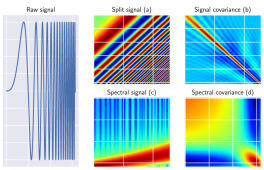
Gyrogroup Batch Normalization

Ziheng Chen[1], Yue Song[2], Xiao-Jun Wu[3], Nicu Sebe[1]

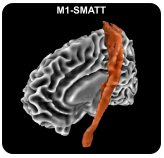
[1]University of Trento [2]Caltech [3]Jiangnan University

ICLR 2025

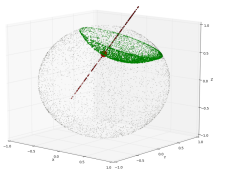
Motivations



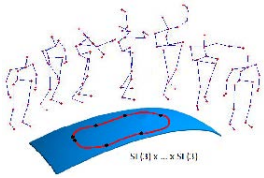
Radar Classification
(Brooks, 2020)



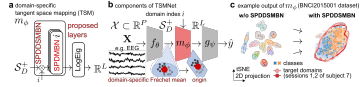
Medical Imaging
(Chakraborty et al., 2020)



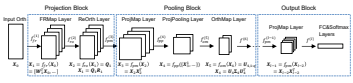
NLP, Graph...
(Ganea et al., 2018)



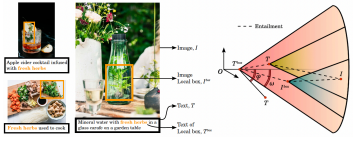
Action Recognition
(Vemulapalli et al., 2014)



EEG
(Kobler et al., 2022a)



GrNet
(Huang et al., 2018)



Vision-Language Models
(Huang et al., 2018)

Overview

Euclidean batch normalization:

$$\forall i \leq N, x_i \leftarrow \gamma \frac{x_i - \mu}{\sqrt{v^2 + \epsilon}} + \beta \qquad (1)$$

- Centering
- Biasing
- Scaling

Table 1: Comparison of previous RBN methods with our GyroBN.

Method	Controllable Statistics	Applied Geometries	Incorporated by GyroBN
SPDBN (Brooks et al., 2019)	M	SPD manifolds under AIM	✓
SPDBN (Kobler et al., 2022b)	M+V	SPD manifolds under AIM	✓
SPDDSMBN (Kobler et al., 2022a)	M+V	SPD manifolds under AIM	✓
ManifoldNorm (Chakraborty, 2020, Algs. 1-2)	N/A	Riemannian homogeneous space	✗
ManifoldNorm (Chakraborty, 2020, Algs. 3-4)	M+V	Matrix Lie groups under the distance $d(X, Y) = \|\text{mlog}(X^{-1}Y)\|$	✓
RBN (Lou et al., 2020, Alg. 2)	N/A	Geodesically complete manifolds	✗
LieBN (Chen et al., 2024a)	M+V	General Lie groups	✓
GyroBN	M+V	Pseudo-reductive gyrogroups with gyro isometric gyrations	N/A

Preliminaries: Gyro Structures

Gyro structure on manifold (Nguyen, 2022a):

$$\text{Gyro addition: } P \oplus Q = \operatorname{Exp}_P \left(\operatorname{PT}_{E \to P} \left(\operatorname{Log}_E(Q) \right) \right), \tag{2}$$

$$\text{Gyro scalar product: } t \odot P = \operatorname{Exp}_E \left(t \operatorname{Log}_E(P) \right), \tag{3}$$

$$\text{Gyro inverse: } \ominus P = -1 \odot P = \operatorname{Exp}_E \left(-\operatorname{Log}_E(P) \right), \tag{4}$$

$$\text{Gyration: } \operatorname{gyr}[P, Q]R = (\ominus(P \oplus Q)) \oplus (P \oplus (Q \oplus R)), \tag{5}$$

$$\text{Gyro inner product: } \langle P, Q \rangle_{\text{gr}} = \langle \operatorname{Log}_E(P), \operatorname{Log}_E(Q) \rangle_E, \tag{6}$$

$$\text{Gyro norm: } \|P\|_{\text{gr}} = \langle P, P \rangle_{\text{gr}}, \tag{7}$$

$$\text{Gyrodistance: } \mathrm{d}_{\text{gry}}(P, Q) = \|\ominus P \oplus Q\|_{\text{gr}}, \tag{8}$$

Insight:

- Axioms of gyrogroups
- Riemannian batch normalization by Gyro addition, scalar product, and distance.

Examples

Table 2: Gyrogroup properties on several geometries.

Geometry	Symbol	$P \oplus Q$ or $x \oplus y$	E	$\ominus P$ or $\ominus x$	Lie group	Gyrogroup	References
AIM $\mathcal{S}^n_{++}$	$\oplus^{\mathrm{AI}}$	$P^{\frac{1}{2}}QP^{\frac{1}{2}}$	I_n	P^{-1}	✗	✓	(Nguyen, 2022b)
LEM $\mathcal{S}^n_{++}$	$\oplus^{\mathrm{LE}}$	$\mathrm{mexp}(\mathrm{mlog}(P)+\mathrm{mlog}(Q))$	I_n	P^{-1}	✓	✓	(Arsigny et al., 2005) (Nguyen, 2022b)
LCM $\mathcal{S}^n_{++}$	$\oplus^{\mathrm{LC}}$	$\psi^{-1}_{\mathrm{LC}}(\psi_{\mathrm{LC}}(P)+\psi_{\mathrm{LC}}(Q))$	I_n	$\psi_{\mathrm{LC}}(-\psi_{\mathrm{LC}}(P))$	✓	✓	(Lin, 2019) (Nguyen, 2022b) (Chen et al., 2024b)
$\widetilde{\mathrm{Gr}}(p,n)$	$\widetilde{\oplus}^{\mathrm{Gr}}$	$\mathrm{mexp}(\Omega)Q\,\mathrm{mexp}(-\Omega)$	$\widetilde{I}_{p,n}$	$\mathrm{mexp}(-\Omega)\widetilde{I}_{p,n}\,\mathrm{mexp}(\Omega)$	✗	Non-reductive	(Nguyen, 2022a)
$\mathrm{Gr}(p,n)$	$\oplus^{\mathrm{Gr}}$	$\mathrm{mexp}(\Omega)V$	$I_{p,n}$	$\mathrm{mexp}(-\Omega)I_{p,n}$			(Nguyen and Yang, 2023)
$\mathcal{M}_K$	$\oplus_K$	$\frac{\left(1-2K\langle x,y\rangle-K\|y\|^2\right)x+\left(1+K\|x\|^2\right)y}{1-2K\langle x,y\rangle+K^2\|x\|^2\|y\|^2}$	0	$-x$	✗(✓for K=0)	✓	(Ungar, 2009) (Ganea et al., 2018) (Skopek et al., 2019)

Isometries

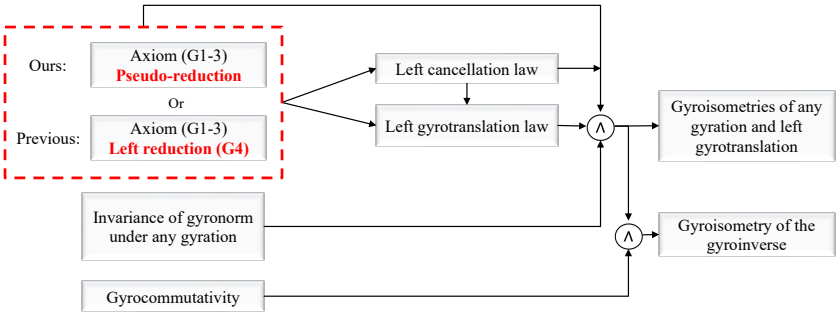


Figure 1: The conceptual comparison of derivation logic of gyroisometries.

Proposition

For every (pseudo-reductive) gyrogroup in Tab. 2, the gyrodistance is identical to the geodesic distance (therefore symmetric). The gyroinverse, any gyration and any left gyrotranslation are gyroisometries.

General Framework

Gyro statistics $\Leftrightarrow$ Riemannian ones: $M = \mathrm{FM}(\{P_i\}) = \underset{Q \in \mathcal{M}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \mathrm{d}^2_{\mathrm{gry}}(P_i, Q)$

Core operations: $x_i \leftarrow \gamma \frac{x_i - \mu}{\sqrt{v^2 + \epsilon}} + \beta \Longrightarrow \tilde{P}_i = \overbrace{B \oplus}^{\text{Biasing}} \left(\overbrace{\frac{s}{\sqrt{v^2 + \epsilon}} \odot}^{\text{Scaling}} \left(\overbrace{\ominus M \oplus P_i}^{\text{Centering}} \right) \right)$

Algorithm 1: Gyrogroup Batch Normalization (GyroBN)

Require : batch of activations $\{P_{1\ldots N} \in \mathcal{M}\}$, small positive constant ϵ, and momentum $\eta \in [0, 1]$, running mean M_r, running variance v_r^2, biasing parameter $B \in \mathcal{M}$, scaling parameter $s \in \mathbb{R}$.

Return : normalized batch $\{\tilde{P}_{1\ldots N} \in \mathcal{M}\}$

if *training* **then**
  Compute batch mean M_b and variance v_b^2 of $\{P_{1\ldots N}\}$;
  Update running statistics $M_r = \mathrm{Bar}_\gamma(M_b, M_r)$, $v_r^2 = \gamma v_b^2 + (1 - \gamma) v_r^2$;
end
$(M, v^2) = (M_b, v_b^2)$ **if** *training* **else** (M_r, v_r^2)
$\forall i \leq N, \tilde{P}_i = B \oplus \left(\frac{s}{\sqrt{v^2 + \epsilon}} \odot (\ominus M \oplus P_i) \right)$

Manifestations

Table 3: Key operators in calculating GyroBN on the Grassmannian and hyperbolic manifolds.

Operator	$\mathrm{Gr}(p,n)$	$\mathbb{P}_K^n$
Identity element	$I_{p,n}$	$0 \in \mathbb{R}^n$
$P \oplus^{\mathrm{Gr}} Q$ or $x \oplus_K y$	$\mathrm{mexp}(\Omega)V$	$\frac{\left(1-2K\langle x,y\rangle-K\|y\|^2\right)x+\left(1+K\|x\|^2\right)y}{1-2K\langle x,y\rangle+K^2\|x\|^2\|y\|^2}$
$\ominus^{\mathrm{Gr}} P$ or $\ominus_K x$	$\mathrm{mexp}(-\Omega)I_{p,n}$	$-x$
$t \odot^{\mathrm{Gr}} P$ or $t \odot_K x$	$\mathrm{mexp}(t\Omega)I_{p,n}$	$\frac{1}{\sqrt{\|K\|}} \tanh\left(t \tanh^{-1}(\sqrt{\|K\|}\|x\|)\right) \frac{x}{\|x\|}$
$\mathrm{Bar}_\gamma^{\mathrm{Gr}}(Q,P)$ or $\mathrm{Bar}_\gamma^{K}(y,x)$	$\mathrm{Exp}_P^{\mathrm{Gr}}(\gamma \mathrm{Log}_P^{\mathrm{Gr}}(Q))$	$x \oplus_K (-x \oplus_K y) \odot_K t$
Fréchet Mean	Karcher Flow (Karcher, 1977)	(Lou et al., 2020, Alg. 1)

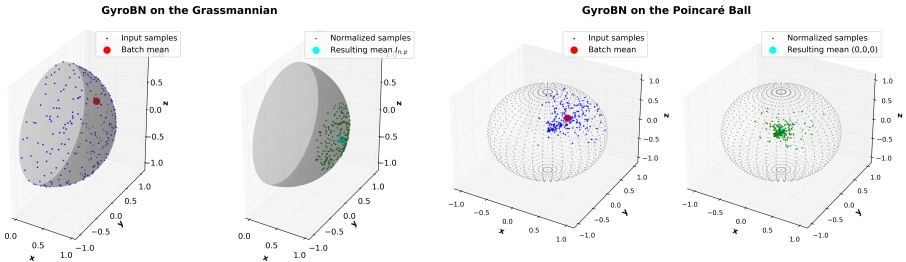


Figure 2: Illustration of GyroBN on the Grassmannian and hyperbolic spaces.

Experiments

Table 4: Comparison of GyroBN against other Grassmannian BNs under GyroGr backbone.

BN	None		ManifoldNorm-Gr		RBN-Gr		GyroBN-Gr	
Acc.	Mean±std	Max	Mean±std	Max	Mean±std	Max	Mean±std	Max
HDM05	48.97 ± 0.24	49.23	49.67 ± 0.76	50.41	48.64 ± 0.77	49.49	$\mathbf{51.89 \pm 0.37}$	**52.43**
NTU60	70.13 ± 0.16	70.32	68.56 ± 0.43	69.14	67.77 ± 0.52	68.35	$\mathbf{72.60 \pm 0.04}$	**72.65**
NTU120	53.76 ± 0.18	53.96	51.41 ± 0.38	51.92	50.56 ± 0.22	50.82	$\mathbf{55.47 \pm 0.10}$	**55.59**

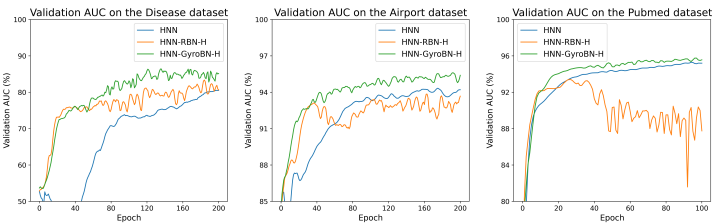


Figure 3: Comparison of HNN with or without GyroBNH or RBN-H.

Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2005). *Fast and simple computations on tensors with log-Euclidean metrics.* PhD thesis, INRIA.

Brooks, D. (2020). *Deep Learning and Information Geometry for Time-Series Classification*. PhD thesis, Sorbonne université.

Brooks, D., Schwander, O., Barbaresco, F., Schneider, J.-Y., and Cord, M. (2019). Riemannian batch normalization for SPD neural networks. In *NeurIPS*.

Chakraborty, R. (2020). ManifoldNorm: Extending normalizations on Riemannian manifolds. *arXiv preprint arXiv:2003.13869*.

Chakraborty, R., Bouza, J., Manton, J. H., and Vemuri, B. C. (2020). Manifoldnet: A deep neural network for manifold-valued data with applications. *IEEE TPAMI*.

Chen, Z., Song, Y., Liu, Y., and Sebe, N. (2024a). A Lie group approach to Riemannian batch normalization. In *ICLR*.

Chen, Z., Song, Y., Xu, T., Huang, Z., Wu, X.-J., and Sebe, N. (2024b). Adaptive Log-Euclidean metrics for SPD matrix learning. *IEEE TIP*.

Ganea, O., Bécigneul, G., and Hofmann, T. (2018). Hyperbolic neural networks. In *NeurIPS*.

Huang, Z., Wu, J., and Van Gool, L. (2018). Building deep networks on Grassmann manifolds. In *AAAI*.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541.

Kobler, R., Hirayama, J.-i., Zhao, Q., and Kawanabe, M. (2022a). SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in EEG. In *NeurIPS*.

Kobler, R. J., Hirayama, J.-i., and Kawanabe, M. (2022b). Controlling the Fréchet variance improves batch normalization on the symmetric positive definite manifold. In *ICASSP*.

Lin, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIMAX*.

Lou, A., Katsman, I., Jiang, Q., Belongie, S., Lim, S.-N., and De Sa, C. (2020). Differentiating through the Fréchet mean. In *ICML*.

Nguyen, X. S. (2022a). The Gyro-structure of some matrix manifolds. In *NeurIPS*.

Nguyen, X. S. (2022b). A Gyrovector space approach for symmetric positive semi-definite matrix learning. In *ECCV*.

Nguyen, X. S. and Yang, S. (2023). Building neural networks on matrix manifolds: A Gyrovector space approach. *arXiv preprint arXiv:2305.04560*.

Skopek, O., Ganea, O.-E., and Bécigneul, G. (2019). Mixed-curvature variational autoencoders. *arXiv preprint arXiv:1911.08411*.

Ungar, A. A. (2009). *A Gyrovector Space Approach to Hyperbolic Geometry*. Springer.

Vemulapalli, R., Arrate, F., and Chellappa, R. (2014). Human action recognition by representing 3D skeletons as points in a Lie group. In *CVPR*.

Thank You