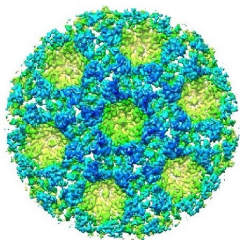


Utilising Symmetry for Dimension Reduction

Louis Christie¹ and John Aston¹

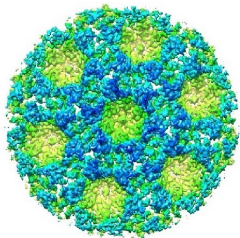
¹Statistical Laboratory, University of Cambridge, UK

Symmetries exist everywhere in statistics

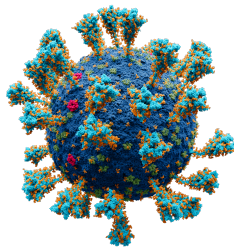


HIV Viral Capsid [JT17]

Symmetries exist everywhere in statistics

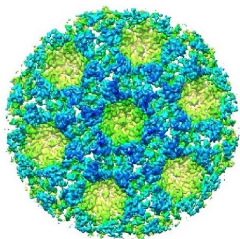


HIV Viral Capsid [JT17]

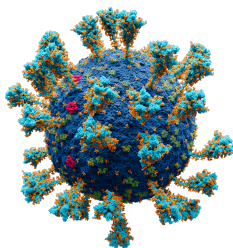


SARS-CoV-2 Capsid [S21]

Symmetries exist everywhere in statistics



HIV Viral Capsid [JT17]



SARS-CoV-2 Capsid [S21]

Main Questions

Suppose that we wish to estimate a function $f : \mathcal{X} \rightarrow \mathbb{R}$.

Can we quantify the benefits of knowing the symmetries of f ?

Can we estimate the symmetries if they are unknown?

What is a symmetry?

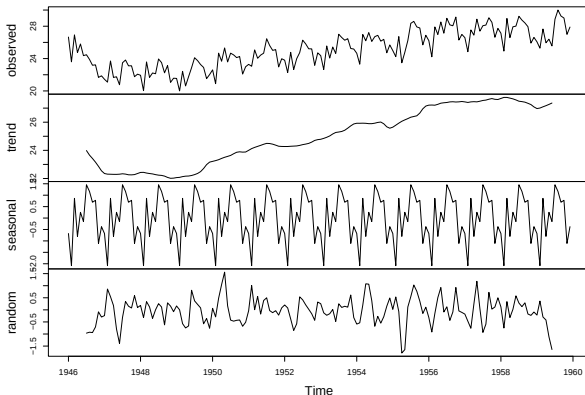
A **symmetry** is a set G of bijections of a domain $\mathcal{X}$ to itself that is closed under composition and inversion.

A function $f : \mathcal{X} \rightarrow \mathbb{R}$ has the symmetry G , or is **G -invariant**, if $f(\phi(x)) = f(x)$ for all $x \in \mathcal{X}$ and all $\phi \in G$.

Seasonal time series exhibit translation symmetries

Births in New York City per Month

Decomposition of additive time series



Sparse functions have translation symmetries too



$$f(x_1, x_2) = \sin(x_2)$$

Sparse functions have translation symmetries too



$$f(x_1, x_2) = \sin(x_2)$$

is invariant to $G = \{x \mapsto x + ae_1 : a \in \mathbb{R}\}$

Regression functions can have other symmetries



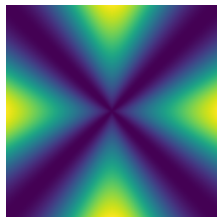
$$f_1(x) = \sin(x_2)$$

Translational
Symmetry



$$f_2(x) = \sin(\|x\|_2)$$

Rotational
Symmetry



$$f_3(x) = \sin(\|x\|_2) \times \cos(2\arctan(x_2/x_1))^2$$

Discrete
Symmetry

Examples of symmetries

- ▶ Multi-index models in nonparametric regression rely on translation symmetries (as with sparse models);
- ▶ Image classifications can be invariant to rotations of the image;
- ▶ Kendall's Shape Spaces rely on invariance to scaling, translations, and/or rotations of point clouds;
- ▶ Graph Neural Networks rely on permutation symmetries on the nodes and edges;
- ▶ Functional data registration relies on symmetries on the domain of the functions.

Symmetry in Nonparametric Regression

Suppose that we collect iid data (X_i, Y_i) from the model:

$$Y_i = f(X_i) + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

where $f : \mathcal{X} \rightarrow \mathbb{R}$, can be any bounded Lipschitz function with maximum slope and value L , and $X_i \stackrel{iid}{\sim} \mu_X$ with support $\mathcal{X}$ of dimension d .

Symmetry in Nonparametric Regression

Suppose that we collect iid data (X_i, Y_i) from the model:

$$Y_i = f(X_i) + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

where $f: \mathcal{X} \rightarrow \mathbb{R}$, can be any bounded Lipschitz function with maximum slope and value L , and $X_i \stackrel{iid}{\sim} \mu_X$ with support $\mathcal{X}$ of dimension d .

Then we know that the minimax rate of convergence is given by:

$$\inf_{f_n} \sup_{f \in \mathcal{F}_{Lips}(L)} \mathbb{E}(\|f - f_n\|_2^2) > c_1 n^{-\frac{2}{2+d}}$$

Symmetry in Nonparametric Regression

Suppose that we collect iid data (X_i, Y_i) from the model:

$$Y_i = f(X_i) + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

where $f : \mathcal{X} \rightarrow \mathbb{R}$, can be any bounded Lipschitz function with maximum slope and value L , and $X_i \stackrel{iid}{\sim} \mu_X$ with support $\mathcal{X}$ of dimension d .

Then we know that the minimax rate of convergence is given by:

$$\inf_{f_n} \sup_{f \in \mathcal{F}_{Lips}(L)} \mathbb{E}(\|f - f_n\|_2^2) > c_1 n^{-\frac{2}{2+d}}$$

Classical Questions

Can we estimate f faster if depends on fewer coordinates of X ?

Can we estimate the important coordinates if they are unknown?

Symmetry in Nonparametric Regression

Suppose that we collect iid data (X_i, Y_i) from the model:

$$Y_i = f(X_i) + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

where $f : \mathcal{X} \rightarrow \mathbb{R}$, can be any bounded Lipschitz function with maximum slope and value L , and $X_i \stackrel{iid}{\sim} \mu_X$ with support $\mathcal{X}$ of dimension d .

Then we know that the minimax rate of convergence is given by:

$$\inf_{f_n} \sup_{f \in \mathcal{F}_{Lips}(L)} \mathbb{E}(\|f - f_n\|_2^2) > c_1 n^{-\frac{2}{2+d}}$$

Main Questions

Can we quantify the benefits of knowing the symmetries of f ?

Can we estimate the symmetries if they are unknown?

Can we construct estimators that use symmetry?

Given a **base estimator** $\hat{f}(x, \mathcal{D})$, we can construct an symmetrised estimator $\hat{f}_G$ for finite G in different ways:

Can we construct estimators that use symmetry?

Given a **base estimator** $\hat{f}(x, \mathcal{D})$, we can construct an symmetrised estimator $\hat{f}_G$ for finite G in different ways:

- Symmetrise the output: $\hat{f}_G(x) = |G|^{-1} \sum_{\phi \in G} \hat{f}(\phi(x), \mathcal{D})$;

Can we construct estimators that use symmetry?

Given a **base estimator** $\hat{f}(x, \mathcal{D})$, we can construct an symmetrised estimator $\hat{f}_G$ for finite G in different ways:

- ▶ Symmetrise the output: $\hat{f}_G(x) = |G|^{-1} \sum_{\phi \in G} \hat{f}(\phi(x), \mathcal{D})$;
- ▶ Symmetrise the data: $\hat{f}_G(x) = \hat{f}(x, \{(\phi_j(X_i), Y_i)\}_{i,j=1}^{n,m})$ for some collection of $\{\phi_j\}_{j=1}^m \in G$;

Can we construct estimators that use symmetry?

Given a **base estimator** $\hat{f}(x, \mathcal{D})$, we can construct an symmetrised estimator $\hat{f}_G$ for finite G in different ways:

- ▶ Symmetrise the output: $\hat{f}_G(x) = |G|^{-1} \sum_{\phi \in G} \hat{f}(\phi(x), \mathcal{D})$;
- ▶ Symmetrise the data: $\hat{f}_G(x) = \hat{f}(x, \{(\phi_j(X_i), Y_i)\}_{i,j=1}^{n,m})$ for some collection of $\{\phi_j\}_{j=1}^m \in G$;
- ▶ Symmetrise the kernel: Replace $K(x, y)$ with $K_G(x, y) = |G|^{-1} \sum_{\phi \in G} K(\phi(x), y)$; or

Can we construct estimators that use symmetry?

Given a **base estimator** $\hat{f}(x, \mathcal{D})$, we can construct an symmetrised estimator $\hat{f}_G$ for finite G in different ways:

- ▶ Symmetrise the output: $\hat{f}_G(x) = |G|^{-1} \sum_{\phi \in G} \hat{f}(\phi(x), \mathcal{D})$;
- ▶ Symmetrise the data: $\hat{f}_G(x) = \hat{f}(x, \{(\phi_j(X_i), Y_i)\}_{i,j=1}^{n,m})$ for some collection of $\{\phi_j\}_{j=1}^m \in G$;
- ▶ Symmetrise the kernel: Replace $K(x, y)$ with $K_G(x, y) = |G|^{-1} \sum_{\phi \in G} K(\phi(x), y)$; or
- ▶ Invariant / Equivariant models for CNNs, GNNs etc. [Bronstein et al. 2021].

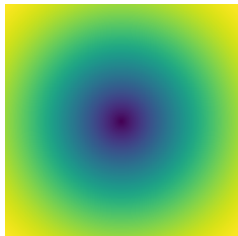
Can we construct estimators that use symmetry?

Given a **base estimator** $\hat{f}(x, \mathcal{D})$, we can construct an symmetrised estimator $\hat{f}_G$ for finite G in different ways:

- ▶ **Symmetrise the output:** $\hat{f}_G(x) = |G|^{-1} \sum_{\phi \in G} \hat{f}(\phi(x), \mathcal{D})$;
- ▶ Symmetrise the data: $\hat{f}_G(x) = \hat{f}(x, \{(\phi_j(X_i), Y_i)\}_{i,j=1}^{n,m})$ for some collection of $\{\phi_j\}_{j=1}^m \in G$;
- ▶ Symmetrise the kernel: Replace $K(x, y)$ with $K_G(x, y) = |G|^{-1} \sum_{\phi \in G} K(\phi(x), y)$; or
- ▶ Invariant / Equivariant models for CNNs, GNNs etc. [Bronstein et al. 2021].

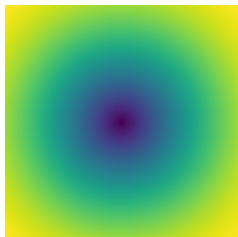
How does this improve the estimator?

Regression
Function
 $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

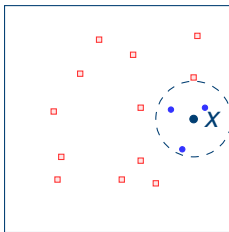


How does this improve the estimator?

Regression
Function
 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$

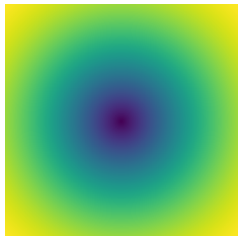


Local Constant
Estimation (LCE)
 $f_n(x)$

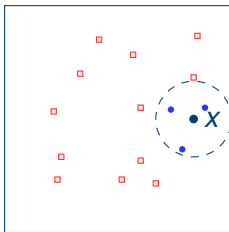


How does this improve the estimator?

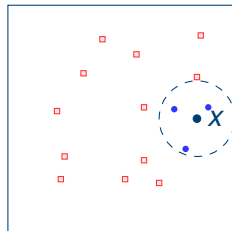
Regression
Function
 $f : \mathbb{R}^2 \rightarrow \mathbb{R}$



Local Constant
Estimation (LCE)
 $f_n(x)$

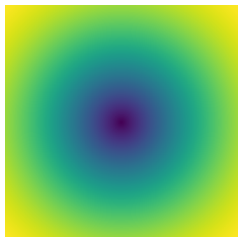


Symmetrised LCE
 $\frac{1}{|G|} \sum_{\phi \in G} \hat{f}(\phi(x))$

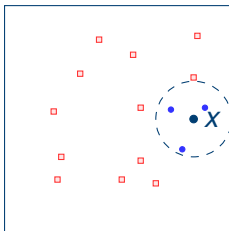


How does this improve the estimator?

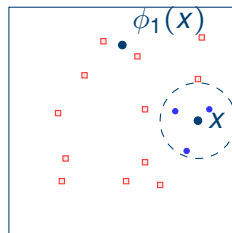
Regression
Function
 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$



Local Constant
Estimation (LCE)
 $f_n(x)$

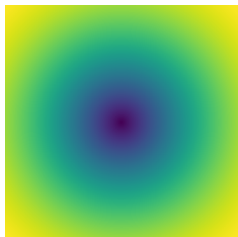


Symmetrised LCE
 $\frac{1}{|G|} \sum_{\phi \in G} \hat{f}(\phi(x))$

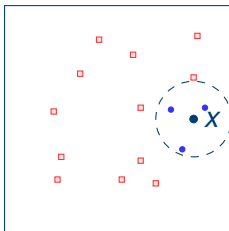


How does this improve the estimator?

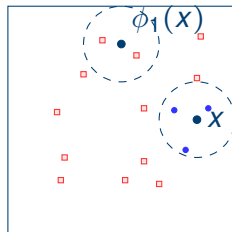
Regression
Function
 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$



Local Constant
Estimation (LCE)
 $f_n(x)$

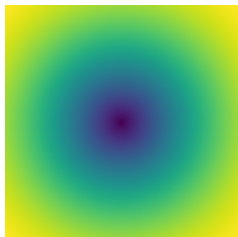


Symmetrised LCE
 $\frac{1}{|G|} \sum_{\phi \in G} \hat{f}(\phi(x))$

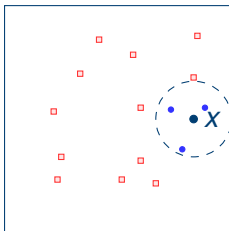


How does this improve the estimator?

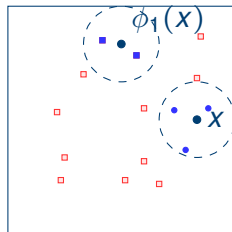
Regression
Function
 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$



Local Constant
Estimation (LCE)
 $f_n(x)$

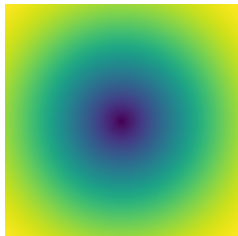


Symmetrised LCE
 $\frac{1}{|G|} \sum_{\phi \in G} \hat{f}(\phi(x))$

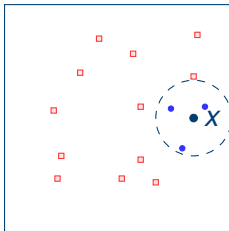


How does this improve the estimator?

Regression
Function
 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$

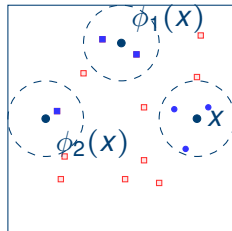


Local Constant
Estimation (LCE)
 $f_n(x)$



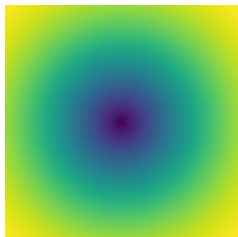
Symmetrised LCE

$$\frac{1}{|G|} \sum_{\phi \in G} \hat{f}(\phi(x))$$

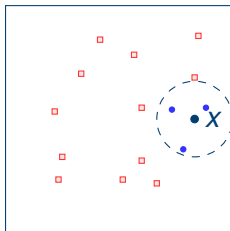


How does this improve the estimator?

Regression
Function
 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$

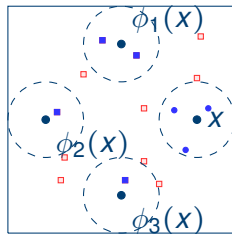


Local Constant
Estimation (LCE)
 $f_n(x)$



Symmetrised LCE

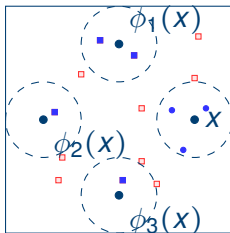
$$\frac{1}{|G|} \sum_{\phi \in G} \hat{f}(\phi(x))$$



How does this improve the estimator?

Symmetrised LCE

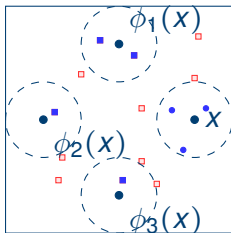
$$\frac{1}{|G|} \sum_{\phi \in G} \hat{f}(\phi(x))$$



How does this improve the estimator?

Symmetrised LCE

$$\frac{1}{|G|} \sum_{\phi \in G} \hat{f}(\phi(x))$$



Each $\hat{f}(\phi(x))$ has:

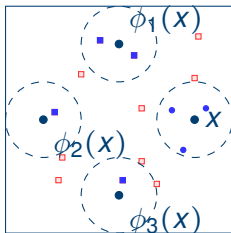
$$\text{Bias}(\hat{f}(\phi(x))) \approx \text{Bias}(\hat{f}(x))$$

$$\text{Var}(\hat{f}(\phi(x))) \approx \text{Var}(\hat{f}(x))$$

How does this improve the estimator?

Symmetrised LCE

$$\frac{1}{|G|} \sum_{\phi \in G} \hat{f}(\phi(x))$$



Each $\hat{f}(\phi(x))$ has:

$$\text{Bias}(\hat{f}(\phi(x))) \approx \text{Bias}(\hat{f}(x))$$

$$\text{Var}(\hat{f}(\phi(x))) \approx \text{Var}(\hat{f}(x))$$

Thus the symmetrised estimator has:

$$\text{Bias}(\hat{f}_G(x)) \approx \text{Bias}(\hat{f}(x))$$

$$\text{Var}(\hat{f}_G(x)) \approx \frac{1}{|G|} \text{Var}(\hat{f}(x))$$

Symmetric functions form closed subspaces

For any class of functions $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ and symmetry G of $\mathcal{X}$, define the subset:

$$\mathcal{F}_G = \{f \in \mathcal{F} : f \text{ is } G\text{-invariant}\} \quad (1)$$

Lemma (Elesedy, 2021)

The subset $L_G^2(\mathcal{X})$ is a closed linear subspace of $L^2(\mathcal{X})$

Symmetric functions form closed subspaces

For any class of functions $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ and symmetry G of $\mathcal{X}$, define the subset:

$$\mathcal{F}_G = \{f \in \mathcal{F} : f \text{ is } G\text{-invariant}\} \quad (1)$$

Lemma (Elesedy, 2021)

The subset $L_G^2(\mathcal{X})$ is a closed linear subspace of $L^2(\mathcal{X})$

When G is compact and $\mu_{\mathcal{X}}$ is G invariant, we can give a closed form orthogonal projection S_G to this subspace:

$$S_G f(x) = \mathbb{E}(f(\phi(x))), \quad \text{where } \phi \sim U(G) \quad (2)$$

Where $U(G)$ is the unique normalised Haar measure on G

How does this symmetrised estimator perform?

If $f \in L^2(\mathcal{X})$ is G -invariant then:

$$\|f_n - f\|_2 \geq \|S_G f_n - f\|_2 \quad (3)$$

almost surely over the data.

How does this symmetrised estimator perform?

If $f \in L^2(\mathcal{X})$ is G -invariant then:

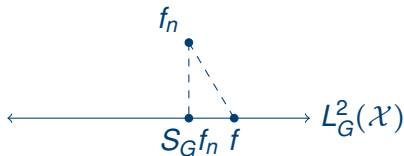
$$\|f_n - f\|_2 \geq \|S_G f_n - f\|_2 \quad (3)$$

almost surely over the data.

Proof.

$$\|f - S_G \hat{f}\| = \|S_G f - S_G \hat{f}\| \leq \|S_G\| \|f - \hat{f}\| \leq \|f - \hat{f}\|.$$

□



Can we achieve dimensionality reduction?

Theorem (cf. Christie and Aston, 2025b+)

If μ_X is G -invariant and f_n is a rate optimal estimator over the Lipschitz class¹, then under **technical conditions**, with a new choice of hyper-parameters for f_n , we have:

$$\mathbb{E}(\|S_G f_n - S_G f\|_2^2) \leq C n^{-\frac{2}{2+d-\dim[x]_G}} \log n \quad (4)$$

where x is a regular point of the action on $\mathcal{X}$.

¹This holds for other functional classes, including on non-Euclidean domains.

Can we achieve dimensionality reduction?

Theorem (cf. Christie and Aston, 2025b+)

If μ_X is G -invariant and f_n is a rate optimal estimator over the Lipschitz class¹, then under **technical conditions**, with a new choice of hyper-parameters for f_n , we have:

$$\mathbb{E}(\|S_G f_n - S_G f\|_2^2) \leq C n^{-\frac{2}{2+d-\dim[x]_G}} \log n \quad (4)$$

where x is a regular point of the action on $\mathcal{X}$.

Corollary

If f is also G -invariant, then $\mathbb{E}(\|S_G f_n - f\|_2^2) \leq C n^{-\frac{2}{2+d-\dim[x]_G}} \log n$.

¹This holds for other functional classes, including on non-Euclidean domains.

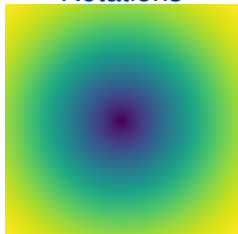
How many dimensions can we remove?

Suppose that G is the smooth action of a Lie group on $\mathcal{X}$.
The set $[x]_G = \{\phi(x) : \phi \in G\}$ is called the **orbit** of x under G .
Each orbit $[x]_G$ is a submanifold of $\mathcal{X}$.

How many dimensions can we remove?

Suppose that G is the smooth action of a Lie group on $\mathcal{X}$.
The set $[x]_G = \{\phi(x) : \phi \in G\}$ is called the **orbit** of x under G .
Each orbit $[x]_G$ is a submanifold of $\mathcal{X}$.

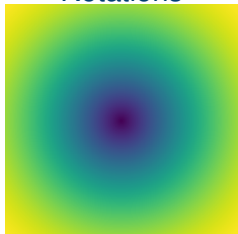
Rotations



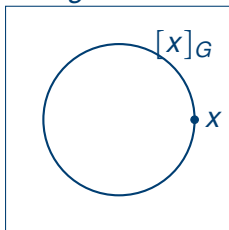
How many dimensions can we remove?

Suppose that G is the smooth action of a Lie group on $\mathcal{X}$.
The set $[x]_G = \{\phi(x) : \phi \in G\}$ is called the **orbit** of x under G .
Each orbit $[x]_G$ is a submanifold of $\mathcal{X}$.

Rotations



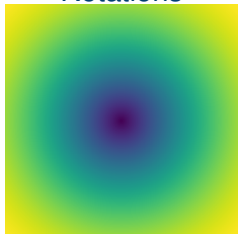
Regular Orbit



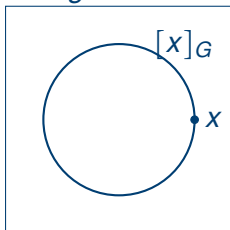
How many dimensions can we remove?

Suppose that G is the smooth action of a Lie group on $\mathcal{X}$.
The set $[x]_G = \{\phi(x) : \phi \in G\}$ is called the **orbit** of x under G .
Each orbit $[x]_G$ is a submanifold of $\mathcal{X}$.

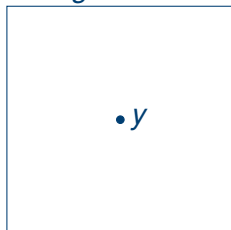
Rotations



Regular Orbit



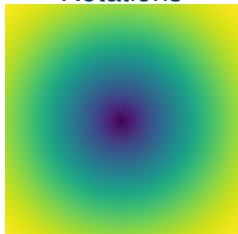
Singular Orbit



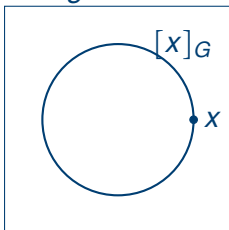
How many dimensions can we remove?

Suppose that G is the smooth action of a Lie group on $\mathcal{X}$.
The set $[x]_G = \{\phi(x) : \phi \in G\}$ is called the **orbit** of x under G .
Each orbit $[x]_G$ is a submanifold of $\mathcal{X}$.

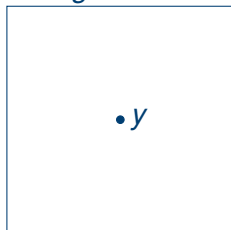
Rotations



Regular Orbit



Singular Orbit



Theorem (Principal Orbit Theorem)

The union of the regular orbits is open and dense in $\mathcal{X}$, and all have the same dimension

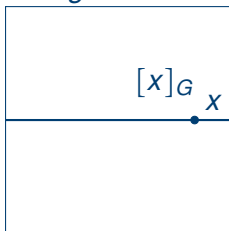
How many dimensions can we remove?

Suppose that G is the smooth action of a Lie group on $\mathcal{X}$.
The set $[x]_G = \{\phi(x) : \phi \in G\}$ is called the **orbit** of x under G .
Each orbit $[x]_G$ is a submanifold of $\mathcal{X}$.

Translations



Regular Orbit



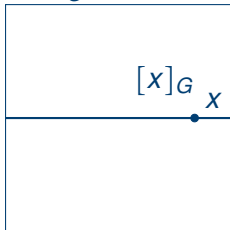
How many dimensions can we remove?

Suppose that G is the smooth action of a Lie group on $\mathcal{X}$.
The set $[x]_G = \{\phi(x) : \phi \in G\}$ is called the **orbit** of x under G .
Each orbit $[x]_G$ is a submanifold of $\mathcal{X}$.

Translations



Regular Orbit



The dimension of the orbit is equal to the number of covariates removed

What are the technical conditions?

- ▶ (Optimality) there exists a bandwidth h such that f_n has $\text{Var}(f_n(x)) \lesssim (nh^d)^{-1}$ and $\text{Bias}(f_n(x)) \lesssim h$;

What are the technical conditions?

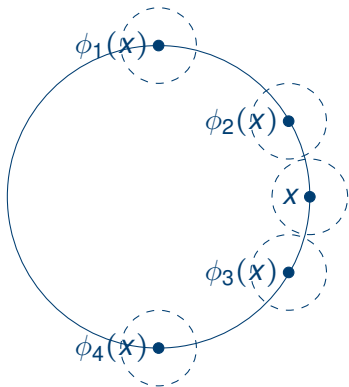
- ▶ **(Optimality)** there exists a bandwidth h such that f_n has $\text{Var}(f_n(x)) \lesssim (nh^d)^{-1}$ and $\text{Bias}(f_n(x)) \lesssim h$;
- ▶ **(Locality)** Conditioned on $\{X_i\}_{i=1}^n$, $\text{Cov}(f_n(x), f_n(y)) \lesssim \exp(-an)$ whenever $d(x, y) \geq 2h$ for some positive a ; and

What are the technical conditions?

- ▶ **(Optimality)** there exists a bandwidth h such that f_n has $\text{Var}(f_n(x)) \lesssim (nh^d)^{-1}$ and $\text{Bias}(f_n(x)) \lesssim h$;
- ▶ **(Locality)** Conditioned on $\{X_i\}_{i=1}^n$, $\text{Cov}(f_n(x), f_n(y)) \lesssim \exp(-an)$ whenever $d(x, y) \geq 2h$ for some positive a ; and
- ▶ **(Anti-Concentration)** $\mathbb{E}\left(\mathcal{R}(X)^{-\dim[X]_G}\right) < \infty$ where $\mathcal{R}(X)$ is related to the injectivity radius of each orbit $[X]_G$.

Sketch of proof of theorem 1

1. Estimate $f(x) \approx \frac{1}{M} \sum_{i=1}^M f_n(\phi_i(x))$
with deterministic $\{\phi_i\}_{i=1}^M$
2. The reduction in variance is on the order M^{-1} using locality of f_n
3. Obtain a lower bound
 $M \geq \max(1, (\mathcal{R}(x)(2h)^{-1})^{\dim[x]_G})$
4. Re-balance the bias-variance trade off achieving faster rate
5. Compare this estimator to $S_G \hat{f}$,
which performs even better
6. Control integrated error with anti-concentration condition.



What if we don't know the symmetries of f ?

What if we don't know the symmetries of f ?

If H is a subset of a symmetry group G on $\mathcal{X}$, then we say that H is a **sub-group** (or sub-symmetry) of G (and write $H \leq G$) if H is also a symmetry of $\mathcal{X}$.

What if we don't know the symmetries of f ?

If H is a subset of a symmetry group G on $\mathcal{X}$, then we say that H is a **sub-group** (or sub-symmetry) of G (and write $H \leq G$) if H is also a symmetry of $\mathcal{X}$.

1. If f is G -invariant and $H \leq G$, then f is H -invariant;

What if we don't know the symmetries of f ?

If H is a subset of a symmetry group G on $\mathcal{X}$, then we say that H is a **sub-group** (or sub-symmetry) of G (and write $H \leq G$) if H is also a symmetry of $\mathcal{X}$.

1. If f is G -invariant and $H \leq G$, then f is H -invariant;
2. If f is not H -invariant and $H \leq G$, then f is not H -invariant; and

What if we don't know the symmetries of f ?

If H is a subset of a symmetry group G on $\mathcal{X}$, then we say that H is a **sub-group** (or sub-symmetry) of G (and write $H \leq G$) if H is also a symmetry of $\mathcal{X}$.

1. If f is G -invariant and $H \leq G$, then f is H -invariant;
2. If f is not H -invariant and $H \leq G$, then f is not H -invariant; and
3. If f is both G - and H -invariant, then f is invariant to the group generated by G and H , written $\langle G, H \rangle$.

What if we don't know the symmetries of f ?

Given any group $\mathcal{G}$ acting continuously and faithfully on $\mathcal{X}$, there exists a unique subgroup $G_{\max}(f, \mathcal{G})$ such that

- 1) f is $G_{\max}(f, \mathcal{G})$ -invariant;
- 2) if $G \leq \mathcal{G}$ and f is G -invariant, then $G \leq G_{\max}(f, \mathcal{G})$;

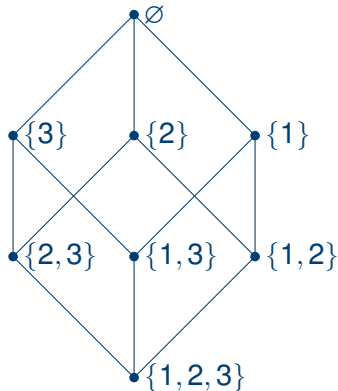
We call $G_{\max}(f, \mathcal{G})$ the **maximal invariant subgroup** of f .

Main Question

Given data $\{(X_i, Y_i)\}_{i=1}^n$ and a search group $\mathcal{G}$, what is $G_{\max}(f, \mathcal{G})$?

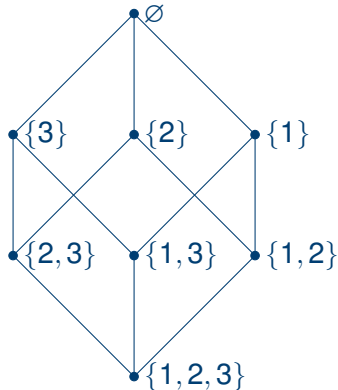
The space of symmetries is larger than sparsities

Variable Index Lattice

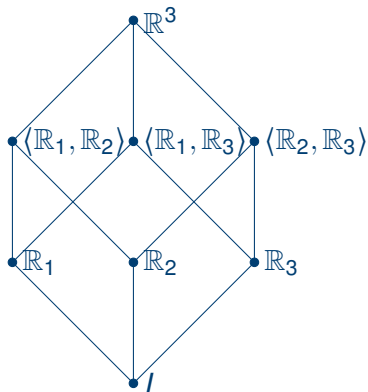


The space of symmetries is larger than sparsities

Variable Index Lattice

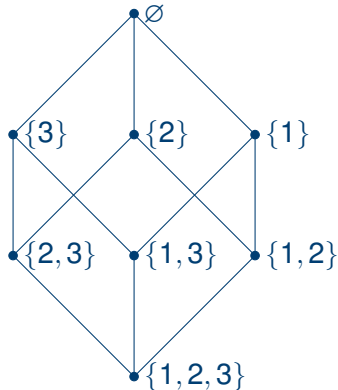


Translation Symmetry Lattice

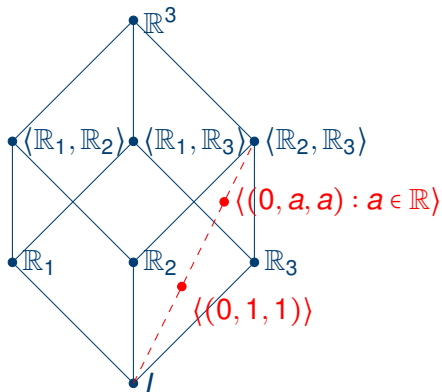


The space of symmetries is larger than sparsities

Variable Index Lattice



Translation Symmetry Lattice



The space of symmetries is larger than sparsities

$$\begin{aligned}\{\text{Translations of co-ord axes}\} &\subseteq \{\text{Translations}\} \\ &\subseteq \{\text{Rigid motions}\} \\ &\subseteq \{\text{Euclidean transformations}\} \\ &\vdots \\ &\subseteq \{\text{Automorphisms}\}\end{aligned}$$

Is a $G_{\max}$ -adaptive estimator statisically possible?

Is a $G_{\max}$ -adaptive estimator statistically possible?

- ▶ We only need to consider closed subgroups
- ▶ We only need to consider connected subgroups (if we ignore constants)

Is a $G_{\max}$ -adaptive estimator statisically possible?

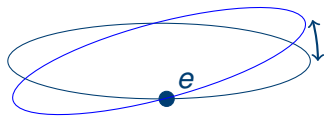
- ▶ We only need to consider closed subgroups
 - ▶ We only need to consider connected subgroups (if we ignore constants)
1. Construct a metric on the set of CC subgroups $\{G \leq \mathcal{G}\}$;
 2. Show that $\|S_G f - f\|_2$ is locally Lipschitz under this metric;
 3. Show that $\{G \leq \mathcal{G}\}$ is totally bounded (or pre-compact);
 4. Construct a δ -cover $\{G_j\}_{j=1}^K$ of this metric space;
 5. Take an M -estimate over this δ -cover.

$$\hat{G}_E = \operatorname{argmin}_{G_j} \sum_{i=1}^n (S_{G_j} f_n(X'_i) - Y'_i)^2 \quad (5)$$

How should we metrize the space of subgroups?

Suppose $\mathcal{G}$ is locally compact Lie group with metric $d_{\mathcal{G}}$, and take a compact neighbourhood U of the identity $e \in \mathcal{G}$. We define a finite semi-metric $d_{\text{Haus}(U)} : K(\mathcal{G}) \times K(\mathcal{G}) \rightarrow [0, \infty)$ given by:

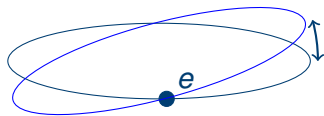
$$d_{\text{Haus}(U)}(G, H) = \max \left(\sup_{g \in G \cap U} \inf_{h \in H \cap U} d_{\mathcal{G}}(g, h), \sup_{h \in H \cap U} \inf_{g \in G \cap U} d_{\mathcal{G}}(g, h) \right)$$



How should we metrize the space of subgroups?

Suppose $\mathcal{G}$ is locally compact Lie group with metric $d_{\mathcal{G}}$, and take a compact neighbourhood U of the identity $e \in \mathcal{G}$. We define a finite semi-metric $d_{\text{Haus}(U)} : K(\mathcal{G}) \times K(\mathcal{G}) \rightarrow [0, \infty)$ given by:

$$d_{\text{Haus}(U)}(G, H) = \max \left(\sup_{g \in G \cap U} \inf_{h \in H \cap U} d_{\mathcal{G}}(g, h), \sup_{h \in H \cap U} \inf_{g \in G \cap U} d_{\mathcal{G}}(g, h) \right)$$



Suppose that $\mathcal{G}$ has a Lipschitz action on $\mathcal{X}$ and f is H invariant, then:

$$|(S_G f)(x) - f(x)| \leq L_{\mathcal{G}} d_{\text{Haus}(U)}(G, H)$$

Is a $G_{\max}$ -adaptive estimator possible?

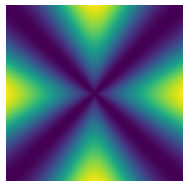
$$\hat{G}_E = \operatorname{argmin}_{G_j} \sum_{i=1}^n (S_{G_j} f_n(X'_i) - Y'_i)^2$$

Theorem (Christie and Aston, 2025b+)

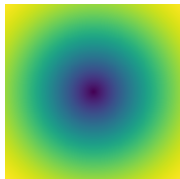
Under technical conditions, if $\mathcal{G}$ is compact and its action is Hölder continuous then we have:

$$\sup_{f \in \mathcal{F}} R(S_{\hat{G}_E} f_n) n^{\frac{2\beta}{2\beta + d - \dim[x]_{G_{\max}(f, \mathcal{G})}}} \leq C \log n$$

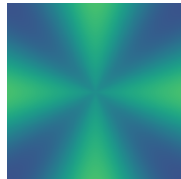
This works from the orthogonal decomposition of f



=



+



$$\sin(r) \cos(2\theta)^2$$

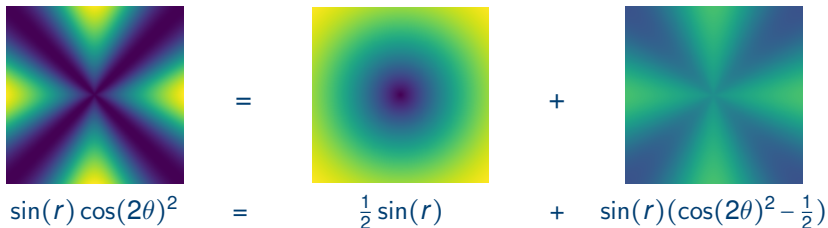
=

$$\frac{1}{2} \sin(r)$$

+

$$\sin(r) (\cos(2\theta)^2 - \frac{1}{2})$$

This works from the orthogonal decomposition of f



The figure illustrates the orthogonal decomposition of a 2D function f into its symmetric and antisymmetric components. It consists of three heatmaps arranged horizontally, separated by an equals sign and a plus sign. The first heatmap on the left shows a function with four-fold symmetry, with dark purple regions along the diagonals and bright yellow-green regions along the axes. Below it is the expression $\sin(r) \cos(2\theta)^2$. The middle heatmap shows a radially symmetric function, with a dark purple center fading to a bright yellow-green outer ring. Below it is the expression $\frac{1}{2} \sin(r)$. The third heatmap on the right shows a function with four-fold symmetry, similar to the first but with a different phase or amplitude distribution, with bright green regions along the diagonals and darker blue-green regions along the axes. Below it is the expression $\sin(r)(\cos(2\theta)^2 - \frac{1}{2})$.

$$\sin(r) \cos(2\theta)^2 = \frac{1}{2} \sin(r) + \sin(r)(\cos(2\theta)^2 - \frac{1}{2})$$

Adaption comes from balancing fast estimation of the symmetric piece with minimisation of the bias of the antisymmetric piece.

Simulated data

Consider the domain $\mathcal{X} = \overline{B_{\mathbb{R}^3}(0, 1)}$ and three regression functions:

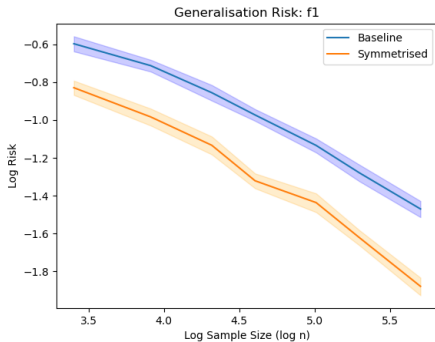
Regression function	Maximal Rotational Symmetry
$f_1(X) = \cos(\ X\ _2)$	$SO(3)$
$f_2(X) = \cos(\sqrt{X_2^2 + X_3^2})$	S_X^1
$f_3(X) = X_1^2 + X_2 - 0.6X_3$	I

With data collected as:

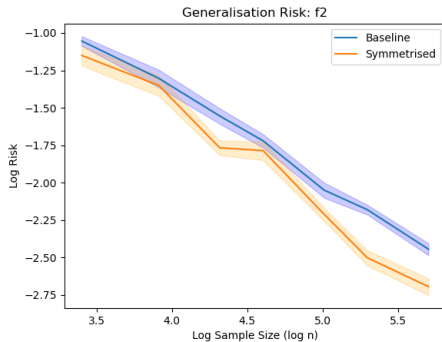
$$Y_i^{(j)} = f_j(X_i^{(j)}) + \epsilon_i^{(j)}; \quad X_i^{(j)} \stackrel{iid}{\sim} U(\mathcal{X}), \epsilon_i^{(j)} \stackrel{iid}{\sim} N(0, 0.05^2) \quad (6)$$

We compare a local constant baseline with one using the estimated rotational symmetry of each dataset

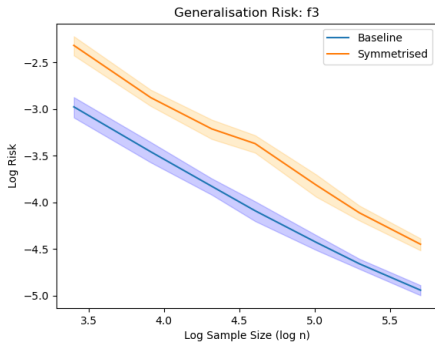
Performance: $f_1(X) = \cos(\|X\|_2)$



Performance: $f_2(X) = \cos(\sqrt{X_2^2 + X_2^3})$



Performance: $f_3(X) = X_1^2 + X_2 - 0.6X_3$



Symmetries work on non-linear domains

Symmetries work on non-linear domains

Sometimes our (covariate) data is constrained to non-linear spaces:

$$X_i \in \mathcal{S}^d$$

Directional Data

$$X_i \in \Delta^d$$

Compositional Data

$$X_i \in \{\text{Space of Digits}\} \subseteq [0, 1]^{\# \text{Pixels}}$$

Image Data

Symmetries work on non-linear domains

Sometimes our (covariate) data is constrained to non-linear spaces:

$$X_i \in \mathcal{S}^d$$

Directional Data

$$X_i \in \Delta^d$$

Compositional Data

$$X_i \in \{\text{Space of Digits}\} \subseteq [0, 1]^{\# \text{Pixels}}$$

Image Data

For these spaces, we cannot select a finite subset of coordinates.

Symmetries work on non-linear domains

Sometimes our (covariate) data is constrained to non-linear spaces:

$X_i \in \mathcal{S}^d$ Directional Data

$X_i \in \Delta^d$ Compositional Data

$X_i \in \{\text{Space of Digits}\} \subseteq [0, 1]^{\# \text{Pixels}}$ Image Data

For these spaces, we cannot select a finite subset of coordinates.
But we can look for symmetries:

- Rotational symmetries of directional data
- Permutation symmetries of compositional data
- Rigid transformations of images

Conclusion

Symmetry generalises sparsity to a non-linear framework.

We can use known symmetries to achieve dimension reduction and can estimate symmetries fast enough to do this adaptively.

Opens up questions for non-linear data spaces: what are the natural symmetries to consider for Wald spaces (for phylogenetic data) or other graph spaces (in drug discovery)?

Can we generalise this to local symmetries?

Can we improve the computational burden through Bayesian methods on the space of subgroups?

References

- (1) Lyle, C., van der Wilk, M., Kwiatkowska, M., Gal, Y., and Bloem-Reddy, B. (2020). On the benefits of invariance in neural networks. arXiv preprint arXiv:2005.00178.
- (2) Bietti, A., Venturi, L. and Bruna, J. (2021) On the sample complexity of learning with geometric stability. Advances in Neural Information Processing.
- (3) Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A. and Vandergheynst, P. (2017) Geometric deep learning: Going beyond Euclidean data. IEEE Signal Processing Magazine.
- (4) Bronstien, M. M., Bruna, J., Cohen, T., Velikovi, P. (2021) Geometric deep learning: Grids, groups, graphs, geodesics, and gauges.
- (5) Christie, L. and Aston, J. (2025a+) Estimating Maximal Symmetries of Regression Functions via Subgroup Lattices. JRSSB, in press, arXiv preprint arXiv:2303.13616
- (6) Christie, L. and Aston, J. (2025b+) Symmetry: a General Structure in Nonparametric Regression. Annals of Statistics, in press arXiv preprint arXiv:2404.12943
- (7) Elesedy, B. (2021) Provably strict generalisation benefit for invariance in kernel methods. Advances in Neural Information Processing.
- (8) Jiang, W. and Tang, L. (2017) Atomic cryo-em structures of viruses. Current Opinion in Structural Biology.