图谱论导引(第十三期)

Lapalce矩阵

> 本节将系统介绍Laplace矩阵相关内容, 并在文末阐述其名称来源, 即揭示∇^2算子与Laplace矩阵之联系.

定义

Laplace矩阵之定义分两派, 其唯一区别即全矩阵正负之别. 本节将L定义为$D-A$.

另一有以元矩阵之和定义者: 图中$v_1\sim v_2$对应之元矩阵为

$$L_{v_1v_2}=\begin{pmatrix}1&-1&\mathbf{0}^T\\-1&1&\mathbf{0}^T\\\mathbf{0}&\mathbf{0}&O\end{pmatrix}$$

Lapalce矩阵即$\dfrac{1}{2}\sum\limits_{v_i\sim v_j}L_{v_iv_j}$. 注意到

$$\begin{pmatrix}v_1&v_2\end{pmatrix}\begin{pmatrix}1&-1\\-1&1\end{pmatrix}\begin{pmatrix}v_1\\v_2\end{pmatrix}=(v_1-v_2)^2,$$

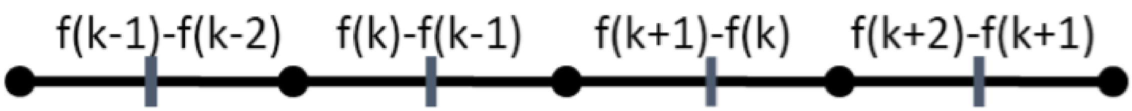
从而$v^TLv=\dfrac{1}{2}\sum\limits_{i\sim j}(v_i-v_j)^2$: 该式亦证明了$L$正定.

Laplace矩阵与Laplace算符

两者同名并非偶然, 下先考虑一维情形的Laplace算子$\dfrac{\mathrm{d}^2}{\mathrm{d}x^2}$.

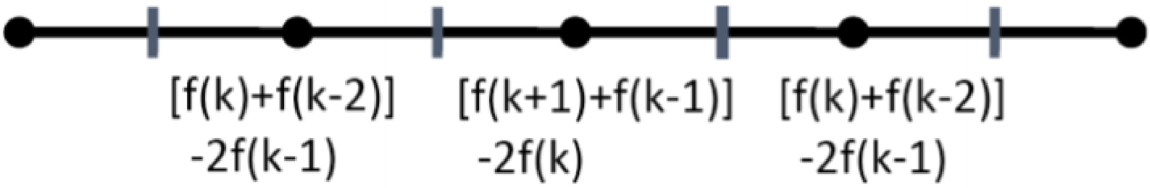
诸位读者最初接触Laplace算子应是在"打点计时器测量加速度"一课上, 物理老师尝授"逐差法"以通过位移计算加速度, 其本质即通过离散化计算$a=\dfrac{\mathrm{d}^2}{\mathrm{d}t}x$. 下假定函数$f:\mathbb{R}\to\mathbb{R}$, $\mathbb{R}$上整点$\{k\}$列分别对应相应的函数值$\{f(k)\}$. 如下图所示

各$k+1/2$点处的一阶导数近似表示为$f(k+1)-f(k)$,



从而各k点处对应的二阶导数近似表示为

$$[f(k+1)-f(k)]-[f(k)-f(k-1)]=f(k+1)+f(k-1)-2f(k).$$



中心差分(central difference)是谓也. 从而$\nabla^2 f \approx L_{P_\infty} f$. 对$d$维空间而言, 记整点$x$在$i$维方向的邻点为$x_i^{\pm}$, 则中心差分为

$$\nabla^2 f(x) \approx \sum_{i=1}^{d} [f(x_i^+) + f(x_i^-)] - 2d f(x).$$

再论聚类

λ_2上界

上一期文已讨论聚类指标: 特征值. 本文将对简单图的2-聚类做进一步的探讨, 并着重介绍Cheeger不等式.

对简单无权图G的2-聚类的本质是求得集合$S \subset V(G)$使得

$$\frac{\sum_{i \in S, j \in (V \subset S)} 1}{|S| \cdot |V \setminus S|}$$

最小. 今不妨设x为$|V|$维指示向量: $x_i = 1$若且仅若$i \in S$, 反之$x_i = -1$. 从而上式等价求解$\dfrac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i<j} (x_i - x_j)^2}$取最小时对应的$S$. 由于上式仅与$x_i - x_j$相关, 故可对$x$进行适当平移使得$\sum_i x_i = 0$, 从而优化问题转化为

$$\begin{aligned} & \min_{x \in \mathbb{R}^n \sum_i x_i = 0} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i<j} (x_i - x_j)^2} \\ = & \min_{x \in \mathbb{R}^n x \perp \mathbf{1}} \frac{x^T L x}{n \sum_i x_i^2} \\ \geq & \frac{\lambda_2}{n} \end{aligned} .$$

定义$\phi(S) := \dfrac{\mathrm{cut}(S, \overline{S})}{\min(|S|, |\overline{S}|)}$为切比(cut ratio), $\phi(G) := \min_{S \subset G} \phi(S)$, 从而$\phi(G) \geq \dfrac{\lambda_2}{2}$.

Cheeger不等式

上文已证明$\lambda_2 \leq 2\phi(G)$; Cheeger不等式给出等式的两端, 即

$$\frac{\phi(G)^2}{2 d_{\max}} \leq \lambda_2 \leq 2\phi(G).$$

其中, 等式两端都是最优的. 以下证明过程中, 我们实将推导出一些略强的结果.

第一步. 我们将于$V(G)$中选取某一参考点, 并将剩余两点尽量均分. 不失性一般性地假设n为奇数, 记$m = \dfrac{n+1}{2}$, 平移得$y_i := x_i - x_m$. 从而

$$
\begin{aligned}
\frac{y^TLy}{x^Tx} &= \frac{\sum_{(i,j)\in E}(y_i-y_j)^2}{y^Ty} \\
&= \frac{\sum_{(i,j)\in E}(y_i+x_m-y_j-m_m)^2}{(x+x_m\mathbf{1})^T(x+x_m\mathbf{1})} \\
&= \frac{\sum_{(i,j)\in E}(x_i-x_j)^2}{x^Tx+nx_m^2} \\
&\geq \frac{x^TLx}{x^Tx}
\end{aligned}.
$$

第二步. 将一切编号符合$i<m<j$式子的边(i,j)分裂为(i,m), (m,j). 记新边集为E', $E'_- := \{(i,j): i,j\leq m\}$, $E'_+ := \{(i,j): i,j\leq m\}$. 注意到$y_m=0$, 故

$$
\begin{aligned}
\frac{\sum_{(i,j)\in E'}(y_i-y_j)^2}{\sum_i y_i^2} &= \frac{\sum_{(i,j)\in E'_-}(y_i-y_j)^2+\sum_{(i,j)\in E'_+}(y_i-y_j)^2}{\sum_{i\leq m}y_i^2+\sum_{i\geq m}y_i^2} \\
&\geq \min\left(\frac{\sum_{(i,j)\in E'_-}(y_i-y_j)^2}{\sum_{i\leq m}y_i^2}, \frac{\sum_{(i,j)\in E'_+}(y_i-y_j)^2}{\sum_{i\geq m}y_i^2}\right)
\end{aligned}
$$

第三步. 回顾$\phi=\phi(G)=\min\limits_{S\subset V}\dfrac{e(G)}{\min(|S|,|\bar{S}|)}$. 不妨设$C_i$为通过$x_i$处的边的总数, 则$i\leq n/2$时$C_i\geq i\phi$; 反之$C_i\geq(n-i)\phi$. 对任意序列$z_1\leq\cdots\leq z_m=0$, 注意到

$$
\begin{aligned}
\sum_{(i,j)\in E'_-}|z_i-z_j| &= \sum_{k=1}^{m-1}C_k(z_{k+1}-z_k) \\
&\geq \phi\sum_{k=1}^{m-1}k(z_{k+1}-z_k) \\
&= \phi(-z_1-z_2-\cdots-z_{m-1}+(m-1)z_m) \\
&= \phi\sum_{i=1}^{m}|z_i|
\end{aligned}.
$$

不妨令z_i为y_i^2, 则

$$
\begin{aligned}
\phi &\leq \frac{\sum_{(i,j)\in E'_-}|y_i^2-y_j^2|}{\sum_{i=1}^m|y_i^2|} \\
&\leq \frac{\sqrt{\sum_{(i,j)\in E'_-}(y_i-y_j)^2\sum_{(i,j)\in E'_-}(y_i+y_j)^2}}{\sum_{i=1}^m|y_i^2|} \\
&\leq \frac{\sqrt{\sum_{(i,j)\in E'_-}(y_i-y_j)^2\cdot 2\sum_{(i,j)\in E'_-}d_{\max}|y_i|^2}}{\sum_{i=1}^m|y_i^2|}
\end{aligned}
$$

从而

$$
\frac{\sum_{(i,j)\in E'_-}(y_i-y_j)^2}{\sum_{i=1}^m y_i^2}\geq\frac{\phi^2}{2d_{\max}}
$$

对E'_+同理, 故Cheeger不等式成立.

Cheeger不等式应用

定义子图的导率$\Phi(S) := \dfrac{e(S)}{\min(\sum_{v\in S}\deg v, \sum_{v\in \overline{S}}\deg v)}$, 图的导率

$$\Phi(G) := \min_{S\subset V}\Phi(S)$$

由Cheeger不等式知$1-\lambda_2$之阶数为$\Phi(G)$至$\Phi(G)^2$之间.