Geometric Deep Learning meets BCI to advance inter-session and -subject transfer

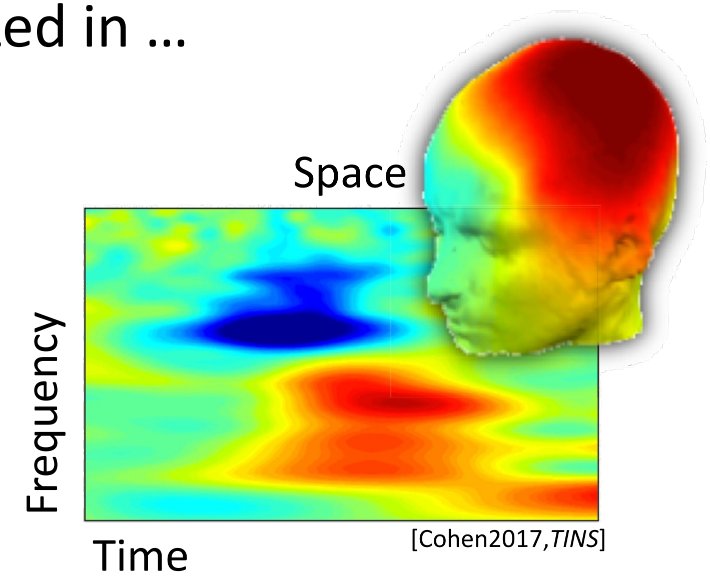
Reinmar J. KOBLER

ATR, Japan
RIKEN AIP, Japan

2023/10/17

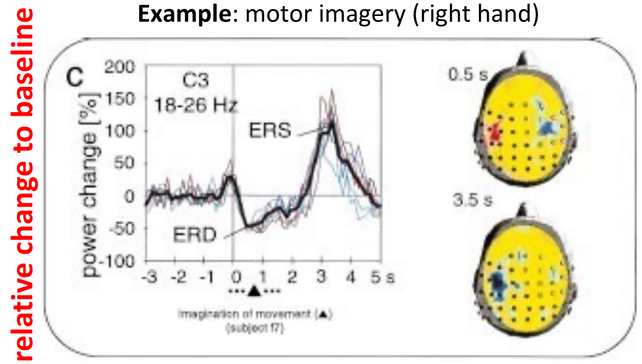
Effects in brain oscillations

Localized in …



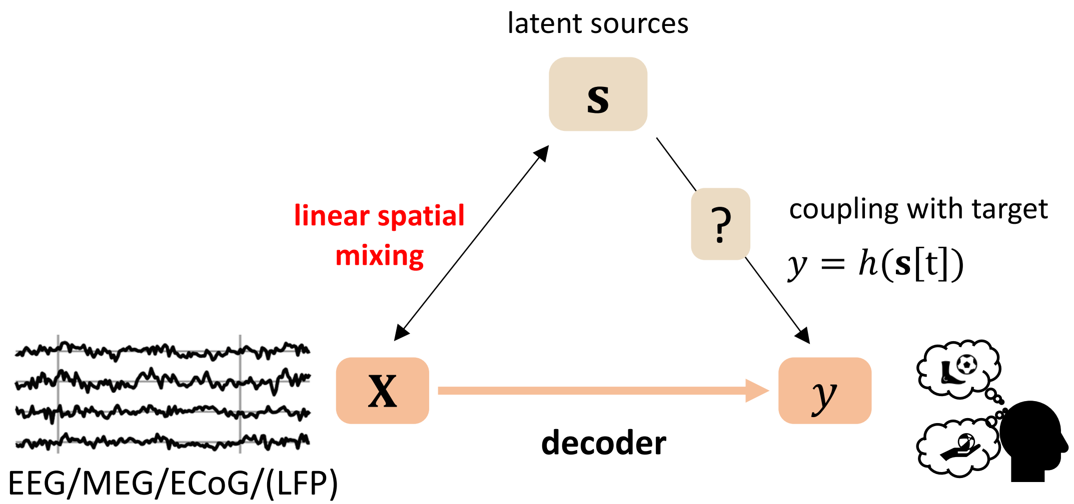
[Cohen2017, *TINS*]

Example: motor imagery (right hand)



[Pfurtscheller&Lopes da Silva1999, *Clin Neurophysiol.*]

We measure a linear mixture of sources



Challenges

- effect localization
  - poor signal-to-noise ratio
  - individual differences
  - non-stationary (changes to baseline!)
- scarce data
  - expensive to record new data
- coupling
  - mechanism (temporal/spectral)
  - specificity
- artifacts (outliers)

What makes a good decoder?

feature extraction learn temporal/spatial/spectral effects

robustness scarce data, outliers, …

interpretation provide insight about localized effects

generalization unseen data e.g., days or subjects

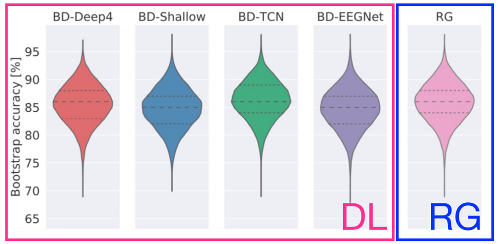
Riemannian geometry aware methods utilize

- dimensionality reduction (e.g., spatial covariance)
- power modulations with log-linear coupling between $\mathbf{X}$ and y [Sabbagh+2019, *NeurIPS*]
- invariance to linear spatial mixing

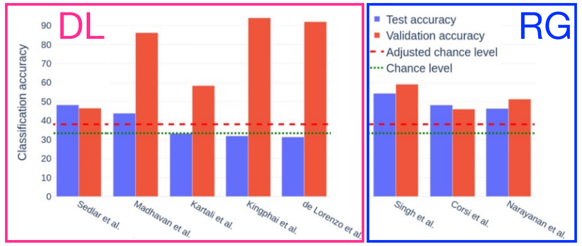
Examples:

Deep learning (DL)
+ feature extraction
- robustness
- interpretation
= generalization

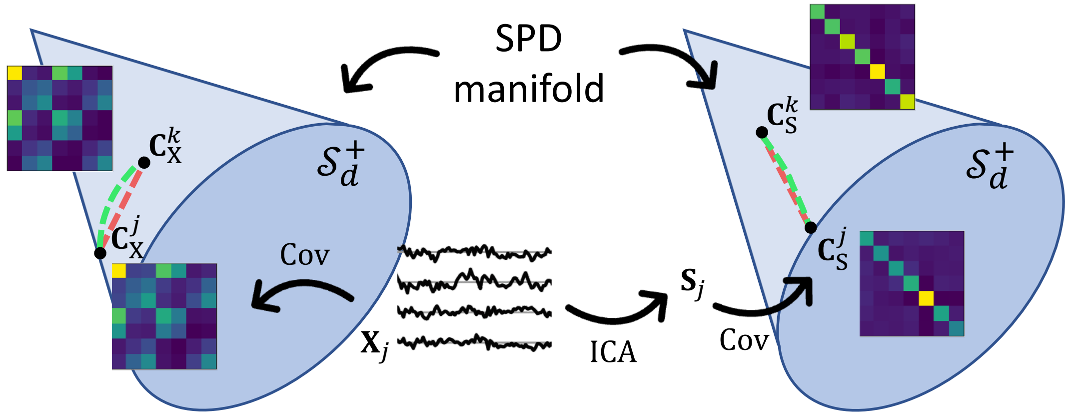
Riemannian geometry (RG)
- feature extraction
+ robustness
+ interpretation
= generalization



[Gemein+2020,*NeuroImage*]



[Roy+2022, *Front. Neuroergonomics*]



invariance to linear mixing

Euclidean distance ($\|\cdot\|_F$)

$$\left\|\mathbf{C}_X^j - \mathbf{C}_X^k\right\|_F = 0.35 \neq 0.60 = \left\|\mathbf{C}_S^j - \mathbf{C}_S^k\right\|_F$$

✗

Affine invariance metric (δ_{AIRM})

$$\delta_{AIRM}\left(\mathbf{C}_X^j, \mathbf{C}_X^j\right) = 0.75 = 0.75 = \delta_R\left(\mathbf{C}_S^j, \mathbf{C}_S^k\right)$$

✓

Generalization across sessions and subjects

Poor generalization of (EEG-based) BCIs across subjects and sessions [Wei+2021,*arXiv*].

Reason: **distribution shifts**

- Internal: mental traits and states
- External: noise sources and artifacts

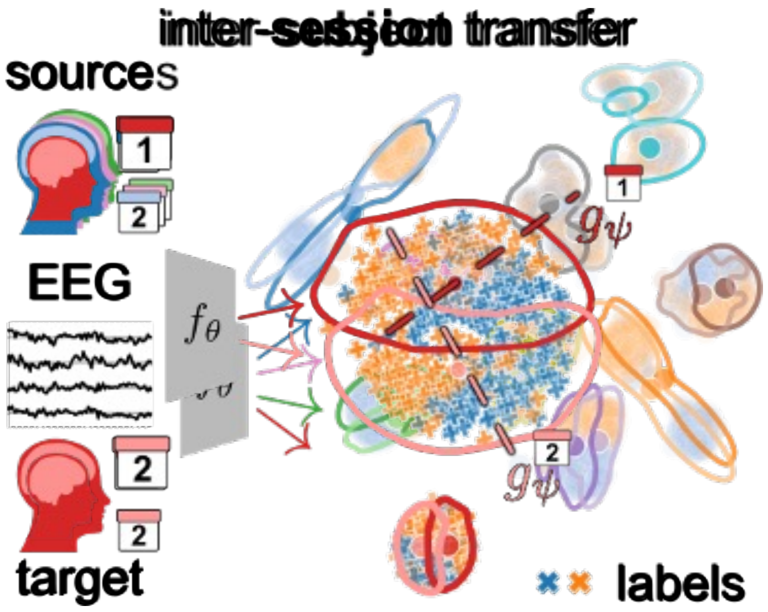
Gold standard:

domain-specific data and model

- poor scalability and utility

Alternative: domain-adaptation (DA)

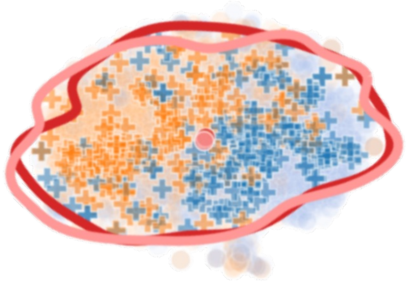
- unsupervised
- source-free
- online



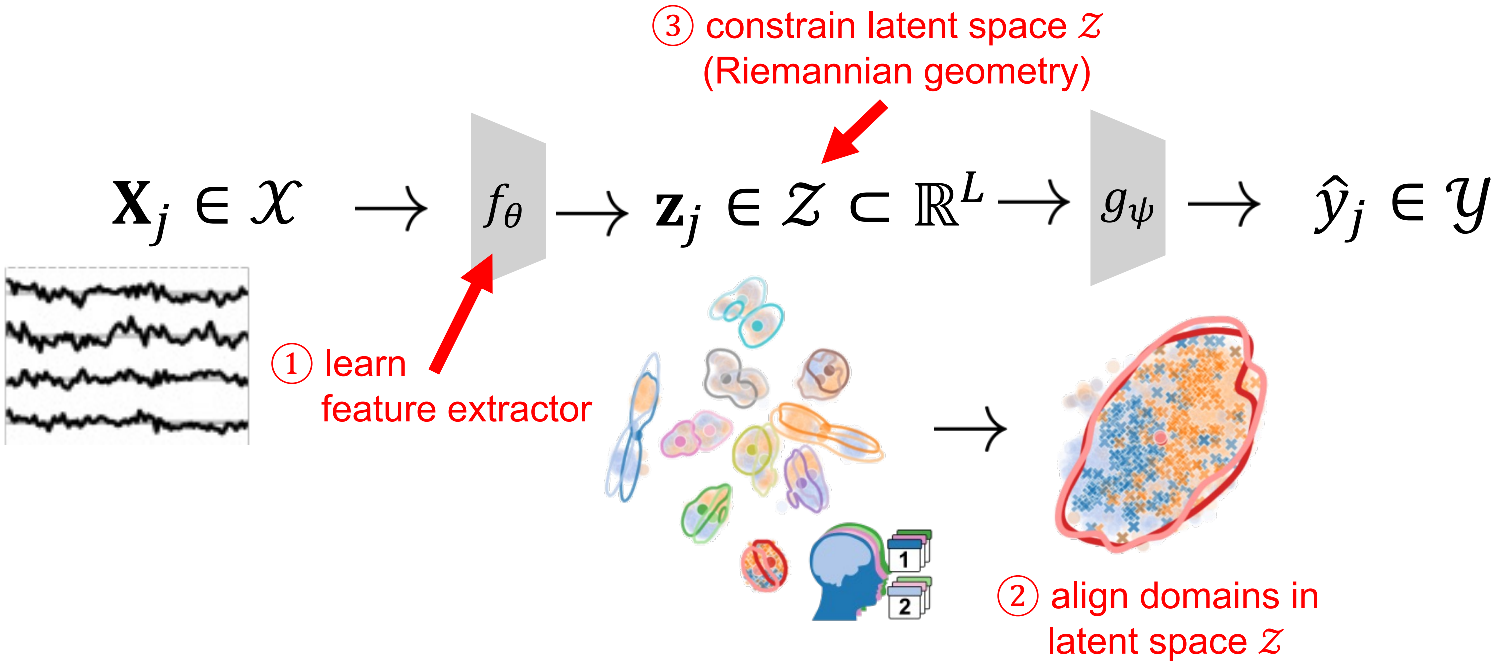
BCNI2015011 dataset (3 class motor-imagery)

unsupervised domain-adaptation (UDA)

→



Why geometric deep learning?



feature extraction 1 standard layers (e.g., conv nets) to extract spatial/temporal/spectral features

generalization 2 explicit alignment of domains (=subject/session) in <u>latent</u> space

robustness 3 constrain latent space; benefit from invariance properties of $(\mathcal{S}_d^+, \delta_{\mathrm{AIRM}})$

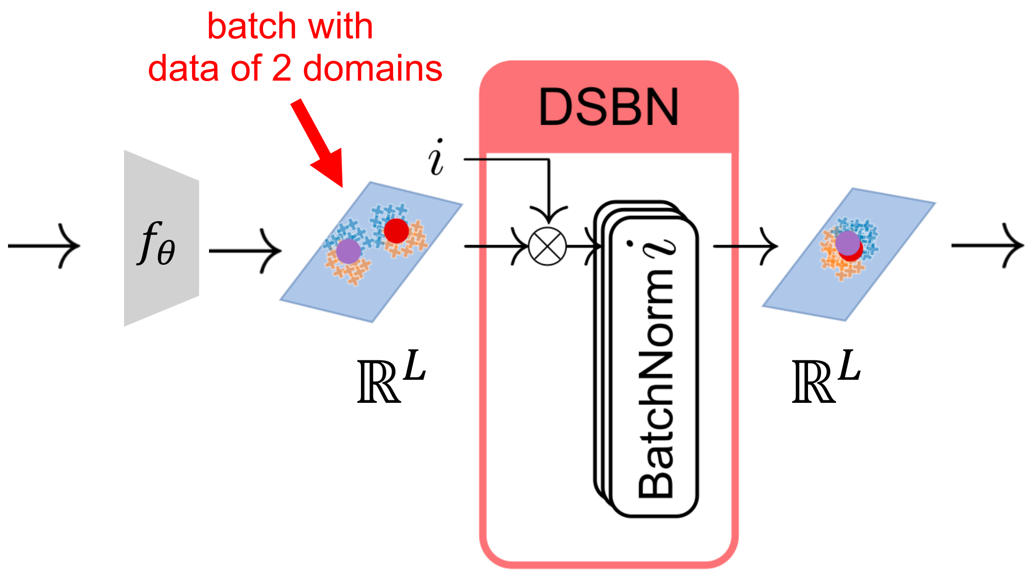
Explicit alignment in latent space $\mathcal{Z}$

Motivation:

align mean and std along **relevant** (=discriminative) dimensions

Constraint:

backpropagate gradients



Domain-specific batch-normalization (DSBN)

[Chang+2019,*CVPR*]

- Source-free domain adaptation
- Online unsupervised domain adaptation

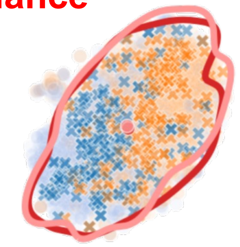
Effective for EEG [Li+2021,*Methods*; Bakas+2022,*arXiv*]

Constrain latent space to SPD manifold $\mathcal{S}_d^+$

$$\mathbf{X}_j \in \mathcal{X} \longrightarrow f_\theta \longrightarrow \mathbf{z}_j \in \mathcal{Z} \subset \mathbb{R}^L \longrightarrow \text{DSBN} \longrightarrow \tilde{\mathbf{z}}_j \in \mathbb{R}^L \longrightarrow g_\psi \longrightarrow y_j \in \mathcal{Y}$$

align mean and variance

$\longrightarrow$



constrain latent space and align Fréchet mean and variance

$$\mathbf{X}_j \in \mathcal{X} \longrightarrow f'_\theta \ \underbrace{\text{Cov-Pool} \ \ \text{ReEig}}_{\text{SPDNet layers [Huang\&Gool2017, } \textit{AAAI}\text{]}} \longrightarrow \mathbf{Z}_j \in \mathcal{S}_d^+ \longrightarrow \underbrace{m_\phi}_{\text{alignment and tangent space projection}} \longrightarrow \tilde{\mathbf{z}}_j \in \mathbb{R}^L \longrightarrow g_\psi \longrightarrow y_j \in \mathcal{Y}$$

Alignment on $\mathcal{S}_d^+$ and projection to $\mathbb{R}^L$

[Kobler+2022,*NeurIPS*]

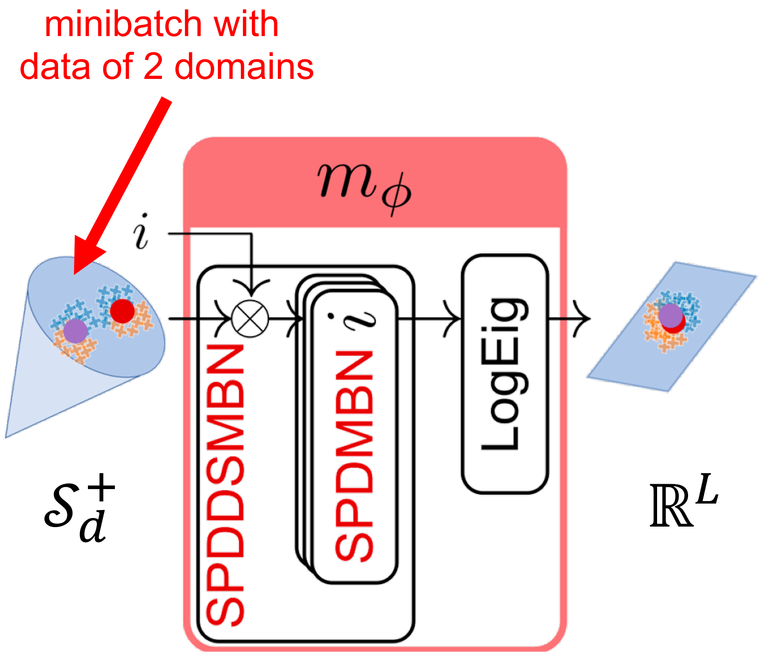
Alignment

- New layer (SPDDSMBN) that combines ideas of DSBN [Chang+2019,*CVPR*], momentum batch norm [Yong+2020,*ECCV*] and SPDBN [Kobler+2022,*ICASSP*]
- Exponential smoothing to track domain's Fréchet mean $\mathbf{G}_i$ and variance ν_i across minibatches, while learning f_θ.
- Standardize and re-bias each domain's data so that the Fréchet mean $\mathbf{G}_i \to \mathbf{I}$ and variance $\nu_i \to \nu_\phi \in \mathbb{R}^+$ (ν_ϕ … learnable parameter).

Projection: LogEig layer [Huang&Gool2017,*AAAI*]

Properties of m_ϕ:

- Estimates $\mathbf{G}_i, \nu_i$ can converge to the data's true Fréchet mean and variance
- Bridge to classic RG methods on $(\mathcal{S}_d^+, \delta_{\mathrm{AIRM}})$ based on tangent space mapping (TSM) [Barachant+2012, *TBME*]



TSM in a nutshell

[Barachant+2012, *TBME*]

Setting

Riemannian manifold $(\mathcal{S}_d^+, \delta_{\mathrm{AIRM}})$

$$\mathcal{S}_d^+ = \{\mathbf{Z} \in \mathbb{R}^{d\times d} : \mathbf{Z} = \mathbf{Z}^T, \mathbf{Z} \succ 0\}$$

$$\delta_{\mathrm{AIRM}}(\mathbf{Z}_1, \mathbf{Z}_2) = \left\| \mathrm{Log}\left(\mathbf{Z}_1^{-\frac{1}{2}} \mathbf{Z}_2 \mathbf{Z}_1^{-\frac{1}{2}}\right) \right\|_F$$

For all invertible $\mathbf{A}$ (linear mixing):

$$\delta_{\mathrm{AIRM}}(\mathbf{A}\mathbf{Z}_1\mathbf{A}^T, \mathbf{A}\mathbf{Z}_2\mathbf{A}^T) = \delta_{\mathrm{AIRM}}(\mathbf{Z}_1, \mathbf{Z}_2)$$

N observations from a single domain i

$$\mathcal{D}_i = \{(\mathbf{X}_j, y_j) \mid j = 1, \dots, N\}$$

Limitations

Expensive to compute Fréchet mean $\mathbf{G}$ and variance ν.
$\Rightarrow$ $\{\theta, \phi, \psi\}$ are typically fitted sequentially

Typical tangent space mapping (TSM) models

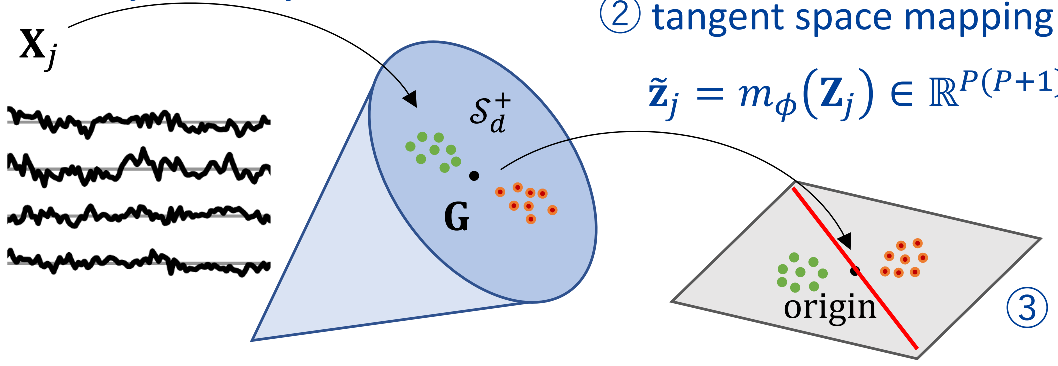
① feature extraction

$$\mathbf{Z}_j = f_\theta(\mathbf{X}_j) \in \mathcal{S}_d^+$$

$$m_{\boldsymbol{\phi}}(\mathbf{Z}_j) = \mathrm{upper} \circ \Gamma_{\mathbf{G}\to\mathbf{I}} \circ \mathrm{Log}_{\mathbf{G}}(\mathbf{Z}_j) = \mathrm{upper} \circ \mathrm{Log}\left(\mathbf{G}^{-\frac{1}{2}} \mathbf{Z}_j \mathbf{G}^{-\frac{1}{2}}\right)$$

② tangent space mapping

$$\tilde{\mathbf{z}}_j = m_\phi(\mathbf{Z}_j) \in \mathbb{R}^{P(P+1)/2}$$



③ classification

$$y_j = g_\psi(\mathbf{v}_j)$$

Fréchet mean $\mathbf{G}$ and variance ν

$$\mathbf{G} = \arg\min_{\mathbf{Z}\in\mathcal{S}_P^+} L(\mathbf{Z}) \qquad \nu = L(\mathbf{G})$$

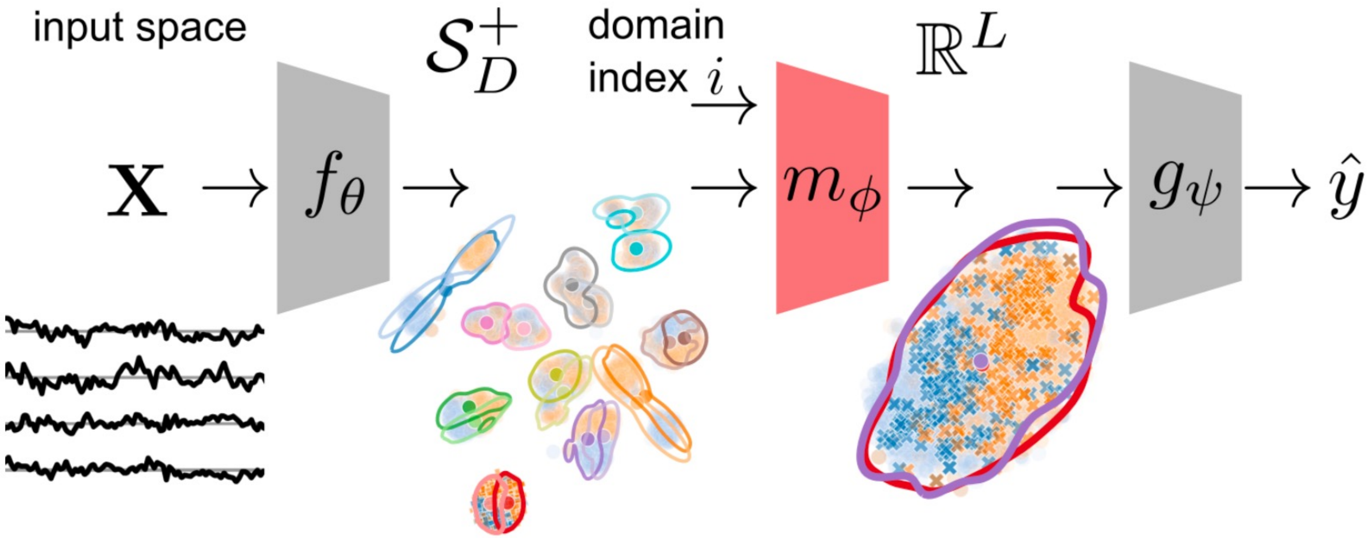
$$L(\mathbf{Z}) = \frac{1}{N}\sum_{j=1}^{N} \delta_{\mathrm{AIRM}}^2(\mathbf{Z}, \mathbf{Z}_j)$$

Tangent space distances are…

- … cheap to compute
- … locally approximate δ_{AIRM}
  $\Rightarrow$ inherit *invariance* properties

TSMNet: learning TSM end-to-end

[Kobler+2022,*NeurIPS*]



Training:

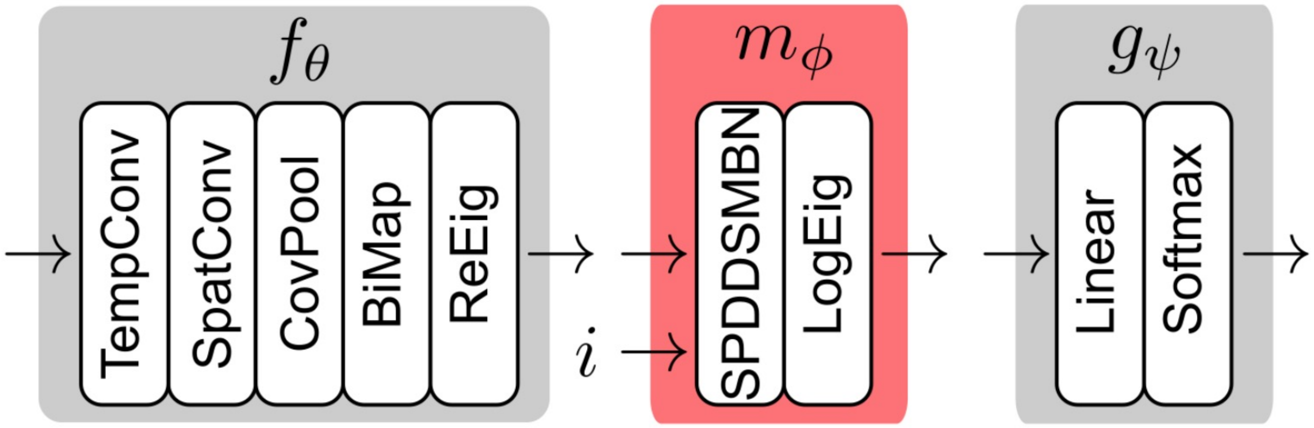
- **Loss**: cross-entropy error (classification setting)
- **Strategy**: minibatch-based gradient descent
- **Optimizer**: Riemannian ADAM [Becigneul+2019,*ICML*]

Inference: keep learnable parameters fixed $\{\theta,\phi,\psi\}$

- Source domains: use learned Fréchet mean and variance
- New domains: latent space projection + estimate Fréchet mean and variance
- Online: update Fréchet mean and variance iteratively

TSMNet: parametrization and interpretation

[Kobler+2022,*NeurIPS*]



linear + temporal covariance TSM **linear**

Equivalent to typical TSM [Barachant+2012, *TBME*]

⇒ **global model interpretation feasible** (i.e., $\{\theta,\phi,\psi\} \rightarrow$ spatial and spectral patterns)

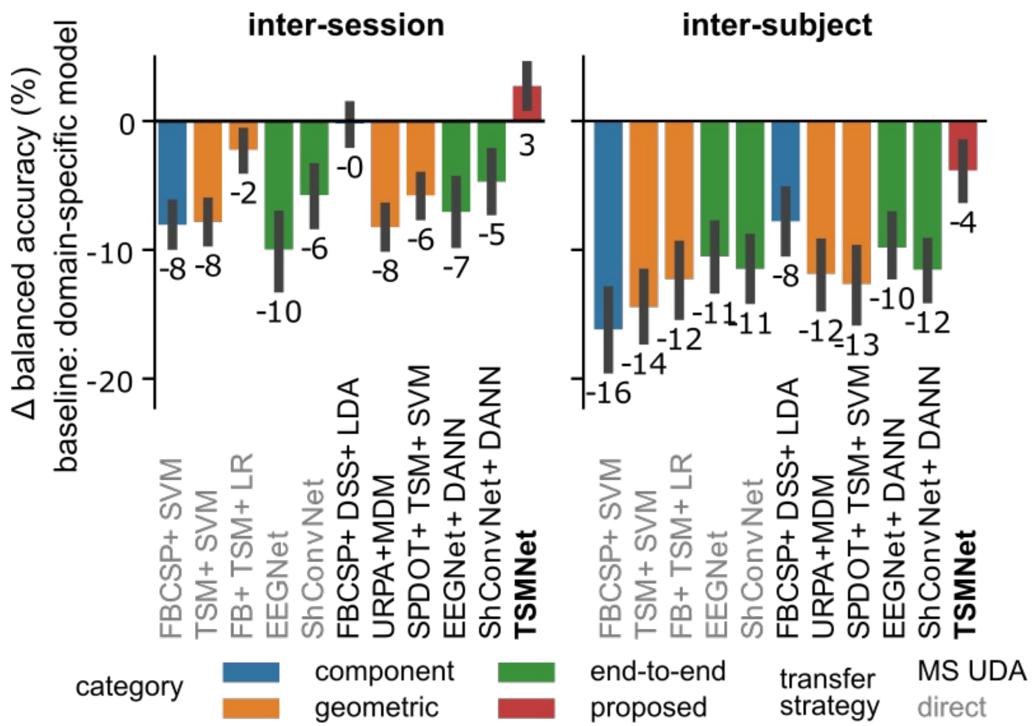
[Xu+2019,*GBCIC*;Kobler+2021, *EMBC*]

Results: inter-subject/-session transfer

[Kobler+2022,*NeurIPS*]

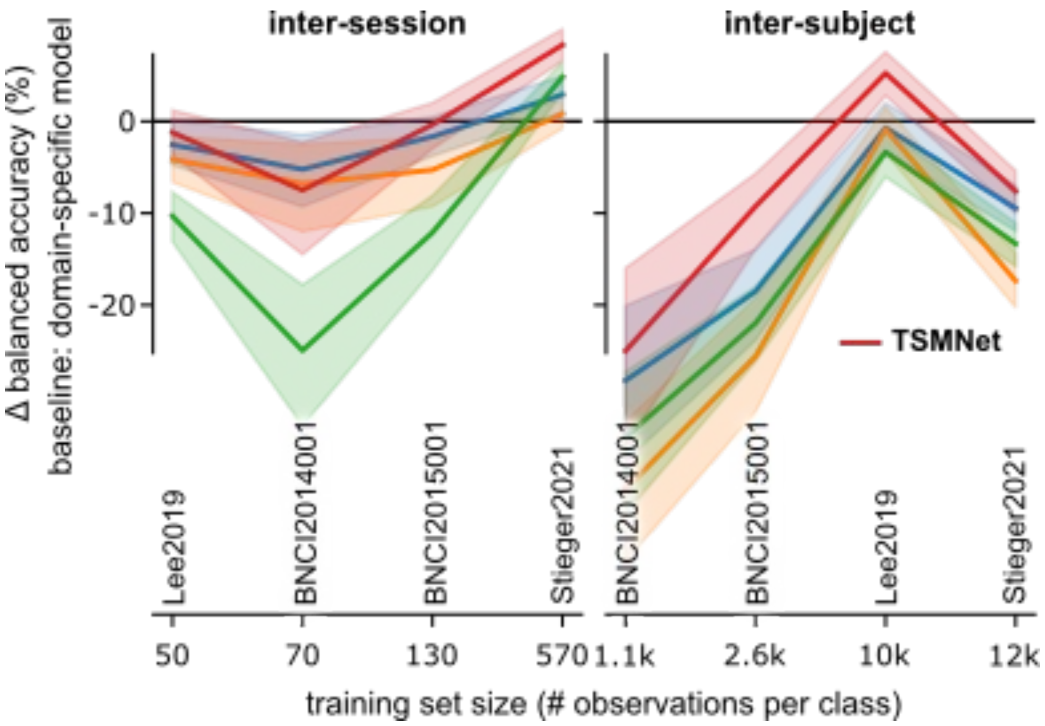
Grand average (motor imagery)

5 dataset, 573 sessions, 138 subjects



Performance over dataset size

Small to large-scale MI datasets



What drives the improvement (ablation)?

[Kobler+2022,*NeurIPS*]

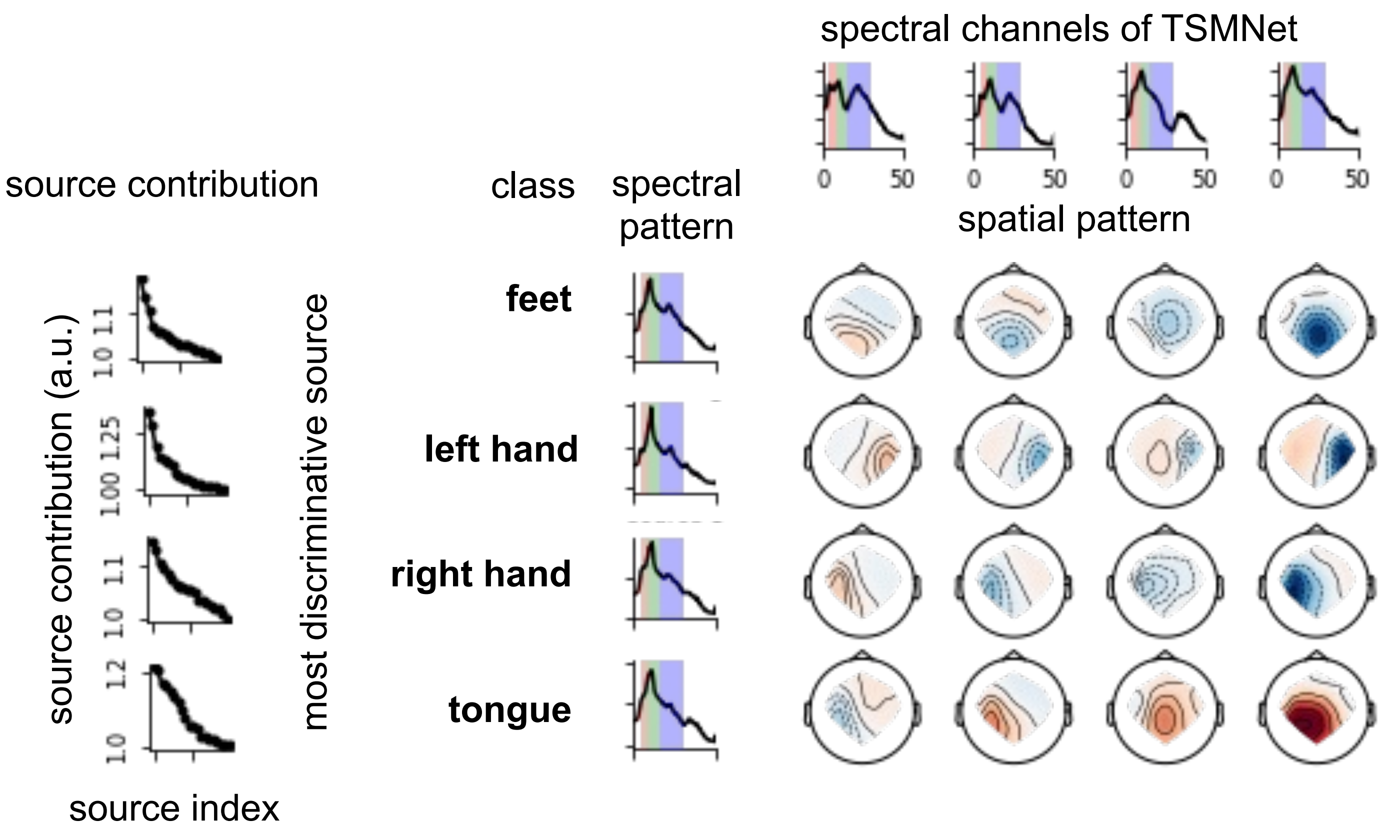
5 dataset, 138 subjects

$\mathcal{S}_d^+$	alignment	Δ balanced accuracy (%)	t-value	fit time (s)
yes	yes	-	-	16.9
	no	-3.9	-10.7	11.3
no	yes	-4.5	-10.1	6.6
	no	-6.9	-13.4	4.4

What effects did the model localize?

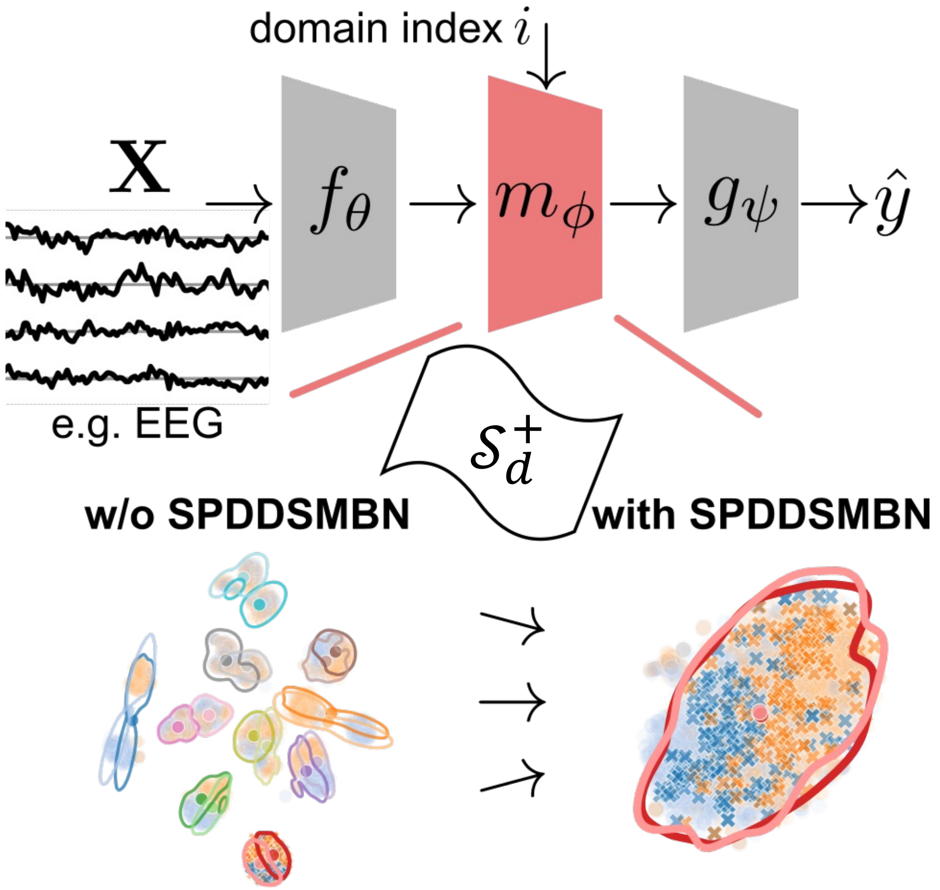
[Kobler+2022,*NeurIPS*]

Patterns for TSMNet fitted to session 1 of subject 9 (BNCI2014001 dataset).



Summary

- ✓ Learn interpretable TSM models on $(\mathcal{S}_d^+, \delta_{\mathrm{AIRM}})$ end-to-end
- ✓ Generalization across sessions and subjects
- ✓ Systematic performance increase compared to conventional, geometry aware methods and conv nets.
- ✓ Seamless transfer to new domains and unsupervised online adaptation



Limitations

- × Computation cost $O(d^3)$ of matrix logarithms and matrix powers (eigen decomposition)
- × Datasets with domain-specific class-imbalances

https://github.com/rkobler/TSMNet

https://github.com/rkobler/TSMNet

Acknowledgments

Acknowledgments
This work was supported by Innovative Science and Technology Initiative for Security Grant Number JPJ004596, ATLA, Japan.

References

Chang et al., “Domain-Specific Batch Normalization for Unsupervised Domain Adaptation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,2019.

Li et al. “Subject matching for cross-subject EEG-based recognition of driver states related to situation awareness,” *Methods*, vol. 202, pp. 136–143, 2022.

Bakas *et al*., “Team Cogitat at NeurIPS 2021: Benchmarks for EEG Transfer Learning Competition.” arXiv, 2022.

Huang and Gool, “A Riemannian Network for SPD Matrix Learning,” in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

Yong et al., “Momentum Batch Normalization for Deep Learning with Small Batch Size,” in *Computer Vision – ECCV,* 2020.

Barachant et al., “Multiclass brain-computer interface classification by Riemannian geometry.,” *IEEE transactions on bio-medical engineering*, vol. 59, no. 4, pp. 920–928, 2012.

Becigneul et al., “Riemannian Adaptive Optimization Methods,” in *International Conference on Learning Representations*, 2019.

Kobler et al., “On the interpretation of linear Riemannian tangent space model parameters in M/EEG,” in *Proceedings of the 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2021.

Kobler et al., “Controlling The Fréchet Variance Improves Batch Normalization on the Symmetric Positive Definite Manifold,” in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

Kobler et al., “SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in EEG,” in *Advances in Neural Information Processing Systems*, 2022.