PARTITIONING SIGNAL CLASSES USING TRANSPORT TRANSFORMS FOR DATA ANALYSIS AND MACHINE LEARNING

AKRAM ALDROUBI, SHIYING LI, GUSTAVO K. ROHDE

ABSTRACT. A relatively new set of transport-based transforms (CDT, R-CDT, LOT) have shown their strength and great potential in various image and data processing tasks such as parametric signal estimation, classification, cancer detection among many others. It is hence worthwhile to elucidate some of the mathematical properties that explain the successes of these transforms when they are used as tools in data analysis, signal processing or data classification. In particular, we give conditions under which classes of signals that are created by algebraic generative models are transformed into convex sets by the transport transforms. Such convexification of the classes simplify the classification and other data analysis and processing problems when viewed in the transform domain. More specifically, we study the extent and limitation of the convexification ability of these transforms under an algebraic generative modeling framework. We hope that this paper will serve as an introduction to these transforms and will encourage mathematicians and other researchers to further explore the theoretical underpinnings and algorithmic tools that will help understand the successes of these transforms and lay the groundwork for further successful applications.

1. Introduction

Recently a set of transforms that have close relationship to the mathematics of optimal transport [13, 8, 3, 26] have been introduced for representing signals, images and data. In this framework, a signal $p$ is a normalized non-negative function, while its transform $\hat{p}$ is a function which is consistent with an optimal transport map. Succinctly, the transform $\hat{p}$ of a function $p$ is the transport map that morphs (transports) a chosen reference function $r$ to $p$. This general idea has led to the development of several types of non-linear transforms for signal and image data, including the Cumulative Distribution Transform (CDT) [17], the Radon CDT [9], and the Linear Optimal Transport (LOT) [27, 11] transforms. Numerous problems in signal and image analysis, and in machine learning have been successfully and efficiently solved in the transport transform domains [10]. At their core, the aforementioned transforms capture the movement (transport) of the signal amplitude (or pixel intensity) along the independent variable (usually time for 1-d signals or space for images) in a mathematically rigorous manner. This transport encoding of signal or image intensity has enabled numerous interesting applications, many of which were not possible with other, more standard, techniques. The transport-based transforms have been successfully applied

2010 Mathematics Subject Classification.

Key words and phrases. convexity, transport-transforms, convex groups, data analysis, classification, machine learning.
to many data analysis problems including parametric signal estimation \cite{rubaiyat2010} (Rubaiyat et al., 2010), signal and image classification \cite{park2017, kolouri2016, kolouri2016}, modeling turbulence \cite{emerson2020} (Emerson et al., 2020), cancer detection \cite{ozolek2014, tosun2015} (Ozolek et al., 2014; Tosun et al., 2015), vehicle-type recognition \cite{guan2019} (Guan et al., 2019), and others \cite{kundu2018}. Although not based on the transport transforms, transport maps have also been used in many other successful applications of data analysis such as \cite{7, 20, 14, 23, 1, 22}.

In the transport transform framework, the set of signals $P_d$ (see (1) below) consists of non-negative functions. Except for the zero signal, they are normalized to have $L_1$-norm equal to 1. For signal analysis, processing or classification problems, the signals are modeled by deformations of template signals. For example, a smiling face-image $p_h$ of an individual can be thought of (approximately) as a deformation of a neutral face-image $p$ of the same individual via the diffeomorphism $h$, see Figure 2. A class of signals $S_{p,H}$ can be generated by applying a set of diffeomorphisms $H$ to a signal $p$. The process of generating such class from $H$ will be called an algebraic generative model. This is different from the statistical generative model in machine learning which assumes that the data is generated from an underlying conditional probability distribution, and the aim is to find this distribution. In contrast, the algebraic generative model creates signal classes from a template and a set of diffeomorphisms that produce the class by morphing the template (no probabilistic assumptions about the classes are used). Under this algebraic generative modeling assumption, the transport transforms can be particularly useful for applications.

Data analysis, and machine learning methods when applied in the transport transform domain work best if the data has formed via a transport type phenomena. For example, take the task of modeling the difference between images of smiling faces versus faces with neutral expression as in Figure 2. Building a mathematical model to automatically recognize the difference between these two classes is still a challenging problem, and learning a classifier in raw image domain is a difficult task. However, a smiling face is formed by muscle mass movement (transport) from a neutral face. Thus, in transport transform domain (LOT), this problem becomes simple as shown in Figure 2. The two underlying classes (smiling vs neutral) seem to cluster into two disjoint convex sets. Thus classification can be easily achieved using the Hyperplane Separation Theorem (or the Hahn-Banach Separation Theorem). In addition, because the transform is invertible, any point in transform space can be inverted (see caveats below), and thus any point along the estimated classifier line (i.e., the line orthogonal to the separating hyperplane in the transform domain) can be visualized in image space. As can be seen from the inverse of the linear classifier, the technique is able to correctly summarize the differences between the two classes by providing an “average” face that goes from neutral to smiling as one traverses along the classifier line. This idea of utilizing transport-based transforms to perform morphometry operations was first introduced in \cite{27, 2} where the goal was to combine learning techniques with transport-based representation techniques to decode important trends in a given dataset.
As in the previous example, it has been demonstrated \cite{16, 9} that one of the main advantages of solving estimation and detection (i.e. classification) problems in transport transform (CDT, R-CDT, LOT) domain relates to rendering signal and image classes convex. The idea is outlined in Figure 1. Two classes in signal space $\mathbb{P}$ and $\mathbb{Q}$ are displayed. They consist of signals containing “one bump” and “two bumps”, respectively, observed under random translation along the independent variable. The right panel shows the same classes expressed in transport transform domain. It is clear the act of transforming the set of 1D signals has rendered the problem easier to solve, given the ensuing manifold (the geometric structure encompassing the transformed data) becomes linear in this example. Recently, Shifat-E-Rabbi et al. \cite{24}, have made use of such properties to propose a transport transform nearest subspace method for image classification. Empirical tests show the method, which is simple to compute and does not require iterative tuning of hyperparameters, can rival that of popular neural network type classifiers in certain problems, for a small fraction of the computational cost and with fewer training samples.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{example_signal_classification.pdf}
\caption{An example of a signal classification problem. In this example, class 1 consists of random translations of a “one bump” signal and class 2 consists of random translations of a “two bump” signal.}
\end{figure}

1.1. Contributions. The ability of the emerging transport-based transforms to facilitate the solution of estimation, classification, data processing and analysis problems is related to the extent to which they are able to partition the data into convex sets. As such, this manuscript is devoted to clarifying the ability of different transport-based transforms to render the classes produced by algebraic generative models into convex sets. More precisely, we specify conditions which will render signal or image classes convex in transform domain.

1.2. Paper organization. The rest of the paper is organised as follows. In Section 2, the transport-based transforms, their connections to the optimal transport theory and an associated generative model are introduced. In Section 3, convexification results (Proposition 3.1, Theorem 3.2 and corollaries) and examples of the CDT for one dimensional generative models are presented. In Section 4, we present the limitations on the generative model with respect to convexification by the LOT in dimension $d \geq 2$ (Theorem 4.2) and a possible relaxation to mitigate the limitations in dimension two.
Figure 2. Automated modeling differences between smiling and neutral facial expressions using transport transforms. Top left: sample images of both neutral and smiling classes. Top right: depiction of data and a linear classifier (line in red) in transform domain. Bottom: given that the transforms are invertible, the inverse of the classifier can be obtained. As shown by the inverse transform of the classifier line, the classifier consists of an “average” image, while different directions along the discriminant line (in red) represent how wide a person smiles. The technique is able to correctly summarize the differences between the two classes by providing an “average” face that goes from neutral to smiling as one traverses along the classifier line.

(Theorem 4.8). A more detailed summary of results and a discussion of open questions are given in Section 5.

2. Preliminaries, Notation and Model Assumptions

In this section, we define the various transforms, their domains, ranges, and their connections to optimal transport theory, and introduce a signal model of interest.

2.0.1. Signal and Transform spaces. The signal spaces we consider consist of non-negative Lebesgue measurable functions that are compactly supported and normalized, i.e., a non-zero signal $p$ is first normalized to have its $L_1$-norm $\|p\|_1 = 1$. The space of all signals is formally described by

$$\mathcal{P}_d := \{ p : \mathbb{R}^d \to \mathbb{R}_+ \mid \text{supp}(p) = \Omega_p \text{ is compact}, \int_{\mathbb{R}^d} p(x)dx = 1 \}.$$


Note that the signals in the above set have finite moments since they are compactly supported. Since signals in $P_d$ have $L_1$-norm equal to 1, it is sometimes useful to think of them as probability density functions. The class of transforms $\hat{P}_d$ belongs to the set of functions $F_d$ that are solutions to the Monge Transport Problem discussed in Subsection 2.1 below:

$$F_d := \{ f : \mathbb{R}^d \to \mathbb{R}^d \mid f = \nabla \phi \text{ for some convex } \phi : \mathbb{R}^d \to \mathbb{R} \}.$$  

In particular, $F_1$ consists of all a.e. non-decreasing functions from $\mathbb{R}$ to $\mathbb{R}$.

### 2.1. Transforms and the Monge problem.

The general theory of optimal transport is a deep and well-developed area that started with a transport problem due to Monge [13]. The theory and solutions to this problem took many detours and gave rise to a general theory of optimal transports which was spearheaded and developed by Kantorovich, Brenier, Villani and many others in the last two hundred years [8, 3, 26, 21]. For our purpose we use one of the simplest results of the theory to introduce the needed transforms.

Given a fixed reference function $r \in P_d$, the transform $\hat{p}$ of $p \in P_d$ is the unique solution to the Monge optimal transport problem:

$$\text{minimize } \mathcal{M}(T) = \int_{\mathbb{R}^d} |x - T(x)|^2 r(x) dx$$

over all $T$ that makes the push-forward relation below hold:

$$\int_B p(y) dy = \int_{T^{-1}(B)} r(x) dx,$$

holds for every measurable set $B$. In measure theory, the relation (4) above is written more compactly as

$$\mu_p = T_# \mu_r,$$

where $d\mu_p = p dx$, $d\mu_r = r dx$. When $T$ is a $C^1$ diffeomorphism, the constraint above becomes

$$r(x) = |\det (J_T)| p(T(x)),$$

where $J_T$ is the Jacobian matrix of $T$.

The existence and uniqueness of solutions to the Monge problem under the assumptions of the signal spaces $P_d$ in this paper is a special case of the well-known Brenier’s Theorem described below. The function $T$ in Equation (5) can be interpreted as a mass preserving map between the reference signal $r$ (a reference mass distribution) and the given signal $p$ (a mass distribution described by $p$). With this point of view, Monge’s Problem can then be interpreted as finding the optimal transport map that will transform a mass distribution to another distribution of equal mass.

Let $r \in P_d$ and $T_r : P_d \to F_d$ be the operator that maps an element $p \in P_d$ to its optimal transport map described above. Under the assumptions on $P_d$, we define the transport transform of a function $p$ with respect to the reference $r$ as

$$\hat{p} = T_r(p).$$
Definition 2.1 (CDT, LOT, and R-CDT).

(1) when \( d = 1 \) in (7), \( T_r \) is called the CDT.
(2) when \( d \geq 2 \), \( T_r \) is called the Linear Optimal Transport (LOT) transforms (although the transform itself is nonlinear).
(3) The R-CDT consists of the composition of the Radon Transform and the CDT transform.

Remark 2.2. In one dimension, as mentioned above, the optimal transport maps are non-decreasing functions defined on \( \mathbb{R} \). In particular, given \( p \in \mathcal{P}_1 \), its CDT transform \( \hat{p} \) with respect to a reference \( r \) can be equivalently defined through the following relation

\[
\hat{p}(x) = \sup \{ t \mid \int_{-\infty}^t p(\xi) d\xi \leq \int_{-\infty}^x r(\xi) d\xi \},
\]

If \( \text{supp}(p) \) is an interval and \( p \) is continuous on \( \text{supp}(p) \), then (8) simplifies to

\[
\int_{-\infty}^{\hat{p}(x)} p(\xi) d\xi = \int_{-\infty}^x r(\xi) d\xi, \quad \forall x \in \text{supp}(r).
\]

The equivalent formulation (9) can be derived by integrating Equation (6), replacing \( T \) by \( \hat{p} \), and making use of the property that \( \hat{p} \) is increasing a.e. \( \mu_r \).

Under the assumption on the signals spaces, all the transforms above are non-linear and are injective a.e. \( \mu_r \). See Figure 3 for examples of CDTs with the reference \( r \) being the characteristic function on \( [0, 1] \).

![Figure 3. CDT examples with \( r = \chi_{[0,1]} \). Left panel: the reference function \( r \) (red), a Gaussian like signal (black), and the sum of two characteristic functions (blue). Right panel: the transform of the Gaussian like function (black) and the transform of the sum of two characteristic functions (blue).](image)

A summary of the various spaces and symbols is provided in Table 1.

---

\(^1\)Note that \( \hat{p}(x) \) defined in (8) is finite \( \mu_r \) a.e., but could be \( +\infty \) for some \( x \).
Table 1. Notation

| Symbols | Description |
|---------|-------------|
| $p$, $q$, $r$ | normalized functions on $\mathbb{R}^d$ |
| $\Omega_p$ | Support of a function $p$ |
| $\mathcal{P}_d$ | the set of compactly supported normalized functions |
| $\mathcal{F}_d$ | the set of optimal transport maps from $\mathbb{R}^d$ to $\mathbb{R}^d$ |
| $\mathcal{F}_G^d$ | the set of $C^1$ diffeomorphisms in $\mathcal{F}_d$ |
| $\mathcal{H}$ | a subgroup of $\mathcal{F}_G^d$ |
| $\mathcal{H}_a$ | the group of translation and isotropic scaling diffeomorphisms on $\mathbb{R}^d$ |
| $\mathcal{S}_{p,\mathcal{H}}$ | a signal class generated by template $p$ under the diffeorphisms in $\mathcal{H}$ |
| $\hat{\mathcal{S}}_{p,\mathcal{H}}$ | the transformed signal class $\{\hat{p}_h \mid p_h \in \mathcal{S}_{p,\mathcal{H}}\}$ |
| $T_r$ | the CDT/LOT transform operator: $T_r(p) = \hat{p}$ |

2.1.1. Connection with the Wasserstein distance. Given two probability measures $\mu$ and $\nu$, the Wasserstein-2 distance between them is defined as

$$W_2(\mu, \nu) := \left( \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi(x, y) \right)^{\frac{1}{2}},$$

where $\Pi(\mu, \nu)$ is the set of measures on $\mathbb{R}^d \times \mathbb{R}^d$ with $\mu$ and $\nu$ as marginals [26]. In the one dimensional case, one can show that the CDT transform defines an embedding from $\mathcal{P}_1$ with the $W_2$-metric to the transformed space $\hat{\mathcal{P}}_1$ [17]. In particular, $W_2(\mu_p, \mu_q) = ||(\hat{p} - \hat{q})\sqrt{T}||_{L^2}$ for any $p, q \in \mathcal{P}_1$. However, when $d \geq 2$, this embedding property does not hold in general. The Euclidean-type distance $|| (\hat{p} - \hat{q}) \sqrt{T} ||_{L^2}$ is referred as the linearized optimal transport (LOT) between $p$ and $q$ when $d \geq 2$ and has been shown useful in image pattern recognition, discrimination and visualization problems [27, 11].

2.2. Optimal transport maps. We start with a special case of Brenier’s Theorem (see e.g., [3, 21, 26] and the references therein) which we need for this investigation. In particular, using Theorem 2.12 in [26] and Theorem 1.48 in [21] we have

**Theorem 2.3** (Brenier’s Theorem). Let $p, r \in \mathcal{P}_d$. Then there exists a unique solution $T \in \mathcal{F}_d$ (up to sets of $\mu_r$-measure zero) to the Monge transport problem associated with $r, p$ and the cost function $c(x, y) = |x - y|^2$. Conversely, let $r \in \mathcal{P}_d$ and $T \in \mathcal{F}_d$ such that $\int_{\mathbb{R}^d} |T(x)|^2 r(x) dx < \infty$. Then $T$ is optimal for the Monge Problem above for the function $p \in \mathcal{P}_d$ satisfying $\mu_p = T_# \mu_r$.

**Remark 2.4.** Brenier’s Theorem is much more general than the version described above. Specifically, Brenier’s general Theorem describes the situation where $r, p$ in the Monge Problem are replaced by two probability measures $\mu, \nu$ that do not necessarily have associated density functions and where the term $|x - y|^2$ is replaced by a more general term $c(x, y)$. 

7
2.3. **Algebraic generative models.** Data analysis and processing often assumes a mathematical model of how the data has been generated. For example, algorithms that work well on the wavelet transforms of images often assume the class of images are well modeled by functions that have Fourier transforms well concentrated in some regions of frequency domain (e.g., near the origin). Similarly, the transport-based transforms described above tend to work best when the data to be processed was generated by physical processes that include transport. Examples where generative models have been used effectively with transport-based transforms include modeling mass (e.g., molecule) concentrations inside cells [2], tissues in human brains [12], photon distributions in turbulent media [18], and others [10, 19].

In contrast with the generative model definition typically used in machine learning, the algebraic models assume that a class of functions $S \subset P_d$ is generated from a function $p \in P_d$ by a mass transport phenomena. Although there are infinitely many ways to transport $p$ to form $q \in S$, absent other information related to the generative process, the least action principle from physics often provides one with plausible solutions. In these circumstances a set $H \subset B$ of optimal transport maps can be used as a generative model, where $B$ is set of $C^1$-diffeomorphisms on $\mathbb{R}^d$. In summary, the set of optimal transport maps used for our generative modeling is

$$(10) \quad \mathcal{F}_d^G := \{ f \in B \mid f = \nabla \phi \text{ for some convex } \phi : \mathbb{R}^d \to \mathbb{R} \}.$$ 

Given a function $p \in P_d$ (signal), a class of functions (signals) is generated by the action of a set of diffeomorphism $\mathcal{H} \subset \mathcal{F}_d^G$ on $p$ via

$$(11) \quad S_{p, \mathcal{H}} := \{ p_h = |\det J_h| \cdot p \circ h \mid h \in \mathcal{H} \},$$

where $J_h$ denotes the Jacobian matrix of $h$. It is not hard to check that $S_{p, \mathcal{H}} \subset P_d$.

One of the simplest examples of such a generative model is described in Figure 1. A natural assumption on the set of diffeomorphisms $\mathcal{H}$ used in the generative model is that it has a group structure (see [17]):

1. $\mathcal{H}$ is closed under composition.
2. $Id \in \mathcal{H}$.
3. if $h \in \mathcal{H}$, then $h^{-1} \in \mathcal{H}$.

In this generative model, the set $\mathcal{H}$ has a group structure that does not depend on $p \in P_d$. However, other generative models are possible where $\mathcal{H}$ does not have a group structure, or where $\mathcal{H} = \mathcal{H}(p)$ depends on the initial function $p$.

3. **CDT AND GENERATIVE MODELS IN ONE DIMENSION**

In this section we consider a set of transformations that act on one-dimensional functions (the signals) and produce classes of functions. For this situation (11) becomes

$$(12) \quad S_{p, \mathcal{H}} := \{ p_h = h'(p \circ h) \mid h \in \mathcal{H} \}.$$ 

Given $p \in P_1$, and a diffeomorphism $h \in F_1^G$ and a function $p \in P_1$, a new function $p_h \in P_1$ is generated via the formula

$$(13) \quad p_h = h'(p \circ h).$$
which is (11) for the one dimensional case. Let \( r \in \mathcal{P}_1 \) be a fixed reference and denote by \( \hat{p} \) the transform of \( p \in \mathcal{P}_1 \) as in (7).

Under the definitions above, if we assume that the generative model uses a subgroup \( \mathcal{H} \) of \( \mathcal{F}_1^G \) to generate \( S_{p,\mathcal{H}} \), then any function in \( q \in S_{q,\mathcal{H}} \) can be used as a template to generate \( S_{p,\mathcal{H}} \) by transport diffeomorphism from \( \mathcal{H} \), i.e., \( S_{p,\mathcal{H}} = S_{q,\mathcal{H}} \) for any \( q \in S_{p,\mathcal{H}} \).

Every subgroup \( \mathcal{H} \subseteq \mathcal{F}_1^G \) will generate a partition of \( \mathcal{P}_1 \). Thus, in principle, the group structure in the generative model allows us to classify the image (functions) in \( \mathcal{P}_1 \), hence can be used in data analysis tasks related to classification.

**Proposition 3.1.** Every subgroup \( \mathcal{H} \subseteq \mathcal{F}_1^G \) partition the set \( \mathcal{P}_1 \) via the equivalence relation \( \sim_\mathcal{H} \) defined by \( p \sim_\mathcal{H} q \) if and only if \( p \in S_{q,\mathcal{H}} \).

Often times, if the set of transforms \( \hat{S}_{p,\mathcal{H}} \) is convex, then the solutions of certain estimation problems in transform domain become simple (e.g. linear least squares [19]). Moreover, this property can also enable easier classification when two disjoint generative classes can be easily separated [24]. Thus, one of the main goals is to identify conditions under which the set \( \hat{S}_{p,\mathcal{H}} \) of transforms of \( S_{p,\mathcal{H}} \) is convex. We have the following theorem.

**Theorem 3.2.** Let \( \mathcal{W} \subset \mathcal{F}_1^G \). Then \( \hat{S}_{p,\mathcal{W}} \) is convex for every \( p \in \mathcal{P}_1 \) if and only if \( \mathcal{W}^{-1} := \{ s^{-1} \mid s \in \mathcal{W} \} \) is convex.

**Corollary 3.3.** Let \( \mathcal{H} \subset \mathcal{F}_1^G \) be a group. Then \( \hat{S}_{p,\mathcal{H}} \) is convex for every \( p \in \mathcal{P}_1 \) if and only if \( \mathcal{H} \) is convex.

Obviously \( \mathcal{F}_1^G \) is itself a convex group. Thus, it partitions \( \hat{\mathcal{P}}_1 \) into equivalent classes that are convex. Two equivalent classes \( \hat{S}_{p,\mathcal{F}_1^G}, \hat{S}_{q,\mathcal{F}_1^G} \) are distinct if the \( \text{supp}(p) \) and \( \text{supp}(q) \) are topologically distinct (i.e., non-homeomorphic). Subgroups of \( \mathcal{F}_1^G \) will further partition each class \( \hat{S}_{p,\mathcal{F}_1^G} \) into sub-classes and so on. Even when a generative model \( \mathcal{H} \subset \mathcal{F}_1^G \) is a subset but not a subgroup of \( \mathcal{F}_1^G \), it is still true that \( \text{conv}(\hat{S}_{p,\mathcal{H}}) \cap \text{conv}(\hat{S}_{q,\mathcal{H}}) = \emptyset \) if \( \text{supp}(p) \) and \( \text{supp}(q) \) are topologically distinct. An illustration is presented in Figure 5, where Linear Discriminant Analysis (LDA) [5] is applied to two generative classes and their corresponding CDTs in the transform domain. The two signal classes are generated with templates (one-bump characteristic function \( p_1 \) and two-bump characteristic function \( p_2 \)) shown in Figure 4 and a set \( \mathcal{H} \) of 500 randomly generated fifth degree polynomials with certain constraints on the coefficients\(^2\). It can be seen from Figure 5 that the transformed signal classes \( \hat{S}_{p_1,\mathcal{H}} \) and \( \hat{S}_{p_2,\mathcal{H}} \) (right) are much better separated than the original signal classes \( S_{p_1,\mathcal{H}} \) and \( S_{p_2,\mathcal{H}} \) (left). Moreover, as predicted by our theory above, \( \text{conv}(\hat{S}_{p_1,\mathcal{H}}) \) and \( \text{conv}(\hat{S}_{p_2,\mathcal{H}}) \) are indeed disjoint. This property of the CDT makes it well-adapted for many applications of data analysis, processing and classification. Thus, one of our goals is to understand the structure of the convex subgroups of \( \mathcal{F}_1^G \).

\(^2\)The coefficients of the polynomials are chosen so that the supports of \( h'(p_1 \circ h) \) and \( h'(p_2 \circ h) \) are inside the interval \([0,1]\) for all \( h \in \mathcal{H} \), see Figure 4 for some randomly chosen samples from each class.
3.1. **Examples of convex subgroups of** $F_1^G$. Note that not every subgroup of $F_1^G$ is convex, for example, the group generated by the integer translation diffeomorphisms $f \in F_1^G$, i.e., $\{ h_i \in F_1^G \mid h_i(x) = x - i, i \in \mathbb{Z} \}$. Examples of convex subgroups of $F_1^G$ are given below.

**Example 1.**

(1) $\{Id\}$ is a convex subgroup of $F_1^G$.

(2) $\{\alpha Id \mid \alpha > 0\}$ is a convex subgroup of $F_1^G$ which is also a cone.

(3) Let $T_1 = \{ h_\mu \in F_1^G \mid h_\mu(x) = x - \mu, \mu \in \mathbb{R} \}$. The set of all translation functions is a convex subgroup of $F_1^G$ but not a cone.
(4) Let \( A_1 = \{ h_{\alpha,\mu} \in F_1^G \mid h_{\alpha,\mu}(x) = \alpha x - \mu, \alpha > 0, \mu \in \mathbb{R} \} \). The set of all increasing affine functions is a convex subgroup of \( F_1^G \) and a cone.

(5) Let \( x_0 \in \mathbb{R} \) and consider the set \( F_{x_0} = \{ f \in F_1^G \mid f(x_0) = x_0 \} \). Then \( F_{x_0} \) is a convex subgroup of \( F_1^G \). It is also a cone.

(6) Let \( \Omega \) define a closed interval in \( \mathbb{R} \). Then the set \( F_{\Omega} = \{ f \in F_1 \mid f(y) = y \forall y \in \Omega \} \) is a convex subgroup of \( F_1^G \).

Remark 3.4. Let \( C_1, C_2 \subset F_1^G \) be convex subgroups of \( F_1^G \). Then it is not difficult to see that if \( C_1 \cap C_2 \neq \emptyset \), then \( C_1 \cap C_2 \) is a convex subgroup of \( F_1^G \). In fact, let \( \{ C_\alpha \mid \alpha \in A \} \) be a family of convex subgroups. If \( \bigcap_{\alpha \in A} C_\alpha \neq \emptyset \), then \( \bigcap_{\alpha \in A} C_\alpha \) is a convex subgroup of \( F_1^G \).

The following proposition shows that the set of subgroups of \( F_1^G \) is sufficiently rich.

Proposition 3.5. There are uncountably many distinct convex subgroups of \( F_1^G \).

By examining the proof of the proposition 3.5, we get the following corollary.

Corollary 3.6. Let \( I \) be an indexing set and let \( X = \{ x_\alpha \subset \mathbb{R} : \alpha \in I \} \). Then

- The set \( F_X := \{ f \in F_1^G \mid f(x_\alpha) = x_\alpha, \alpha \in I \} \) is convex subgroup of \( F_1^G \).
- If \( X_1 \subsetneq X_2 \), then \( F_{X_2} \subsetneq F_{X_1} \).

Remark 3.7. The results above are useful for the one dimensional transform CDT transform and the multidimensional R-CDT transforms.

3.2. Proofs of Section 3.

3.2.1. Proof of Theorem 3.2. We start by proving the following lemma:

Lemma 3.8. Let \( p \in \mathcal{P}_1 \), then

\[
\hat{p}_h = h^{-1} \circ \hat{p}, \quad \forall h \in F_1^G
\]

where \( \hat{p} \) denotes the CDT with respect to the reference \( r \).

Proof. Let \( r \in \mathcal{P}_1 \), and \( \hat{p}, \hat{p}_h \in F_1 \) be the optimal transports from \( r \) to \( p, p_h \) respectively. We have that by definition \( \mu_p = \hat{p} \# \mu_r \), \( \mu_p = h \# \mu_{p_h} \) and that \( \mu_{p_h} = (\hat{p}_h) \# \mu_r \). Using the well-known relation \( (S \circ T) \# \mu_q = S_\#(T_\# \mu_q) \) for any maps \( S, T \) and the fact that \( h \circ \hat{p}_h \in F_1 \) (since both \( h \) and \( \hat{p} \) are non-decreasing) and the property that \( h \circ \hat{p}_h \) is square-integrable with respect to \( \mu_p \), we conclude that \( h \circ \hat{p}_h = \hat{p} \). It follows that \( \hat{p}_h = h^{-1} \circ \hat{p} \) for all \( h \in F_1^G \).

Proof of Theorem 3.2. Assume that \( \mathcal{W}^{-1} \) is convex. Then for \( p_{h_1}, p_{h_2} \in S_{p,S} \), and \( 0 \leq \alpha \leq 1 \), we have

\[
\alpha \hat{p}_{h_1}(x) + (1 - \alpha) \hat{p}_{h_2}(x) = \alpha h_{1}^{-1} \circ \hat{p}(x) + (1 - \alpha) h_{2}^{-1} \circ \hat{p}(x)
= \alpha h_{1}^{-1}(\hat{p}(x)) + (1 - \alpha) h_{2}^{-1}(\hat{p}(x))
= (\alpha h_{1}^{-1} + (1 - \alpha) h_{2}^{-1}) \circ \hat{p}(x).
\]

Thus, \( \hat{S}_{p,\mathcal{W}} \) is convex.
For the converse statement, assume that \( \hat{S}_{p,W} \) is convex, i.e., \((\alpha h_1^{-1} + (1-\alpha) h_2^{-1}) \circ \hat{p} \in \hat{S}_{p,W}\) for all \( h_1, h_2 \in W \). Since \( \hat{S}_{p,W} \) is convex for every \( p \in P_1 \), by varying \( p \) (e.g., by choosing \( p \) as translations of \( r \)) one can conclude that \( \alpha h_1^{-1} + (1-\alpha) h_2^{-1} \in W \) for all \( h_1, h_2 \in W \).

\[ \square \]

3.2.2. Proof of Proposition 3.5.

Proof of Proposition 3.5. Let \( x_0, x_1 \in \mathbb{R} \) with \( x_0 \neq x_1 \). Choose \( f, g \in F_{d}^{G} \) such that \( f(x_0) = x_0, f(x_1) \neq x_0, \) and \( g(x_0) \neq x_0, g(x_1) = x_1 \). Using the notation of (5) in Example (1), we get \( f \in F_{x_0} \) but \( f \neq F_{x_1} \) and similarly, \( g \in F_{x_1} \) but \( g \neq F_{x_0} \). Extending the previous argument over an uncountable set \( X = \{x_i \mid i \in I\} \subset \mathbb{R} \) provides an uncountable set \( \{F_{x_i}\} \) of distinct subgroups of \( F_{d}^{G} \).

\[ \square \]

4. LOT AND GENERATIVE MODELS IN MULTI-DIMENSIONS

As in the one dimensional case, a group \( H \subseteq F_{d}^{G} \) partitions the set \( P_d \) into equivalent classes (as in Proposition (3.1)) that are useful for classification problems. However, unlike the one dimensional case, Equation (14) does not hold in general. Thus, in order to obtain convexity of \( \hat{S}_{p,H} \) from the convexity of \( H \) as in Theorem 3.3, we need to find conditions on \( H \) such that for the given \( p \in P_d \) the equation

\[ \hat{p}_h = h^{-1} \circ \hat{p} \]

holds for all \( h \in H \) and hence to generate convex partitions of \( \hat{P}_d \) when \( H \) is convex. In particular we have

**Theorem 4.1.** Let \( p \in P_d \) and \( H \subseteq F_{d}^{G} \) be a subgroup. If (15) holds for all \( h \in H \), and \( H \) is a convex set then \( \hat{S}_{p,H} \) is also convex.

Note that, unlike \( F_{d}^{G}, F_{d}^{G} \) is not a group for \( d \geq 2 \). Our next goal is to find conditions on \( H \) such that \( \hat{S}_{p,H} \) is convex for all \( p \in P_d \). On the other hand, there are groups that are subsets of \( F_{d}^{G} \). For examples, the diffeomorphisms group of translations and isotropic scaling \( H_a := \{L_{a,u}(x) := ax + u \mid a > 0, u \in \mathbb{R}^d\} \). To see this, let \( h \in H_a \), and let \( h^{-1}(x) := \alpha x + u \) for some \( \alpha > 0 \) and \( u \in \mathbb{R}^d \). For \( p \in P_d \), we have that \( \hat{p} = \nabla \phi_p \) for some convex function \( \phi_p \). Hence

\[ h^{-1} \circ \hat{p} = h^{-1} \circ \nabla \phi_p = \alpha \nabla \phi_p + u. \]

To see that \( h^{-1} \circ \hat{p} = \hat{p}_h \), we use the second part of Brenier’s Theorem above 2.3 and simply note that \( h^{-1} \circ \hat{p} = \alpha \nabla \phi_p + u \) is the gradient of the convex function \( \psi(x) = \alpha \phi_p(x) + u \cdot x \). We have the following Theorem.

**Theorem 4.2.** Let \( d \geq 2 \) and \( W \subseteq F_{d}^{G} \). If for any \( p \in P_d \) Equation (15) holds for all \( h \in W \), then \( W \subseteq H_a \).

**Remark 4.3.** Note that in the previous theorem, \( W \) does not need to be a subgroup of \( F_{d}^{G} \).

**Corollary 4.4.** Let \( d \geq 2 \) and \( H \subseteq F_{d}^{G} \) be a subgroup. If for any \( p \in P_d \) Equation (15) holds for all \( h \in H \), then \( H \subseteq H_a \).
Remark 4.5. Theorem 4.2 and its corollary show that the partitioning of the set of transformed signals \( \mathcal{P}_d \) into convex sets is much more constrained than in the one-dimensional case. However, by allowing condition (15) to hold only on a subset of \( \mathcal{P}_d \), one can enlarge the set of convex generative models.

4.1. Relaxation in dimension \( d = 2 \). As mentioned in Remark 4.5, one can relax the condition in (15) to hold on a subset of \( \mathcal{P}_d \) rather than all of \( \mathcal{P}_d \). This group is strictly larger than \( \mathcal{H}_d \) satisfying Equality (15). In this section, we show how to construct such subset \( \mathcal{P}_r \subset \mathcal{P}_d \) and group \( \mathcal{H}_r \) strictly larger than \( \mathcal{H}_d \) such that Equation (15) holds for all \( h \in \mathcal{H}_r \) and \( p \in \mathcal{P}_r \).

Definition 4.6 (Restrictive sets of transformations and PDFs).

\[
\mathcal{H}_r = \left\{ h(x,y) := \frac{1}{2} \left[ f'(x+y) + g'(x-y) \right] \mid f, g \in \mathcal{R} \right\}
\]

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

\[
\mathcal{P}_r := \{ p \in \mathcal{P}_2 : \mid \text{det} J_h(p \circ h) = r \text{ for some } h \in \mathcal{H}_r \}.
\]

Remark 4.7. \( \mathcal{H}_r \) has the following properties:

i) for any \( h \in \mathcal{H}_r \), \( h = \nabla \phi \) for the convex function \( \phi(x,y) = f(x+y) + g(x-y) \). The fact that \( \phi \) is a convex function on \( \mathbb{R}^2 \) follows from the fact that \( f', g' \) are strictly increasing.

ii) for any \( h_1, h_2 \in \mathcal{H}_r \), \( h_1 \circ h_2 = \nabla \psi \) for some convex function \( \psi : \mathbb{R}^2 \rightarrow \mathbb{R} \). To see this, one can check that

\[
(h_1 \circ h_2)(x,y) = \frac{1}{2} \left[ f'_1(f'_2(x+y)) + g'_1(g'_2(x-y)) \right],
\]

where \( h_i(x,y) = \frac{1}{2} \nabla \left( f_i(x+y) + g_i(x-y) \right) \) and \( f_i, g_i \in \mathcal{R} \) for \( i = 1, 2 \). Since \( f'_i, g'_i \) are strictly increasing, so are \( f'_2 \circ f'_1 \) and \( g'_2 \circ g'_1 \). The conclusion that \( h_1 \circ h_2 = \nabla \psi \) for some convex function \( \psi \) follows from Part i) and identity (18) above.

iii) for any \( h \in \mathcal{H}_r \), \( h^{-1} \in \mathcal{H}_r \). To see this, a direct computation gives

\[
h^{-1}(z,w) = \frac{1}{2} \left[ (f')^{-1}(z+w) + (g')^{-1}(z-w) \right].
\]

Since \( f', g' \) are strictly increasing bijections on \( \mathbb{R} \), so are \( (f')^{-1}, (g')^{-1} \), and the conclusion then follows from Part i).

By part i) of the previous remark, it follows that every \( h \in \mathcal{H}_r \) is a conservative vector field (i.e., \( h = \nabla \phi \) for some \( C^1 \) function \( \phi \)) and hence is irrotational since \( \nabla \times h = \nabla \times \nabla \phi = 0 \). Figure 6 shows an example of such vector fields generated by some \( f, g \) which are invertible and whose derivatives are strictly increasing on \([-5, 5]\).

By part iii) of Remark 4.7, we have that \( \mathcal{H}_r \) is a group. From the definition of \( \mathcal{P}_r \) and Brenier’s Theorem 2.3, we have that \( \hat{\mathcal{P}} \in \mathcal{H}_r \). The fact that \( h \circ \hat{\mathcal{P}} \) is the gradient of some convex function follows from part ii) of Remark 4.7. Thus, for any \( p \in \mathcal{P}_r \), Equation (15) holds for all \( h \in \mathcal{H}_r \). In summary, we have the following theorem:
Theorem 4.8. For any \( p \in \mathcal{P}_r \), we have \( \hat{p}_h = h^{-1} \circ \hat{p} \) for any \( h \in \mathcal{H}_r \) where \( p_h = | \det J_h | \cdot p \circ h \).

By an argument similar to the one for Theorem 3.3, we have the following corollary of Theorem 4.8:

Corollary 4.9. Let \( \mathcal{H} \subseteq \mathcal{H}_r \) be a group. Then \( \hat{S}_{p,\mathcal{H}} \) is convex for any \( p \in \mathcal{P}_r \) if and only if \( \mathcal{H} \) is convex.

4.1.1. Convex Subgroups of \( \mathcal{H}_r \). We remark first that the convex group \( \mathcal{H}_a \) of translation and isotropic scaling diffeomorphisms in dimension 2 is a subgroup of \( \mathcal{H}_r \). In particular, by choosing \( f_{a,b_1}(t) = \frac{1}{2}at^2 + b_1 t \) and \( g_{a,b_2}(t) = \frac{1}{2}at^2 + b_2 t \) where \( a > 0 \) and \( b_1, b_2 \in \mathbb{R} \), one obtains that \( h(x,y) = \frac{1}{2} \nabla (f_{a,b_1}(x+y) + g_{a,b_2}(x-y)) = a \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2} \begin{bmatrix} b_1 + b_2 \\ b_1 - b_2 \end{bmatrix} \).

Indeed, we can construct other examples of convex subgroups of \( \mathcal{H}_r \) by judiciously choosing \( f, g \).

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

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

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.

4.2. Proofs of Section 4.
4.2.1. **Proof of Theorems 4.2.** The proof of Theorem 4.2 relies on the following proposition which will be proved toward the end of this section. Through the rest of the section, \( \phi \) will denote a convex function such that \( \nabla \phi \) is the optimal transport map between a fixed reference \( r \) and a function \( p \in \mathcal{P}_d^* \) \( (d \geq 2) \), where

\[
\mathcal{P}_d^* := \{ p \in \mathcal{P}_d \mid r = |\det \nabla f(p \circ f)\} \text{ for some } f \in \mathcal{F}_d^C \}.
\]

**Proposition 4.10.** Let \( \varphi : \mathbb{R}^d \to \mathbb{R} \) be a convex function in \( C^2(\mathbb{R}^d) \) such that for any \( p \in \mathcal{P}_d^* \), \( \nabla \varphi \circ \nabla \phi \) can be written as \( \nabla \varphi \circ \nabla \phi = \nabla \gamma \) for some function \( \gamma = \gamma(p) \). Then \( \nabla^2 \varphi(x) \equiv \alpha I_d \), where \( I_d \) is the identity matrix in \( \mathbb{R}^{d \times d} \).

We are now ready to prove Theorem 4.2.

**Proof of Theorem 4.2.** Recall that, by the assumption on \( \mathcal{W} \subset \mathcal{F}_d^C \), for any \( p \in \mathcal{P}_d \) the following holds
\[
\hat{p}_h = h^{-1} \circ \hat{p}, \quad \text{for all } h \in \mathcal{W}.
\]
Using Brenier’s Theorem 2.3 above, a transport map is optimal in the sense of (3) if and only if it is the gradient of a convex function. Accordingly, there exist convex functions \( \phi_{ph}, \varphi, \phi \) such that \( \nabla \phi_{ph} = \hat{p}_h \), \( \nabla \phi = h^{-1} \) and \( \nabla \phi_p = \hat{p} \) (note that like \( h \), \( h^{-1} \) is also an optimal transport map). In particular, we have that \( \nabla \varphi \circ \nabla \phi_p = \nabla \phi_{ph} \) for every \( p \in \mathcal{P}_d^* \). By proposition 4.10, it follows that \( \nabla^2 \varphi \equiv \alpha I_d \) \( (\alpha > 0) \). Hence \( \nabla \varphi(x) = \alpha x + b \), where \( b \in \mathbb{R}^d \), i.e., \( h^{-1} \in \mathcal{H}_a \), which also implies \( h \in \mathcal{H}_a \). Thus \( \mathcal{W} \subset \mathcal{H}_a \). □

4.2.2. **Proof of Proposition 4.10.** We start by proving the following two lemmas.

**Lemma 4.11.** Let \( \varphi, \phi : \mathbb{R}^d \to \mathbb{R} \) be two functions in \( C^2(\mathbb{R}^d) \). If \( \nabla \varphi \circ \nabla \phi = \nabla \gamma \) for some function \( \gamma \in C^2(\mathbb{R}^d) \), then the matrix-valued functions \( (\nabla^2 \varphi) \circ \nabla \phi \) and \( \nabla^2 \phi \) must commute, i.e., \( \nabla^2 \varphi(\nabla \phi(x)) \nabla^2 \phi(x) = \nabla^2 \phi(x) \nabla^2 \varphi(\nabla \phi(x)) \) for all \( x \in \mathbb{R}^d \).

**Proof.** Since \( \varphi, \phi, \gamma \) are continuous twice differentiable, by Schwarz’s theorem, their Hessian matrices \( \nabla^2 \varphi, \nabla^2 \phi, \nabla^2 \gamma \) are all symmetric. It follows from \( (\nabla \varphi) \circ \nabla \phi = \nabla \gamma \) that \( \nabla^2 \varphi(\nabla \phi(x)) \nabla^2 \phi(x) = \nabla^2 \gamma(x) \) for all \( x \in \mathbb{R}^d \) by multivariate chain rule. Since the product of two real symmetric matrices is symmetric if and only if they commute, \( \nabla^2 \varphi(\nabla \phi(x)) \) and \( \nabla^2 \phi(x) \) must commute. □

**Lemma 4.12.** Let \( A \) be an \( \mathbb{R}^{d \times d} \) matrix such that \( A \Delta = \Delta A \) for some diagonal matrix \( \Delta \) with distinct diagonal entries. Then \( A \) is a diagonal matrix.

**Proof.** Let \( \Delta = \begin{bmatrix} \delta_1 & & \\ & \delta_2 & \\ & & \ddots \\ & & & \delta_d \end{bmatrix} \) where \( \delta_i \neq \delta_j \) whenever \( i \neq j \). Since \( A \Delta = \Delta A \), by comparing of the \((i, j)\)-th entry of both sides we have

\[
\delta_j a_{ij} = \delta_i a_{ij},
\]
where \( a_{ij} \) denotes the \((i, j)\)-th entry of matrix \( A \). Since \( \delta_i \neq \delta_j \), for \( i \neq j \), it follows that for \( i \neq j \), \( a_{ij} = 0 \). □
Lemma 4.13. Let $D$ be a diagonal matrix in $\mathbb{R}^{d \times d}$ such that $DM = MD$ for some matrix $M$ having the property that it has an eigen-space $E_\lambda = \text{span}\{u\}$ and such that all entries of $u$ are non-zero. Then $D = \alpha I$ where $\alpha \in \mathbb{R}$ and $I_d$ is the identity matrix in $\mathbb{R}^{d \times d}$.

Proof. Since $DM = MD$, it follows that

\begin{equation}
MDu = DMu = D\lambda u = \lambda Du.
\end{equation}

Hence $Du \in E_\lambda(M)$ and $Du = \alpha u$ for some $\alpha \in \mathbb{R}$ since $\dim E_\lambda(M) = 1$. Using the fact all entries of $u$ are non-zero and comparing the entries of $Du$ and $\alpha u$, one immediately gets that $D = \alpha I_d$. \hfill \square

As stated at the beginning of this section, $\phi_p$ will denote a convex function such that $\nabla \phi_p$ is the optimal transport map between $r$ and $p$.

Proof of Proposition 4.10. By Lemma 4.11, we have that the matrices $\nabla^2 \varphi(\nabla \phi_p(x))$ and $\nabla^2 \phi_p(x)$ must commute for every $x \in \mathbb{R}^d$, and for every $p \in \mathcal{P}_d^*$. To prove the proposition, we make judicious choices for $p$. First, we choose $p = p(\delta_1, \cdots, \delta_d)$ with $\phi_p(x) = \frac{1}{2} \sum_{i=1}^n \delta_i x_i^2$ where $\delta_j > 0$ for $j = 1, \cdots, d$. In particular, $r(x) = |\det \nabla^2 \phi_p(x)|p(\nabla \phi_p(x))$ where $r \in \mathcal{P}_d$ is the reference. Since $\phi_p$ is convex and quadratic, it is not difficult to show that $p \in \mathcal{P}_d^*$. It is easy to see that $\nabla \phi_p(x) = \begin{bmatrix} \delta_1 x_1 \\
\vdots \\
\delta_d x_d \end{bmatrix}$ and $\nabla^2 \phi_p(x) = \begin{bmatrix} \delta_1 \\
\ddots \\
\delta_d \end{bmatrix}$.

By Lemma 4.12, setting $A = \nabla^2 \varphi(\nabla \phi_p(x))$ and $\Delta = \phi_p(x)$, we conclude that $A = \nabla^2 \varphi(\nabla \phi_p(x))$ is a diagonal matrix for every $x \in \mathbb{R}^d$. Since $\nabla \phi_p(x) : \mathbb{R}^d \to \mathbb{R}^d$ is bijective, it follows that $\nabla^2 \varphi(x)$ is also diagonal for every $x \in \mathbb{R}^d$. Next, we choose $p = p(M)$ such that $\phi_p(x) = \frac{1}{2} x^t M x$, where $M$ is a constant positive definite matrix with an eigenvector $u$ whose entries are non-zero and whose corresponding eigenspace has dimension 1. A simple construction using the spectral decomposition of symmetric matrices show that such an $M$ exists. For these choices of $p$, $\phi_p$, we have $\nabla^2 \phi_p(x) = M$ which is a constant matrix independent of $x$. Again, it is not difficult to show that $p \in \mathcal{P}_d^*$. Hence $\nabla^2 \varphi(Mx)$ and $M$ commute. Since $\nabla^2 \varphi(Mx)$ is diagonal for every $x \in \mathbb{R}^d$, using Lemma 4.13 with $D = \nabla^2 \varphi(x)$, we have that $\nabla^2 \varphi(Mx) = \beta_x I$, where $\beta_x$ is a constant depending on $x$. Since $M$ is an invertible matrix, it follow that $\nabla^2 \varphi(x) = \alpha_x I_d$ where $\alpha_x$ is a constant depending on $x$. Since $\frac{\partial^2 \varphi}{\partial x_i \partial x_j} \equiv 0$ for $i \neq j$, we have that $\varphi(x_1, \ldots, x_d) = F_1(x_1) + \cdots + F_n(x_d)$ for some univariate functions $F_1, \ldots, F_d$. Hence $F''_i(x_1) = F''_2(x_2) = \cdots = F''_d(x_d)$ for all $(x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$. If follows that $F''_i$ must be the same constant function for $i = 1, \ldots, d$ and that $\alpha_x = \alpha$ is independent of $x$, which implies that $\nabla^2 \varphi(x) \equiv \alpha I_d$ for some constant $\alpha$. \hfill \square

16
5. Summary and Open questions

In this paper we have worked to highlight and clarify certain properties of an emerging set of transport-based signal transforms. More specifically, we have worked to show that for certain types of signals and generative models, the transport transforms discussed earlier have the ability to render signal classes convex in transform space. As convex signal classes render solutions to estimation and detection problems much simpler to solve (e.g. via linear least squares, or linear classification), the topic is important. While the picture is far from complete, we hope that this paper serves a starting point on guiding meaningful applications of these transforms and opens the doors to deeper understanding about when and how to apply these transforms.

5.1. List of contributions. More specifically, we make the connection between the convex group structure in our generative model and convex partitions of the transformed signal space via the formula (14), which holds naturally in the 1D case (see Lemma 3.8). In 1D, we give examples of convex groups of diffeomorphisms in 1D and show that there are infinitely many such groups.

In dimension \(d \geq 2\), we show that the only groups \(H\) of diffeomorphisms that validate formula (14) for all \(p \in \mathcal{P}_d\) and all \(h \in H\) are subgroups of \(H_a\), which is group of translations and isotropic scaling diffeomorphisms. In particular, any convex subgroup of \(H_a\) generates a convex partition of the transformed signal space \(\hat{\mathcal{P}}_d\).

Moreover, in dimension two, we show how to construct a group \(H_r \subset \mathcal{F}_2\) which is larger than \(H_a\) such that (14) holds for all \(h \in H_r\) and \(p\) in \(\mathcal{P}_r\) which is a subset of \(\mathcal{P}_d\). In particular, any convex subgroup of \(H_r\) generates a convex partition of \(\hat{\mathcal{P}}_r\).

5.2. Open questions. In 1D, a characterization of convex subgroups of \(\mathcal{F}_1\) is missing. More specifically, one can ask the following:

1. Besides \(\mathcal{F}_1\), are there subgroups of \(\mathcal{F}_1\) that are not of the form of those in Example 1?
2. If the answer to the previous question is yes, can we give more examples of convex subgroups, and can we characterize the subgroups into a few concrete categories?
3. For every group \(H\) the set \(\hat{\mathcal{P}}_1\) is tiled by a convex structure in \(\hat{\mathcal{P}}_1\). What is the geometry of this structure?

In multi-dimensions, while it is convenient to make use of formula (14) to derive convexity results similar to the ones in one dimension, it is not necessary. In particular, one can ask the following questions:

1. Can one derive that the convexity of \(\hat{\mathcal{S}}_{p,H}\) from the convexity of \(H\) without formula (14) being true for all \(h \in H\)?
2. Are there other conditions one can impose on the model \(S_{p,H}\) other than that \(H\) is convex so that \(\hat{\mathcal{S}}_{p,H}\) is convex?

There are several natural questions related to the relaxation in dimension two:

1. Are there more interesting examples of subgroups of \(H_r\) that have connections to possible applications?
For dimension $d > 2$, using Example 2 as a guide, what are the set of convex subgroups $\mathcal{H}$ of $\mathcal{F}_d$ when we restrict the signals to some subsets of $\mathcal{P}_d$?

The above discussions fall under the framework of a algebraic generative model $\mathcal{S}_{p, \mathcal{H}}$ with a single template $p$ and a convex group $\mathcal{H}$ of diffeomorphisms. In fact, it might be suitable, to consider multiple templates for the generative modeling in certain applications, or to not assume a group structure for $\mathcal{H}$ when modeling certain image classes. We leave such extensions for future research topics and believe that delving into these questions could potentially lead to more thoughtful engineering modeling, algorithmic design and new interesting mathematics.

ACKNOWLEDGMENT

This work is supported by NIH award R01 GM130825. The authors would also like to thank Longxiu Huang, Soheil Kolouri, Armenak Petrosyan, and Mohammad Shifat-E-Rabbi for their careful reading of our paper and their pertinent suggestions.

REFERENCES

[1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. arXiv preprint arXiv:1701.07875, 2017.
[2] S. Basu, S. Kolouri, and G. Rohde. Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry. Proc. Natl. Acad. Sci. U.S.A., 111(9):3448–3453, 2014.
[3] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. Commun. Pure Appl. Math., 44(4):375–417, 1991.
[4] T. H. Emerson and J. M. Nichols. Fitting local, low-dimensional parameterizations of optical turbulence modeled from optimal transport velocity vectors. Pattern Recognition Letters, 133:123–128, 2020.
[5] R. A. Fisher. The use of multiple measurements in taxonomic problems. Annals of eugenics, 7(2):179–188, 1936.
[6] S. Guan, B. Liao, Y. Du, and X. Yin. Vehicle type recognition based on Radon-CDT hybrid transfer learning. In 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), pages 1–4, 2019.
[7] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. Int. J. Comput. Vis., 60(4):225–240, 2004.
[8] L. V. Kantorovich. On translation of mass (in Russian), C R. Doklady. Acad. Sci. USSR, 37:199–201, 1942.
[9] S. Kolouri, S. Park, and G. Rohde. The Radon cumulative distribution transform and its application to image classification. IEEE Trans. Image Process., 25(2):920–934, 2016.
[10] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde. Optimal mass transport: Signal processing and machine-learning applications. IEEE Signal Processing Magazine, 34(4):43–59, 2017.
[11] S. Kolouri, A. Tosun, J. Ozolek, and G. Rohde. A continuous linear optimal transport approach for pattern analysis in image datasets. Pattern Recognit., 51:453–462, 2016.
[12] S. Kundu, S. Kolouri, K. I. Erickson, A. F. Kramer, E. McAuley, and G. K. Rohde. Discovery and visualization of structural biomarkers from mri using transport-based morphometry. NeuroImage, 167:256–275, 2018.
[13] G. Monge. Mémoire sur la théorie des déblais et des remblais. De l’Imprimerie Royale, 1781.
[14] K. Ni, X. Bresson, T. Chan, and S. Esedoglu. Local histogram based segmentation using the wasserstein distance. International journal of computer vision, 84(1):97–111, 2009.
[15] J. Ozolek, A. Tosun, W. Wang, C. Chen, S. Kolouri, S. Basu, H. Huang, and G. Rohde. Accurate diagnosis of thyroid follicular lesions from nuclear morphology using supervised learning. *Med. Image. Anal.*, 18(5):772–780, 2014.

[16] S. Park, L. Cattell, J. Nichols, A. Watnik, T. Doster, and G. Rohde. De-multiplexing vortex modes in optical communications using transport-based pattern recognition. *Opt. Express*, 26(4):4004–4022, 2018.

[17] S. Park, S. Kolouri, S. Kundu, and G. Rohde. The cumulative distribution transform and linear pattern classification. *Appl. Comput. Harmon. Anal.*, 2017.

[18] S. R. Park, L. Cattell, J. M. Nichols, A. Watnik, T. Doster, and G. K. Rohde. De-multiplexing vortex modes in optical communications using transport-based pattern recognition. *Optics express*, 26(4):4004–4022, 2018.

[19] A. H. M. Rubaiyat, K. M. Hallam, J. M. Nichols, M. N. Hutchinson, S. Li, and G. K. Rohde. Parametric signal estimation using the cumulative distribution transform. *IEEE Transactions on Signal Processing*, 68:3312–3324, 2020.

[20] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover’s distance as a metric for image retrieval. *IJCV*, 40(2):99–121, 2000.

[21] F. Santambrogio. *Optimal transport for applied mathematicians*. Springer, 2015.

[22] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

[23] J. Shen, Y. Qu, W. Zhang, and Y. Yu. Wasserstein distance guided representation learning for domain adaptation. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[24] M. Shifat-E-Rabbi, X. Yin, A. H. M. Rubaiyat, S. Li, S. Kolouri, A. Aldroubi, J. M. Nichols, and G. K. Rohde. Radon cumulative distribution transform subspace modeling for image classification. [Submitted to] *IEEE Transactions on Image Processing*, 2020. Preprint available at https://arxiv.org/abs/2004.03669.

[25] A. B. Tosun, O. Yergiyev, S. Kolouri, J. F. Silverman, and G. K. Rohde. Detection of malignant mesothelioma using nuclear structure of mesothelial cells in effusion cytology specimens. *Cytometry Part A*, 87(4):326–333, 2015.

[26] C. Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[27] W. Wang, D. Slepčev, S. Basu, J. Ozolek, and G. Rohde. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *IJCV*, 101(2):254–269, 2013.