

p-voltages: Laplacian Regularization for Semi-Supervised Learning on High-Dimensional Data

Nick Bridle
Department of Computer Sciences
University of Wisconsin-Madison
nbridle@cs.wisc.edu

Xiaojin Zhu
Department of Computer Sciences
University of Wisconsin-Madison
jerryzhu@cs.wisc.edu

ABSTRACT

We investigate the p -voltages algorithm, which labels nodes in a graph based on their theoretical voltages in a reformulated system of electricity. Building on previous work concerning p -electric networks, we prove that the p -voltage solution is well-formed and has desirable properties for semi-supervised learning. Our experiments confirm that the p -voltages algorithm does not suffer from the same weaknesses as the Laplacian Regularization algorithm (equivalent to $p = 2$) and therefore improves classification performance. However, our p -voltages algorithm does not outperform the state-of-the-art iterated Laplacian algorithm.

1. INTRODUCTION

A popular method for semi-supervised learning (SSL) on graphs is Laplacian Regularization [1]. It has underlying connections to random walks and electric networks and works well in practice on low dimensional data. However, Nadler et al [2] showed that in higher dimensions when the number of unlabeled vertices increases, the resulting label function becomes constant almost everywhere with extremely thin “spikes” at the labeled points, which is undesirable for both classification and regression. In this paper, we examine the p -voltages algorithm which seeks to combat this issue.

From an electric network perspective, we can think of the graph as an electric circuit with edge weights interpreted as conductance. The Laplacian Regularization solution is exactly the set of voltages that would occur in a physical circuit if the voltages of the labeled vertices were held constant. Intuitively, the larger the resistance between two vertices, the more energy that must be expended for current to travel between them, and so the larger their difference in voltage. It turns out that as the number of vertices grows, the current becomes too widely distributed to encounter any significant resistance in the majority of the graph, which leads to the unwanted flatness in the solution.

Alamgir et al [3] attempted to address this problem with a concept of p -electric networks, which reformulates the defini-

tion of power to favor flows which are concentrated on fewer paths in the graph. Their study focuses on p -resistances, which are the analog of effective resistance in p -electric networks. However, they did not examine closely the *voltages* resulting from such a p -electric which we call “ p -voltages”. They did propose to label the vertices by p -voltages (which they call q -Laplacian Regularization), but they connect it to p -resistances via a conjecture which we prove to be false in this paper. p -voltages has appeared in other previous work but did not receive sufficient attention. Specifically, Herbster et al [4] formulated a family of p -electric networks, and proposed labeling by p -voltages (which they call p -seminorm interpolation) in an online learning setting. However, they advocated using a value of p that suffers from flatness, leading to poor empirical performance. In contrast, we argue that labeling by p -voltages with the correct setting for p fixes the flaws and is suitable for semi-supervised learning on graphs.

This paper makes three major contributions. First, we disprove the conjecture of [3]. Next, we prove two properties of p -voltages which demonstrate that they do not suffer from the issues in [2]. In doing so, we utilize the machinery about p -electric networks developed by [3] and [4]. Finally, we present empirical results of p -voltages on both real and synthetic data and compare them to other state-of-the-art graph-based semi-supervised learning algorithms.

2. BACKGROUND

There is a set of data points $X = \{X_1, \dots, X_n\}$ drawn from an underlying smooth density $p(x)$ in \mathbb{R}^d . We know the labels of some of the points $X_l \subset X$, and would like to predict the labels of the unlabeled points $X_u := X \setminus X_l$. To capture the similarities between the points, we construct an undirected graph $G = (V, E)$ where the points are vertices and the edge weights correspond to some similarity function between the points. A common choice for the similarity function is the RBF kernel:

$$w_{ij} = e^{-\frac{\|X_i - X_j\|^2}{2\sigma^2}} \quad (1)$$

We can view this graph as an electric circuit where the edges are resistors with resistance $r_e = 1/w_e$. That is, the more similar two points are, the lower the resistance in the edge between them. Consider a binary semi-supervised classification problem with one labeled point per class. The two labeled points are named the source s and the sink t , respectively. To address the flatness issue pointed out by [2], Alamgir et al [3] proposed the following optimization problem for computing the effective p -resistance on the graph:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Eleventh Workshop on Mining and Learning with Graphs. Chicago, Illinois, USA

Copyright 2013 ACM 978-1-4503-2322-2 ...\$15.00.

$$R_p(s, t) = \min_i \left\{ \sum_{e \in E} r_e |i_e|^p \mid i = \{i_e\} \text{ is a unit flow } s \rightarrow t \right\} \quad (2)$$

This simulates running a unit current from s to t and selecting the current which minimizes the p -energy shown above. There is an equivalent optimization problem which computes the effective conductance of the graph:

$$C_p(s, t) = \min_v \left\{ \sum_{(a,b) \in E} \frac{|v_a - v_b|^{\frac{p}{p-1}}}{r_{ab}^{\frac{1}{p-1}}} \mid v_s - v_t = 1 \right\} \quad (3)$$

This simulates imposing a unit difference in voltage between s and t and selecting voltages for the other vertices which minimize the p -energy shown above. Notice that when $p = 2$, these reduce to standard Harmonic energy minimization in an electric circuit. The solutions i^* from $R_2(s, t)$ and v^* from $C_2(s, t)$ are exactly the currents and voltages, respectively, that would arise in a real circuit, and v^* is the solution to the original Laplacian Regularization algorithm of [1]. (2) and (3) generalize electric networks to p -electric networks, where the current flows differently because the definition of energy to be minimized is changed. We only consider values of $p > 1$. Alamgir et al [3] showed that when $p < p^* := \frac{d}{d-1}$, the p -resistance contains meaningful information about the graph as the size of the graph grows to infinity, and when $p > p^{**} := \frac{d-1}{d-2}$ the p -resistance converges to a meaningless quantity.

The result of solving $R_p(s, t)$ on a graph is a set of currents i_R which form a unit flow and which minimize the p -energy of the graph. It can be shown that there is a unique mapping of points to values $v : V \rightarrow \mathbb{R}$ such that every edge $e = (a, b)$ in the graph has the following property [4]:

$$v_a - v_b = \text{sgn}(i_{ab}) |i_{ab}|^{p-1} r_{ab} \quad (4)$$

This can be viewed as p -Ohm's Law, Ohm's Law for general p -electric networks. Call the set of voltages derived in this way v_R , with $v_R(s) = R_p(s, t)$ and $v_R(t) = 0$.

When we solve the equivalent $C_p(s, t)$ problem (3) with $v_s = 1$ and $v_t = 0$ on the same graph, we find a set of p -voltages v_C that minimizes the p -energy of the graph. p -Ohm's Law holds in this setting as well, and the corresponding set of currents i_C can be calculated using (4). We will develop the relationship between $R_p(s, t)$ and $C_p(s, t)$ later as Equation (13), and show that for every node $u \in V$,

$$v_R(u) = R_p(s, t) \cdot v_C(u) \quad (5)$$

That is, v_R and v_C are proportional to each other with factor $R_p(s, t)$. In the following proofs we will discuss properties of v_R , and the same properties will follow for v_C (p -voltages) directly since they are proportional.

The p -voltages classification algorithm proceeds as follows. First, we construct a similarity graph over the dataset such as a knn - or ϵ -graph. Specific conditions for graph construction are covered in Section 3.2. Next, we set voltages $v_i = 1$ for all labeled points X_i in the "positive" class, and $v_j = 0$ for all labeled points X_j in the "negative" class. We solve the $C_p(s, t)$ problem (3) with $p \in \left(1, \frac{d}{d-1}\right)$, where d is the underlying dimension of the data. A standard numerical solver can be used to calculate the answer. $C_p(s, t)$ is a

convex optimization problem over a convex set, so it has a unique minimizer, which yields a set of p -voltages $\{v_i^*\}$ over all points in the graph. Finally, we threshold the p -voltages: all points with $v_i^* > 0.5$ are predicted as "positive", and all points with $v_i^* \leq 0.5$ are predicted as "negative". In the regression setting, we follow the same steps but skip the final thresholding step.

3. THEORETICAL RESULTS

Our first main result shows that while p -voltages and p -resistances are related, using them to label graphs in the SSL setting results in different solutions.

3.1 Disproving the p -resistance conjecture

The authors of [3] propose the following conjecture in Section 5.

CONJECTURE 1. *Given a graph $G = (V, E)$, consider a semi-supervised classification problem with one labeled point per class: $v_s = 1, v_t = -1$. Let the solution to the conductance problem (3) be v^* , and let u be an unlabeled point. Then, for all $p > 1$,*

$$v_u^* - v_t^* > v_s^* - v_u^* \iff R_p(u, t) > R_p(s, u) \quad (6)$$

They showed that this property holds for $p = 2$, and their conjecture is that it holds for general $p > 1$. If true, this property (6) would be extremely useful because the p -voltages algorithm would be an efficient surrogate method for classification with p -resistances. Classification with p -resistances as proposed in [3] is the right-hand-side of (6): each point u is in class 1 iff $R_p(u, t) > R_p(s, u)$. Alamgir et al. proved that classification with p -resistances captures cluster structure in the graph and does not suffer from the flatness issue, if $p < p^*$. Note, however, that this requires computing two effective p -resistances $R_p(u, t), R_p(s, u)$ for each unlabeled point u by solving (2), which is not practical for large graphs. The value of Conjecture 1, if it were true, would be to solve v_C only once and obtain all the classifications that effective p -resistances would produce.

However, we can show that this conjecture does not hold for general p . In numerical experiments over random geometric graphs, we found that it often fails. Below we present an analytic counterexample where the conjecture does not hold.

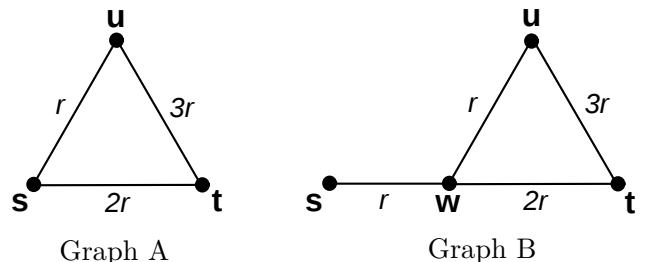


Figure 1: Two sample graphs. The resistance of each edge is shown as a multiple of r . In both graphs we fix $v_s = 1$ and $v_t = -1$.

First we consider Graph A, shown in Figure 1. For generality, we set the edge resistances to multiples of r where

r can be any finite positive real number. To compute the value of v_u^* , we directly minimize (3):

$$\begin{aligned} C_p(s, t) &\equiv \min_v P_C \\ &= \min_v \frac{|v_s - v_u|^{\frac{p}{p-1}}}{r^{\frac{1}{p-1}}} + \frac{|v_u - v_t|^{\frac{p}{p-1}}}{(3r)^{\frac{1}{p-1}}} + \frac{|v_s - v_t|^{\frac{p}{p-1}}}{(2r)^{\frac{1}{p-1}}} \\ \frac{\partial P_C}{\partial v_u} &= \frac{p}{p-1} \left(-\frac{|v_s^* - v_u^*|^{\frac{1}{p-1}}}{r^{\frac{1}{p-1}}} + \frac{|v_u^* - v_t^*|^{\frac{1}{p-1}}}{(3r)^{\frac{1}{p-1}}} \right) = 0 \\ v_s^* - v_u^* &= \frac{1}{3}(v_s^* - v_t^*) \\ v_u^* &= \frac{3}{4}v_s^* + \frac{1}{4}v_t^* = \frac{3}{4}(1) + \frac{1}{4}(-1) = \frac{1}{2} \end{aligned}$$

To solve for $R_p(s, u)$ and $R_p(u, t)$, we use the definition from (2):

$$\begin{aligned} R_p(s, u) &= \min_i P_R \equiv r(1-i)^p + 5r(i)^p \\ \frac{\partial P_R}{\partial i} &= -rp(1-i^*)^{p-1} + 5rp(i^*)^{p-1} = 0 \\ i^* &= \frac{1}{5^{\frac{1}{p-1}} + 1} \\ R_p(s, u) &= r(1-i^*)^p + 5r(i^*)^p \\ &= \frac{5\mathbf{r}}{(5^{\frac{1}{p-1}} + 1)^{p-1}} \\ R_p(u, t) &= \min_i P_R \equiv 3r(1-i)^p + 3r(i)^p \\ \frac{\partial P_R}{\partial i} &= -3rp(1-i^*)^{p-1} + 3rp(i^*)^{p-1} = 0 \\ i^* &= \frac{1}{2} \\ R_p(u, t) &= 3r(1-i^*)^p + 3r(i^*)^p \\ &= \frac{3\mathbf{r}}{2^{p-1}} \end{aligned}$$

$R_p(s, u) < R_p(u, t)$ for $p \in (1, \infty)$ and v_u^* is closer to v_s^* than it is to v_t^* , so the conjecture holds for Graph A for finite p .

Now let us consider the slightly modified Graph B, also shown in Figure 1. We can show that the conjecture does not hold on this graph for node u . Since the structure of the graph is very similar, we can reuse some of our intermediate results above. Specifically, we know that $v_u^* = \frac{3}{4}v_s^* + \frac{1}{4}v_t^*$. Again, we hold the labeled points fixed: $v_s = 1, v_t = -1$. For notational convenience, we define the constant $\alpha := \left[\frac{2}{4^{\frac{1}{p-1}}} + \frac{1}{2^{\frac{1}{p-1}}} \right]$.

$$\begin{aligned} C_p(s, t) &= \frac{|v_s - v_w|^{\frac{p}{p-1}}}{r^{\frac{1}{p-1}}} + \frac{|v_w - v_u|^{\frac{p}{p-1}}}{r^{\frac{1}{p-1}}} + \\ &\quad \frac{|v_u - v_t|^{\frac{p}{p-1}}}{(3r)^{\frac{1}{p-1}}} + \frac{|v_w - v_t|^{\frac{p}{p-1}}}{(2r)^{\frac{1}{p-1}}} \\ \frac{\partial P_C}{\partial v_w} &= \frac{p}{(p-1)(r^{\frac{1}{p-1}})} \left(-\left| v_s^* - v_w^* \right|^{\frac{1}{p-1}} + \left| \frac{1}{4}v_w^* - \frac{1}{4}v_t^* \right|^{\frac{1}{p-1}} + \right. \\ &\quad \left. \frac{1}{3^{\frac{1}{p-1}}} \left| \frac{3}{4}v_w^* - \frac{3}{4}v_t^* \right|^{\frac{1}{p-1}} + \frac{1}{2^{\frac{1}{p-1}}} \left| v_w^* - v_t^* \right|^{\frac{1}{p-1}} \right) = 0 \\ &= -|v_s^* - v_w^*|^{\frac{1}{p-1}} + \alpha |v_w^* - v_t^*|^{\frac{1}{p-1}} \end{aligned}$$

$$\begin{aligned} v_w^* &= \frac{1 - \alpha^{p-1}}{1 + \alpha^{p-1}} \\ v_u^* &= \frac{3}{4}v_w^* - \frac{1}{4} \end{aligned}$$

To solve for $R_p(s, u)$ and $R_p(u, t)$, we can re-use our results from Graph A:

$$R_p(s, u) = \min_i P_R = r(1-i)^p + r(1-i)^p + 5r(i)^p$$

The optimal flow is identical to the one calculated for $R_p(s, u)$ in Graph A, with a unit flow through the new edge from s to w .

$$R_p(s, u) = \mathbf{r} + \frac{5\mathbf{r}}{(5^{\frac{1}{p-1}} + 1)^{p-1}}$$

Similarly, the optimal flow for $R_p(u, t)$ is identical to the one calculated for $R_p(u, t)$ in Graph A. No current flows on the new edge from s to w .

$$R_p(u, t) = \min_i P_R = 3r(1-i)^p + 3r(i)^p = \frac{3\mathbf{r}}{2^{p-1}}.$$

Over the range $p \in (1, 2]$, α^{p-1} takes values in $(\frac{1}{2}, 1]$, which means that $v_w^* \in [0, \frac{1}{3})$ and $v_u^* \in [-\frac{1}{4}, 0)$. Therefore, $v_u^* - v_t^* < v_s^* - v_u^*$ over this range of p . However, for many values of $p \in (1, 2]$, including all $p \leq 1.6$, $R_p(s, u) < R_p(u, t)$, so the conjecture is violated for Graph B.

Thus, we have shown that Conjecture 1 does not hold for many values of p on this counterexample graph. In practice, we have found that many random graphs fail the conjecture for several of their nodes. Unfortunately, this means that p -voltages cannot be used as a surrogate to efficiently obtain the comparison of p -resistances.

3.2 p -voltages Solutions are Well-Formed

While p -voltages are not a surrogate of p -resistances, this does not mean they lack desirable properties for classification. Perhaps surprisingly, prior work did not study the properties of p -voltage in its own right. Here we will prove two important properties of p -voltages which demonstrate that they could be potentially useful for semi-supervised learning over graphs. The first result shows that with the proper selection of parameter p , p -voltage solutions are not “spiky” like those which potentially arise from Laplacian Regularization (when $p = 2$). The second result shows that p -voltages interpolate smoothly over the graph and do not exhibit undesirable “flatness” behavior in the graph.

Our results apply to unweighted geometric graphs in \mathbb{R}^d , $d > 2$, which satisfy the general graph assumptions stated in Section 4 of [3]. These are mild assumptions which hold for many classes of graphs including appropriately-constructed knn - and ϵ -graphs. An important fact is that we assume there exist constants $0 < r \leq R$ such that every sample point x is directly connected to all sample points inside ball $B(x, r)$ and x is not directly connected to any sample points outside $B(x, R)$. There also must exist a function $\tau_n := \tau(n)$ such that all degrees in the graph are of order $\Theta(\tau_n)$ and both r and R are of order $\Theta((\tau_n/n)^{1/d})$, where n is the number of sample points.

3.2.1 p -Voltage Solutions Are Not Spiky

Zhu et al [1] developed the Laplacian regularization method which corresponds to using p -voltages with $p = 2$. Nadler et

al [2] identified a problem with this method. As the number of unlabeled points grows to infinity, there exist functions which minimize the energy equation and yet have constant value almost everywhere and narrow “spikes” at the source and sink points. These pathological functions are clearly not useful for predicting the labels in the graph and suggest that the $C_2(s, t)$ optimization problem is ill-posed. Here we show that p -voltages address this problem with appropriate choice of p . Another approach that addresses this deficiency is the Iterated Laplacians approach [5], to which we will compare in the experiments.

Consider a smooth density $p(x)$ over \mathbb{R}^d , $d > 2$. Define $\|\cdot\|$ as the Euclidean distance metric, and fix a source s and a sink t with $\|s - t\| \geq 1$. An example of a Nadler “spiky” function is:

$$y_\epsilon(x) = \min\left(\frac{\|x - t\|}{\epsilon}, 1\right) \quad (7)$$

As $n \rightarrow \infty$, $\epsilon \rightarrow 0$, the infinite limit of $C_2(y_\epsilon) \rightarrow 0$, so y_ϵ is a solution to C_2 . It is also clear that $y_\epsilon(t) = 0$ and $y_\epsilon(x) = 1$ almost everywhere else, so y_ϵ is not a reasonable estimate for a smooth function on $p(x)$ for the purpose of semi-supervised learning.

However, we can show that this type of function cannot be the solution to the p -voltages problem when $1 < p < p^* = \frac{d}{d-1}$. As in [3], we define the local neighborhood $\mathcal{N}(s)$ of vertex s as the ball with radius $C \cdot r$ around s , $C \geq R/r$. By the definition of R , $\mathcal{N}(s)$ must contain at least all vertices adjacent to s . Let $E_{\mathcal{N}(s)}$ be the set of edges which have both endpoints in $\mathcal{N}(s)$.

Consider a graph $G = (V, E)$ which satisfies the general assumptions, and $n \rightarrow \infty$. We next show that, if we solve the p -voltages problem $C_p(s, t)$ over the graph with $1 < p < \frac{d}{d-1}$, the maximum voltage drop in $\mathcal{N}(s)$ and $\mathcal{N}(t)$ will be negligible compared to the voltage drop across the entire graph. Thus, Nadler’s “spiky” functions cannot be solutions to the p -voltage problem.

Specifically, let us define the following quantities:

$$\begin{aligned} V_s &:= \max\{|v_s - v_u|, u \in \mathcal{N}(s)\} \\ V_t &:= \max\{|v_u - v_t|, u \in \mathcal{N}(t)\} \\ V &:= v_s - v_t = R_p(s, t). \end{aligned}$$

THEOREM 3.1. $\frac{V_s + V_t}{V} \rightarrow 0$ as $n \rightarrow \infty$.

PROOF. We can examine the contribution of the local neighborhoods of s and t to the overall p -resistance in the graph. Suppose we compute $R_p(s, t)$ from (2) on G and obtain the optimal unit flow $i^* = i_R$. By (4), we know that the voltages $\{v_R\}$ from the optimal solution to (2) have the property that for all edges $e = (a, b)$, $|v_a - v_b| := |\Delta v_e| = |i_e^*|^{p-1} r_e = |i_e^*|^{p-1}$ where the last step follows because the graph is unweighted. Consider a vertex $u \in \mathcal{N}(s)$. Denote by $\pi_{s,u}$ the shortest path from s to u which contains edges only in $E_{\mathcal{N}(s)}$, and the length of that path as $l_{s,u}$. Then

$$|v_s - v_u| \leq \sum_{e \in \pi_{s,u}} |\Delta v_e| = \sum_{e \in \pi_{s,u}} |i_e^*|^{p-1} \stackrel{(a)}{\leq} \sum_{e \in \pi_{s,u}} 1 = l_{s,u} \quad (8)$$

where the step labeled (a) holds because i is a unit flow and $p > 1$. By Proposition 9 of [3], for every vertex $u \in \mathcal{N}(s)$, the shortest path from s to u is smaller than $2C$. Thus

we can conclude from (8) that $|v_s - v_u| < 2C$, $\forall u \in \mathcal{N}(s)$. Identical reasoning applies to the sink t .

We know that $V_s < 2C$, $V_t < 2C$, and thus $V_s + V_t < 4C$. Consider the ratio $\frac{V_s + V_t}{V}$, which is the portion of the total graph-wide change in p -voltage which occurs in the local neighborhoods of s and t . By Theorem 5 of [3],

$$V = R_p(s, t) \geq \frac{1}{d_s^{p-1}} + \frac{1}{d_t^{p-1}} + T_1 \geq T_1 \quad (9)$$

where d_u is the degree of vertex u and

$$T_1 = \Theta\left(\frac{1}{n^{p(1-1/d)-1} \tau_n^{p(1+1/d)-1}}\right)$$

Now, if we used $p < p^*$ in (2) and define $c_2 := 1 - p(1 - 1/d) > 0$, we can bound the ratio in a manner similar to their Proof of Theorem 3:

$$\frac{V_s + V_t}{V} \leq \frac{4C}{T_1} \leq \frac{4C \cdot \tau_n^{2/(d-1)}}{n^{c_2}} \quad (10)$$

This ratio converges to 0 as $n \rightarrow \infty$ if τ is sub-polynomial in n , which is true under the general graph assumptions. Thus in the limit, the p -voltage solution has almost no voltage drop in the local neighborhoods of s and t , so arbitrarily thin spikes do not manifest themselves at the labeled points. We can immediately see that Nadler’s problematic y_ϵ function cannot be the solution since all of its change in voltage occurs in the local neighborhoods of s and t . \square

3.2.2 p -Voltage Solutions Are Not Flat

We now know that almost no p -voltage drop occurs in the immediate neighborhoods of s and t . However, we would also like to understand how p -voltages behave in the remainder of the graph. After all, there could still be large regions of constant voltage, which would still be detrimental for predicting labels on the graph. It turns out that we can show that p -voltages drop gradually over the graph and are not “flat” over large regions. First, we need to prove a lemma establishing a connection between change in p -voltage over a region and the effective p -resistance of that region.

LEMMA 3.2. Fix a source vertex s and sink vertex t in graph $G = (V, E)$. Let $M \in V$ be a region of G which has edge set E_M , and let ∂M be the set of vertices (“the boundary”) in M which are either s , t , or connected to a vertex not in M . Define $R_p^M(s, t)$ as the portion of the solution of $R_p(s, t)$ that passes through the edges in E_M . Then we know that

$$R_p^M(s, t) = \sum_{u \in \partial M} v_u i_u^* \quad (11)$$

where

$$i_u^* := \sum_{x \in M} i_{ux}^* \quad (12)$$

PROOF. As mentioned, i^* is the optimal flow solution resulting from $R_p(s, t)$. We can see that

$$\begin{aligned}
R_p^M(s, t) &:= \sum_{e \in E_M} r_e |i_e^*|^p \stackrel{(a)}{=} \sum_{e \in E_M} |i_e^*|^{p-1} |i_e^*| \\
&\stackrel{(b)}{=} \sum_{(a,b) \in E_M} |v_a - v_b| |i_{ab}^*| \\
&\stackrel{(c)}{=} \sum_{(a,b) \in E_M} (v_a - v_b) i_{ab}^*
\end{aligned}$$

where (a) holds because the graph is unweighted, (b) follows from p -Ohm's Law, and (c) holds because i_{ab}^* and $(v_a - v_b)$ must have the same sign by (4). Assume without loss of generality that there is some ordering over the vertices of M . Now, using a derivation in the spirit of [6], page 49,

$$\begin{aligned}
R_p^M(s, t) &= \sum_{a < b \in M} v_a i_{ab}^* - \sum_{a < b \in M} v_b i_{ab}^* \\
&= \sum_{u \in M} v_u \left(\sum_{x > u \in M} i_{ux}^* - \sum_{x < u \in M} i_{xu}^* \right) \\
&= \sum_{u \in M} v_u \left(\sum_{x > u \in M} i_{ux}^* + \sum_{x < u \in M} i_{ux}^* \right) \\
&= \sum_{u \in M} v_u \left(\sum_{x \in M} i_{ux}^* \right) \\
&= \sum_{u \in M} v_u i_u^*
\end{aligned}$$

We took advantage of that fact that in any flow, $i_{ab} = -i_{ba}$. By the flow definition of i^* , $i_u^* = 0$ for any vertex u on the interior of M , $M \setminus \partial M$. Therefore we can conclude that

$$R_p^M(s, t) = \sum_{u \in \partial M} v_u i_u^*$$

which is what was to be proved. \square

When M is the entire set of vertices in the graph, $\partial M = \{s, t\}$ and

$$R_p^M(s, t) = R_p(s, t) = v_s i_s + v_t i_t = (v_s - v_t) i_s \quad (13)$$

which coincides with Lemma 4 of [4]. It is clear why when we solve $R_p(s, t)$ with $i_s = 1$ and $v_t = 0$, we know that $v_s = R_p(s, t)$.

We select our region M according to the following procedure:

PROCEDURE 3.3. *Given a graph $G = (V, E)$ with fixed source vertex s and sink vertex t , solve the $R_p(s, t)$ problem over the graph to recover the optimal flow. Use p -Ohm's Law to calculate the p -voltage for each vertex. Sort the vertices in descending order of voltage: $V_{\text{sorted}} = \{v_s^{(1)}, \dots, v^{(i)}, \dots, v_t^{(n)}\}$. Ties can be broken in any order, as long as when $v^{(i)}$ is chosen, one of the vertices in $\{v^{(1)}, \dots, v^{(i-1)}\}$ is adjacent to it. As shown in the appendix in Lemma A.1, this is always possible. Define region M by picking two endpoints in V_{sorted} , exactly one of which is s or t . All vertices between the two endpoints in V_{sorted} , inclusive, are considered part of M . Denote by \bar{M} the remainder of the graph not in M .*

We can lower bound the maximum p -voltage drop across a region selected in this manner.

LEMMA 3.4. *Consider a region M selected according to Procedure 3.3. Define $V_M := \max_{a,b \in M} |v_a - v_b|$. Then*

$$V_M \geq R_p^M(s, t) \quad (14)$$

PROOF. Without loss of generality, let us say that t is part of M and that s is not. By Lemma A.1, M is contiguous: every vertex in \bar{M} has a path to s not passing through M , because M cannot "surround" vertices with voltage outside its range. ∂M consists of the vertices of M connected to \bar{M} . Consider the total flow between \bar{M} and M . We know that the total flow must equal 1, because M and \bar{M} form an s - t cut; that is, $\sum_{u \in \partial M} i_u = 1$. We also know that the flow on every edge from \bar{M} to M is non-negative because $\forall a \in \bar{M}, b \in M, v_a \geq v_b$ by the construction of M . Equation (11) gives us:

$$\begin{aligned}
R_p^M(s, t) &= v_t i_t + \sum_{u \in \partial M} v_u i_u \\
&= -v_t + \text{weighted-average}(v_u) \quad (15)
\end{aligned}$$

Given a set of real numbers $X = \{x_1, \dots, x_t\}$ and their weighted average A , it is self-evident that there must exist some element in X greater than or equal to A , and some element less than or equal to A . Thus, we can conclude from (15) that:

$$R_p^M(s, t) \leq \max_{u \in \partial M} |v_u - v_t| \leq \max_{a,b \in M} |v_a - v_b| =: V_M \quad (16)$$

That is, the p -voltage drop across a region M constructed in this manner is at least as large as its contribution to the p -resistance of the graph, $R_p^M(s, t)$. A symmetric argument can be made if M contains s instead of t . In that case,

$$R_p^M(s, t) \leq \max_{u \in \partial M} |v_s - v_u| \leq \max_{a,b \in M} |v_a - v_b| =: V_M \quad (17)$$

\square

Our next goal is to lower bound p -resistance in regions of the graph, so that we can lower bound the change in p -voltage over those regions. Specifically, we will improve the lower bound of Theorem 5 of [3]. We take advantage of Rayleigh's Monotonicity Principle, which states that the effective resistance can only decrease when we add edges of finite resistance to the graph. This was shown to hold for p -electric networks in Lemma 9 of [4], and can be easily adapted to our formulation of p -electric networks. In a similar approach to [3], we will "short" some vertices of the network together to form a new graph G' . We can then calculate the p -resistance of G' to establish a lower bound for G .

We construct G' as follows, in a similar manner to the graph setup of Proposition 8 of [3]. We add edges of zero resistance between all neighbors of s , merging them into one vertex. Then, expanding out from s , we merge all vertices together that fall into "rings" of width $r/2$. There are $1/r$ of these rings around s . We perform the exact same construction around t . Between the two sets of rings, we divide the

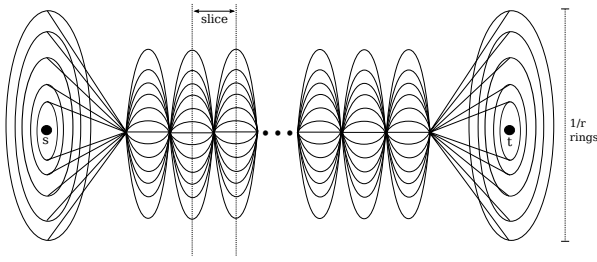


Figure 2: Construction of G'

graph into “slices” of width $3R$ which run perpendicular to the line connecting s and t , and merge all of the vertices in each slice together. For a visualization of G' , see Figure 2.

First, we can lower bound the p -resistance contribution of one of the ring regions, $R_p^{ring(s)}(s, t)$. The number of vertices in ring k , N_k , is proportional to n times the volume of the ring. As shown in [3], $N_k = \Theta(\tau_n k^{d-2})$. Therefore, between rings k and $k+1$, there are $\Theta(N_k \tau_n)$ edges in edge set E_k . Denote by $R_p^{E_k}(s, t)$ the contribution of the p -resistance flowing between the rings. By the generalized mean inequality for values p and 1, we know that:

$$\begin{aligned} \left(\frac{1}{N_k \tau_n} \sum_{e \in E_k} i_e^p \right)^{\frac{1}{p}} &\geq \frac{1}{N_k \tau_n} \sum_{e \in E_k} i_e \\ \left(\frac{1}{N_k \tau_n} R_p^{E_k}(s, t) \right)^{\frac{1}{p}} &\geq \frac{1}{N_k \tau_n} \sum_{e \in E_k} i_e \\ \left(\frac{1}{N_k \tau_n} R_p^{E_k}(s, t) \right)^{\frac{1}{p}} &\geq \frac{1}{N_k \tau_n} \\ R_p^{E_k}(s, t) &\geq (N_k \tau_n)^{1-p} \end{aligned}$$

where we used the fact that the flow between rings k and $k+1$ must be a unit flow because they form an s - t cut.

The p -resistance of all of the rings around s is simply the sum of the contributions of all the rings.

$$\begin{aligned} R_p^{ring(s)}(s, t) &= \sum_k R_p^{E_k}(s, t) \geq \sum_k (N_k \tau_n)^{1-p} \\ &= \Theta \left(\frac{1}{\tau_n^{2(p-1)}} \sum_{k=1}^{1/r} \frac{1}{k^{(d-2)(p-1)}} \right) =: T_2 \end{aligned} \quad (18)$$

The rings around t have the same lower bound by symmetry.

Next, we can lower bound the p -resistance contribution of one of the slices. Consider a single slice S . We say an edge e is in E_S if it has both endpoints in S . Denote by $|E_S|$ the number of edges in E_S . Then,

$$R_p^{E_S}(s, t) = \sum_{e \in E_S} |i_e|^p \stackrel{(a)}{\geq} \frac{1}{|E_S|^{p-1}} \left(\sum_{e \in E_S} |i_e| \right)^p \stackrel{(b)}{\geq} \frac{1}{|E_S|^{p-1}} \quad (19)$$

Step (a) is a result of the generalized mean inequality for values p and 1, and step (b) results because slice S is a cross-section of the whole graph G and so must have at least unit total flow traveling through it.

Define by N_S the number of vertices which fall at least R distance from the boundaries of S . Since S has width $3R$, these vertices occupy a region of width R . We can upper bound $|E_S|$ by N_S times the maximum degree of the graph.

$$\begin{aligned} |E_S| &\leq N_S \cdot d_{max} = \Theta(nR)\Theta(\tau_n) \\ &= \Theta \left(n^{\frac{d-1}{d}} \tau_n^{\frac{1}{d}} \right) \Theta(\tau_n) = \Theta \left(n^{\frac{d-1}{d}} \tau_n^{\frac{d+1}{d}} \right) \end{aligned}$$

From (19) we have that:

$$R_p^{E_S}(s, t) \geq \Theta \left(\frac{1}{\left(n^{\frac{d-1}{d}} \tau_n^{\frac{d+1}{d}} \right)^{p-1}} \right) \quad (20)$$

There are on the order of $K = \Theta(\frac{1}{3R}) = \Theta((n/\tau_n)^{1/d})$ of these slices in the graph. Suppose region M contains some number of slices such that it contains some fraction c of the total K slices. That is, it contains cK slices, with $0 < c \leq 1$. Using (20), we can see that:

$$R_p^M(s, t) \geq \sum_{S \in M} R_p^S(s, t) \geq \Theta \left(\frac{cK}{\left(n^{\frac{d-1}{d}} \tau_n^{\frac{d+1}{d}} \right)^{p-1}} \right) = cT_1 \quad (21)$$

For convenience, we define a property of regions in which we are interested.

DEFINITION 3.5. A **substantial region** is one that is constructed according to Procedure 3.3, and which contains all vertices in the rings around s or t as well as cK slices of the graph as defined above.

If we have a substantial region M , then we know by (18) and (21) that

$$R_p^M(s, t) \geq cT_1 + T_2 \quad (22)$$

We are ready to prove our main result.

THEOREM 3.6. Consider the ratio $\frac{V_M}{V}$, which is the portion of the total graph-wide change in p -voltage which occurs in region M . If M is a substantial region for constant $0 < c \leq 1$, and $p < \frac{d}{d-1}$, then we know that $\frac{V_M}{V} \geq c$.

PROOF. If we examine the reciprocal of the ratio, and use (14) and (22),

$$\frac{V}{V_M} \leq \frac{R_p(s, t)}{R_p^M(s, t)} \leq \frac{4C + T_1 + T_2}{cT_1 + T_2}$$

where the numerator comes from the bound in Theorem 5 in [3]. When we let the number of vertices $n \rightarrow \infty$,

$$\frac{V}{V_M} \leq \frac{4C}{cT_1 + T_2} + \frac{T_1 + T_2}{cT_1 + T_2} \rightarrow \frac{T_1 + T_2}{cT_1 + T_2} \leq \frac{1}{c}$$

since $T_1, T_2 > 0$. We can take the reciprocal to arrive at our result: $\frac{V_M}{V} \geq c$. \square

This result has significant implications. If we label the graph with p -voltages using a suitable value for p , any substantial region will have significant voltage drop across it.

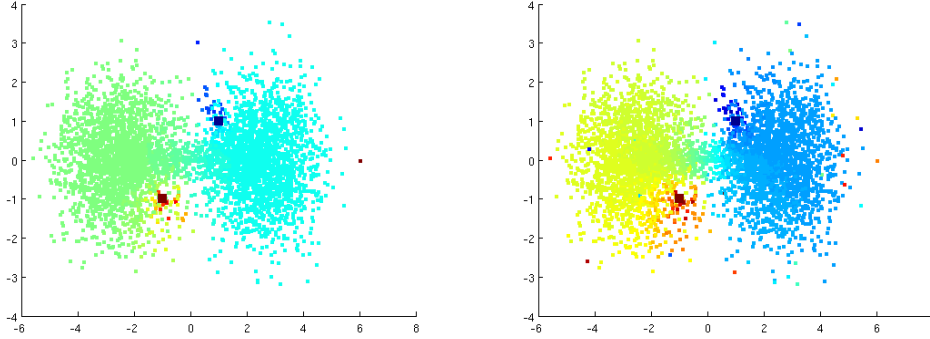


Figure 3: Barbell graph labeled with Laplacian Regularization (left, $p = 2$) and p -voltages (right, $p = 1.4$)

This means that there cannot be large areas of constant p -voltage in the graph, and that the voltage drop is distributed somewhat evenly over the graph. Thus p -voltages with a properly selected p do not suffer from the same limitations as Laplacian Regularization of [1].

4. EXPERIMENTS

We conducted several numerical experiments to test the performance of the p -voltages algorithm. We were interested in two questions: in practice, do p -voltage solutions exhibit the pathological behavior of Laplacian Regularization solutions? Additionally, how does the p -voltage algorithm perform on both real and synthetic datasets compared to other algorithms such as the Iterated Laplacians algorithm of [5]?

Our theoretical guarantees apply to geometric graphs such as knn - and ϵ -graphs. However, we found in practice that it was difficult to create “balanced” forms of these graphs on high-dimensional datasets. Symmetric knn -graphs resulted in points that had many more neighbors than others, while mutual knn -graphs were poorly connected, requiring random neighbors and these “hubs” to be added back in. Similar problems occurred for ϵ -graphs. These issues violate the local neighborhood assumptions of our theory. We found that complete graphs with edge weights from the RBF kernel (1) provided the best empirical performance. In one sense, the complete graph can be thought of as a “soft” version of a neighborhood graph where the great majority of a point’s edge weight is still connected to its local neighbors.

4.1 Toy Example

To examine the p -voltage solution, we generated a series of toy graphs and compared the results of the Laplacian Regularization and p -voltage algorithms. Here we share a sample “barbell” graph, which is meant to simulate a graph with defined cluster structure.

To generate the barbell in \mathbb{R}^3 , we sampled 2000 data points from each of two three-dimensional Gaussians, centered at $(-2.5, 0, 0)$ and $(2.5, 0, 0)$. The “handle” of the barbell consisted of another 500 points sampled from a narrow cylinder spanning the means of the Gaussians. A single source point was placed in one end of the barbell at $(-1, -1, 0)$, and a single sink point was placed in the other end at $(1, 1, 0)$. A complete graph was constructed over the data points using the RBF kernel (1) with bandwidth

$\sigma = 0.1386$.

We ran both Laplacian Regularization and p -voltages with source label $v_s = 1$ and sink label $v_t = 0$. Since p must be set to less than $\frac{d}{d-1} = \frac{3}{3-1} = 1.5$, we chose $p = 1.4$. The two solutions are shown in Figure 3. The points are shown projected onto the x - y plane, and their predicted labels are indicated by their color: red shows labels close to 1, and blue shows labels close to 0.

Note several aspects of these examples. While the Laplacian Regularization solution does seem to label based on the cluster structure of the graph, the range of labels in the majority of the graph is extremely small. This is the “flatness” problem that we discussed. Also, the appropriate threshold between the two classes is not at 0.5 despite the equally-sized clusters; if we thresholded at 0.5, all but a few points around the source point would be labeled as negative, resulting in poor classification performance. On the other hand, the p -voltage solution appears to be much more desirable. There is a more gradual and constant drop in label throughout the graph, and the threshold between the clusters is much closer to 0.5 (between yellow and green). On similar graphs with more dimensions or more data points, the difference between the algorithms is even more pronounced. Thus, on these examples, our empirical results agree with our theoretical ones and suggest that p -voltages do not suffer from the same limitations as Laplacian Regularization.

4.2 Benchmark Datasets

We tested the p -voltages algorithm on several synthetic and real world datasets from the binary classification benchmark from Chapter 21 of [7]. Each of the datasets contains 1500 points and is provided with twelve different random training-test set splits (100 labeled training set points and 1400 test set points). The exception is the BCI dataset, which has 400 data points and 100 and 300 training and test set sizes, respectively. We constructed a complete graph over the points and used the RBF kernel (1) with the bandwidth σ fixed according to the heuristic described in [7]: $\sigma = \frac{d}{3}$, where d is the average distance between a point and its tenth nearest neighbor.

Our experiments compare the performance of four algorithms. The first is standard Laplacian Regularization of [1]. The second is a version of Laplacian Regularization which uses class mass normalization (CMN). CMN is a heuristic proposed in [1] which attempts to shift the classification

Dataset	LapReg	LapReg + CMN	IterLap	p-voltages
g241c	48.95 ± 4.38	22.46 ± 1.42	19.40 ± 4.88	33.74 ± 7.20
g241n	49.93 ± 1.20	30.88 ± 3.43	13.15 ± 0.97	29.91 ± 3.67
Digit1	8.81 ± 0.56	3.74 ± 1.05	2.24 ± 0.81	3.14 ± 0.95
USPS	19.19 ± 0.67	10.91 ± 1.06	4.58 ± 0.86	7.54 ± 1.98
BCI	46.89 ± 2.33	46.19 ± 2.14	45.67 ± 2.75	45.03 ± 2.78

Table 1: Average Test Set Classification Error Percentage and Standard Error on Chapelle Benchmark

threshold to an appropriate value by using the size of the labeled classes as a prior. In practice this significantly improves the performance of Laplacian Regularization. Third, we test Iterated Laplacians of [5], which seeks to fix Laplacian Regularization using an exponentiated Laplacian to provide a higher order norm. Finally, we test the p -voltages algorithm.

To determine the other parameters of the algorithms, we performed a grid search and chose the best setting by validation on the first training set. A random tuning set of size 10 was withheld several times from the training set and the parameters which provided the lowest average classification error were selected. The p parameter for p -voltages was selected among $\{\frac{128}{127}, \frac{64}{63}, \frac{32}{31}, \frac{16}{15}, \frac{8}{7}, \frac{6}{5}, \frac{4}{3}, \frac{3}{2}\}$. The m parameter of Iterated Laplacians was chosen among $\{2, 3, 4, 6, 8, 16, 32, 64, 128\}$, the μ parameter among $\{10^{-2}, 10^{-4}, 10^{-6}\}$ and the Laplacian type among the four types listed in [5].

Once the final parameters were set, we ran the algorithms on each of the twelve splits using the training sets as the labeled points and measuring classification error on the test sets. The mean classification error rate and standard error over the twelve splits for each algorithm and dataset is shown in Table 1.

There is a mixture of positive and negative conclusions from these results. On one hand, the p -voltages algorithm outperformed Laplacian Regularization significantly on every dataset. On all but one of the datasets, p -voltages also outperformed the version with class mass normalization, which suggests that not only does the p -voltage solution have “better shape”, it also provides a better ranking of the data points. However, the current state-of-the-art Iterated Laplacians algorithm performed better on all but one of the datasets. From a practical standpoint, one advantage of the p -voltages algorithm is that it has only a single parameter p to set as compared to Iterated Laplacians, which has three (m , μ , and L). This made the parameter search significantly more difficult for Iterated Laplacians. It also may have contributed to the success of Iterated Laplacians since the algorithm had more degrees of freedom to tune.

5. CONCLUSION

We have investigated many of the properties as well as the performance of the p -voltages algorithm for semi-supervised learning. While we proved that p -voltage solutions are not equivalent to p -resistance solutions, we also showed that p -voltage solutions are potentially valuable as they do not suffer from the same limitations as Laplacian Regularization. Numerical experiments confirmed these results, though empirically p -voltage did not outperform Iterated Laplacian.

Acknowledgments

This research is supported in part by National Science Foundation grant IIS-0916038.

6. REFERENCES

- [1] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. In *International Conference on Machine Learning (ICML)*, 2003.
- [2] B. Nadler, N. Srebro, and X. Zhou. Semi-Supervised Learning with the Graph-Laplacian: The limit of infinite unlabelled data. In *Neural Information Processing Systems (NIPS)*, 2009.
- [3] M. Alamyir and U. von Luxburg. Phase transition in the family of p -resistances. In *Neural Information Processing Systems (NIPS)*, 2011.
- [4] M. Herbster and G. Lever. Predicting the Labelling of a Graph via Minimum p -Seminorm Interpolation. In *Conference on Learning Theory (COLT)*, 2009.
- [5] X. Zhou and M. Belkin. Semi-supervised Learning by Higher Order Regularization. In *Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.
- [6] P. Doyle and J. L. Snell. *Random walks and electric networks*. 2000.
- [7] O. Chapelle, B. Schölkopf, and A. Zien, editors. *Semi-Supervised Learning*. MIT Press, Cambridge, MA, 2006. <http://www.kyb.tuebingen.mpg.de/ss1-book>.

APPENDIX

A. HELPFUL P-VOLTAGE RESULTS

To prove that ties can always be resolved in Procedure 3.3, we can use the following lemma:

LEMMA A.1. *No vertex besides the source vertex s can have a p -voltage greater than all of its neighbors. Likewise, no vertex besides the sink t can have a p -voltage less than all of its neighbors.*

PROOF. We will prove the first statement by contradiction, and the second will follow by similar logic. Assume there is some non-source node u which has a p -voltage greater than all of its neighbors. By definition, the set of p -voltages minimizes equation (3) on the graph. However, if we lower v_u so that it is equal to the maximum p -voltage of its neighbors, then all terms in (3) involving edges connected to u will decrease, and no other terms will be affected. Thus, this new set of voltages leads to a lower value of $C_p(s, t)$, and the original set of voltages could not be the set of p -voltages. \square

Note that this statement holds for the voltages calculated from $R_p(s, t)$ as well because the two sets of voltages are proportional to each other. Thus, according to Lemma A.1, every $u \neq s, t$ is connected to some vertex with greater or equal p -voltage, and some vertex with less or equal p -voltage. As a result, ties can always be resolved when constructing V_{sorted} .