Optimal transport in biomedical imaging

A Brief Overview

Soheil Kolouri[*] Liam Cattell[†] Gustavo K. Rohde[†]

[*] HRL Laboratories, LLC, Malibu, CA

[†] University of Virginia, Charlottesville, VA

April 25, 2019

https://github.com/skolouri/BAMC2019

Outline

Outline

Optimal Transport (OT)

- Find the most efficient way of transforming one distribution of mass into another.

Notation

- Let $P_p(\Omega)$ denote the set of Borel probability measures with finite p'th moment defined on a given metric space (Ω, d).
- Let $\mu_i \in P_p(X)$ represent probability measures defined on $X \subset \Omega$.
- We use I_i to denote the probability density function corresponding to measure μ_i, i.e., $d\mu_i(x) = I_i(x)dx$.
- We treat normalized positive signals and images as probability density.
- We denote transport maps as f.

Notation

- Let $P_p(\Omega)$ denote the set of Borel probability measures with finite p'th moment defined on a given metric space (Ω, d).
- Let $\mu_i \in P_p(X)$ represent probability measures defined on $X \subset \Omega$.
- We use I_i to denote the probability density function corresponding to measure μ_i, i.e., $d\mu_i(x) = I_i(x)dx$.
- We treat normalized positive signals and images as probability density.
- We denote transport maps as f.

Notation

- Let $P_p(\Omega)$ denote the set of Borel probability measures with finite p'th moment defined on a given metric space (Ω, d).
- Let $\mu_i \in P_p(X)$ represent probability measures defined on $X \subset \Omega$.
- We use I_i to denote the probability density function corresponding to measure μ_i, i.e., $d\mu_i(x) = I_i(x)dx$.
- We treat normalized positive signals and images as probability density.
- We denote transport maps as f.

Notation

- Let $P_p(\Omega)$ denote the set of Borel probability measures with finite p'th moment defined on a given metric space (Ω, d).
- Let $\mu_i \in P_p(X)$ represent probability measures defined on $X \subset \Omega$.
- We use I_i to denote the probability density function corresponding to measure μ_i, i.e., $d\mu_i(x) = I_i(x)dx$.
- We treat normalized positive signals and images as probability density.
- We denote transport maps as f.

Notation

- Let $P_p(\Omega)$ denote the set of Borel probability measures with finite p'th moment defined on a given metric space (Ω, d).
- Let $\mu_i \in P_p(X)$ represent probability measures defined on $X \subset \Omega$.
- We use I_i to denote the probability density function corresponding to measure μ_i, i.e., $d\mu_i(x) = I_i(x)dx$.
- We treat normalized positive signals and images as probability density.
- We denote transport maps as f.

OT - Monge Formulation

- Find mass-preserving (MP) mapping f between two probablity density functions I_0 and I_1.

$$\int_{f^{-1}(A)} I_0(x)dx = \int_A I_1(y)dy, \quad A \subseteq X$$

$$I_0(x) = \det(D_f(x))I_1(f(x)) \rightarrow I_0 = \det(D_f)I_1 \circ f$$

OT - Monge Formulation

- Find mass-preserving (MP) mapping f between two probablity density functions I_0 and I_1.

$$\int_{f^{-1}(A)} I_0(x)dx = \int_A I_1(y)dy, \quad A \subseteq X$$

$$I_0(x) = \det(D_f(x))I_1(f(x)) \to I_0 = \det(D_f)I_1 \circ f$$

OT - Monge Formulation

- Find mass-preserving (MP) mapping f between two probablity density functions I_0 and I_1.

$$\int_{f^{-1}(A)} I_0(x)dx = \int_A I_1(y)dy, \quad A \subseteq X$$

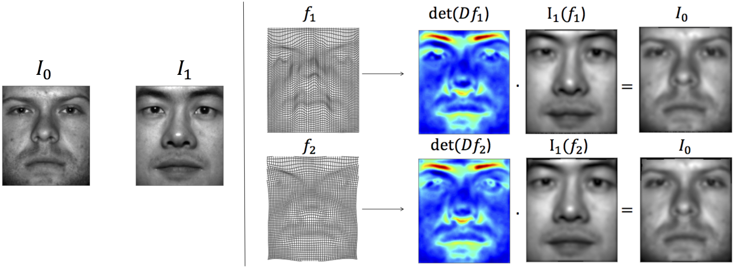
$$I_0(x) = \det(D_f(x))I_1(f(x)) \rightarrow I_0 = \det(D_f)I_1 \circ f$$

OT - Monge Formulation

- Find mass-preserving (MP) mapping f between two positive density functions I_0 and I_1.

$$\int_{f^{-1}(A)} I_0(x)dx = \int_A I_1(y)dy, \quad A \subseteq X$$

$$I_0(x) = \det(D_f(x))I_1(f(x)) \rightarrow I_0 = \det(D_f)I_1 \circ f$$



OT - Monge Formulation

- Find mass-preserving (MP) mapping f between two positive density functions I_0 and I_1.

$$\int_{f^{-1}(A)} I_0(x)dx = \int_A I_1(y)dy, \quad A \subseteq X$$

$$I_0(x) = \det(D_f(x)) I_1(f(x)) \rightarrow I_0 = \det(D_f) I_1 \circ f$$

- Mapping f must be optimal in some sense.

$$M(f) = \inf_{f \in \text{MP}} \int_X c(x, f(x)) I_0(x) dx$$

- Cost function: $c(x, f(x)) = \|f(x) - x\|^2$
- Find the MP mapping that provides the least amount of expected translation to morph I_0 into I_1.

OT - Monge Formulation

- Find mass-preserving (MP) mapping f between two positive density functions I_0 and I_1.

$$\int_{f^{-1}(A)} I_0(x)dx = \int_A I_1(y)dy, \quad A \subseteq X$$

$$I_0(x) = \det(D_f(x)) I_1(f(x)) \rightarrow I_0 = \det(D_f) I_1 \circ f$$

- Mapping f must be optimal in some sense.

$$M(f) = \inf_{f \in \mathrm{MP}} \int_X c(x, f(x)) I_0(x) dx$$

- Cost function: $c(x, f(x)) = \|f(x) - x\|^2$
- Find the MP mapping that provides the least amount of expected translation to morph I_0 into I_1.

OT - Monge Formulation

- Find mass-preserving (MP) mapping f between two positive density functions I_0 and I_1.

$$\int_{f^{-1}(A)} I_0(x)dx = \int_A I_1(y)dy, \quad A \subseteq X$$

$$I_0(x) = \det(D_f(x))I_1(f(x)) \rightarrow I_0 = \det(D_f)I_1 \circ f$$

- Mapping f must be optimal in some sense.

$$M(f) = \inf_{f \in \mathrm{MP}} \int_X c(x, f(x))I_0(x)dx$$

- Cost function: $c(x, f(x)) = \|f(x) - x\|^2$
- Find the MP mapping that provides the least amount of expected translation to morph I_0 into I_1.

OT - Monge Formulation

- Find mass-preserving (MP) mapping f between two positive density functions I_0 and I_1.

$$\int_{f^{-1}(A)} I_0(x)dx = \int_A I_1(y)dy, \quad A \subseteq X$$

$$I_0(x) = \det(D_f(x))I_1(f(x)) \rightarrow I_0 = \det(D_f)I_1 \circ f$$

- Mapping f must be optimal in some sense.

$$M(f) = \inf_{f \in \mathrm{MP}} \int_X c(x, f(x))I_0(x)dx$$

- Cost function: $c(x, f(x)) = \|f(x) - x\|^2$
- Find the MP mapping that provides the least amount of expected translation to morph I_0 into I_1.

OT - Monge Formulation

- Find mass-preserving (MP) mapping f between two positive density functions I_0 and I_1.

$$\int_{f^{-1}(A)} I_0(x)dx = \int_A I_1(y)dy, \quad A \subseteq X$$

$$I_0(x) = \det(D_f(x))I_1(f(x)) \to I_0 = \det(D_f)I_1 \circ f$$

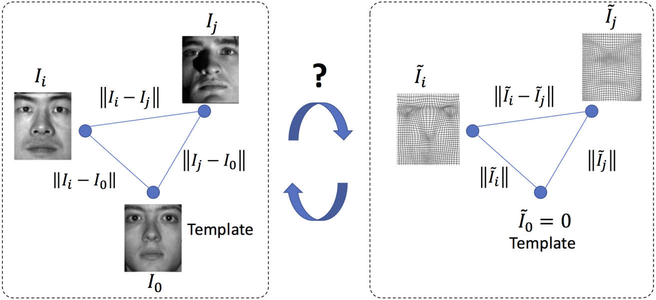
- Mapping f must be optimal in some sense.

$$M(f) = \inf_{f \in \mathrm{MP}} \int_X c(x, f(x))I_0(x)dx$$

- Cost function: $c(x, f(x)) = \|f(x) - x\|^2$
- Find the MP mapping that provides the least amount of expected translation to morph I_0 into I_1.

Transport-Based Transforms

- Tranport-based transform provides an embedding for signals and images, in which the Euclidean distance is equal (or it approximates) the 2-Wasserstein distance.



- such that, $\|\tilde{I}_i\| = W_2(I_i, I_0)$ and $\|\tilde{I}_i - \tilde{I}_j\| \approx W_2(I_i, I_j)$ for $\forall i, j \neq 0$.

Transport-Based Transforms

- 1D: Cumulative distribution transform (CDT)

$$\tilde{I}_i = (f - \mathrm{Id})\sqrt{I_0}$$

$$I_i = (f^{-1})' I_0 \circ f^{-1}$$

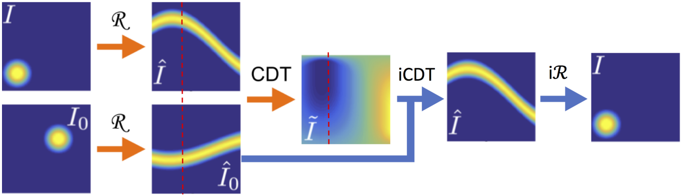
Transport-Based Transforms

- 1D: Cumulative distribution transform (CDT)

$$\tilde{I}_i = (f - \mathrm{Id})\sqrt{I_0}$$

$$I_i = (f^{-1})' I_0 \circ f^{-1}$$

- nD: Radon-CDT



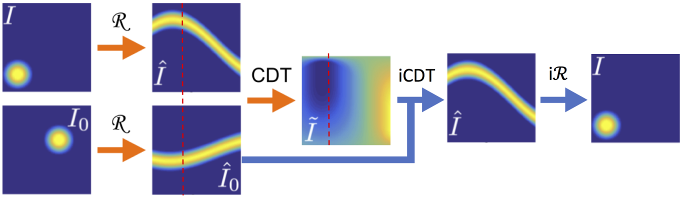
Transport-Based Transforms

- 1D: Cumulative distribution transform (CDT)

$$\tilde{I}_i = (f - \mathrm{Id})\sqrt{I_0}$$

$$I_i = (f^{-1})' I_0 \circ f^{-1}$$

- nD: Radon-CDT

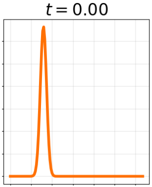


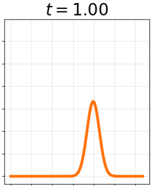
- nD: Linear optimal transport (LOT)

Outline

Why use transport-based transforms?

- Linear interpolation in signal space.

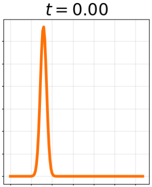


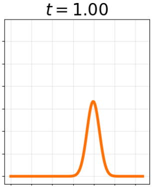


$$I_t = (1-t)I_0 + tI_1$$

Why use transport-based transforms?

- Linear interpolation in signal space.

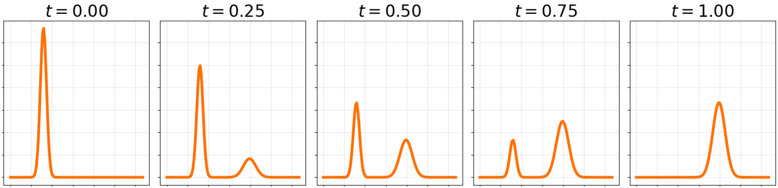




$$I_t = (1 - t)I_0 + tI_1$$

Why use transport-based transforms?

- Linear interpolation in signal space.

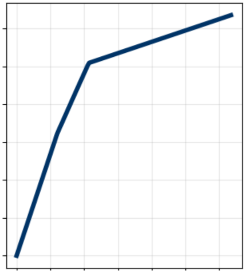


$$I_t = (1 - t)I_0 + tI_1$$

Why use transport-based transforms?

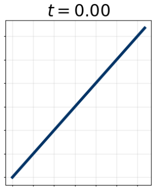
- Linear interpolation in transport (CDT) space.
- Compute transport map f between I_0 and I_1.

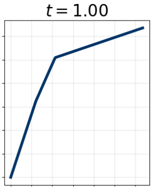
$$I_0 = f' I_1 \circ f$$



Why use transport-based transforms?

- Linear interpolation in transport (CDT) space.

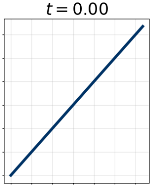


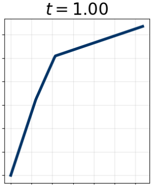


$$f_t = (1 - t)x + tf$$

Why use transport-based transforms?

- Linear interpolation in transport (CDT) space.

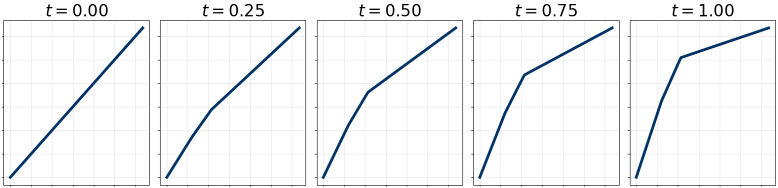




$$f_t = (1 - t)x + tf$$

Why use transport-based transforms?

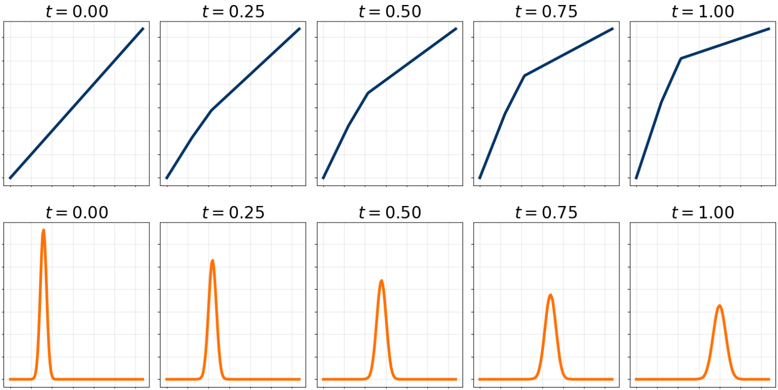
- Linear interpolation in transport (CDT) space.



$$f_t = (1 - t)x + tf$$

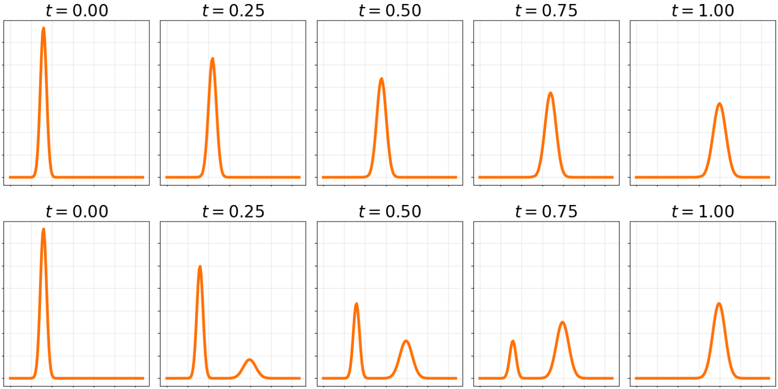
Why use transport-based transforms?

- Apply inverse to recover signal: $I_t = (f^{-1})' I_0 \circ f^{-1}$



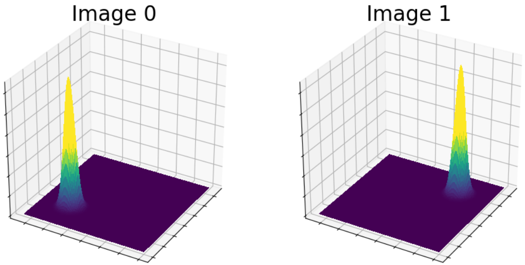
Why use transport-based transforms?

- Compare transport space and signal space.



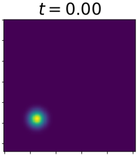
Why use transport-based transforms?

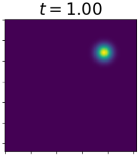
- Can extend this to n dimensions.



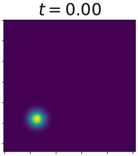
Why use transport-based transforms?

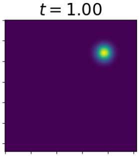
$$I_t = (1 - t)I_0 + tI_1$$





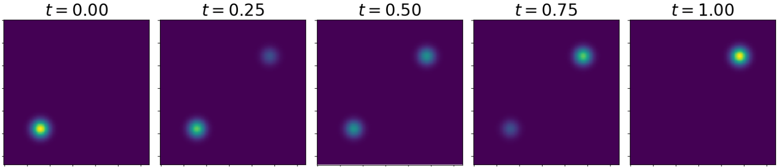
$$f_t = (1 - t)\mathrm{Id} + tf$$



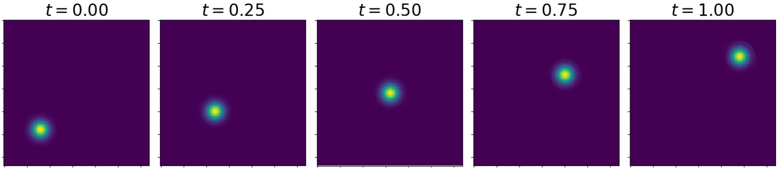


Why use transport-based transforms?

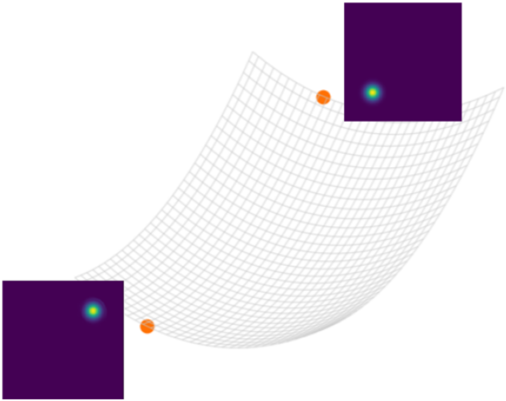
$$I_t = (1-t)I_0 + tI_1$$



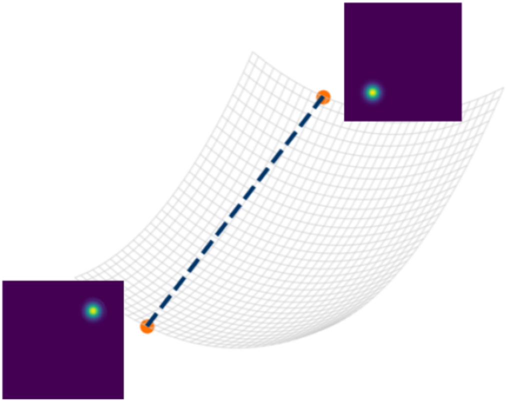
$$f_t = (1-t)\mathrm{Id} + tf$$



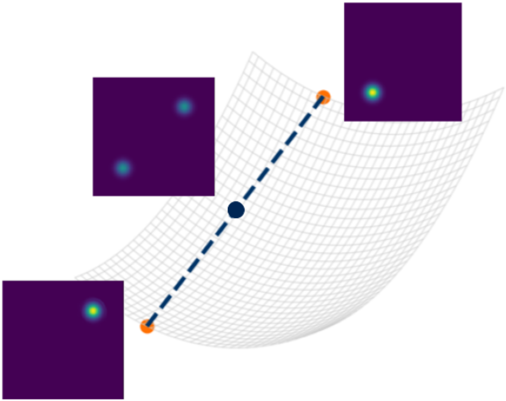
Geodesics



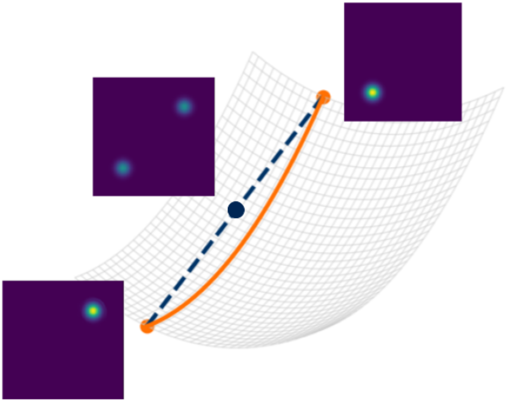
Geodesics



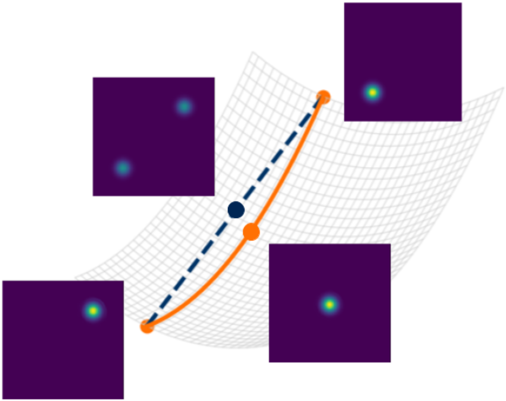
Geodesics



Geodesics



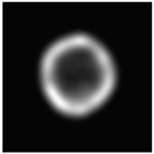
Geodesics



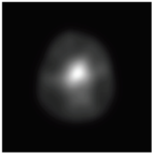
Geodesics

- Can use this to model changes due to cancer.

Normal



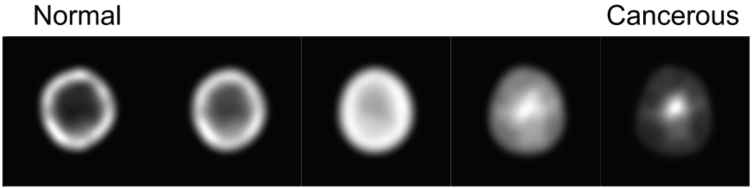
Cancerous



Basu et al., *PNAS*, 111(9):3448-3453, 2014

Geodesics

- Can use this to model changes due to cancer.



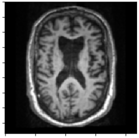
Basu et al., *PNAS*, 111(9):3448-3453, 2014

Geodesics

- Can use this to model changes over time.

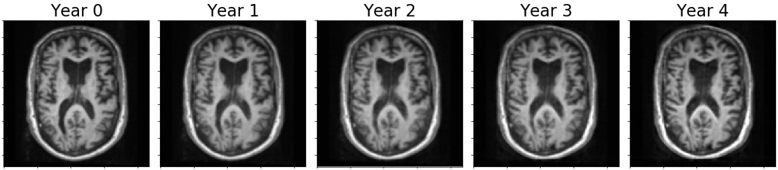
Year 0

Year 4



Geodesics

- Can use this to model changes over time.



Geodesics

Can compute the geodesics using *optimaltransport* package.

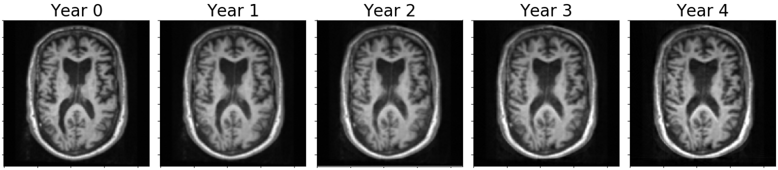
```
from optrans.utils import signal_to_pdf
from optrans.continuous import RadonCDT

# Load images
...

# Smooth the images and convert to PDFs
img0 = signal_to_pdf(img0, sigma=.8)
img1 = signal_to_pdf(img1, sigma=.8)
```

Geodesics

```python
# Compute the Radon-CDT of img1 w.r.t. img0
radoncdt = RadonCDT()
rcdt = radoncdt.forward(img0, img1)

# Get the domain of our signal
x = radoncdt.transport_map_ + radoncdt.displacements_

# Linearly interpolate in Radon-CDT space
img_recon = []
for t in np.linspace(0, 1, 5):
    u = radoncdt.displacements_ * t
    f = x - u
    img_recon.append(radoncdt.apply_inverse_map(f,
        img0))
```

Geodesics

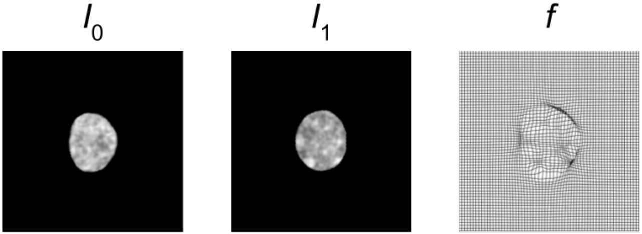
- Plot resulting images.



Outline

Image morphing

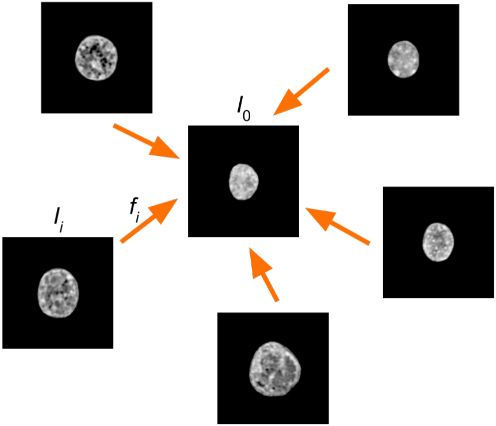
- Image morphing through optimal transport maps



W. Wang et al., *IJCV*, 101:254–269, 2013
S. Basu et al., *PNAS*, 111(9):3448–3453, 2014
S. Kolouri et al., *Pattern Recognit.*, 51:453–462, 2016

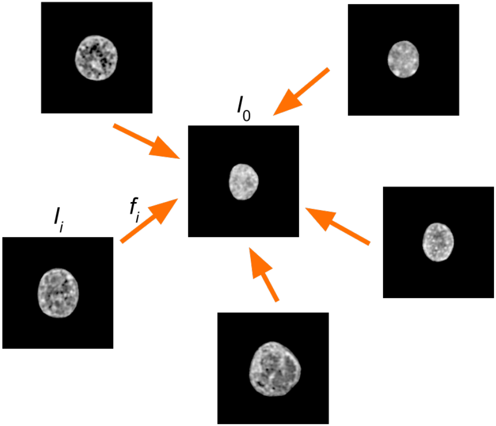
Shape modeling

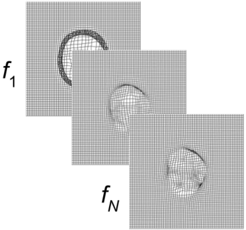
- Image morphing through optimal transport maps



Shape modeling

- Image morphing through optimal transport maps

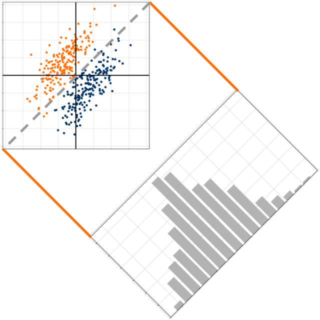




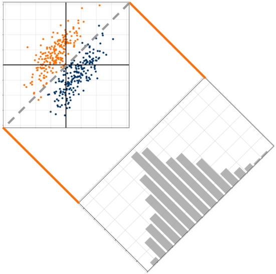
Shape modeling

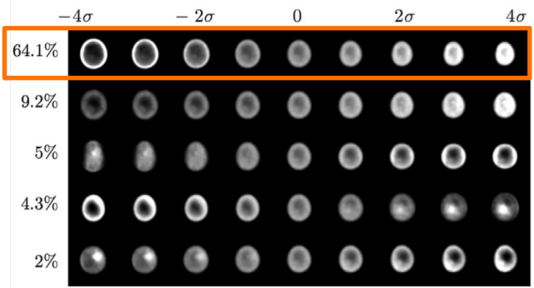
- Perform principal component analysis (PCA) on transport maps.

$$\boldsymbol{w}_{\mathrm{PCA}} = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^T \Sigma_T \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{w}}$$



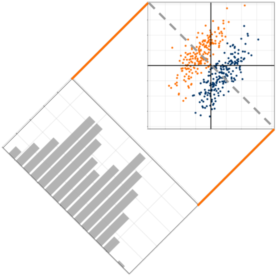
Shape modeling

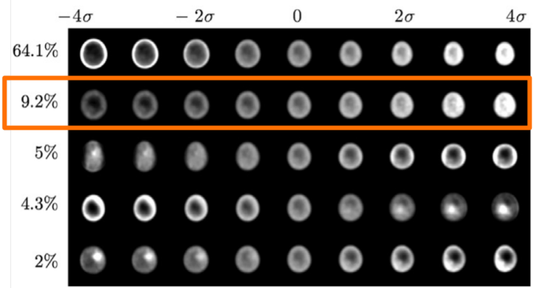




S. Kolouri et al., *Pattern Recognit.*, 51:453-462, 2016

Shape modeling

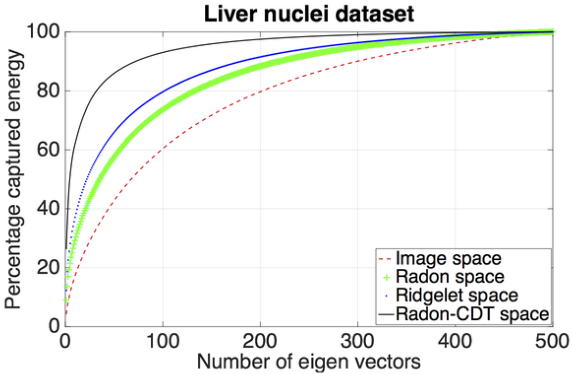




S. Kolouri et al., *Pattern Recognit.*, 51:453-462, 2016

Better representation

- OT can provide a more parsimonious data model.

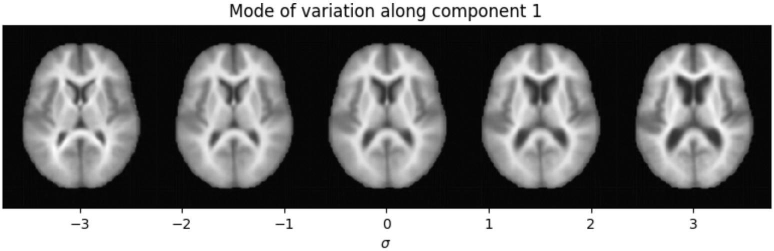


S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

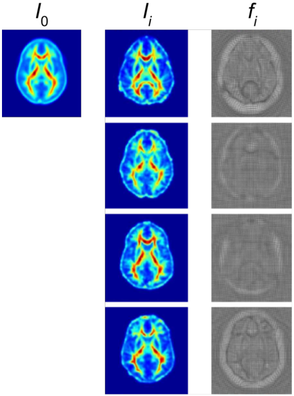
PCA modes

```
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from optrans.continuous import RadonCDT
from optrans.visualization import plot_mode_image

# Compute transport maps (f) & ref. image (img0)...

# Data dimensions
n_samples, h, w = f.shape

# Fit PCA transform
pca = PCA(n_components=3)
f_pca = pca.fit_transform(f.reshape(n_samples,-1))

ax = plot_mode_image([pca], component=0, shape=(h,w),
    transform=RadonCDT(), img0=img0, n_std=6, n_steps
    =5, cmap='gray')
plt.show()
```

PCA modes

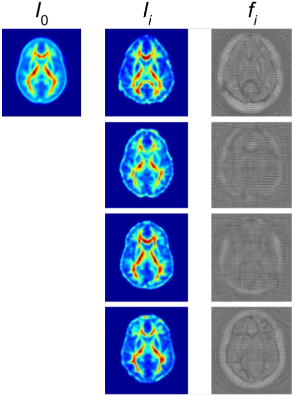


Correlation



S. Kundu, *PhD thesis*, CMU, 2016

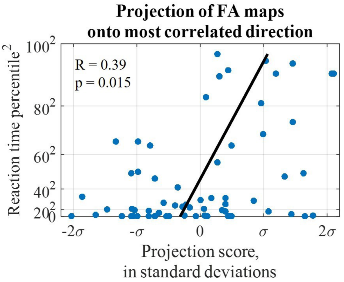
Correlation



$$\boldsymbol{w}_{\text{corr}} = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^T X \boldsymbol{v}}{\boldsymbol{w}^T \boldsymbol{w}}$$

X: OT data

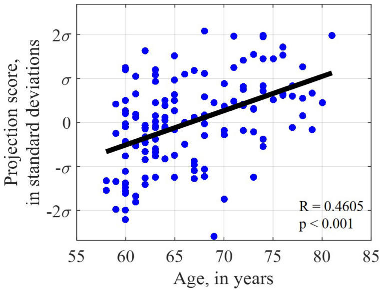
$\boldsymbol{v}$: reaction time

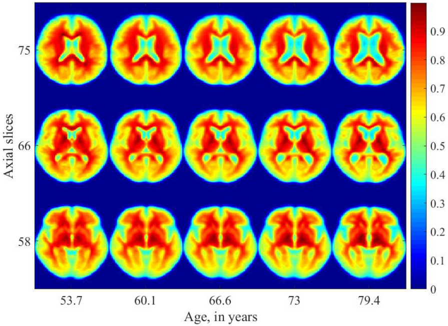


S. Kundu, *PhD thesis*, CMU, 2016

Correlation

- Correlation using transport maps.





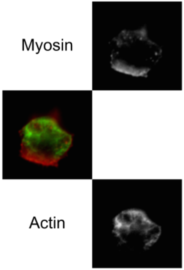
S. Kundu et al., *NeuroImage*, 167:256–275, 2018

X: OT data

$\boldsymbol{v}$: age

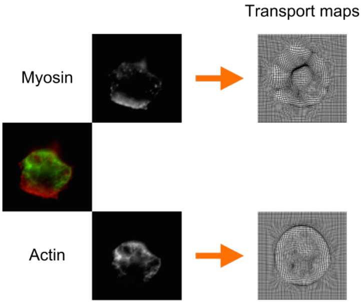
Correlation

- Canonical correlation analysis (CCA).



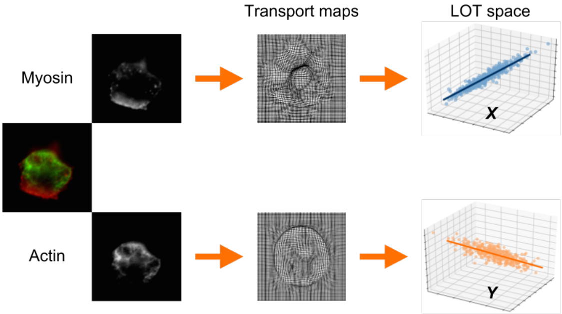
Correlation

- Canonical correlation analysis (CCA).



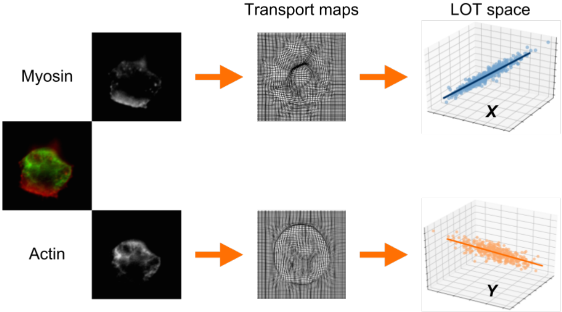
Correlation

- Canonical correlation analysis (CCA).



Correlation

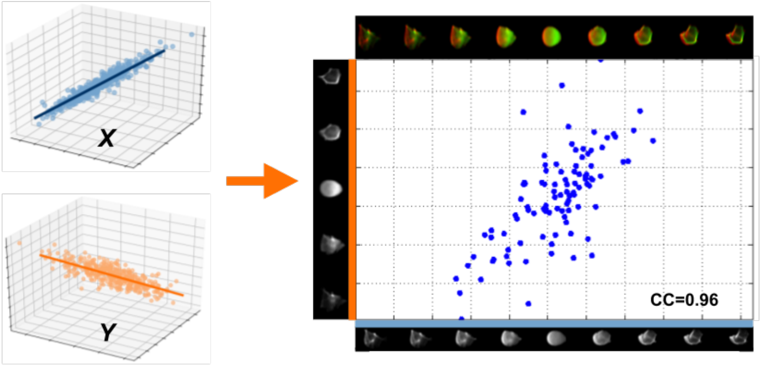
- Canonical correlation analysis (CCA).



$$\arg\max_{\boldsymbol{u},\boldsymbol{v}} \frac{\boldsymbol{u}^T X Y^T \boldsymbol{v}}{\sqrt{(\boldsymbol{u}^T X X^T \boldsymbol{u})(\boldsymbol{v}^T Y Y^T \boldsymbol{v})}}$$

Correlation

- Canonical correlation analysis (CCA).

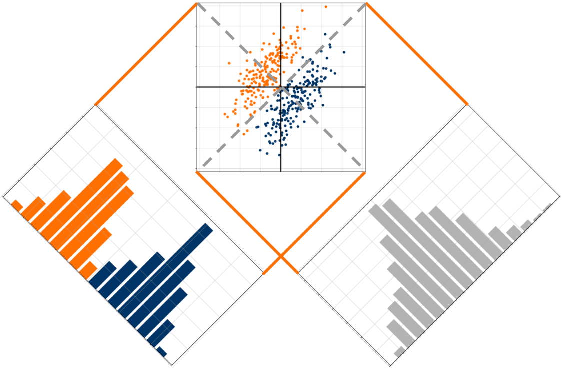


S. Kolouri et al., *ISBI*, 2014

Outline

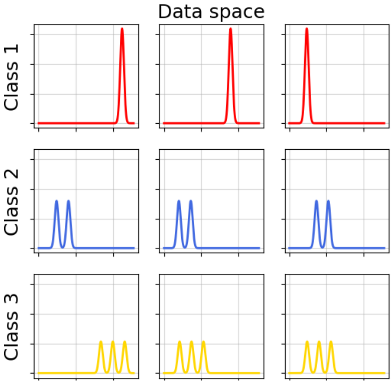
Linear discriminant analysis

$$\boldsymbol{w}_{\mathrm{LDA}} = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^T \Sigma_b \boldsymbol{w}}{\boldsymbol{w}^T \Sigma_w \boldsymbol{w}} \qquad \boldsymbol{w}_{\mathrm{PCA}} = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^T \Sigma_T \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{w}}$$



Classification

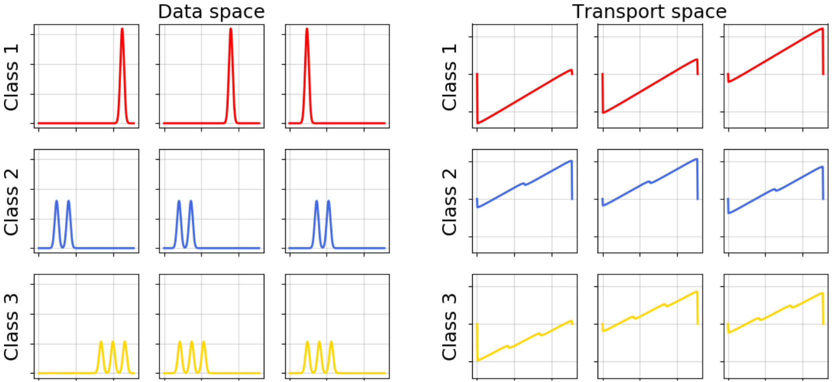
- Classify 1D signals using linear classifier.



S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Signal Process. Mag.*, 34(4):43–59, 2017

Classification

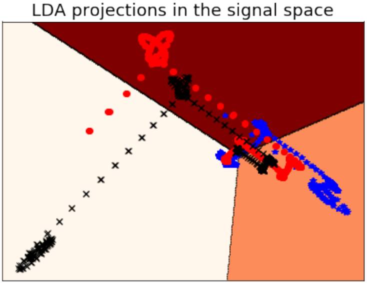
- Classify 1D signals using linear classifier.

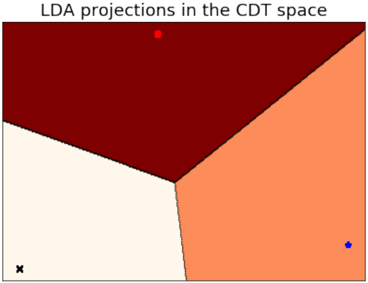


S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Signal Process. Mag.*, 34(4):43–59, 2017

Classification

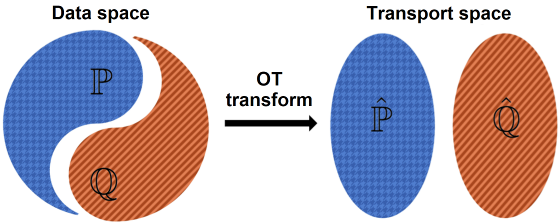
- Classify 1D signals using linear classifier.





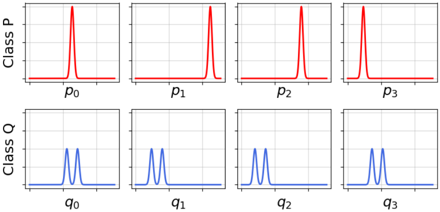
Linear separability theorem

- Under certain conditions, data can become linearly separable in transport space.



L. Cattell et al., *arXiv:1802.07163*, 2018
S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Linear separability theorem

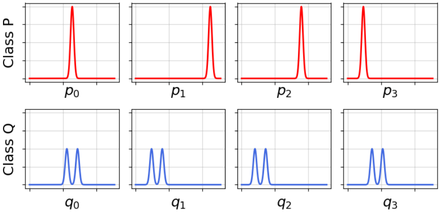


- Let $\mathbb{F}$ be a subset of transport maps
- $p_i = f_i' p_0 \circ f_i, \quad f_i \in \mathbb{F}$
- $q_j = f_j' q_0 \circ f_j, \quad f_j \in \mathbb{F}$
- $q_0 \neq f' p_0 \circ f, \quad \forall f \in \mathbb{F}$

1. $\forall f \in \mathbb{F}, \quad f^{-1} \in \mathbb{F}$
2. $\forall f \in \mathbb{F}, \quad \sum_i \alpha_i f_i \in \mathbb{F},$ where $\sum_i \alpha_i = 1$
3. $\forall f_1, f_2 \in \mathbb{F}, \quad f_1 \circ f_2 \in \mathbb{F}$

L. Cattell et al., *arXiv:1802.07163*, 2018
S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Linear separability theorem

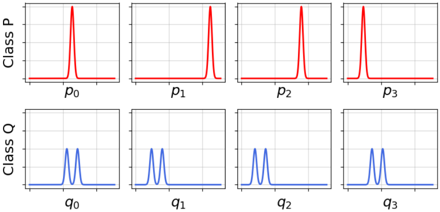


- Let $\mathbb{F}$ be a subset of transport maps
- $p_i = f_i' p_0 \circ f_i, \quad f_i \in \mathbb{F}$
- $q_j = f_j' q_0 \circ f_j, \quad f_j \in \mathbb{F}$
- $q_0 \neq f' p_0 \circ f, \quad \forall f \in \mathbb{F}$

1. $\forall f \in \mathbb{F}, \quad f^{-1} \in \mathbb{F}$
2. $\forall f \in \mathbb{F}, \quad \sum_i \alpha_i f_i \in \mathbb{F}$, where $\sum_i \alpha_i = 1$
3. $\forall f_1, f_2 \in \mathbb{F}, \quad f_1 \circ f_2 \in \mathbb{F}$

L. Cattell et al., *arXiv:1802.07163*, 2018
S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Linear separability theorem

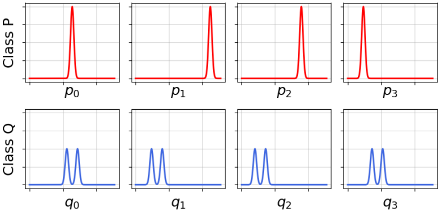


- Let $\mathbb{F}$ be a subset of transport maps
- $p_i = f_i' p_0 \circ f_i, \quad f_i \in \mathbb{F}$
- $q_j = f_j' q_0 \circ f_j, \quad f_j \in \mathbb{F}$
- $q_0 \neq f' p_0 \circ f, \quad \forall f \in \mathbb{F}$

1. $\forall f \in \mathbb{F}, \quad f^{-1} \in \mathbb{F}$
2. $\forall f \in \mathbb{F}, \quad \sum_i \alpha_i f_i \in \mathbb{F}$, where $\sum_i \alpha_i = 1$
3. $\forall f_1, f_2 \in \mathbb{F}, \quad f_1 \circ f_2 \in \mathbb{F}$

L. Cattell et al., *arXiv:1802.07163*, 2018
S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Linear separability theorem

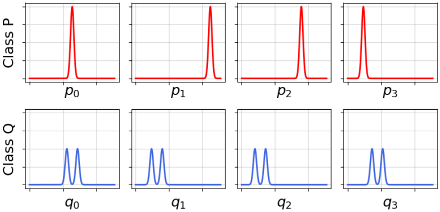


- Let $\mathbb{F}$ be a subset of transport maps
- $p_i = f_i' p_0 \circ f_i, \quad f_i \in \mathbb{F}$
- $q_j = f_j' q_0 \circ f_j, \quad f_j \in \mathbb{F}$
- $q_0 \neq f' p_0 \circ f, \quad \forall f \in \mathbb{F}$

1. $\forall f \in \mathbb{F}, \quad f^{-1} \in \mathbb{F}$
2. $\forall f \in \mathbb{F}, \quad \sum_i \alpha_i f_i \in \mathbb{F}$, where $\sum_i \alpha_i = 1$
3. $\forall f_1, f_2 \in \mathbb{F}, \quad f_1 \circ f_2 \in \mathbb{F}$

L. Cattell et al., *arXiv:1802.07163*, 2018
S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Linear separability theorem

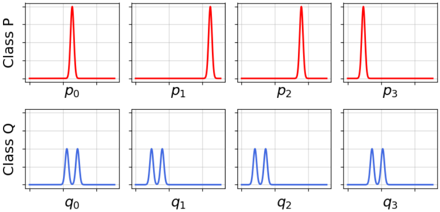


- Let $\mathbb{F}$ be a subset of transport maps
- $p_i = f_i' p_0 \circ f_i, \quad f_i \in \mathbb{F}$
- $q_j = f_j' q_0 \circ f_j, \quad f_j \in \mathbb{F}$
- $q_0 \neq f' p_0 \circ f, \quad \forall f \in \mathbb{F}$

1. $\forall f \in \mathbb{F}, \quad f^{-1} \in \mathbb{F}$
2. $\forall f \in \mathbb{F}, \quad \sum_i \alpha_i f_i \in \mathbb{F}$, where $\sum_i \alpha_i = 1$
3. $\forall f_1, f_2 \in \mathbb{F}, \quad f_1 \circ f_2 \in \mathbb{F}$

L. Cattell et al., *arXiv:1802.07163*, 2018
S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Linear separability theorem

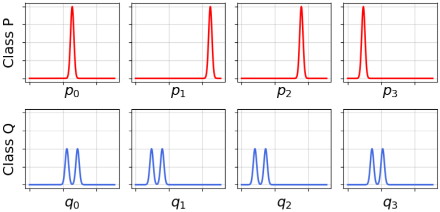


- Let $\mathbb{F}$ be a subset of transport maps
- $p_i = f_i' p_0 \circ f_i, \quad f_i \in \mathbb{F}$
- $q_j = f_j' q_0 \circ f_j, \quad f_j \in \mathbb{F}$
- $q_0 \neq f' p_0 \circ f, \quad \forall f \in \mathbb{F}$

1. $\forall f \in \mathbb{F}, \quad f^{-1} \in \mathbb{F}$
2. $\forall f \in \mathbb{F}, \quad \sum_i \alpha_i f_i \in \mathbb{F}$, where $\sum_i \alpha_i = 1$
3. $\forall f_1, f_2 \in \mathbb{F}, \quad f_1 \circ f_2 \in \mathbb{F}$

L. Cattell et al., *arXiv:1802.07163*, 2018
S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Linear separability theorem

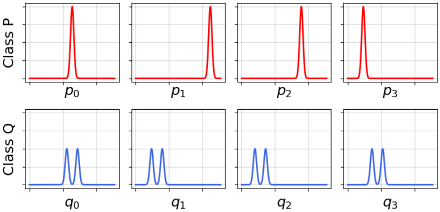


- Let $\mathbb{F}$ be a subset of transport maps
- $p_i = f_i' p_0 \circ f_i, \quad f_i \in \mathbb{F}$
- $q_j = f_j' q_0 \circ f_j, \quad f_j \in \mathbb{F}$
- $q_0 \neq f' p_0 \circ f, \quad \forall f \in \mathbb{F}$

1. $\forall f \in \mathbb{F}, \quad f^{-1} \in \mathbb{F}$
2. $\forall f \in \mathbb{F}, \quad \sum_i \alpha_i f_i \in \mathbb{F}$, where $\sum_i \alpha_i = 1$
3. $\forall f_1, f_2 \in \mathbb{F}, \quad f_1 \circ f_2 \in \mathbb{F}$

L. Cattell et al., *arXiv:1802.07163*, 2018
S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Linear separability theorem



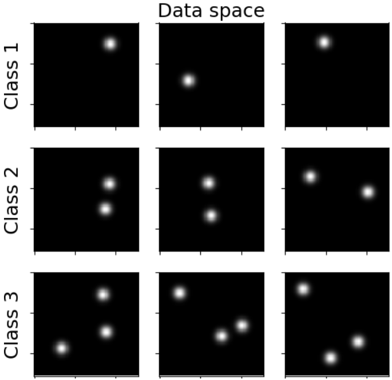
- Let $\mathbb{F}$ be a subset of transport maps
- $p_i = f_i' p_0 \circ f_i, \quad f_i \in \mathbb{F}$
- $q_j = f_j' q_0 \circ f_j, \quad f_j \in \mathbb{F}$
- $q_0 \neq f' p_0 \circ f, \quad \forall f \in \mathbb{F}$

1. $\forall f \in \mathbb{F}, \quad f^{-1} \in \mathbb{F}$
2. $\forall f \in \mathbb{F}, \quad \sum_i \alpha_i f_i \in \mathbb{F},$ where $\sum_i \alpha_i = 1$
3. $\forall f_1, f_2 \in \mathbb{F}, \quad f_1 \circ f_2 \in \mathbb{F}$

L. Cattell et al., *arXiv:1802.07163*, 2018
S.R. Park et al., *Appl. Comput. Harmon. Anal.*, 2017, in press
S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

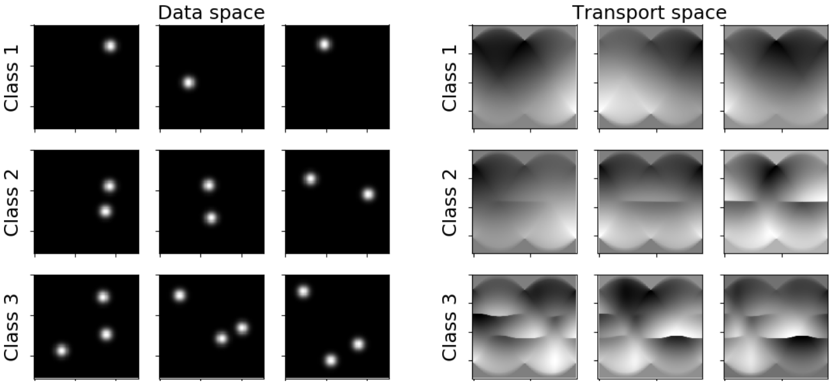
Classification

- Can extend to 2D data using Radon-CDT.



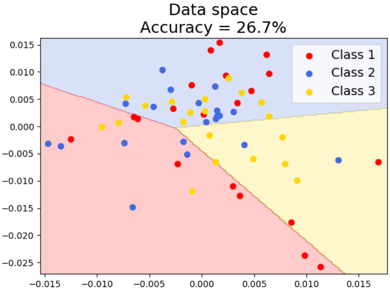
Classification

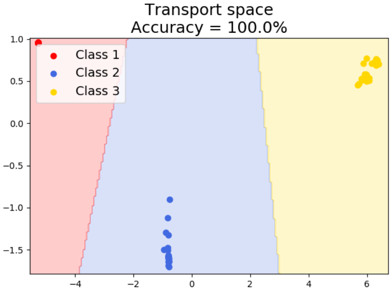
- Can extend to 2D data using Radon-CDT.



Classification

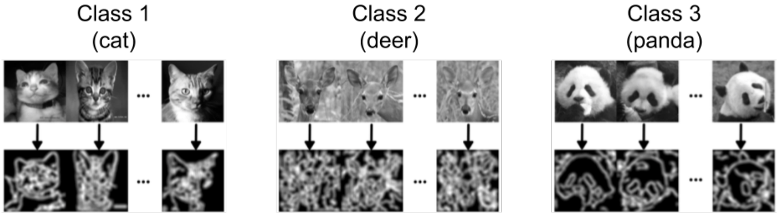
- Can extend to 2D data using Radon-CDT.





Classification

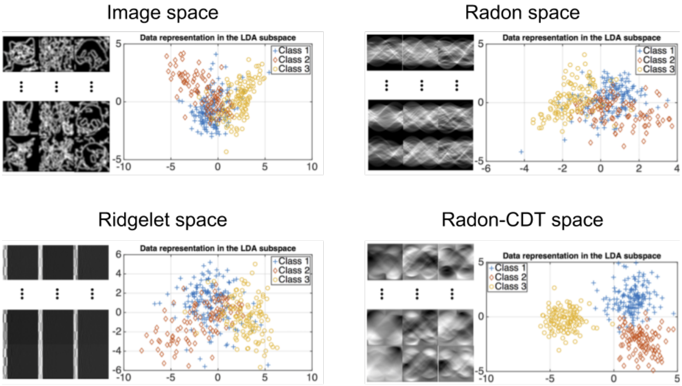
- What about more realistic applications?



S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Classification

- What about more realistic applications?

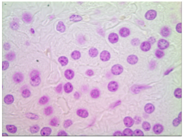


S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Penalized LDA

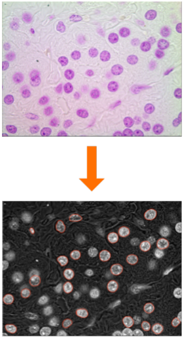
```
from sklearn.model_selection import train_test_split
from optrans.decomposition import PLDA

# Load data X and class labels y
...

# Split data into training and test sets
Xtr, Xte, ytr, yte = train_test_split(X, y, test_size=0.2, stratify=y)

# Train PLDA classifier
plda = PLDA(n_components=2, alpha=1.)
Xtr_plda = plda.fit_transform(Xtr, ytr)

# Test classifier
acc = plda.score(Xte, yte)
print("ACC: {:.3f}".format(acc))
```

Histology example



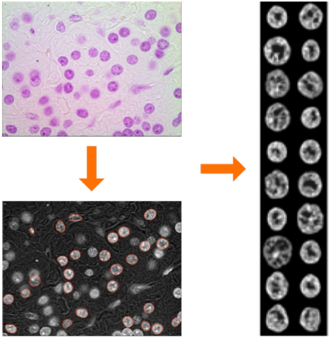
J.A. Ozolek, Dept. of Pathology, UPMC
S. Basu et al., *PNAS*, 11(9):3448–3453, 2014

Histology example



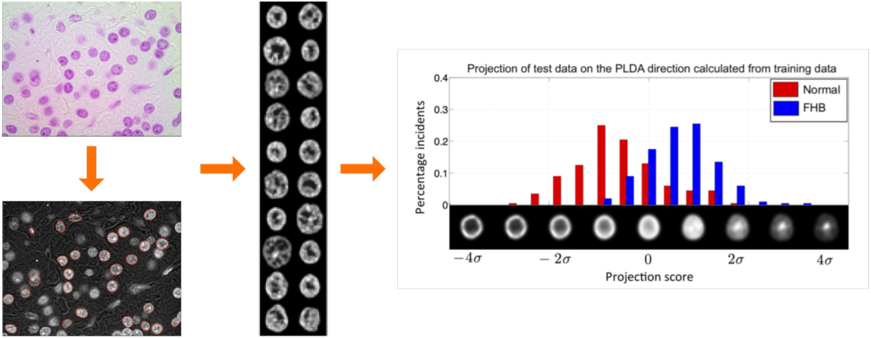
J.A. Ozolek, Dept. of Pathology, UPMC
S. Basu et al., *PNAS*, 11(9):3448–3453, 2014

Histology example



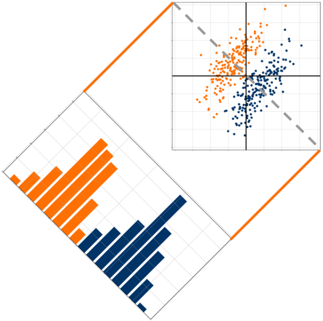
J.A. Ozolek, Dept. of Pathology, UPMC
S. Basu et al., *PNAS*, 11(9):3448–3453, 2014

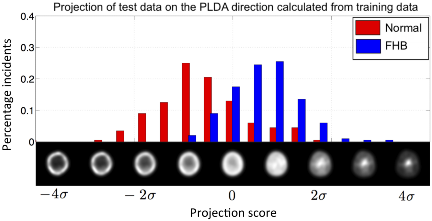
Histology example



J.A. Ozolek, Dept. of Pathology, UPMC
S. Basu et al., *PNAS*, 11(9):3448–3453, 2014

Histology example

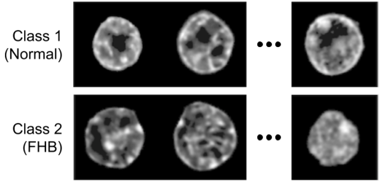




J.A. Ozolek, Dept. of Pathology, UPMC
S. Basu et al., *PNAS*, 11(9):3448–3453, 2014

Histology example

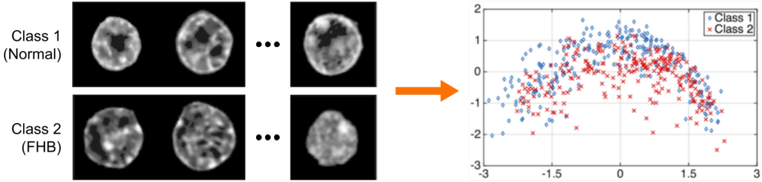
- Classify dataset using LDA classifier.



S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Histology example

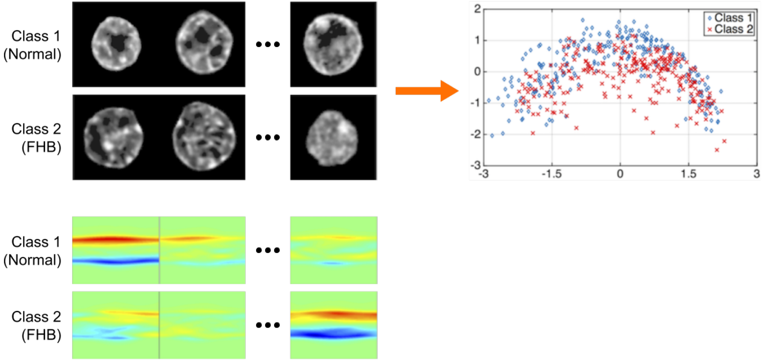
- Classify dataset using LDA classifier.



S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Histology example

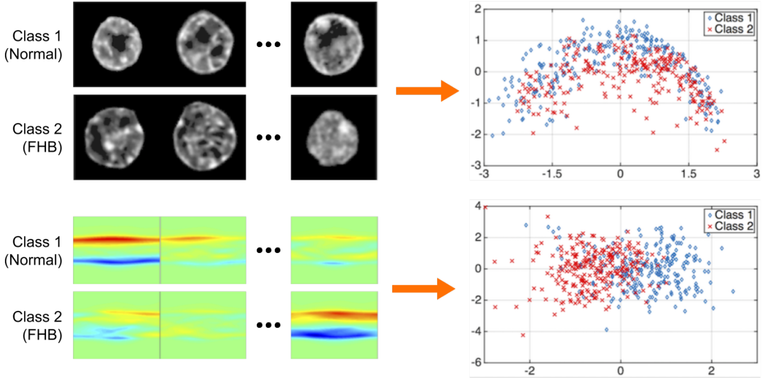
- Classify dataset using LDA classifier.



S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Histology example

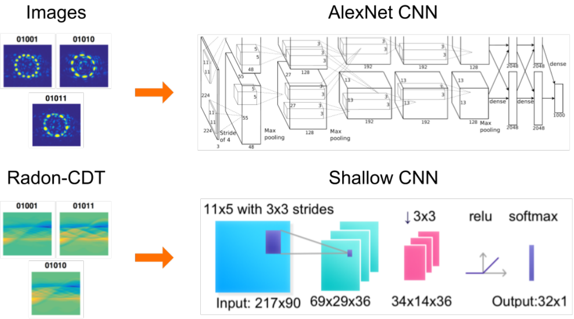
- Classify dataset using LDA classifier.



S. Kolouri et al., *IEEE Trans. Image Process.*, 25(2):920–934, 2015

Convolutional neural networks

- Allows for simpler classifiers (fewer parameters).



- Accuracy for high turbulence: 99.48% vs 99.60%.
- Radon-CDT + shallow CNN requires 90x fewer FLOPS.

S.R. Park et al., *Optics Express*, 26(4):4004-4022, 2018

Outline

Optimal Transport

Visualizing transformations

Transport-based morphometry

Classification

Inverse problems

Conclusions

Super-resolution

- $J = \phi(I) + n$
- $\phi(\cdot)$: downsampling function (known)
- n: noise

$$I^* = \arg\min_I \frac{1}{2}\|J - \phi(I)\|^2 + \frac{\lambda}{2}\|I - M(\alpha)\|^2$$

- $M(\alpha)$: model for the high-res images

S. Kolouri et al., *CVPR*, 2015

Super-resolution

- $J = \phi(I) + n$
- $\phi(\cdot)$: downsampling function (known)
- n: noise

$$I^* = \arg\min_I \frac{1}{2}\|J - \phi(I)\|^2 + \frac{\lambda}{2}\|I - M(\alpha)\|^2$$

- $M(\alpha)$: model for the high-res images

S. Kolouri et al., *CVPR*, 2015

Super-resolution

- $J = \phi(I) + n$
- $\phi(\cdot)$: downsampling function (known)
- n: noise

$$I^* = \arg\min_I \frac{1}{2}\|J - \phi(I)\|^2 + \frac{\lambda}{2}\|I - M(\alpha)\|^2$$

- $M(\alpha)$: model for the high-res images

S. Kolouri et al., *CVPR*, 2015

Super-resolution

- $J = \phi(I) + n$
- $\phi(\cdot)$: downsampling function (known)
- n: noise

$$I^* = \arg\min_I \frac{1}{2}\|J - \phi(I)\|^2 + \frac{\lambda}{2}\|I - M(\alpha)\|^2$$

- $M(\alpha)$: model for the high-res images

S. Kolouri et al., *CVPR*, 2015

Super-resolution

- $J = \phi(I) + n$
- $\phi(\cdot)$: downsampling function (known)
- n: noise

$$I^* = \arg\min_I \frac{1}{2}\|J - \phi(I)\|^2 + \frac{\lambda}{2}\|I - M(\alpha)\|^2$$

- $M(\alpha)$: model for the high-res images

S. Kolouri et al., *CVPR*, 2015

Super-resolution

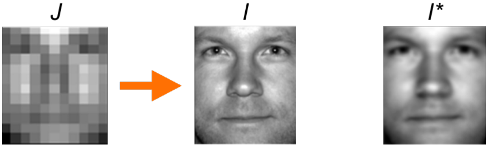
- $J = \phi(I) + n$
- $\phi(\cdot)$: downsampling function (known)
- n: noise

$$I^* = \arg\min_{I} \frac{1}{2}\|J - \phi(I)\|^2 + \frac{\lambda}{2}\|I - M(\alpha)\|^2$$

- $M(\alpha)$: model for the high-res images

S. Kolouri et al., *CVPR*, 2015

Super-resolution



- $J = \phi(I) + n$
- $\phi(\cdot)$: downsampling function (known)
- n: noise

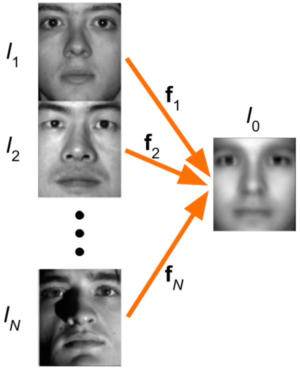
$$I^* = \arg\min_{I} \frac{1}{2}\|J - \phi(I)\|^2 + \frac{\lambda}{2}\|I - M(\alpha)\|^2$$

- $M(\alpha)$: model for the high-res images

S. Kolouri et al., *CVPR*, 2015

Super-resolution

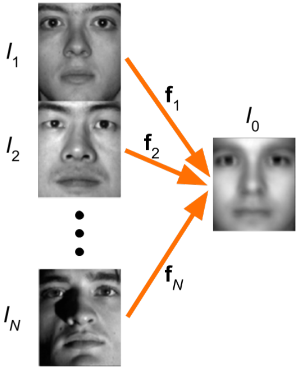
- Learning model $M(\alpha)$



S. Kolouri et al., *CVPR*, 2015

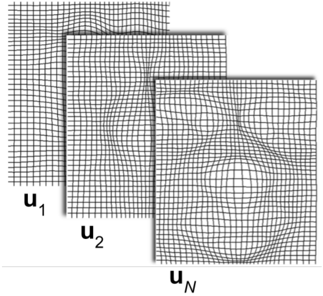
Super-resolution

- Learning model $M(\alpha)$



S. Kolouri et al., *CVPR*, 2015

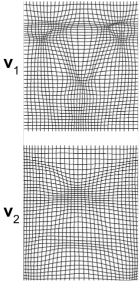
$$\boldsymbol{f}_i = \boldsymbol{x} + \boldsymbol{u}_i$$



$$\boldsymbol{u}_i = \sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k$$

Super-resolution

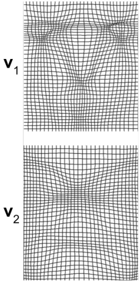
- $\boldsymbol{v}_k$ are PCA components.

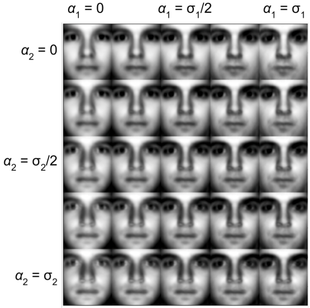


S. Kolouri et al., *CVPR*, 2015

Super-resolution

- $\boldsymbol{v}_k$ are PCA components.



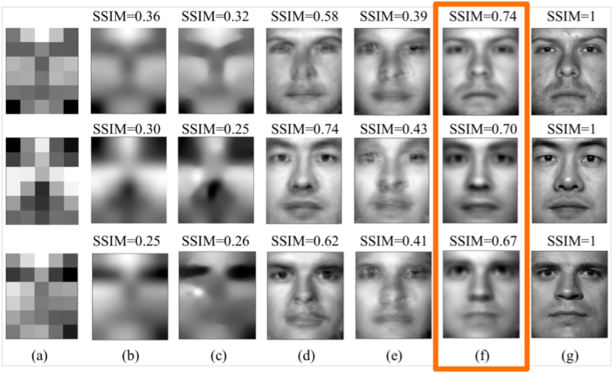


$$\boldsymbol{f} = \boldsymbol{x} + \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2$$

$$\det(D_{\boldsymbol{f}}) I_0 \circ \boldsymbol{f}$$

S. Kolouri et al., *CVPR*, 2015

Super-resolution



$$I^* = \arg\min_I \frac{1}{2}\|J - \phi(I)\|^2 + \frac{\lambda}{2}\|I - M(\alpha)\|^2$$

S. Kolouri et al., *CVPR*, 2015

Outline

Optimal Transport

Visualizing transformations

Transport-based morphometry

Classification

Inverse problems

Conclusions

Final thoughts

Optimal transport can provide:

- Simultaneous statistical analysis and visualization
  - Principal component analysis
  - Canonical correlation analysis
- Better representation of data
  - Higher classification accuracy
  - Better reconstructions

Final thoughts

Optimal transport can provide:

- Simultaneous statistical analysis and visualization
  - Principal component analysis
  - Canonical correlation analysis
- Better representation of data
  - Higher classification accuracy
  - Better reconstructions

Final thoughts

Optimal transport can provide:

- Simultaneous statistical analysis and visualization
  - Principal component analysis
  - Canonical correlation analysis
- Better representation of data
  - Higher classification accuracy
  - Better reconstructions

Final thoughts

Optimal transport can provide:

- Simultaneous statistical analysis and visualization
  - Principal component analysis
  - Canonical correlation analysis
- Better representation of data
  - Higher classification accuracy
  - Better reconstructions

Final thoughts

Optimal transport can provide:

- Simultaneous statistical analysis and visualization
  - Principal component analysis
  - Canonical correlation analysis
- Better representation of data
  - Higher classification accuracy
  - Better reconstructions

Final thoughts

Optimal transport can provide:

- Simultaneous statistical analysis and visualization
  - Principal component analysis
  - Canonical correlation analysis
- Better representation of data
  - Higher classification accuracy
  - Better reconstructions

Final thoughts

Optimal transport can provide:

- Simultaneous statistical analysis and visualization
    - Principal component analysis
    - Canonical correlation analysis
- Better representation of data
    - Higher classification accuracy
    - Better reconstructions

Thanks for listening

Code, slides, and references

- https://github.com/skolouri/BAMC2019