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:
  \[
  \left\{ X_i^{(1)}, y_i^{(1)} \right\}_{i=1}^{N_1}, \left\{ X_i^{(2)}, y_i^{(2)} \right\}_{i=1}^{N_2}
  \]
  (target) (source)

- Each set contains \( N_k \) matrices of observations
  \[
  X_i^{(k)} \in \mathbb{R}^{d \times T_i^{(k)}}
  \]

- \( d \) - dimension (\# of EEG electrodes)

- \( T_i^{(k)} \) - observation length

- \( y_i^{(k)} \) - hidden label (imagined motor task)

- Let \( P_i^{(k)} \in \mathbb{R}^{d \times d} \) be the (sample) covariance of \( X_i^{(k)} \)
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 et 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 \( T_x \mathcal{M} \) be the tangent plane at the point \( x \)

• Define the following operations:

<table>
<thead>
<tr>
<th></th>
<th>Vector Space</th>
<th>Riemannian Manifold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtraction</td>
<td>( \overrightarrow{xy} = y - x )</td>
<td>( \overrightarrow{xy} = \text{Log}_x (y) )</td>
</tr>
<tr>
<td>Addition</td>
<td>( y = x + \overrightarrow{xy} )</td>
<td>( y = \text{Exp}_x (\overrightarrow{xy}) )</td>
</tr>
<tr>
<td>Mean</td>
<td>( \arg \min_x \sum_i | x - x_i |_i^2 )</td>
<td>( \arg \min_x \sum_i d_R^2 (x, x_i) )</td>
</tr>
</tbody>
</table>

\[ \arg \min_P d_R^2 (P, X) + d_R^2 (P, Y) \]
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

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

- $\mathcal{T}_P \mathcal{M}$ 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.

$s_i = \text{vec}(S_i)$

With $\sqrt{2}$ scaling on the off-diagonal elements.
The SPD Cone – Properties

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

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

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

  $$d_R^2(P_1, P_2) = \left\| \log \left( P_2^{-\frac{1}{2}} P_1 P_2^{-\frac{1}{2}} \right) \right\|_F^2$$

  $$= \sum_{i=1}^{n} \log^2 \left( \lambda_i \left( P_2^{-\frac{1}{2}} P_1 P_2^{-\frac{1}{2}} \right) \right)$$

- $\lambda_i(P)$ is the i-th eigenvalue of $P$
- Scale-invariant
The SPD Cone – Properties

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

\[ \Rightarrow \begin{cases} \text{Tr}(M) = x + z > 0 \\ |(M)| = xz - y^2 > 0 \end{cases} \Rightarrow \begin{cases} x > 0 \\ z > 0 \end{cases} \Rightarrow |y| < \sqrt{xz} \]
The SPD Cone – Properties

- **Logarithm map:**
  \[ S_1 = \text{Log}_P (P_1) = P^{\frac{1}{2}} \log \left( P^{-\frac{1}{2}} P_1 P^{-\frac{1}{2}} \right) P^{\frac{1}{2}} \in T_P \mathcal{M} \]

- **Exponential map:**
  \[ P_1 = \text{Exp}_P (S_1) = P^{\frac{1}{2}} \exp \left( P^{-\frac{1}{2}} S_1 P^{-\frac{1}{2}} \right) P^{\frac{1}{2}} \in \mathcal{M} \]

- **Relation to the (unique) geodesic curve** \( \varphi(t) \) **from** \( P_1 \) **to** \( P_2 \)** is given by the initial velocity
  \[ \varphi'(0) = \text{Log}_{P_1} (P_2) \in T_{P_1} \mathcal{M} \]
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 $\overline{P}^{(1)}$.
- $S^{(k)}_i$ - the symmetric matrix (or vector) in $\mathcal{T}_{\overline{P}^{(k)}}\mathcal{M}$.

\[
S^{(k)}_i = \text{Log}_{\overline{P}^{(k)}}(P^{(k)}_i)
\]
Domain Adaptation

Formulation

• **Goal:**
  Derive a new representation $\Gamma(S_i^{(2)})$:
  $$\Gamma: \mathcal{T}_P^{(2)} \mathcal{M} \rightarrow \mathcal{T}_P^{(1)} \mathcal{M}$$
  so that $\{S_i^{(1)}\}$ and $\{\Gamma(S_i^{(2)})\}$ live in the same linear space

• **Benefit:**
  Relate samples from the two subsets
  • Compute quantities such as $\langle S^{(1)}_i, \Gamma(S^{(2)}_j) \rangle_{P^{(1)}}$
Domain Adaptation

Constraints:

• Zero mean:

\[
\frac{1}{N_2} \sum_{i=1}^{N_2} \Gamma(S_{i}^{(2)}) = \frac{1}{N_1} \sum_{i=1}^{N_1} S_{i}^{(1)} = 0
\]

• Inner product preservation:

\[
\langle \Gamma(S_{i}^{(2)}), \Gamma(S_{j}^{(2)}) \rangle_{P^{(1)}} = \langle S_{i}^{(2)}, S_{j}^{(2)} \rangle_{P^{(2)}}
\]

• Geodesic velocity preservation:

\[
\Gamma(\varphi'(0)) = \varphi'(1)
\]
Domain Adaptation

Formulation

• **Constraints:**
  - Zero mean:
    \[
    \frac{1}{N_2} \sum_{i=1}^{N_2} \Gamma(S^{(2)}_i) = \frac{1}{N_1} \sum_{i=1}^{N_1} S^{(1)}_i = 0
    \]
  - Inner product preservation:
    \[
    \langle \Gamma(S^{(2)}_i), \Gamma(S^{(2)}_j) \rangle_{P^{(1)}} = \langle S^{(2)}_i, S^{(2)}_j \rangle_{P^{(2)}}
    \]

Imply that the map preserves inter-sample relations
Domain Adaptation

Formulation

• **Constraints:**
  • Zero mean:
    \[
    \frac{1}{N_2} \sum_{i=1}^{N_2} \Gamma(S_i^{(2)}) = \frac{1}{N_1} \sum_{i=1}^{N_1} S_i^{(1)} = 0
    \]
  • Inner product preservation:
    \[
    \langle \Gamma(S_i^{(2)}), \Gamma(S_j^{(2)}) \rangle_{P^{(1)}} = \langle S_i^{(2)}, S_j^{(2)} \rangle_{P^{(2)}}
    \]

Not unique:
If \( \Gamma \) admits to these properties, then also \( R \circ \Gamma \)
where \( R \) is an arbitrary rotation.
Domain Adaptation

Formulation

• Constraints:
  • Zero mean:
    \[
    \frac{1}{N_2} \sum_{i=1}^{N_2} \Gamma(S_i^{(2)}) = \frac{1}{N_1} \sum_{i=1}^{N_1} S_i^{(1)} = 0
    \]
  • Inner product preservation:
    \[
    \langle \Gamma(S_i^{(2)}), \Gamma(S_j^{(2)}) \rangle_{\mathcal{P}^{(1)}} = \langle S_i^{(2)}, S_j^{(2)} \rangle_{\mathcal{P}^{(2)}}
    \]

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) \( \varphi' \left( 0 \right) \in T_{p\left( 2 \right)} M \) and \( \varphi' \left( 1 \right) \in T_{p\left( 1 \right)} M \) are used to fix the rotation.
Domain Adaptation

Formulation

• **Constraints:**
  • Geodesic velocity preservation:
    \[ \Gamma (\varphi' (0)) = \varphi' (1) \]

Unsupervised – no labels are used

Present a closed-form expression (no optimization)
**Lemma** (Parallel Transport)

Let $A, B \in \mathcal{M}$.
The PT from $B$ to $A$ of any $S \in \mathcal{T}_B \mathcal{M}$ is:

$$\Gamma_{B \rightarrow A} (S) \triangleq E S E^T$$

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

**Theorem.**
The representation $\Gamma_{\overline{P}^{(2)} \rightarrow \overline{P}^{(1)}} (S_i^{(2)})$, i.e., the unique PT of $S_i^{(2)}$ from $\overline{P}^{(2)}$ to $\overline{P}^{(1)}$, is well defined and satisfies properties $(1) - (3)$. 
Parallel Transport
Algorithm:

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

$$S_i^{(2)} = \text{Log}_{\mathbf{P}^{(2)}}(\mathbf{P}_i^{(2)})$$

2. Parallel transport $S_i^{(2)}$ from $\mathbf{P}^{(2)}$ to $\mathbf{P}^{(1)}$ by computing

$$S_i^{(2)\rightarrow(1)} = \Gamma_{\mathbf{P}^{(2)}\rightarrow\mathbf{P}^{(1)}}\left(S_i^{(2)}\right)$$

3. Project the symmetric matrix $S_i^{(2)\rightarrow(1)} \in \mathcal{T}_{\mathbf{P}^{(1)}}\mathcal{M}$ back to the manifold using $\text{Exp}_{\mathbf{P}^{(1)}}\left(S_i^{(2)\rightarrow(1)}\right)$. 
Domain Adaptation with PT

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

$$\Psi(P_i^{(2)}) = \text{Exp}_{\mathcal{P}^{(1)}} \left( \Gamma_{\mathcal{P}^{(2)}} \rightarrow \mathcal{P}^{(1)} \left( \text{Log}_{\mathcal{P}^{(2)}} (P_i^{(2)}) \right) \right)$$

for any $P_i^{(2)} \in \mathcal{P}^{(2)}$
Theorem.
Let $A, B, P \in \mathcal{M}$ and let $S = \text{Log}_B(P) \in \mathcal{T}_B \mathcal{M}$. Then,

$$\text{Exp}_A \left( \Gamma_{B \rightarrow A} (S) \right) = EPET,$$

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

• $\Psi$ can be efficiently implemented

$$\Psi \left( P_i^{(2)} \right) = \Gamma_{P^{(2)} \rightarrow P^{(1)}} \left( P_i^{(2)} \right) = EP_i^{(2)} ET$$

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

- **Important consequence:**
  - The covariance adaptation:
    \[ EP_i^{(2)} E^T \]
    where
    \[ E \triangleq \left( \overline{P}^{(1)} \left( \overline{P}^{(2)} \right)^{-1} \right)^{\frac{1}{2}} \]
  - Can be applied directly to data by:
    \[ EX_i^{(2)} \]
Toy Problem

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

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

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 \( \phi_i \) (the initial phase of the oscillations)
Toy Problem

• The population covariance of $s_i[n]$ is

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

which depends only on $\phi_i$

• Note: 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

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

such that:

\[ x^{(k)}_i[n] = M^{(k)} s_i[n] \]

where

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

* 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

One subject
Two sessions

After PT

tSNE [v. d. Maanen & Hinton, 08] representation of Covariance matrices
Brain Computer Interface

Two subjects

After PT

Mean subtraction [Barachant et al, 13]
Intrinsicness

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

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

- Equivalent to parallel transport to the identity:

\[
\Psi \left( P^{(k)}_i \right) = \Gamma_{\overline{P}^{(k)} \to I} \left( P^{(k)}_i \right) = EP^{(k)}_i E^T
\]

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

- 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 \((A_1, B_1)\) and \((A_2, B_2)\), such that \(A_1, B_1, A_2, B_2 \in \mathcal{M}\), are *equivalent* if there exists an invertible matrix \(E\) such that

\[
A_2 = \Gamma(A_1) = EA_1E^T \\
B_2 = \Gamma(B_1) = EB_1E^T
\]

We denote this relation by

\[(A_1, B_1) \sim (A_2, B_2)\]
**Definition**  (Equivalent Pairs)
Two pairs $(A_1, B_1)$ and $(A_2, B_2)$, such that $A_1, B_1, A_2, B_2 \in \mathcal{M}$, are *equivalent* if there exists an invertible matrix $E$ such that

\[
A_2 = \Gamma(A_1) = EA_1E^T \\
B_2 = \Gamma(B_1) = EB_1E^T
\]

We denote this relation by

\[
(A_1, B_1) \sim (A_2, B_2)
\]

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

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

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

\[\Gamma \circ \Gamma_{B_1 \rightarrow A_1} = \Gamma_{B_2 \rightarrow A_2} \circ \Gamma,\]

where \(\Gamma (P) = EPET\) and \(E\) satisfies the equivalence relation.
Intrinsicness

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

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

Input: \( \{ P_i^{(1)} \}_{i=1}^{N_1}, \{ P_i^{(2)} \}_{i=1}^{N_2}, \ldots, \{ P_i^{(K)} \}_{i=1}^{N_K} \)

Output: \( \{ \tilde{S}_i^{(1)} \}_{i=1}^{N_1}, \{ \tilde{S}_i^{(2)} \}_{i=1}^{N_2}, \ldots, \{ \tilde{S}_i^{(K)} \}_{i=1}^{N_K} \)

1. **For** each \( k \in \{1, 2, \ldots, K\} \), compute \( \bar{P}^{(k)} \) the Riemannian mean of the subset \( \{ P_i^{(k)} \} \).

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

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

\[
\Gamma_i^{(k)} = \Gamma_{\bar{P}^{(k)} \rightarrow \hat{P}} \left( P_i^{(k)} \right).
\]

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

\[
\tilde{S}_i^{(k)} = \log \left( \hat{P}^{-\frac{1}{2}} \Gamma_i^{(k)} \hat{P}^{-\frac{1}{2}} \right).
\]
Brain Computer Interface

Five subjects

After PT
Brain Computer Interface

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

![Bar chart showing classification accuracy for different test subjects. The x-axis represents test subjects (1, 3, 7, 8, 9), and the y-axis represents classification accuracy. The chart compares Baseline, Mean Transport, and Algorithm 1.]
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

<table>
<thead>
<tr>
<th>Predicted Class</th>
<th>True Class</th>
<th>Baseline</th>
<th>Mean Subtraction</th>
<th>Parallel Transport</th>
</tr>
</thead>
<tbody>
<tr>
<td>REM</td>
<td>Stage 3</td>
<td>34</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>Stage 3</td>
<td></td>
<td>23</td>
<td>41</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>REM</td>
<td>1</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Stage 3</td>
<td>26</td>
<td>41</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>True Class</td>
<td>97.1%</td>
<td>77.1%</td>
<td>88.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53.1%</td>
<td>83.7%</td>
<td>98.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>71.4%</td>
<td>81.0%</td>
<td>85.7%</td>
</tr>
</tbody>
</table>

- **Baseline**
  - REM Stage 3: 34, 23, 59.6%
  - Stage 3: 26, 96.3%
- **Mean Subtraction**
  - REM Stage 3: 27, 8, 77.1%
  - Stage 3: 8, 41, 83.7%
- **Parallel Transport**
  - REM Stage 3: 24, 1, 96.0%
  - Stage 3: 11, 48, 81.4%
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

Baseline

Parallel Transport
Mental Arithmetic Identification

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

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Mean Subtraction</th>
<th>Parallel Transport</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>74%</td>
<td>73%</td>
<td>78%</td>
</tr>
</tbody>
</table>
Extensions and Outlook

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

![Baseline](image1)

![Parallel Transport](image2)

![Mean Subtraction](image3)

![Parallel Transport & Moments Alignment](image4)
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