FedSPDnet: Geometry-Aware Federated Deep Learning with SPDnet

Thibault Pautrel [1], Florent Bouchard [1], Guillaume Ginolhac [2] and Ammar Mian [2]

[1]L2S, CentraleSupélec [2]LISTIC

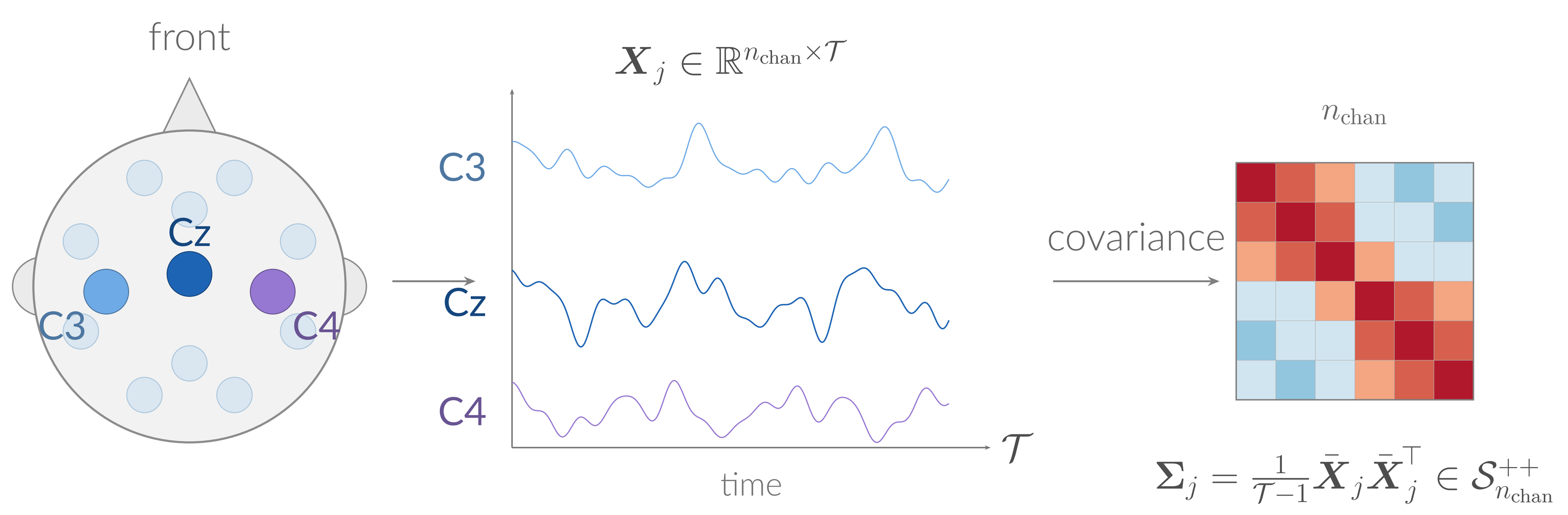
Overview

Task: classify EEG trials (motor imagery), e.g. left/right hand, tongue, feet, rest.
Models: two architectures, two geometries:

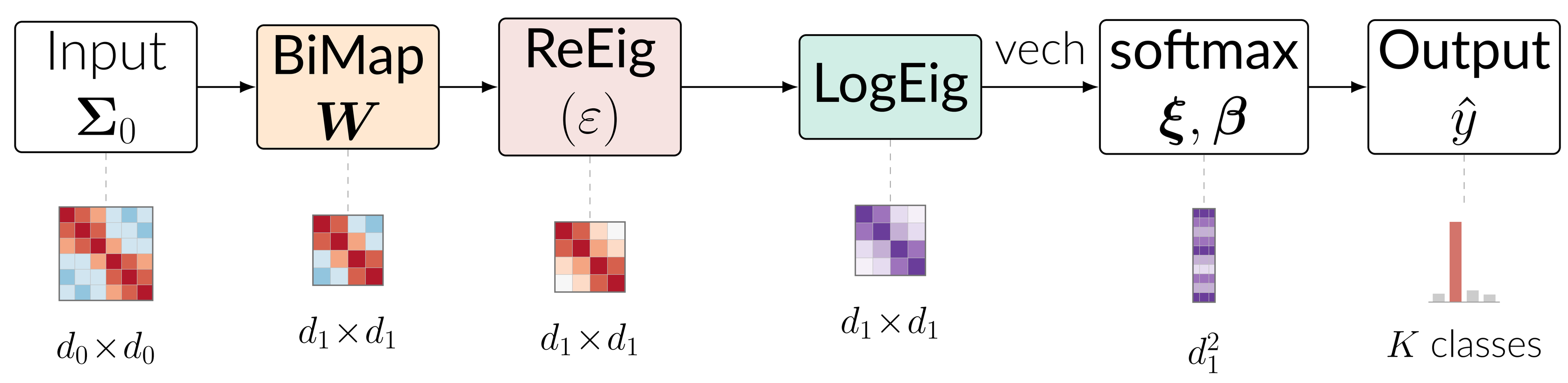
EEGnet [1]	vs	SPDnet [2]
Euclidean architecture on raw signals		*Riemannian* architecture on covariance matrices [3]

Challenge: EEG recordings **cannot be centralised** (confidentiality) $\longrightarrow$ **federated learning**.
Proposal: Geometry-preserving **FedSPDnet** vs **FedEEGnet**.

Why Riemannian geometry for EEG?



SPDnet: a Riemannian network that classifies SPD matrices



- BiMap: $\boldsymbol{\Sigma} \leftarrow \boldsymbol{W}^\top \boldsymbol{\Sigma} \boldsymbol{W}$, $\boldsymbol{W} \in \mathrm{St}_{d_0,d_1}$ orthogonal $(d_1 < d_0)$ *(reduction)*
- ReEig: $\boldsymbol{\Sigma} \leftarrow \boldsymbol{U} \max(\varepsilon \boldsymbol{I}, \boldsymbol{\Lambda}) \boldsymbol{U}^\top$ *(non-linearity)*
- LogEig: lift to the tangent space at the identity, then softmax. *(classif)*

Parameters $\theta = (W, \xi, \beta)$ live on a **product manifold** $\mathcal{M} = \mathrm{St}_{d_0,d_1} \times \mathbb{R}^{K \times d_1(d_1+1)/2} \times \mathbb{R}^K$.

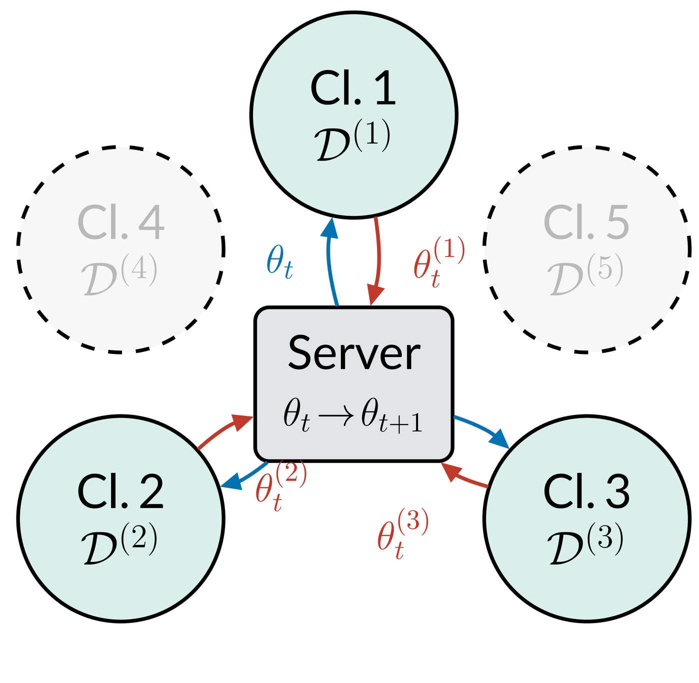
Federated learning: where the geometry breaks

Goal: find

$$\theta^\star \in \operatorname*{arg\,min}_{\theta \in \mathcal{M}} \left\{ F(\theta) = \tfrac{1}{N} \textstyle\sum_{i=1}^{N} F_i(\theta; \mathcal{D}^{(i)}) \right\}$$

At each round $t = 0, \dots, T-1$:

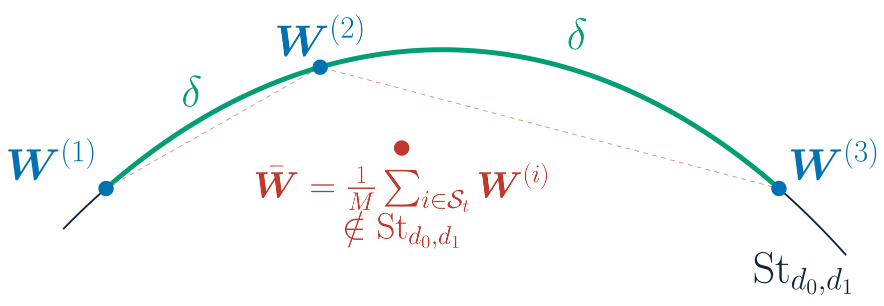
1. **Sample** $\mathcal{S}_t \subset \{1, \dots, N\}$, $|\mathcal{S}_t| = k$, uniformly at random *w.o.r.* (green).
2. **Broadcast** θ_t to every client $i \in \mathcal{S}_t$.
3. Local Riemannian optimisation: E epochs, staying on $\mathcal{M}$
4. **Return** $\theta_t^{(i)} \in \mathcal{M}$ — raw data never moves.
5. Aggregation: form $\theta_{t+1} \in \mathcal{M}$. FedAvg [4]: $\theta_{t+1} = \frac{1}{k} \sum_{i \in \mathcal{S}_t} \theta_t^{(i)}$

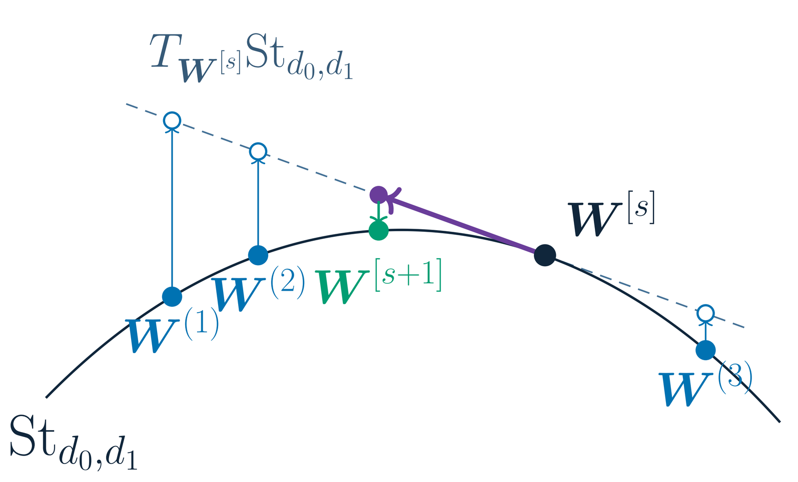


- An arithmetic mean of *orthogonal* matrices is **not orthogonal**.
- The Riemannian mean $\bar{\boldsymbol{W}}^{\mathcal{R}}$ has **no closed form**, only an **iterative** solve.

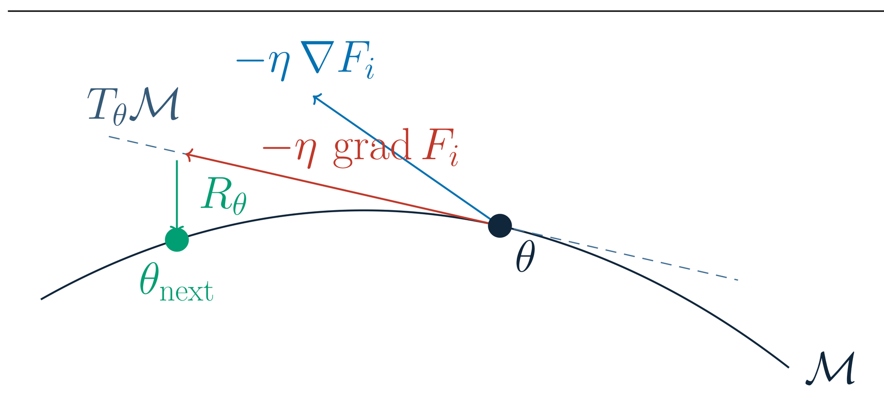
$$\bar{\boldsymbol{W}}^{\mathcal{R}} = \operatorname*{arg\,min}_{\boldsymbol{W} \in \mathrm{St}} \sum_{i \in \mathcal{S}_t} \delta^2(\boldsymbol{W}, \boldsymbol{W}^{(i)})$$

$$\boldsymbol{W}^{[s+1]} = \mathrm{Exp}_{\boldsymbol{W}^{[s]}} \Big(\tfrac{1}{M} \sum_{i \in \mathcal{S}_t} \mathrm{Log}_{\boldsymbol{W}^{[s]}} (\boldsymbol{W}^{(i)}) \Big)$$





Local Riemannian optimisation



Gradient-based descent

- Euclidean: $\theta \leftarrow \theta - \eta \nabla F_i(\theta)$
- Riemannian: $\theta \leftarrow R_\theta\big(-\eta \operatorname{grad} F_i(\theta)\big) \in \mathcal{M}$

Proposed closed-form Stiefel aggregations

Following [5], we *approximate* the Riemannian mean. For every BiMap weight:

ProjAvg — average in the ambient space, project back:

$$\boldsymbol{W}_{t+1} = \Pi \left(\tfrac{1}{M} \textstyle\sum_{i \in \mathcal{S}_t} \boldsymbol{W}_t^{(i)} \right)$$

RLAvg — lift, average in the tangent space, retract:

$$\boldsymbol{W}_{t+1} = R_{\boldsymbol{W}_t} \Big(\tfrac{1}{M} \textstyle\sum_{i \in \mathcal{S}_t} L_{\boldsymbol{W}_t}(\boldsymbol{W}_t^{(i)}) \Big)$$

- **Closed-form**: one polar decomposition per layer per round, cost $\mathcal{O}(M n_{\text{chan}} d + n_{\text{chan}} d^2)$.
- Orthogonality **preserved**.
- **Optimizer-agnostic**: clients may use *any* manifold-aware solver.

Experiments and results

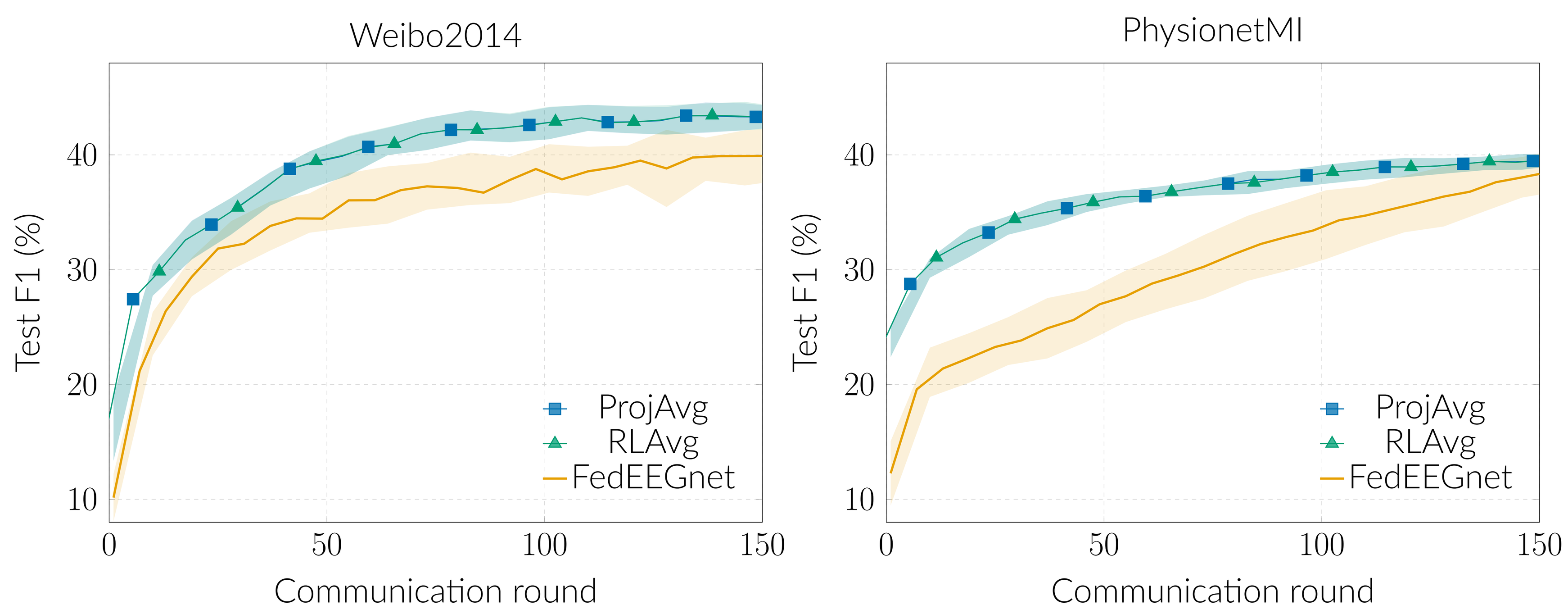
Two MOABB motor-imagery datasets [6], one population-level model per dataset.

Weibo2014	
Subjects / channels / classes	10 / 60 / 7
Clients N (2 subj. each)	5
SPDnet d_1 / ε	22 / 10^{-1}

PhysionetMI	
Subjects / channels / classes	103 / 64 / 4
Clients N (2 subj. each)	53
SPDnet d_1 / ε	18 / 10^{-2}

Band-pass $[8, 32]$ Hz; local split 75/10/15; Adam (lr 10^{-3}, batch 64); 300 centralised epochs, T=150 rounds $\times$ E=2 local epochs; full and half participation ($M = \lfloor N/2 \rfloor$); 10 seeds.

Baseline: **EEGnet** [1] operating on raw signals + FedAvg.



Full participation; shaded bands: $\pm$1 std over 10 seeds.

Weibo2014

	FedEEGnet	FedSPDnet	Δ
# params	5 303	**4 715**	-11%
Centralised	50.7 ± 1.5	$\mathbf{51.7 \pm 0.8}$	
Federated, full	39.9 ± 2.4	$\mathbf{43.3 \pm 1.0}$	$+3.4$
Federated, half	36.4 ± 4.1	$\mathbf{41.2 \pm 0.8}$	$+4.8$

PhysionetMI

	FedEEGnet	FedSPDnet	Δ
# params	3 284	**2 452**	-25%
Centralised	$\mathbf{50.8 \pm 0.7}$	43.1 ± 0.5	
Federated, full	38.3 ± 1.8	$\mathbf{39.5 \pm 0.6}$	$+1.2$
Federated, half	38.0 ± 2.7	$\mathbf{39.5 \pm 0.9}$	$+1.5$

Test F1 (%), mean $\pm$ std over 10 seeds. FedSPDnet uses ProjAvg.

- **ProjAvg $\approx$ RLAvg.** Both markers sit on the same curve; ProjAvg is preferred, it stores no previous iterate.
- **Less is lost to federation.** FedSPDnet gives up 8.4 / 3.6 F1 points from its centralised score, against 10.8 / 12.5 for FedEEGnet — on PhysionetMI it *loses* centrally yet *wins* once federated.
- **Robust.** Half participation costs FedSPDnet ≤ 2 points against 3.5 for FedEEGnet, with lower variance.

Conclusion

- FedSPDnet: two closed-form, geometry-preserving aggregations — **ProjAvg** is computationally cheaper.
- **FedSPDnet** beats **FedEEGnet** on two EEG benchmarks: **less lost** to federation, **faster** convergence, **robust** to partial participation, **fewer** parameters per round.

References

[1] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.

[2] Zhiwu Huang and Luc Van Gool. A riemannian network for spd matrix learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

[3] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012.

[4] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[5] Florent Bouchard, Nils Laurent, Salem Said, and Nicolas Le Bihan. Beyond r-barycenters: an effective averaging method on stiefel and grassmann manifolds. *arXiv preprint arXiv:2501.11555*, 2025.

[6] Sylvain Chevallier, Igor Carrara, Bruno Aristimunha, Pierre Guetschel, Sara Sedlar, Bruna Lopes, Sébastien Velut, Salim Khazem, and Thomas Moreau. The largest EEG-based BCI reproducibility study for open science: the MOABB benchmark. *arXiv preprint arXiv:2404.15319*, 2024.

Riemannian toolbox on Stiefel

$\mathrm{St}_{p,k} = \{\boldsymbol{W} \in \mathbb{R}^{p \times k} : \boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}_k\}$, with the ambient Frobenius metric $\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle = \mathrm{tr}(\boldsymbol{\xi}^\top \boldsymbol{\eta})$.

- *Riemannian gradient:* $\operatorname{grad} F(\boldsymbol{W}) = \nabla F(\boldsymbol{W}) - \boldsymbol{W}\big(\boldsymbol{W}^\top \nabla F(\boldsymbol{W})\big)_{\text{sym}}$
- *Projection-based retraction:* $R_{\boldsymbol{W}}(\boldsymbol{X}) = \Pi(\boldsymbol{W} + \boldsymbol{X})$ (Π = polar projection)
- *Lifting:* $L_{\boldsymbol{W}}(\boldsymbol{W}') = \boldsymbol{W}' - \boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{W}')_{\text{sym}}$