The Cumulative Distribution Transform For Data Analysis And Machine Learning

Akram Aldroubi
Vanderbilt University

Jubilee of Fourier Analysis and Applications
In Celebration of John Benedetto 80th Birthday

– Typeset by FoilTEX –
Supported by NIH Grant (Gustavo Rohde PI)

Gustavo Rohde
Transports Transform

Transforms: Fourier Transform, Wavelet transform, Zak Transform, Shearlets, Scattering “transform,”...
Transports Transform

Transforms: Fourier Transform, Wavelet transform, Zak Transform, Shearlets, Scattering “transform,”... 

Transport Transforms: Non-linear transforms based on transport theory: Monge and Kantorovich Transport theory
The Monge Problem (1781)

The Monge Problem Let $\mu$ be a pile of sand on $X \subset \mathbb{R}^n$, find the "most efficient way" to transport it to the hole in the ground $\nu$ on $X \subset \mathbb{R}^n$. 
The Monge Problem (1781)

The Monge Problem Let $\mu$ be a pile of sand on $X \subset \mathbb{R}^n$, find the "most efficient way" to transport it to the hole in the ground $\nu$ on $X \subset \mathbb{R}^n$.

Let $\mu, \nu$ be probability measures on $\mathbb{R}^n$ find a map $T^\dagger : \mathbb{R}^n \to \mathbb{R}^n$ such

$$T^\dagger = \arg \min_{\nu=T^\#\mu} \int_{\mathbb{R}^n} \|x - T(x)\|^2 d\mu(x),$$

where $\nu(B) = \mu(T^{-1}(B))$ for all measurable measurable sets $B$. 
The Monge Problem (1781)

The Monge Problem Let \( \mu \) be a pile of sand on \( X \subset \mathbb{R}^n \), find the ”most efficient way” to transport it to the hole in the ground \( \nu \) on \( X \subset \mathbb{R}^n \).

Let \( \mu, \nu \) be probability measures on \( \mathbb{R}^n \) find a map \( T^\dagger : \mathbb{R}^n \rightarrow \mathbb{R}^n \) such

\[
T^\dagger = \arg \min_{\nu = T_\# \mu} \int_{\mathbb{R}^n} \| x - T(x) \|^2 d\mu(x),
\]

where \( \nu(B) = \mu(T^{-1}(B)) \) for all measurable sets \( B \).

Weak version of the Monge problem: The Kantorovich problem (1939).
Brenier’s Theorem (1991)

(Brenier’s Theorem) Let $\mu, \nu$ be two probability measures on $\mathbb{R}^n$ (finite 2nd moments) that are absolutely continuous w.r.t Lebesgue measure. Then there exists a map $T^\dagger : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that

$$T^\dagger = \arg \min_{\nu = T_\# \mu} \int_{\mathbb{R}^n} \|x - T(x)\|^2 d\mu(x),$$

where $\nu(B) = \mu(T^{-1}(B))$ for all measurable measurable sets $B$. Moreover, $T^\dagger$ is unique. (Generalization by Gangbo and McCann 1996, Villani 2006)
Brenier’s Theorem (1991)

(Brenier’s Theorem) Let \( \mu, \nu \) be two probability measures on \( \mathbb{R}^n \) (finite 2nd moments) that are absolutely continuous w.r.t Lebesgue measure. Then there exists a map \( T^\dagger : \mathbb{R}^n \rightarrow \mathbb{R}^n \) such that

\[
T^\dagger = \arg \min_{\nu = T_\#\mu} \int_{\mathbb{R}^n} \|x - T(x)\|^2 d\mu(x),
\]

where \( \nu(B) = \mu(T^{-1}(B)) \) for all measurable measurable sets \( B \). Moreover, \( T^\dagger \) is unique. (Generalization by Gangbo and McCann 1996, Villani 2006)

The transport transform: Let \( d\mu(x) = s(x)dx \) and \( d\nu(x) = s_0(x)dx \), where \( r \) is a fixed reference signal, then the transform \( \tilde{s} \) of \( s \) is the unique solution to the Monge problem above, i.e., \( \tilde{s} = T^\dagger \).
The CDT transform

Let $s$ a smooth probability density function on $[0, 1] \subset \mathbb{R}$, and $s_0$ a reference probability density function on $[0, 1] \subset \mathbb{R}$.

The Cumulative Distribution Transform

$$\tilde{s}(x) = \int_0^x s(\xi) d\xi = \int_0^x s_0(\xi) d\xi, \quad x \in [0, 1].$$
The CDT transform

Let $s$ a smooth probability density function on $[0, 1] \subset \mathbb{R}$, and $s_0$ a reference probability density function on $[0, 1] \subset \mathbb{R}$.

The Cumulative Distribution Transform

$$\tilde{s}(x) = \int_0^x s(\xi) d\xi = \int_0^x s_0(\xi) d\xi, \quad x \in [0, 1].$$

Typically $s_0 = \chi[0, 1]$. 

The CDT transform

Let \( s \) a smooth probability density function on \([0, 1] \subset \mathbb{R}\), and \( s_0 \) a reference probability density function on \([0, 1] \subset \mathbb{R}\).

The Cumulative Distribution Transform

\[
\tilde{s}(x) = \int_{0}^{x} s(\xi) d\xi = \int_{0}^{x} s_0(\xi) d\xi, \quad x \in [0, 1].
\]

Typically \( s_0 = \chi[0, 1] \).

Inverse Transform:

\[
\tilde{s}'(x)s(\tilde{s}(x)) = s_0(x).
\]
CDT and its inverse

[Image: Diagram showing signal (pdf) with match to fixed reference and transform.]

\[ s(x) = \int s(f(x)) \, df(x) \]

[Nonlinear map]

[Forward and inverse]

[Reference: Park, Kolouri, Kundu, Rohde, ACHA 2018]
Radon-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
\[ \mathcal{R}_I^{-1}(x) = \int_{S^{d-1}} (\mathcal{R}I(\cdot, \theta) * \eta(\cdot)) \circ (x \cdot \theta) \, d\theta \]

Radon Cumulative Distribution Transform

[Kolouri, Park, Rohde, IEEE TIP 2016]
Convexification properties

\[ T_s : \mathbb{R} \rightarrow L^2(\mathbb{R}) : T_s(\mu)(x) = s_\mu(x) = s(x - \mu). \]
Convexification properties

\[ p_\mu(x) = p_0(x - \mu); \mu \in \mathbb{R} \]

\[ q_\mu(x) = q_0(x - \mu); \mu \in \mathbb{R} \]

\[ \hat{p}_\mu(x) = \hat{p}_0(x) + \mu; \mu \in \mathbb{R} \]

\[ \hat{q}_\mu(x) = \hat{q}_0(x) + \mu; \mu \in \mathbb{R} \]
Convexification properties

- Nonlinear signal transform
- Smooth PDFs
- Diffeomorphisms
- Nonconvex, nonlinear
  - [Wang et al., UCV, 2014]
  - [Basu et al., PNAS 2014]
  - [Kolouri et al IEEE TIP 16]
  - [Kolouri et al, Pat. Rec, 17]
  - [Park et al, ACHA 18]
  - [Kolouri et al. IEEE SPM 17]
  - Imagedatascience.com/transport

- Convex, linearly separable
Wasserstein distance between two measures

Let $\Pi(\mu, \nu)$ be the set of all probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals $\mu$ and $\nu$.

2-Wasserstein distance between $\mu$ and $\nu$:

$$W_2^2(\mu, \nu) = \min_{\pi \in \Pi} \int_{\mathbb{R}^n} ||x - y||^2 d\pi(x, y).$$
Wasserstein distance between two measures

Let $\Pi(\mu, \nu)$ be the set of all probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals $\mu$ and $\nu$.

2-Wasserstein distance between $\mu$ and $\nu$:

$$W_2^2(\mu, \nu) = \min_{\pi \in \Pi} \int_{\mathbb{R}^n} \|x - y\|^2 d\pi(x, y).$$

Existence of a minimizer is due to Kantorovich.
Wasserstein distance between two measures

Let $\Pi(\mu, \nu)$ be the set of all probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals $\mu$ and $\nu$.

2-Wasserstein distance between $\mu$ and $\nu$:

$$W_2^2(\mu, \nu) = \min_{\pi \in \Pi} \int_{\mathbb{R}^n} \|x - y\|^2 \, d\pi(x, y).$$

Existence of a minimizer is due to Kantorovich.

The solution $T^\dagger$ of the transport map of the Monge problem give rise to the minimizer to the Kantorovich problem: $d(\mu(x)\delta(y = T^\dagger(x)) \in \Pi$ and minimizes $W_2(\mu, \nu)$. 
Wasserstein distance between two measures

Theorem (Kolouri, Rohde) Let \( s_1 \) and \( s_2 \) be two signals (PDFs) and \( \tilde{s}_1, \tilde{s}_2 \) be their CDT transform with respect to fixed reference \( s_0 \). Then

\[
\|\tilde{s}_2 - \tilde{s}_1\|_{L^2(\mathbb{R}^n)}^2 = W_2(\mu, \nu) = \min_{\pi \in \Pi} \int_{\mathbb{R}^n} \|x - y\|^2 d\pi(x, y)
\]

where \( d\mu(x) = s_1(x)dx \) and \( d\nu = s_2(x)dx \).
signal space, Wasserstein geodesics

transform space Euclidean geodesics

[Kolouri, Rohde, CVPR 16]
Applications

Example: classifying facial expressions

Radon Cumulative Distribution Transform

Image space

Class 1

Class 2

Transform space

Class 1

Class 2

Projection of the data onto a 2-dimensional discriminant subspace

Projection of the transformed data onto a 2-dimensional discriminant subspace
Applications

(TBM): modeling discriminant information

Class 1: benign cells

Class 2: malignant cells

LDA in transport space

Fetal-type hepatoblastoma

Actual classifier inverse

CANCER MODELING

- Liver: Basu et al, PNAS 2014
- Thyroid: Ozolek et al, Medical Image Analysis, 2014
- Skin melanoma: Liu et al, J. Pathology Informatics, 2016
- Breast carcinoma: unpublished
Applications

Application: single image super-resolution

\[ M_\alpha = D_{f_\alpha}(x)I_0(f_\alpha(x)) \]

\[ f_\alpha(x) = \sum_k \alpha_k \phi_k(x) \]

Fit:
\[ \min_\alpha \| I - VM_\alpha \|^2 \]

Transport PCA trained on high res database

Happy Birthday Caro