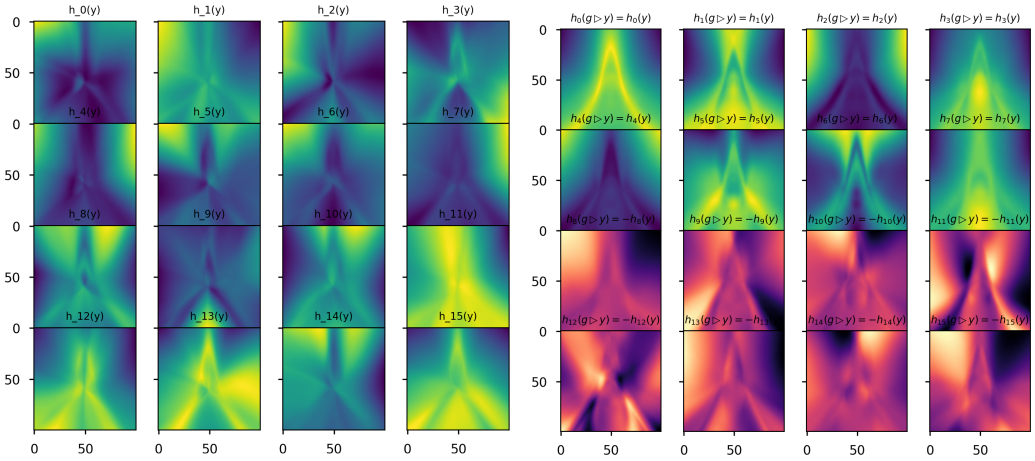

Figure 8: **Left:** Learned basis functions from the NCP model for $\mathbf{y} = [\mathrm{y}_0, \mathrm{y}_1]$. **Right:** Learned basis functions from the eNCP model for $\mathbf{y}$. The marginal distribution of $\mathbf{y}$ exhibits reflection symmetry $g_r \triangleright_{\mathcal{Y}} \mathbf{y} = [-\mathrm{y}_0, \mathrm{y}_1]$ under $\mathbb{G} = \mathbb{C}_2$. Incorporating this prior, the eNCP model decomposes its latent space as $\mathcal{F}_{\mathbf{y}} = \mathcal{F}_{\mathbf{y}}^{\mathrm{inv}} \oplus \mathcal{F}_{\mathbf{y}}^{(2)}$, with the first subspace capturing $\mathbb{C}_2$-invariant functions and the second capturing those that change sign under reflection. The orthogonality of these subspaces allows independent optimization of the basis functions.

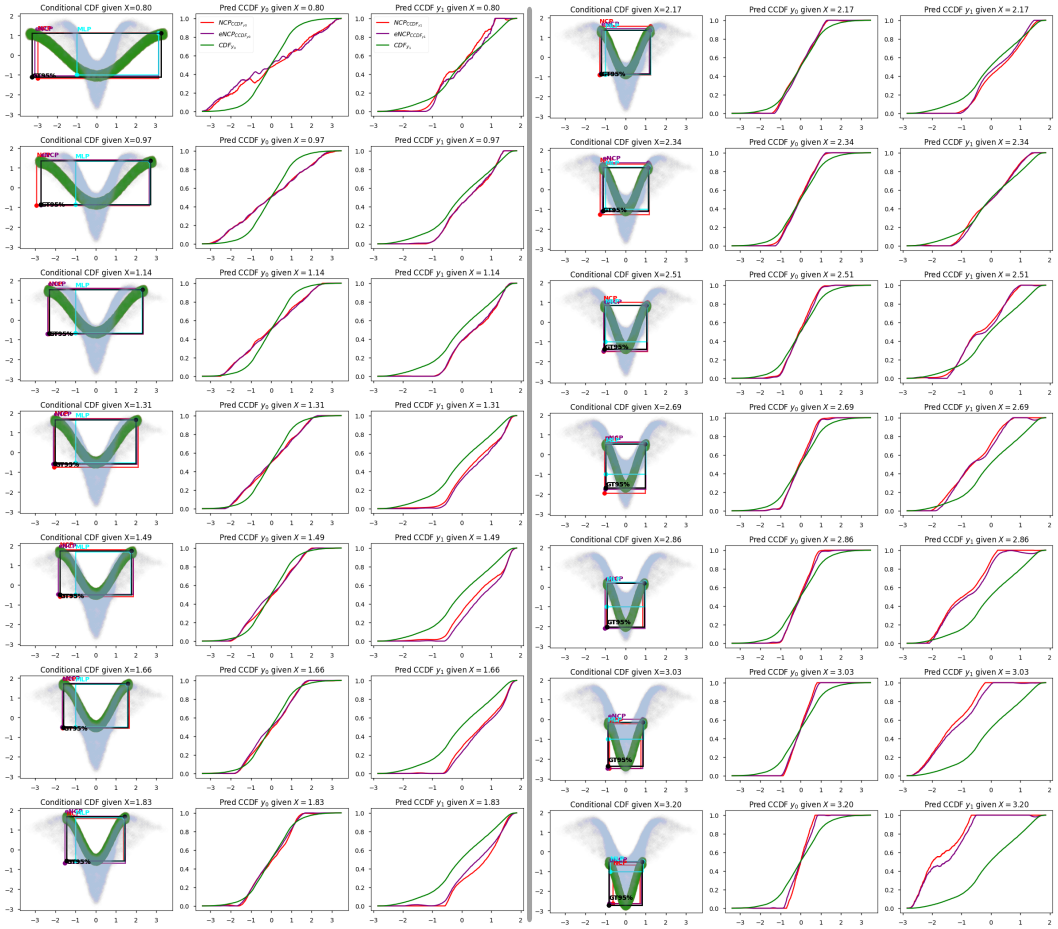

Figure 9: Results of a synthetic experiment in uncertainty quantification comparing CQR, NCP, and eNCP models. The task, originally proposed by Feldman et al. [7], is to predict the 95% confidence intervals of a random variable $\mathbf{y} \in \mathbb{R}^2$ conditioned on a scalar random variable $\mathbf{x} \in \mathbb{R}$. The conditional distributions $\mathbb{P}(\mathbf{y}|\mathbf{x} = \cdot)$ are shown in the left and fourth columns for different conditioning values, while the second-third and fifth-sixth columns display the CCDF predicted by the eNCP and NCP models, respectively. The CQR model directly regresses the upper and lower quantiles for each dimension of $\mathbf{y}$ and must be retrained if the confidence interval probability changes. In contrast, since the NCP and eNCP models estimate the CCDF for each dimension, these predictions can be easily adapted to any confidence interval probability by simply changing the threshold value.

	Validation			Test		
	r-Coverage ↑	Coverage ↑	Set Size ↓	r-Coverage ↑	Coverage ↑	Set Size ↓
eNCP	99.3±0.0%	94.1±0.4%	$2.4\pm0.4\times10^{10}$	99.5±0.1%	95.0±0.4%	$4.3\pm3.6\times10^{9}$
NCP	96.4±0.0%	56.9±0.1%	$3.9\pm4.5\times10^{10}$	99.5±0.0%	56.9±0.3%	$2.6\pm1.4\times10^{10}$
eCQR	70.7±0.6%	7.3±1.7%	$3.7\pm2.6\times10^{8}$	84.2±0.7%	6.7±1.2%	$1.7\pm1.7\times10^{7}$
CQR	67.6±1.8%	7.6±0.4%	$2.5\pm2.4\times10^{9}$	80.5±3.7%	8.5±0.9%	$1.4\pm0.1\times10^{8}$

Table 3: Validation and test set metrics for the prediction of 95% confidence intervals on observables of a quadruped robot traversing rough terrains (see App. D.2.2). Model performance is evaluated using three metrics: **(i)** relaxed coverage (r-Coverage) (Eq. (13)), **(ii)** coverage (Eq. (12)), and **(iii)** set size (Eq. (14)). The best results are highlighted in blue. Note that although the confidence interval volumes (set size) of the eCQR and CQR models are significantly smaller than those of the NCP and eNCP models, the former fail to achieve the expected 95% coverage on both the validation and test sets. In contrast, the eNCP model attains the best overall coverage, proving its effectiveness for uncertainty quantification. Importantly, the eNCP and NCP models can be adjusted, **without retraining**, to provide confidence intervals for any desired coverage level, whereas the CQR and eCQR models must be retrained for each new level.

D.2.2 Uncertainty quantification in quadruped legged locomotion

We test how well conditional-quantile models can recover the conditional 95% confidence regions of three physically meaningful observables produced by a simulated AlienGo quadruped walking over rough terrain (see Fig. 2) under varying friction coefficients. The dataset was collected using the Quadruped-PyMPC simulation framework and model predictive controller from [33].

The observables for which state-dependent uncertainty estimates are desired are $\boldsymbol{y}_t = [U_t,\ T_t,\ \boldsymbol{\tau}_t^{\text{grf}}]^\top$, with each component defined as follows:

- **$\mathbb{G}$-invariant Kinetic Energy.** $T(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \frac{1}{2}\,\dot{\boldsymbol{q}}^\top M(\boldsymbol{q})\,\dot{\boldsymbol{q}} \in \mathbb{R}$, where $M(\boldsymbol{q})$ is the configuration-dependent inertia matrix. Noise is introduced through sensor measurement errors on the robot's degree of freedom (DoF) position $\boldsymbol{q} \in \mathbb{R}^{12}$ and velocity $\dot{\boldsymbol{q}} \in \mathbb{R}^{12}$.
- **$\mathbb{G}$-invariant Instantaneous Mechanical Work.** $U(\boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{\tau}) \in \mathbb{R}$, representing the instantaneous mechanical work exerted or absorbed by the robot. This quantity depends on the actuator torques (typically measured with noisy, biased sensors) as well as the external forces (e.g. gravity, contact forces) that are not reliably measurable due to unobserved terrain parameters.
- **$\mathbb{G}$-equivariant Ground-Reaction Forces** $\boldsymbol{\tau}_{\text{grf}} \in \mathbb{R}^{12}$, a fundamental quantity in quadruped control, whose reliable estimation and uncertainty quantification are critical for downstream tasks in robotics [17, 22].

The observables of interest are predicted using a suit of onboard proprioceptive sensory signals available at time t:

$$\boldsymbol{x}_t \;=\; \Big[\boldsymbol{q}_t,\ \dot{\boldsymbol{q}}_t,\ \boldsymbol{a}_t,\ \boldsymbol{v}_t,\ \boldsymbol{v}_{t,\text{err}},\ \boldsymbol{\omega}_t,\ \boldsymbol{\omega}_{t,\text{err}},\ \boldsymbol{g}_t,\ \dot{\boldsymbol{p}}_{t,\text{feet}},\ \boldsymbol{\tau}_t^{\text{cmd}}\Big]^\top.$$

Specifically, $\boldsymbol{q}_t \in \mathbb{R}^{n_q}$ and $\dot{\boldsymbol{q}}_t \in \mathbb{R}^{n_q}$ are the joint positions and velocities, respectively; $\boldsymbol{a}_t \in \mathbb{R}^3$ is the linear acceleration of the robot's base frame measured by the IMU; $\boldsymbol{v}_t \in \mathbb{R}^3$ is the base linear velocity, while $\boldsymbol{v}_{t,\text{err}} \in \mathbb{R}^3$ the command error base linear velocity; $\boldsymbol{\omega}_t \in \mathbb{R}^3$ and $\boldsymbol{\omega}_{t,\text{err}} \in \mathbb{R}^3$ are the base angular velocity and its command error; $\boldsymbol{g}_t \in \mathbb{R}^3$ is the gravity vector expressed in the base frame; $\dot{\boldsymbol{p}}_{t,\text{feet}} \in \mathbb{R}^{12}$ stacks the linear velocities of the four feet (three components each); and $\boldsymbol{\tau}_t^{\text{cmd}} \in \mathbb{R}^{n_q}$ contains the commanded joint torques.

Hence we design the experiments to compare models of similar footprint in number of parameters, while the loss used for training differs between the NCP and eNCP models w.r.t to the CQR and eCQR models.

NN architectures We configure all models considered eNCP, NCP, eCQR, and CQR to have an inference-time NN architecture of the similar footprint. The backbone of all architectures is a standard multilayer perceptron consisting of three hidden layers, each with 512 units, followed by

a final hidden layer containing 128 units. This final layer serves to encode the feature vector r for the NCP and eNCP models. Crucially, since $\mathbb{G}$-equivariance enforces weight sharing in the NN architecture the encoder NN for eNCP, eCQR have $\times 2$ less parameters than their symmetry-agnostic counterparts.

Results. Given sensory input $\mathbf{x}$, the model predicts a set $\mathbb{C}_{0.95}(\mathbf{x}) \subseteq \mathbb{R}^{14}$ satisfying $\mathbb{P}(\mathbf{y} \in \mathbb{C}_{0.95}(\mathbf{x}) \mid \mathbf{x}) \approx 0.95$, while minimizing its volume $\widehat{\mathbb{E}}_{\mathbf{x}}[\mathrm{vol}(\mathbb{C}_{0.95}(\mathbf{x}))]$. Empirically high coverage implies that the true $\mathbb{G}$-invariant kinetic energy, instantaneous mechanical work, and the $\mathbb{G}$-equivariant 12-dimensional ground-reaction forces lie within the predicted confidence set. In contrast, relaxed coverage (r-Coverage) quantifies the reliability of the estimates on a per-dimension basis. Tab. 3 summarizes the validation and test results for the eNCP, NCP, CQR, and eCQR models, and Fig. 10 illustrates a trajectory of GRF and their respective 90% confidence intervals for each model. Both CQR and eCQR tend to produce confidence intervals of smaller volume but fail to achieve the desired coverage on the testing set, implying that the models' confidence intervals are not reliable and require further calibration through retraining or conformal calibration [7]. In contrast, the eNCP model achieves the desired coverage on the test set while producing confidence intervals of larger volume, hence yielding reliable confidence intervals.

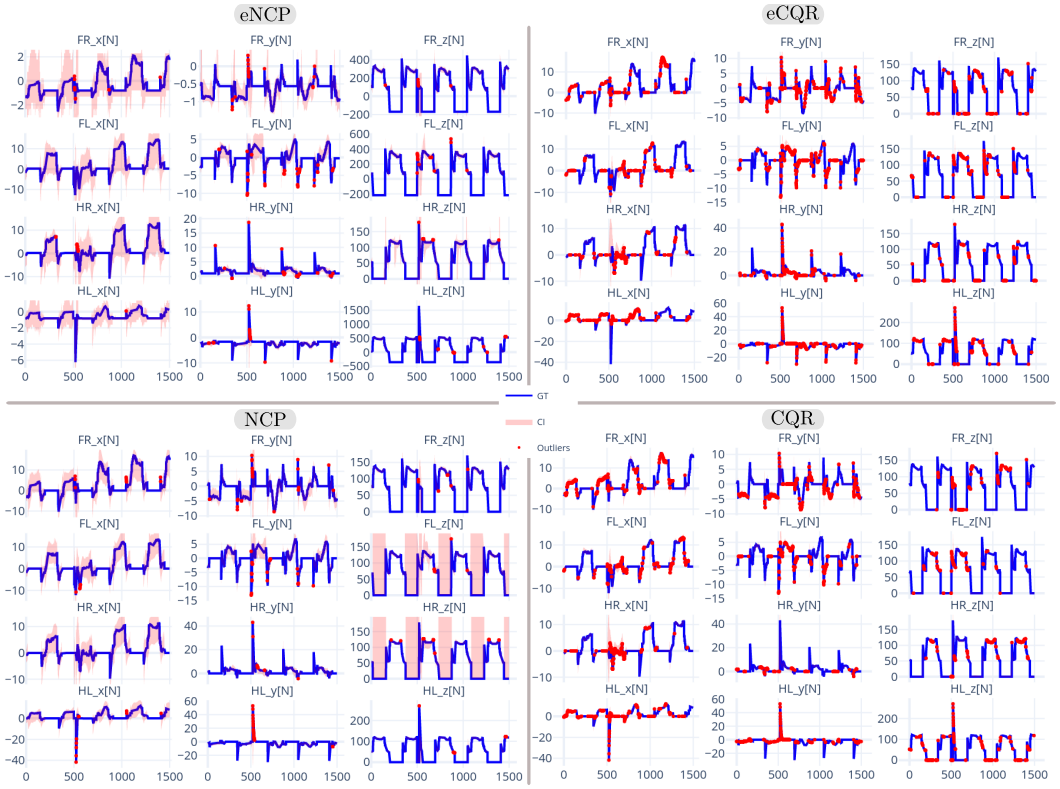

Figure 10: Prediction of 90% confidence intervals (CI) for the ground-reaction forces $\boldsymbol{\tau}_{\text{grf}} \in \mathbb{R}^{12}$ of a quadruped robot on rough terrain with varying friction. We compare the eNCP, NCP, eCQR, and CQR models based on relaxed coverage and set size (see Tab. 3). CIs are computed for each leg—front-right (FR), front-left (FL), hind-right (HR), and hind-left—along the x, y, and z axes. Forces outside the CI are highlighted in red, while those within appear in blue. Terrain variations cause significant variability in the x and y components due to differences in surface orientation and friction, whereas the z component is mainly influenced by local height changes that alter contact timing and produce short-duration high-impact forces.

E Background on group and representation theory

Group actions and representations This section provides a concise overview of the fundamental concepts in group and representation theory, which are used to define the symmetries of the random variables we consider in this work. For a comprehensive background on these topics in finite-dimensional vector spaces, see Weiler et al. [36]; for the infinite-dimensional case, consult Knapp [14]. These concepts will be referenced as needed in the main text. To begin, we define a group as an abstract mathematical object.

Definition E.1 (Group)**.** *A group is a set $\mathbb{G}$, endowed with a binary composition operator defined as:*

$$\begin{array}{rccc} (\circ): & \mathbb{G}\times\mathbb{G} & \longrightarrow & \mathbb{G} \\ & (g_1, g_2) & \longrightarrow & g_1 \circ g_2, \end{array} \tag{15a}$$

such that the following axioms hold:

$$\textit{Associativity:} \quad (g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3), \quad \forall\, g_1, g_2, g_3 \in \mathbb{G}, \tag{15b}$$

$$\textit{Identity:} \quad \exists\, e \in \mathbb{G} \textit{ such that } e \circ g = g = g \circ e, \quad \forall\, g \in \mathbb{G}, \tag{15c}$$

$$\textit{Inverses:} \quad \forall\, g \in \mathbb{G},\ \exists\, g^{-1} \in \mathbb{G} \textit{ such that } g \circ g^{-1} = e = g^{-1} \circ g. \tag{15d}$$

We are primarily interested in symmetry groups, i.e., groups of transformations acting on a set $\mathcal{X}$. Each transformation is a bijection that leaves a fundamental property invariant. For example, if $\mathcal{X}$ represents states of a dynamical system, the invariant property is the state energy (see Fig. 5); if $\mathcal{X}$ is a data space, the preserved quantity is typically the probability density/distribution (see Fig. 3).

Definition E.2 (Group action on a set [36])**.** *Let $\mathcal{X}$ be a set endowed with symmetry group $\mathbb{G}$. The (left) group action of the group $\mathbb{G}$ on the set $\mathcal{X}$ is a map:*

$$\begin{array}{rccc} (\triangleright): & \mathbb{G}\times\mathcal{X} & \longrightarrow & \mathcal{X} \\ & (g, \boldsymbol{x}) & \longrightarrow & g \triangleright \boldsymbol{x} \end{array} \tag{16a}$$

that is compatible with the group composition and identity element $e \in \mathbb{G}$, in the sense that:

$$\textit{Identity:} \quad e \triangleright \boldsymbol{x} = \boldsymbol{x}, \qquad \forall\, \boldsymbol{x} \in \mathcal{X} \tag{16b}$$

$$\textit{Associativity:} \quad (g_1 \circ g_2) \triangleright \boldsymbol{x} = g_1 \triangleright (g_2 \triangleright \boldsymbol{x}), \qquad \forall\, g_1, g_2 \in \mathbb{G}, \forall\, \boldsymbol{x} \in \mathcal{X}. \tag{16c}$$

We are primarily interested in studying symmetry transformations on sets with a vector space structure. In most practical cases, the group action on a vector space is linear, allowing symmetry transformations to be represented as linear invertible maps. These maps can be expressed in matrix form once a basis for the space is chosen.

Definition E.3 (Linear group representation)**.** *Let $\mathcal{X}$ be a vector space endowed with symmetry group $\mathbb{G}$. A* linear representation *of $\mathbb{G}$ on $\mathcal{X}$ is a map, denoted by $\boldsymbol{\rho}_{\mathcal{X}}$, between symmetry transformation and invertible linear maps on $\mathcal{X}$ (i.e., elements of the general linear group $\mathbb{GL}(\mathcal{X})$):*

$$\begin{array}{rccc} \boldsymbol{\rho}_{\mathcal{X}}: & \mathbb{G} & \longrightarrow & \mathbb{GL}(\mathcal{X}) \\ & g & \longrightarrow & \boldsymbol{\rho}_{\mathcal{X}}(g), \end{array} \tag{17a}$$

such that the following properties hold:

$$\textit{composition}: \boldsymbol{\rho}_{\mathcal{X}}(g_1 \circ g_2) = \boldsymbol{\rho}_{\mathcal{X}}(g_1)\boldsymbol{\rho}_{\mathcal{X}}(g_2), \qquad \forall\, g_1, g_2 \in \mathbb{G}, \tag{17b}$$

$$\textit{inversion}: \boldsymbol{\rho}_{\mathcal{X}}(g^{-1}) = \boldsymbol{\rho}_{\mathcal{X}}(g)^{-1}, \qquad \forall\, g \in \mathbb{G}. \tag{17c}$$

$$\textit{identity}: \boldsymbol{\rho}_{\mathcal{X}}(g \circ g^{-1}) = \boldsymbol{\rho}_{\mathcal{X}}(e) = \boldsymbol{I}, \tag{17d}$$

Whenever the vector space is of finite dimension $n < \infty$, linear maps admit a matrix form $\boldsymbol{\rho}_{\mathcal{X}}(g) \in \mathbb{R}^{n\times n}$, once a basis set $\mathbb{I}_{\mathcal{X}}$ for the vector space $\mathcal{X}$ is chosen. In this case, Eqs. (17b) to (17d) show how the composition and inversion of symmetry transformations translate to matrix multiplication and inversion, respectively. Moreover, $\boldsymbol{\rho}_{\mathcal{X}}$ allows to express a (linear) group action (Def. E.2) as a matrix-vector multiplication:

$$\begin{array}{rccc} (\triangleright): & \mathbb{G}\times\mathcal{X} & \longrightarrow & \mathcal{X} \\ & (g, \boldsymbol{x}) & \longrightarrow & g \triangleright \boldsymbol{x} := \boldsymbol{\rho}_{\mathcal{X}}(g)\boldsymbol{x}. \end{array} \tag{17e}$$

Since the matrix form of linear maps depends on the choice of basis, we can relate different matrix representations of the same linear map through changes of basis. This leads us to the concept of equivalent group representations.

Definition E.4 (Equivalent group representations)**.** *Let $\mathcal{X}$ be a vector space endowed with symmetry group $\mathbb{G}$, and let $\boldsymbol{\rho}'_{\mathcal{X}}$ and $\boldsymbol{\rho}_{\mathcal{X}}$ be two group representations of $\mathbb{G}$ on $\mathcal{X}$. They are said to be* equivalent*, denoted by $\boldsymbol{\rho}'_{\mathcal{X}} \sim \boldsymbol{\rho}_{\mathcal{X}}$, if there exists a change of basis $\boldsymbol{Q} : \mathcal{X} \to \mathcal{X}$ such that*

$$\boldsymbol{\rho}'_{\mathcal{X}}(g) = \boldsymbol{Q}\boldsymbol{\rho}_{\mathcal{X}}(g)\boldsymbol{Q}^{-1}, \quad \forall\, g \in \mathbb{G}. \tag{18}$$

Equivalent representations arise when the same group action $(\triangleright) : \mathbb{G} \times \mathcal{X} \to \mathcal{X}$ is expressed in different coordinate frames or bases. For instance, let $\mathbb{A}_{\mathcal{X}}$ and $\mathbb{B}_{\mathcal{X}}$ be two bases for $\mathcal{X} = span(\mathbb{A}_{\mathcal{X}}) = span(\mathbb{B}_{\mathcal{X}})$, and let $\boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}} : \mathcal{X} \to \mathcal{X}$ denote the change of basis from $\mathbb{A}_{\mathcal{X}}$ to $\mathbb{B}_{\mathcal{X}}$, so that $\boldsymbol{x}^{\mathbb{B}} = \boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}\boldsymbol{x}^{\mathbb{A}}$ for all $\boldsymbol{x}^{\mathbb{A}} \in \mathcal{X}$. Then the group action admits equivalent representations, $\boldsymbol{\rho}^{\mathbb{A}}_{\mathcal{X}} \sim \boldsymbol{\rho}^{\mathbb{B}}_{\mathcal{X}}$, since

$$\begin{aligned} g \triangleright \boldsymbol{x}^{\mathbb{B}} &:= \boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}(g \triangleright \boldsymbol{x}^{\mathbb{A}}), \qquad \forall g \in \mathbb{G}, \\ \boldsymbol{\rho}^{\mathbb{B}}_{\mathcal{X}}(g)\boldsymbol{x}^{\mathbb{B}} &= \boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}\big(\boldsymbol{\rho}^{\mathbb{A}}_{\mathcal{X}}(g)\boldsymbol{x}^{\mathbb{A}}\big) = \Big(\boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}\boldsymbol{\rho}^{\mathbb{A}}_{\mathcal{X}}(g){\boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}}^{-1}\Big)\boldsymbol{x}^{\mathbb{B}}, \\ \boldsymbol{\rho}^{\mathbb{B}}_{\mathcal{X}}(g) &= \boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}\boldsymbol{\rho}^{\mathbb{A}}_{\mathcal{X}}(g){\boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}}^{-1}. \end{aligned} \tag{19}$$

To reveal the modular structure of symmetric vector spaces, we often change bases to decompose them into subspaces stable under the action of the group $\mathbb{G}$, termed $\mathbb{G}$-stable subspaces. This decomposition mirrors how a symmetry group breaks down into subgroups and is essential for analyzing and simplifying group representations. We introduce the following definition.

Definition E.5 ($\mathbb{G}$-stable and irreducible subspaces)**.** *Let $\mathcal{X}$ be a vector space endowed with a group action $(\triangleright)$ of the symmetry group $\mathbb{G}$. A subspace $\mathcal{X}' \subseteq \mathcal{X}$ is said to be* $\mathbb{G}$-stable *if the action of any group element on any vector in the subspace remains within the subspace, that is,*

$$g \triangleright \boldsymbol{x} \in \mathcal{X}', \quad \forall\, \boldsymbol{x} \in \mathcal{X}' \subseteq \mathcal{X}, \forall\, g \in \mathbb{G}.$$

If the only $\mathbb{G}$-stable subspaces of $\mathcal{X}$ are $\{\mathbf{0}\}$ and $\mathcal{X}$ itself, then $\mathcal{X}$ is a irreducible $\mathbb{G}$-stable space*.*

Decomposing symmetric vector spaces into $\mathbb{G}$-stable subspaces corresponds to decomposing the group representation associated with $\triangleright$ into smaller representations acting on these $\mathbb{G}$-stable subspaces:

Definition E.6 (Decomposable representation)**.** *Let $\mathcal{X}$ be a vector space with a group action $(\triangleright)$ defined by the representation $\boldsymbol{\rho}_{\mathcal{X}}$ in a chosen basis $\mathbb{A}_{\mathcal{X}}$. The representation is* decomposable *if it is equivalent to a direct sum of two lower-dimensional representations, $\boldsymbol{\rho}_{\mathcal{X}} \sim \boldsymbol{\rho}_{\mathcal{X}_1} \oplus \boldsymbol{\rho}_{\mathcal{X}_2}$, where $\mathcal{X}_1$ and $\mathcal{X}_2$ are $\mathbb{G}$-stable subspaces of $\mathcal{X}$. Equivalently, there exists a change of basis $\boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}} : \mathcal{X} \to \mathcal{X}$ such that*

$$\boldsymbol{\rho}^{\mathbb{B}}_{\mathcal{X}} = \begin{bmatrix} \rho_{\mathcal{X}_1} & \mathbf{0} \\ \mathbf{0} & \rho_{\mathcal{X}_2} \end{bmatrix} = \boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}\boldsymbol{\rho}_{\mathcal{X}}{\boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}}^{-1}, \text{ and } \quad g \triangleright \boldsymbol{x}^{\mathbb{B}} := \boldsymbol{\rho}^{\mathbb{B}}_{\mathcal{X}}(g)\boldsymbol{x}^{\mathbb{B}} = \begin{bmatrix} \rho_{\mathcal{X}_1}(g)\boldsymbol{x}^{\mathbb{B}}_1 \\ \rho_{\mathcal{X}_2}(g)\boldsymbol{x}^{\mathbb{B}}_2 \end{bmatrix}, \text{ where } \boldsymbol{Q}^{\mathbb{B}}_{\mathbb{A}}\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}^{\mathbb{B}}_1 \in \mathcal{X}_1 \\ \boldsymbol{x}^{\mathbb{B}}_2 \in \mathcal{X}_2 \end{bmatrix}$$

This shows that the decomposition $\boldsymbol{\rho}_{\mathcal{X}} \sim \boldsymbol{\rho}_{\mathcal{X}_1} \oplus \boldsymbol{\rho}_{\mathcal{X}_2}$ corresponds to splitting the vector space into $\mathbb{G}$-stable subspaces, $\mathcal{X} = \mathcal{X}_1 \oplus \mathcal{X}_2$. Moreover, if the representation is block-diagonal in some basis set $\mathbb{B}_{\mathcal{X}}$, then $\mathbb{B}_{\mathcal{X}}$ is the union of disjoint basis sets $\mathbb{B}_{\mathcal{X}_1}$ and $\mathbb{B}_{\mathcal{X}_2}$ for $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively.

Definition E.7 (Irreducible representation)**.** *Let $\mathcal{X}$ be a vector space endowed with a group action $(\triangleright)$ of a symmetry group $\mathbb{G}$. A representation $\boldsymbol{\rho}_{\mathcal{X}}$ of $\mathbb{G}$ on $\mathcal{X}$ is said to be* irreducible *if it cannot be decomposed into smaller representations acting on proper $\mathbb{G}$-stable subspaces (Def. E.5). That is, the only $\mathbb{G}$-stable subspaces $\mathcal{X}' \subseteq \mathcal{X}$ are $\mathcal{X}' = \{\mathbf{0}\}$ and $\mathcal{X}' = \mathcal{X}$ itself.*

We have now equipped all the necessary tools to decompose symmetric vector spaces into their smallest building blocks: irreducible $\mathbb{G}$-stable subspaces.

Irreducible representations are the fundamental building blocks for all representations of the group $\mathbb{G}$. Any unitary representation can be decomposed into a direct sum of irreducible representations, analogous to the prime factorization of integers. In terms of the vector spaces on which the group acts, this decomposition of the representation corresponds to decomposing the space into $\mathbb{G}$-irreducible subspaces (Def. E.5):

Theorem E.8 (Isotypic decomposition of symmetric Hilbert spaces [14])**.** *Let $\mathbb{G}$ be a compact group and $\mathcal{H}$ a separable Hilbert space with a unitary group representation $\boldsymbol{\rho}_{\mathcal{H}} : \mathbb{G} \to \mathbb{U}(\mathcal{H})$. Then we can identify $n_{\text{iso}} \leq |\mathbb{G}|$ irreducible representations $\bar{\boldsymbol{\rho}}_k : \mathbb{G} \to \mathbb{U}(\bar{\mathcal{H}}_k)$ that allow us to decompose $\mathcal{H}$ into a sum of orthogonal subspaces, denoted* isotypic subspaces*: $\mathcal{H} = \bigoplus^{\perp}_{1 \leq k \leq n_{\text{iso}}} \mathcal{H}_k$ where each $\mathcal{H}_k = \bigoplus_{j=1}^{m_k} \mathcal{H}_{k,j}$ is the sum of at most $m_k \leq \infty$ countably many subspaces isometrically isomorphic to $\bar{\mathcal{H}}_k$.*

Isotypic decomposition and disentangled representations Whenever the symmetric vector space of interest defines a vector valued representation of some data, the isotypic decomposition of the representation space is intricately linked with the concept of *disentangled representations*

Definition E.9 (Disentangled representation (Higgins et al. [11])). *A vector representation is called a disentangled representation with respect to a particular decomposition of a symmetry group into subgroups, if it decomposes into independent subspaces, where each subspace is affected by the action of a single subgroup, and the actions of all other subgroups leave the subspace unaffected.*

The *subspaces* of Def. E.9 reefer to each of the isotypic subspaces $\mathcal{H}_i$, and the symmetry subgroups refer to the effective (matrix) group encoded by each irreducible representation $\bar{\boldsymbol{\rho}}_k : \mathbb{G} \mapsto \mathbb{U}(\bar{\mathcal{H}}_k)$. Which we denote in the main body as $\mathbb{G}^{(k)}$.

E.1 Maps between symmetric vector spaces

We will frequently study and use linear and non-linear maps between symmetric vector spaces. Our focus is on maps that preserve entirely or partially the group structure of the vector spaces. These types of maps can be classified as $\mathbb{G}$-equivariant, $\mathbb{G}$-invariant maps:

Definition E.10 ($\mathbb{G}$-equivariant and $\mathbb{G}$-invariant maps). *Let $\mathcal{X}$ and $\mathcal{Y}$ be two vector spaces endowed with the same symmetry group $\mathbb{G}$, with the respective group actions $\triangleright_{\mathcal{X}}$ and $\triangleright_{\mathcal{Y}}$. A map $f : \mathcal{X} \mapsto \mathcal{Y}$ is said to be $\mathbb{G}$-equivariant if it commutes with the group action, such that:*

$$\begin{aligned} g \triangleright_{\mathcal{Y}} \boldsymbol{y} = g \triangleright_{\mathcal{Y}} f(\boldsymbol{x}) &= f(g \triangleright_{\mathcal{X}} \boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathcal{X}, g \in \mathbb{G}. \\ \boldsymbol{\rho}_{\mathcal{Y}}(g) f(\boldsymbol{x}) &= f(\boldsymbol{\rho}_{\mathcal{X}}(g)\boldsymbol{x}) \end{aligned} \quad \Longleftrightarrow \quad \begin{tikzcd} \mathcal{X} \arrow[r, dash, "\triangleright_{\mathcal{X}}"] \arrow[d, "f"] & \mathcal{X} \arrow[d, "f"] \\ \mathcal{Y} \arrow[r, dash, "\triangleright_{\mathcal{Y}}"] & \mathcal{Y} \end{tikzcd} \tag{20a}$$

A specific case of $\mathbb{G}$-equivariant maps are the $\mathbb{G}$-invariant ones, which are maps that commute with the group action and have trivial output group actions $\triangleright_{\mathcal{Y}}$ such that $\boldsymbol{\rho}_{\mathcal{Y}}(g) = \boldsymbol{I}$ for all $g \in \mathbb{G}$. That is:

$$\begin{aligned} \boldsymbol{y} = g \triangleright_{\mathcal{Y}} f(\boldsymbol{x}) &= f(g \triangleright_{\mathcal{X}} \boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathcal{X}, g \in \mathbb{G}. \\ \boldsymbol{y} = \boldsymbol{\rho}_{\mathcal{Y}}(g) f(\boldsymbol{x}) &= f(\boldsymbol{\rho}_{\mathcal{X}}(g)\boldsymbol{x}) \end{aligned} \quad \Longleftrightarrow \quad \begin{tikzcd} \mathcal{X} \arrow[r, dash, "\triangleright_{\mathcal{X}}"] \arrow[dr, "f"'] & \mathcal{X} \arrow[d, "f"] \\ & \mathcal{Y} \arrow[loop right, "\triangleright_{\mathcal{Y}}"] \end{tikzcd} \tag{20b}$$

E.2 Structure of $\mathbb{G}$-equivariant linear maps

Definition E.11 (Homomorphism, Isomorphism, and $\mathbb{G}$-equivariant linear maps). *Let $\mathcal{X}$ and $\mathcal{Y}$ be two vector spaces endowed with the same symmetry group $\mathbb{G}$, with the respective group actions $\triangleright_{\mathcal{X}}$: $\mathbb{G} \times \mathcal{X} \mapsto \mathcal{X}$ and $\triangleright_{\mathcal{Y}}$: $\mathbb{G} \times \mathcal{Y} \mapsto \mathcal{Y}$. The spaces are said to be $\mathbb{G}$-homomorphic if there exists a linear map $\mathbb{A} : \mathcal{X} \mapsto \mathcal{Y}$ that commutes with the group action, such that $g \triangleright_{\mathcal{Y}} (\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}(g \triangleright_{\mathcal{X}} \boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}$. They are said to be $\mathbb{G}$-isomorphic if the linear map is invertible. Graphically, $\mathcal{X}$ and $\mathcal{Y}$ are $\mathbb{G}$-homomorphic or $\mathbb{G}$-isomorphic if the following diagrams commute:*

$$\underbrace{\begin{tikzcd} \mathcal{X} \arrow[r, dash, "\triangleright_{\mathcal{X}}"] \arrow[d, "\boldsymbol{A}"] & \mathcal{X} \arrow[d, "\boldsymbol{A}"] \\ \mathcal{Y} \arrow[r, dash, "\triangleright_{\mathcal{Y}}"] & \mathcal{Y} \end{tikzcd}}_{\textit{Homomorphism}} \quad \boldsymbol{A} \in \textit{Homo}_{\mathbb{G}}(\mathcal{X}, \mathcal{Y}) \qquad \textit{or} \qquad \underbrace{\begin{tikzcd} \mathcal{X} \arrow[r, dash, "\triangleright_{\mathcal{X}}"] \arrow[d, shift right, "\boldsymbol{A}^{-1}"'] & \mathcal{X} \arrow[d, shift right, "\boldsymbol{A}^{-1}"'] \\ \mathcal{Y} \arrow[r, dash, "\triangleright_{\mathcal{Y}}"] \arrow[u, shift right, "\boldsymbol{A}"'] & \mathcal{Y} \arrow[u, shift right, "\boldsymbol{A}"'] \end{tikzcd}}_{\textit{Isomorphism}} \quad \boldsymbol{A} \in \textit{Iso}_{\mathbb{G}}(\mathcal{X}, \mathcal{Y}). \tag{21}$$

Here, $\textit{Homo}_{\mathbb{G}}(\mathcal{X}, \mathcal{Y})$ denotes the space of $\mathbb{G}$-equivariant linear maps between $\mathcal{X}$ and $\mathcal{Y}$, and $\textit{Iso}_{\mathbb{G}}(\mathcal{X}, \mathcal{Y})$ denotes the space of $\mathbb{G}$-equivariant invertible linear maps between $\mathcal{X}$ and $\mathcal{Y}$.

Lemma E.12 (Schur's Lemma for unitary representations [14, Prop 1.5]). *Consider two Hilbert spaces, $\mathcal{H}$ and $\mathcal{H}'$, endowed with the irreducible unitary representations $\bar{\boldsymbol{\rho}}_{\mathcal{H}} : \mathbb{G} \mapsto \mathbb{U}(\mathcal{H})$ and $\bar{\boldsymbol{\rho}}_{\mathcal{H}'} : \mathbb{G} \mapsto \mathbb{U}(\mathcal{H}')$, respectively. Let $\mathsf{T} : \mathcal{H} \mapsto \mathcal{H}'$ be a linear $\mathbb{G}$-equivariant operator such that $\bar{\boldsymbol{\rho}}_{\mathcal{H}'}\mathsf{T} = \mathsf{T}\bar{\boldsymbol{\rho}}_{\mathcal{H}}$. If the irreducible representations are not equivalent, i.e., $\bar{\boldsymbol{\rho}}_{\mathcal{H}} \nsim \bar{\boldsymbol{\rho}}_{\mathcal{H}'}$, then T is the*

trivial (or zero) map. Conversely, if $\bar{\boldsymbol{\rho}}_{\mathcal{H}} \sim \bar{\boldsymbol{\rho}}_{\mathcal{H}'}$, then T is a constant multiple of an isomorphism (Def. E.11). Denoting I as the identity operator, this can be expressed as:

$$\bar{\boldsymbol{\rho}}_{\mathcal{H}} \nsim \bar{\boldsymbol{\rho}}_{\mathcal{H}'} \iff \quad \mathbf{0}_{\mathcal{H}'} = \mathsf{T}\boldsymbol{h} \mid \forall\, \boldsymbol{h} \in \mathcal{H} \tag{22a}$$

$$\bar{\boldsymbol{\rho}}_{\mathcal{H}} \sim \bar{\boldsymbol{\rho}}_{\mathcal{H}'} \iff \quad \mathsf{T} = \alpha\mathsf{U}, \alpha \in \mathbb{C}, \mathsf{U} \cdot \mathsf{U}^H = \mathsf{I} \tag{22b}$$

$$\bar{\boldsymbol{\rho}}_{\mathcal{H}} = \bar{\boldsymbol{\rho}}_{\mathcal{H}'} \iff \quad \mathsf{T} = \alpha\mathsf{I} \tag{22c}$$

For intiution refeer to the following blog post

F Representation theory of symmetric function spaces

In this section, we study symmetry group actions on infinite-dimensional function spaces and specify the conditions needed to approximate these spaces in finite dimensions. Specifically, given a set $\mathcal{X}$ with a compact symmetry group $\mathbb{G}$ acting via $(\triangleright)$ (Def. E.2), the space of scalar-valued functions on $\mathcal{X}$, $\mathcal{F} = \{f \mid f : \mathcal{X} \mapsto \mathbb{R}\}$, becomes a symmetric function space. The action of a symmetry transformation on a function is defined as:

Definition F.1 (Group action on a function space)**.** *Let $\mathcal{X}$ be a set endowed with the symmetry group $\mathbb{G}$, and let $\mathcal{F}$ be the space of scalar-valued functions on $\mathcal{X}$. The (left) action of $\mathbb{G}$ on a function $f \in \mathcal{F}$ is defined as the composition of f with the inverse of the group element g^{-1}:*

$$\begin{array}{rccc} (\triangleright_{\mathcal{F}}): & \mathbb{G} \times \mathcal{F} & \longrightarrow & \mathcal{F} \\ & (g, f) & \longrightarrow & [g \triangleright_{\mathcal{F}} f](\boldsymbol{x}) := [f \circ g^{-1}](\boldsymbol{x}) = f(g^{-1} \triangleright \boldsymbol{x}), \quad \forall\, \boldsymbol{x} \in \mathcal{X}. \end{array} \tag{23a}$$

In other words, the point-wise evaluation of f on a g^{-1}-transformed set $\mathcal{X}$ is equivalent to the evaluation of the transformed function $g \triangleright_{\mathcal{F}} f \in \mathcal{F}$ on the original set $\mathcal{X}$ (see simple examples in Fig. 11). Any function space that is stable under the group action Eq. (23a) is refereed to as a symmetric function space. Note that this action is compatible with the group composition and identity element $e \in \mathbb{G}$, such that the following properties hold:

$$\textit{Identity:} \quad e \triangleright_{\mathcal{F}} f(\cdot) = f(\cdot), \tag{23b}$$

$$\textit{Associativity:} \quad [(g_2 \circ g_1) \triangleright_{\mathcal{F}} f](\cdot) = [g_2 \triangleright_{\mathcal{F}} [g_1 \triangleright_{\mathcal{F}} f]](\cdot), \qquad \forall\, g_1, g_2 \in \mathbb{G}. \tag{23c}$$

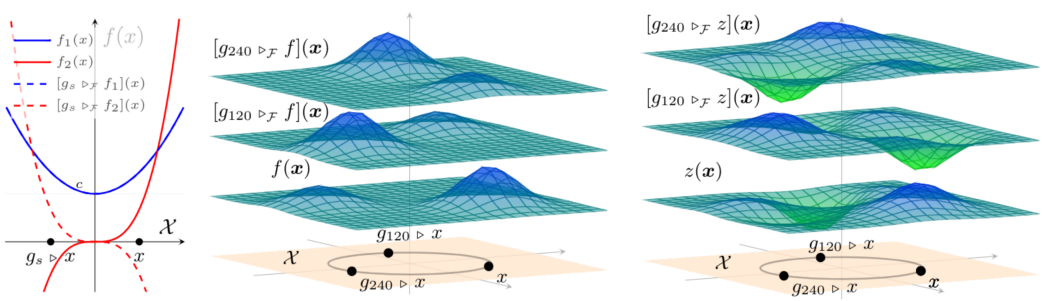

Figure 11: **Left:** Diagram of the group action $\triangleright_{\mathcal{F}}$ on functions $f_1(x) = x^2 + c$ and $f_2(x) = x^3$ defined on the domain $\mathcal{X} := \mathbb{R}$ endowed with the reflectional symmetry group $\mathbb{G} := \mathbb{C}_2 = \{e, g_s\}$, with the reflection action acting on the domain by $g_s \triangleright x = -x$ and on the function space $\mathcal{F} := \{f \mid f : \mathcal{X} \mapsto \mathbb{R}\}$ by $[g \triangleright_{\mathcal{F}} f](\boldsymbol{x}) = f(g \triangleright_{\mathcal{X}} \boldsymbol{x}) = f(-\boldsymbol{x})$. Hence we have that f_1 is a $\mathbb{G}$-invariant function, $g_s \triangleright_{\mathcal{F}} f_1(x) = f_1(x)$ and f_2 a $\mathbb{G}$-equivariant function $g_s \triangleright_{\mathcal{F}} f_2(x) = -x^3$. **Center:** Diagram representing the action $\triangleright_{\mathcal{F}}$ on the (arbitrarily chosen) function $f(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{c}_1, 2) + \mathcal{N}(\boldsymbol{x}; \boldsymbol{c}_2, 1)$ defined over the symmetric domain $\mathcal{X} = \mathbb{R}^2$ with the cyclic symmetry group $\mathbb{G} = \mathbb{C}_3 = \{e, g_{120}, g_{240}\}$ and group action $g \triangleright \boldsymbol{x} = \boldsymbol{\rho}_{\mathcal{X}}(g)\boldsymbol{x} = \boldsymbol{R}_g\boldsymbol{x}$, where $\boldsymbol{R}_g$ is a rotation matrix in 2D. Here, $g_{120} \triangleright_{\mathcal{F}} f$ is equivalent to evaluating f on a domain rotated by $-120°$. The same holds for $g_{240} \triangleright_{\mathcal{F}} f$. Note that the z-offsets are added for visualization purposes. **Right:** Diagram representing the action $\triangleright_{\mathcal{F}}$ on the function $z \in \widehat{\mathcal{F}}$, defined to be a member of the finite-dimensional symmetric function space $\widehat{\mathcal{F}} := \text{span}(\mathbb{I}_{\widehat{\mathcal{F}}})$, constructed from a basis set composed of the group orbit of the (arbitrarily chosen) function $f \in \mathcal{F}$, that is $\mathbb{I}_{\widehat{\mathcal{F}}} := \mathbb{G}f = \{f, g_{120} \triangleright_{\mathcal{F}} f, g_{240} \triangleright_{\mathcal{F}} f\}$. This function space is $\mathbb{G}$-stable by construction, since $\mathbb{G}\mathbb{I}_{\widehat{\mathcal{F}}} = \mathbb{I}_{\widehat{\mathcal{F}}}$. Note that the z-offsets are added for visualization purposes.

Remark F.2. From an algebraic perspective, the inversion g^{-1} (*contragredient representation*) emerges to ensure that the associativity property of the group action (Eq. (23c)) holds:

$$
\begin{aligned}
&[(g_2 \circ g_1) \triangleright_{\mathcal{F}} f](\boldsymbol{x}) = [g_2 \triangleright_{\mathcal{F}} [g_1 \triangleright_{\mathcal{F}} f]](\boldsymbol{x}), \qquad \forall\, \boldsymbol{x} \in \mathcal{X} \\
&f((g_2 \circ g_1)^{-1} \triangleright \boldsymbol{x}) = [g_1 \triangleright_{\mathcal{F}} f](g_2^{-1} \triangleright \boldsymbol{x}) = f(g_1^{-1} \triangleright (g_2^{-1} \triangleright \boldsymbol{x})) \\
&f((g_2 \circ g_1)^{-1} \triangleright \boldsymbol{x}) = f((g_1 \circ g_2)^{-1} \triangleright \boldsymbol{x}).
\end{aligned}
$$

In the context of this work, we will study the scenario where the function space $\mathcal{F}$ is a separable Hilbert space and the group action of $\mathbb{G}$ on $\mathcal{F}$ is unitary, i.e., it preserves the inner product of the function space. This setup is crucial to enable us to approximate $\mathcal{F}$ and the group action on $\mathcal{F}$ in finite dimensions.

F.1 Unitary group representation on function spaces

Assume our symmetric set $\mathcal{X}$ is endowed with a measure space structure $(\mathcal{X}, \Sigma_{\mathcal{X}}, P_{\mathbf{x}})$, where $P_{\mathbf{x}} : \Sigma_{\mathcal{X}} \mapsto \mathbb{R}$ is the space measure. Then, consider a function space with a separable Hilbert space structure $\mathcal{F} := \mathcal{L}^2_{P_{\mathbf{x}}}\mathcal{X}, \mathbb{R}$, and inner product $\langle f_1, f_2 \rangle_{P_{\mathbf{x}}} = \int_{\mathcal{X}} f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) P_{\mathbf{x}}(d\boldsymbol{x})$ for all $f_1, f_2 \in \mathcal{F}$. Then, the action $\triangleright_{\mathcal{F}}$ of the group $\mathbb{G}$ on the function space $\mathcal{F}$ is termed unitary if it preserves the inner product of the function space:

$$
\begin{aligned}
\langle f_1, f_2 \rangle_{P_{\mathbf{x}}} &= \langle g \triangleright_{\mathcal{F}} f_1, g \triangleright_{\mathcal{F}} f_2 \rangle_{P_{\mathbf{x}}} \qquad \forall\, f_1, f_2 \in \mathcal{F}, g \in \mathbb{G} \\
\int_{\mathcal{X}} f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) P_{\mathbf{x}}(d\boldsymbol{x}) &= \int_{\mathcal{X}} (g \triangleright_{\mathcal{F}} f_1)(\boldsymbol{x})(g \triangleright_{\mathcal{F}} f_2)(\boldsymbol{x}) P_{\mathbf{x}}(d\boldsymbol{x}) \\
&= \int_{\mathcal{X}} f_1(g^{-1} \triangleright \boldsymbol{x}) f_2(g^{-1} \triangleright \boldsymbol{x}) P_{\mathbf{x}}(d\boldsymbol{x}) \\
&= \int_{g \triangleright \mathcal{X} = \mathcal{X}} f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) P_{\mathbf{x}}(g \triangleright d\boldsymbol{x}).
\end{aligned} \tag{24}
$$

That is, the group action is unitary if $P_{\mathbf{x}}$ is a $\mathbb{G}$-invariant measure $P_{\mathbf{x}}(g \triangleright d\boldsymbol{x}) = P_{\mathbf{x}}(d\boldsymbol{x}),\ \forall\, g \in \mathbb{G}, d\boldsymbol{x} \subseteq \mathcal{X}$. Note that an $\mathbb{G}$-invariant measure (and inner product) exists whenever $\mathbb{G}$ is finite, because for any measure $\eta : \Sigma_{\Omega} \mapsto \mathbb{R}$, we can use the group-average trick to obtain one, given by $P_{\mathbf{x}}(\mathbb{X}) = \Sigma_{g \in \mathbb{G}} \eta(g \triangleright \mathbb{X})$.[2]

The importance of the Hilbert space structure is that it enables the definition of a unitary group representation. Unitary representations have a well-studied modular structure that allows their decomposition (Thm. E.8) into $\mathbb{G}$-stable subspaces (Def. E.5), which is crucial for approximating symmetric function spaces using a finite set of basis elements. Let $\mathbb{I}_{\mathcal{F}} = \{\phi_i \mid \phi_i \in \mathcal{L}^2_{\mathbf{x}}\}_{i \in \mathbb{N}}$ be an orthogonal basis for the function space $\mathcal{F} = \text{span}(\mathbb{I}_{\mathcal{F}})$, so that any function $f \in \mathcal{F}$ can be represented by its basis expansion coefficients $\boldsymbol{\alpha} = [\langle \phi_i \rangle_{P_{\mathbf{x}}} f]_{i \in \mathbb{N}}$, since $f_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \sum_{i \in \mathbb{N}} \langle \phi_i, f \rangle_{P_{\mathbf{x}}} \phi_i(\boldsymbol{x})$. In this basis, the group action of $\mathbb{G}$ on $\mathcal{F}$ defines a unitary group representation mapping group elements to unitary linear integral operators on $\mathcal{F}$, which can be expressed in matrix form.

Definition F.3 (Unitary group representation on a function space)**.** *Let $\mathcal{F} = \mathcal{L}^2_{P_{\mathbf{x}}}\mathcal{X}, \mathbb{R}$ be a separable Hilbert space of scalar-valued functions on a set $\mathcal{X}$ endowed with the symmetry group $\mathbb{G}$. Let $\mathbb{I}_{\mathcal{F}}$ be an orthogonal basis set spanning $\mathcal{F}$. Then, the group action of $\mathbb{G}$ on $\mathcal{F}$ (Def. F.1) defines a unitary group representation mapping group elements to unitary linear integral operators on $\mathcal{F}$:*

$$
\boldsymbol{\rho}_{\mathcal{F}} : \begin{array}{ccc} \mathbb{G} & \longrightarrow & \mathbb{U}(\mathcal{F}) \\ g & \longrightarrow & \boldsymbol{\rho}_{\mathcal{F}}(g) \end{array}, \qquad \text{s.t.} \quad \boldsymbol{\rho}_{\mathcal{F}}(g)^* = \boldsymbol{\rho}_{\mathcal{F}}(g^{-1}). \tag{25}
$$

Each unitary operator $\boldsymbol{\rho}_{\mathcal{F}}(g) : \mathcal{F} \mapsto \mathcal{F}$ admits an infinite-dimensional matrix representation with entries $[\boldsymbol{\rho}_{\mathcal{F}}(g)]_{i,j} := \langle \hat{f}_i, g \triangleright_{\mathcal{F}} \hat{f}_j \rangle_{P_{\mathbf{x}}}$, which characterize how the group action transforms the chosen basis functions. Consequently, once the group representation for a chosen basis set is defined, the group action on a function $f_{\boldsymbol{\alpha}} \in \mathcal{F}$ can be expressed as an (infinite-dimensional) matrix transformation of its basis expansion coefficients $\boldsymbol{\alpha}$, given by:

$$
[g \triangleright_{\mathcal{F}} f_{\boldsymbol{\alpha}}](\cdot) := \sum_{i \in \mathbb{N}} \langle \hat{f}_i, g \triangleright_{\mathcal{F}} f_{\boldsymbol{\alpha}} \rangle_{P_{\mathbf{x}}} \hat{f}_i(\cdot) = \sum_{i \in \mathbb{N}} \left(\sum_{j \in \mathbb{N}} \langle \hat{f}_i, g \triangleright_{\mathcal{F}} \hat{f}_j \rangle_{P_{\mathbf{x}}} \underbrace{\langle \hat{f}_j, f \rangle_{P_{\mathbf{x}}}}_{\alpha_j} \right) \hat{f}_i(\cdot). \tag{26}
$$

[2] Such a $\mathbb{G}$-invariant measure exists for any (finite or continuous) compact group. See discussion.