Sampling Theory, Signal Processing, and Data Analysis (2024) 22:15
https://doi.org/10.1007/s43670-024-00089-7

ORIGINAL ARTICLE

Approximation properties of slice-matching operators

Shiying Li[1] · **Caroline Moosmüller**[1]

Received: 20 October 2023 / Accepted: 23 April 2024 / Published online: 17 May 2024

Abstract

Iterative slice-matching procedures are efficient schemes for transferring a source measure to a target measure, especially in high dimensions. These schemes have been successfully used in applications such as color transfer and shape retrieval, and are guaranteed to converge under regularity assumptions. In this paper, we explore approximation properties related to a single step of such iterative schemes by examining an associated slice-matching operator, depending on a source measure, a target measure, and slicing directions. In particular, we demonstrate an invariance property with respect to the source measure, an equivariance property with respect to the target measure, and Lipschitz continuity concerning the slicing directions. We furthermore establish error bounds corresponding to approximating the target measure by one step of the slice-matching scheme and characterize situations in which the slice-matching operator recovers the optimal transport map between two measures. We also investigate connections to affine registration problems with respect to (sliced) Wasserstein distances. These connections can be also be viewed as extensions to the invariance and equivariance properties of the slice-matching operator and illustrate the extent to which slice-matching schemes incorporate affine effects.

Keywords Optimal transport · Sliced Wasserstein distance · Slice-matching · Measure approximation · Registration

Mathematics Subject Classification 49Q22 · 68T10 · 41A65 · 65D18

Communicated by Mike Neamtu.

✉ Shiying Li
shiyl@unc.edu

Caroline Moosmüller
cmoosm@unc.edu

1 Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

1 Introduction

Optimal transport and Wasserstein distances play a crucial role in machine learning and related applications [1–4]. For example, these methods have gained prominence in generative modeling [5], aiming to find transport maps or their approximations [6, 7] that align posterior and prior distributions. A basic problem can be described as follows: Given a random variable $X \sim \sigma$ where σ is a prior probability distribution and a target probability distribution μ of interest, find a transformation T such that $T(X) \sim \mu$.

Transforming one probability distribution into another is a fundamental problem that also has implications for flow matching [8], probability flows [9, 10], particle evolutions [11], and the general learning of underlying distributions in complex data sets [1, 12].

While Wasserstein distances have proven successful in modeling probability distributions [1, 7], their computational expense, especially in high-dimensional scenarios, necessitates more efficient approaches. The computation can be intensive for large-scale problems; specifically, the cost of calculation via linear programs comes with a complexity of $O(m^3 \log(m))$, while the Sinkhorn version [13] provides a faster approximation at $O(m^2 \log(m))$, where m is the number of particles used to approximate a measure. Sliced-Wasserstein-based generative models [14, 15] provide scalable alternatives. Our study focuses on one such model, namely, slice-matching schemes [16, 17]. These schemes utilize projections onto lines (slices) as well as the computational benefit of one-dimensional optimal transport to offer effective approximations of target distributions. Furthermore, they are closely aligned with the broader context of normalizing flows [12, 18] and variational inference [7]. Beyond their computational advantages, these schemes also demonstrate promising convergence results, as shown in [16, 19, 20].

While convergence analysis of iterative slice-matching schemes is the primary focus of [20], the present manuscript aims at establishing a comprehensive understanding of one step of such schemes, focusing on recovery and stability properties. We explore structural relationships between measures and address two closely related questions: (A) when can closed-form formulations serve as robust alternatives to optimal transport maps? (B) how effectively do slice-matching approximations represent the target measure? Additionally, our study examines the ability to handle transformations like shifts and scalings in the initial step, demonstrated through the analysis of some basic registration problems. Registration problems are not the main focus of this paper; we rather use them to understand affine effects of slice-matching maps. Nonetheless, we would like to point out that optimal transport is a useful tool for registration problems, see for example [21–23].

1.1 Slice-matching maps

In this paper, we are interested in maps defined by a slicing and matching procedure [16], which is closely related to the sliced Wasserstein distance. These are maps of the form

$$T_{\sigma,\mu;P}(x) = \sum_{i=1}^{n} T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}(x \cdot \theta_i)\,\theta_i, \quad x \in \mathbb{R}^n$$

involving a source measure σ, a target measure μ and an orthogonal matrix $P = [\theta_1, \ldots, \theta_n]$. Here σ^{θ_i} denotes the 1-dimensional measure obtained by projecting σ onto the line θ, and $T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}$ is the optimal transport map between the 1-dimensional measures σ^{θ_i} and μ^{θ_i}. Note that the 1-dimensional optimal transport map can be computed explicitly, see Sect. 2 for more details.

This maps allows to define the iterative *slice-matching procedure* [16, 20]

$$\sigma_{k+1} = (T_{\sigma_k,\mu;P_k})_\sharp \sigma_k, \quad k \geq 0, \tag{1}$$

which have been successfully used in applications such as color transfer [16], texture mixing [24] and shape retrieval [25]. Convergence results of (1) in special cases have been obtained in [16, 19]. More general almost sure (a.s.) convergence of $\sigma_k \to \mu$ for a stochastic variant of (1) have been established in [20].

The procedure (1) also defines an iteration on the level of maps through $T_{\sigma_k,\mu;P_k}$, $k \geq 0$ (though the mentioned convergence results only hold for measures). Note that bounds which are valid for maps directly carry over to the associated pushforward measures through the well-known stability result [26, Eq. (2.1)]

$$W_2(F_\sharp\sigma, G_\sharp\sigma) \leq \|F - G\|_\sigma. \tag{2}$$

In this paper, we are interested in the approximation power of one step of (1), both for measures and maps. This means that we study (A) the relation between $\sigma_1 = (T_{\sigma,\mu;P})_\sharp\sigma$ and the target μ and (B) the relation between $T_{\sigma,\mu;P}$ and T, with $\mu = T_\sharp\sigma$.

1.2 Main contributions

The contributions of this paper are twofold and summarized in Theorem 1 and 2. To formulate our contributions, we use the notation $\mathcal{S} = \{x \mapsto ax + b : a > 0, b \in \mathbb{R}^n\}$ for the set of shifts and scaling, and $\mathfrak{S}(P) = \{x \mapsto \sum_{i=1}^n f_i(x \cdot \theta_i)\theta_i : f_i : \mathbb{R} \to \mathbb{R} \text{ increasing}\}$ with $P \in O(n)$ for the set of P-compatible maps [27]. Note that $\mathcal{S} \subset \mathfrak{S}(P)$ for any P.

Theorem 1 [Recovery and approximation: informal implications of Corollary 6, Proposition 10, and Proposition 11] *Consider two measures σ, μ with $\mu = T_\sharp\sigma$. Then we get*

1. *If $\|T - S\| \leq \varepsilon$ for some $S \in \mathcal{S}$, then one step of the scheme (1) reconstructs T up to an error of order ε, i.e. $\|T - T_{\sigma,\mu;P}\| \leq 2\varepsilon$ for any $P \in O(n)$. In particular, if $T \in \mathcal{S}$, then $T_{\sigma,\mu;P} = T$ for any $P \in O(n)$.*

2. *If $T \in \mathfrak{S}(P)$ for some orthogonal matrix P, then one step of the scheme (1) using Q reconstructs T up to an error of order $\|P - Q\|_F$, i.e. $\|T - T_{\sigma,\mu;Q}\| \leq C\|P - Q\|_F$. In particular, if $Q = P$, then $T_{\sigma,\mu;Q} = T$.*

Through the stability bound of (2), all results of Theorem 1 also hold for the reconstruction of the target measure μ through σ_1.

Our results shows that basic transformations relating σ to μ can be recovered easily, not needing any optimization scheme. This relates to recent efforts in trying to approximate the optimal transport map (or fully replace it) by simpler maps such as the Knothe-Rosenblatt construction [6].

To formulate our second contribution, we define the *slice-matching operator* $\mathcal{U}$, which assigns the first step of (1) to a given source σ, target μ, and slicing directions $P \in O(n)$:

$$\mathcal{U} : (\sigma, \mu, P) \mapsto (T_{\sigma,\mu;P})_\sharp\sigma.$$

Theorem 2 [Encoding of special affine effects: informal implications of Proposition 5, Proposition 13, Proposition 12, Corollary 15] *Consider a source measure σ and a target measure μ. One step of the slice-matching procedure (1) encodes basic transformations in the following sense:*

1. *$\mathcal{U}$ is invariant to $\mathfrak{S}(P)$-actions on the source measure σ:*

$$\mathcal{U}(T_\sharp\sigma, \mu, P) = \mathcal{U}(\sigma, \mu, P), \quad T \in \mathfrak{S}(P).$$

2. *$\mathcal{U}$ is equivariant to $\mathfrak{S}(P)$-actions on the target measure μ:*

$$\mathcal{U}(\sigma, T_\sharp\mu, P) = T_\sharp\mathcal{U}(\sigma, \mu, P), \quad T \in \mathfrak{S}(P).$$

3. *$\mathcal{U}$ encodes translation effects between σ and μ by matching means: $E(\mathcal{U}(\sigma, \mu, P)) = E(\mu)$ for any $P \in O(n)$.*
4. *$\mathcal{U}$ encodes translation-and-scaling effects in the following sense: Let S be the best map in $\mathcal{S}$ that aligns σ and μ, and let S^* be the best map in $\mathcal{S}$ that aligns σ and $\mathcal{U}(\sigma, \mu, P)$. Then, by choosing P randomly, in expectation we get*

$$\mathbb{E}\|S - S^*\|_\sigma = C_\sigma \left(W_2^2(\sigma, \mu) - nSW_2^2(\sigma, \mu) \right) \geq 0,$$

where C_σ depends on the mean and second moment of σ, W_2 denotes the Wasserstein distance, and SW_2 denotes the sliced Wasserstein distance.

Note that (4) means the following: If $(W_2^2(\sigma, \mu) - nSW_2^2(\sigma, \mu)) \leq \varepsilon$, then one step of the iterative scheme (1) removes global translation-and-scaling effects between the source σ and the target μ up to an error of size $C_\sigma\varepsilon$.

1.3 Structure of the paper

This paper is organized as follows: Sect. 2 provides essential background information on optimal transport. Section 3 delves into the details of slice-matching maps, its relations to compatibility, as well as moment-matching properties. In Sect. 4, we present invariance, equivariance, and Lipschitz properties associated with the slice-matching operator, which lead to recovery and stability results. In Sect. 5, we further explore how the slice-matching procedure handles affine effects by studying basic registration problems. The paper closes with a concluding remark in Sect. 6.

2 Preliminaries

We use the notation $\mathcal{P}(\mathbb{R}^n)$ and $\mathcal{P}_{ac}(\mathbb{R}^n)$ for the spaces of probability measures on $\mathbb{R}^n$ and absolutely continuous measures with respect to the Lebesgue measure, respectively. We consider the quadratic Wasserstein space, denoted by $W_2(\mathbb{R}^n)$, which includes probability measures σ with finite second moments, i.e. σ satisfying $M_2(\sigma) = \int_{\mathbb{R}^n} \|x\|^2 d\sigma(x) < \infty$. In addition, let $\mathcal{W}_{2,ac}(\mathbb{R}^n) = \mathcal{W}_2(\mathbb{R}^n) \cap \mathcal{P}_{ac}(\mathbb{R}^n)$. The mean of a measure σ is denoted by $E(\sigma) = \int x d\sigma(x)$.

On $W_2(\mathbb{R}^n)$ we consider the quadratic Wasserstein distance [28]:

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

where $\Gamma(\sigma, \mu) := \{\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : \pi(A \times \mathbb{R}^n) = \sigma(A), \pi(\mathbb{R}^n \times B) = \mu(B), A, B \subseteq \mathbb{R}^n \text{ measurable}\}$ represents the set of couplings between σ and μ.

When $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\mu \in \mathcal{W}_2(\mathbb{R}^n)$, the optimization problem:

$$\min_{T:T_\sharp\sigma=\mu} \int_{\mathbb{R}^n} \|T(x) - x\|^2 \, d\sigma(x),$$

with T a map in $L^2(\mathbb{R}^n, \sigma)$, has a unique (up to constants) solution [29], which we denote by T_σ^μ. Here $\sharp$ is the pushforward operator. The map T_σ^μ takes the form $T_\sigma^\mu = \nabla\varphi$ where φ is convex [29]. Maps which are the gradients of convex functions will be referred to as *Brenier maps*.

If T_σ^μ exists, the optimal coupling has the form $\pi = (\mathrm{id}, T_\sigma^\mu)_\sharp\sigma$. In this case, the Wasserstein-2 distance can then be written as: $W_2(\sigma, \mu) = \|T_\sigma^\mu - \mathrm{id}\,\|_\sigma$, where $\|\cdot\|_\sigma$ is the L^2-norm with respect to σ.

For 1-dimensional measures, the exist explicit formulae for the optimal transport map and the Wasserstein distance. With $\sigma \in \mathcal{P}_{ac}(\mathbb{R})$ and $\mu \in \mathcal{P}(\mathbb{R})$ we get $T_\sigma^\mu = F_\mu^{-1} \circ F_\sigma$, where F_σ is the cumulative distribution function (CDF) of σ, and F_μ^{-1} is the pseudo-inverse of the CDF of μ. This leads to:

$$W_2(\sigma, \mu) = \left(\int_0^1 |F_\mu^{-1}(x) - F_\sigma^{-1}(x)|^2 \, dx \right)^{1/2}. \tag{3}$$

Throughout this paper, we use the same symbols to denote Wasserstein distance and optimal transport maps for probability measures on both $\mathbb{R}^n$ and $\mathbb{R}$, with the context clarifying the dimension of the measures.

We also use the *sliced-Wasserstein distance* between $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\mu \in \mathcal{W}_2(\mathbb{R}^n)$:

$$SW_2^2(\sigma, \mu) = \int_{S^{n-1}} W_2^2(\sigma^\theta, \mu^\theta)\, du(\theta), \tag{4}$$

with $\sigma^\theta = \mathcal{P}_{\theta\sharp}\sigma$, where $\mathcal{P}_\theta(x) = x \cdot \theta$ denotes the projection onto the line defined by θ, and u denotes the uniform measure on S^{n-1}. In (4), W_2 denotes the Wasserstein distance between the 1-dimensional projected measures σ^θ, μ^θ.

It is known that $SW_2 \leq W_2$ while these two distances are equivalent for measures with compact supports [19].

3 Slice-matching maps, compatibility and moment matching

Slice-matching schemes were first introduced by [16] to iteratively transport an initial measure to target measure. An almost sure convergence result of such iterative schemes has been shown in [20]. In this paper, we are interested in approximation properties of one step of this slice-matching procedure. In what follows, we present the definitions of the schemes and the associated *slice-matching maps*. We furthermore show that the mean and second moments of σ and μ are matched through one step of such schemes.

3.1 Slice-matching maps and compatibility

Definition 1 (Single-slice and matrix-slice matching, [16, 20]) Consider $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$, $\mu \in \mathcal{W}_2(\mathbb{R}^n)$ and a vector $\theta \in S^{n-1}$. The single-slice matching map from σ to μ is defined by

$$T_{\sigma,\mu;\theta}(x) = x + (T_{\sigma^\theta}^{\mu^\theta}(x \cdot \theta) - x \cdot \theta)\,\theta \tag{5}$$

where $T_{\sigma^\theta}^{\mu^\theta}$ is the optimal transport map between the 1-dimensional measures σ^θ and μ^θ obtained through projection by θ. If an orthonormal basis of $\mathbb{R}^n$ is used, the matrix-slice matching map from σ to μ is defined by

$$T_{\sigma,\mu;P}(x) = x + P \begin{bmatrix} T_{\sigma^{\theta_1}}^{\mu^{\theta_1}}(x \cdot \theta_1) - x \cdot \theta_1 \\ T_{\sigma^{\theta_2}}^{\mu^{\theta_2}}(x \cdot \theta_2) - x \cdot \theta_2 \\ \vdots \\ T_{\sigma^{\theta_n}}^{\mu^{\theta_n}}(x \cdot \theta_n) - x \cdot \theta_n, \end{bmatrix} = \sum_{i=1}^n T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}(x \cdot \theta_i)\,\theta_i \tag{6}$$

where $P = [\theta_1, \cdots, \theta_n]$ is an orthogonal matrix.

Remark 1 The motivation for the name slice-matching map is the following: If $\nu = (T_{\sigma,\mu;P})_\sharp\sigma$, then $\nu^{\theta_i} = \mu^{\theta_i}$ for $1 \leq i \leq n$, i.e. all slices are matched. Similar properties hold for $T_{\sigma,\mu;\theta}$. Moreover, the following relation between n-dimensional Wasserstein distance and one-dimensional Wasserstein distance of the corresponding slices holds:

$$W_2^2(\sigma, (T_{\sigma,\mu;P})_\sharp\sigma) = \sum_{i=1}^{n} W_2^2(\sigma^{\theta_i}, \mu^{\theta_i}), \tag{7}$$

see [20, Lemma 3.9]. An analogous result for empirical measures can be found in [19, Proposition 5.2.7].

Remark 2 The matrix-slice matching maps (as well as the single-slice and generalizations to $1 \leq j \leq n$ slices, see [20]) can be used to approximate a target measure μ by iteratively pushing a source measure σ_0:

$$\sigma_{k+1} = ((1-\gamma_k)\,\mathrm{id} + \gamma_k\, T_{\sigma_k,\mu;P_k})_\sharp\sigma_k, \quad k \geq 0, \tag{8}$$

where γ_k is a sequence of step-sizes and P_k are matrices in $O(n)$. When γ_k satisfies classical stochastic gradient descent assumptions [30] and P_k are chosen i.i.d. form the Haar measure on $O(n)$ (and some technical details are satisfied), then $\sigma_k \to \mu$ in both W_2 and SW_2 a.s. [20].

The paper [16] considers the above iterative scheme with $\gamma_k = 1$, whose convergence is however not covered by results of [20]. [16] shows convergence for special measures (the target is Gaussian) and in the KL-divergence.

In this paper, we study approximation properties of one step of (8) with $\gamma_k = 1$, i.e. we are interested in the relation between $\sigma_1 = (T_{\sigma_0,\mu;P})_\sharp\sigma_0$ and the target μ. An illustration of such approximations using different orthogonal matrices P is given in Fig. 1v.

It follows easily that both $T_{\sigma,\mu;\theta}$ and $T_{\sigma,\mu;P}$ are Brenier maps, though they are not necessarily optimal transport maps between σ and μ. The maps $T_{\sigma,\mu;\theta}$ and $T_{\sigma,\mu;P}$ are furthermore a special type of P*-compatible maps* [27, 31], which are defined by

$$\mathfrak{S}(P) = \left\{x \mapsto \sum_{i=1}^{n} f_i(x\cdot\theta_i)\theta_i \,:\, f_i : \mathbb{R} \to \mathbb{R} \text{ is increasing}\right\}, \tag{9}$$

for any fixed $P = [\theta_1, \ldots, \theta_n] \in O(n)$. An analog to (7) holds for P-compatible maps

$$W_2^2(\sigma, T_\sharp\sigma) = \sum_{i=1}^{n} W_2^2(\sigma^{\theta_i}, (T_\sharp\sigma)^{\theta_i}), \quad T \in \mathfrak{S}(P).$$

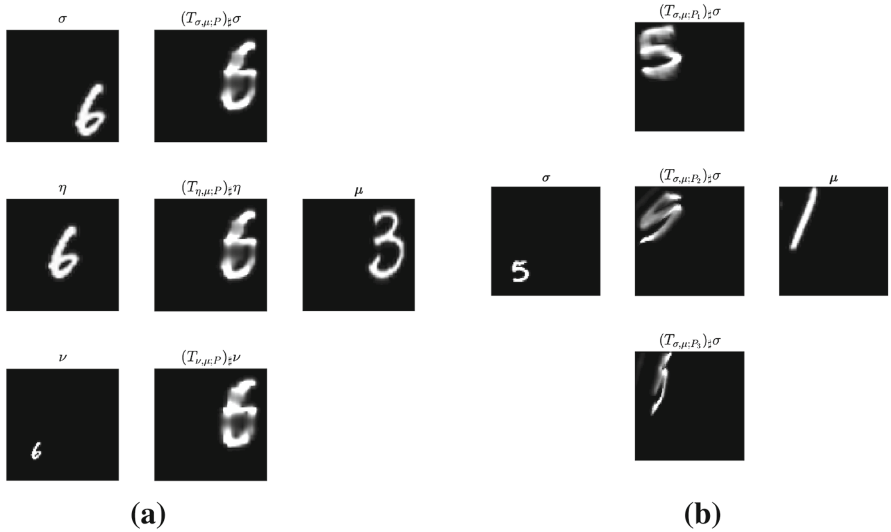

Fig. 1 Effects of matrix-slice matching. Left panel: Illustration of invariance to translation and scaling of the initial measures. Right panel: Illustration of slice-matching using different orthogonal matrices

Moreover, the slice-matching map can be viewed as the minimizer in $\mathfrak{S}(P)$ associated with the following minimization problem

$$T_{\sigma,\mu;P} = \underset{T\in\mathfrak{S}(P)}{\arg\min} \sum_{i=1}^{n} W_2^2((T_\sharp\sigma)^{\theta_i}, \mu^{\theta_i}).$$
$$\in \underset{T \text{ is Brenier}}{\arg\min} \sum_{i=1}^{n} W_2^2((T_\sharp\sigma)^{\theta_i}, \mu^{\theta_i}), \tag{10}$$

where (10) follows from the fact that for $T \in \mathfrak{S}(P)$, and therefore

$$\sum_{i=1}^{n} W_2^2((T_\sharp\sigma)^{\theta_i}, \mu^{\theta_i}) = \|T - T_{\sigma,\mu;P}\|_\sigma^2, \text{ which is minimal iff } T = T_{\sigma,\mu;P}.$$

The details of this statement are presented in Corollary 8.

P-compatible maps have been used in tangent space embeddings, which allow for linear separability of two classes of measures, see [27, 31–33]. These maps satisfy

$$\mathcal{P}_{\theta_i} \circ T = f_i \circ \mathcal{P}_{\theta_i}, \quad i = 1, \ldots, n, \tag{11}$$

where $T(x) = \sum_{i=1}^n f_i(x \cdot \theta_i)\theta_i \in \mathfrak{S}(P)$ and $P = [\theta_1, \ldots, \theta_n]$.

Next, we show that the property (11) also characterizes the set of P-compatible maps given in (9).

Proposition 3 *Let* $T \in \mathfrak{S}(P)$ *with* $T(x) = \sum_{i=1}^{n} f_i(x \cdot \theta_i)\theta_i$*, where* $[\theta_1, \ldots, \theta_n] = P \in O(n)$ *and* $f_i, i = 1, \ldots, n$ *are increasing functions. Then* $T = T_\sigma^\mu$ *for some measures* $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ *and* $\mu \in \mathcal{P}_{ac}(\mathbb{R}^n)$*, if and only if*

$$\mathcal{P}_{\theta_i} \circ T_\sigma^\mu = f_i \circ \mathcal{P}_{\theta_i}, \quad i = 1, \ldots, n. \tag{12}$$

Furthermore, in this case, $f_i = T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}$ *and* T *is a matrix-slice matching map.*

Proof For the equivalence, note that relation (12) is equivalent to

$$T(x) = \sum_{i=1}^{n} f_i(x \cdot \theta_i)\,\theta_i = \sum_{i=1}^{n} (\theta_i \cdot T_\sigma^\mu(x))\,\theta_i = T_\sigma^\mu(x),$$

since the columns of P are orthonormal.

For the other statement note that (11) implies that

$$\begin{aligned}\mu^{\theta_i} &= \left(\mathcal{P}_{\theta_i} \circ T\right)_\sharp \sigma = \left(f_i \circ \mathcal{P}_{\theta_i}\right)_\sharp \sigma \\ &= (f_i)_\sharp \sigma^{\theta_i}.\end{aligned}$$

Since f_i are increasing we obtain $f_i = T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}$. □

Remark 3 Note that the proof of Proposition 3 relies on the columns of P forming an orthonormal basis of $\mathbb{R}^n$. Therefore, this equivalence is not true for single-slice matching maps. We only have the implication $T_{\sigma,\mu;\theta} = T_\sigma^\mu \implies \mathcal{P}_\theta \circ T_\sigma^\mu = T_{\sigma^\theta}^{\mu^\theta} \circ \mathcal{P}_\theta$.

3.2 Moment matching

We show that slice-matching maps push the source measure to a measure that has the same mean and second moments as the target measure.

Proposition 4 *Let* $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ *and* $\mu \in \mathcal{W}_2(\mathbb{R}^n)$*. Then for any* $P \in O(n)$*, the following holds:*

$$E((T_{\sigma,\mu;P})_\sharp \sigma) = E(\mu) \tag{13}$$

$$M_2((T_{\sigma,\mu;P})_\sharp \sigma) = M_2(\mu) \tag{14}$$

Proof A direct computation shows that $\int T_{\sigma,\mu;P}(x)d\sigma(x) = \int y d\mu(y)$ and $\int \|T_{\sigma,\mu;P}(x)\|^2 d\sigma(x) = \int \|y\|^2 d\mu(y)$, see Lemma 22 for more details. □

The equal mean property (13) gives a first hint towards the shift-eliminating phenomenon of the slice-matching procedure. In particular, it can be verified that if T_σ^μ is a shift, then $(T_{\sigma,\mu;P})_\sharp \sigma = \mu$. A more comprehensive perspective of the shift-eliminating effect will be presented in (16) and (23).

4 Invariance, equivariance, and Lipschitz properties

We consider the following operator induced by the slice-matching maps of Definition 1:

Definition 2 (Slice-matching operator) We define the following operator based on the slice-matching approximation, concerning a source measure, a target measure, and slicing directions given by an orthogonal matrix:

$$\begin{aligned}\mathcal{U}:\ \mathcal{W}_{2,ac}(\mathbb{R}^n)\times\mathcal{W}_2(\mathbb{R}^n)\times O(n)&\to\mathcal{W}_2(\mathbb{R}^n)\\ (\sigma,\ \mu,\ P)&\mapsto(T_{\sigma,\mu;P})_\sharp\sigma.\end{aligned}$$

Note that $(T_{\sigma,\mu;P})_\sharp\sigma$ has finite second moment when $\mu\in\mathcal{W}_2(\mathbb{R}^n)$ by (14), which implies that $\mathcal{U}$ maps into $\mathcal{W}_2(\mathbb{R}^n)$. If absolute continuity of $(T_{\sigma,\mu;P})_\sharp\sigma$ is desired, one can further assume that both σ, μ are absolutely continuous, in which case $\mathcal{U}:\mathcal{W}_{2,ac}(\mathbb{R}^n)\times\mathcal{W}_{2,ac}(\mathbb{R}^n)\times O(n)\to\mathcal{W}_{2,ac}(\mathbb{R}^n)$, see Lemma 29.

We first illustrate the invariance and equivariance properties of the slice-matching operator in terms of basic transformations, i.e., shifts and scalings, which can be viewed as special cases of compatible transformations, as shown in Sect. 4.2. We also show how such properties are related to the recovery of optimal transport maps using matrix-slice matching. Moreover, as a complementary remark to the unifying convergence framework [20] of the single-slice and matrix-slicing matching schemes, we illustrate their differences via different recovery properties. A Lipschitz property in terms of the third component is shown in Sect. 4.3.

4.1 Invariance and equivariance with respect to shifts and scalings

We show that the slice-matching operator is invariant to actions induced by push-forward operations of shifts and scalings on any initial measure $\sigma\in\mathcal{W}_{2,ac}(\mathbb{R}^n)$ and is equivariant to actions of these maps on any target measure $\mu\in\mathcal{W}_2(\mathbb{R}^n)$, regardless of the orthogonal matrix P. More specifically,

Proposition 5 *Let $\sigma\in\mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\mu\in\mathcal{W}_2(\mathbb{R}^n)$. Then for any $P\in O(n)$ and $S(x)=ax+b$ where $a>0$, $b\in\mathbb{R}^n$, we get*

$$\mathcal{U}(S_\sharp\sigma,\mu,P)=\mathcal{U}(\sigma,\mu,P),\tag{15}$$

$$\mathcal{U}(\sigma,S_\sharp\mu,P)=S_\sharp\mathcal{U}(\sigma,\mu,P).\tag{16}$$

Proof Since $S\in\bigcap_{P\in O(n)}\mathfrak{S}(P)$, the conclusion follows from the corresponding invariance and equivariance properties in terms the group $\mathfrak{S}(P)$ of compatible transformations, see Proposition 7. ◻

An illustration of the invariance property the $\mathcal{U}$ with respect to translation-and-scaling transformations of the source measure is presented in Fig. 1a. Approximations with different choices of orthogonal matrices is illustrated in Fig. 1b.

Note that isotropic scalings and translations $S(x)=ax+b$ with $a>0$ and $b\in\mathbb{R}^n$ are special types of compatible maps. They satisfy $S\in\mathfrak{S}(P)$ for all $P\in O(n)$. An

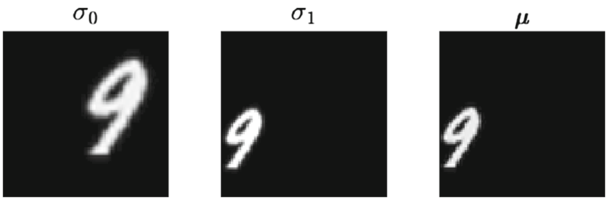

Fig. 2 Recovery of basic transformation (shift and scaling) using matrix-slice matching. The initial σ_0 (left) and target image μ (right) are related $T_\sharp\mu = \sigma_0$ where $T(x) = 1.6(x + [-35, 20]^t)$, where each image is of size 84 × 84. Here $\sigma_1 = (T_{\sigma_0,\mu;P})_\sharp\sigma_0$ (middle) is the image obtained via slice-matching transport between σ_0 and μ for an arbitrary orthogonal matrix P. Note $\sigma_1 \approx \mu$ up to numerical errors, as implied by Corollary 6

immediate corollary of the above proof is that, given two measures that are related by shifts and scalings, the target is recovered exactly by push the initial measure with the slice matching map $T_{\sigma,\mu;P}$ for any P.

Corollary 6 (Recovery of basic transformations with one step of slice matching map) *Given $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$, $\mu \in \mathcal{W}_2(\mathbb{R}^n)$ with $T_\sigma^\mu(x) = ax + b$ for some $a > 0$, $b \in \mathbb{R}^n$. Then we have $T_{\sigma,\mu;P} = T_\sigma^\mu$ and $\mathcal{U}(\sigma, \mu, P) = \mu$ for any $P \in O(n)$.*

Proof Let $S = T_\sigma^\mu$ in (15). Then $S_\sharp\sigma = \mu$ and $\mathcal{U}(S_\sharp\sigma, \mu, P) = (T_{\mu,\mu;P})_\sharp\mu = (\mathrm{id})_\sharp\mu = \mu$. Hence by (15), $\mathcal{U}(\sigma, \mu, P) = \mu$. The fact that $T_{\sigma,\mu;P} = T_\sigma^\mu$ follows from the fact that they are both Brenier maps pushing σ to μ. □

An illustration of Corollary 6 is presented in Fig. 2, the target image μ is matched (up to numerical errors) by its slice-matching approximation $(T_{\sigma,\mu;P})_\sharp\sigma$ for any P, if T_σ^μ is a translation-scaling function.

In addition, we can show that a differentiable map T connecting σ and μ can only be recovered with one step of the slice-matching scheme with any choice of P if and only if T is an isotropic scaling with translation:

Remark 4 Under the assumptions in Corollary 6, and if we further assume that T_σ^μ is differentiable, we obtain the following: $T = T_{\sigma,\mu;P}$ for any choice of $P \in O(n)$ if and only if $T(x) = ax + b$ for some $a > 0$ and $b \in \mathbb{R}^n$. One direction follows from Corollary 6. See Proposition 16 for details of the other direction.

The above recovery result holds for the matrix-slice scheme, but in general does not hold for single-slice schemes:

Example 4.1 Let $\mu = T_\sharp^b\sigma$ with $T^b(x) = x + b$, $b \neq 0 \in \mathbb{R}^n$. Unlike the matrix-slice matching, the target measure μ cannot always be recovered via $(T_{\sigma,\mu;\theta})_\sharp\sigma$ by the observing that $T_{\sigma,\mu;\theta}(x) = x + \theta(\theta \cdot b)$, for $\theta \in S^{n-1}$. To recover μ exactly, $\theta = \frac{1}{\|b\|}b$ is the only choice. Therefore, in general, we will not recover μ after one step.

Discussion A possible remedy is through an iterative scheme by repeating the above slice-matching procedure with different θ as introduced in Remark 2. In [20] we show that the iterative scheme $\sigma_{k+1} = ((1 - \gamma_k)\,\mathrm{id} + \gamma_k T_{\sigma_k,\mu;\theta_k})_\sharp\sigma_k$ with step size γ_k

satisfying $\sum \gamma_k = \infty$ and $\sum \gamma_k^2 < \infty$ converges a.s. ($\theta_k \overset{\text{i.i.d}}{\sim} u$, u uniform measure on S^{n-1}).

However, for the particular example of a shift, a.s. convergence can also be achieved for $\gamma_k = 1$. In Proposition 28 we give an elementary proof that $\sigma_{k+1} = (T_{\sigma_k,\mu;\theta_k})_\sharp \sigma_k$ converges to $\mu = T^b_\sharp \sigma$ a.s. with respect to the W_2-distance.

4.2 Invariance and equivariance with compatible maps

In this subsection, we discuss the invariance and equivariance properties of the slice-matching operator $\mathcal{U}$ as defined in Definition 2 concerning compatible transformations, defined in (9).

Proposition 7 *Let $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\mu \in \mathcal{W}_2(\mathbb{R}^n)$. For any $T \in \mathfrak{S}(P)$, where $P \in O(n)$, we have*

$$\mathcal{U}(T_\sharp\sigma, \mu, P) = \mathcal{U}(\sigma, \mu, P), \tag{17}$$

$$\mathcal{U}(\sigma, T_\sharp\mu, P) = T_\sharp\mathcal{U}(\sigma, \mu, P). \tag{18}$$

Proof Let $T(x) = Pf(P^t x)$, where $f(x) = (f_1(x_1), \cdots, f_n(x_n))$ with each f_i being increasing and $P = [\theta_1, \cdots, \theta_n]$. Let $\sigma_T = T_\sharp\sigma$. By (11), we get $\sigma_T^{\theta_i} = f_{i\,\sharp}\sigma^{\theta_i}$. It follows from the fact that f_i is increasing that $T_{\sigma^{\theta_i}}^{\mu^{\theta_i}} = T_{\sigma_T^{\theta_i}}^{\mu^{\theta_i}} \circ f_i$. Denote $g(x) = [(T_{\sigma_T^{\theta_1}}^{\mu^{\theta_1}} \circ f_1)(x_1), \cdots, (T_{\sigma_T^{\theta_n}}^{\mu^{\theta_n}} \circ f_n)(x_n)]^t$ and $g_T(x) = [T_{\sigma_T^{\theta_1}}^{\mu^{\theta_1}}(x_1), \cdots, T_{\sigma_T^{\theta_n}}^{\mu^{\theta_n}}(x_n)]^t$. Then

$$\begin{aligned}(T_{\sigma,\mu;P})_\sharp\sigma &= (Pg \circ P^t)_\sharp\sigma = (Pg_T \circ f \circ P^t)_\sharp\sigma \\ &= (pg_T \circ P^t \circ P \circ f \circ P^t)_\sharp\sigma = (T_{\sigma_T,\mu;P} \circ T)_\sharp\sigma \\ &= (T_{\sigma_T,\mu;P})_\sharp\sigma_T.\end{aligned}$$

This proves (17). For (18), denote $\mu_T = T_\sharp\mu$. With similar reasoning as before and with the observations that $T_{\sigma^{\theta_i}}^{\mu_T^{\theta_i}} = f_i \circ T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}$, we get

$$T_{\sigma,\mu_T;P} = Pf \circ P^t \circ T_{\sigma,\mu;P} = T \circ T_{\sigma,\mu;P}.$$

□

As a direct conclusion, define $\mathfrak{S}(P)_\sharp\eta := \{T_\sharp\eta : T \in \mathfrak{S}(P)\}$, then $\mathcal{U}(\mathfrak{S}(P)_\sharp\sigma, \mu, P) = \{\mathcal{U}(\sigma, \mu, P)\}$ is a singleton set, as illustrated in Fig. 1a, and $\mathcal{U}(\sigma, \mathfrak{S}(P)_\sharp\mu, P) = \mathfrak{S}(P)_\sharp\mathcal{U}(\sigma, \mu, P)$.

Corollary 8 *Let $\sigma, \mu \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$. Then*

$$\sum_{i=1}^n W_2^2((T_\sharp\sigma)^{\theta_i}, \mu^{\theta_i}) = \|T - T_{\sigma,\mu;P}\|_\sigma^2,$$

where $T \in \mathfrak{S}(P)$ *and* $P \in O(n)$.

Proof Rewriting (7), we have

$$\sum_{i=1}^{n} W_2^2(\sigma^{\theta_i}, \mu^{\theta_i}) = W_2^2(\sigma, \mathcal{U}(\sigma, \mu, P)).$$

Hence

$$\begin{aligned}\sum_{i=1}^{n} W_2^2((T_\sharp\sigma)^{\theta_i}, \mu^{\theta_i}) &= W_2^2(T_\sharp\sigma, \mathcal{U}(T_\sharp\sigma, \mu, P)) = W_2^2(T_\sharp\sigma, \mathcal{U}(\sigma, \mu, P)) \\ &= W_2^2(T_\sharp\sigma, (T_{\sigma,\mu;P})_\sharp\sigma) = \|T - T_{\sigma,\mu;P}\|_\sigma^2,\end{aligned}$$

where the second step uses the invariance property (17) and the last steps makes uses of the isometry property with respect to P-compatible maps T and $T_{\sigma,\mu;P}$, see [27]. □

4.3 Recovery and stability properties of matrix-slicing matching

When two measures are related by a P-compatible map, the optimal map between them can be recovered exactly by the (P)-matrix-slice matching procedure. This is in contrast to shifts and scalings, where an arbitrary orthogonal matrix can be used for recovery. The following corollary to Proposition 7 summarized this result.

Corollary 9 (Recovery of P-compatible transformations with one step of P-slice-matching) *Given* $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$, $\mu \in \mathcal{W}_2(\mathbb{R}^n)$ *with* $T_\sigma^\mu \in \mathfrak{S}(P)$ *for some* $P \in O(n)$. *Then we have* $\mathcal{U}(\sigma, \mu, P) = \mu$ *and* $T_{\sigma,\mu;P} = T_\sigma^\mu$.

Proof With Proposition 7, the above result follows from similar arguments as in Corollary 6. □

To recover a compatible map T with one step of the iteration, Corollary 6 implies that we need to know the orthogonal matrix P. The following Lipschitz continuity of the slice-matching operator $\mathcal{U}$ with respect to P establishes a stability result on the choice of P:

Proposition 10 *Let* $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ *and* $\mu \in \mathcal{W}_2(\mathbb{R}^n)$. *Assume that there exists* $L > 0$ *such that* $T_{\sigma^\theta}^{\mu^\theta}$ *is L-Lipschitz on* $\mathbb{R}$ *for all* $\theta \in S^{n-1}$. *Then*

$$\begin{aligned}W_2(\mathcal{U}(\sigma, \mu, P), \mathcal{U}(\sigma, \mu, Q)) &= \|T_{\sigma,\mu;P} - T_{\sigma,\mu;Q}\|_\sigma \\ &\le (3L+1)C\|P - Q\|_F,\end{aligned} \tag{19}$$

where $C = \max\{M_2(\sigma), M_2(\mu)\}$ *and* $\|\cdot\|_F$ *denotes the Frobenius norm.*

Proof See A.2. □

Remark 5 Inequality (19) can be viewed a stability result for one step of the iterative schemes described in Remark 2. If $T_\sigma^\mu \in \mathfrak{S}(P)$ for some $P \in O(n)$, then the push-forward measure of σ using a slice-matching map associated with $Q \in O(n)$ is within $(3L+1)C\|P-Q\|_F$ in Wasserstein distance to the target μ. Picking a Q close to P is good enough to obtain an approximation of μ by $\sigma_1 := (T_{\sigma,\mu;Q})_\sharp\sigma$. Note that if $Q = P$, then $\sigma_1 = \mu$, which also follows from Corollary 9.

The stability result in Proposition 10 shows how well the target μ can be approximated by the slice-matching approximation $(T_{\sigma,\mu;P})_\sharp\sigma$ when σ and μ are related by some compatible map $T_{\sigma,\mu;Q}$. Additionally, we will now show that if σ and μ are related by a map which is an ε-perturbation of shifts and scalings, then μ can be approximated by its slice-matching approximation with at most 2ε error (Remark 6). This can also be viewed as an extension to the recovery result in Corollary 6.

Proposition 11 *Let $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\mu \in \mathcal{W}_2(\mathbb{R}^n)$. Then for any $P \in O(n)$,*

$$W_2((T_{\sigma,\mu;P})_\sharp\sigma, \mu) \leq 2 \inf_{S\in\mathcal{S}} W_2(S_\sharp\sigma, \mu),$$

where $\mathcal{S} := \{S(x) = ax + b \mid a > 0, b \in \mathbb{R}^n\}$.

Proof Since $W_2(S_\sharp\sigma, \mu) = \|S - T_\sigma^\mu\|_\sigma$, it suffices to show that for any $S \in \mathcal{S}$,

$$W_2((T_{\sigma,\mu;P})_\sharp\sigma, \mu) \leq 2\|S - T_\sigma^\mu\|_\sigma.$$

By the Lipschitz property (see e.g., [26, Eq. (2.1)]) associated with W_2 and triangle inequality, we have

$$\begin{aligned} W_2((T_{\sigma,\mu;P})_\sharp\sigma, \mu) &\leq \|T_{\sigma,\mu;P} - T_\sigma^\mu\|_\sigma \\ &\leq \|T_{\sigma,\mu;P} - S\|_\sigma + \|S - T_\sigma^\mu\|_\sigma. \end{aligned} \tag{20}$$

Next we bound the first term. Since $S \in \mathfrak{S}(P)$ for any $P \in O(n)$, it follows from Corollary 8 that

$$\|S - T_{\sigma,\mu;P}\|_\sigma^2 = \sum_{i=1}^n W_2^2((S_\sharp\sigma)^{\theta_i}, \mu^{\theta_i}) \leq W_2^2(S_\sharp\sigma, \mu) = \|S - T_\sigma^\mu\|_\sigma^2$$

where the bound follows from Lemma 24 and the last equality follows from isometry properties with respect to transformations in $\mathcal{S}$ [33]. The desired inequality hence follows from (20). □

Remark 6 Assume that $\mu = (f \circ S)_\sharp\sigma$, where $f : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\|f - \mathrm{id}\|_{S_\sharp\sigma} \leq \varepsilon$ for some $\varepsilon > 0$ and $S(x) = ax + b$, $a > 0$, $b \in \mathbb{R}^n$. Then

$$W_2((T_{\sigma,\mu;P})_\sharp\sigma, \mu) \leq 2\varepsilon.$$

Remark 7 Using essentially the same arguments as in the proof of Proposition 11, one can show that for any $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $\mu = T_\sharp\sigma$,

$$\|T - T_{\sigma,\mu;P}\|_\sigma \le 2 \inf_{S\in\mathcal{S}} \|T - S\|_\sigma.$$

Note T is not necessarily the optimal transport map between σ and μ.

5 Affine effects and registration problems

We study two basic image and point-cloud registration problems to understand the effects of the slice-matching maps (6). Image registration [34] involves matching images with variations caused by differences in acquisition, object growth or other changes. It plays a fundamental role in image processing, particularly in medical image applications [35]. Modeling image data and shape with probability measures has paved the way for robust and scalable algorithms by leveraging the theory optimal transport, such as diffeomorphic registration methods [21, 22], including point cloud registration [23].

We have shown that slice-matching maps can be used to register translation-and-scaling deformations exactly, see Corollary 6 and Fig. 2. We also showed that if the two measures are related by perturbations of translations and scalings, the registration error is bounded by the the size of this perturbation, see Proposition 11. To gain a better understanding of how the slice-matching procedure incorporates affine effects, particularly when the initial and target measures significantly differ–meaning they are not merely small perturbations of translations and scalings– we compare the registration maps aimed at the target μ with those directed at its slice-matching approximation $\mathcal{U}(\sigma, \mu, P)$, see Fig. 3. Specifically, we demonstrate that registration maps involving only translations are identical (Proposition 12), and those involving translations and isotropic scalings are comparable (Proposition 13).

Let $D(\cdot, \cdot)$ be a distance between probability measures, e.g., W_2 or SW_2. We study registration problems with the following subsets of affine transformation: $\mathcal{S}_t := \{S(x) = x + b \mid b \in \mathbb{R}^n\}$ (the set of translations), $\mathcal{S} := \{S(x) = ax + b \mid a > 0, b \in \mathbb{R}^n\}$ (the set of compositions of isotropic scalings and translations):

$$S_t^{\sigma,\eta,D} := \arg\min_{S\in\mathcal{S}_t} D(S_\sharp\sigma, \eta), \tag{21}$$

$$S^{\sigma,\eta,D} := \arg\min_{S\in\mathcal{S}} D(S_\sharp\sigma, \eta), \tag{22}$$

the existence and uniqueness of which will be addressed later.

We show that the optimal translation registration maps from the initial measure to the target and to slice-matching approximation of the target are identical.

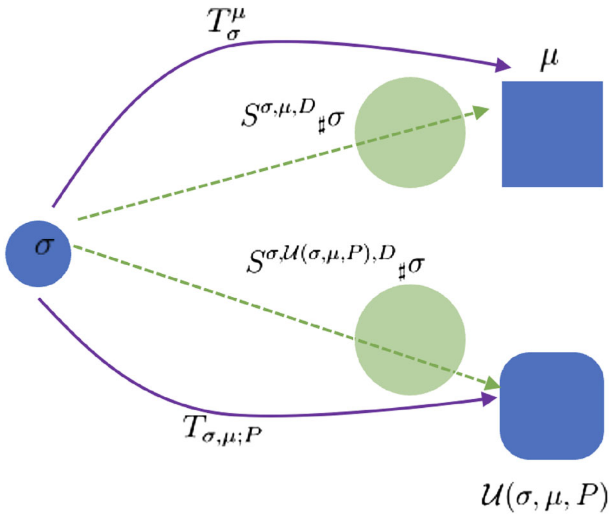

Fig. 3 Illustration of registration problems with translation-and-scaling function $S^{\sigma,\eta,D}$ from Proposition 13, where $\eta = \mu$ or $\eta = \mathcal{U}(\sigma, \mu, P)$. As indicated in (25), $S^{\sigma,\eta,D}{}_\sharp\sigma$ and η have the same mean; however for better visualization, we intentionally kept them separate in this Fig.

Proposition 12 (Registration with translations (21)) *Let $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\mu \in \mathcal{W}_2(\mathbb{R}^n)$. Then for any $P \in O(n)$, the unique minimizer in (21) satisfies*

$$S_t^{\sigma,\mu,W_2} = S_t^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}. \tag{23}$$

Proof The minimization problem (21) is a quadratic problem in the parameter b, which can be solved by taking partial derivatives with respect to b, and setting them to 0. From this, the existence and uniqueness follows immediately. Calculations are summarized in the proof of Proposition 18 and Corollary 20. In particular, the arguments show that the optimal parameters $b^{W_2}, \widetilde{b}^{W_2}$ for S_t^{σ,μ,W_2} and $S_t^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}$ respectively, satisfy

$$b^{W_2} = E(\mu) - E(\sigma), \quad \widetilde{b}^{W_2} = E(\mathcal{U}(\sigma, \mu, P)) - E(\sigma).$$

The conclusion hence follows from the fact that $E(\mathcal{U}(\sigma, \mu, P)) = E(\mu)$, see (13). □

Remark 8 By similar calculations, one can also show that

$$S_t^{\sigma,\mu,SW_2} = S_t^{\sigma,\mathcal{U}(\sigma,\mu,P),SW_2}.$$

See expressions b^{SW_2} and $\widetilde{b}^{SW_2}$ in Proposition 18 and Corollary 20.

The above registration result illustrates the idea of how shifts are eliminated by the slice-matching procedure. When considering registration involving translations and isotropic scalings measured by the W_2 distance, the following comparison holds:

Proposition 13 (Registration with translation-and-scalings (22)) *Let $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\mu \in \mathcal{W}_2(\mathbb{R}^n)$. Assume that (i) The convex potential ϕ such that $\nabla\phi = T_\sigma^\mu$ given by Brenier's theorem is differentiable at $E(\sigma)$, and (ii) For any $\lambda \in (0,1)$, $\phi((1-\lambda)y+\lambda E(\sigma)) < (1-\lambda)\phi(y)+\lambda\phi(E(\sigma))$ for all y in some ball $B(x,r)$, where x lies in the support of σ. Then S^{σ,μ,W_2} and $S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}$ in (22) are well-defined and unique, and satisfy the following*

$$\begin{aligned} W_2(S^{\sigma,\mu,W_2}{}_\sharp\sigma, S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}{}_\sharp\sigma) &= \|S^{\sigma,\mu,W_2} - S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}\|_\sigma \\ &= \frac{W_2^2(\sigma,\mu) - \sum_{i=1}^n W_2^2(\sigma^{\theta_i},\mu^{\theta_i})}{2\sqrt{M_2(\sigma) - \|E(\sigma)\|^2}}, \end{aligned} \tag{24}$$

where $P = [\theta_1, \ldots, \theta_n]$. In particular,

$$S^{\sigma,\mu,W_2} = S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2} \quad \textit{iff} \quad W_2^2(\sigma,\mu) = \sum_{i=1}^n W_2^2(\sigma^{\theta_i},\mu^{\theta_i}).$$

Moreover, the registrations eliminate the effects of translation in the following sense

$$E(S^{\sigma,\mu,W_2}{}_\sharp\sigma) = E(S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}{}_\sharp\sigma) = E(\mu). \tag{25}$$

Proof Similar to the proof of Proposition 12, the minimization problems (22) are quadratic in terms of parameters a and b, where $S(x) = ax + b$. By taking the partial derivatives and setting them to zero, checking the Hessian matrix, together with the assumption of the proposition, we obtain existence and uniqueness of the minimizers. The equalities then follow from direct computations. See Proof 1 for details. □

Remark 9 With (7), Proposition 13 implies

$$S^{\sigma,\mu,W_2} = S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2} \quad \text{iff} \quad W_2^2(\sigma,\mu) = W_2^2(\sigma,\mathcal{U}(\sigma,\mu,P)).$$

Corollary 6 and Corollary 9 show special cases where maps S^{σ,μ,W_2} and $S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}$ are equal. However, for the case of registration with only shifts, $S_t^{\sigma,\mu} = S_t^{\sigma,\mathcal{U}(\sigma,\mu,P)}$ always holds, see (23).

We further establish a connection between the Wasserstein distance and sliced-Wasserstein distance by comparing the registration maps in (22). This insight holds the potential to enhance our understanding of the distinction between sliced-Wasserstein and Wasserstein flows, as demonstrated in a special case by Bonet et al. [15, p.7, Eq. 19].

Corollary 14 *Let $\mathbf{P}$ be a random variable corresponding to the Haar probability measure u_n on the orthogonal group $O(n)$. For fixed $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\mu \in \mathcal{W}_2(\mathbb{R}^n)$, we have*

$$\|S^{\sigma,\mu,W_2} - \mathbb{E}S^{\sigma,\mathcal{U}(\sigma,\mu,\mathbf{P}),W_2}\|_\sigma = \mathbb{E}\|S^{\sigma,\mu,W_2} - S^{\sigma,\mathcal{U}(\sigma,\mu,\mathbf{P}),W_2}\|_\sigma$$

$$= \frac{W_2^2(\sigma,\mu) - nSW_2^2(\sigma,\mu)}{2\sqrt{M_2(\sigma) - \|E(\sigma)\|^2}} \geq 0.$$

Proof By (27), (28), (24) and calculations in Proposition 18, both equations reduce to the following:

$$\int_{O(n)} \sum_{i=1}^{n} W_2^2(\sigma^{\theta_i}, \mu^{\theta_i}) du_n(P) = nSW_2^2(\sigma,\mu), \tag{26}$$

which can be observed by an explicit geometric construction of the Haar measure on $O(n)$ (see e.g., [36, p.19]). □

In light of the inherent connection between $\sum_{i=1}^n W_2^2(\sigma^{\theta_i}, \mu^{\theta_i})$ and $SW_2^2(\sigma,\mu)$ as shown in (26), and the minimization problem (10) associated with $T_{\sigma,\mu;P}$, we demonstrate a similar connection between the registration maps associated with two distinct registration problems concerning the W_2 and SW_2 distances respectively:

Corollary 15 *Given the same assumptions as in Proposition 13, we have*

$$\|S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2} - S^{\sigma,\mu,SW_2}\|_\sigma = \frac{nSW_2^2(\sigma,\mu) - \sum_{i=1}^n W_2^2(\sigma^{\theta^i}, \mu^{\theta_i})}{\sqrt{M_2(\sigma) - \|E(\sigma)\|^2}}.$$

Moreover, let $\mathbf{P}$ *be a random variable corresponding to the Haar probability measure* u_n *on the orthogonal group* $O(n)$*, then*

$$\mathbb{E}S^{\sigma,\mathcal{U}(\sigma,\mu,\mathbf{P}),W_2} = S^{\sigma,\mu,SW_2}.$$

Proof The proof follows directly from Proposition 18 and Corollary 20 and similar calculations as in (27) and (28). □

In simpler terms, the map that optimally aligns σ with $\mathcal{U}(\sigma,\mu,\mathbf{P})$ considering shifts and scalings in the W_2 distance is, on average, the same map that optimally aligns σ with μ in the SW_2 distance.

Remark 10 Similar to the registration problems (22) and (21), one can study registration in terms of maps $\mathcal{S}(P) := \{x \mapsto P\Lambda P^t x + b : \Lambda \text{ is positive and diagonal}, b \in \mathbb{R}^n\}$. We provide a summary in Proposition 21.

6 Conclusion

Optimal transport-based slice-matching schemes benefit from closed-form formulations, computational efficiency and convergence guarantees. In the present paper, we are interested in the approximation power of one step of such schemes, both on the level of measures and maps. This can be considered as a step towards understanding to what extend slice-matching maps can serve as effective alternatives to optimal transport maps.

We investigate the exact recovery of basic transformations, such as translations and scalings, as well as the approximate recovery of perturbations of such transformations. These results are derived by studying invariance properties of an associated slice-matching operator. In addition, we explore equivariance and Lipschitz properties of the same operator, to understand how it incorporates actions of basic transformations on the target measure, as well as perturbations on the slicing directions.

We provide a quantitative perspective on how slice-matching procedures encode special affine transformations in their approximations through the study of basic registration problems. These registration problems potentially also offer insights into the relationship between Wasserstein and sliced-Wasserstein flows, which is an interesting problem for future research.

Appendix A Proofs for Sect. 4

A.1 Key facts for proof of Remark 4

Proposition 16 *Let $\mathcal{D}(\mathbb{R}^n)$ be the set of differentiable vector fields from $\mathbb{R}^n$ to $\mathbb{R}^n$.*

$$\Big(\bigcap_{P\in O(n)} \mathfrak{S}(P)\Big) \cap \mathcal{D}(\mathbb{R}^n) = \{x \mapsto ax + b : a > 0 \text{ and } b \in \mathbb{R}^n\}.$$

Proof For the proof, we need to show that a differentiable vector field $S \in \bigcap_{P\in O(n)} \mathfrak{S}(P)$ is an isotropic scaling with translation. Choose $P \in O(n)$ and write $S(x) = \sum_{i=1}^n f_i^P(x\cdot\theta_i)\theta_i$ with $P = [\theta_1, \ldots, \theta_n]$. Note that using the standard basis, we can also write $S(x) = \sum_{i=1}^n g_i(x_i)e_i$. Computing the Jacobian of S with respect to the two basis representations, we obtain

$$\begin{bmatrix} g_1'(x_1) & & \\ & \ddots & \\ & & g_n'(x_n) \end{bmatrix} = P \begin{bmatrix} f_1^{P'}(x\cdot\theta_1) & & & \\ & f_2^{P'}(x\cdot\theta_2) & & \\ & & \ddots & \\ & & & f_n^{P'}(x\cdot\theta_n) \end{bmatrix} P^t.$$

Hence the two diagonal matrices above have the same diagonal entries, allowing for a possible reordering of the entries. Without loss of generality, we assume that $g_i'(x_i) = f_i^{P'}(x\cdot\theta_i)$, $i = 1, ..., n$ by possibly performing a column permutation of P and renaming f_i^P's. Choosing an orthogonal matrix P such that one of its column θ_i with all entries being non-zero, one can immediately derive that the diagonal entries $g_i'(x_i)$'s are the same for any fixed x. In summary,

$$g_1'(x_1) = \cdots = g_n'(x_n) = f_1^{P'}(x\cdot\theta_1) = \cdots = f_n^{P'}(x\cdot\theta_n) = a_x,$$

where a_x is a constant depending on $x = [x_1, \cdots, x_n]^t \in \mathbb{R}^n$. Since the diagonal element $g_i'(x_i)$ only depends on x_i, it follows that a_x is a constant independent of x. Hence $S(x) = ax + b$ for some $a > 0, b \in \mathbb{R}^n$. □

Remark 11 In general, if $T \in \bigcap_{P \in O(n)} \mathfrak{S}(P)$ is differentiable on an open set $\Omega \subseteq \mathbb{R}^n$, the $T|_\Omega : \Omega \to \mathbb{R}^n$ is an isotropic scaling with translation. In particular, $\bigcap_{P \in O(n)} \mathfrak{S}(P)$ include some piecewise isotropic scalings with translations.

A.2 Proof of Proposition 10

We need the following proposition to derive the proof of Proposition 10:

Proposition 17 *Consider two angles $\theta, \nu \in S^{n-1}$, and assume that $T_{\sigma^\nu}^{\mu^\nu}$ is L-Lipschitz for all ν, i.e. there exists $L > 0$ such that $|T_{\sigma^\nu}^{\mu^\nu}(x) - T_{\sigma^\nu}^{\mu^\nu}(y)| \le L|x - y|$ for $x, y \in \mathbb{R}$ and $\nu \in S^{n-1}$, then*

$$\|T_{\sigma^\theta}^{\mu^\theta} \circ \mathcal{P}_\theta - T_{\sigma^\nu}^{\mu^\nu} \circ \mathcal{P}_\nu\|_\sigma \le (2L + 1)C\|\theta - \nu\|_2,$$

where C is the max over the second moments of σ resp. μ.

Proof

$$\|T_{\sigma^\theta}^{\mu^\theta} \circ \mathcal{P}_\theta - T_{\sigma^\nu}^{\mu^\nu} \circ \mathcal{P}_\nu\|_\sigma \le \|T_{\sigma^\theta}^{\mu^\theta} \circ \mathcal{P}_\theta - T_{\sigma^\theta}^{\mu^\nu} \circ \mathcal{P}_\theta\|_\sigma + \|T_{\sigma^\theta}^{\mu^\nu} \circ \mathcal{P}_\theta - T_{\sigma^\nu}^{\mu^\nu} \circ \mathcal{P}_\nu\|_\sigma = (\diamond).$$

We bound these separately.

$$\begin{aligned}
\|T_{\sigma^\theta}^{\mu^\theta} \circ \mathcal{P}_\theta - T_{\sigma^\theta}^{\mu^\nu} \circ \mathcal{P}_\theta\|_\sigma &= \|T_{\sigma^\theta}^{\mu^\theta} - T_{\sigma^\theta}^{\mu^\nu}\|_{\sigma^\theta} = W_2\left(\mu^\theta, \mu^\nu\right) \\
&\le \|\mathcal{P}_\theta - \mathcal{P}_\nu\|_\mu = \left(\int_{\mathbb{R}^n} |\mathcal{P}_\theta(x) - \mathcal{P}_\nu(x)|^2 \, d\mu(x)\right)^{1/2} \\
&= \left(\int_{\mathbb{R}^n} |(\theta - \nu) \cdot x|^2 \, d\mu(x)\right)^{1/2} \\
&\le \|\theta - \nu\|_2 \left(\int_{\mathbb{R}^n} \|x\|^2 \, d\mu(x)\right)^{1/2} \\
&\le C\|\theta - \nu\|_2,
\end{aligned}$$

with C max of the second moments, which is bounded by assumption. Now for the second part, note that on $\mathbb{R}$ we have $T_{\sigma^\theta}^{\mu^\nu} = T_{\sigma^\nu}^{\mu^\nu} \circ T_{\sigma^\theta}^{\sigma^\nu}$

$$\begin{aligned}
&\|T_{\sigma^\theta}^{\mu^\nu} \circ \mathcal{P}_\theta - T_{\sigma^\nu}^{\mu^\nu} \circ \mathcal{P}_\nu\|_\sigma \\
&= \left(\int_{\mathbb{R}^n} |T_{\sigma^\nu}^{\mu^\nu}(T_{\sigma^\theta}^{\sigma^\nu}(\mathcal{P}_\theta(x))) - T_{\sigma^\nu}^{\mu^\nu}(\mathcal{P}_\nu(x))|^2 \, d\sigma(x)\right)^{1/2} = (\star)
\end{aligned}$$

Since $T_{\sigma^\nu}^{\mu^\nu}$ is L-Lipschitz, we get

$$
\begin{aligned}
(\star) &\leq L \left(\int_{\mathbb{R}^n} |T_{\sigma^\theta}^{\sigma^\nu}(\mathcal{P}_\theta(x)) - \mathcal{P}_\nu(x)|^2 \, d\sigma(x) \right)^{1/2} = L \| T_{\sigma^\theta}^{\sigma^\nu} \circ \mathcal{P}_\theta - \mathcal{P}_\nu \|_\sigma \\
&\leq L \left(\| T_{\sigma^\theta}^{\sigma^\nu} \circ \mathcal{P}_\theta - \mathcal{P}_\theta \|_\sigma + \| \mathcal{P}_\theta - \mathcal{P}_\nu \|_\sigma \right) \\
&\leq L \left(\| T_{\sigma^\theta}^{\sigma^\nu} - \mathrm{id} \, \|_{\sigma^\theta} + C \| \theta - \nu \|_2 \right) \\
&= L \left(W_2(\sigma^\theta, \sigma^\nu) + C \| \theta - \nu \|_2 \right) \\
&\leq L \left(\| \mathcal{P}_\theta - \mathcal{P}_\nu \|_\sigma + C \| \theta - \nu \|_2 \right) \\
&\leq 2LC \| \theta - \nu \|_2
\end{aligned}
$$

This implies

$$
(\diamond) \leq (2L+1) C \| \theta - \nu \|_2 .
$$

□

Proof of Proposition 10 Based on (9), we let $T_{\sigma,\mu;P} = PD \circ P^t$ where $D(x) = [T_{\sigma^{\theta_1}}^{\mu^{\theta_1}}(x_1), T_{\sigma^{\theta_2}}^{\mu^{\theta_2}}(x_2), \cdots, T_{\sigma^{\theta_n}}^{\mu^{\theta_n}}(x_n)]^t$ for $x \in \mathbb{R}^n$ and $P = [\theta_1, \ldots, \theta_n]$. Similarly, we let and $T_{\sigma,\mu;Q} = Q\widetilde{D} \circ Q^t$, with $Q = [\nu_1, \ldots, \nu_n]$. We continue with deriving the bound:

$$
\begin{aligned}
\| T_{\sigma,\mu;P} - T_{\sigma,\mu;Q} \|_\sigma &= \| PDP^t - Q\widetilde{D}Q^t \|_\sigma \\
&\leq \| PDP^t - P\widetilde{D}Q^t \|_\sigma + \| P\widetilde{D}Q^t - Q\widetilde{D}Q^t \|_\sigma \\
&= (1) + (2).
\end{aligned}
$$

We bound the two terms seperately. For (1), using Proposition 17, we get

$$
\begin{aligned}
\| PDP^t - P\widetilde{D}Q^t \|_\sigma^2 &= \int_{\mathbb{R}^n} \| D(P^t x) - \widetilde{D}(Q^t x) \|_2^2 \, d\sigma(x) \\
&= \sum_{i=1}^n \int_{\mathbb{R}^n} | T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}((P^t x)_i) - T_{\sigma^{\nu_i}}^{\mu^{\nu_i}}((Q^t x)_i) |^2 \, d\sigma(x) \\
&= \sum_{i=1}^n \| T_{\sigma^{\theta_i}}^{\mu^{\theta_i}} \circ \mathcal{P}_{\theta_i} - T_{\sigma^{\nu_i}}^{\mu^{\nu_i}} \circ \mathcal{P}_{\nu_i} \|_\sigma^2 \\
&\leq ((2L+1)C)^2 \sum_{i=1}^n \| \theta_i - \nu_i \|_2^2 \\
&= ((2L+1)C)^2 \| P - Q \|_F^2 .
\end{aligned}
$$

For (2) we get

$$\begin{aligned}\|P\widetilde{D}Q^t - Q\widetilde{D}Q^t\|_\sigma^2 &= \int_{\mathbb{R}^n} \|(P-Q)\widetilde{D}(Q^t x)\|_2^2\, d\sigma(x)\\ &\le \|P-Q\|_2^2 \int_{\mathbb{R}^n} \|\widetilde{D}(Q^t x)\|_2^2\, d\sigma(x)\\ &\le \|P-Q\|_2^2\, L^2 \int_{\mathbb{R}^n} \|Q^t x\|_2^2\, d\sigma(x) \le \|P-Q\|_2^2\, L^2 C^2\\ &\le \|P-Q\|_F^2\, L^2 C^2\end{aligned}$$

Combining (1) and (2) gives the final bound. □

Appendix B Proofs for Sect. 5

Proof of Proposition 13 Let $S^{\sigma,\mu,W_2}(x) = a^{W_2}x + b^{W_2}$ and $S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}(x) = \widetilde{a}^{W_2}x + \widetilde{b}^{W_2}$ be the critical functions for the associated minimization problem (22). By Proposition 18 and Corollary 20, we have

$$\widetilde{a}^{W_2} - a^{W_2} = \frac{W_2^2(\sigma,\mu) - \sum_{i=1}^n W_2^2(\sigma^{\theta_i},\mu^{\theta_i})}{2(M_2(\sigma) - \|E(\sigma)\|^2)}, \tag{27}$$

$$\widetilde{b}^{W_2} - b^{W_2} = -(\widetilde{a}^{W_2} - a^{W_2})E(\sigma), \tag{28}$$

and the norm bound $\|S^{\sigma,\mu,W_2} - S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}\|_\sigma$ in (24) can be obtained via direct computation and the fact that the RHS is non-negative, see Lemma 24. It is left to show that these critical functions are indeed the minimizers by verifying

1. $\widetilde{a}^{W_2} \ge a^{W_2} > 0$, see Lemmas 24, 26, and 27.
2. The Hessian associated $H(a,b)$ with both the minimization problems are positive definite by a direct calculation and Lemma 26, where

$$H(a,b) = 2\begin{bmatrix} M_2(\sigma) & (E(\sigma))^t \\ E(\sigma) & I_{n-1}\end{bmatrix}.$$

   Here I_{n-1} denotes the identity matrix of size $(n-1)\times(n-1)$.

The equality concerning the means follows from Corollary 19. □

Proposition 18 *Let S^{σ,η,W_2} and S^{σ,η,SW_2} correspond to the critical points of the minimization problems in (22) and (30), respectively. Then the corresponding parameters satisfy*

$$\begin{aligned} a^{W_2} &= \frac{\frac{1}{2}(M_2(\eta) + M_2(\sigma) - W_2^2(\sigma,\eta)) - E(\sigma)\cdot E(\eta)}{M_2(\sigma) - \|E(\sigma)\|^2},\\ b^{W_2} &= E(\eta) - a^{W_2}E(\sigma),\end{aligned}$$

$$
\begin{aligned}
a^{SW_2} &= \frac{\frac{1}{2}(M_2(\eta) + M_2(\sigma) - nSW_2^2(\sigma,\eta)) - E(\sigma)\cdot E(\eta)}{M_2(\sigma) - \|E(\sigma)\|^2},\\
b^{SW_2} &= E(\eta) - a^{SW_2}E(\sigma),
\end{aligned}
$$

where $S^{\sigma,\eta,W_2}(x) = a^{W_2}x + b^{W_2}$ *and* $S^{\sigma,\eta,SW_2}(x) = a^{SW_2}x + b^{SW_2}$.

Proof Given $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\eta \in \mathcal{W}_2(\mathbb{R}^n)$, let $M_2(\sigma) = \int \|x\|^2 d\sigma(x)$ (similarly define $M_2(\eta)$), $E(\sigma) = \int x d\sigma(x)$ (similarly define $E(\eta)$). For $S(x) = ax + b$, by the changes of variables formula and the fact that $T_\sigma^\eta = T_{S_\sharp\sigma}^\eta \circ S$, we have

$$
\begin{aligned}
W_2^2(S_\sharp\sigma,\eta) = \|T_\sigma^\eta - (ax+b)\|_\sigma^2 &= M_2(\eta) + a^2M_2(\sigma) + 2ab\cdot E(\sigma)\\
&\quad - 2a\int T_\sigma^\eta(x)\cdot x d\sigma(x) - E(\sigma)\cdot E(\eta) + \|b\|^2 - 2E(\eta)\cdot b.
\end{aligned}
$$

Taking the partial derivatives gives

$$
\begin{aligned}
\frac{\partial}{\partial a} &= 2aM_2(\sigma) + 2b\cdot E(\sigma) - 2\int T_\sigma^\eta(x)\cdot x d\sigma(x),\\
\frac{\partial}{\partial b} &= 2b + 2aE(\sigma) - 2E(\eta).
\end{aligned}
$$

Setting the above equations to zero and with the observation that $\int T_\sigma^\eta(x)\cdot x d\sigma(x) = \frac{1}{2}(M_2(\eta) + M_2(\sigma) - W_2^2(\sigma,\eta))$, we get the the desired formulas for a^{W_2} and b^{W_2}. Similarly,

$$
\begin{aligned}
SW_2^2(S_\sharp\sigma,\eta) &= \int_{S^{n-1}} W_2^2((S_\sharp\sigma)^\theta,\eta^\theta)du(\theta)\\
&= \int_{S^{n-1}}\int_{\mathbb{R}} |T_{\sigma^\theta}^{\eta^\theta}(t) - (at + b\cdot\theta)|^2 dt du(\theta)\\
&= \frac{1}{n}\Big(M_2(\eta) + a^2M_2(\sigma) + 2ab\cdot E(\sigma)\\
&\quad - 2na\int_{S^{n-1}}\int_{\mathbb{R}} tT_{\sigma^\theta}^{\eta^\theta}(t)d\sigma^\theta(t)du(\theta) - E(\sigma)\cdot E(\eta) + \|b\|^2 - 2E(\eta)\cdot b\Big).
\end{aligned}
$$

Taking the partial derivatives gives

$$
\begin{aligned}
\frac{\partial}{\partial a} &= \frac{1}{n}\Big(2aM_2(\sigma) + 2b\cdot E(\sigma) - 2n\int_{S^{n-1}}\int_{\mathbb{R}} tT_{\sigma^\theta}^{\eta^\theta}(t)dt du(\theta)\Big),\\
\frac{\partial}{\partial b} &= \frac{1}{n}\Big(2b + 2aE(\sigma) - 2E(\eta)\Big).
\end{aligned}
$$

Setting the above equations to zero and with the observation that $\int_{S^{n-1}}\int_{\mathbb{R}} tT_{\sigma^\theta}^{\eta^\theta}(t)d\sigma^\theta(t)du(\theta) = \frac{1}{2n}(M_2(\sigma) + M_2(\eta) - nSW_2^2(\sigma,\eta))$, we get the desired formulas for a^{SW_2} and b^{SW_2}. We provide computational details in Appendix C. □

Corollary 19 *Given the same assumptions as in Proposition 18, for $D = W_2$ or SW_2*

$$E({S^{\sigma,\eta,D}}_{\sharp}\sigma) = E(\eta). \tag{29}$$

Proof Upon direct calculation, we have $E({S^{\sigma,\eta,D}}_{\sharp}\sigma = a^D E(\sigma) + b^D$, where a^D, b^D are as in Proposition 18. The conclusion can be derived from the expressions for b^{W_2} and b^{SW_2}. □

Corollary 20 *Let $\eta = \mathcal{U}(\sigma, \mu, P)$ in Proposition 18. Then the parameters corresponding to $S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}$ and $S^{\sigma,\mathcal{U}(\sigma,\mu,P),SW_2}$ satisfy*

$$\begin{aligned}
\widetilde{a}^{W_2} &= \frac{\frac{1}{2}(M_2(\mu) + M_2(\sigma) - \sum_{i=1}^n W_2^2(\sigma^{\theta_i}, \mu^{\theta_i})) - E(\sigma)\cdot E(\mu)}{M_2(\sigma) - \|E(\sigma)\|^2},\\
\widetilde{b}^{W_2} &= E(\mu) - \widetilde{a}^{W_2} E(\sigma),\\
\widetilde{a}^{SW_2} &= \frac{\frac{1}{2}(M_2(\mu) + M_2(\sigma) - nSW_2^2(\sigma, \mathcal{U}(\sigma,\mu,P))) - E(\sigma)\cdot E(\mu)}{M_2(\sigma) - \|E(\sigma)\|^2},\\
\widetilde{b}^{SW_2} &= E(\mu) - a^{SW_2} E(\sigma),
\end{aligned}$$

where $S^{\sigma,\mathcal{U}(\sigma,\mu,P),W_2}(x) = \widetilde{a}^{W_2}x + \widetilde{b}^{W_2}$ and $S^{\sigma,\mathcal{U}(\sigma,\mu,P),SW_2}(x) = \widetilde{a}^{SW_2}x + \widetilde{b}^{SW_2}$.

Proof The above formulas follows directly from Proposition 18, the fact that $\mathcal{U}(\sigma, \mu, P)$ and μ have the same mean (see (13)), and the formula (7) for $W_2^2(\sigma, \mathcal{U}(\sigma, \mu, P))$. □

Proposition 21 *Let*

$$\mathcal{S}(P) := \{x \mapsto P\Lambda P^t x + b : \Lambda \text{ is positive and diagonal}, b \in \mathbb{R}^n\}.$$

Consider the minimization problem

$$S_P^{\sigma,\eta} := \underset{S_P \in \mathcal{S}(P)}{\arg\min} \|S_P - T_\sigma^\eta\|_\sigma. \tag{30}$$

Let $S_P^{\sigma,\mu}$ and $S_P^{\sigma,\mathcal{U}(\sigma,\mu;P)}$ be the minimizers of (30) with $\eta = \mu$ and $\eta = \mathcal{U}(\sigma, \mu, P)$, respectively. We denote the diagonal entries of the corresponding Λ by a_i and $\widetilde{a}_i$, respectively. Similar notation holds for b_i and $\widetilde{b}_i$. Then

$$\begin{aligned}
\widetilde{a}_i - a_i &= \frac{\int |\theta_i \cdot (T_\sigma^\mu(x) - x)|^2 d\sigma(x) - W_2^2(\sigma^{\theta_i}, \mu^{\theta_i})}{2(M_2^{\sigma^{\theta_i}} - (E^{\theta_i})^2)} \geq 0,\\
\widetilde{b}_i - b_i &= -\sum_{i=1}^n \theta_i E^{\sigma^{\theta_i}}(\widetilde{a}_i - a_i).
\end{aligned}$$

Proof The proof uses similar arguments in Proposition 18 and Corollary 20 except the partial derivatives are with respect to a_i and $\widetilde{a}_i$ instead of a and $\widetilde{a}$. Note that following these arguments, we use the equations presented in Lemma 23. □

Appendix C Other technical details

Lemma 22 *Let $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\eta, \mu \in \mathcal{W}_2(\mathbb{R}^n)$. Then we get*

$$E((T_{\sigma,\mu;P})_\sharp\sigma) = \int T_{\sigma,\mu;P}(x)d\sigma(x) = \int yd\mu(y) = E(\mu),$$
$$\int T_\sigma^\eta(x) \cdot xd\sigma(x) = \frac{1}{2}(M_2(\eta) + M_2(\sigma) - W_2^2(\sigma,\eta))$$
$$M_2((T_{\sigma,\mu;P})_\sharp\sigma) = \int \|T_{\sigma,\mu;P}(x)\|^2 d\sigma(x) = M_2(\mu)$$

Proof By the change of variables formula, we have

$$\begin{aligned}\int T_{\sigma,\mu;P}(x)d\sigma(x) &= \sum_{i=1}^n \theta_i \int T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}(x\cdot\theta_i)d\sigma(x) = \sum_{i=1}^n \theta_i \int T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}(t)d\sigma^{\theta_i}(t)\\ &= \sum_{i=1}^n \theta_i \int zd\mu^{\theta_i}(z) = \sum_{i=1}^n \theta_i \int y\cdot\theta_i d\mu(y)\\ &= \int yd\mu(y),\end{aligned}$$

$$\begin{aligned}\int T_\sigma^\eta(x)\cdot xd\sigma(x) &= \frac{1}{2}\Big(\int \|T_\sigma^\eta(x)\|^2 d\sigma(x)\\ &\quad + \int \|x\|^2 d\sigma(x) - \int \|T_\sigma^\eta(x) - x\|^2 d\sigma(x)\Big)\\ &= \frac{1}{2}(M_2(\eta) + M_2(\sigma) - W_2^2(\sigma,\eta)),\end{aligned}$$

$$\begin{aligned}\int \|T_{\sigma,\mu;P}(x)\|^2 d\sigma(x) &= \int \sum_{i=1}^n |T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}(x\cdot\theta)|^2 d\sigma(x) = \sum_{i=1}^n \int |T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}(t)|^2 d\sigma^{\theta_i}(t)\\ &= \sum_{i=1}^n \int |w|^2 d\mu^{\theta_i}(w) = \sum_{i=1}^n \int |y\cdot\theta_i|^2 d\mu(y)\\ &= \int \|y\|^2 d\mu(y) = M_2(\mu),\end{aligned}$$

where the last steps make use of the fact that $P = [\theta_1, \cdots, \theta_n]$ is an orthogonal matrix. □

Lemma 23 *Let $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$, $\eta \in \mathcal{W}_2(\mathbb{R}^n)$, and $b \in \mathbb{R}^n$. Then*

$$\int_{S^{n-1}} \int_{\mathbb{R}} tT_{\sigma^\theta}^{\eta^\theta}(t)d\sigma^\theta(t)du(\theta) = \frac{M_2(\sigma) + M_2(\eta) - nSW_2^2(\sigma,\eta)}{2n} \tag{31}$$

$$\int_{S^{n-1}} \int_{\mathbb{R}} t^2 d\sigma^\theta(t) du(\theta) = \frac{M_2(\sigma)}{n} \tag{32}$$

$$\int_{S^{n-1}} \int_{\mathbb{R}} |T_{\sigma^\theta}^{\eta^\theta}(t)|^2 d\sigma^\theta(t) du(\theta) = \frac{M_2(\eta)}{n} \tag{33}$$

$$\int_{S^{n-1}} \int_{\mathbb{R}} (b \cdot \theta) t d\sigma^\theta(t) du(\theta) = \frac{E(\sigma) \cdot b}{n} \tag{34}$$

$$\int_{S^{n-1}} \int_{\mathbb{R}} (b \cdot \theta) T_{\sigma^\theta}^{\eta^\theta}(t) t d\sigma^\theta(t) du(\theta) = \frac{E(\eta) \cdot b}{n} \tag{35}$$

Proof We note that (32) and (33) are analogous by the change of variables formula, so are (34) and (35). We will first show (32).

$$\begin{aligned} \int_{S^{n-1}} \int_{\mathbb{R}} t^2 d\sigma^\theta(t) du(\theta) &= \int_{S^{n-1}} \int_{\mathbb{R}^n} |x \cdot \theta|^2 d\sigma(x) du(\theta) \\ &\overset{\text{Fubini}}{=} \int_{\mathbb{R}^n} \int_{S^{n-1}} |x \cdot \theta|^2 du(\theta) d\sigma(x) \\ &= \int_{\mathbb{R}^n} \frac{\|x\|^2}{2} d\sigma(x) \\ &= \frac{M_2(\sigma)}{n}. \end{aligned}$$

For (34), we have

$$\begin{aligned} &\int_{S^{n-1}} \int_{\mathbb{R}} (b \cdot \theta) t d\sigma^\theta(t) du(\theta) \\ &= \int_{S^{n-1}} b \cdot \theta \int_{\mathbb{R}^n} x \cdot \theta d\sigma(x) du(\theta) \\ &= \int_{S^{n-1}} (b \cdot \theta)(E(\sigma) \cdot \theta) du(\theta) \\ &= \int_{S^{n-1}} \frac{1}{2}\Big(|b \cdot \theta)|^2 + |E(\sigma) \cdot \theta|^2 - |(b - E(\sigma)) \cdot \theta|^2\Big) du(\theta) \\ &= \frac{1}{2n}\Big(\|b\|^2 + \|E(\sigma)\|^2 - \|b - E(\sigma)\|^2\Big) \\ &= \frac{E(\sigma) \cdot b}{n}. \end{aligned}$$

With (32) and (33), we have (31):

$$\begin{aligned} &\int_{S^{n-1}} \int_{\mathbb{R}} t T_{\sigma^\theta}^{\eta^\theta}(t) d\sigma^\theta(t) du(\theta) \\ &= \frac{1}{2} \int_{S^{n-1}} \int_{\mathbb{R}} \Big(t^2 + (T_{\sigma^\theta}^{\eta^\theta}(t))^2 - (t - T_{\sigma^\theta}^{\eta^\theta}(t))^2\Big) d\sigma^\theta(t) du(\theta) \\ &= \frac{1}{2n}\Big(M_2(\sigma) + M_2(\eta) - n \int_{S^{n-1}} W_2^2(\sigma^\theta, \eta^\theta) du(\theta)\Big) \end{aligned}$$

$$= \frac{M_2(\sigma) + M_2(\eta) - nSW_2^2(\sigma, \eta)}{2n}.$$

□

Lemma 24 *Let* $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ *and* $\mu \in \mathcal{W}_2(\mathbb{R}^n)$ *and* $P = [\theta_1, \cdots, \theta_n] \in O(n)$. *Then*

$$W_2^2(\sigma, \mu) \geq \sum_{i=1}^{n} W_2^2(\sigma^{\theta_i}, \mu^{\theta_i}).$$

Proof By [19, Proposition 5.1.3],

$$W_2^2(\sigma^\theta, \mu^\theta) \leq \int |\theta \cdot x - \theta \cdot y|^2 d\gamma^*(x, y),$$

where γ^* is the optimal transport plan between σ and μ. Then

$$\begin{aligned}\sum_{i=1}^{n} W_2^2(\sigma^{\theta_i}, \mu^{\theta_i}) &\leq \int \sum_{i=1}^{n} |\theta_i \cdot (x - y)|^2 d\gamma^*(x, y)\\ &= \int \|x - y\|^2 d\gamma^*(x, y)\\ &= W_2^2(\sigma, \mu).\end{aligned}$$

□

Lemma 25 *Let* $h : \mathbb{R}^n \to \mathbb{R}^n$ *and* $\sigma(\mathbb{R}^n) = 1$. *Then*

$$\int \|h(x)\|^2 d\sigma(x)\| \geq \| \int h(x) d\sigma(x)\|^2,$$

where equality holds if and only if $h(x) = v$ σ*-a.e. for some* $v \in \mathbb{R}^n$.

Proof Let $h(x) = [h_1(x), \cdots, h_n(x)]^t$. By Hölder's inequality,

$$\begin{aligned}\int |h_i(x)| d\sigma(x) &\leq \left(\int |h_i(x)|^2 d\sigma(x)\right)^{1/2} \left(\int 1^2 d\sigma(x)\right)^{1/2}\\ &= \left(\int |h_i(x)|^2 d\sigma(x)\right)^{1/2}.\end{aligned}$$

Squaring the above inequality and summing over i gives the desired inequality. Observe that equality holds if and only if $h_i(x) = v_i$ for some constant $v_i \in \mathbb{R}$. □

Lemma 26 *Let* $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$, *and* $M_2(\sigma)$, $E(\sigma)$ *be defined as in Proposition 18. Then*

$$M_2(\sigma) - \|E(\sigma)\|^2 > 0.$$

Proof Since x is not a constant vector σ-a.e. ($\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$), it follows from Lemma 25 with $h(x) = x$ that

$$\int \|x\|^2 d\sigma(x) > \| \int x d\sigma(x) \|^2.$$

□

Lemma 27 *Let $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$, $\mu \in \mathcal{W}_2(\mathbb{R}^n)$ and ϕ be a convex function such that $\nabla\phi = T_\sigma^\mu$ given by Brenier's theorem (see e.g., [37, Theorem 1.48]). If ϕ is differentiable at $E(\sigma)$, where $E(\sigma) = \int x d\sigma(x)$, then*

$$\int T_\sigma^\mu(x) \cdot x d\sigma(x) - \Big(\int x d\sigma(x) \Big) \cdot \Big(\int T_\sigma^\mu(x) d\sigma(x) \Big) \geq 0. \tag{36}$$

Proof Let $A = \{x \in \mathbb{R}^n : \phi \text{ is differentiable at } x\}$. Since ϕ is σ-a.e. differentiable, we have $\sigma(A) = 1$. Then it follows from the convexity of ϕ that

$$(\nabla\phi(x) - \nabla\phi(E(\sigma))) \cdot (x - E(\sigma)) \geq 0, \quad \forall x \in A. \tag{37}$$

Hence

$$\int_A (T_\sigma^\mu(x) - T_\sigma^\mu(E(\sigma))) \cdot (x - E(\sigma)) d\sigma(x) \geq 0,$$

which is exactly the desired inequality (36) by a direct computation using $\sigma(A) = 1$:

$$\begin{aligned}
&- \int T_\sigma^\mu(E(\sigma)) \cdot x d\sigma(x) - \int T_\sigma^\mu(x) \cdot E(\sigma) d\sigma(x) + T_\sigma^\mu(E(\sigma)) \cdot E(\sigma) \\
&= -T_\sigma^\mu(E(\sigma)) \cdot E(\sigma) - \Big(\int x d\sigma(x) \Big) \cdot \Big(\int T_\sigma^\mu(x) d\sigma(x) \Big) + T_\sigma^\mu(E(\sigma)) \cdot E(\sigma) \\
&= \Big(\int x d\sigma(x) \Big) \cdot \Big(\int T_\sigma^\mu(x) d\sigma(x) \Big).
\end{aligned}$$

□

Remark 12 The same conclusion holds if the assumption were "$E(\sigma)$ lies in the support of σ" instead of ϕ being differentiable at $E(\sigma)$, which can be proved using the fact that the support of optimal transport plan is cyclically monotone.

Remark 13 Given the assumptions in:Lemma 27, one can show that the inequality is strict if in addition, there exists a ball $B(x, r)$, where x lies in the support of σ, such that for any $\lambda \in (0, 1)$ and $y \in B(x, r)$

$$\phi((1 - \lambda)y + \lambda E(\sigma)) < (1 - \lambda)\phi(y) + \lambda\phi(E(\sigma)),$$

which guarantees that the inequality (37) is strict for y in a set with positive measure. In particular, if furthermore ϕ in Lemma 27 is strictly convex, the strict inequality holds.

Proposition 28 *Let $\sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$ and $\mu = T^b_\sharp \sigma$ with $T^b(x) = x + b$, $b \neq 0 \in \mathbb{R}^n$. Consider iteration $\sigma_{k+1} = (T_{\sigma_k,\mu;\theta_k})_\sharp \sigma_k$, with $\sigma_0 = \sigma$ and where θ_k is chosen i.i.d. according to the uniform measure on S^{n-1}. Then*

$$\sigma_k \xrightarrow{a.s.} \mu \quad \text{in } W_2.$$

Proof By a direct computation, $T^\mu_{\sigma_k}(x) = x + b_k$, where

$$b_{k+1} = b_k - \theta_k(\theta_k \cdot b_k).$$

To show $\sigma_k \to \mu$ almost surely, it suffices to show that $b_k \to 0$ almost surely. By symmetry of S^{n-1}, we assume without of generality that $b_0 = [1, 0, \cdots, 0]^t$. Note that $\|b_1\|^2 = 1 - |\theta_0 \cdot b_0|^2$. Consider the spherical coordinates for S^{n-1} with $\phi_1, \dots, \phi_{n-2} \in [0, \pi]$ and $\phi_{n-1} \in [0, 2\pi]$:

$$\begin{aligned} &x_1 = \cos(\varphi_1), \quad x_2 = \sin(\varphi_1)\cos(\varphi_2), \quad x_3 = \sin(\varphi_1)\sin(\varphi_2)\cos(\varphi_3) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\cdots \\ &x_{n-1} = \sin(\varphi_1)\cdots\sin(\varphi_{n-2})\cos(\varphi_{n-1}),\ x_n = \sin(\varphi_1)\cdots\sin(\varphi_{n-2})\sin(\varphi_{n-1}). \end{aligned}$$

The corresponding Jacobian is $\sin^{n-2}(\varphi_1)\sin^{n-3}(\varphi_2)\cdots\sin\varphi_{n-2}$. A direct computation gives

$$\begin{aligned} \mathbb{E}[|\theta_0 \cdot b_0|^2] &= \frac{\int_0^\pi \sin^{n-2}(\varphi_1)\cos^2(\varphi_1)d\varphi_1}{\int_0^\pi \sin^{n-2}(\varphi_1)d\varphi_1} \\ &= 1 - \frac{\int_0^\pi \sin^{n}(\varphi_1)d\varphi_1}{\int_0^\pi \sin^{n-2}(\varphi_1)d\varphi_1} \\ &= \rho < 1. \end{aligned}$$

Hence $\mathbb{E}[\|b_1\|^2] = 1 - \rho \in (0, 1)$. By symmetry and induction, one can show that

$$\mathbb{E}[\|b_k\|^2] = (1-\rho)^k \xrightarrow{k\to\infty} 0.$$

Since $\|b_{k+1}\| \leq \|b_k\|$, by the monotone convergence theorem, we have

$$\mathbb{E}[\|b_k\|^2] \longrightarrow \mathbb{E}[\alpha_\infty^2],$$

where $\alpha_\infty = \lim \alpha_k$ and $\alpha_k = \|b_k\|$, which implies $\alpha_\infty = 0$ almost surely and hence $b_k \to 0$ almost surely. □

Lemma 29 *Let $\sigma, \mu \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$. Then $(T_{\sigma,\mu;P})_\sharp \sigma \in \mathcal{W}_{2,ac}(\mathbb{R}^n)$, for any $P \in O(n)$.*

Proof Let $P = [\theta_1, \cdots, \theta_n]$. A direct computation shows

$$\nabla T_{\sigma,\mu;P}(x) = P \begin{bmatrix} (T_{\sigma^{\theta_2}}^{\mu^{\theta_i}})'(x \cdot \theta_1) & & & \\ & (T_{\sigma^{\theta_2}}^{\mu^{\theta_i}})'(x \cdot \theta_2) & & \\ & & \ddots & \\ & & & (T_{\sigma^{\theta_n}}^{\mu^{\theta_i}})'(x \cdot \theta_n) \end{bmatrix} P^t.$$

Following similar arguments as in [38, Proof of Lemma 1, p. 949], it suffices to show that there exists a set Σ such that (i) $\sigma(\mathbb{R}^n \setminus \Sigma) = 0$ (ii) $T_{\sigma,\mu;P}|_\Sigma$ is injective and $\nabla T_{\sigma,\mu;P}$ is positive definite on Σ. To this end, it suffices to observe that $T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}$ is injective and $(T_{\sigma^{\theta_i}}^{\mu^{\theta_i}})' > 0$ outside a set U_i that is σ^{θ_i}-negligible, i.e., $\sigma^{\theta_i}(U_i) = 0$. Here we have used the fact that $T_{\sigma^{\theta_i}}^{\mu^{\theta_i}}$ exists and is unique given that $\sigma \in \mathcal{P}_{ac}(\mathbb{R}^n)$ (and hence σ^{θ_i} is absolutely continuous, see e.g., Box 2.4. in [37, p. 82]). The fact that $M_2((T_{\sigma,\mu;P})_\sharp\sigma)$ is finite follows from (14) and that $M_2(\mu) < \infty$. □

Acknowledgements The authors appreciate helpful discussions with Dr. Hengrong Du regarding Example 4.1 and Proposition 28.

Funding CM is supported by NSF award DMS-2306064 and by a seed grant from the School of Data Science and Society at UNC.

Data Availability The data that support the findings of this study are available from the corresponding author, [SL], upon reasonable request.

Declarations

Conflict of interest The authors have no competing or conflicting interests to declare that are relevant to the content of this article.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Precup, D., Teh, Y.W. (eds.) Proceedings of Machine Learning Research, vol. 70, pp. 214–223 (2017)
2. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vis. **40**(2), 99–121 (2000)
3. Kolouri, S., Park, S.R., Thorpe, M., Slepcev, D., Rohde, G.K.: Optimal mass transport: Signal processing and machine-learning applications. IEEE Signal Process. Mag. **34**(4), 43–59 (2017)
4. Peyré, G., Cuturi, M.: Computational optimal transport. Foundations Trends Mach. Learn. **11**(5–6), 355–607 (2019)
5. Bousquet, O., Gelly, S., Tolstikhin, I., Simon-Gabriel, C.-J., Schoelkopf, B.: From optimal transport to generative modeling: the VEGAN cookbook (2017). arXiv preprint arXiv:1705.07642
6. Baptista, R., Hosseini, B., Kovachki, N.B., Marzouk, Y.M., Sagiv, A.: An Approximation Theory Framework for Measure-Transport Sampling Algorithms. arXiv:2302.13965 (2023)
7. Lambert, M., Chewi, S., Bach, F., Bonnabel, S., Rigollet, P.: Variational inference via Wasserstein gradient flows. Adv. Neural Inform. Process. Syst. **35**, 14434–14447 (2022)
8. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: ICLR 2023 (2023)

9. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (ICLR) (2021)
10. Chen, R.T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural ordinary differential equations. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 31 (2018)
11. Taghvaei, A., Mehta, P.G.: An optimal transport formulation of the linear feedback particle filter. In: 2016 American Control Conference (ACC), pp. 3614–3619 (2016). IEEE
12. Kobyzev, I., Prince, S.J., Brubaker, M.A.: Normalizing flows: an introduction and review of current methods. IEEE Trans. Pattern Anal. Mach. Intell. **43**(11), 3964–3979 (2020)
13. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 26, pp. 2292–2300 (2013)
14. Kolouri, S., Pope, P.E., Martin, C.E., Rohde, G.K.: Sliced wasserstein auto-encoders. In: International Conference on Learning Representations (2018)
15. Bonet, C., Courty, N., Septier, F., Drumetz, L.: Efficient gradient flows in sliced-Wasserstein space. Transactions on Machine Learning Research (2022)
16. Pitié, F., Kokaram, A.C., Dahyot, R.: Automated colour grading using colour distribution transfer. Comput. Vis. Image Understanding **107**(1–2), 123–137 (2007)
17. Bonneel, N., Rabin, J., Peyré, G., Pfister, H.: Sliced and Radon Wasserstein barycenters of measures. J. Math. Imaging Vis. **51**, 22–45 (2015)
18. Papamakarios, G.: Neural density estimation and likelihood-free inference. PhD thesis, University of Edinburgh (2019)
19. Bonnotte, N.: Unidimensional and evolution methods for optimal transportation. PhD thesis, Université Paris-Sud, Scuola Normale Superiore (December 2013)
20. Li, S., Moosmueller, C.: Measure transfer via stochastic slicing and matching (2023). arXiv:2307.05705
21. Feydy, J., Charlier, B., Vialard, F.-X., Peyré, G.: Optimal transport for diffeomorphic registration. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) Medical Image Computing and Computer Assisted Intervention-MICCAI 2017, pp. 291–299. Springer, Cham (2017)
22. De Lara, L., González-Sanz, A., Loubes, J.-M.: Diffeomorphic registration using sinkhorn divergences. SIAM J. Imaging Sci. **16**(1), 250–279 (2023)
23. Shen, Z., Feydy, J., Liu, P., Curiale, A.H., San Jose Estepar, R., Niethammer, M.: Accurate point cloud registration with robust optimal transport. Advances in Neural Information Processing Systems **34**, 5373–5389 (2021)
24. Rabin, J., Peyré, G., Delon, J., Bernot, M.: Wasserstein barycenter and its application to texture mixing. In: Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3, pp. 435–446 (2012). Springer
25. Rabin, J., Peyré, G., Cohen, L.D.: Geodesic shape retrieval via optimal mass transport. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) Computer Vision - ECCV 2010, pp. 771–784. Springer, Berlin, Heidelberg (2010)
26. Ambrosio, L., Gigli, N.: A User's Guide to Optimal Transport, pp. 1–155. Springer, Berlin, Heidelberg (2013)
27. Khurana, V., Kannan, H., Cloninger, A., Moosmüller, C.: Supervised learning of sheared distributions using linearized optimal transport. Sampling Theory, Signal Processing, and Data Analysis **21**(1) (2023)
28. Villani, C.: Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften, vol. 338. Springer, Berlin (2009)
29. Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Commun. Pure Appl. Math. **44**(4), 375–417 (1991)
30. Robbins, H., Monro, S.: A stochastic approximation method. Annal. Math. Stat. 400–407 (1951)
31. Aldroubi, A., Li, S., Rohde, G.K.: Partitioning signal classes using transport transforms for data analysis and machine learning. Sampl. Theory Signal Process. Data Anal. **19**(6) (2021)
32. Park, S.R., Kolouri, S., Kundu, S., Rohde, G.K.: The cumulative distribution transform and linear pattern classification. Appl. Comput. Harmonic Anal. **45**(3), 616–641 (2018)

33. Moosmüller, C., Cloninger, A.: Linear optimal transport embedding: Provable Wasserstein classification for certain rigid transformations and perturbations. Inform. Inference **12**(1), 363–389 (2023)
34. Brown, L.G.: A survey of image registration techniques. ACM Comput. Surv. **24**(4), 325–376 (1992)
35. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: a survey. IEEE Trans. Med. Imaging **32**(7), 1153–1190 (2013)
36. Meckes, E.S.: The Random Matrix Theory of the Classical Compact Groups. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (2019)
37. Santambrogio, F.: Optimal Transport for Applied Mathematicians. Birkäuser Cham, NY (2015)
38. Zemel, Y., Panaretos, V.M.: Fréchet means and procrustes analysis in wasserstein space. Bernoulli **25**(2), 932–976 (2019)