# PolySHAP: Extending KernelSHAP with Interaction-Informed Polynomial Regression

**Fabian** Fumagalli[1], R. **Teal** Witter[2], Christopher Musco[3]

## Motivation

- **Computational Bottleneck:** Exact Shapley values are **exponentially expensive ($2^d$)**.
- **Interaction Gap:** KernelSHAP assumes linearity, missing **complex, non-linear feature interactions**.
- **Theoretical Mystery:** "Paired sampling" is a ubiquitous heuristic lacking **rigorous theoretical justification**.

## Contribution

- **PolySHAP:** Extends KernelSHAP to capture **higher-order interactions** using polynomials.
- **Theoretical Bridge:** Proves equivalence between **paired KernelSHAP** and **second-order PolySHAP** (Theorem 5.1).
- **Impact:** Theoretical foundation for the **paired sampling** heuristic for all games.

## Theoretical foundation

**KernelSHAP:** Shapley values via least squares

$$\boldsymbol{\phi}^{\mathrm{SV}}[\nu] := \operatorname*{arg\,min}_{\boldsymbol{\phi}\in\mathbb{R}^d:\langle\boldsymbol{\phi},\mathbf{1}\rangle=\nu(D)} \sum_{S\subseteq D} \mu(S)\left(\nu(S) - \sum_{i=1}^{d}\phi_i \mathbf{1}[i\in S]\right)^2$$

**PolySHAP:** Linear to polynomial regression

The **PolySHAP representation** $\boldsymbol{\phi}^{\mathcal{I}}\in\mathbb{R}^{d'}$ with $d' = d + |\mathcal{I}|$ is given by

$$\boldsymbol{\phi}^{\mathcal{I}}[\nu] := \operatorname*{arg\,min}_{\boldsymbol{\phi}\in\mathbb{R}^{d'}:\langle\boldsymbol{\phi},\mathbf{1}\rangle=\nu(D)} \sum_{S\subseteq D} \mu(S)\left(\nu(S) - \sum_{T\in D\cup\mathcal{I}} \phi_T \prod_{j\in T}\mathbf{1}[j\in S]\right)^2.$$

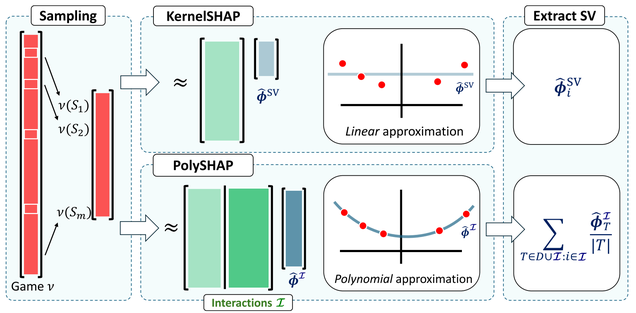
Here, we abuse notation with $\phi_i := \phi_{\{i\}}$ and $\mathbf{1}[j\in i] := \mathbf{1}[j=i]$ for $i, j \in D$.

Recovering Shapley values from interactions

The Shapley values of $\nu$ are recovered from the PolySHAP representation as

$$\phi_i^{\mathrm{SV}}[\nu] = \phi_i^{\mathcal{I}} + \sum_{S\in\mathcal{I}:i\in S} \frac{\phi_S^{\mathcal{I}}}{|S|} \quad \text{for } i\in D$$

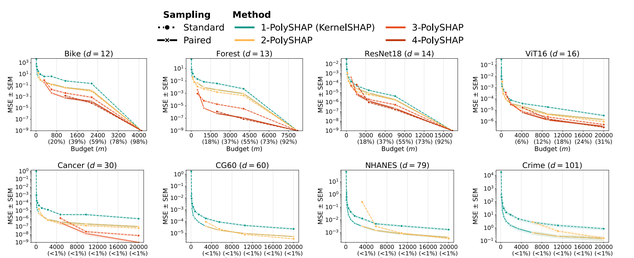
## Algorithm



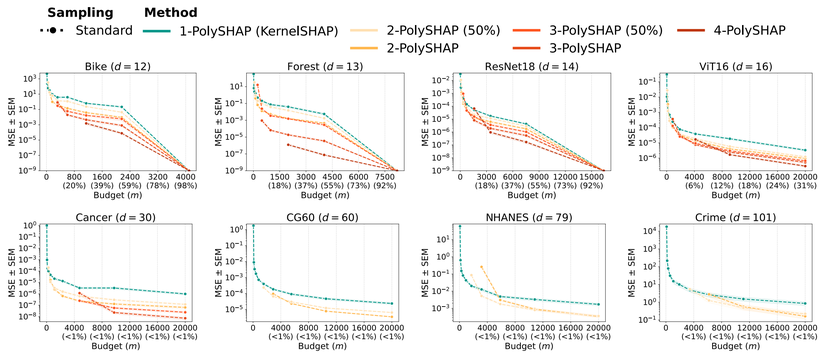

## Paired KernelSHAP is 2-PolySHAP

**Theorem** (Paired KernelSHAP is Paired 2-PolySHAP). *Suppose that subsets are sampled in pairs i.e., if $S$ is sampled then so is its complement $D\setminus S$, and, the matrix $\tilde{\mathbf{X}}$ has full column rank for interaction frontier $D$ and $\mathcal{I}_{\leq 2}$. Then*

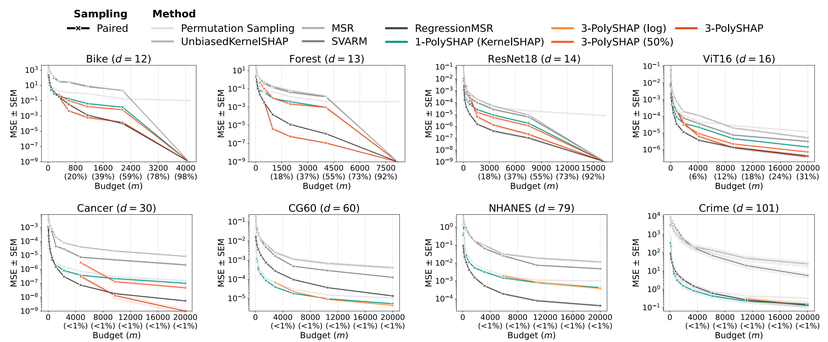
$$\hat{\phi}^{SV} = \textsc{PolySHAPtoSV}(\hat{\phi}^{\mathcal{I}_{\leq 2}})$$




## Interactions improve estimation quality




## PolySHAP outperforms KernelSHAP




paper

code

shapiq

1) Bielefeld University

2) Claremont McKenna College

3) New York University