A Unified Framework for Discovering Discrete Symmetries

Pavan Karjol, Rohan Kashyap, Aditya Gopalan, Prathosh A P

Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru, Karnataka

Problem statement

We consider the problem of learning an H-invariant function $f : X \to \mathbb{R}$, where $X = [0,1]^n \setminus E$, $E = \left\{[x_1, x_2 \ldots, x_n]^T \in [0,1]^n : x_i = x_j \text{ for some } i, j \in [n] \text{ with } i \neq j\right\}$ and H is the unknown subgroup of S_n, i.e., $H \in \bigcup_{\mathcal{I} \subseteq [n], |\mathcal{I}|>1} \{\mathbb{Z}_\mathcal{I}, D_\mathcal{I}, S_\mathcal{I}\}$. In particular, we develop a unified framework to automatically discover the data symmetry across a broad range of subgroups by leveraging the multi-armed bandits (MAB) and gradient descent method (SGD).

Theorem: Learning $\mathbb{Z}_k$-invariant function

Let $\psi : [0,1]^k \to \mathbb{R}$ be $\mathbb{Z}_k$-invariant. There exists an S_k-invariant function $\phi : [0,1]^{k\times 2} \to \mathbb{R}$ and $\rho : [0,1]^k \to [0,1]^{k\times 2}$, such that

$$\psi = \phi \circ \rho,$$

where ρ is defined as:

$$\left[x_1 \ldots x_k\right] \mapsto \left[(x_1, x_2), (x_2, x_3), \ldots \ldots, (x_k, x_1)\right]$$

Regularity: ϕ is continuous $\left(C^0\right)$ or smooth (C^∞) whenever ψ is C^0 or C^∞ respectively.

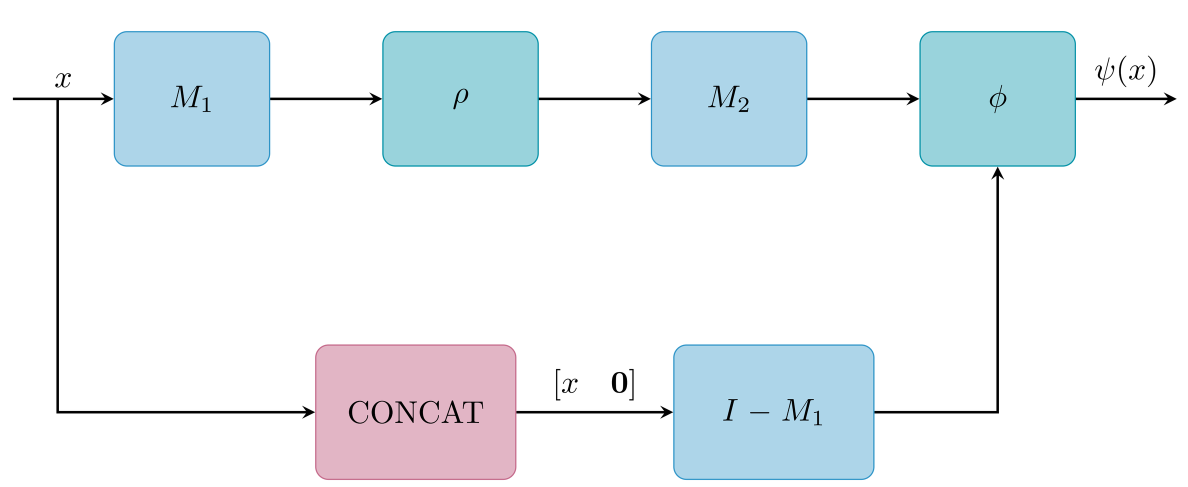
Note: Similar results hold for ψ as a D_{2k} or S_k-invariant function, with modifications to the definition of the function ρ as described in the table below.

Matrix-Valued Function

	S_k	$\mathbb{Z}_k$	D_{2k}
$\rho(x)$	$\begin{bmatrix} & \vdots & \\ x_i & & x_i \\ & \vdots & \end{bmatrix}_{i\in[k]}$	$\begin{bmatrix} & \vdots & \\ x_i & & x_{\tau(i)} \\ & \vdots & \end{bmatrix}_{i\in[k]}$	$\begin{bmatrix} & \vdots & \\ x_i & & x_{\tau(i)} \\ x_{\tau(i)} & & x_i \\ & \vdots & \end{bmatrix}_{i\in[k]}$

Table 1. Subgroups of S_n and corresponding definitions of the matrix-valued function ρ, where τ is cyclic right shift by 1 element.

Method



1. Our framework: G-invariant network with linear (M_1, M_2), matrix-valued ρ, and non-linear ϕ functions.
2. Explicit characterization of (M_1, M_2) matrices for various subgroups.
3. Efficient training algorithm: MAB with SGD leveraging specific structures.
4. Optimal (M_1, M_2) search: linear parametric Thompson Sampling (LinTS).
5. ϕ function approximation: neural network learned through SGD.

Theorem: Unified Framework

Let $\mathcal{B}$ denote the class of all functions from $X \to \mathbb{R}$ of the form:

$$x \mapsto \phi\left(\begin{bmatrix} (M_2 \circ \rho \circ M_1)(x) \\ (I - M_1)([x \quad \mathbf{0}]) \end{bmatrix}\right)$$

where,

- M_1 and M_2 are matrices of size $n \times n$ and $n^2 \times n^2$ respectively.
- $\phi : [0,1]^{n(n+1)\times 2} \to \mathbb{R}$ is an S_{n^2}-invariant function where the invariance pertains to the initial n^2 rows out of a total of $n(n+1)$.
- $\rho : X \to [0,1]^{n^2\times 2}$ is a matrix-valued function given as:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mapsto \begin{bmatrix} & \vdots & \\ x_i & & x_j \\ & \vdots & \end{bmatrix}_{i,j\in[n]}$$

Let $\mathcal{I} = \{i_1, i_2, \ldots i_k\} \subseteq [n]$ $(k > 1)$ and τ be the permutation (cyclic shift). Then, the following hold:

a) Any $S_\mathcal{I}$-invariant function belongs to $\mathcal{B}$. Moreover, the matrices M_1 and M_2 in its decomposition have the forms:

$$M_1[u, v] = \begin{cases} 1, & \text{if } u \in [k] \text{ and } v = i_u \\ 0, & \text{otherwise.} \end{cases}$$

$$M_2[u, v] = \begin{cases} 1, & \text{if } u \in [k^2], \quad u = v \text{ and} \\ & (\rho \circ M_1)(x)[v] = (x_i, x_i) \\ & \text{for some } i \in \mathcal{I} \\ 0, & \text{otherwise.} \end{cases}$$

b) Any $\mathbb{Z}_\mathcal{I}$-invariant function belongs to $\mathcal{B}$. Moreover, M_1 is of the form as given above and M_2 is as follows:

$$M_2[u, v] = \begin{cases} 1, & \text{if } u \in [k] \text{ and} \\ & (\rho \circ M_1)(x)[v] = (x_{i_u}, x_{\tau(i_u)}) \\ 0, & \text{otherwise.} \end{cases}$$

c) Any $D_\mathcal{I}$-invariant function belongs to $\mathcal{B}$. Moreover, M_1 is of the form as given in part (a) and (b) and M_2 is as follows:

$$M_2[u, v] = \begin{cases} 1, & \text{if } u \in [k] \text{ and} \\ & (\rho \circ M_1)(x)[v] = (x_{i_u}, x_{\tau(i_u)}) \\ 1, & \text{else if } u \in [2k] \setminus [k] \text{ and} \\ & (\rho \circ M_1)(x)[v] = (x_{\tau(i_{u-k})}, x_{i_{u-k}}) \\ 0, & \text{otherwise.} \end{cases}$$

Theorem: Product Groups

Let $[n] = \bigcup_{j=1}^{L} \mathcal{I}_j$ be a partition of $[n]$, $G_i \in \{S_{\mathcal{I}_j}, D_{\mathcal{I}_j}, \mathbb{Z}_{\mathcal{I}_j}\}, \forall j \in [L]$ and $G = G_1 \times G_2 \times \cdots G_L$ such that no two groups G_i, G_j are isomorphic and only one of the component groups is of the type $S_\mathcal{I}$. Let ψ be a G-invariant function, then there exists an S_I-invariant function ϕ and a specific matrix-valued function ρ, such that,

$$\psi = \phi \circ \rho.$$

Theorem: Error probability bound for LinTS

Let the set of arms $\mathcal{A} \subset \mathbb{R}^d$ be finite. Suppose that the reward from playing an arm $a \in \mathcal{A}$ at any iteration, conditioned on the past, is sub-Gaussian with mean $a^\top \mu^\star$. After T iterations, let the guessed best arm A_T be drawn from the empirical distribution of all arms played in the T rounds, i.e., $\mathbb{P}[A_T = a] = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{a^{(t)} = a\}$ where $a^{(t)}$ denotes the arm played in iteration t. Then, $\mathbb{P}[A_T \neq a^\star] \leq \frac{c \log(T)}{T}$, where $c \equiv c(\mathcal{A}, \mu^\star, \nu)$ is a quantity that depends on the problem instance $(\mathcal{A}, \mu^\star)$ and algorithm parameter (ν).

Interpretability

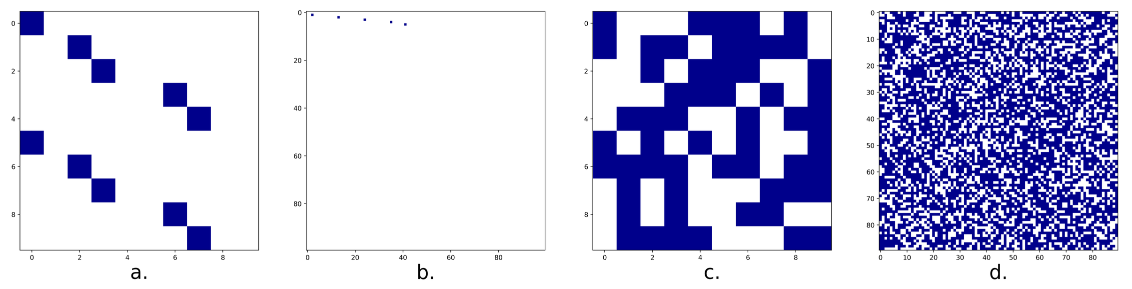


Figure 1. Visualization of reference (bandit) matrices M_1 (a) and M_2 (b), along with those obtained through training our method entirely using SGD for polynomial regression of $\mathbb{Z}_\mathcal{I}$-invariant functions. $n = 10$ and $\mathcal{I} = \{0, 2, 3, 6, 7\}$ (c, d).

Results

Symmetric Polynomial Regression: We evaluate the performance of our method on symmetric polynomial regression tasks for subgroups $\mathbb{Z}_\mathcal{I}, D_\mathcal{I}, S_\mathcal{I}$. We experiments with different values of $k = |\mathcal{I}|\,(k < n, n = 10)$ and randomly sampled index sets $\mathcal{I}$. These accuracies indicate the successful identification of the underlying subgroup within the top 3 bandit arms.

Table 2. Model Accuracy [%]

Task	G	Accuracy
Polynomial Regression	$\mathbb{Z}_\mathcal{I}$	100
Polynomial Regression	$D_\mathcal{I}$	100
Image-Digit Sum	$S_\mathcal{I}$	100
Convex Area	$D_\mathcal{I}$	100
$S_\mathcal{I}$ (4)	$S_\mathcal{I}$	100

Table 3. MAE $[\times 10^{-2}]$ Regression task

G	$\mathbb{Z}_\mathcal{I}(5)$	$\mathbb{Z}_\mathcal{I}(7)$	$D_\mathcal{I}(5)$	$D_\mathcal{I}(7)$
$\mathbb{Z}_\mathcal{I}$	**4.2**	**6.1**	8.2	15.2
$D_\mathcal{I}$	4.7	7.9	**6.3**	**10.1**
$S_\mathcal{I}$	11.7	18.5	21.3	34.3
$M + H$-INV	12.3	-	23.2	-
SGD	14.4	17.7	26.5	34.4

Image-Digit Sum: This task aims to sum the digits of k images from a set of n input images. The positions of the images are fixed but unknown to the model, as is the value of k. Thus, it corresponds to learning an $S_\mathcal{I}$-invariant function.

Limitations

Our framework is a proof of concept. Further extensive experiments are needed, and it's only applicable for discrete groups. The complete collection of all possible permutation subgroups discoverable through our method is unknown.

References

1 Pavan Karjol, Rohan Kashyap, and AP Prathosh. Neural discovery of permutation subgroups.
2 Piotr Kicki, Mete Ozay, and Piotr Skrzypczyński. A computationally efficient neural network invariant to the action of symmetry subgroups.