DOMAIN ADAPTATION USING RIEMANNIAN GEOMETRY OF SPD MATRICES

SIPL Signal and Image Processing Lab

Gal Maman[1], Or Yair[1], Danny Eytan[2], Ronen Talmon[1]

[1]Viterbi Faculty of Electrical Engineering [2]Faculty of Medicine
Technion, Israel Institute of Technology
Haifa, 3200003, Israel

TECHNION Israel Institute of Technology

Goal

- **Goal:** Feature extraction and classification of a dataset with multiple domains.
- **Objective:** Obtain a new representation using an unsupervised domain adaptation method based on parallel transport (PT) and moments alignment.
- **Applicability:** Domain adaptation and classification of high-dimensional noisy electrophysiological (EEG) signals collected from different subjects, where each subject is considered as a different domain.

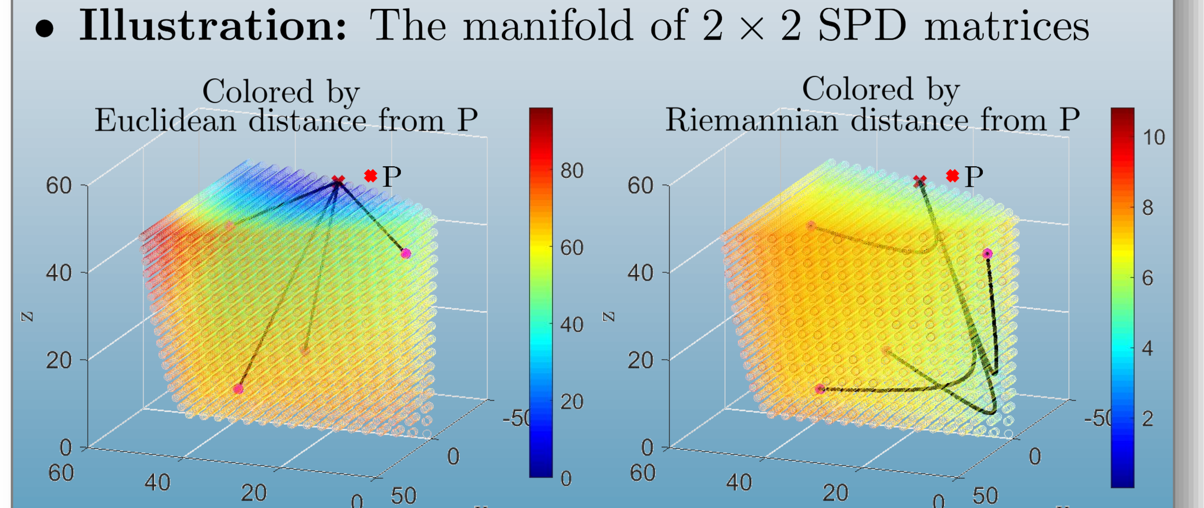
Problem Formulation

- Let $\{X_i^{(1)}, y_i^{(1)}\}_{i=1}^{N_1}$ be the source set and let $\{X_i^{(2)}, y_i^{(2)}\}_{i=1}^{N_2}$ be the target set.

  Each set $k = 1, 2$ contains N_k matrices of data observations $X_i^{(k)} \in \mathbb{R}^{d \times T_i^{(k)}}$, such that d is the observation dimension, $T_i^{(k)}$ is the observation length and $y_i^{(k)}$ is the corresponding label.
- Let $P_i^{(k)} \in \mathbb{R}^{d \times d}$ denote the sample covariance of the data matrix $X_i^{(k)}$, given by $P_i^{(k)} \triangleq \frac{1}{T_i^{(k)}} X_i^{(k)} \left(X_i^{(k)}\right)^T$, where for simplicity we assume zero mean.
- We aim to find a new representation for $\{P_i^{(1)}\}$ and $\{P_i^{(2)}\}$, that will allow to train a classifier on $\{P_i^{(1)}\}$ and to apply it to $\{P_i^{(2)}\}$ (domain adaptation).

SPD Matrices

- A symmetric matrix $P \in \mathbb{R}^{d \times d}$ is positive-definite (SPD) if all its eigenvalues are strictly positive, or equivalently, if $v^T P v > 0$ for every $v \neq 0$.
- The set of all SPD matrices is an open convex cone, constituting a differential Riemannian manifold $\mathcal{M}$.
- **Illustration:** The manifold of 2×2 SPD matrices



Each point (x, y, z) represents a matrix $M = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$.
$M > 0$ iff $x > 0, z > 0$ and $y^2 < xz$.

The 4 magenta points represent matrices with equal Riemannian distance from P. The curves are the Euclidean and Riemannian geodesic respectively.

This project has received funding from the European Union's Horizon 2020 research grant agreement DIFFOP 802735

Or Yair is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities

The Riemannian Cone Manifold of SPD Matrices

- Let $\mathcal{T}_P\mathcal{M}$ be the tangent space at the point $P \in \mathcal{M}$, equipped with the following inner product:

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

  where $S_1, S_2 \in \mathcal{T}_P\mathcal{M}$ are two symmetric matrices.
- Let $P_1, P_2 \in \mathcal{M}$; the Riemannian geodesic distance is defined by:

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

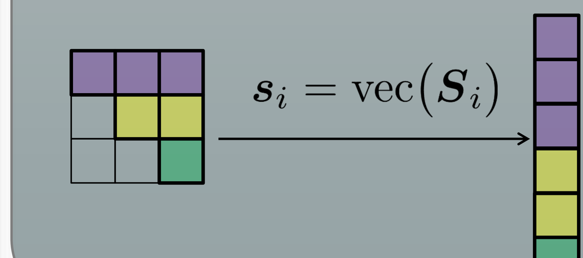
- The Riemannian mean of a set $\{P_i \in \mathcal{M}\}_{i=1}^N$ is defined using the Fréchet mean:

$$\overline{P} \triangleq \arg\min_{P \in \mathcal{M}} \sum_{i=1}^N \delta_R^2(P, P_i)$$

- The logarithmic mapping from $P_i \in \mathcal{M}$ to $S_i \in \mathcal{T}_P\mathcal{M}$ is defined by:

$$S_i = \mathrm{Log}_P(P_i) \triangleq P^{\frac{1}{2}} \log\left(P^{-\frac{1}{2}} P_i P^{-\frac{1}{2}}\right) P^{\frac{1}{2}}$$

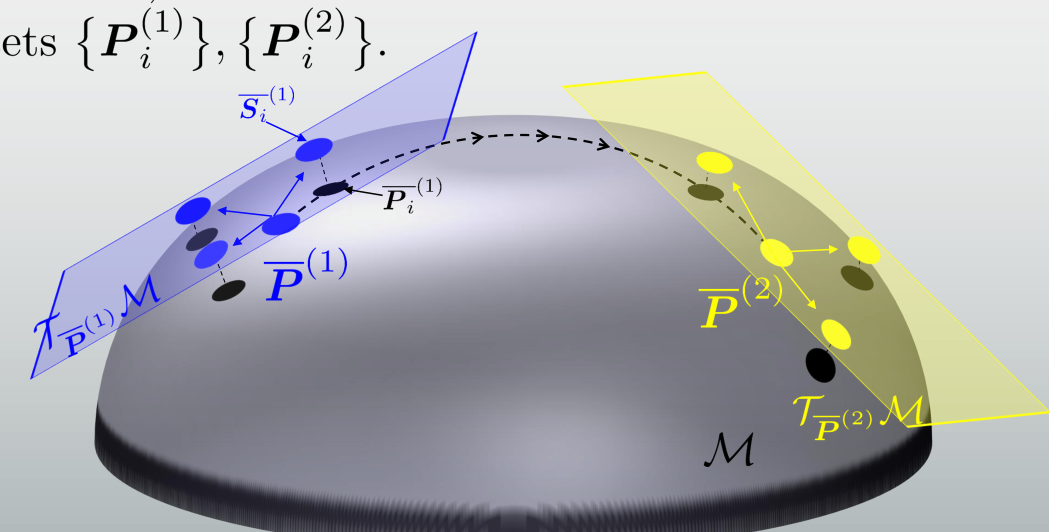
- We denote the vector representation $s_i = \mathrm{vec}(S_i)$



Where $\mathrm{vec}(\cdot)$ is the vectorized upper triangular of a symmetric matrix, with $\sqrt{2}$ weights for its off-diagonal elements.

Parallel Transport (PT)

- Let $\overline{P}^{(1)}, \overline{P}^{(2)}$ be the Riemannian means of the two sets $\{P_i^{(1)}\}, \{P_i^{(2)}\}$.



- The PT from $\overline{P}^{(1)}$ to $\overline{P}^{(2)}$ of any SPD matrix in $\{P_i^{(1)}\}$ is given by:

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

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

- 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.
- X. Pennec, P. Fillard, and N. Ayache, A riemannian framework for tensor computing, International Journal of computer vision, 2006.
- A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, Multiclass brain–computer interface classification by Riemannian geometry, IEEE Transactions on Biomedical Engineering, 2012.

PT and Moments Alignment

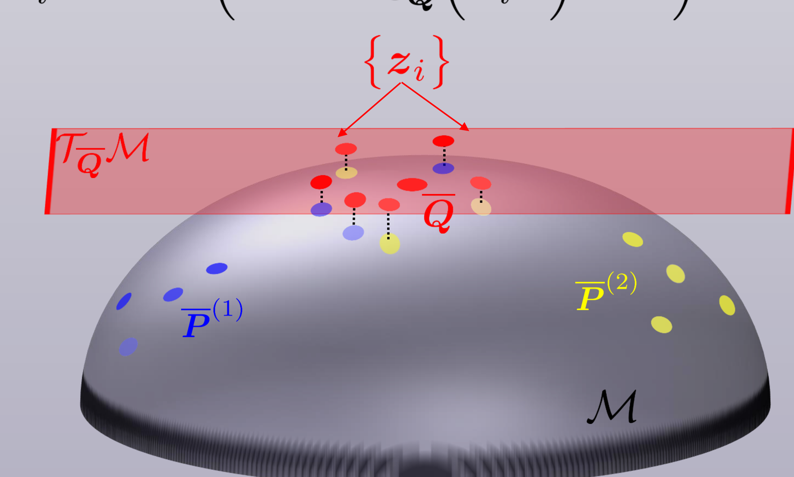
Algorithm 1

Input: K sets $\left\{P_i^{(k)} \in \mathbb{R}^{d \times d}\right\}_{i=1}^{N_k}$, $k = 1, \ldots, K$.

Output: K matrices $\left\{Y^{(k)} \in \mathbb{R}^{D \times N_k}\right\}_{k=1}^{K}$, $D = d(d+1)/2$

Where column $y_i^{(k)}$ is the aligned vector representation of the matrix $P_i^{(k)}$.

1. For each set $\left\{P_i^{(k)}\right\}_{i=1}^{N_k}$, compute the Riemannian mean $\overline{P}^{(k)}$.
2. Compute $\overline{Q}$, the Riemannian mean of $\left\{\overline{P}^{(k)}\right\}_{k=1}^{K}$.
3. For $k = 1 \ldots, K$:
   (a) For $i = 1, \ldots, N_k$:
      i. Apply PT: $\Gamma_i^{(k)} = \Gamma_{\overline{P}^{(k)} \to \overline{Q}}\left(P_i^{(k)}\right)$
      ii. Project $\Gamma_i^{(k)}$ to a Euclidean space:
      $$z_i^{(k)} = \mathrm{vec}\left(\overline{Q}^{-\frac{1}{2}} \mathrm{Log}_{\overline{Q}}\left(\Gamma_i^{(k)}\right) \overline{Q}^{-\frac{1}{2}}\right)$$



   (b) Collect $Z^{(k)} = \left[z_1^{(k)}, z_2^{(k)}, \ldots, z_{N_k}^{(k)}\right]$
   (c) Apply SVD $Z^{(k)} = U^{(k)} \Lambda^{(k)} V^{(k)}$.
   (d) If $k \neq 1$, for each $j \in \{1, 2, \ldots, D\}$ update the columns of $U^{(k)}$: $u_j^{(k)} \leftarrow \mathrm{sign}\left(\left\langle u_j^{(1)}, u_j^{(k)} \right\rangle\right) u_j^{(k)}$
   (e) Compute $Y^{(k)} = \left(U^{(k)}\right)^T Z^{(k)}$.

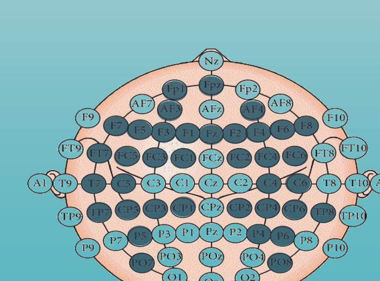
Experimental Results

EEG Data

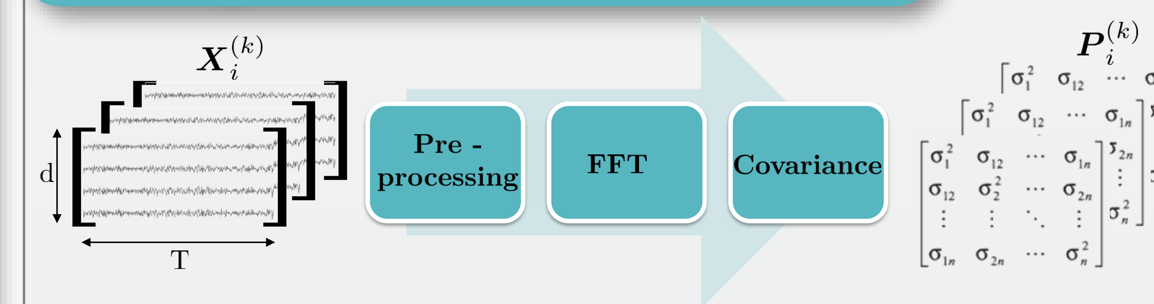
Raw data:
- EEG recordings from 11 subjects of ages ranging from 7 to 16 years.
- 64 EEG electrodes.
- 3 types of stimuli: visual, auditory and somatosensory.

Pre-processing:
- Organizing the recordings into trials of 1 second, and downsampling to 1KHz.
- Excluding of malfunctioning electrodes and highly noisy trials.
- Applying FFT and taking the absolute value.

After pre-processing, the data contain recordings from 37 electrodes, and include 80-500 repeated trials per stimulus for each subject.



Experimental Results



2 Subjects

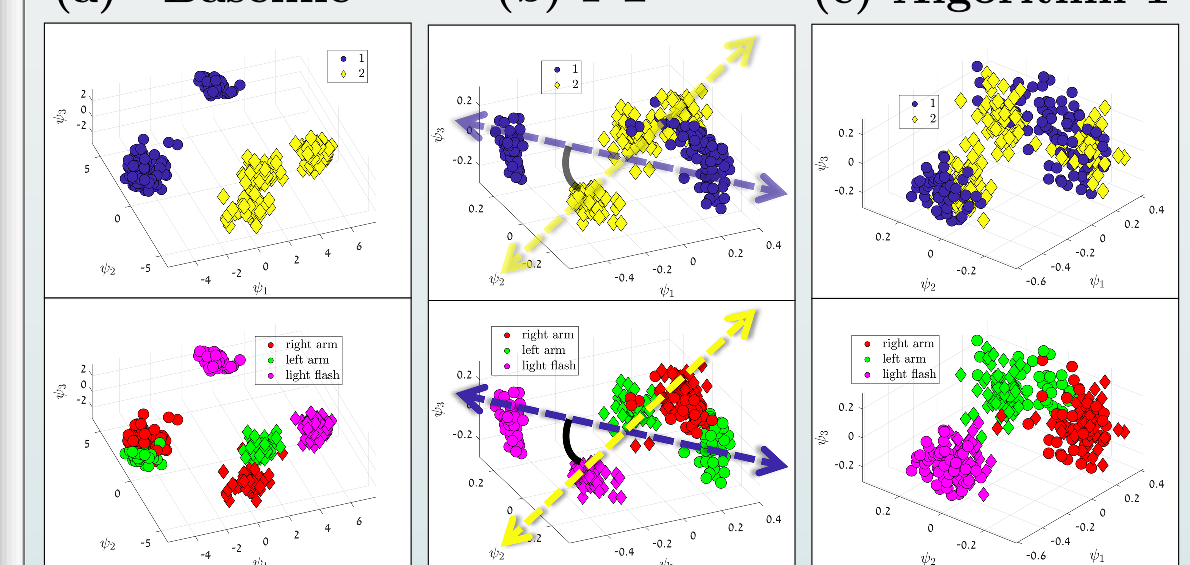


Fig.1. The first three principal components of the representation of the trials obtain by the different algorithms: (a) "Baseline", (b) "PT" and (c) Algorithm 1.

Multiple Subjects

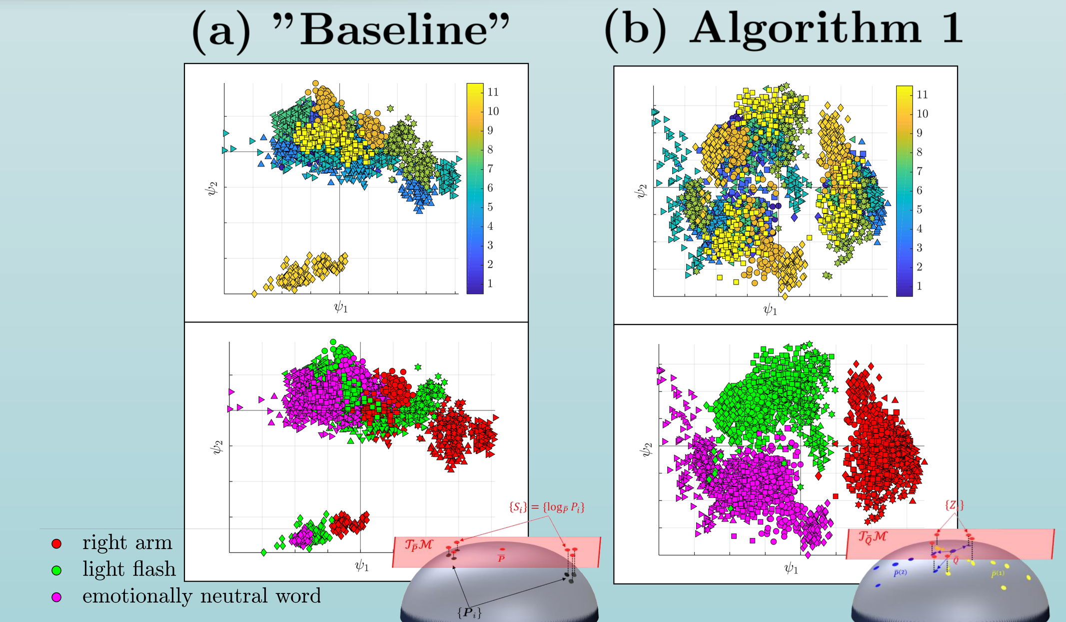
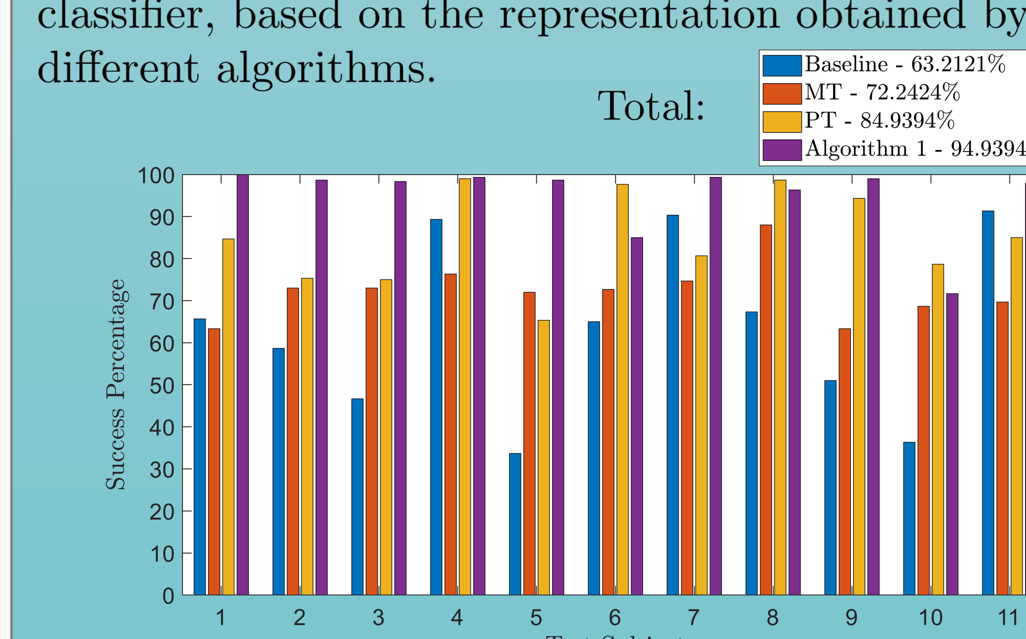


Fig.2. The first two principal components of the representation of the trials obtain by: (a) "Baseline" and (b) Algorithm 1.

Classification Results

Fig.3. Classification results of the responses to three different stimuli from 11 subjects, using a cubic SVM classifier, based on the representation obtained by the different algorithms.



We remark that the sign of the second PC of subjects 5 and 6 was manually flipped (note that their respective angles were 89° and 95°).