Data Analysis with the Riemannian Geometry of SPD Matrices

Ronen Talmon

Technion – IIT

*Joint work with Or Yair and Mirela Ben-Chen

IPAM Workshop III: Geometry of Big Data
April 2019

Domain Adaptation

- **P:** Datasets often comprise multiple domains:
  - Sessions
  - Subjects
  - Batches
- **Q:** How to adapt a given model that is well performing on a particular domain to a different yet related domain?
- **Q:** How to train a classifier based on data from one domain and apply it to data from another domain?

Domain Adaptation

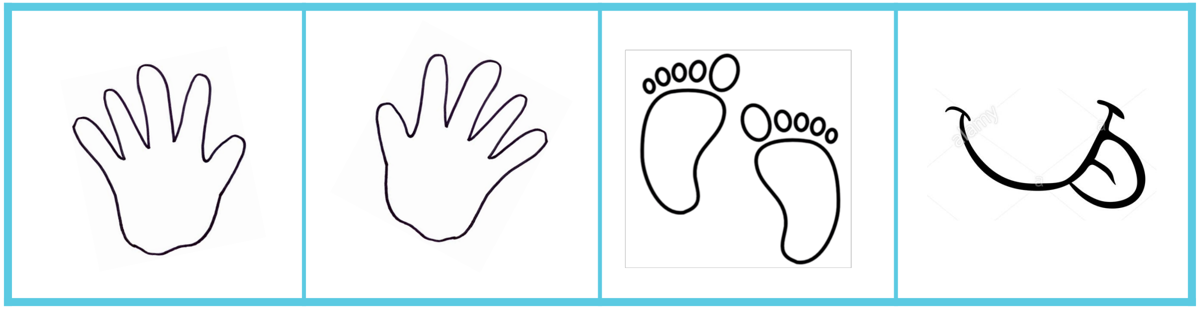
- Highly researched subject
- Many previous studies, e.g.:
  - [Ben-David et al, 07]
  - [Raina et al, 07]
  - [Dai et al, 09]
  - [Pan et al, 10]
  - …

Data Analysis in High Dimension

- **P:** The data do not live in a Euclidean space
  - Multiple modalities
  - Dependence between coordinates
- **Q:** What is the proper non-Euclidean metric?
- **Q:** How to find an embedding into a Euclidean space?
- **P:** Unsupervised

Illustrative Application

- Brain Computer Interface (BCI) [Barachant et al, 13]
- Data: [recent BCI competition]
  - EEG from 9 subjects
  - 22 electrodes
  - 2 days of experiments
  - 288 trials per subject
- In each trial, imagine performing 1 of 4 motor tasks:
  - Left hand
  - Right hand
  - Both feet
  - Tongue



Illustrative Application

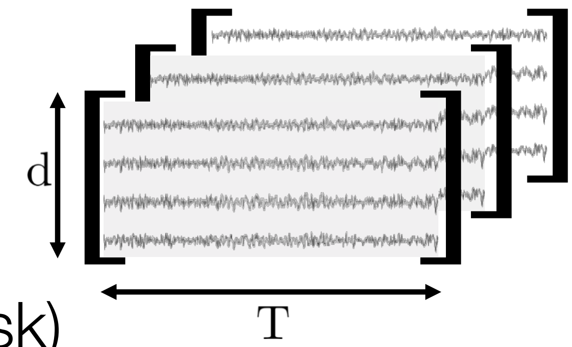
- Consider two datasets from two subjects:

$$\left\{ \boldsymbol{X}_i^{(1)}, y_i^{(1)} \right\}_{i=1}^{N_1} \qquad \left\{ \boldsymbol{X}_i^{(2)}, y_i^{(2)} \right\}_{i=1}^{N_2}$$

(target) (source)

  - Each set contains N_k matrices of observations

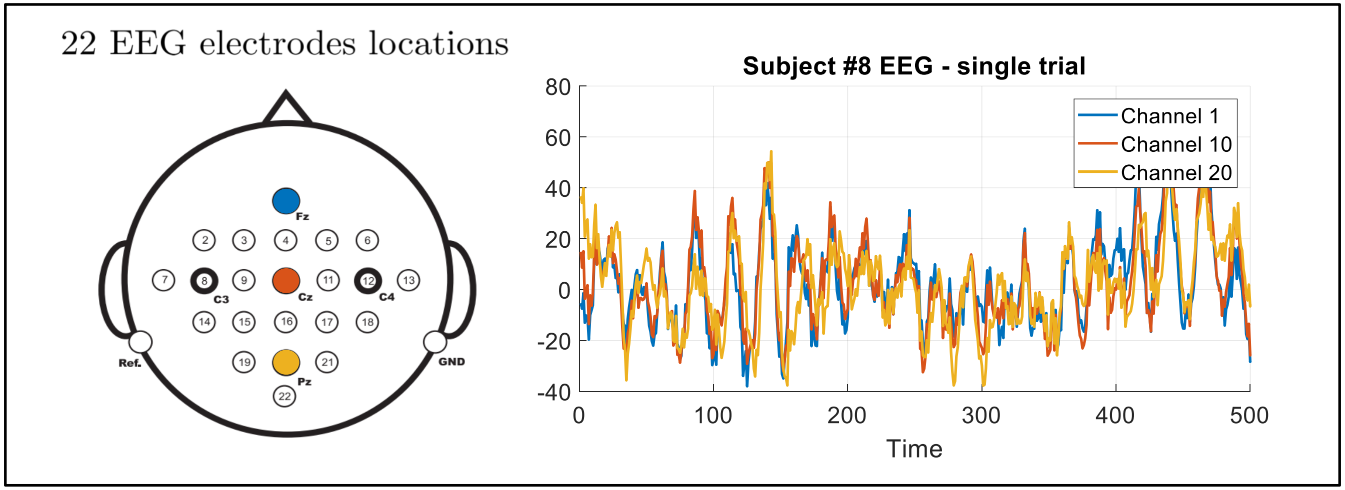
$$\boldsymbol{X}_i^{(k)} \in \mathbb{R}^{d \times T_i^{(k)}}$$



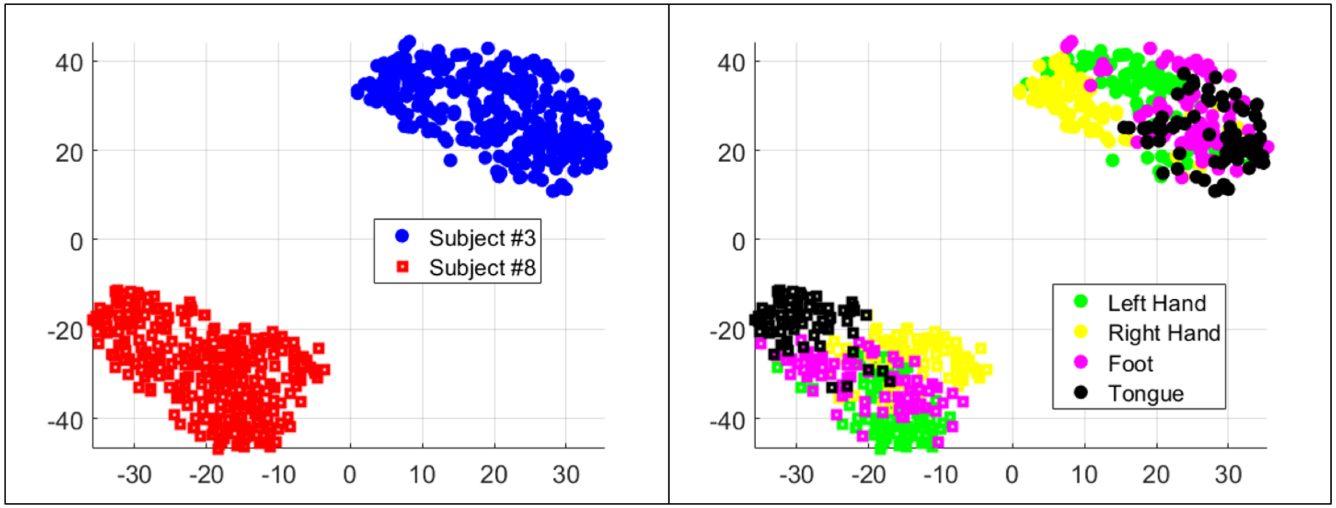
  - d - dimension (# of EEG electrodes)
  - $T_i^{(k)}$ - observation length
  - $y_i^{(k)}$ - hidden label (imagined motor task)

- Let $\boldsymbol{P}_i^{(k)} \in \mathbb{R}^{d \times d}$ be the (sample) covariance of $\boldsymbol{X}_i^{(k)}$

Illustrative Application



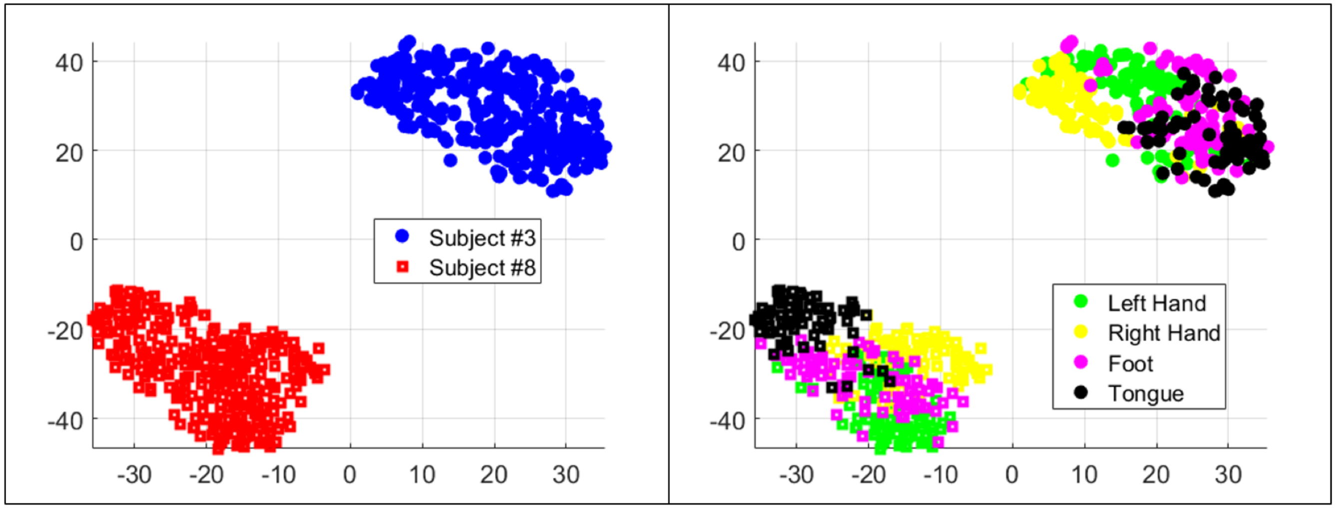
Low dimensional representation of the covariance matrices from **two** subjects



Illustrative Application

Identify the imagined motor activity per trial *from multiple subjects*

- → Training a classifier from one subject and testing on another subject
- → Unsupervised



Our Solution

Riemannian Geometry of SPD matrices

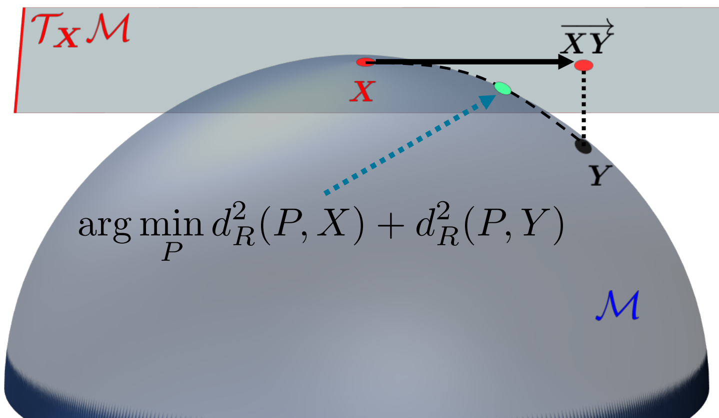
Benefits

- *Known* non-Euclidean space
  facilitating comparisons, additions, subtractions
- Joint representation from multiple domains
- Following recent work:
  - Theory [Pennec et al, 06], [Sra & Hosseini, 15]
  - Applications in BCI [Barachant et al, 13]
  - Applications in computer vision [Tuzel el al, 08], [Freifeld et al, 14], [Bergman et al, 17]

Preliminaries on Riemannian Geometry

- Let $x, y \in \mathcal{M}$ be two points on a Riemannian manifold
- Let $\mathcal{T}_x\mathcal{M}$ be the tangent plane at the point x
- Define the following operations:

	Vector Space	Riemannian Manifold
Subtraction	$\overrightarrow{\boldsymbol{x}\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{x}$	$\overrightarrow{\boldsymbol{x}\boldsymbol{y}} = \mathrm{Log}_{\boldsymbol{x}}(\boldsymbol{y})$
Addition	$\boldsymbol{y} = \boldsymbol{x} + \overrightarrow{\boldsymbol{x}\boldsymbol{y}}$	$\boldsymbol{y} = \mathrm{Exp}_{\boldsymbol{x}}(\overrightarrow{\boldsymbol{x}\boldsymbol{y}})$
Mean	$\arg\min_{\boldsymbol{x}} \sum_i \lVert \boldsymbol{x} - \boldsymbol{x}_i \rVert_i^2$	$\arg\min_{\boldsymbol{x}} \sum_i d_R^2(\boldsymbol{x}, \boldsymbol{x}_i)$

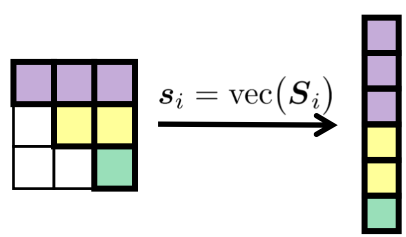


The SPD Cone

- The SPD matrices constitute a convex half-cone in the space of real symmetric matrices
- This cone forms a differentiable Riemannian manifold $\mathcal{M}$ equipped with the inner product

$$\langle S_1, S_2 \rangle_{\mathcal{T}_P\mathcal{M}} = \langle P^{-\frac{1}{2}} S_1 P^{-\frac{1}{2}}, P^{-\frac{1}{2}} S_2 P^{-\frac{1}{2}} \rangle$$

  - $\mathcal{T}_P\mathcal{M}$ is the tangent plane at $P \in \mathcal{M}$
  - $S_1, S_2 \in \mathcal{T}_P\mathcal{M}$
  - $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product
- The symmetric matrices $S \in \mathcal{T}_P\mathcal{M}$ live in a linear space
  - We can view them as vectors



With $\sqrt{2}$ scaling on the off-diagonal elements

The SPD Cone – Properties

- There exists a **unique geodesic curve** between any two SPD matrices $P_1, P_2 \in \mathcal{M}$:

$$\varphi(t) = P_1^{\frac{1}{2}} \left(P_1^{-\frac{1}{2}} P_2 P_1^{-\frac{1}{2}}\right)^t P_1^{\frac{1}{2}}, \qquad 0 \leq t \leq 1$$

- Define a **Riemannian distance** on the manifold as the arc-length of the geodesic curve:

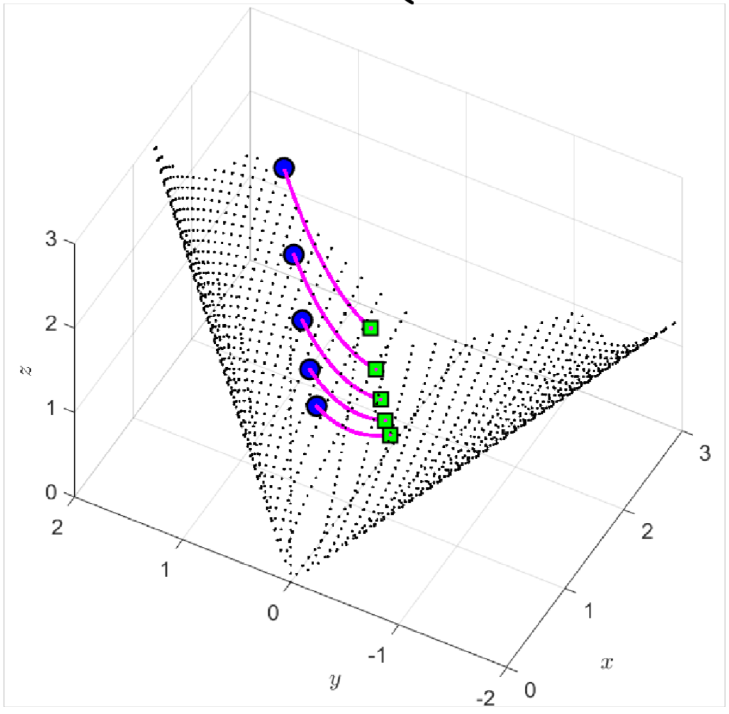
$$\begin{aligned} d_R^2(P_1, P_2) &= \left\| \log \left(P_2^{-\frac{1}{2}} P_1 P_2^{-\frac{1}{2}}\right) \right\|_F^2 \\ &= \sum_{i=1}^n \log^2 \left(\lambda_i \left(P_2^{-\frac{1}{2}} P_1 P_2^{-\frac{1}{2}}\right)\right) \end{aligned}$$

  - $\lambda_i(P)$ is the i-th eigenvalue of P
  - Scale-invariant

The SPD Cone – Properties

$$M = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$$

$$\Rightarrow \begin{cases} \mathrm{Tr}\,(M) = x + z > 0 \\ |(M)| = xz - y^2 > 0 \end{cases} \Rightarrow \begin{cases} x > 0 \\ z > 0 \end{cases} \Rightarrow \boxed{|y| < \sqrt{xz}}$$



The SPD Cone – Properties

- Logarithm map:

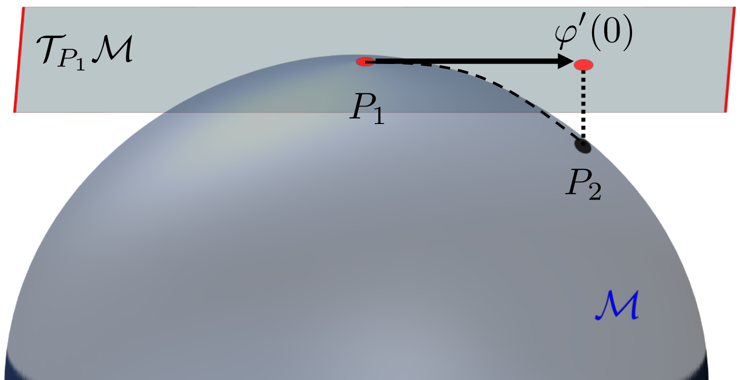
$$S_1 = \mathrm{Log}_P(P_1) = P^{\frac{1}{2}} \log\left(P^{-\frac{1}{2}} P_1 P^{-\frac{1}{2}}\right) P^{\frac{1}{2}} \in \mathcal{T}_P \mathcal{M}$$

- Exponential map:

$$P_1 = \mathrm{Exp}_P(S_1) = P^{\frac{1}{2}} \exp\left(P^{-\frac{1}{2}} S_1 P^{-\frac{1}{2}}\right) P^{\frac{1}{2}} \in \mathcal{M}$$

- Relation to the (unique) geodesic curve $\varphi(t)$ from P_1 to P_2 is given by the initial velocity

$$\varphi'(0) = \mathrm{Log}_{P_1}(P_2) \in \mathcal{T}_{P_1}\mathcal{M}$$



Domain Adaptation

Formulation

- Consider two subsets $\mathcal{P}^{(1)}$ (target) and $\mathcal{P}^{(2)}$ (source) of SPD matrices from *two different domains*
  - $\overline{P}^{(1)}$ and $\overline{P}^{(2)}$- Riemannian means
  - $\varphi(t)$ - the unique geodesic from $\overline{P}^{(2)}$ to $\overline{P}^{(1)}$
  - $S_i^{(k)}$ - the symmetric matrix (or vector) in $\mathcal{T}_{\overline{P}^{(k)}}\mathcal{M}$

$$S_i^{(k)} = \mathrm{Log}_{\overline{P}^{(k)}}\left(P_i^{(k)}\right)$$

Domain Adaptation

Formulation

- **Goal:**
  Derive a new representation $\Gamma\left(S_i^{(2)}\right)$:
  $$\Gamma : \mathcal{T}_{\overline{P}^{(2)}}\mathcal{M} \to \mathcal{T}_{\overline{P}^{(1)}}\mathcal{M}$$
  so that $\left\{S_i^{(1)}\right\}$ and $\left\{\Gamma\left(S_i^{(2)}\right)\right\}$ live in the same linear space
- **Benefit:**
  Relate samples from the two subsets
  - Compute quantities such as $\left\langle S_i^{(1)}, \Gamma\left(S_j^{(2)}\right)\right\rangle_{\overline{P}^{(1)}}$

Domain Adaptation

Formulation

- **Constraints:**
  - Zero mean:

$$\frac{1}{N_2}\sum_{i=1}^{N_2}\Gamma\left(S_i^{(2)}\right) = \frac{1}{N_1}\sum_{i=1}^{N_1} S_i^{(1)} = 0$$

  - Inner product preservation:

$$\left\langle \Gamma\left(S_i^{(2)}\right), \Gamma\left(S_j^{(2)}\right)\right\rangle_{\overline{P}^{(1)}} = \left\langle S_i^{(2)}, S_j^{(2)}\right\rangle_{\overline{P}^{(2)}}$$

  - Geodesic velocity preservation:

$$\Gamma\left(\varphi'\left(0\right)\right) = \varphi'\left(1\right)$$

Domain Adaptation

Formulation

- **Constraints:**
  - Zero mean:

$$\frac{1}{N_2}\sum_{i=1}^{N_2}\Gamma\left(S_i^{(2)}\right) = \frac{1}{N_1}\sum_{i=1}^{N_1} S_i^{(1)} = 0$$

  - Inner product preservation:

$$\left\langle \Gamma\left(S_i^{(2)}\right), \Gamma\left(S_j^{(2)}\right)\right\rangle_{\overline{P}^{(1)}} = \left\langle S_i^{(2)}, S_j^{(2)}\right\rangle_{\overline{P}^{(2)}}$$

Imply that the map preserves inter-sample relations

Domain Adaptation

Formulation

- **Constraints:**
  - Zero mean:

$$\frac{1}{N_2}\sum_{i=1}^{N_2}\Gamma\left(S_i^{(2)}\right) = \frac{1}{N_1}\sum_{i=1}^{N_1} S_i^{(1)} = 0$$

  - Inner product preservation:

$$\left\langle \Gamma\left(S_i^{(2)}\right), \Gamma\left(S_j^{(2)}\right)\right\rangle_{\overline{P}^{(1)}} = \left\langle S_i^{(2)}, S_j^{(2)}\right\rangle_{\overline{P}^{(2)}}$$

Not unique:
If Γ admits to these properties, then also $R \circ \Gamma$
where R is an arbitrary rotation

Domain Adaptation

Formulation

- **Constraints:**
  - Zero mean:

$$\frac{1}{N_2}\sum_{i=1}^{N_2}\Gamma\left(S_i^{(2)}\right)=\frac{1}{N_1}\sum_{i=1}^{N_1}S_i^{(1)}=0$$

  - Inner product preservation:

$$\left\langle\Gamma\left(S_i^{(2)}\right),\Gamma\left(S_j^{(2)}\right)\right\rangle_{\overline{P}^{(1)}}=\left\langle S_i^{(2)},S_j^{(2)}\right\rangle_{\overline{P}^{(2)}}$$

Simple implementation by mean subtraction

Domain Adaptation

Formulation

- **Constraints:**
  - Geodesic velocity preservation:

$$\Gamma\left(\varphi'\left(0\right)\right) = \varphi'\left(1\right)$$

Use the *unique* geodesic to resolve the arbitrary degree of freedom

The two intrinsic symmetric matrices (vectors) $\varphi'(0) \in \mathcal{T}_{\overline{P}^{(2)}}\mathcal{M}$ and $\varphi'(1) \in \mathcal{T}_{\overline{P}^{(1)}}\mathcal{M}$ are used to fix the rotation

Domain Adaptation

Formulation

- **Constraints:**
  - Geodesic velocity preservation:

$$\Gamma\left(\varphi'\left(0\right)\right) = \varphi'\left(1\right)$$

Unsupervised – no labels are used

Present a closed-form expression (no optimization)

Parallel Transport

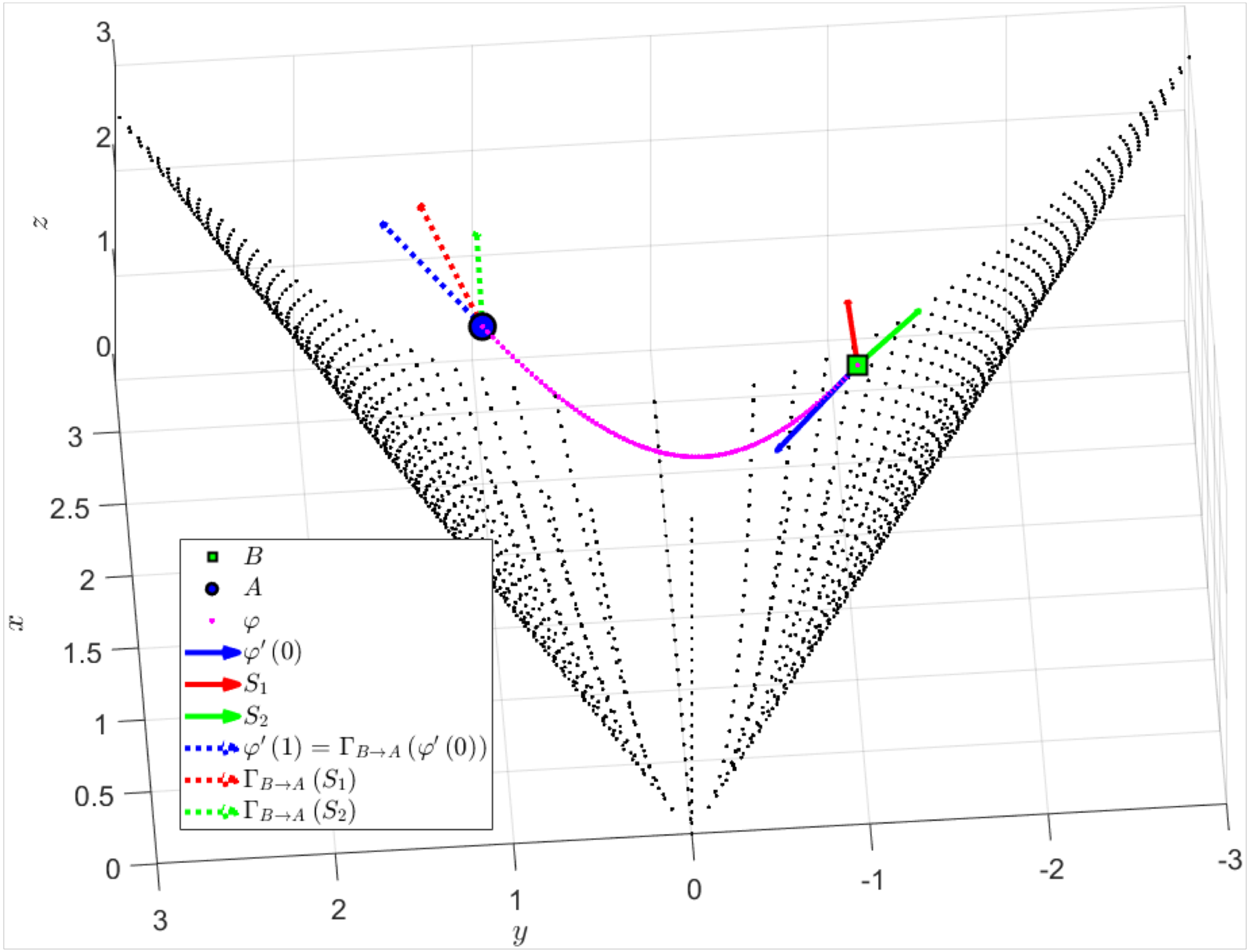
Lemma (Parallel Transport)
Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}$.
The PT from $\boldsymbol{B}$ to $\boldsymbol{A}$ of any $\boldsymbol{S} \in \mathcal{T}_{\boldsymbol{B}}\mathcal{M}$ is:

$$\Gamma_{\boldsymbol{B} \to \boldsymbol{A}}\left(\boldsymbol{S}\right) \triangleq \boldsymbol{E}\boldsymbol{S}\boldsymbol{E}^T$$

where $\boldsymbol{E} = \left(\boldsymbol{A}\boldsymbol{B}^{-1}\right)^{\frac{1}{2}}$.

Theorem.
The representation $\Gamma_{\overline{\boldsymbol{P}}^{(2)} \to \overline{\boldsymbol{P}}^{(1)}}\left(\boldsymbol{S}_i^{(2)}\right)$,
i.e., the unique PT of $\boldsymbol{S}_i^{(2)}$ from $\overline{\boldsymbol{P}}^{(2)}$ to $\overline{\boldsymbol{P}}^{(1)}$,
is well defined and satisfies properties $(1) - (3)$.

Parallel Transport



Domain Adaptation with PT

Algorithm:

1. Project the SPD matrix $\boldsymbol{P}_i^{(2)}$ to the tangent plane $\mathcal{T}_{\overline{\boldsymbol{P}}^{(2)}}\mathcal{M}$

$$\boldsymbol{S}_i^{(2)} = \mathrm{Log}_{\overline{\boldsymbol{P}}^{(2)}}\left(\boldsymbol{P}_i^{(2)}\right)$$

2. Parallel transport $\boldsymbol{S}_i^{(2)}$ from $\overline{\boldsymbol{P}}^{(2)}$ to $\overline{\boldsymbol{P}}^{(1)}$ by computing

$$\boldsymbol{S}_i^{(2)\to(1)} = \Gamma_{\overline{\boldsymbol{P}}^{(2)}\to\overline{\boldsymbol{P}}^{(1)}}\left(\boldsymbol{S}_i^{(2)}\right)$$

3. Project the symmetric matrix $\boldsymbol{S}_i^{(2)\to(1)} \in \mathcal{T}_{\overline{\boldsymbol{P}}^{(1)}}\mathcal{M}$ back to the manifold using $\mathrm{Exp}_{\overline{\boldsymbol{P}}^{(1)}}\left(\boldsymbol{S}_i^{(2)\to(1)}\right)$.

Domain Adaptation with PT

Define the map $\Psi : \mathcal{M} \to \mathcal{M}$ that adapts the domain of $\mathcal{P}^{(2)}$ to the domain of $\mathcal{P}^{(1)}$

$$\Psi\big(P_i^{(2)}\big) = \mathrm{Exp}_{\overline{P}^{(1)}}\left(\Gamma_{\overline{P}^{(2)} \to \overline{P}^{(1)}}\left(\mathrm{Log}_{\overline{P}^{(2)}}\big(P_i^{(2)}\big)\right)\right)$$

for any $P_i^{(2)} \in \mathcal{P}^{(2)}$

Implementation

Theorem.
Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P} \in \mathcal{M}$ and let $\boldsymbol{S} = \mathrm{Log}_{\boldsymbol{B}}(\boldsymbol{P}) \in \mathcal{T}_{\boldsymbol{B}}\mathcal{M}$.
Then,

$$\mathrm{Exp}_{\boldsymbol{A}}\left(\Gamma_{\boldsymbol{B}\to\boldsymbol{A}}\left(\boldsymbol{S}\right)\right) = \boldsymbol{E}\boldsymbol{P}\boldsymbol{E}^T,$$

where $\boldsymbol{E} = \left(\boldsymbol{A}\boldsymbol{B}^{-1}\right)^{\frac{1}{2}}$.

- Ψ can be efficiently implemented

$$\Psi\left(P_i^{(2)}\right) = \Gamma_{\overline{P}^{(2)}\to\overline{P}^{(1)}}\left(P_i^{(2)}\right) = EP_i^{(2)}E^T$$

$$E \triangleq \left(\overline{P}^{(1)}\left(\overline{P}^{(2)}\right)^{-1}\right)^{\frac{1}{2}}$$

Implementation

- **Important consequence:**
  - The covariance adaptation:

$$E P_i^{(2)} E^T$$

    where

$$E \triangleq \left(\overline{P}^{(1)} \left(\overline{P}^{(2)} \right)^{-1} \right)^{\frac{1}{2}}$$

  - Can be applied directly to data by:

$$E X_i^{(2)}$$

Toy Problem

- Consider the set of hidden multi-dimensional times series $\{s_i[n]\}_{i=1}^{100}$:

$$\boldsymbol{s}_i\left[n\right]=\begin{bmatrix}\sin\left(2\pi f_0 n/T\right)\\ \cos\left(2\pi f_0 n/T+\phi_i\right)\end{bmatrix}, n=0,\ldots,T-1$$

where $f_0=10$, $T=500$, and $\phi_i \sim U\left[-\pi/2,0\right]$

- Short segments of two oscillatory signals
- Governed by a 1-dimensional hidden variable ϕ_i (the initial phase of the oscillations)

Toy Problem

- The population covariance of $\boldsymbol{s}_i[n]$ is

$$\frac{1}{2}\begin{bmatrix} 1 & -\sin\left(\phi_i\right) \\ -\sin\left(\phi_i\right) & 1 \end{bmatrix}$$

  which depends only on ϕ_i

- <u>Note</u>:
  when presenting the population covariances as vectors in $\mathbb{R}^3$, two coordinates are fixed and only one varies

Toy Problem

- We generate two observable subsets

$$\mathcal{X}^{(1)} = \{\boldsymbol{x}_i^{(1)}[n]\}_{i=1}^{100}, \mathcal{X}^{(2)} = \{\boldsymbol{x}_i^{(2)}[n]\}_{i=1}^{100}$$

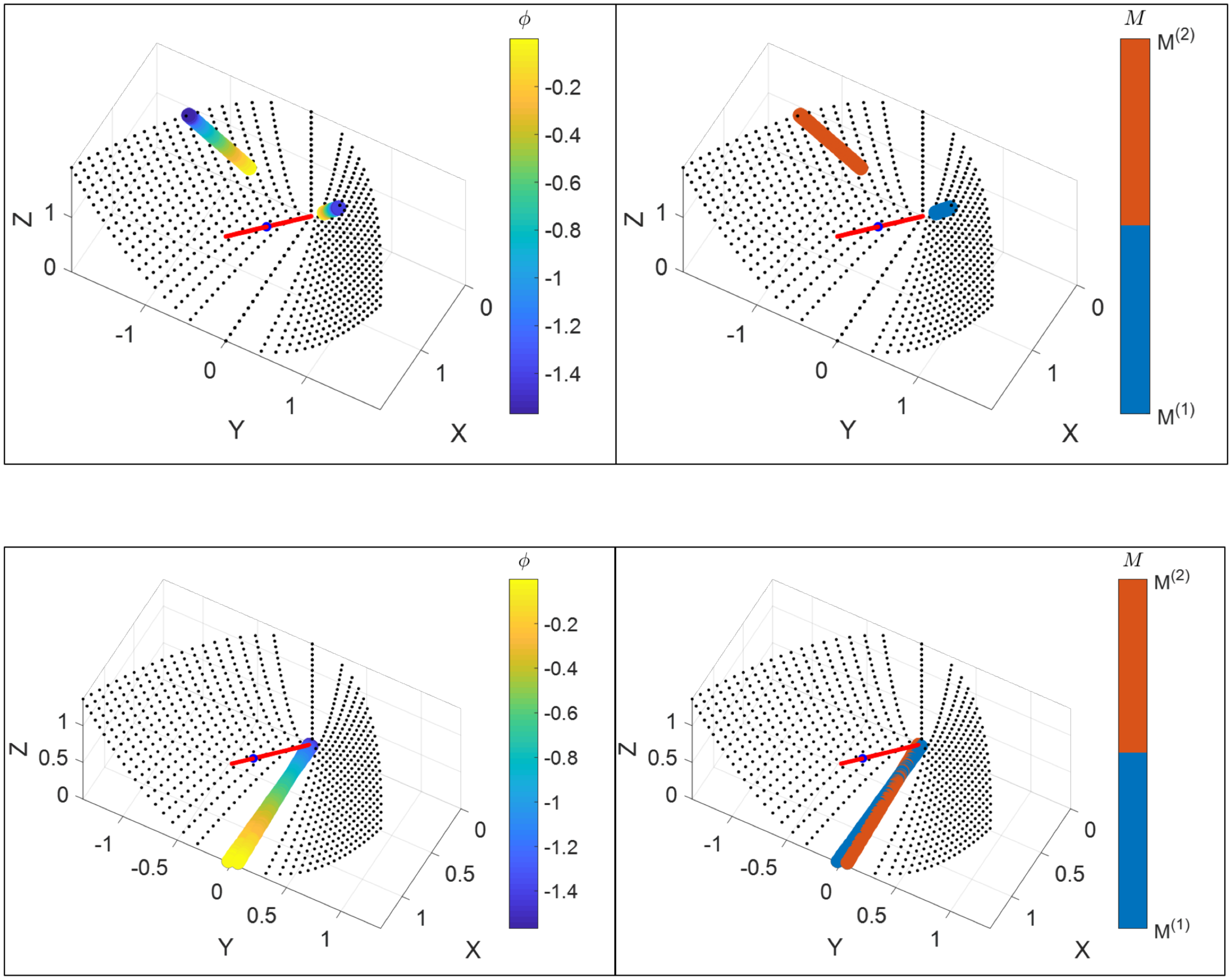
such that:

$$x_i^{(k)}[n] = M^{(k)} s_i[n]$$

where

- $M^{(1)}$ is randomly chosen
- $M^{(2)} = 1.5 \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} M^{(1)}$

Toy Problem



After PT*

* PT of the source and target sets to the mid-point

Brain Computer Interface

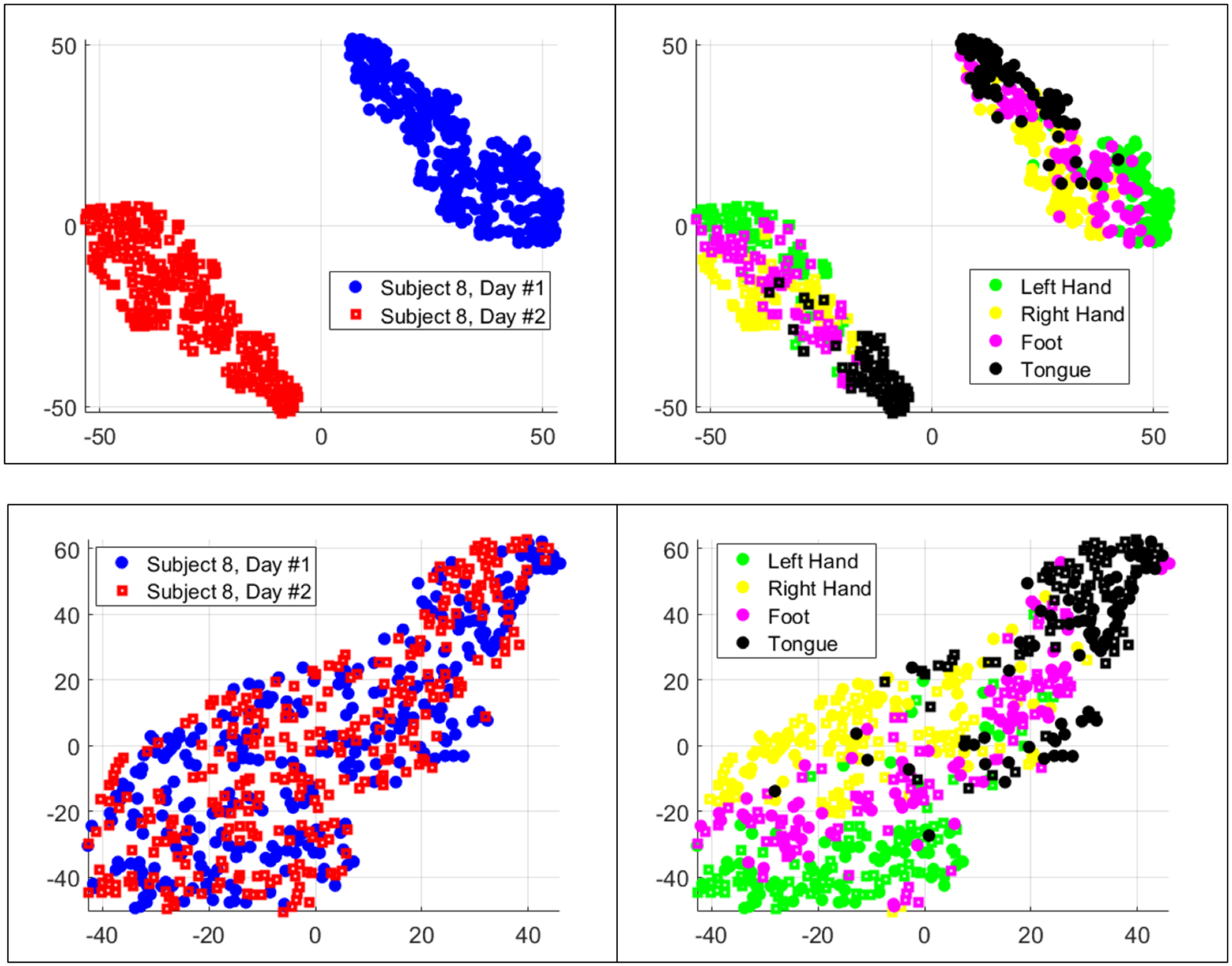
- Data: [BCI competition]
  - 22 EEG electrodes
  - 9 subjects
  - 2 days of experiments
  - 288 trials per subject
  - In each trial the subject is asked to imagine performing 1 out of 4 motor tasks:
    left hand, right hand, both feet, tongue

Brain Computer Interface

tSNE [v. d. Maaen & Hinton, 08] representation of Covariance matrices

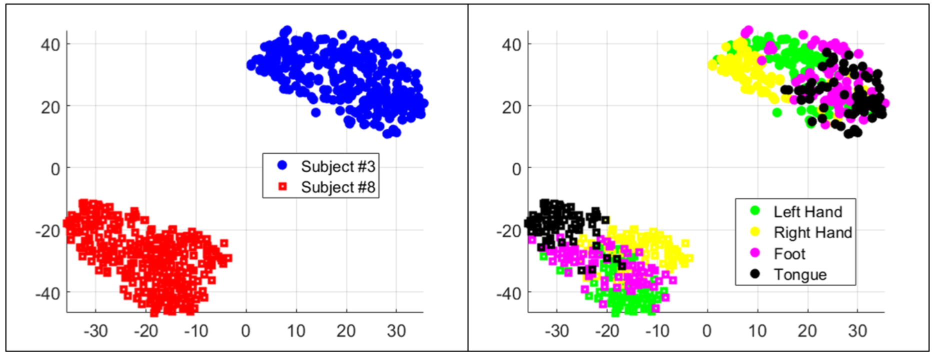
One subject
Two sessions

After PT

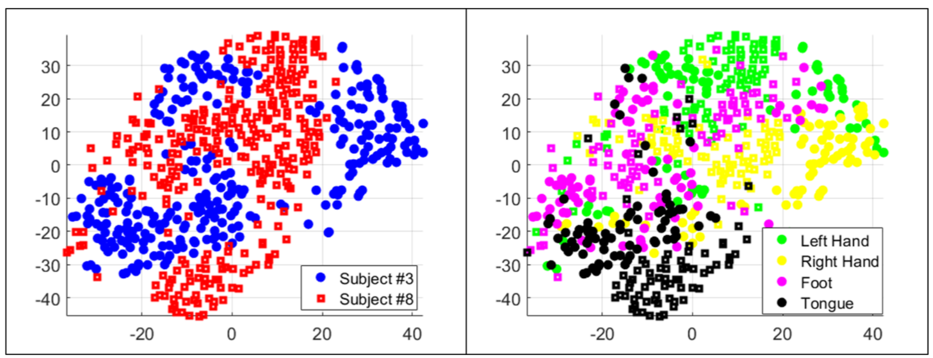


Brain Computer Interface

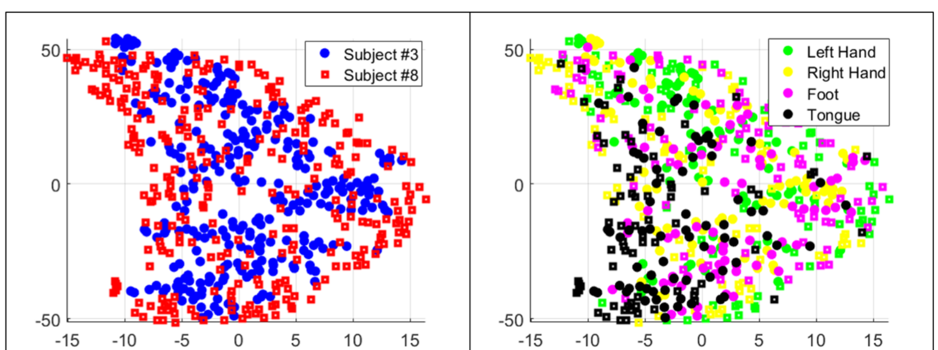
Two subjects



After PT



Mean subtraction
[Barachant et al, 13]



Intrinsicness

- Using *affine transformation* [Zanini et al, 18]:

$$\left(\overline{P}^{(k)}\right)^{-\frac{1}{2}} P_i^{(k)} \left((\overline{P}^{(k)})^{-\frac{1}{2}}\right)^T$$

  - Equivalent to parallel transport to the identity:

$$\Psi\left(P_i^{(k)}\right) = \Gamma_{\overline{P}^{(k)} \to I}\left(P_i^{(k)}\right) = E P_i^{(k)} E^T$$

$$E \triangleq \left(I(\overline{P}^{(k)})^{-1}\right)^{\frac{1}{2}}$$

  - When there are two sets, it is equivalent to

$$\Gamma_{I \to \overline{P}^{(1)}} \circ \Gamma_{\overline{P}^{(2)} \to I}$$

  - Coincides with our work when

$$\Gamma_{\overline{P}^{(2)} \to \overline{P}^{(1)}} = \Gamma_{I \to \overline{P}^{(1)}} \circ \Gamma_{\overline{P}^{(2)} \to I}$$

    - $\overline{P}^{(1)}$ and $\overline{P}^{(2)}$ commute and have the same eigenvectors (PC)

- **Q:** what is special about the proposed transport?

Intrinsicness

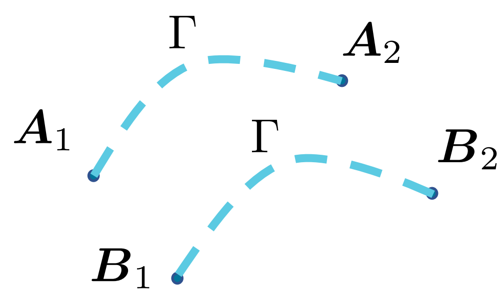
Definition (Equivalent Pairs)
Two pairs $(\boldsymbol{A}_1, \boldsymbol{B}_1)$ and $(\boldsymbol{A}_2, \boldsymbol{B}_2)$, such that $\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{A}_2, \boldsymbol{B}_2 \in \mathcal{M}$, are *equivalent* if there exists an invertible matrix $\boldsymbol{E}$ such that

$$\boldsymbol{A}_2 = \Gamma(\boldsymbol{A}_1) = \boldsymbol{E}\boldsymbol{A}_1\boldsymbol{E}^T$$

$$\boldsymbol{B}_2 = \Gamma(\boldsymbol{B}_1) = \boldsymbol{E}\boldsymbol{B}_1\boldsymbol{E}^T$$

We denote this relation by

$$(\boldsymbol{A}_1, \boldsymbol{B}_1) \sim (\boldsymbol{A}_2, \boldsymbol{B}_2)$$



Intrinsicness

Definition (Equivalent Pairs)
Two pairs $(\boldsymbol{A}_1, \boldsymbol{B}_1)$ and $(\boldsymbol{A}_2, \boldsymbol{B}_2)$, such that $\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{A}_2, \boldsymbol{B}_2 \in \mathcal{M}$, are *equivalent* if there exists an invertible matrix $\boldsymbol{E}$ such that

$$\boldsymbol{A}_2 = \Gamma(\boldsymbol{A}_1) = \boldsymbol{E}\boldsymbol{A}_1\boldsymbol{E}^T$$

$$\boldsymbol{B}_2 = \Gamma(\boldsymbol{B}_1) = \boldsymbol{E}\boldsymbol{B}_1\boldsymbol{E}^T$$

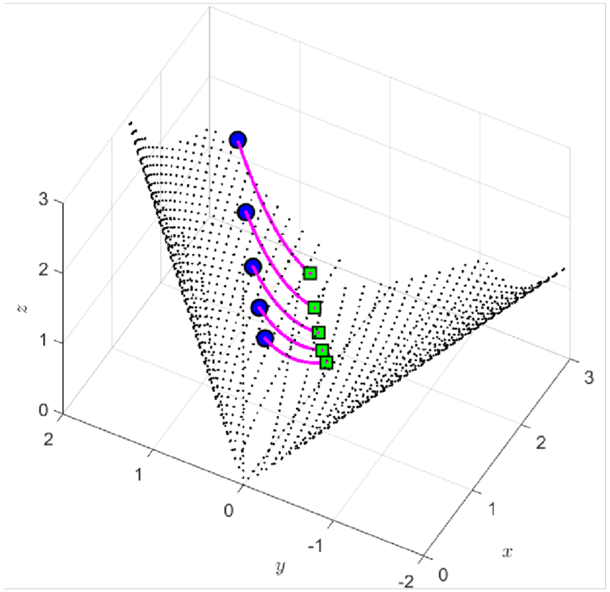
We denote this relation by

$$(\boldsymbol{A}_1, \boldsymbol{B}_1) \sim (\boldsymbol{A}_2, \boldsymbol{B}_2)$$

Lemma.
The relation $\sim$ is an equivalence relation, satisfying reflexivity, symmetry, and transitivity.

Intrinsicness

- Interpretation:
  - Equivalent pairs are matrices with equivalent intra-relations
  - E.g. if $(A_1, B_1) \sim (A_2, B_2)$ then
$$d_R(A_1, B_1) = d_R(A_2, B_2)$$
    but with a different global position on the manifold



Intrinsicness

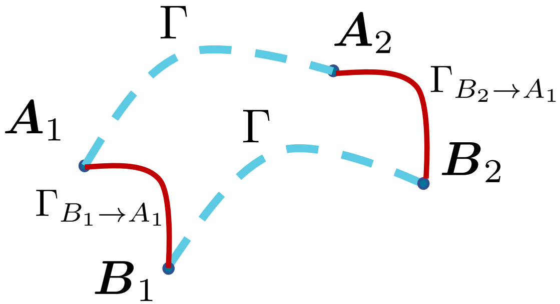
Proposition.
Let (A_1, B_1) be a pair of SPD matrices $A_1, B_1 \in \mathcal{M}$, and let $\big[(A_1, B_1)\big]$ denote the equivalence class

$$\big[(A_1, B_1)\big] = \big\{(A_2, B_2) \in \mathcal{M} \times \mathcal{M} \,|\, (A_2, B_2) \sim (A_1, B_1)\big\},$$

of all matrix pairs that are equivalent to (A_1, B_1).
Then, for any $(A_2, B_2) \in \big[(A_1, B_1)\big]$:

$$\Gamma \circ \Gamma_{B_1 \to A_1} = \Gamma_{B_2 \to A_2} \circ \Gamma\,,$$

where $\Gamma(P) = EPE^T$ and E satisfies the equivalence relation.



Intrinsicness

- Direct consequence:
  - Domain adaptation via Ψ is invariant to the relative position of $\overline{P}^{(1)}$ and $\overline{P}^{(2)}$ on the manifold
  - It is constructed equivalently for every pair in the equivalence class

$$[(\overline{P}^{(1)}, \overline{P}^{(2)})]$$

- Guarantees consistence
  - For example, two subjects in two sessions in the BCI problem

Extension to Multiple Subsets

Algorithm.
Input: $\{\boldsymbol{P}_i^{(1)}\}_{i=1}^{N_1}, \{\boldsymbol{P}_i^{(2)}\}_{i=1}^{N_2}, \dots, \{\boldsymbol{P}_i^{(K)}\}_{i=1}^{N_K}$
Output: $\{\tilde{\boldsymbol{S}}_i^{(1)}\}_{i=1}^{N_1}, \{\tilde{\boldsymbol{S}}_i^{(2)}\}_{i=1}^{N_2}, \dots, \{\tilde{\boldsymbol{S}}_i^{(K)}\}_{i=1}^{N_K}$

1. **For** each $k \in \{1, 2, \dots, K\}$, compute $\overline{\boldsymbol{P}}^{(k)}$ the Riemannian mean of the subset $\{\boldsymbol{P}_i^{(k)}\}$.

2. Compute $\hat{\boldsymbol{P}}$, the Riemannian mean of $\{\overline{\boldsymbol{P}}^{(k)}\}_{k=1}^{K}$.

3. **For** all k and all i, apply Parallel Transport using:

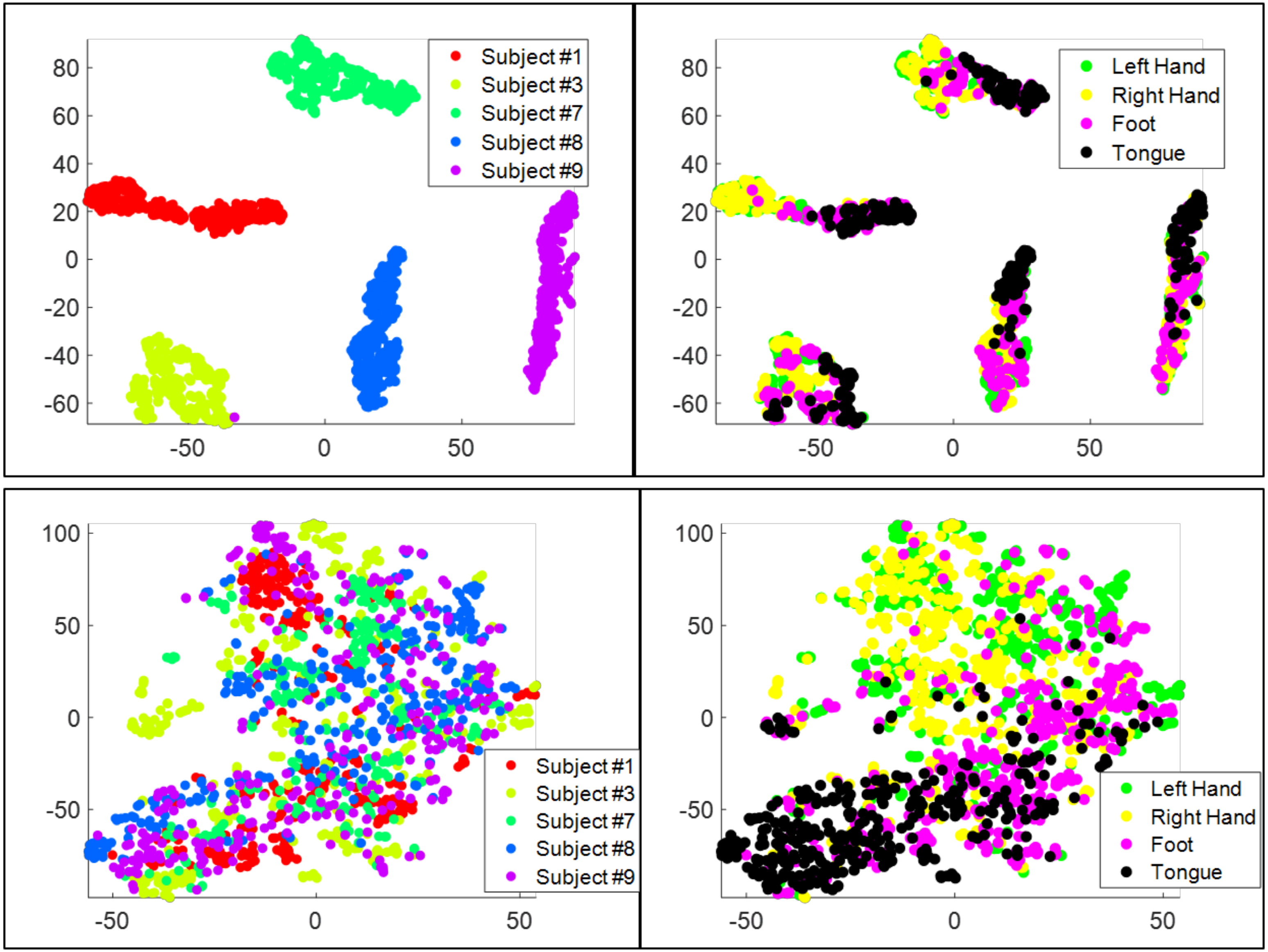
$$\boldsymbol{\Gamma}_i^{(k)} = \Gamma_{\overline{\boldsymbol{P}}^{(k)} \to \hat{\boldsymbol{P}}}\left(\boldsymbol{P}_i^{(k)}\right).$$

4. **For** all k and all i, project the transported matrix to the tangent space via:

$$\tilde{\boldsymbol{S}}_i^{(k)} = \log\left(\hat{\boldsymbol{P}}^{-\frac{1}{2}} \boldsymbol{\Gamma}_i^{(k)} \hat{\boldsymbol{P}}^{-\frac{1}{2}}\right).$$

Brain Computer Interface

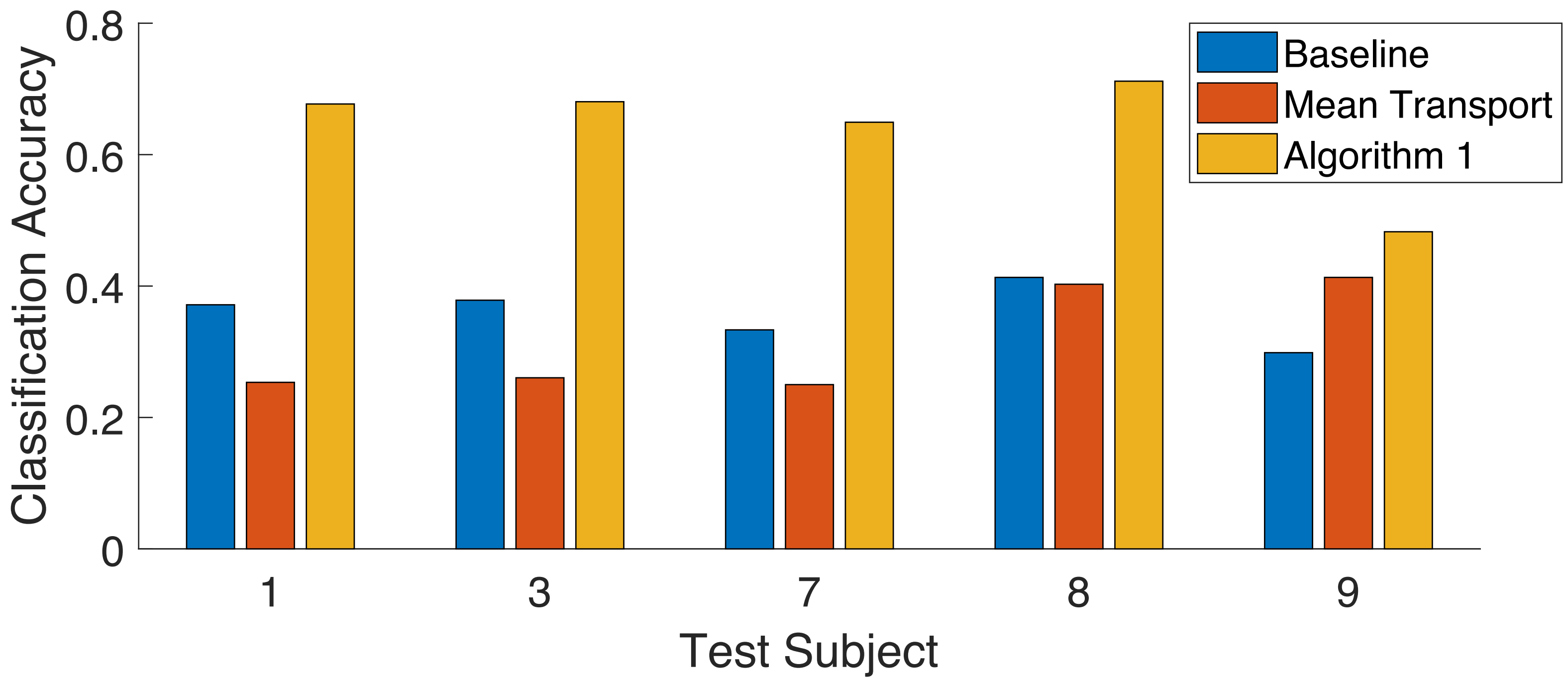
Five subjects



After PT

Brain Computer Interface

- Objective evaluation via classification
  - Leave-one-subject-out
  - Linear SVM



Sleep Stage Identification

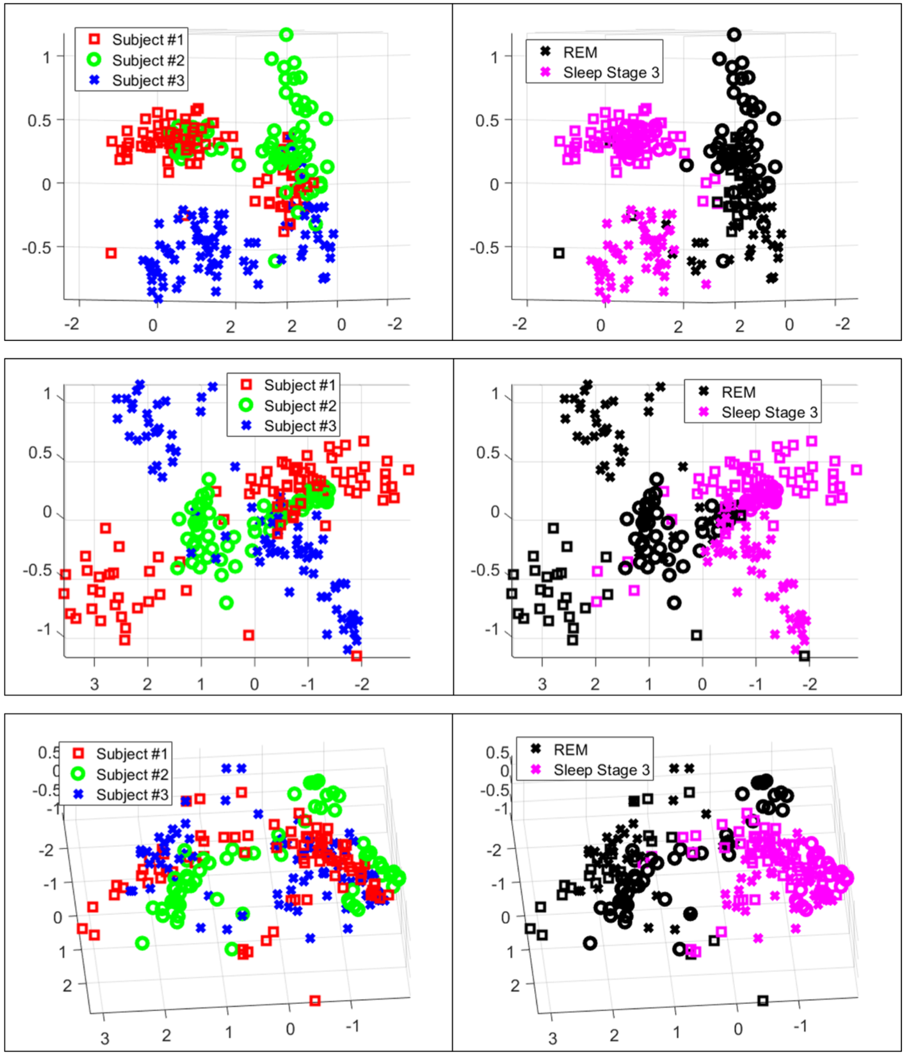
- Six different sleep stages:
  awake, REM, and sleep stages 1-4
- Recordings [PhysioNet.org]:
  - Two EEG channels
  - One electrooculography (EOG) channel
- Data from three subjects

Sleep Stage Identification

PCA of the covariance matrices[1]

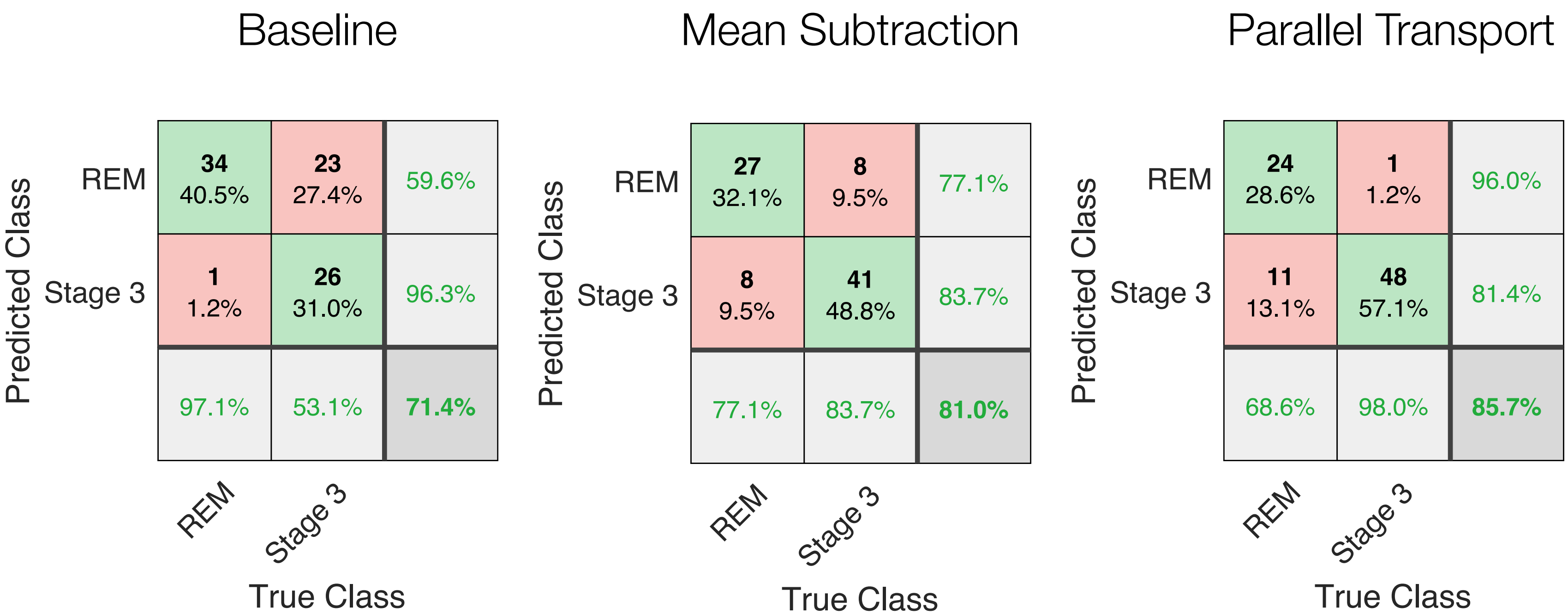
Mean subtraction

Parallel transport



[1]Since the covariance matrices are 3x3, dimension reduction using PCA was sufficient

Sleep Stage Identification

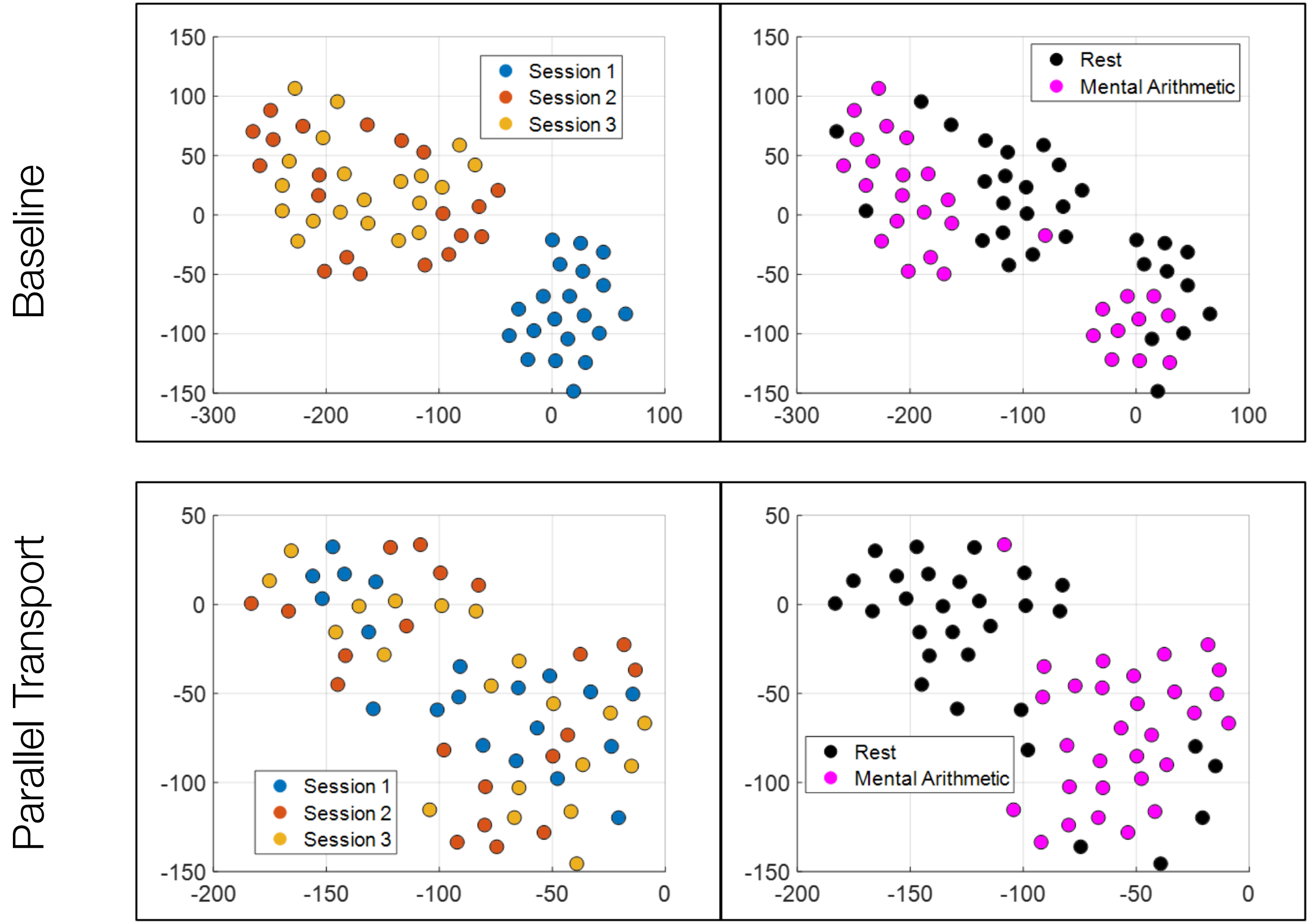


Mental Arithmetic Identification

- Recordings [Shin et al, 17]:
  - EEG from 29 subjects
  - 30 electrodes at 1000Hz
  - 3 sessions per subject
  - 20 repetitions/trials per session
- Two mental states:
  - Performing repeated simple arithmetic calculations
  - Baseline resting state

Mental Arithmetic Identification

tSNE representation of trials from subject #1



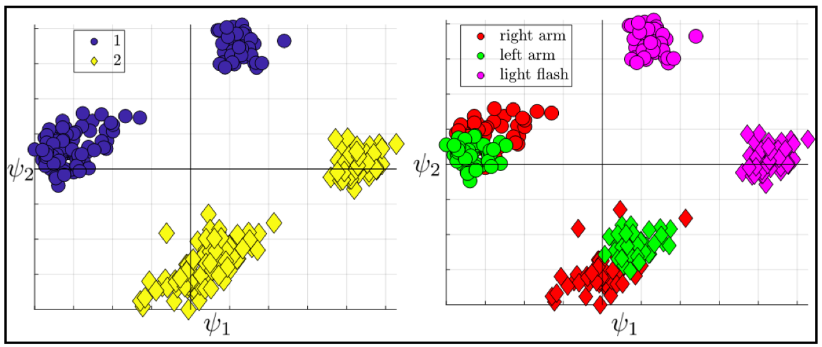
Mental Arithmetic Identification

- Average classification results over all 29 subjects
  - Leave-one-session-out cross-validation
  - Linear SVM

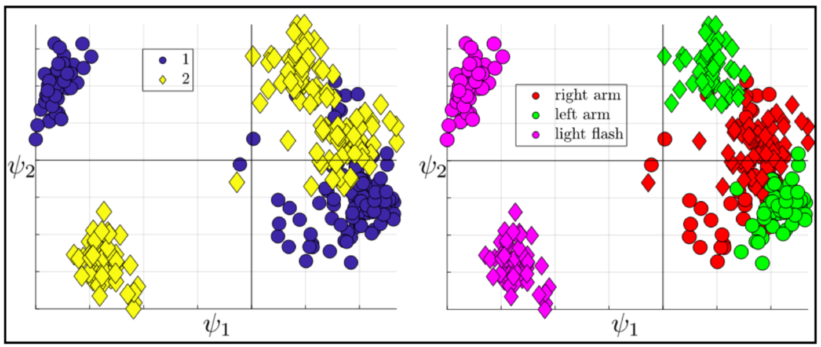
Baseline	Mean Subtraction	Parallel Transport
74%	73%	78%

Extensions and Outlook

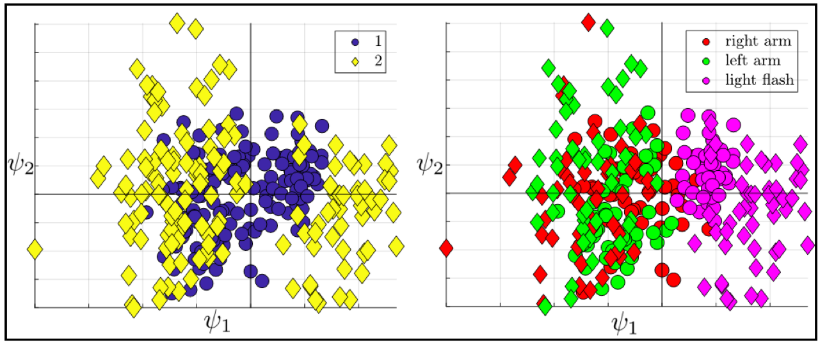
- To facilitate the internal structure of each subset:
  - Rotation following the PT
  - *Unsupervised moments alignment*



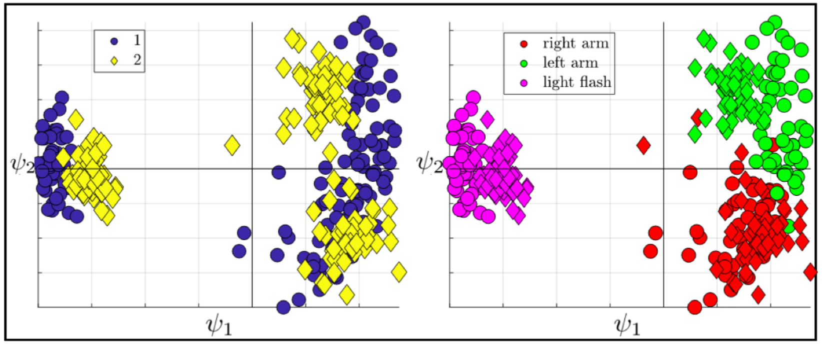
Baseline



Parallel Transport



Mean Subtraction



Parallel Transport & Moments Alignment

Extensions and Outlook

- Take home message
  - High-dimensional data live in a non-Euclidean space
  - Covariance matrices are informative features
  - They live in a non-Euclidean space with operations given in closed-form
- Covariance matrices might be insufficient features
- Instead, we could use:
  - Correlation and Partial Correlation matrices
  - Positive Kernels
  - Graph Laplacians
  - Transition probability matrices of random walks on graphs

Thank you

O. Yair, M. Ben-Chen and R. Talmon
"Parallel Transport on the Cone Manifold of SPD Matrices for Domain Adaptation"
IEEE Transactions on Signal Processing, 2019

We acknowledge funding from the European Union's Horizon 2020 research grant agreement 802735