Harmonic minimization (continued) and spectral representation of data

MATH 697 AM:ST

November 14th, 2017

Warm up

Laplacian on hypersurfaces and other submanifolds

A classical topic: the Laplacian in spherical coordiantes
If $f : \mathbb{R}^d \to \mathbb{R}$, and $x = (r, \theta)$ where $\theta \in \mathbb{S}^{d-1}$, then

$$\Delta f = \partial_r^2 f + \frac{d-1}{r}\partial_r f + \frac{1}{r^2}\Delta_\theta f$$

The operator Δ_θ (which acts only along the θ "directions") is an operator that acts on any function defined on the sphere $\mathbb{S}^{d-1}$. This is a special case of what is known as the Laplace-Beltrami operator of a manifold.

Warm up

Laplacian on hypersurfaces and other submanifolds

In coordinates $(x_1, \ldots, x_k)$, we can write Δ_Σ as

$$\Delta_\Sigma u = \frac{1}{\sqrt{|g|}} \sum_{i,j=1}^{k} \partial_i \left(\sqrt{|g|} g^{ij} \partial_j u \right)$$

(here k denotes the intrinsic dimension of Σ)

Warm up

Laplacian on hypersurfaces

Consider the heat equation

$$\partial_t u = \Delta_\Sigma u$$

The solution is given via the heat kernel by

$$u(x,t) = \int_\Sigma u(y) H(x,y,t) d\sigma(y)$$

and

$$\Delta_\Sigma u(x) = \lim_{t \to 0^+} \left\{ \frac{1}{t} \int_\Sigma (u(y) - u(x)) H(x,y,t) d\sigma(y) \right\}$$

Warm up

Laplacian on hypersurfaces

Using the Gaussian asymptotics for the heat kernel, we have

$$\begin{aligned}\Delta_\Sigma u(x) &\approx \frac{1}{t}\int_\Sigma (u(y)-u(x))H(x,y,t)d\sigma(y)\\ &\approx \frac{1}{t(4\pi t)^{\frac{k}{2}}}\sum_j (u(x_j)-u(x))e^{-\frac{|x_i-x_j|^2}{4t}}\end{aligned}$$

This is one motivation for the use of exponential weights: if a finite graph is expected to be a **discretization** of some limiting manifold, then the graph Laplacian for Gaussian weights ought to be close to the Laplace-Beltrami operator.

(It should be noted, however, that this property is not exclusive to Gaussian weights!)

Measuring Smoothness

RECAP

Let $f : G \mapsto \mathbb{R}$, and define the norms

$$\|f\|_{L^2} := \left(\sum_{x \in G} |f(x)|^2\right)^{\frac{1}{2}}$$

$$\|f\|_{H_w} := \left(\sum_{x \in G} |f(x)|^2 + \sum_{x,y \in G} |f(x) - f(y)|^2 \omega_{xy}\right)^{\frac{1}{2}}$$

as well as

$$\|f\|_{\dot{H}_w} := \left(\sum_{x,y \in G} |f(x) - f(y)|^2 \omega_{xy}\right)^{\frac{1}{2}}$$

Measuring Smoothness

RECAP

While $\|f\|_{L^2}$ measures simply the average "size" of f, $\|f\|_{\dot{H}_w}$ measures the average size of its oscillations.

Observe that if $\|f\|_{\dot{H}_w} = 0$, then f must be constant in each connected component of the graph.

Thus, the larger $\|f\|_{\dot{H}_w}$ is with respect to $\|f\|_{L^2}$, the more the function f is oscillating with respect to its size.

Eigenfunctions of Δ
RECAP

There is a family of functions

$$\{\phi_n\}_{n=0}^{N}, \quad \phi_n : G \mapsto \mathbb{R}$$

as well as numbers $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$, such that

$$-\Delta\phi_n(x) = \lambda_n \phi_n(x) \quad \forall\, x \in G.$$

Moreover, the ϕ_n are orthonormal,

$$\langle \phi_n, \phi_m \rangle = \delta_{nm}.$$

Eigenfunctions of Δ

Decrasing Smoothness of ϕ_n

RECAP

Observe that

$$\begin{aligned} \|\phi_n\|^1_{\dot{H}_w} &= \sum_{x,y \in G} w_{xy}(\phi_n(y) - \phi_n(x))^2 \\ &= -2\langle \Delta\phi_n, \phi_n\rangle \end{aligned}$$

Therefore,

$$\|\phi_n\|^1_{\dot{H}_w} = 2\lambda_n \|\phi_n\|^2_{L^2}$$

As the λ_n is increasing with n, we see that the ϕ_n become increasingly more oscillatory as n gets larger and larger.

The Heat Equation
RECAP

We consider the heat equation in all of G, with some initial datum $u_0 : G \mapsto \mathbb{R}$,

$$\begin{aligned} \dot{u} &= \Delta u \quad \text{in } G \times (0, \infty), \\ u &= u_0 \quad \text{at } t = 0. \end{aligned}$$

The Heat Equation
RECAP

By the orthogonality of the ϕ_n, it follows that u_0 can be expressed as

$$u_0 = \sum_{n=0}^{N} \alpha_n \phi_n \quad \text{where } \alpha_n = \langle \phi_n, u_0 \rangle.$$

Therefore, for $t > 0$, the solution to the heat equation is

$$u(t) = \sum_{n=0}^{N} \alpha_n e^{-t\lambda_n} \phi_n$$

The Heat Equation
RECAP

This leads to an analogue of the heat kernel $H(t, x, y)$, namely, a function such that

$$u(x, t) = \sum_{y \in G} H(t, x, y) u_0(y)$$

This $H(t, x, y)$ is given via the eigenfunction decomposition

$$H(t, x, y) = \sum_{n=0}^{N} e^{-t\lambda_n} \phi_n(x) \phi_n(y)$$

The Dirichlet Problem Revisited

RECAP

Fix some subset $D \subset G$, and suppose we are given a function

$$g : G \setminus D \mapsto \mathbb{R}$$

Then, we aim to find $f : G \mapsto \mathbb{R}$, such that

$$f = g \text{ in } G \setminus D,$$

and with f minimizing the discrete Dirichlet energy

$$J(f) := \frac{1}{2} \sum_{x,y \in G} w_{xy}(f(x) - f(y))^2$$

The Dirichlet Problem Revisited

RECAP

There is a unique minimizer f, and it is characterized by

$$\begin{cases} Lf(x) & = 0 \qquad\quad \text{if } x \in D, \\ f(x) & = g(x) \text{ if } x \in G \setminus D. \end{cases}$$

Compare with the Dirichlet problem for the Laplacian

$$\min \int_D |\nabla f(x)|^2 \, dx, \;\; f = g \text{ on } \partial D.$$

and the one for the fractional Laplacian ($\alpha \in (0, 2)$),

$$\min \int_{\mathbb{R}^{2d}} (f(x) - f(y)|^2 |x - y|^{-d-2\alpha} \, dx, \;\; f = g \text{ in } \mathbb{R}^d \setminus D.$$

The Dirichlet Problem Revisited

RECAP

Thanks to the variational characterization of harmonic functions and the comparison theorem, we have the following

Theorem

Let $D \subset G$ be a connected subset of G. Then, given $g : G \setminus D \mapsto \mathbb{R}$ there exists one, and exactly one, function f which solves

$$\begin{cases} Lf(x) &= 0 \qquad\quad \text{if } x \in D, \\ f(x) &= g(x) \text{ if } x \in G \setminus D. \end{cases}$$

From data sets to graphs

Local Similarities

If the data set amounts to points $\{x_1, \ldots, x_N\}$ in some metric space (X, ρ), we set $G_N = \{x_1, \ldots, x_N\}$ with weights

$$W_\sigma(x, y) = h\left(\frac{1}{\sigma^{\frac{1}{2}}}\rho(x, y)\right)$$

Most of the time, the metric space is modeled in $\mathbb{R}^p$ and the function h is most usually a Gaussian, this is known as a graph with Gaussian weights.

From data sets to graphs

Supervised and semi-supervised

In *supervised learning*, we are also given scalars $y_1, \ldots, y_N$, this is the same as saying we are given a function

$$f_N : G_N \mapsto \mathbb{R}$$

Then, we aim at extrapolating from this data a function $\hat{f}$ defined in all of $\mathbb{R}^p$. A variation on this is *semi-supervised learning*.

From data sets to graphs

Supervised and semi-supervised

In semi-supervised learning, we are given again input data $G = \{x_1, \dots, x_N\}$, as well as output data y_i but only **partially**, that is y_i is given only for i in some subset $\mathcal{I} \subset \{1, \dots, N\}$. The goal is then to fill in the values of f at the remaining points.

In other words, we want to determine a function

$$f : D \mapsto \mathbb{R} \text{ for some } D \subset \mathbb{R},$$

for some subset $D \subset G$, using the geometric information about the input sets, and given values for the function in the rest of the graph $g : G \setminus D \mapsto \mathbb{R}$.

Harmonic functions and (semi) supervised learning

Zhu, Gharamani, and Lafferty (2003) idea: Apply the Dirichlet problem to the semi-supervised learning problem.

The setup is as follows: denote by G the set of input points

$$G = \{x_1, \ldots, x_N\} \subset \mathbb{R}^p$$

Denote by D the subset of G of points that are "unlabeled" i.e. where we are not given the value of the unknown function.

Harmonic functions and (semi) supervised learning

For $\sigma > 0$, they use the weights given by

$$w_{xy} := e^{-\frac{|x-y|^2}{\sigma^2}}$$

Thus, the graph G is connected, with points x_i and x_j which are close in $\mathbb{R}^p$ having a weight close to 1, and which is nearly zero as soon as $|x_i - x_j|$ is large respect to σ.

The parameter σ is a tuning parameter and it is usually chosen using criteria such as the generalization error according to what other statistical assumptions we are dealing with.

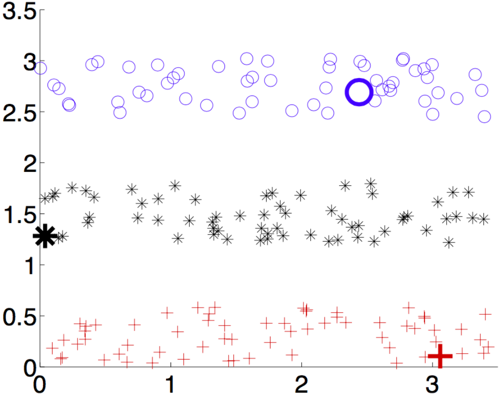
Harmonic functions and (semi) supervised learning

Then, let $\hat{f} : G \mapsto \mathbb{R}$ be defined as the unique solution to

$$\begin{cases} \Delta \hat{f}(x) & = 0 \text{ in } D, \\ \hat{f}(x) & = f_\ell(x) \text{ in } G \setminus D. \end{cases}$$

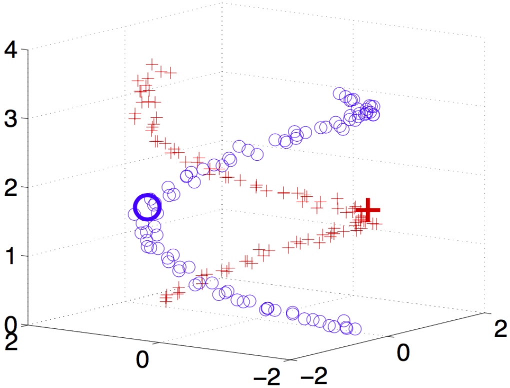
Harmonic functions and (semi) supervised learning

Example



Harmonic functions and (semi) supervised learning

Example



The Dirichlet Problem Revisited

The solution to

$$\begin{cases} Lf(x) & = 0 \qquad \text{if } x \in D, \\ \ \ f(x) & = g(x) \text{ if } x \in G \setminus D. \end{cases}$$

Is given by

$$f(x) = \sum_{y \in G \setminus D} P_D(x,y) g(y)$$

for a certain kernel $P_D(x,y)$ computable from the w_{xy}.

The Dirichlet Problem Revisited

Let $G = \{x_1, \ldots, x_N\}$ the indices arranged so that

$$D = \{x_1, \ldots, x_m\}$$
$$G \setminus D = \{x_{m+1}, \ldots, x_N\}$$

Then, we the graph Laplacian corresponds to the matrix

$$L_{ij} = w_{ij} - d_i \delta_{ij}$$

which can be decomposed in various blocks as

$$L = \begin{pmatrix} L_{uu} & L_{u\ell} \\ L_{\ell u} & L_{\ell\ell} \end{pmatrix}$$

The Dirichlet Problem Revisited

The Dirichlet problem then can be written as

$$\begin{pmatrix} L_{uu} & L_{u\ell} \\ L_{\ell u} & L_{\ell\ell} \end{pmatrix} \begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} 0 \\ \ldots \end{pmatrix}$$

(the $(\ldots)$ are irrelevant terms) and we see that it comes down to finding the remaining unknowns in the linear above equation

$$L_{uu} f + L_{u\ell} g = 0$$

Therefore, we obtain a formula for $f_u = f\mid_D$, and it yields

$$f = (-L_{uu})^{-1} L_{u\ell} g$$

The Dirichlet Problem Revisited

If one wants to use the random walk Laplacian, then

$$f = (-L_{uu})^{-1} L_{u\ell} g$$

can be written as

$$f = (I - P_{uu})^{-1} P_{u\ell} g$$

The Dirichlet Problem Revisited

Lemma

The matrix $(I - P_{uu})^{-1} P_{u\ell}$ *is well defined.*

The Dirichlet Problem Revisited

Lemma

The matrix $(I - P_{uu})^{-1} P_{u\ell}$ *is well defined.*

Note: This matrix however does not necessarily have full rank!

The Dirichlet Problem Revisited

Let $y \in G \setminus D$, and let $P_D(\cdot, y) : G \to \mathbb{R}$ be the solution to

$$\begin{cases} LP_D(\cdot, y) & = 0 \quad \text{in } D, \\ P_D(\cdot, y) & = \delta_y \quad \text{in } G \setminus D. \end{cases}$$

where

$$\delta_y(x) = \begin{cases} 1 \text{ if } y = x, \\ 0 \text{ otherwise.} \end{cases}$$

Thus, we can think of having constructed a function $P_D(x, y)$

$$P_D : D \times (G \setminus D) \to \mathbb{R}$$

The Dirichlet Problem Revisited

Then, we have that

$$f(x) := \sum_{y \in G \setminus D} P_D(x,y) g(y)$$

Solves the Dirichlet Problem

$$\begin{cases} Lf & = 0 \quad \text{in } D, \\ f & = g \quad \text{in } G \setminus D. \end{cases}$$

The Dirichlet Problem Revisited

If we think of the matrix $(I - P_{uu})^{-1} P_{u\ell}$ as a function

$$D \times (G \setminus D) \to \mathbb{R}$$

then this is simply the function $P_D(x, y)$ we just constructed.

Inverting the Laplacian in all of G

Often times we are given a function $f : G \mapsto \mathbb{R}$ and seek u solving (L the "combinatorial" Laplacian)

$$Lu = f \text{ in } G$$

Lemma

Assume G is connected. For the existence of a solution to the above equation, it is sufficient and necessary that

$$\sum_{y \in G} f(y) = 0.$$

Spectral representation of data and graph cuts

MATH 697 AM:ST

November 16th, 2017

Warmup

Let G be a connected graph with weights w_{xy}.

Suppose we want to minimize

$$\frac{\langle \Delta u, u \rangle}{\langle u, u \rangle}$$

over all $u : G \mapsto \mathbb{R}$ such that

$$\langle u, \mathbf{1} \rangle = 0$$

Then, the solution u is given by the eigenfunction of Δ with second-smallest eigenvalue.

Warmup

Variations on the Laplacian and Densities

A different inner product for functions in G is given by

$$\langle f, g \rangle_d := \sum_{x \in G} f(x)g(x)d(x)$$

where $d(x) = \sum_y w_{xy}$ is the degree of x.

This different inner product alters the symmetry properties of the Laplacian!.

Warmup

Variations on the Laplacian and Densities

For instance, what happens if one tries to minimize

$$\frac{\langle Lf, f\rangle}{\langle f, f\rangle_d}$$

(note the different product above and below)

Warmup

Variations on the Laplacian and Densities

Consider the linear map D on functions on G given by

$$(D^{\frac{1}{2}}f)(x) := d(x)^{\frac{1}{2}}f(x)$$

This is clearly and invertible map (G is connected) and

$$(D^{-\frac{1}{2}}f)(x) := d(x)^{-\frac{1}{2}}f(x).$$

Observe that, for any $u, v : G \mapsto \mathbb{R}$, we have

$$\langle u, v\rangle_d = \langle Du, v\rangle = \langle D^{\frac{1}{2}}u, D^{\frac{1}{2}}v\rangle$$

Warmup

Variations on the Laplacian and Densities

An elementary computation also shows that

$$\langle Lf, f \rangle = \langle \hat{L} D^{\frac{1}{2}} f, D^{\frac{1}{2}} f \rangle$$

where $\hat{L}$ denotes the *symmetric normalized Laplacian*

$$\hat{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$$

Warmup

Variations on the Laplacian and Densities

Therefore, we see that if $\hat{f} = D^{\frac{1}{2}} f$, then

$$\frac{\langle Lf, f \rangle}{\langle f, f \rangle_d} = \frac{\langle \hat{L}\hat{f}, \hat{f} \rangle}{\langle \hat{f}, \hat{f} \rangle}$$

Warmup

Variations on the Laplacian and Densities
Eigenvalue Characterization

Let L be the (combinatorial) Laplacian for (G, w_{ij}) and

$$-Lu_k = \lambda_k u_k \;\; 0 = \lambda_0 < \lambda_1 \leq \lambda_2 \dots$$

Then, for $k = 1, \dots,$ we have

$$\lambda_k = \operatorname{argmin} \left\{ \sum_{ij} w_{ij}(u_i - u_j)^2 \mid u \in [u_0, u_1, \dots, u_{k-1}]^\perp, \;\|u\|_2 = 1 \right\}$$

Warmup

Variations on the Laplacian and Densities
Eigenvalue Characterization

This is equivalent to seeking a vector map

$$\bar{u} : G \to \mathbb{R}^{N-1}$$

(we write $\bar{u}_i = ((u_1)_i, \ldots, (u_{N-1})_i)$ for each $i = 1, \ldots, N$)
and the map $\bar{u}$ is meant to minimize the energy functional

$$\sum_{ij} w_{ij} \|\bar{u}_i - \bar{u}_j\|^2 = \sum_{ij} \sum_{k=1}^{N-1} w_{ij} ((\bar{u}_k)_i - (\bar{u}_k)_j)^2$$

while satisfying the "integral" constraints

$$\sum_i (\bar{u}_k)_i (\bar{u}_{k'})_i = \delta_{kk'} \ \ \forall\, k, k'$$

Dimensionality Reduction Via Eigenfunctions

Let us discuss another important role for harmonic functions and graph Laplacians: **data representation and dimensionality reduction.**

Given $G = \{x_1, \ldots, x_N\} \subset \mathbb{R}^p$, set

$$w_{ij} = w_{x_i x_j} = e^{-\frac{|x_i - x_j|^2}{t}}$$

Compute the first m eigenvectors of the Laplacian

$$-Lu_m = \lambda_m u_m \text{ in } D$$

Dimensionality Reduction Via Eigenfunctions

Then, we define a map

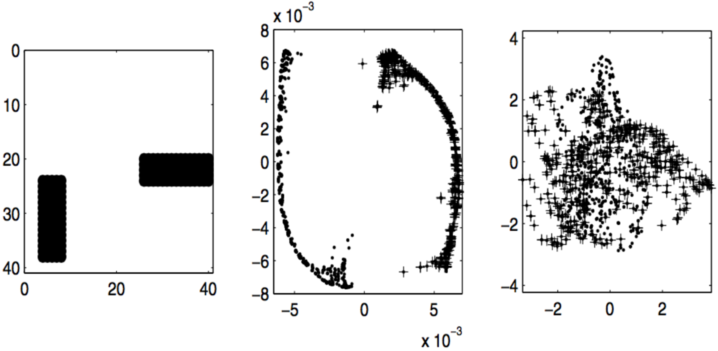
$$\bar{u} : G \mapsto \mathbb{R}^m$$

$$\bar{u}(x) = (u_1(x), \ldots, u_m(x))$$

The idea is that in general m is much smaller than p, and while $\bar{u}$ may not be necessarily injective for small m, the map $\bar{u}$ captures the relevant features of the data set G.

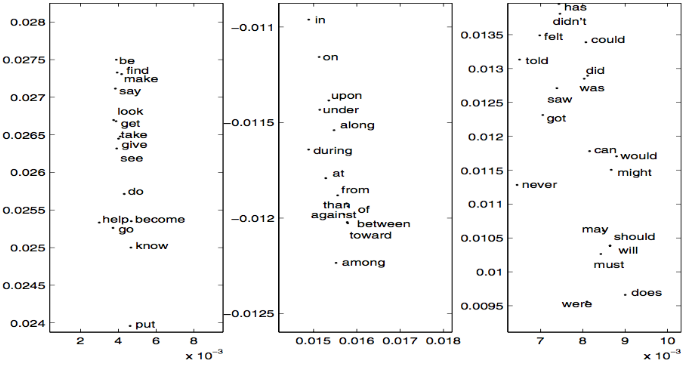
Dimensionality Reduction Via Eigenfunctions

Example



Dimensionality Reduction Via Eigenfunctions

Example



Clustering

A simple and informative definition of clustering in ML is given in Hastie, Tibsharani, and Friedman, Chapter 14:

Cluster analysis, also called data segmentation, has a variety of goals. All relate to grouping or segmenting a collection of objects into subsets or "clusters," such that those within each cluster are more closely related to one another than objects assigned to different clusters.

Clustering

In clustering, one is given **disorganized** data

Clustering

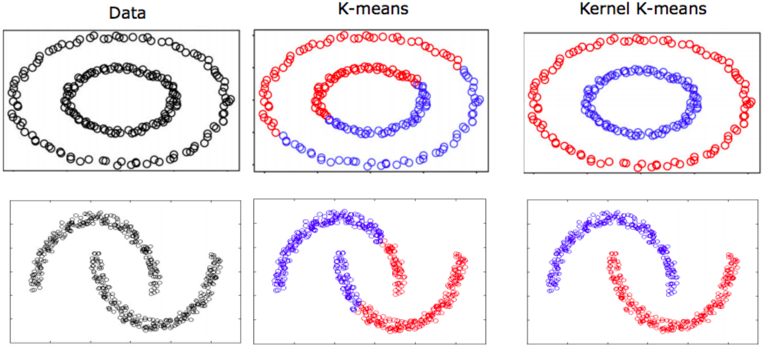
Examples

. . . and then one seeks to organize it in terms of groups of similar elements

Clustering

Examples

Ultimately, this problem comes down to dividing discrete sets in a proper metric space



Key point: No training labels were given a priori!

Clustering

Challenges

1. Typically, these are **unsupervised** problems: no labels are given in advance.
2. In fact, unclear a priori what is the "right" number of clusters and what their nature is (e.g. I give you a data set of all the music available on some streaming service, can one automatically generate playlists with similar songs?)
3. Clustering in itself is a form of dimensional reduction for a data set by way of "coarsening".
4. Often used in combination with other learning algorithms.

Clustering

Similarity measure / Proximity matrices

One then has a set $\{x_1, \ldots, x_N\}$, where one is given quantitative data measuring how dissimilar two elements are

$$w(x_i, x_j) = \begin{cases} \text{the greater this number,} \\ \text{the greater the similarity} \\ \text{between } x_i \text{ and } x_j \end{cases}$$

Clearly, this amounts to a weighted graph!!

Clustering

Most popular algorithms

1. K-means clustering
2. Gaussian mixture models
3. Spectral clustering / Kernel K-means
4. Nearest neighbor

From Harmonic Functions to Spectral Clustering

For a weighted (G, w_{ij}), we want to partition G in two classes

$$G = A \cup B, \quad A \cap B = \emptyset$$

We want to make this partition as optimal as possibe: one criterium, the "connectivity" between the two components of the partition must be minimal –but in what sense?

Given two sets $A, B \subset G$ (disjoint or not) we define their **cut**

$$\mathrm{cut}(A, B) = \sum_{x \in A} \sum_{y \in B} w_{xy}$$

From Harmonic Functions to Spectral Clustering

Meanwhile, the volume of a set A is defined by

$$\begin{aligned}\mathrm{Vol}(A) &:= \sum_{x \in A} d(x) \\ &= \sum_{x \in A} \sum_{y \in G} w_{xy} = \mathrm{cut}(A, G)\end{aligned}$$

Another term for the quantity $\mathrm{cut}(A, G)$ is the *association* of A, since it increases as elements of A share more and stronger connections with other elements of G.

From Harmonic Functions to Spectral Clustering

In 1993, Wu and Leahy proposed a method to detect relevant clusters in data:

$$\text{minimize } \mathrm{cut}(A, G \setminus A)$$

That is: cut the graph in two components in a way that minimizes the total "similarity" between elements in different classes.

From Harmonic Functions to Spectral Clustering

In 1993, Wu and Leahy proposed a method to detect relevant clusters in data:

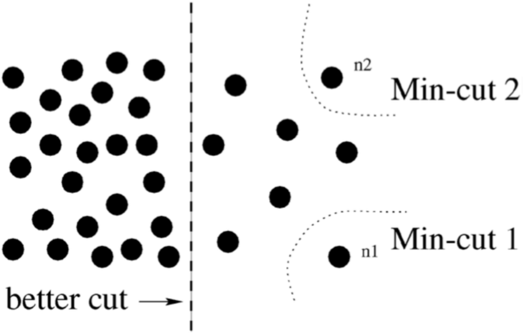
$$\text{minimize } \operatorname{cut}(A, G \setminus A)$$

That is: cut the graph in two components in a way that minimizes the total "similarity" between elements in different classes.

This intuitive idea, however, has a drawback...

From Harmonic Functions to Spectral Clustering

…this minimization problem tends to pick up **small**, disconnected subsets



From Harmonic Functions to Spectral Clustering

A fix for this, is related to a key idea in graph theory.

A central object of study in geometry and graph theory is the **Cheeger constant** of a graph

$$\min\left\{\frac{\mathrm{cut}(A, G\setminus A)}{\min(\mathrm{Vol}(A), \mathrm{Vol}(G\setminus A))} \;\mid A\subset G\right\}$$

also sometimes expressed as

$$\min\left\{\frac{\mathrm{cut}(A, G\setminus A)}{\mathrm{Vol}(A)} \;\mid A\subset G \text{ with } 0 < \mathrm{Vol}(A) \leq \frac{1}{2}\mathrm{Vol}(G)\right\}$$

From Harmonic Functions to Spectral Clustering

Normalized Cuts

Shi and Malik (2000) proposed the following minimization problem, modifying Leahy and Wu's proposal:

$$\mathrm{Ncut}(A,B) = \mathrm{cut}(A,B)\left(\frac{1}{\mathrm{Vol}(A)} + \frac{1}{\mathrm{Vol}(B)}\right)$$

This gives a criterium to divide a data set G into **two** distinguished sets, which avoids the tendency to pick small, disconnected subsets.

The hope is that the minimizing partition (A_0, B_0) captures two important features of the data set. This is a special instance of a **clustering problem**.

$$\mathrm{Ncut}(A,B) = \mathrm{cut}(A,B)\left(\frac{1}{\mathrm{Vol}(A)} + \frac{1}{\mathrm{Vol}(B)}\right)$$

This minimization problem, it must be noted, is **NP-hard**.

If one is willing to give up a small amount of accuracy, it can be approximated via a problem which is solvable in polynomial time (or, as fast as one compute find graph eigenfunctions).

From Harmonic Functions to Spectral Clustering

Ncut and indicator functions

Take $A \subset G$, and consider the function

$$f(x) = \chi_A(x)$$

Then, observe that

$$\begin{aligned}
\mathrm{cut}(A, A^c) &= \sum_{x \in A, y \in A^c} w_{xy} \\
&= \sum_{x,y \in G} w_{xy} f(x)(1 - f(y)) \\
&= \frac{1}{2} \sum_{x,y \in G} w_{xy} (f(x) - f(y))^2
\end{aligned}$$

The last identity followed from the fact that

$$
\begin{aligned}
& f(x)(1 - f(y)) + f(y)(1 - f(x)) \\
& = f(x) + f(y) - 2f(x)f(y) = (f(x) - f(y))^2
\end{aligned}
$$

and the fact that $w_{xy} = w_{yx}$.

On the other hand,

$$\mathrm{Vol}(A) = \sum_{x \in A} \sum_{y \in G} w_{xy} = \sum_{x \in G} f(x) d(x)$$

Therefore

$$\begin{aligned} \mathrm{Ncut}(A, A^c) &= \frac{1}{2} \langle \Delta f, f \rangle \left(\frac{1}{\langle f, \mathbf{1} \rangle_d} + \frac{1}{\langle \mathbf{1} - f, \mathbf{1} \rangle_d} \right) \\ &= \frac{\langle \mathbf{1}, \mathbf{1} \rangle_d}{2} \frac{\langle \Delta f, f \rangle_d}{\langle f, \mathbf{1} \rangle_d \langle \mathbf{1} - f, \mathbf{1} \rangle_d} \end{aligned}$$

From Harmonic Functions to Spectral Clustering

A relaxation of the graph cut problem

Shi and Malik observed that if we define

$$u = f - b(1 - f), \quad \text{where } b = \frac{\mathrm{Vol}(A)}{\mathrm{Vol}(G)}$$

Then, A minimizes $\mathrm{Ncut}(A, A^c)$ if and only if u minimizes

$$\frac{\langle \Delta u, u \rangle}{\langle u, u \rangle_d}$$

over the set of u's satisfying the constraints

$$u(x) \in \{1, -b\} \; \forall \, x, \;\; \langle u, \mathbf{1} \rangle_d = 0.$$

From Harmonic Functions to Spectral Clustering

A relaxation of the graph cut problem

Now, this last problem is precisely equivalent to the Ncut minimization problem.

The **relaxation** consists in looking for a minimizer of

$$\frac{\langle \Delta u, u \rangle}{\langle u, u \rangle_d}$$

over all functions $u : G \mapsto \mathbb{R}$, **ignoring the constraints.**

This relaxed problem is the same as finding the first non-zero eigenfunction of the associated Laplacian. This problem can be solved in polynomial time.

From Harmonic Functions to Spectral Clustering

A relaxation of the graph cut problem

Of course, there is a drawback: the relaxed problem is of course not the same as the original one.

Having solved the relaxed problem, however, we can approximate the original solution to a great degree.

Simply, one takes as a guess

$$A = \{x \in G \mid \phi_2(x) > 0\}$$

(one can also take instead $\{\phi_2(x) > \delta\}$ for some other value δ)

Shi-Malik Ncut minimization

A relaxation of the graph cut problem

Shi-Malik Ncut minimization

A relaxation of the graph cut problem

Shi-Malik Ncut minimization

A relaxation of the graph cut problem