Factorized Diffusion Map Approximation: Detailed Proofs

Saeed Amizadeh
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, PA 15213

Hamed Valizadegan
Department of Computer Science
University of Pittsburgh
Pittsburgh, PA 15213

Milos Hauskrecht
Department of Computer Science
University of Pittsburgh
Pittsburgh, PA 15213

Proof of Lemma 1:

Proof. Let $\boldsymbol{\mathcal{T}_k} = \{T_1, T_2, \dots, T_k\}$ be a partition of the variables in V and $q(V) = \prod_{i=1}^k p_i(T_i)$. Now assume $x = [x_1, x_2, \dots, x_k]^T$ and $y = [y_1, y_2, \dots, y_k]^T$ are two random vectors partitioned according to $\boldsymbol{\mathcal{T}_k}$ (note that x_i's and y_i's are sub-vectors). Then we will have:

$$a_\varepsilon(x, y) = \prod_{i=1}^k a_\varepsilon(x_i, y_i)$$

and therefore,

$$a_t(x, y) = \lim_{\varepsilon\to 0} a_{\varepsilon,t/\varepsilon}(x, y) = \lim_{\varepsilon\to 0} \prod_{i=1}^k a_{\varepsilon,t/\varepsilon}(x_i, y_i) = \prod_{i=1}^k a_t(x_i, y_i)$$

Now let $\psi^t(x) = \prod_{i=1}^k \psi^t_{i,m_i}(x_i)$ and $\lambda^{(t)} = \prod_{i=1}^k \lambda^{(t)}_{i,m_i}$ where $A^t_{p_i}[\psi^t_{i,m_i}(x_i)] = \lambda^{(t)}_{i,m_i}\psi^t_{i,m_i}(x_i)$, then we will have:

$$\begin{aligned} A^t_q[\psi^t(x)] &= \int a_t(x, y)\psi^t(y)q(y)dy \\ &= \prod_{i=1}^k \int a_t(x_i, y_i)\psi^t_{i,m_i}(y_i)p_i(y_i)dy_i \\ &= \prod_{i=1}^k \lambda^{(t)}_{i,m_i}\psi^t_{i,m_i}(x_i) = \lambda^{(t)}\psi^t(x) \end{aligned}$$

Therefore, $\psi^t(x)$ is an eigenfunction of A^t_q with eigenvalue $\lambda^{(t)}$. ☐

Proof of Lemma 2:

Proof. Suppose ψ is the m-th eigenfunction of A^t_p associated with the m-th largest eigenvalue λ constructed using Lemma 1. That is, we have that $\psi = \prod_{i=1}^k \psi_i$ and $\lambda = \prod_{i=1}^k \lambda_i$ where ψ_i is an eigenfunction of $A^t_{p_i}$ in the subspace T_i associated with eigenvalue λ_i. Now suppose, ψ consists of eigenfunctions from only $\ell < k$ subspaces; that is, only ℓ of the eigenfunctions in the product above are non-constant (non-trivial) with eigenvalues strictly less than 1, while the rest of them are constant with eigenvalues equal to 1. Now if any of these ℓ eigenfunctions is replaced by the constant eigenfunction (and its corresponding eigenvalue with 1) we will have a new valid pair of eigenvalue and eigenfunction $\langle\lambda', \psi'\rangle$ for $A^t_{p_i}$ where $\lambda' > \lambda$. Using this replacement method, we can generate 2^ℓ new pairs with eigenvalues all greater than λ. However, since λ is the m-th largest eigenvalue of A^t_p, we must have $m \geq 2^\ell$ or equivalently $\ell \leq \lceil\lg m\rceil$. On the other hand, the number involved subspaces ℓ cannot be greater than k which means that $\ell \leq \min(k, \lceil\lg m\rceil)$. ☐

Proof of Theorem 1:

Proof. From [2], we have that:

$$\|\psi^t_{p,m} - \psi^t_{q,m}\|^2_2 \leq \frac{16}{\delta^2_m}\|A^t_p - A^t_q\|^2 \tag{1}$$

where

$$\|A^t_p - A^t_q\|^2 = \sup_{\|f\|\leq 1}\|A^t_p[f(x)] - A^t_q[f(x)]\|^2_2 =$$

$$\sup_{\|f\|\leq 1}\left\|\int a_t(x, y)f(y)p(y)dy - \int a_t(x, y)f(y)q(y)dy\right\|^2_2$$

$$= \sup_{\|f\|\leq 1}\left\|\int a_t(x, y)f(y)[p(y) - q(y)]dy\right\|^2_2 =$$

$$\sup_{\|f\|\leq 1}\int\left(\int a_t(x, y)f(y)[p(y) - q(y)]dy\right)^2 p(x)dx \leq$$

$$\sup_{\|f\|\leq 1}\int\left(\int |a_t(x, y)||f(y)||p(y) - q(y)|dy\right)^2 p(x)dx$$

$$= \sup_{\|f\|\leq 1}\int\left(\int |p(y') - q(y')|dy' \times\right.$$

$$\left.\int |a_t(x, y)||f(y)|\frac{|p(y) - q(y)|}{\int |p(y') - q(y')|dy'}dy\right)^2 p(x)dx$$

$$\leq \sup_{\|f\|\leq 1} \int \Bigg(\Big(\int |p(y') - q(y')| dy' \Big)^2 \times$$

$$\int a_t^2(x,y) f^2(y) \frac{|p(y)-q(y)|}{\int |p(y')-q(y')|dy'} dy \Bigg) p(x) dx$$

$$\leq \int \Bigg(\Big(\int |p(y') - q(y')| dy' \Big)^2 \times$$

$$\int \jmath^2 \ell^2 \frac{|p(y)-q(y)|}{\int |p(y')-q(y')|dy'} dy \Bigg) p(x) dx$$

$$= \int \jmath^2 \ell^2 \|p-q\|_1^2 p(x) dx = \jmath^2 \ell^2 \|p-q\|_1^2 \quad (2)$$

On the other hand we have the following inequality [1]:

$$\|p-q\|_1 \leq \sqrt{2 \ln 2 \cdot D_{KL}(p\|q)} \quad (3)$$

Therefore, we have:

$$\|\psi_{p,m}^t - \psi_{q,m}^t\|_2^2 \leq \frac{16}{\delta_m^2} \|A_p^t - A_q^t\|^2 \leq \frac{16}{\delta_m^2} \jmath^2 \ell^2 \|p-q\|_1^2$$

$$\leq \frac{32 \ln 2}{\delta_m^2} \jmath^2 \ell^2 \cdot D_{KL}(p\|q) \quad (4)$$

□

Proof of Theorem 2:

Proof. Let $\boldsymbol{T_k} = \{T_1, T_2, \ldots, T_k\}$ be a partition of the variables in V into k subspaces. Also, let $\boldsymbol{\Lambda_i^t} = \{\lambda_{i,m}^{(t)} \mid 1 \leq m \leq \infty\}$ be the set of eigenvalues of the marginal diffusion operator $A_{p_i}^t$ on subspace T_i for all $1 \leq i \leq k$. Assume the members of $\boldsymbol{\Lambda_i^t}$ are sorted in the decreasing order with the first (the largest) eigenvalue $\lambda_{i,1}^{(t)} = 1$ associated with the constant eigenfunction $\psi_{i,1}^t = 1$.

Using Lemma 1, the eigenvalues (and their associated eigenfunctions) of A_q^t are constructed by picking one eigenvalue from each set $\boldsymbol{\Lambda_i^t}$ for all $1 \leq i \leq k$ and multiply them together; that is, the $\lambda_{q,m}^{(t)} = \prod_{i=1}^k \lambda_{i,j_i}^{(t)}$ is the m-th largest eigenvalue of A_q^t. For each m, we can find the index tuple $(j_1, \ldots, j_k)$ indicating which eigenvalue is exactly picked in each subspace to construct the m-th largest eigenvalue of A_q^t. If we know the index tuple $(j_1, \ldots, j_k)$ for the m-th eigenfunction, we can find the upper bound on the estimation error as follows:

$$\|\psi_{q,m}^t - \hat{\psi}_{q,m}^t\|_2^2 = \Big\| \prod_{i=1}^k \psi_{i,j_i}^t - \prod_{i=1}^k \hat{\psi}_{i,j_i}^t \Big\|_2^2$$

$$\leq 2 \Big\| \prod_{i=1}^k \psi_{i,j_i}^t - \psi_{1,j_1}^t \prod_{i=2}^k \hat{\psi}_{i,j_i}^t \Big\|_2^2$$

$$+ 2 \Big\| \psi_{1,j_1}^t \prod_{i=2}^k \hat{\psi}_{i,j_i}^t - \prod_{i=1}^k \hat{\psi}_{i,j_i}^t \Big\|_2^2$$

$$= 2 \Big\| \psi_{1,j_1}^t \Big(\prod_{i=2}^k \psi_{i,j_i}^t - \prod_{i=2}^k \hat{\psi}_{i,j_i}^t \Big) \Big\|_2^2$$

$$+ 2 \Big\| (\psi_{1,j_1}^t - \hat{\psi}_{1,j_1}^t) \prod_{i=2}^k \hat{\psi}_{i,j_i}^t \Big\|_2^2$$

$$\leq 2\ell^2 \cdot \Big\| \prod_{i=2}^k \psi_{i,j_i}^t - \prod_{i=2}^k \hat{\psi}_{i,j_i}^t \Big\|_2^2$$

$$+ 2\ell^{2(k-1)} \cdot \|\psi_{1,j_1}^t - \hat{\psi}_{1,j_1}^t\|_2^2 \quad (5)$$

Using the above derivation recursively, we get:

$$\|\psi_{q,m}^t - \hat{\psi}_{q,m}^t\|_2^2 \leq \ell^{2(k-1)} \sum_{i=1}^k 2^i \|\psi_{i,j_i}^t - \hat{\psi}_{i,j_i}^t\|_2^2$$

$$= \ell^{2(k-1)} \sum_{\substack{i=1 \\ j_i \neq 1}}^k 2^i \|\psi_{i,j_i}^t - \hat{\psi}_{i,j_i}^t\|_2^2 \quad (6)$$

The second equality in Eq. (6) comes from the fact that for $j_i = 1$, $\psi_{i,j_i}^t = \hat{\psi}_{i,j_i}^t = 1$. Since we don't know the true eigenvalues in each subspace, we cannot identify the index tuple $(j_1, \ldots, j_k)$ for a given index m. As a result the above bound is replaced by the worst case scenario across all possible index tuples, that is:

$$\|\psi_{q,m}^t - \hat{\psi}_{q,m}^t\|_2^2 \leq \max_{(j_1,\ldots,j_k)} \ell^{2(k-1)} \sum_{\substack{i=1 \\ j_i \neq 1}}^k 2^i \|\psi_{i,j_i}^t - \hat{\psi}_{i,j_i}^t\|_2^2$$

However, because $\psi_{q,m}^t$ is associated with the m-th largest eigenvalue of A_q^t (i.e. $\lambda_{i,1}^{(t)}$), not all combinations for the index tuple should be considered in taking the maximum. More precisely, if we replace any of the indices j_i in the index tuple $(j_1, \ldots, j_k)$ with a smaller index $j'_i < j_i$, the resulted multiplicative eigenvalue will become larger; this is because of the fact that smaller indices in each set $\boldsymbol{\Lambda_i^t}$ correspond to larger eigenvalues. The total number of such replacements for the index tuple $(j_1, \ldots, j_k)$ is $\prod_{i=1}^k j_i$. This means that if the index tuple for the m-th largest eigenvalue of A_q^t is $(j_1, \ldots, j_k)$, m must be greater than $\prod_{i=1}^k j_i$. In other words, the valid index tuples for the m-th largest eigenvalue must satisfy $\prod_{i=1}^k j_i < m$. If S_m denotes the set of such tuples, we can improve the bound as:

$$\|\psi_{q,m}^t - \hat{\psi}_{q,m}^t\|_2^2 \leq \max_{(j_1,\ldots,j_k)\in S_m} \ell^{2(k-1)} \sum_{\substack{i=1 \\ j_i \neq 1}}^k 2^i \|\psi_{i,j_i}^t - \hat{\psi}_{i,j_i}^t\|_2^2$$

Now, using Eq. (5) in the paper we get:

$$\|\psi_{q,m}^t - \hat{\psi}_{q,m}^t\|_2^2$$

$$= O_P\Bigg(\max_{(j_1,\ldots,j_k)\in S_m} \ell^{2(k-1)} \sum_{\substack{i=1 \\ j_i \neq 1}}^k \frac{2^i t \sqrt{d_i}}{\mu_{i,j_i}^{(t)}} \Big[\frac{\log n}{n} \Big]^{2/(d_i+8)} \Bigg)$$

Moreover, using Lemma 2, there at most $\min(k, \lceil \lg m \rceil)$ non-constant eigenvectors contributing in constructing $\psi^t_{q,m}$ which means the sum in the above bound has at most $\min(k, \lceil \lg m \rceil)$ terms. $\square$

References

[1] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, USA, 2000.

[2] Laurent Zwald and Gilles Blanchard. On the convergence of eigenspaces in kernel principal component analysis. In *NIPS*, 2005.