

Linear independent component analysis in Wasserstein space

Shiying Li and Caroline Moosmüller and Chuxiangbo Wang

Abstract This paper introduces and analyzes a set-up for linear independent component analysis (ICA) in Wasserstein space. This is motivated by applications in which an instance of data is naturally interpreted as a probability measure or point-cloud, such as gene expression data, and the need to meaningfully analyze this type of data. ICA is a well-developed method for identifying independent components in multivariate data, but mainly focuses on data in Euclidean space. We propose an extension to Wasserstein space by viewing the linear ICA problem in this space as a deviation of the “classical” Euclidean setting. We then show how spectral methods based on the Wasserstein distance can be used to identify independent components in point-cloud data.

1 Introduction

Independent Component Analysis (ICA) is a computational and statistical technique used to uncover independent components from multivariate data, also known as blind source separation [15, 8, 16]. It was first introduced in [1] and has gained much interest since then, see e.g. [14, 2, 10]. Specifically, [7] highlighted the potential of ICA in Mathematics and Statistics. ICA has since become an essential tool in various fields, including signal separation of biological data [9, 25], MRI data [23], audio and image noise reduction [21, 13].

Shiying Li

Department of Mathematics, University of North Carolina at Chapel Hill, e-mail: shiyil@unc.edu

Caroline Moosmüller

Department of Mathematics, University of North Carolina at Chapel Hill, e-mail: cmoosm@unc.edu

Chuxiangbo Wang

Department of Mathematics, University of North Carolina at Chapel Hill, e-mail: chuxianw@unc.edu

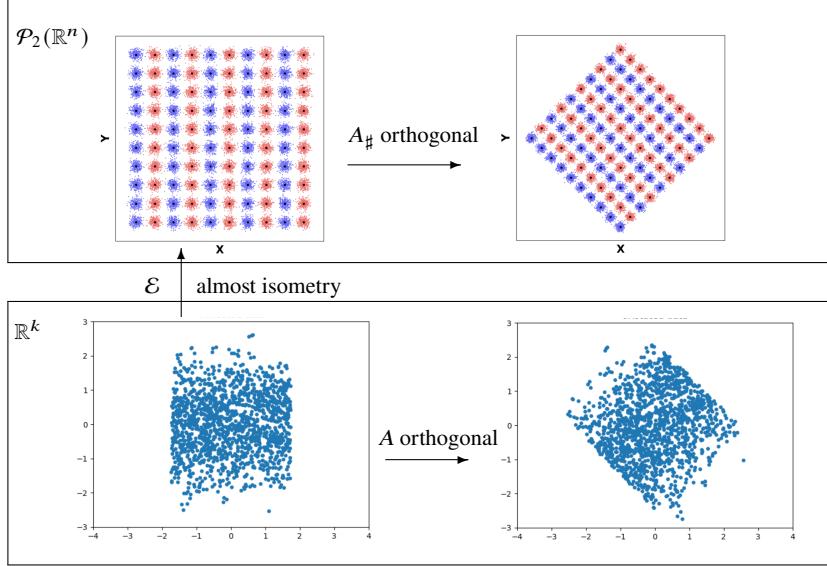


Fig. 1: Set-up for linear ICA in Wasserstein space. *Bottom panel (“classical” linear setting in \mathbb{R}^k):* Independent components are drawn from $\mathcal{U}(-\sqrt{3}, \sqrt{3}) \times \mathcal{N}(0, 1)$ (left plot). An unknown orthogonal transformation A is applied (right plot). The eigenvectors of the Laplacian built from the observed data (right plot) is used to uncover the independent components [28]. *Top panel (proposed setting in $\mathcal{P}_2(\mathbb{R}^n)$):* An almost isometric map $\mathcal{E} : \mathbb{R}^k \rightarrow \mathcal{P}_2(\mathbb{R}^n)$ is assumed; in this example, the map is $\mathcal{E}(\omega) = \mathcal{N}(\omega, cI)$ for $c > 0$ fixed. An unknown orthogonal transformation $A_{\#}$ is applied to obtain the observable data in Wasserstein space (right plot). The eigenvectors of the Laplacian built from the observed data (right plot) with Wasserstein distances is used to uncover the independent components (bottom left), see Section 3. *Note:* The plots in the top panel are sketches for visualization purposes, i.e. we sampled a uniform grid with only a small number of points so that the Gaussians are visible. The blue-red coloring scheme is for visualization purposes only. For the actual numerical experiments, we sampled the means of the Gaussian from the bottom left plot, see Section 4 for details.

The classical linear ICA problem assumes n independent random variables, which have been “mixed” by the application of an orthogonal matrix, and one only has access to N observations of this mixing process. From these observations, the aim is to identify the independent components and the matrix. This problem is usually formulated in Euclidean space, i.e. the independent components and the observations are elements of some \mathbb{R}^k . In this paper we study a version of linear ICA in the Wasserstein space, which is the space of probability measures, see [30]. In particular, we assume that the observed data consists of probability measures or point-clouds,

which have been obtained by a linear mixing through Euclidean independent components. This set-up is motivated by applications in which an instance of data is not naturally interpreted as a vector in some \mathbb{R}^k , but rather as a probability measure or point-cloud. Examples include imaging data [26], text documents [33], gene expression data [19, 5], and flow cytometry [34, 3].

While there exists a large body of literature on linear ICA (in the Euclidean setting), we follow the ideas of [28], which uses the eigenvectors of a graph Laplacian built from the observed data to identify the independent components and the mixing matrix. This method naturally adapts to our setting, as we only need to reinterpret the graph Laplacian for point-cloud data. Essentially, we replace the Euclidean distance by the Wasserstein distance when building the graph Laplacian, which has shown success in other methods as well [19, 5, 18, 32, 20].

The contributions of this paper are twofold. We first describe a natural setting for linear ICA in Wasserstein space, where the observed data consists of probability measures. This idea mimics the classical linear ICA in Euclidean space, and is outlined in Figure 1. We then show that our method is successful in identifying the independent components as long as the observed point-cloud data is “close to” (almost isometric to) Euclidean data by using results on eigenvector perturbations. We present toy examples where the observed data are rotated Gaussians, and the independent components are their means.

The paper is organized as follows. Section 2 presents the preliminaries on spectral linear ICA as introduced by [28] and gives a basic introduction to optimal transport and the Wasserstein distance. In Section 3 we show a natural setting for linear ICA in the Wasserstein space and provide the main result on recovery of the independent components in the almost isometric setting. Section 4 contains numerical toy examples to showcase our proposed method.

2 Preliminaries

2.1 Linear ICA via the graph Laplacian

For the linear ICA problem, we follow the set-up and results from [28]. Here we briefly summarize the main results needed.

The linear ICA problem is formulated as follows. Let $S = (S_1, S_2, \dots, S_n)$ be n unknown independent components (random variables) with zero mean and unit variance. Let $A \in \mathbb{R}^{n \times n}$ be an unknown orthogonal mixing matrix. Consider observations of these random variables, denoted as $\bar{S}_i \in \mathbb{R}^N$. The observed data under the mixing matrix A is given by

$$X = A\bar{S}^T. \quad (1)$$

To recover the independent components S from X , [28] interprets the observed data points x_1, \dots, x_N (column vectors of X) as the nodes of a graph. The weights

of this graph are defined by

$$W_{i,j} = e^{\frac{-\|x_i - x_j\|^2}{2h}}, \quad (2)$$

where $\|\cdot\|$ denotes the Euclidean distance between x_i and x_j , and h is the width parameter of the kernel. From this weight matrix, the normalized graph Laplacian,

$$L = I - D^{-1}W, \quad (3)$$

is constructed, where D is the diagonal degree matrix defined by $D = \text{diag}\left(\sum_{j=1}^N W_{i,j}\right)$.

It is proved in [28] that the eigenvectors of the graph Laplacian approximate the independent components S_i . The main argument concerns the convergence of the graph Laplacian L to the backward Fokker-Planck operator as the number of samples $N \rightarrow \infty$, and the fact that the Fokker-Planck operator separates into n one-dimensional operators when S_i are independent, see [28].

2.2 Optimal transport and Wasserstein space

In this manuscript, we focus on recovering the underlying independent components when probability measures or point-clouds undergo a linear mixing transformation. The optimal transport (OT) theory [22, 17] provides a natural framework for comparing probability measures. We introduce the necessary background here and refer the readers to [30, 27] for a thorough treatment of the subject. For an overview of the computational aspects of OT, see [24].

Let $\mathcal{P}(\mathbb{R}^n)$ be a set of Borel probability measures on \mathbb{R}^n . Consider the space of probability measures with bounded second moments, denoted by $\mathcal{P}_2(\mathbb{R}^n)$ where

$$\mathcal{P}_2(\mathbb{R}^n) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^n) : \int_{\mathbb{R}^n} |x|^2 d\mu(x) < +\infty \right\}. \quad (4)$$

For two probability measures $\alpha, \beta \in \mathcal{P}_2(\mathbb{R}^n)$, the 2-Wasserstein distance is defined as:

$$W_2(\alpha, \beta) = \min_{\pi \in \Pi(\alpha, \beta)} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\pi(x, y) \right)^{\frac{1}{2}}, \quad (5)$$

where $\Pi(\alpha, \beta)$ denotes the set of transport plans (couplings) between α and β , i.e., $\pi \in \Pi(\alpha, \beta)$ is a probability measure on $\mathbb{R}^n \times \mathbb{R}^n$ with first marginal α and second marginal β . Here we refer to the metric space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ as the Wasserstein space. Note that we focus on the 2-Wasserstein distance and space since this will be convenient for our setting (the 2-Wasserstein distance is invariant under orthogonal transformations, see Lemma 1).

In the case when α is absolutely continuous, the minimizer π^* of (5) is unique and of the form $(\text{id}, T^*)_\# \alpha$, where T^* is called the optimal transport map between α and β (see e.g., [29, Theorem 2.12]). Here $\#$ denotes push-forward operation between

probability measures. Specifically, given $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g_{\sharp}\alpha$, often referred to as the push-forward measure of α by g , is a measure in $\mathcal{P}(\mathbb{R}^m)$ defined via

$$g_{\sharp}\alpha(B) := \alpha(g^{-1}(B)), \quad \forall \text{ Borel sets } B \subseteq \mathbb{R}^m. \quad (6)$$

3 Linear ICA in Wasserstein space

We now describe a natural set-up for linear ICA in the Wasserstein space. This is similar to ideas related to manifold learning in the Wasserstein space, see e.g. [12, 6, 11]. As introduced in Section 2.1, the linear ICA problem for Euclidean data can be solved by analyzing the spectral properties of the normalized graph Laplacian L (eq. (3)). In the context of Wasserstein space, this process involves analyzing a Wasserstein-based graph Laplacian by leveraging the optimal transport (OT) framework, and more specifically, using the 2-Wasserstein distance to compare probability measures.

Consider $\mathcal{E} : \Omega \rightarrow \mathcal{P}_2(\mathbb{R}^n)$ where $\Omega \subseteq \mathbb{R}^n$ represents a set of governing parameters and the map \mathcal{E} describes a nonlinear process of generating probability measures from the parameters. Assume $\{\omega_j\}_{j=1}^N$ are the underlying parameters sampled from n independent components $S = (S_1, \dots, S_n)$ from Ω and consider observed data of the form $\beta_j := A_{\sharp}\mathcal{E}(\omega_j)$, $j = 1, \dots, N$. The probability measures β_j are obtained via the push-forward of $\mathcal{E}(\omega_j)$ by a mixing orthogonal transformation A . As in the linear setting (Section 2.1), the task is to uncover the independent components S_1, \dots, S_n from the observed data $\{\beta_j\}_{j=1}^N$. See Figure 1 for an overview of this construction.

Example 1 To illustrate the Wasserstein ICA set-up we give a basic example, which is discussed in more detail in Section 4 and is visualized in Figure 1. Denote by $\mathcal{N}(m, \Sigma)$ the Gaussian in \mathbb{R}^n with mean m and covariance Σ . A possible map \mathcal{E} is $\omega \mapsto \mathcal{N}(\omega, cI)$, where $c > 0$ fixed. This is a simple way of generating probability measures from parameters. The observed data would then be $A_{\sharp}\mathcal{N}(\omega, cI)$, i.e. each of the Gaussians is pushed with an orthogonal transformation.

Remark 1 We assume that $\Omega \subseteq \mathbb{R}^n$ and that support space for the probability measures is \mathbb{R}^n . It is not necessary for those spaces to have the same dimension n ; this is mostly for convenience of presentation. The results that follow still hold if the dimensions differ.

To recover the independent components S_1, \dots, S_n , the idea is to utilize the graph Laplacian with Wasserstein distances between the observed probability measures $\{\beta_j\}_{j=1}^N$ rather than the Euclidean distance used in (3). This is natural since we are dealing with objects in the metric space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$. In particular, we construct the normalized graph Laplacian

$$L_{W_2} = I - \tilde{D}^{-1}\tilde{W}, \quad (7)$$

where

$$\tilde{W}_{ij} = e^{-\frac{W_2(\beta_i, \beta_j)^2}{2h}}, \quad (8)$$

and \tilde{D} is the degree matrix associated with \tilde{W} .

Our goal is to understand to which extent the eigenvectors of L_{W_2} approximate the independent components under the assumption that the parameter space $(\Omega, \|\cdot\|)$ is “almost” isometric to $(\mathcal{E}(\Omega), W_2)$. The recovery result follows from combining eigenvector perturbation results with results from [28]. The main theorem of this paper concerns the eigenvector perturbation under an almost isometry.

Theorem 1 *Let S_1, \dots, S_n be (real-valued) independent random variables and $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be an orthogonal mixing matrix. Assume that $\Omega \subseteq \mathbb{R}^n$ such that $S \in \Omega$, ¹ where $S = (S_1, \dots, S_n)$. Let $\mathcal{E} : \Omega \rightarrow \mathcal{P}_2(\mathbb{R}^n)$. Assume that there exists $\eta \geq 0$ such that*

$$\left| W_2^2(\mathcal{E}(\omega), \mathcal{E}(\kappa)) - \|\omega - \kappa\|^2 \right| \leq \eta, \quad \forall \omega, \kappa \in \Omega. \quad (9)$$

Let $\{\omega_j\}_{j=1}^N$ be N instances of S and $\beta_j := A_{\#}\mathcal{E}(\omega_j)$. Let L and L_{W_2} be the normalized graph Laplacian associated with $\{A\omega_j\}_{j=1}^N$ (see (3)) and associated with $\{\beta_j\}_{j=1}^N$ (see (7)), respectively. Let $\lambda_1 \leq \dots \leq \lambda_N$ and $\tilde{\lambda}_1 \leq \dots \leq \tilde{\lambda}_N$ be the eigenvalues of L and L_{W_2} , respectively. Fix $1 \leq j \leq N$, assume that $\delta_j := \min\{\lambda_j - \lambda_{j-1}, \lambda_{j+1} - \lambda_j\} > 0$. Then for the eigenvectors ϕ_j and $\tilde{\phi}_j$ satisfying $L\phi_j = \lambda_j\phi_j$ and $L_{W_2}\tilde{\phi}_j = \tilde{\lambda}_j\tilde{\phi}_j$, the following holds

$$\cos \angle(\phi_j, \tilde{\phi}_j) \geq 1 - \frac{\varepsilon_{\max}^{1/2}}{2} r_D \left(b + 2^{3/2} \delta_j^{-1} \left(a \varepsilon_{\min}^{-1} r_D + b r_D \varepsilon_{\min}^{-1/2} + b r_D^{1/2} \right) \right)^2. \quad (10)$$

Here $\varepsilon_{\min} = e^{-\frac{\eta}{2h}}$ and $\varepsilon_{\max} = e^{\frac{\eta}{2h}}$, $a = \max\{|\varepsilon_{\max} - 1|, |\varepsilon_{\min} - 1|\}$, $b = \max\{|\varepsilon_{\max}^{-1/2} - 1|, |\varepsilon_{\min}^{-1/2} - 1|\}$, and $r_D = \frac{D_{\max}}{D_{\min}}$ with $D_{\max} = \max_i D_{ii}$, $D_{\min} = \min_i D_{ii}$. Here D is the degree matrix associated with $\{\omega_j\}_{j=1}^N$.

Proof We start by comparing the distances used in the kernel in the Wasserstein and Euclidean settings (see (8) and (2)). The former uses the Wasserstein distance of the observed measures, i.e., $W_2^2(\beta_i, \beta_j)$, while the latter uses the Euclidean distances between the mixed parameters, i.e., $\|A\omega_i - A\omega_j\|^2$. Since A is an orthogonal matrix, by Lemma 1 we have

$$W_2(\beta_i, \beta_j) = W_2(A_{\#}\mathcal{E}(\omega_i), A_{\#}\mathcal{E}(\omega_j)) = W_2(\mathcal{E}(\omega_i), \mathcal{E}(\omega_j)).$$

Since $\|A\omega_i - A\omega_j\| = \|\omega_i - \omega_j\|$, it follows from Equation (9) that

$$\left| W_2^2(\beta_i, \beta_j) - \|A\omega_i - A\omega_j\|^2 \right| \leq \eta, \quad \forall i, j = 1, \dots, N. \quad (11)$$

¹ Here we abuse notation and do not differentiate the measurable function S from its function value. $S \in \Omega$ means that the function values of S are in the set Ω .

Denoting by \tilde{W}, W the weight matrices associated with L_{W_2} and L , respectively, we have that

$$e^{-\eta/2h} \leq \frac{\tilde{W}_{ij}}{W_{ij}} \leq e^{\eta/2h}. \quad (12)$$

The relationship between the corresponding eigenvectors of L_{W_2} and L then follows from an eigenvector perturbation result, which we summarize in the Appendix, see Proposition 1. \square

Remark 2 When \mathcal{E} defines an isometry, i.e., $\eta = 0$ in (9), we have that $\varepsilon_{\min} = \varepsilon_{\max} = 1$, which implies that $a = b = 0$ and hence $\angle(\phi_j, \tilde{\phi}_j) = 0$. Therefore, as expected in the isometric case, there is no difference between the Wasserstein and the Euclidean setting, hence the angle between the eigenvectors is 0 (see Corollary 1 below).

Remark 3 In general, the lower bound of $\cos \angle(\phi_j, \tilde{\phi}_j)$ given by the RHS of eq. (10) depends on an interplay between the constants $\varepsilon_{\min}, \varepsilon_{\max}, \delta_j$, and r_D . In particular, when \mathcal{E} is an almost-isometry, i.e., the perturbation η between the distances is “small” such that $\varepsilon_{\min}, \varepsilon_{\max} \approx 1$, one can expect that $\cos \angle(\phi_j, \tilde{\phi}_j) \approx 1$ (or equivalently, $\angle(\phi_j, \tilde{\phi}_j) \approx 0$), as long as δ_j is reasonably large and r_D is reasonably small. However, the numerical experiments seem to be more robust than what this lower bound can predict. Remark 6 shows that the independent components can be recovered, even when the RHS in eq. (10) is below -1 , in which case this bound is not useful in predicting the recovery performance. We leave the improvement of this lower bound for future work.

Remark 4 When $0 \leq W_2^2(\mathcal{E}(\omega), \mathcal{E}(\kappa)) - \|\omega - \kappa\|^2 \leq \eta$, we have that the constant $\varepsilon_{\max} \leq 1$, and hence $a \leq 1$.

Corollary 1 Let $S, A, \Omega, \mathcal{E}$, and $\{\omega_j\}_{j=1}^N$ be as defined in Theorem 1. Assume that

$$W_2(\mathcal{E}(\omega), \mathcal{E}(\kappa)) = \|\omega - \kappa\|, \quad \forall \omega, \kappa \in \Omega. \quad (13)$$

Then $L_{W_2} = L$ in Theorem 1.

Proof In this case $\tilde{W}_{ij} = W_{ij}$, which implies $L_{W_2} = L$. \square

Remark 5 (Recovery of independent components)

From [28, Section 4.1] we know that in the Euclidean setting, the eigenvectors of the graph Laplacian L approximate the independent components S_1, \dots, S_n up to errors coming from the sampling process (in the limit $N \rightarrow \infty$). Our Theorem 1 now states that in the almost-isometric setting, the eigenvectors of the Wasserstein-based Laplacian L_{W_2} approximate the eigenvectors of L up to the error (10). Putting this together, in the almost-isometric setting, the Wasserstein ICA recovers the independent components S_1, \dots, S_n up to these two approximation errors combined.

4 Examples and Numerical Experiments

In the following two numerical experiments, we use point-clouds drawn from Gaussian distributions that use independent sources as means, with fixed (isometric case, see Example 3) and varying (almost isometric case, see Example 4) covariance matrices (Figure 2a, Figure 3a). These Gaussians then undergo an unknown orthogonal transformation (Figure 2b, Figure 3b). The Wasserstein-based ICA method is then applied to the observed “linearly mixed” point-clouds, and the recovery of the independent components (the means) is presented.

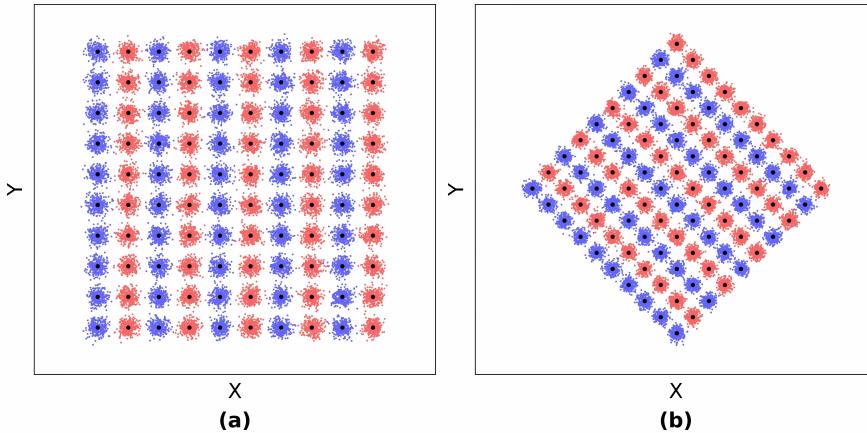


Fig. 2: An illustration of the isometric case (Section 4.1). This is a sketch for illustration purposes, the actual numerical set-up is described in Section 4.1. (a) Independent components (the Gaussian means) are sampled on a square and Gaussians with these means and the same covariance (are multiple of I) are considered (b) Gaussians from (a) are transformed with an orthogonal matrix A .

Figures 2 and 3 are illustrations of the two settings we consider (isometric and almost-isometric). We note that as with Figure 1 these are sketches for visualization purposes and do not represent the actual data used to carry out the numerical experiments. The reason for using sketches only is to make sure individual Gaussians are visible (when means are densely and non-uniformly sampled, different Gaussians easily overlap making it hard to identify individual instances of data).

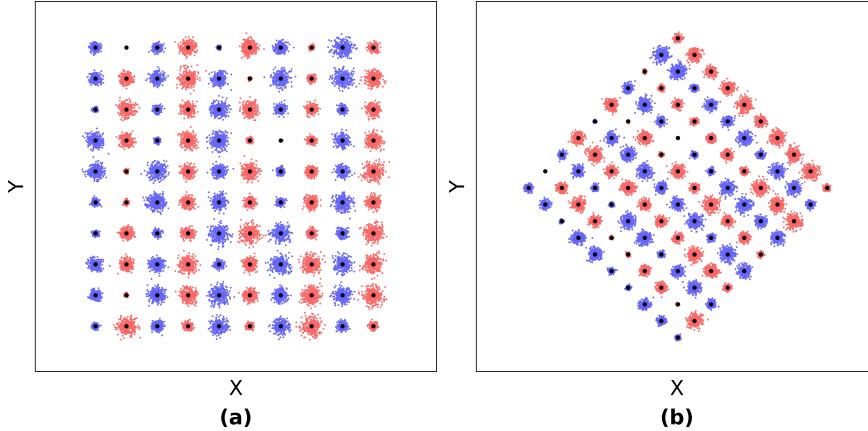


Fig. 3: An illustration of the almost-isometric case (Section 4.2). This is a sketch for illustration purposes, the actual numerical set-up is described in Section 4.2. (a) Independent components (the Gaussian means) are sampled on a square and Gaussians with these means and the covariances of varying sizes are considered (b) Gaussians from (a) are transformed with an orthogonal matrix A .

4.1 Isometric case

We first look at the case when \mathcal{E} defines an isometry from the parameter space $(\Omega, \|\cdot\|)$ to the space of probability measures $(\mathcal{P}_2(\mathbb{R}^n), W_2)$, i.e., when $\eta = 0$ in Theorem 1. In this case, the Wasserstein ICA problem reduces to the Euclidean linear ICA problem, as the graph Laplacian using the Wasserstein distances of observed measures coincides with the graph Laplacian in the parameter space. The same approximation and recovery results for the independent components hence follow from [28], see Corollary 1 and Remark 5.

One way of generating measures for which isometry (e.g. (13)) holds can be obtained by the translation of a base measure.

Example 2 (Isometric \mathcal{E})

Let $\alpha_0 \in \mathcal{P}_2(\mathbb{R}^n)$ and $\Omega \subseteq \mathbb{R}^n$. Define $\mathcal{E} : \Omega \rightarrow \mathcal{P}_2(\mathbb{R}^n)$ by $\mathcal{E}(\omega) = T_{\omega\sharp}\alpha_0$, where $T_\omega(x) = x - \omega$ is a translation. It follows that \mathcal{E} is an isometry since $W_2(\mathcal{E}(\omega), \mathcal{E}(\kappa)) = W_2(T_{\omega\sharp}\alpha_0, T_{\kappa\sharp}\alpha_0) = \|\omega - \kappa\|$.

We now consider a related example which is built from an example in [28].

Example 3 Consider $n = 2$, and generate parameters in Ω by $S = (S_1, S_2)$ with the independent components S_1, S_2 given by

$$S_1 \sim \mathcal{U}(-\sqrt{3}, \sqrt{3}), \quad S_2 \sim \mathcal{N}(0, 1), \quad (14)$$

where $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ denotes uniform distribution on $[-\sqrt{3}, \sqrt{3}]$ and $\mathcal{N}(0, 1)$ is the standard normal distribution ².

Let $\{\omega_j\}_{j=1}^N$ be N instances of S and generate point-clouds $\widehat{\beta}_j$ sampled from $A_{\sharp}\mathcal{N}(\omega_j, cI)$, $j = 1, \dots, N$. Here we choose $c = 0.003$ and

$$A = \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix}, \quad (15)$$

which describes the orthogonal mixing matrix. The number of point-clouds is $N = 600$, each of which contains 30 points.

The eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ of the normalized graph Laplacian L_{W_2} are computed from $\{\widehat{\beta}_j\}_{j=1}^N$ using the kernel (2) based on the W_2 -distance, i.e., $\widetilde{W}_{ij} = e^{\frac{-W_2(\widehat{\beta}_i, \widehat{\beta}_j)^2}{2h}}$, where h is set to 0.2. Here \widetilde{D} is the degree matrix associated with the weight matrix \widetilde{W} . We use the eigenvectors $\widetilde{\phi}_2$ and $\widetilde{\phi}_3$, which correspond to the first two non-trivial eigenvalues to recover the independent components. In Figure 4c and d we plot $\widetilde{\phi}_2$ and $\widetilde{\phi}_3$ and color them by the original independent components S_1 (Figure 4c) and S_2 (Figure 4d), respectively. We observe that the independent components are recovered since the eigenvectors are in one-to-one correspondence with the independent components (Figure 4a, Figure 4b), as is expected from Corollary 1.

² In all the numerical experiments performed, we have applied a filter on samples from S to remove the isolated outliers, similar to [28].

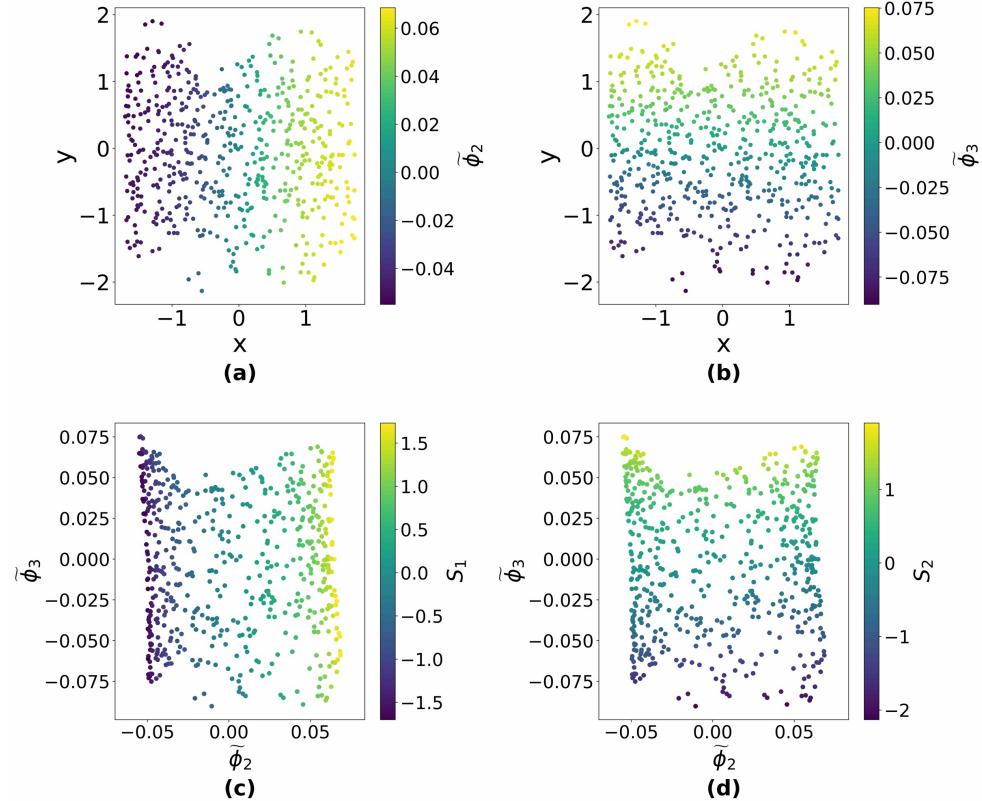


Fig. 4: Illustration of the one-to-one correspondence between the eigenvectors of the graph Laplacian and the independent components described in Example 3 (isometric example). (a) independent components $S = (S_1, S_2)$ colored by the first non-trivial eigenvector $\tilde{\phi}_2$, (b) independent components $S = (S_1, S_2)$ colored by the second non-trivial eigenvector $\tilde{\phi}_3$, (c) eigenvectors $(\tilde{\phi}_2, \tilde{\phi}_3)$ colored by the first independent component S_1 , and (d) eigenvectors $(\tilde{\phi}_2, \tilde{\phi}_3)$ colored by the second independent components S_2 .

4.2 Almost isometric case

A more interesting case is when η in Theorem 1 is small, i.e., the almost isometric case. One way of generating measures such that (9) holds is by varying an isotropic Gaussian by its mean and variance.

Example 4 Let $\Omega \subseteq \mathbb{R}^n$. Let $c : \Omega \rightarrow [c_1, c_2]$ where $0 < c_1 < c_2$. Define $\mathcal{E} : \Omega \rightarrow \mathcal{P}_2(\mathbb{R}^n)$ by $\mathcal{E}(\omega) = \mathcal{N}(\omega, c(\omega)I)$. Using the Wasserstein distance formula for Gaussians (see Lemma 3), we have

$$W_2^2(\mathcal{E}(\omega), \mathcal{E}(\kappa)) - \|\omega - \kappa\|^2 = n \left(\sqrt{c(\omega)} - \sqrt{c(\kappa)} \right)^2 \quad (16)$$

$$\leq n(\sqrt{c_2} - \sqrt{c_1})^2 \quad (17)$$

$$\leq \min\{n(c_2 - c_1), \frac{n(c_2 - c_1)^2}{4c_1}\}, \quad (18)$$

where the last inequality follows from the simple facts that $\sqrt{c_2} - \sqrt{c_1} \leq \sqrt{c_2 - c_1}$ and $\sqrt{c_2} - \sqrt{c_1} \leq \frac{c_2 - c_1}{2\sqrt{c_1}}$. Similar to (12), we get

$$e^{-\eta/2h} \leq \frac{\tilde{W}_{ij}}{W_{ij}} \leq 1, \quad (19)$$

where $\eta = \min\{n(c_2 - c_1), \frac{n(c_2 - c_1)^2}{4c_1}\}$ can be made small by choosing $c_2 - c_1$ small. By Proposition 1, the relationship between the eigenvectors of L_{W_2} and L is given by (10) with constants $a = |1 - e^{-\eta/2h}| \leq 1$ and $b = |1 - e^{\eta/4h}|$.

In our numerical experiments, we again choose empirical measures corresponding to point-clouds $\hat{\beta}_j$ sampled from the observed Gaussians $A_{\#}\mathcal{N}(\omega_j, c(\omega_j)I)$, where ω_j are instances of parameters sampled from the independent component vector $S = (S_1, S_2, \dots, S_n)$.³

We follow a similar numerical set-up as in Example 3. The dimension is $n = 2$, the parameters $\{\omega_j\}_{j=1}^N$ are generated from $S = (S_1, S_2)$ specified in (14), $N = 600$, A is in (15), and $h = 0.2$. Each point cloud $\hat{\beta}_j$ contains 30 points sampled from $A_{\#}\mathcal{N}(\omega_j, c(\omega_j)I)$, where $c(\omega_j)$ is chosen uniformly in $[0.998, 1.002]$.

As in Example 3, the eigenvectors $\tilde{\phi}_2$ and $\tilde{\phi}_3$ of L_{W_2} corresponding to the first two non-trivial eigenvalues are used for the independent components recovery.

In Figure 5, we plot $\tilde{\phi}_2$ and $\tilde{\phi}_3$ colored by the original independent components S_1 (Figure 5c) and S_2 (Figure 5d), respectively. In Figure 5a and Figure 5b, the original parameters $S = (S_1, S_2)$ are colored by $\tilde{\phi}_2$ (Figure 5a) and $\tilde{\phi}_3$ (Figure 5b), respectively. We observe that the eigenvectors are in one-to-one correspondence with the independent components, as expected from Remark 5.

To obtain the theoretical bound in (10) from Theorem 1, we estimate ε_{\min} , ε_{\max} by (16) and hence use the min and max of $\left(\sqrt{c(\omega_i)} - \sqrt{c(\omega_j)} \right)^2$. We observe that $\varepsilon_{\min}, \varepsilon_{\max} \approx 1$ for the chosen parameter interval. The remaining constants $r_D, \delta_j, j = 2, 3$ are computed directly using W and L (see (3)) associated with $\{A\omega_j\}_{j=1}^N$. Taking the average of multiple numerical outputs, we obtain

$$\cos \angle(\phi_2, \tilde{\phi}_2) \geq 0.993, \quad \cos \angle(\phi_3, \tilde{\phi}_3) \geq 0.989, \quad (20)$$

with standard deviation 0.006 and 0.012. The two angles are around 6.8° and 8.5° , respectively. Based on the chosen example, small angles were expected, see Remark 3.

³ Note here $\mathcal{E}(\omega_j)$ is the empirical measure of a point-cloud sampled from $\mathcal{N}(\omega_j, c(\omega_j)I)$.

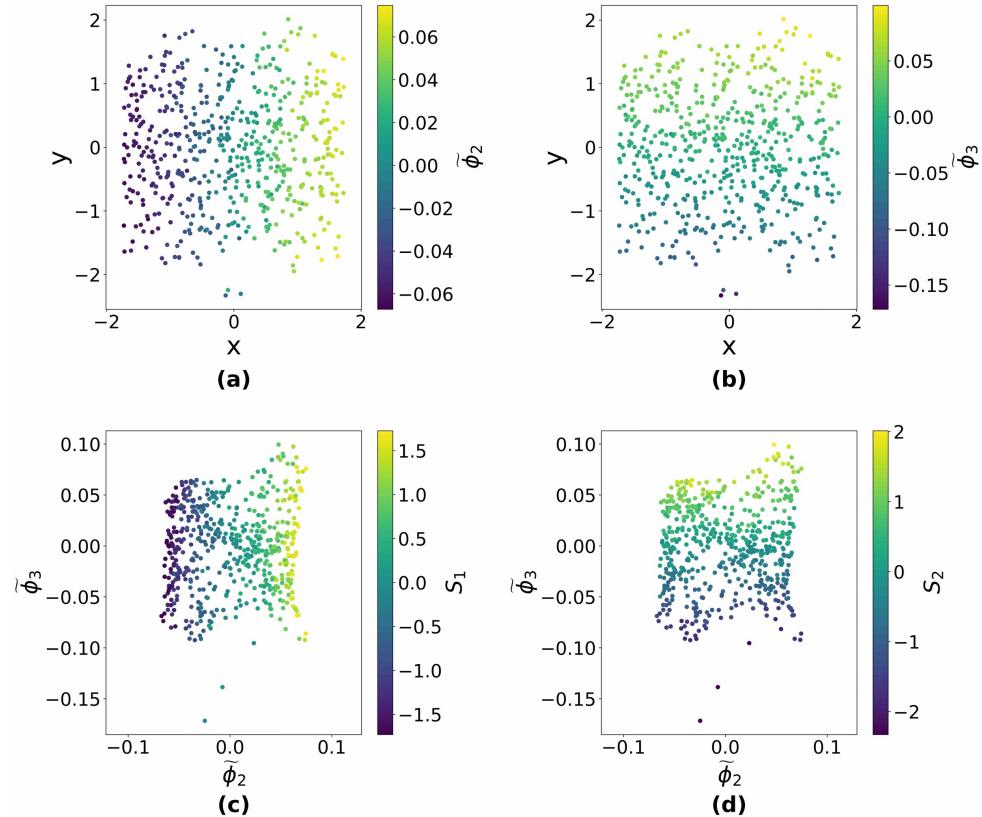


Fig. 5: Illustration of the one-to-one correspondence between the eigenvectors of the graph Laplacian and the independent components described in Example 4 (almost-isometric example). (a) independent components $S = (S_1, S_2)$ colored by the first non-trivial eigenvector $\tilde{\phi}_2$, (b) independent components $S = (S_1, S_2)$ colored by the second non-trivial eigenvector $\tilde{\phi}_3$, (c) eigenvectors $(\tilde{\phi}_2, \tilde{\phi}_3)$ colored by the first independent component S_1 , and (d) eigenvectors $(\tilde{\phi}_2, \tilde{\phi}_3)$ colored by the second independent components S_2 .

The preceding example shows that when the covariance of the Gaussians varies by small constants, i.e. when $\frac{\max c(\omega)}{\min c(\omega)} \approx 1$, then the independent components S_1 and S_2 are well approximated by the eigenvectors of the graph Laplacian and the error established in Theorem 1, (10), can be explicitly computed, compare (20). Even when the error bound (10) is not meaningful (for example, when the lower bound is negative), the Wasserstein ICA method may still be successful in recovering the independent components. We now discuss one such case.

Remark 6 Following the exact same set up as Example 4, we choose $c(\omega_j)$ uniformly from $[0.00003, 0.3]$ such that $\frac{\max c(\omega)}{\min c(\omega)} \approx 10^4$, which indicates a significant size

difference in $\{\widehat{\beta}_j\}_{j=1}^N$. The error established in (10) is computed but exceeds the range of cosine function due to small ε_{\min} defined in Theorem 1, and is thus not insightful. However, the first two non-trivial eigenvectors $\widehat{\phi}_2$ and $\widehat{\phi}_3$ computed from L_{W_2} (7) are nevertheless in one-to-one correspondence with the independent components S_1 and S_2 , as illustrated by Figure 6 in the Appendix.

5 Discussion

We have presented a framework for applying linear independent component analysis when the observed data consists of probability measures or point-clouds. Our method mimics the classical Euclidean setting and shows that when the observed point-cloud data is almost isometric to Euclidean data, comparable recovery results can be achieved. We consider this paper a first step towards the development of a complete theory for ICA in the Wasserstein space. Topics of future interest concern going beyond the almost-isometry assumption, and studying non-linear ICA problems.

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

Appendix

Proposition 1 Let W, \tilde{W} be $N \times N$ weight matrices built from (2) and (8). Let $L = I - D^{-1}W$ and $\tilde{L} = I - \tilde{D}^{-1}\tilde{W}$ be the corresponding normalized graph Laplacians, with D and \tilde{D} being the associated degree matrices, respectively. Assume that $\tilde{W}_{ij} = \varepsilon_{ij} W_{ij}$ such that

$$0 < \varepsilon_{\min} \leq \varepsilon_{ij} \leq \varepsilon_{\max}, \quad i, j = 1, \dots, N. \quad (21)$$

Suppose $\lambda_1 \leq \dots \leq \lambda_N$ and $\tilde{\lambda}_1 \leq \dots \leq \tilde{\lambda}_N$ are the eigenvalues of L and \tilde{L} , respectively. Fix $j \in \{1, \dots, N\}$, and assume that $\delta_j := \min\{\lambda_j - \lambda_{j-1}, \lambda_{j+1} - \lambda_j\} > 0$. Then for eigenvectors ϕ_j and $\tilde{\phi}_j$ satisfying $L\phi_j = \lambda_j \phi_j$ and $\tilde{L}\tilde{\phi}_j = \tilde{\lambda}_j \tilde{\phi}_j$,

$$\cos \angle(\phi_j, \tilde{\phi}_j) \geq 1 - \frac{\varepsilon_{\max}^{1/2}}{2} r_D \left(b + 2^{3/2} \delta_j^{-1} \left(a \varepsilon_{\min}^{-1} r_D + b r_D \varepsilon_{\min}^{-1/2} + b r_D^{1/2} \right) \right)^2. \quad (22)$$

where $a = \max\{|\varepsilon_{\max} - 1|, |\varepsilon_{\min} - 1|\}$, $b = \max\{|\varepsilon_{\max}^{-1/2} - 1|, |\varepsilon_{\min}^{-1/2} - 1|\}$, and $r_D = \frac{D_{\max}}{D_{\min}}$ with $D_{\max} = \max_i D_{ii}$, $D_{\min} = \min_i D_{ii}$. Here D is the degree matrix associated with W .

Proof Let $M = D^{-1}W$ and $\tilde{M} = \tilde{D}^{-1}\tilde{W}$. Since L and M (similarly, \tilde{L} and \tilde{M}) have the same eigenvectors, it is equivalent to analyze the eigenvectors for M and \tilde{M} . We first look at eigenvectors for the symmetric matrices $S = D^{1/2}MD^{-1/2}$ and $\tilde{S} = \tilde{D}^{1/2}\tilde{M}\tilde{D}^{-1/2}$. It is not hard to verify that, if V is an orthogonal matrix whose columns are eigenvectors of S , then the columns of $D^{-1/2}V$ are eigenvectors of M corresponding to the same eigenvalue. Without loss of generality, assume that $\phi_j = D^{-1/2}v_j$ and $\tilde{\phi}_j = \tilde{D}^{-1/2}\tilde{v}_j$, where v_j and \tilde{v}_j are unit eigenvectors of S and \tilde{S} , corresponding to eigenvalues λ_j and $\tilde{\lambda}_j$, respectively. We will first bound $\|\tilde{v}_j - v_j\|$ using Lemma 2. Observe that $S = D^{-1/2}WD^{-1/2}$ and $\tilde{S} = \tilde{D}^{-1/2}\tilde{W}\tilde{D}^{-1/2}$. By a direct computation, we have the following bounds,

$$\|D^{-1/2}\| \leq D_{\min}^{-1/2}, \quad (23)$$

$$\|\tilde{D}^{-1/2}\| \leq \varepsilon_{\min}^{-1/2} D_{\min}^{-1/2}, \quad (24)$$

$$\|W\| \leq \|D\| \|D^{-1}W\| \leq D_{\max}, \quad (25)$$

$$\|\tilde{W} - W\| \leq \sqrt{\|\tilde{W} - W\|_1 \|\tilde{W} - W\|_\infty} = \|\tilde{W} - W\|_1 \leq aD_{\max}, \quad (26)$$

where $a = \max\{|\varepsilon_{\max} - 1|, |\varepsilon_{\min} - 1|\}$. Here $\|\cdot\|$ denotes the matrix 2-norm, and we have used the fact that $\|D^{-1}W\| \leq 1$ in (24) (since DW^{-1} is non-negative and row stochastic), and the fact that $\tilde{W} - W$ is symmetric in (26). Similarly, since

$$(\varepsilon_{\max}^{-1/2} - 1)(D_{ii})^{-1/2} \leq (\tilde{D}^{-1/2})_{ii} - (D^{-1/2})_{ii} \leq (\varepsilon_{\min}^{-1/2} - 1)(D_{ii})^{-1/2},$$

we obtain

$$\|\tilde{D}^{-1/2} - D^{-1/2}\| \leq bD_{\min}^{-1/2}, \quad (27)$$

where $b = \max\{|\varepsilon_{\max}^{-1/2} - 1|, |\varepsilon_{\min}^{-1/2} - 1|\}$. By the triangle inequality

$$\begin{aligned} \|\tilde{S} - S\| &\leq \|\tilde{D}^{-1/2}\tilde{W}\tilde{D}^{-1/2} - D^{-1/2}W\tilde{D}^{-1/2}\| \\ &\quad + \|D^{-1/2}W\tilde{D}^{-1/2} - D^{-1/2}WD^{-1/2}\|. \end{aligned} \quad (28)$$

For the first term in (28), we have

$$\begin{aligned} \|\tilde{D}^{-1/2}\tilde{W}\tilde{D}^{-1/2} - D^{-1/2}W\tilde{D}^{-1/2}\| &\leq \|\tilde{D}^{-1/2}\tilde{W} - D^{-1/2}W\| \|\tilde{D}^{-1/2}\| \\ &\leq \left(\|\tilde{D}^{-1/2}\tilde{W} - \tilde{D}^{-1/2}W\| + \|\tilde{D}^{-1/2}W - D^{-1/2}W\| \right) \|\tilde{D}^{-1/2}\| \\ &\leq \|\tilde{D}^{-1/2}\|^2 \|\tilde{W} - W\| + \|\tilde{D}^{-1/2} - D^{-1/2}\| \|W\| \|\tilde{D}^{-1/2}\| \\ &\leq a\varepsilon_{\min}^{-1} D_{\min}^{-1} D_{\max} + b r_D \varepsilon_{\min}^{-1/2}, \end{aligned}$$

where the last inequality follows from (24)-(27). For the second term in (28), we have

$$\begin{aligned}
\|D^{-1/2}W\tilde{D}^{-1/2} - D^{-1/2}WD^{-1/2}\| &\leq \|D^{-1/2}W\|\|\tilde{D}^{-1/2} - D^{-1/2}\| \\
&\leq \|D^{1/2}\|\|D^{-1}W\|(bD_{\min}^{-1/2}) \\
&\leq bD_{\max}^{1/2}D_{\min}^{-1/2}.
\end{aligned}$$

Hence

$$\|\tilde{S} - S\| \leq a\varepsilon_{\min}^{-1}r_D + br_D\varepsilon_{\min}^{-1/2} + br_D^{1/2}, \quad (29)$$

where $r_D = \frac{D_{\max}}{D_{\min}}$.

Without loss of generality, assume that $\tilde{v}_j^T v_j \geq 0$ (otherwise reverse the direction of one of the vectors). Then by Lemma 2, we have

$$\|\tilde{v}_j - v_j\| \leq \frac{2^{3/2}\|\tilde{S} - S\|}{\delta_j} \leq 2^{3/2}\delta_j^{-1} \left(a\varepsilon_{\min}^{-1}r_D + br_D\varepsilon_{\min}^{-1/2} + br_D^{1/2} \right). \quad (30)$$

Here we have used the fact S has the same ‘‘eigenvalue gaps’’ (δ_j ’s) as L (in the reversed order) since $\{1 - \lambda_j\}_{j=1}^N$ are eigenvalues of S . It follows that

$$\begin{aligned}
\|\tilde{\phi}_j - \phi_j\| &= \|\tilde{D}^{-1/2}\tilde{v}_j - D^{-1/2}v_j\| \\
&\leq \|\tilde{D}^{-1/2}\tilde{v}_j - D^{-1/2}\tilde{v}_j\| + \|D^{-1/2}\tilde{v}_j - D^{-1/2}v_j\| \\
&\leq \|\tilde{D}^{-1/2} - D^{-1/2}\| + \|D^{-1/2}\|\|\tilde{v}_j - v_j\| \\
&\leq D_{\min}^{-1/2} \left(b + 2^{3/2}\delta_j^{-1} \left(a\varepsilon_{\min}^{-1}r_D + br_D\varepsilon_{\min}^{-1/2} + br_D^{1/2} \right) \right).
\end{aligned} \quad (31)$$

Moreover, it is not hard to show that $\|D^{-1/2}v\| \geq D_{\max}^{-1/2}\|v\|$ for any $v \in \mathbb{R}^N$, which implies that $\|\phi_j\| = \|D^{-1/2}v_j\| \geq D_{\max}^{-1/2}$ as $\|v_j\| = 1$. Similarly $\|\tilde{\phi}_j\| \geq (\varepsilon_{\max}D_{\max})^{-1/2}$. Let θ_j be the angle between $\tilde{\phi}_j$ and ϕ_j . Then

$$\begin{aligned}
\cos \theta_j &\geq 1 - \frac{\|\tilde{\phi}_j - \phi_j\|^2}{2\|\tilde{\phi}_j\|\|\phi_j\|} \\
&\geq 1 - \frac{\varepsilon_{\max}^{1/2}}{2}r_D \left(b + 2^{3/2}\delta_j^{-1} \left(a\varepsilon_{\min}^{-1}r_D + br_D\varepsilon_{\min}^{-1/2} + br_D^{1/2} \right) \right)^2.
\end{aligned} \quad \square$$

Lemma 1 Let $\alpha, \beta \in \mathcal{P}_2(\mathbb{R}^n)$ and A be a $n \times n$ orthogonal matrix. Then $W_2(\alpha, \sigma) = W_2(A_{\sharp}\alpha, A_{\sharp}\sigma)$ for any $\alpha, \sigma \in \mathcal{P}_2(\mathbb{R}^n)$.

Proof Let $\alpha, \sigma \in \mathcal{P}_2(\mathbb{R}^n)$. It suffices to show that $W_2(\alpha, \sigma) \geq W_2(A_{\sharp}\alpha, A_{\sharp}\sigma)$ for any orthogonal matrix A , which implies the the reversed inequality by starting with $W_2(A_{\sharp}\alpha, A_{\sharp}\sigma)$ and applying A^{-1} . Let γ be an optimal transport plan between α and σ . It is not hard to see that $(A, A)_{\sharp}\gamma$ is a transport plan between $A_{\sharp}\alpha$ and $A_{\sharp}\sigma$, where $(A, A) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^n$ is defined by $(A, A)(x, y) = (Ax, Ay)$. By the change of variables formula, we have

$$\begin{aligned}
W_2^2(A_{\#}\alpha, A_{\#}\sigma) &\leq \int_{\mathbb{R}^n \times \mathbb{R}^n} \|\tilde{x} - \tilde{y}\|^2 d((A, A)_{\#}\gamma)(\tilde{x}, \tilde{y}) \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} \|Ax - Ay\|^2 d\gamma(x, y) \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|^2 d\gamma(x, y) \\
&= W_2^2(\alpha, \sigma). \quad \square
\end{aligned}$$

Lemma 2 [31, Corollary 3] Let $\Sigma, \hat{\Sigma} \in \mathbb{R}^{n \times n}$ be symmetric, with eigenvalues $\lambda_1 \geq \dots \geq \lambda_p$ and $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p$ respectively. Fix $j \in \{1, \dots, n\}$, and assume that $\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}) > 0$, where $\lambda_0 := \infty$ and $\lambda_{n+1} := -\infty$. If $v, \hat{v} \in \mathbb{R}^n$ satisfy $\Sigma v = \lambda_j v$ and $\hat{\Sigma} \hat{v} = \hat{\lambda}_j \hat{v}$, then

$$\sin \angle(\hat{v}, v) \leq \frac{2\|\hat{\Sigma} - \Sigma\|}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}.$$

Moreover, if $\hat{v}^T v \geq 0$, then

$$\|\hat{v} - v\| \leq \frac{2^{3/2}\|\hat{\Sigma} - \Sigma\|}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}.$$

Here $\|\cdot\|$ denotes the matrix 2-norm.

Lemma 3 (Wasserstein Distance between isotropic Gaussians)

Let $\alpha_i = \mathcal{N}(\omega_i, c_i I)$ and $\alpha_j = \mathcal{N}(\omega_j, c_j I)$ be two Gaussians supported on \mathbb{R}^n . The W_2 distance between α_i and α_j is

$$W_2^2(\alpha_i, \alpha_j) = \|\omega_i - \omega_j\|^2 + n(\sqrt{c_i} - \sqrt{c_j})^2.$$

Proof By [24, Remark 2.31], the Wasserstein distance between two Gaussians $\alpha = \mathcal{N}(m_\alpha, \Sigma_\alpha)$ and $\beta = \mathcal{N}(m_\beta, \Sigma_\beta)$ is given by

$$W_2^2(\alpha, \beta) = \|m_\alpha - m_\beta\|_2^2 + B(\Sigma_\alpha, \Sigma_\beta)^2$$

where

$$B(\Sigma_\alpha, \Sigma_\beta)^2 = \text{Tr}(\Sigma_\alpha + \Sigma_\beta - 2(\Sigma_\alpha^{1/2} \Sigma_\beta \Sigma_\alpha^{1/2})^{1/2})$$

is the Bures distance, see e.g. [4]. We let $\alpha_i = \mathcal{N}(\omega_i, c_i I)$ and $\alpha_j = \mathcal{N}(\omega_j, c_j I)$, which implies

$$W_2^2(\alpha_i, \alpha_j) = \|\omega_i - \omega_j\|^2 + B(c_i I, c_j I)^2.$$

It is easy to see that $B(c_i I, c_j I)^2 = n(\sqrt{c_i} - \sqrt{c_j})^2$, from which the result follows. \square

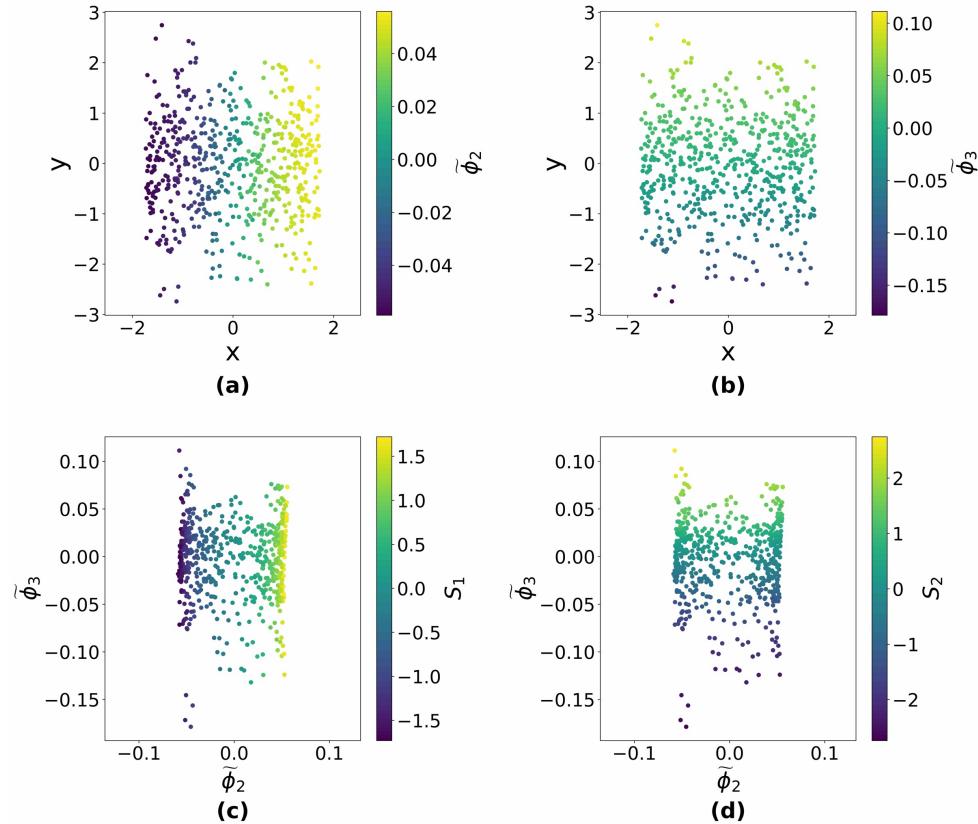


Fig. 6: Illustration of the one-to-one correspondence between the eigenvectors of the graph Laplacian and the independent components described in Remark 6. (a) independent components $S = (S_1, S_2)$ colored by the first non-trivial eigenvector $\tilde{\phi}_2$, (b) independent components $S = (S_1, S_2)$ colored by the second non-trivial eigenvector $\tilde{\phi}_3$, (c) eigenvectors $(\tilde{\phi}_2, \tilde{\phi}_3)$ colored by the first independent component S_1 , and (d) eigenvectors $(\tilde{\phi}_2, \tilde{\phi}_3)$ colored by the second independent components S_2 .

References

1. B. Ans, J. Hérault, and C. Jutten. Architectures neuromimétiques adaptatives: Détection de primitives. In *Proceedings of Cognitiva 85*, pages 593–597, Paris, 1985.
2. A. D. Back and A. S. Weigend. A first application of independent component analysis to extracting structure from stock returns. *International Journal of Neural Systems*, 08(04):473–484, 1997.
3. R. V. Bruggner, B. Bodenmiller, D. L. Dill, R. J. Tibshirani, and G. P. Nolan. Automated identification of stratifying signatures in cellular subpopulations. *Proceedings of the National Academy of Sciences*, 111(26):E2770–E2777, 2014.

4. D. Bures. An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite *-algebras. *Transactions of the American Mathematical Society*, 135:199–212, 1969.
5. Y. Chen, F. D. Cruz, R. Sandhu, A. L. Kung, P. Mundi, J. O. Deasy, and A. Tannenbaum. Pediatric sarcoma data forms a unique cluster measured via the earth mover's distance. *Scientific Reports*, 7(1):7035, 2017.
6. A. Cloninger, K. Hamm, V. Khurana, and C. Moosmüller. Linearized Wasserstein dimensionality reduction with approximation guarantees, 2023. arXiv:2302.07373.
7. P. Comon. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.
8. P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, Inc., USA, 1st edition, 2010.
9. A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, Mar 2004.
10. X. Giannakopoulos, J. Karhunen, and E. Oja. An experimental comparison of neural ICA algorithms. In *ICANN 98: Proceedings of the 8th International Conference on Artificial Neural Networks, Skövde, Sweden*, pages 651–656. Springer, 1998.
11. K. Hamm, N. Henscheid, and S. Kang. Wassmap: Wasserstein isometric mapping for image manifold learning. *SIAM Journal on Mathematics of Data Science*, 5(2):475–501, 2023.
12. K. Hamm, C. Moosmüller, B. Schmitzer, and M. Thorpe. Manifold learning in Wasserstein space, 2023. arXiv:2311.08549.
13. S. Haykin and B. Kosko. Image denoising by sparse code shrinkage. In *Intelligent Signal Processing*, pages 554–568. IEEE, 2001.
14. A. Hyvärinen. New approximations of differential entropy for independent component analysis and projection pursuit. *Advances in neural information processing systems*, 10, 1997.
15. A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2004.
16. A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.
17. L. Kantorovich. On the transfer of masses. *Doklady Akademii Nauk*, 37(2):227–229, 1942.
18. J. Kileel, A. Moscovich, N. Zelesko, and A. Singer. Manifold learning with arbitrary norms. *Journal of Fourier Analysis and Applications*, 27(5):1–56, 2021.
19. J. Mathews, M. Pouryahya, C. Moosmüller, I. G. Kevrekidis, J. Deasy, and A. Tannenbaum. Molecular phenotyping using networks, diffusion, and topology: soft-tissue sarcoma. *Scientific Reports*, 2019.
20. G. Mishne, R. Talmon, R. Meir, J. Schiller, M. Lavzin, U. Dubin, and R. R. Coifman. Hierarchical coupled-geometry analysis for neuronal structure and activity pattern discovery. *IEEE Journal of Selected Topics in Signal Processing*, 10(7):1238–1253, 2016.
21. K. Mohanaprasad, A. Singh, K. Sinha, et al. Noise reduction in speech signals using adaptive independent component analysis (ICA) for hands free communication devices. *International Journal of Speech Technology*, 22:169–177, 2019.
22. G. Monge. *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale, 1781.
23. M. K. Nath and J. Sahambi. Independent component analysis of functional mri data. In *TENCON 2008 - 2008 IEEE Region 10 Conference*, pages 1–6, 2008.
24. G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
25. S. Raychaudhuri, P. Sutphin, J. Chang, and R. Altman. Basic microarray analysis: grouping and feature reduction. *Trends in Biotechnology*, 19:189–193, Jun 2001.
26. Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *Int J Comput Vis*, 40(2):99–121, 2000.
27. F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing, 2015.

28. A. Singer. Spectral independent component analysis. *Applied and Computational Harmonic Analysis*, 21(1):135–144, 2006.
29. C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003.
30. C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 2009.
31. Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis—Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.
32. N. Zelesko, A. Moscovich, J. Kileel, and A. Singer. Earthmover-based manifold learning for analyzing molecular conformation spaces. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1715–1719, 2020.
33. Y. Zhang, R. Jin, and Z.-H. Zhou. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52, 2010.
34. J. Zhao, A. Jaffe, H. Li, O. Lindenbaum, E. Sefik, R. Jackson, X. Cheng, R. A. Flavell, and Y. Kluger. Detection of differentially abundant cell subpopulations in scRNA-seq data. *Proceedings of the National Academy of Sciences*, 118(22), 2021.