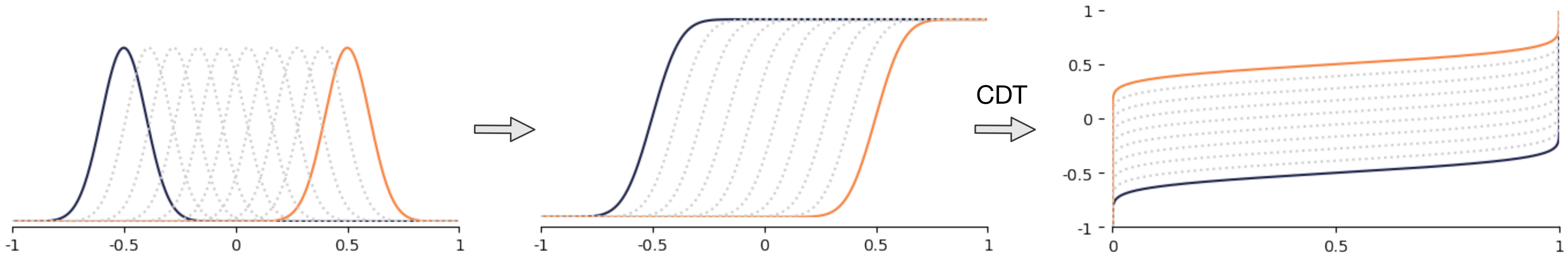
Transport transforms for signal analysis and machine learning

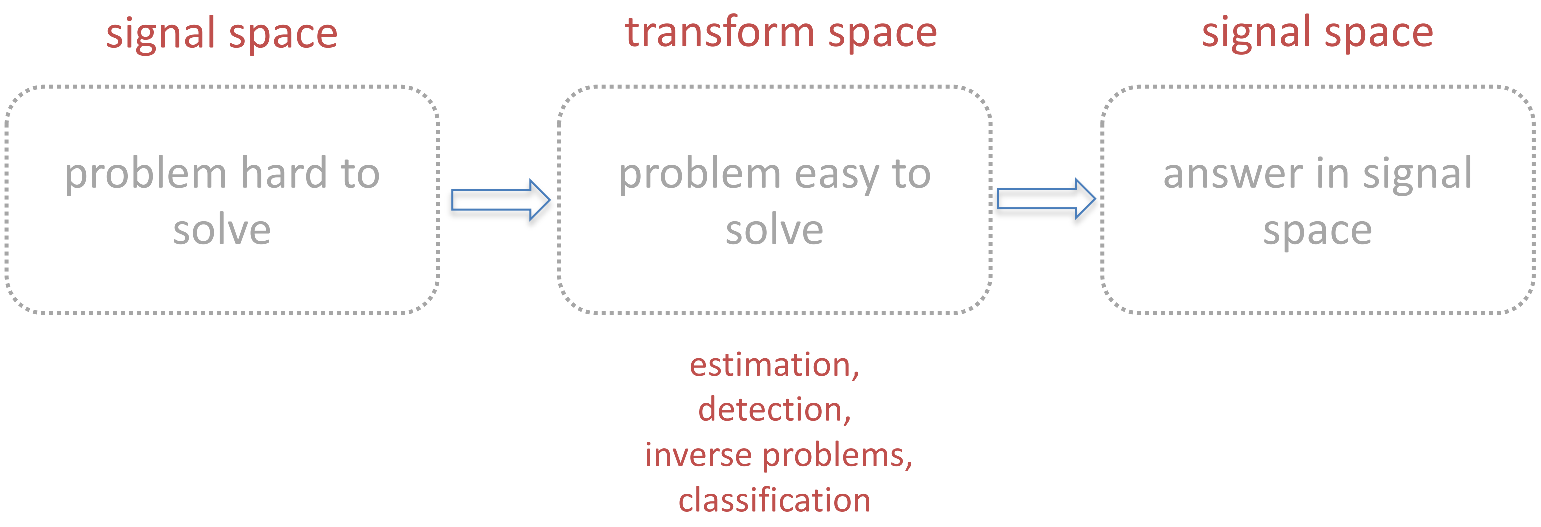
Gustavo K. Rohde

Imaging and Data Science Laboratory

imagedatascience.com/transport
github.com/rohdelab/PyTransKit



Transport Transforms



[Wang, Slepcev, Basu, Ozolek, Rohde, IJCV 2013]
[Kolouri, Tosun, Ozolek, Rohde, Pattern Recognition 2016]
[Kolouri, Park, Rohde, IEEE TIP, 2016]
[Park, Kolouri, Kundu, Rohde, ACHA 18]
[Rubaiyat et al, IEEE TSP, 20]
[Aldroubi, Li, Rohde, SaSiDa, 21]
[Shifat-E-Rabbi et al, Rohde, JMIV 21]
[Aldroubi et al, AIMS Foundations of Data Science, 22]

Fourier transform:

forward

$$\hat{s}(\omega) = \int s(t) e^{-jwt} dt$$

inverse

$$s(t) = \frac{1}{2\pi} \int \hat{s}(\omega) e^{k\omega t} d\omega$$

$$\int s(u)h(t-u)du = f(t)$$

signal domain

difficult to solve

⇒

$$\hat{s}(\omega)\hat{h}(\omega) = \hat{f}(\omega)$$

$$\hat{s}(\omega) = \hat{f}(\omega)/\hat{h}(\omega)$$

transform domain

easier to solve

⇒

$$s$$

signal domain

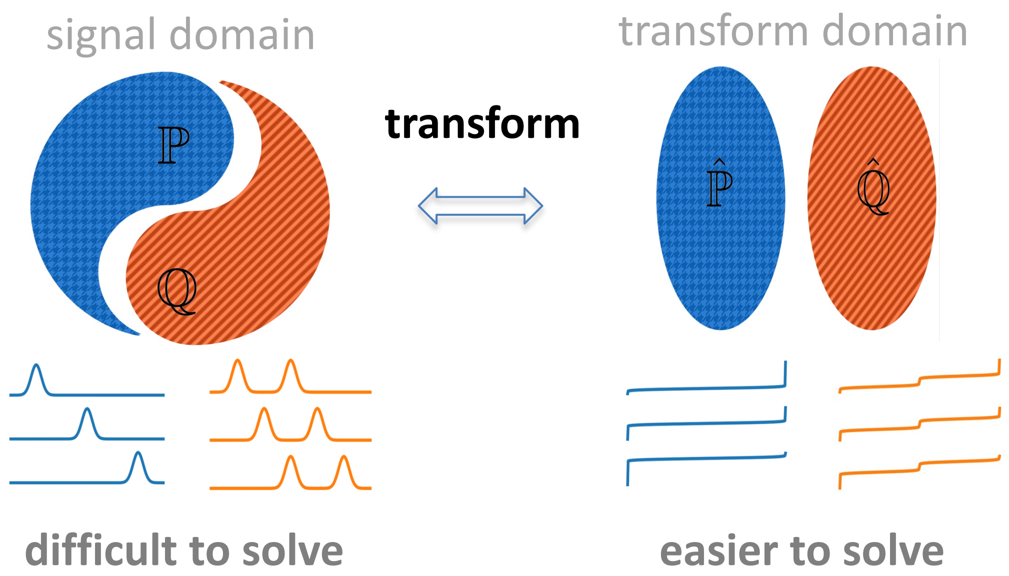
New signal transform

forward

$$\hat{s}(x) := S^{-1}(x)$$

inverse

$$s(t) = \frac{d}{dt}(\hat{s})^{-1}$$



Outline

- 1D case
  - CDT, SCDT
  - Connections with optimal transport
- N-D case
  - R-CDT, LOT
  - Connections with optimal transport
  - Related work

1-D Transport transform

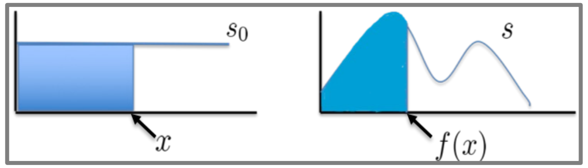
cumulative distribution transform (CDT)

[Park, Kolouri, Kundu, Rohde, ACHA 2018]

Let $s(x), s_0(x) > 0$ and $\int s(x)dx = \int s_0(x)dx = 1$ (PDFs)

Consider a map $f(x)$, such that

$$\underbrace{\int_{-\infty}^{x} s_0(u)du}_{} = \underbrace{\int_{-\infty}^{f(x)} s(u)du}_{}$$

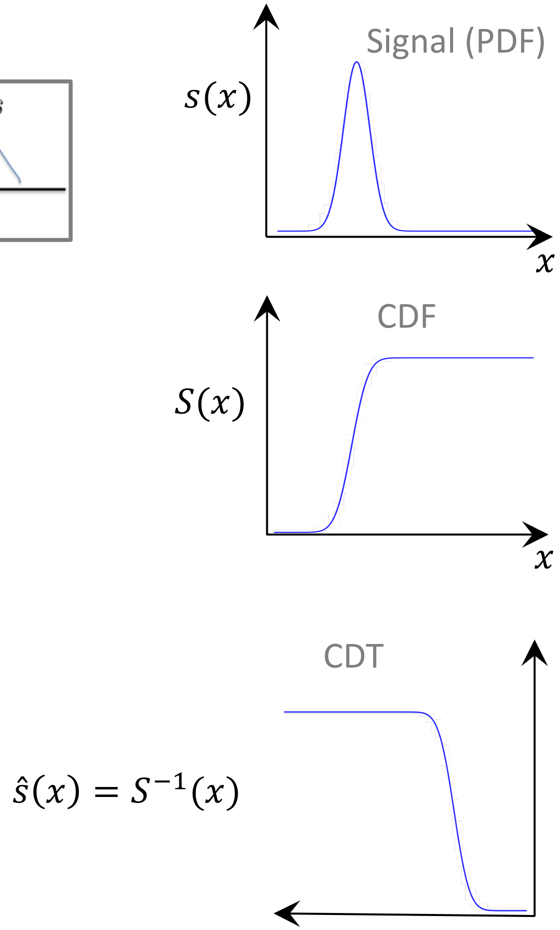


antiderivative

$$S_0(x) = S(f(x)) \Longrightarrow f(x) = S^{-1}(S_0(x))$$

transform of s

Eg.: Let $s_0 = \chi_{[0,1]}$. Then, $S_0(x) = x$ and $f(x) = S^{-1}(x)$.

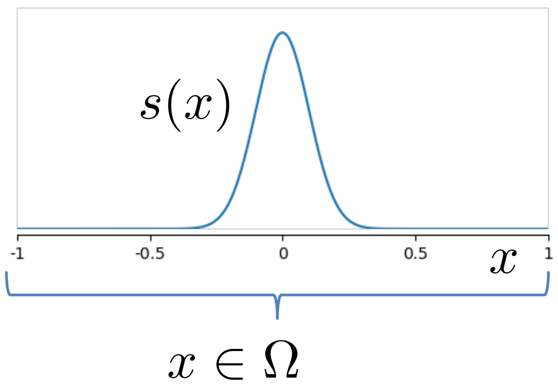


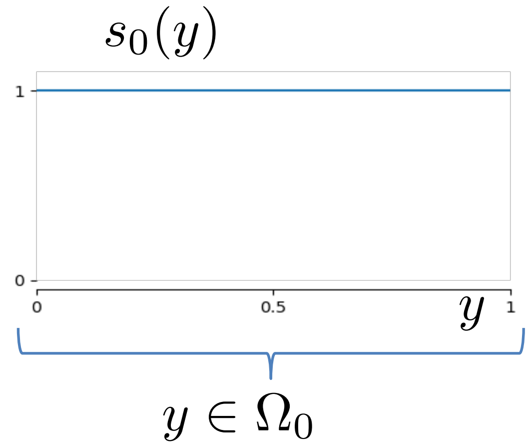
transform equations

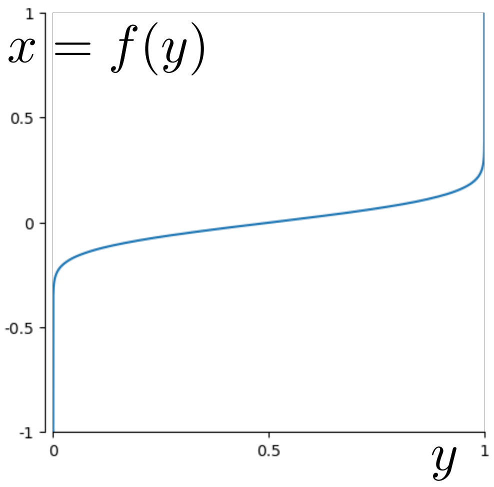
$$\hat{s}(x) := S^{-1}(x) \quad \text{forward}$$

$$s(x) = \left(\hat{s}^{-1}(x)\right)' \quad \text{inverse}$$

CDT: invertible (nonlinear) signal transform







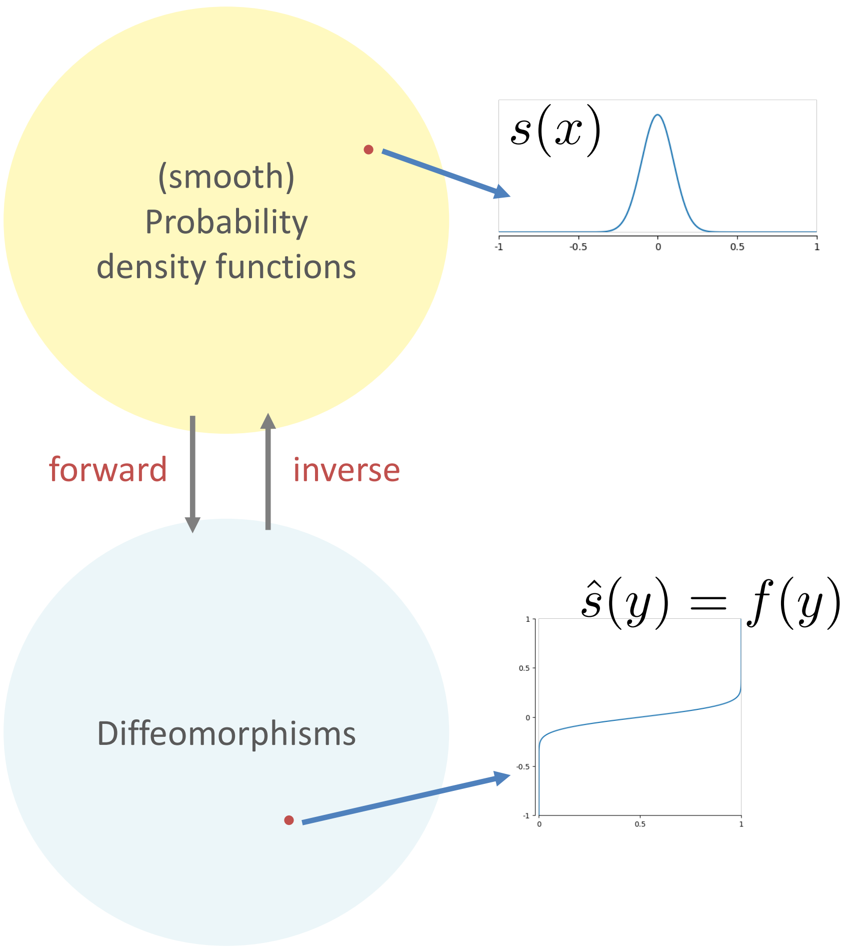
$$f : \Omega_0 \rightarrow \Omega$$

$$f'(y)s(f(y)) = s_0(y)$$

$$S(f(y)) = S_0(y)$$

$$y \in \Omega_0$$

$$\hat{s}(y) = f(y)$$



CDT properties: composition

Composition: $s(x) \;\; x \in \Omega \qquad \hat{s}(y) \;\; y \in \Omega_0$ what is the CDT of $s_g(x) = g'(x)s(g(x)), x \in \Omega_g$?

$$\int_{-\infty}^{\widehat{s}(y)} s(u)du = \int_{-\infty}^{\widehat{s}_g(y)} s_g(u)du = \int_{-\infty}^{y} s_0(u)du$$

$$\int_{-\infty}^{\widehat{s}(y)} s(u)du = \int_{-\infty}^{\widehat{s}_g(y)} g'(u)s(g(u))du$$

change of variables $g(u) = v \quad g'(u)du = dv$

$$\int_{-\infty}^{\widehat{s}(y)} s(u)du = \int_{-\infty}^{g(\widehat{s}_g(y))} s(v)dv \Rightarrow g\left(\widehat{s}_g(x)\right) = \widehat{s}(x) \implies \widehat{s}_g(x) = g^{-1}\left(\widehat{s}(x)\right) \text{ or, } \widehat{s}_g = g^{-1} \circ \widehat{s}$$

Conclusion: the CDT of $s_g(x) = g'(x)s(g(x)), x \in \Omega_g$ is given by: $\widehat{s}_g(y) = g^{-1}(\widehat{s}(y)), y \in \Omega_0$

CDT Properties

Translation

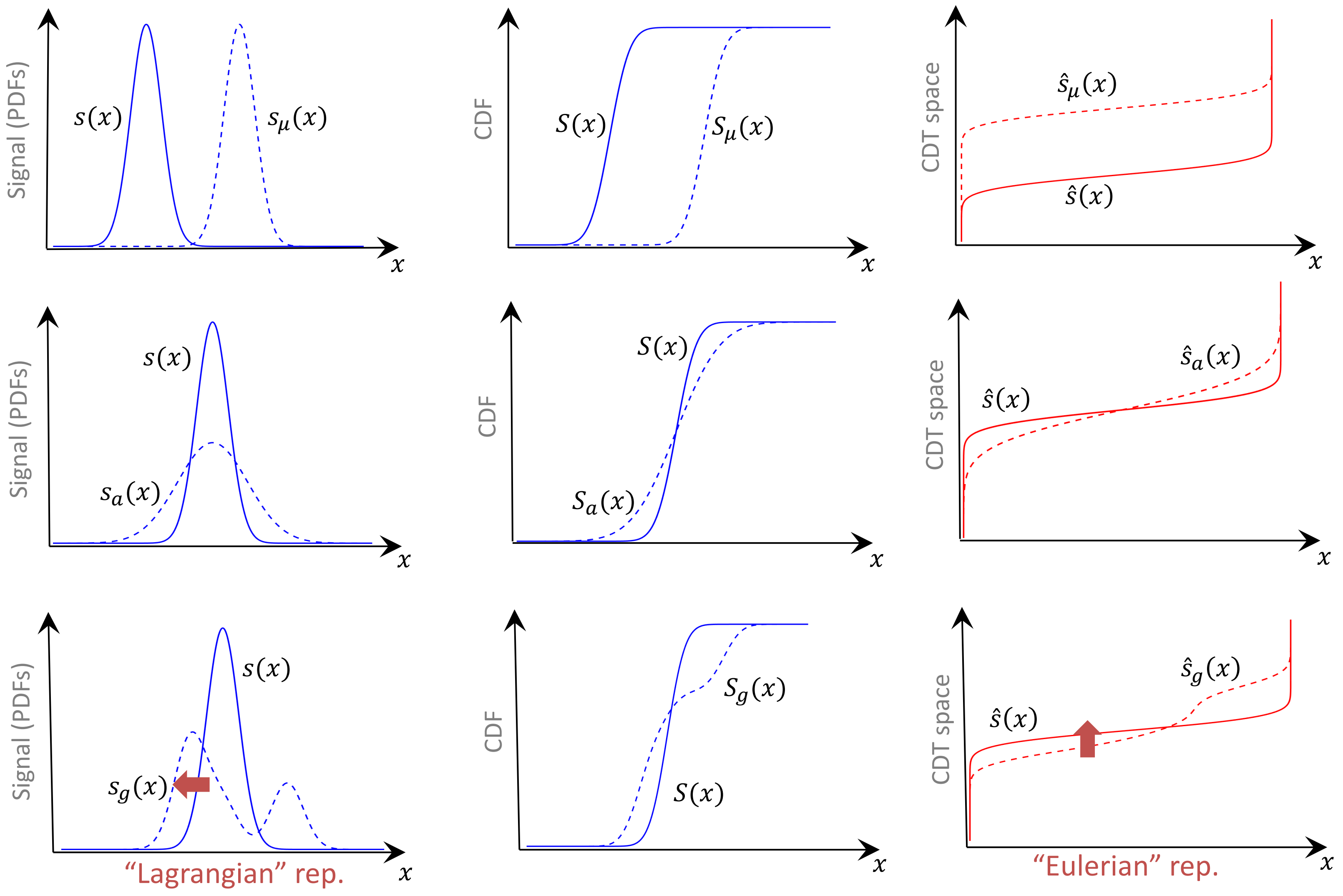
$$s_\mu(x) = s(x - \mu)$$
$$\hat{s}_\mu(x) = \hat{s}(x) + \mu$$

Scaling

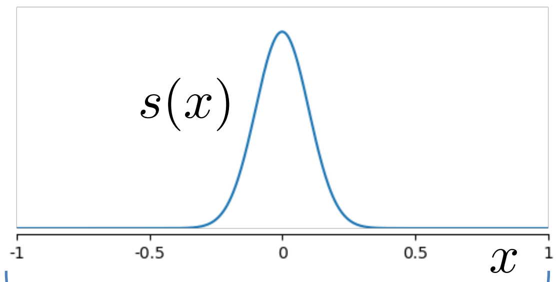
$$s_a(x) = as(ax)$$
$$\hat{s}_a(x) = \frac{\hat{s}(x)}{a}$$

Composition

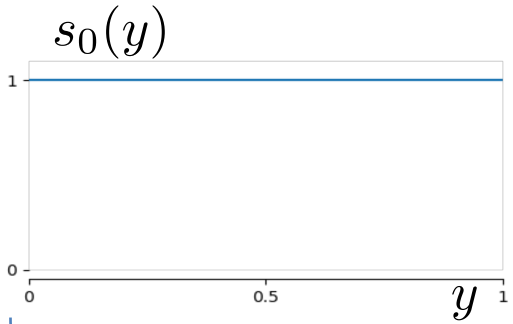
$$s_g(x) = g'(x)s(g(x))$$
$$\hat{s}_g(x) = g^{-1}(\hat{s}(x))$$



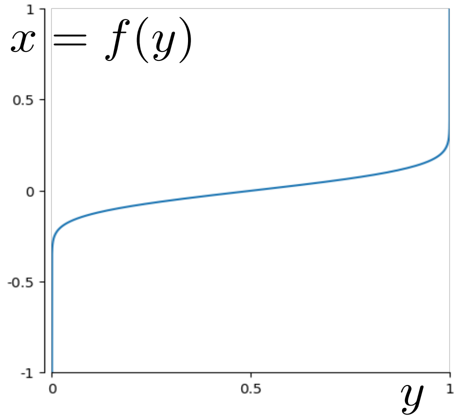
1D optimal transport



$x \in \Omega$



$y \in \Omega_0$



Integrate: $S(f(y)) = S_0(y) \quad y \in \Omega_0$

Differentiate: $f'(y)s(f(y)) = s_0(y)$

Mapping function $f : \Omega_0 \to \Omega$ is unique.

Matching (optimal transport) cost:

$$\text{Monge}(f) = \int_{\Omega_0} |f(y) - y|^2 s_0(y)dy$$

$\hat{s}(y)$

Squared Wasserstein distance $W_2^2(s, s_0)$

Claim: $W_2^2(s_1, s_2) = \|(\hat{s_1} - \hat{s_2})\sqrt{s_0}\|_2^2$

Proof: Let $\hat{s}_1$ transform of s_1 that is

$\hat{s}_2$ transform of s_2 $\hat{s}_1'(y)s_1(\hat{s}_1(y)) = \hat{s}_2'(y)s_2(\hat{s}_2(y)) = s_0(y)$ $y \in \Omega_0$ $\hat{s}_1 : \Omega_0 \to \Omega_1$

Let $s_1(x) = h'(x)s_2(h(x))$ that is $\hat{s}_1 = h^{-1} \circ \hat{s}_2$ → $\hat{s}_2 = h \circ \hat{s}_1$

$$W_2^2(s_1, s_2) = \int_{\Omega_1} (x - h(x))^2 s_1(x)dx$$

change of variables: $x = \hat{s}_1(y)$

$$W_2^2(s_1, s_2) = \int_{\hat{s}_1^{-1}(\Omega_1)} (\hat{s}_1(y) - h(\hat{s}_1(y)))^2 s_1(\hat{s}_1(y))\hat{s}_1'(y)dy$$

use: $\hat{s}_1 : \Omega_0 \to \Omega_1$

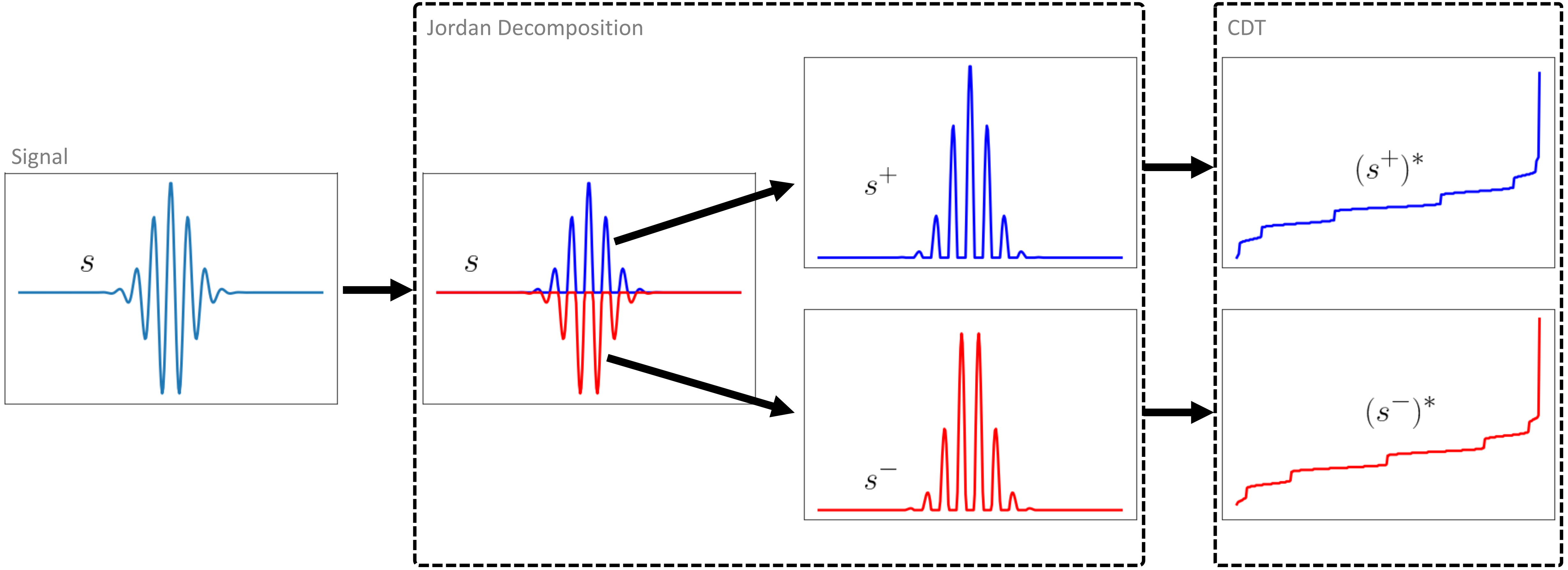
$\hat{s}_1'(y)s_1(\hat{s}_1(y)) = s_0(y)$

$\hat{s}_2 = h \circ \hat{s}_1$

$$W_2^2(s_1, s_2) = \int_{\Omega_0} (\hat{s}_1(y) - \hat{s}_2(y))^2 s_0(y)dy = \|(\hat{s}_1 - \hat{s}_2)\sqrt{s_0}\|_2^2$$

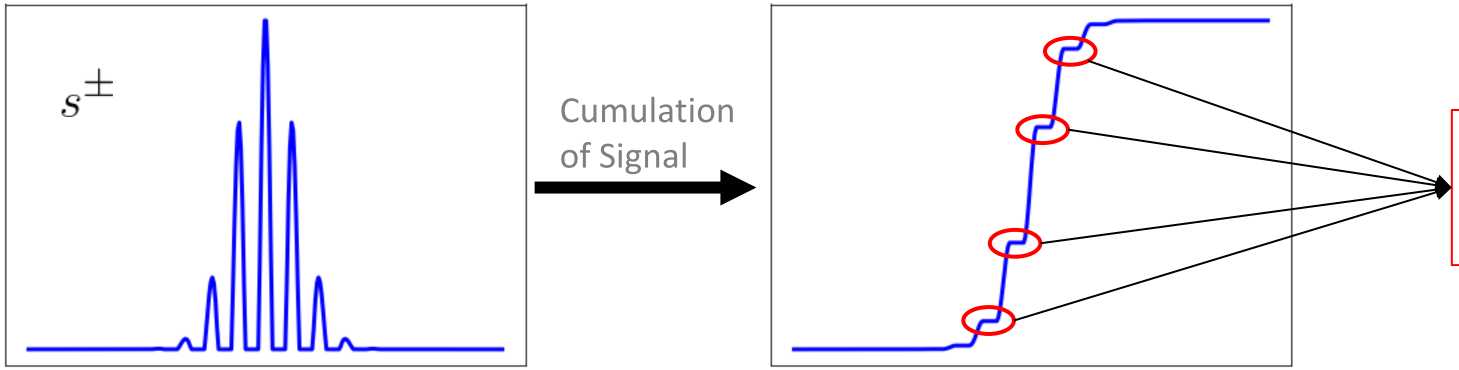
Signed Cumulative Distribution Transform (SCDT)

$$\hat{s} = \left((s^+)^*, \|s^+\|_{L_1}, (s^-)^*, \|s^-\|_{L_1}\right)$$



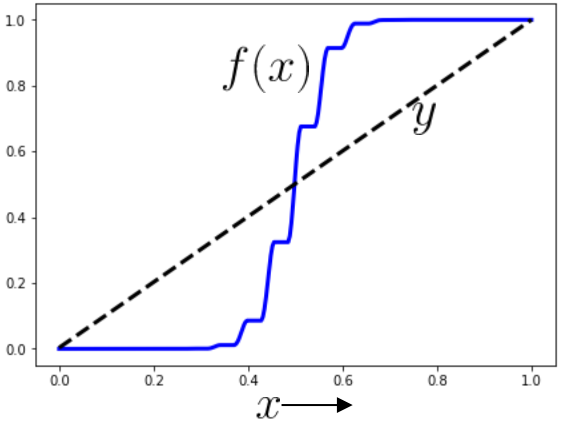
Aldroubi, Diaz Martin, Medri, Rohde, Thareja, The Signed Cumulative Distribution Transform for 1D signal analysis and classification, AIMS Foundations of Data Science, 2022

Generalized Inverse to Calculate SCDT

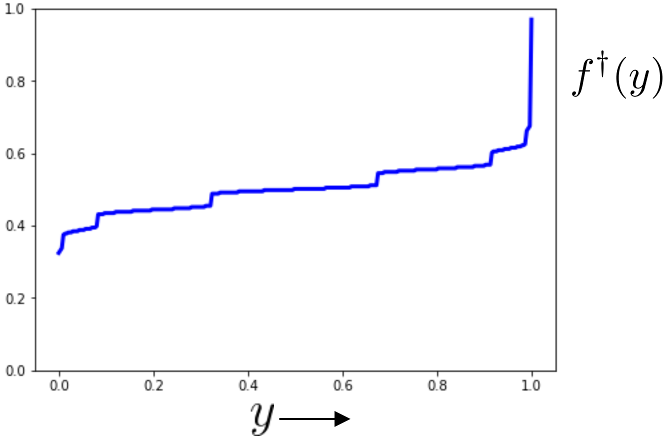


Lack of monotonicity because of signal being zero due to Jordan Decomposition. Therefore, function inverse is not well defined.

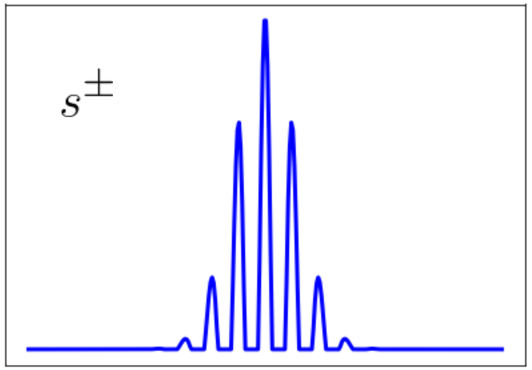
Generalized Inverse:



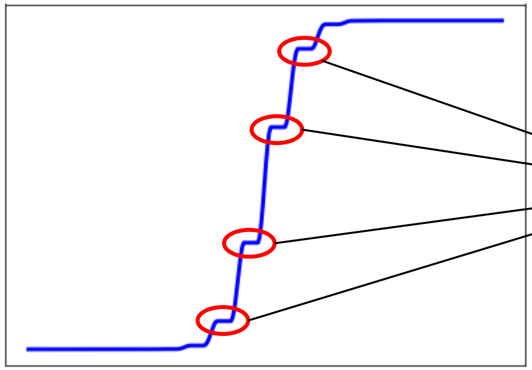
$$f^{\dagger}(y) = \inf\{x \in \mathbb{R} : f(x) \geq y\}, y \in \mathbb{R}$$



Generalized Inverse to Calculate SCDT

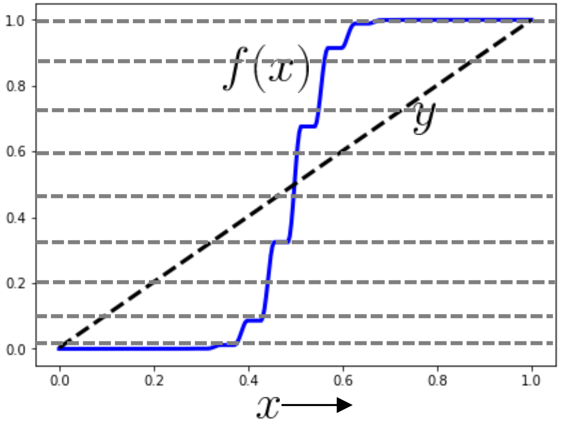


Cumulation of Signal

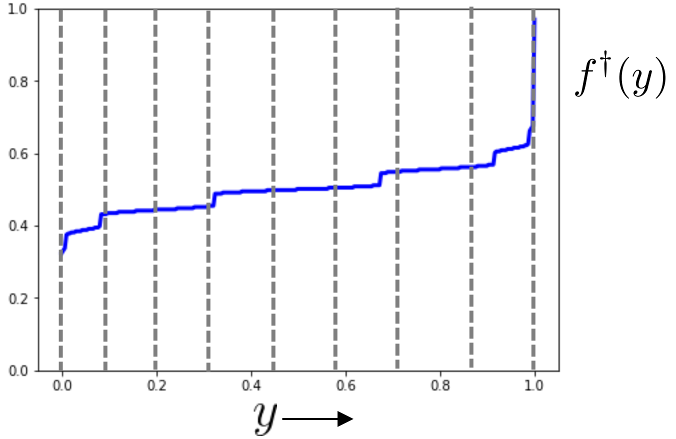


Lack of monotonicity because of signal being zero due to Jordan Decomposition. Therefore, function inverse is not well defined.

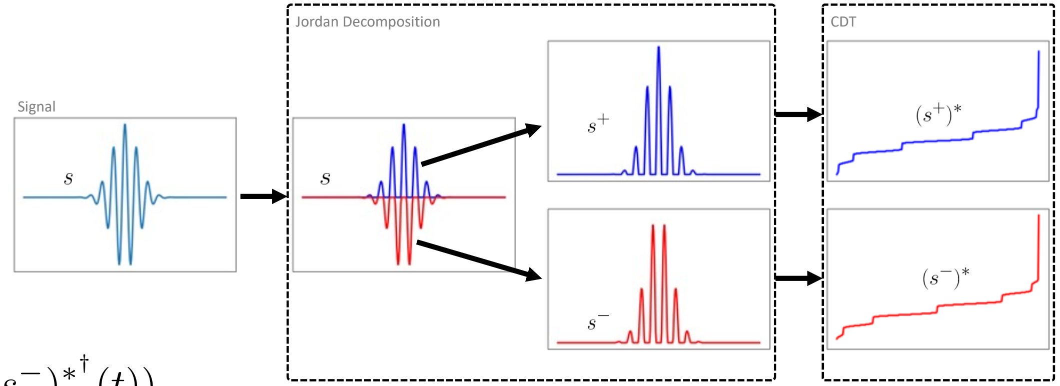
Generalized Inverse:



$$f^{\dagger}(y) = \inf\{x \in \mathbb{R} : f(x) \geq y\}, y \in \mathbb{R}$$



Signed Cumulative Distribution Transform (SCDT)



Forward:

$$\hat{s} = \left((s^+)^*, \|s^+\|_{L_1}, (s^-)^*, \|s^-\|_{L_1}\right)$$

Inverse:

$$s(t) = \|s^+\|_{L_1} \left((s^+)^{*^\dagger}(t)\right)' s_0((s^+)^{*^\dagger}(t)) - \|s^-\|_{L_1} \left((s^-)^{*^\dagger}(t)\right)' s_0((s^-)^{*^\dagger}(t))$$

Composition:

$$s_g = g' s \circ g \implies \widehat{s}_g = \left(g^{-1} \circ (s^+)^*, \|s^+\|_{L_1}, g^{-1} \circ (s^-)^*, \|s^-\|_{L_1}\right)$$

Convexity:

Let $\mathbb{S} = \{s_j | s_j = g_j' \varphi \circ g_j, \forall g_j \in \mathcal{G}\}$ $\widehat{\mathbb{S}} = \{\widehat{s}_j : s_j \in \mathbb{S}\}$ convex IFF $\mathcal{G}^{-1} = \{g_j^{-1} : g_j \in \mathcal{G}\}$ convex.

for any reference s_0

New metric/distance:

$$D_S^2(s_1, s_2) := \|\hat{s}_2 - \hat{s}_2\|_{(L^2 \times \mathbb{R}^2)} = W_2^2(s_1^+, s_2^+) + W_2^2(s_1^-, s_2^-) + \lambda(\|s_1^+|_{L_1} - \|s_2^+|_{L_1})^2 + \lambda(\|s_1^-|_{L_1} - \|s_2^-|_{L_1})^2$$

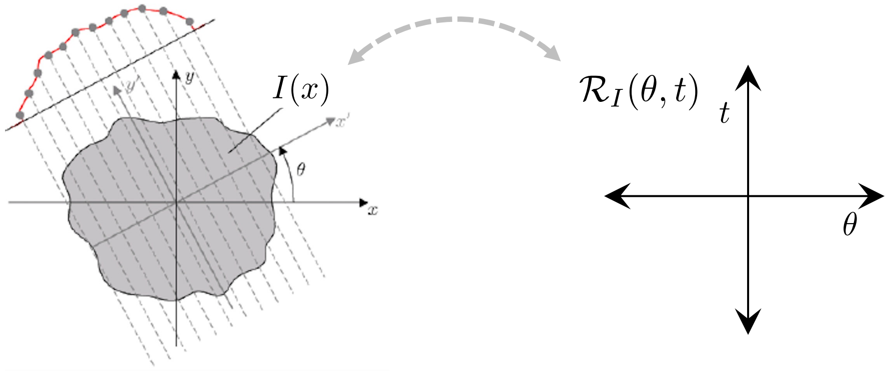
Aldroubi, Diaz Martin, Medri, Rohde, Thareja, The Signed Cumulative Distribution Transform for 1D signal analysis and classification, AIMS Foundations of Data Science, 2022

Extension to 2d, 3d, …: Radon Cumulative distribution transform (R-CDT)

Radon Transform

Forward

$$\mathcal{R}_I(\theta, t) = \int I(x)\delta(t - \omega \cdot x)dx$$
$$\omega = [\sin\theta, \cos\theta]^T$$



Extension to 2d, 3d, …: Radon Cumulative distribution transform (R-CDT)

Radon Transform

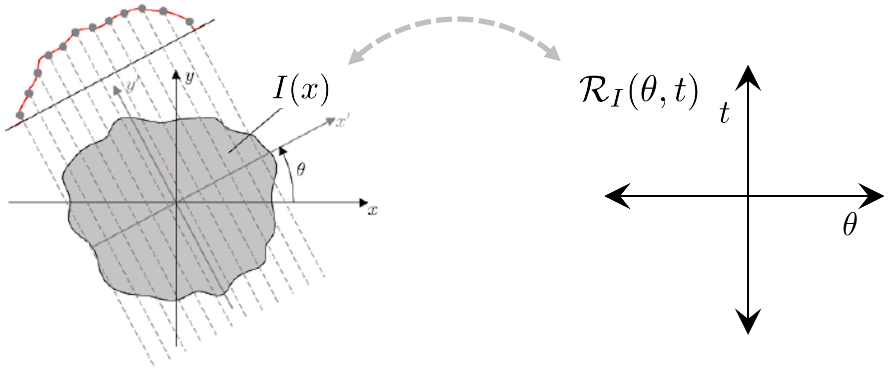
Forward

$$\mathcal{R}_I(\theta, t) = \int I(x)\delta(t - \omega \cdot x)dx$$

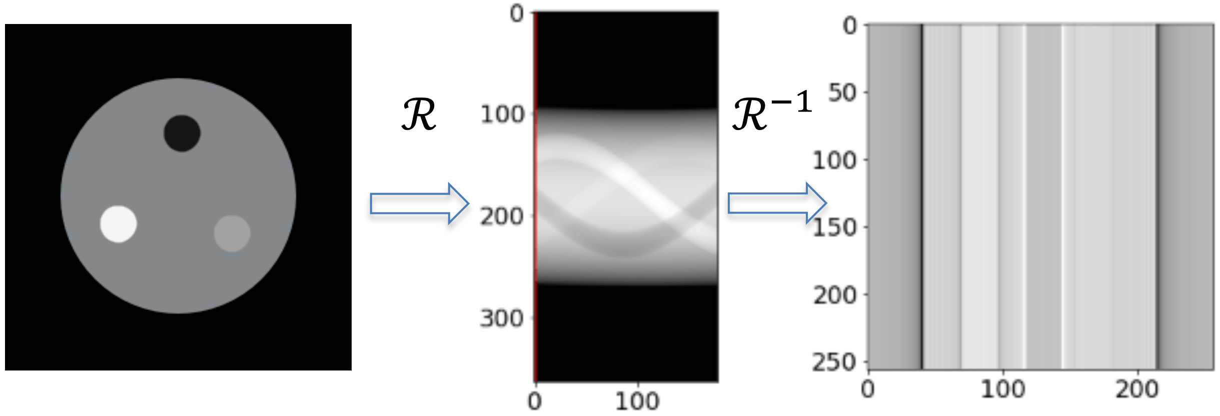
$$\omega = [\sin\theta, \cos\theta]^T$$

Inverse

$$\mathcal{R}_I^{-1}(x) = \int_{\mathbb{S}^{d-1}} (\mathcal{R}I(\cdot,\theta) * \eta(\cdot)) \circ (x \cdot \theta) d\theta$$



Filtered back-projection algorithm



Extension to 2d, 3d: Radon Cumulative distribution transform (R-CDT)

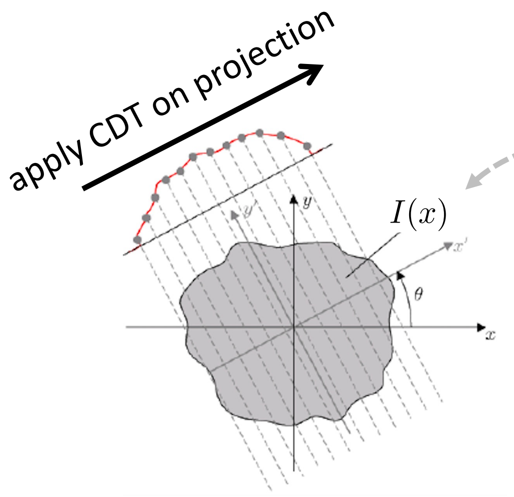
Radon Transform

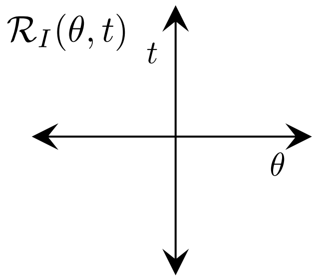
Forward

$$\mathcal{R}_I(\theta, t) = \int I(x)\delta(t - \omega \cdot x)dx$$
$$\omega = [\sin\theta, \cos\theta]^T$$

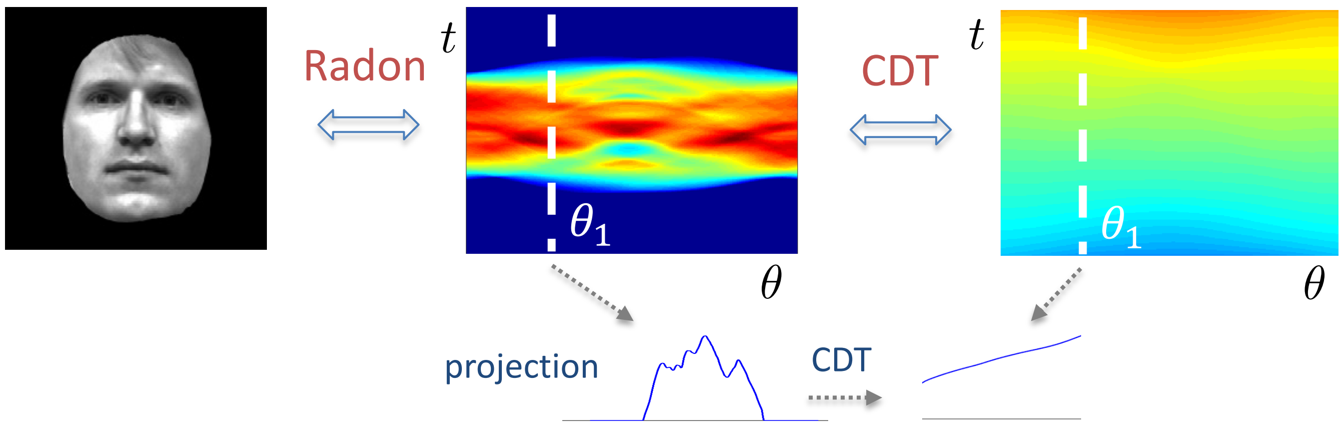
Inverse

$$\mathcal{R}_I^{-1}(x) = \int_{\mathbb{S}^{d-1}} (\mathcal{R}I(\cdot, \theta) * \eta(\cdot)) \circ (x \cdot \theta) d\theta$$





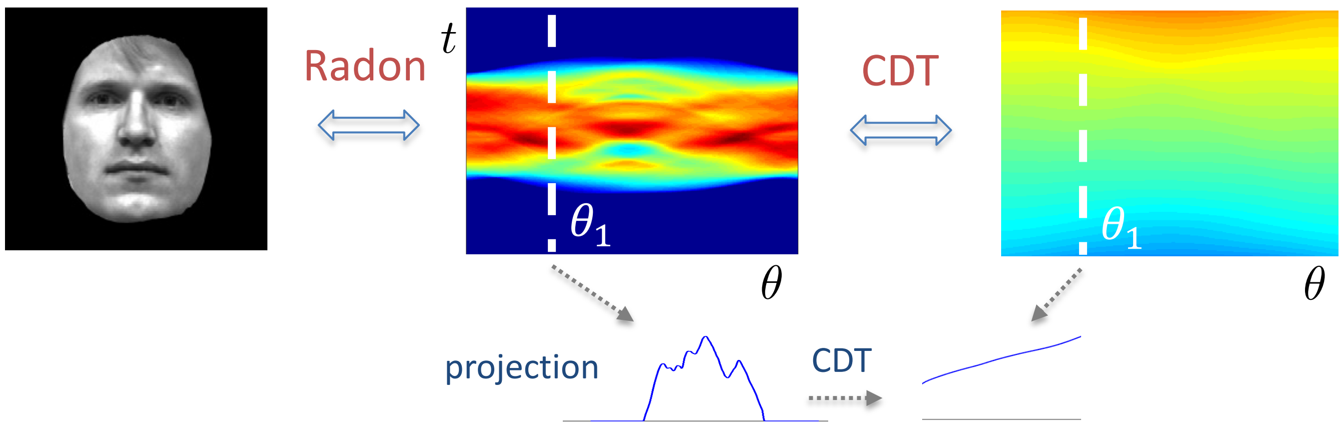
Radon Cumulative Distribution Transform



[Kolouri, Park, Rohde, IEEE TIP 2016]

Radon Cumulative distribution transform (R-CDT)

$\widetilde{s} = \mathcal{R}(s)$ $\quad$ $\widehat{s}(t,\theta)$



$$\int_{-\infty}^{\widehat{s}(t,\theta)} \widetilde{s}(u,\theta)du = \int_{-\infty}^{t} \widetilde{s_0}(u,\theta)du, \quad \forall \theta \in [0,\pi]$$

Inverse:

$$s(\mathbf{x}) = \mathcal{R}^{-1}\left(\frac{\partial \widehat{s}^{-1}(t,\theta)}{\partial t} \widetilde{s_0}\left(\widehat{s}^{-1}(t,\theta),\theta\right)\right)$$

Composition:

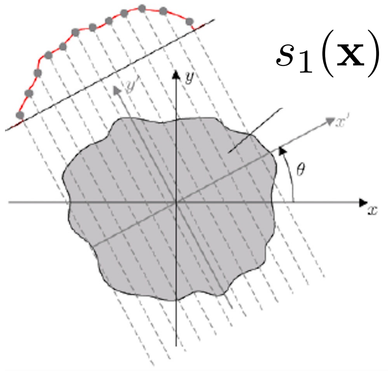
Let $s_{g^\theta}(\mathbf{x}) = \mathcal{R}^{-1}\left(\frac{\partial \widehat{s}^{-1}(g^\theta(t),\theta)}{\partial t} \widetilde{s_0}\left(\widehat{s}^{-1}(g^\theta(t),\theta),\theta\right)\right)$

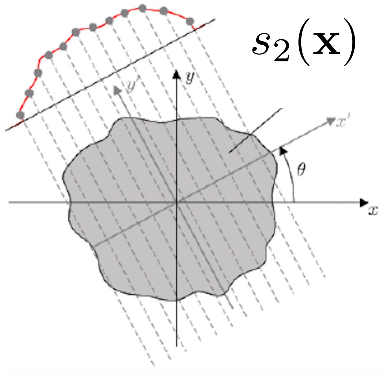
Then $\widehat{s}_{g^\theta}(t,\theta) = (g^\theta)^{-1}(\widehat{s}(t,\theta))$

[Kolouri, Park, Rohde, IEEE TIP 2016]

Sliced Wasserstein Distance

$$SW_2^2(s_1, s_2) = \int_{\Omega_{\widetilde{s}_0}} \int_0^{\pi} \left(\widehat{s}_1(t,\theta) - \widehat{s}_2(t,\theta)\right)^2 \widetilde{s}_0(t,\theta) d\theta dt$$





Sliced Wasserstein Embedding:

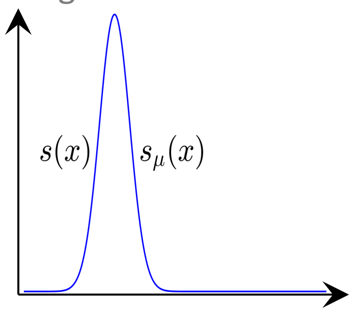
$$SW_2^2(s_1, s_2) = \left\| \left(\widehat{s}_1 - \widehat{s}_2\right) \sqrt{\widetilde{s}_0} \right\|^2_{L^2(\Omega_{\widetilde{s}_0})}$$

R-CDT

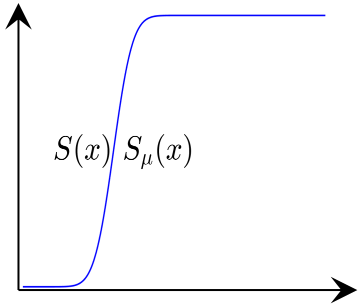
Data Geometry

Consider the data set generated by a signal and all of its translates

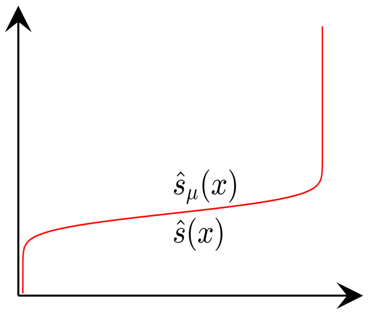
Signal



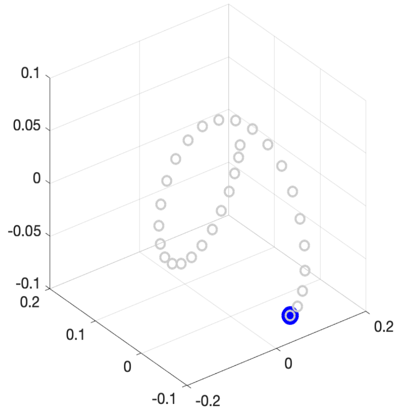
CDF



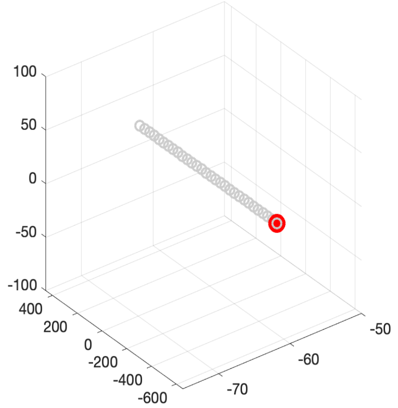
CDT



PCA (3 components) of dataset



nonlinear/nonconvex



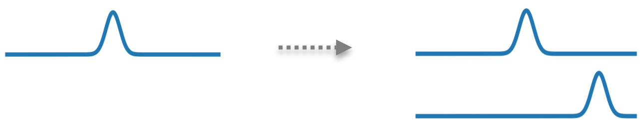
linear/convex

Signal classes: algebraic generative model

signal space

Class $\mathbb{P}$

template 1 p_0



$p_0(x-\mu)$

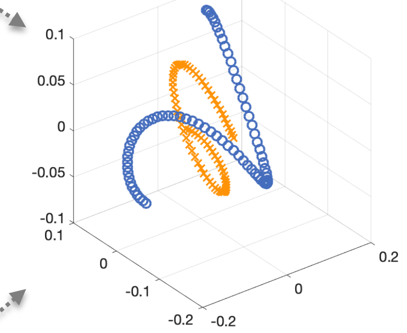
$$\mathbb{P} = \{p_g | p_g(x) = g'(x) p_0(g(x))\}$$

Class $\mathbb{Q}$

template 2 q_0

$q_0(x-\mu)$

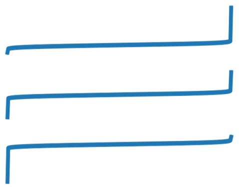
$$\mathbb{Q} = \{q_g \, | q_g(x) = g'(x) q_0(g(x))\}$$



low dimensional projection

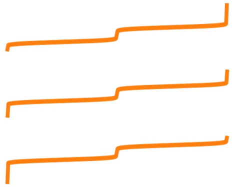
transform space

$\hat{\mathbb{P}}$

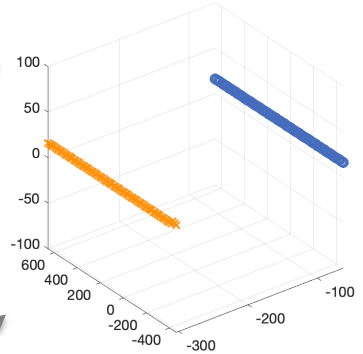


$\hat{p_g} = \hat{p} + \mu$

$\hat{\mathbb{Q}}$



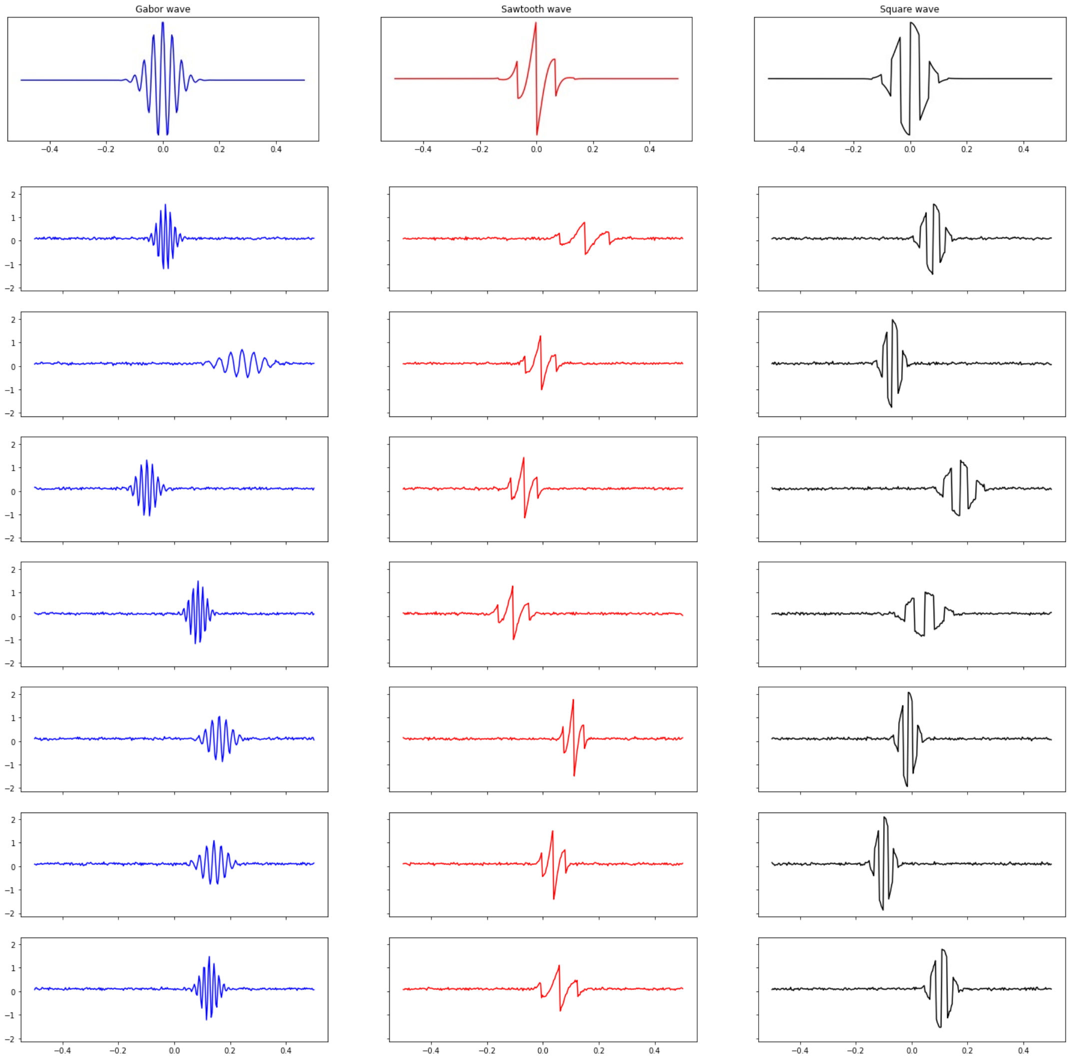
$\hat{q}_g = \hat{q} + \mu$

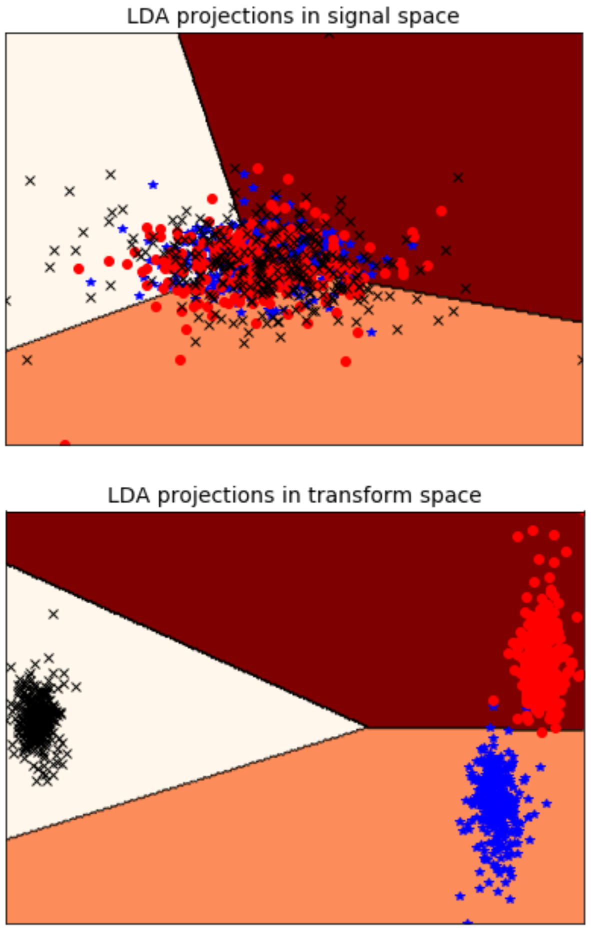


low dimensional projection

"confound" $g \in \mathcal{C}$ e.g.: translation $g(x) = x - \mu$

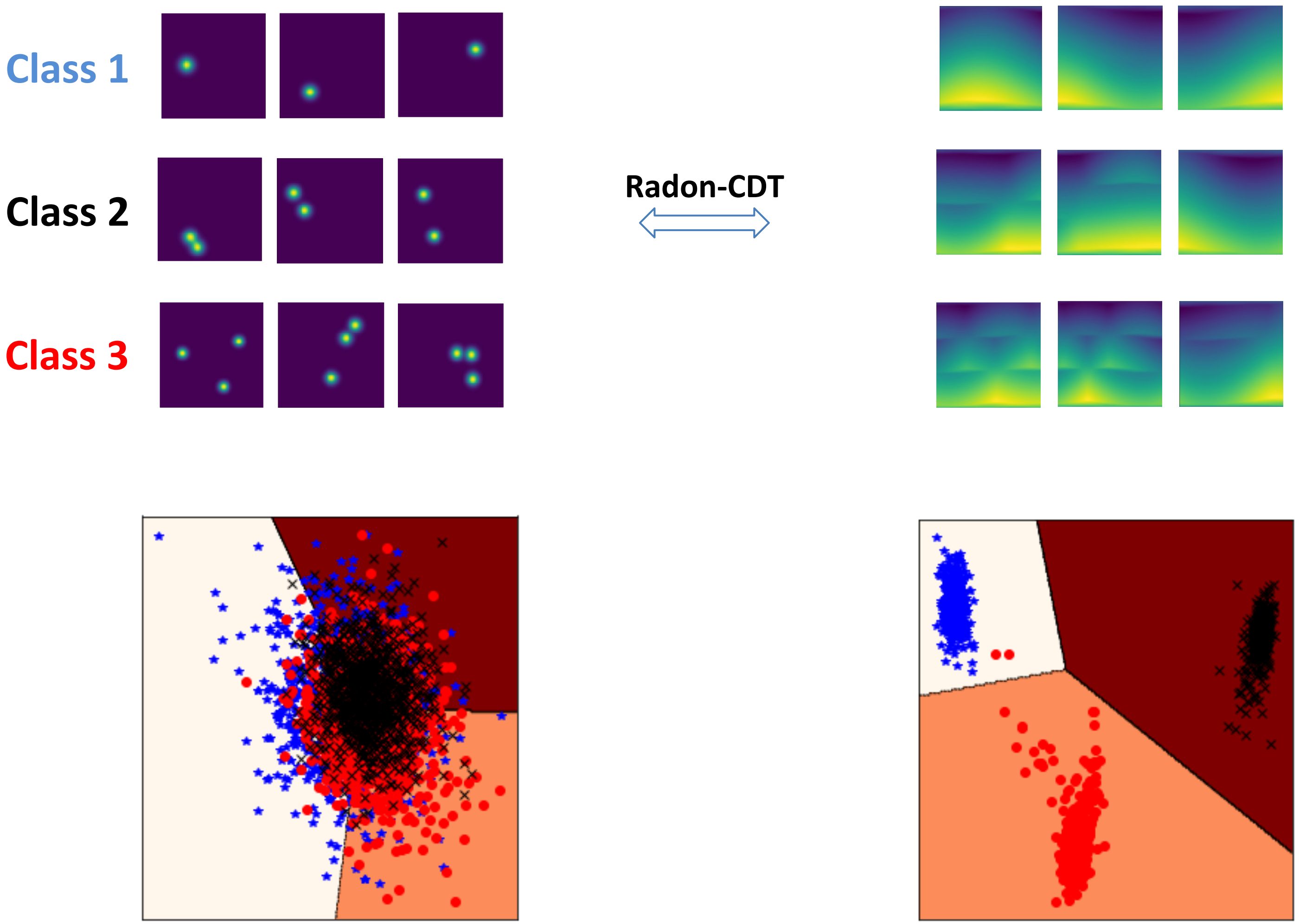
SCDT & linear separability





Aldroubi, Diaz Martin, Medri, Rohde, Thareja, The Signed Cumulative Distribution Transform for 1D signal analysis and classification, AIMS Foundations of Data Science, 2022

2D Example (R-CDT)



LDA subspace of original data

LDA subspace of transformed data

CDT: linear separation theorem

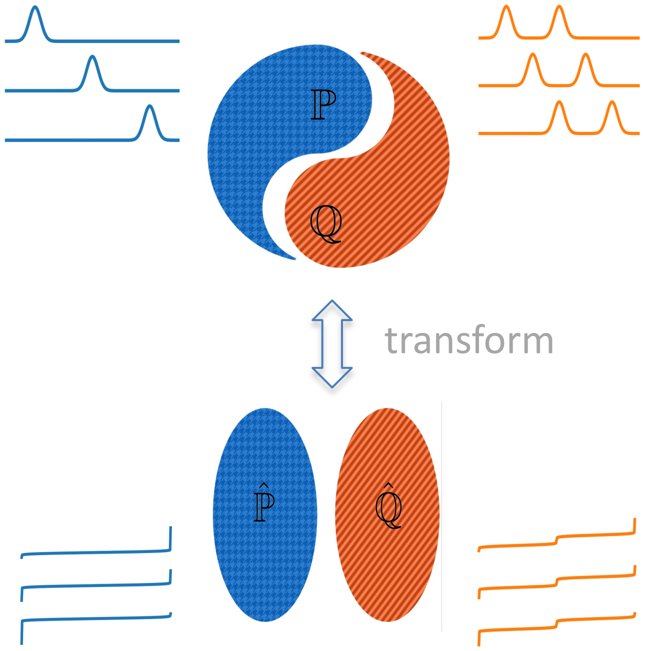
Consider signal classes:

$\mathbb{P} = \{p_g : g \in \mathcal{C}\}$ and $\mathbb{Q} = \{q_g : g \in \mathcal{C}\}$

where $p_g(x) = g'(x)p_0(g(x))$

$q_g(x) = g'(x)q_0(g(x))$

$\mathbb{P} \cap \mathbb{Q} = \emptyset$



Theorem:

$\hat{\mathbb{P}}, \hat{\mathbb{Q}}$ will be *convex*

iff $\mathcal{C}^{-1}$ is convex.

for any reference s_0

does not require knowledge of p_0 q_0

Corollary: $\hat{\mathbb{P}}, \hat{\mathbb{Q}}$ linearly separable

if $\mathcal{C}^{-1}$ is convex.

[Park, Kolouri, Kundu, Rohde, ACHA 18]
[Aldroubi, Li, Rohde, SaSiDa, 21]
[Shifat-E-Rabbi, …, Rohde, JMIV 21]
[Aldroubi et al, AIMS Foundations of Data Science, 22]

To see this: $\widehat{\mathbb{P}} = \{g^{-1}(\hat{p}(y))|g^{-1} \in \mathcal{C}^{-1}\}$

$\widehat{\mathbb{Q}} = \{g^{-1}(\hat{q}(y))|g^{-1} \in \mathcal{C}^{-1}\}$

$\Longrightarrow$ convex iff $\mathcal{C}^{-1}$ convex

R-CDT: linear separation theorem

Consider signal classes:

$\mathbb{P} = \{p_g : g \in \mathcal{C}\}$ and $\mathbb{Q} = \{q_g : g \in \mathcal{C}\}$

where

$$p_g = \mathcal{R}^{-1}\left((g^\theta)' \widetilde{p_0} \circ g^\theta\right)$$
$$q_g = \mathcal{R}^{-1}\left((g^\theta)' \widetilde{q_0} \circ g^\theta\right)$$
$$\mathbb{P} \cap \mathbb{Q} = \emptyset$$

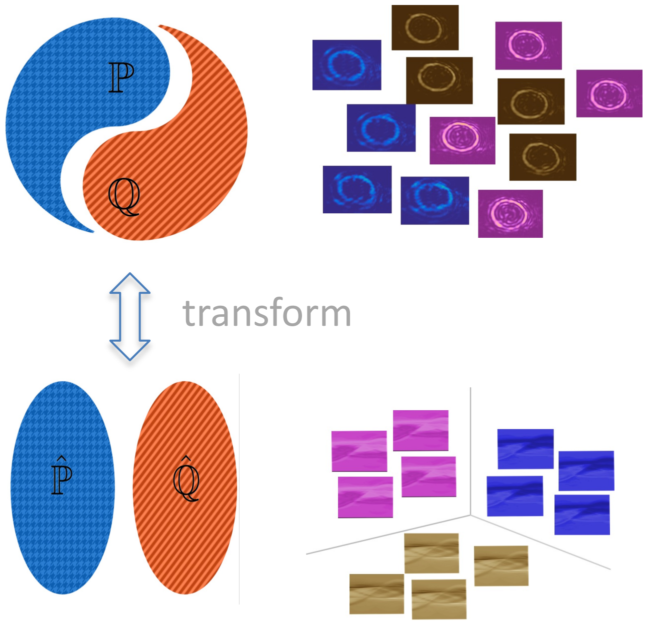
Theorem:

$\hat{\mathbb{P}}$, $\hat{\mathbb{Q}}$ will be *linearly separable*

if $\mathcal{C}^{-1}$ is convex.

for any reference s_0

does not require knowledge of p_0 q_0



[Park, Kolouri, Kundu, Rohde, ACHA 18]
[Aldroubi, Li, Rohde, SaSiDa, 21]
[Shifat-E-Rabbi, ..., Rohde, JMIV 21]
[Aldroubi et al, AIMS Foundations of Data Science, 22]

Linear Optimal Transport (LOT)

Wasserstein space $(P(\Omega), W_2)$ is a Riemannian manifold

geodesic

$$f_\alpha(x) = (1-\alpha)x + \alpha f(x)$$

$$s_0(x) = D_f(x) s_1(f(x)), \; f = \nabla\phi = \hat{s}_1$$

Linear optimal transport (LOT):

[Wang, Slepcev, Basu, Ozolek, Rohde, IJCV 13]

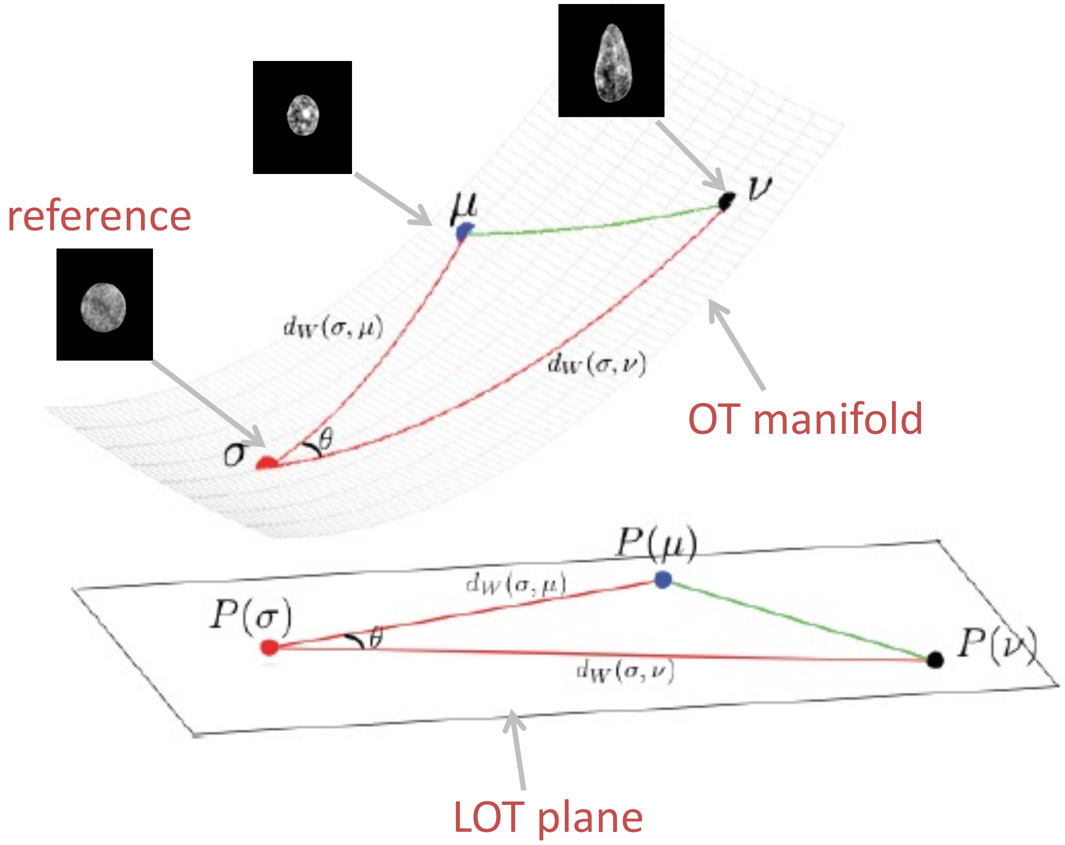
projection (invertible):

$$P(s) = \hat{s}$$

LOT distance:

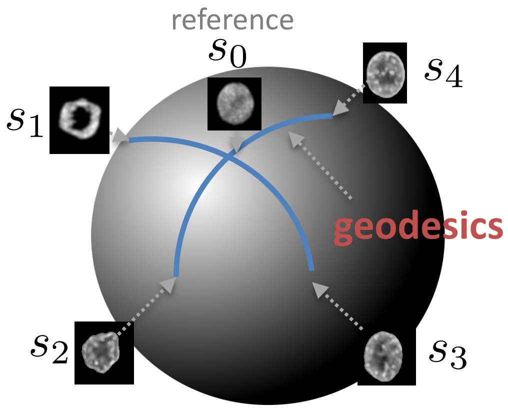
$$\|P(s_1)\sqrt{s_0} - P(s_2)\sqrt{s_0}\|^2$$

CDT is simply LOT projection

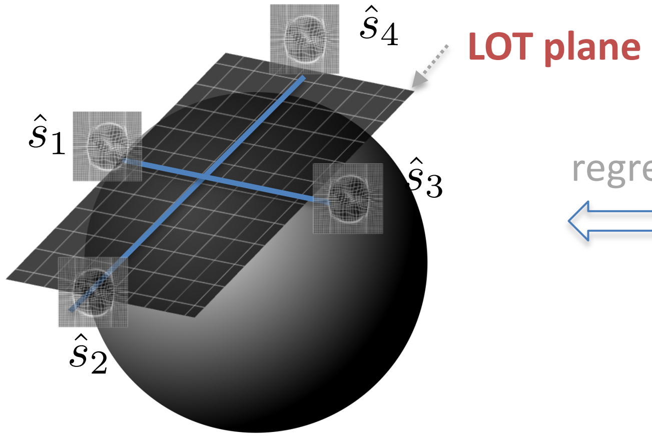


Azimuthal equidistant projection

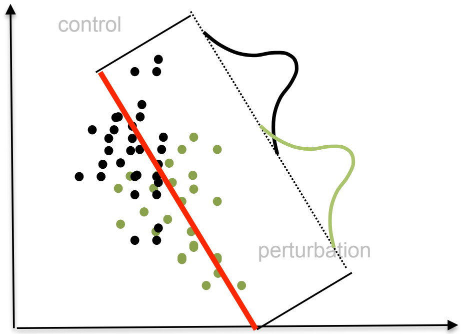
Linear Optimal Transport (LOT)



transform



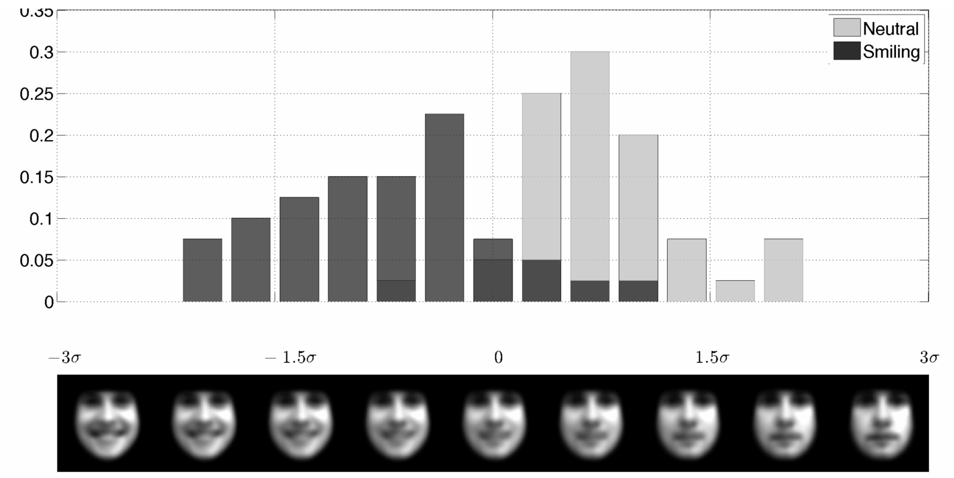
regression



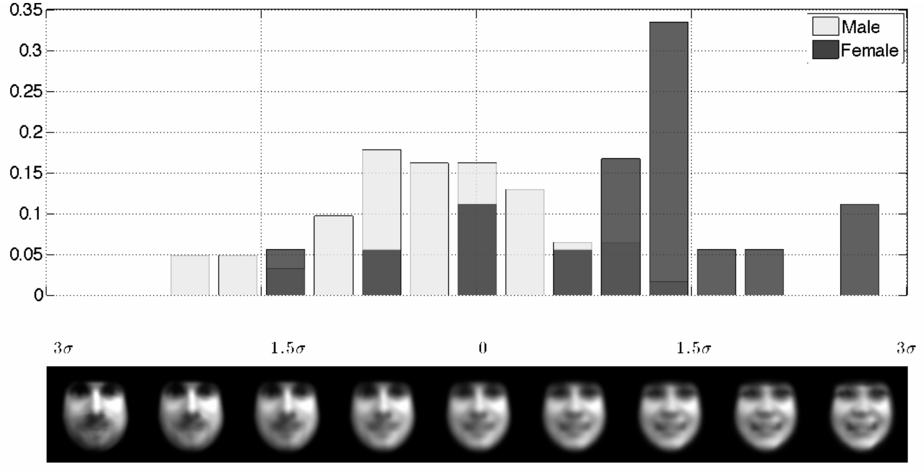
Face database (smiling, neutral)

vision.ucsd.edu

Smiling vs. neutral



Men vs. women



LOT computed with variational OT solver

[Kolouri et al, Pattern Recognition, 2016]

Convexity property of CDT (1D):

Recall that for $\mathbb{P}(p,\mathcal{C}) = \{p_g = g' p \circ g : g \in \mathcal{C}\}$ (p - template, $\mathcal{C}$ - a set of increasing diffeomorphisms),

$\hat{\mathbb{P}}$ is convex for every p iff $\mathcal{C}^{-1}$ is convex.

Key: composition property $\widehat{p}_g = g^{-1} \circ \widehat{p}$

Corollary: If $\mathcal{C}$ is a convex group, the transform space is partitioned equivalent classes $\bigcup_p \widehat{\mathbb{P}}(p,\mathcal{C})$, where each $\widehat{\mathbb{P}}(p,\mathcal{C})$ is convex.

Note: group $\mathcal{C}$ defines an equivalence relation by $p \sim_{\mathcal{C}} q$ iff $q \in \mathbb{P}(p,\mathcal{C})$
i.e., $\widehat{\mathbb{P}}(p,\mathcal{C}) \cap \widehat{\mathbb{P}}(q,\mathcal{C}) \neq \emptyset \Rightarrow \widehat{\mathbb{P}}(p,\mathcal{C}) = \widehat{\mathbb{P}}(q,\mathcal{C})$

Given a subgroup $\mathcal{C}_1 \subseteq \mathcal{C}$, each $\widehat{\mathbb{P}}(p,\mathcal{C})$ can be further partitioned into a collection of convex sets

Examples of convex groups:

- scaling group: $\{\alpha \, \mathrm{id} : \alpha > 0\}$
- translation group: $\{g_\mu(x) = x - \mu : \mu \in \mathbb{R}\}$
- affine group: $\{g_{\alpha,\mu}(x) = \alpha x - \mu : \alpha > 0, \mu \in \mathbb{R}\}$
- Fixed point group: $\{f : f(x_0) = x_0 : f \text{ is an increasing diffeomorphism}\}, \quad x_0 \in \mathbb{R}$
- It can be shown there are uncountable many distinct convex subgroups

[Aldroubi, Li, Rohde, SaSiDa, 21]

Convexity property of LOT (n-D)

Unlike the 1-d case, the convexity of $\widehat{\mathbb{P}}(p,\mathcal{C})$ does not follow from the convexity of $\mathcal{C}^{-1}$ using the same proof techniques since the composition property does not hold in general

Let $\mathcal{P}_d := \{p : \mathbb{R}^d \to \mathbb{R}^+ \mid \mathrm{supp}(p) = \Omega_p \text{ is compact}, \int_{\mathbb{R}^d} p(x)dx = 1\}$ be the set of non-negative normalized signals on $\mathbb{R}^d$

Theorem: Let $d \geq 2$ and $\mathcal{W}_p \subseteq \mathcal{F}_d$ be a set of deformations. If $\widehat{p}_g = g^{-1} \circ \widehat{p}$ holds for all $g \in \mathcal{W}_p$, then $\widehat{\mathbb{P}}(p,\mathcal{W}_p)$ is convex if the set $\mathcal{W}_p^{-1}$ is convex.

Theorem: Let $d \geq 2$ and $\mathcal{W} \subseteq \mathcal{F}_d$ be a set of deformations. If for any $p \in \mathcal{P}_d$, $\widehat{p}_g = g^{-1} \circ \widehat{p}$ holds all $g \in \mathcal{W}$ then $\mathcal{W} \subseteq \mathcal{H}_a := \{h(x) = ax + u : a > 0, u \in \mathbb{R}^d\}$, i.e., if the composition property holds regardless of p, $\mathcal{W}$ has to be a subset of compositions of translation and scaling diffeomorphisms.

Remark: one can relax the condition $p \in \mathcal{P}_d$ above to hold on a subset of $\mathcal{P}_d$ and enlarge $\mathcal{W}$.

Open questions: Can one derive the convexity of $\widehat{\mathbb{P}}(p,\mathcal{W})$ from the convexity of $\mathcal{W}^{-1}$ without the composition property being true for all $g \in \mathcal{W}$? Are there other conditions one can impose on the model $\mathbb{P}(p,\mathcal{W})$ other than $\mathcal{W}^{-1}$ being convex so that $\widehat{\mathbb{P}}(p,\mathcal{W}_p)$ is convex ?

[Aldroubi, Li, Rohde, SaSiDa, 21]

A relaxation in dimension d=2

Definition 4.8 (Restrictive sets of transformations and PDFs).

(16) $$\mathcal{H}_r = \left\{ h(x,y) := \frac{1}{2}\begin{bmatrix} f'(x+y) + g'(x-y) \\ f'(x+y) - g'(x-y) \end{bmatrix} \mid f, g \in \mathcal{R} \right\}$$

where $\mathcal{R} = \{f \in C^2(\mathbb{R}) \mid f'$ *is a strictly increasing bijection on* $\mathbb{R}\}$.

(17) $$\mathcal{P}_r := \{p \in \mathcal{P}_2 : |\det J_h|(p \circ h) = r \text{ for some } h \in \mathcal{H}_r\}.$$

Note: $\mathcal{H}_r$ is a convex group .

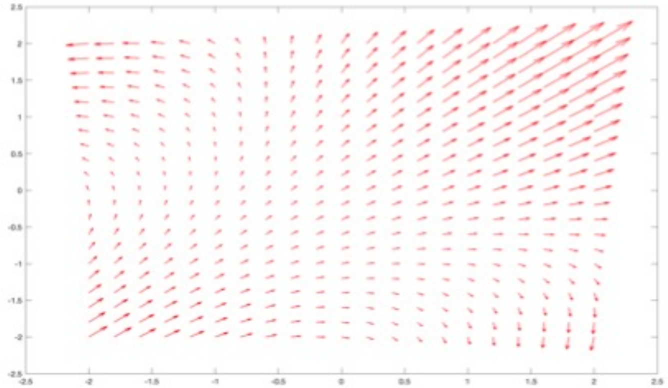


FIGURE 6. A vector field h on a grid $[-2,2] \times [-2,2]$ generated with $f'(t) = t + 0.1t^2$ and $g'(t) = t$.

Theorem 4.10. *For any* $p \in \mathcal{P}_r$, *we have* $\widehat{p_h} = h^{-1} \circ \widehat{p}$ *for any* $h \in \mathcal{H}_r$ *where* $p_h = |\det J_h| \cdot p \circ h$.

Note: here the generating template r for $\mathcal{P}_r$ is the same as reference r for computing the LOTs

Example 2. *Let* $\mathcal{R}_s = \{f(t) = at^2 + bt \mid a > 0, b \in \mathbb{R}\}$ *and* $\mathcal{H}_s = \{h(x,y) = \frac{1}{2}\nabla(f(x+y) + g(x-y)) \mid f, g \in \mathcal{R}_s\}$

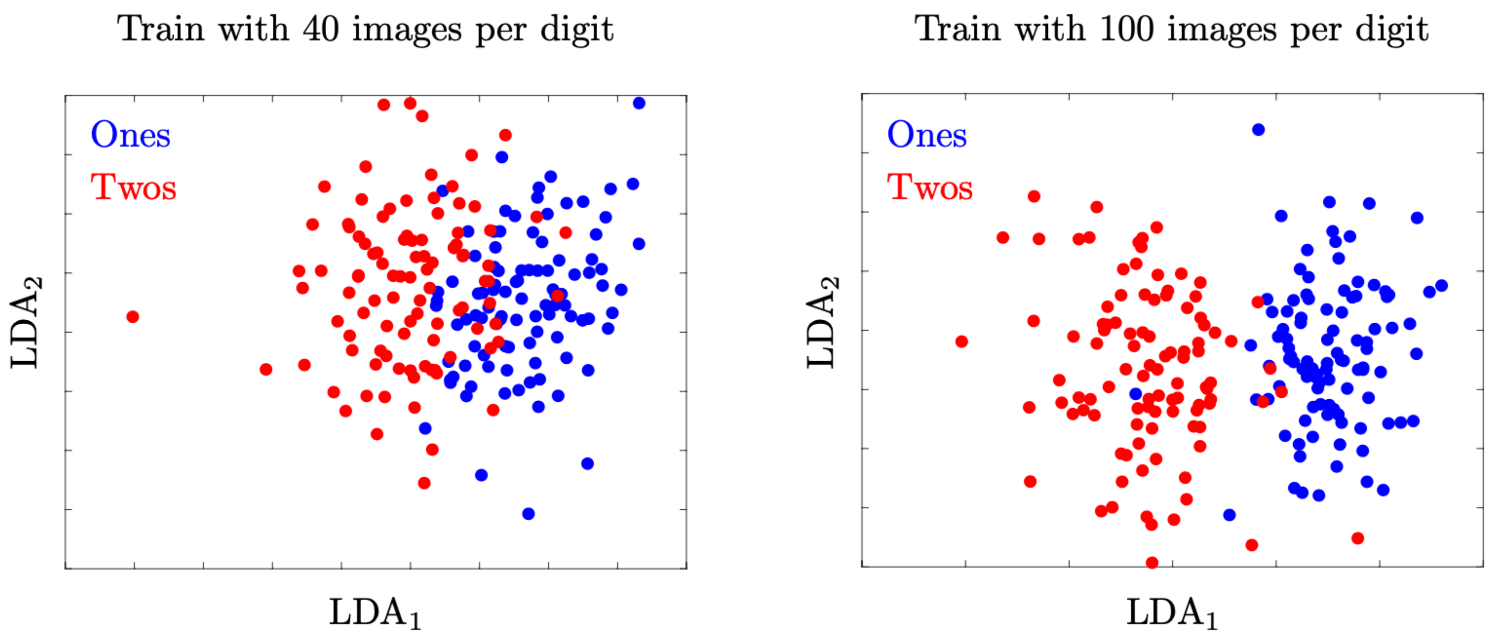
By direct computation, it is easy to see that every $h \in \mathcal{H}_s$ is of the form

(20) $$h(x,y) = (a_1 + a_2)\begin{bmatrix} x \\ y \end{bmatrix} + (a_1 - a_2)\begin{bmatrix} y \\ x \end{bmatrix} + \begin{bmatrix} b_1 - b_2 \\ b_1 + b_2 \end{bmatrix},$$

and vice versa, where $a_1, a_2 > 0$ and $b_1, b_2 \in \mathbb{R}$. Equivalently, $h(x,y) = A\begin{bmatrix} x \\ y \end{bmatrix} + u$, where $A = \begin{bmatrix} a_1 + a_2 & a_1 - a_2 \\ a_1 - a_2 & a_1 + a_2 \end{bmatrix}$ and $u = \begin{bmatrix} b_1 - b_2 \\ b_1 + b_2 \end{bmatrix}$. It is not difficult to show that $\mathcal{H}_s$ is a convex group of diffeomorphisms under the composition operation.

[Moosmüller, Cloninger, 21] **Linear Optimal Transport Embedding: Provable Wasserstein classification for certain rigid transformations and perturbations**
https://arxiv.org/abs/2008.09165

- explored settings when LOT embeds family of distributions into linearly separable sets
- linear separability: if $\mathcal{P} = \{\mu_i : y_i = 1\}$ are ϵ-perturbations of shifts and scalings of μ and $\mathcal{Q} = \{\nu_i : y_i = -1\}$ are are ϵ-perturbations of shifts and scalings of ν, and $\mathcal{P}$ and $\mathcal{Q}$ have a small minimal distance depending on ϵ, then $\mathcal{P}$ and $\mathcal{Q}$ are linearly separable in the LOT embedding space.
- provided conditions when the LOT distance between two measures is nearly isometric to the Wasserstein distance



LDA embedding plots for the MNIST classification of digits 1 and 2 using LOT.

figure credit: Moosmüller et al.

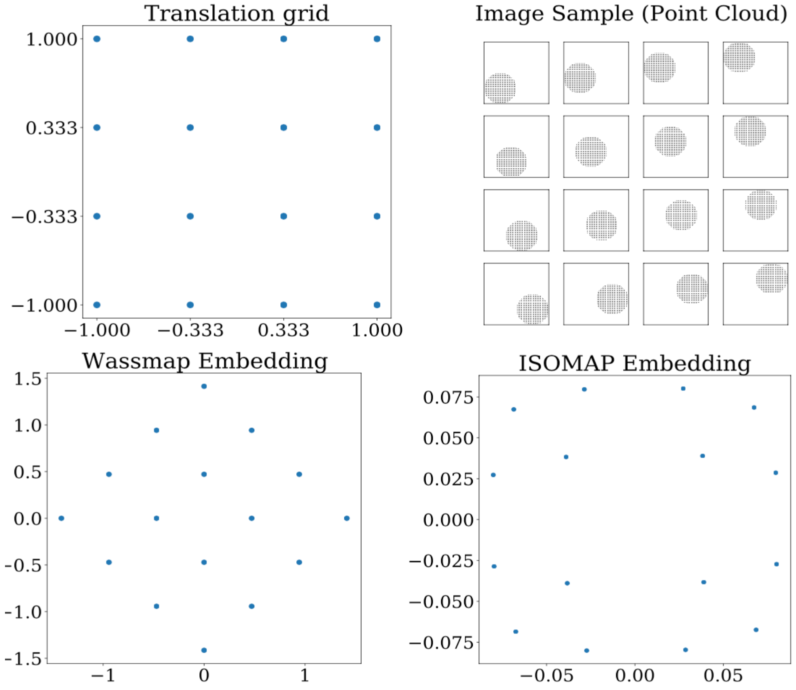
[Hamm, Henscheid, Kang, 22] **Wassmap: Wasserstein Isometric Mapping for Image Manifold Learning**
https://arxiv.org/abs/2204.06645

- Proposed a variant of ISOMAP called Wassmap, where classical multi-dimensional scaling (MDS) is applied to pairwise Wasserstein distances of data

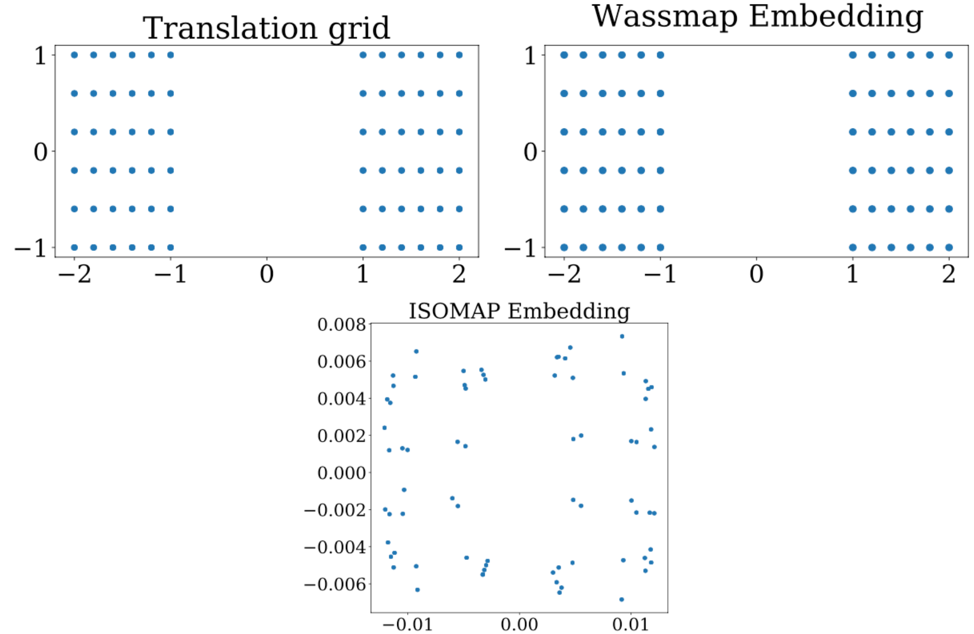
Algorithm 3.1 Functional Wasserstein Isometric Mapping (Functional Wassmap)

Input: Probability measures $\{\mu_i\}_{i=1}^N \subset \mathbb{W}_2(\mathbb{R}^m)$; embedding dimension d
Output: Low-dimensional embedding points $\{z_i\}_{i=1}^N \subset \mathbb{R}^d$
Compute pairwise Wasserstein distance matrix $W_{ij} = W_2^2(\mu_i, \mu_j)$
$B = -\frac{1}{2}HWH$, where $(H = I - \frac{1}{N}\mathbb{1}_N)$
(Truncated SVD): $B_d = V_d \Lambda_d V_d^T$
$z_i = (V_d \Lambda_d^{\frac{1}{2}})(i, :)$, for $i = 1, \ldots, N$
Return: $\{z_i\}_{i=1}^N$

- Recover manifold parametrizations



THEOREM A. *Let $\Theta \subset \mathbb{R}^d$ be a parameter set that generates a smooth submanifold $\mathcal{M}(\Theta) \subset \mathbb{W}_2(\mathbb{R}^m)$ such that $(\mathcal{M}, W_2)$ is isometric up to a constant to $(\Theta, |\cdot|_{\mathbb{R}^d})$. If $\{\theta_i\}_{i=1}^N \subset \Theta$, and $\{\mu_{\theta_i}\}_{i=1}^N \subset \mathcal{M}$ are the corresponding measures on the manifold, then the Functional Wassmap Algorithm (Algorithm 3.1) with embedding dimension d recovers $\{\theta_i\}$ up to rigid transformation and global scaling.*



[CAI, CHENG, SCHMITZER, THORPE, SIIMS 22] **The Linearized Hellinger–Kantorovich Distance** https://arxiv.org/pdf/2102.08807.pdf

- studied the local linearization of the Hellinger–Kantorovich distance via its Riemannian structure.
- Hellinger–Kantorovich distance defines a metric for non-negative measures which allows for the creation/destruction of mass within the optimal transport framework and can be seen as a particular variant of an 'unbalanced' transport problem

DEFINITION 3.1 (Continuity equation with source). *For $\mu_0, \mu_1 \in \mathcal{M}(\Omega)$ we denote by $\mathcal{CES}(\mu_0, \mu_1)$ the set of solutions for the continuity equation with source on $[0,1] \times \Omega$, i.e. the set of triplets of measures $(\rho, \omega, \zeta) \in \mathcal{M}([0,1] \times \Omega)^{1+d+1}$ where ρ interpolates between μ_0 and μ_1 and that solve*

$$\partial_t \rho + \operatorname{div} \omega = \zeta$$

[Beier, Beinert, Steidl, 21] **On a linear Gromov–Wasserstein distance** https://arxiv.org/pdf/2112.11964.pdf

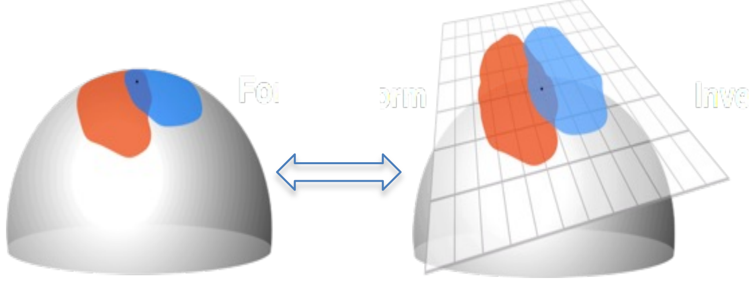
- Gromov-Wasserstein (GW) distance is invariant under translation and rotation, is practical for applications such as shape comparison
- defined a generalized LOT based on barycentric projection maps of transport plans, which coincides with LOT when the reference (base) measure is absolutely continuous
- defined a notion of generalized linear Gromov-Wasserstein (gLGW) distance to alleviate computational cost of GW

[Nenna, Pass, 22] **Transport type metrics on the space of probability measures involving singular base measures**

- Semi-metric inspired by generalized geodesics $W_\nu(\mu_1, \mu_0) := \sqrt{\inf_{\gamma \in \mathcal{P}(X \times X \times X) | \gamma_{yx_i} \in \Pi_{opt}(\mu_i, \nu), i=0,1} \int_{X \times X \times X} |x_0 - x_1|^2 d\gamma(y, x_0, x_1)}.$
- the base measure ν can be singular, e.g., concentrated on a lower dimensional manifold -- related to the <u>unequal dimensional OT</u>; If ν is concentrated on a line segment – <u>layerwise-Wasserstein metric;</u> LOT distance can be seen as a special case
- If ν abs. continuous w.r.t Lebesgue, $W_\nu(\mu_0, \mu_1) = \int_X |\tilde{T}_0(y) - \tilde{T}_1(y)|^2 d\nu(y)$ -- LOT
- If $\nu = \delta_y$, $W_\nu^2(\mu_0, \mu_1) = \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{X \times X} |x_0 - x_1|^2 d\pi(x_0, x_1)$ -- standard Wasserstein

Summary: Transport & other Lagrangian Transforms

Nonlinear signal transform



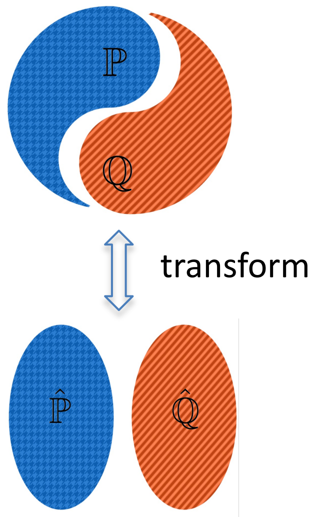
smooth PDFs diffeomorphisms

(LOT, CDT, R-CDT):

invertible transform framework for
Lagrangian modeling of signals & images
Connections to optimal transport

Properties

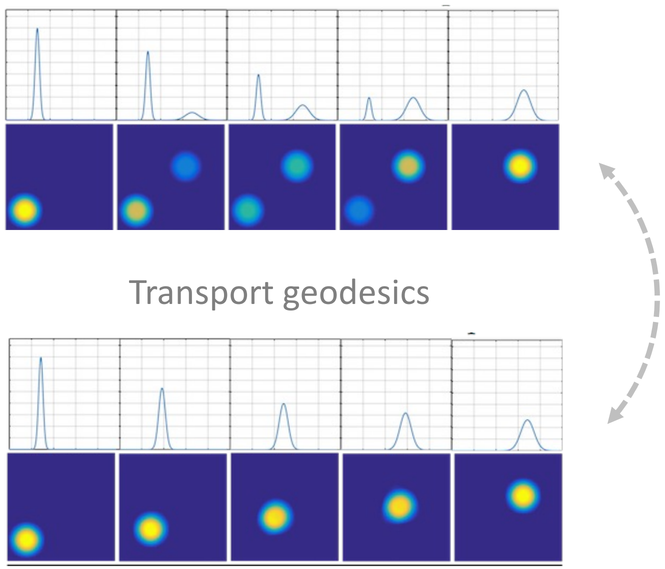
nonconvex,
nonlinear



convex,
linearly separable

Intuition

Euclidean geodesics



“looks around,”
moves intensities

Imagedatascience.com/transport