Vector-valued Distance and Gyrocalculus on the Space of SPD Matrices

Federico López
Beatrice Pozzetti
Steve Trettel
Michael Strube
Anna Wienhard

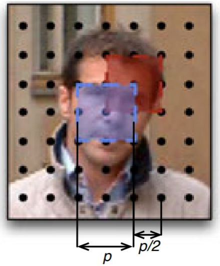
HITS Heidelberg Institute for Theoretical Studies
UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386
Stanford University

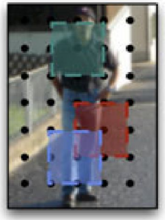
SPD Matrices FTW!

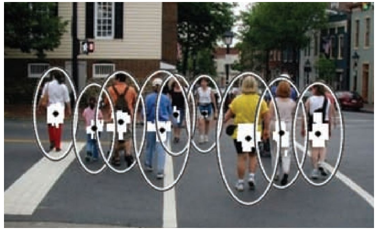
- SPDs have been used for:
  - Pedestrian detection
  - Action and face recognition
  - Image classification
  - Visual tracking
  - Medical image analysis

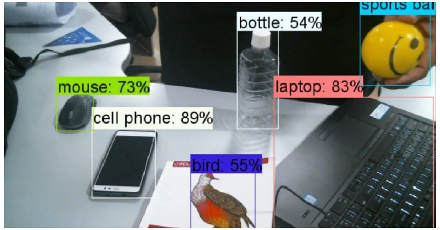
- They capture statistical notions
  - Gaussian distributions
  - Covariance

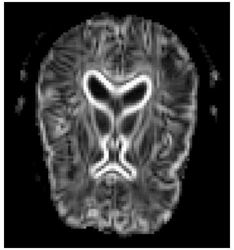
- Convenient trade-off between **structural richness** and **computational tractability**

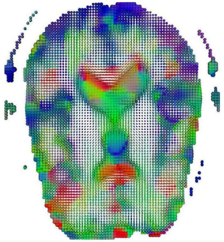


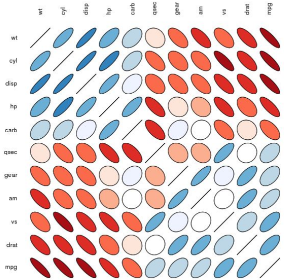


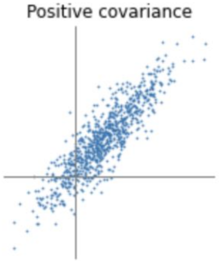


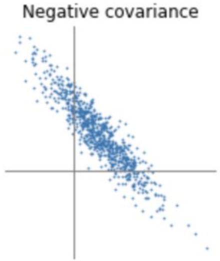


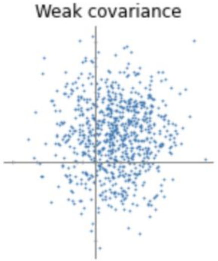










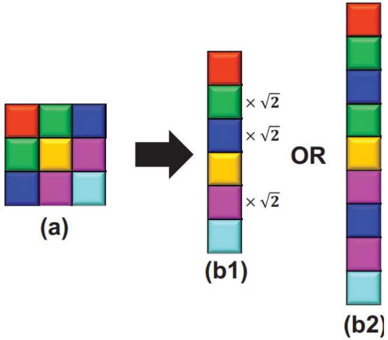


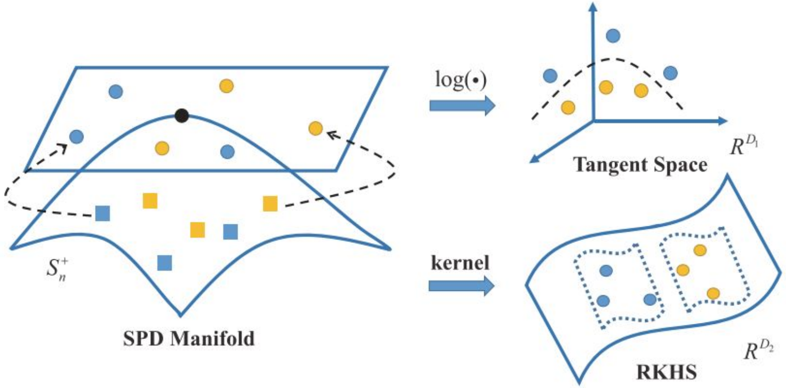
Previous Work

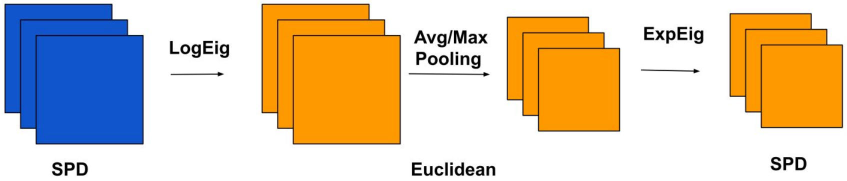
To operate with SPD matrices, previous work:

- Maps matrices to vectors
- Projects points into their tangent space
- Embedding the manifold into high-dimensional Hilbert spaces

These methods **distort the geometrical structure** of the manifold



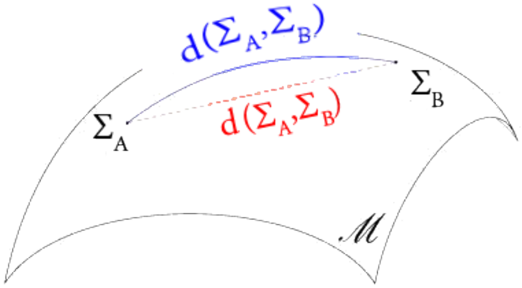




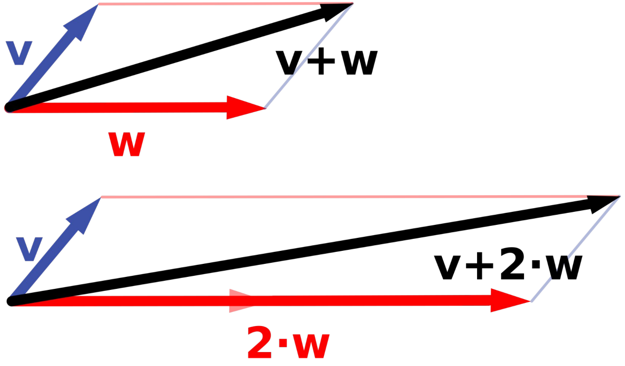
Previous Work

- Several distances proposed
  - Affine invariant metric
  - Stein metric
  - Bures–Wasserstein metric
  - Log-Euclidean Metric

  They do not fully exploit the representational power of SPD

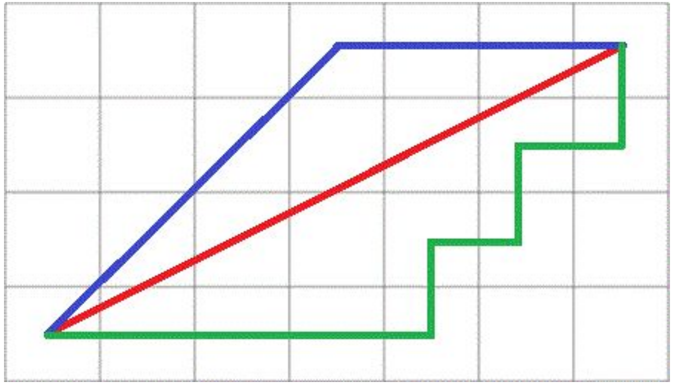


- Growing need to generalize basic operations into the SPD space
  - Hard to translate operations given the lack of closed-form expressions

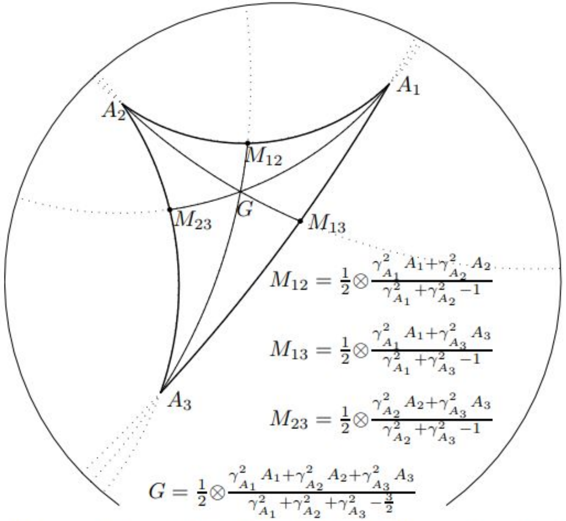


We Propose!

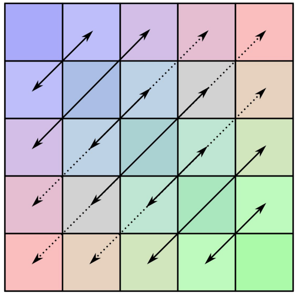
- Vector-valued distance function in SPD
  - Compute only one vector and derive multiple distances
  - Much more information than just the distance
    - Analysis and visualization tool



- Gyrocalculus on SPD
  - Arithmetic operations in the space
    - Addition
    - Scaling
    - Rotations
    - Reflections

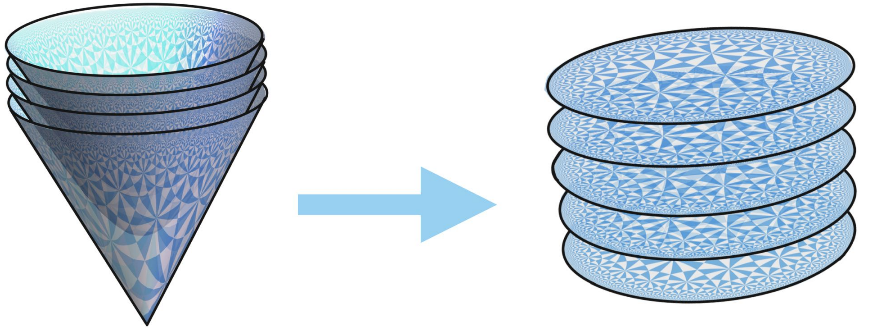


The Space SPD

$Z =$ 

- Points are positive definite real symmetric **n × n** matrices

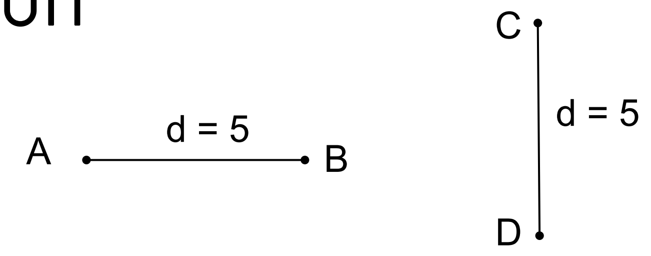
- SPD_n: Riemannian manifold of non-positive curvature of **n(n + 1)/2** dimensions
  - n-dimensional Euclidean subspaces
  - n−1 dimensional hyperbolic subspaces
  - $\lfloor \frac{n}{2} \rfloor$ hyperbolic planes



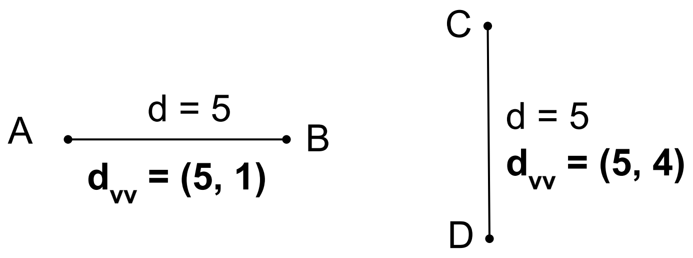
- They **combine hyperbolic and Euclidean geometry** thus they can accommodate:
  - Hierarchical structures -> hyperbolic subspaces
  - Flat structures -> Euclidean subspaces

Vector-valued Distance Function

- In Euclidean or hyperbolic spaces the only invariant of two points is their distance



- In the SPD_n space the invariant between two points is an **n**-dims distance vector



- To assign this vector we employ the vector-valued distance (VVD) function:

$$d_{vv}: \ \mathrm{SPD}_n \times \mathrm{SPD}_n \ \ \rightarrow \ \ \mathbb{R}^n$$

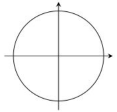
$$d_{vv}(P, Q) = \log(\lambda_1(P^{-1}Q), \ldots, \lambda_n(P^{-1}Q))$$

Advantages of the VVD

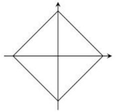
- By taking different norms on the VVD we can compute **several metrics**
  - Riemannian distance
  - Finsler distances
  - Generalize previous metrics

“One vector to rule them all”

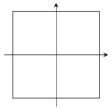
$$d^{R}(Z_1, Z_2) := \sqrt{\sum_{i=1}^{n} v_i^2}$$



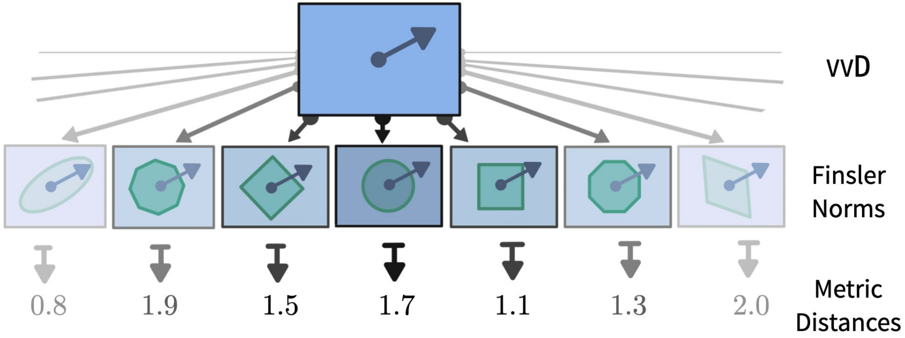
$$d^{F1}(Z_1, Z_2) := \sum_{i=1}^{n} v_i$$



$$d^{F\infty}(Z_1, Z_2) := \max\{v_i\}$$



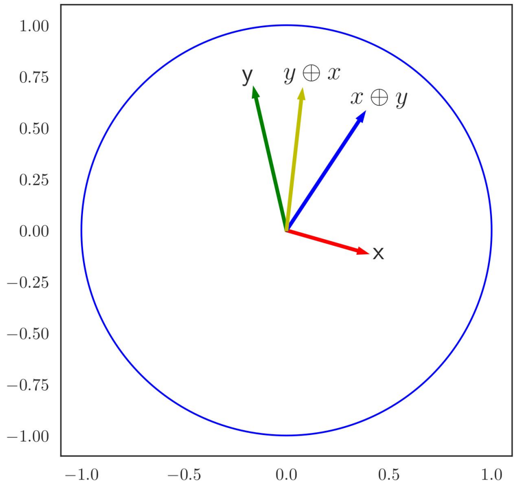
- VVD provides **much more information** than just the distance
  - Read the regularity of the geodesics joining two points
  - **Visualize and analyze** high-dimensional embeddings

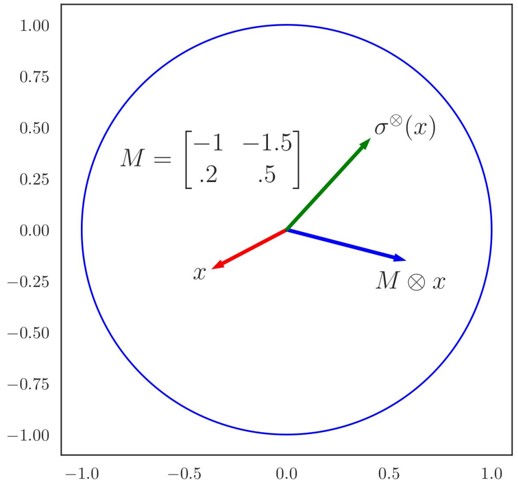


Gyrocalculus

- It is an algebraic formalisms to translate Euclidean operations to other spaces in a geometrically meaningful way

- Successful applications with Hyperbolic geometry [Ganea et al, 2018]
  - Addition
  - Matrix-vector multiplication
  - Pointwise non-linearities

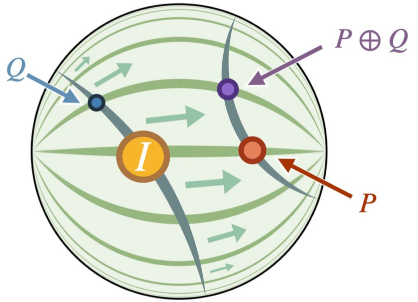




Gyrocalculus in SPD

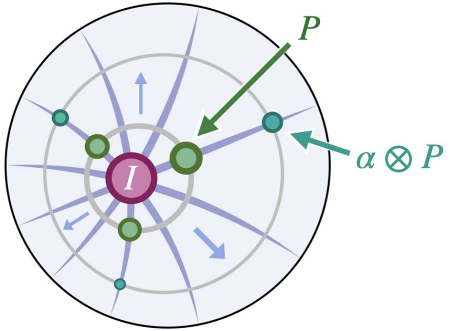
- Addition / subtraction

$$P \oplus Q = \sqrt{P} Q \sqrt{P}$$
$$\ominus P = P^{-1}$$



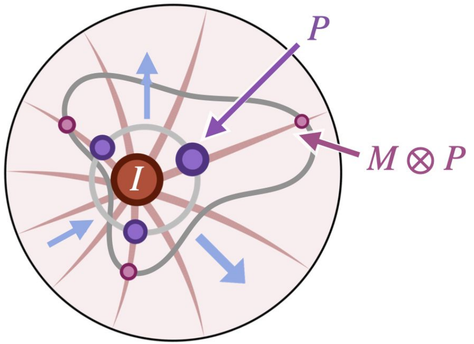
- Scalar multiplication

$$\alpha \otimes P = P^{\alpha} = \exp(\alpha \log(P))$$



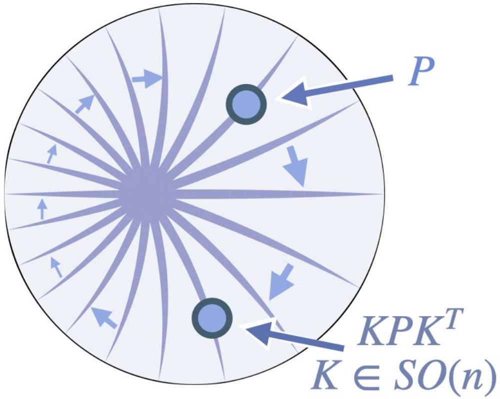
- Matrix scaling

$$A \otimes P = \exp(A \odot \log(P))$$



Isometries

- We also provide a way to learn rotations and reflections in SPD

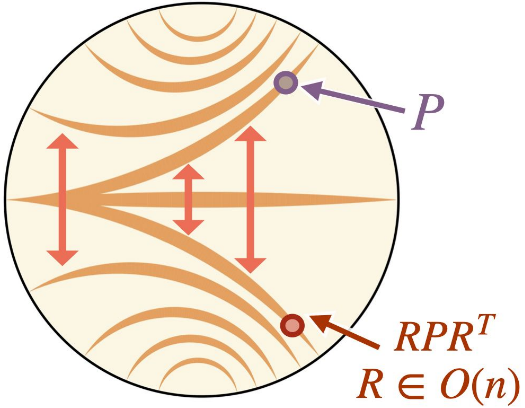


- Rotations

$$\mathrm{Rot}(\vec{\theta}) = \prod_{i<j} R^{+}_{ij}(\theta_{ij})$$

- Reflections

$$\mathrm{Refl}(\vec{\theta}) = \prod_{i<j} R^{-}_{ij}(\theta_{ij})$$



Knowledge Graph Completion

- Operations
  - Scaling

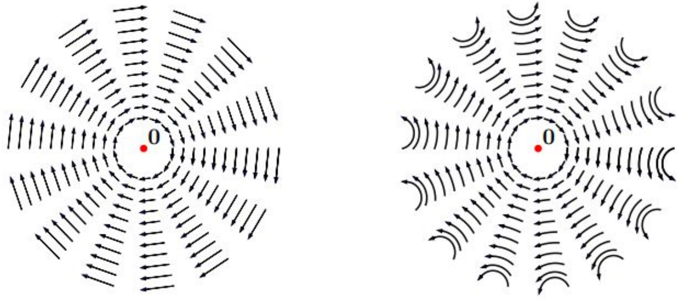
$$\phi(h,r,t) = -d((\mathbf{M}_r \otimes \mathbf{H}) \oplus \mathbf{R}, \mathbf{T})^2 + b_h + b_t$$

  - Rotations
  - Reflections

$$\phi(h,r,t) = -d((\mathbf{M}_r \circledcirc \mathbf{H}) \oplus \mathbf{R}, \mathbf{T})^2 + b_h + b_t$$

- Distances
  - Riemannian
  - Finsler One

- Baselines: scaling, rotations and reflections
  - Euclidean space
  - Hyperbolic spaces
  - Complex space



Results for Knowledge Graph Completion

		WN18RR				FB15k-237			
Operation	Model	MRR	HR@1	HR@3	HR@10	MRR	HR@1	HR@3	HR@10
Scaling	MuRE	47.5	43.6	48.7	55.4	33.6	24.5	37.0	52.1
	MuRP	48.1	**44.0**	49.5	56.6	33.5	24.3	36.7	51.8
	$\text{SPD}^{R}_{\text{Sca}}$	48.1	43.1	50.1	57.6	**34.5**	**25.1**	**38.0**	**53.5**
	$\text{SPD}^{F_1}_{\text{Sca}}$	**48.4**	42.6	**51.0**	**59.0**	32.9	23.6	36.3	51.5

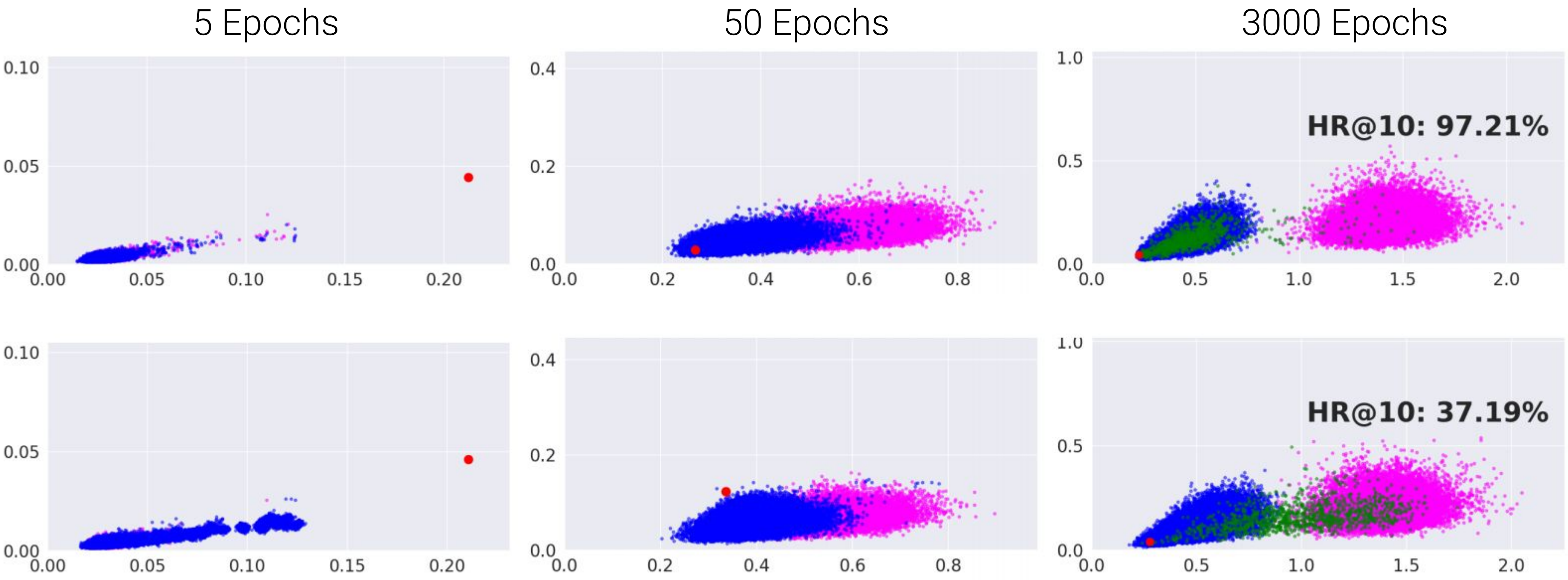
- SPD scaling **outperforms** Euclidean and hyperbolic models

Results for Knowledge Graph Completion

Operation	Model	WN18RR MRR	WN18RR HR@1	WN18RR HR@3	WN18RR HR@10	FB15k-237 MRR	FB15k-237 HR@1	FB15k-237 HR@3	FB15k-237 HR@10
Scaling	MuRE	47.5	43.6	48.7	55.4	33.6	24.5	37.0	52.1
	MuRP	48.1	**44.0**	49.5	56.6	33.5	24.3	36.7	51.8
	$\text{SPD}^{R}_{\text{Sca}}$	48.1	43.1	50.1	57.6	**34.5**	**25.1**	**38.0**	**53.5**
	$\text{SPD}^{F_1}_{\text{Sca}}$	**48.4**	42.6	**51.0**	**59.0**	32.9	23.6	36.3	51.5
Rotations	RotC	47.6	42.8	49.2	57.1	33.8	24.1	37.5	53.3
	RotE	49.4	44.6	51.2	58.5	**34.6**	**25.1**	**38.1**	**53.8**
	RotH	**49.6**	**44.9**	**51.4**	**58.6**	34.4	24.6	38.0	53.5
	$\text{SPD}^{R}_{\text{Rot}}$	46.2	39.7	49.6	57.8	32.9	23.6	36.3	51.6
	$\text{SPD}^{F_1}_{\text{Rot}}$	40.9	30.5	48.2	57.3	32.1	22.9	35.4	50.5
Reflections	RefE	47.3	43.0	48.5	56.1	**35.1**	**25.6**	**39.0**	**54.1**
	RefH	46.1	40.4	48.5	56.8	34.6	25.2	38.3	53.6
	$\text{SPD}^{R}_{\text{Ref}}$	48.3	44.0	49.7	56.7	32.5	23.4	35.6	51.0
	$\text{SPD}^{F_1}_{\text{Ref}}$	**48.7**	**44.3**	**50.1**	**57.4**	31.6	22.5	34.6	50.0

- Competitive results for rotations and reflections with **significantly less dimensions**
- In many cases **Finsler one metric** outperforms the Riemannian distance

Visualization of Embeddings

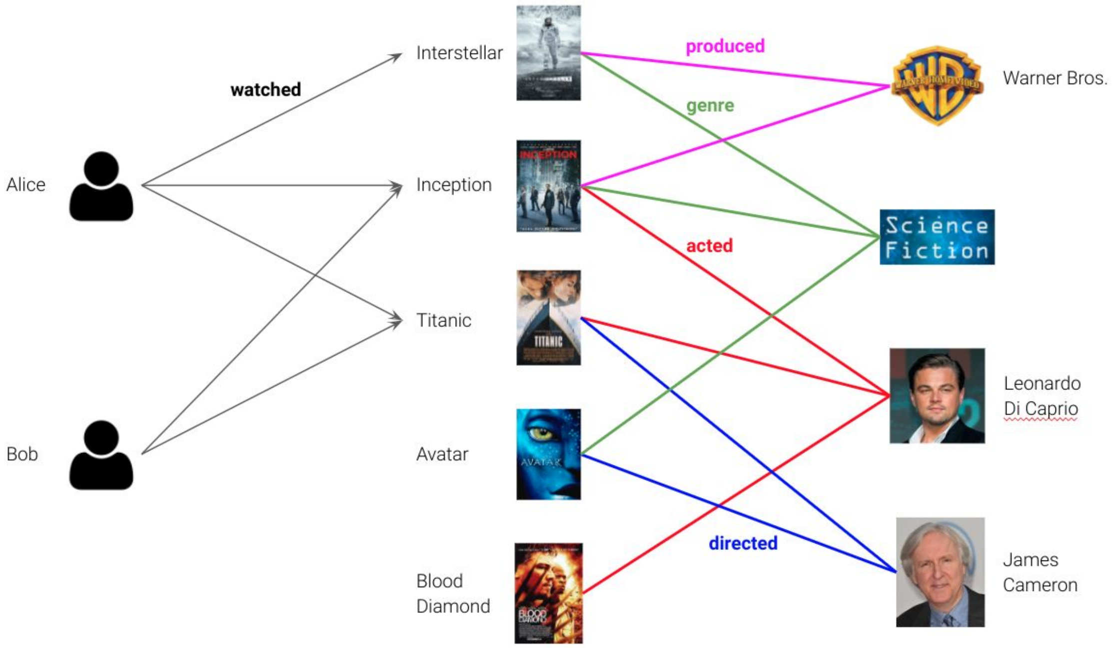


- **Train**, **negative** and **validation** triples for WN18RR relationships
- Position of **validation** triples directly correlates with performance

Knowledge Graph-based Recommender Systems

- Recommendation problem as a link prediction task over users and items enhanced with side information

- Same models than for Knowledge graph completion



Results for KG Recommender Systems

Model	Software		Luxury		Pantry		MindReader	
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
TransE	28.5±0.1	47.2±0.5	35.6±0.1	52.3±0.1	16.6±0.0	35.3±0.1	19.1±0.4	37.6±0.1
RotC	28.5±0.3	45.4±1.4	33.0±0.1	49.8±0.2	14.5±0.0	31.3±0.2	25.3±0.3	50.3±0.6
MuRE	29.4±0.4	47.1±0.4	35.6±0.7	54.0±0.3	19.4±0.1	39.5±0.2	25.2±0.3	49.9±0.6
MuRP	29.6±0.3	47.9±0.3	**37.5±0.1**	**55.2±0.3**	19.4±0.1	39.8±0.2	25.3±0.3	49.3±0.2
$\mathrm{SPD}^{R}_{\mathrm{Sca}}$	29.4±0.4	48.1±0.8	**37.5±0.2**	55.1±0.2	19.5±0.0	39.6±0.3	25.4±0.1	49.8±0.3
$\mathrm{SPD}^{F_1}_{\mathrm{Sca}}$	28.8±0.1	46.9±0.5	37.3±0.3	54.1±0.9	19.0±0.1	38.8±0.2	**25.7±0.5**	49.5±0.1
$\mathrm{SPD}^{R}_{\mathrm{Rot}}$	**30.3±0.2**	48.6±0.9	37.2±0.1	54.8±0.4	**20.0±0.1**	**40.3±0.1**	25.3±0.0	**50.5±0.3**
$\mathrm{SPD}^{F_1}_{\mathrm{Rot}}$	30.1±0.1	**49.1±0.3**	36.9±0.1	54.5±0.6	19.2±0.0	39.3±0.1	**25.7±0.0**	49.5±0.2
$\mathrm{SPD}^{R}_{\mathrm{Ref}}$	29.6±0.2	48.0±0.5	37.3±0.2	55.0±0.2	19.3±0.0	39.7±0.3	25.3±0.0	49.1±0.1
$\mathrm{SPD}^{F_1}_{\mathrm{Ref}}$	29.3±0.1	47.5±0.6	36.8±0.0	54.8±0.1	18.6±0.2	38.3±0.3	24.8±0.2	47.9±1.8

- SPD models **tie or outperform baselines** in all cases
- **Rotations**, both in Riemannian and Finsler metrics, seem to be the most effective operator

Linear Layers on SPD

- Linear layer: feature transformation + bias addition
  - Scaling
  - Rotation
  - Reflection

$$y = Wx + b$$

$$y = W \otimes x \oplus b$$

- Experiments on Question Answering
  - Training word embeddings on SPD

- Model
  - Model question/answer as word embeddings summation followed by a linear transformation

$$\mathbf{Q} = T(\bigoplus_{i=1}^{n} t_i^q) \oplus B$$

  - Rank question-answer similarity

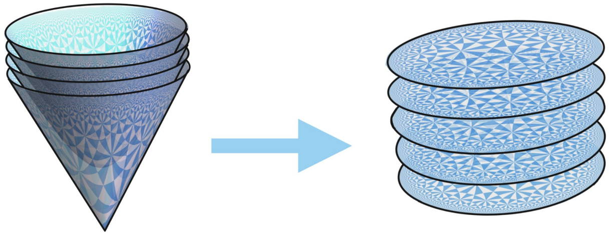
$$\text{sim}(q, a) = -w_f d(\mathbf{Q}, \mathbf{A}) + w_b$$

Results for Question Answering

Model	TrecQA		WikiQA	
	MRR	H@1	MRR	H@1
Euclidean	55.9±2.0	41.0±2.0	43.4±0.3	22.4±1.1
Hyperbolic	58.0±1.3	39.3±2.0	44.0±0.4	22.8±0.6
$\mathrm{SPD}^{R}_{\mathrm{Sca}}$	55.4±0.1	37.1±0.1	**45.5±0.5**	24.4±1.1
$\mathrm{SPD}^{F_1}_{\mathrm{Sca}}$	57.1±0.7	38.6±0.2	44.8±0.5	24.0±0.6
$\mathrm{SPD}^{R}_{\mathrm{Rot}}$	58.7±1.5	41.4±2.9	44.6±0.6	23.6±0.6
$\mathrm{SPD}^{F_1}_{\mathrm{Rot}}$	58.1±0.5	**43.6±1.0**	43.7±0.4	23.8±0.8
$\mathrm{SPD}^{R}_{\mathrm{Ref}}$	57.3±0.3	40.7±1.1	43.9±0.7	23.4±2.0
$\mathrm{SPD}^{F_1}_{\mathrm{Ref}}$	**59.6±0.5**	42.1±1.0	44.7±1.2	**25.0±2.5**

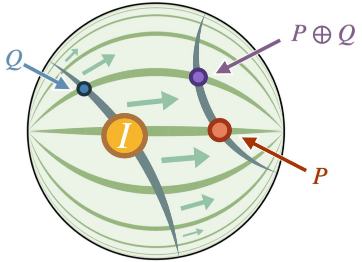
- SPD models **outperform baselines** in all cases
- **Embeddings in SPD manifolds** exploited for downstream tasks
- We showcase how to build **linear layers on SPD**

Summary

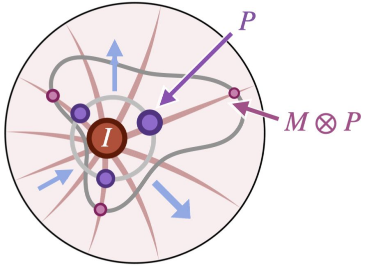


- Growing need to **generalize tools** for SPD manifolds

- Introduce the **vector-valued distance for SPD**
  - Riemannian, Finsler and more distances
  - Visualization and analysis tools



- **Gyrocalculus** on SPD
  - Arithmetic operations that respect the geometry of the space



- Experiments on three tasks and eight datasets
  - **Versatility** of the approach for different types of data
  - **Ease of integration** with downstream tasks
  - Reflect the **superior expressivity** of SPD