Supplemental to
'Context-guided diffusion for label propagation on graphs'

Kwang In Kim
Lancaster University

James Tompkin
Harvard Paulson SEAS

Hanspeter Pfister
Harvard Paulson SEAS

Christian Theobalt
MPI for Informatics

We present a procedure to construct a consistent graph diffusivity operator $D^{\mathcal{C}}$. We reproduce some contents from the main paper to minimize cross-referencing.

Suppose we are given a graph $G=(X,E,W)$ where X is the set of nodes $\{\mathbf{x}_1,...,\mathbf{x}_n\}$, E is the set of edges $\{e_{ij}\}\subset X\times X$, and weights $W=\{w_{ij}\}$ represent non-negative node similarities at each edge:

$$w_{ij}=\begin{cases}\exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{\sigma_{\mathbf{x}}}\right) & \text{if } \mathbf{x}_i\in N_K(\mathbf{x}_j) \\ & \text{ or } \mathbf{x}_j\in N_K(\mathbf{x}_i) \\ 0 & \text{otherwise,}\end{cases} \quad (1)$$

where $N_K(\mathbf{x}_i)$ is the K-nearest neighborhood of $\mathbf{x}_i$.

For each node $\mathbf{x}_i\in X$, a subgraph $G_i=(X_i,E_i,W_i)$ is defined as the set of nodes that are connected to $\mathbf{x}_i$ and the corresponding edges, i.e., $X_i=\{\mathbf{x}_j|e_{ij}\in E\}$ and $E_i=\{e_{ij}|\mathbf{x}_j\in X_i\}$. The graph gradient operator $\nabla_i: H(X_i)\rightarrow H(E_i)$ and the divergence operator $\nabla_i^*: H(E_i)\rightarrow H(X_i)$ on G_i are respectively defined as:

$$\begin{aligned}[\nabla_i f](e_{ij}) &= \sqrt{w_{ij}}(f(j)-f(i)), \forall e_{ij}\in E_i, f\in H(X_i)\\ [\nabla_i^* S](i) &= \frac{1}{2d_i}\sum_{j=1}^{n}\sqrt{w_{ji}}(S(j,i)-S(i,j)), \forall S\in H(E_i).\end{aligned} \quad (2)$$

Now, the (anisotropic) diffusivity coefficients $\{q_{ij}\}$ are defined as eigenvalues of each local graph diffusivity operator D_i:

$$\begin{aligned}D_i &:= \sum_{j\neq i,\mathbf{x}_j\in X_i} q_{ij}\mathbf{b}_{ij}\otimes\mathbf{b}_{ij}\\ \Leftrightarrow D_i(S)(e_{ij}) &= q_{ij}\mathbf{b}_{ij}\langle\mathbf{b}_{ij},S\rangle, \forall S\in H(E_i),\end{aligned} \quad (3)$$

where $\otimes$ is the tensor product on $H(E_i)$ and the eigenfunction $\mathbf{b}_{ij}$ is the indicator function of the edge e_{ij}.

Given the inner-product structures and the diffusivity operator $D=\{D_i\}$, our anisotropic graph Laplacian is obtained as:

$$\begin{aligned}[L^D f](i) &:= [\nabla_i^* D_i\nabla_i f](i),\\ &= \left(\frac{1}{d_i}\sum_{j=1}^{n} w_{ij}^D\right) f(i)-\frac{1}{d_i}\sum_{j=1}^{n} w_{ij}^D f(j),\end{aligned} \quad (4)$$

with:

$$w_{ij}^D = w_{ij}q_{ij}. \quad (5)$$

With the anisotropic graph Laplacian L^D, our diffusion algorithm is stated as:

$$\Leftrightarrow f^{t+1}=f^t-\delta L^D f^t. \quad (6)$$

Our strategy to build a consistent graph diffusivity operator $D=\{D_i\}$ is to enforce each D_i to agree with an explicit form of the analytic diffusivity operator $\mathcal{D}:\mathcal{T}(M)\rightarrow\mathcal{T}(M)$ with $\mathcal{T}(M)$ being the tangent bundle of M: The restriction $\mathcal{D}_i: T_{\mathbf{x}_i}(M)\rightarrow T_{\mathbf{x}_i}(M)$ of $\mathcal{D}$ at each point $\mathbf{x}_i\in M$, is defined as

$$\mathcal{D}_{\mathbf{x}_i} := \sum_{k,l=1}^{m} q_{kl}^{\mathcal{D}}\mathbf{b}^k\otimes\mathbf{b}'^l \quad (7)$$

$$\Leftrightarrow \mathcal{D}_{\mathbf{x}_i}(\mathbf{v},\mathbf{w}) = \sum_{k,l=1}^{m} q_{kl}^{\mathcal{D}}\langle\mathbf{b}^k,\mathbf{w}\rangle_g\otimes\langle\mathbf{b}^l,\mathbf{v}\rangle_g, \quad (8)$$

where $\mathbf{v},\mathbf{w}\in T_{\mathbf{x}_i}(M)$, $\otimes$ is the tensor product on $\mathcal{T}(M)$, $\{\mathbf{b}^k\}$ is an orthonormal basis vectors with respect to g in $T_{\mathbf{x}_i}(M)$, and $\mathbf{b}'^k$ is the dual co-vector of $\mathbf{b}^k$. The coefficients $\{q_{kl}^{\mathcal{D}}\}$ are given in *Riemannian normal coordinates* where the metric g becomes Euclidean up to second order [4]. The second line is obtained by converting $\mathbf{b}^k$ to $\mathbf{b}'^k$ using the metric g in this coordinate.

The tangent space $T_{\mathbf{x}_i}(M)$ and the orthonormal basis $\{\mathbf{b}^k\}$ therein can be estimated by performing the principal component analysis on $N_K(\mathbf{x}_i)$ [1]. Due to the positive definiteness requirement, $\mathcal{D}_{\mathbf{x}_i}$ has $\frac{m(m+1)}{2}$ degree of freedom with m being the dimensionality of M: $\frac{m(m-1)}{2}$ rotations for its eigenvectors plus m-parameters for the corresponding eigenvalues.[1] The construction of the coefficients $\{q_{kl}^{\mathcal{D}}\}$ given D_i will be presented shortly.

Now, given $\mathcal{D}_{\mathbf{x}_i}$, an $\frac{m(m+1)}{2}$-dimensional graph diffusivity operator $D_i: H(E_i)\rightarrow H(E_i)$ is *restored* by enforcing the distance measure as induced by D_i to agree with the corresponding distance measured with $\mathcal{D}_{\mathbf{x}_i}$. First, each input difference vector $\mathbf{x}_j-\mathbf{x}_i$ (aligned with the edge $e_{ij}\in E_i$) is projected onto $T_{\mathbf{x}_i}(M)$. Then, the length of the resulting projection $\text{proj}_{\mathbf{x}_i}[\mathbf{x}_j-\mathbf{x}_i]$ as measured in $T_{\mathbf{x}_i}(M)$ based on $\mathcal{D}_{\mathbf{x}_i}$ is *assigned* as the induced distance on e_{ij} in the graph:

$$q_{ij}^{\mathcal{C}} \Leftarrow \mathcal{D}_{\mathbf{x}_i}(\text{proj}_{\mathbf{x}_i}[\mathbf{x}_j-\mathbf{x}_i],\text{proj}_{\mathbf{x}_i}[\mathbf{x}_j-\mathbf{x}_i]), \quad (9)$$

[1] Note $\{\mathbf{b}^k\}$ is an orthonormal basis in $T_{\mathbf{x}_i}(M)$ which is in general, different from the set of eigenvectors of $\mathcal{D}_{\mathbf{x}_i}$.

where the assignment (denoted with '$\Leftarrow$') is completed by equating the right hand side with:

$$d^2(\mathbf{x}_j,\mathbf{x}_i) = \|\mathbf{x}_j - \mathbf{x}_i\|^2 - \log q^{\mathcal{C}}_{ij}\sigma^2_{\mathbf{x}}, \tag{10}$$

which is obtained by representing w^D_{ij} (Eqs. 5) as the Gaussian envelop of the new distance $d^2(\mathbf{x}_j,\ \mathbf{x}_i)$, i.e., $w^D_{ij} = \exp\left(-\frac{d^2(\mathbf{x}_j,\mathbf{x}_i)}{\sigma^2_{\mathbf{x}}}\right)$ (see Eq. 1). Finally, in this setting, the coefficients $\{q^{\mathcal{D}}_{kl}\}$ of $\mathcal{D}_{\mathbf{x}_i}$ are defined as the minimizer of the squared loss:

$$L(\{q^{\mathcal{D}}_{kl}\}) = \sum_{j=1,\ldots,K} (q_{ij} - q^{\mathcal{C}}_{ij})^2, \tag{11}$$

with positive definiteness condition. This can be formulated as a convex optimization on a cone and it can be solved in various ways, for example, by using projected gradient descent.

As $n \to \infty$ and the diameter of neighborhood size $N_K(\mathbf{x}_i)$ shrink to zero, the difference vector $\mathbf{x}_j - \mathbf{x}_i$ defined in the ambient Euclidean space converges the tangent vector $\text{proj}_{\mathbf{x}_i}[\mathbf{x}_j - \mathbf{x}_i]$ [2]. Now, using Proposition 1 of the main paper, the point-wise convergence of the corresponding anisotropic Laplacian L_D to $\Delta_{\mathcal{D}}$ is obtained as a corollary of Theorem 3 of [2].

In general, for nonlinear anisotropic diffusion processes, $\mathcal{D}$ (and accordingly $\overline{g}$) depends on f, and the corresponding anisotropic Laplacian $\Delta^{\mathcal{D}}$ also depends on f. In this case, the convergence of L^D to $L^{\mathcal{D}}$ is not defined as $L^{\mathcal{D}}$ itself is not defined without making reference to f. Instead, the convergence of the anisotropic difference equation (Eq. 6) to the continuous differential equation can be established:

$$\frac{\partial f}{\partial t} = -\Delta_{\mathcal{D}} f. \tag{12}$$

See [3] for the analysis of this type of convergence.

References

[1] D. L. Donoho and C. Grimes. Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. *PNAS*, 100(10):5591–5596, 2003. 1

[2] M. Hein, J.-Y. Audibert, and U. von Luxburg. From graphs to manifolds - weak and strong pointwise consistency of graph Laplacians. In *Proc. COLT*, pages 470–485, 2005. 2

[3] M. Hein and M. Maier. Manifold denoising. In *NIPS*, pages 561–568, 2007. 2

[4] J. Jost. *Riemannian Geometry and Geometric Analysis*. Springer, New York, 6th edition, 2011. 1