MAX-NORMALIZED RADON CUMULATIVE DISTRIBUTION TRANSFORM FOR LIMITED DATA CLASSIFICATION

Matthias Beckmann[1], Robert Beinert[2], Jonas Bresch[2]

[1] Universität Bremen, [2] Technische Universität Berlin

Key Takeaways

- We introduce a novel image feature extractor, called $_{\mathrm{m}}$NR-CDT, which is invariant under arbitrary affine transformations.
- Our $_{\mathrm{m}}$NR-CDT extends the Radon cumulative distribution transform [1] and is related to the sliced Wasserstein distance [2].
- $_{\mathrm{m}}$NR-CDT allows for (near to perfect) classification, especially, in small data regimes.

Radon Transform of Measures

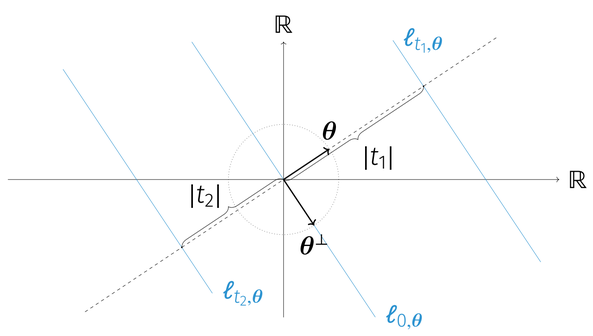
- For $\boldsymbol{\theta} \in \mathbb{S}_1 := \{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\| = 1\}$, we consider the *slicing operator*

$$S_{\boldsymbol{\theta}} \colon \mathbb{R}^2 \to \mathbb{R}, \quad S_{\boldsymbol{\theta}}(\mathbf{x}) := \langle \mathbf{x}, \boldsymbol{\theta} \rangle,$$

- and define the *restricted Radon transform* for measures via

$$\mathcal{R}_{\boldsymbol{\theta}} \colon \mathcal{M}(\mathbb{R}^2) \to \mathcal{M}(\mathbb{R}), \quad \mathcal{R}_{\boldsymbol{\theta}}[\mu] := (S_{\boldsymbol{\theta}})_{\#}\mu = \mu \circ S_{\boldsymbol{\theta}}^{-1}.$$

$\rightsquigarrow$ $\mathcal{R}_{\boldsymbol{\theta}}$ corresponds to integration along lines $\ell_{t,\boldsymbol{\theta}} := S_{\boldsymbol{\theta}}^{-1}(t)$.



- For $\mu_{\mathbf{A},\mathbf{y}} := (\mathbf{A} \cdot + \mathbf{y})_{\#}\mu$, the restricted Radon transform satisfies

$$\mathcal{R}_{\boldsymbol{\theta}}[\mu_{\mathbf{A},\mathbf{y}}] = (\|\mathbf{A}^{\top}\boldsymbol{\theta}\| \cdot + \langle \mathbf{y}, \boldsymbol{\theta} \rangle)_{\#} \mathcal{R}_{\frac{\mathbf{A}^{\top}\boldsymbol{\theta}}{\|\mathbf{A}^{\top}\boldsymbol{\theta}\|}}[\mu].$$

Radon Cumulative Distribution Transform

- For $\mu \in \mathcal{P}(\mathbb{R})$, we consider the *cumulative distribution function*

$$F_{\mu} \colon \mathbb{R} \to [0,1], \quad F_{\mu}(t) := \mu((-\infty, t]),$$

- the *cumulative distribution transform* (CDT)

$$\widehat{\mu} \colon [0,1] \to \mathbb{R}, \quad \widehat{\mu}(t) := \inf\{s \in \mathbb{R} \mid F_{\mu}(s) > t\} = F_{\mu}^{[-1]}(t),$$

- and the *Radon cumulative distribution transform* (R-CDT)

$$\widehat{\mathcal{R}}[\mu] \colon [0,1] \times \mathbb{S}_1 \to \mathbb{R}, \quad \widehat{\mathcal{R}}[\mu](t, \boldsymbol{\theta}) := \widehat{\mathcal{R}}_{\boldsymbol{\theta}}[\mu](t).$$

$\rightsquigarrow$ For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^2)$, we recover the sliced Wasserstein-2 distance

$$\|\widehat{\mathcal{R}}[\mu] - \widehat{\mathcal{R}}[\nu]\| := \left(\int_{\mathbb{S}_1} \int_0^1 |\widehat{\mathcal{R}}[\mu](t, \boldsymbol{\theta}) - \widehat{\mathcal{R}}[\nu](t, \boldsymbol{\theta})|^2 \, \mathrm{d}t \, \mathrm{d}u_{\mathbb{S}_1}(\boldsymbol{\theta}) \right)^{\frac{1}{2}}.$$

Normalized R-CDT

- For $g \in \mathrm{L}^2([0,1])$, we set

$$\mathsf{mean}(g) := \int_0^1 g(s) \, \mathrm{d}s, \quad \mathsf{std}(g) := \left(\int_0^1 |g(s) - \mathsf{mean}(g)|^2 \, \mathrm{d}s \right)^{\frac{1}{2}}.$$

- We restrict ourselves to non-degenerate measures

$$\begin{aligned} \mathcal{P}_c^*(\mathbb{R}^2) &:= \{\mu \in \mathcal{P}(\mathbb{R}^2) \mid \mathsf{supp}(\mu) \subset\subset \mathbb{R}^2 \wedge \dim(\mathsf{supp}(\mu)) > 1\} \\ &= \{\mu \in \mathcal{P}(\mathbb{R}^2) \mid \exists\, c > 0 \colon \mathsf{std}(\widehat{\mathcal{R}}_{\boldsymbol{\theta}}[\mu]) \geq c \;\forall\, \boldsymbol{\theta} \in \mathbb{S}_1\} \end{aligned}$$

- and introduce the *normalized R-CDT* (NR-CDT)

$$\mathcal{N}_{\boldsymbol{\theta}}[\mu] \colon [0,1] \to \mathbb{R}, \quad \mathcal{N}_{\boldsymbol{\theta}}[\mu](t) := \frac{\widehat{\mathcal{R}}_{\boldsymbol{\theta}}[\mu](t) - \mathsf{mean}(\widehat{\mathcal{R}}_{\boldsymbol{\theta}}[\mu])}{\mathsf{std}(\widehat{\mathcal{R}}_{\boldsymbol{\theta}}[\mu])}.$$

The NR-CDT satisfies

$$\mathcal{N}_{\boldsymbol{\theta}}[\mu_{\mathbf{A},\mathbf{y}}] = \mathcal{N}_{\frac{\mathbf{A}^{\top}\boldsymbol{\theta}}{|\mathbf{A}^{\top}\boldsymbol{\theta}|}}[\mu].$$

- Finally, we define the *max-normalized R-CDT* ($_{\mathrm{m}}$NR-CDT)

$$\mathcal{N}_{\mathrm{m}}[\mu] \colon [0,1] \to \mathbb{R}, \quad \mathcal{N}_{\mathrm{m}}[\mu](t) := \sup_{\boldsymbol{\theta} \in \mathbb{S}_1} \mathcal{N}_{\boldsymbol{\theta}}[\mu](t).$$

The $_{\mathrm{m}}$NR-CDT satisfies

$$\mathcal{N}_{\mathrm{m}}[\mu_{\mathbf{A},\mathbf{y}}] = \mathcal{N}_{\mathrm{m}}[\mu].$$

Theorem (Linear Separability in $_{\mathrm{m}}$NR-CDT Space)

For template measures $\mu_0, \nu_0 \in \mathcal{P}_c^*(\mathbb{R}^2)$ with

$$\mathcal{N}_{\mathrm{m}}[\mu_0] \neq \mathcal{N}_{\mathrm{m}}[\nu_0]$$

consider the classes

$$\mathbb{F} = \left\{\mu \in \mathcal{P}(\mathbb{R}^2) \mid \exists\, \mathbf{A} \in \mathrm{GL}(2),\, \mathbf{y} \in \mathbb{R}^2 \colon \mu = (\mathbf{A} \cdot + \mathbf{y})_{\#}\mu_0\right\},$$

$$\mathbb{G} = \left\{\nu \in \mathcal{P}(\mathbb{R}^2) \mid \exists\, \mathbf{A} \in \mathrm{GL}(2),\, \mathbf{y} \in \mathbb{R}^2 \colon \nu = (\mathbf{A} \cdot + \mathbf{y})_{\#}\nu_0\right\}.$$

Then, $\mathbb{F}$ and $\mathbb{G}$ are linearly separable in $_{\mathrm{m}}$NR-CDT space.

Numerical Experiments

Datasets

1. *Academic:* Up to three synthetic template symbols are randomly translated, rotated, dilated, and sheared.
2. *LinMNIST:* MNIST digits (here 1, 5 and 7) are affinely transformed.

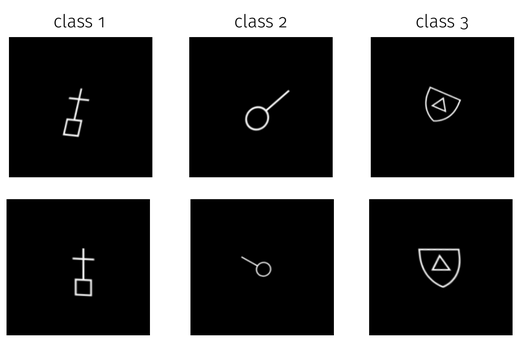


Figure: Academic samples.

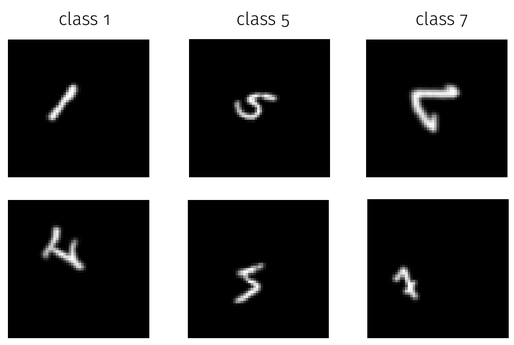


Figure: LinMNIST samples.

Nearest Neighbour Classification

- $_{\mathrm{m}}$NR-CDT maps affine classes to single points.
- Academic: Use only the original templates for classification.
- LinMNIST: Use five random samples for classification.

Table: NN accuracies for equispaced angles and different norms.

num. angles	academic $\|\cdot\|_\infty$	academic $\|\cdot\|_2$	LinMNIST $\|\cdot\|_\infty$	LinMNIST $\|\cdot\|_2$
2	0.5666	0.6333	0.5814 ± 0.0890	0.5970 ± 0.0467
4	0.9333	1.0000	0.6444 ± 0.0613	0.7155 ± 0.0492
8	0.9333	1.0000	0.7533 ± 0.0500	0.8200 ± 0.0454
16	0.9667	1.0000	0.7844 ± 0.0333	0.8362 ± 0.0401
32	0.9000	1.0000	0.8014 ± 0.0380	0.8429 ± 0.0409
64	0.8333	1.0000	0.8059 ± 0.0468	0.8503 ± 0.0377
128	0.8333	1.0000	0.8022 ± 0.0392	0.8570 ± 0.0300

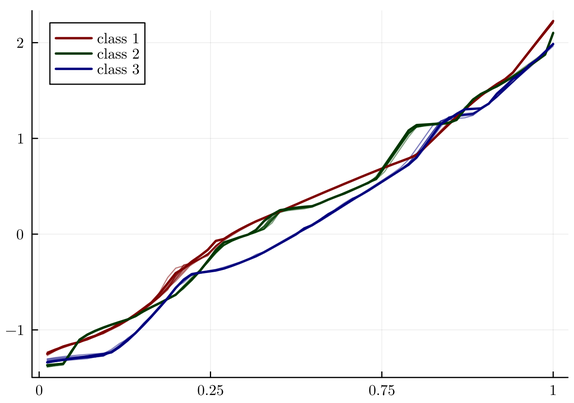


Figure: $_{\mathrm{m}}$NR-CDTs of academic dataset.

Support Vector Machine Classification

- $_{\mathrm{m}}$NR-CDT promotes linear separability.
- Cross-validation: 1 tenth for training, 9 tenths for testing.
- Comparison with naïve Euclidean and R-CDT ansatz.

Table: SVM accuracy for classes 1 vs. class 2 of academic dataset.

class size	Euclidean	R-CDT 2	R-CDT 4	R-CDT 8	$_{\mathrm{m}}$NR-CDT 2	$_{\mathrm{m}}$NR-CDT 4	$_{\mathrm{m}}$NR-CDT 8
10	0.5277 ± 0.1118	0.5555 ± 0.0868	0.5555 ± 0.0785	0.5500 ± 0.0714	0.6666 ± 0.1458	0.8777 ± 0.0819	**0.9944 ± 0.0175**
30	0.4777 ± 0.0697	0.5314 ± 0.0771	0.6111 ± 0.0740	0.6111 ± 0.0872	0.8129 ± 0.0512	0.9537 ± 0.0463	**1.0000 ± 0.0000**
90	0.5283 ± 0.0422	0.8074 ± 0.0673	0.9024 ± 0.0765	0.9259 ± 0.0770	0.9111 ± 0.0578	0.9728 ± 0.0184	**1.0000 ± 0.0000**
270	0.5413 ± 0.0286	0.9806 ± 0.0144	0.9983 ± 0.0045	0.9979 ± 0.0065	0.9938 ± 0.0073	0.9952 ± 0.0061	**1.0000 ± 0.0000**

Table: SVM accuracy for class 1 vs. class 7 of LinMNIST dataset.

class size	Euclidean	R-CDT 2	R-CDT 4	R-CDT 8	$_{\mathrm{m}}$NR-CDT 2	$_{\mathrm{m}}$NR-CDT 4	$_{\mathrm{m}}$NR-CDT 8
10	0.5055 ± 0.1841	0.5277 ± 0.1315	0.5333 ± 0.1233	0.5277 ± 0.1178	0.7000 ± 0.1340	0.7944 ± 0.1229	**0.8500 ± 0.0457**
20	0.4472 ± 0.0606	0.5166 ± 0.0644	0.5194 ± 0.0435	0.5222 ± 0.0520	0.8111 ± 0.0597	0.8250 ± 0.1134	**0.8666 ± 0.0772**
50	0.4911 ± 0.0214	0.5744 ± 0.6200	0.6200 ± 0.0755	0.6400 ± 0.0944	0.8088 ± 0.0536	0.9055 ± 0.0328	**0.9311 ± 0.0533**
250	0.5291 ± 0.0318	0.8346 ± 0.0252	0.9057 ± 0.0274	0.9206 ± 0.0191	0.8997 ± 0.0128	0.9493 ± 0.0109	**0.9551 ± 0.0105**
500	0.5408 ± 0.0181	0.8596 ± 0.0130	0.9296 ± 0.0093	0.9433 ± 0.0095	0.9031 ± 0.0101	0.9479 ± 0.0119	**0.9598 ± 0.0122**
1.000	0.5348 ± 0.0213	0.8675 ± 0.0090	0.9437 ± 0.0053	0.9538 ± 0.0067	0.9016 ± 0.0124	0.9602 ± 0.0042	**0.9675 ± 0.0055**
5.000	0.5421 ± 0.0068	0.8843 ± 0.0055	0.9544 ± 0.0024	0.9587 ± 0.0020	0.9110 ± 0.0042	0.9650 ± 0.0015	**0.9739 ± 0.0012**

References

[1] S. Kolouri, S. R. Park, and G. K. R. Rohde: The Radon Cumulative Distribution Transform and Its Application to Image Classification. *IEEE Transactions on Image Processing* **25** (2016).

[2] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister: Sliced and Radon Wasserstein barycenters of measures *Journal of Mathematical Imaging and Vision* **51** (2015).

[3] M. Beckmann, R. Beinert, and J. Bresch: Max-Normalized Radon Cumulative Distribution Transform for Limited Data Classification. *SSVM.* (2025) – arXiv:2411.16282.

research@mbeckmann.de, beinert@math.tu-berlin.de, bresch@math.tu-berlin.de