Sliced-Wasserstein on Symmetric Positive Definite Matrices for M/EEG Signals

Clément Bonet*[1], Benoît Malézieux*[2], Alain Rakotomamonjy[3], Lucas Drumetz[4], Thomas Moreau[2], Matthieu Kowalski[5], Nicolas Courty [6]

[1]Université Bretagne Sud, LMBA; [2]Université Paris-Saclay, Inria, CEA; [3]Criteo AI Lab; Université de Rouen, LITIS; [4]IMT Atlantique, Lab-STICC; [5]Université Paris-Saclay, CNRS, LISN; [6]Université Bretagne Sud, Irisa

Symmetric Positive Definite matrices (SPD) for multivariate time series analysis

M/EEG data: Multivariate time series $X \in \mathbb{R}^{S\times T}$
S sensors and T time samples
Covariance matrices work well for prediction

After dimension reduction → SPD matrices in $S_d^{++}(\mathbb{R})$
$S_d^{++}(\mathbb{R}) = \{M \in S_d(\mathbb{R}),\ \forall x \in \mathbb{R}^d \setminus \{0\},\ x^T M x > 0\}$
Euclidean ML techniques don't work well on manifolds

Log-Euclidean distance:

$$\forall X, Y \in S_d^{++}(\mathbb{R}),\ d_{LE}(X,Y) = \|\log X - \log Y\|_F$$

Tangent space: $TS_d^{++}(\mathbb{R}) \cong S_d(\mathbb{R})$
Geodesics passing through the origin I_d in direction $A \in S_d(\mathbb{R})$:

$$\mathcal{G}_A = \{\exp(tA),\ t \in \mathbb{R}\}.$$

Optimal Transport

Wasserstein distance: Let $p \geq 1$, $\mu, \nu \in \mathcal{P}_p(S_d^{++}(\mathbb{R}))$, then

$$W_p^p(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int d_{LE}(x,y)^p \,\mathrm{d}\gamma(x,y).$$

Complexity *w.r.t* number of samples n: $O(n^3 \log n)$
Wasserstein distance on $\mathcal{P}(\mathbb{R})$:

$$\forall p \geq 1, \forall \mu,\nu \in \mathcal{P}_p(\mathbb{R}),\ W_p^p(\mu,\nu) = \int_0^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)|^p \,\mathrm{d}u$$

Complexity *w.r.t* number of samples n: $O(n \log n)$
Sliced-Wasserstein distance:

$$\forall \mu,\nu \in \mathcal{P}_p(\mathbb{R}^d),\ SW_p^p(\mu,\nu) = \int_{S^{d-1}} W_p^p(t_\#^\theta \mu, t_\#^\theta \nu)\,\mathrm{d}\lambda(\theta),$$

where $t^\theta(x) = \langle \theta, x\rangle$ and λ is the uniform distribution on S^{d-1}.
Slicing reduces the problem to multiple one dimensional problems and decreases the computational cost

Contributions

- Sliced-Wasserstein distance on SPD matrices (SPDSW)
- Properties of SPDSW
- Application to M/EEG data: Brain-age regression and BCI

SPDSW

Sliced-Wasserstein with geodesic projections onto geodesics sampled uniformly.

SPDSW

Let λ_S be the uniform distribution on $\{A \in S_d(\mathbb{R}),\ \|A\|_F = 1\}$. Let $p \geq 1$ and $\mu, \nu \in \mathcal{P}_p(S_d^{++}(\mathbb{R}))$, then the SPDSW discrepancy is defined as

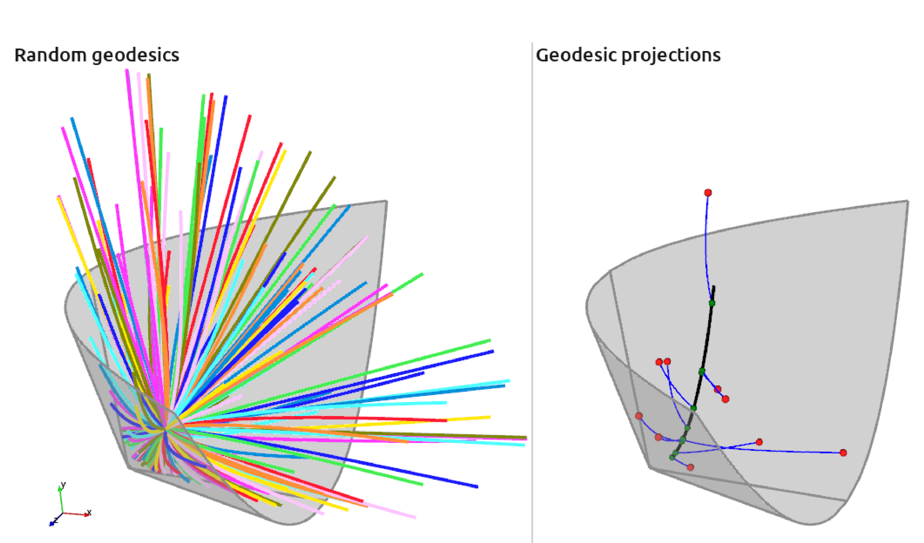
$$\mathrm{SPDSW}_p^p(\mu,\nu) = \int_{S_d(\mathbb{R})} W_p^p(t_\#^A \mu, t_\#^A \nu)\,\mathrm{d}\lambda_S(A),$$

with $t^A(M) = \langle A, \log M\rangle_F = \mathrm{Tr}(A \log M) = \mathrm{argmin}_{t\in\mathbb{R}}\ d_{LE}(M, \exp(tA))$.

Properties.
For $p \geq 1$, SPDSW_p is a distance.

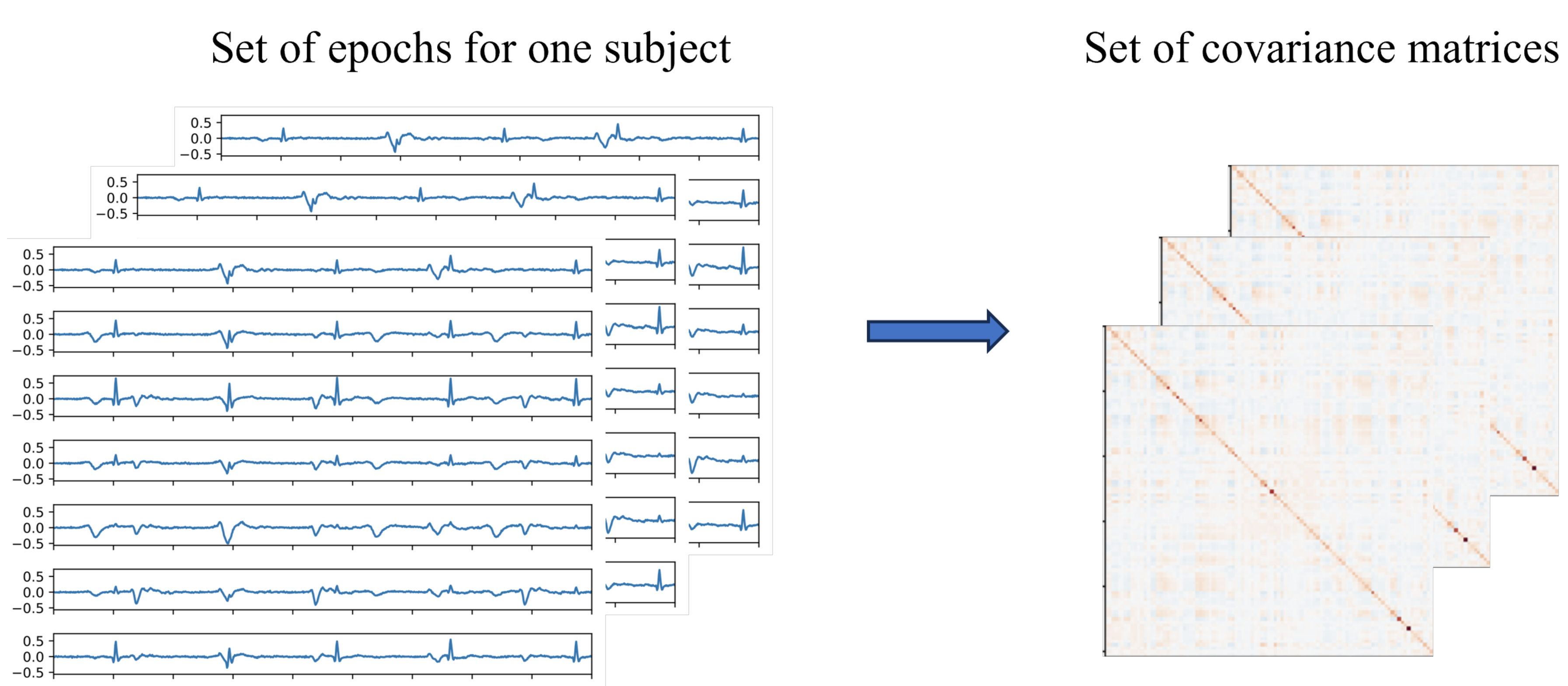
SPDSW_p metrizes the weak convergence.

Complexity *w.r.t* number of samples n and the number of projections L: $O(Ln(\log n + d^2) + (L+n)d^3)$.



Brain-age: subjects as distributions in $\mathcal{P}_p(S_d^{++}(\mathbb{R}))$

One subject's recording can be split into multiple frames (epochs) to obtain a set of covariance matrices, and thus an **empirical distribution of SPD matrices** after rank reduction.



Brain-age: Kernel Ridge regression with SPDSW

Positive definite Gaussian Kernel with SPDSW

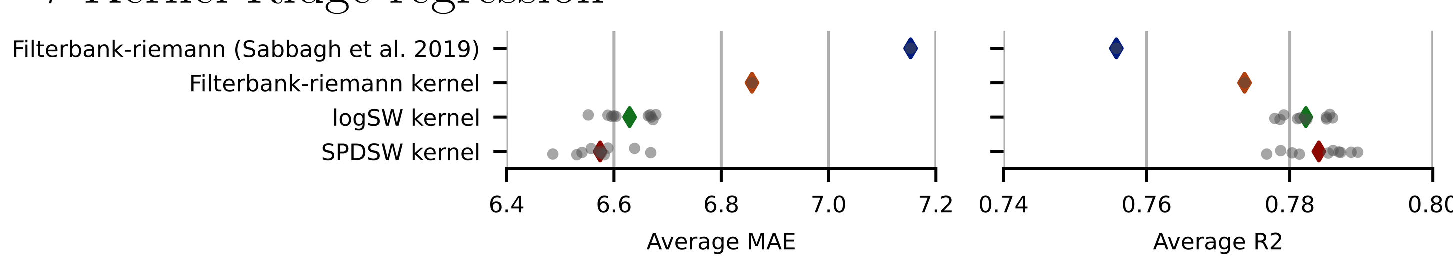
$$K(\mu,\nu) = e^{-\gamma \mathrm{SPDSW}_2^2(\mu,\nu)} = e^{-\gamma\|\Phi(\mu)-\Phi(\nu)\|_{\mathcal{H}}^2}$$

Known feature map Φ, no need for expensive quadratic computations

Application to MEG data. Computation of a Kernel K_f for each frequency band f of interest, and sum

$$K = \sum_f K_f$$

→ Kernel Ridge regression



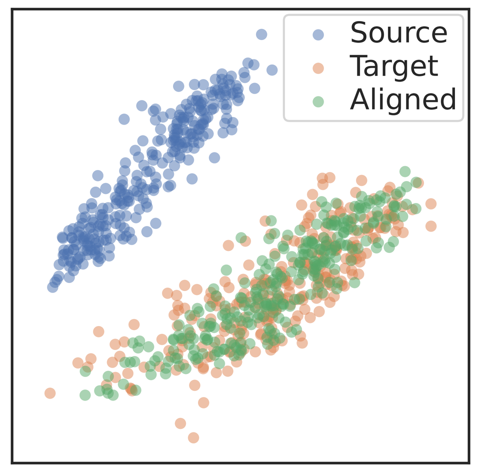
Domain adaptation for BCI

Learning a map f_θ between a source μ and a target ν

$$\min_\theta\ \mathrm{SPDSW}_2^2((f_\theta)_\#\mu, \nu)$$

Minimizing over the particles

$$\min_{(x_i)_{i=1}^n}\ \mathrm{SPDSW}_2^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \nu\right)$$



	Source	AISOTDA	SPDSW	LogSW	LEW	LES	SPDSW	LogSW	LEW	LES
			Transformations in $S_d^{++}(\mathbb{R})$				Descent over particles			
Avg. acc.	77.87	82.93	82.43	80.76	82.24	82.54	83.71	83.12	79.13	78.63
Avg. time (s)	-	-	**4.34**	**4.32**	11.41	12.04	**3.68**	**3.67**	8.50	11.43

References

Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, D., Engemann, D. A. Manifold-regression to Predict from MEG/EEG Brain Signals without Source Modeling. *Advances in Neural Information Processing Systems*, 32, 2019.