OPTIMAL TRANSPORT APPLIED TO BCI

Nathalie GAYRAUD   Maureen CLERC   Alain RAKOTOMAMONJY

PhD Seminars of Inria

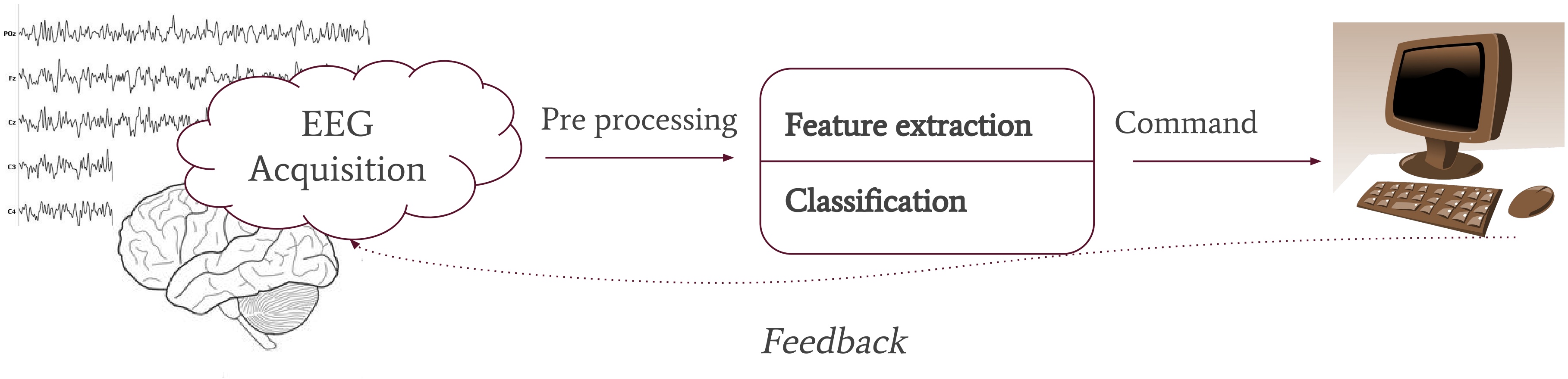
OPTIMAL TRANSPORT APPLIED TO BCI

OUTLINE

1. BRAIN COMPUTER INTERFACES (THE P300 SPELLER)
2. OPTIMAL TRANSPORT
3. APPLYING OT TO BCI
4. RESULTS
5. DISCUSSION

BRAIN COMPUTER INTERFACES

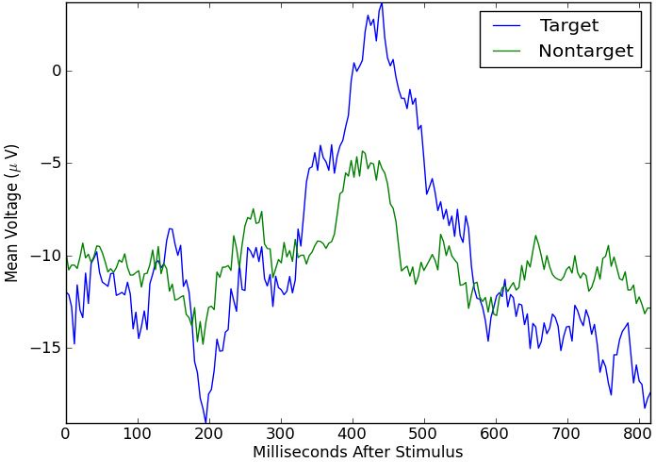
Brain Computer Interfaces



The P300 Speller

P300 Speller session

- User looks at a keyboard on the screen.
  - Letters are flashing.
- User counts the number of times “his” letter flashed.
- Each time “his” letter flashes, it elicits a response.
- Amplitudes are typically highest over parietal brain areas



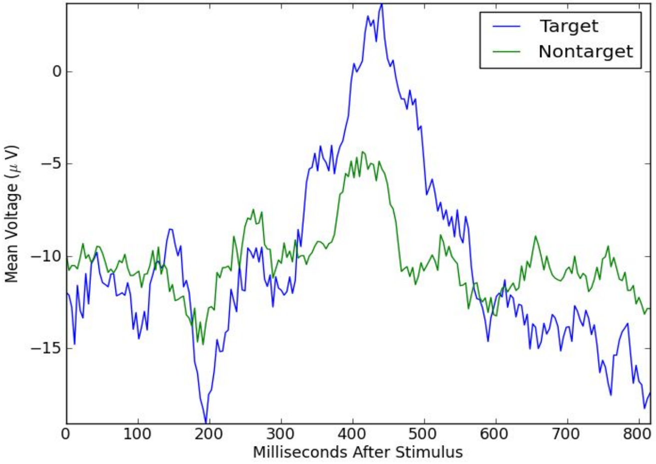
The P300 Speller

Feature extraction

- ➢ Time windowing
- ➢ Signal pre-processing
  - Frequency filtering
  - Spatial Filtering / Component analysis (Xdawn)[1]
  - Downsampling

Classification

- ➢ Feature vectors $\{\mathbf{x}_i\}_{i=1}^{N} = \mathbf{X}$ are classified into two classes: *Target / Nontarget*



[1] Rivet, B., Souloumiac, A., Attina, V., & Gibert, G. xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. IEEE Transactions on Biomedical Engineering, 2009, vol. 56, no 8, p. 2035-2043.

Variability

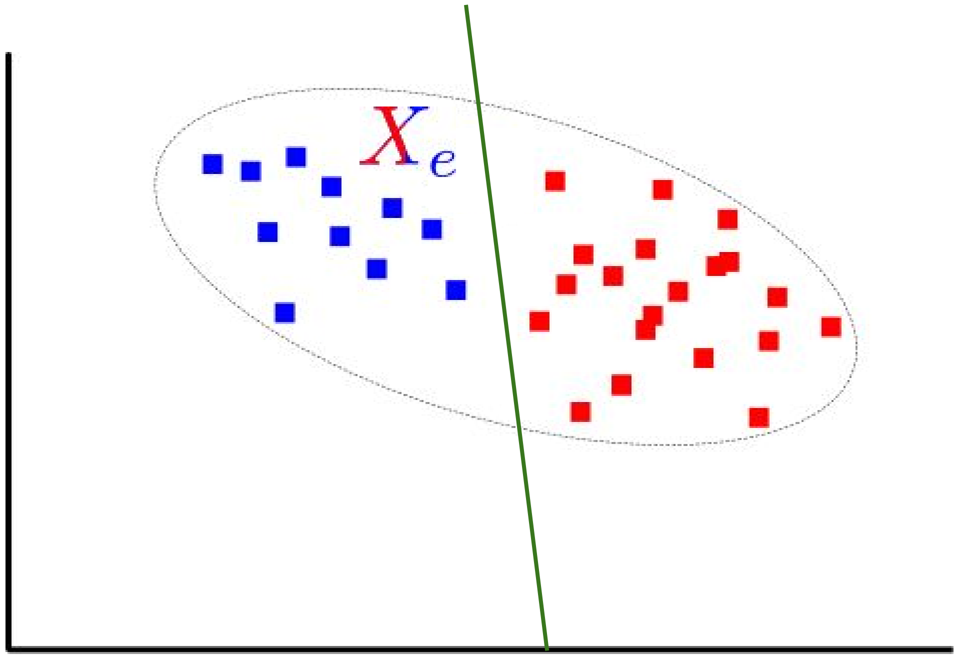
Types

- Same-session
- Cross-session
- Cross-patient

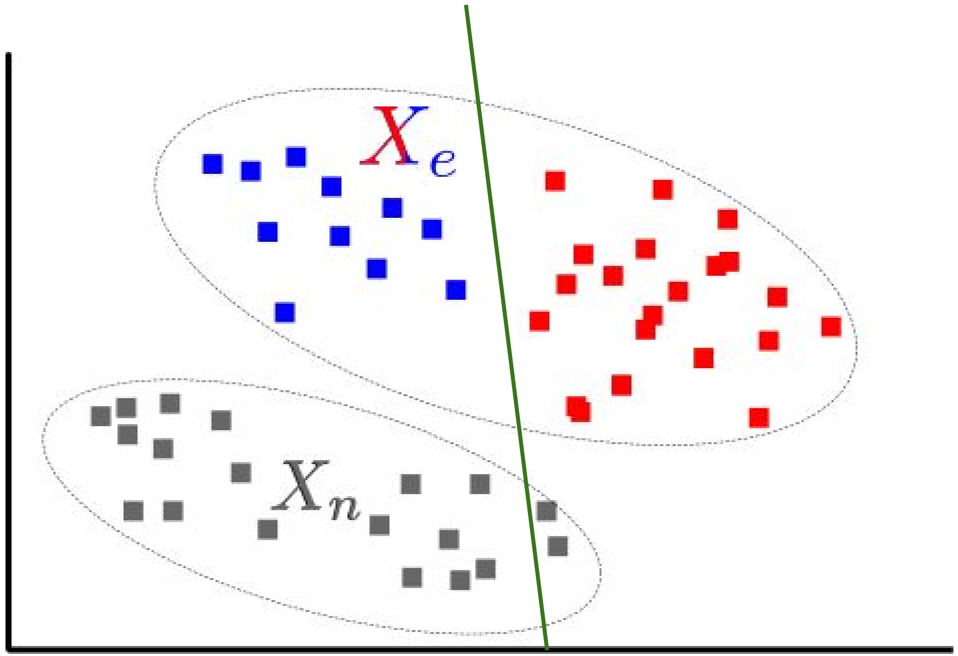
Sources

- Equipment
- Electrode position
- Mental state
- Physiological differences
- Environmental

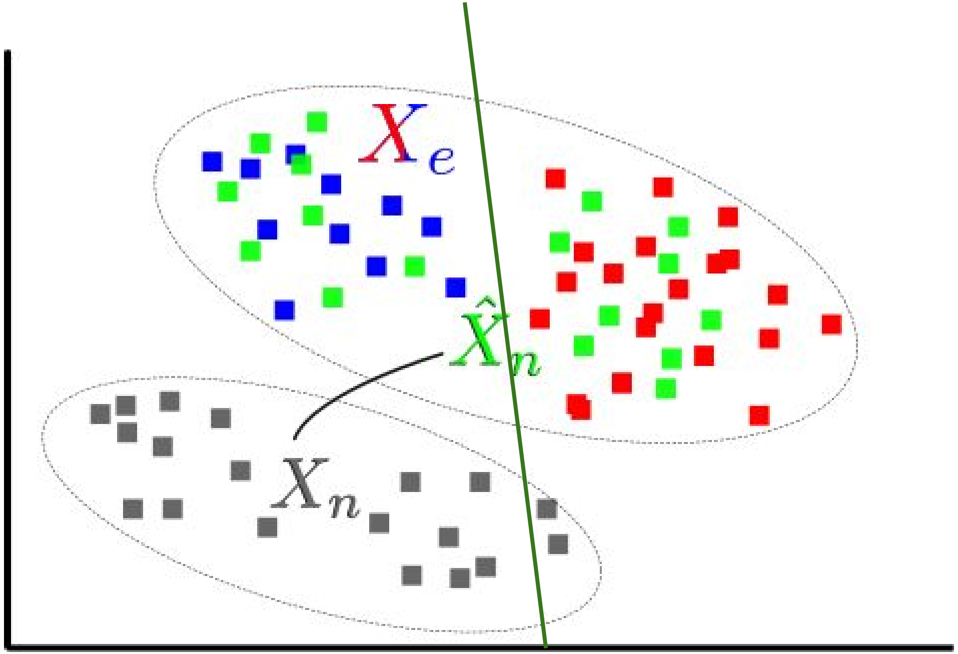
Towards a zero-calibration BCI



Towards a zero-calibration BCI



Towards a zero-calibration BCI



REGULARISED DISCRETE OPTIMAL TRANSPORT WITH CLASS LABELS

Optimal Transport

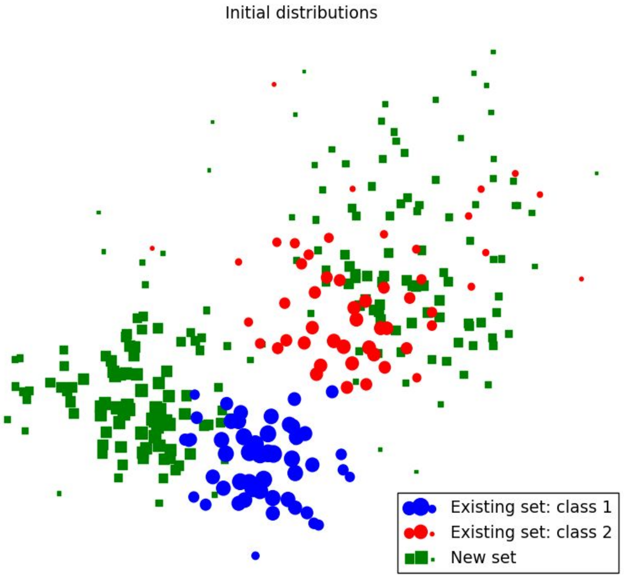
Find a transportation that minimises a cost function.

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\operatorname{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}}$$

$$\mathcal{B} = \left\{ \gamma \in (\mathbb{R}^+)^{N_e \times N_n} \mid \gamma \mathbf{1}_{N_n} = \mu_e \, , \gamma^{\mathbf{T}} \mathbf{1}_{N_e} = \mu_n \right\}$$

$$C = d(x_i^e, x_j^n) = \|x_i^e - x_j^n\|_2^2$$

"What is the optimal way to transport mass from domain A to domain B?"



$$\hat{\mathbf{X}}^n = \operatorname{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

Optimal Transport

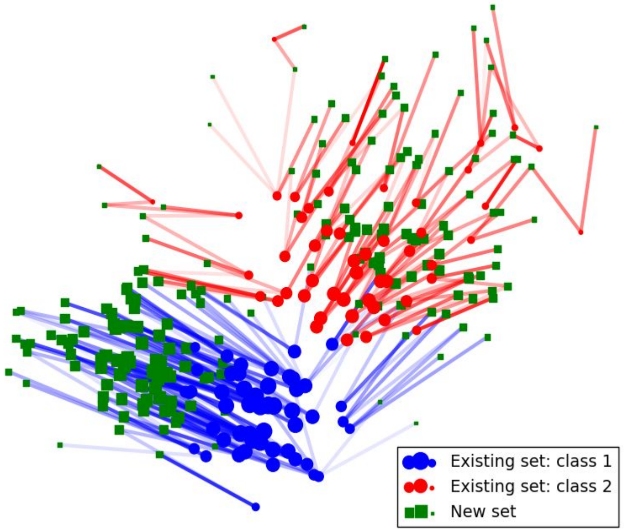
Find a transportation that minimises a cost function.

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\mathrm{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}}$$

$$\mathcal{B} = \left\{ \gamma \in (\mathbb{R}^+)^{N_e \times N_n} \mid \gamma \mathbf{1}_{N_n} = \mu_e \, , \gamma^{\mathbf{T}} \mathbf{1}_{N_e} = \mu_n \right\}$$

$$C = d(x_i^e, x_j^n) = \|x_i^e - x_j^n\|_2^2$$

"What is the optimal way to transport mass from domain A to domain B?"



$$\hat{\mathbf{X}}^n = \mathrm{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

Optimal Transport

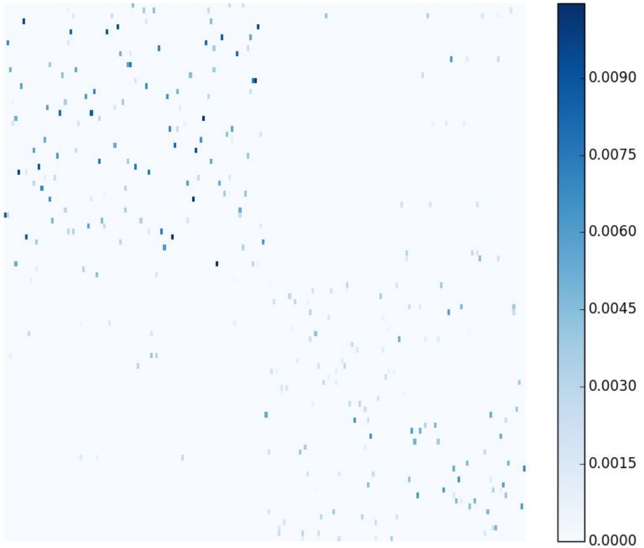
Find a transportation that minimises a cost function.

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\mathrm{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}}$$

$$\mathcal{B} = \left\{ \gamma \in (\mathbb{R}^+)^{N_e \times N_n} \mid \gamma \mathbf{1}_{N_n} = \mu_e \,, \gamma^{\mathbf{T}} \mathbf{1}_{N_e} = \mu_n \right\}$$

$$C = d(x_i^e, x_j^n) = \|x_i^e - x_j^n\|_2^2$$

“What is the optimal way to transport mass from domain A to domain B?”



$$\hat{\mathbf{X}}^n = \mathrm{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

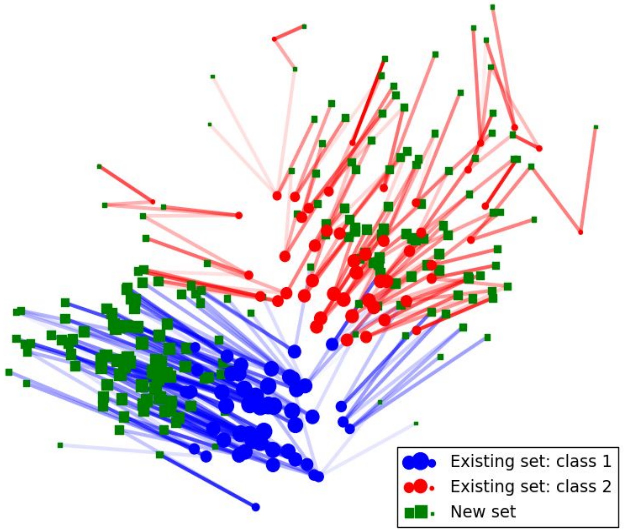
Regularised Optimal Transport

Entropic regularisation

Allows for a faster implementation using Sinkhorn algorithm [2]

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\mathrm{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}} + \lambda \mathbf{R}_s(\gamma)$$

$$\mathbf{R}_s(\gamma) = \lambda \sum_{i,j} \gamma(i,j) \log \gamma(i,j)$$



$$\hat{\mathbf{X}}^n = \mathrm{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

“What is the optimal way to transport mass from domain A to domain B?”

[2] Cuturi, Marco. "Sinkhorn distances: Lightspeed computation of optimal transport." Advances in Neural Information Processing Systems. 2013.

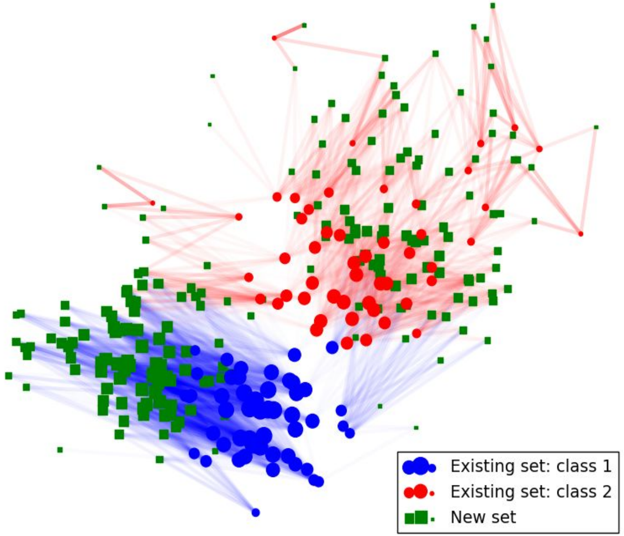
Regularised Optimal Transport

Entropic regularisation

Allows for a faster implementation using Sinkhorn algorithm [2]

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\operatorname{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}} + \lambda \mathbf{R}_s(\gamma)$$

$$\mathbf{R}_s(\gamma) = \lambda \sum_{i,j} \gamma(i,j) \log \gamma(i,j)$$



$$\hat{\mathbf{X}}^n = \operatorname{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

“What is the optimal way to transport mass from domain A to domain B?”

[2] Cuturi, Marco. "Sinkhorn distances: Lightspeed computation of optimal transport." Advances in Neural Information Processing Systems. 2013.

Regularised Optimal Transport

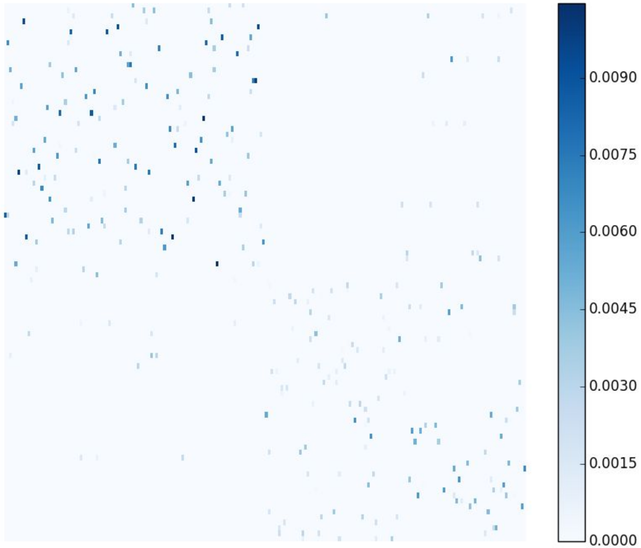
Entropic regularisation

Allows for a faster implementation using Sinkhorn algorithm [2]

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\operatorname{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}} + \lambda \mathbf{R}_s(\gamma)$$

$$\mathbf{R}_s(\gamma) = \lambda \sum_{i,j} \gamma(i,j) \log \gamma(i,j)$$

“What is the optimal way to transport mass from domain A to domain B?”



$$\hat{\mathbf{X}}^n = \operatorname{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

Regularised Optimal Transport

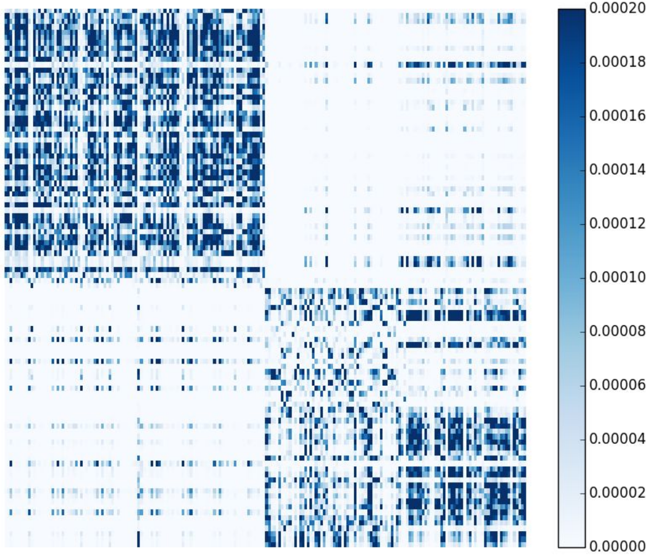
Entropic regularisation

Allows for a faster implementation using Sinkhorn algorithm [2]

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\mathrm{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}} + \lambda \mathbf{R}_s(\gamma)$$

$$\mathbf{R}_s(\gamma) = \lambda \sum_{i,j} \gamma(i,j) \log \gamma(i,j)$$

"What is the optimal way to transport mass from domain A to domain B?"



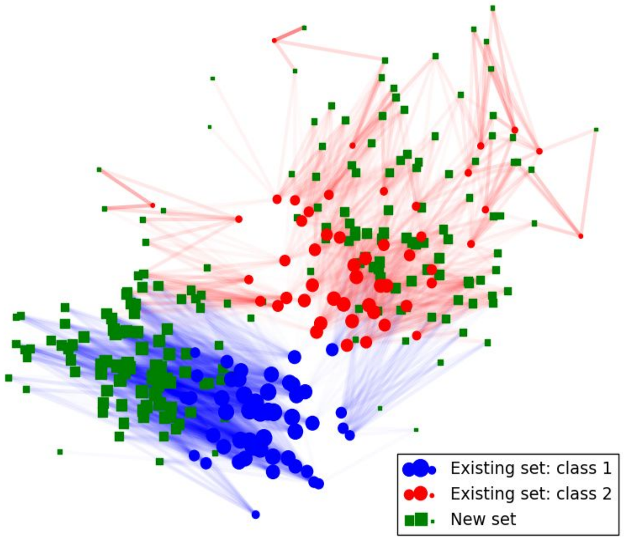
$$\hat{\mathbf{X}}^n = \mathrm{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

Regularised Optimal Transport with Class Labels

Regularise by class label when available for one of the two datasets [3]

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\operatorname{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}} + \lambda \mathbf{R}_s(\gamma) + \eta \mathbf{R}_c(\gamma)$$

$$\mathbf{R}_c(\gamma) = \sum_j \sum_{cl} \|\gamma(\mathcal{I}_{cl}, j)\|_2$$



$$\hat{\mathbf{X}}^n = \operatorname{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

“What is the optimal way to transport mass from domain A to domain B?”

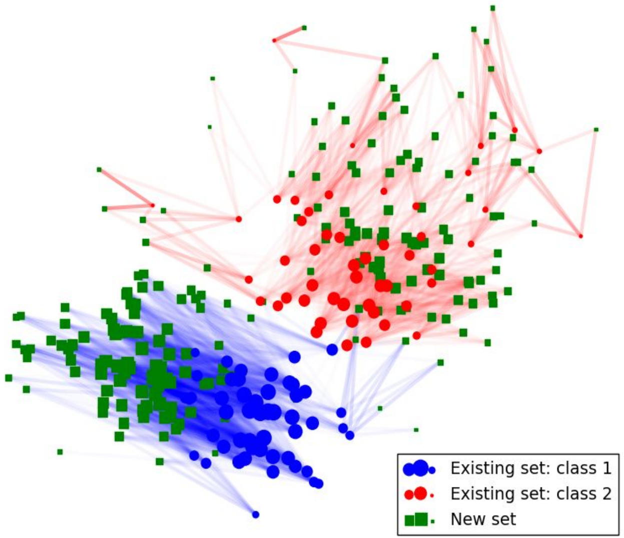
[3] Courty, Nicolas, Rémi Flamary, and Devis Tuia. "Domain adaptation with regularized optimal transport." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2014.

Regularised Optimal Transport with Class Labels

Regularise by class label when available for one of the two datasets [3]

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\mathrm{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}} + \lambda \mathbf{R}_s(\gamma) + \eta \mathbf{R}_c(\gamma)$$

$$\mathbf{R}_c(\gamma) = \sum_j \sum_{cl} \|\gamma(\mathcal{I}_{cl}, j)\|_2$$



$$\hat{\mathbf{X}}^n = \mathrm{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

“What is the optimal way to transport mass from domain A to domain B?”

[3] Courty, Nicolas, Rémi Flamary, and Devis Tuia. "Domain adaptation with regularized optimal transport." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2014.

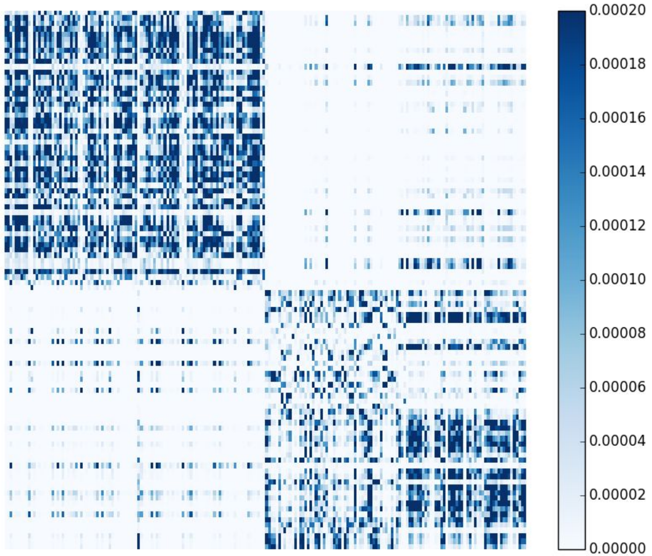
Regularised Optimal Transport with Class Labels

Regularise by class label when available for one of the two domains

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\operatorname{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}} + \lambda \mathbf{R}_s(\gamma) + \eta \mathbf{R}_c(\gamma)$$

$$\mathbf{R}_c(\gamma) = \sum_j \sum_{cl} \|\gamma(\mathcal{I}_{cl}, j)\|_2$$

“What is the optimal way to transport mass from domain A to domain B?”



$$\hat{\mathbf{X}}^n = \operatorname{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

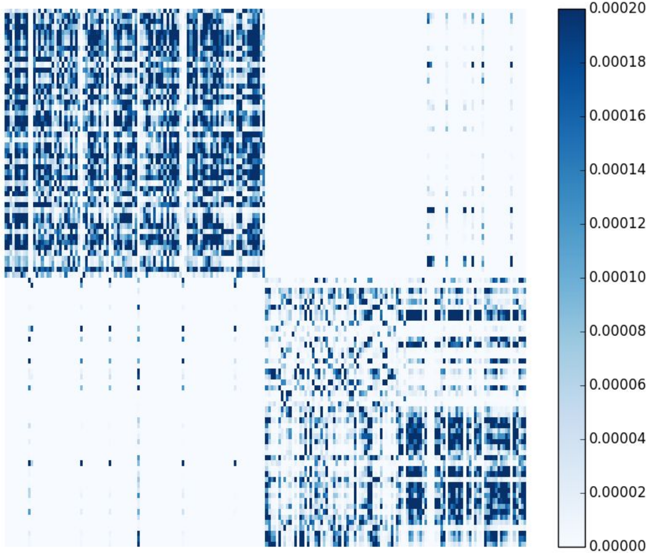
Regularised Optimal Transport with Class Labels

Regularise by class label when available for one of the two domains

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\operatorname{argmin}} \langle \gamma, \mathbf{C} \rangle_{\mathbf{F}} + \lambda \mathbf{R}_s(\gamma) + \eta \mathbf{R}_c(\gamma)$$

$$\mathbf{R}_c(\gamma) = \sum_j \sum_{cl} \|\gamma(\mathcal{I}_{cl}, j)\|_2$$

"What is the optimal way to transport mass from domain A to domain B?"



$$\hat{\mathbf{X}}^n = \operatorname{diag}(\gamma_0^\top \mathbf{1}_{N_e})^{-1} \gamma_0^\top \mathbf{X}_e$$

APPLYING CL-REGULARISED OT TO P300

Applying CL-Regularised OT to P300 Speller Data

Transport a set of unlabeled feature vectors onto the domain of a set of labeled feature vectors.

Training

Input: Sets X^e, Y^e

- Train classifier
- Compute probability vector μ_e

Testing

Input: Set X^n

- Compute probability vector μ_n
- Compute γ
- Transport feature vectors: $\hat{X}^n$
  - Input to classifier

Output: Set Y^n

Experiments

Dataset

- ➢ EEG signals recorded during P300 speller sessions at the CHU of Nice.
- ➢ Adult patients suffering from Amyotrophic Lateral Sclerosis (ALS).
- ➢ 12 Subjects
  - 1 Session per subject (calibration)
- ➢ 12 electrodes

Pairwise Transfer Learning

Train with one session

$$\{\mathbf{X}_i^e, Y_i^e\}, i \in \{A1, \ldots, A12\}$$

Test with one session

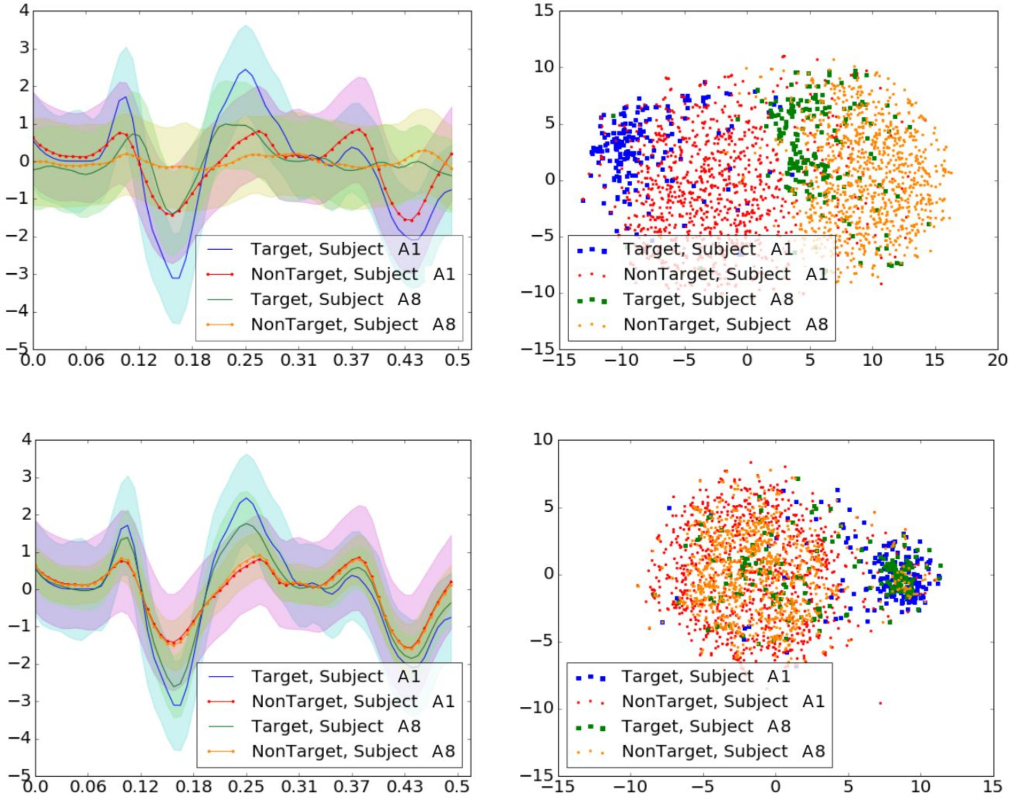
$$\mathbf{X}_j^n, j \in \{A1, \ldots, A12\}, j \neq i$$

132 experiments

Example

1st Xdawn Component

2d Projection



Before Transportation

After Transportation

RESULTS

Results

Existing Session	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	**Avg.**
XD+LDA	0.535 ± 0.05	0.562 ± 0.07	0.598 ± 0.09	0.600 ± 0.07	0.591 ± 0.10	0.595 ± 0.08	0.570 ± 0.06	0.578 ± 0.06	0.516 ± 0.02	0.566 ± 0.07	0.553 ± 0.06	0.526 ± 0.02	**0.566** ± **0.03**
XD+OT+LDA	0.627 ± 0.07	0.539 ± 0.02	0.567 ± 0.06	0.548 ± 0.04	0.611 ± 0.11	0.598 ± 0.07	0.560 ± 0.06	0.490 ± 0.17	0.518 ± 0.01	0.551 ± 0.04	0.583 ± 0.05	0.585 ± 0.06	**0.565** ± **0.04**

Average performance (area under ROC curve) of an existing classifier over 11 experiments.

- ➢ Best performance before transport: 60%
- ➢ Best performances after transport: 62,7%

CONCLUSIONS & FUTURE WORKS

Conclusions & Future Works

- Optimal Transport application enhances the generalisation capacity of existing classifiers
- Computation is fast enough to allow online simulations

- Include more information in the existing set
- Combination of more than one existing set
- Use Optimal Transport theory to reduce dimensionality

Thank you !

Python toolbox used for Optimal Transport Computation: http://pot.readthedocs.io/en/latest/