Note on the John Transform for Object Classification

Jonas Bresch*
Technische Universität Berlin
based on Inspiration during the IPMS at Malta,
inspired by the works with M. Quellmalz, R. Beinert, M. Beckmann

Abstract

The John transform, or also called X-ray transform, is a useful tool for analyzing data in imaging beside the closely related Radon transform. We prove a closed-form expression for the John transform of an axis-aligned rectangular voxel and use the relation of John and Radon transform via one- and multi-stage projection to introduce the John-CDT (J-CDT) and its normalized variant (NJ-CDT). The latter has theoretically provable invariance properties for translated and isotropically scaled measures in arbitrary dimensions. Numerically, a multi-slicing strategy that iteratively applies John projections before CDT produces compact, invariant features that are highly effective in limited-data regimes. Comparisons to related Radon-CDT embeddings, LOT, and momentum-based methods demonstrate competitive or superior accuracy with simple classifiers.

1 Introduction

We study a family of integral transforms and associated feature maps for object and image classification based on one- and multi-stage projections. Classical tomography uses the Radon transform to convert a function or a measure on $\mathbb{R}^d$ into its integrals over hyperplanes; this transform and its computational and stability properties are well understood and form the mathematical background of many imaging modalities [16]. For certain low-data or geometric applications it is natural to replace hyperplane integrals by lower-dimensional projections: the John (or X-ray) transform integrates along parallel lines to a given direction and records the resulting measures on the orthogonal complement.

In this work we exploit the interplay between the Radon and John transforms to build stable, low-dimensional representations that are tailored to classification tasks with limited training data. Furthermore, we derive a closed form of the John transform of a hypercube (Thm. 2), similar to [5: Thm. 1] (cf. Thm. 1), for the numerical realizations. The approach is motivated by recent work on *Radon Cumulative Distribution Transforms* (R-CDT) and their normalized variants. Those are demonstrating

*Webpage and Email
17th July 2026

invariance and competitive empirical performance for limited data image classification [2–4]. Here the study of affine measure classes in the sense that each class originates from one (unique) template measure under affine transformations is considered. The latter are crucial, for instance, in the domains of isotropic position of convex bodies [11; 14; 17], since every full-dimensional log-concave probability measure has a unique isotropic representative, or for John and Löwner positions [1; 10; 12], in statistics and machine learning for *Independent Component Analysis* (ICA) [6; 9], in optimal transport and measure geometry by the standard correspondence in the Wasserstein-2 distance [18; 19] or filigranolology, the study of historic watermarks [8; 13].

Our central idea is to use iterative John projections to reduce the ambient dimension and then apply one-dimensional Radon and CDT techniques on the projected measures. This "multi-slicing" strategy yields a *John-CDT* (J-CDT) and a *Normalized J-CDT* (NJ-CDT) that (i) admit closed-form expressions for simple templates (rectangular voxels), (ii) are equivariant under translations and isotropic scalings, and (iii) produce compact features that are effective in limited-data classification experiments. The analytical relations between affine changes of variables and the induced action on John and Radon transforms that underlie these properties are elementary but, to our knowledge, have not been exploited systematically for design of classification features. We therefore connect classical transform theory [16] with recent optimal-transport-inspired feature maps [2–4; 15; 20] and with moment-based descriptors used in shape recognition [7].

2 Radon and John transform

Let $\mu \in \mathcal{M}(\mathbb{R}^d)$ be a signed, regular, finite measure. We define the slicing map

$$S_{\boldsymbol{\theta}} : \mathbb{R}^d \to \mathbb{R}, \qquad \boldsymbol{x} \mapsto \langle \boldsymbol{x}, \boldsymbol{\theta} \rangle \qquad \text{in direction} \qquad \boldsymbol{\theta} \in \mathbb{S}^{d-1} \coloneqq \{\boldsymbol{x} \in \mathbb{R}^d : \|x\| = 1\}$$

with the Euclidean norm $\|\cdot\|$ on $\mathbb{R}^d$ induced by the standard inner product $\langle \cdot, \cdot \rangle$ and the restricted Radon transform

$$\mathcal{R}_{\boldsymbol{\theta}} : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}), \qquad \mu \mapsto (S_{\boldsymbol{\theta}})_{\#}\mu =: \big(\mathcal{B}(\mathbb{R}) \ni A \to \mu(S_{\boldsymbol{\theta}}^{-1}(A))\big). \tag{1}$$

Here $\mathcal{B}$ denotes the Borel set. By definition, the Radon transform is mass preserving, i.e. it holds $\mathcal{R}_{\boldsymbol{\theta}}[\mu](\mathbb{R}) = \mu(\mathbb{R}^d)$ for any direction $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$. The Radon transform for $\mu \in \mathcal{M}(\mathbb{R}^d)$ may be defined by integrating along all directions $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$. We define the gluing operator

$$\mathcal{I}_{\mathcal{R}} : \mathbb{S}^{d-1} \times \mathbb{R}^d \to \mathbb{S}^{d-1} \times \mathbb{R}, \qquad (\boldsymbol{\theta}, \boldsymbol{x}) \mapsto (\boldsymbol{\theta}, S_{\boldsymbol{\theta}}(\boldsymbol{x}))$$

and the Radon transform by

$$\mathcal{R} : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{S}^{d-1} \times \mathbb{R}), \qquad \mu \mapsto (\mathcal{I}_{\mathcal{R}})_{\#}[u_{\mathbb{S}^{d-1}} \times \mu] =: (u_{\mathbb{S}^{d-1}} \times \mu) \circ (\mathrm{id}, S_{\boldsymbol{\theta}}^{-1}),$$

where $u_{\mathbb{S}^{d-1}}$ denotes the uniform measure on $\mathbb{S}^{d-1}$ and $\mathrm{id} : X \to X$ the identity on arbitrary spaces X.

Remark 1 (Radon transform for functions). The definition in (1) is equivalent to the definition of the Radon transform of functions $f \in \mathrm{L}^1(\mathbb{R}^d)$ given by the integral restricted to the hyperplane, i.e.

$$\mathcal{R}_{\boldsymbol{\theta}}[f] : \mathbb{R} \to \mathbb{R}, \quad t \mapsto \int_{H_{\boldsymbol{\theta}}(t)} f(\boldsymbol{x}) \, \mathrm{d}\mathcal{H}^{d-1}(\boldsymbol{x}),$$

and

$$H_{\boldsymbol{\theta}}(t) \coloneqq \{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{\theta} \rangle = t\} = S_{\boldsymbol{\theta}}^{-1}(t)$$

the $(d-1)$-dimensional hyperplane, where $\mathcal{H}^{d-1}$ denotes the Hausdorff measure on $H_{\boldsymbol{\theta}}(t)$. ∘

The John transform uses the orthogonal projection

$$\pi_{\boldsymbol{\theta}} : \mathbb{R}^d \mapsto \boldsymbol{\theta}^\perp, \quad \boldsymbol{x} \mapsto \boldsymbol{x} - S_{\boldsymbol{\theta}}(\boldsymbol{x})\boldsymbol{\theta} = (\boldsymbol{I}_d - \boldsymbol{\theta}\boldsymbol{\theta}^\top)\boldsymbol{x},$$

where

$$\boldsymbol{\theta}^\perp \coloneqq \{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{\theta} \rangle = 0\} = S_{\boldsymbol{\theta}}^{-1}(0)$$

denotes the $(d-1)$-dimensional space orthogonal to the direction $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$. The restricted John transform, or also called (restricted) X-ray transform, of a measure $\mu \in \mathcal{M}(\mathbb{R}^d)$ is defined by

$$\mathcal{X}_{\boldsymbol{\theta}} : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(\boldsymbol{\theta}^\perp), \qquad \mu \mapsto (\pi_{\boldsymbol{\theta}})_\# \mu =: \big(\mathcal{B}(\boldsymbol{\theta}^\perp) \ni A \mapsto \mu(\pi_{\boldsymbol{\theta}}^{-1}(A))\big). \tag{2}$$

Similar to the Radon transform (1), the John transform (2) is mass preserving, i.e. it holds $\mathcal{X}_{\boldsymbol{\theta}}[\mu](\boldsymbol{\theta}^\perp) = \mu(\mathbb{R}^d)$. In analogy to the definition of the Radon transform, we introduce for the John transform the gluing map

$$\mathcal{I}_{\mathcal{X}} : \mathbb{S}^{d-1} \times \mathbb{R}^d \to \mathbb{S}^{d-1} \times \mathbb{R}^d, \qquad (\boldsymbol{\theta}, \boldsymbol{x}) \mapsto (\boldsymbol{\theta}, \pi_{\boldsymbol{\theta}}(\boldsymbol{x})).$$

Note that the projection is by definition in the multi-dimensional Euclidean space $\mathbb{R}^d$; but for each direction $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ it can be restricted to $\boldsymbol{\theta}^\perp$ to avoid overparameterization. The John transform is defined by

$$\mathcal{X} : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{S}^{d-1} \times^* \mathbb{R}^d), \qquad \mu \mapsto (\mathcal{I}_{\mathcal{X}})_\#(u_{\mathbb{S}^{d-1}} \times \mu) =: (u_{\mathbb{S}^{d-1}} \times \mu) \circ (\mathrm{id}, \pi_{\boldsymbol{\theta}}^{-1}),$$

where $\mathbb{S}^{d-1} \times^* \mathbb{R}^d \coloneqq \{(\boldsymbol{\theta}, \boldsymbol{x}) \in \mathbb{S}^{d-1} \times \mathbb{R}^d : \boldsymbol{x} \in \boldsymbol{\theta}^\perp\}$ reducing the overparametrization; we have $\mathbb{S}^{d-1} \times^* \mathbb{R}^d \simeq \mathbb{S}^{d-1} \times \mathbb{R}^{d-1}$.

Remark 2 (John transform for functions). The definition in (2) is equivalent to the definition for functions $f \in \mathrm{L}^1(\mathbb{R}^d)$ given by the integral restricted to $\mathrm{span}(\boldsymbol{\theta}) \coloneqq \{t \cdot \boldsymbol{\theta} \mid t \in \mathbb{R}\}$, cf. [16], $\mathcal{X}_{\boldsymbol{\theta}}[f] : \boldsymbol{\theta}^\perp \to \mathbb{R}, \boldsymbol{x} \mapsto \int_{\mathbb{R}} f(\boldsymbol{x} + t\boldsymbol{\theta})\ \mathrm{d}t$. ∘

Remark 3 (Correspondence of Radon and John transform). If $d = 2$, the hyperplane $H_{\boldsymbol{\theta}}(t)$ is given be the line $t\boldsymbol{\theta} \oplus \boldsymbol{\theta}^\perp$ and hence the John transform, modulo the overparametrization discussed in Rem. 2 equals the Radon transform. ∘

We can represent the d dimensional Radon transform by applying first the John transform followed by a $(d-1)$ dimensional Radon transform, cf. Fig. 1. Denote the Radon transform on an ℓ-dimensional vector space $X \subset \mathbb{R}^d$ by

$$\mathcal{R}_{\boldsymbol{\theta}}^X : \mathcal{M}(X) \to \mathcal{M}(\mathbb{R}), \qquad \boldsymbol{\theta} \in \mathbb{S}^{\ell-1}(X) \coloneqq \{\boldsymbol{\eta} \in X : \|\boldsymbol{\eta}\| = 1\}.$$

Lemma 1. *Let $\mu \in \mathcal{M}(\mathbb{R}^d)$ and $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$. Then, it holds $\mathcal{R}_{\boldsymbol{\eta}}^{\boldsymbol{\theta}^\perp}[\mathcal{X}_{\boldsymbol{\theta}}[\mu]] = \mathcal{R}_{\boldsymbol{\eta}}[\mu]$ for any $\boldsymbol{\eta} \in \mathbb{S}^{d-2}(\boldsymbol{\theta}^\perp)$.*

Proof. Let $\boldsymbol{\eta} \in \mathbb{S}^{d-2}(\boldsymbol{\theta}^\perp) \subset \mathbb{R}^d$ and $B \in \mathcal{B}(\mathbb{R})$. By definition, we obtain

$$\begin{aligned} \mathcal{R}_{\boldsymbol{\eta}}^{\boldsymbol{\theta}^\perp}[\mathcal{X}_{\boldsymbol{\theta}}[\mu]](B) &= \mathcal{X}_{\boldsymbol{\theta}}[\mu](\{\boldsymbol{y} \in \boldsymbol{\theta}^\perp : \langle \boldsymbol{\eta}, \boldsymbol{y} \rangle \in B\}) \\ &= \mu(\pi_{\boldsymbol{\theta}}^{-1}(\{\boldsymbol{y} \in \boldsymbol{\theta}^\perp : \langle \boldsymbol{\eta}, \boldsymbol{y} \rangle \in B\})). \end{aligned}$$

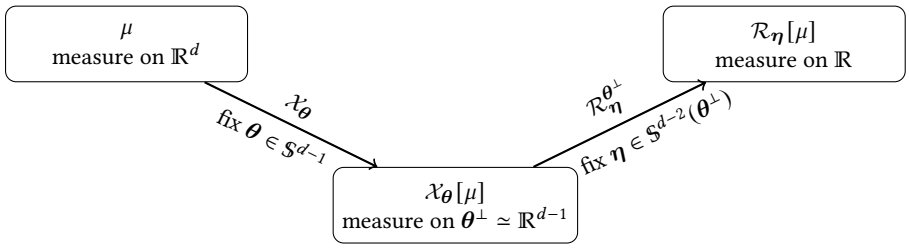

Figure 1: *Visualization of the procedure to capture the (restricted) Radon transform for measure on $\mathbb{R}^d$ via Lem. 1.*

Now, for any $\boldsymbol{x} \in \mathbb{R}^d$ we have $\langle \pi_{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\eta} \rangle = \langle (\boldsymbol{x} - \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle \boldsymbol{\theta}), \boldsymbol{\eta} \rangle = \langle \boldsymbol{x}, \boldsymbol{\eta} \rangle$, since $\langle \boldsymbol{\eta}, \boldsymbol{\theta} \rangle = 0$. Hence, we obtain

$$\begin{aligned} \pi_{\boldsymbol{\theta}}^{-1}(\{\boldsymbol{y} \in \boldsymbol{\theta}^\perp : \langle \boldsymbol{\eta}, \boldsymbol{y} \rangle \in B\}) &= \{\boldsymbol{x} \in \mathbb{R}^d : \pi_{\boldsymbol{\theta}}(\boldsymbol{x}) \in \boldsymbol{\theta}^\perp, \langle \pi_{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\eta} \rangle \in B\} \\ &= \{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{\eta} \rangle = t\} \end{aligned}$$

such that

$$\mathcal{R}_{\boldsymbol{\eta}}^{\boldsymbol{\theta}^\perp}[\mathcal{X}_{\boldsymbol{\theta}}[\mu]](B) = \mu(\{\boldsymbol{x} \in \mathbb{R}^d : S_{\boldsymbol{\eta}}(\boldsymbol{x}) \in B\}) = \mathcal{R}_{\boldsymbol{\eta}}[\mu](B), \quad \forall B \in \mathcal{B}(\mathbb{R}). \qquad \square$$

Corollary 1. *Let $\mu \in \mathcal{M}(\mathbb{R}^d)$ and $\{\boldsymbol{\theta}_0, ..., \boldsymbol{\theta}_{d-1}\} \subset \mathbb{S}^{d-1}$ be an orthonormal basis of $\mathbb{R}^d$. Define $\mathcal{X}_{\boldsymbol{\theta}_1}^{(1)}[\mu] \coloneqq \mathcal{X}_{\boldsymbol{\theta}_1}[\mu]$, and iteratively $\mathcal{X}_{\boldsymbol{\theta}_k}^{(k)}[\mu] \coloneqq \mathcal{X}_{\boldsymbol{\theta}_k}[\mathcal{X}_{\boldsymbol{\theta}_{k-1}}^{(k-1)}[\mu]]$ for $k \in \{2, ..., d-1\}$. Then*

i) $\mathcal{X}_{\boldsymbol{\theta}_k}^{(k)}[\mu]$ is a measure on $\operatorname{span}(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k)^\perp$,

ii) $\mathcal{X}_{\boldsymbol{\theta}_{d-1}}^{(d-1)}[\mu] = \mathcal{R}_{\boldsymbol{\theta}_0}[\mu]$.

Proof. Apply Lem. 1 iteratively and utilize Rem. 3. $\square$

Transformations of uniform measure For $\boldsymbol{a} \in \mathbb{R}^d$, let $u_{\boldsymbol{a}}$ be the Lebesgue measure λ^d restricted to the d-dimensional cube $(-\boldsymbol{a}, \boldsymbol{a}] \coloneqq \times_{i=1}^d (-a_i, a_i]$, i.e. we have $u_{\boldsymbol{a}} = \chi_{\boldsymbol{a}} \lambda^d$ for some density function $\chi_{\boldsymbol{a}}|_{(-\boldsymbol{a},\boldsymbol{a}]} = 1$ and zero elsewhere. Furthermore, we denote by $\boldsymbol{x} \odot \boldsymbol{y}$ the componentwise product in $\mathbb{R}^d$, by $P : \mathbb{R}^d \to \mathbb{R}, \boldsymbol{x} \mapsto \prod_{i=1}^d x_i$ the product of all entries of a vector, and by $\boldsymbol{x}^\circ \coloneqq \boldsymbol{x}|_{\operatorname{supp}(\boldsymbol{x})} \in \mathbb{R}^\ell$ the vector whose i-th component is the i-th non-zero component of $\boldsymbol{x}$, where $\ell = \|\boldsymbol{x}\|_0 \coloneqq |\operatorname{supp}(\boldsymbol{x})|$ denotes the ℓ_0 pseudo norm. In the following, we provide closed forms for the Radon and John transform of a rectangular voxel. This view-point is based on Rem. 1 and Rem. 2 and is a basic first step for numerical computations afterwards. Notably, the following is the density function of $\mathcal{R}_{\boldsymbol{\theta}}[u_{\boldsymbol{a}}]$ w.r.t. the Lebesgue measure λ^1.

Theorem 1 (Radon transform [5: Thm. 1])**.** *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, $t \in \mathbb{R}$, and $\boldsymbol{a} \in \mathbb{R}_{>0}^d$. Furthermore, let $\ell \coloneqq \|\boldsymbol{\theta}\|_0$. It holds that*

$$\mathcal{R}_{\boldsymbol{\theta}}[\chi_{\boldsymbol{a}}](t) = \frac{2^{d-\ell} P(\boldsymbol{a})}{P((\boldsymbol{a} \odot \boldsymbol{\theta})^\circ)\,(\ell-1)!} \sum_{\boldsymbol{k} \in \{-1,1\}^\ell} P(\boldsymbol{k}) \left(t + \langle \boldsymbol{k}, (\boldsymbol{a} \odot \boldsymbol{\theta})^\circ \rangle\right)_+^{\ell-1}, \qquad t \in \mathbb{R}. \quad (3)$$

The following is the density function of $\mathcal{X}_{\boldsymbol{\theta}}[u_{\boldsymbol{a}}]$ w.r.t. the Lebesgue Measure λ^{d-1} i.e. $\lambda^d|_{\boldsymbol{\theta}^\perp}$ restricted to $\boldsymbol{\theta}^\perp \simeq \mathbb{R}^{d-1}$.

Theorem 2 (John transform). *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, $\boldsymbol{x} \in \boldsymbol{\theta}^\perp \subset \mathbb{R}^d$, and $\boldsymbol{a} \in \mathbb{R}^d_{\geq 0}$. It holds that*

$$\mathcal{X}_{\boldsymbol{\theta}}[\chi_{\boldsymbol{a}}](\boldsymbol{x}) = \sum\nolimits_{\boldsymbol{k}\in\{-\boldsymbol{a},\boldsymbol{a}\}} P(\operatorname{sign}(\boldsymbol{k})) \max\Big\{0, \min_{i:\theta_i<0}\{\tfrac{k_i-x_i}{\theta_i}\} - \max_{i:\theta_i>0}\{\tfrac{k_i-x_i}{\theta_i}\}\Big\}.$$

Proof. Denote by $H(x)$ the Heavyside function, which is $H(x) = 0$ for all $x \leq 0$ and $H(x) = 1$ for all $x > 0$. The indicator function satisfies the identity

$$\chi_a(x) = H(x+a) - H(x-a) \qquad \text{for} \quad x \in \mathbb{R} \quad \text{and} \quad a > 0.$$

Therefore, it is

$$\chi_{\boldsymbol{a}}(\boldsymbol{x}) = \prod_{i=1}^d \chi_{a_i}(x_i) = \prod_{i=1}^d H(x+a) - H(x-a) = \sum_{\boldsymbol{k}\in\{-\boldsymbol{a},\boldsymbol{a}\}^d} (-1)^{|\boldsymbol{k}|} \prod_{i=1}^d H(x_i+k_i).$$

By definition, we obtain

$$\mathcal{X}_{\boldsymbol{\theta}}[\chi_{\boldsymbol{a}}](\boldsymbol{x}) = \int_{\mathbb{R}} \chi_{\boldsymbol{a}}(\boldsymbol{x}+t\boldsymbol{\theta})\ \mathrm{d}t = \sum_{\boldsymbol{k}\in\{-\boldsymbol{a},\boldsymbol{a}\}^d} (-1)^{|\boldsymbol{k}|} \int_{\mathbb{R}} \prod_{i=1}^d H(x_i+t\vartheta_i+k_i)\ \mathrm{d}t.$$

Note that the product of $H(\cdot + k_i)$ defines the area, after substituting $\boldsymbol{x} \mapsto \boldsymbol{x} + t\boldsymbol{\theta}$, where

$$\begin{aligned} x_i + t\vartheta_i \geq k_i \quad &\Leftrightarrow \quad \begin{cases} t \geq \frac{k_i-x_i}{\vartheta_i} & : \vartheta_i > 0, \\ t \leq \frac{k_i-x_i}{\vartheta_i} & : \vartheta_i < 0 \end{cases} \\ &\Leftrightarrow \quad \max_{i:\vartheta_i>0}\{\tfrac{k_i-x_i}{\vartheta_i}\} \leq t \leq \min_{i:\vartheta_i<0}\{\tfrac{k_i-x_i}{\vartheta_i}\}, \quad 1 \leq i \leq d. \end{aligned}$$

□

Remark 4. More generally, i.e. for the rectangular cube $(\boldsymbol{a}, \boldsymbol{b}] \coloneqq \times_{i=1}^d (a_i, b_i]$ where $\boldsymbol{a} < \boldsymbol{b}$ componentwise, holds

$$\mathcal{X}_{\boldsymbol{\theta}}[\chi_{\boldsymbol{a},\boldsymbol{b}}](\boldsymbol{x}) = \max\{0, t^{\uparrow}(\boldsymbol{x})_{\boldsymbol{a}}^{\boldsymbol{b}} - t^{\downarrow}(\boldsymbol{x})_{\boldsymbol{a}}^{\boldsymbol{b}}\},$$

where $\chi_{\boldsymbol{a},\boldsymbol{b}}|_{(\boldsymbol{a},\boldsymbol{b}]} \coloneqq 1$, and zero elsewhere with

$$t^{\uparrow}(\boldsymbol{x})_{\boldsymbol{a}}^{\boldsymbol{b}} \coloneqq \min_{i\in\{1,\dots,d\}} t_i^{\uparrow}(\boldsymbol{x})_{\boldsymbol{a}}^{\boldsymbol{b}} \qquad \text{and} \qquad t^{\downarrow}(\boldsymbol{x})_{\boldsymbol{a}}^{\boldsymbol{b}} \coloneqq \max_{i\in\{1,\dots,d\}} t_i^{\downarrow}(\boldsymbol{x})_{\boldsymbol{a}}^{\boldsymbol{b}},$$

and for any $i \in \{1, \dots, d\}$

$$t_i^{\uparrow}(\boldsymbol{x})_{\boldsymbol{a}}^{\boldsymbol{b}} \coloneqq \begin{cases} \max\left\{\frac{a_i-x_i}{\theta_i}, \frac{b_i-x_i}{\theta_i}\right\} & \text{if } \theta_i \neq 0 \\ +\infty, & \text{if } \begin{cases} \theta_i = 0 \text{ and} \\ x_i \in [a_i, b_i), \end{cases} \\ -\infty, & \text{else,} \end{cases}$$

and

$$t_i^{\downarrow}(\boldsymbol{x})_{\boldsymbol{a}}^{\boldsymbol{b}} \coloneqq \begin{cases} \min\left\{\frac{a_i-x_i}{\theta_i}, \frac{b_i-x_i}{\theta_i}\right\} & \text{if } \theta_i \neq 0 \\ -\infty & \text{if } \begin{cases} \theta_i = 0 \text{ and} \\ x_i \in [a_i, b_i), \end{cases} \\ +\infty & \text{else.} \end{cases}$$

∘

Affine measure classes In the manner of [2–4], we define an affinely transformed measure $\mu_{\boldsymbol{A},\boldsymbol{y}}$ in terms of a template measure $\mu \in \mathcal{M}(\mathbb{R}^d)$, a regular matrix $\boldsymbol{A} \in \mathrm{GL}(d)$ and bias $\boldsymbol{y} \in \mathbb{R}^d$ by

$$\mu_{\boldsymbol{A},\boldsymbol{y}} \coloneqq (\boldsymbol{A} \cdot + \boldsymbol{y})_{\#}\mu = \mu(\boldsymbol{A}^{-1}(\cdot - \boldsymbol{y})). \tag{4}$$

It is well known [3: Prop. 3] that the Radon transform of an affinely transformed measure $\mu_{\boldsymbol{A},\boldsymbol{y}}$ has a close correspondence to a Radon transform of the underlying template measure μ:

$$\mathcal{R}_{\boldsymbol{\theta}}[\mu_{\boldsymbol{A},\boldsymbol{y}}] = (\|\boldsymbol{A}^\top\boldsymbol{\theta}\| \cdot + \langle \boldsymbol{y}, \boldsymbol{\theta}\rangle)_{\#}\mathcal{R}_{\frac{\boldsymbol{A}^\top\boldsymbol{\theta}}{\|\boldsymbol{A}^\top\boldsymbol{\theta}\|}}[\mu] = \mathcal{R}_{\frac{\boldsymbol{A}^\top\boldsymbol{\theta}}{\|\boldsymbol{A}^\top\boldsymbol{\theta}\|}}[\mu] \circ \left(\tfrac{\cdot - \langle \boldsymbol{y},\boldsymbol{\theta}\rangle}{\|\boldsymbol{A}^\top\boldsymbol{\theta}\|}\right). \tag{5}$$

The following similar correspondence holds for the John transform.

Lemma 2. *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, $\boldsymbol{A} \in \mathrm{GL}(d)$, $\boldsymbol{y} \in \mathbb{R}^d$ and define $\boldsymbol{\theta}_{\boldsymbol{A}} \coloneqq {}^{\boldsymbol{A}^{-1}\boldsymbol{\theta}}/_{\|\boldsymbol{A}^{-1}\boldsymbol{\theta}\|} \in \mathbb{S}^{d-1}$. Furthermore, let $M_{\boldsymbol{\theta}} \colon \boldsymbol{\theta}_{\boldsymbol{A}}^{\perp} \to \boldsymbol{\theta}^{\perp}, \boldsymbol{x} \mapsto \pi_{\boldsymbol{\theta}}(\boldsymbol{A}\pi_{\boldsymbol{\theta}_{\boldsymbol{A}}}(\boldsymbol{x}))$. Then the John transform satisfies $\mathcal{X}_{\boldsymbol{\theta}}[\mu_{\boldsymbol{A},\boldsymbol{y}}] = (\boldsymbol{z} \mapsto M_{\boldsymbol{\theta}}(\boldsymbol{z}) + \pi_{\boldsymbol{\theta}}(\boldsymbol{y}))_{\#}\mathcal{X}_{\boldsymbol{\theta}_{\boldsymbol{A}}}[\mu]$.*

Proof. We consider the direction $\boldsymbol{\theta}_{\boldsymbol{A}}$ and observe that $\pi_{\boldsymbol{\theta}}\boldsymbol{A}\boldsymbol{x} = 0$ if and only if $\pi_{\boldsymbol{\theta}_{\boldsymbol{A}}}\boldsymbol{x} = 0$. This means that $\ker(\pi_{\boldsymbol{\theta}} \circ (\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x})) = \ker(\pi_{\boldsymbol{\theta}_{\boldsymbol{A}}})$. By definition and linearity, we have for all $\boldsymbol{x} \in \mathbb{R}^d$ that

$$\begin{aligned} M_{\boldsymbol{\theta}}(\boldsymbol{x}) &= \pi_{\boldsymbol{\theta}}(\boldsymbol{A}\pi_{\boldsymbol{\theta}_{\boldsymbol{A}}}(\boldsymbol{x})) = \pi_{\boldsymbol{\theta}}(\boldsymbol{A}(\boldsymbol{x} - \langle \boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{A}}\rangle\boldsymbol{\theta}_{\boldsymbol{A}})) = \pi_{\boldsymbol{\theta}}(\boldsymbol{A}\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{A}}\rangle\pi_{\boldsymbol{\theta}}(\boldsymbol{A}\boldsymbol{\theta}_{\boldsymbol{A}}) \\ &= \pi_{\boldsymbol{\theta}}(\boldsymbol{A}\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{A}}\rangle \tfrac{1}{\|\boldsymbol{A}^{-1}\boldsymbol{\theta}\|}\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \pi_{\boldsymbol{\theta}}(\boldsymbol{A}\boldsymbol{x}), \end{aligned}$$

and hence $M_{\boldsymbol{\theta}} = \pi_{\boldsymbol{\theta}} \circ (\boldsymbol{z} \mapsto \boldsymbol{A}\boldsymbol{z})$. Then, we have by the linearity of $\pi_{\boldsymbol{\theta}}$

$$\begin{aligned} \mathcal{X}_{\boldsymbol{\theta}}[\mu_{\boldsymbol{A},\boldsymbol{y}}] &= (\pi_{\boldsymbol{\theta}})_{\#}[(\boldsymbol{z} \mapsto \boldsymbol{A}\boldsymbol{z} + \boldsymbol{y})_{\#}\mu] \\ &= (\boldsymbol{z} \mapsto \pi_{\boldsymbol{\theta}}(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{y}))_{\#}\mu \\ &= (\boldsymbol{z} \mapsto \boldsymbol{z} + \pi_{\boldsymbol{\theta}}(\boldsymbol{y}))_{\#}[(\boldsymbol{z} \mapsto \pi_{\boldsymbol{\theta}}(\boldsymbol{A}\boldsymbol{z}))_{\#}\mu] \end{aligned}$$

and hence

$$\mathcal{X}_{\boldsymbol{\theta}}[\mu_{\boldsymbol{A},\boldsymbol{y}}] = (\boldsymbol{z} \mapsto \boldsymbol{z} + \pi_{\boldsymbol{\theta}}(\boldsymbol{y}))_{\#}(M_{\boldsymbol{\theta}})_{\#}\underbrace{(\pi_{\boldsymbol{\theta}_{\boldsymbol{A}}})_{\#}\mu}_{=\mathcal{X}_{\boldsymbol{\theta}_{\boldsymbol{A}}}[\mu]} = (\boldsymbol{z} \mapsto M_{\boldsymbol{\theta}}(\boldsymbol{z}) + \pi_{\boldsymbol{\theta}}(\boldsymbol{y}))_{\#}\mathcal{X}_{\boldsymbol{\theta}_{\boldsymbol{A}}}[\mu]. \quad \square$$

Corollary 2 (Special case for $A \in \mathrm{O}(d)$)**.** *If $\boldsymbol{A} \in \mathrm{O}(d)$, which means that $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}_d$, then it holds*

$$\mathcal{X}_{\boldsymbol{\theta}}[\mu_{\boldsymbol{A},\boldsymbol{y}}] = (\boldsymbol{A} \cdot + \pi_{\boldsymbol{\theta}}(\boldsymbol{y}))_{\#}\mathcal{X}_{\boldsymbol{A}^\top\boldsymbol{\theta}}[\mu] = \mathcal{X}_{\boldsymbol{A}^\top\boldsymbol{\theta}}[\mu] \circ \left(\boldsymbol{A}^\top(\cdot - \pi_{\boldsymbol{\theta}}(\boldsymbol{y}))\right).$$

Proof. It holds

$$\begin{aligned} \mathcal{X}_{\boldsymbol{\theta}}[\mu_{\boldsymbol{A},\boldsymbol{y}}] &= (\pi_{\boldsymbol{\theta}})_{\#}[(\boldsymbol{A} \cdot + \boldsymbol{y})_{\#}\mu] = (\pi_{\boldsymbol{\theta}}(\boldsymbol{A} \cdot + \boldsymbol{y}))_{\#}\mu = (\pi_{\boldsymbol{\theta}}(\boldsymbol{A}\cdot) + \pi_{\boldsymbol{\theta}}(\boldsymbol{y}))_{\#}\mu \\ &= (\boldsymbol{A}\pi_{\boldsymbol{A}^\top\boldsymbol{\theta}} \cdot + \pi_{\boldsymbol{\theta}}(\boldsymbol{y}))_{\#}\mu = (\boldsymbol{A} \cdot + \pi_{\boldsymbol{\theta}}(\boldsymbol{y}))_{\#}\mathcal{X}_{\boldsymbol{A}^\top\boldsymbol{\theta}}[\mu], \end{aligned}$$

using that $\|\boldsymbol{A}^\top\boldsymbol{\theta}\| = \|\boldsymbol{\theta}\| = 1$ and

$$\begin{aligned} \pi_{\boldsymbol{\theta}}(\boldsymbol{A}\boldsymbol{x}) &= (\boldsymbol{I}_d + \boldsymbol{\theta}\boldsymbol{\theta}^\top)\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} + \underbrace{\boldsymbol{A}\boldsymbol{A}^\top}_{=\boldsymbol{I}_d}\boldsymbol{\theta}(\boldsymbol{A}^\top\boldsymbol{\theta})^\top\boldsymbol{x} \\ &= \boldsymbol{A}(\boldsymbol{I}_d - \boldsymbol{A}^\top\boldsymbol{\theta}(\boldsymbol{A}^\top\boldsymbol{\theta})^\top)\boldsymbol{x} = \boldsymbol{A}\pi_{\boldsymbol{A}^\top\boldsymbol{\theta}}(\boldsymbol{x}) \end{aligned}$$

for any $\boldsymbol{x} \in \mathbb{R}^d$. $\square$

Multi slicing for multi-dimensional data Motivated by Cor. 1, we aim to combine for applications the John transform twice, i.e. first the John transform with respect to a direction $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ for reducing the dimensions of the directions and afterwards applying the Radon transform on the resulting measure on the orthogonal hyperplane $\operatorname{span}\{\boldsymbol{\theta}\}^\perp \simeq \mathbb{R}^{d-1}$, which has a closed form and is much easier to evaluate than the 3D Radon transform.

Corollary 3 (3D John transform and Projecting Twice)**.** *Let $\{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\} \subset \mathbb{S}^2$ be an orthonormal basis of $\mathbb{R}^3$. Then, for any $\boldsymbol{A} \in \mathrm{GL}(3)$ and $\boldsymbol{y} \in \mathbb{R}^3$, it holds*

$$\mathcal{X}_{\boldsymbol{\theta}_2}[\mathcal{X}_{\boldsymbol{\theta}_1}[\mu_{\boldsymbol{A},\boldsymbol{y}}]] = \mathcal{X}_{\boldsymbol{\eta}_2}[\mathcal{X}_{\boldsymbol{\eta}_1}[\mu]] \circ \left(\tfrac{\cdot - \langle \boldsymbol{y}, \boldsymbol{\theta}_0 \rangle}{\|\boldsymbol{A}^\top \boldsymbol{\theta}_0\|}\right),$$

where

$$\boldsymbol{\eta}_0 \coloneqq \tfrac{\boldsymbol{A}^\top \boldsymbol{\theta}_0}{\boldsymbol{A}^\top \boldsymbol{\theta}_0}, \quad \boldsymbol{\eta}_1 \coloneqq \tfrac{\pi_{\boldsymbol{\eta}_0}(\boldsymbol{\theta}_1)}{\|\pi_{\boldsymbol{\eta}_0}(\boldsymbol{\theta}_1)\|}, \quad \textit{and} \quad \boldsymbol{\eta}_2 \coloneqq \boldsymbol{\eta}_0 \times \boldsymbol{\eta}_1.$$

Proof. Since $\boldsymbol{x} = \sum_{j=0}^2 \boldsymbol{\theta}_j \boldsymbol{\theta}_j^\top \boldsymbol{x} = \sum_{j=0}^2 S_{\boldsymbol{\theta}_j}(\boldsymbol{x}) \boldsymbol{\theta}_j$, we have

$$(\pi_{\boldsymbol{\theta}_1} \circ \pi_{\boldsymbol{\theta}_2})(\boldsymbol{x}) = \big(I_d - \sum\nolimits_{j=1}^2 \boldsymbol{\theta}_j \boldsymbol{\theta}_j^\top\big)\boldsymbol{x} = S_{\boldsymbol{\theta}_0}(\boldsymbol{x})\boldsymbol{\theta}_0, \qquad \forall \boldsymbol{x} \in \mathbb{R}^d.$$

The assertion follows by applying Cor. 1 and (5). □

In the following, we denote $\mathcal{X}_\Theta[\mu] \coloneqq \mathcal{X}_{\boldsymbol{\theta}_{d-1}}[\ldots[\mathcal{X}_{\boldsymbol{\theta}_1}[\mu]]]$ for the iterative application of the John transform from Cor. 1.

Corollary 4. *Let $\boldsymbol{A} \in \{c \cdot \boldsymbol{I}_d : c > 0\}$ and $\boldsymbol{y} \in \mathbb{R}^d$, then $\mathcal{X}_\Theta[\mu_{\boldsymbol{A},\boldsymbol{y}}] = (\|\boldsymbol{A}\boldsymbol{\theta}_0\| \cdot + \langle \boldsymbol{\theta}_0, \boldsymbol{y} \rangle)_\# \mathcal{X}_\Theta$ on $\Theta^\perp = \operatorname{span}(\boldsymbol{\theta}_0)$, where $\{\boldsymbol{\theta}_0\} \cup \Theta \subset \mathbb{S}^{d-1}$ is an orthonormal basis of $\mathbb{R}^d$.*

Proof. Utilize Lem. 2, with $\boldsymbol{\theta}_{\boldsymbol{A}} = \boldsymbol{\theta}$, the idempotenc of $\pi_{\boldsymbol{\theta}_i}$, and the decompositions from the proof of Cor. 3. □

3 Classification in John and Radon CDT spaces

Due to the multi slicing procedure and Cor. 3 and 4, the analogue to the Radon Cumulative Distribution Transform (R-CDT) [3], say John Cumulative Distribution Transform (J-CDT), can be defined via the *cumulative distribution transform* (CDT) $\hat{\nu} \coloneqq F_\nu^{[-1]}$, with the *cumulative distribution function* $F_\nu \coloneqq \nu((-\infty, \cdot])$ for $\nu \in \mathcal{P}(\mathbb{R})$, and the *generalized inverse* $g^{-1} \coloneqq \inf\{s : g(s) > \cdot\}$ applied on $\mathcal{X}_\Theta[\mu] \in \mathcal{P}(\mathbb{R})$ where $\mu \in \mathcal{P}(\mathbb{R}^d)$ by Cor. 1. Notably, it holds

$$\hat{\nu} = \operatorname*{arg\,max}_{T_\# u_{[0,1]} = \nu} \int_0^1 c(s - T(s)) \, \mathrm{d}s, \quad \text{for some convex cost function} \quad c : \mathbb{R} \to [0, \infty),$$

where $T : \mathbb{R} \to \mathbb{R}$ is measurable and $u_{[0,1]}$ denotes the uniform measure on $[0, 1]$. By [4], we have for $\mu \in \mathcal{P}_c^*(\mathbb{R}^d)$, where

$$\mathcal{P}_c^*(\mathbb{R}^d) \coloneqq \big\{\mu \in \mathcal{P}(\mathbb{R}^d) \mid \dim(\operatorname{conv}(\operatorname{supp}(\mu))) > d - 1 \\ \wedge \ \operatorname{conv}(\operatorname{supp}(\mu)) \Subset \mathbb{R}^d\big\} \subset \mathcal{P}_2(\mathbb{R}^2),$$

and $\Subset$ denotes a compact subset, that normalizing yields features, say *Normalized John-CDT* (NJ-CDT), which are invariant under isotropic scaling as well as shifting given by

$$\mathcal{J}_\Theta : \mathcal{P}(\mathbb{R}^d) \to L^\infty([0, 1]), \quad \mu \to (\widehat{\mathcal{X}_\Theta[\mu]} - \operatorname{mean}(\widehat{\mathcal{X}_\Theta[\mu]}))\operatorname{std}(\widehat{\mathcal{X}_\Theta[\mu]})^{-1},$$

where $\operatorname{mean}(g) \coloneqq \int_0^1 g(s) \, \mathrm{d}s$ and $\operatorname{std}(g)^2 \coloneqq \int_0^1 (g(s) - \operatorname{mean}(g))^2 \, \mathrm{d}s$ for $g \in L^\infty([0, 1])$.

Proposition 1. *Let $\boldsymbol{\theta}_0 \in \mathbb{S}^{d-1}$ and $\Theta \subset \mathbb{S}^{d-1}$ such that $\{\boldsymbol{\theta}_0\} \cup \Theta$ is an orthonormal basis of $\mathbb{R}^d$. Further, let $\boldsymbol{A} \in \{\mathrm{diag}(c \cdot \mathbf{1}) : c \in \mathbb{R} \setminus \{0\}\}$ and $\boldsymbol{y} \in \mathbb{R}^d$. Then, it holds $\mathcal{J}_\Theta[\mu] = \mathcal{J}_\Theta[(\boldsymbol{A} \cdot + \boldsymbol{y})_\# \mu]$.*

Proof. The well-definitness of the NJ-CDT is given by the continuity of $\boldsymbol{\theta} \to \mathrm{mean}(\widehat{\mathcal{X}_\Theta[\mu]})$ and $\boldsymbol{\theta} \to \mathrm{std}(\widehat{\mathcal{X}_\Theta[\mu]})$, cf. [3: Lem. 1]. Furthermore, for $\mu \in \mathcal{P}_c^*(\mathbb{R}^d)$ it is $\mathrm{std}(\widehat{\mathcal{X}_\Theta[\mu]}) \geq c$ for some $c > 0$ for any $\Theta \subset \mathbb{S}^{d-1}$ with $|\Theta| = d - 1$. Translating Lem. 2, i.e. Cor. 4, into the CDT space, we obtain

$$\widehat{\mathcal{X}_\Theta[\mu_{\boldsymbol{A},\boldsymbol{y}}]}(t) = \|\boldsymbol{A}\boldsymbol{\theta}_0\| \widehat{\mathcal{X}_\Theta[\mu]}(t) + \|\textstyle\prod_{i=1}^{d-1} \pi_{\boldsymbol{\theta}_i} \boldsymbol{y}\| = c_{\boldsymbol{A}} \widehat{\mathcal{X}_\Theta[\mu]}(t) + \langle \boldsymbol{\theta}_0, \boldsymbol{y} \rangle, \qquad t \in \mathbb{R},$$

where $\boldsymbol{A} = c_{\boldsymbol{A}} \boldsymbol{I}_d$, such that

$$\mathrm{mean}(\widehat{\mathcal{X}_\Theta[\mu_{\boldsymbol{A},\boldsymbol{y}}]}) = \|\boldsymbol{A}\boldsymbol{\theta}_0\| \mathrm{mean}(\widehat{\mathcal{X}_\Theta[\mu]}) + \langle \boldsymbol{\theta}_0, \boldsymbol{y} \rangle,$$

and

$$\mathrm{std}(\widehat{\mathcal{X}_\Theta[\mu_{\boldsymbol{A},\boldsymbol{y}}]}) = \|\boldsymbol{A}\boldsymbol{\theta}_0\| \mathrm{std}(\widehat{\mathcal{X}_\Theta[\mu]}) + \langle \boldsymbol{\theta}_0, \boldsymbol{y} \rangle,$$

and hence $\mathcal{J}_\Theta[\mu_{\boldsymbol{A},\boldsymbol{y}}] = \mathcal{J}_\Theta[\mu]$, which yields the assertion. □

We directly obtain the (linear) separability of arbitrary finitely many measure classes of the form

$$\mathbb{C}_i \coloneqq \{(c \cdot \boldsymbol{I}_d \cdot + \boldsymbol{y})_\# \mu_i \mid c \in \mathbb{R} \setminus \{0\}, \boldsymbol{y} \in \mathbb{R}^d\}, \quad \mu_i \in \mathcal{P}_c^*(\mathbb{R}^d), \quad \text{with} \quad \mathcal{J}_\Theta[\mu_i] \neq \mathcal{J}_\Theta[\mu_j],$$

for $i \neq j$.

Remark 5 (Normalized Radon-CDT). For arbitrary $\boldsymbol{A} \in \mathrm{GL}(d)$ and $\boldsymbol{y} \in \mathbb{R}^d$, the so-called *Normalized Radon-CDT* (NR-CDT) [2] provides features

$$\mathcal{N}_{\boldsymbol{\theta}_0}[(\boldsymbol{A} \cdot + \boldsymbol{y})_\# \mu] = \mathcal{J}_\Theta[(\boldsymbol{A} \cdot + \boldsymbol{y})_\# \mu] = \mathcal{N}_{\boldsymbol{A}\boldsymbol{\theta}_0 / \|\boldsymbol{A}\boldsymbol{\theta}_0\|}[\mu]$$

and resorting of the angles—by taking the maximum—yields the assertion [2; 3] as

$$\mathcal{N}_\mathrm{m}[\mu](t) \coloneqq \sup_{\boldsymbol{\theta} \in \mathbb{S}^d} \mathcal{N}_{\boldsymbol{\theta}}[\mu](t), \quad t \in \mathbb{R},$$

such that $\mathcal{N}_\mathrm{m}[(\boldsymbol{A} \cdot + \boldsymbol{y})_\# \mu] = \mathcal{N}_\mathrm{m}[\mu]$. ○

4 Numerical results

We start with simple classification techniques for data recognition in limited data regimes. For numerical evidence of the theoretical results we consider a 2D and 3D dataset of template images and objects which become translated and isotropically scaled to fulfill the conditions of Prop. 1, similar to [15; 20]. We further add rotations and anisotropic scaling violating the theoretical assumptions to study the robustness of the feature extractors.

We implemented the multidimensional discrete John transform and the experiments in Julia.[1] They are performed on a MacBookPro 2020 with an Intel Core i5 CPU (4 cores, 1.4 GHz) and 8 GB RAM.

[1] The code is available at https://github.com/JJEWBresch/3dxNRCDT.

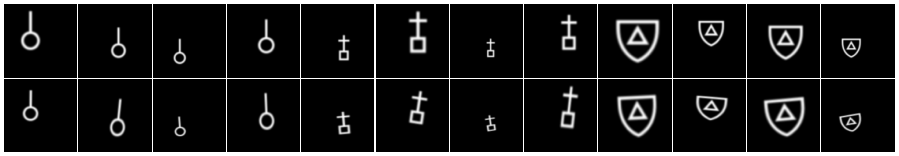

Figure 2: *The 2D acdemic datset. i): Idealized setting; ii): Non-idealized setting (with additional rotating and anisotropic scaling).*

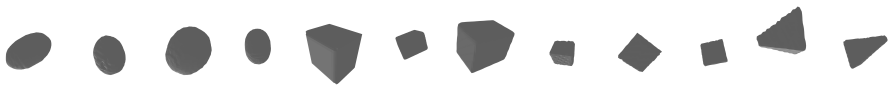

Figure 3: *The 3D academic dataset. Non-idealized setting (with additional rotating and anisotropic scaling).*

Image classification We consider a dataset of three template images of size 128×128 which get randomly isotropically scaled within a range of $[0.5, 1.5]$ and translated by pixel positions within the range $[-20, 20]$, see for visualization Fig. 2. Afterwards the John transform is applied with a random angle $\boldsymbol{\theta}_0 \in \mathbb{S}^1$, uniformly chosen from the circle, with a uniform grid of size 128 in the tangent space $\mathbb{R} \hookleftarrow (\boldsymbol{\theta}_0)^\perp \subset \mathbb{R}^2$ inside the box (centered image) of (pixel)size $[-64, 64]^2$. The generalized inverse is calculated by applying a linear interpolation on a uniform grid of size 128.

The resulting features $\mathcal{J}_{\boldsymbol{\theta}_0^\perp}$ can be compared: here, we, firstly, use the underlying three template measures as the reference to calculate the nearest template for the NJ-CDT transformations of each sample, see Tab. 1 (second columns, respectively); secondly, we only use the NJ-CDT transformations of the samples and perform a nearest neighbor calculation, with an random split of $\{1, 3, 5\}$ samples out of 10 for each class for training, against $\{9, 7, 5\}$, respectively, for testing, and repeating it 20 times. The means and standard deviations are reported in Tab. 1 (third to fives columns, respectively, top). For both examples, we utilize different norms in the feature space for calculating the distances. The same experiments are performed on the non-idealized setting, i.e. Fig. 2 ii). As the assumptions in Prop. 1 are similar to the once on the *linear OT* (LOT) [15; 20], a comparison with our proposed feature extractor is provided in Tab. 1 (bottom).

We observe perfect classification results on the idealized dataset, and fairly good accuracies (up to 93%) in the non-idealized setting substantiating significant robustness. In both settings, the LOT embedding becomes clearly outperformed.

Object recognition In the second part, we consider the task of 3D shape recognition. Here, the experiments consits of one academic dataset based three synthetic

Table 1: Top: *NT and 1-NN calculation on $\mathcal{J}_{\boldsymbol{\theta}_0^\perp}$ embedding, where $\boldsymbol{\theta}_0 \in \mathbb{S}^1$ is randomly sampled, uniformly.* Bottom: *NT and 1-NN calculation on LOT embedding [15; 20].*

dataset	Fig. 2 i)				Fig. 2 ii)			
method	NT	1-NN (10%)	1-NN (33%)	1-NN (50%)	NT	1-NN (10%)	1-NN (33%)	1-NN (50%)
$\|\cdot\|_1$	**1.0000**	**1.0000±0.0000**	**1.0000±0.0000**	**1.0000±0.0000**	0.8333	0.7241±0.1246	0.8881±0.0864	0.9434±0.0817
$\|\cdot\|_2$	**1.0000**	**1.0000±0.0000**	**1.0000±0.0000**	**1.0000±0.0000**	0.8333	0.7074±0.1529	0.8905±0.0726	0.9100±0.0845
$\|\cdot\|_\infty$	**1.0000**	**1.0000±0.0000**	**1.0000±0.0000**	**1.0000±0.0000**	**0.9333**	**0.8259±0.1147**	**0.9190±0.0466**	**0.9434±0.0622**
$\|\cdot\|_1$	0.5660	0.5324±0.1390	0.7380±0.0780	0.8130±0.0932	0.5000	0.5314±0.1122	0.5833±0.0858	0.6266±0.0687
$\|\cdot\|_2$	0.5377	0.5574±0.1198	0.7119±0.1064	0.7966±0.1091	0.5333	0.5351±0.1268	0.5833±0.0999	0.6495±0.1121
$\|\cdot\|_\infty$	0.5000	0.4570±0.0949	0.4328±0.0983	0.5566±0.1150	0.4666	0.4370±0.0896	0.4309±0.0920	0.5300±0.0929

Table 2: Left: *1-NN calculation for the non-idealized dataset on* $\mathcal{J}_{(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^\perp}$*, where* $\boldsymbol{\theta}_1 \in \mathbb{S}^2$ *and* $\boldsymbol{\theta}_2 \in \mathbb{S}^1(\boldsymbol{\theta}_1^\perp)$ *is randomly sampled.* Middle: *NT and 1-NN calculation on* $\mathcal{N}_{\mathrm{m}}$ *calculated by* $\mathcal{R}_\Theta[\mathcal{X}_{\boldsymbol{\theta}_1}]$ *where* $\Theta = \mathbb{S}^1(\boldsymbol{\theta}_1^\perp)$ *and* $\boldsymbol{\theta}_1 \in \mathbb{S}^2$ *is randomly chosen.* Right: *NT and 1-NN calculation with momentum based features on* $\mathcal{X}_{\boldsymbol{\theta}_1^\perp}$.

	method	1-NN (20%)	1-NN (40%)	1-NN (60%)	1-NN (20%)	1-NN (40%)	1-NN (60%)	1-NN (20%)	1-NN (40%)	1-NN (60%)
Fig. 3	$\|\cdot\|_1$	0.6500±0.1420	0.7222±0.2555	0.8333±0.1529	0.9541±0.0425	0.9722±0.0493	0.9666±0.0418	0.4916±0.1642	0.6333±0.1616	0.6333±0.1999
	$\|\cdot\|_2$	0.6375±0.1124	0.8222±0.1503	0.8750±0.1417	**0.9208±0.0632**	**0.9833±0.0407**	**1.0000±0.0000**	0.5208±0.1452	0.6388±0.1833	0.7083±0.1863
	$\|\cdot\|_\infty$	**0.7875±0.0875**	**0.8500±0.1263**	**0.9000±0.0837**	0.9208±0.0875	0.9722±0.0493	0.9916±0.0372	0.5125±0.1123	0.5055±0.1549	0.7983±0.2529

template objects (sphere, cube, pyramid). First it is generated in an idealized way, i.e. we randomly scaled the object, isotropically, within the range of $[0.5, 1.5]$ and translated in all three directions within the range of $[-20, 20]$. Additionally, we consider rotations within $[-180°, 180°]$ for all Eulerian angles, and add anisotropic scaling effect within the range of $[0.75, 1.25]$.

For classification, we use, as in the previous paragraph, 1-NN calculation, firstly, on the NJ-CDT, via the John transform w.r.t. the angle pairs $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{S}^2 \times \mathbb{S}^1(\boldsymbol{\theta}_1^\perp)$, which corresponds to $\mathcal{R}_{\boldsymbol{\theta}_0}$ in the manner of Cor. 1, and accuracy results in Tab. 2 (left block). Secondly, we perform the same techniques on the $_{\mathrm{m}}$NR-CDT via the coupling from Cor. 1; the accuracy results are reported in Tab. 2 (middle block) Furthermore, we combine the John transform $\mathcal{X}_{\boldsymbol{\theta}_1}$ with the momentum-based method from [7]; the accuracy results are reported in Tab. 2 (right block).

Theoretically, for the moment-based feature extractors on the measures $\mathcal{X}_{\boldsymbol{\theta}_1}[\mu] \in \mathcal{P}_c^*(\mathbb{R}^2)$ the setting would be idealized, at least for the idealized dataset, since isotropic scaling and shifting result into affine transformations of the measures after applying the John transform once; however, we observe accuracies up to 71%, and the other proposed combinations yield perfect classification results in all settings. In the non-idealized setting, cf. Fig. 3, slight uncertainties can be observed, cf. Tab. 2, which can be significantly reduced via the NN calculation with an appropriate rate of test samples.

References

[1] Ball, K., "An elementary introduction to modern convex geometry", in S. Levy (ed.), *Flavors of Geometry*, xxxi (MSRI Publications; Cambridge University Press, 1997), 1–58.

[2] Beckmann, M., Beinert, R., and Bresch, J., *Generalizations of the Normalized Radon Cumulative Distribution Transform for Limited Data Recognition*, arXiv:2512.08099, 2025. doi: 10.48550/arXiv.2512.08099.

[3] —— "Max-Normalized Radon Cumulative Distribution Transform for Limited Data Classification", in *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)* (Springer, 2025), 241–54. doi: 10.1007/978-3-031-92366-1_19.

[4] —— "Normalized Radon Cumulative Distribution Transforms for Invariance and Robustness in Optimal Transport Based Image Classification", *SIAM Journal on Imaging Sciences*, 19/2 (2026), 677–709. doi: 10.1137/25M1767741.

[5] Beinert, R., Bresch, J., and Quellmalz, M., *A Discrete Radon Transform Based on the Area of Cube-Plane Intersection*, arXiv:2603.13503, 2026. doi: 10.48550/arXiv.2603.13503.

[6] Comon, P., "Independent component analysis, a new concept?", *Signal Processing*, 36 (1994), 287–314.

[7] Flusser, J., and Suk, T., "Pattern recognition by affine moment invariants", *Pattern Recognition*, 26/1 (1993), 167–74. doi: 10.1016/0031-3203(93)90098-H.

[8] Hauser, D., Beckmann, M., Koliander, G., and Stiehl, H. S., "On Image Processing and Pattern Recognition for Thermograms of Watermarks in Manuscripts – A First Proof-of-Concept", in *International Conference on Document Analysis and Recognition (ICDAR)* (Springer, 2024), 91–107. doi: 10.1007/978-3-031-70543-4_6.

[9] Hyvärinen, A., Karhunen, J., and Oja, E., *Independent Component Analysis* (Wiley, 2001).

[10] John, F., "Extremum problems with inequalities as subsidiary conditions", in *Studies and Essays Presented to R. Courant on His 60th Birthday* (Interscience, 1948), 187–204.

[11] Klartag, B., "On convex perturbations with a bounded isotropic constant", *Geom. Funct. Anal.* 16 (2006), 1274–90.

[12] Löwner, K., "Über monotone Matrixfunktionen", *Mathematische Zeitschrift*, 38 (1934), 177–216.

[13] Mazzola, R., and Moschini, U., "Automatic watermark recognition and classification", *Digital Scholarship in the Humanities* (2012).

[14] Milman, V. D., and Pajor, A., "Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space", in J. Lindenstrauss and V. Milman (eds.), *GAFA Seminar 1987–89*, mccclxxvi (Lecture Notes in Mathematics; Springer, 2006), 64–104.

[15] Moosmüller, C., and Cloninger, A., "Linear optimal transport embedding: provable Wasserstein classification for certain rigid transformations and perturbations", *Information and Inference: A Journal of the IMA* 12/1 (2023), 363–89. doi: 10.1093/imaiai/iaac023.

[16] Natterer, F., and Wübbeling, F., *Mathematical Methods in Image Reconstruction* (Philadelphia, PA: SIAM, 2001). doi: 10.1137/1.9780898718324.

[17] Vershynin, R., *High-Dimensional Probability* (Cambridge University Press, 2018).

[18] Villani, C., *Optimal Transport: Old and New* (Springer, 2009).

[19] —— *Topics in Optimal Transportation* (American Mathematical Society, 2003).

[20] Wang, W., Slepčev, D., Basu, S., Ozolek, J. A., and Rohde, G. K., "A linear optimal transportation framework for quantifying and visualizing variations in sets of images", *International Journal of Computer Vision*, 101/2 (2013), 254–69. doi: 10.1007/s11263-012-0566-z.