

Robust and scalable manifold learning via landmark diffusion for long-term medical signal processing

Chao Shen

CHAO.SHEN@DUKE.EDU

*Department of Mathematics
Duke University
Durham, NC 27708-0320, USA*

Yu-Ting Lin

LINYUTING@HOTMAIL.COM.TW

*Department of Anesthesiology
Taipei Veterans General Hospital
School of Medicine
National Yang Ming Chiao Tung University Taipei, Taiwan.*

Hau-Tieng Wu

HAUWU@MATH.DUKE.EDU

*Department of Mathematics and Department of Statistical Science
Duke University
Durham, NC 27708-0320, USA*

Editor: Benjamin Marlin

Abstract

Motivated by analyzing long-term physiological time series, we design a robust and scalable spectral embedding algorithm that we refer to as ROBust and Scalable Embedding via LANdmark Diffusion (Roseland). The key is designing a diffusion process on the dataset where the diffusion is done via a small subset called the *landmark set*. Roseland is theoretically justified under the manifold model, and its computational complexity is comparable with commonly applied subsampling scheme such as the Nyström extension. Specifically, when there are n data points in \mathbb{R}^d and n^β points in the landmark set, where $\beta \in (0, 1)$, the computational complexity of Roseland is $O(n^{1+2\beta} + qn^{1+\beta})$, while that of Nystrom is $O(n^{2.81\beta} + qn^{1+2\beta})$. To demonstrate the potential of Roseland, we apply it to three datasets and compare it with several other existing algorithms. First, we apply Roseland to the task of spectral clustering using the MNIST dataset (70,000 images), achieving 85% accuracy when the dataset is clean and 78% accuracy when the dataset is noisy. Compared with other subsampling schemes, overall Roseland achieves a better performance. Second, we apply Roseland to the task of image segmentation using images from COCO. Finally, we demonstrate how to apply Roseland to explore long-term arterial blood pressure waveform dynamics during a liver transplant operation lasting for 12 hours. In conclusion, Roseland is scalable and robust, and it has a potential for analyzing large datasets.

1. Introduction

Learning from data has been an intriguing topic in many scientific fields, particularly the biomedical field. While there are many different types of datasets in the biomedical field, in this paper we focus on the biomedical waveforms (or time series, signals); for example, the electrocardiogram (Barrett et al. (2014)), the arterial blood pressure signal (Vlachopoulos et al. (2011)), etc. It has been well known that rich information is available in these signals, particularly the nonlinear dynamics.

However, to our knowledge, how to extract useful information from biomedical waveforms for clinical usage is relatively less discussed compared with other types of datasets, and most of existing literature on the biomedical signal processing focus on simplifying the waveform information into scalars for clinical purposes (Chen et al. (1997); O’Rourke and Nichols (2004); Avolio et al. (2009); Rudnick et al. (2015); Teboul et al. (2016)). While this has been successfully applied to clinical medicine, we may lose information encoded in the original waveform via the simplification steps. 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 (Hatib et al. (2018)). Another solution is applying manifold learning algorithms to the *original* physiological waveforms (Lin et al. (2021a); Wang et al. (2020)). The basic idea in (Lin et al. (2021a); Wang et al. (2020)) is truncating the physiological waveform into pieces according to some rules, and then apply the spectral embedding algorithm, like the diffusion maps (DM) (Coifman and Lafon (2006)), 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.

We hypothesize that if the technique shown in Lin et al. (2021a); Wang et al. (2020) could be efficiently applied to analyze long-term physiological waveforms of length on the order of days or weeks, it would be beneficial to help recognize fine structures/patterns that are undetectable even to the human eyes. Note that while the idea in Lin et al. (2021a); Wang et al. (2020) is natural and potential, unfortunately, it is not possible to further explore its clinical applicability and apply it to time series of lengths actually encountered in real clinical data sets due to its poor computational scalability. Specifically, due to the prohibiting computational complexity inherited in most spectral decomposition based machine learning algorithms, like DM applied in Lin et al. (2021a); Wang et al. (2020), we are not able to efficiently apply these algorithms to study an ABP waveform lasting for 12 hours or longer.

1.1 Challenges and related work – scalability and robustness

Spectral based nonlinear dimension reduction algorithms are challenged by the scalability issue, for example, the diffusion map (DM) (Coifman and Lafon (2006)). Take electrocardiogram (ECG) into account. Suppose we are interested in studying the dynamics of the electrophysiology of each heartbeat. Note that there are $10^3 \sim 10^4$ heartbeats in 1 hour long ECG, and hence roughly 10^6 heartbeats in 14 days. Therefore, the amount of data points cannot be handled efficiently by naive spectral based algorithms. In practice, it is natural to consider subsampling the dataset. However, we may lose information by the subsampling scheme. As a result, although DM works well and provides useful clinical information, for example, a visualization of fine details hidden in the time series that are undetectable to the human eye (Wang et al. (2020)), since it depends on the eigendecomposition, it is limited to analyzing signals of length about one hour that is sampled at high frequency, like 500Hz.

There have been several solutions toward this scalability challenge. One usual technique is the k-nearest neighbor (kNN) scheme. See Czaja et al. (2017) for a summary and a recently proposed randomized kNN approach (Jaffe et al. (2020)). While the kNN scheme has been widely applied, in general it is not robust to noise. 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 (Cheng

et al. (2009)). If the tangent plane is known, it can help us determine neighbors (Wang et al. (2005)); however, when the dataset is noisy, the local principle component analysis approach to estimate the tangent space is biased (Johnstone (2007)). In short, the kNN scheme is useful when we have a clean dataset, or when neighbors can be correctly determined. Otherwise, erroneous nearest neighbors due to noises could hurt the algorithm performance.

Another practical solution is directly subsampling the dataset, and then recovering the information of interest by the *Nyström extension* (Czaja et al. (2017)). This approach is also called the *Nyström low-rank approximation* (Chang et al. (2013)), the *kernel extension method* (Fowlkes et al. (2004)), or in general the *interpolative decomposition* (Martinsson et al. (2011)). There have been several theoretical justifications for this approach, for example Chang et al. (2013), and has been widely applied. While it works well for some tasks, this approach is limited by the information loss during the subsampling process.

Yet another approach is to speed up the matrix decomposition. For example, we can approximate the kernel decomposition by classical iteration-based algorithms (Halko et al. (2011)). We can also evaluate the matrix decomposition by designing a randomized algorithm (Rokhlin et al. (2009); Martinsson et al. (2011)). While this approach has been widely applied, to the best of our knowledge, there is limited work investigating how this can help spectral embedding algorithms, and how robust it is to the inevitable noise.

For the robustness issue, one naive idea is “denoising” the dataset before applying any algorithms. 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 (Hein and Maier (2007)) and the manifold fitting scheme (Fefferman et al. (2018)). We mention that the algorithm is not scalable, as is mentioned in the paper that it is hard to choose the number of iterations if the data and noise are of high dimensional. Also, one needs some prior knowledge. Another approach is to modify 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 (Steinerberger (2016)), which is not possible in some applications. To our knowledge, the general theory for the robustness of kernel methods was first studied in El Karoui (2010), and the analysis was extended to the large noise setup (El Karoui and Wu (2016)). 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; that is, removing the diagonal entries from the affinity matrix before constructing the transition matrix. Unfortunately, while it could help stabilize the noise impact, the algorithm is not scalable since we still need to handle a dense transition matrix associated with a complete graph.

1.2 Our contribution

Unlike the above, in this paper we propose a novel algorithm that simultaneously 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, so we coin the proposed algorithm the *RObust and Scalable Embedding via LANdmark Diffusion* (Roseland). We will show that the proposed Roseland algorithm can be applied to efficiently and robustly handle large datasets that cannot be efficiently handled by the traditional algorithm like DM. We emphasize that the solution is generic and is not limited to analyzing physiological waveforms.

In addition to comparing Roseland with other algorithms in the spectral clustering and image segmentation problems, we demonstrate how to apply Roseland to explore the long-term arterial blood pressure waveform dynamics during a liver transplant operation lasting longer than 12 hours. For the sake of self-containedness, we also summarize and explain the theoretical results shown in Shen and Wu (2022), particularly the spectral convergence and robustness of Roseland under the manifold setup.

2. The proposed Roseland algorithm

Before introducing the proposed Roseland algorithm, we start with some well developed algorithms that are closely related to Roseland from different angles, including DM, Nyström extension, and locally linear landmarks (LLL). In this section, we assume that we have a dataset $\mathcal{X} = \{x_i\}_{i=1}^n \subseteq (\mathfrak{M}, d)$, where \mathfrak{M} is a metric space with the metric d . Also, take a set $\mathcal{Y} = \{y_k\}_{k=1}^m \subseteq (\mathfrak{M}, d)$, which might or might not be a subset of \mathcal{X} . In Roseland and LLL, we call \mathcal{Y} the *landmark set*. In Nyström extension, $\mathcal{Y} \subset \mathcal{X}$.

2.1 Related methods

2.1.1 DIFFUSION MAPS

We start with the well-known DM (Coifman and Lafon (2006)), which we shall see to be closely related to Roseland. First, pre-fix a non-negative kernel function K and a bandwidth parameter $\epsilon > 0$; for example, a Gaussian kernel. Then, compute the affinity matrix $W \in \mathbb{R}^{n \times n}$ by

$$W_{ij} := K\left(\frac{d(x_i, x_j)}{\sqrt{\epsilon}}\right) \quad (1)$$

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 $W^{(\alpha)} \in \mathbb{R}^{n \times n}$ (Coifman and Lafon (2006)) 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 (1). 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 $D_{ii}^{(\alpha)} := \sum_{j=1}^n W_{ij}^{(\alpha)}$. $D_{ii}^{(\alpha)}$ is called the degree of the i -th data point x_i . Intuitively, it represents how strong x_i is attached to the dataset. The graph Laplacian (GL) is defined as $L^{(\alpha)} := I - A^{(\alpha)}$, where

$$A^{(\alpha)} := (D^{(\alpha)})^{-1} W^{(\alpha)} \quad (2)$$

is the associated transition matrix. Clearly, $A^{(\alpha)}$ is row stochastic, and it defines a random walk on the dataset \mathcal{X} . Now let $\tilde{U}^{(\alpha)} \in \mathbb{R}^{n \times n}$ be eigenvectors of $A^{(\alpha)}$ with corresponding eigenvalues $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$.

With the spectral decomposition of the GL, the chosen normalization α , embedding dimension q' and *diffusion time* t , DM embeds \mathcal{X} via the map

$$\Phi_t : x_i \mapsto e_i^\top \tilde{U}_{q'}^{(\alpha)} (\Lambda_{q'}^{(\alpha)})^t \in \mathbb{R}^{q'}, \quad (3)$$

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 $\Lambda_{q'}^{(\alpha)} := \text{diag}(\lambda_2, \dots, \lambda_{q'+1})$. The diffusion distance (DD) with the diffusion time $t > 0$ is defined as

$$D_t(x_i, x_j) := \|\Phi_t(x_i) - \Phi_t(x_j)\|_{\mathbb{R}^{q'}}. \quad (4)$$

Note that in DM, we do not consider any subset \mathcal{Y} .

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 $t > 1$.

2.1.2 NYSTRÖM EXTENSION

Subsampling is a common approach to handle a large dataset. A typical and widely applied algorithm in this direction is the Nyström extension (Belabbas and Wolfe (2009); Fowlkes et al. (2004); Williams and Seeger (2001)). The basic idea is running the eigen-decomposition on a small subset \mathcal{Y} of the whole dataset \mathcal{X} , and then extending the eigenvectors to \mathcal{X} before embedding the dataset. In this work, for the sake of a fair comparison, we apply the Nyström extension (Lafon et al. (2006); Singer and Wu (2012); Shen and Wu (2022)) that respects the diffusion property considered in DM but not the vanilla Nyström extension. For the sake of self-containedness, we provide details about this Nyström extension algorithm. First, apply DM to \mathcal{Y} and obtain eigenvectors $\tilde{U}_{\mathcal{Y}}^{(\alpha)} \in \mathbb{R}^{m \times m}$ with corresponding eigenvalues $1 = l_1 > l_2 \geq \dots \geq l_m$. Set $\Lambda_{\mathcal{Y}}^{(\alpha)} = \text{diag}[l_1 \dots l_m] \in \mathbb{R}^{m \times m}$. Next, calculate $E \in \mathbb{R}^{(n-m) \times m}$, where E_{ij} is the similarity between the point $x_i \in \mathcal{X} \setminus \mathcal{Y}$ and the point $y_j \in \mathcal{Y}$. Let D_{n-m} be the degree matrix of E . With E and D_{n-m} , extend $\tilde{U}_{\mathcal{Y}}^{(\alpha)}$ to the rest $n - m$ points in $\mathcal{X} \setminus \mathcal{Y}$ by:

$$\tilde{U}_{\mathcal{X} \setminus \mathcal{Y}}^{(\alpha)} = D_{n-m}^{-1} E \tilde{U}_{\mathcal{Y}}^{(\alpha)} (\Lambda_{\mathcal{Y}}^{(\alpha)})^{-1} \in \mathbb{R}^{(n-m) \times m}. \quad (5)$$

The column vectors in the matrix

$$\tilde{U}_{\text{nys}}^{(\alpha)} = \begin{bmatrix} \tilde{U}_{\mathcal{Y}}^{(\alpha)} \\ \dots \\ \tilde{U}_{\mathcal{X} \setminus \mathcal{Y}}^{(\alpha)} \end{bmatrix} \in \mathbb{R}^{n \times m} \quad (6)$$

are used to embed the whole dataset in the following way. Set $\Lambda_{\text{nys}}^{(\alpha)} = \text{diag}[l_1 \dots l_m \ 0 \dots 0] \in \mathbb{R}^{n \times n}$. For $q' < m$, the associated Nyström embedding is then defined by

$$\Phi_t^{(\text{Nyström})} : x_i \mapsto e_i^\top \tilde{U}_{\text{nys}, q'}^{(\alpha)} (\Lambda_{\text{nys}, q'}^{(\alpha)})^t, \quad (7)$$

where $t > 0$ is the chosen diffusion time, $\tilde{U}_{\text{nys}, q'}^{(\alpha)} \in \mathbb{R}^{n \times q'}$ comes from the second to the $(q' + 1)$ -th columns of $\tilde{U}_{\text{nys}}^{(\alpha)}$, and $\Lambda_{\text{nys}, q'}^{(\alpha)} = \text{diag}[l_2 \dots l_{q'+1}] \in \mathbb{R}^{q' \times q'}$. Note that $\tilde{U}_{\text{nys}, q'}^{(\alpha)}$ and $\Lambda_{\text{nys}, q'}^{(\alpha)}$ are in parallel of the top q' non-trivial eigen-pairs used in DM. In short, in this Nyström extension algorithm, a point $x \in \mathcal{X} \setminus \mathcal{Y}$ is embedded by averaging all of the embeddings in \mathcal{Y} with a weight determined by the similarity between x and \mathcal{Y} .

2.1.3 LOCALLY LINEAR LANDMARKS

Working with a subset of the whole database, called a landmark or a reference set, is another common approach when the dataset is large. We consider the Locally Linear Landmarks (LLL) (Vladymyrov and Carreira-Perpinán (2013)) as an example for comparison in this work. In this case, $\mathfrak{M} = \mathbb{R}^q$ and $d(x_i, x_j) = \|x_i - x_j\|_{\mathbb{R}^q}$; that is, the dataset and the landmark set are both in the Euclidean space. The main idea of LLL is approximating the dataset by a globally nonlinear but locally linear geometric structure, which is assumed to be a manifold in Vladymyrov and Carreira-Perpinán (2013), around the chosen landmark set, and then constraining the solution to follow this locally linear structure as a linear combination of some points in the chosen landmark set. Explicitly, solve

$$\arg \min_{Z \in \mathbb{R}^{n \times m}, Z\mathbf{1}=\mathbf{1}} \|ZY - X\|^2,$$

where $X \in \mathbb{R}^{n \times q}$ is the data matrix formed from \mathcal{X} , $Y \in \mathbb{R}^{m \times q}$ is the landmark data matrix formed from \mathcal{Y} , and Z consists of the weights of linear combination, which is suggested to be sparse so that only the closest $K_Z \in \mathbb{N}$ points in the landmark set is chosen. Here K_Z is usually assumed to be smaller than the size of \mathcal{Y} . Z can be determined by solving the linear system $\sum_{k=1}^m (x_n - y_i)^\top (x_n - y_j) z_{nk}$, and then normalize to have sum 1.

Next, preprocess the dataset and find the affinity matrix $W \in \mathbb{R}^{n \times n}$ defined in (1) and its corresponding degree matrix $D \in \mathbb{R}^{n \times n}$. Note that while a more general setup is possible (See Equation 1 in Vladymyrov and Carreira-Perpinán (2013)), we focus on this commonly applied framework. Unlike performing the eigen-decomposition of $D^{-1}W$ in DM, instead, one runs the eigen-decomposition of the matrix $\hat{D}^{-1}\hat{W}$ in LLL, where $\hat{W} = Z^\top WZ \in \mathbb{R}^{m \times m}$ and $\hat{D} = Z^\top DZ \in \mathbb{R}^{m \times m}$. Let $\hat{U} \in \mathbb{R}^{m \times m}$ be the solution to this smaller problem, where the eigenvectors in \hat{U} are ordered by the corresponding eigenvalues. One can then extend it to the whole set by setting $U = Z\hat{U} \in \mathbb{R}^{n \times m}$. For any downstream tasks, one can use \hat{U}_d which consists of the $d \leq m$ trailing eigenvectors in \hat{U} associated with the smallest eigenvalues. We have some remarks. First, since $\hat{W}\mathbf{1} = Z^\top WZ\mathbf{1} = Z^\top W\mathbf{1} = Z^\top D \neq \hat{D}$, in LLL, $\hat{D}^{-1}\hat{W}$ is not a transition matrix. Second, LLL uses eigenvalues to select eigenvectors, but the eigenvectors are not re-scaled by their corresponding eigenvalues. So the eigenvalues in LLL are used in a different way compared with DM and Nyström extension we considered in this paper. Third, the calculation of the affinity matrix W and the degree matrix D is similar to DM. While it is possible to apply a kNN scheme to speed up the algorithm, this approach is limited, particularly when the metric is not Euclidean.

2.2 The proposed Roseland algorithm

Fix a non-negative kernel function $K : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_+$ with proper decay and regularity; for example, a Gaussian function. First, construct a *landmark-set affinity matrix* $W^{(r)} \in \mathbb{R}^{n \times m}$, which is defined as

$$W_{ik}^{(r)} = K_\epsilon(x_i, y_k) := K\left(\frac{d(x_i, y_k)}{\sqrt{\epsilon}}\right). \quad (8)$$

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

$$D_{ii}^{(R)} := e_i^\top W^{(r)}(W^{(r)})^\top \mathbf{1}, \quad (9)$$

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. Note that the superscripts (r) and (R) are purely symbolic indicating that they are not the usual affinity or degree matrices considered in DM. With $W^{(r)}$ and $D^{(R)}$, we evaluate the SVD of $(D^{(R)})^{-1/2}W^{(r)}$:

$$(D^{(R)})^{-1/2}W^{(r)} = U\Lambda V^\top, \quad (10)$$

where the singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m \geq 0$ are on the diagonal of the diagonal matrix Λ , which is a parallel step of the eigen-decomposition of $A^{(0)}$ in DM. Set $\bar{U} := (D^{(R)})^{-1/2}U$. 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 $L_{q'} := \text{diag}(\sigma_2^2, \dots, \sigma_{q'+1}^2)$. Finally we define the Roseland embedding as

$$\Phi_t^{(R)} : x_i \mapsto e_i^\top \bar{U}_{q'} (L_{q'})^t, \quad (11)$$

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-wisely rescaled by $[\sigma_2^{2t}, \dots, \sigma_{q'+1}^{2t}]$. See Algorithm 1 for a summarization of the Roseland algorithm. We thus define the associated *Roseland diffusion distance* (RDD) by

$$D_t^{(R)}(x_i, x_j) := \|\Phi_t^{(R)}(x_i) - \Phi_t^{(R)}(x_j)\|_{\mathbb{R}^{q'}}. \quad (12)$$

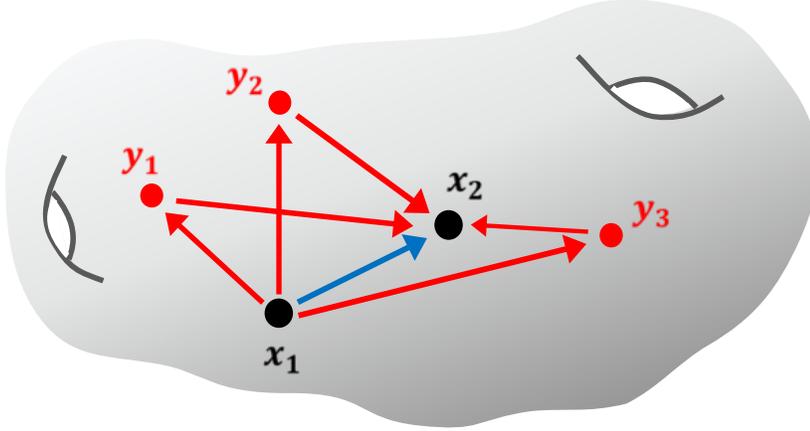


Figure 1: Main idea of Roseland: to measure the similarity between x_1 to x_2 , instead of diffuse from x_1 to x_2 directly, we take a detour and first diffuse x_1 to the landmarks y_1, y_2, y_3 , and then diffuse from the landmarks back to x_2 .

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

$$W^{(R)} := W^{(r)}(W^{(r)})^\top \in \mathbb{R}^{n \times n}, \quad (13)$$

where $W^{(r)}$ is the landmark-set affinity matrix (8). We call $W^{(R)}$ 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 $W_{ij}^{(R)} = \bar{K}(d(x_i, x_j)/\sqrt{\bar{\epsilon}})$ for all i, j in general. Also

Algorithm 1 The pseudo-code of Roseland.

- 1: **procedure** INPUT(dataset $\mathcal{X} = \{x_i\}_{i=1}^n \subset (\mathfrak{M}, d)$, landmark set $\mathcal{Y} = \{y_i\}_{i=1}^m \subset (\mathfrak{M}, d)$, kernel K , bandwidth $\epsilon > 0$, embedding dimension $q' \in \mathbb{N}$, and diffusion time $t > 0$)
 - 2: Construct the affinity matrix $W^{(t)}$ w.r.t. the landmark set.
 - 3: Construct the degree matrix $D^{(R)}$.
 - 4: Run SVD $(D^{(R)})^{-1/2}W^{(t)} = U\Lambda V^\top$, where U is an $n \times n$ orthonormal matrix and V is an $m \times m$ orthonormal matrix. Denote singular values as $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$.
 - 5: Set $\bar{U} = (D^{(R)})^{-1/2}U$.
 - 6: Let $\bar{U}_{q'} \in \mathbb{R}^{n \times q'}$ be the second to the $(q' + 1)$ -th columns of \bar{U} . Set $L_{q'} := \text{diag}(\sigma_2^2, \sigma_3^2, \dots, \sigma_{q'+1}^2)$.
 - 7: Embed \mathcal{X} via $\Phi_t^{(R)} : x_i \mapsto e_i^\top \bar{U}_{q'} L_{q'}^t \in \mathbb{R}^{q'}$
 - 8: Output $\{\Phi_t^{(R)}(x_i)\}_{i=1}^n$.
 - 9: **end procedure**
-

note that $A^{(R)} := (D^{(R)})^{-1}W^{(R)}$ is a transition matrix on \mathcal{X} . Hence, the Roseland algorithm can be viewed as an alternative way of constructing a Markov process on the dataset \mathcal{X} .

In practice, we would suggest to use Roseland when DM cannot be efficiently carried out, and choose $m = n^\beta$ for some $\beta < 1/2$; for example, $\beta = 0.45$. The landmark set \mathcal{Y} could be chosen uniformly from the dataset \mathcal{X} or collected separately. 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, note that \mathcal{Y} is independent of $\mathcal{X} \setminus \mathcal{Y}$. Thus, we apply 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.

2.3 Computational complexity

Let n be the size of the dataset, q be the dimension of the ambient space, and $[n^\beta]$ the size of the landmark set used in both Roseland and LLL or the subset used in Nyström extension, where $0 < \beta < 1$ and $[x]$ means the largest integer not exceeding $x > 0$. The computational complexity can be divided into three parts. First, construct the affinity matrix and its corresponding degree matrix (DM, LLL, Roseland, Nyström) or the coordinate matrix (LLL); second, apply the eigen-decomposition to the transition matrix or the coordinate matrix; third, extend the eigen-decomposition to the whole database in Nyström and LLL for the purpose of embedding the whole dataset. Below, we calculate the computational complexity by assuming that the metric is the ordinary Euclidean distance.

For the first part, in DM and LLL, the construction of the full affinity matrix is $O(qn^2)$, where the dependence on q comes from the calculation of pairwise distance, and the construction of the degree matrix is $O(n^2)$. In practice, computing the full $n \times n$ affinity matrix is inefficient from the perspective of both memory and speed. One can construct a sparse affinity matrix by using the kNN scheme to save memory, where a sparse affinity matrix with $k = n^\gamma$, where $\gamma \in (0, 1)$, and hence the degree matrix, is constructed in $O(qn^{1+\gamma})$ on average. In LLL, the extra step of calculating Z takes $O(nqK_Z^2 + nK_Z^3)$, where nqK_Z^2 is needed to calculate the Gram matrix and nK_Z^3 is solving the linear system, and the extra step of calculating \hat{W} and \hat{D} takes $O(K_Z n^2)$ if the affinity matrix is full, or

$O(K_Z n^{1+\gamma})$ if the kNN scheme is applied. Since $K_Z \leq n^\beta$ is suggested to be small, it can be ignored in the following calculation. In the Nyström extension, the construction of the affinity matrix is $O(qn^{2\beta})$ and the construction of the degree matrix is $O(n^{2\beta})$. In Roseland, the construction of the landmark-set affinity matrix and its associated degree matrix is $O(qn^{1+\beta})$.

For the second part, recall that the eigen-decomposition complexity for a full symmetric kernel matrix $M \in \mathbb{R}^{N \times N}$ can practically achieve $O(N^{2.81})$ if the Strassen algorithm is chosen for the matrix multiplication (Demmel et al. (2007)). When M is k -sparse, where k is much smaller than N , the complexity can be improved to $O(N^{2+\eta})$ for an arbitrary $\eta > 0$ when $k \leq N^{0.14}$ (Yuster and Zwick (2004)). The SVD complexity for a matrix of size $N \times N'$, where $N \geq N'$ is $O(NN'^2)$. We thus have the complexity of this part for all algorithms.

For the Nyström and LLL, we need the third part. This extension takes $O(qn^{1+2\beta})$ for Nyström, where the dependence on q comes from the calculation of pairwise distances between points in the landmark set and those in the remaining dataset, and $O(n^{1+2\beta})$ for LLL.

With the above calculation, the overall computational complexity is $O(qn^2 + n^{2.81})$ for DM with the full affinity matrix and $O(qn^{1+\gamma} + n^{2+\eta})$ if the kNN scheme is applied; $O(n^{2.81\beta} + qn^{1+2\beta})$ for Nyström extension; $O(qn^2 + n^{2.81\beta} + n^{1+2\beta})$ for LLL with the full affinity matrix and $O(qn^{1+\gamma} + n^{2.81\beta} + n^{1+2\beta})$ with the kNN scheme; $O(qn^{1+\beta} + n^{1+2\beta})$ for Roseland.

To see the relationship among different algorithms, suppose $\gamma = \beta$ and q is a fixed small constant. We see that the Nyström extension, LLL and Roseland are faster than DM, and Nyström runs the fastest in practice. On the other hand, although LLL with the kNN scheme is of the same computational complexity order as Roseland, due to more intermediate steps, like solving the linear system for weights, LLL runs more slowly compared with Roseland in practice. In the high dimensional setup; that is, when $q = q(n)$ so that $q/n \rightarrow \gamma \in (0, \infty)$ when $n \rightarrow \infty$, the impact of q is not negligible. In this case, we can see clearly the benefit of Roseland.

Finally, see Table 1 for a comparison of different algorithms, and see Section 4 for more details on runtime comparisons in different tasks.

	Sparse affinity?	Sparse matrix for EVD or SVD?	Subset & extension?	Use Eigenvalue in Embedding?	Diffusion Process?	Space Complexity	Time Complexity
DM	Full	Full EVD	No	Yes	Yes	$O(qn + n^2)$	$O(qn^2 + n^{2.81})$
	kNN	Sparse EVD				$O(qn + n^{1+\beta})$	$O(qn^{1+\beta} + n^{2+\eta}), \eta > 0$
Nyström	Full	Full EVD	Yes	Yes	Yes	$O(qn + n^{1+\beta})$	$O(n^{2.81\beta} + qn^{1+2\beta})$
LLL	Full	Full EVD	Yes	Yes	No	$O(qn + n^2)$	$O(qn^2 + n^{2.81\beta} + n^{1+2\beta})$
	kNN	Full EVD				$O(qn + n^{1+\beta})$	$O(qn^{1+\beta} + n^{2.81\beta} + n^{1+2\beta})$
Roseland	Full	Full SVD	No	Yes	Yes	$O(qn + n^{1+\beta})$	$O(qn^{1+\beta} + n^{1+2\beta})$

Table 1: Comparison between the algorithms. We assume the number of landmarks is $m = n^\beta$, for $0 < \beta < 1$ and the number of kNN is $k = n^\beta$ in this table.

3. Theoretical support

In this section, we summarize two theoretical results shown in the accompanying paper (Shen and Wu (2022)) that describe the asymptotic behavior and the robustness of Roseland. The first one is the spectral convergence and the second one is the robustness of Roseland; both results are under the manifold setup. We refer readers with interest in details to Shen and Wu (2022).

3.1 Manifold model

Denote the dataset by $\mathcal{X} = \{x_i\}_{i=1}^n \subseteq \mathbb{R}^q$. Assume that the observed dataset \mathcal{X} is independently and identically (i.i.d.) sampled from a random vector $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}^q$, where the range of X is assumed to be supported on a d -dimensional compact smooth Riemannian manifold (M^d, g) without boundary and isometrically embedded in \mathbb{R}^q via $\iota : M^d \hookrightarrow \mathbb{R}^q$. Suppose the density function of X on M is 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 $Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}^q$, where the range of Y is assumed to be (M^d, g) as well. Suppose the density function of Y on M is p_Y . Furthermore, assume that X and Y are independent. Assume $m = n^\beta$, where $\beta \in (0, 1)$. When we run Roseland, assume the kernel in (8) is Gaussian in the theoretical analysis; that is, $K_\epsilon(x, y) = e^{-\|x-y\|^2/\epsilon}$.

For the analysis, consider the following discretization of a function defined on M . Take $f \in C(M)$ and let $\mathbf{f} \in \mathbb{R}^n$ be the discretization of f over \mathcal{X} ; in other words, we have $f_i = f(x_i)$. Denote by Δ the Laplace-Beltrami operator on (M, g) . Let (λ_i, u_i) be the i -th eigenpair of $-\Delta$, where the eigenvalues $\lambda_i \geq 0$ are in the ascending order and $\|u_i\|_2 = 1$. Under our manifold setup and by the well known elliptic theory, the spectrum of $-\Delta$ is discrete with ∞ as the only accumulation point, the eigenfunctions u_i are smooth, and each eigenspace is of finite dimension. To simplify the discussion, we assume that λ_i is simple, which holds for generic smooth manifolds. Here is a practical issue regarding this model that deserves a comment. Note that in practice we are not sure if the dataset fulfills the assumption, like if the manifold assumption holds or if the density functions are regular. To our knowledge, if we do not have sufficient prior knowledge, it is an open problem to check if such assumptions are correct for the dataset we want to analyze. Solving this problem is out of the scope of this paper. In practice, when we cannot directly verify the assumptions, the method can be applied directly and compared to other approaches to see which gives the best performance for a given task, and the algorithm performs well for several real datasets.

3.2 Spectral convergence

Let $\{v_l\}_{l=1}^n$ be the set of eigenvectors of the matrix $(D^{(R)})^{-1}W^{(R)}$ evaluated from the dataset \mathcal{X} and the landmark \mathcal{Y} . We now summarize the result that the eigenvectors $\{v_l\}_{l=1}^n$ asymptotically converge to the eigenfunctions of the Laplace-Beltrami operator of (M, g) as $n \rightarrow \infty$, and in what sense.

First, note that the eigenvectors v_l are in difference Euclidean spaces for different n , hence they cannot be compared directly with the eigenfunctions of the Laplace-Beltrami operator Δ , which are smooth functions on M . To make sense of the comparison, the main idea of the proof is introducing quantities summarized in Table 2 that bridge the discrete and continuous setups. For example, the quantity $\widehat{K}_{\text{ref}, \epsilon, n}(x, y)$ could be viewed as an extrapolation of the similarity $W_{ij}^{(R)}$ between x_i and x_j to any two points x and y via the kernel function. The other terms follow the same extrapolation scheme.

Based on this extrapolation, we could find a function $f_l \in C(M)$, such that the restriction of f_l on the data \mathcal{X} is v_l ; that is, $f_l(x_i) = v_l(i)$, for $i = 1, \dots, n$, and study the convergence of f_l as $n \rightarrow \infty$. It turns out that there is an one-to-one correspondence between the spectrum of the graph Laplacian $I - (D^{(R)})^{-1}W^{(R)}$ and its continuous counterpart $I - \widehat{T}_{\text{ref}, \epsilon, n}$. Hence, to study and make sense of the spectral convergence of $\frac{I - (D^{(R)})^{-1}W^{(R)}}{\epsilon} \rightarrow \Delta$, we study the spectral convergence of $\frac{I - \widehat{T}_{\text{ref}, \epsilon, n}}{\epsilon} \rightarrow \Delta$. Let $(\lambda_{\epsilon, n, i}, u_{\epsilon, n, i})$ be the i -th eigenpair of $\frac{I - \widehat{T}_{\text{ref}, \epsilon, n}}{\epsilon}$, where the eigenvalues $\lambda_{\epsilon, n, i}$ are in the ascending order, and assume $\|u_{\epsilon, n, i}\|_2 = 1$. Fix $q' \in \mathbb{N}$, where q' is the embedding dimension. If we further

	Discrete	Continuous counterpart
Similarity between two points	$W_{ij}^{(R)} = \sum_{k=1}^m K_\epsilon(x_i, y_k) K_\epsilon(y_k, x_j)$	$\widehat{K}_{\text{ref}, \epsilon, n}(x, y) = \frac{1}{m} \sum_{j=1}^m K_\epsilon(x, y_j) K_\epsilon(y_j, y)$
Degree of x_i	$D_{ii}^{(R)} = \sum_j W_{ij}^{(R)}$	$\widehat{d}_{\text{ref}, \epsilon, n}(x) = \frac{1}{n} \sum_{i=1}^n \widehat{K}_{\text{ref}, \epsilon, n}(x, x_i)$
Transition operator	$[(D^{(R)})^{-1} W^{(R)} f](i) = \frac{\sum_{j=1}^n W_{ij}^{(R)} f_j}{\sum_{j=1}^n W_{ij}^{(R)}}$	$\widehat{T}_{\text{ref}, \epsilon, n} f(x) = \frac{1}{n} \sum_{i=1}^n \frac{\widehat{K}_{\text{ref}, \epsilon, n}(x, x_i)}{\widehat{d}_{\text{ref}, \epsilon, n}(x)} f(x_i)$

Table 2: Commonly used notations

take $\epsilon = \epsilon(m) = (\frac{\log(m)}{m})^{1/(4d+11/2)}$, and assume a technical bound that $\epsilon \leq C(q')$, where $C(q')$ is a constant depends on q' , then, with another technical assumption 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 $\{a_n\}$ with $a_n \in \{1, -1\}$ such that with probability $1 - \mathcal{O}(m^{-2})$, for all $i < q'$, we have

$$\|a_n u_{\epsilon, n, i} - u_i\|_{L^\infty} = \mathcal{O}(\epsilon^{1/2}), \quad |\lambda_{\epsilon, n, i} - \lambda_i| = \mathcal{O}(\epsilon^{3/4}),$$

where the implied constants depend on the kernel, the curvature of M , p_X and p_Y . We comment that the result only guarantees the spectral convergence, while the convergence rate is not optimal. We mention that when λ_i is not simple, this result still holds, but we need to use the eigenprojection to describe the eigenvector convergence statement. Hence, by the well established spectral geometry theory (Bérard et al. (1994); Portegies (2016)), the Roseland embedding recovers the underlying manifold, and the embedding can be close to an isometric one when m is sufficiently large (Portegies (2016)).

3.3 Robustness

In real applications, the dataset is often corrupted by noise, or the dataset is not located precisely on a manifold. In this case, we shall worry that spectral embedding algorithms might lead us to low quality, or even misleading results. The robustness issue for DM was studied in El Karoui (2010); El Karoui and Wu (2016) and some solutions are proposed. It is suggested in El Karoui and Wu (2016) to work with the complete graph or a graph with a large number of kNN, and force the random walk non-lazy; that is, set the diagonal entries of the affinity matrix to 0. While this solution works, however, it is not scalable. Roseland, on the other hand, automatically enjoys the desired robustness property (Shen and Wu (2022)), so that it can handle the case when the dataset is contaminated by noise, or if the dataset is not precisely located on the manifold. We now summarize this robustness result. 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 dataset and the landmark set are corrupted by additive ambient space noise that could be colored and heterogeneous. That is, we observe $\tilde{X} = \{\tilde{x}_i\}_{i=1}^n$ and $\tilde{Y} = \{\tilde{y}_j\}_{j=1}^m$:

$$\tilde{x}_i = x_i + \xi_i, \quad \tilde{y}_j = y_j + \eta_j,$$

where ξ_i and η_j are Gaussian noise. We assume that $\xi_i \sim \mathcal{N}(0, \Sigma_i)$ with mean 0 and covariance Σ_i and the noise contaminating the landmark set is $\eta_j \sim \mathcal{N}(0, \bar{\Sigma}_j)$ with mean 0 and covariance $\bar{\Sigma}_j$. Suppose $\|\Sigma_i\|_2 \leq \sigma_q^2$ and $\|\bar{\Sigma}_j\|_2 \leq \bar{\sigma}_q^2$ for $i = 1, \dots, n$ and $j = 1, \dots, m$, where $\sigma_q \geq 0$ and $\bar{\sigma}_q \geq 0$. Assume x_i, y_j, ξ_i and η_j are independent. For more practically challenging purposes, we consider the large q large n setup; that is, the ambient space dimension grows asymptotically as the dataset size n grows. Mathematically, assume $p = p(n)$ and $p/n \rightarrow \gamma > 0$ when $n \rightarrow \infty$. We further assume that

$$\delta_q := \sqrt{\log nm} \sqrt{\sigma_q^2 + \bar{\sigma}_q^2} \left[\sqrt{q(\sigma_q^2 + \bar{\sigma}_q^2)} + \log nm + \mathcal{K} \right] \rightarrow 0, \quad (14)$$

where $\mathcal{K} := \max_{x, y \in M} \|\iota(x) - \iota(y)\|_{\mathbb{R}^q}$.

Remark 1 *To have a closer look at the condition (14), we consider the following simplified case. Suppose $\Sigma_i = \bar{\Sigma}_j = \sigma_q^2 I$ for $i = 1, \dots, n$ and $j = 1, \dots, m$, $\sigma_q = \bar{\sigma}_q$, and $m = n^\beta$. Then, when n is sufficiently large, we have*

$$\delta_q = \sqrt{2(1 + \beta)} \sqrt{\log n} \sigma_q \left[\sqrt{2p/n} \sqrt{n} \sigma_q + (1 + \beta) \log n + \mathcal{K} \right] = O(\sqrt{\log n} \sigma_q [\sqrt{n} \sigma_q + \log(n)])$$

since \mathcal{K} is finite by the compactness assumption of M . The assumption $\delta_q \rightarrow 0$ says that the dominant term $\sigma_q^2 \sqrt{n \log n}$ must go to zero. In other words, the noise level σ_q could be slightly smaller than $(n \log(n))^{-1/4}$ as long as $\sigma_q^2 \sqrt{n \log n} \rightarrow 0$; for example, it holds when $\sigma_q = n^{-1/4} \log(n)^{-1}$. Thus, while the entrywise variance of the noise goes to 0, the total noise energy, defined as the sum of variances in all axes, is $\sigma_q^2 q = \gamma n^{1/2} / \log(n)^2 \rightarrow \infty$ when $n \rightarrow \infty$. An interpretation of the condition (14) is that the algorithm is robust to noise even when the total noise energy blows up.

Denote the Roseland embedding with diffusion time $t > 0$ and the embedding dimension $q' \in \mathbb{N}$ from the clean data and the noisy data by $\Phi L^t \in \mathbb{R}^{n \times q'}$ and $\tilde{\Phi} \tilde{L}^t \in \mathbb{R}^{n \times q'}$ respectively. By the spectral convergence results, we can pick $\epsilon = \epsilon(q') > 0$ so that the first q' non-trivial singular values are sufficiently away from zero when n is sufficiently large. Denote $W^{(r)}$ and $\tilde{W}^{(r)}$ to be the landmark-set affinity matrices from clean and noisy datasets respectively. Denote $\Phi L^t \in \mathbb{R}^{n \times q'}$ and $\tilde{\Phi} \tilde{L}^t \in \mathbb{R}^{n \times q'}$ to be Roseland embeddings from $W^{(r)}$. Then, we have

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

for some orthogonal matrix $O \in \mathbb{R}^{q' \times q'}$, and s_2 are the largest non-trivial singular value of Roseland from the clean dataset. The rotation O here deserves a discussion. Note that when the eigenvalue is simple, the associated eigenvector is free up to a positive and negative sign. When the eigenvalue is not simple, the associated eigenvectors are non-identifiable up to a rotation inside the eigenspace. In practice, this rotation is not possible to recover unless more information is available. Since the purpose is measuring how accurate the embedding is when the dataset is noisy, we introduce the rotation O to eliminate this freedom. Note that this rotation is different from the non-identifiability of the global embedding up to a translation and rotation, which is caused by the nature of the nonlinear algorithm that only takes local relationship into account.

This result 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 $\delta_q \rightarrow 0$ when $n \rightarrow \infty$, it does not mean that the noise impact is small when $n \rightarrow \infty$. We shall quantify the noise impact by $q(\sigma_q^2 + \bar{\sigma}_q^2)$ in δ_q , which can be viewed as the *total noise energy* 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 noise energy blows up since $\sigma_q^2 q \rightarrow \infty$.

Moreover, when the landmark set is noise free or $\bar{\sigma}_q = 0$, we could achieve a better convergence. 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 size is large. For a fair comparison between Nyström extension and Roseland, both algorithms use the same landmark set. All of the simulations were done on a Linux machine with 4-core 3.5Ghz i5 CPUs and 16GB memory. The Matlab code for the reproducibility purpose can be found in https://github.com/shenchaojerry/Roseland_numerical, where the Matlab code for LLL is downloaded from <https://eng.ucmerced.edu/people/vladymyrov>.

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 (LeCun and Cortes (2010)). There are 60,000 training images and 10,000 testing images. For the spectral clustering purpose, we include all 70,000 data points. We set the landmark set to have size m , where $m = \lceil n^\beta \rceil$ for some $\beta \in (0, 1)$ and $\lceil x \rceil$ means the largest integer not exceeding $x > 0$.

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 this work, we set $K = 10$ since we know there are ten digits. Below, we compare the performance of the spectral clustering based on different spectral embedding algorithms, including DM, Nyström extension, LLL, and Roseland. In DM, in order to fit our 16G memory, we use kNN with $k = 200$ to construct a sparse affinity matrix. The kernel used in DM, Nyström and Roseland is Gaussian, the metric d is the ordinary Euclidean distance, and the bandwidth ϵ is chosen to be the square root of the median of all pairwise distances. We follow the traditional spectral clustering approach and only use eigenvectors; that is, the diffusion time t is chosen to be 0. The chosen parameters for LLL is different from those used in the original LLL paper (Vladymyrov and Carreira-Perpinán (2013)), where the task was different from ours. Indeed, the authors examined the relative error between the LLL embedding of 60,000 MNIST data points and the exact embedding by eigenmap. Thus, the reported parameters in Vladymyrov and Carreira-Perpinán (2013) may not be optimal for our purpose. So we run a grid search to optimize the performance of LLL for our spectral clustering task. Since one do not know the optimal embedding dimension beforehand, for each β , we select the parameters of LLL such that they maximize the median accuracy across all considered embedding dimensions and report the results. We also consider the widely used linear dimension reduction

algorithm, principal component analysis (PCA), as an extra baseline. In addition to comparing the above algorithms on the clean MNIST dataset, we also compare them on a noisy MNIST dataset 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 in Figure 2. While it is not easy to identify ten digits from the embedding, we can see that when the data is noisy, the embedding quality of LLL is not ideal.

We consider the following metric to quantitatively evaluate the performance of different algorithms. Suppose $\{\hat{C}_k\}_{k=1}^{10}$ are the clusters obtained by a chosen spectral clustering. Denote the dataset by \mathcal{X} , so $\mathcal{X} = \bigsqcup_{k=1}^{10} \hat{C}_k$, where \bigsqcup indicates disjoint union. Let $f : \mathcal{X} \rightarrow \{0, 1, \dots, 9\}$ be the true predictor function and $\hat{f} : \mathcal{X} \rightarrow \{0, 1, \dots, 9\}$ be the estimated predictor function that is constant on \hat{C}_k for $\forall k$. Here, \hat{f} is defined in the following way:

$$\hat{f}|_{\hat{C}_k} = \arg \max_{y \in \{0, 1, \dots, 9\}} \sum_{x \in \hat{C}_k} \mathbb{I}\{f(x) = y\},$$

for $k = 1, \dots, 10$, 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:

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

where $|\cdot|$ is the size of the cluster.

The accuracies using different algorithms with different embedding dimensions are plotted in Figure 3. In this experiment, we focus on small landmark sets when $\beta \leq 0.5$ that is suitable for large datasets. Note that in Vladymyrov and Carreira-Perpinán (2013), the number of eigenvectors is suggested to be 50, which cannot be achieved when $\beta = 0.3$ since $70,000^{0.3} \sim 28$. Thus, when $\beta = 0.3$, we only report the results up to dimension 28 for all algorithms. We note that when the embedding dimension is low by using only the top few eigenvectors, LLL has better performance than Roseland and Nyström s, and when $\beta = 0.4, 0.5$, LLL even outperforms DM and almost achieve 74% accuracy when $\beta = 0.5$ and the embedding dimension is less than 8. As the dimension increases, however, Roseland and Nyström surpass LLL and Roseland has the best performance among the three. When $\beta = 0.5$, that is, the number of landmarks is 264, Roseland achieves a comparable performance as DM, while Roseland runs nearly 100× faster than DM. See Table 3 for detailed runtime reports, which contains a comparison of different algorithms with different landmark sets of different sizes by reporting the overall runtime measured in the unit of second. It is clear that while Roseland is slower than Nyström, the difference is acceptable even when $\beta = 0.5$. While LLL is significantly faster than DM, it is slower than Roseland, which mainly comes from the construction of the affinity matrix. As expected, Roseland is slower than the Nyström extension.

The performance of LLL deserves more discussion. Note that the weights of all eigenvectors in LLL are equal. While there is no theoretical support, we conjecture that the eigenvector “quality” in LLL decreases when the associated eigenvalues increases. Suppose this conjecture is true, the equally weighted eigenvectors explain why the more eigenvectors we use, the worse the accuracy is, particularly when the landmark set is small. While LLL performs well when the embedding dimension is all, exploring LLL and its theoretical property is of potential interest. However, this topic is out of the scope of this paper, so we leave this conjecture to future work.

We also investigate the effect of the landmark number and the results are shown in Figure 4. We notice that the performance of Roseland benefits if we have access to clean landmarks. We remark that this setup is realistic in real world. In practice, one may collect a large amount of noisy data by a cheap device, and collect some clean data by an expensive device.

We mention that other than LLL (Vladymyrov and Carreira-Perpinán (2013)), the two largest experiments we are aware of with the MNIST dataset are Günter et al. (2007) and Li et al. (2016). In Günter et al. (2007), the authors improved the convergence of the kernel Hebbian algorithm to accelerate the kernel PCA, and the total run time is more than 50 hours on 60,000 images. No 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 Li et al. (2016), the authors applied a similar idea to Roseland, but with a different normalization. Similar to Yan et al. (2009), 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 Ouimet and Bengio (2005), 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 (Shi and Malik (2000)). We consider two grey-level images, one is the well-known 256×256 cameraman and one is a 481×321 hawks image taken from the Berkeley Segmentation Dataset and Benchmark (Martin et al. (2001)). To make the task more challenging, we corrupt the images with additive noise. Specifically, we first normalized all pixel values, then entry-wise add Gaussian noise with mean 0 and standard deviation 0.2. See Figure 5 for two example images.

We slightly modified the algorithm proposed in Shi and Malik (2000). Specifically, for each pixel, we attach a 5×5 patch surrounding the pixel. Thus, for a gray-level image of size 480×640 , we have about 305,000 points in \mathbb{R}^{25} . Since DM is not efficient for a dataset of this size even with a kNN scheme, we only compare LLL, Nyström extension and Roseland. For Nyström and Roseland, the landmark set is sampled from the “edge points” in the image; in other words, we sample those pixel values that have high contrast to the neighbors. For LLL, we use the reported parameters to construct the affinity matrix in the paper (Vladymyrov and Carreira-Perpinán (2017)). See Figures 6, 7, 8 and 9 for the image segmentation results of clean and noisy images for a visual inspection.

A comparison of different algorithms with landmark sets of different sizes in the sense of runtime per second is evaluated on the COCO dataset (Lin et al. (2014)). For RGB images, we take the mean of 3 channels and we reshape the images to have size 480×640 when necessary. The result is reported in Table 3. We see that while Roseland is slower than Nyström extension, they are on the same order. Also, it is faster than LLL. This result suggests that Roseland is efficient and can be applied to the image segmentation task. A systematic exploration of this application is out of the scope of this paper and will be studied in our future work.

5. Liver transplant analysis

The last example is a case study of a long-term and high-frequency biomedical signal collected from the whole liver transplant procedure. Since liver transplant usually takes several hours, the collected

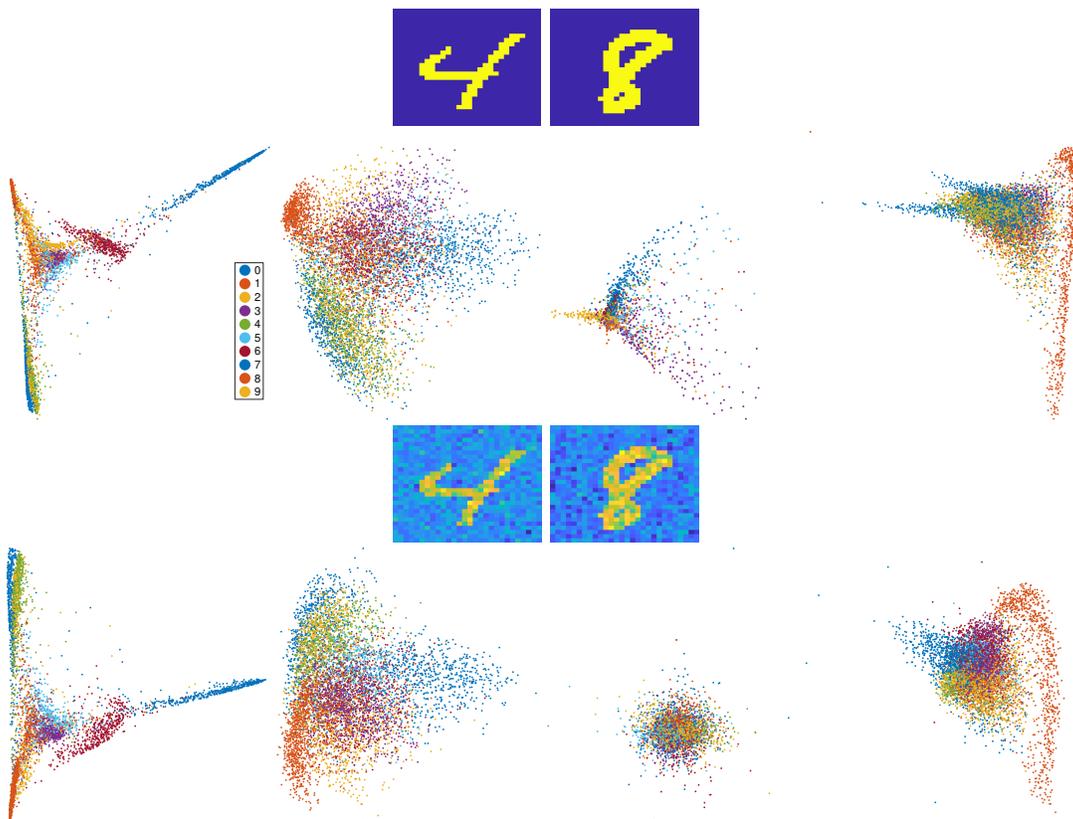


Figure 2: First row: Examples of clean digits. Second row: embeddings of clean MNIST, from left to right: the DM embedding, the Nyström extension embedding, the LLL embedding and the Roseland embedding. Third row: Examples of noisy digits. Fourth row: embeddings of noisy MNIST, from left to right: the DM embedding, the Nyström extension embedding, the LLL embedding and the Roseland embedding. In this figure, $\beta = 0.5$, so the number of landmarks is 264 for Nyström, LLL and Roseland.

signal is long and it is challenging to analyze it without a numerically efficient algorithm. In this example, we demonstrate how to apply Roseland to study such a long biomedical signal.

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 (Rudnick et al. (2015)). 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 have been questioned subsequently for their performance in liver transplant surgery (Biancofiore et al. (2011); Tsai et al. (2012); Shih et al.

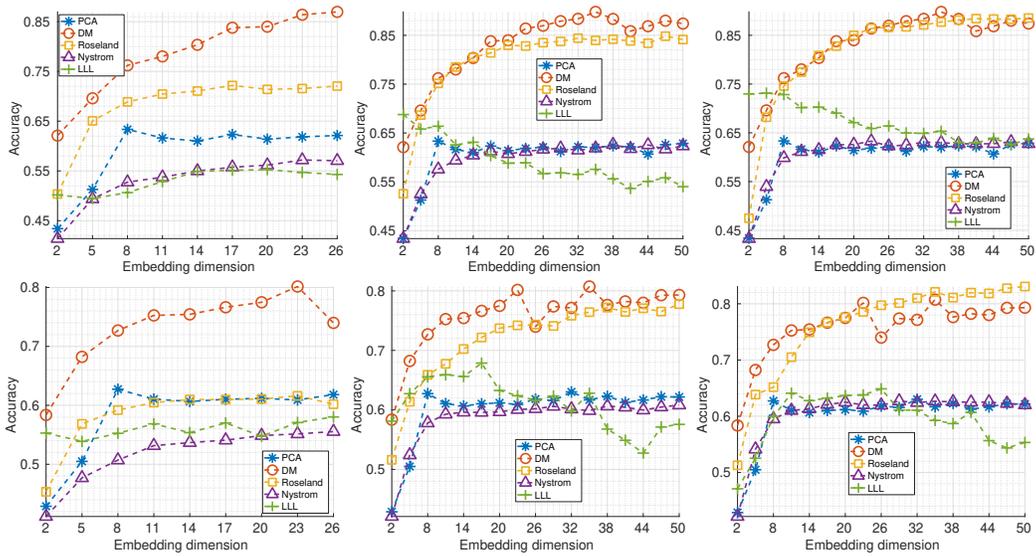


Figure 3: Accuracies of spectral clustering in MNIST, where data size $n = 70,000$. First row, clean MNIST, from left to right we let $\beta = 0.3$, $\beta = 0.4$ and $\beta = 0.5$, so the number of landmarks used is $m = \lceil n^\beta \rceil = 28, 87, 264$ respectively. Second row, noisy MNIST, from left to right we let $\beta = 0.3$, $\beta = 0.4$ and $\beta = 0.5$.

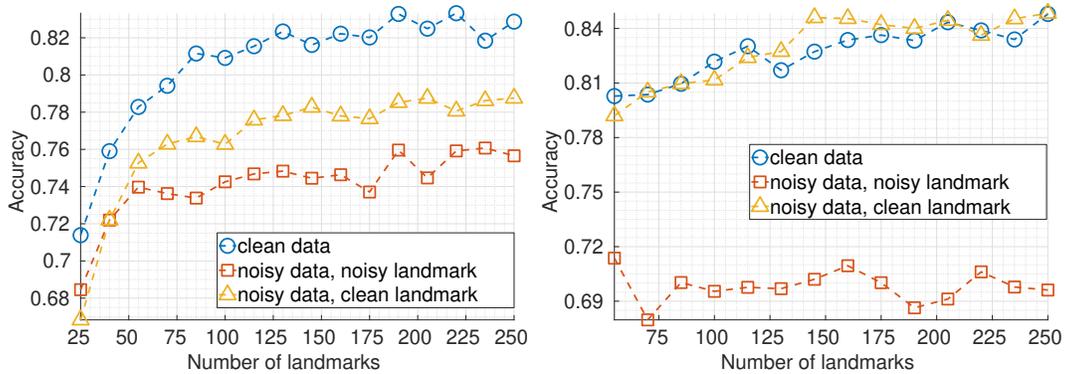


Figure 4: Accuracies of spectral clustering in MNIST, where data size $n = 70,000$ and the number of landmarks ranging from 25 to 264, that is, $0.25 < \beta < 0.5$. Left: use top 20 eigenvectors. Right: use top 50 eigenvectors.

(2016)). 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 (Mitchell et al. (2010)) or quantities in the frequency domain (Wang et al. (2010)), 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.



Figure 5: We run spectral based image segmentation on two images and their noisy versions. Left two: the 256×256 cameraman image. Right two: the 481×321 hawks image taken from the Berkeley Segmentation Dataset and Benchmark.



Figure 6: Clean cameraman. Top row, LLL: first image is the image segmentation result using top 4 eigenvectors, the second to the fifth images are top 1 to 4 eigenvectors. Second row is from Nyström and the last row is from Roseland.

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 hypothesize that taking the whole waveform into account might provide more complimentary information compared with those traditional parameters. On the other hand, due to the complicated features/patterns caused by the nonstationarity, 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 DM can extract rich information directly from the *raw cardiovascular waveform* (Wang et al. (2020)). The novelty in Wang et al.



Figure 7: Noisy cameraman. Top row, LLL: first image is the image segmentation result using top 4 eigenvectors, the second to the fifth images are top 1 to 4 eigenvectors. Second row is from Nyström and the last row is from Roseland.

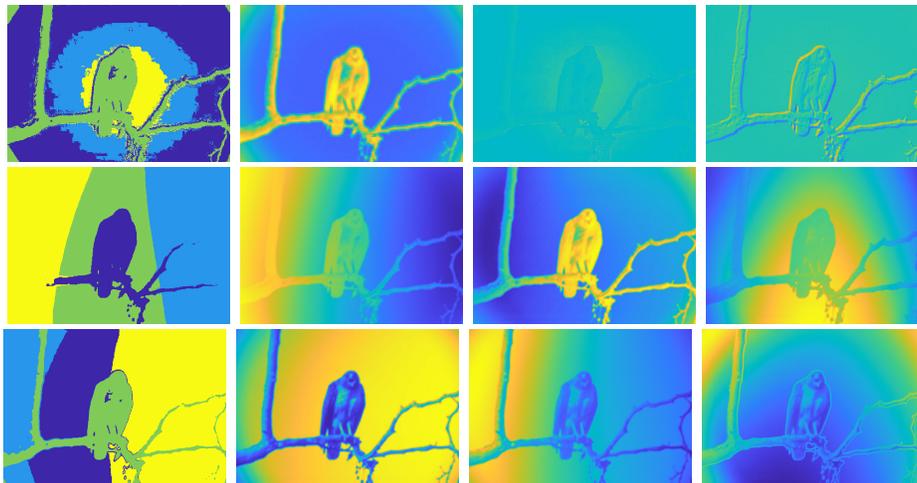


Figure 8: Clean hawks. Top row, LLL: first image is the image segmentation result using top 3 eigenvectors, the second to the forth images are top 1 to 3 eigenvectors. Second row is from Nyström and the last row is from Roseland.

(2020) is 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

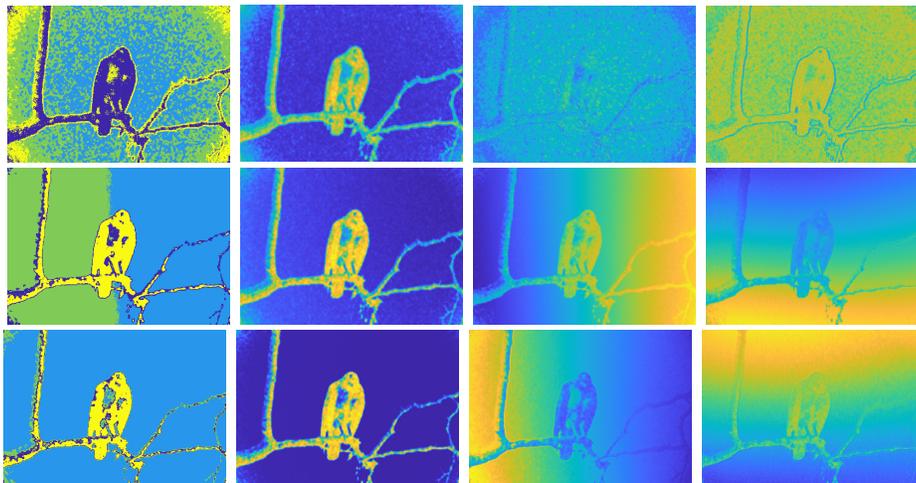


Figure 9: Noisy hawks. Top row, LLL: first image is the image segmentation result using top 3 eigenvectors, the second to the fourth images are top 1 to 3 eigenvectors. Second row is from Nyström and the last row is from Roseland.

	MNIST			Image segmentation		
PCA	4.27±0.04			N/A		
DM	1468.4±5.05			N/A		
	$\beta=0.3$	$\beta=0.4$	$\beta=0.5$	$\beta=0.2$	$\beta=0.3$	$\beta=0.4$
LLL	317.75±0.59	328.28±0.80	332.68±0.41	33.19±0.43	46.92±0.55	57.45±0.76
Nyström	0.60±0.04	1.57±0.02	4.78±0.41	0.44±0.03	0.63±0.01	1.52±0.05
Roseland	1.09±0.04	4.14±0.06	15.24±0.85	0.95±0.11	2.29±0.02	6.82±0.09

Table 3: Runtime in seconds. MNIST consists of 70,000 images, each of size 28×28 ; images used in image segmentation are of size 480×640 .

small dataset, and applying it to study ABP waveforms during liver transplant surgery is difficult. Note that the whole period of a typical surgery yields more than 10^5 consecutive pulses associated with associated heart beats, and we view each of these pulses as a data point in a high dimensional Euclidean space. As a result, we have more than 10^5 points in a high dimensional Euclidean space, where the dimension depends on the data that is detailed before, for a pulse-to-pulse waveform analysis.

In this study, we hypothesize that with the help of Roseland, the manifold learning approach shown in Wang et al. (2020) can be applied to study the ABP waveform during the liver transplant procedure, and provide hemodynamic information on both the small and large scales.

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™B850 (GE Health-

care, Chicago, Illinois, United States). The collected ABP signal recorded during the whole surgery was analyzed offline after the surgery finished. 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 upsampling is carried out to enhance the peak detection performance (Laguna and Sörnmo (2000)). The signal is of 78,350s long spanning the whole surgical procedure and contains 120,725 pulses. While the liver transplant surgery is a complicated one, we consider three major phase transitions, including

1. vascular clamp; that is, 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;
2. reperfusion; that is, 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;
3. artery anastomosis; that is, the start of the circulatory connection between hepatic artery and the graft.

All these transitions drastically affect the cardiovascular system via the changes of fluid volume and electrolyte composition. To better observe the dynamics, we consider the phases before and after each phase transition. Thus, we focus on the following main steps during the surgery procedure: dissection phase before vascular clamp; vascular clamp; anhepatic phase after vascular clamp; anhepatic phase before reperfusion; reperfusion; neohepatic phase after reperfusion; before artery anastomosis; artery anastomosis; after artery anastomosis.

5.3 Data analysis

Denote the ABP waveform as $x^A \in \mathbb{R}^N$, where N is the length of the ABP time series. We used the maximum of the first derivative 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 $\bar{x}_i^A := [x^A(n_i), x^A(n_i + 1), \dots, x^A(n_{i+1})]^T$. Since the duration of each pulse is not constant, we followed the common procedure (Cerutti et al. (1986); Wang et al. (2020)) and truncated them to be of an uniform size according to their minimal length $q = \min\{n_{i+1} - n_i + 1\} \in \mathbb{N}$, and get $\hat{x}_i^A := [x^A(n_i), x^A(n_i + 1), \dots, x^A(n_i + q - 1)]^T$. 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 $x_i^A \in \mathbb{R}^q$ and its derivative as $x_i^{dA} = 500 \times [x_i^A(2) - x_i^A(1), \dots, x_i^A(q) - x_i^A(q - 1)]^T \in \mathbb{R}^{q-1}$. Derived from the ABP signal, we get the dataset $\mathcal{X}^A = \{[(x_i^A)^T (x_{i-1}^A)^T (x_{i-2}^A)^T (x_i^{dA})^T (x_{i-1}^{dA})^T (x_{i-2}^{dA})^T]^T\}_{i=3}^L \subseteq \mathbb{R}^{6q-3}$ to capture the temporal relationship in the ABP signal. Note that this concatenation is related to the Takens' embedding map that recovers the underlying phase space hosting the intrinsic dynamics (Takens (1981)). We assume that \mathcal{X}^A can be well approximated by a low dimensional manifold, referred to as the *wave-shape manifold* (Lin et al. (2021a)). Since n is roughly 1.2×10^5 in a typical liver transplant surgery, if we choose $\beta = 0.45$, $m \sim 193$ and $n/m \sim 621$. To apply Roseland, we thus

construct the landmark set by uniformly sampling points from the whole database by picking one every 600 beats, which is about choosing one landmark every 10 minutes. Specifically, the landmark set $\mathcal{Y}^A = \{y_k^A\}_{k=1}^m$ comes from setting $y_k^A = x_{600k}^A$.

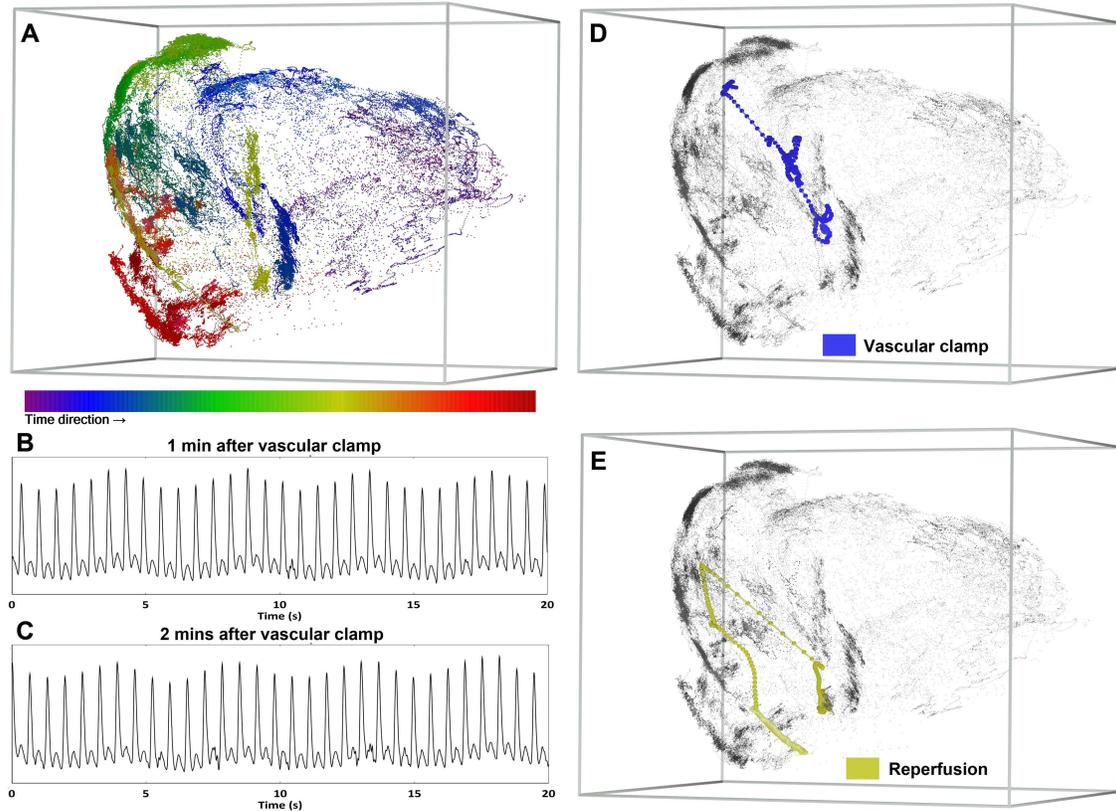


Figure 10: The 3D embedding (panel A) of pulse-to-pulse pressure waveforms collected from a 78,350s ABP signal (120,725 pulses) during the liver transplantation procedure. The embedded pulses are labeled by colors encoding the time. The color helps visualize the ever-changing trajectory formed from successive 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. See <https://www.youtube.com/watch?v=XDisVP1twRA> for details.

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 computational time using traditional DM algorithm based on eigendecomposition would be more than one day.

The embedding result is shown in Figure 10. The successive pulses evolve with time and constitute a trajectory on the manifold presented as a 3D embedding (Fig.10, panel A). The trajectory visits different locations during different steps of the liver transplant procedure. This is reasonable since a subject’s hemodynamic status differ from one surgery step to another. 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, it is not easy to perceive the long term dynamics and organize them with only human eyes and brain (Fig.10, panel B and C).

We further quantify the trajectory in different surgical phases as well as the phase transition periods in which the trajectory moves in fast pace. Specifically, we consider different hemodynamic phases during the liver transplant procedure, particularly those that phase transition happens with violent physiological changes take place, including vascular clamp, reperfusion, and artery anastomosis. The quantification consists of the following steps. First, for each hemodynamic phase, the geometric center of all embedded ABP waveforms in the 10-dim Euclidean space by Roseland is evaluated. Second, we define the distance between two hemodynamic phases by measuring the Roseland diffusion distance (RDD) (12) between their geometric centers. The quantitative measurement is expressed as mean and 95 % confidence interval after a bootstrap resampling without replacement in 100,000 samples. Third, we evaluate two quantities in light of the Newtonian mechanics, including the *two-point velocity* and the *trajectory speed*. The two-point velocity is defined as the ratio of the RDD between two points and the time difference between the two points. The two-point velocity measures the dynamics *on the large scale* (macroscopically). 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 trajectory speed measures the dynamics *on the small scale* (microscopically). In other words, the two-point velocity cares how fast the hemodynamic phase shifts from one to the other without caring about its local dynamics, while the trajectory speed cares about its local dynamics.

The quantitative result is shown in Table 4. We see that during each phase transition, the trajectory moves faster 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 small scale, are similar during the phase transition and within each surgical phase. The numeric results are consistent with the visualization of the 3D embedding (Figures 10 and 11). This quantitative result suggests that while the exterior force drives the hemodynamic status from one to another, “locally” the variation is less impacted. This interesting finding needs to be further validated from both clinical and physiological perspectives. We shall also correlate it with the clinical outcomes so that we could determine its clinical application scenario.

Table 4: 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 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. There are at least two potential clinical application scenarios. First, as there is room to be improved in monitoring the hemodynamic status in liver transplant surgery (Biancofiore et al. (2011); Tsai et al. (2012); Shih et al. (2016)), we expect that the proposed waveform analysis would lead to more insights into the hemodynamic status of liver transplant surgery, and hence improve the patient’s outcome. Second, the visualization of the ultra-long term waveform could help healthcare givers better depict the long term dynamics of the patient with fine structure, either for the liver transplant surgery or others, so that the health status could be better quantified from a different perspective. The user environment, user experience and other clinical needs of the proposed algorithm are all under exploration. See Lin et al. (2021b) for some preliminary results. 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. Last but not the least, it would be interesting to consider analyzing the waveform during the surgery in the real-time manner and obtain useful information to, for example, guide the surgery. How to re-fit continuously when we obtain new data to obtain real-time information, and its clinical application, will be explored in the future work.

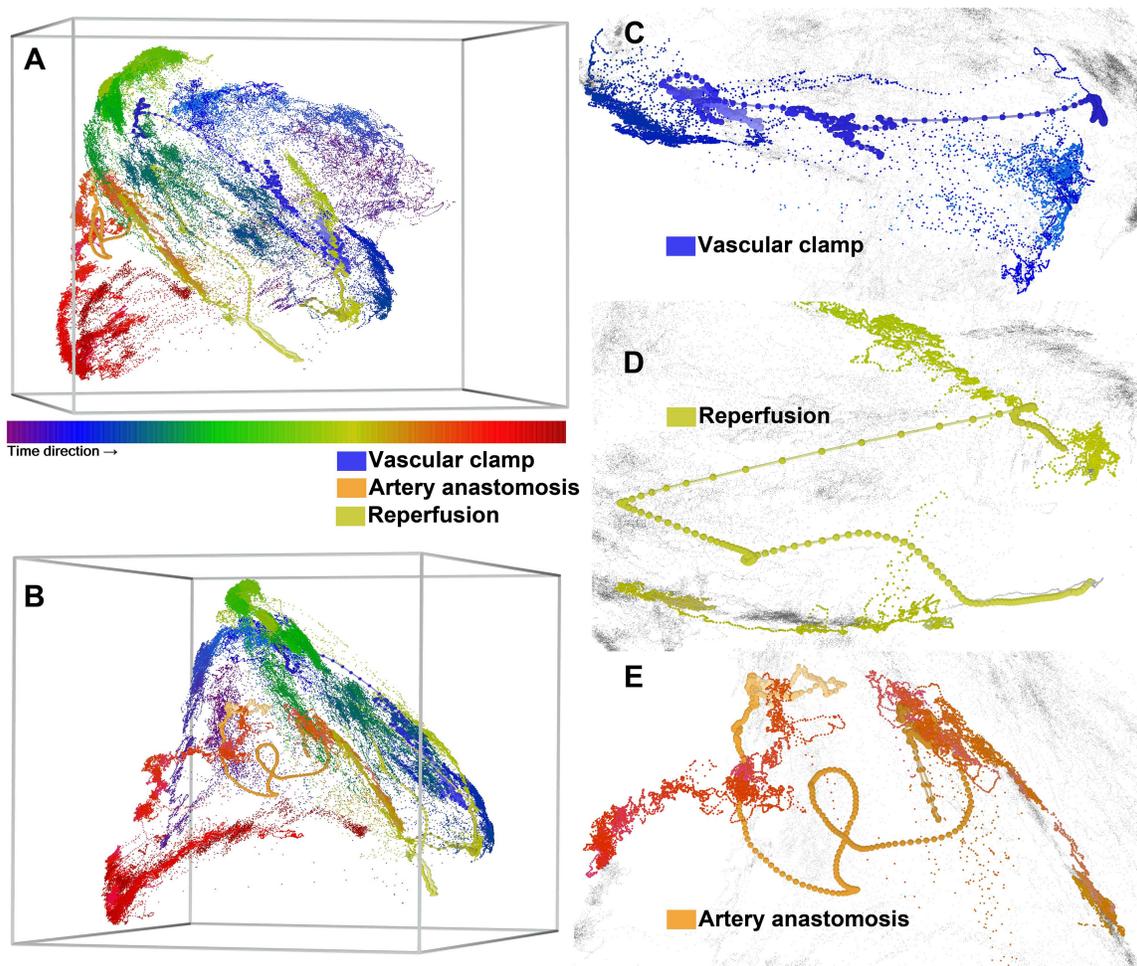


Figure 11: The 3D embedding highlights the locations of transition phases (in linked dot) in relation to the whole period of ABP waveforms (smaller dot) in a liver transplant surgery. The color labels the time sequence. Panel B is the horizontally rotated Panel A by 60 degrees for a better visualization 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.

6. Discussion and conclusion

Motivated by analyzing long-term, high-frequency and nonstationary physiological time series, we propose a landmark diffusion based algorithm, Roseland, that is numerically efficient and robust to

various types of noises with theoretical support under the manifold model. The key idea is controlling a diffusion process on the dataset so that the diffusion is forced to go through a small subset called the landmark set. Compared with the commonly applied speed-up approaches like kNN, Nystrom or recently proposed LLL, Roseland is balanced from the perspectives of computational efficiency and robustness. Specifically, while Roseland is slightly slower compared with Nystrom, it is more efficient compared with kNN and LLL. On the other hand, Roseland is robust to noise while kNN is not and we do not have theoretical support for Nystrom or LLL. These claims are confirmed by applying Roseland to various datasets, including the MNIST database, images and ABP waveforms recorded from a liver transplant surgery. We conclude that Roseland has potential in handling big and noisy real-world databases.

6.1 Application of the Roseland

The idea of landmark set have several applications. Here we mention two of them. The vector diffusion map (VDM) (Singer and Wu (2012)) 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 (Alzate and Suykens (2008)). It is well known that the more clusters we need to determine, the more eigenvectors we need (Alzate and Suykens (2008); Lee et al. (2014)). As is shown in the numerical section, Roseland has the ability to recover more and better eigenvectors, at least compared with the Nystrom extension. This shows the potential of applying 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. The authors also want to thank the associate editor and three anonymous reviewers for their constructive and helpful comments.

References

- 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.
- 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.

- Paddy M Barrett, Ravi Komatireddy, Sharon Haaser, Sarah Topol, Judith Sheard, Jackie Encinas, Angela J Fought, and Eric J Topol. Comparison of 24-hour holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring. *The American journal of medicine*, 127(1): 95–e11, 2014.
- 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.
- P. Bérard, G. Besson, and S. Gallot. Embedding riemannian manifolds by their heat kernel. *Geometric & Functional Analysis GAFA*, 4(4):373–398, 1994.
- 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.
- S. Cerutti, G. B. Baselli, S. Civardi, E. Ferrazzi, A. M. Marconi, M. Pagani, and G. Pardi. Variability analysis of fetal heart rate signals as obtained from abdominal electrocardiographic recordings. *Journal of Perinatal Medicine*, 14(6):445–452, 1986.
- L.B. Chang, Z. Bai, S.Y. Huang, and C.R. Hwang. Asymptotic error bounds for kernel-based nystrom low-rank approximation matrices. *Journal of Multivariate Analysis*, 120:102–119, 2013.
- 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.
- 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.
- R.R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21(1):5–30, 2006.
- W. Czaja, T. Doster, and A. Halevy. An overview of numerical acceleration techniques for non-linear dimension reduction. In *Recent Applications of Harmonic Analysis to Function Spaces, Differential Equations, and Data Science*, pages 797–829. Springer, 2017.
- J. Demmel, I. Dumitriu, and O. Holtz. Fast linear algebra is stable. *Numer. Math.*, 108:59–91, 2007.
- N. El Karoui. On information plus noise kernel random matrices. *Ann. Statist.*, 38(5):3191–3216, 2010.
- 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.
- 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.
- 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.

- 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.
- 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.
- 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.
- 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.
- A Jaffe, Y Kluger, GC Linderman, G Mishne, and S Steinerberger. Randomized near-neighbor graphs, giant components and applications in data science. *Journal of applied probability*, 57(2): 458–476, 2020.
- 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.
- 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.
- P Laguna and Leif Sörnmo. Sampling rate and the estimation of ensemble variability for repetitive signals. *Medical and Biological Engineering and Computing*, 38(5):540–546, 2000.
- Y. LeCun and C. Cortes. MNIST handwritten digit database. 2010. URL <http://yann.lecun.com/exdb/mnist/>.
- 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.
- Y. Li, J. Huang, and W. Liu. Scalable sequential spectral clustering. In *Thirtieth AAAI conference on artificial intelligence*, 2016.
- 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.
- Y-T Lin, J. Malik, and H.-T. Wu. Wave-shape oscillatory model for nonstationary periodic time series analysis. *Foundations of Data Science*, 3(2):99–131, 2021a.
- Y.-T. Lin, H.-T. Wu, S.-C. Wang, C.-K. Ting, C. Liu, N.-C. Lin, C.-Y. Chen, and C.-C. Loong. Intraoperative arterial pressure waveforms shows temporal structure complexity correlated with acuity of liver transplant by pulse wave manifold learning analysis. *Society for Technology in Anesthesia, Virtual Annual Meeting*, 2021b.
- 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.

- 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.
- 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.
- 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.
- M. Ouimet and Y. Bengio. Greedy spectral embedding. In *AISTATS*. Citeseer, 2005.
- J.W. Portegies. Embeddings of riemannian manifolds with heat kernels and eigenfunctions. *Communications on Pure and Applied Mathematics*, 69(3):478–518, 2016.
- 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.
- MR Rudnick, L De Marchi, and JS Plotkin. Hemodynamic monitoring during liver transplantation: A state of the art review. *World journal of hepatology*, 7(10):1302, 2015.
- C. Shen and H.-T. Wu. Scalability and robustness of spectral embedding: landmark diffusion is all you need. *Information and Inference: A Journal of the IMA*, 2022.
- J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.
- 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.
- A. Singer and H.-T. Wu. Vector diffusion maps and the connection laplacian. *Communications on pure and applied mathematics*, 65(8):1067–1144, 2012.
- S. Steinerberger. A Filtering Technique for Markov Chains with Applications to Spectral Embedding. *Applied and Computational Harmonic Analysis*, 40:575–587, 2016.
- F. Takens. Detecting strange attractors in turbulence. In David Rand and Lai-Sang Young, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381. Springer Berlin Heidelberg, 1981.
- 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.
- 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.

- C Vlachopoulos, M O'Rourke, and W Nichols. *McDonald's blood flow in arteries: theoretical, experimental and clinical principles*. CRC press, 2011.
- M. Vladymyrov and M. Carreira-Perpinán. Locally linear landmarks for large-scale manifold learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 256–271. Springer, 2013.
- M. Vladymyrov and M. Carreira-Perpinán. Fast, accurate spectral clustering using locally linear landmarks. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 3870–3879. IEEE, 2017.
- J. Wang, Z. Zhang, and H. Zha. Adaptive manifold learning. In *Advances in neural information processing systems*, pages 1473–1480, 2005.
- 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.
- 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.
- 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.
- 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.
- R. Yuster and U. Zwick. Fast sparse matrix multiplication. *Lect. Notes Comput. Sci.*, 322:604–15, 2004.