Symmetry-Based Structured Matrices for Efficient Approximately Equivariant Networks

Ashwin Samudre[1], Mircea Petrache[2], Brian D. Nord[3], Shubhendu Trivedi[4]

[1]SFU [2]PUC Chile & CENIA [3]FNAL & UChicago [4]Independent

AISTATS 2025

Introduction

- **Equivariance in Neural Networks**: Output transforms predictably under input symmetry (e.g., rotation of images).

Introduction

- **Equivariance in Neural Networks**: Output transforms predictably under input symmetry (e.g., rotation of images).
- **Challenge**: Strict equivariance can fail with imperfect real-world symmetries.

Introduction

- **Equivariance in Neural Networks**: Output transforms predictably under input symmetry (e.g., rotation of images).
- **Challenge**: Strict equivariance can fail with imperfect real-world symmetries.
- **Solution**: Approximately equivariant networks balance flexibility and symmetry.

Introduction

- **Equivariance in Neural Networks**: Output transforms predictably under input symmetry (e.g., rotation of images).
- **Challenge**: Strict equivariance can fail with imperfect real-world symmetries.
- **Solution**: Approximately equivariant networks balance flexibility and symmetry.
- **Our Approach**: Use Group Matrices (GMs) to generalize CNNs efficiently.

Structured Matrices and Equivariance

- **Structured Matrices**: Matrices with patterns (e.g., Toeplitz, Hankel, circulant) that reduce parameters.

Dense Vs. Structured Matrix

a	b	c
d	o	f
g	h	i

Dense Matrix

j	k	l
l	j	k
k	l	j

Toeplitz (structured)

Structured Matrices and Equivariance

- **Structured Matrices**: Matrices with patterns (e.g., Toeplitz, Hankel, circulant) that reduce parameters.
- **Displacement Rank (DR)**: Measures structure via $\nabla_{A,B}(M) = AM - MB$. DR = rank of $\nabla_{A,B}(M)$.

Structured Matrices and Equivariance

- **Structured Matrices**: Matrices with patterns (e.g., Toeplitz, Hankel, circulant) that reduce parameters.
- **Displacement Rank (DR)**: Measures structure via $\nabla_{A,B}(M) = AM - MB$. DR = rank of $\nabla_{A,B}(M)$.
- Low DR preserves structure under operations.

Structured Matrices and Equivariance

- **Structured Matrices**: Matrices with patterns (e.g., Toeplitz, Hankel, circulant) that reduce parameters.
- **Displacement Rank (DR)**: Measures structure via $\nabla_{A,B}(M) = AM - MB$. DR = rank of $\nabla_{A,B}(M)$.
- Low DR preserves structure under operations.
- **Equivariance**: Circulant matrices (used in CNNs) are GMs for cyclic groups, enabling shift equivariance.

Group Matrices

- **Definition**: Matrices encoding group symmetries, $M_{gh} = M_{g'h'}$ if $gh = g'h'$.

Example: Dihedral group matrix (D_3)

e	r	r^2	f	fr	fr^2
r^2	e	r	fr	fr^2	f
r	r^2	e	fr^2	f	fr
f	fr	fr^2	e	r	r^2
fr	fr^2	f	r^2	e	r
fr^2	f	fr	r	r^2	e

Group Matrix (GM)

Group Matrices

- **Definition**: Matrices encoding group symmetries, $M_{gh} = M_{g'h'}$ if $gh = g'h'$.
- **Group Diagonals**: Basis matrices B_g, sparse, with $(B_g)_{h,h'} = \delta(h = gh')$.

Group Matrices

- **Definition**: Matrices encoding group symmetries, $M_{gh} = M_{g'h'}$ if $gh = g'h'$.
- **Group Diagonals**: Basis matrices B_g, sparse, with $(B_g)_{h,h'} = \delta(h = gh')$.
- **Example**: For cyclic group C_3, B_g shifts entries cyclically.

Example: Group Diagonal for C_3

For $C_3 = \{e, a, a^2\}$ with $a^3 = e$, the group diagonal B_a is:

$$B_a = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

where rows and columns are labeled by e, a, a^2.

Group Convolution

- **Standard Convolution**: $\phi \star \psi(x) = \sum_{g \in G} \phi(g)\psi(g^{-1}x)$.

Group Convolution

- **Standard Convolution**: $\phi \star \psi(x) = \sum_{g \in G} \phi(g)\psi(g^{-1}x)$.
- **GM Formulation**: $\mathrm{Conv}_{\psi}\vec{\phi} = \sum_{g \in G} \phi(g) B_g \vec{\psi}$.

Group Convolution

- **Standard Convolution**: $\phi \star \psi(x) = \sum_{g \in G} \phi(g)\psi(g^{-1}x)$.
- **GM Formulation**: $\text{Conv}_{\psi}\vec{\phi} = \sum_{g \in G} \phi(g) B_g \vec{\psi}$.
- **Efficient kernels:** Restrict ϕ to a radius k neighborhood in G's Cayley graph, similar to $k \times k$ CNN kernels.

Group Convolution

- **Standard Convolution**: $\phi \star \psi(x) = \sum_{g \in G} \phi(g)\psi(g^{-1}x)$.
- **GM Formulation**: $\text{Conv}_{\psi}\vec{\phi} = \sum_{g \in G} \phi(g) B_g \vec{\psi}$.
- **Efficient kernels:** Restrict ϕ to a radius k neighborhood in G's Cayley graph, similar to $k \times k$ CNN kernels.
- **Complex groups**: Compute GMs for direct/semi-direct products via Kronecker products, e.g., $B^{G \times H}_{(g,h)} = B^G_g \otimes B^H_h$.

Generalizing CNN Operations

- **Strides**: Restrict output to a subgroup $H \subseteq G$.

Generalizing CNN Operations

- **Strides**: Restrict output to a subgroup $H \subseteq G$.
- **Pooling**: $\mathrm{Pool}_H^{\square}(\psi)(h) = \square\{\psi(g') : g' \in P_h\}$, over cosets.

Generalizing CNN Operations

- **Strides**: Restrict output to a subgroup $H \subseteq G$.
- **Pooling**: $\mathrm{Pool}_H^{\square}(\psi)(h) = \square\{\psi(g') : g' \in P_h\}$, over cosets.
- **Closure**: M^T, MM', $M \otimes N$ remain GMs.

Generalizing CNN Operations

- **Strides**: Restrict output to a subgroup $H \subseteq G$.
- **Pooling**: $\text{Pool}_H^{\square}(\psi)(h) = \square\{\psi(g') : g' \in P_h\}$, over cosets.
- **Closure**: M^T, MM', $M \otimes N$ remain GMs.
- **Extensions**: Applies to homogeneous spaces and infinite groups.

Approximate Equivariance

- **Goal**: Relax strict equivariance for flexibility.

Approximate Equivariance

- **Goal**: Relax strict equivariance for flexibility.
- **Method**: $M = \sum_{g \in G} \mathrm{diag}(F_g) B_g$, with learnable F_g.

Approximate Equivariance

- **Goal**: Relax strict equivariance for flexibility.
- **Method**: $M = \sum_{g \in G} \mathrm{diag}(F_g) B_g$, with learnable F_g.
- **Displacement Operator**: $\mathrm{D}(M) = F(M) - PF(M)$, where P is a permutation. $\mathrm{D}(M) = 0$ for exact GMs.

Approximate Equivariance

- **Goal**: Relax strict equivariance for flexibility.
- **Method**: $M = \sum_{g \in G} \mathrm{diag}(F_g) B_g$, with learnable F_g.
- **Displacement Operator**: $\mathrm{D}(M) = F(M) - PF(M)$, where P is a permutation. $\mathrm{D}(M) = 0$ for exact GMs.
- **Low Displacement Rank**: $\mathrm{rank}(\mathrm{D}(M)) \leq r$ reduces parameters while controlling error.

Experimental Setup

- **Tasks**: Dynamics prediction (Plumes, JetFlow), image classification (MNIST, CIFAR-10).

Experimental Setup

- **Tasks**: Dynamics prediction (Plumes, JetFlow), image classification (MNIST, CIFAR-10).
- **Baselines**: CNN, E2-CNN, RPP, LIFT, structured matrix methods.

Experimental Setup

- **Tasks**: Dynamics prediction (Plumes, JetFlow), image classification (MNIST, CIFAR-10).
- **Baselines**: CNN, E2-CNN, RPP, LIFT, structured matrix methods.
- **Metrics**: RMSE, accuracy, parameter count.

Results: JetFlow

Task	GM-CNN RMSE ↓	Best Baseline RMSE ↓	Param Ratio
Translation (F)	0.15 ± 0.01	0.15 ± 0.00 (Lift)	**1:75**
Rotation (F)	0.17 ± 0.01	0.16 ± 0.01 (RSteer)	**1:32**
Scaling (F)	0.15 ± 0.01	0.14 ± 0.01 (RSteer)	**1:255**

Note: F = Future.

Results: Plumes

Task	GM-CNN RMSE ↓	Best Baseline RMSE ↓	Param Ratio
Translation (F)	0.72 ± 0.01	0.72 ± 0.01 (RGroup)	**1:100**
Rotation (F)	0.92 ± 0.01	0.80 ± 0.01 (RSteer)	**1:50**
Scaling (F)	0.92 ± 0.01	0.82 ± 0.01 (RSteer)	**1:350**

Note: F = Future.

Results: Image Classification

Dataset	GM-CNN Acc (%) ↑ (Params) ↓	Best Baseline Acc (%) (Params)
M-bg-rot	**55.29** (5.9k)	45.81 (14k) [LDR]
M-noise	**90.20** (5.9k)	82.55 (25k) [GM(n=3,E)]
CIFAR-10	**58.31** (14k)	45.33 (18k) [LDR]
SmallNORB	85.22 (25k)	83.23 (14k) [LDR]
Rot-MNIST	**84.50** (15k)	83.78 (49k) [GM(n=3)]
Rect	**99.89** (5.9k)	99.60 (12k) [GM(n=3)]

Note: n = neighborhood radius, E = error addition.

Conclusion

- **Contributions**: GM-CNNs generalize CNN operations using GMs, enabling efficient approximate equivariance.

Conclusion

- **Contributions**: GM-CNNs generalize CNN operations using GMs, enabling efficient approximate equivariance.
- **Impact**: Achieves high accuracy with 1-2 orders fewer parameters, ideal for resource-limited settings.

Conclusion

- **Contributions**: GM-CNNs generalize CNN operations using GMs, enabling efficient approximate equivariance.
- **Impact**: Achieves high accuracy with 1-2 orders fewer parameters, ideal for resource-limited settings.
- **Future Work**: Extend to continuous groups, incorporate steerability, and improve scalability.

Thank You

Questions?

Paper

Code