Synchronizing Probability Measures on Rotations via Optimal Transport

Tolga Birdal [1]   Michael Arbel [2]   Umut Şimşekli [3,4]   Leonidas Guibas [1]

[1] Department of Computer Science, Stanford University, USA
[2] Gatsby Computational Neuroscience Unit, University College London, UK
[3] LTCI, Télécom Paris, Institut Polytechnique de Paris, Paris, France
[4] Department of Statistics, University of Oxford, Oxford, UK

Abstract

We introduce a new paradigm, 'measure synchronization', for synchronizing graphs with measure-valued edges. We formulate this problem as maximization of the cycle-consistency in the space of probability measures over relative rotations. In particular, we aim at estimating marginal distributions of absolute orientations by synchronizing the 'conditional' ones, which are defined on the Riemannian manifold of quaternions. Such graph optimization on distributions-on-manifolds enables a natural treatment of multimodal hypotheses, ambiguities and uncertainties arising in many computer vision applications such as SLAM, SfM, and object pose estimation. We first formally define the problem as a generalization of the classical rotation graph synchronization, where in our case the vertices denote probability measures over rotations. We then measure the quality of the synchronization by using Sinkhorn divergences, which reduces to other popular metrics such as Wasserstein distance or the maximum mean discrepancy as limit cases. We propose a nonparametric Riemannian particle optimization approach to solve the problem. Even though the problem is non-convex, by drawing a connection to the recently proposed sparse optimization methods, we show that the proposed algorithm converges to the global optimum in a special case of the problem under certain conditions. Our qualitative and quantitative experiments show the validity of our approach and we bring in new perspectives to the study of synchronization.

1. Introduction

Synchronization [70, 30, 71], is the art of consistently recovering absolute quantities from a collection of ratios or pairwise relationships. An equivalent definition describes the problem as simultaneously upgrading from pairwise local information onto the global one: an *averaging* [34, 36, 8]. At this point, synchronization is a fundamental piece of most multi-view reconstruction / multi-shape analysis pipelines [68, 17, 18] as it heavy lifts the global constraint satisfaction while respecting the geometry
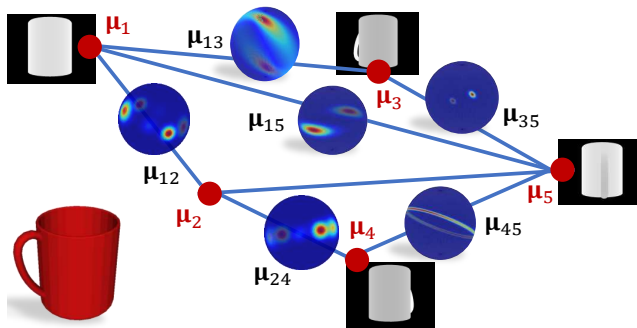

Figure 1. We coin **measure synchronization** to refer to the problem of recovering distributions over absolute rotations $\{\boldsymbol{\mu}_i\}_i$ (shown in red) from a given set of potential relative rotations $\{\boldsymbol{\mu}_{ij}\}_{(i,j)}$ per pair that are estimated in an independent fashion.

of the parameters. In a wide variety of applications, these parameters are composed of the elements of a rotation expressed either as a point on a Riemannian manifold (forms a Lie group) or its tangent space (Lie algebra) [33, 14].

Unfortunately, for our highly complex environments, symmetries are inevitable and they may render it impossible to hypothesize an ideal single candidate to explain the configuration of capture, be it pairwise or absolute. In those cases, there is no single best solution to the 3D perception problem. One might then speak of a set of plausible solutions that are only collectively useful in characterizing global positioning of the cameras / objects [42, 54, 14]. Put more concretely, imagine a 3D sensing device capturing a rotating untextured, cylindrical coffee mug with a handle (*c.f.* Fig. 1). The mug's image is, by construction, identical under the rotations around its symmetry axis whenever the handle is occluded. Any relative or absolute measurement is then subject to an *ambiguity* which prohibits the estimation of a unique best alignment. Whenever the handle appears, mug's pose can be determined uniquely, leading to a single mode. Such phenomena are ubiquitous in real scenarios and have recently lead to a paradigm shift in computer vision towards reporting the estimates in terms of either empirical or continuous multimodal probability distributions on the parameter spaces [41, 54, 14, 13], rather than single entities as done in the conventional methods [61, 55, 37].

In the aforementioned situations, to the best of our knowledge, all of the existing synchronization algorithms [33, 65, 25, 34, 79, 12] are doomed to failure and a new notion of synchronization is required that rather takes into account the distributions that are defined on the parameter space. In other words, what are to be synchronized are now the probability measures and not single data points[1].

In this paper, considering the Sinkhorn divergence, which has proven successful in many application domains [23, 26, 28], we introduce the concept of *measure synchronization*. After formally defining the graph synchronization problem over probability measures by using the Sinkhorn divergence, we propose a novel algorithm to solve this problem, namely the Riemannian particle gradient descent (RPGD), which is based on the recently developed nonparametric 'particle-based' methods [51, 5, 52, 44] and Riemannian optimization methods [1]. In particular, we parameterize rotations by quaternions and use the Sinkhorn divergence [26] between the estimated joint rotation distribution and the input empirical relative rotation distribution as a measure of consistency.

We further investigate the special case of the proposed problem, where the loss function becomes the *maximum mean discrepancy* (MMD) combined with a simple regularizer. By using this special structure and imposing some structure on the weights of the measures to be estimated, we draw a link between our approach and the recently proposed sparse optimization of [20]. This link allows us to establish a theoretical guarantee showing that RPGD converges to the global optimum under certain conditions.

Our synthetic and real experiments demonstrate that both of our novel, non-parametric algorithms are able to faithfully discover the probability distribution for each of the absolute cameras with varying modes, given the multimodal rotation distribution for each edge in the graph. This is the first time the two realms of synchronization and optimal transport are united and we hope to foster further research leading to enhanced formulations and algorithms. To this end, we will make our code as well as the datasets publicly available. In a nutshell, our contributions are:

1. We introduce the problem of measure synchronization that aims for synchronization of probability distributions defined on Riemannian manifolds.
2. We address the particular case of multiple rotation averaging and solve this new synchronization problem via minimizing a collection of regularized optimal transport costs over the camera pose graph.
3. We propose the first two solutions resulting in two algorithms: *Sinkhorn-Sync* and *MMD-Sync*. These algorithms differ both by the topology they induce and by their descent regimes.

4. We show several use-cases of our approach under synthetic and real settings, demonstrating the applicability to a general set of challenges.

We release our source code under: `https://synchinvision.github.io/probsync/`.

2. Prior Art

Synchronization. The emergence of the term *synchronization* dates back to Einstein & Poincaré and the principles of relativity [27]. It was first used to recover clocks [70]. Since then the ideas have been extended to many domains other than space and time, leaping to communication systems [31, 30], finally arriving at the graphs in computer vision with Singer's angular synchronization [71]. Various aspects of the problem have been vastly studied: different group structures [33, 13, 11, 8, 38], closed form solutions [12, 11, 8], robustness [19], certifiability [66], global optimality [15], learning-to-synchronize [39] and uncertainty quantification (UQ) [84, 14]. The latter recent work of Birdal *et al.* [14] and the K-best synchronization [79] are probably what relates to us the most. In the former a probabilistic framework to estimate an empirical absolute distribution and thereby the uncertainty per camera is proposed. The latter uses a deterministic approach to yield K hypotheses with the hope that ambiguities are captured. Yet, none of the works in the synchronization field including [14, 13, 79] are able to handle multiple relative hypotheses and thus are prone to failure when ambiguities are present in the input. This is what we overcome in this work by proposing to align the observed and computed relative probability measures: *measure synchronization*.

Optimal transport and MMD. Comparing probability distributions has even a longer history and is a very well studied subject [53, 47]. Taking into account the geometry of the underlying space, optimal transport (OT) [86], proposed by Monge in 1781 [57] and advanced by Kantorovich [40] seeks to find the minimum amount of effort needed to morph one probability measure into another one. This geometric view of comparing empirical probability distributions metrizes the convergence in law better capturing the topology of measures [26]. Thanks to the efficient algorithms developed recently [23, 64, 75], OT has begun to find use-cases in various sub-fields related to computer vision such as deep generative modeling [7], non-parametric modeling [52], domain adaptation [22], 3D shape understanding [76], graph matching [87], topological analysis [48], clustering [56], style transfer [43] or image retrieval [67] to name a few.

Only recently Feydy *et al.* [26] formulated a new geometric Sinkhorn [74, 72] divergence that can relate OT distances, in particular the Sinkhorn distance [23] to *maximum mean discrepancy* (MMD) [35], a cheaper to compute frequentist norm with a smaller sample complexity. MMD

---

[1]We note that our framework extends the classical one in the sense that single data points can be seen as degenerate probability distributions represented by Dirac-delta functions.

has been successfully used in domain adaptation [85], in GANs [50, 6], as a critic [80] as well as to build auto-encoders [82]. Furthermore, Arbel *et al.* [5] have established the Wasserstein gradient flow of the MMD and demonstrated global convergence properties based on a novel noise injection scheme. Jointly, these advancements allow us to tackle the proposed measure synchronization problem using MMDs with global optimality guarantees.

To the best of our knowledge, no prior work addresses the measure synchronization problem we consider here. The closest work is by Solomon *et al.* [77] where a Wasserstein metric is minimized so as to perform a graph transduction: the labels are propagated from the *absolute* entities along the edges (relative), not the other way around.

3. Background

We now define and review the necessary concepts required for the grasp of our method.

3.1. Rotations and Synchronization

Definition 1 (Rotation Graph). *We consider a finite simple directed graph $G = (\mathcal{V}, \mathcal{E})$, where vertices correspond to reference frames and edges to the available relative measurements. Both vertices and edges are labeled with absolute and relative orientations, respectively. Each absolute frame is described by a rotation matrix $\{\mathbf{R}_i \in SO(3)\}_{i=1}^n$. Each relative pose is expressed as the transformation between frames i and j, $\mathbf{R}_{ij}$, where $(i,j) \in \mathcal{E} \subset [n] \times [n]$. G is undirected such that if $(i,j) \in \mathcal{E}$, then $(j,i) \in \mathcal{E}$ and $\mathbf{R}_{ij} = \mathbf{R}_{ji}^{-1}$.*

Definition 2 (Rotation Synchronization). *Synchronizing pose graphs, also known as multiple rotation averaging, involves the task of relating different coordinate frames by satisfying the compatibility constraint [32, 36]:*

$$\mathbf{R}_{ij} \approx \mathbf{R}_j \mathbf{R}_i^{-1}, \; \forall i \neq j \tag{1}$$

Definition 3 (Quaternion). *In the rest of the paper, we will represent a 3D rotation by a* quaternion $\mathbf{q}$, *an element of Hamilton algebra $\mathbb{H}$, extending the complex numbers with three imaginary units $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ in the form $\mathbf{q} = q_1\boldsymbol{1} + q_2\boldsymbol{i} + q_3\boldsymbol{j} + q_4\boldsymbol{k} = (q_1, q_2, q_3, q_4)^T$, with $(q_1, q_2, q_3, q_4)^T \in \mathbb{R}^4$ and $\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = \boldsymbol{ijk} = -\boldsymbol{1}$. We also write $\mathbf{q} := [a, \boldsymbol{v}]$ with the scalar part $a = q_1 \in \mathbb{R}$ and the vector part $\boldsymbol{v} = (q_2, q_3, q_4)^T \in \mathbb{R}^3$. The conjugate $\bar{\mathbf{q}}$ of the quaternion $\mathbf{q}$ is given by $\bar{\mathbf{q}} := q_1 - q_2\boldsymbol{i} - q_3\boldsymbol{j} - q_4\boldsymbol{k}$. A versor or unit quaternion $\mathbf{q} \in \mathbb{H}_1$ with $1 \stackrel{!}{=} \|\mathbf{q}\| := \mathbf{q} \cdot \bar{\mathbf{q}}$ and $\mathbf{q}^{-1} = \bar{\mathbf{q}}$, gives a compact and numerically stable parametrization to represent orientation of objects in $\mathbb{S}^3$, avoiding gimbal lock and singularities [16]. Identifying antipodal points $\mathbf{q}$ and $-\mathbf{q}$ with the same element, the unit quaternions form a double covering group of $SO(3)$. The non-commutative multiplication of two quaternions $\mathbf{p} := [p_1, \mathbf{v}_p]$ and $\mathbf{r} :=$* $[r_1, \mathbf{v}_r]$ *is defined to be $\boldsymbol{p} \otimes \boldsymbol{r} = [p_1 r_1 - \mathbf{v}_p \cdot \mathbf{v}_r, p_1 \mathbf{v}_r + r_1 \mathbf{v}_p + \mathbf{v}_p \times \mathbf{v}_r]$. For simplicity we use $\boldsymbol{p} \otimes \boldsymbol{r} := \boldsymbol{p} \cdot \boldsymbol{r} := \boldsymbol{pr}$.*

Definition 4 (Relative Quaternion). *The relative rotation between two frames can also be conveniently written in terms of quaternions: $\mathbf{q}_{ij} = \mathbf{q}_i \otimes \bar{\mathbf{q}}_j$. Note that for quaternions we will later omit the symbol $\otimes$.*

Riemannian geometry of quaternions We will now briefly explain the Lie group structure of the quaternions $\mathbb{H}$ that are essential for deriving the update equations on the manifold. Our convention follows [4].

Definition 5 (Exponential Map). *The exponential map $\text{Exp}_{\mathbf{q}}(\cdot)$ maps any vector in $\mathcal{T}_{\mathbf{q}}\mathbb{H}$, the tangent space of $\mathbb{H}$, onto $\mathbb{H}$:*

$$\text{Exp}_{\mathbf{q}}(\boldsymbol{\eta}) = \mathbf{q}\exp(\boldsymbol{\eta}) = \mathbf{q}\left(e^w\left(\cos(\theta), \mathbf{v}\frac{\sin(\theta)}{\theta}\right)\right), \tag{2}$$

where $\boldsymbol{\eta} = (w, \mathbf{v}) \in \mathcal{T}_{\mathbf{q}}\mathbb{H}$ and $\theta = \|\mathbf{v}\|$.

Definition 6 (Logarithmic Map). *The inverse of* exp-map, *$\text{Log}_{\mathbf{q}}(\mathbf{p}) : \mathbb{H} \mapsto \mathcal{T}_{\mathbf{q}}\mathbb{H}$ is called the* log-map *and defined as:*

$$\text{Log}_{\mathbf{q}}(\mathbf{p}) = \log(\mathbf{q}^{-1}\mathbf{p}) = \left(0, \frac{\mathbf{v}}{\|\mathbf{v}\|}\arccos(w)\right), \tag{3}$$

this time with a slight abuse of notation $\mathbf{q}^{-1}\mathbf{p} = (w, \mathbf{v})$.

Definition 7 (Riemannian Distance). *Let us rephrase the Riemannian distance between two unit quaternions using the logarithmic map, whose norm is the length of the shortest geodesic path. Respecting the antipodality:*

$$d(\mathbf{q}_1, \mathbf{q}_2) = \begin{cases} \|\text{Log}_{\mathbf{q}_1}(\mathbf{q}_2)\| &= \arccos(w), & w \geq 0 \\ \|\text{Log}_{\mathbf{q}_1}(-\mathbf{q}_2)\| &= \arccos(-w), & w < 0 \end{cases}$$

where $\mathbf{q}_1^{-1}\mathbf{q}_2 = (w, \mathbf{v})$.

3.2. Optimal Transport

Definition 8 (Discrete Probability Measure). *We define a discrete probability measure $\boldsymbol{\mu}$ on the simplex $\Sigma_d \triangleq \{\mathbf{x} \in \mathbb{R}_+^n : \mathbf{x}^\top \mathbf{1}_n = 1\}$ with weights $\mathbf{a} = \{a_i\}$ and locations $\{x_i\}, i = 1 \ldots n$ as:*

$$\boldsymbol{\mu} = \sum\nolimits_{i=1}^n a_i \delta_{x_i}, \quad a_i \geq 0 \; \wedge \; \sum\nolimits_{i=1}^n a_i = 1 \tag{4}$$

where δ_x is a Dirac delta function at x.

We further define the product of two measures $\boldsymbol{\mu} = \sum_{i=1}^{n_\mu} a_i^{\boldsymbol{\mu}} \delta_{x_i^{\boldsymbol{\mu}}}$ and $\boldsymbol{\nu} = \sum_{i=1}^{n_\nu} a_i^{\boldsymbol{\nu}} \delta_{x_i^{\boldsymbol{\nu}}}$, *as follows:*

$$\boldsymbol{\mu} \otimes \boldsymbol{\nu} \triangleq \sum\nolimits_{i=1}^{n_\mu} \sum\nolimits_{j=1}^{n_\nu} a_i^{\boldsymbol{\mu}} a_j^{\boldsymbol{\nu}} \delta_{x_i^{\boldsymbol{\mu}}} \delta_{x_j^{\boldsymbol{\nu}}}. \tag{5}$$

Definition 9 (Transportation Polytope). *For two probability measures $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ in the simplex, let $U(\boldsymbol{\mu}, \boldsymbol{\nu})$ denote the polyhedral set of $n_{\boldsymbol{\mu}} \times n_{\boldsymbol{\nu}}$ matrices:*

$$U(\boldsymbol{\mu}, \boldsymbol{\nu}) = \{\mathbf{P} \in \mathbb{R}_+^{n_{\boldsymbol{\mu}} \times n_{\boldsymbol{\nu}}} : \mathbf{P}\mathbf{1}_{n_{\boldsymbol{\nu}}} = \boldsymbol{\mu} \wedge \mathbf{P}^\top \mathbf{1}_{n_{\boldsymbol{\mu}}} = \boldsymbol{\nu}\}$$

$\mathbf{1}_d$ *is a d-dimensional ones-vector.*

Definition 10 (Kantorovich's Optimal Transport Problem). *Let $\mathbf{C}$ be a $n_{\boldsymbol{\mu}} \times n_{\boldsymbol{\nu}}$ cost matrix that is constructed from the ground cost function $c(x_i^{\boldsymbol{\mu}}, x_j^{\boldsymbol{\nu}})$. Kantorovich's problem seeks to find the transport plan optimally mapping $\boldsymbol{\mu}$ to $\boldsymbol{\nu}$:*

$$\mathcal{W}_c(\boldsymbol{\mu}, \boldsymbol{\nu}) = d_c(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\mathbf{P} \in U(\boldsymbol{\mu}, \boldsymbol{\nu})} \langle \mathbf{P}, \mathbf{C} \rangle \quad (6)$$

with $\langle \cdot \rangle$ denoting the Frobenius dot-product. Note that this expression is also known as the Wasserstein distance (WD).

Definition 11 (Maximum Mean Discrepancy). *Given a characteristic positive semi-definite kernel k, the maximum mean discrepancy (MMD)([35]) defines a distance between probability distributions. In particular, for discrete measures $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ and when k is induced from a ground cost function c: $k(x, y) = \exp(-c(x, y))$, the squared MMD is given by:*

$$\mathcal{MMD}_k^2(\boldsymbol{\mu}, \boldsymbol{\nu}) = \bar{K}_{\boldsymbol{\mu}, \boldsymbol{\mu}} + \bar{K}_{\boldsymbol{\nu}, \boldsymbol{\nu}} - 2\bar{K}_{\boldsymbol{\mu}, \boldsymbol{\nu}} \quad (7)$$

where $\bar{K}_{\boldsymbol{\mu}, \boldsymbol{\nu}}$ is an expectation of k under $\boldsymbol{\mu} \otimes \boldsymbol{\nu}$:

$$\bar{K}_{\boldsymbol{\mu}, \boldsymbol{\nu}} = \sum_{i=1}^{n_{\boldsymbol{\mu}}} \sum_{j=1}^{n_{\boldsymbol{\mu}}} a_i^{\boldsymbol{\mu}} a_j^{\boldsymbol{\nu}} \exp\left(-c(x_i^{\boldsymbol{\mu}}, x_j^{\boldsymbol{\nu}})\right) \quad (8)$$

Definition 12 (Sinkhorn Distances). *Cuturi [23] introduced an alternative convex set consisting of joint probabilities with small enough mutual information:*

$$U_\alpha(\boldsymbol{\mu}, \boldsymbol{\nu}) = \{\mathbf{P} \in U(\boldsymbol{\mu}, \boldsymbol{\nu}) : \mathrm{KL}(\mathbf{P} \parallel \boldsymbol{\mu}\boldsymbol{\nu}^\top) \leq \alpha\} \quad (9)$$

where $\mathrm{KL}(\cdot)$ refers to the Kullback Leibler divergence [47] between $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. This restricted polytope leads to a tractable distance between discrete probability measures under the cost matrix $\mathbf{C}$ constructed from the function c:

$$d_{c,\alpha}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\mathbf{P} \in U_\alpha(\boldsymbol{\mu}, \boldsymbol{\nu})} \langle \mathbf{P}, \mathbf{C} \rangle. \quad (10)$$

$d_{c,\alpha}$ can be computed by a modification of simple matrix scaling algorithms, such as Sinkhorn's algorithm [73].

The discrepancy $d_{c,\alpha}$ suffers from an entropic bias, that is corrected by [26, 29] giving rise to a new family of divergences as we define below:

Definition 13 (Sinkhorn Divergences). *$\mathcal{S}_{c,\alpha}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is defined to be the Sinkhorn divergence:*

$$\mathcal{S}_{c,\alpha}(\boldsymbol{\mu}, \boldsymbol{\nu}) \triangleq 2d_{c,\alpha}(\boldsymbol{\mu}, \boldsymbol{\nu}) - d_{c,\alpha}(\boldsymbol{\mu}, \boldsymbol{\mu}) - d_{c,\alpha}(\boldsymbol{\nu}, \boldsymbol{\nu}).$$

Unlike Dfn. 12, this satisfies $\mathcal{S}_{c,\alpha}(\boldsymbol{\mu}, \boldsymbol{\mu}) = \mathcal{S}_{c,\alpha}(\boldsymbol{\nu}, \boldsymbol{\nu}) = 0$ and interpolates between the WD and MMD with kernel $k = -c$:

$$\mathcal{W}_c(\boldsymbol{\mu}, \boldsymbol{\nu}) \xleftarrow{\alpha \to 0} \mathcal{S}_{c,\alpha}(\boldsymbol{\mu}, \boldsymbol{\nu}) \xrightarrow{\alpha \to \infty} \frac{1}{2}\mathcal{MMD}_{-c}^2(\boldsymbol{\mu}, \boldsymbol{\nu}).$$

4. Measure Synchronization

In this section, we will introduce the measure synchronization problem. In our problem setting, for each camera pair $(i, j) \in \mathcal{E}$, we will assume that we *observe* a discrete probability measure $\boldsymbol{\mu}_{ij}$, denoting the distribution of the relative poses between the cameras i and j. More formally, such *relative pose measures* will consist of a collection of weighted Dirac masses (cf. (4)), given as follows:

$$\boldsymbol{\mu}_{ij} \triangleq \sum_{k=1}^{K_{ij}} w_{ij}^{(k)} \delta_{\mathbf{q}_{ij}^{(k)}}, \quad (11)$$

where $\{\mathbf{q}_{ij}^{(k)}\}_{k=1}^{K_{ij}}$ denote all the possible *relative pose masses* and K_{ij} denotes the total number of potential relative poses for a given camera pair (i, j), whereas $w_{ij}^{(k)} \in \mathbb{R}_+$ denotes the weight of each point mass $\mathbf{q}_{ij}^{(k)}$. We will assume that these measures can be estimated (in a somewhat noisy manner) by using an external algorithm, such as [55, 61, 24].

Given all the 'noisy' relative pose measures $\{\boldsymbol{\mu}_{ij}\}_{(i,j) \in \mathcal{E}}$, similar to the conventional pose synchronization problem, our goal becomes estimating the *absolute pose measures* $\boldsymbol{\mu}_i$, which are a-priori unknown and denote the probability distribution of the absolute pose corresponding to camera i. Similar to the relative pose measures, we define the absolute pose measures as follows:

$$\boldsymbol{\mu}_i \triangleq \sum_{k=1}^{K_i} w_i^{(k)} \delta_{\mathbf{q}_i^{(k)}}, \quad (12)$$

where $\{\mathbf{q}_i^{(k)}, w_i^{(k)}\}_{k=1}^{K_i}$ are the point mass-weight pairs in $\boldsymbol{\mu}_i$ and K_i denotes the number of points in $\boldsymbol{\mu}_i$. Semantically, $\mathbf{q}_i^{(k)}$ and $w_i^{(k)}$ will respectively denote all the 'plausible' poses and their weights for camera i. Our ultimate goal is to estimate the absolute pose measures such that they will produce 'compatible' relative measures in the sense of Eq (1). In other words, we aim at estimating the pairs $\{\{\mathbf{q}_i^{(k)}, w_i^{(k)}\}_{k=1}^{K_i}\}_{i=1}^n$, that achieve a good graph consistency at a probability distribution level, a notion which will be precised in the sequel.

Let us introduce some notation that will be useful for defining the ultimate problem. We denote by $\boldsymbol{\mu}$ a coupling between $(\mathbf{q}_1, \ldots, \mathbf{q}_n)$ which is of the form:

$$\boldsymbol{\mu} = \sum_{k_1=1}^{K_1} \ldots \sum_{k_n=1}^{K_n} v^{(k_1, \ldots, k_n)} \prod_{i=1}^n \delta_{q_i^{k_i}}. \quad (13)$$

where $v^{(k_1, \ldots, k_n)}$ are non-negative weights that sum to 1. $\boldsymbol{\mu}$ represents the joint distribution over $(\mathbf{q}_1, ..., \mathbf{q}_n)$ and recovers each marginal distributions $\boldsymbol{\mu}_i$ after marginalizing over all the other ones. Although one could, in principle, learn the general form of Eq (13), we will be interested in two extreme cases for $\boldsymbol{\mu}$, a *high entropy* case (HE) and a *low*

entropy case (LE) for which Eq (13) simplifies. In the (HE) case, we assume that $\boldsymbol{\mu}$ is fully factorized, i.e:

$$v^{(k_1,...,k_n)} = \prod_{i=1}^{n} w_i^{(k_i)}. \tag{14}$$

This case occurs when all information about the pairings between the absolute poses measures $\boldsymbol{\mu}_i$ is lost. In the (LE) case, $v^{(k_1,...,k_n)} = 0$ whenever $k_1, ...k_n$ are not all equal to each other. This implies in particular that all marginals $\boldsymbol{\mu}_i$ have the same number of point masses $K_1 = ... = K_n = K$ and share the same weights:

$$w_i^{(k)} = v^{(k,...,k)} \triangleq v^{(k)}. \tag{15}$$

This setting is 'simpler' in the sense that it carries more information about the absolute poses. We will illustrate that both of these settings will be useful in different applications.

We propose to compute a joint distribution $\boldsymbol{\mu}$ in either case that is consistent with the relative pose measures. As a first step towards this, we further introduce the 'composition functions' g_{ij}, which are simply given by:

$$g_{ij}(\mathbf{q}_1, ..., \mathbf{q}_n) \triangleq \mathbf{q}_i \overline{\mathbf{q}}_j. \tag{16}$$

We use g_{ij} to construct 'ideal relative pose measures' from the joint absolute measure $\boldsymbol{\mu}$ by a push-forward operation, again denoted by $g_{ij}(\boldsymbol{\mu})$:

$$g_{ij}(\boldsymbol{\mu}) \triangleq \sum_{k_1=1}^{K_1} ... \sum_{k_n=1}^{K_n} v^{(k_1,...,k_n)} \delta_{\mathbf{q}_i^{(k_i)} \overline{\mathbf{q}_j^{(k_j)}}}. \tag{17}$$

In both cases (HE) and (LE), g_{ij} admits a simpler expression which is provided in the supplementary document.

We are now ready to formally the define the measure synchronization problem.

Definition 14. *Let $\mathcal{L}(\boldsymbol{\mu})$ be a* loss *functional of the form:*

$$\mathcal{L}(\boldsymbol{\mu}) \triangleq \sum_{(i,j) \in \mathcal{E}} \mathcal{D}(g_{ij}(\boldsymbol{\mu}), \boldsymbol{\mu}_{ij}) \tag{18}$$

where $\mathcal{D}$ denotes a divergence that measures the discrepancy between two probability distributions and let $\mathcal{R}$ be a regularization *functional imposing structure on $\boldsymbol{\mu}$. The measure synchronization problem is formally defined as:*

$$\min_{\boldsymbol{\mu}} \quad \mathcal{L}(\boldsymbol{\mu}) + \mathcal{R}(\boldsymbol{\mu})$$
$$\text{s.t.} \quad [w_i^1, ..., w_i^{K_i}]^\top \in \mathcal{C}_i, \quad \forall i \in \{1, ..., n\} \tag{19}$$

where $\mathcal{C}_i \subseteq \mathbb{R}_+^{K_i}$ denotes a constraint set where the weights of the absolute measure $\boldsymbol{\mu}_i$ are restricted to live in.

Note that in both (HE) and (LE), $\boldsymbol{\mu}$ can be completely recovered from the marginals $\boldsymbol{\mu}_i$ and that an additional constraint on the weights is needed in (LE): $w_i^k = w_j^k$ for all $1 \leq i, j \leq n$.

Our problem formulation resembles the recent optimal transport-based implicit generative modeling problems [7, 46, 81]. The main differences in our setting are as follows: (i) the choice of the composition functions g_{ij}, which is specific to the pose synchronization problem, (ii) the problem is purely nonparametric in the sense that the absolute measures $\{\boldsymbol{\mu}_i\}_i$ are not defined through any parametric mapping. We will be interested in two choices for the loss $\mathcal{D}$:

Case 1 - *Sinkhorn-Sync*.   Due to its recent success in various applications in machine learning [7] and its favorable theoretical properties [28], in our first configuration, we choose the loss functional $\mathcal{D}$ to be the squared Sinkhorn divergence $\mathcal{S}_{c,\alpha}^2$ with finite α, as defined in Dfn. 12. In this setting, we impose the absolute measures $\boldsymbol{\mu}_i$ to be probability measures, i.e. we set each $\mathcal{C}_i$ to be the probability simplex Σ_{K_i}, hence imposing the sum $\sum_{k=1}^{K_1} w_i^{(k)}$ to be equal to 1. Thanks to these constraints, in this setup we do not need to use regularization, i.e. we set $\mathcal{R}(\mu) = 0$.

Case 2 - *MMD-Sync*.   In our second configuration, we consider a setting which will enable us to analyze the proposed Riemannian particle descent algorithm, which will be defined in the next section. In particular, we consider the loss functional is given by the MMD distance. In this setup, we constrain the absolute measures $\boldsymbol{\mu}_i$ to be positive measures by choosing the constraint sets to be $\mathcal{C}_i = \mathbb{R}_+^{K_i}$. Accordingly, we use regularization on the weights of $\boldsymbol{\mu}_i$, that is given as follows:

$$\mathcal{R}(\boldsymbol{\mu}) = \lambda \sum_{i=1}^{n} \sum_{k_i=1}^{K_i} w_i^{(k_i)} \tag{20}$$

where $\lambda > 0$ is a hyper-parameter.

By relaxing the constraint on the weights, this problem can be written as an unconstrained optimization problem, which favors 'sparse' solutions, in the sense that we enforce the weights to have small values, which in turn forces the solution to have only a few point masses with large weights. We will show that in the (LE) setting and by using the structure of MMD, this setup is closely related to the recent sparse optimization problem in the space of measures [20] and hence inherits its theoretical properties.

Choice of the ground cost.   As for the ground cost function c required for the Sinkhorn divergence and MMD, we investigate two options. The first option is simply choosing the squared Euclidean distance between the point masses, $c(\hat{\mathbf{q}}_{ij}^{(k)}, \mathbf{q}_{ij}^{(l)}) = \|\hat{\mathbf{q}}_{ij}^{(k)} - \mathbf{q}_{ij}^{(l)}\|_2^2$, where $\hat{\mathbf{q}}_{ij}^{(k)}$ and $\mathbf{q}_{ij}^{(k)}$ are given in (17) and (11), respectively. This yields a positive definite MMD kernel. On the other hand, for the quaternions which are naturally endowed with a Riemannian metric, the second option is the geodesic distances as defined in Dfn. 7, i.e. $c(\hat{\mathbf{q}}_{ij}^{(k)}, \mathbf{q}_{ij}^{(l)}) = d(\hat{\mathbf{q}}_{ij}^{(k)}, \mathbf{q}_{ij}^{(l)})^p$, $1 \leq p \leq 2$. This second choice leads to a positive definite MMD kernel k (*c.f.* Dfn. 11) only when $p = 1$. Positivity of the
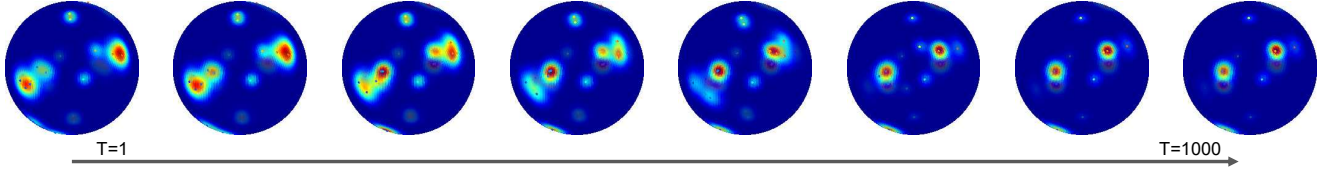
Figure 2. Density estimates for a single camera visualized over time. We sum up the probability density functions of Bingham distributions centered on each quaternion and marginalize over the angles to reduce the dimension to 3. This way we are able to plot the density estimate for all particles of a given camera and track it across iterations. Note that in here we use MMD distance with an unconstrained RPGD.

kernel is a theoretical requirement for both the Sinkhorn divergence and the MMD to be a distance metric, however, it turns out that the computation of the gradients when $p = 1$ becomes numerically unstable. Hence, as a remedy, we follow the robust optimization literature [2] and consider $p > 1$. This approach finds a good balance in terms of numerical stability: for p slightly larger than 1, we lose from the positive definiteness of the kernel k; however, the numerical computation of the gradients becomes numerically stable. We have observed in practice that small values of p i.e. $1 + \epsilon < p \leq 1.4$ would not harm our algorithm.

Remarks. We note that in the (HE) case, the cost functional requires a quadratic pairing of the elements in the absolute measures to estimate the relative distribution. That is, any quaternions $\mathbf{q}_i \sim \boldsymbol{\mu}_i$ and $\mathbf{q}_j \sim \boldsymbol{\mu}_j$ are assumed to be potentially correct candidates for the orientation of cameras i and j, respectively and any relative rotation computed from those would be a valid candidate for explaining $\boldsymbol{\mu}_{ij}$.

5. Riemannian Particle Gradient Descent

Due to the non-convex nature of our problem, solving (19) is a highly challenging task. In this section, we will present our method built upon the recently proposed particle gradient descent (PGD) algorithms which are designed to solve optimization problems in the space of measures [20, 52, 5, 51]. Such PGD algorithms are often derived by first considering a continuous-time gradient flow in the Wasserstein spaces (i.e., the metric space associated with the Wasserstein distance), given as follows:

$$\mathrm{d}\boldsymbol{\mu}(t) = -\nabla_{\mathcal{W}}\Big(\mathcal{L}(\boldsymbol{\mu}(t)) + \mathcal{R}(\boldsymbol{\mu}(t))\Big)\mathrm{d}t, \quad (21)$$

where t denotes the (artificially introduced) 'time' and $\nabla_{\mathcal{W}}$ denotes a notion of gradient in the Wasserstein space [69]. Under appropriate conditions on $\mathcal{L}$ and $\mathcal{R}$, one can show that the trajectories of the flow $(\boldsymbol{\mu}(t))_{t \geq 0}$ converge to stable points of the optimization problem given in (19) [69, 3].

Since the trajectories of (21) are continuous in time, in general, they cannot be directly simulated on a computer. Hence, similar to the conventional Euclidean gradient descent algorithm that can be viewed as a discretization of a gradient flow in the Euclidean space, we can discretize the gradient flow in (21) to be able to simulate the trajectories in discrete-time. Such discretizations give rise to the

PGD algorithms, which are in general applicable to measures that are defined in Euclidean spaces. Unfortunately, standard PGD algorithms often cannot be directly applied to our problem since our measures are defined on a Riemannian manifold. Therefore, similar to [20], we consider a modified PGD algorithm, which we call the Riemannian PGD (RPGD). RPGD consists of two coupled equations. The first equation updates the location of the particles $\mathbf{q}_i^{(k)}$ on the quaternion manifold and doesn't depend on $\mathcal{R}(\boldsymbol{\mu})$:

$$\mathbf{q}_i^{(k)} \leftarrow Exp_{\mathbf{q}_i^{(k)}}\Big(-\eta_q(w_i^{(k)})\nabla_{\mathbf{q}_i^{(k)}}\mathcal{L}(\boldsymbol{\mu})\Big). \quad (22)$$

Here, Exp denotes the exponential map defined in Section 3 and $\eta_q(w_i^{(k)})$ is an adaptive step-size that depends on the weight of the particle $\mathbf{q}_i^{(k)}$ only. We consider either a constant step-size $\eta_q(w_i^{(k)}) = \eta_q$ for the Sinkhorn divergence or a step-size that is inversely proportional to the weight, i.e. $\eta_q(w_i^{(k)}) = \frac{\eta_q^0}{w_i^{(k)}}$ for the MMD as proposed in [20]. We will discuss this choice in more details later in this section. We also provide the analytical form of the gradients with respect to $\mathbf{q}_i^{(k)}$ in the supplementary document.

The second equation updates assumes a parametrization of the weights $w_i^{(k)}$ of the form $w_i^{(k)} = (\beta_i^{(k)})^2$ and uses a smaller step-size η_β. We recall the two cases considered: the *constrained case* (Sinkhorn-Sync) and the *unconstrained case* (MMD-Sync). In the latter, the weights are not constrained to sum to 1 and the penalty term $\mathcal{R}(\boldsymbol{\mu}(t))$ ensures they don't diverge:

$$\beta_i^{(k)} \leftarrow \beta_i^{(k)} - \eta_\beta \nabla_{\beta_i^{(k)}}\big(\mathcal{L}(\boldsymbol{\mu}) + \mathcal{R}(\boldsymbol{\mu})\big) \quad (23)$$

On the ohter hand, Sinkhorn-Sync enforces the weights to sum to 1 by performing optimization on the unit hypersphere:

$$\beta_i^{(k)} \leftarrow Exp_{\beta_i^{(k)}}\Big(-\eta_\beta \nabla_{\beta_i^{(k)}}\big(\mathcal{L}(\boldsymbol{\mu})\big)\Big). \quad (24)$$

We have implemented the proposed algorithm in Python using PyTorch [63] for GPU support. We provide a pseudocode in the supplementary document and release the full source code at synchinvision.github.io/probsync/.

Convergence result for regularized MMD. In this section, we will analyze the convergence behavior of the proposed RPGD algorithm on a special case of MMD-Sync. In
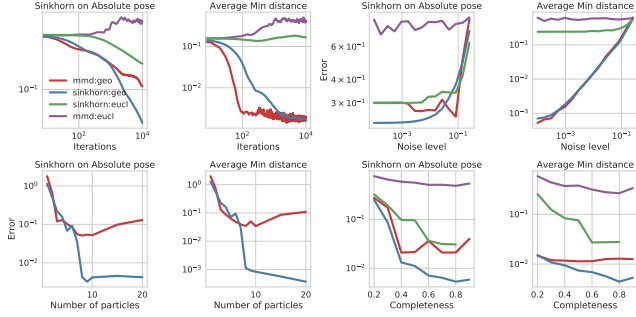
Figure 3. Controlled experiments in synthetic data: We show the performance (Sinkhorn and minimum distances) of our method against varying factors of noise, graph sparsity and number of particles. In all experiments the ground truth distribution has 3 modes.

particular, we restrict ourselves to the *Low entropy* case and where we choose MMD as the loss function and we choose the regularizer as in (20). This formulation enables us to analyze RPGD as a special case of the sparse optimization framework considered in [20]. Thanks to the convexity of the MMD loss, we will be able to show that RPGD will converge to the global optimum under the following conditions:

H1. *The kernel k is twice Fréchet differentiable, with locally Lipschitz second-order derivatives.*

H2. *Problem Eq* (19) *admits a unique global minimizer which is of the form* $\boldsymbol{\mu}^\star = \sum_{k=1}^{K^\star} v_k^\star \delta_{\mathbf{q}_k^\star}$ *with* $K^\star < K = K_1 = \cdots = K_n$.

H3. *The minimizer* $\boldsymbol{\mu}^\star$ *is non-degenerate in the sense of [20] Assumption 5.*

H4. *The weights* $w_i^{(k)}$ *are parametrized as:* $w_i^{(k)} = w_j^{(k)} = (\beta^{(k)})^2$ *for all* $1 \leq i, j \leq n$ *and the adaptive step-size in Eq* (22) *is chosen to be* $\eta_q^0/w_i^{(k)}$ *for some* $\eta_q^0 > 0$.

Theorem 1. *Consider the LE setting and Case 2 defined above. Assume that **H**1-4 hold. Then, the RPGD algorithm converges to the global optimum of* (19) *with a linear rate.*

Proof. The result is a consequence of [20], Thm. 3.9. □

We provide a more detailed definition of **H** 3 in the supplementary. This result shows that, even though our problem is non-convex, the RPGD algorithm will be able to find the global optimum thanks to over-parametrization (cf.**H** 2). We also note that the uniqueness condition in **H** 2 does not directly hold due to the antipodal symmetry of the quaternions. We circumvent this issue by making sure that the scalar part of each quaternion is positive.

6. Experimental Evaluation

We evaluate our algorithm qualitatively and quantitatively on a series of synthetic and real test beds. We use the HE case in all the settings except the PnP ambiguity application. Unless stated otherwise, we fix $\alpha = 0.05$ and

$p = 1.2$. We note that in all the settings, in order to eliminate the gauge ambiguity, we set the absolute measure of the first camera to its true distribution. This, for the case of a single particle, is $\mathbf{q}_1^{(1)} = [1, 0, 0, 0]^\top$ without loss of generality, while for the K-best prediction scenarios, we are obliged to use the ground truth particles. We leave it as a future study to remove this requirement.

6.1. Evaluations on synthetic data.

Similar to [14], we use a controlled setup to carefully observe the behaviour of our algorithm and its variants. To do so, we begin by synthesizing a dataset of $N = 10$ cameras where we could have access to ground truth absolute and relative pose distributions. We use $K_i = 3$ ground truth modes and from those generate $K_{ij} = 9$ relative rotation particles. Optionally, we corrupt the ground truth with gradual noise. To this end, a perturbation quaternion is applied on the relative quaternion. We set the learning rate to $\eta = 0.01$. We report the Sinkhorn-error attained as well as the actual geodesic distance (*Average Min distance*) between the matching particles. It is reassuring that these metrics behave similarly (*c.f*. Fig. 2 and Fig. 3). We run our algorithm for $t_{max} = 10000$ iterations until no improvement in the error is recorded. We analyze four variants of our algorithm: (1) MMD loss with Euclidean kernel (mmd:euc), (2) MMD loss with Geodesic kernel (mmd:geo), (3) Sinkhorn divergence with Euclidean kernel (sinkhorn:euc) and (4) Sinkhorn divergence with Geodesic kernel (sinkhorn:geo).

Iterations. We see that despite the theoretical guarantees, in practice minimizing the Sinkhorn loss yields better minimum. This is due to the fact that sometimes it is not possible to satisfy all the assumptions presented above Thm. 1. In fact, for all the other experiments, we have observed a better behaviour of Sinkhorn divergences and will argue for this method of choice. It is also noteworthy that MMD with a Euclidean kernel is not guaranteed to converge with the learning rates tested.

Noise resilience. We see that it is important to use the true metric (geo) when it comes to dealing with noise. Both MMD and Sinkhorn losses can handle reasonable noise levels. Note that while small levels of noise (~ 0.01) are tolerable, the measure synchronization problem gets more difficult with the increasing noise: we observe an exponential-like increase in the error attained after $\sigma > 0.01$.

How do we scale with the number of particles? The behavior of our algorithm when we don't know the exact number of modes in the data is of curiosity. We see that to we can recover the $K_{ij} = 9$ better when we use $K_i > 9$ absolute particles during estimation. This is not surprising and strongly motivated by the theory [20] which can give provable guarantees only when this number is over-estimated. Nevertheless, our experiments show that as long as we over-

Table 1. Results for finding the single-best solution on EPFL Benchmark [78]. We run a specific case of our algorithm where both $k_{abs} = 1$ and $k_{rel} = 1$, while all the other methods are specialized for this special case. We also use random initialization whereas [33, 83, 14] are initialized from a minimum spanning tree solution. The table reports the well accepted mean angular errors in degrees.

	Ozyesil *et al*. [62]	R-GODEC [10]	Govindu [33]	Torsello [83]	EIG-SE(3) [9]	TG-MCMC [14]	Ours
HerzJesus-P8	0.060	0.040	0.106	0.106	0.040	0.106	0.057
HerzJesus-P25	0.140	0.130	0.081	0.081	0.070	0.081	0.092
Fountain-P11	0.030	0.030	0.071	0.071	0.030	0.071	0.024
Entry-P10	0.560	0.440	0.101	0.101	0.040	0.090	0.190
Castle-P19	3.690	1.570	0.393	0.393	1.480	0.393	0.630
Castle-P30	1.970	0.780	0.631	0.629	0.530	0.622	0.753
Average	1.075	0.498	0.230	0.230	0.365	0.227	0.291

estimate, the algorithm is insensitive to the particular choice of this number, and what is affected is the runtime.

Robustness against graph sparsity. We observe that graphs of 50% completeness (connectedness) are sufficient for our algorithm to get good results. We have diminishing returns from making the graph more connected. Nevertheless, Sinkhorn:geo seems to benefit the most from the increasing number of edges.

6.2. Evaluations on real data

Single best solution. Classical multiple rotation averaging (or synchronization) is a special case of the problem we pose. When our algorithm is run with a single particle both for the absolute and relative rotations, we recover the solution to the classical problem. While our aim is not to explicitly tackle this challenge, being able to handle such special case is a sanity check for us. We present our results in Tab. 1 where we report the mean angular errors in degrees on the standard EPFL Benchmark dataset [78]. We observe that our performance is on par with state of the art on this real dataset, even when starting from a random initialization. Moreover, unlike TG-MCMC [14] our approach is non-parametric and we do not perform any scene specific tuning. Note that we do not process the translation information which is an additional cue for the other methods.

Resolving the PnP ambiguity. Our algorithm can be used to treat many multiview vision problems with their inherent ambiguities. One such ambiguity arises in the Perspective N-point problem [49] where the single view camera pose estimation from a planar target admits two solutions. While targeted synchronization approaches do exist [21], our framework can be a natural fit. We test our algorithm on the multiview dataset of MarkerMapper [58]. We first estimate the pairwise poses between cameras from randomly selected markers and keep both of the possible solutions. We then use 2 particles per camera with joint coupling to recover the entirety of the plausible solution space. Further details are given in the suppl. document. We attain a minimum error of $7°$, a quantity that is on par with PnP-specific synchronizers [21]. We present qualitative results from our estimations in Fig. 4.
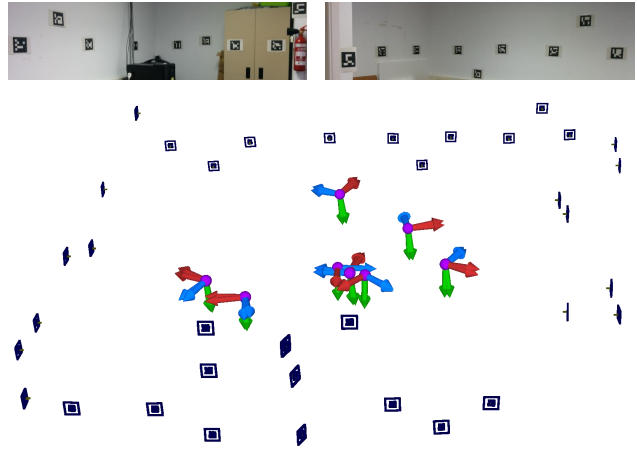

Figure 4. Our estimation vs. ground truth particles in the markers dataset. Estimated orientations are displayed at common camera centers of RGB frames. On the top row we show two images from the dataset of [58]. 3D location of the markers are shown as black squares. The small deviation is the error we make.

7. Conclusion

We introduced the problem of *measure syncronization*: synchronizing graphs whose edges are probability measures. We then extended it to Riemannian measures and formulated the problem as maximization of a novel cycle-consistency notion over the space of measures. After formally defining the problem, we developed a Riemannian particle gradient descent algorithm for estimating the marginal distributions on absolute poses. In addition to being a practical algorithm, by drawing connections to recent optimization methods, we showed that the proposed algorithm converges to the global optimum of a special case of the problem. We showed the validity of our approach via qualitative and quantitative experiments. Our future work will address better design of composition functions g, exploration of sliced OT distances [60, 59, 45], and applications to tackling more ambiguities in vision such as essential matrix estimation.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009. 2

[2] K. Aftab, R. Hartley, and J. Trumpf. Generalized weiszfeld algorithms for lq optimization. *IEEE transactions on pattern analysis and machine intelligence*, 37(4):728–745, 2014. 6

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008. 6

[4] J. Angulo. Riemannian l p averaging on lie group of nonzero quaternions. *Advances in Applied Clifford Algebras*, 24(2):355–382, 2014. 3

[5] M. Arbel, A. Korba, A. Salim, and A. Gretton. Maximum mean discrepancy gradient flow. *arXiv preprint arXiv:1906.04370*, 2019. 2, 3, 6

[6] M. Arbel, D. Sutherland, M. Bińkowski, and A. Gretton. On gradient regularizers for mmd gans. In *Advances in Neural Information Processing Systems*, pages 6700–6710, 2018. 3

[7] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017. 2, 5

[8] F. Arrigoni and A. Fusiello. Synchronization problems in computer vision with closed-form solutions. *International Journal of Computer Vision*, Sep 2019. 1, 2

[9] F. Arrigoni, A. Fusiello, and B. Rossi. Spectral motion synchronization in se (3). *arXiv preprint arXiv:1506.08765*, 2015. 8

[10] F. Arrigoni, L. Magri, B. Rossi, P. Fragneto, and A. Fusiello. Robust absolute rotation estimation via low-rank and sparse matrix decomposition. In *3D Vision (3DV), 2014 2nd International Conference on*, volume 1. IEEE, 2014. 8

[11] F. Arrigoni, E. Maset, and A. Fusiello. Synchronization in the symmetric inverse semigroup. In *International Conference on Image Analysis and Processing*, pages 70–81. Springer, 2017. 2

[12] F. Arrigoni, B. Rossi, and A. Fusiello. Spectral synchronization of multiple views in se (3). *SIAM Journal on Imaging Sciences*, 9(4):1963–1990, 2016. 2

[13] T. Birdal and U. Simsekli. Probabilistic permutation synchronization using the riemannian structure of the birkhoff polytope. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019. 1, 2

[14] T. Birdal, U. Simsekli, M. O. Eken, and S. Ilic. Bayesian pose graph optimization via bingham distributions and tempered geodesic mcmc. In *Advances in Neural Information Processing Systems*, pages 308–319, 2018. 1, 2, 7, 8

[15] J. Briales and J. Gonzalez-Jimenez. Cartan-sync: Fast and global se (d)-synchronization. *IEEE Robotics and Automation Letters*, 2(4):2127–2134, 2017. 2

[16] B. Busam, T. Birdal, and N. Navab. Camera pose filtering with local regression geodesics on the riemannian manifold of dual quaternions. In *IEEE International Conference on Computer Vision Workshop (ICCVW)*, October 2017. 3

[17] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics*, 32(6):1309–1332, 2016. 1

[18] L. Carlone, R. Tron, K. Daniilidis, and F. Dellaert. Initialization techniques for 3d slam: a survey on rotation estimation and its use in pose graph optimization. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 4597–4604. IEEE, 2015. 1

[19] A. Chatterjee and V. M. Govindu. Robust relative rotation averaging. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):958–972, 2017. 2

[20] L. Chizat. Sparse optimization on measures with over-parameterized gradient descent. *arXiv preprint arXiv:1907.10300*, 2019. 2, 5, 6, 7

[21] S.-F. Ch'ng, N. Sogi, P. Purkait, T.-J. Chin, and K. Fukui. Resolving marker pose ambiguity by robust rotation averaging with clique constraints. *arXiv preprint arXiv:1909.11888*, 2019. 8

[22] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2016. 2

[23] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013. 2, 4

[24] H. Deng, T. Birdal, and S. Ilic. 3d local features for direct pairwise registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3244–3253, 2019. 4

[25] A. Eriksson, C. Olsson, F. Kahl, and T.-J. Chin. Rotation averaging and strong duality. In *The IEEE Conference on Comptuter Vision and Pattern Recognition (CVPR)*, June 2018. 2

[26] J. Feydy, T. Séjourné, F.-X. Vialard, S.-i. Amari, A. Trouve, and G. Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In K. Chaudhuri and M. Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*. PMLR, 16–18 Apr 2019. 2, 4

[27] P. Galison. *Einstein's clocks, Poincaré's maps: Empires of time*. WW Norton & Company, 2003. 2

[28] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré. Sample complexity of sinkhorn divergences. *arXiv preprint arXiv:1810.02733*, 2018. 2, 5

[29] A. Genevay, G. Peyré, and M. Cuturi. Learning generative models with sinkhorn divergences. *arXiv preprint arXiv:1706.00292*, 2017. 4

[30] A. Giridhar and P. R. Kumar. Distributed clock synchronization over wireless networks: Algorithms and analysis. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 4915–4920. IEEE, 2006. 1, 2

[31] S. Golomb, J. Davey, I. Reed, H. Van Trees, and J. Stiffler. Synchronization. *IEEE Transactions on Communications Systems*, 11(4):481–491, 1963. 2

[32] V. M. Govindu. Combining two-view constraints for motion estimation. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE, 2001. 3

[33] V. M. Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004*

IEEE Computer Society Conference on, volume 1, pages I–I. IEEE, 2004. 1, 2, 8

[34] V. M. Govindu and A. Pooja. On averaging multiview relations for 3d scan registration. *IEEE Transactions on Image Processing*, 23(3):1289–1302, 2014. 1, 2

[35] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola. A kernel method for the two-sample-problem. In *Advances in neural information processing systems*, pages 513–520, 2007. 2, 4

[36] R. Hartley, J. Trumpf, Y. Dai, and H. Li. Rotation averaging. *International journal of computer vision*, 103(3), 2013. 1, 3

[37] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 1

[38] Q. Huang, Z. Liang, H. Wang, S. Zuo, and C. Bajaj. Tensor maps for synchronizing heterogeneous shape collections. *ACM Transactions on Graphics (TOG)*, 38(4):106, 2019. 2

[39] X. Huang, Z. Liang, X. Zhou, Y. Xie, L. J. Guibas, and Q. Huang. Learning transformation synchronization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8082–8091, 2019. 2

[40] L. V. Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133(4):1381–1382, 2006. 2

[41] A. Kendall and R. Cipolla. Modelling uncertainty in deep learning for camera relocalization. In *2016 IEEE international conference on Robotics and Automation (ICRA)*, pages 4762–4769. IEEE, 2016. 1

[42] A. Kendall, M. Grimes, and R. Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938–2946, 2015. 1

[43] N. Kolkin, J. Salavon, and G. Shakhnarovich. Style transfer by relaxed optimal transport and self-similarity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10051–10060, 2019. 2

[44] S. Kolouri, K. Nadjahi, U. Simsekli, R. Badeau, and G. Rohde. Generalized sliced Wasserstein distances. In *Advances in Neural Information Processing Systems*, 2019. 2

[45] S. Kolouri, K. Nadjahi, U. Simsekli, and S. Shahrampour. Generalized sliced distances for probability distributions. *arXiv preprint arXiv:2002.12537*, 2020. 8

[46] S. Kolouri, P. E. Pope, C. E. Martin, and G. K. Rohde. Sliced-wasserstein autoencoder: an embarrassingly simple generative model. *arXiv preprint arXiv:1804.01947*, 2018. 5

[47] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951. 2, 4

[48] T. Lacombe, M. Cuturi, and S. Oudot. Large scale computation of means and clusters for persistence diagrams using optimal transport. In *Advances in Neural Information Processing Systems*, pages 9770–9780, 2018. 2

[49] V. Lepetit, F. Moreno-Noguer, and P. Fua. Epnp: An accurate o (n) solution to the pnp problem. *International journal of computer vision*, 81(2):155, 2009. 8

[50] C.-L. Li, W.-C. Chang, Y. Cheng, Y. Yang, and B. Póczos. Mmd gan: Towards deeper understanding of moment matching network. In *Advances in Neural Information Processing Systems*, pages 2203–2213, 2017. 3

[51] Q. Liu. Stein variational gradient descent as gradient flow. In *Advances in neural information processing systems*, pages 3118–3126, 2017. 2, 6

[52] A. Liutkus, U. Simsekli, S. Majewski, A. Durmus, and F.-R. Stöter. Sliced-wasserstein flows: Nonparametric generative modeling via optimal transport and diffusions. In *International Conference on Machine Learning*, pages 4104–4113, 2019. 2, 6

[53] P. C. Mahalanobis. On the generalized distance in statistics. National Institute of Science of India, 1936. 2

[54] F. Manhardt, D. M. Arroyo, C. Rupprecht, B. Busam, T. Birdal, N. Navab, and F. Tombari. Explaining the ambiguity of object detection and 6d pose from visual data. In *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2019. 1

[55] I. Melekhov, J. Ylioinas, J. Kannala, and E. Rahtu. Relative camera pose estimation using convolutional neural networks. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 675–687. Springer, 2017. 1, 4

[56] L. Mi, W. Zhang, X. Gu, and Y. Wang. Variational wasserstein clustering. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 322–337, 2018. 2

[57] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781. 2

[58] R. Munoz-Salinas, M. J. Marin-Jimenez, E. Yeguas-Bolivar, and R. Medina-Carnicer. Mapping and localization from planar markers. *Pattern Recognition*, 73:158–171, 2018. 8

[59] K. Nadjahi, A. Durmus, L. Chizat, S. Kolouri, S. Shahrampour, and U. Şimşekli. Statistical and topological properties of sliced probability divergences. *arXiv preprint arXiv:2003.05783*, 2020. 8

[60] K. Nadjahi, A. Durmus, U. Simsekli, and R. Badeau. Asymptotic guarantees for learning generative models with the sliced-wasserstein distance. In *Advances in Neural Information Processing Systems*, pages 250–260, 2019. 8

[61] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6):0756–777, 2004. 1, 4

[62] O. Özyeşil, A. Singer, and R. Basri. Stable camera motion estimation using convex programming. *SIAM Journal on Imaging Sciences*, 8(2):1220–1262, 2015. 8

[63] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in PyTorch. In *NIPS Autodiff Workshop*, 2017. 6

[64] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011. 2

[65] D. Rosen, L. Carlone, A. Bandeira, and J. Leonard. SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group. Technical Report MIT-CSAIL-TR-2017-002, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, Feb. 2017. 2

[66] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard. Se-sync: A certifiably correct algorithm for synchronization

over the special euclidean group. *The International Journal of Robotics Research*, 38(2-3):95–125, 2019. 2

[67] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000. 2

[68] R. F. Salas-Moreno, R. A. Newcombe, H. Strasdat, P. H. Kelly, and A. J. Davison. Slam++: Simultaneous localisation and mapping at the level of objects. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1352–1359, 2013. 1

[69] F. Santambrogio. {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences*, 7(1):87–154, 2017. 6

[70] B. Simons. An overview of clock synchronization. In *Fault-Tolerant Distributed Computing*. Springer, 1990. 1, 2

[71] A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36, 2011. 1, 2

[72] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879, 1964. 2

[73] R. Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4):402–405, 1967. 4

[74] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967. 2

[75] J. Solomon, F. De Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics*, (4), 2015. 2

[76] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher. Earth mover's distances on discrete surfaces. *ACM Transactions on Graphics (TOG)*, 33(4):67, 2014. 2

[77] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher. Wasserstein propagation for semi-supervised learning. In *International Conference on Machine Learning*, 2014. 3

[78] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. Ieee, 2008. 8

[79] Y. Sun, J. Zhuo, A. Mohan, and Q. Huang. K-best transformation synchronization. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, 2019. 2

[80] D. J. Sutherland, H.-Y. Tung, H. Strathmann, S. De, A. Ramdas, A. Smola, and A. Gretton. Generative models and model criticism via optimized maximum mean discrepancy. In *International Conference on Learning Representations (ICLR)*. OpenReview.net, 2017. 3

[81] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf. Wasserstein auto-encoders. *arXiv preprint arXiv:1711.01558*, 2017. 5

[82] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schölkopf. Wasserstein auto-encoders. In *International Conference on Learning Representations (ICLR 2018)*. OpenReview. net, 2018. 3

[83] A. Torsello, E. Rodola, and A. Albarelli. Multiview registration via graph diffusion of dual quaternions. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2441–2448. IEEE, 2011. 8

[84] R. Tron and K. Daniilidis. Statistical pose averaging with non-isotropic and incomplete relative measurements. In *European Conference on Computer Vision*. Springer, 2014. 2

[85] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014. 3

[86] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008. 2

[87] H. Xu, D. Luo, H. Zha, and L. C. Duke. Gromov-Wasserstein learning for graph matching and node embedding. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6932–6941, Long Beach, California, USA, 09–15 Jun 2019. PMLR. 2