Supplementary material for *Geometric fluid approximation for general continuous-time Markov chains*

Michalis Michaelides, Jane Hillston, Guido Sanguinetti

August 19, 2019

Journal: Proceedings of the Royal Society A

A Diffusion maps for undirected graphs

There exists extensive literature examining the implications of diffusion maps, as well as their limitations and strengths [1, 2, 3, 4, 5]. What follows is therefore not an attempt to re-derive these results or convince the reader of the validity of the method, but rather to set notation and highlight the aspects that are relevant to our purposes. The exposition below is also necessary to act as a foundation for the results of Perrault-Joncas and Meilă that build upon the original concept of diffusion maps as put forth by Coifman, Lafon, Nadler, and Kevrekidis.

In [4, 5], the authors consider a family of density-normalised (i.e. anisotropic) symmetric kernels

$$k_\epsilon^{(\alpha)}(\mathbf{x},\mathbf{y}) = \frac{k_\epsilon(\mathbf{x},\mathbf{y})}{p_\epsilon^\alpha(\mathbf{x})p_\epsilon^\alpha(\mathbf{y})}$$

characterising the distance between high-dimensional points $\mathbf{x},\mathbf{y} \in \mathcal{M} \subseteq \mathbb{R}^p$. The kernel used here is the radial basis function $k_\epsilon(\mathbf{x},\mathbf{y}) = \exp\big(-d(\mathbf{x},\mathbf{y})^2/\epsilon\big)$, which provides a similarity between points based on the Euclidean distance d in the original space. The density-normalising factor $p_\epsilon^\alpha(\mathbf{x})$ depends on the manifold density, $p_\epsilon(\mathbf{x}) = \int k_\epsilon(\mathbf{x},\mathbf{y}p_\epsilon(y)dy$, and the choice of the power α leads to transition kernels of different diffusion process operators (see below). The hyperparameter ϵ is the kernel width, which corresponds to the time elapsed between observations of a putative diffusion process (see below). For a finite set of points we can construct an adjacency matrix whose elements are given by the kernel, for a network with points as nodes and weighted undirected edges.

Assuming that the points were sampled by observing a diffusion process in the space $\mathcal{M}$, the authors then take the forward Markov transition probability kernel to be

$$M_f^{(\alpha)}(\mathbf{x}|\mathbf{y}) = \Pr\left[\mathbf{x}(t+\epsilon) \mid \mathbf{x}(t) = \mathbf{y}\right] = \frac{k_\epsilon^{(\alpha)}(\mathbf{x},\mathbf{y})}{d_\epsilon^{(\alpha)}(\mathbf{y})},$$

where $d_\epsilon^{(\alpha)}(\mathbf{y}) = \int_{\mathcal{M}} k_\epsilon^{(\alpha)}(\mathbf{x},\mathbf{y})p(\mathbf{x})d\mathbf{x}$ is the graph Laplacian normalisation factor. Since this is the transition probability for the putative continuous diffusion process evolving

in the space $\mathcal{M}$, the (forward) infinitesimal diffusion operator of the process is given by

$$\frac{\partial}{\partial t} = \mathcal{H}_f^{(\alpha)} = \lim_{\epsilon \to 0} \left[\frac{T_f^{(\alpha)} - I}{\epsilon} \right],$$

where I is the identity operator, and $T_f^{(\alpha)}$ is a (forward) transport operator defined as $T_f^{(\alpha)}[\phi](\mathbf{x}) = \int_{\mathcal{M}} M_f^{(\alpha)}(\mathbf{x}|\mathbf{y})\phi(\mathbf{y})p(\mathbf{y})d\mathbf{y}$, which evolves a function $\phi : \mathcal{M} \to \mathbb{R}$ according to $M_f^{(\alpha)}$ and the manifold measure $p(\mathbf{y}) = e^{-U(\mathbf{x})}$.

By asymptotic expansion of the relevant integrals, they show that the forward and backward operator pair is

$$\mathcal{H}_f^{(\alpha)} = \Delta - 2\alpha\nabla U \cdot \nabla + (2\alpha - 1)(\|\nabla U\|^2 - \Delta U), \qquad \text{and} \tag{1}$$

$$\mathcal{H}_b^{(\alpha)} = \Delta - 2(1-\alpha)\nabla U \cdot \nabla \tag{2}$$

respectively.

We then regard the adjacency matrix W of a given network to be a discrete approximation of the transition kernel k_ϵ defined over continuous space. From that, we can construct discrete (in time and space) approximations to the diffusion operators $\mathcal{H}^\alpha$ above by performing the necessary normalisations. To retrieve the embedding coordinates for each network vertex one needs to spectrally analyse the approximation to the diffusion operator, taking the 1 to $k+1$ eigenvectors $\{\psi_j\}_{j=1}^d$ ordered by the associated eigenvalues $\{-\lambda_j\}_{j=1}^d$ with $\lambda_0 = 0 > -\lambda_1 \geq -\lambda_2 \geq \cdots \geq -\lambda_d$, to be the vertices' coordinates in the first $k < d$ dimensions of the embedding. The first eigenvector is discarded as a trivial dimension where every vertex has the same coordinate by construction. Thus, the k-dimensional diffusion map at time t is defined as:

$$\Psi_k^t(\mathbf{x}) := \left(e^{-\lambda_1 t}\psi_1(\mathbf{x}), e^{-\lambda_2 t}\psi_2(\mathbf{x}), \ldots, e^{-\lambda_k t}\psi_k(\mathbf{x})\right),$$

where we have discarded ψ_0 associated with $\lambda_0 = 0$ as a trivial dimension. The time parameter t refers to the diffusion distance after time t which is preserved as Euclidean distance in the embedding space. Trivially, as $t \to \infty$ all network nodes are mapped to the same point since the diffusion distance vanishes.

The parameter α adjusts the effect that the manifold density has on the diffusion process. Choosing $\alpha = 1$ recovers the Laplace-Beltrami operator Δ as the backward diffusion operator, if the points approximately lie on a manifold $\mathcal{M} \subset \mathbb{R}^d$. Thus, the diffusion map corresponds to an embedding of the points unaffected by the manifold density (such that if two different networks were sampled from the same manifold $\mathcal{M}$ but with different densities, we would recover consistent positions of the points on $\mathcal{M}$). Choosing $\alpha = 0$ is equivalent to the *Laplacian eigenmaps* method which preceded diffusion maps [6]. If the vertices are sampled uniformly from the hidden manifold, Laplacian eigenmap becomes equivalent to analysing the Laplace-Beltrami operator, and so constructing a diffusion map with $\alpha = 1$ and with $\alpha = 0$ will recover the same embedding [2].

Consider now an Itô stochastic differential equation (SDE) of the form

$$\dot{\mathbf{x}} = \boldsymbol{\mu}(\mathbf{x}) + \boldsymbol{\sigma}\dot{\mathbf{w}}, \tag{3}$$

where $\mathbf{w}_t$ is the d-dimensional Brownian motion. A probability distribution over the state-space of this system $\phi(\mathbf{x}, t)$ with condition $\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x})$, evolves forward in

time according to the Fokker-Planck equation (FPE), also known as the Kolmogorov forward equation (KFE):

$$\partial_t \phi(\mathbf{x}, t) = -\sum_i \partial_i \left[\mu_i(\mathbf{x})\phi(\mathbf{x}, t)\right] + \sum_i \sum_j \partial_i \partial_j \left[\frac{1}{2}\sigma_i \sigma_j \phi(\mathbf{x}, t)\right], \tag{4}$$

with the sums running over all d dimensions and ∂_i denoting partial derivatives with respect to the ith dimension ($\partial_i = \partial/\partial x_i$) [7]. Similarly, the probability distribution $\psi(\mathbf{y}, s)$ for $s \leq t$ and condition $\psi(\mathbf{y}, t) = \psi_t(\mathbf{x})$ satisfies

$$-\partial_s \psi = \boldsymbol{\mu} \cdot \nabla \psi + \frac{1}{2}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top \Delta \psi, \tag{5}$$

where the differentiations are with respect to $\mathbf{y}$. Terms in the backward FPE become directly identifiable with the backward operator $\mathcal{H}_b^{(\alpha)}$ if we take $\boldsymbol{\sigma} = \sqrt{2}\mathbf{I}$ and $\boldsymbol{\mu} = 2(1-\alpha)\nabla U$.

The original formulation of diffusion maps, as described above, assumes a symmetric kernel $k_\epsilon(\mathbf{x}, \mathbf{y}) = k_\epsilon(\mathbf{y}, \mathbf{x})$. Given a CTMC with a symmetric generator matrix Q, the methodology laid out so far would be sufficient to recover an embedding for the states on a continuous compact manifold $\mathcal{M}$, on which we can define an SDE approximation to the Markov jump process of the CTMC. Encouragingly, it has also been shown that the jump process would satisfy the reflecting (no flux) conditions on the manifold boundary $\partial\mathcal{M}$, as required by a diffusion FP operator defined on such a manifold — i.e. for a point $\mathbf{x} \in \partial\mathcal{M}$ where $\mathbf{n}$ is a normal unit vector at $\mathbf{x}$, and a function $\psi : \mathcal{M} \to \mathbb{R}$,

$$\left.\frac{\partial \psi(\mathbf{x})}{\partial \mathbf{n}}\right|_{\partial\mathcal{M}} = 0.$$

B Embedding unweighted, undirected, grid graphs

Taking the case of a pCTMC produced by a particular class of chemical reaction networks, we show that the embedding produced by *Laplacian eigenmaps* [6] (equivalent to diffusion maps with $\alpha = 0$) for the unweighted, undirected transition matrix, is consistent in some respect to the canonical (manual) embedding for the fluid limit of chemical reaction systems. This implies that we ignore any density information of the vertices (states) on the manifold, and any directional component. We will later return to how this information affects our results.

Laplacian eigenmaps embedding Assume that we have symmetric similarity matrix W between n points. We construct the Laplacian matrix $L = D - W$, with $D_{ii} = \sum_j W_{ji}$. The Laplacian eigenmaps algorithm solves the minimisation problem

$$\underset{\Upsilon^\top D \Upsilon = I}{\operatorname{argmin}} \frac{1}{2}\sum_{i,j} \|\mathbf{y}^{(i)} - \mathbf{y}^{(j)}\|_2^2 \, W_{ij} \tag{6}$$

$$= \underset{\Upsilon^\top D \Upsilon = I}{\operatorname{argmin}} \operatorname{Tr}(\Upsilon^\top L \Upsilon), \tag{7}$$

where $\mathbf{y}^{(i)}$ is the ith row of Υ, and the constraint $\Upsilon^\top D \Upsilon = I$ serves to exclude the trivial solution of mapping everything to the same point. The solution $\Upsilon \in \mathbb{R}^{n \times m}$ is a matrix

with each column vector corresponding to the m-dimensional coordinate embedding of each datum ($m < n$). It is shown that the solution to the problem is the eigenvector matrix corresponding to the m lowest eigenvalues of $L\mathbf{y} = \lambda D\mathbf{y}$, excluding the $\lambda = 0$ solution.

This emphasis on preserving local information allows us to appropriate the algorithm for embedding the network of states without having to calculate global state separation — i.e. by using only neighbouring state similarities as represented in Q. For a CTMC described by a transition matrix Q, we transform Q to be an adjacency matrix between the nodes (states) of the network (CTMC) by placing an undirected edge of weight 1 between states which are separated by a single transition and 0 otherwise:

$$W_{ij} = 1 - \delta_{0,Q_{ij}}\delta_{0,Q_{ji}}. \tag{8}$$

If the network is connected and m (the dimensionality for the embedding space) is picked appropriately, the algorithm will attempt to preserve local dimensions and therefore global ones if the network fits in that m space. If m is chosen higher than necessary, some states which are far apart might be placed closer together in the embedding, but local distances will still be preserved.

The unweighted Laplacian fluid approximation The proof for Theorem B.1 is laid out here. It involves the construction of an undirected, unweighted graph with adjacency matrix W from the Q matrix of a specific kind of pCTMCs, as shown above. Explicit eigenvectors of the Laplacian L of this graph give analytic coordinates for the vertices of Q in some space $\mathbb{R}^d$. A drift vector field is inferred on this space using Gaussian process regression, from Q and the embedding coordinates. We show from these how conditions for a fluid approximation are met, as stated in Section 2(b). Specifically, we show how *initial conditions converge*, *mean dynamics converge*, and *noise converges to zero* (via Taylor expansion of the relevant analytic coordinates), in the same way as in the canonical embedding of such a pCTMC resulting from *hydrodynamic scaling*.

Theorem B.1. *Let $\mathcal{C}$ be a pCTMC, whose underlying transition graph maps to a multi-dimensional grid graph in Euclidean space using the canonical* hydrodynamic scaling *embedding. The* unweighted Laplacian fluid approximation *of $\mathcal{C}$ coincides with the canonical fluid approximation in the hydrodynamic scaling limit.*

Proof. We examine a particular case of pCTMCs, produced by allowing reactions that only change the count of a single species per reaction. This produces an adjacency matrix W for the network of states describing a grid network in d dimensions. Following the derivation for the eigenvectors of the Laplacian L of such a network presented in [8], we find that the lowest eigenvalue λ_1 (excluding $\lambda_0 = 0$) is degenerate ($\lambda_1 = \lambda_{\{2,\ldots,d\}}$), and associated with d eigenvectors $\mathbf{v}_j,\ j \in \{1,\ldots,d\}$. Their elements are

$$\mathbf{v}_{j,[x_1,\ldots,x_d]} = \cos\left(\frac{\pi}{n_j}\left(x_j - \frac{1}{2}\right)\right) \tag{9}$$

where the index $[x_1,\ldots,x_d]$ is the mapping of the node to its integer grid coordinates. Therefore, the embedded jth coordinate of a node is $\cos(\pi/n_j(x_j - 1/2))$, where $x_j \in \{1,\ldots,n_j\}$ is the integer grid position of the node in that j dimension. We observe that away from the boundaries (i.e. near the centre of the grid $x \approx n/2$) and for large n, the

argument of cos is near $\pi/2$, so we approach the linear part of cos. This means that near the centre states are almost uniformly distributed, as in the canonical embedding.

We define the volume $\Omega_U([x_1,\ldots,x_d])$ for a state with grid coordinates $[x_1,\ldots,x_d]$ in the network, to be the volume of the polygon (n-orthotope) whose vertices are that state and the next state along each grid dimension:

$$\Omega_U([x_1,\ldots,x_d]) = \prod_j \left(\mathbf{v}_{j,[x_1,\ldots,x_j+1,\ldots,x_d]} - \mathbf{v}_{j,[x_1,\ldots,x_j,\ldots,x_d]}\right) \tag{10}$$

$$= \prod_j \left[\cos\left(\frac{\pi}{2n_j}\left(2x_j+1\right)\right) - \cos\left(\frac{\pi}{2n_j}\left(2x_j-1\right)\right)\right]. \tag{11}$$

We then observe that $\lim_{n\to\infty}\Omega_U = 0$ for all states; this satisfies the convergence condition of initial states for a fluid approximation.

We define dynamics by means of a drift field $\langle b\rangle : U \to \mathbb{R}^d$. The function is inferred using Gaussian process regression, $b(\cdot) \mid Q \sim \mathcal{GP}(m(\cdot) \mid Q, k(\cdot,\cdot) \mid Q)$, such that it is a Lipschitz field. This satisfies the convergence condition of mean dynamics for a fluid approximation. In the canonical embedding of a pCTMC, the drift vector field is a polynomial function $f_p \in L^2(U)$ over the concentration space. Away from the boundaries, the Laplacian embedding approaches this canonical embedding. As $n \to \infty$, the inferred field in this region will tend to the same polynomial function:

$$\langle b\rangle \to f_p \quad ,$$

as the Gaussian process can approximate any function in $L^2(U)$ arbitrarily well.

Finally, the conditions for noise converging to zero are trivially met, since embedding distances γ are at most $\mathcal{O}\big(n^{-1}\big)$:

$$\begin{aligned}
\gamma &= \cos\left(\frac{\pi}{2n_j}\left(2x_j+1\right)\right) - \cos\left(\frac{\pi}{2n_j}\left(2x_j-1\right)\right) \\
&= 1 - \frac{1}{2!}\left(\frac{\pi}{2n_j}\left(2x_j+1\right)\right)^2 + \frac{1}{4!}\left(\frac{\pi}{2n_j}\left(2x_j+1\right)\right)^4 - \ldots \\
&\quad - 1 + \frac{1}{2!}\left(\frac{\pi}{2n_j}\left(2x_j-1\right)\right)^2 - \frac{1}{4!}\left(\frac{\pi}{2n_j}\left(2x_j-1\right)\right)^4 + \ldots \\
&= \mathcal{O}\big(n_j^{-1}\big),
\end{aligned}$$

and $n = \sum_j n_j$, such that $\gamma^2 = \mathcal{O}\big(n^{-2}\big)$.

Thus the criteria for *fluid approximation* of this pCTMC are satisfied. Further, for some region of the state-space and in the limit of infinite states, this construction is consistent with the embedding and dynamics recovered by *hydrodynamic scaling*, the canonical *fluid approximation* of a pCTMC. This concludes our proof. □

C Diffusion maps for directed graphs

Our focus necessarily shifts on embedding an arbitrary CTMC with no symmetry condition on Q. Following Perrault-Joncas and Meilă [9] assume that we observe a graph G, with nodes sampled from a diffusion process on a manifold $\mathcal{M}$ with density $p = e^{-U}$

and edge weights given by the (non-symmetric) kernel k_ϵ. The directional component of the kernel is further assumed to be derived from a vector field $\mathbf{r}$ on $\mathcal{M}$ without loss of kernel generality. As the authors saliently put it: "The question is then as follows: can the generative process' geometry $\mathcal{M}$, distribution $p = e^{-U}$, and directionality $\mathbf{r}$, be recovered from G?"

In the same manner as for the original formulation of diffusion maps a set of backward evolution operators are derived, the two relevant ones being:

$$-\partial_t = \mathcal{H}^{(\alpha)}_{aa} = \Delta + (\mathbf{r} - 2(1-\alpha)\nabla U)\cdot\nabla, \qquad \text{and} \tag{12}$$

$$-\partial_t = \mathcal{H}^{(\alpha)}_{ss} = \Delta - 2(1-\alpha)\nabla U\cdot\nabla. \tag{13}$$

To construct this family of operators, the kernel is first decomposed into its symmetric h_ϵ and anti-symmetric a_ϵ parts,

$$k^{(\alpha)}_\epsilon(\mathbf{x},\mathbf{y}) = \frac{k_\epsilon(\mathbf{x},\mathbf{y})}{p^\alpha_\epsilon(\mathbf{x})p^\alpha_\epsilon(\mathbf{y})} = \frac{1}{p^\alpha_\epsilon(\mathbf{x})p^\alpha_\epsilon(\mathbf{y})}\left[h_\epsilon(\mathbf{x},\mathbf{y}) + a_\epsilon(\mathbf{x},\mathbf{y})\right],$$

and further normalised according to either the asymmetric $d^{(\alpha)}_\epsilon(\mathbf{x}) = \int_{\mathcal{M}} k^{(\alpha)}_\epsilon(\mathbf{x},\mathbf{y})p(\mathbf{y})d\mathbf{y}$, or symmetric outdegree distribution $\tilde{d}^{(\alpha)}_\epsilon(\mathbf{x}) = \int_{\mathcal{M}} h^{(\alpha)}_\epsilon(\mathbf{x},\mathbf{y})p(\mathbf{y})d\mathbf{y}$. The subscript indices denote the type of kernel used to construct the operator and the outdegree distribution used to normalise it (such that $\mathcal{H}_{aa}$ associates to the full asymmetric kernel $k^{(\alpha)}_\epsilon$ normalised with asymmetric degree distribution p_ϵ, and so on).

Discrete approximations for these operators can be constructed for an asymmetric kernel matrix of distances between N high-dimensional points, $\mathbf{W} \in \mathbb{R}^{N\times N}$. The symmetric matrix $H^{(1)}_{ss} \in \mathbb{R}^{N\times N}$ can be extracted and the necessary eigen-decomposition carried out to yield an embedding, where $\lim_{N\to\infty} H^{(1)}_{ss} = \mathcal{H}^{(1)}_{ss} = \Delta$. However, given the infinitesimal generator of a CTMC Q, we do not have access to $\mathbf{W}$, but rather to the discrete approximation of the final evolution operator, $\lim_{N\to\infty} Q = \mathcal{H}^{(\alpha)}_{aa}$. In order to recover the initial kernel matrix $\mathbf{W}$ that gave rise to Q, we take $\alpha = 0$, a uniform measure on the manifold $U(\mathbf{x}) = 0 \implies p(\mathbf{x}) = 1$, and a small value for ϵ. This makes the transformations from $\mathbf{W}_\epsilon$ to Q reversible, since

$$Q = \lim_{\epsilon\to 0}\left[\frac{T^{(\alpha=0)}_\epsilon - \mathbf{I}}{\epsilon}\right], \text{ and} \tag{14}$$

$$T^{(\alpha=0)}_\epsilon = D^{-1}\mathbf{W}_\epsilon, \text{ such that} \tag{15}$$

$$\mathbf{W}_\epsilon = D(\mathbf{I} + \epsilon Q.) \tag{16}$$

In the above, D is a diagonal matrix which forces the diagonal of $\mathbf{W}_\epsilon$ to be 1, as expected from a distance-based kernel matrix. The final step is the familiar *uniformisation* procedure which approximates a CTMC with a DTMC. The choice of $\epsilon < (\max_i |Q_{ii}|)^{-1}$ determines the quality of approximation (the smaller the better).

Once the kernel matrix $\mathbf{W}_\epsilon$ is recovered we can proceed to construct the operators $\Delta = \mathcal{H}^{(1)}_{ss}$ and $\left(\mathcal{H}^{(0)}_{aa} - \mathcal{H}^{(1)}_{ss}\right) = (\mathbf{r} - 2\nabla U)\cdot\nabla$, which are used to embed the state-space on a manifold $\mathcal{M} \in \mathbb{R}^d$, and endow it with the advective field $\boldsymbol{\mu} = (2\nabla U + \mathbf{r})$ in the Kolmogorov backward equation, respectively. See Algorithm 1 for procedural details of the *geometric fluid approximation.*

References

[1] Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, Zucker SW. 2005 Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences* **102**, 7426–7431.

[2] Coifman RR, Lafon S. 2006 Diffusion maps. *Applied and Comp. Harm. Anal.* **21**, 5–30.

[3] Nadler B, Lafon S, Kevrekidis I, Coifman RR. 2006a Diffusion Maps, Spectral Clustering and Eigenfunctions of Fokker-Planck Operators. In Weiss Y, Schölkopf B, Platt JC, editors, *Advances in NeurIPS 18* pp. 955–962. MIT Press.

[4] Nadler B, Lafon S, Coifman RR, Kevrekidis IG. 2006b Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Comp. Harm. Anal.* **21**, 113–127.

[5] Coifman RR, Kevrekidis IG, Lafon S, Maggioni M, Nadler B. 2008 Diffusion Maps, Reduction Coordinates, and Low Dimensional Representation of Stochastic Systems. *Multiscale Modeling & Simulation* **7**, 842–864.

[6] Belkin M, Niyogi P. 2003 Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation* **15**, 1373–1396.

[7] Gardiner CW. 2009 *Stochastic methods: a handbook for the natural and social sciences*. Springer series in synergetics. Berlin: Springer 4th ed edition.

[8] Kłopotek MA. 2017 Spectral Analysis of Laplacian of a Multidimensional Grid Graph. *arXiv:1707.05210 [math]*. arXiv: 1707.05210.

[9] Perrault-Joncas DC, Meilă M. 2011 Directed Graph Embedding: an Algorithm based on Continuous Limits of Laplacian-type Operators. In Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira F, Weinberger KQ, editors, *Advances in NeurIPS 24* pp. 990–998. Curran Associates, Inc.