Local High-order Regularization on Data Manifolds

Kwang In Kim[1], James Tompkin[2] Hanspeter Pfister[2] Christian Theobalt[3]
[1]Lancaster University [2]Harvard SEAS [3]MPI for Informatics

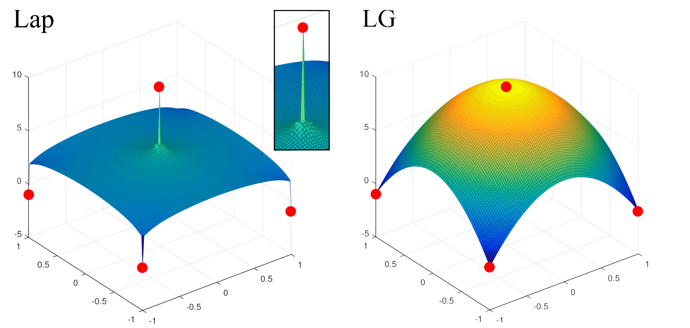

Figure 1: The graph Laplacian regularizer (*Lap*) fails to regularize this toy example, leading to spiky functions. Our *local Gaussian* or *LG* regularizer leads to smooth functions.

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:

$$\mathcal{R}_L(\mathbf{f}) := \mathbf{f}^\top L\mathbf{f} = \sum_{i,j=1}^{u} [W]_{ij}(f_i - f_j)^2, \tag{1}$$

This converges to the Laplace-Beltrami operator Δ 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 Δ reveals an important shortcoming: For high-dimensional manifolds ($m > 1$), the null space of Δ includes discontinuous functions on M, e.g., "spiky" Dirac delta-like functions f, with norm $\|f\|_\Delta^2 = 0$ (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 graph Laplacian (with $p > \frac{m}{2}$):

$$\mathcal{R}_{L^p}(\mathbf{f}) := \mathbf{f}^\top L^p\mathbf{f}, \tag{2}$$

While this adds higher order terms to the regularization matrix, it also makes L denser, which leads to higher computational cost.

We build a global regularization matrix G based on local high-order (p) differential operators D evaluated at each point in $\mathcal{X}$.

$$\|f\|_D^2 := \int_M \sum_{p=1}^{\infty} c_p |D^p f|_X|^2 dV(X), \tag{3}$$

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:

$$D_0^2 f|_X = \Delta_0 f|_X = \sum_{r=1}^{m} \frac{\partial}{\partial X^r}^2 f|_X. \tag{4}$$

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

Algorithm 1: The construction of the regularization functional $\mathcal{R}_{G'}$ from a point cloud $\mathcal{X}$.

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

Table 1: Mean L2-reconstruction error on the MoCap dataset.

Algorithm	*Lap*	*i-Lap*	*LG*
Joint angles error	1.62	1.24	**1.16**
Joint locations error	1.22	0.72	**0.50**

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

$$\|q^i\|_D^2 := \sum_{k=1}^{\infty} c_k \int_{T_{X_i}(M)} |D^k q^i|_{\mathbf{x}}|^2 d\mathbf{x} = \|q^i\|_K^2, \tag{5}$$

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

$$\mathcal{R}_G(\mathbf{f}) = \sum_{i=1,\ldots,u} \mathbf{f}^{i\top} \mathbf{G}^i \mathbf{f}^i \tag{6}$$

$$\mathbf{f}^{i\top} \mathbf{G}^i \mathbf{f}^i = \|f(X_i) - q^i(\cdot)\|_K^2 = \mathbf{f}^{i\top} (I - \mathbf{1}\mathbf{1}^i)^\top (\mathbf{K}^i)^+ (I - \mathbf{1}\mathbf{1}^i)\mathbf{f}^i, \tag{7}$$

where $[\mathbf{K}]_{lm} = K(\mathbf{x}_l, \mathbf{x}_m)$, $\mathbf{f}^i = [f(X_1), \ldots, f(X_k)]^\top$, $\mathbf{K}^+$ is the Moore-Penrose pseudoinverse of $\mathbf{K}$, and $\mathbf{1}\mathbf{1}^i$ is an indicator matrix which is zero except for the $l(i)$-th column of ones with $l(i)$ being the index of X_i in $N_k(X_i)$.

$\mathcal{R}_G$ construction pseudocode is in Algorithm 6. In the main paper, we further augment the null space of our regularizer with the m-dimensional space of geodesic functions.

For estimating pose in a MoCap database with u =50,000, our new regularizer is 6× faster than iterated graph Laplacian, and 2.5× slower than graph Laplacian; furthermore, both pose angle and location errors decrease vs. both standard and iterated graph Laplacian (Table 1). The full paper contains further experiments on the CAESAR body shape database. These show improvements in both regression performance and computation time over the iterated graph Laplacian regularizer.

[1] M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, 2005.

[2] A. L. Yuille and N. M. Grzywacz. The motion coherence theory. In *Proc. ICCV*, pages 344–353, 1988.

[3] X. Zhou and M. Belkin. Semi-supervised learning by higher order regularization. *JMLR W&CP (Proc. AISTATS)*, pages 892–900, 2011.