Local High-order Regularization on Data Manifolds

Kwang In Kim1, James Tompkin2 Hanspeter Pfister2 Christian Theobalt3
1Lancaster University 2Harvard SEAS 3MPI for Informatics

The graph Laplacian regularizer is a first-order regularizer which can lead to degenerate functions in high-dimensional manifolds. The iterated graph Laplacian enables high-order regularization, but it has a high computational cost. Our new regularizer is globally high order and so not degenerate, and is also sparse for efficient computation. We build a local first-order approximation of the manifold as a surrogate geometry, and construct our high-order regularizer based on local derivative evaluations therein. Supplemental MATLAB code is available.

One of the best established regularizers for semi-supervised learning is the graph Laplacian $L$:

$$\mathcal{R}_L(f) := \|f\|_2^2 := \sum_{i,j=1}^n |W|_{ij} (f_i - f_j)^2,$$  \hspace{1cm} (1)

This converges to the Laplace-Beltrami operator $\Delta$ on the underlying data-generating $M$ of dimension $m$ [1], which measures the first-order variations of a continuously differentiable function $f$ on $M$. However, the convergence to $\Delta$ reveals an important shortcoming: For high-dimensional manifolds ($m > 1$), the null space of $\Delta$ includes discontinuous functions on $M$, e.g., “spiky” Dirac delta-like functions $f$, with norm $\|f\|_{\Delta}$ = 0 (Fig. 1). This is important because we commonly minimize the regularized risk of attaining a zero value by such a function, and so no generalization is obtained.

Zhou and Belkin [3] prevent this degeneracy by iterating powers of the iterated graph Laplacian with $p$ = 0, which measures the first-order variations of $f$ on $M$, and is also sparse for efficient computation. We build a local first-order approximation of the manifold as a surrogate geometry, and construct our high-order regularizer based on local derivative evaluations therein. Supplemental MATLAB code is available.

AlGORITHM 1: The construction of the regularization functional $\mathcal{R}_G$ from a point cloud $\mathcal{X}$.

\textbf{Input:} $\mathcal{X} = \{X_1, \ldots, X_n\}$, manifold dimension $n$, $k$.
\textbf{Output:} $G$.
1 Initialization: Find $k$ nearest neighbors, e.g., build KD-tree;
2 for $i = 1, \ldots, n$ do
3 \hspace{1cm} Construct the local approximation $M_{k}(X_i)$ using $n$-dimensional PCA of $N_k(X_i)$;
4 \hspace{1cm} Calculate the local regularization matrix $G_i$ for $N_k(X_i)$ in the PCA representation: $G_i = (L_i^T (K_i)^T + L_i)$ (Eqs. 6 and 7);
5 end
6 Re-arrange $\{G_i\}$ according to the indices of $\{f\}$ in $f$ to construct matrix $G$ s.t. $f^T G f = \mathcal{R}_G(f)$.

This still requires explicit derivative calculation. However, for the special case of Eq. 3 with coefficients $\{c_k\}$ given as $c_k = \frac{\sigma^2}{\sigma_k^2}$ with $\sigma^2$ defined as the bandwidth in a Gaussian kernel interpolation, we can efficiently calculate an approximation: First, the local energy $q'$ over $T_X$, defined as

$$\|q'\|^2_2 := \sum_{k=1}^n \sum_{x \in T_X} |Dk'q'(x)|^2 dx = \|q'\|_{K}^2,$$  \hspace{1cm} (5)

This can be analytically evaluated as the corresponding Gaussian reproducing kernel Hilbert space (RKHS) norm $\|q\|_{K}$: The second equality is one of the central results in regularization theory [2]. This is always possible as $q'$ has $k$ degrees of freedom, and leads to an Euler-Lagrange equation that renders $k$ as Green’s result of our operator $D$. Then, we build a new regularizer $\mathcal{R}_G$ as a combination of local regularizers on $q' - f(X_i)$ for $i = 1, \ldots, n$:

$$\mathcal{R}_G(f) = \sum_{i=1}^n f^T G f,$$  \hspace{1cm} (6)

$$f^T G f = \|f(X_i) - q'(i)\|^2_{K} = f^T (I - 11_i^T K_i^T (I - 11_i)^T f,$$  \hspace{1cm} (7)

We use a local first-order approximation $T_X(M)$ of manifold $M$ at each point $X$ as a proxy geometry for $M$ near $X$. Since $T_X(M)$ is identified with $\mathbb{R}^m$, evaluating the derivative operators on $X$ boils down to the calculation of the derivative operators in Euclidean geometry. Evaluating the Laplace-Beltrami operator becomes the calculation of the Laplacian operator:

$$D_0^2 f|X = \Delta_0 f|X = \sum_{r=1}^m \frac{\partial^2 f|X}{\partial X^r \partial X^r} = \sum_{r=1}^m \frac{\partial^2 f|X}{\partial X^r \partial X^r},$$  \hspace{1cm} (4)

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.