Sheaf Neural Networks on SPD Manifolds: Second-Order Geometric Representation Learning

ICML International Conference On Machine Learning

Yuhan Peng*[1] Junwen Dong*[2] Yuzhi Zeng†[1] Hao Li†[3] Ce Ju[4] Huitao Feng[2] Diaaeldin Taha[5,6] Anna Wienhard[5,6] Kelin Xia[1]

[1]Nanyang Technological University [2]Chern Institute of Mathematics, Nankai University [3]National University of Singapore [4]Inria, CEA, Uni versit ´e Paris-Saclay
[5]Max Planck Institute for Mathematics in the Sciences [6]Center for Scalable Data Analytics and Artificial Intelligence

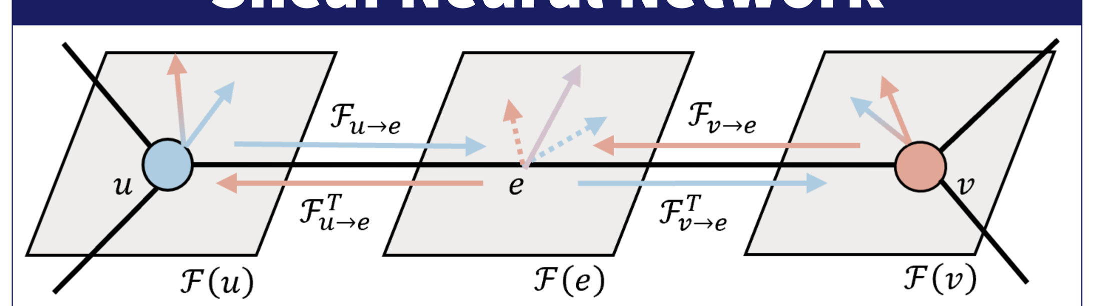
Contacts: xiakelin@ntu.edu.sg | anna.wienhard@mis.mpg.de | diaaeldin.taha@mis.mpg.de

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE · NUS National University of Singapore · 陈省身数学研究所 Chern Institute of Mathematics · Inria · ScaDS.AI DRESDEN LEIPZIG · MAX PLANCK INSTITUTE FOR MATHEMATICS

Motivation

- **Why sheaf?** Standard GNNs share one information-combination rule across all edges; sheaves make it **edge-specific**.
- **Why SPD stalks?** Existing Sheaf NNs use vector stalks, but second-order geometry is **matrix-valued** → SPD_n **(symmetric positive definite manifold)**.
- **Our solution**: First sheaf neural network native on SPD_n with Lie group structure → well-posed operators & **strictly more expressive**.
- **Why is it hard?** SPD_n has **no** vector subtraction, Euclidean adjoint, or classical sheaf Laplacian.
- **Core problem:** How do we construct SPD counterparts of $\delta, \delta^T, \boldsymbol{L}$ that play the same role as classical sheaf neural network operators?

Sheaf Neural Network



Cellular sheaf:
- Vector space (**stalk**) $\mathcal{F}(u)$ and $\mathcal{F}(v)$
- Linear **restriction map** $\mathcal{F}_{v\to e}: \mathcal{F}(u) \to \mathcal{F}(e)$

Coboundary operator measures *local inconsistency*: $\delta(\mathbf{x})_e = \mathcal{F}_{v\to e}\mathbf{x}_v - \mathcal{F}_{u\to e}\mathbf{x}_u$

Sheaf Laplacian: $\boldsymbol{L}_{\mathcal{F}} = \delta^T\delta$

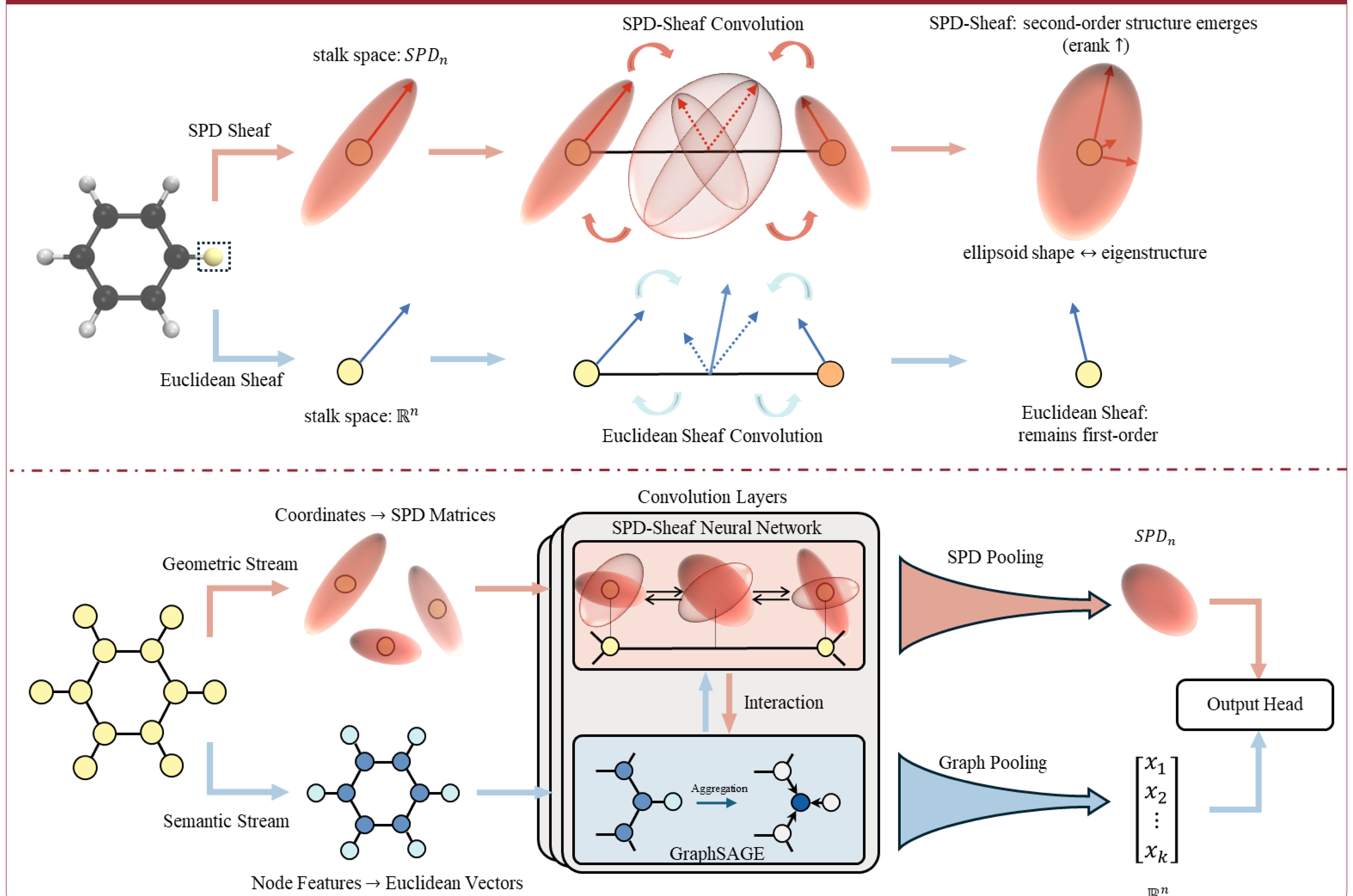
Sheaf neural network:

$$H^{(l+1)} = \sigma((I - \boldsymbol{L}_{\mathcal{F}})H^{(l)}\boldsymbol{W}^{(l)})$$

Lie Group Structure on SPD

$P \odot Q := \exp(\log P + \log Q)$ forms an **Abelian lie group** with identity I_n and inverse $P^{-1} = \exp(-\log P)$

Method Overview



Experiments on MoleculeNet

Model	BACE	BBBP	ClinTox	SIDER	Tox21	HIV	MUV
No. molecules	1,513	2,039	1,478	1,427	7,831	41,127	93,808
No. avg atoms	65	46	50.58	65	36	46	43
No. tasks	1	1	2	27	12	1	17
D-MPNN	$80.9_{0.6}$	$71.0_{0.3}$	$90.6_{0.7}$	$57.0_{0.7}$	$75.9_{0.7}$	$77.1_{0.5}$	$78.6_{1.4}$
AttentiveFP	$78.4_{2.2}$	$66.3_{1.8}$	$84.7_{0.3}$	$60.6_{3.2}$	$78.1_{0.5}$	$75.7_{1.4}$	$78.6_{1.5}$
N-GramRF	$77.9_{1.5}$	$69.7_{0.6}$	$77.5_{4.0}$	$66.8_{0.7}$	$74.3_{0.9}$	$77.2_{0.4}$	$76.9_{0.2}$
N-GramXGB	$79.1_{1.3}$	$69.1_{0.8}$	$87.5_{2.7}$	$65.5_{0.7}$	$75.8_{0.9}$	$78.7_{0.4}$	$74.8_{0.2}$
PretrainGNN	$84.5_{0.7}$	$68.7_{1.3}$	$72.6_{1.5}$	$62.7_{0.8}$	$78.1_{0.6}$	$79.9_{0.7}$	$81.3_{2.1}$
$GROVE_{base}$	$82.1_{0.7}$	$70.0_{0.1}$	$81.2_{3.0}$	$64.8_{0.6}$	$74.3_{0.1}$	$62.5_{0.9}$	$67.3_{1.8}$
$GROVE_{large}$	$81.0_{1.4}$	$69.5_{0.1}$	$76.2_{3.7}$	$65.4_{0.1}$	$73.5_{0.1}$	$68.2_{1.1}$	$67.3_{1.8}$
GraphMVP	$81.2_{0.9}$	$72.4_{1.6}$	$79.1_{2.8}$	$63.9_{1.2}$	$75.9_{0.5}$	$77.0_{1.2}$	$77.7_{0.6}$
MolCLR	$82.4_{0.9}$	$72.2_{2.1}$	$91.2_{3.5}$	$58.9_{1.4}$	$75.0_{0.2}$	$78.1_{0.5}$	$79.6_{1.9}$
GEM	$85.6_{1.1}$	$72.4_{0.4}$	$90.1_{1.3}$	$67.2_{0.4}$	$78.1_{0.1}$	$80.6_{0.9}$	$81.7_{0.5}$
Mol-GDL	$86.3_{1.9}$	$72.8_{1.9}$	$96.6_{0.2}$	$83.1_{0.2}$	$79.4_{0.5}$	$80.8_{0.7}$	$67.5_{1.4}$
Uni-Mol	$85.7_{0.2}$	$72.9_{0.6}$	$91.9_{1.8}$	$65.9_{1.3}$	$79.6_{0.5}$	$80.8_{0.7}$	$82.1_{1.3}$
SMPT	$\underline{87.3}_{1.5}$	$\underline{73.4}_{0.3}$	$92.7_{0.2}$	$67.6_{5.0}$	$\underline{79.7}_{0.1}$	$\mathbf{81.2}_{0.1}$	$\underline{82.2}_{0.8}$
SchNet	$73.7_{0.9}$	$70.8_{1.4}$	$97.8_{0.3}$	$83.1_{0.0}$	$72.4_{0.2}$	$68.8_{0.1}$	$68.2_{0.6}$
EGNN	$79.4_{2.3}$	$72.1_{2.0}$	$\underline{98.6}_{0.5}$	$\underline{83.2}_{0.1}$	$74.5_{0.3}$	$70.4_{0.2}$	$68.6_{0.5}$
SPD-Sheaf (Ours)	$\mathbf{89.0}_{1.4}$	$\mathbf{77.4}_{2.7}$	$\mathbf{99.4}_{0.2}$	$\mathbf{84.3}_{0.4}$	$\mathbf{80.1}_{0.7}$	$\underline{80.9}_{0.7}$	$\mathbf{82.3}_{1.4}$

Geometric Module	BBBP	BACE	ClinTox	SIDER	Tox21	HIV	MUV
GCN (baseline)	$73.9_{1.9}$	$87.0_{1.0}$	$98.4_{0.2}$	$82.6_{0.1}$	$78.7_{0.3}$	$80.4_{1.0}$	$75.7_{3.0}$
Effect of sheaf structure (Euclidean stalks):							
Euclidean SheafNN (Barbero et al., 2022)	$74.1_{1.8}$	$87.5_{0.9}$	$99.0_{0.4}$	$83.9_{0.2}$	$79.4_{0.6}$	$80.9_{0.7}$	$75.3^{\dagger}_{1.1}$
Effect of SPD geometry (no sheaf structure):							
SPD4GNN (Zhao et al., 2023)	$71.9_{2.8}$	$87.5_{0.9}$	$98.7_{0.4}$	$83.9_{0.3}$	$79.3_{0.8}$	$79.6_{1.2}$	$76.2_{4.6}$
HyperbolicGCN (Chami et al., 2019)	$73.3_{2.3}$	$87.3_{0.8}$	$98.6_{0.6}$	$84.1_{0.2}$	$79.3_{0.9}$	$80.7_{1.2}$	$79.2_{2.6}$
Combined: SPD geometry + sheaf structure:							
SPD Sheaf (Ours)	$\mathbf{77.4}_{2.7}$	$\mathbf{89.0}_{1.4}$	$\mathbf{99.4}_{0.2}$	$\mathbf{84.3}_{0.4}$	$\mathbf{80.1}_{0.7}$	$\mathbf{80.9}_{0.7}$	$\mathbf{82.3}_{1.4}$

Variant	BBBP	BACE	ClinTox	SIDER	Tox21	HIV	MUV
Full model	$77.4_{2.7}$	$89.0_{1.4}$	$99.4_{0.2}$	$84.3_{0.4}$	$80.1_{0.7}$	$80.9_{0.7}$	$82.3_{1.4}$
w/o Semantic stream	$71.8_{2.0}$	$71.6_{4.1}$	$97.9_{0.6}$	$83.2_{0.5}$	$68.0_{0.7}$	$66.2_{1.5}$	$58.4_{0.9}$
Δ	−5.6	−17.4	−1.5	−1.1	−12.1	−14.7	−23.9
w/o Geometric stream	$73.3_{3.2}$	$82.1_{1.4}$	$98.1_{0.7}$	$84.3_{0.2}$	$76.3_{0.3}$	$74.8_{1.4}$	$69.8_{0.5}$
Δ	−4.1	−6.9	−1.3	0.0	−3.8	−6.1	−12.5
w/o Cross-modal interaction	$75.2_{0.6}$	$86.9_{1.3}$	$99.4_{0.3}$	$84.3_{0.3}$	$79.8_{0.8}$	$80.3_{0.6}$	$80.8_{2.0}$
Δ	−2.2	−2.1	0.0	0.0	−0.3	−0.6	−1.5

Paper:

Group:

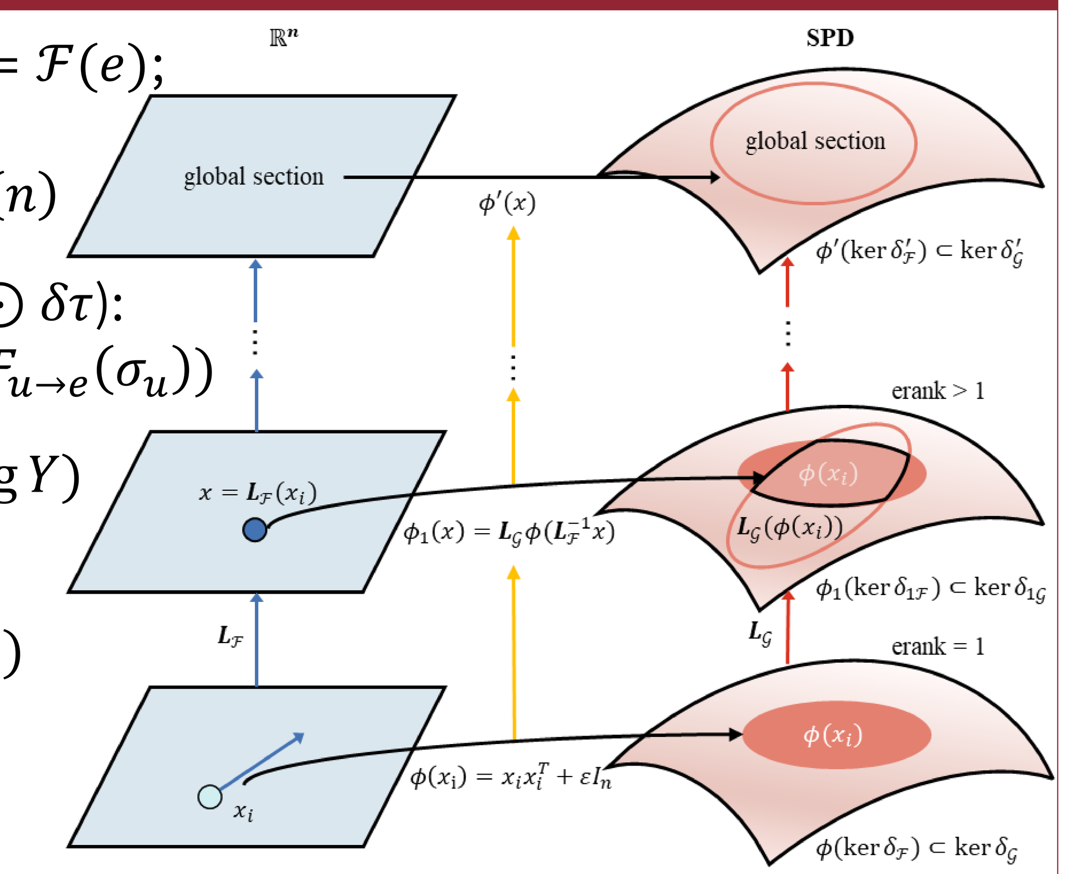
SPD-Valued Sheaf Theory

SPD sheaf: SPD-valued stalk $\mathcal{F}(v) = \mathcal{F}(e)$; restriction map (**isometry**) : $\mathcal{F}_{v\to e}(P) = M_{ve}PM_{ve}^T$ with $M_{ve} \in O(n)$

SPD coboundary ($\delta(\sigma \odot \tau) = \delta\sigma \odot \delta\tau$): $(\delta\sigma)_e = \exp(\log(\mathcal{F}_{v\to e}(\sigma_v) - log(\mathcal{F}_{u\to e}(\sigma_u))$

SPD Pairing: $\langle X, Y\rangle := g_{I_n}(\log X, \log Y)$

Adjoint Operator: $(\delta^T\tau)_v = \exp(\sum_{v\to e}(-1)^{I(v,e)}M_{ve}^T(\log\tau_e)M_{ve})$ satisfying $\langle\delta\sigma, \tau\rangle = \langle\sigma, \delta^T\tau\rangle$



SPD Sheaf Laplacian: $\boldsymbol{L}_{\mathcal{F}} = \delta^T\delta$

$$(\boldsymbol{L}_{\mathcal{F}}X)_v = \exp(\sum_{(v,u)=e}(-1)^{I(v,e)}M_{ve}^T(\log(\mathcal{F}_{v\to e}X_v) - \log(\mathcal{F}_{u\to e}X_u))M_{ve})$$

Hodge-type Decomposition for SPD-valued Sheaves: $\ker\boldsymbol{L}_{\mathcal{F}} = \ker\delta$

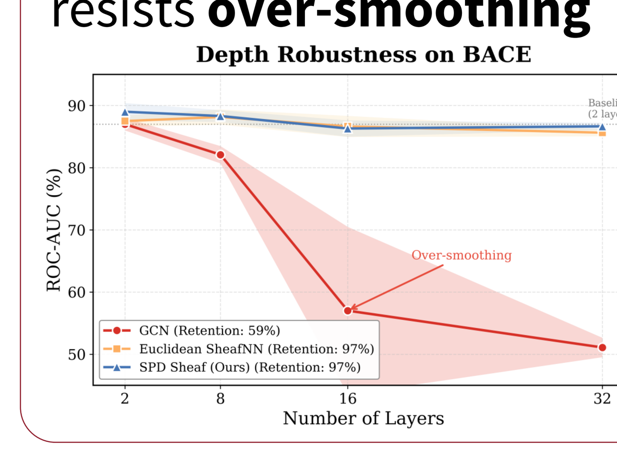
Euclidean-to-SPD Embedding: $\Phi_\varepsilon(x) = xx^{\mathrm{T}} + \varepsilon I_n$

Strict generalization: $\phi'(\ker \boldsymbol{L}_{euclidean}) \subsetneq \ker(\boldsymbol{L}_{SPD})$

SPD-Sheaf convolution: $X_v^{(l+1)} = \sigma(X_v^{(l)} \odot (\boldsymbol{L}_{\mathcal{F}}^{(l)}X^{(l)})_v)$

Analysis

Depth robustness: 97% performance at 32 layers → resists **over-smoothing**



Second-order emergence: SPD-Sheaf lifts erank-1 SPD matrix to full-erank, **boundary (first-order) → interior (second-order)** of SPD_n

Dataset	Layers	Initial Erank	Final Erank	Δ Erank	λ_2: Init→Final
BACE	2	1.00	2.28	+1.28	0.00 → 1.16
BBBP	2	1.00	2.16	+1.16	0.00 → 1.15
SIDER	2	1.00	2.22	+1.22	0.00 → 1.14
ClinTox	5	1.00	1.83	+0.83	0.00 → 1.32
HIV	5	1.00	1.87	+0.87	0.00 → 1.42
Tox21	7	1.00	1.75	+0.75	0.00 → 1.46
MUV	7	1.00	1.72	+0.72	0.00 → 1.55

Acknowledgements

This work was supported in part by the Singapore Ministry of Education Academic Research Fund Tier 1 (RG16/23) and Tier 2 (MOE-T2EP20125-0004, MOE-T2EP50223 0036). It was further supported by the National Key R&D Program of China (Grant No. 2024YFA1013202), the National Natural Science Foundation of China (Grant No. 11221091), the Fundamental Research Funds for the Central Universities, and the Nankai Zhide Foundation. A.W. thanks the Hector Fellow Academy for support.