ROBUST AND SCALABLE MANIFOLD LEARNING VIA LANDMARK DIFFUSION FOR LONG-TERM MEDICAL SIGNAL PROCESSING

CHAO SHEN, YU-TING LIN, AND HAU-TIENG WU

Abstract. Motivated by analyzing long-term physiological time series, we design a robust and scalable spectral embedding algorithm, coined the algorithm RObust and Scalable Embedding via LANdmark Diffusion (ROSELAND). The key is designing a diffusion process on the dataset, where the diffusion is forced to interchange on a small subset called the *landmark set*. In addition to demonstrating its application to spectral clustering and image segmentation, the algorithm is applied to study the long-term arterial blood pressure waveform dynamics during a liver transplant operation lasting for 12 hours long.

1. Introduction

Learning from data has been an intriguing topic in many scientific fields, particularly the biomedical field. A learning procedure is taking existing data as the "knowledge" or "experience" to interpret incoming data and help field experts make decision. In the clinical medicine, it has been widely debated that eventually a well trained artificial intelligence (AI) by a proper learning procedure could replace physicians. While it is not impossible, we believe that there is still a long way to go. No matter how, a general consensus is that such a system would augment physicians, and release physicians' precious time by automatizing some routine and time-consuming workflow so that physicians.

Motivated by its importance, various datasets of different data types from clinics have been extensively explored from this learning perspective, like electronic health record [30], medical imaging [17], genomic data [38], biomedical waveforms (or time series, signals) [15], etc. To our knowledge, however, how to extract intrinsic dynamics underlying biomedical waveforms for clinical usage is relatively less discussed, and most of existing literature focus on simplifying the waveform information into few scalars [7, 34, 2, 45]. While this has been successfully applied to clinical medicine, we may loss information encoded in the original waveform. One solution to depict intrinsic dynamics directly from the original waveform is obtaining as many features as possible, and selecting suitable parameters for the learning purpose [19]. Another solution is applying manifold learning algorithms to the *original* physiological waveforms [28, 48]. The basic idea in [28, 48] is truncating the physiological waveform into pieces according to some rules, and then apply the spectral embedding algorithm, like the diffusion maps (DM) [9], to embed those pieces into a finite dimensional Euclidean space, which represents the intrinsic dynamics. If the physiological waveform is embedded into the three dimensional Euclidean

space, the physiological waveform is converted into a three dimensional image so that users can visualize the waveform from a different perspective.

It is commonly believed that more data lead to more comprehensive "knowledge" in the learning procedure, and hence more informative features and better visualization of the intrinsic dynamics. We thus hypothesize that if the technique shown in [28, 48] could be applied to analyze long-term physiological waveforms of length on the order of days or weeks, it will be beneficial to supporting the short-term memory of the human brain when handling long-term physiological waveforms. Specifically, due to the short-term memory limitation, it is easy to overlook information hidden in the long-term waveform; for example, what is the relationship between the waveform in the first hour of the operator and the 7th hour? Will the surgeon remember the useful information in the first hour after 5 hours? While this idea is natural, unfortunately, it is prohibited by the computational complexity inherited in most spectral decomposition based machine learning algorithm, including the DM applied in [48].

Recall that the DM is based on the eigendecomposition of the graph Laplacian (GL) matrix. The algorithm has been shown to perform well when the database is "tiny", like in the order of $10^3 \sim 10^4$. However, when the database gets larger, like in the order of 10^6 or above, the algorithm is challenged by the scalability issue. Take electrocardiogram (ECG) into account. There are $10^3 \sim 10^4$ cycles in 1 hour long ECG, and roughly 10^6 cycles in 14 days. In practice, it is natural to consider subsampling the dataset, however, we may loss information. As a result, although the DM works well and provides alternative clinical information [48], it is limited to datasets of length about one hour.

There have been several solutions toward this scalability challenge. One usual technique is the k-nearest neighbor (kNN) scheme. However, it is not robust to noise. Specifically, when the dataset is noisy and the neighboring information is *not* provided, obtaining a reliable kNN information is challenging. A randomized kNN approach is recently considered in [29]. Another practical solution is directly subsampling the dataset, and then recovering the information of interest by the *Nyström extension* [10]. This approach is also called the *Nyström low-rank approximation* [6], the *kernel extension method* [14], or in general the *interpolative decomposition* [32]. This approach has several theoretical backups, for example [6], and has been widely applied. While it works well for some missions, this approach is limited by the information loss during the subsampling process. Yet another approach is speeding up the matrix decomposition by randomization. For example, we can speed up the algorithm by constructing a thin matrix by taking a random subset of columns of the GL matrix and evaluating the singular value decomposition (SVD) [32]. While this approach has been widely applied, to the best of our knowledge, we have limited knowledge about how it helps the spectral embedding algorithms, and how robust it is to the inevitable noise.

Unlike the above, in this paper we propose a novel algorithm that resolves two common challenges when we apply spectral embedding algorithms—robustness and scalability. The algorithm is intuitive and can be summarized in three steps. First, we find a "small" subset of points from the whole dataset, either randomly or by design, or collect a separate clean point cloud of small size, which we call a *landmark set*. Second, we construct an affinity matrix recording the affinities between points

in the whole dataset and the landmark set, and normalize it properly. This normalized affinity matrix is thin; that is, there are fewer columns than rows. Third, evaluate the singular vectors and singular values of the normalized affinity matrix, and embed the dataset using the singular vectors and singular values. As we will make clear soon, this algorithm is directly related to the diffusion process. We coin the proposed algorithm the *RObust and Scalable Embedding via LANdmark Diffusion* (Roseland). The solution is generic and not limited to analyze the motivating physiological waveform problem.

In addition to demonstrating its application to spectral clustering and image segmentation, the Roseland is applied to study the long-term arterial blood pressure waveform dynamics during a liver transplant operation lasting for 12 hours long. For the sake of self-containedness, we also summarize and explain the theoretical results shown in [40], particularly the spectral convergence and robustness of Roseland under the manifold setup.

2. The proposed Roseland algorithm

We assume that we have a data set $\mathcal{X} = \{x_i\}_{i=1}^n \subseteq \mathbb{R}^q$. Take a set $\mathcal{Y} = \{y_k\}_{k=1}^m$, which might or might not be a subset of $\mathcal{X}$. We call $\mathcal{Y}$ the *landmark set*. Fix a non-negative kernel function $K : \mathbb{R}_{\geq 0} \to \mathbb{R}_+$ with proper decay and regularity; for example, a Gaussian function.

First we construct a *landmark-set affinity matrix* $W^{(\mathrm{r})} \in \mathbb{R}^{n\times m}$, which is defined as

$$W_{ik}^{(\mathrm{r})} = K_\epsilon(x_i, y_k) := K\left(\frac{\|x_i - y_k\|_{\mathbb{R}^q}}{\sqrt{\epsilon}}\right). \tag{1}$$

That is, the (i,k)-th entry of $W^{(\mathrm{r})}$ is the similarity between the i-th data point and j-th landmark, and clearly the larger the distance between two points (or the two points are less similar), the smaller the similarity. Next compute a diagonal matrix $D^{(\mathrm{R})}$ as

$$D_{ii}^{(\mathrm{R})} := e_i^\top W^{(\mathrm{r})}(W^{(\mathrm{r})})^\top \mathbf{1}\,, \tag{2}$$

where $\mathbf{1}$ is a $n\times 1$ vector with all entries 1, and e_i is the unit vector with 1 in the i-th entry. $D_{ii}^{(\mathrm{R})}$ is called the degree of the i-th data point x_i. Intuitively, it represents how strong x_i is attached to the data set. With $W^{(\mathrm{r})}$ and $D^{(\mathrm{R})}$, we evaluate the SVD of $(D^{(\mathrm{R})})^{-1/2}W^{(\mathrm{r})}$:

$$(D^{(\mathrm{R})})^{-1/2}W^{(\mathrm{r})} = U\Lambda V^\top, \tag{3}$$

where the singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m \geq 0$ are on the diagonal of the diagonal matrix Λ. Set $\bar{U} := (D^{(\mathrm{R})})^{-1/2}U$. Take $q' \in \mathbb{N}$ so that $q' \leq m$. Let $\bar{U}_{q'} \in \mathbb{R}^{n\times q'}$ to be a matrix consisting of the second to the $(q'+1)$-th columns of $\bar{U}$ and $L_{q'} := \mathtt{diag}(\sigma_2^2, \ldots, \sigma_{q'+1}^2)$. Finally we define the Roseland embedding as

$$\Phi_t^{(\mathrm{R})} : x_i \mapsto e_i^\top \bar{U}_{q'}(L_{q'})^t\,, \tag{4}$$

where $t > 0$ is the chosen diffusion time, in other words, the i-th data point x_i is embedded using the i-th row of $\bar{U}_{q'}$ entry-wisly rescaled by $[\sigma_2^{2t}, \cdots, \sigma_{q'+1}^{2t}]$. See Algorithm 1 for a summarization of the Roseland algorithm. We thus define the associated *Roseland diffusion distance* (RDD) by

$$D_t^{(\mathrm{R})}(x_i, x_j) := \|\Phi_t^{(\mathrm{R})}(x_i) - \Phi_t^{(\mathrm{R})}(x_j)\|_{\mathbb{R}^{q'}}\,. \tag{5}$$

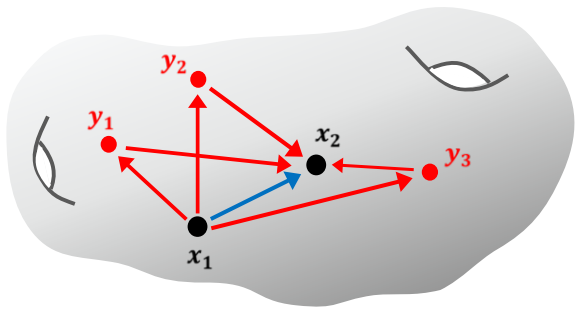

FIGURE 1. Main idea of Roseland: to measure the similarity between x_1 to x_2, instead of diffuse from x_1 to x_2 directly, we take a detour and first diffuse x_1 to the landmarks y_1, y_2, y_3, and then diffuse from the landmarks to x_2.

Algorithm 1 The pseudo-code of Roseland.

procedure INPUT(data set $\mathcal{X} = \{x_i\}_{i=1}^n \subset \mathbb{R}^q$, landmark set $\mathcal{Y} = \{y_i\}_{i=1}^m \subset \mathbb{R}^q$, kernel K, bandwidth $\sigma > 0$, embedded dimension $q' \in \mathbb{N}$, $0 \le \alpha \le 1$, and diffusion time $t > 0$)
  Construct the affinity matrix $W^{(\mathrm{r})}$ w.r.t. the landmark set.
  Construct the degree matrix $D^{(\mathrm{R})}$.
  Run SVD $(D^{(\mathrm{R})})^{-1/2} W^{(\mathrm{r})} = U \Lambda V^\top$, where $U \in O(n)$ and $V \in O(m)$. Denote singular values as $\sigma_1 \ge \sigma_2 \ge \ldots \ge 0$.
  Set $\bar{U} = (D^{(\mathrm{R})})^{-1/2} U$.
  Let $\bar{U}_{q'} \in \mathbb{R}^{n \times q'}$ be the second to the $(q'+1)$-th columns of $\bar{U}$. Set $L_{q'} := \mathtt{diag}(\sigma_2^2, \sigma_3^2, \ldots, \sigma_{q'+1}^2)$.
  Embed $\mathcal{X}$ via $\Phi_t^{(\mathrm{R})} : x_i \mapsto e_i^\top \bar{U}_{q'} L_{q'}^t$
  Output $\{\Phi_t^{(\mathrm{R})}(x_i)\}_{i=1}^n \subset \mathbb{R}^{q'}$.
end procedure

Note that the Roseland induces a new affinity matrix on the data set $\mathcal{X}$ via

$$W^{(\mathrm{R})} := W^{(\mathrm{r})} (W^{(\mathrm{r})})^\top \in \mathbb{R}^{n \times n}, \tag{6}$$

where $W^{(\mathrm{r})}$ is the landmark-set affinity matrix (1). We call $W^{(\mathrm{R})}$ the *landmark-affinity matrix*, which is positive and positive-definite. We remark that traditional affinity matrices between data points are often constructed from *one* global pre-fixed kernel K, while in Roseland we cannot find a global fixed kernel $\bar{K}$ and a bandwidth $\bar{\epsilon} > 0$ so that $W_{ij}^{(\mathrm{R})} = \bar{K}(\|x_i - x_j\|/\bar{\epsilon})$ for all i, j in general. Hence Roseland can provide us with more dynamic similarity measurements between data points based on the landmarks. Also note that $A^{(\mathrm{R})} := (D^{(\mathrm{R})})^{-1} W^{(\mathrm{R})}$ is a transition matrix on $\mathcal{X}$, which means each of the row in $A^{(\mathrm{R})}$ sums to 1. This is because $D^{(\mathrm{R})}$ is the degree matrix associated with the landmark-affinity matrix $W^{(\mathrm{R})}$ by construction. Hence, we can view the similarity measure between data points x_i and x_j as a Markov or diffusion process through the landmarks.

2.1. **Reference set as subset of the data.** In cases that we may not be able to acquire additional data points as landmark set but have to select the landmark set

from the available dataset, we propose to first sample m landmarks, denoted by $\mathcal{Y}$, from $\mathcal{X}$ so that $\mathcal{Y}$ is independent of $\mathcal{X} \setminus \mathcal{Y}$. Then, we apply the Roseland on $\mathcal{X} \setminus \mathcal{Y}$ using $\mathcal{Y}$ as landmarks, and extend the embedding to $\mathcal{Y}$ by the Nyström extension. When $|\mathcal{Y}| \ll |\mathcal{X}|$, the discrepancy of this approach and the independence setup with original Roseland is negligible, and will asymptotically vanish.

2.2. **Related methods.**

2.2.1. *Graph Laplacian and Diffusion Maps.* We now review the GL and compare Roseland with the well-known algorithm diffusion map (DM) [9]. First, pre-fix a kernel function K and a bandwidth parameter $\epsilon > 0$. Then, compute the affinity matrix $W \in \mathbb{R}^{n\times n}$ by

$$W_{ij} := K\left(\frac{\|x_i - x_j\|_{\mathbb{R}^q}}{\sqrt{\epsilon}}\right) \tag{7}$$

and the corresponding *degree matrix* $D \in \mathbb{R}^{n\times n}$, which is a diagonal matrix defined as $D_{ii} := \sum_{j=1}^n W_{ij}$. For a fixed $\alpha \in [0,1]$, the α-*normalized* affinity matrix $W^{(\alpha)} \in \mathbb{R}^{n\times n}$ [9] is defined as $W_{ij}^{(\alpha)} := \frac{W_{ij}}{D_{ii}^\alpha D_{jj}^\alpha}$, where $W_{ij}^{(\alpha)}$ is called the α-*normalized affinity* between x_i and x_j. Note that $W^{(0)} = W$ defined in (7). In some applications when we want to remove the density effect caused by data sampling, we set $\alpha = 1$. With the α-normalized affinity matrix $W^{(\alpha)}$, one can analogously define the associated degree matrix $D^{(\alpha)} \in \mathbb{R}^{n\times n}$ by $D_{ii}^{(\alpha)} := \sum_{j=1}^n W_{ij}^{(\alpha)}$. The GL is defined as $L^{(\alpha)} := I - A^{(\alpha)}$, where

$$A^{(\alpha)} := (D^{(\alpha)})^{-1} W^{(\alpha)} \tag{8}$$

is the associated transition matrix. Clearly, $A^{(\alpha)}$ is row stochastic, and it defines a random walk on the dataset $\mathcal{X}$. We mention that the transition matrix $(D^{(\mathrm{R})})^{-1}W^{(\mathrm{R})}$ in the Roseland algorithm can be viewed as an alternative way of constructing a Markov process on the dataset $\mathcal{X}$. Since $A^{(\alpha)}$ is similar to $(D^{(\alpha)})^{-1/2}W^{(\alpha)}(D^{(\alpha)})^{-1/2}$ we can find its eigendecomposition with eigenvalues $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$ and the associated eigenvectors $\phi_1, \ldots, \phi_n$. Denote ϕ_i the i-th right eigenvector of $A^{(\alpha)}$. Recall in Roseland, we perform the SVD decomposition in (3), which is a parallel step of the eigen-decomposition of $\bar{A}^{(0)}$.

Among various algorithms, we focus on the well-known DM algorithm that we shall see to be closely related to Roseland. With the spectral decomposition of the GL, the chosen normalization α, embedding dimension q' and *diffusion time* t, the DM embeds $\mathcal{X}$ via the map

$$\Phi_t : x_i \mapsto e_i^\top \tilde{U}_{q'}^{(\alpha)} (\Lambda_{q'}^{(\alpha)})^t \in \mathbb{R}^{q'}, \tag{9}$$

where $\tilde{U}_{q'}^{(\alpha)} \in \mathbb{R}^{n\times q'}$ to be a matrix consisting of the second to the $(q'+1)$-th columns of $\tilde{U}^{(\alpha)}$ and $\Lambda_{q'}^{(\alpha)} := \mathtt{diag}(\lambda_2, \ldots, \lambda_{q'+1})$. The diffusion distance (DD) with the diffusion time $t > 0$ is defined as

$$D_t(x_i, x_j) := \|\Phi_t(x_i) - \Phi_t(x_j)\|_{\mathbb{R}^{q'}}. \tag{10}$$

Recall the Roseland embedding (4) and the RDD (5) to notice the close relation between the DM embedding and the DD.

As it is expensive to perform eigen-decomposition of dense matrices, one common practice of DM or general spectral embedding methods is to use the kNN scheme to construct a rather sparse affinity matrix; that is, set $W_{ij} = 0$ when x_j is not within

the first k nearest neighbors of x_i, where k is chosen by the user. Another way is to use a compactly supported kernel K. For example, $K(t)$ is 1 when $t \in [0,1]$ and 0 when $t > 1$.

2.2.2. *Nyström Extension.* Another directly related algorithm is the common Nyström extension [3, 14, 50]. The basic idea is running the eigen-decompostion on a small subset of the whole database, and then extending the eigenvectors to the whole dataset. In this work, we apply the modified Nyström extension [22, 43] that is suitable for the diffusion process. See [40] for a summary of this algorithm.

3. Theoretical results

In this section, we summarize two theoretical results shown in the accompanying paper [40], one is the asymptotic spectral convergence and one is the robustness of Roseland under the manifold setup.

3.1. **Manifold model.** Denote our observed dataset by $\mathcal{X} = \{x_i\}_{i=1}^n \subseteq \mathbb{R}^q$. Assume $\mathcal{X}$ is independently and identically (i.i.d.) sampled from a random vector $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^q$, where the range of X is assumed to be supported on a d-dimensional compact smooth Riemannian manifold (M^d, g) without boundary isometrically embedded in $\mathbb{R}^q$ via $\iota : M^d \hookrightarrow \mathbb{R}^q$. Suppose the density function of X on M is denoted as p_X. Similarly, denote the landmark set by $\mathcal{Y} = \{y_i\}_{i=1}^m \subseteq \mathbb{R}^q$. Assume $\mathcal{Y}$ is i.i.d. sampled from a random vector $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^q$, where the range of Y is assumed to be (M^d, g) as well. Suppose the density function of Y on M is denoted as p_Y. Furthermore, assume that X and Y are independent.

3.2. **Spectral convergence.** Let $\{v_n\}_{n\in\mathbb{N}}$ be the set of eigenvectors of the matrix $(D^{(\mathrm{R})})^{-1}W^{(\mathrm{R})}$ evaluated from the dataset $\mathcal{X}$ and the landmark $\mathcal{Y}$. It is shown in [40] that the eigenvectors $\{v_n\}_{n\in\mathbb{N}}$ asymptotically converge to the eigenfunctions of the Laplace-Beltrami operator of (M, g) as $n \to \infty$. Hence, by the well established spectral geometry theory [4, 36], the Roseland embedding recover the underlying manifold.

First, note that the eigenvectors v_n are in difference Euclidean spaces for different n, hence they cannot be compared directly with the eigenfunctions of the Laplace-Beltrami operator, which are smooth functions on M. To make sense of the comparison, we find a sequence of functions $f_n \in C(M)$, such that the restriction of f_n on the data $\mathcal{X}$ is v_n; that is, $f_n(x_i) = v_n(i)$, for $i = 1, \ldots, n$. After that, we study the convergence of $\{f_n\}$ as $n \to \infty$. We start with some definitions.

Definition 1. *Take $\mathcal{X} = \{x_i\}_{i=1}^n \subset M$. Define the following functions*

$$\widehat{K}_{\mathrm{ref},\epsilon,n}(x,y) := \frac{1}{m}\sum_{j}^{m} K_\epsilon(x, z_j) K_\epsilon(z_j, y) \in C(M \times M),$$

$$\widehat{d}_{\mathrm{ref},\epsilon,n}(x) := \frac{1}{n}\sum_{i=1}^{n} \widehat{K}_{\mathrm{ref},\epsilon,n}(x, x_i) \in C(M), \quad \widehat{M}_{\mathrm{ref},\epsilon,n}(x,y) := \frac{\widehat{K}_{\mathrm{ref},\epsilon,n}(x,y)}{\widehat{d}_{\mathrm{ref},\epsilon,n}(x)} \in C(M \times M)\,.$$

Let

$$\widehat{T}_{\mathrm{ref},\epsilon,n} f(x) := \frac{1}{n}\sum_{i=1}^{n} \widehat{M}_{\mathrm{ref},\epsilon,n}(x, x_i) f(x_i)\,. \tag{11}$$

Moreover, define the restriction operator $\rho_n : C(M) \to \mathbb{R}^n$ by $\rho_n : f \mapsto [f(x_1), f(x_2), \ldots, f(x_n)]^\top$.

The following Lemma essentially says that if we restrict the eigenfunctions of $\widehat{T}_{\mathrm{ref},\epsilon,n}$ on the data, we obtain the eigenvectors of $(D^{(\mathrm{R})})^{-1}W^{(\mathrm{R})}$.

Proposition 1. [40] *Let $U_n := (D^{(\mathrm{R})})^{-1}W^{(\mathrm{R})}$, then $U_n \circ \rho_n = \rho_n \circ \widehat{T}_{\mathrm{ref},\epsilon,n}$. Moreover, we have the following one to one correspondence.*

(1) If $f \in C(M)$ is an eigenfunction of $\widehat{T}_{\mathrm{ref},\epsilon,n}$ with the eigenvalue λ, then the vector $v := \rho_n f$ is an eigenvector of U_n with the eigenvalue λ. Moreover, suppose $\lambda \neq 0$ is an eigenvalue of $\widehat{T}_{\mathrm{ref},\epsilon,n}$ with the eigenfunction f. If we let $v := \rho_n f$, then f satisfies

$$f(x) = \frac{\sum_{j=1}^n \widehat{K}_{\mathrm{ref},\epsilon,n}(x,x_j)v_j}{\lambda \sum_{j=1}^n \widehat{K}_{\mathrm{ref},\epsilon,n}(x,x_j)}. \tag{12}$$

(2) If v is an eigenvector of U_n with the eigenvalue $\lambda \neq 0$, then f defined in (12) is an eigenfunction of $\widehat{T}_{\mathrm{ref},\epsilon,n}$ with the eigenvalue λ.

Denoted by Δ the Laplace-Beltrami operator on (M,g). Let (λ_i, u_i) be the i-th eigenpair of $-\Delta$, where the eigenvalues λ_i are in the ascending order. Under our manifold setup and by the well known elliptic theory, the spectrum of $-\Delta$ is discrete with ∞ as the only accumulation point, and each eigenspace is of finite dimension. Correspondingly, let $(\lambda_{\epsilon,n,i}, u_{\epsilon,n,i})$ be the i-th eigenpair of $\frac{I-T_{\mathrm{ref},\epsilon,n}}{\epsilon}$, where the eigenvalues $\lambda_{\epsilon,n,i}$ are in the ascending order as well. We assume that $\|u_i\|_2 = \|u_{\epsilon,n,i}\|_2 = 1$. With the above preparation, we have the following theorem.

Theorem 1 (Spectral convergence)**.** [40] *Fix $K \in \mathbb{N}$. Suppose the kernel is Gaussian; that is, $K_\epsilon(x,y) = e^{-\|x-y\|^2/\epsilon}$. Suppose λ_i is simple. Suppose $m = n^\beta$, where $\beta \in (0,1)$, $\epsilon = \epsilon(n)$ so that $\epsilon \to 0$ and $\frac{\sqrt{-\log\epsilon}+\sqrt{\log m}}{\sqrt{m}\epsilon^d} \to 0$, as $n \to \infty$, and $\sqrt{\epsilon} \leq \mathcal{K}_1 \min\left(\left(\frac{\min(\Gamma_K,1)}{\mathcal{K}_2+\lambda_K^{d/2+5}}\right)^2, \frac{1}{(2+\lambda_K^{d+1})^2}\right)$, where Γ_K, $\mathcal{K}_1$ and $\mathcal{K}_2 > 1$ are introduced in [40]. Then, when p_Y is properly chosen so that $\frac{2\nabla p_X(x)}{p_X(x)} + \frac{\nabla p_Y(x)}{p_Y(x)} = 0$, there exists a sequence of signs $\{a_n\}$ such that with probability $1-\mathcal{O}(m^{-2})$, for all $i < K$, we have*

$$\|a_n u_{\epsilon,n,i} - u_i\|_{L^\infty} = \mathcal{O}(\epsilon^{1/2}) + \mathcal{O}\left(\frac{\sqrt{-\log\epsilon}+\sqrt{\log m}}{\sqrt{m}\epsilon^{2d+2}}\right), \tag{13}$$

$$|\lambda_{\epsilon,n,i} - \lambda_i| = \mathcal{O}(\epsilon^{3/4}) + \mathcal{O}\left(\frac{\sqrt{-\log\epsilon}+\sqrt{\log m}}{\sqrt{m}\epsilon^{2d+2}}\right),$$

where the implied constants depend on the kernel, the curvature of M, p_X and p_Y

We see that the error term converges to zero if we further assume that $\frac{\sqrt{-\log\epsilon}+\sqrt{\log m}}{\sqrt{m}\epsilon^{2d+2}} \to 0$. We mention that when λ_i is not simple, the theorem still holds, but we need to use the eigenprojection to describe the eigenvector convergence statement.

3.3. **Robustness.** In real applications, the dataset is often corrupted by noise. In this case, spectral embedding algorithms might lead us to a low quality, or even misleading result. This robustness issue for DM was studied in [11, 12] and some solutions are proposed. It is suggested in [12] to work with the complete graph or a graph with a large number of kNN, and force the random walk non-lazy. While this solution works, however, it is not scalable. Roseland, on the

other hand, automatically enjoys the desired robustness property. Note that we measure similarities between data points by diffusing through *all* landmarks in Roseland. This seemingly simple step has a significant consequence. It can be viewed as a surrogate of knowing *true neighbors* in the kNN scheme, and it explains the robustness of Roseland. To appreciate this significance, recall that when the neighboring information is not provided and if the kNN approach is considered, we need to estimate the neighbors. However, estimating neighbors from noisy data is error-prone and wrong neighboring information corrupts the DM.

To show the robustness of Roseland, on top of the manifold model in Section 3.1, we assume that the data set and the landmark set are corrupted by additive ambient space noise. That is, we observe $\widetilde{\mathcal{X}} = \{\tilde{x_i}\}_{i=1}^n$ and $\widetilde{\mathcal{Y}} = \{\tilde{y_j}\}_{j=1}^m$:

$$\tilde{x_i} = x_i + \xi_i, \quad \tilde{y_j} = y_j + \eta_j \,,$$

where ξ_i and η_j are noise. We assume that x_i, y_j, ξ_i and η_j are independent. For more practically challenging purpose, we consider large q large n setup, that is, the ambient space dimension grows asymptotically as the dataset size n. To simplify the discussion, we focus on Gaussian noise, and refer readers with interest to [40] for more complicated non-Gaussian noise model and results.

Assumption 3.1 (Gaussian noise model)**.** *Suppose the noise contaminating the dataset is $\xi_i \sim \mathcal{N}(0, \Sigma)$ with mean 0 and covariance Σ and the noise contaminating the landmark set is $\eta_i \sim \mathcal{N}(0, \overline{\Sigma})$ with mean 0 and covariance $\overline{\Sigma}$. Assume ξ_i and η_j are independent. Suppose $\|\Sigma\|_2 \leq \sigma_q^2$ and $\|\overline{\Sigma}\|_2 \leq \overline{\sigma}_q^2$, where $\sigma_q \geq 0$ and $\overline{\sigma}_q \geq 0$.*

Recall the Roseland algorithm in Section 2. Denote the Roseland embedding with diffusion time t from clean data and noisy data by $\Phi L^t \subset \mathbb{R}^{n\times q'}$ and $\widetilde{\Phi}\tilde{L}^t \subset \mathbb{R}^{n\times q'}$ respectively. We analyze the discrepancy between these two embeddings. Since the embeddings are free up to rotations and reflections, we quantify $\|\Phi O L^t - \widetilde{\Phi}\widetilde{L}^t\|_F$, where $O \in \mathbb{R}^{q'\times q'}$ is some orthogonal matrix that aligns the embeddings, and $\|\cdot\|_F$ is the Frobenius norm. We have the following theorem.

Theorem 2 (Robustness of Roseland)**.** [40] *Assume the point clouds $\{x_i\}_{i=1}^n$ and $\{y_j\}_{j=1}^m \subseteq \mathbb{R}^q$ are i.i.d. sampled from the underlying manifold. Let $\tilde{x_i} = x_i + \xi_i$ and $\tilde{y_j} = y_j + \eta_i$, where the noises ξ_i and y_j satisfy Assumption 3.1 are independent of x_i and y_j. Also, assume*

$$\delta_q := \sqrt{\log nm}\sqrt{\sigma_q^2 + \overline{\sigma}_q^2}\left[\sqrt{q(\sigma_q^2 + \overline{\sigma}_q^2)} + K\right] . \tag{14}$$

Fix $q' \in \mathbb{N}$. According to Theorem 1, pick a sufficiently small $\epsilon = \epsilon(q') > 0$ so that the first q' non-trivial singular values are sufficiently away from zero when n is sufficiently large. Denote $W^{(\mathrm{r})}$ and $\widetilde{W}^{(\mathrm{r})}$ to be the landmark-set affinity matrices from clean and noisy datasets respectively. Denote $\Phi L^t \in \mathbb{R}^{n\times q'}$ and $\widetilde{\Phi}\widetilde{L}^t \in \mathbb{R}^{n\times q'}$ to be Roseland embeddings from $W^{(\mathrm{r})}$ and $\widetilde{W}^{(\mathrm{r})}$ respectively. Then, for fixed $t \geq 0$ and $q' \in \mathbb{N}$, we have

$$\left\|\Phi O L^t - \widetilde{\Phi}\widetilde{L}^t\right\|_F = \mathcal{O}_P\left(\frac{\delta_q}{\sqrt{m}}\frac{q' t s_2^{2t-2} + \sqrt{q'} s_2^{2t}}{\epsilon^{2d+1}}\right) .$$

where $O \in \mathbb{R}^{q'\times q'}$ is an orthogonal matrix, and s_2 are the largest non-trivial singular value of Roseland from the clean.

The Theorem says that the bandwidth ϵ should chose "large" enough so that the noise impact on the embedding is alleviated. This fulfills the intuition that we can tame the noise by aggregating more independent noise, since a larger bandwidth indicates including more noisy data locally. Note that while $\delta_q \to 0$ when $n \to \infty$, it does not mean that the noise impact is small when $n \to \infty$. We shall quantify the noise impact by $q(\sigma_q^2 + \overline{\sigma}_q^2)$ in δ_q, which can be viewed as the *total energy of noise* in the data. From this perspective, we mention that Roseland can tolerate large noise, with the noise level up to $\sigma_q = q^{-(1/4+a)}$ for arbitrary constant $a > 0$. In this setup, the total energy in noise blows up since $\sigma_q^2 q \to \infty$. For more discussion, see [40].

Moreover, when the landmark set is noise free, we could achieve a better convergence since $\overline{\sigma}_q = 0$. This condition is related to the situation that we are able to collect a small clean dataset as the landmark set in addition to the large but noisy dataset. This situation is commonly encountered in real life. For example, in the medical field, collecting a clean dataset of high quality is usually labor-intensive and expensive. However, it is relatively easy to collect a large dataset from a rather cheap equipment, in exchange of the data quality.

4. Numerical Results

In this section, we show how to apply Roseland to speed up spectral clustering and image segmentation when the dataset is big.

4.1. **Spectral clustering.** MNIST is a dataset consisting of labelled handwritten digits from 0 to 9, in the form of 28×28 gray-scale images [23]. There are $60,000$ training images and $10,000$ testing images. For the spectral clustering purpose, we include all $70,000$ data points. For a fair comparison between Nyström extension and Roseland, we fix the landmark set of size 250; that is $\beta \approx 0.5$.

A spectral clustering algorithm consists of two steps: the first step is performing a chosen spectral embedding, and the second step is applying the K-mean clustering. In our experiment, we set $K = 10$. Let $\{C_k\}_{k=1}^{10}$ be the clusters obtained by the spectral clustering. To evaluate the performance of the spectral clustering based on different spectral embedding algorithms, including the diffusion map, Roseland and Nyström extension, we consider the following procedure. Denote the data set by $\mathcal{X}$, so $\mathcal{X} = \bigsqcup_{k=1}^{10} C_k$. Let $f : \mathcal{X} \to y$ be the true predictor function and $\hat{f} : \mathcal{X} \to \hat{y}$ be the estimated predictor function that is constant on C_k for $\forall k$. Here, $\hat{f}$ is defined in the following way:

$$\hat{f}|_{C_k} = \arg\max_y \sum_{x \in C_k} \mathbb{I}\{f(x) = y\}\,,$$

for $k = 1, \cdots, 10$, where $\mathbb{I}$ is the indicator function. In other words, for each cluster, we take a vote to decide the label it is associated with. As MNIST is almost balanced, we compute the accuracy as follows:

$$Acc(\hat{f}) = \frac{1}{10} \sum_{k=1}^{10} \sum_{x \in C_k} \frac{\mathbb{I}\{\hat{f}(x) = y\}}{|C_k|}\,,$$

where $|\cdot|$ is the size of the cluster. In addition to running the above on the clean MNIST, we also run the algorithm on the MNIST with additive Gaussian noise.

Specifically, we first normalized all pixel values, then entry-wisely add Gaussian noise with mean 0 and standard deviation 0.2.

We plot 3 dimensional spectral embeddings for a visual comparison. To enhance the color contrast for the visualization purpose, we only show the first 5 digits. See Figure 2 for the embedding result. In our simulations, the DM (with the sparse kernel matrix) of the whole MNIST dataset took approximately 1,500 seconds, the Nyström extension took about 5 seconds, and the Roseland took about 15 seconds. The accuracies are reported in Figure 3. All of the simulations were done on a Linux machine with 4-core 3.5Ghz i5 CPUs and 16GB memory.

We mention that the two largest experiments we are aware of with the MNIST dataset are [16] and [26]. In [16], the authors improved the convergence of the kernel Hebbian algorithm to accelerate the kernel Principle Component Analysis algorithm, and the total run time is more than 50 hours on $60,000$ images. No classification accuracy was reported. The experiments were performed on an AMD Athlon 2.4 GHz CPU with 2 GB main memory and 512 kB cache. In [26], the authors applied a similar idea to HKC, but with a different normalization. Similar to [51], the algorithm output clusters rather than embeddings. The algorithm took about 100 seconds on $70,000$ MNIST images, and the overall accuracy achieved 70%. In [35], the authors adapted the Nyström idea and proposed a greedy algorithm to build the subset via projection. In their experiments, they restricted the run time to be less than 14 seconds, and computed the embeddings of 1,300 MNIST images consisting of only 0's and 1's with image sizes scaled down from 28×28 to 14×14.

4.2. **Image segmentation.** Next, we apply the spectral clustering algorithm to the image segmentation problem [41]. The first two images shown in Figure 4 are of size 481×321 from the Berkeley Segmentation Dataset and Benchmark [31] and the third image is of size 582×640 from the COCO data set [27]. To make the task more challenging, we corrupted the images with additive noise. Specifically, we first normalized all pixel values, then entry-wisely add Gaussian noise with mean 0 and standard deviation 0.2, see Figure 5.

Due to the data size, we only applied the Roseland to perform image segmentation. We slightly modified the algorithm proposed in [41]. Specifically, for each pixel, we attach a 3×3 patch surrounding the pixel. For all images, we fix the landmark set size to be 150. For the two images of size 481×321, the Roseland based image segmentation took about 5 seconds, and for the image of size 582×640, it took about 12 seconds. The results are shown in Figures 4 and 5.

5. Liver transplant analysis

5.1. **Background.** Liver transplant surgery is the only life-saving treatment for patients in certain medical conditions. It is a challenging surgical procedure, and significant medical resource, experience and dedication are needed. During the surgery, the clamping of major vessels and the subsequent vascular anastomosis bring huge impacts on the recipient's circulation system [39]. As a better understanding of the cardiovascular dynamics during the procedure may help optimize the intraoperative management, commercial monitoring instruments based on real-time arterial blood pressure (ABP) waveform analysis has been introduced. However, they been questioned subsequently for its performance in liver transplant surgery

[5, 46, 42]. Thus, obtaining useful information from the ABP waveform in liver transplant is still a challenging problem.

Traditionally, various extracted features, either landmark measurements in the time domain [33] or quantities in the frequency domain [49], serve as the input for the subsequent pulse waveform analysis. These designated features are supposed to reflect underlying physiological information, or those parameters driving the network interaction. However, it is reasonable to suspect that information hidden in the finer scale might be ignored via the above approach, and hence finer structure of network dynamics is overlooked, particularly when the physiology is disturbed. It is thus reasonable to argue that taking the whole waveform into account might provide more complimentary information compared with those traditional parameters. On the other hand, due to the short-term memory nature of human brain, it is challenging to visualize and directly utilize the dynamics encoded in the ABP waveform on the large scale. Motivated by handling the above challenges, including finding finer information on the short scale, and exploring the dynamics on the large scale, in our previous research, we reported a solution under the manifold learning framework, and showed that the DM can extract rich information directly from the *raw cardiovascular waveform* [48]. The novelty in [48] capturing subtle morphological changes that might be overlooked by the designed features. However, due to the computational barrier intrinsic to the DM, the approach is limited to relatively small dataset.

In this study, we hypothesize that with the help of Roseland, the manifold learning approach shown in [48] can be applied to study the ABP waveform during the liver transplant procedure, and provide hemodynamic information on both the small and large scales. Note that the Roseland is critical in this analysis since the whole period of the surgery can yield more than 10^5 continuous pulses as data points in high dimensional space for a pulse-to-pulse waveform analysis.

5.2. **Material.** The data was collected from an observational study per institutional ethic regulation. We collected physiological signals via the data collection software, S5 collect (GE Healthcare, Chicago, Illinois, United States) from the standard patient monitor instrument, GE CARESCAPE^{TM}B850 (GE Healthcare, Chicago, Illinois, United States). The recorded ABP signal was uniformly sampled at 300 Hz in the instrument and resampled at 500 Hz via the cubic spline interpolation for off-line processing. The signal is of 78,350s long spanning the whole surgical procedure and contains 120,725 pulses.

5.3. **Data analysis.** Denote the ABP waveform as $x^A \in \mathbb{R}^N$. We used the maximum of the first derivation during the ascent of each ABP pulse waveform as a fiducial point. A legitimate ABP pulse is determined by a two-pass algorithm using the following measurements automatically: the peak maximum, the trough minimum, the minimum of difference between the maximum and minimum within the pulse, the pulse width, and the duration to the previous pulse. The thresholds for those measurements are automatically adjusted by a feedback mechanism. Suppose there are L legitimate cycles in x^A. Denote the i-th fiducial point as n_i. Break x^A into $L-1$ segments so that the i-th segment is the i-th ABP pulse containing one waveform cycle. Denote the i-th segment as $\bar{x}_i^A := [x^A(n_i), x^A(n_i+1), \ldots, x^A(n_{i+1})]^T$. Since the duration of each pulse is not constant, we truncated them into an uniform size according to their minimal length

$q = \min\{n_{i+1} - n_i + 1\} \in \mathbb{N}$, and get $\hat{x}_i^A := [x^A(n_i), x^A(n_i+1), \ldots, x^A(n_i+q-1)]^T$. Next, normalize $\hat{x}_i^A$ by removing the mean and setting the variance to 1 to separate the blood pressure information from the normalized ABP pulse, and denote the normalized ABP pulse as $x_i^A \in \mathbb{R}^q$. Derived from the ABP signal, we get the data set $\mathcal{X}^A = \{[(x_i^A)^T\ (x_{i-1}^A)^T\ (x_{i-2}^A)^T]^T\}_{i=3}^L \subseteq \mathbb{R}^{3q}$. We assume that $\mathcal{X}^A$ can be well approximated by a low dimensional manifold, referred to as the *wave-shape manifold* [28]. To apply the Roseland, the landmark set $\mathcal{Y}^A = \{y_k^A\}_{k=1}^m$, where $m = \lfloor n/600 \rfloor$, was chosen from $\mathcal{X}^A$ via setting $y_k^A = x_{600k}^A$.

5.4. **Results.** The dataset consists of $L = 120,725$ legitimate cycles of length $q = 334$, and hence $120,723$ data points in $\mathcal{X}^A$. Thus, $m = 201$ and note that $201 = 120,723^{0.453}$. The total computation time of the Roseland algorithm is less than 40s on an ordinary personal computer (CPU: Intel Core i5-7500, operation system: Microsoft Windows 10 64-bit home edition, programming platform and language: Microsoft Visual Studio Community version 2019, .NET framework 4.8, and C#, LAPACK software library: Intel Math Kernel Library 2020 Initial Release), while the estimated computation time using traditional DM algorithm based on eigendecomposition would be more than one day.

The embedding result is shown in Figure 6. The successive pulses evolve with time and constitute a trajectory on the manifold presented as a 3D embedding (Fig.6, panel A). The trajectory visits different locations during different steps of the liver transplant procedure. Moreover, there is a "clustering" effect in the embedding, which is enhanced by the imposed color that encodes the temporal information. We can thus visualize the relationship among different hemodynamic status during different surgery steps. This relationship provides physiological dynamics on the large scale. We emphasize that while we can easily read the waveform, but it is not easy to perceive the dynamics and organize them with only human eyes and brain (Fig.6, panel B and C), particularly when the signal is long.

We further quantify the trajectory in different surgical phases as well as the phase transition periods in which the trajectory moves in fast pace. We consider the following different hemodynamic phases during the liver transplant procedure, particularly those that phase transition happens with violent physiological changes take place—First, the occlusion of blood inflow to the "old" liver (to be replaced), performed by the cross-clamp of the inferior vena cava, the largest vein of the human body; second, the start of the circulatory connection from the graft (the new liver organ) to the circulation system as the blood flow starts in the portal vein; third, the start of the connection between hepatic artery and the graft. All these transitions drastically affect the cardiovascular system via the changes of fluid volume and electrolyte composition.

For each hemodynamic phase, we embed the ABP waveforms by the Roseland embedding into 10-dim Euclidean space; that is, $q' = 10$, and evaluate the geometric center of all beats during that phase. Then, we define the distance between two hemodynamic phases groups by measuring the RDD between their geometric centers. The quantitative measurement is expressed as mean and 95 % confidence interval after bootstrap resampling without replacement in 100,000 samples. In light of the Newtonian mechanics, we consider two quantities. The *two-point velocity* measures the dynamics on the large scale. The two-point velocity is defined as the ratio of the RDD between two points and the time difference between the two points. We also consider the *trajectory speed*, which measures the dynamics on

TABLE 1. Quantitative results of phases and phase transitions from liver transplantation data

Surgical phases	two point velocity	trajectory speed	RDD to previous phase
dissection phase before clamp	0.59	32.77	na
vascular clamp (transition)	5.34	32.32	na
anhepatic phase after clamp	0.84	36.58	2152.6 (2127.9, 2177.2)
anhepatic phase before reperfusion	0.74	17.53	409.8 (386.1, 434.0)
reperfusion (transition)	6.27	41.41	na
neohepatic phase after reperfusion	0.64	13.26	984.8 (976.4, 993.3)
before artery anastomosis	0.21	16.65	439.3 (422.5, 456.3)
artery anastomosis (transition)	2.47	22.23	na
after artery anastomosis	0.34	11.87	1818.9 (1802.8, 1834.9)

All numbers expressed by 10^{-5}; speed and velocity unit: s^{-1}

the small scale. The trajectory speed is defined as the ratio of the path length of the trajectory and the time different between the beginning and ending of the trajectory. The quantitative result (Table 1) shows that during the phase transition, the trajectory moves faster macroscopically (on the large scale) when quantified by the two-point velocity, particularly when compared with that within each surgical phase. However, the trajectory speeds, which represents hemodynamics on the microscopic scale (small scale), are similar during the phase transition and within each surgical phase. The numeric results are consistent with visualization from the 3D embedding (Figure 6 and 7).

The liver-transplant example shows the benefit and potential of the Roseland algorithm. As more (longer) data leads to a richer knowledge base, we now have an unprecedented signal processing tool for a long-period signal with complex underlying physiology. In the liver transplant example, the 3D embedding reflects the complex relationship among different surgical phases without *ad hoc* pulse waveform knowledge. This suggests the practicability of handling data governed by complex physiological mechanism. As there is room to be improved in monitoring the hemodynamic status in liver transplant surgery [5, 46, 42], we expect that the proposed waveform analysis would lead to more insights into the hemodynamic status of liver transplant surgery to improve the patient's outcome. We will report research in this direction in our future work. Certainly, the source of knowledge base is not limited to the ABP waveform. Different physiological waveforms can be considered to further enrich the knowledge base. How to simultaneously utilize multimodal physiological waveforms, particularly when the recording is long, is a relatively white area, and we expect that the proposed waveform analysis would form a base toward this goal. We also expect that the similar principal could be applied to study other medical datasets for different medical problems, for example, the long-term outcome of the patient underwent organ transplantation with respect to the immune function, or the genetic predisposition and environment factors with respect to the cancer occurrence.

6. Discussion

6.1. **More related work − scalability and robustness.** The recent paper [10] contains a comprehensive review of numerical acceleration techniques for nonlinear dimension reduction, and we refer readers with interest to that work. To handle scalability, an intuitive approach is accelerating the kNN search step. See [10] for a summary and a recently proposed randomized kNN approach [29]. However, it is well known that the kNN scheme is not robust when the dataset is noisy when the neighboring information is not provided. Specifically, it is challenging to estimate pairwise distance robustly, unless we have extra structure to design a robust metric, for example, in the image analysis [8]. If the tangent plane is known, it can help us determine neighbors [47]; however, when the dataset is noisy, the local principle component analysis approach to estimate the tangent space is biased [21]. In short, the kNN is only useful when we have an accurate information about the neighbors.

Another natural approach to handle scalability is accelerating the eigen-decomposition step. For example, we can approximate the kernel decomposition by classical iteration-based algorithms [18]. We can also evaluate the matrix decomposition by designing a randomized algorithm [37].

For the robustness issue, one naive idea is "denoising" the dataset before applying any algorithm. However, it is in general an independent challenging problem. Under the manifold setup, researchers have proposed several algorithms to denoise the dataset. For example, the "reverse diffusion" scheme [20] and the manifold fitting scheme [13]. We mention that the algorithm might not be scalable, but not too much is known at this moment. Another approach is modifying the random walk scheme to a non-lazy random walk via diffusion to obtain a self-consistency Markov chain. But it is under the assumption that the edge information is known [44], which is not possible in many applications. To our knowledge, the general theory for the robustness of kernel methods was first studied in [11], and the analysis was extended to the large noise setup [12]. The authors proved that the spectral embedding methods can be efficiently stabilized by forcing the random walk to be a non-lazy one on the complete graph. Unfortunately, while it could help stabilize the noise impact, the algorithm is not scalable.

6.2. **Relationship with the alternating diffusion algorithms.** Note that the Roseland is related to the recently developed alternating DM (ADM) algorithm [24]. The ADM is developed to deal with the sensor fusion problem, when we have multiple data-sets simultaneously acquired by multimodal sensors. In short, suppose we have 2 aligned data sets $\mathcal{X}^{(1)} = \{x_i^{(1)}\}_{i=1}^n$ and $\mathcal{X}^{(2)} = \{x_i^{(2)}\}_{i=1}^n$, then we build two transition matrices $K^{(j)} = D^{-1}W$ from $\mathcal{X}^{(j)}$, where $j = 1, 2$. Next the alternating diffusion operator is defined as $K^{(1)}K^{(2)}$. It is easy to check that $K^{(1)}K^{(2)}$ is also row stochastic, so it can be considered as a transition probability matrix of a new Markov chain that alternates between the two data sets. The idea of the Roseland is closely related to the ADM, in the sense that we can consider the landmark set as a new data set on its own right. So, the idea of measuring similarity between two data points via the landmark set can be understood as diffusing between the original data set and the landmark set from the ADM point of view.

6.3. **Application of the Roseland.** The idea of landmark set have several applications. Here we mention two of them. The VDM [43] is a generalization of

DM that aims to encode the group structure when comparing objects. The VDM suffers from the expensive computational cost more than the DM, since the group structure is usually represented as a matrix, which inflates the matrix size. Specifically, if the group structure is represented as a $q \times q$ matrix and we have n objects to compare, then we need to eigendecompose a $nq \times nq$ kernel matrix in the VDM. We expect the landmark idea can be generalized to accelerate the VDM. We will explore this possibility in our future work.

Spectral clustering methods are known to perform well when the classical clustering methods such as k-means and linkage fail [1]. It is well known that the more clusters we need to determine, the more eigenvectors we need [1, 25]. As is shown in the numerical section, the Roseland has the ability to recover more and better eigenvectors, at least compared with the Nystöm extension. This shows the potential of applying the Roseland for the multiway spectral clustering purpose.

Acknowledgement

The authors acknowledge Dr. Shen-Chih Wang for the fruitful discussion and suggestion. The work of Yu-Ting Lin was supported by the National Science and Technology Development Fund (MOST 108-2115-M-075-001) of Ministry of Science and Technology, Taipei, Taiwan.

References

[1] C. Alzate and J.A.K. Suykens. Multiway spectral clustering with out-of-sample extensions through weighted kernel pca. *IEEE transactions on pattern analysis and machine intelligence*, 32(2):335–347, 2008.

[2] A. P. Avolio, L. M. Van Bortel, P. Boutouyrie, and others. Role of pulse pressure amplification in arterial hypertension: experts opinion and review of the data. *Hypertension*, 54(2):375–383, 2009.

[3] M.A. Belabbas and P.J. Wolfe. On landmark selection and sampling in high-dimensional data analysis. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367(1906):4295–4312, 2009.

[4] P. Bérard, G. Besson, and S. Gallot. Embedding riemannian manifolds by their heat kernel. *Geometric & Functional Analysis GAFA*, 4(4):373–398, 1994.

[5] G. Biancofiore, L.A.H. Critchley, A. Lee, X.-x. Yang, L. M. Bindi, M. Esposito, M. Bisà, L. Meacci, R. Mozzo, and F. Filipponi. Evaluation of a new software version of the flotrac/vigileo (version 3.02) and a comparison with previous data in cirrhotic patients undergoing liver transplant surgery. *Anesthesia & Analgesia*, 113(3):515–522, 2011.

[6] L.B. Chang, Z. Bai, S.Y. Huang, and C.R. Hwang. Asymptotic error bounds for kernel-based nyström low-rank approximation matrices. *Journal of Multivariate Analysis*, 120:102–119, 2013.

[7] C.-H. Chen, E. Nevo, B. Fetics, P. H. Pak, F. C.P. Yin, W. L. Maughan, and D. A. Kass. Estimation of central aortic pressure waveform by mathematical transformation of radial tonometry pressure: validation of generalized transfer function. *Circulation*, 95(7):1827–1836, 1997.

[8] B. Cheng, J. Yang, S. Yan, Y. Fu, and T.S. Huang. Learning with ℓ^1-graph for image analysis. *IEEE transactions on image processing*, 19(4):858–866, 2009.

[9] R.R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21(1):5–30, 2006.

[10] W. Czaja, T. Doster, and A. Halevy. An overview of numerical acceleration techniques for nonlinear dimension reduction. In *Recent Applications of Harmonic Analysis to Function Spaces, Differential Equations, and Data Science*, pages 797–829. Springer, 2017.

[11] N. El Karoui. On information plus noise kernel random matrices. *Ann. Statist.*, 38(5):3191–3216, 2010.

[12] N. El Karoui and H.-T. Wu. Graph connection laplacian methods can be made robust to noise. *The Annals of Statistics*, 44(1):346–372, 2016.

[13] C. Fefferman, S. Ivanov, Y. Kurylev, M. Lassas, and H. Narayanan. Fitting a putative manifold to noisy data. In *Conference On Learning Theory*, pages 688–720, 2018.

[14] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the nystrom method. *IEEE transactions on pattern analysis and machine intelligence*, 26(2):214–225, 2004.

[15] N. Gambarotta, F. Aletti, G. Baselli, and M. Ferrario. A review of methods for the signal quality assessment to improve reliability of heart rate and blood pressures derived parameters. *Medical & biological engineering & computing*, 54(7):1025–1035, 2016.

[16] S. Günter, N.N. Schraudolph, and S.V.N. Vishwanathan. Fast iterative kernel principal component analysis. *Journal of Machine Learning Research*, 8(Aug):1893–1918, 2007.

[17] S. S. Halabi, L. M. Prevedello, J. Kalpathy-Cramer, A. B. Mamonov, A. Bilbily, M. Cicero, I. Pan, L. A. Pereira, R. T. Sousa, and N. Abdala. The rsna pediatric bone age machine learning challenge. *Radiology*, 290(2):498–503, 2019.

[18] N. Halko, P.G. Martinsson, and J.A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.

[19] F. Hatib, Z. Jian, S. Buddi, C. Lee, J. Settels, K. Sibert, J. Rinehart, and M. Cannesson. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology*, 129(4):663–674, 2018.

[20] M. Hein and M. Maier. Manifold denoising. In B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 561–568. MIT Press, 2007.

[21] I. M. Johnstone. High dimensional statistical inference and random matrices. In *Proceedings of the International Congress of Mathematicians Madrid, August 22–30, 2006*, pages 307–333, 2007.

[22] S. Lafon, Y. Keller, and R.R. Coifman. Data fusion and multicue data matching by diffusion maps. *IEEE Transactions on pattern analysis and machine intelligence*, 28(11):1784–1797, 2006.

[23] Y. LeCun and C. Cortes. MNIST handwritten digit database. 2010.

[24] R.R. Lederman and R. Talmon. Learning the geometry of common latent variables using alternating-diffusion. *Applied and Computational Harmonic Analysis*, 44(3):509 – 536, 2018.

[25] J.R. Lee, S.O. Gharan, and L. Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *Journal of the ACM (JACM)*, 61(6):37, 2014.

[26] Y. Li, J. Huang, and W. Liu. Scalable sequential spectral clustering. In *Thirtieth AAAI conference on artificial intelligence*, 2016.

[27] T-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C.L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[28] Y.-T. Lin, J. Malik, and H.-T. Wu. Wave-shape oscillatory model for biomedical time series with applications. *arXiv preprint arXiv:1907.00502*, 2019.

[29] G. C. Linderman, G. Mishne, Y. Kluger, and S. Steinerberger. Randomized near neighbor graphs, giant components, and applications in data science. *arXiv preprint arXiv:1711.04712*, 2017.

[30] J. L. Marcus, L. B. Hurley, D. S. Krakower, S. Alexeeff, M. J. Silverberg, and J. E. Volk. Use of electronic health record data and machine learning to identify candidates for hiv pre-exposure prophylaxis: a modelling study. *The Lancet HIV*, 6(10):e688–e695, 2019.

[31] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[32] P.G. Martinsson, V. Rokhlin, and M. Tygert. A randomized algorithm for the decomposition of matrices. *Applied and Computational Harmonic Analysis*, 30(1):47–68, 2011.

[33] G. F. Mitchell, S.-J. Hwang, R. S. Vasan, M. G. Larson, M. J. Pencina, N. M. Hamburg, J. A. Vita, D. Levy, and E. J. Benjamin. Arterial stiffness and cardiovascular events: the framingham heart study. *Circulation*, 121(4):505, 2010.

[34] M. F. O'Rourke and W. W. Nichols. Changes in wave reflection with advancing age in normal subjects. *Hypertension (Dallas, Tex.: 1979)*, 44(6):e10–1, 2004.

[35] M. Ouimet and Y. Bengio. Greedy spectral embedding. In *AISTATS*. Citeseer, 2005.

[36] J.W. Portegies. Embeddings of riemannian manifolds with heat kernels and eigenfunctions. *Communications on Pure and Applied Mathematics*, 69(3):478–518, 2016.

[37] V. Rokhlin, A. Szlam, and M. Tygert. A randomized algorithm for principal component analysis. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1100–1124, 2009.

[38] C. F. Rowlands, D. Baralle, and J. M. Ellingford. Machine learning approaches for the prioritization of genomic variants impacting pre-mrna splicing. *Cells*, 8(12):1513, 2019.

[39] M. R. Rudnick, L. De Marchi, and J. S. Plotkin. Hemodynamic monitoring during liver transplantation: A state of the art review. *World journal of hepatology*, 7(10):1302, 2015.

[40] C. Shen and H.-T. Wu. Scalability and robustness of spectral embedding: landmark diffusion is all you need. *arXiv preprint arXiv:2001.00801*, 2020.

[41] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

[42] B-F Shih, P-H Huang, H-P Yu, F-C Liu, C-C Lin, PC-H Chung, C-Y Chen, C-J Chang, and Y-F Tsai. Cardiac output assessed by the fourth-generation arterial waveform analysis system is unreliable in liver transplant recipients. In *Transplantation proceedings*, volume 48, pages 1170–1175. Elsevier, 2016.

[43] A. Singer and H.-T. Wu. Vector diffusion maps and the connection laplacian. *Communications on pure and applied mathematics*, 65(8):1067–1144, 2012.

[44] S. Steinerberger. A Filtering Technique for Markov Chains with Applications to Spectral Embedding. *Applied and Computational Harmonic Analysis*, 40:575–587, 2016.

[45] J.-L. Teboul, B. Saugel, M. Cecconi, et al. Less invasive hemodynamic monitoring in critically ill patients. *Intensive care medicine*, 42(9):1350–1359, 2016.

[46] Y-F Tsai, B-C Su, C-C Lin, F-C Liu, W-C Lee, and H-P Yu. Cardiac output derived from arterial pressure waveform analysis: validation of the third-generation software in patients undergoing orthotopic liver transplantation. In *Transplantation proceedings*, volume 44, pages 433–437. Elsevier, 2012.

[47] J. Wang, Z. Zhang, and H. Zha. Adaptive manifold learning. In *Advances in neural information processing systems*, pages 1473–1480, 2005.

[48] S.-C. Wang, H.-T. Wu, P.-H. Huang, C.-H. Chang, C.-K. Ting, and Y.-T. Lin. Novel imaging revealing inner dynamics for cardiovascular waveform analysis via unsupervised manifold learning. *Anesthesia & Analgesia*, 130(5):1244–1254, 2020.

[49] Y.-Y L. Wang, T.-L. Hsu, M.-Y. Jan, W.-K. Wang, et al. Theory and applications of the harmonic analysis of arterial pressure pulse waves. *Journal of Medical and Biological Engineering*, 30(3):125–131, 2010.

[50] C.K.I. Williams and M. Seeger. Using the nyström method to speed up kernel machines. In *Advances in neural information processing systems*, pages 682–688, 2001.

[51] D. Yan, L. Huang, and M.I. Jordan. Fast approximate spectral clustering. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 907–916. ACM, 2009.

Department of Mathematics, Duke University, Durham, NC, USA

Department of Anesthesiology, Taipei Veteran General Hospital, Taipei, Taiwan

Department of Mathematics, Duke University, Durham, NC, USA; Department of Statistical Science, Duke University, Durham, NC, USA; Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan

E-mail address: hauwu@math.duke.edu

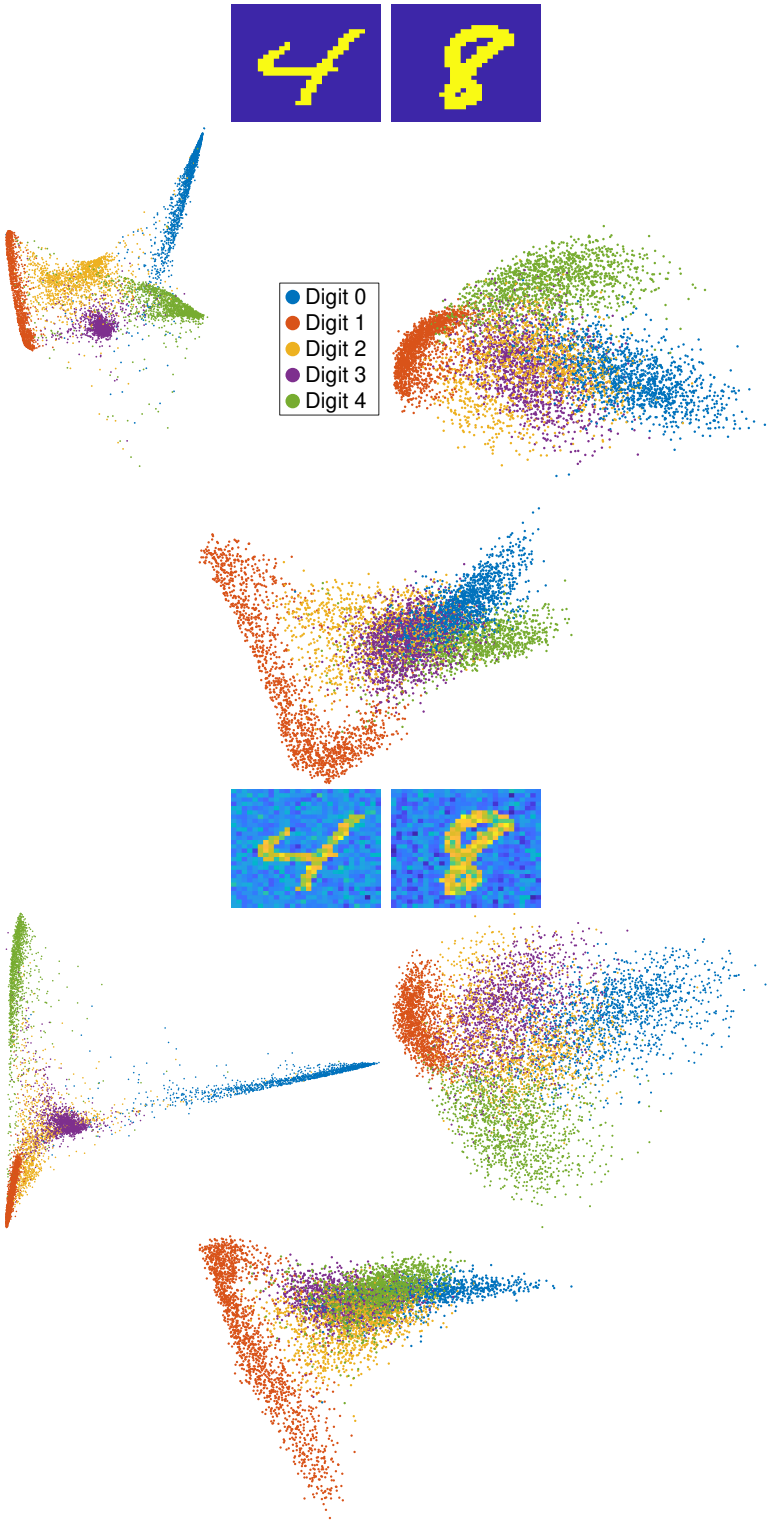

FIGURE 2. First row: Examples of clean digits. Second row: embeddings of digits 0 to 4 using clean MNIST, from left to right: the DM embedding, the Nyström extension embedding, and the Roseland embedding. Third row: Examples of noisy digits. Fourth row: embeddings of digits 0 to 4 using noisy MNIST, from left to right: the DM embedding, the Nyström extension embedding use clean landmars, and the Roseland embedding use clean landmarks.

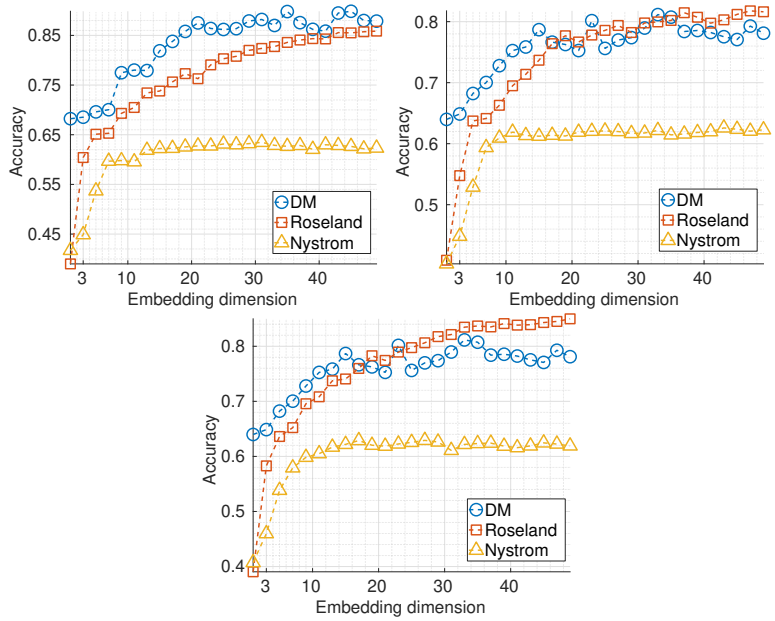

FIGURE 3. Accuracies of three spectral clustering. From left to right: clean MNIST, noisy MNIST with noisy landmarks, noisy MNIST with clean landmarks.

FIGURE 4. Top row: first and second images are from the Berkeley Segmentation Dataset and Benchmark, of size 481×321, the third one is from COCO, of size 582×640. Bottom row: Image segmentations by spectral clustering based on Roseland using three eigenvectors.

FIGURE 5. Top row: first and second images are from the Berkeley Segmentation Dataset and Benchmark, of size 481×321, the third one is from COCO, of size 582×640 corrupted by Gaussian noise. Bottom row: Image segmentations by spectral clustering based on Roseland using three eigenvectors.

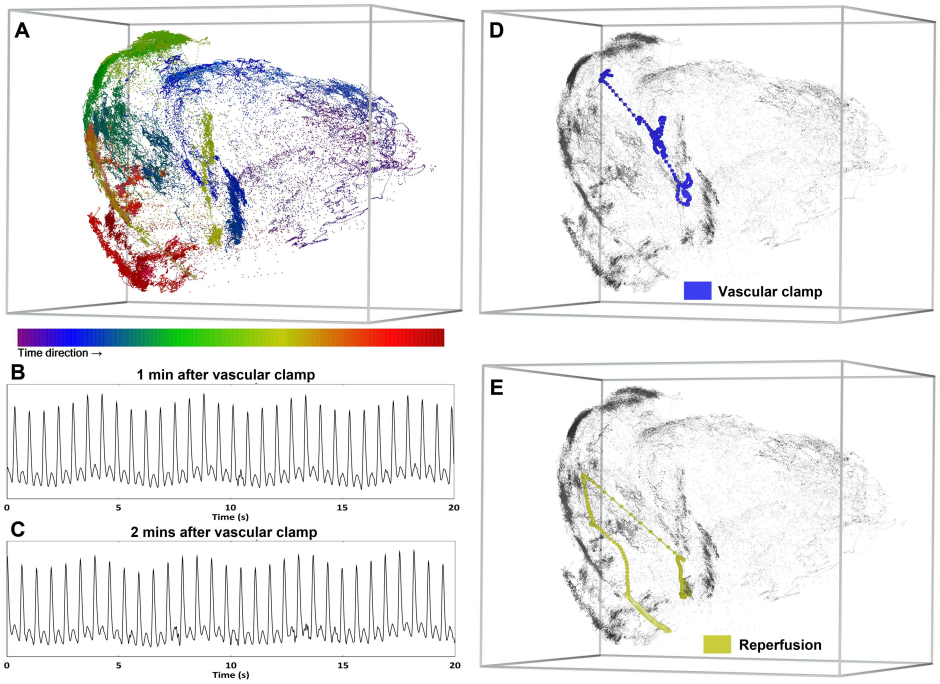

FIGURE 6. The 3D embedding (panel A) of pulse-to-pulse pressure waveforms collected from a 78,350s ABP signal (120,725 pulses) during the liver transplantation procedure. The embedded pulses are labeled by colors encoding the time. The color helps visualize the ever-changing trajectory formed from successive pulse waveforms. The embedding is clustered and different clusters are related to different stages. During transition phases of inferior vena cava cross clamping (panel B, C), the ABP tracings provide little clues with respect to the subtle waveform information and its long-term evolving, while the Roseland algorithm reveals the fast paced movements (panel D). To signify the physiological dynamics associated with the vascular clamp (panel D) and reperfusion (panel E) events, pulses in transition phases are labeled with colored linked dots while the rest pulses are not colored. In panels A, D, and E, the grids are drawn to enhance the 3D visualization, and an online supplementary video is provided for more details.

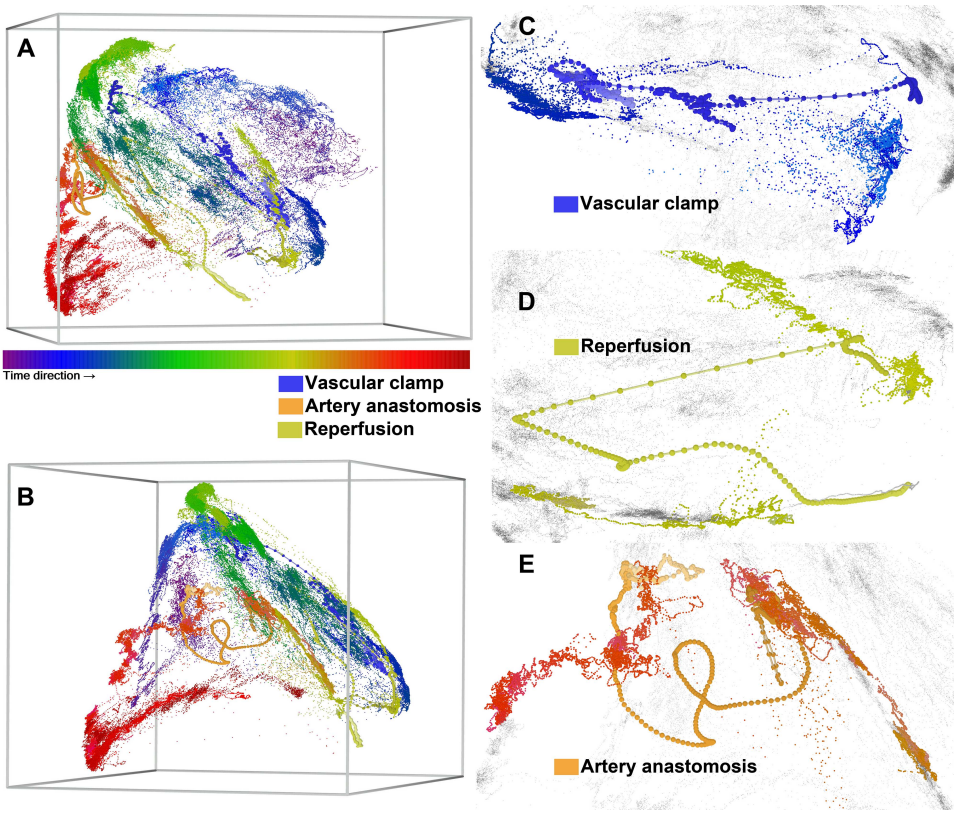

FIGURE 7. The 3D embedding highlights the locations of transition phases (in linked dot) in relation to the whole period of ABP waveforms (smaller dot) in a liver transplant surgery. The color labels the time sequence. Panel B is the horizontally rotated Panel A by 60 degrees for a better visualization of the artery anastomosis transition. These views show geographic relationship between surgical stages and transition phases. Zoom-in views of the 3D embedding of pulse-to-pulse ABP waveform include transition phases of major vascular cross-clamping (panel A), new liver graft reperfusion (panel B), and hepatic artery anastomosis (panel C), which shows fast paced movement in transition phases (enlarged colored dots with line-link). On the other hand, in the statuses immediate before and after transition (colored dots without line-link), we see less movement and the embeddings are clustered. The rest pulses (uncolored small dots) appears in the background.