Convergence and Energy Landscape for Cheeger Cut Clustering

Xavier Bresson[1], Thomas Laurent[2], David Uminsky[3] and James von Brecht [4]

[1] City University of Hong Kong Dept. of Computer Science [2] UC Riverside Dept. of Mathematics [3] University of San Francisco Dept. of Mathematics [4] UCLA Dept. of Mathematics

Neural Information Processing Systems Foundation

Summary

We provide both theoretical and algorithmic results for the ℓ_1-relaxation of the Cheeger cut problem. The ℓ_1-relaxation is non-convex and non-differentiable, but is provably equivalent to the original problem. The ℓ_1-relaxation therefore trades convexity for exactness, yielding improved clustering results at the cost of a more challenging optimization. We provide the first complete proof of convergence for algorithms that minimize the ℓ_1-relaxation. We also characterize the ℓ_1-energy landscape, i.e. the set of possible points to which an algorithm might converge. We show that ℓ_1-algorithms can get trapped in local minima that are not globally optimal and we provide a classification theorem to interpret these local minima. This classification gives meaning to these sub-optimal solutions and helps to explain, in terms of graph structure, when the ℓ_1-relaxation provides the solution of the original Cheeger cut problem.

Steepest Descent Algorithm

$$E(f) = \frac{\|f\|_{TV}}{\|f - \text{median}(f)\|_{L_1}} = \frac{\|f\|_{TV}}{B(f - \text{median}(f))}$$

$$B(f) = \|f\|_{\ell_1} = \sum_i |f_i|$$

$$\partial_0 B(f) := \{v \in \mathbb{R}^n : B(g) - B(f) \geq \langle v, g - f \rangle \,, \langle \mathbf{1}, v \rangle = 0\}$$

$$\mathcal{S}_0^{n-1} := \{v \in \mathbb{R}^n : ||v||_{\ell_2} = 1 \,, \langle \mathbf{1}, v \rangle = 0\}$$

$f^0 \in \mathcal{S}_0^{n-1}$ nonzero function with median zero.

while $|E(f^{k+1}) - E(f^k)| \geq \text{TOL}$ **do**

$g^k = f^k + v^k$ where $v^k \in \partial_0 B(f^k)$

$\hat{h}^k = \arg\min_f \left\{ \|f\|_{TV} + \frac{1}{2} E(f^k) \, ||f - g^k||_2^2 \right\}$ (ROF)

$h^k = \hat{h}^k - \text{median}(\hat{h}^k)\mathbf{1}$

$f^{k+1} = \frac{h^k}{\|h^k\|_2}$

end while

Monotonicity:

$$E(f^k) \geq E(f^{k+1}) + \frac{E(f^k)}{B(h^k)} \, \|\hat{h}^k - f^k\|_2^2$$

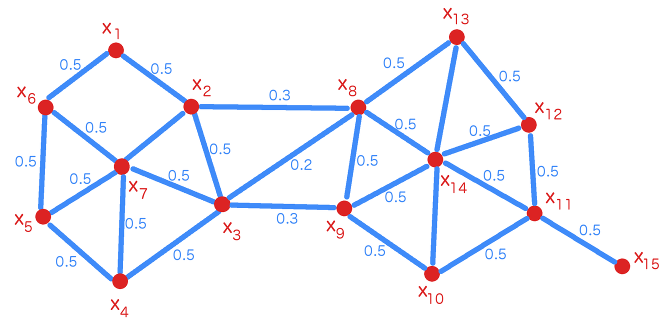
Stopping criteria for the inner loop:

$$E(f^k) \geq E(f^{k+1}) + \theta \frac{E(f^k)}{B(h^k)} \, \|\hat{h}^k - f^k\|_2^2 \qquad \theta \in (0, 1)$$

The Cheeger Cut Problem

Find the subset $A \subsetneq V$ which minimize

$$\frac{\text{cut}(A, A^c)}{\min(|A|, |A^c|)}$$



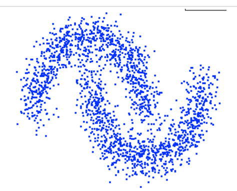
$V = \{x_1, x_2, \ldots, x_{15}\}$

NP hard!

Separate the data set into two groups of roughly equal size while cutting as few links as possible.

Total Variation Clustering

Spectral Clustering



Total Variation Clustering

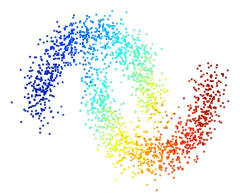
$$\min_{A \subsetneq V} \frac{\text{cut}(A, A^c)}{|A||A^c|} \qquad \min_{A \subsetneq V} \frac{\text{cut}(A, A^c)}{\min(|A|, |A^c|)}$$

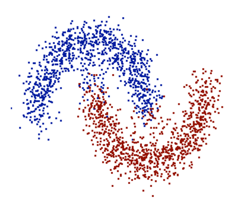
$$\Updownarrow \qquad \Updownarrow$$

$$\min_{f:V \to \{0,1\}} \frac{\sum_{ij} w_{ij} |f_i - f_j|^2}{\sum_i |f_i - m_2(f)|^2} \qquad \min_{f:V \to \{0,1\}} \frac{\sum_{ij} w_{ij} |f_i - f_j|}{\sum_i |f_i - m_1(f)|}$$

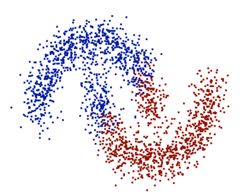
continuous relaxation

$$\min_{f:V \to \mathbb{R}} \frac{\sum_{ij} w_{ij} |f_i - f_j|^2}{\sum_i |f_i - m_2(f)|^2} \qquad \min_{f:V \to \mathbb{R}} \frac{\sum_{ij} w_{ij} |f_i - f_j|}{\sum_i |f_i - m_1(f)|}$$





treshold



The Continuous ℓ_1-Relaxation

Find the function $f : V \to \mathbb{R}$ which minimize

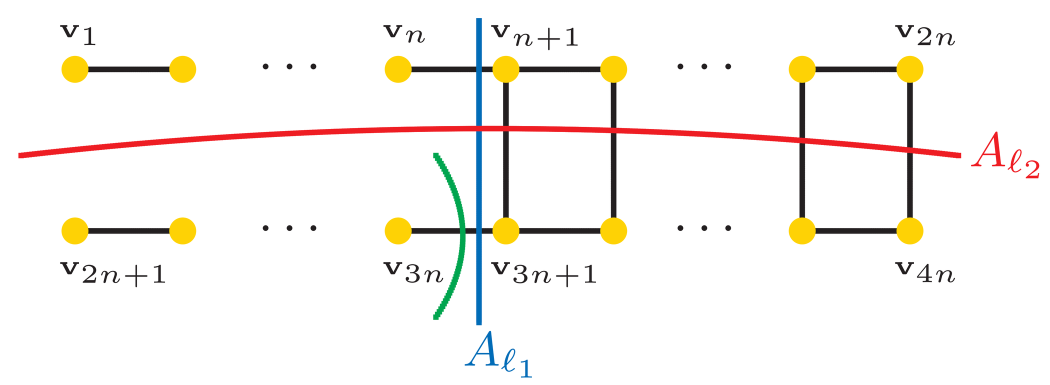
$$E(f) = \frac{\|f\|_{TV}}{\|f - \text{median}(f)\|_{\ell_1}}$$

where $\| \cdot \|_{TV}$ is the Total Variation semi-norm:

$$\|f\|_{TV} = \sum_{i,j} w_{i,j} |f_i - f_j|$$

The energy E is non-differentiable and non-convex

Exactness of the ℓ_1-Relaxation



$$\min_{A \subsetneq V} \frac{\text{cut}(A, A^c)}{|A||A^c|} = \frac{1}{3k^2} << \frac{1}{4k} = \frac{\text{cut}(A_{\ell_2}, A_{\ell_2}^c)}{|A_{\ell_2}||A_{\ell_2}^c|}$$

$$\min_{A \subsetneq V} \frac{\text{cut}(A, A^c)}{\min\{|A|, |A^c|\}} = \frac{\text{cut}(A_{\ell_1}, A_{\ell_1}^c)}{\min\{|A_{\ell_1}|, |A_{\ell_1}^c|\}}$$

The "cockroach" graph from Guattery and Miller [1]. The ℓ_2-relaxation of RatioCut, i.e. spectral clustering, gives the partition in red. The resulting RatioCut energy exhibits arbitrarily large deviations from the optimal solution of the RatioCut in green. The optimal solutions of the ℓ_1-relaxation and the Cheeger cut, shown in blue, coincide. The steepest descent algorithm finds the optimal ℓ_1 solution and therefore solves the Cheeger cut problem.

Mathematical Properties

"Continuity"
This allows to prove that all the accumulation points of the sequence $\{f^k\}$ are critical points of the energy.

$\|f^{k+1} - f^k\|_{L^2} \to 0$
Either the sequence converges, or the set of accumulation points is a connected subset of $\mathcal{S}_0^{n-1}$

Full convergence near local min:
If f^* is an isolated local minima of the energy, then $f^k \to f^*$ if the initial iterate is close enough to f^*.

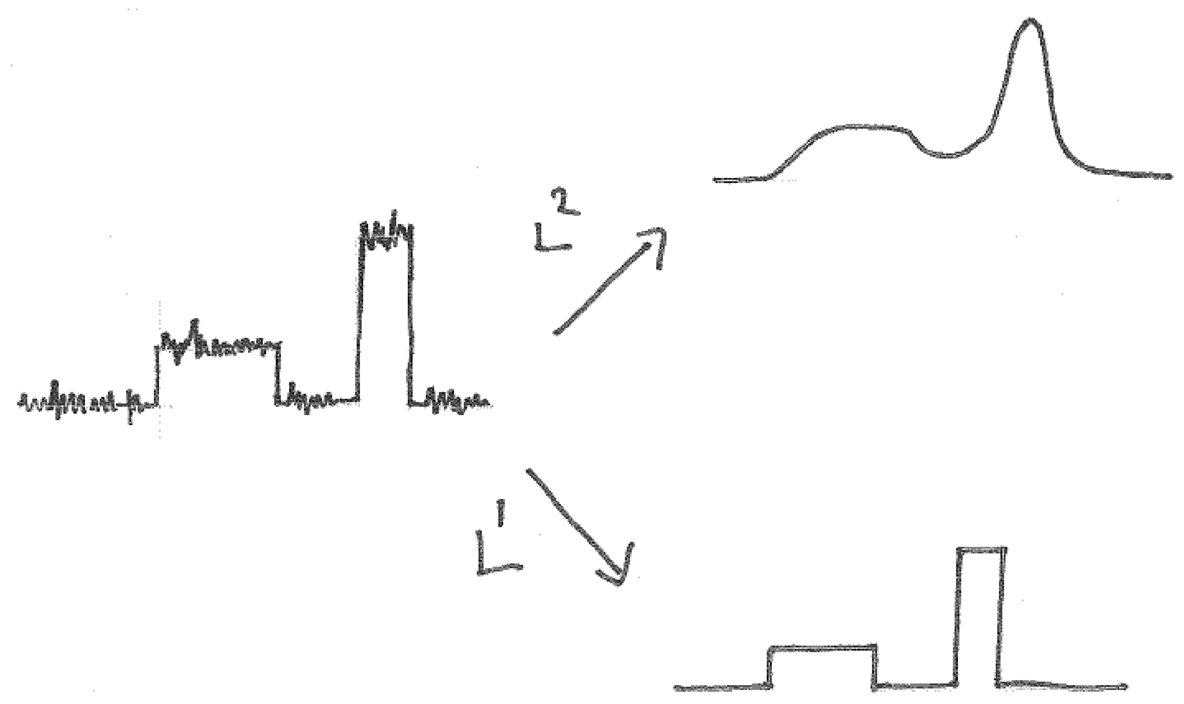
TV in Image Processing

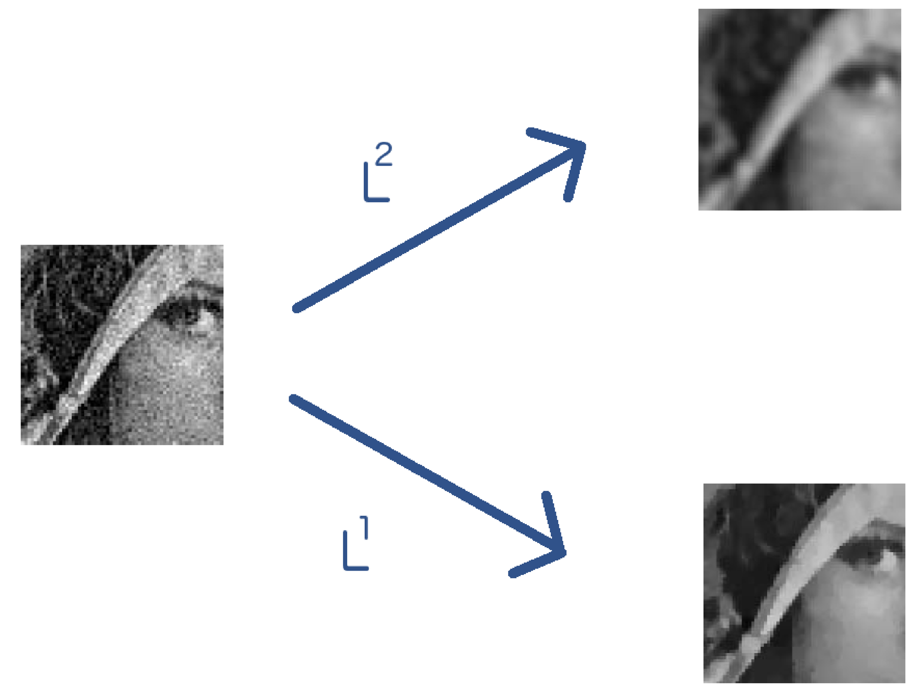
f_0

f_{denoized}

$$f_{\text{denoized}} = \arg\min_f \left\{ \|\nabla f\|_{L^2}^2 + \lambda \|f - f_0\|_{L^2}^2 \right\} \quad \text{(Heat eqn)}$$

$$f_{\text{denoized}} = \arg\min_f \left\{ \|\nabla f\|_{L^1} + \lambda \|f - f_0\|_{L^2}^2 \right\} \quad \text{(ROF model)}$$

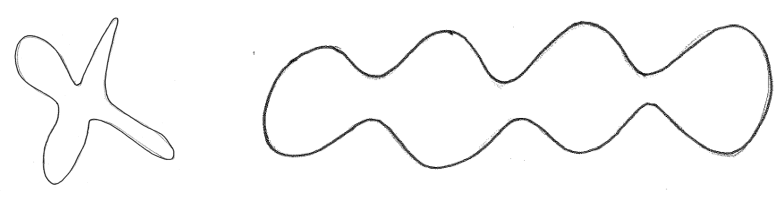




Total Variation preserves sharp edges

Dirichlet Energy does not preserve sharp edges

Characterization of Local Minima



A cut is a local minima of the non-convex energy E iff it has smaller energy than all the non-intersecting cuts. Therefore local minima are not "sub-classifications," but instead provide classifications based on different criteria.

Acknowledgements

This work supported by AFOSR MURI grant FA9550-10-1-0569 and Hong Kong GRF grant # 110311.

[1] Guattery, S. and Miller, G. L. *SIAM J. Matrix Anal. Appl.*, 1998.