Barriers to Learning Symmetries

Vasco Portilheiro, Apr. 2023

⓪ Quick reminder

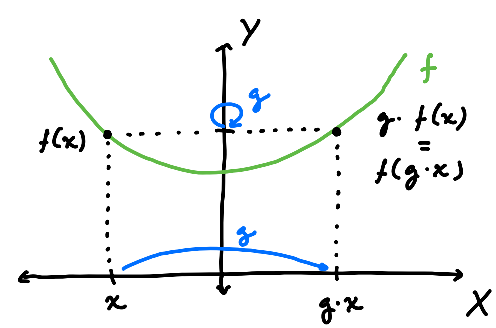
A group G :

- there is $id \in G$: $id \cdot g = g \cdot id = g$
- there are g^{-1} : $g^{-1} g = g g^{-1} = id$
- $(gh)k = g(hk)$

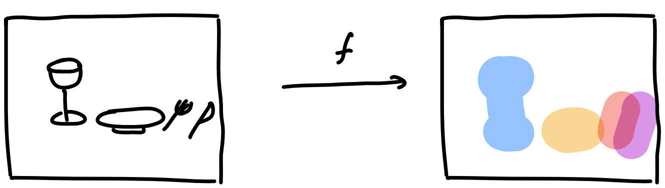
A G - equivariant $f : X \to Y$: $g \cdot f(x) = f(g \cdot x)$

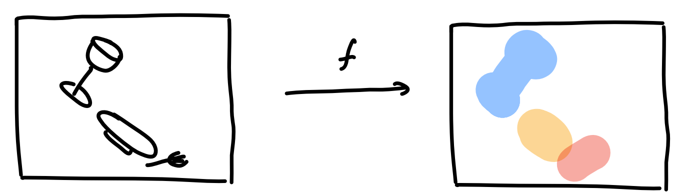
Example :

- $G = \{-1, 1\}$ (w/ usual multiplication)
- $X = \mathbb{R}$, $Y = \mathbb{R}$, $f(x) = x^2$
- G acts on X by multiplication, on Y trivially

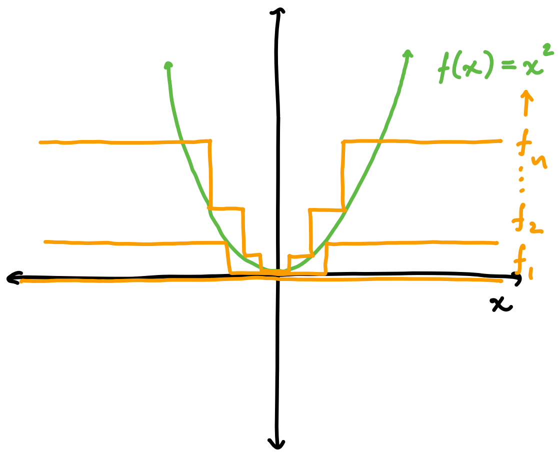


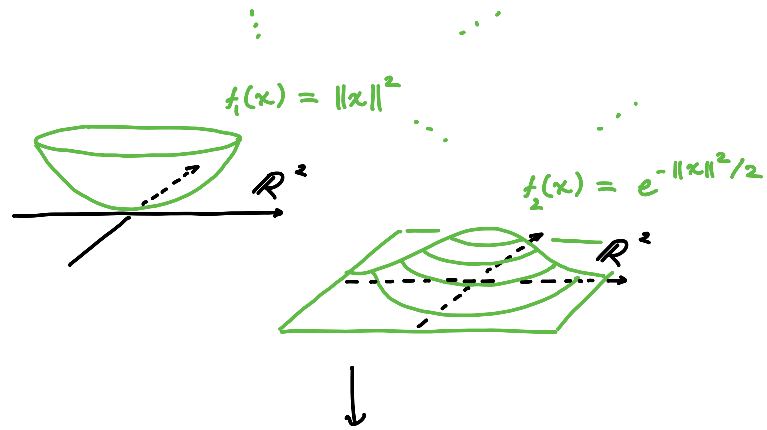
Note : semigroups (no inverses) are also useful





Approximation by equivariant functions **vs** Identifiability of groups given equivariant functions





$G = O(2)$? $G = SO(2)$?

Outline

1. General setup

2. Repurposing EMLPs (Finzi et al. 2021) : a failed (?) experiment
   ↳ • need "approximate equivariance"
     • the "failure" is already worst-case

3. Symmetry non-uniqueness : the "failure" is a special case of a general result
   ↳ what does "learning a group" mean ?

4. GCNNs : they can't "fail"
   ↳ but semigroup convolutions can

① General setup

Let F be a set of functions, Γ a set of groups.

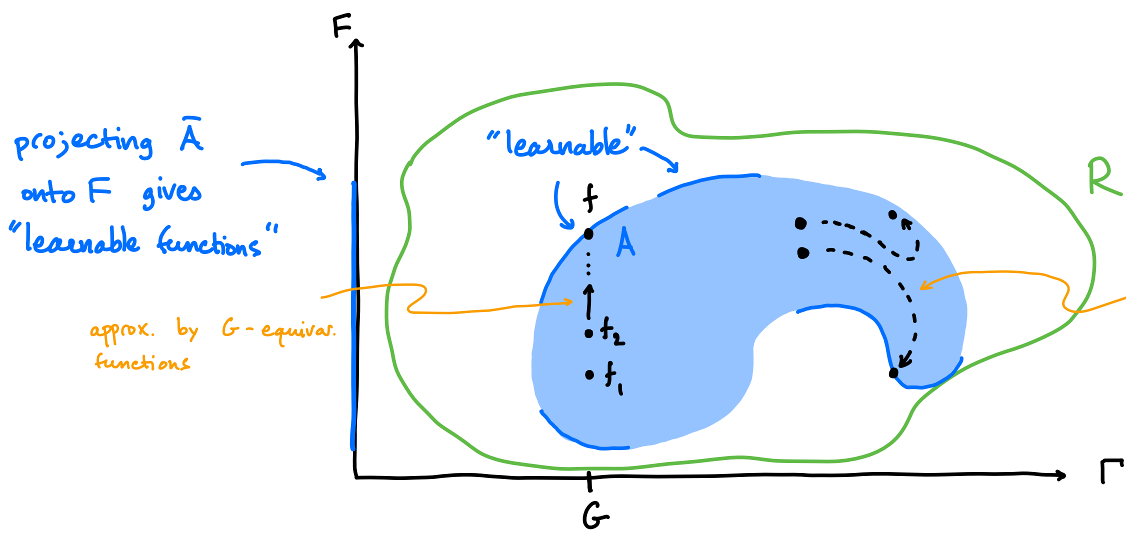
We want to "learn" in F and Γ by approximation.

↳ suppose we have notions of convergence (e.g. topologies) on F, Γ

In particular, we want to learn symmetries

$$\boxed{R \subseteq F \times \Gamma} = \{(f, G) : f \text{ is } G\text{-equivariant}\}$$

using an equivariant hypothesis class : $\boxed{A \subseteq R}$ imposing constraints, e.g. layer-wise equivariant NNs



How do you design NNs with learnable symmetries?

Idea: Fix a class of groups Γ.
For any $G \in \Gamma$, a layer is of the form

$$\text{input} \quad \boxed{x} \xrightarrow{G\text{-equivar.}} Lx \xrightarrow[\text{equivar. on } \Gamma]{\text{non-lin. } \sigma} y = \sigma(Lx)$$

This is the GCNN design pattern. (see Zhou et al. 2021, Dehmamy et al. 2021)
(Γ = "space groups", σ = any point wise nonlinearity)

Problem: if Γ is too large, no non-trivial σ exist. (see also Sergeant-Perthuis et al. 2023)

② EMLPs (Finzi et al. 2021)

Briefly, for a fixed G, an EMLP layer is:

$$\begin{tikzcd}
\text{input} & \boxed{x} \arrow[r] & \tilde{W} x \arrow[r, "\text{equivariant } \sigma \;(\text{non-lin.})"] & \sigma(\tilde{W} x) \\
& & \tilde{W} \arrow[u] & \\
\text{free param.} & & \boxed{W} \arrow[u, "\text{``make } G\text{-equivariant''}"'] &
\end{tikzcd}$$

Based on scaling + tensor contraction

by projecting onto a subspace depending on the generators of G:

discrete: $h_1 \cdots h_m$

Lie algebra: $A_1 \cdots A_n$

Formally, $g\tilde{W} = \tilde{W}_g \Leftrightarrow \tilde{W} = \text{ProjectOntoNullspace}(W, C_{h,A})$

where

$$C_{h,A} = \begin{pmatrix} h_1 \otimes h_1^{-1^T} - I \\ \vdots \\ h_m \otimes h_m^{-1^T} - I \\ A_1 \otimes I - I \otimes A_1^T \\ \vdots \\ A_n \otimes I - I \otimes A_n^T \end{pmatrix}$$

(Finzi et al. 2021, Theorem 1)

<u>Idea</u>: learn the generators simultaneously with W

Approximate equivariance is needed

Problem : no gradient signal, since $C_{h,A}$ "usually" has trivial nullspace

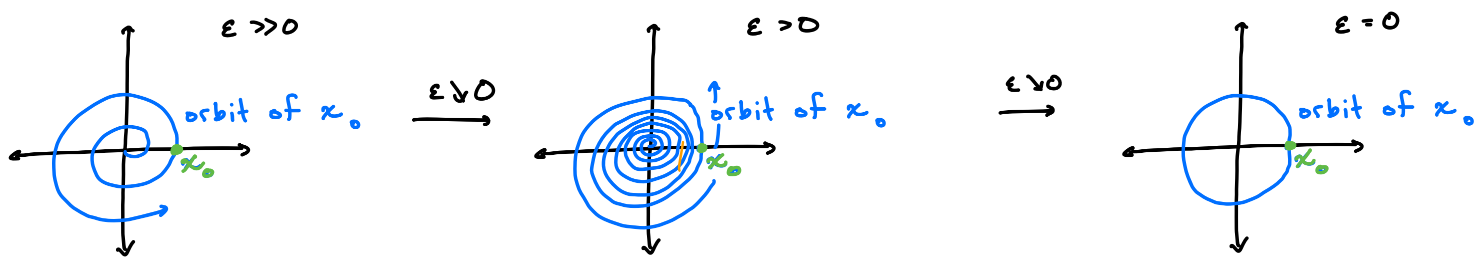
Formal statements can be made ...

Prop : for Lebesgue-almost-every $A \in GL(\mathbb{R}^d)$ there exist no non-constant uniformly continuous $\{A^k : k \in \mathbb{Z}\}$-invariant $f : \mathbb{R}^d \to \mathbb{R}$

Prop : for (product-) Lebesgue a.e. $(A, B) \in GL(\mathbb{R}^d)^2$, there exist no non-trivial linear $W : \mathbb{R}^d \to \mathbb{R}^d$ s.t. $AW = WA$ and $BW = WB$.

More generally, the map from group to group orbits is "discontinuous".

Example: G_ε : the group generated by ~ generator of rotatitions $A_\varepsilon = \begin{pmatrix} \varepsilon & -1 \\ 1 & \varepsilon \end{pmatrix}$.



② EMLP results

no tensor contraction ... others don't converge.

Consider $G = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\} \cong S_2$. Using "simplified" nonlinearities, when trying to learn f:

- if f is non-linear, do not learn f unless $\hat{G} \simeq \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$.
- if f is linear, $\hat{f} \approx f$ but $\hat{G} \simeq \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}$

always a solution ... need priors/regularization

This is exactly the group algebra $\mathbb{R}[G]$:

- elements $a \in \mathbb{R}[G]$ are $a : G \to \mathbb{R}$ with finite support
- $(r \cdot a)(g) = r \cdot a(g)$, $(a+b)(g) = a(g) + b(g)$, $(ab)(g) = \sum_{h_1 h_2 = g} a(h_1) b(h_2)$

We write $a = \sum_g a(g)\, g$, thinking of $\mathbb{R}[G]$ as a "vector space" with basis G.

Fact: a linear $\tilde{W}$ is G-equivariant $\Leftrightarrow$ $\tilde{W}$ is $\mathbb{R}[G]$-equivariant

↳ $\sum_g a(g)\, g \tilde{W} = \sum_g a(g)\, \tilde{W} g = \tilde{W} \sum_g a(g)\, g$

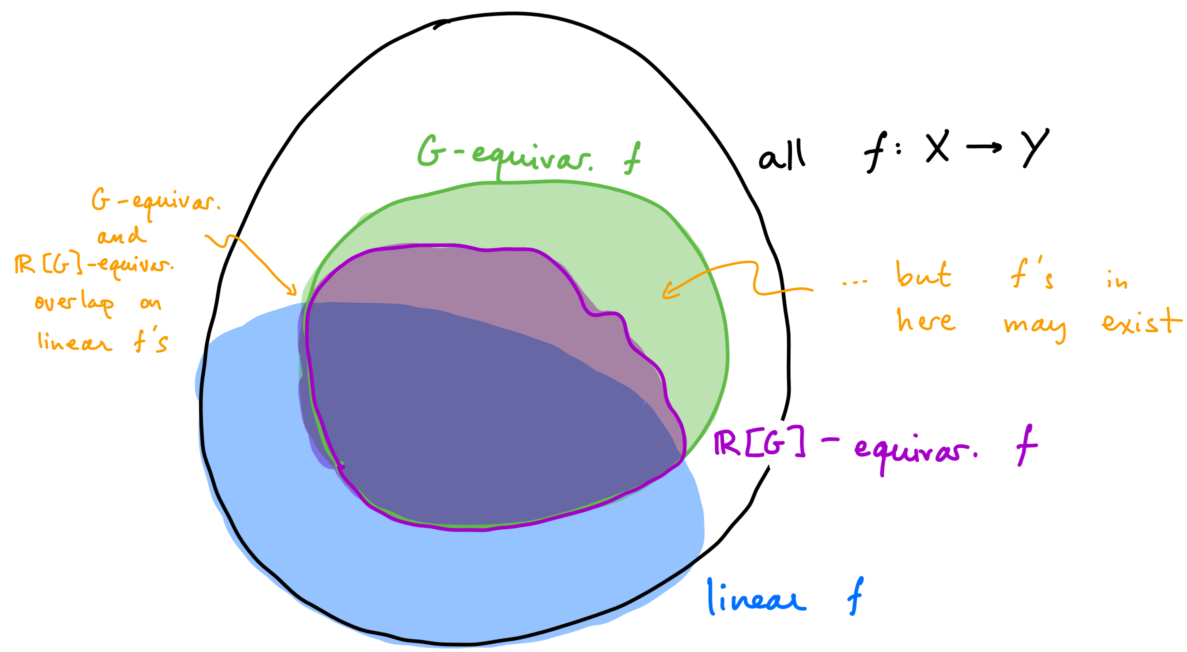
Rabbit hole: but why do we learn $\mathbb{R}[G]$, rather than an even larger structure?

↳ • for semisimple groups, Schur/Jacobson means $A\tilde{W} = \tilde{W}A$ for all G-equiv. $\tilde{W}$ $\Rightarrow A \in \mathbb{R}[G]$

- this generalizes to all unitarizable G of type I, and maps $\tilde{W} : X \to Y$, $X \neq Y$.

. . . given all linear G - equivar. functions . . .

. . . we cannot distinguish between G and $\mathbb{R}[G]$.

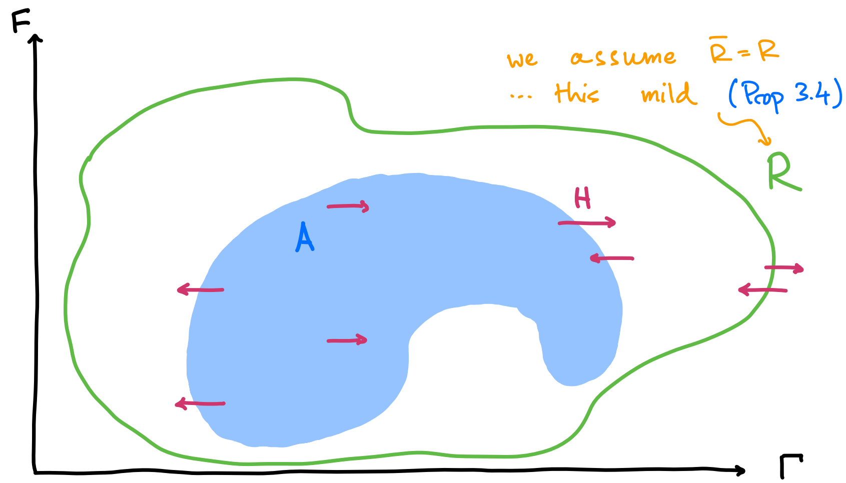


Remark : linear maps can only approximate linear maps

Question : how does this picture generalize to

- maps which can approximate a larger family (e.g. NNs) ?
- approximating G, rather than identifying it ?

③ Symmetry non-uniqueness



A **symmetry non-uniqueness** $H : \Gamma \to \Gamma$ preserves equivariance on A:

$$(f, G) \in A \Rightarrow (f, H(G)) \in R.$$

Theorem (Cor. 3.2): if H is cts. at G, and f is G-equivar. but not $H(G)$-equivar., then (f, G) is not learnable.

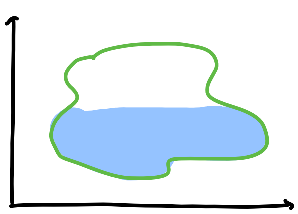
↳ Proof: suppose $(f, G) \in \bar{A}$, so $(f_n, G_n) \in A \to (f, G)$. Then $(f_n, H(G_n)) \in R \to (f, H(G)) \in \bar{R} = R$.

Theorem (Cor. 3.4): if $H(A) \subseteq A$ and $R \subseteq H(R)$, and H is cts. at G
then only one of the following can hold

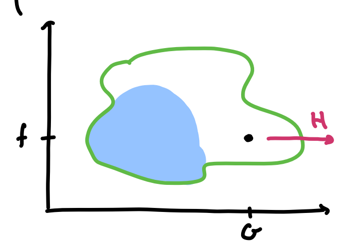
A is "expressive"

H is like "taking a supergroup"

(i) for any G-equivar. learnable f, (f, G) is learnable



(ii) there exists a learnable f equivariant under only one of G, $H(G)$



Rabbit hole : what is a natural convergence / topology on groups ?

$f : X \to Y$

(**Remark** : ideas generalize to approx. equivar.)

$f_n \to f,\ x_n \to x \Rightarrow f_n(x_n) \to f(x)$

e.g.: · compact-open topology
· uniform conv. on compact sets
· Hausdorff metric

For a subset of $\mathrm{Aut}(X) \times \mathrm{Aut}(Y)$, fix an admissible convergence η : $g_n \xrightarrow{\eta} g$.

for proofs / Prop. 3.4

We want groups with elements from the subset , and want a convergence $G_n \to G$ to "respect limit elements" : $\forall\, g \in G$ there are $g_n \in G_n \xrightarrow{\eta} g$.

Option 1 (Thurston) : (closed groups) $G_n \to G$ geometrically if

- $\forall\, g \in G$ there are $g_n \in G_n \xrightarrow{\eta} g$.
- if $g_{n_k} \in G_{n_k} \xrightarrow{\eta} g$ then $g \in G$.

→ if 2^{subset} is a (locally) compact metric space, this is the Hausdorff metric convergence

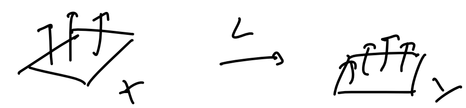
(also Chabauty - Fell / Vietoris topology

Option 2 (Thurston) : for representations of G on $X \times Y$, $\rho_n \to \rho$ algebraically if $\rho(G)$ is the η-limits of $\rho_n(g),\ g \in G$

→ Easy to verify from convergence of group generators

(also makes continuity of non-uniqueness H easy to verify

④ GCCNs



We call integral operators maps L between "signals" $f : X \to \mathbb{R}$ and $Lf : Y \to \mathbb{R}$ of the form

$$(Lf)(y) = \int k(x,y)\, f(x)\, \mu(dx)$$

some measure on X ... usually G-invar.

where k is the kernel function (i.e. "filter").

Fact : let $t_X : X \to X$ be μ-preserving and invertible, and $t_Y : Y \to Y$.

$$(Lf) \circ t_Y \equiv L(f \circ t_X) \iff k(t_X^{-1} x, y) = k(x, t_Y y) \quad \forall y \text{ for } \mu\text{-a.e. } x .$$

e.g. G is compact

Under conditions, μ decomposes as the product of Haar measure λ and $\mu_{X/G}$.

$$(Lf)(g_y, o_y) = \int \ell(g_y^{-1} g_x, o_x, o_y)\, f(g_x, o_x)\, \lambda(dg_x)\, \mu_{X/G}(do_x) \quad \text{with} \quad \ell(g,o,p) = k((g,o),(id,p))$$

Theorem (Thm 4.12) : [under conditions] If μ is $H(G)$-invariant, TFAE:

(i) any G-equivar. integral $L : L^1(X) \to L^\infty(Y)$ is $H(G)$-equivar.

(ii) $H(G)$ acts on X, Y as a subgroup of G

↳ Proof idea : (ii) ⇒ (i) trivially. (i) ⇒ ("Fact" above) "$\ell((h g_y)^{-1} g_x, o_x, h o_y) = \ell(g_y^{-1} h^{-1} g_x, h^{-1} o_x, o_y)$"

So $(h g_y)^{-1} g_x = g_y^{-1}(h^{-1} g_x)$, $o_x = h^{-1} o_x$, $o_y = h^{-1} o_y$.

Rabbit hole : semigroup convolutions can have non-uniquenesses

(e.g. Worrall & Welling 2019)

T acts on the set S and needn't act by homomorphism

If $(Lf)(s_1) = \int \ell(s_2)\, f(s_2 s_1)\ \lambda(ds_2)$ then for any T acting on the right on S

$$(L(t\cdot f))(s_1) = \int \ell(s_2)\, f(s_2 s_1 t)\ \lambda(ds_2) = ((Lf)\cdot t)(s_1)$$

so <u>any</u> super-semigroup T of S of which S is a right-ideal gives a non-uniqueness.

Moral : recognize whether model has non-uniqueness
↳ though strong priors can overcome this

(... also : approximate symmetry alleviates discontinuities)