关于“On the eigenvectors of p-Laplacian”目标函数的优化问题

作者：凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

图p-拉普拉斯 (Graph p-Laplacian) / p-谱聚类算法 (p-spectral clustering) 从提出到现在有一些年景了，但关于目标函数的优化问题却很少被提及，而是直接引用前人[2]的结论。这篇博客，我们追根溯源，从最初提出图p-拉普拉斯开始，来探讨目标函数的优化(最小化)问题。

1. Graph p-Laplacian

给定一个带权无向图$G=(V, E)$，其中V是边集，E是点集。

$H(V)$: The Hilbert space of real-valued functions on each vertex.

$H(E)$: The Hilbert space of real-valued functions on each edge.

The graph p-Laplacian ${{\Delta }_{p}}:H(V)\to H(V)$ 为：

$${{\Delta }_{p}}f=-\frac{1}{2}div({{\left\| \nabla f \right\|}^{p-2}}\nabla f)$$

其中$\Delta $是拉普拉斯算子，$\nabla $是梯度，$div$为散度。当$p=2$时，${{\Delta }_{p}}f={{\Delta }_{2}}f=-\frac{1}{2}div(\nabla f)$，此时，图$p$-拉普拉斯退化为标准的图拉普拉斯。

通过引入函数f的二次型，标准图拉普拉斯算子${{\Delta }_{2}}$满足：

$$\left\langle f,{{\Delta }_{2}}f \right\rangle =\frac{1}{2}\sum\limits_{i,j\in V}{{{w}_{ij}}{{({{f}_{i}}-{{f}_{j}})}^{2}}}$$

与图拉普拉斯类似，p-拉普拉斯算子[1]满足：

$$\left\langle f,{{\Delta }_{p}}f \right\rangle =\frac{1}{2}\sum\limits_{i,j\in V}{{{w}_{ij}}{{\left| {{f}_{i}}-{{f}_{j}} \right|}^{p}}}$$

对于每一个节点$i\in V$，未规范化的图p-拉普拉斯算子为：

$${{(\Delta _{p}^{(w)})}_{i}}=\sum\limits_{j\in V}{{{w}_{ij}}{{\phi }_{p}}({{f}_{i}}-{{f}_{j}})},i\in V$$

其中${{\phi }_{p}}(x)={{\left| x \right|}^{p-1}}sig(x)$。

定义一个特征向量

$${{(\Delta _{p}^{w}f)}_{i}}={{\lambda }_{p}}{{\phi }_{p}}({{f}_{i}}),i\in V$$

注：特征向量在缩放时是不变的。

定理1：f是p-拉普拉斯的特征向量，当且仅当下列函数${{F}_{p}}$在f处有临界点：

${{F}_{p}}(f)=\frac{\left\langle f,{{\Delta }_{p}}f \right\rangle }{\left\| f \right\|_{p}^{p}}=\frac{\sum\nolimits_{ij}{{{w}_{ij}}{{\left| {{f}_{i}}-{{f}_{j}} \right|}^{p}}}}{2\left\| f \right\|_{p}^{p}}$(广义Rayleigh-Ritz原理)

其中$\left\| f \right\|_{p}^{p}=\sum\nolimits_{i}{{{\left| {{f}_{i}} \right|}^{p}}},\text{ }i,j\in V$，相应地特征值${{\lambda }_{p}}$通过等式${{\lambda }_{p}}={{F}_{p}}(f)$得出。

2. 用图p-拉普拉斯进行K聚类

Luo等人[2]通过求解下面的p-拉普拉斯嵌入问题，引入了对p-拉普拉斯全特征向量的一种近似，目标函数为：

$$\underset{F}{\mathop{\min }}\,{{J}_{E}}(F)=\sum\limits_{k}{\frac{\sum\nolimits_{ij}{{{w}_{ij}}{{\left| f_{i}^{k}-f_{j}^{k} \right|}^{p}}}}{\left\| {{f}^{k}} \right\|_{p}^{p}}}$$

$$s.t.\text{ }{{F}^{T}}F=I.$$

但是，在求导过程中出现错误，原文截图为：

4 p-Laplacian embedding

Since $F_p(f) = F_p(\alpha f)$ for $\alpha \neq 0$, we can always scale f without any change. Thus, we propose the following p-Laplacian Embedding problem.

$$\min_{\mathcal{F}} J_E(\mathcal{F}) = \sum_k \frac{\sum_{ij} w_{ij} |f_i^k - f_j^k|^p}{\|f^k\|_p^p}, \tag{20}$$

$$\text{s.t.} \quad \mathcal{F}^T \mathcal{F} = I. \tag{21}$$

4.1 Optimization

The gradient of J_E w.r.t. f_i^k can be written as,

$$\frac{\partial J_E}{\partial f_i^k} = \frac{1}{\|f^k\|_p^p} \left[\sum_j w_{ij} \phi_p(f_i^k - f_j^k) - \frac{\phi_p(f_i^k)}{\|f^k\|_p^p} \right]. \tag{22}$$

If we simply use the gradient descend approach, the solution f^k might not be orthogonal. We modify the gradient as following to enforce the orthogonality,

$$\frac{\partial J_E}{\partial \mathcal{F}} \leftarrow \frac{\partial J_E}{\partial \mathcal{F}} - \mathcal{F} \left(\frac{\partial J_E}{\partial \mathcal{F}} \right)^T \mathcal{F}.$$

也就是从这里起，后面的结果已经无效。

我推导的为：

$$\frac{\partial {{J}_{E}}}{\partial f_{i}^{k}}=\frac{1}{\left\| {{f}^{k}} \right\|_{p}^{p}}\left[p\sum\nolimits_{j}{{{w}_{ij}}{{\phi }_{p}}(f_{i}^{k}-f_{j}^{k})}-p\sum\nolimits_{j}{{{w}_{ji}}{{\phi }_{p}}(f_{j}^{k}-f_{i}^{k})} \right]-\sum\nolimits_{ij}{{{w}_{ij}}{{\left| f_{i}^{k}-f_{j}^{k} \right|}^{p}}}\cdot \frac{p\cdot {{\phi }_{p}}(f_{i}^{k})}{{{\left(\sum\nolimits_{i}{{{\left| {{f}_{i}} \right|}^{p}}} \right)}^{2}}}$$

$$=\frac{p}{\left\| {{f}^{k}} \right\|_{p}^{p}}\left[2\sum\nolimits_{j}{{{w}_{ij}}{{\phi }_{p}}(f_{i}^{k}-f_{j}^{k})}-\frac{{{\phi }_{p}}(f_{i}^{k})}{\left\| {{f}^{k}} \right\|_{p}^{p}}\sum\nolimits_{ij}{{{w}_{ij}}{{\left| f_{i}^{k}-f_{j}^{k} \right|}^{p}}} \right]$$

不管怎样，在求导的过程中，总会有系数p，而[2]中没有。可以认为这篇文章在求偏导的过程中出现问题。

算法总体流程：

Input: Pairwise graph similarity W, number of embedding dimension K
Output: Embedding space $\mathcal{F}$
Compute $L = D - W$, where D is a diagonal matrix with $D_{ii} = d_i$.
Compute eigenvector decomposition of L: $L = USU^T$,
Initialize $\mathcal{F} \leftarrow U(:, 1 : K)$
while not converged do
    $G \leftarrow \frac{\partial J_E}{\partial \mathcal{F}} - \mathcal{F}\left(\frac{\partial J_E}{\partial \mathcal{F}}\right)^T \mathcal{F}$, where $\frac{\partial J_E}{\partial \mathcal{F}}$ is computed using Eq. (22)
    $F \leftarrow F - \alpha G$.
end

Algorithm 1: The p-Laplacian embedding algorithm.

注：(22)公式出错。

有趣的是，直接引用这篇文章结论的大有论文在。例如，这一篇[4]

$$\min_{\mathcal{F}} J_E(\mathcal{F}) = \sum_k \frac{\sum_{ij} w_{ij} \left| f_i^k - f_j^k \right|^p}{\|f^k\|_p^p}$$
$$\text{s.t. } \mathcal{F}^T \mathcal{F} = I.$$

Algorithm 1 Approximation of Graph p-Laplacian

Input: Training sample set X, p

Output: graph p-Laplacian L_p

Step 1: Construct adjacency matrix W, standard graph Laplacian $L = D - W$, where D is diagonal matrix with $D_{ii} = \sum_{j=1}^{n} W_{ij}$.

Step 2: Compute eigenvector decomposition of graph Laplacian $L = USU^T$.

Step 3: Initialize $\mathcal{F} = U(:, 1{:}K)$

repeat

$$G = \frac{\partial J_E}{\partial \mathcal{F}} - \mathcal{F}\left(\frac{\partial J_E}{\partial \mathcal{F}}\right)^T \mathcal{F},$$

where $\frac{\partial J_E}{\partial f_i^k} = \frac{1}{\|f^k\|_p^p}\left[\sum_j w_{ij}\phi_p\left(f_i^k - f_j^k\right) - \frac{\phi_p(f_i^k)}{\|f^k\|_p^p}\right].$

$$\mathcal{F} = \mathcal{F} - \eta G$$

until convergence

Step 4: $\lambda_k = \frac{\sum_{ij} w_{ij}|f_i^k - f_j^k|^p}{\|f^k\|_p^p}$

return $L_p = \mathcal{F}\Lambda\mathcal{F}^T$

更有趣的是，时隔整整十年，[3]明确指出[2]中的推导错误：

For a vector $\mathbf{u}^k \in \mathbb{R}^n$ the entries of the Euclidean gradient approximation used by (Luo et al., 2010) for the k-way p-norm functional F, defined later in (27a), read:[1]

$$\frac{\partial F}{\partial u_i^k} = \frac{1}{\|\mathbf{u}^k\|_p^p} \left[\sum_j w_{ij} \phi_p \left(u_i^k - u_j^k \right) - \frac{\phi_p \left(u_i^k \right)}{\|\mathbf{u}^k\|_p^p} \right] \tag{26}$$

with k being the cluster index number and with the function $\phi(\cdot)$ defined as in (21). However, this approximated gradient suffers from inaccuracies because the second term in (26) misses a factor of $w_{ij}|u_i^k - u_j^k|^p$. The actual (corrected) gradient is shown in (32) in Subsection 3.2. This problem is illustrated in Figure 3a, where the ratio of directional derivative F' obtained using a first order Taylor expansion[2] is compared to that of the computed gradient $\nabla(F)$, using (26), for the UMIST dataset (Graham and Allinson, 1998) with $p = 1.8$. The ratio of $(F(\mathbf{u}+\eta) - F(\mathbf{u}))/\langle \eta, \nabla F(\mathbf{u}) \rangle$ should ideally approach one as the step size η in the Taylor expansion decreases. However, with the gradient defined in (26) this is not the case (see Figure 3a).

Due to this gradient inaccuracy, fundamental properties of the spectrum of Δ_p, outlined in Theorem 2, are no longer valid for the approximation presented in (Luo et al., 2010). For example, the degeneracy of the eigenvalues, corresponding to the constant eigenvectors $\mathbf{v} = c \cdot \mathbf{e}$, no longer indicates the number of connected components in the graph. These inaccuracies lead to noncompetitive clustering results in some widely used machine learning datasets[3]. In contrast, our Grassmann approach, analyzed in Subsec-

[3]给出了他自己提的目标函数与求导公式：

$$\underset{\mathbf{U}\in\mathcal{G}r(k,n)}{\text{minimize}}\ F_p(\mathbf{U}) = \sum_{l}^{k}\sum_{ij}^{N}\frac{w_{ij}|u_i^l - u_j^l|^p}{2\|\mathbf{u}^l\|_p^p},\quad p\in(1,2]. \tag{31}$$

ROPTLIB requires the Euclidean gradient as input when performing optimization routines, with the library converting it internally to the Riemannian counterpart. The entries of the Euclidean gradient ($\mathbf{g}^k$) of F_p (31) with respect to u_i^k read

$$g_i^k = \frac{\partial F_p}{\partial u_i^k} = \frac{p}{\|\mathbf{u}^k\|_p^p}\left[\sum_j w_{ij}\phi_p\left(u_i^k - u_j^k\right) - \phi_p\left(u_i^k\right)\sum_{ij}\frac{w_{ij}|u_i^k - u_j^k|^p}{2\|\mathbf{u}^k\|_p^p}\right]. \tag{32}$$

可以看出，[3]的推导结论与我的推导是一致的，只是目标函数分母部分有无系数2。

3. 参考文献

[1] Bühler, Thomas & Hein, Matthias. (2009). Spectral clustering based on the graph Laplacian. Proceedings of the 26th International Conference On Machine Learning, ICML 2009. 382. 11-88. 10.1145/1553374.1553385. Code: p-Spectral Clustering

[2] Luo, Dijun & Huang, Heng & Ding, Chris & Nie, Feiping. (2010). On the eigenvectors of p-Laplacian. Machine Learning. 81. 37-51. 10.1007/s10994-010-5201-z.

[3] Pasadakis, Dimosthenis & Alappat, Christie & Schenk, Olaf & Wellein, Gerhard. (2020). K-way p-spectral clustering on Grassmann manifolds.

[4] W. Liu, X. Ma, Y. Zhou, D. Tao and J. Cheng, "p-Laplacian Regularization for Scene Recognition," in IEEE Transactions on Cybernetics, vol. 49, no. 8, pp. 2927-2940, Aug. 2019, doi: 10.1109/TCYB.2018.2833843.

[5] 拉普拉斯矩阵与拉普拉斯算子的关系 - 知乎

最后的思考：做研究切忌浮躁，要追根溯源，明白公式的来龙去脉，自己动手，丰衣足食。