Wasserstein K-Means for Clustering Tomographic Projections

Rohan Rao
Princeton University
rohanr@princeton.edu

Amit Moscovich
Princeton University
amit@moscovich.org

Amit Singer
Princeton University
amits@math.princeton.edu

Abstract

Motivated by the 2D class averaging problem in single-particle cryo-electron microscopy, we present a k-means algorithm based on a rotationally-invariant Wasserstein metric for images. Unlike existing methods that are based on Euclidean (L_2) distances, we prove that the Wasserstein metric better accommodates for the out-of-plane angular differences between different particle views. We demonstrate on a synthetic dataset that our method gives superior results compared to an L_2 baseline. Furthermore, there is little computational overhead, thanks to the use of a fast linear-time approximation to the Wasserstein-1 metric, also known as the Earthmover's distance.

1 Introduction

Single particle cryo-electron microscopy (cryo-EM) is a powerful method for reconstructing the 3D structure of individual proteins and other macromolecules (Frank, 2006; Cheng, 2018). In a cryo-EM experiment, a sample of the molecule of interest is rapidly frozen, forming a thin sheet of vitreous ice, and then imaged using a transmission electron microscope. This results in a set of images that contain noisy tomographic projections of the electrostatic potential of the molecule at random orientations. The images then undergo a series of processing steps, including:

- Motion correction.
- Estimation of the contrast transfer function (CTF) or point-spread function.
- Particle picking, where the position of the molecules in the ice sheet is identified.
- 2D class averaging, where similar particle images are clustered together and averaged.
- Filtering of bad images (typically, entire clusters from the previous step).
- Reconstruction of a 3D molecular model (or models).

All of these steps are done using special-purpose software, for example: Tang et al. (2007); Scheres (2012); Lyumkis et al. (2013); de la Rosa-Trevín et al. (2013); Rohou and Grigorieff (2015); Punjani et al. (2017); Heimowitz et al. (2018); Grant et al. (2018).

Cryo-EM has been gaining popularity and its importance has been recognized in the 2017 Nobel prize in Chemistry (Cressey and Callaway, 2017). An examination of the Protein Data Bank (Berman, 2000) shows that in 2020, 16% of entries had structures that were determined by cryo-EM, compared to 12% in 2019, and 7% in 2018. For more on cryo-EM and its current challenges, see the reviews of Vinothkumar and Henderson (2016); Lyumkis (2019); Singer and Sigworth (2020).

1.1 Image formation in cryo-EM

Given a 3D molecule with an electrostatic potential function $\rho : \mathbb{R}^3 \to \mathbb{R}$ (henceforth "particle"), the images captured by the electron microscope are modeled as noisy tomographic projections along

Machine Learning for Structural Biology Workshop, NeurIPS 2020, Vancouver, Canada.

different orientations. A tomographic projection of the particle onto the x-y plane is defined as $(\mathcal{T}\rho)(x,y) = \int_{\mathbb{R}} \rho(x,y,z)dz$. More general tomographic projections are given by a rotation R of the potential function ρ followed by projection onto the x-y plane,

$$\mathcal{T}_R\rho = \int_{\mathbb{R}} (R\rho)(x,y,z)dz \tag{1}$$

where $R \in \mathrm{SO}(3)$, and we define $(R\rho)(x,y,z) = \rho(R^T(x,y,z))$. Particle images in cryo-EM are typically modeled as follows,

$$\text{Image} = h * \mathcal{T}_R\rho + \boldsymbol{\epsilon}, \tag{2}$$

where h is a point spread function that is convolved with the projection and $\boldsymbol{\epsilon}$ is Gaussian noise. See Figure 1 for an example of a tomographic projection. We define the viewing angle of our particle as a unit vector $\mathbf{v}$ that represents orientation of our particle $R \in \mathrm{SO}(3)$ modulo in-plane rotations (which can be represented as members of $\mathrm{SO}(2)$).

1.2 Clustering tomographic projections

The imaging process in cryo-EM involves high-levels of noise. To improve the signal-to-noise ratio (SNR), it is common to cluster the noisy projection images with ones that are similar up to an in-plane rotations. This clustering task is called "2D classification and averaging" in the cryo-EM literature and the clusters are called classes. The results of this stage have multiple uses, including:

- Template selection from 2D class averages for particle picking (Frank and Wagenknecht, 1983; Chen and Grigorieff, 2007; Scheres, 2015).
- Particle sorting (Zhou et al., 2020), discarding images that belong to bad clusters.
- Visual assessment and symmetry detection.
- Ab-initio modeling, where an initial 3D model (or set of models) is constructed to be refined in later stages (Greenberg and Shkolnisky, 2017; Punjani et al., 2017).

Existing methods for 2D clustering include the reference-free alignment algorithm that tries to find a global alignment of all images (Penczek et al., 1992), clustering based on invariant functions and multivariate statistical analysis (Schatz and Van Heel, 1990), rotationally invariant k-means (Penczek et al., 1996) and the Expectation-Maximization (EM) based approach of Scheres et al. (2005). All of these methods use the L_2 distance metric on rotated images.

Let us consider a simple centroid+noise model, where each image in a cluster is generated by adding Gaussian noise to a particular clean view $\boldsymbol{\mu}$ that is the (oracle) centroid,

$$\text{Image}_i \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I). \tag{3}$$

In that case the L_2 metric is (up to constants) nothing but the log-likelihood of the image patch, conditioned on $\boldsymbol{\mu}$,

$$\log \mathcal{L}(\text{Image}_i|\boldsymbol{\mu}) = -\|\text{Image}_i - \boldsymbol{\mu}\|^2/2\sigma^2 + C_\sigma. \tag{4}$$

Under the centroid+noise model, the commonly-used L_2 distance metric seems perfectly suitable. However, real particle images have many other sources of variability, including angular differences, in-plane shifts and various forms of molecular heterogeneity. For these sources of variability, the L_2 distance metric is ill-suited. See Section 4 for more on this.

In this work we propose to use the Wasserstein-1 metric W_1, also known as the Earthmover's distance, as an alternative to the L_2 distance for comparing particle images. In Section 2 we describe a variant of the rotationally invariant k-means that is based on a fast linear-time approximation of W_1 and in Section 3.2 we demonstrate superior performance on a synthetic dataset of Ribosome projections. In Section 4 we analyze the behavior of the W_1 and L_2 metrics theoretically with respect to angular differences of tomographic projections. In particular we show that the rotationally-invariant W_1 metric has the nice property that it is bounded from above by the angular difference of the projections. On the other hand, the L_2 metric shows no such relation.

2 Methods

In cryo-EM, in-plane rotations of the molecular projections are assumed to be uniformly distributed, hence is is desirable for the distance metric to be invariant to in-plane rotations. A natural candidate is the rotationally-invariant Euclidean distance,

$$L_2^R(I_1, I_2) := \min_{R \in \mathrm{SO}(2)} ||I_1 - RI_2||_2. \tag{5}$$

A drawback of this metric is that visually similar projection images that have a small viewing angle difference can have an arbitrarily large L_2 distance. See discussion in Section 4.

To address this, we define a metric based on the Wasserstein-p Metric between two probability distributions (Villani, 2003). This metric measures the "work" it takes to transport the mass of one probability distribution to the other, where work is defined as the amount of mass times the distance (on the underlying metric space) between the origin and destination of the mass. More formally, the Wasserstein-p distance between two normalized greyscale $N \times N$ images is defined as

$$W_p(I_1, I_2) = \inf_{\pi \in \Pi(I_1, I_2)} \sum_{u \in [N]^2} \sum_{v \in [N]^2} ||u - v||^p \pi(u, v), \tag{6}$$

where $\Pi(I_1, I_2)$ is the set of joint distributions on $[N]^2$ with marginals I_1, I_2 respectively.

In cryo-EM, the Wasserstein-1 metric has been used to understand the conformation space of 3D volumes of molecules (Zelesko et al., 2020). For clustering tomographic projections, we construct a rotationally-invariant W_p distance to be our clustering metric analogously to Eq. (5)

$$W_p^R(I_1, I_2) := \min_{R \in \mathrm{SO}(2)} W_p(I_1, RI_2). \tag{7}$$

2.1 Clustering with rotationally-invariant metrics

Algorithm 1 describes a generic rotationally-invariant k-means algorithm based on an arbitrary image patch distance metric d. The choice $d = L_2$ gives the rotationally-invariant k-means of Penczek et al. (1996). By supplying $d = W_1$ we get a new rotationally-invariant k-means algorithm based on the W_1 distance. We choose $p = 1$ for the W_p distance as it admits a fast linear-time wavelet approximation as we will see in the next section. For both choices of the metric, we initialize the centroids using a rotationally-invariant k-means++ initialization which we describe in Algorithm 2. When we denote $r \in \mathrm{rot}$ we mean that r performs an in-plane rotation of an image. We approximate the space of all in-plane rotation angles by a discrete set of angles.

Algorithm 1: Rotationally invariant k-means

Parameters: k, n, d
Output Cluster Averages: $\{C_j\}_{j \in [k]}$
Input Images: $\{I_i\}_{i \in [n]}$

$\{C_j\}_{j \in [k]} := \mathrm{InitializeCenters}(\{I_i\}_{i \in [n]})$
while loss decreases **do**
    loss $= 0$
    for $i \in [n]$ **do**
        $(j_i, r_i) := \mathrm{argmin}_{(j,r): j \in [k], r \in \mathrm{rot}}\, d(I_i, r(C_j))$
        loss $=$ loss $+ d(I_i, r_i(C_{j_i}))^2$
    end
    for $j \in [k]$ **do**
        $C_j := \mathrm{mean}(\{r_i^{-1}(I_i) : j_i = j\})$
    end
end

return $\{C_j\}_{j \in [k]}$

Algorithm 2: Rotationally invariant k-means++ initialization

Parameters: k, n, d
Output Cluster Averages: $\{C_j\}_{j\in[k]}$
Input Images: $\{I_i\}_{i\in[n]}$

$C_1 = \text{RandomSelect}(\{I_i\}_{i\in[n]})$
for $j \in \{2, ..., k\}$ **do**
  for $i \in [n]$ **do**
    $p_i := \left(\min_{m\in[j-1], r\in\text{rot}} d(I_i, r(C_m))\right)^2$
  end
  $j = \text{DrawFrom}(\mathbf{p})$ `// Draw index` j `with probability proportional to` p_j
  $C_j = I_j$
end

return $\{C_j\}_{j\in[k]}$

2.2 Computing the Earthmover's distance

Computing the distance $W_1(I_1, I_2)$ (also known as the Earthmover's distance) between two $N \times N$ pixels can be formulated as a linear program in $O(N^2)$ variables and $O(N)$ constraints. Given n images, k centers, and m rotations for each image and t iterations of k-means, we have to compute $O(n \times m \times k \times t)$ distances. Computing the W_1 distance between two 100×100 size images using a standard LP solver takes on average of 5 seconds to compute using the Python Optimal Transport Library of Flamary and Courty (2017) on a single core of a 7th generation 2.8 GHz Intel Core i7 processor. Computing the exact W_1 distance over all the rotations of all the images over all the iterations is prohibitively slow.

Fortunately, the W_1 distance admits a fast linear-time approximation based on the two-dimensional fast wavelet transform. Let $\mathcal{W}I_i(\lambda)$ be the 2D wavelet transform of I_i, where $\lambda = (k, s, (n_x, n_y))$ is an index to the wavelet coefficients (Mallat, 2009, Ch. 7.7.2)). In this notation $k \in \{1, 2, 3\}$ selects between one of three mother wavelets, the scale (or dilation) parameter is s and the two-dimensional translation of the wavelet is specified by (n_x, n_y). We can approximate the W_1 distance between two images I_1, I_2 using a weighted ℓ_1 distance between their wavelet coefficients, with an exponential weight given by the scale parameter (Shirdhonkar and Jacobs, 2008),

$$\widetilde{W}_1(I_1, I_2) := \sum_{\lambda} 2^{-2\lambda_s} |(\mathcal{W}I_1)(\lambda) - (\mathcal{W}I_2)(\lambda)|. \tag{8}$$

This metric is strongly equivalent to W_1. i.e. there exist $0 < c < C$ such that for all $I_1, I_2 \in \mathbb{R}^{N^2}$,

$$c \cdot \widetilde{W}_1(I_1, I_2) \leq W_1(I_1, I_2) \leq C \cdot \widetilde{W}_1(I_1, I_2). \tag{9}$$

Different choices of the wavelet basis give different ratios C/c. We chose the symmetric Daubechies wavelets of order 5 due to the quality of their approximation (Shirdhonkar and Jacobs, 2008). The sum in Eq. (8) was computed with level up to 6 using the PyWavelets package (Lee et al., 2019).

3 Experimental results

3.1 Dataset generation

We built a synthetic dataset of 10,000 tomographic projections of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine, EMD-2660 (Wong et al., 2014), shown in Figure 1. To generate each image, we randomly rotated the ribosome around its center of mass using a uniform draw of $\text{SO}(3)$, projected it to 2D by summing over the z axis and resized the resulting image to 128×128. Finally we added i.i.d. $\mathcal{N}(0, \sigma^2)$ noise to the image at different signal-to-noise (SNR) levels. Given a dataset of images $S \in \mathbb{R}^{N\times N\times n}$, the SNR is defined as

$$\text{SNR} := \frac{\|D\|^2}{nN^2\sigma^2}. \tag{10}$$

We produce three datasets to run experiments on by adding noise at SNR values $\{1/8, 1/12, 1/16\}$.

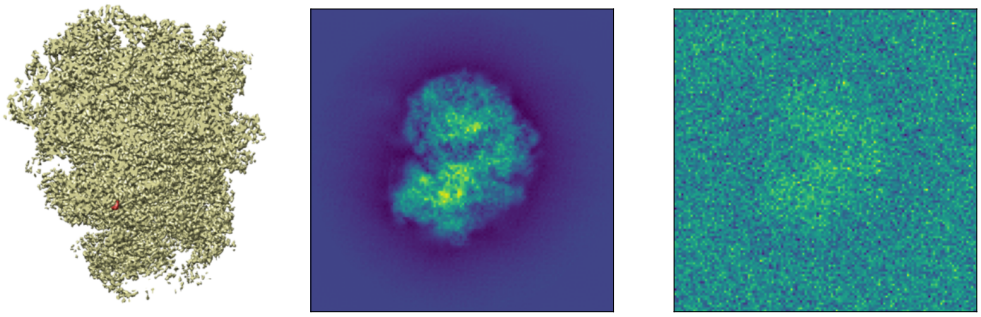

Figure 1: (left) Surface plot of the 3D electrostatic potential of the 80S ribosome; (middle) Example tomographic projection; (right) The same projection with Gaussian noise at SNR=1/16.

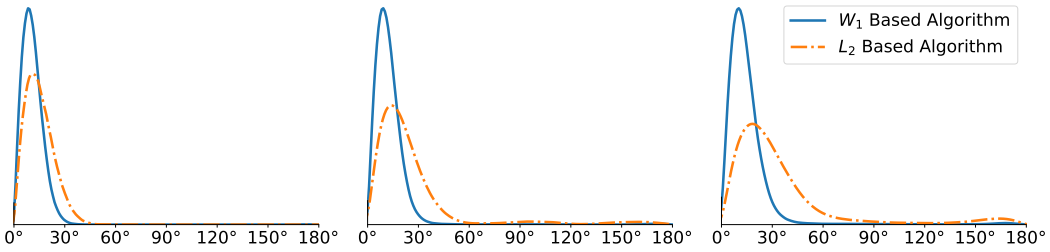

Figure 2: Distribution of within-cluster pairwise angular differences (narrower is better) at SNR values $1/8,\ 1/12,\ 1/16$ from left to right.

3.2 Simulation results

We performed 2D clustering on our three datasets using rotationally-invariant k-means++ (Algorithms 1, 2) with the $\widetilde{W}_1$ and L_2 distance using $k = 150$ clusters. All in-plane rotations at increments of $\pi/100$ radians were tested in the Algorithm 1. To quantify the difference in performance, we computed the distribution of within-cluster viewing angle differences for all pairs of images assigned to the same cluster which we visualize in Figure 2. For a molecule like the Ribosome that has no symmetries, we expect these angular differences to be concentrated around zero, since large angular differences typically give large differences in the projection images. For all SNR levels, we see that $\widetilde{W}_1$ gives better angular coherence than L_2. Some of the distributions have a small peak near $180°$ which corresponds to images where our algorithms is confusing antipodal viewing angles.

In Figure 3 we show the cluster means of the 8 largest clusters at SNR $1/16$. Visually, we can see that the algorithm based on $\widetilde{W}_1$ produces sharper mean images. Finally, we examine the occupancy of the clusters. The $\widetilde{W}_1$ algorithm provides more evenly sized clusters, whereas for the L_2 algorithm we see a few very populated centers and a large drop-off in occupancy in the other clusters. When clustering images with Gaussian noise, the averages of larger clusters will tend to be less noisy, since the noise variance is inversely proportional to the cluster size. Due to the lower noise levels, more of the images will be assigned to the larger clusters, making them even larger. This "rich get richer" phenomenon has been observed in cryo-EM (Sorzano et al., 2010). It can explain the large differences in occupancy visible in the top panels of Figure 4, despite the fact that the angles were drawn uniformly. The Wasserstein distance is more resilient to i.i.d. noise and this may explain the uniformity in the resulting cluster sizes seen in Figure 4.

Finally, clustering with the $\widetilde{W}_1$ distance does not increase the runtime of the clustering algorithm by much. We include timing results per iteration of K-means with each metric, along with the number of iterations it took for the algorithm to converge in Table 1.

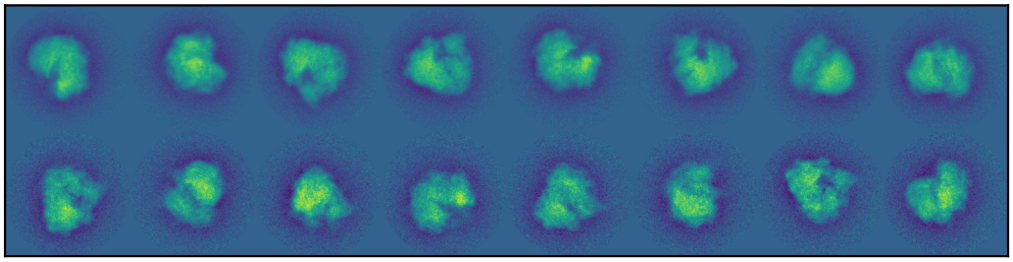

Figure 3: Means of the largest 8 clusters for the SNR=$1/16$ dataset, sorted by cluster size from left to right. (top) L_2 distance; (bottom) W_1 metric. We can see that the W_1 metric preserves more details than the L_2 distance. The portion of the image that does not contain our particle appears less noisy in the L_2 averages, however this is just an artifact of the cluster size. By averaging a large number of images, the noise decreases but the signal also deteriorates.

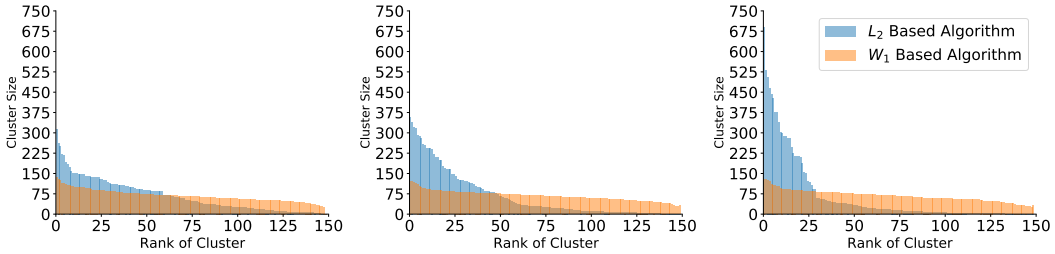

Figure 4: Cluster sizes for the datasets with signal-to-noise $1/8$, $1/12$, $1/16$ from left to right.

3.3 Sensitivity to noise

To examine the effect of noise on the $\widetilde{W}_1^R$ and L_2^R distances, we plot the $\widetilde{W}_1^R$ and L_2^R distances against the viewing angle difference between projections of the ribosome. In Figure 5 we can see that $\widetilde{W}_1^R$ continues to give meaningful distances under noise compared to the L_2^R distance.

4 Theory

For a given particle, two tomographic projections from similar viewing angles typically produce images that are similar. Hence, it is desirable that a metric for comparing tomographic projections will assign a small distance to projections that have a small difference in viewing angle. We show that the rotationally-invariant Wasserstein metric satisfies this property.

Proposition 1. *Let $\rho : \mathbb{R}^3 \to \mathbb{R}_{\geq 0}$ be a probability distribution supported on the 3D unit ball and let I_1 and I_2 be its tomographic projections along the vectors $\mathbf{u}$ and $\mathbf{v}$ respectively. Denote by $\measuredangle(\mathbf{u}, \mathbf{v}) \in [0, \pi]$ the angle between the vectors, then*

$$W_p^R(I_1, I_2) \leq [2\sin(\measuredangle(\mathbf{u}, \mathbf{v})/2)]^p \leq \measuredangle(\mathbf{u}, \mathbf{v})^p \tag{11}$$

where W_p^R is the rotationally-invariant Wasserstein metric defined in Eq. (7).

Table 1: Seconds per iteration averaged (over two runs) for the L_2 and $\widetilde{W}_1$ metrics and the number of iterations before convergence at different SNRs (for one run). The programs were run using 32 Intel core i7 CPU cores which allow us to compute the distances from images in our dataset to all the cluster averages in a parallelized fashion across the averages.

Metric	k-means (sec/iteration)	n_{iter} (SNR $= 1/8$)	n_{iter} (SNR $= 1/12$)	n_{iter} (SNR $= 1/16$)
L_2	1204 s	13	31	22
W_1	1676 s	23	27	26

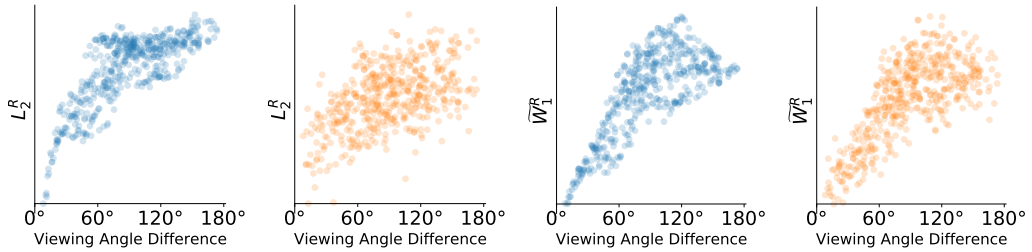

Figure 5: $\widetilde{W}_1^R$ and L_2^R distances versus the viewing angle difference between a fixed random clean projection and 500 random noisy projections (orange) and 500 clean images (blue). SNR = $1/16$

See Appendix A for the proof. A similar upper-bound for the rotationally-invariant L_2 distance cannot hold for all densities ρ. To see why, consider an off-center point mass. Any two projections at slightly different viewing angles will have a large L_2^R distance no matter how small their angular difference is. However, for densities with bounded gradients it is possible to produce upper bounds.

Proposition 2. *Let $B = \sup_{\mathbf{x}} |\nabla \rho(\mathbf{x})|$ be an upper bound on the absolute gradient of the density. Under the same conditions of Proposition 1 we have $L_2^R(I_1, I_2) \leq 2\sqrt{\pi} B \measuredangle(\mathbf{u}, \mathbf{v})$.*

The proof is in Appendix A. This bound suggests that L_2^R is a reasonable metric to use for smooth signals. For non-smooth signals, or signals with very large B, this means that there is no guarantee that the L_2^R distance will assign a small distance between projections with a small viewing angle.

5 Discussion

From our numerical experiments, we see that the rotationally-invariant Wasserstein-1 k-means clustering algorithm produces clusters that have more angular coherence and sharper cluster means than the rotationally-invariant L_2 clustering algorithm. Furthermore, at low SNRs we see that the Wasserstein-1 distance continues to give meaningful differences between projections. Thus, we believe it is a promising alternative to commonly used rotationally-invariant clustering algorithms based on the L_2 distance. In future work we would like to investigate what component of the success of our method comes from having a metric that is robust to noise, and what component comes from better reflecting angular differences.

Recently, there has been an explosion of interest in the analysis of molecular samples with continuous variability (Nakane et al., 2018; Sorzano et al., 2019; Moscovich et al., 2020; Lederman et al., 2020; Zhong et al., 2020; Dashti et al., 2020). In the presence of continuous variability, it is much more challenging to employ 2D class averaging and related techniques. This is due to the fact that the clusters need to account not only for the entire range of viewing directions but also for all the possible 3D conformations of the molecule of interest. In future work we would like to test the performance of Wasserstein-based k-means clustering for datasets with continuous variability. Wasserstein metrics seem like a natural fit since, by definition, motion of a part of the molecule incurs a distance no greater than the mass of the moving part multiplied by its displacement.

There are many other directions for future work, including incorporating Wasserstein metrics into EM-ML style algorithms (Scheres, 2015), Wasserstein barycenters for estimating the cluster centers (Cuturi and Doucet, 2014), rotationally-invariant denoising procedures (Zhao and Singer, 2014), incorporating translations, experiments on more realistic datasets with a contrast transfer function and analysis of real datasets.

6 Broader impact

Our work demonstrates that the use of Wasserstein metrics can improve the quality of clustering tomographic projections in the context of cryo-electron microscopy. This can reduce the number of images required to produce high quality cluster averages. Improved data efficiency in the cryo-EM pipeline will accelerate discovery in structural biology (with broad implications to medicine) and lower the barriers of entry to cryo-EM for resource-constrained labs around the world.

A Appendix: proofs

A.1 Proof of Proposition 1

Consider a probability measure μ on some space X and let $T : X \to Y$ be a measurable mapping. it naturally induces a measure on its image known as the *push-forward measure* or *image measure* $T\#\mu$, defined by

$$(T\#\mu)(S) = \mu(T^{-1}(S)). \tag{12}$$

Lemma 1. *Given two probability distributions on the unit ball $B \subset \mathbb{R}^3$ μ, η,*

$$W_1(P\#\mu, P\#\eta) \leq W_1(\mu, \eta), \tag{13}$$

where $P\#\mu$ is the tomographic projection in terms of a pushforward measure of μ defined on B onto $\mathbb{D} \subset \mathbb{R}^2$, and the Earthmover's distance for each is defined for the underlying probability space.

Proof. Let Γ be any transportation measure on $B \times B$ such that $\Gamma(A, B) = \mu(A), \Gamma(B, A) = \nu(A)$. Let T map Lesbegue measurable subsets of $B \times B$ to measurable subsets of $\mathbb{D} \times \mathbb{D}$ such that each element $((x_1, y_1, z_1), (x_2, y_2, z_2))$ in a measurable set of $B \times B$ is mapped to $((x_1, y_2), (x_2, y_2))$. Let $J = T\#\Gamma$ where $\#$ denotes the push-forward operator. $J(A, \mathbb{D}) = P\#\mu(A), J(\mathbb{D}, A) = P\#\nu(A)$. By the change of variables formula for push-forward measures:

$$\int_{\mathbb{D}\times\mathbb{D}} ||(x_1, y_1) - (x_2, y_2)|| dJ((x_1, y_1), (x_2, y_2)) \tag{14}$$

$$= \int_{B\times B} ||(x_1, y_1) - (x_2, y_2)|| d\Gamma((x_1, y_1, z_1), (x_2, y_2, z_2)) \tag{15}$$

$$\leq \int_{B\times B} ||(x_1, y_1, z_1) - (x_2, y_2, z_2)|| d\Gamma((x_1, y_1, z_1), (x_2, y_2, z_2)). \tag{16}$$

We conclude that

$$W_1(\mu, \nu) = \inf_{\Gamma} \int_{B\times B} ||u - v|| d\Gamma(u, v) \geq \inf_{T\#\Gamma} \int_{\mathbb{D}\times\mathbb{D}} ||u - v|| d(T\#\Gamma)(u, v) \geq W_1(P\#\mu, P\#\nu), \tag{17}$$

where the last inequality is because any $J = T\#\Gamma$ is a valid transportation measure for $(P\#\mu, P\#\nu)$. □

Lemma 2. *Let I_1, I_2 be tomographic projections of ρ at viewing angles $\mathbf{u}, \mathbf{v}$ respectively. Then*

$$W_1^R(I_1, I_2) \leq 2 \sin(\measuredangle(\mathbf{u}, \mathbf{v})/2) \leq \measuredangle(\mathbf{u}, \mathbf{v}). \tag{18}$$

Proof. The second inequality is immediate, since for all $\theta \geq 0$ we have $\sin(\theta) \leq \theta$. We denote the tomographic projection of ρ at orientation $R \in \mathrm{SO}(3)$ as $\mathcal{T}_R\rho$. Without loss of generality, $I_1 = \mathcal{T}_I\rho$ where I is the identity matrix and so $u = (0, 0, -1)$ and $I_2 = \mathcal{T}_R\rho = \mathcal{T}_I(\rho \circ R^T)$. decompose $R = R_1 \circ R_2$ to an in-plane rotation $R_1 \in \mathrm{SO}(2)$ and R_2 an out-of-plane rotation (has axis in $\mathbb{R}^2$).

We observe the chain of inequalities:

$$W_1^R(I_1, I_2) = W_1^R(I_1, \mathcal{T}_I((\rho \circ R_2^T) \circ R_1^T)) \tag{19}$$

$$= W_1^R(I_1, \mathcal{T}_I(\rho \circ R_2^T)) \tag{20}$$

$$\leq W_1(I_1, \mathcal{T}_I(\rho \circ R_2^T)) \tag{21}$$

$$= W_1(P\#\rho, P\#(\rho \circ R_2^T)) \tag{22}$$

$$\leq W_1(\rho, (\rho \circ R_2^T)) \tag{23}$$

$$\leq 2 \sin(\measuredangle(\mathbf{u}, \mathbf{v})/2). \tag{24}$$

The first equality comes from the decomposition of R, the second comes from the invariance of W_1^R to in-plane rotations, the third inequality comes from the definition of W_1^R, and the fourth equality rewrites $\mathcal{T}_I\rho = P\#\rho$. The fifth inequality comes from Lemma 1. The final step of this chain comes

from the Monge formulation of the Wasserstein metric (Peyré and Cuturi, 2019), from which it follows that for any two probability distributions on B, P, Q we have

$$W_1(P,Q) \leq \int_B \|x - F(x)\| dP(x). \tag{25}$$

For $F : B \to B$, $F\#P = Q$. It is immediate from this statement that using the map $F((x,y,z)) = R_2(x,y,z)$ we obtain that

$$W_1(\rho, \rho \circ R_2^T) \leq \int_B ||(x,y,z) - R_2(x,y,z)|| d\rho(x,y,z). \tag{26}$$

Furthermore, for any vector (x,y,z) with $\|(x,y,z)\| \leq 1$, we have

$$\|(x,y,z) - R_2(x,y,z)\| \leq 2\sin(\theta/2), \tag{27}$$

where θ is the angle of rotation of R_2 in its axis-angle representation. This gives,

$$W_1(\rho, \rho \circ R_2^T) \leq 2\sin\left(\measuredangle(\mathbf{u},\mathbf{v})/2\right). \tag{28}$$

which completes our proof. □

Corollary 1. *Let I_1, I_2 be tomographic projections of ρ at viewing angles $\mathbf{u}, \mathbf{v}$ respectively. Then*

$$W_p^R(I_1, I_2) \leq \left[2\sin\left(\measuredangle(\mathbf{u},\mathbf{v})/2\right)\right]^p \leq \measuredangle(\mathbf{u},\mathbf{v})^p. \tag{29}$$

This comes from the fact that $W_p(I_1, I_2) \leq W_1(I_1, I_2)^p$ (Peyré and Cuturi, 2019) from which we obtain $W_p^R(I_1, I_2) \leq W_1^R(I_1, I_2)^p$ and the subsequent chain of inequalities.

A.2 Proof of Proposition 2

Let ρ be differentiable, and $|\nabla\rho| \leq B$. Then for two tomographic projections of ρ, I_1, I_2 at angles $\mathbf{v}, \mathbf{u}$ we have

$$L_2^R(I_1, I_2) \leq 2\sqrt{\pi}B\measuredangle(\mathbf{u},\mathbf{v}). \tag{30}$$

Proof. This is a consequence of the mean value theorem. Without loss of generality, let us assume that $I_1 = \mathcal{T}_I(\rho)$ where I is the identity matrix and so $v = (0,0,-1)$. Let $I_2 = \mathcal{T}_R(\rho)$. We can decompose R into an in-plane component $R_1 \in \mathrm{SO}(2)$ and an out-of-plane component R_2 such that $R = R_1 \circ R_2$ and R_2 has its axis of rotation in $\mathbb{R}^2$. Because $L_2^R(I_1, I_2)$ is invariant to rotations of I_2, $L_2^R(I_1, I_2) = L_2^R(I_1, \mathcal{T}_R(\rho \circ R_1))$ which means that since $R^T R_1 = R_2^T$,

$$\begin{aligned}
L_2^R(I_1, I_2) &= L_2^R(I_1, \mathcal{T}_I(\rho \circ R_2^T)) && (31)\\
&\leq L_2(I_1, \mathcal{T}_I(\rho \circ R_2^T)) && (32)\\
&= \sqrt{\int_{\mathbb{D}} \left(\int_{-1}^{1} (\rho(x,y,z) - \rho \circ R_2^T(x,y,z)) dz\right)^2} && (33)\\
&\leq \sqrt{\int_{\mathbb{D}} \left(\int_{-1}^{1} (\sup|\nabla\rho|)|\phi| dz\right)^2} && (34)\\
&\leq 4\sqrt{\pi}B|\sin(\phi/2)| \leq 2\sqrt{\pi}B|\phi|. && (35)
\end{aligned}$$

The second to last line holds because by the mean value theorem

$$||\rho(x,y,z) - \rho \circ R_2^T(x,y,z)|| \leq (\sup|\nabla\rho|)||(x,y,z) - R_2^T(x,y,z)||. \tag{36}$$

We observe that the distance the point (x,y,z) travels when acted on by R_2^T is equal to the out-of-plane angle ϕ that R_2^T rotates by multiplied by the distance from (x,y,z) to the axis of R_2^T which is ≤ 1. Thus we can upper bound $||(x,y,z) - R_2^T(x,y,z)|| \leq |2\sin(\phi/2)| \leq |\phi|$. Since $|\phi| = \measuredangle(\mathbf{u},\mathbf{v})$ we achieve the desired upper bound. □

References

Helen M. Berman. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi:10.1093/nar/28.1.235.

James Z. Chen and Nikolaus Grigorieff. SIGNATURE: A single-particle selection system for molecular electron microscopy. *Journal of Structural Biology*, 157(1):168–173, 2007. doi:10.1016/j.jsb.2006.06.001.

Yifan Cheng. Single-particle cryo-EM—How did it get here and where will it go. *Science*, 361 (6405):876–880, 2018. doi:10.1126/science.aat4346.

Daniel Cressey and Ewen Callaway. Cryo-electron microscopy wins chemistry Nobel. *Nature*, 550 (7675):167–167, 2017. doi:10.1038/nature.2017.22738.

Marco Cuturi and Arnaud Doucet. Fast computation of Wasserstein barycenters. *International Conference on Machine Learning (ICML)*, 32(2):685–693, 2014.

Ali Dashti et al. Retrieving functional pathways of biomolecules from single-particle snapshots. *Nature Communications*, 11(1):4734, 2020. doi:10.1038/s41467-020-18403-x.

J.M. de la Rosa-Trevín et al. Xmipp 3.0: An improved software suite for image processing in electron microscopy. *Journal of Structural Biology*, 184(2):321–328, 2013. doi:10.1016/j.jsb.2013.09.015.

Rémi Flamary and Nicolas Courty. POT Python Optimal Transport library, 2017. `http://pythonot.github.io`.

Joachim Frank. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press, 2006. ISBN 9780195182187. doi:10.1093/acprof:oso/9780195182187.001.0001.

Joachim Frank and Terence Wagenknecht. Automatic selection of molecular images from electron micrographs. *Ultramicroscopy*, 12(3):169–175, 1983. doi:10.1016/0304-3991(83)90256-5.

Timothy Grant, Alexis Rohou and Nikolaus Grigorieff. cisTEM, user-friendly software for single-particle image processing. *eLife*, 7(3):377–388, 2018. doi:10.7554/eLife.35383.

Ido Greenberg and Yoel Shkolnisky. Common lines modeling for reference free Ab-initio reconstruction in cryo-EM. *Journal of Structural Biology*, 200(2):106–117, 2017. doi:10.1016/j.jsb.2017.09.007.

Ayelet Heimowitz, Joakim Andén and Amit Singer. APPLE picker: Automatic particle picking, a low-effort cryo-EM framework. *Journal of Structural Biology*, 204(2):215–227, 2018. doi:10.1016/j.jsb.2018.08.012.

Roy R. Lederman, Joakim Andén and Amit Singer. Hyper-molecules: on the representation and recovery of dynamical structures for applications in flexible macro-molecules in cryo-EM. *Inverse Problems*, 36(4):044005, 2020. doi:10.1088/1361-6420/ab5ede.

Gregory Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt and Aaron O'Leary. PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019. doi:10.21105/joss.01237.

Dmitry Lyumkis. Challenges and opportunities in cryo-EM single-particle analysis. *Journal of Biological Chemistry*, 294(13):5181–5197, 2019. doi:10.1074/jbc.REV118.005602.

Dmitry Lyumkis, Axel F. Brilot, Douglas L. Theobald and Nikolaus Grigorieff. Likelihood-based classification of cryo-EM images using FREALIGN. *Journal of Structural Biology*, 183(3): 377–388, 2013. doi:10.1016/j.jsb.2013.07.005.

Stéphane Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Elsevier, third edition, 2009. ISBN 9780123743701. doi:10.1016/B978-0-12-374370-1.X0001-8.

Amit Moscovich, Amit Halevi, Joakim Andén and Amit Singer. Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes. *Inverse Problems*, 36(2):024003, 2020. doi:10.1088/1361-6420/ab4f55.

Takanori Nakane, Dari Kimanius, Erik Lindahl and Sjors HW Scheres. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife*, 7:1–18, 2018. doi:10.7554/eLife.36861.

Pawel Penczek, Michael Radermacher and Joachim Frank. Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*, 40(1):33–53, 1992. doi:10.1016/0304-3991(92)90233-A.

Pawel A. Penczek, Jun Zhu and Joachim Frank. A common-lines based method for determining orientations for N>3 particle projections simultaneously. *Ultramicroscopy*, 63(3-4):205–218, 1996. doi:10.1016/0304-3991(96)00037-X.

Gabriel Peyré and Marco Cuturi. Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. doi:10.1561/2200000073.

Ali Punjani, John L. Rubinstein, David J. Fleet and Marcus A. Brubaker. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3):290–296, 2017. doi:10.1038/nmeth.4169.

Alexis Rohou and Nikolaus Grigorieff. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221, 2015. doi:10.1016/j.jsb.2015.08.008.

Michael Schatz and Marin Van Heel. Invariant classification of molecular views in electron micrographs. *Ultramicroscopy*, 32(3):255–264, 1990. doi:10.1016/0304-3991(90)90003-5.

Sjors H.W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, 2012. doi:10.1016/j.jsb.2012.09.006.

Sjors H.W. Scheres. Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of Structural Biology*, 189(2):114–122, 2015. doi:10.1016/j.jsb.2014.11.010.

Sjors H.W. Scheres, Mikel Valle and J.-M. Carazo. Fast maximum-likelihood refinement of electron microscopy images. *Bioinformatics*, 21(Suppl 2):ii243–ii244, 2005. doi:10.1093/bioinformatics/bti1140.

Sameer Shirdhonkar and David W. Jacobs. Approximate earth mover's distance in linear time. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008. doi:10.1109/CVPR.2008.4587662.

Amit Singer and Fred J. Sigworth. Computational Methods for Single-Particle Electron Cryomicroscopy. *Annual Review of Biomedical Data Science*, 3(1):163–190, 2020. doi:10.1146/annurev-biodatasci-021020-093826.

Carlos Oscar S. Sorzano et al. A clustering approach to multireference alignment of single-particle projections in electron microscopy. *Journal of Structural Biology*, 171(2):197–206, 2010. doi:10.1016/j.jsb.2010.03.011.

Carlos Oscar S. Sorzano et al. Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. *Acta Crystallographica Section F Structural Biology Communications*, 75(1):19–32, 2019. doi:10.1107/S2053230X18015108.

Guang Tang et al. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1):38–46, 2007. doi:10.1016/j.jsb.2006.05.009.

Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 2003. ISBN 9780821833124. doi:10.1090/gsm/058.

Kutti R. Vinothkumar and Richard Henderson. Single particle electron cryomicroscopy: trends, issues and future perspective. *Quarterly Reviews of Biophysics*, 49:1–25, 2016. doi:10.1017/S0033583516000068.

Wilson Wong et al. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife*, 3(3):1–20, 2014. doi:10.7554/eLife.03080.

Nathan Zelesko, Amit Moscovich, Joe Kileel and Amit Singer. Earthmover-Based Manifold Learning for Analyzing Molecular Conformation Spaces. In *International Symposium on Biomedical Imaging (ISBI)*, pages 1715–1719. IEEE, 2020. doi:10.1109/ISBI45749.2020.9098723.

Zhizhen Zhao and Amit Singer. Rotationally invariant image representation for viewing direction classification in cryo-EM. *Journal of Structural Biology*, 186(1):153–166, 2014. doi:10.1016/j.jsb.2014.03.003.

Ellen D Zhong, Tristan Bepler, Joseph H Davis and Bonnie Berger. Reconstructing continuous distributions of 3D protein structure from cryo-EM images. In *International Conference on Learning Representations (ICLR)*, pages 1–20, 2020.

Ye Zhou, Amit Moscovich, Tamir Bendory and Alberto Bartesaghi. Unsupervised particle sorting for high-resolution single-particle cryo-EM. *Inverse Problems*, 36(4), 2020. doi:10.1088/1361-6420/ab5ec8.