Equivariant Neural Tangent Kernels

Learning on graphs and geometry meetup

by Philipp Misof

joint work with Jan Gerken, Pan Kessel

Department of Mathematical Sciences, Division of Algebra and Geometry

February 11, 2025

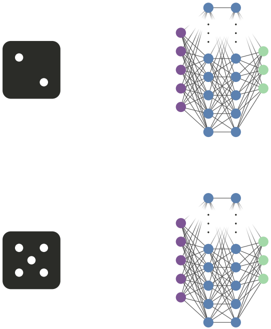
UNIVERSITY OF GOTHENBURG

CHALMERS UNIVERSITY OF TECHNOLOGY

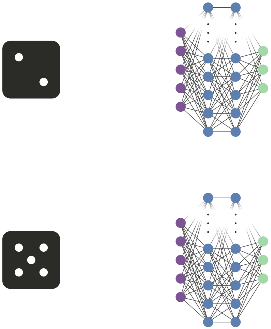
WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

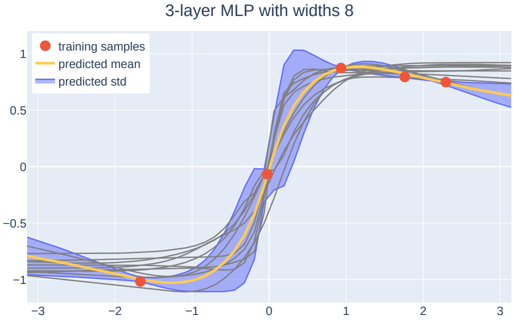
What can we say about the dynamics of NNs without specifying their initial parameters?

What can we say about the dynamics of NNs without specifying their initial parameters?

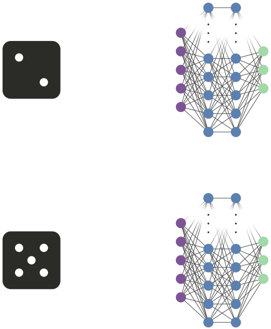


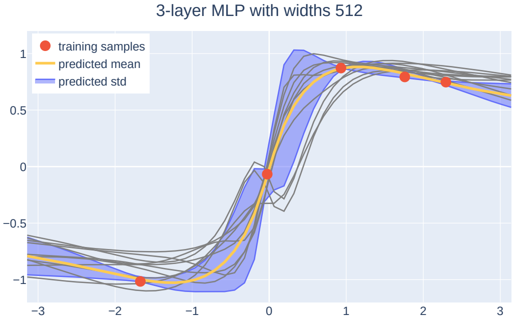
What can we say about the dynamics of NNs without specifying their initial parameters?



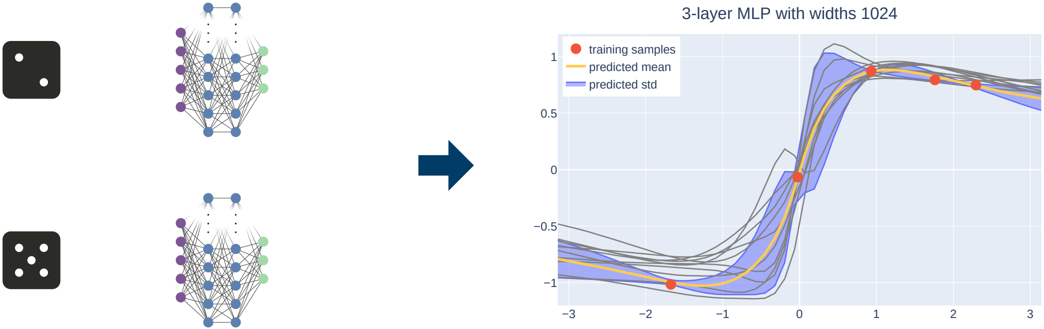


What can we say about the dynamics of NNs without specifying their initial parameters?





What can we say about the dynamics of NNs without specifying their initial parameters?



Training dynamics

(Jacot, Gabriel, and Hongler 2018)

Training dynamics

(Jacot, Gabriel, and Hongler 2018)

$$\frac{\mathrm{d}\mathcal{N}}{\mathrm{d}t} = -\eta \sum_{i=1}^{n_{\text{train}}} \Theta_t(x, x_i) \frac{\partial \mathcal{L}}{\partial \mathcal{N}(x_i)}$$

Training dynamics

(Jacot, Gabriel, and Hongler 2018)

Neural Network output

Learning rate

Loss

$$\frac{\mathrm{d}\mathcal{N}}{\mathrm{d}t} = -\eta \sum_{i=1}^{n_{\text{train}}} \Theta_t(x, x_i) \frac{\partial \mathcal{L}}{\partial \mathcal{N}(x_i)}$$

Training time

Neural Tangent Kernel

Training dynamics

(Jacot, Gabriel, and Hongler 2018)

$$\frac{\mathrm{d}\mathcal{N}}{\mathrm{d}t} = -\eta \sum_{i=1}^{n_{\mathrm{train}}} \Theta_t(x, x_i) \frac{\partial \mathcal{L}}{\partial \mathcal{N}(x_i)}$$

Neural Network output

Learning rate

Loss

Training time

Neural Tangent Kernel

$$\Theta_t \xrightarrow[\text{layer width}]{\text{increasing}} \Theta$$

Training dynamics

(Jacot, Gabriel, and Hongler 2018)

Neural Network output — Learning rate — Loss

$$\frac{\mathrm{d}\mathcal{N}}{\mathrm{d}t} = -\eta \sum_{i=1}^{n_{\text{train}}} \Theta_t(x, x_i) \frac{\partial \mathcal{L}}{\partial \mathcal{N}(x_i)}$$

Training time — Neural Tangent Kernel

$$\Theta_t \xrightarrow[\text{layer width}]{\text{increasing}} \Theta$$

➡ Simple ODE

We extend this to equivariant NNs

We extend this to equivariant NNs

Group Convolution

We extend this to equivariant NNs

Group Convolution

Group Pooling

We extend this to equivariant NNs

Group Convolution

Group Pooling

Elementwise Non-linearity

We extend this to equivariant NNs

Group Convolution

Group Pooling

Elementwise Non-linearity

Explicit expressions for

- **roto-translations** $G = C_4 \ltimes \mathbb{R}^2$ in the plane
- **3d-rotations** $G = \mathrm{SO}(3)$ on the sphere S^2
  via **spherical convolutions**

We extend this to equivariant NNs

Group Convolution

Group Pooling

Elementwise Non-linearity

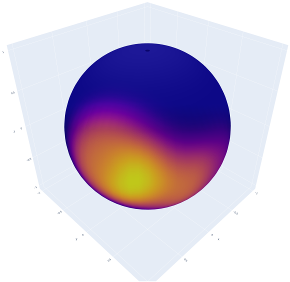
Explicit expressions for

- **roto-translations** $G = C_4 \ltimes \mathbb{R}^2$ in the plane
- **3d-rotations** $G = \mathrm{SO}(3)$ on the sphere S^2 via **spherical convolutions**

Provide **implementations** in the `neural-tangents` library *(Novak et al. 2020)*

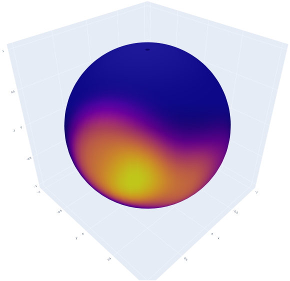
∞-width limit in practice: Molecular Energy Prediction on QM9

∞-width limit in practice: Molecular Energy Prediction on QM9



Atoms' environments are represented as signals on the sphere *(Esteves, Slotine, and Makadia 2023)*

∞-width limit in practice: Molecular Energy Prediction on QM9



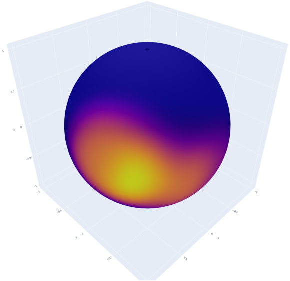
Atoms' environments are represented as signals on the sphere *(Esteves, Slotine, and Makadia 2023)*

MLP

vs.

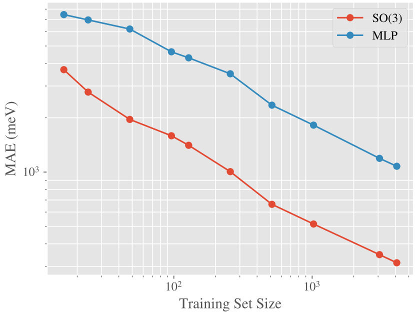
$\mathrm{SO}(3)$ **GCNN**

∞-width limit in practice: Molecular Energy Prediction on QM9

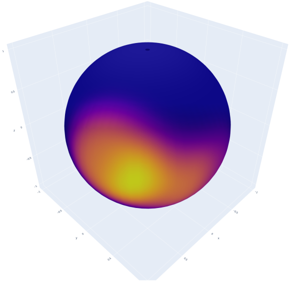


Atoms' environments are represented as signals on the sphere *(Esteves, Slotine, and Makadia 2023)*

MLP
vs.
$\mathrm{SO}(3)$ **GCNN**

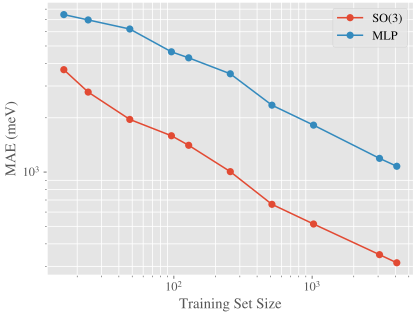


∞-width limit in practice: Molecular Energy Prediction on QM9



Atoms' environments are represented as signals on the sphere *(Esteves, Slotine, and Makadia 2023)*

MLP
vs.
$\mathrm{SO}(3)$ **GCNN**



➡ Performance boost due to 3d-rotation invariance extends to the ∞-width limit

Data Augmentation $\leftrightarrow$ Group Convolutional (GC) NNs

Data Augmentation $\leftrightarrow$ Group Convolutional (GC) NNs

For some specific GCNN,

$$\Theta^{\mathrm{GC}}(f, f') = \frac{1}{\mathrm{vol}(G)} \int_G \mathrm{d}g \; \Theta^{\mathrm{MLP}}(f, \rho_{\mathrm{reg}}(g) f')$$

Data Augmentation $\leftrightarrow$ Group Convolutional (GC) NNs

For some specific GCNN,

$$\Theta^{\mathrm{GC}}(f, f') = \underbrace{\frac{1}{\mathrm{vol}(G)} \int_G \mathrm{d}g \; \Theta^{\mathrm{MLP}}(f, \rho_{\mathrm{reg}}(g) f')}_{\text{Corresponds to data augmentation}}$$

Data Augmentation $\leftrightarrow$ Group Convolutional (GC) NNs

For some specific GCNN,

$$\Theta^{\mathrm{GC}}(f, f') = \underbrace{\frac{1}{\mathrm{vol}(G)} \int_G \mathrm{d}g \; \Theta^{\mathrm{MLP}}(f, \rho_{\mathrm{reg}}(g) f')}_{\text{Corresponds to data augmentation}}$$

➡ At ∞-width: **mean** of an ensemble of data augmented MLPs equals the **mean** of an ensemble of GCNNs **at all training times** t.

Data Augmentation $\leftrightarrow$ Group Convolutional (GC) NNs

For some specific GCNN,

$$\Theta^{\mathrm{GC}}(f, f') = \underbrace{\frac{1}{\mathrm{vol}(G)} \int_G \mathrm{d}g \; \Theta^{\mathrm{MLP}}(f, \rho_{\mathrm{reg}}(g) f')}_{\text{Corresponds to data augmentation}}$$

➡ At ∞-width: **mean** of an ensemble of data augmented MLPs equals the **mean** of an ensemble of GCNNs **at all training times** t.

Similar result for data augmented CNN $\leftrightarrow$ $C_4 \ltimes \mathbb{R}^2$ GCNN

Data Augmentation vs. Group Convolutions **at finite width**

Data Augmentation vs. Group Convolutions **at finite width**

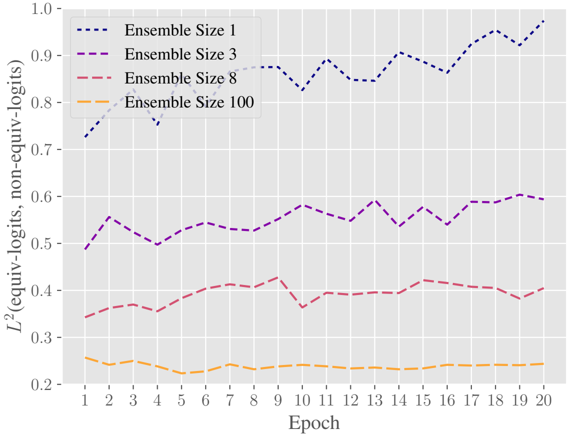
- Data augmented CNN vs $C_4 \ltimes \mathbb{R}^2$ GCNN on **MNIST**

Data Augmentation vs. Group Convolutions **at finite width**

- Data augmented CNN vs $C_4 \ltimes \mathbb{R}^2$ GCNN on **MNIST**
- Compare L_2-difference of mean logits on **out-of-distribution** data

Data Augmentation vs. Group Convolutions **at finite width**

- Data augmented CNN vs $C_4 \ltimes \mathbb{R}^2$ GCNN on **MNIST**
- Compare L_2-difference of mean logits on **out-of-distribution** data



Summing up

Summing up

- **Extended** the Neural Tangent Kernel theory **to equivariant neural networks**

Summing up

- **Extended** the Neural Tangent Kernel theory **to equivariant neural networks**
- Provide explicit expressions for **roto-translations** and **3d-rotations**

Summing up

- **Extended** the Neural Tangent Kernel theory **to equivariant neural networks**
- Provide explicit expressions for **roto-translations** and **3d-rotations**
- Showed how this analytic tool can be used to **analyze the connection between data augmentation and GCNNs**

Summing up

- **Extended** the Neural Tangent Kernel theory **to equivariant neural networks**
- Provide explicit expressions for **roto-translations** and **3d-rotations**
- Showed how this analytic tool can be used to **analyze the connection between data augmentation and GCNNs**

What's next?

Summing up

- **Extended** the Neural Tangent Kernel theory **to equivariant neural networks**
- Provide explicit expressions for **roto-translations** and **3d-rotations**
- Showed how this analytic tool can be used to **analyze the connection between data augmentation and GCNNs**

What's next?

- equivariant graph NNs

Summing up

- **Extended** the Neural Tangent Kernel theory **to equivariant neural networks**
- Provide explicit expressions for **roto-translations** and **3d-rotations**
- Showed how this analytic tool can be used to **analyze the connection between data augmentation and GCNNs**

What's next?

- equivariant graph NNs
- Trainability and generalization regimes of equivariant NNs *(Xiao, Pennington, and Schoenholz 2020)*

Summing up

- **Extended** the Neural Tangent Kernel theory **to equivariant neural networks**
- Provide explicit expressions for **roto-translations** and **3d-rotations**
- Showed how this analytic tool can be used to **analyze the connection between data augmentation and GCNNs**

What's next?

- equivariant graph NNs
- Trainability and generalization regimes of equivariant NNs *(Xiao, Pennington, and Schoenholz 2020)*
- Relations to Quantum Field Theory *(Banta et al. 2024)*

Do you want to know more?

arXiv:2406.06504