Stochastic neural network symmetrisation in Markov categories

Rob Cornish

Department of Statistics, University of Oxford

June 19, 2024

Motivation

Top-down, formal constraints for machine learning models

Motivation

Top-down, formal constraints for machine learning models

Want new methodology as well as a better framework

Paper

arXiv > stat > arXiv:2406.11814

Search... | All fields | Search

Help | Advanced Search

Statistics > Machine Learning

[Submitted on 17 Jun 2024]

Stochastic Neural Network Symmetrisation in Markov Categories

Rob Cornish

We consider the problem of symmetrising a neural network along a group homomorphism: given a homomorphism $\varphi : H \to G$, we would like a procedure that converts H-equivariant neural networks into G-equivariant ones. We formulate this in terms of Markov categories, which allows us to consider neural networks whose outputs may be stochastic, but with measure-theoretic details abstracted away. We obtain a flexible, compositional, and generic framework for symmetrisation that relies on minimal assumptions about the structure of the group and the underlying neural network architecture. Our approach recovers existing methods for deterministic symmetrisation as special cases, and extends directly to provide a novel methodology for stochastic symmetrisation also. Beyond this, we believe our findings also demonstrate the utility of Markov categories for addressing problems in machine learning in a conceptual yet mathematically rigorous way.

Subjects: **Machine Learning (stat.ML)**; Machine Learning (cs.LG); Category Theory (math.CT)

Cite as: arXiv:2406.11814 **[stat.ML]**

(or arXiv:2406.11814v1 **[stat.ML]** for this version)

https://doi.org/10.48550/arXiv.2406.11814

Submission history

From: Rob Cornish [view email]

[v1] Mon, 17 Jun 2024 17:54:42 UTC (103 KB)

Access Paper:

- View PDF
- TeX Source
- Other Formats

Current browse context:
stat.ML

< prev | next >

new | recent | 2024-06

Change to browse by:

cs
  cs.LG
math
  math.CT
stat

References & Citations

- NASA ADS
- Google Scholar
- Semantic Scholar

Export BibTeX Citation

Bookmark

Paper

arXiv > stat > arXiv:2406.11814

Search... All fields Search

Help | Advanced Search

Statistics > Machine Learning

[Submitted on 17 Jun 2024]

Stochastic Neural Network Symmetrisation in Markov Categories

Rob Cornish

We consider the problem of symmetrising a neural network along a group homomorphism: given a homomorphism $\varphi : H \to G$, we would like a procedure that converts H-equivariant neural networks into G-equivariant ones. We formulate this in terms of Markov categories, which allows us to consider neural networks whose outputs may be stochastic, but with measure-theoretic details abstracted away. We obtain a flexible, compositional, and generic framework for symmetrisation that relies on minimal assumptions about the structure of the group and the underlying neural network architecture. Our approach recovers existing methods for deterministic symmetrisation as special cases, and extends directly to provide a novel methodology for stochastic symmetrisation also. Beyond this, we believe our findings also demonstrate the utility of Markov categories for addressing problems in machine learning in a conceptual yet mathematically rigorous way.

Subjects: **Machine Learning (stat.ML)**; Machine Learning (cs.LG); Category Theory (math.CT)

Cite as: arXiv:2406.11814 **[stat.ML]**

(or arXiv:2406.11814v1 **[stat.ML]** for this version)

https://doi.org/10.48550/arXiv.2406.11814

Submission history

From: Rob Cornish [view email]

[v1] Mon, 17 Jun 2024 17:54:42 UTC (103 KB)

Access Paper:

- View PDF
- TeX Source
- Other Formats

Current browse context:
stat.ML

< prev | next >

new | recent | 2024-06

Change to browse by:

cs
cs.LG
math
math.CT
stat

References & Citations

- NASA ADS
- Google Scholar
- Semantic Scholar

Export BibTeX Citation

Bookmark

Builds heavily on work in Markov categories, and CT more generally

Background: group equivariance

Often want a function $f : X \to Y$ to be equivariant with respect to the action of a group G, so that

$$f(g \cdot x) = g \cdot f(x) \qquad \text{for all } x \in X \text{ and } g \in G$$

Background: group equivariance

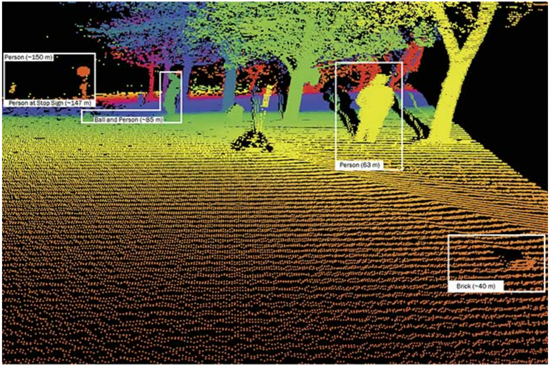
Often want a function $f : X \to Y$ to be equivariant with respect to the action of a group G, so that

$$f(g \cdot x) = g \cdot f(x) \qquad \text{for all } x \in X \text{ and } g \in G$$

Invariance is a special case: $f(g \cdot x) = f(x)$

Examples

Very important for geometric data, such as point clouds:



Taken from www.photonics.com

Other examples: 2D images, sets, graphs, and more

Parametrising equivariance

How can we parameterise an equivariant neural network f?

Parametrising equivariance

How can we parameterise an equivariant neural network f?

Instance of a more general problem: how to ensure a neural network satisfies some "top down" algebraic constraint?

Intrinsic equivariance

Major strategy is intrinsic equivariance:

- Constrain individual layers to obtain equivariance, then compose

Intrinsic equivariance

Major strategy is intrinsic equivariance:

- Constrain individual layers to obtain equivariance, then compose

Some problems with this approach:

- Often quite specific to particular groups and actions
- Hand engineering (e.g. nonlinear layers are often ad hoc)
- Can be brittle and hard to scale

Symmetrisation

Recent interest in symmetrisation approaches: modify an arbitrary neural network to obtain equivariance

Symmetrisation

Recent interest in symmetrisation approaches: modify an arbitrary neural network to obtain equivariance

Example: when it makes sense, the following is invariant:

$$x \mapsto \frac{1}{|G|} \sum_{g \in G} f(g^{-1} \cdot x)$$

Symmetrisation

Recent interest in symmetrisation approaches: modify an arbitrary neural network to obtain equivariance

Example: when it makes sense, the following is invariant:

$$x \mapsto \frac{1}{|G|} \sum_{g \in G} f(g^{-1} \cdot x)$$

Various examples in the literature, including very recently, e.g. [Murphy et al., 2019, Puny et al., 2022, Kaba et al., 2023, Kim et al., 2023]

Stochastic equivariance

More general model: f depends on some additional randomness

Stochastic equivariance

More general model: f depends on some additional randomness

A natural equivariance condition is then:

$$f(g \cdot x, \boldsymbol{U}) \stackrel{\mathrm{d}}{=} g \cdot f(x, \boldsymbol{U})$$

Idea: equivariance across repeated executions

Stochastic equivariance

More general model: f depends on some additional randomness

A natural equivariance condition is then:

$$f(g \cdot x, \boldsymbol{U}) \stackrel{\mathrm{d}}{=} g \cdot f(x, \boldsymbol{U})$$

Idea: equivariance across repeated executions

Practical motivation

More general than deterministic symmetrisation

Practical motivation

More general than deterministic symmetrisation

Of interest in e.g. generative modelling and reinforcement learning, and for uncertainty quantification

Practical motivation

More general than deterministic symmetrisation

Of interest in e.g. generative modelling and reinforcement learning, and for uncertainty quantification

Stochastic symmetrisation does not appear to have been studied in the literature

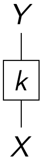
Markov kernels

Convenient to model $(f, \boldsymbol{U})$ as a single entity, a Markov kernel $k : X \to Y$

Markov kernels

Convenient to model $(f, \boldsymbol{U})$ as a single entity, a Markov kernel $k : X \to Y$

Leads to Markov categories as a natural framework:

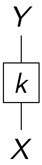


i.e. symmetric monoidal categories with copy and deletion maps

Markov kernels

Convenient to model $(f, \boldsymbol{U})$ as a single entity, a Markov kernel $k : X \to Y$

Leads to Markov categories as a natural framework:



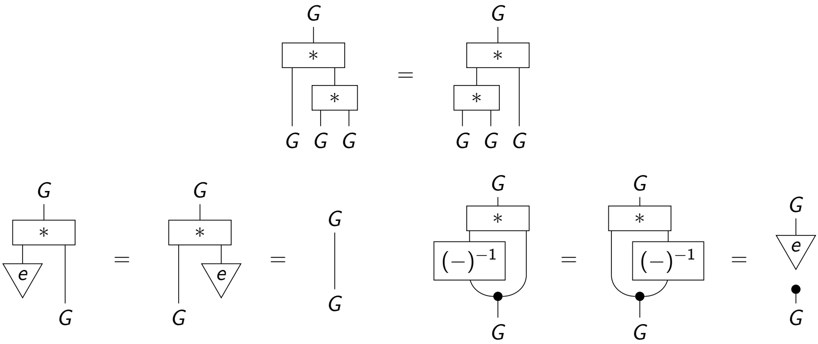
i.e. symmetric monoidal categories with copy and deletion maps

Covers deterministic case also

Group theory in Markov categories

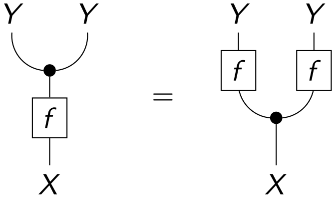
Basic axioms

A group is a Markov category C is an object G together with deterministic morphisms such that



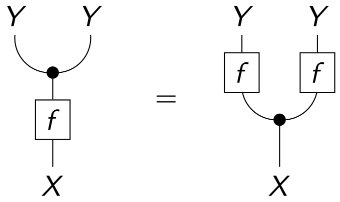
Determinism

Recall: a morphism $f : X \to Y$ in C is deterministic if



Determinism

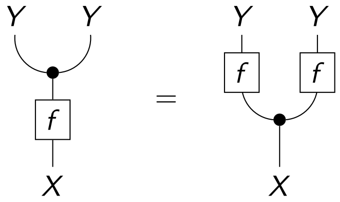
Recall: a morphism $f : X \to Y$ in C is deterministic if



The deterministic morphisms in C form a cartesian monoidal subcategory called $\mathsf{C}_{\mathrm{det}}$

Determinism

Recall: a morphism $f : X \to Y$ in C is deterministic if

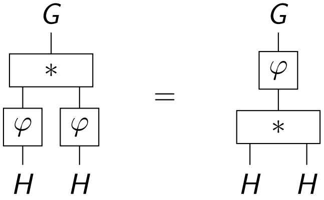


The deterministic morphisms in C form a cartesian monoidal subcategory called $\mathsf{C}_{\mathrm{det}}$

This lets us lift set-theoretic results (e.g. $(g^{-1})^{-1} = g$) to C via the Yoneda Embedding

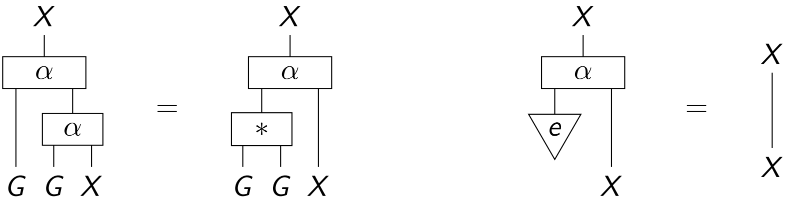
Homomorphisms

Also obtain natural definitions of homomorphism $\varphi : H \to G$ in $\mathsf{C}_{\mathrm{det}}$:



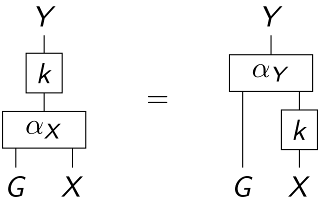
Actions

Similarly, action $\alpha : G \otimes X \to X$ in $\mathsf{C}_{\mathrm{det}}$:



Equivariance

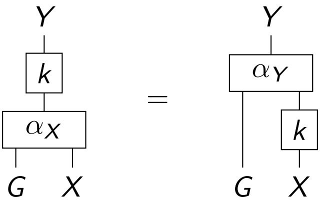
This gives rise to a natural notion of equivariance (now in C): a morphism $k : X \to Y$ is equivariant with respect to α_X and α_Y if



Becomes invariance when α_Y is trivial.

Equivariance

This gives rise to a natural notion of equivariance (now in C): a morphism $k : X \to Y$ is equivariant with respect to α_X and α_Y if



Becomes invariance when α_Y is trivial.

Recovers the desired definitions:

- For C = Set, $k(g \cdot x) = g \cdot k(x)$
- For C = Stoch, $k(dy|g \cdot x) = (g \cdot k)(dy|x)$ as desired

Orbits

Given an action $\alpha : G \otimes X \to X$, a orbit map is a deterministic coequaliser in C as follows:

$$G \otimes X \underset{\varepsilon}{\overset{\alpha}{\rightrightarrows}} X \xrightarrow{q} X/G$$

that is moreover preserved by every functor $(-) \otimes Y$

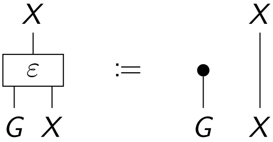
Orbits

Given an action $\alpha : G \otimes X \to X$, a orbit map is a deterministic coequaliser in C as follows:

$$G \otimes X \underset{\varepsilon}{\overset{\alpha}{\rightrightarrows}} X \xrightarrow{q} X/G$$

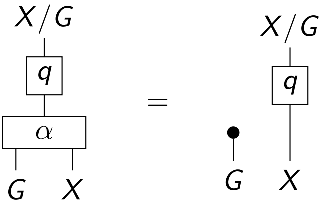
that is moreover preserved by every functor $(-) \otimes Y$

Here ε is the trivial action



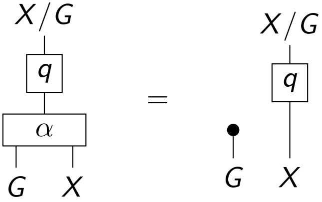
Orbits: intuition

Key idea is that q is initial among invariant maps



Orbits: intuition

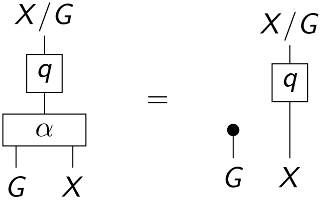
Key idea is that q is initial among invariant maps



Preservation condition ensures: if q is an orbit map, then $q \otimes \mathrm{id}_Y$ is, where G acts trivially on Y

Orbits: intuition

Key idea is that q is initial among invariant maps



Preservation condition ensures: if q is an orbit map, then $q \otimes \mathrm{id}_Y$ is, where G acts trivially on Y

Existence of orbits

Theorem

The Markov category TopStoch *of topological spaces and continuous Markov kernels admits all orbit maps.*

Not sure about the general case (although they seem forthcoming in practice)

Cosets

Recall from classical theory that a subgroup $H \subseteq G$ induces a space of cosets:

$$G/H := \{gH \mid g \in G\}$$

Corresponds to orbits under action $h \cdot g = gh^{-1}$

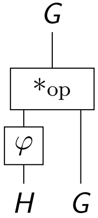
Cosets

Recall from classical theory that a subgroup $H \subseteq G$ induces a space of cosets:

$$G/H := \{gH \mid g \in G\}$$

Corresponds to orbits under action $h \cdot g = gh^{-1}$

Given a homomorphism $\varphi : H \to G$, can define a φ-coset map as an orbit map of the following action:



where $*_{\mathrm{op}}$ does right-multiplication (by inverse)

Recap

In a general Markov category theory, we can talk about:

- Groups
- Homomorphisms
- Actions
- Orbits
- Cosets
- Also semidirect and direct products

Symmetrisation

Markov category of equivariant maps

Given a group G in C, always obtain a Markov category C^G as follows:

- Objects are pairs (X, α_X), where α_X is an action on X in C

Markov category of equivariant maps

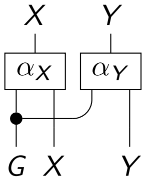
Given a group G in C, always obtain a Markov category C^G as follows:

- Objects are pairs (X, α_X), where α_X is an action on X in C
- Morphisms are equivariant morphisms in C

Markov category of equivariant maps

Given a group G in C, always obtain a Markov category C^G as follows:

- Objects are pairs (X, α_X), where α_X is an action on X in C
- Morphisms are equivariant morphisms in C
- Monoidal product $(X, \alpha_X) \otimes (Y, \alpha_Y)$ is equipped with

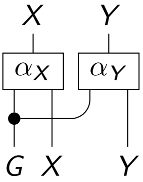


- Other components inherited from C

Markov category of equivariant maps

Given a group G in C, always obtain a Markov category C^G as follows:

- Objects are pairs (X, α_X), where α_X is an action on X in C
- Morphisms are equivariant morphisms in C
- Monoidal product $(X, \alpha_X) \otimes (Y, \alpha_Y)$ is equipped with



- Other components inherited from C

(*Almost* Eilenberg-Moore category of action monad)

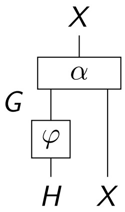
Restriction

Given a homomorphism $\varphi : H \to G$, obtain a functor $R_\varphi : \mathsf{C}^G \to \mathsf{C}^H$ by restriction

Restriction

Given a homomorphism $\varphi : H \to G$, obtain a functor $R_\varphi : \mathsf{C}^G \to \mathsf{C}^H$ by restriction

Idea: $R_\varphi(X, \alpha)$ is X equipped with the H-action



Symmetrisation procedures

In this context, can define a symmetrisation procedure as a function

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \to \mathsf{C}^G(X, Y)$$

(where actions are denoted implicitly)

Symmetrisation procedures

In this context, can define a symmetrisation procedure as a function

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \to \mathsf{C}^G(X, Y)$$

(where actions are denoted implicitly)

Key idea: sends "less equivariant" morphisms to "more equivariant" ones

- E.g. consider $H = I$ the trivial group

Compositionality

Can compose procedures sequentially: given homomorphisms

$$K \xrightarrow{\phi} H \xrightarrow{\varphi} G$$

Compositionality

Can compose procedures sequentially: given homomorphisms

$$K \xrightarrow{\phi} H \xrightarrow{\varphi} G$$

can symmetrise as follows:

$$\mathsf{C}^K(R_\phi R_\varphi X, R_\phi R_\varphi Y) \xrightarrow{\mathsf{sym}_\phi} \mathsf{C}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\mathsf{sym}_\varphi} \mathsf{C}^G(X, Y).$$

where here also $R_\phi R_\varphi = R_{\varphi \circ \phi}$

Compositionality

Can compose procedures sequentially: given homomorphisms

$$K \xrightarrow{\phi} H \xrightarrow{\varphi} G$$

can symmetrise as follows:

$$\mathsf{C}^K(R_\phi R_\varphi X, R_\phi R_\varphi Y) \xrightarrow{\quad \mathsf{sym}_\phi \quad} \mathsf{C}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\quad \mathsf{sym}_\varphi \quad} \mathsf{C}^G(X, Y).$$

where here also $R_\phi R_\varphi = R_{\varphi \circ \phi}$

Can therefore "build up" complex equivariance constraints in a structured way

Methodology

Motivating idea

Suppose R_φ has a left adjoint E (often true classically [May et al., 1997])

Motivating idea

Suppose R_φ has a left adjoint E (often true classically [May et al., 1997])

Obtain directly

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \cong \mathsf{C}^G(ER_\varphi X, Y)$$

Motivating idea

Suppose R_φ has a left adjoint E (often true classically [May et al., 1997])

Obtain directly

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \cong \mathsf{C}^G(ER_\varphi X, Y)$$

This recharacterises the problem of symmetrisation:

- Before, H-equivariance $\mapsto$ G-equivariance
- Now, G-equivariance $\mapsto$ G-equivariance (of another kind)

General approach

When a left adjoint $E \dashv R_\varphi$ exists, obtain the following methodology:

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\quad\cong\quad} \mathsf{C}^G(ER_\varphi X, Y) \xrightarrow{\text{Precompose}} \mathsf{C}^G(X, Y)$$

General approach

When a left adjoint $E \dashv R_\varphi$ exists, obtain the following methodology:

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\quad\cong\quad} \mathsf{C}^G(ER_\varphi X, Y) \xrightarrow{\text{Precompose}} \mathsf{C}^G(X, Y)$$

In the second step, require a morphism $X \to ER_\varphi X$ in C^G

General approach

When a left adjoint $E \dashv R_\varphi$ exists, obtain the following methodology:

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\quad\cong\quad} \mathsf{C}^G(ER_\varphi X, Y) \xrightarrow{\text{Precompose}} \mathsf{C}^G(X, Y)$$

In the second step, require a morphism $X \to ER_\varphi X$ in C^G

This must be already G-equivariant, as for other symmetrisation approaches

- But now this can be very trivial compared with overall model

More general idea

A full left adjoint is quite onerous to obtain, and more than we need

More general idea

A full left adjoint is quite onerous to obtain, and more than we need

Classically in Set, have the following natural isomorphism:

$$ER_{\varphi} \cong G/H \otimes (-)$$

where G/H is the coset space (with a canonical G-action)

More general idea

A full left adjoint is quite onerous to obtain, and more than we need

Classically in Set, have the following natural isomorphism:

$$ER_\varphi \cong G/H \otimes (-)$$

where G/H is the coset space (with a canonical G-action)

Our idea: show directly that previous isomorphism of hom sets hold when ER_φ is replaced like this

- Now G/H is the codomain of a φ-coset map

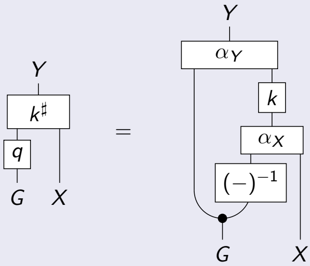
Main result

Theorem

Suppose a φ-coset map $q : G \to G/H$ exists. Then for all X and Y in C^G there is a bijection

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\cong} \mathsf{C}^G(G/H \otimes X, Y)$$

that sends $k : R_\varphi X \to R_\varphi Y$ in C^H to the unique $k^\sharp$ such that



Some comments

Theoretically, corresponds to an equivalence of categories between:

- The full image of R_φ
- The co-Kleisli category of the reader comonad $G/H \otimes -$

Some comments

Theoretically, corresponds to an equivalence of categories between:

- The full image of R_φ
- The co-Kleisli category of the reader comonad $G/H \otimes -$

By earlier result, a φ-coset map always exists in TopStoch

- Seems quite forthcoming in other contexts as well

Some comments

Theoretically, corresponds to an equivalence of categories between:

- The full image of R_φ
- The co-Kleisli category of the reader comonad $G/H \otimes -$

By earlier result, a φ-coset map always exists in TopStoch

- Seems quite forthcoming in other contexts as well

Practically, can compute bijection finding a section of φ-coset map

Resulting procedure

Overall procedure becomes as follows:

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\quad\cong\quad} \mathsf{C}^G(G/H \otimes X, Y) \xrightarrow{\text{Precompose}} \mathsf{C}^G(X, Y)$$

where now precomposing by some $X \to G/H \otimes X$ in C^G

Resulting procedure

Overall procedure becomes as follows:

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\quad\cong\quad} \mathsf{C}^G(G/H \otimes X, Y) \xrightarrow{\text{Precompose}} \mathsf{C}^G(X, Y)$$

where now precomposing by some $X \to G/H \otimes X$ in C^G

Recovers all existing deterministic symmetrisation techniques I am aware of (in a sense inevitably) when combined with further averging step

$$\mathsf{C}^G(X, Y) \longrightarrow \mathsf{C}^G_{\mathrm{det}}(X, Y)$$

Resulting procedure

Overall procedure becomes as follows:

$$\mathsf{C}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\quad\cong\quad} \mathsf{C}^G(G/H \otimes X, Y) \xrightarrow{\text{Precompose}} \mathsf{C}^G(X, Y)$$

where now precomposing by some $X \to G/H \otimes X$ in C^G

Recovers all existing deterministic symmetrisation techniques I am aware of (in a sense inevitably) when combined with further averging step

$$\mathsf{C}^G(X, Y) \longrightarrow \mathsf{C}^G_{\mathrm{det}}(X, Y)$$

Directly gives rise to a novel procedure for stochastic symmetrisation also

Implementation

Application

The method of Kim et al. [2023] requires a stochastically equivariant subcomponent

Application

The method of Kim et al. [2023] requires a stochastically equivariant subcomponent

Kim et al. [2023] use an intrinsically equivariant neural network for this

Application

The method of Kim et al. [2023] requires a stochastically equivariant subcomponent

Kim et al. [2023] use an intrinsically equivariant neural network for this

We apply our stochastic symmetrisation approach instead

Application

The method of Kim et al. [2023] requires a stochastically equivariant subcomponent

Kim et al. [2023] use an intrinsically equivariant neural network for this

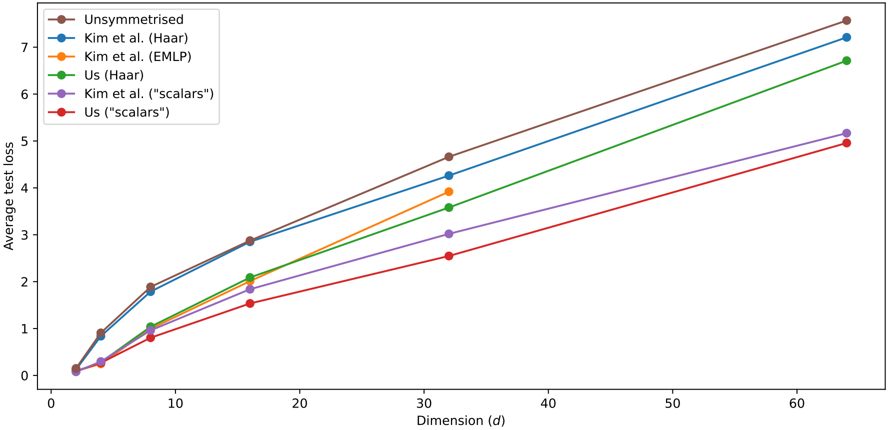
We apply our stochastic symmetrisation approach instead

Consider learning the matrix inversion function $A \mapsto A^{-1}$, which is equivariant with respect to orthogonal group:

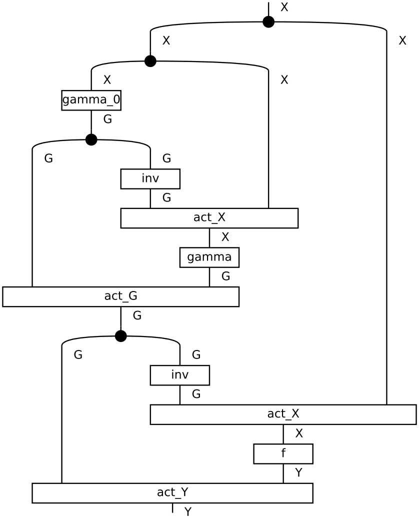
$$(QA)^{-1} = A^{-1}Q^{-1} = A^{-1}Q^T$$

More examples needed!

Results


Unsymmetrised
Kim et al. (Haar)
Kim et al. (EMLP)
Us (Haar)
Kim et al. ("scalars")
Us ("scalars")
Average test loss
Dimension (d)
0
1
2
3
4
5
6
7
0
10
20
30
40
50
60

Implementation - DisCoPy



Thank you!

References I

Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak Rao, and Bruno Ribeiro. Janossy pooling: Learning deep permutation-invariant functions for variable-size inputs. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=BJluy2RcFm`.

Omri Puny, Matan Atzmon, Edward J. Smith, Ishan Misra, Aditya Grover, Heli Ben-Hamu, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=zIUyj55nXR`.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 15546–15566. PMLR, 23–29 Jul 2023. URL `https://proceedings.mlr.press/v202/kaba23a.html`.

Jinwoo Kim, Dat Nguyen, Ayhan Suleymanzade, Hyeokjun An, and Seunghoon Hong. Learning probabilistic symmetrization for architecture agnostic equivariance. In A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 18582–18612. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/3b5c7c9c5c7bd77eb73d0baec7a07165-Paper-Conference.pdf.

J.P. May, R.J. Piacenza, and M. Cole. *Equivariant Homotopy and Cohomology Theory: Dedicated to the Memory of Robert J. Piacenza*. Regional conference series in mathematics. American Mathematical Society, 1997. ISBN 9780821803197. URL https://books.google.co.uk/books?id=KOcZYVxkQO9C.