CPSC 490a Thesis Report

Xinwei (David) Yao
Advisor: Daniel Spielman

May 18, 2016

Abstract

The thesis considers the hardness of the l_0-regularization problem of finding smooth extensions of functions on graphs, where one is given the function value at some initial vertices and an integer k and need to compute the function that minimizes the Laplacian quadratic form while coinciding with all but k of the initial values. The thesis focuses on its relation with the problem of finding the minimum bisection of a graph, which is known to be NP-hard, and gives a new reduction from MIN-BISECTION that yields further result on the relation between hardness of approximating the regularization problem and hardness of solving or approximating MIN-BISECTION for special graphs. Specifically, $\frac{1}{n^4}$-approximation of the regularization problem is NP-hard; $\frac{1}{n^2}$-approximation is at least as hard as MIN-BISECTION for bounded degree graphs; constant-factor approximation is at least as hard as constant-factor approximation of MIN-BISECTION for bounded degree graphs with large bisections.

Contents

1 Introduction

Solving regressions on graphs is a very important problem for graph-based semi-supervised learning. In such a problem, one is given the value of a function on some of the vertices of the graph and needs to find an extension of the function to all vertices in a most continuous manner respecting the edge weights. A common criteria for such continuity is to use the Laplacian quadratic form, as described below.

1.1 2-Laplacian minimization

Given a weighted undirected graph $G = (V, E, l)$ and a function $v_0 : T \to \mathbb{R}$ on $T \subseteq V$. $|V| = n$ and $|E| = m$. Consider weights of the edges as a measure for the disparity between the two endpoints, so that less weight means greater similarity. Then the 2-Laplacian minimization problem asks for an extension $v : V \to \mathbb{R}$ of the function v_0 so that it solves the following:

$$\min_{\substack{v \in \mathbb{R}^n \\ v|_T = v_0}} v^\intercal L v,$$

where L is the Laplacian matrix of G.

Note that

$$v^\intercal L v = \sum_{(x,y)\in E} l(x,y)(v(x) - v(y))^2,$$

so one can think of the function v as assigning **voltages** to each vertex that minimize the overall **energy** on the graph.

The set T is also called the **boundary** of the problem since it gives the boundary condition, and the function values on vertices are also called **labels** in the language of Machine Learning.

1.2 Regularization

When solving the regression on data sets, one would also like to be able to perform regularization on the data to ignore noises in the initial labels v_0 by removing the **outliers** from the boundary. The regularization that I consider is the l_0 regularization of vertex values, which allows $\|v|_T - v_0\|_0$, the 0-norm in $\mathbb{R}^{|T|}$, to be at most k. In this way, the problem of l_0-regularization for l_2 can be stated as finding the solution to the following:

$$\min_{\substack{v \in \mathbb{R}^n \\ \|v|_T - v_0\|_0 \le k}} v^\intercal L v$$

Call this problem 2-REGULARIZATION.

1.3 Prior Work

[1] proves the following theorem by a Karp reduction from MIN-BISECTION.

Theorem 1.1. 2-REGULARIZATION *is NP-Hard.*

I will present a new polynomial reduction from MIN-BISECTION that implies not only Theorem 1.1 but also results on the approximability of the problem.

2 Reduction from Min-Bisection

Let $G = (V, E)$ be any graph, and I will reduce MIN-BISECTION on G to 2-REGULARIZATION. Let $n = |V|$ and $m = |E|$.

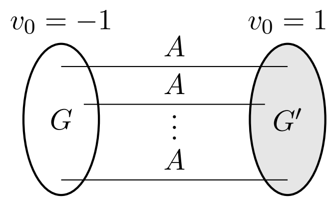

Figure 1: Reduction Gadget

2.1 Setup and notation

The graph $H = (V_H, E_H, w)$ on which the regularization is performed will be two copies of G with a matching of the corresponding vertices in between. $V_H = V \cup V'$. The edges in the two copies have edge weight 1, and the n cross-edges have weight A for some large A to be determined later.

Call the two copies G and G'. I will consistently use lower-case letters to denote vertices in G and upper-case letters to denote sets of vertices in G, and use letters with the prime symbol to denote their corresponding copies in G'. Let *deg* denote the degree of vertices in the original G.

Take any $k \leq n$, I will include all the vertices in the boundary. Let the initial label assignment v_0 be -1 on V and 1 on V'. Then regularization problem becomes changing k labels to achieve the lowest overall energy.

Define a function $F : \mathcal{P}(V_H) \to \mathbb{R}$. For any $K \subseteq V'_H$ be any set of vertices in H, let $T = V_H \setminus K$ be its complement in H. $F(K)$ is then defined to be the minimal energy achieved by relaxing the set K from the boundary:

$$F(K) = \min_{\substack{f \in \mathbb{R}^{2n} \\ v|_T = v_0|_T}} v^\intercal L_H v \tag{2.1}$$

where L_H is the Laplacian of the augmented graph H.

In this way, 2-REGULARIZATION becomes finding K^* and v^* so that $|K^*| = k$ and

$$F(K^*) = \min_{\substack{K \subseteq V_H \\ |K|=k}} F(K) = F^*, \tag{2.2}$$

for some $F^* \in \mathbb{R}$ that is the lowest possible energy. However, K^* may not be unique, so for my argument I will choose a K^* that is inside V as much as possible, that is:

$$|K^* \cap V| = \max_{\substack{|K|=k \\ F(K)=F^*}} |K \cap V| \tag{2.3}$$

v^* is just the function v_0 with change of labels in K^* that achieves F^*, that is:

$$\forall a \notin K^*, \; v^*(a) = v_0(a)$$

and

$$v^{*\intercal} L_H v^* = F(K^*) = F^*$$

The entire augmented graph H for regularization is shown in Figure 1. I will prove that when A is sufficiently large, K^* is a minimum-bisection of the original graph. The following fact I will use is from [2].

Proposition 2.1. *If v is the optimal function for relaxing set K, i.e. achieving the energy $F(K)$, then v is harmonic on H for the changed vertices. In other words,*

$$\forall a \notin K, v(a) = \frac{\sum_{(a,b)\in E_H} w_{a,b} v(b)}{\sum_{(a,b)\in E_H} w_{a,b}}$$

In particular, the boundary prescribes the minimum and the maximum value of v:

$$\forall a \in V_H, \min_{b\notin K} v_0(b) \leq v(a) \leq \max_{b\notin K} v_0(b)$$

2.2 Relaxation on one side only

In this subsection I will prove the following theorem.

Theorem 2.2. *For any Graph $G = (V, E)$ and any integer $0 \leq k \leq n$, $K^* \subseteq V$ if $A \geq 3d_{max}$ where d_{max} is the maximum degree of a node in G.*

I will first show the following Lemma.

Lemma 2.3. *For $A \geq 3d_{max}$, and some candidate K, $|K| = k \leq n$, if $\exists a \in K$ s.t. $a' \in K$, then $\exists K'$ s.t. $|K'| = |K|$, $F(K') \leq F(K)$, and $|K' \cap V| > |K \cap V|$.*

Proof. Let v be the function for K achieving $F(K)$. Since $k \leq n$, there are $k-2$ elements other than a and a' in K. But there are $n-1$ remaining pairs from the matching in H, then by Pigeonhole Principle we can find $b \in V$ s.t. $b, b' \notin K$.

Let $K' = K \setminus \{a'\} \cup \{b\}$, and consider a test vector v' for K', where

$$v'(x) = \begin{cases} 1, & x = a \\ v_0(a') = 1, & x = a' \\ 1, & x = b \\ v(x), & o.w. \end{cases}$$

By Equation 2.1, we have $F(K') \leq v'^{\intercal} L_H v'$, so $\Delta = F(K) - F(K') \geq v^{\intercal} L_H v - v'^{\intercal} L_H v'$. Since v and v' are mostly the same, this difference only comes from edges that are attached to a, a' and b.

The contribution from b is then

$$\begin{aligned} \Delta_b &= A(v(b') - v(b))^2 - 0 + \sum_{(b,c)\in E} (v(b) - v(c))^2 - (v'(b) - v'(c))^2 \\ &\geq 4A + \sum_{(b,c)\in E} -(v'(b) - v'(c))^2 \\ &= 4A + \sum_{(b,c)\in E,\ c\neq a} -(1 - v(c))^2, \text{ since } v'(a) = v'(b) = 1 \end{aligned}$$

By Proposition 2.1, $-1 \leq v(c) \leq 1$, so $(1 - v(c)) \leq 2$, and

$$\Delta_b \geq 4A - 4deg(b) \geq 4A - 4d_{max}$$

Similarly, the contribution from a and a' is

$$\begin{aligned}\Delta_{aa'} &= A(v(a)-v(a'))^2 - 0 \\ &\quad + \sum_{(a,c)\in E} (v(a)-v(c))^2 - (v'(a)-v'(c))^2 + (v(a')-v(c'))^2 - (v'(a')-v'(c'))^2 \\ &\geq \sum_{(a,c)\in E} -(v'(a)-v'(c))^2 - (v'(a')-v'(c'))^2 \\ &\geq \sum_{(a,c)\in E,\ c\neq b} -(1-v(c))^2 - (1-v(c'))^2, \text{ since } v'(b)=v'(b')=v'(a)=v'(a')=1 \\ &\geq -8deg(a) \\ &\geq -8d_{max}\end{aligned}$$

If $A \geq 3d_{max}$, combining the two terms we have $\Delta \geq \Delta_b + \Delta_{aa'} \geq 4(A-3d_{max}) \geq 0$. Therefore $F(K') \leq F(K)$ and by construction $|K'| = k$ and $|K' \cap V| > |K \cap V|$ □

Lemma 2.4. *For some candidate K, $|K| = k \leq n$. If $\forall a \in K \cap V$, $a' \notin K$, then $\exists K' \subseteq V$ s.t. $|K'| = k$ and $F(K') \leq F(K)$.*

Proof. Suppose $K \nsubseteq V$, since otherwise we can set $K' = K$. Then K is partitioned into S and T' so that $S \cup T' = K$, $S \subseteq V$ and $T' \subseteq V'$. We also have that $S \cap T = \emptyset$ where T is the copy of T' in G. Let v be the function for K that achieves $F(K)$.

Let $K' = K \setminus T' \cup T \subseteq V$, and consider a test vector v' for K', where

$$v'(x) = \begin{cases} v_0(x) = 1, & x \in V' \\ v(x), & x \in S \\ -v(x'), & x \in T \\ v_0(x) = -1, & o.w. \end{cases}$$

By Equation 2.1, we then have $F(K') \leq v'^{\intercal} L_H v'$, so $\Delta = F(K) - F(K') \geq v^{\intercal} L_H v - v'^{\intercal} L_H v'$.

Notice that contributions of each cross edge stays the same, as $(-1-v(x'))^2 = (1-v'(x))^2$ whenever $x \in T$. Contributions of edges within T for v' is also the same as contributions of edges within T' for v. Furthermore, contributions of edges from T to $V \setminus K'$ for v' is the same as contributions of edges from T' to $V' \setminus (S' \cup T')$. Therefore the difference only comes from edges from T to S and from T' to S'.

$$\begin{aligned}\Delta &\geq \sum_{\substack{(a,b)\in E \\ a\in S,\ b\in T}} (v(a')-v(b'))^2 + (v(a)-v(b))^2 - (v'(a)-v'(b))^2 - (v'(a')-v'(b'))^2 \\ &= \sum (1-v(b'))^2 + (v(a)+1)^2 - (v(a)+v(b'))^2 \\ &= \sum 2 + 2v(a) - 2v(b') - 2v(a)v(b') \\ &= \sum 2(1+v(a))(1-v(b'))\end{aligned}$$

By Proposition 2.1, each term in the sum is non-negative, so $\Delta \geq 0$. Therefore $\Delta \geq 0$, $F(K') \leq F(K)$ and $|K'| = k$ and $K' \subseteq V$. □

Remark 2.5. When $K \subseteq V'$, the K' and v' constructed in the proof is the symmetric reflection of v and K, and $F(K) = F(K')$ since $S = S' = \emptyset$.

Now I can prove Theorem 2.2 from Lemma 2.3 and Lemma 2.4.

Proof of Theorem 2.2. By Lemma 2.3 and Equation 2.3, we conclude that $\forall a \in K^* \cap V$, $a' \notin K^*$. Then we can apply Lemma 2.4 and conclude that $K^* \subseteq V$. □

2.3 Bounding the lowest energy value

Theorem 2.6. *For any graph G and any integer $0 \leq k \leq n$, $K \subseteq V$. Let $\delta = |\partial K|$ be the number of edges cut in G. Let $T = V \setminus K$ be the complement of K in G. Then $F(K)$ is between $4A(n-k) + 4\delta - \frac{4d_{max}}{A+d_{max}}\delta$ and $4A(n-k) + 4\delta$.*

Proof. The upper bound is given by a test vector v' for K s.t. $v'(K) = 1$.

$$F(K) \leq v'^{\intercal} L_H v' = 4A * |T| + 4\delta = 4A(n-k) + 4\delta$$

For any $S \subseteq V$, $U \subseteq V$, write $E(S,U)$ to denote the edges across the two sets, that is $E(S,U) = \{(a,b) \in E : a \in S, b \in U\}$.

For $a \in K$, let $E(a,\mathrm{T}) = \{(a,b) \in E : b \in T\}$ be the set of edges going from a to outside of K. Then we have

$$\delta = \sum_{a \in K} |E(a,T)| = |E(K,T)| \tag{2.4}$$

To get a lower bound for $F(K)$, we split it into contributions from three sets of edges: edges within K, edges from K to T, and cross edges between G and G'. Let v be the function for K that achieves $F(K)$.

$$\begin{aligned}
v^{\intercal} L_H v = & \sum_{(a,b) \in E(K,K)} (v(a) - v(b))^2 \\
& + \sum_{(a,b) \in E(K,T)} (v(a)+1)^2 + \sum_{a \in K} A(v(a)-1)^2 \\
& + 4A * |T| \\
\geq & \; 4A(n-k) + \sum_{a \in K} |E(a,T)|(v(a)+1)^2 + A(v(a)-1)^2 \\
= & \; 4A(n-k) + \sum_{a \in K} (|E(a,T)| + A)(v(a)^2 + \frac{2(|E(a,T)| - A)}{|E(a,T)| + A} v(a) + 1) \\
\geq & \; 4A(n-k) + \sum_{a \in K} (|E(a,T)| + A)(1 - \left(\frac{|E(a,T)| - A}{|E(a,T)| + A}\right)^2) \\
= & \; 4A(n-k) + \sum_{a \in K} \frac{4|E(a,T)|A}{|E(a,T)| + A} \\
\geq & \; 4A(n-k) + \sum_{a \in K} \frac{4A|E(a,T)|}{A + d_{max}} \\
= & \; 4A(n-k) + \frac{4A}{A + d_{max}} \sum_{a \in K} |E(a,T)| \\
= & \; 4A(n-k) + \left(4 - \frac{4d_{max}}{A + d_{max}}\right)\delta, \text{ by Equation 2.4}
\end{aligned}$$

Therefore $F(K)$ is between $4A(n-k)+4\delta-\frac{4d_{max}}{A+d_{max}}\delta$ and $4A(n-k)+4\delta$. □

2.4 Min-Bisection

The relation with MIN-BISECTION can be shown by applying Theorem 2.2 and Theorem 2.6.

Let $k=\frac{n}{2}$. Since $d_{max}\le n-1$ and $\delta\le(n-k)k=\frac{n^2}{4}$, choose $A=n^3>3d_{max}$, then by Theorem 2.2, we have $K^*\subseteq V$. Furthermore, $\frac{d_{max}\delta}{A+d_{max}}<\frac{1}{4}$, so by Theorem 2.6,

$$F^*\in(2n^4+4\delta-1,2n^4+4\delta]$$

which strictly deceases with δ, so K^* has to be a minimal bisection.

In summary, the reduction is as follows. Given any graph $G=(V,E)$, construct 2-REGULARIZATION problem on $H=(V_H,E_H,w)$ takes polynomial time. Solving 2-REGULARIZATION for $k=\frac{n}{2}$ gives the set K^*, which is a min-bisection of G.

3 Hardness Implications

3.1 Hardness of exact result

The polynomial reduction described in Section 2.4 implies Theorem 1.1.

3.2 Hardness of $\frac{1}{n^4}$-approximation

Theorem 3.1. $\frac{1}{(\frac{1}{2}+\epsilon)n^4}$*-approximation of* 2-REGULARIZATION *is NP-hard*

Proof. Suppose there is a $\frac{1}{(\frac{1}{2}+\epsilon)n^4}$ approximation, set $A=\frac{n^3}{2}$. Let K be the min-bisection and $\delta=|\partial K|$.

By Lemma 2.3 and Lemma 2.4, we can transform the output set of the approximation algorithm into a set $K'\subseteq V$ which is at least as good an approximation, i.e. $F(K')\le(1+\frac{1}{(\frac{1}{2}+\epsilon)n^4})F^*$. Let $\delta'=|\partial K'|$. By Theorem 2.6, we have

$$n^4+4\delta'-\frac{4d_{max}}{n^3/2+d_{max}}\delta'\le(1+\frac{1}{(\frac{1}{2}+\epsilon)n^4})(n^4+4\delta)$$

which simplifies to

$$\begin{aligned}\delta'-\delta&\le\frac{d_{max}\delta'}{n^3/2+d_{max}}+\frac{1}{2+4\epsilon}+\frac{\delta}{(\frac{1}{2}+\epsilon)n^4}\\&<\frac{(n-1)n^2/4}{n^3/2}+\frac{1}{2+\epsilon}+\frac{n^2/4}{2n^4}\\&<\frac{1}{2}+\frac{1}{2+\epsilon}+\frac{1}{2n^2}\\&<1,\text{ when } n \text{ is sufficiently large}\end{aligned}$$

So for sufficiently large n, $\delta'=\delta$ and K' is the min-bisection of G. Since solving MIN-BISECTION is NP-hard, $\frac{1}{(\frac{1}{2}+\epsilon)n^4}$-approximation of 2-REGULARIZATION is NP-hard. □

3.3 Hardness of $\frac{1}{n^2}$-approximation

Consider MIN-BISECTION on graphs with constant bounded degree d, and call this problem MIN-BISECTION$_d$.

Theorem 3.2. *There is a polynomial reduction from* MIN-BISECTION$_d$ *to a* $\frac{1}{(d^2+\epsilon)n^2}$*-approximation of* 2-REGULARIZATION

Proof. Set $A = d^2 n$. Since G has bounded degree d, with the same set up as in Theorem 3.1, we have $\delta \leq \delta' \leq \frac{dn}{2}$. Using Theorem 2.6, we get

$$\begin{aligned}\delta' - \delta &\leq \frac{d\delta'}{d^2 n + d} + \frac{d^2 n}{2(d^2+\epsilon)n} + \frac{\delta}{(d^2+\epsilon)n^2} \\ &< \frac{1}{2} + \frac{1}{2 + 2d^{-2}\epsilon} + \frac{1}{2dn} \\ &< 1, \text{ when } n \text{ is sufficiently large}\end{aligned}$$

Therefore for sufficiently large n, $\delta' = \delta$ and K' is the min-bisection of G. □

Remark 3.3. This implies that getting $\frac{1}{(d^2+\epsilon)n^2}$-approximation of 2-REGULARIZATION is at least as hard as solving MIN-BISECTION on graphs with bounded degree d.

3.4 Hardness of constant-factor approximation

Consider the problem MIN-BISECTION$_{d,\beta}$. It is MIN-BISECTION$_d$ on graphs with not only constant bounded degree d, but also that any half of the vertices will have good expansion. More precisely, they are graphs G with $\delta(G) \geq \beta nd$, where $\delta(G)$ is the number of edges cut in the minimum bisection.

Theorem 3.4. *There is a polynomial reduction from ϵ-approximation of* MIN-BISECTION$_{d,\beta}$ *to a* $\frac{\beta\epsilon^2}{\alpha}$*-approximation of* 2-REGULARIZATION*, where* $\alpha = 2 + \epsilon + \frac{3}{2}\epsilon^2$.

Proof. Set $A = \frac{2d}{\epsilon}$. Use the same set up as in Theorem 3.1. We have $\beta dn \leq \delta \leq \delta' \leq \frac{dn}{2}$. Using Theorem 2.6, we get

$$
\begin{aligned}
2nA + 4\delta' - \frac{4d}{A+d}\delta' &\leq (1 + \frac{\beta\epsilon^2}{\alpha})(2nA + 4\delta) \\
\frac{4A}{A+d}\delta' &\leq 4\delta + \frac{\beta\epsilon^2}{\alpha}(2nA + 4\delta) \\
\frac{A}{A+d}\delta' &\leq (1 + \frac{\beta\epsilon^2}{\alpha})\delta + \frac{\beta\epsilon nd}{\alpha} \\
\frac{A}{A+d}\frac{\delta'}{\delta} &\leq (1 + \frac{\beta\epsilon^2}{\alpha}) + \frac{\beta\epsilon nd}{\alpha\beta nd} \\
\frac{\delta'}{\delta} &\leq (1 + \frac{\beta\epsilon^2}{\alpha} + \frac{\epsilon}{\alpha})(1 + \frac{d}{A}) \\
&\leq (1 + \frac{\beta\epsilon^2}{\alpha} + \frac{\epsilon}{\alpha})(1 + \frac{\epsilon}{2}) \\
&\leq 1 + \frac{\epsilon}{2} + \frac{\beta\epsilon^2 + \beta\epsilon^3/2 + \epsilon + \epsilon^2/2}{\alpha} \\
&\leq 1 + \frac{\epsilon}{2} + \frac{\epsilon}{2}\frac{2\beta\epsilon + \beta\epsilon^2 + 2 + \epsilon^2}{\alpha} \\
&\leq 1 + \epsilon, \text{ since } \beta \leq \tfrac{1}{2}
\end{aligned}
$$

Therefore δ' is a ϵ-approximation of δ, the min-bisection of G. □

Remark 3.5. This implies that getting $\frac{2\beta\epsilon^2}{9}$-approximation of 2-REGULARIZATION is at least as hard as getting ϵ-approximation for MIN-BISECTION$_{d,\beta}$ for $\epsilon \leq 1$

Acknowledgement

I would like to thank Professor Spielman for the immensely helpful weekly discussions throughout the semester, without which this thesis would not have been possible. I would also like to thank Sushant Sachdeva for his insightful ideas.

References

[1] Rasmus Kyng, Anup Rao, Sushant Sachdeva, and Daniel A. Spielman. Algorithms for lipschitz learning on graphs. In *Proceedings of The 28th Conference on Learning Theory*, pages 1990–1223, 2015.

[2] Xiaojin Zhu, Zoubin Ghaharamani, and John Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML)*, pages 912–919, 2003.