Projection-Based Federated Learning on Compact Submanifolds with Geometry-Independent Differential Privacy

Thibault Pautrel[1], Florent Bouchard[1], Guillaume Ginolhac[2], Ammar Mian[2] and Aurélien Bellet[3]

[1]CentraleSupélec, L2S [2]Université Savoie Mont Blanc, LISTIC [3]Inria, Montpellier

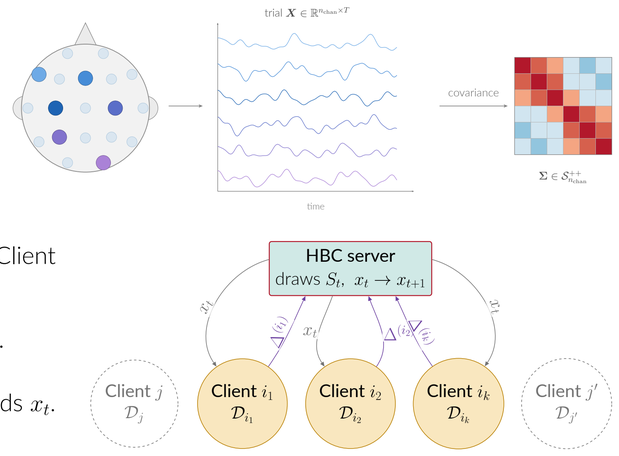
Overview

Goal. Train a shared model whose parameters lie on a **compact submanifold** $\mathcal{M}$, across n clients.

- **One map does all the geometry.** The projection on the manifold Π keeps local steps on $\mathcal{M}$ and aggregates the client iterates, in closed form. Clients may run *any* first-order optimiser.
- **Geometry-free privacy.** A record-level f-DP guarantee for the whole transcript, in closed form. It ignores the manifold, the aggregator, the optimiser and the number of participants.
- **Convergence in the regime FL runs in.** Several local steps *and* partial participation.

Setting: private Riemannian federated learning

Motivation. Covariance matrices act as popular descriptor for EEG signals. SPDNet [1] operates on SPD matrices and has weights $\boldsymbol{W}$ on the Stiefel manifold.



Problem. $(\mathcal{M}, x \mapsto \langle \cdot, \cdot \rangle_x)$ is a Riemannian manifold. Client i holds $\mathcal{D}_i$, of size m_i.

$$\min_{x \in \mathcal{M}} F(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x), \qquad f_i(x) = \frac{1}{m_i} \sum_{z \in \mathcal{D}_i} f(x; z).$$

Round t. The server draws S_t, with $|S_t| = k$, and sends x_t. Two steps: client side and server side.

(i) Local optimisation, client $i \in S_t$:

1. start $\quad x_0^{(i)} = x_t$
2. τ steps $\quad x_{j+1}^{(i)} = \mathrm{Opt}^{(i)}\big(x_j^{(i)}; \mathcal{B}_j\big), \ \mathcal{B}_j \subset \mathcal{D}_i$
3. send $\quad \Delta^{(i)} = \mathrm{Lift}_{x_t}\big(x_\tau^{(i)}\big)$

(ii) Aggregation. Average, then map back:

$$x_{t+1} = \mathrm{Map}_{x_t}\Big(\tfrac{1}{k} \textstyle\sum_{i \in S_t} \Delta^{(i)}\Big).$$

FedAvg [2] (flat): $\mathrm{Lift}_y(x) = x - y$, $\mathrm{Map}_y(v) = y + v$.
KFAvg [3]: $\mathrm{Lift} = \mathrm{Log}$, $\mathrm{Map} = \mathrm{Exp}$.
RetAvg [4]: $\mathrm{Lift} = R^{-1}$, $\mathrm{Map} = R$.

Threat model. The server is **honest but curious**. It keeps the transcript $\mathcal{V}_T = \big(\{x_t, S_t, \{\Delta^{(i)}\}_{i \in S_t}\}_{t<T}, x_T\big)$, and submissions leak records. We want to protect **records**.

Record-level DP. Two databases are **adjacent**, $D \simeq D'$, when one record of one client is replaced. The training $\mathcal{A}$, with output $\mathcal{V}_T$, is (ε, δ)-**DP** if, for all $D \simeq D'$ and every event E,

$$\Pr\big(\mathcal{A}(D) \in E\big) \le e^{\varepsilon} \Pr\big(\mathcal{A}(D') \in E\big) + \delta.$$

How to privatise: clip and noise the gradients

Records enter only through per-sample gradients, so we privatise there [5].

- **Clip** to bound the influence of one record: $\mathrm{clip}_C(v) = v \, \min\big(1, C/\|v\|\big)$, so that

$$\hat{g}_{\mathcal{B}} = \frac{1}{B} \sum_{z \in \mathcal{B}} \mathrm{clip}_C\big(\nabla f(x; z)\big) \quad \text{satisfies} \quad \big\|\hat{g}_{\mathcal{B}} - \hat{g}_{\mathcal{B}'}\big\| \le \frac{2C}{B} \quad \text{(one record replaced).}$$

- **Add noise** to hide it: $\hat{g}_{\mathcal{B}} + \mathcal{N}(0, \sigma_{\mathrm{dp}}^2 I)$, with σ_{dp} calibrated to $2C/B$.

Why intrinsic pipelines fall short

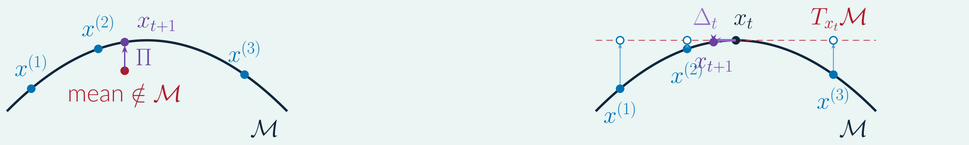
PriRFed [4], the only prior private method, works on a general manifold with an iterate-dependent metric $\langle \cdot, \cdot \rangle_x$. It privatises and aggregates in that geometry [6].

L1 Moving geometry: norm and noise rebuilt at each x.
L2 Loose accounting: per-round (ε, δ) pairs.
L3 Costly maps: R and R^{-1}, no closed form on Stiefel.
L4 Narrow rate: many steps or many clients.

DP-RFedProj: privatise in the ambient space, then project

- (P1) Induced metric. $\|v\|_x = \|v\|$ on $T_x\mathcal{M}$, so clipping and noise ignore x. → lifts L1.
- (P2) Compactness. It supplies $\Pi(y) = \arg\min_{x \in \mathcal{M}} \|y - x\|$. → lifts L3, L4.
- f-DP calculus [7] for the accounting. → lifts L2.

Server side: two projection-based aggregators [8]



ProjAvg: $\mathrm{Lift} = \mathrm{Id}$, $\mathrm{Map} = \Pi$

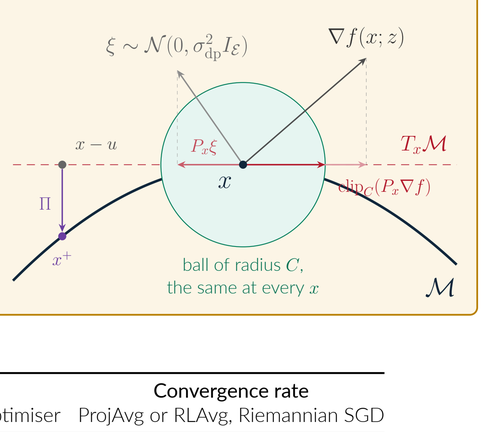
RLAvg: $\mathrm{Lift} = P_{x_t}(\cdot - x_t)$, $\mathrm{Map} = \Pi(x_t + \cdot)$

Client side: private local steps

At local iterate x, with a minibatch $\mathcal{B}$ of size B:

$$\tilde{g} = \frac{1}{B} \sum_{z \in \mathcal{B}} \mathrm{clip}_C\big(P_x \nabla f(x; z)\big) + P_x \xi, \quad \xi \sim \mathcal{N}(0, \sigma_{\mathrm{dp}}^2 I_{\mathcal{E}}).$$

- $P_x\xi$ is the tangent Gaussian of [6]. No orthonormal basis, no transport.
- Any optimiser, $u = \mathrm{Opt}^{(i)}(\tilde{g})$, then $x^+ = \Pi(x - u)$.



Scope of the two guarantees

Parameter space	Metric	Privacy: any aggregator, any optimiser	Convergence rate: ProjAvg or RLAvg, Riemannian SGD
Sphere, Stiefel, Grassmann	induced	✓	✓
Flat $\mathcal{E}$, incl. $\log \mathcal{S}^{++}$	Euclidean	✓	×
$\mathcal{S}^{++}$	affine-invariant	not runnable, no Π available	

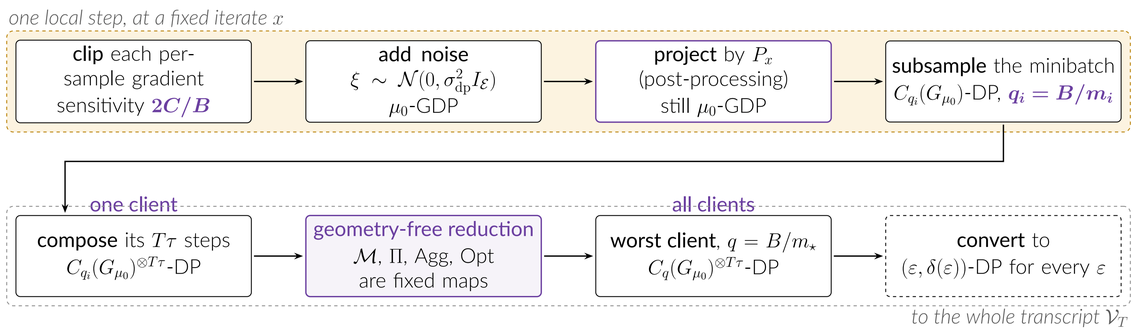
Privacy guarantee

f-**DP** [7]. The server tests D against an adjacent D'. Trade-off function: $T(P, Q)(\alpha) = \inf_\phi \big\{\mathbb{E}_Q[1-\phi] \ : \ \mathbb{E}_P[\phi] \le \alpha\big\}$. A mechanism is f-**DP** if $T(\mathcal{A}(D), \mathcal{A}(D')) \ge f$ for all $D \simeq D'$. It is μ-**GDP** if $f = G_\mu$, with $G_\mu(\alpha) = \Phi\big(\Phi^{-1}(1-\alpha) - \mu\big)$ the trade-off of $\mathcal{N}(0,1)$ against $\mathcal{N}(\mu, 1)$. Larger μ, weaker privacy. Subsampling a fraction q of the records maps f to $C_q(f) \ge f$. Composing N releases gives $f^{\otimes N}$.

Theorem (record-level f-DP). Let $\mu_0 = 2C/(B\sigma_{\mathrm{dp}})$ and $q = B/\min_i m_i$. The transcript $\mathcal{V}_T$ of DP-RFedProj is

$$C_q(G_{\mu_0})^{\otimes T\tau}\text{-DP},$$

whatever k, $\mathcal{M}$, Π, the optimisers and the aggregator.



Reading the budget

With $T\tau$ large and $q\sqrt{T\tau}$ fixed, at leading order in μ_0

$$\mu_{\mathrm{tot}} \approx \frac{2C\sqrt{T\tau}}{m_\star \, \sigma_{\mathrm{dp}}}, \qquad m_\star = \min_i m_i.$$

- Grows as $\sqrt{T\tau}$, against linearly in T for PriRFed.
- Set by the ratio C/σ_{dp}. The batch size cancels.
- The smallest client fixes the budget.

Convergence guarantee

Assumptions. Π is single-valued on a tube around $\mathcal{M}$. Every per-sample loss has an ℓ_f-Lipschitz gradient. Bounded within-client variance σ_{st}^2. Bounded heterogeneity σ_{het}^2.

With ProjAvg or RLAvg and Riemannian SGD, for T large enough,

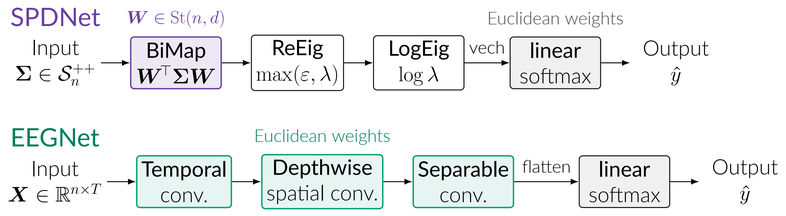
$$\frac{1}{T} \sum_{t<T} \mathbb{E} \, \| \operatorname{grad} F(x_t) \|^2 = O\left(\frac{1}{\sqrt{T}} \Big[1 + \underbrace{\tfrac{1}{k} \tfrac{n-k}{n-1} (\sigma_{\mathrm{het}} + \sigma_{\mathrm{clip}})^2}_{\text{partial participation}} + \underbrace{\frac{\bar{\rho} \, \sigma_{\mathrm{st}}^2}{k\tau}}_{\text{minibatch}} + \underbrace{\frac{d_{\mathcal{M}} \, \sigma_{\mathrm{dp}}^2}{k\tau}}_{\text{DP noise}} \Big] + \underbrace{\beta_C^2}_{\text{clip bias}} \right).$$

- The DP term scales with the **intrinsic** dimension $d_{\mathcal{M}}$, not with $\dim \mathcal{E}$.
- The clipping bias is a floor. It vanishes when the clip is inactive.
- **Why it works.** $\Pi(x + u) = x + P_x u + O(\|u\|^2)$. All local displacements live in the single space $\mathcal{E}$ and telescope. No transport, no curvature constants.

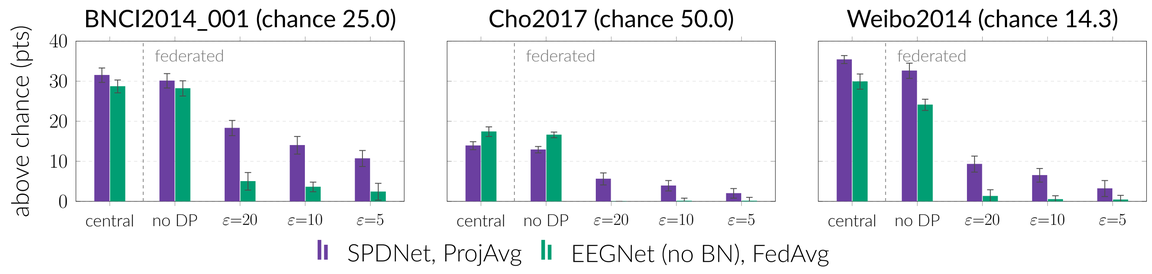
Experiments: private EEG motor-imagery decoding

Three MOABB datasets, full participation, 10 seeds, $\delta = 10^{-5}$. EEGNet [9]: Euclidean, FedAvg, no batch norm.

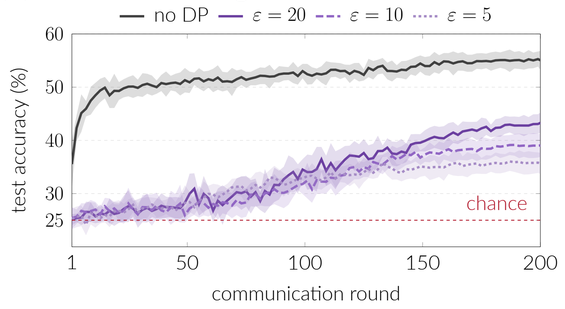
Dataset	chan.	classes	clients
BNCI2014_001	22	4	3
Cho2017	64	2	26
Weibo2014	60	7	5



Results

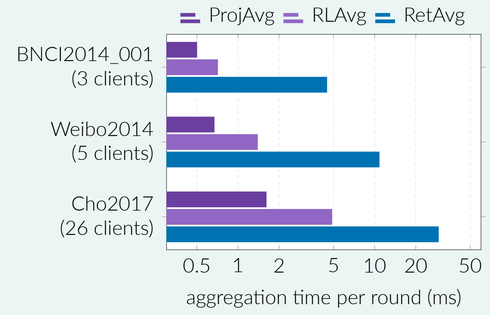


(b) SPDNet learning curves



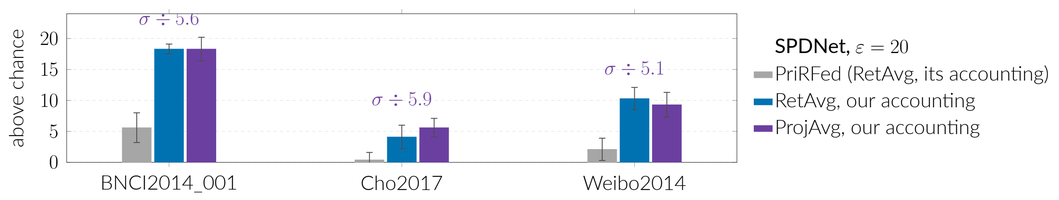
BNCI2014_001, ProjAvg, mean ±1 std over 10 seeds.

(c) Cheap aggregation, same accuracy



RetAvg is 9 to 18× slower, for the same accuracy.

(d) Tighter accounting than PriRFed: same mechanism, 5 to 6× less noise



Conclusion

- **Ambient privatisation** detaches the accounting from the geometry.
- **Projection-based aggregation** gives the first rate with several local steps and partial participation, with or without privacy.
- **Limits.** Compact manifolds with the induced metric only. Rate proved for Riemannian SGD. The clipping bias leaves a floor.

References

[1] 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.

[2] 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.

[3] Jiaxiang Li and Shiqian Ma. Federated learning on riemannian manifolds. *arXiv preprint arXiv:2206.05668*, 2022.

[4] Zhenwei Huang, Wen Huang, Pratik Jawanpuria, and Bamdev Mishra. Federated learning on riemannian manifolds with differential privacy. *Machine Learning*, 115(4):84, 2026.

[5] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.

[6] Andi Han, Bamdev Mishra, Pratik Jawanpuria, and Junbin Gao. Differentially private riemannian optimization. *Machine Learning*, 113(3):1133–1161, 2024.

[7] Jinshuo Dong, Aaron Roth, and Weijie J Su. Gaussian differential privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):3–37, 2022.

[8] Thibault Pautrel, Florent Bouchard, Ammar Mian, and Guillaume Ginolhac. FedSPDnet: Geometry-aware federated deep learning with SPDnet. In *European Signal Processing Conference (EUSIPCO)*, 2026.

[9] 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.

thibault.pautrel@centralesupelec.fr