Consistency, Robustness and Sparsity for Learning Algorithms

Konsistenz, Robustheit und Dünnbesetztheit von Lern-Algorithmen

zur Erlangung des Doktorgrades

Dr. rer. nat.

im Studiengang Mathematik

am Department Mathematik der
Friedrich-Alexander-Universität Erlangen-Nürnberg

vorgelegt am **14.Juni 2023**

von **Tim Roith**

Prüfer: Martin Burger
Betreuer: M.Sc. C

Für eine besondere Person,
ohne sie hätte ich es nie geschafft.

Acknowledgement

I would like to thank my supervisors for helping me finish this thesis and supporting me with their profound knowledge..

Contents

List of Figures

Preface

This work is structured into two main parts, Part I the presentation and explanation of the topics and results presented in Part II, the peer-reviewed articles.

Part I: Exposition	Part II: Prints
Chapter 1: Learning Paradigms	
Chapter 2: Consistent Semi-Supervised Learning on Sparse Graphs	????
Chapter 3: Supervised Learning	????

Part I consists of three chapters, of which the first explains the paradigms, *unsupervised*, *semi-supervised* and *supervised* learning. The other chapters are the split up thematically, concerning the topics semi-supervised and supervised learning respectively. In each of these chapters a short introduction provides the necessary framework allowing us to explain the main contributions. The following publications are reprinted in Part II:

[LIP-I] T. Roith and L. Bungert. "Continuum limit of Lipschitz learning on graphs." In: *Foundations of Computational Mathematics* (2022), pp. 1–39.

[LIP-II] L. Bungert, J. Calder, and T. Roith. "Uniform convergence rates for Lipschitz learning on graphs." In: *IMA Journal of Numerical Analysis* (Sept. 2022). DOI: 10.1093/imanum/drac048.

[CLIP] L. Bungert et al. "CLIP: Cheap Lipschitz training of neural networks." In: *Scale Space and Variational Methods in Computer Vision: 8th International Conference, SSVM 2021, Virtual Event, May 16–20, 2021, Proceedings*. Springer. 2021, pp. 307–319.

[BREG-I] L. Bungert et al. "A bregman learning framework for sparse neural networks." In: *Journal of Machine Learning Research* 23.192 (2022), pp. 1–43.

[FNO] S. Kabri et al. "Resolution-Invariant Image Classification based on Fourier Neural Operators." In: *Scale Space and Variational Methods in Computer Vision: 9th International Conference, SSVM 2023, Proceedings*. Springer. 2023, pp. 307–319.

The following two works that are not part of this thesis but provide an additional insight.

[LIP-III] L. Bungert, J. Calder, and T. Roith. *Ratio convergence rates for Euclidean first-passage percolation: Applications to the graph infinity Laplacian.* 2022. arXiv: 2210.09023 [math.PR].

[BREG-II] L. Bungert et al. "Neural Architecture Search via Bregman Iterations." In: (2021). arXiv: 2106.02479 [cs.LG].

TR's Contribution

Here we list TR's contribution to the publications included in the thesis.

[LIP-I]: This work builds upon the findings in TR's master thesis [Roi22]. It is however important to note that the results constitute a significant extension and are conceptually stronger than the ones in [Roi22], see **??**. TR adapted the continuum limit framework to the L^∞ case, worked out most of the proofs and wrote a significant part of the paper. In collaboration with LB, he identified the crucial domain assumptions that allow to work on non-convex domains and proved convergence for approximate boundary conditions.

[LIP-II]: In collaboration with LB, TR worked on the convergence proofs building upon the ideas of JC. He contributed to both the numeric and the analysis conducted in the paper.

[CLIP]: TR worked out the main algorithm proposed in the paper together with LB, based on his idea. Together with LS and RR he conducted the numerical examples and also wrote most of the source code. Furthermore, he wrote large parts of the paper.

[BREG-I]: TR expanded LB's ideas of employing Bregman iteration for sparse training. Together with MB and LB he worked out the convergence analysis of stochastic Bregman iterations. Here, he also proposed a profound sparse initialization strategy. Furthermore, he conducted the numerical examples and wrote most of the source code.

[FNO]: This work is based on SK's masters thesis, employing the initial ideas of MB for resolution invariance with FNOs. In the paper TR worked out the proofs for well-definedness and Fréchet-differentiability, together with SK. He wrote large parts of the paper and the source code. Here, he conducted the numerical studies in collaboration with SK.

Part I.

Exposition

Chapter 1

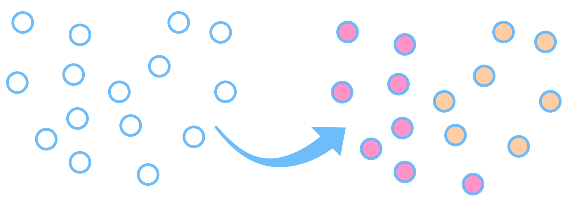
Learning Paradigms

Throughout this thesis, we assume to be given data $\Omega_n \subset \Omega \subset \mathbb{R}^d$ consisting of n data points. We consider task of *learning* a function $u : \tilde{\Omega} \to \Upsilon$ from the given data, where the two most important cases for us are

- **classification**: u assigns a label to each $x \in D$ out of a total of $C \in \mathbb{N}$ possible classes, i.e. $\Upsilon = \{1, \dots, C\}$.
- **image denoising**: u outputs a denoised version of an input image x, i.e. here $\tilde{\Omega} = \Upsilon$.

The set $\tilde{\Omega} \subset \mathbb{R}^d$ is usually either the set of data points Ω_n or the whole space Ω. The learning paradigms we consider in this thesis, differ by their use of labeled data. We review the concepts in the following.

1.1. Unsupervised Learning



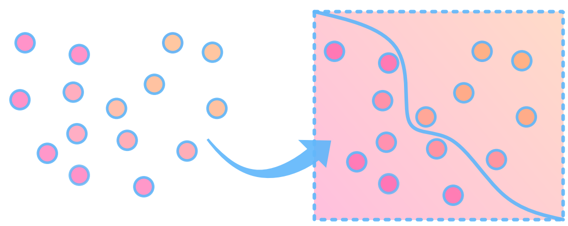
In this case we are not given any labeled data. In our context the most important application is data clustering. Other tasks involve dimensionality reduction or density estimation, see [ST14]. The clustering task consists of grouping data based on some similarity criterion. In this sense, clustering can also be interpreted as classification, i.e., the desired function is a mapping $u : \tilde{\Omega} \to \{1, \dots, C\}$ where $C \in \mathbb{N}$ denotes the number of clusters. Typically, one wants to obtain a clustering of the given data set, i.e., $\otimes = \Omega$. We list some of the typical clustering methods below.

- K-means algorithm, [Ste+56],

- Expectation Maximization, [DLR77],
- Cheeger cuts, [GS15; SB09; Gar+16; GMT22],
- spectral clustering, [GS18].

Unsupervised learning is not the main focus of this present work. However, we note that especially the concepts developed in [GS15] for Cheeger cuts are crucial for the continuum limit framework in Section 2.4.

1.2. Supervised Learning

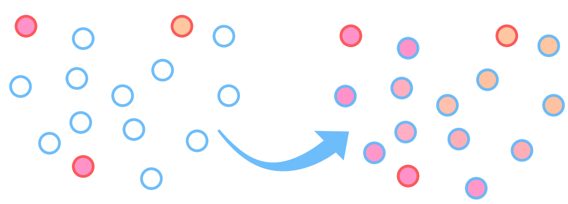


In this setting, each data point $x \in \Omega_n$ is labeled, via a given function $g : \Omega_n \to \Upsilon$ such that we have a finite training set $\mathcal{T} = \{(x, g(x)) : x \in \Omega_n\}$. The task is then to infer a function defined on the underlying space, i.e. $u : \Omega \to \Upsilon$, i.e. we want to assign a label to unseen $x \in \Omega$ that are not necessarily part of the given data. In order to *learn* the function u from the given data, one needs to choose a parameterized class of functions $\mathcal{U}$, where typically each element can be describe by a finite number of parameters. Among others, common methods or parametrizations include

- Support vector machines [CV95; SS05],
- decision Trees [MS63; Bre+84],
- neural networks [Tur04; Ros58; MP69].

In **??** we exclusively focus on supervised learning algorithms employing neural networks. We refer to [Sch15] for an exhaustive historical overview. The concrete setting and learning framework is given in **??**.

1.3. Semi-Supervised Learning



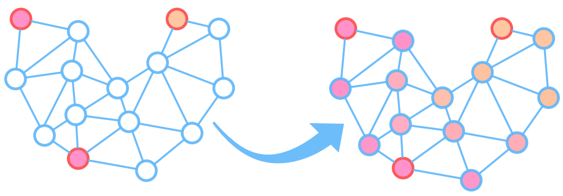
In the semi-supervised setting we assume that only a fraction of the data Ω_n is labeled, i.e., we are given a a function $g : \mathcal{O}_n \to \Upsilon$ where $\mathcal{O}_n \subset \Omega_n$ is the set of labeled data. Typically the labeled data constitutes only a small fraction of all available points, i.e. $|\mathcal{O}_n| << |\Omega_n|$. In this thesis we restrict ourselves to the *transductive setting*, i.e. we ant to infer a function acting only on the data $u : \Omega_n \to \Upsilon$. This is opposed to the inductive setting, where u also classifies unseen points $x \in \Omega$, [Zhu05]. Common algorithms and methods include

- expectation maximization and mixture models [**??**],
- self-training and co-training [**??**],
- graph-based learning [Zhu05].

In Chapter 2 we focus on graph-based learning algorithms, however we refer to [Zhu05] for a wonderful overview of semi-supervised learning algorithms.

Chapter 2

Consistent Semi-Supervised Learning on Sparse Graphs



This chapter considers the consistency of graph-based semi-supervised learning methods. We contextualize the topics provided in the prints [LIP-I; LIP-II] together with additional input from the pre-print [LIP-III]. In Section 2.1 we introduce the concrete setting and task. In Section 2.2 we consider the p Laplacian both in the continuum and the graph setting. This is the cornerstone for topics connected to the ∞ Laplacian and Lipschitz extensions in Section 2.3, the main points of interest in this chapter. We then continue...

the field [LIP-I; LIP-II; LIP-III] contribute to. We then

Section 2.1: Graph-based SSL and Consistency

Section 2.2: The p Laplacian

Section 2.2.1: Continuum Setting

Section 2.2.2: Graph Setting

Section 2.3 Lipschitz Extensions

Section 2.3.1: Continuum Setting

Section 2.3.2: Graph Setting

Section 2.4: Γ-Convergence [LIP-I]

Section 2.5: Uniform Convergence and Convergence of Distance Functions [LIP-II]

2.1. Graph-based SSL and Consistency

p-Lplacian bad for p too small: Numerics and theory

Regularity results p-Laplace

In the semi-supervised learning setting of Section 1.3, we are given a finite set $\Omega_n \subset \tilde{\Omega}$ consisting of n points, where $\tilde{\Omega}$ denotes the input space. We assume that a non-empty subset $\mathcal{O}_n \subset \Omega_n$ is labeled, i.e., we are given a labeling function $g : \mathcal{O}_n \to \mathbb{R}$. The semi-supervised learning problem consists in finding an extension of g to the whole set Ω_n, i.e.

$$\begin{gathered}\text{find } u : \Omega_n \to \mathbb{R},\\ \text{such that } u(x) = g(x) \text{ for all } x \in \mathcal{O}_n.\end{gathered} \tag{SSL}$$

In order to obtain meaningful solutions, one usually incorporates the *smoothness assumption* [**<empty citation>**] which can be informally stated as follows:

"*Points that are close together are more likely to share a similar label.*"

Using the previous insights, we now develop the framework for semi-supervised learning on graphs. The key starting point is to introduce weighted graphs, that allow us to measure the closeness of points. This concept yields a practical way to enforce the smoothness assumption **??**.

Definition 2.1 (Weighted Graphs). For a finite set V and a weight function $w : V \times V \to \mathbb{R}$, the tuple (V, w) is called a *weighted graph.*

Remark 2.2. Typically, a graph is defined as a pair (V, E) where V is a finite set and E denotes the set of edges. Here, one has two cases:
Undirected graph: $E = \{\{x, y\} : \text{there is an edge between } x \in V \text{ and } y \in V\}$, i.e., $E \subset 2^V$ and each edge is undirected,

Directed graph: $E = \{(x, y) : \text{there is an edge from } x \text{ to } y\}$, i.e., $E \subset V \times V$ and each edge is directed. Additionally, one then considers a weight function $W : E \to \mathbb{R}$ assigning a weight to each edge, and then defines the triple (V, E, W) as a weighted graph. However, we note that this information can be represented by a weight function $w : V \times V \to \mathbb{R}$. A directed edge set $E \subset V \times V$ can be equivalently expressed by a weight function $w : V \times V \to \mathbb{R}$ where $w(x, y) > 0$ if and only if $(x, y) \in E$, a weighting $W : E \to \mathbb{R}$ naturally transfers to w. In the case of an undirected graph, one simply requires the weight function to be symmetric, i.e., $w(x, y) = w(y, x)$ for all $x, y \in V$. Furthermore, in the above definition the set V does not enter the definition other than there as a bijection $V \to \{1, \dots, n\}$ for some $n \in \mathbb{N}$. Therefore, a graph could be entirely represented by a weight function $w : \{1, \dots, n\} \times \{1, \dots, n\} \to \mathbb{R}$. However, since the definition of w will incorporate information about points $\Omega_n \subset \mathbb{R}^d$ we use the notation (Ω_n, w_n) for graphs in the following. △

In the following we consider functions defined on graphs $\mathbf{u} : \Omega_n \to \mathbb{R}$ for which we employ bold symbols to distinguish them from functions in the continuum $u : \Omega \to \mathbb{R}$. In particular, the space of graph functions on Ω_n can simply be identified with $\mathbb{R}^n$, since

$$\{\mathbf{u} : \Omega \to \mathbb{R}\} \sim \mathbb{R}^n.$$

2.2. The p-Laplacian: continuum and graph

In **??** we explore concepts that borrow ideas from the theory of partial differential equations, and in particular the p-Laplace equation. In this section we briefly review important ideas and results and provide necessary preliminaries for the following sections.

2.2.1. The p-Laplacian: continuum setting

We follow the exposition in [Lin17]. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, then we consider the p-Dirichlet energy for functions $u \in W^{1,p}(\Omega)$,

$$\mathcal{E}_p(u) := \int_\Omega |\Delta_p u|^p \, dx, \tag{2.1}$$

and the associated variational problem.

Problem 2.3 (Variational Formulation). For $p \in (1, \infty)$ and $V \subset W^{1,p}(\Omega)$ find $u \in W^{1,p}(\Omega)$ such that

$$\mathcal{E}_p(u) \leq \mathcal{E}_p(v)$$

for all v, such that $(u - v) \in W^{1,p}_0$.

If $u \in V$ is a minimizer of the above problem, then its first variation must vanish, i.e., for all $\phi \in C_0^\infty(\Omega)$ one has

$$\int_\Omega \langle |\Delta_p u|^p \, \nabla u, \nabla \phi \rangle \, dx = 0. \tag{2.2}$$

A function $u \in V$ satisfying Eq. (2.2) is called a *weak solution* of the p-Laplace equation. In fact, if u is smooth enough one has that

$$\Delta_p u := \operatorname{div}(|\Delta_p u|^{p-2} \nabla u) = 0 \tag{2.3}$$

where Δ_p is called the p-Laplacian. For most of our applications we want to prescribe boundary conditions on $\partial\Omega$. For a given function $g \in W^{1,p}(\Omega)$ we therefore consider the set $V_g := \{u \in W^{1,p}(\Omega) : u - g \in W_0^{1,p}(\Omega)\}$ for which we have the following result.

Theorem 2.4 (Existence and Uniqueness). For $p \in (1, \infty)$ and $g \in W^{1,p}(\Omega)$ there exists a unique minimizer $u \in V_g$ of the p-Dirichlet energy, i.e.,

$$\operatorname{argmin}_{u \in V_g} \mathcal{E}_p(u) = u.$$

Moreover, u is a weak solution of the p-Laplace equation and there exists a function $\tilde{u} \in C(\Omega)$ such that $u = \tilde{u}$ a.e. in Ω. If $g \in C(\Omega)$ and Ω is sufficiently smooth, then $\tilde{u}|_{\partial\Omega} = g|_{\partial\Omega}$.

Proof. The proof can be found in [Lin17, Thm. 2.16]. □

Local minimization property

2.2.2. Laplacian Learning

While there are various techniques to solve the semi-supervised learning problem (SSL) [**<empty citation>**], in this work we focus on so-called *Laplacian learning* [**belkin2006laplacian**]. This method had one of its first appearances in [ZGL03], where the associated problem was given as

$$\begin{aligned} &\min_{\mathbf{u}:\Omega_n \to \mathbb{R}} \sum_{x,y \in \Omega_n} w_n(x,y)^2 \left(\mathbf{u}(y) - \mathbf{u}(x)\right)^2 \\ &\text{subject to } \mathbf{u}(x) = \mathbf{g}(x) \text{ for all } x \in \mathcal{O}_n. \end{aligned} \tag{2.4}$$

A natural extension of this problem is obtained by substituting the target functional by

$$\mathbf{E}_p^{w_n}(\mathbf{u}) := \sum_{x,y \in \Omega_n} w_n(x,y)^p \left|\mathbf{u}(y) - \mathbf{u}(x)\right|^p,$$

which we refer to as the graph p-Dirichlet energy. Indeed, we notice structural similarities to the p-Dirichlet energy $\mathcal{E}_p$ in Eq. (2.1), replacing the integral by a finite sum and derivatives by weighted finite differences

$$w_n(x,y)^p \left|\mathbf{u}(y) - \mathbf{u}(x)\right|^p.$$

This naturally leads to the following minimization problem.

Problem 2.5 (Graph Energy Minimization). Given a weighted graph (Ω_n, w_n) and a labeling function $\mathbf{g} : \mathcal{O}_n \to \mathbb{R}$, for $\mathcal{O}_n \subset \Omega_n$ we consider the problem

$$\min_{\mathbf{u}:\Omega_n \to \mathbb{R}} J^{w_n,p}(\mathbf{u}) \text{ subject to } \mathbf{u}(x) = \mathbf{g}(x) \text{ for all } x \in \mathcal{O}_n.$$

Since every function $\mathbf{u} : \Omega_n \to \mathbb{R}$ can be identified with a vector $\mathbf{u} \in \mathbb{R}^n$, the above problem is in fact an optimization problem in $\mathbb{R}^n$. Therefore one can prove unique existence of solutions via standard methods.

Theorem 2.6 (Existence and Uniqueness). Problem 2.5 admits a unique solution $\mathbf{u} : \Omega_n \to \mathbb{R}$.

Proof. This and that, a gecko with a hat. □

While Laplacian learning

Graph Laplacian In the continuum case one considers the Euler–Lagrange equation for the functional $\mathcal{E}_p$, which yields the p-Laplacian, see Section 2.2. Analogously, the optimality conditions for the graph p-Dirichlet energy $\mathbf{E}_p^{w_n}$ yield

$$\Delta_p^{w_n} \mathbf{u}(x) := \sum_{y \in \Omega_n} w_n(x,y)^p \left|\mathbf{u}(y) - \mathbf{u}(x)\right|^{p-2} (\mathbf{u}(y) - \mathbf{u}(x)) = 0, \text{ for all } x \in \Omega_n,$$

where $\Delta_p^{w_n}$ is referred to as thegraph p-Laplacian operator. This yields the graph p-Laplacian problem.

Problem 2.7 (Graph p-Laplacian). Given a weighted graph (Ω_n, w_n) and a labeling function $\mathbf{g} : \mathcal{O}_n \to \mathbb{R}$ with $\mathcal{O}_n \subset \Omega_n$, find a function $\mathbf{u} : \Omega_n \to \mathbb{R}$ such that

$$\begin{aligned} \Delta_p^{w_n} \mathbf{u} &= 0, \text{ in } \Omega_n \setminus \mathcal{O}_n, \\ \mathbf{u} &= \mathbf{g} \text{ on } \mathcal{O}_n. \end{aligned}$$

Since the functional $\mathbf{E}_p^{w_n}$ has a unique minimizer subject to the constraints given by $\mathbf{g}$ and the graph p-Laplacian is derived via optimality conditions, one expects that the Problem 2.5 and Problem 2.7 are equivalent. This is formulated in the following theorem.

Theorem 2.8 (Existence and Uniqueness). There exists a unique solution $\mathbf{u} : \Omega_n \to \mathbb{R}$ to Problem 2.7, which also is the unique minimizer of Problem 2.5.

Proof. This and that a gecko with a hat. □

2.3. Lipschitz extensions and the infinity Laplacian: continuum and graph

2.3.1. The continuum setting

This chapter studies the limit $p \to \infty$ of the p-Laplace equation. We first recall, that for $u \in W^{1,\infty}(\Omega)$ we have that

$$\lim_{p\to\infty} \mathcal{E}_p(u)^{1/p} = \operatorname*{ess\,sup}_{x\in\Omega} |\nabla u(x)| =: \mathcal{E}_\infty(u),$$

see [Jen93]. The functional $\mathcal{E}_\infty$ is weak*-lower semicontinuous over $W^{1,\infty}(\Omega)$, (see e.g. [BJW01, Thm. 2.6]). In the classical theory developed by Jensen in [Jen93] one considers the following problem, which is tries to "minimize the *sup-norm* of the gradient" as described by Jensen.

Problem 2.9 (Variational gradient-sup problem). For an open domain $\Omega \subset \mathbb{R}^d$ find a function $u \in W^{1,\infty}(\Omega)$ such that

$$\|\nabla u\|_\infty \leq \|\nabla v\|_\infty \text{ for every } v, \text{s.t. } (u - v) \in W_0^{1,\infty}.$$

The above problem becomes more relevant when we additionally impose boundary values $g : \partial\Omega \to \mathbb{R}$. Here, [Jen93] draws the connection to so-called *Lipschitz extensions*, which are the driving concept in this section. We introduce a more general viewpoint—that does not require the notion of a gradient—later on, but first introduce the variational problem.

The intrinsic metric and the Lipschitz constant. As noticed in [Jen93] working with the Lipschitz constant and the sup-norm of the gradient requires a careful treatment of the distance measurement. Let $\tilde\Omega$ be a set and let d be a semi-metric on $\tilde\Omega$, that is d fulfills the requirement of a metric up to triangle inequality. Then we define the Lipschitz constant of a function $u : V \to \mathbb{R}$ on a subset $V \subset \tilde\Omega$ as

$$\operatorname{Lip}_d(u; V) := \sup_{x,y\in V, x\neq y} \frac{|u(x) - u(y)|}{d(x, y)}.$$

If d denotes the Euclidean distance we use omit the subscript. i.e. $\operatorname{Lip}_d = \operatorname{Lip}$. Additionally, we can introduce the space of Lipschitz functions $\operatorname{Lip}_d(V)$ on V via $u \in \operatorname{Lip}_d(V) \Leftrightarrow \operatorname{Lip}_d(u; V) < \infty$.

Remark 2.10 (Lipschitz and Sobolev functions). If $\Omega \subset \mathbb{R}^d$ is sufficiently regular, e.g., it has Lipschitz boundary then we have that

$$\operatorname{Lip}(\Omega) = W^{1,\infty}(\Omega),$$

where this identity is of course to be understood in the sense of equivalence classes in $Ł^p$ spaces. We refer to [**evansgariepy**] for a proof of this result. △

The above remark already relates Lipschitz with $W^{1,\infty}$ functions. Often however, we need a quantitative comparison between the Lipschitz constant and the sup-norm of the gradient of a function. Here, it is essential which distance measure is chosen for the Lipschitz constant. For an open domain $\Omega \subset \mathbb{R}^d$ we have the inequality

$$\|\nabla u\|_\infty \leq \sup_{x\neq y} \frac{|u(x)-u(y)|}{|x-y|}$$

which can be proven via the definition of the gradient. For the reverse inquality, one has to respect the geometry of the domain, namely for $x, y \in \Omega$ we have that

$$|u(x)-u(y)| \leq \|\nabla u\|_\infty \, d_\Omega(x,y) \tag{2.5}$$

see, e.g., [BB11, Prop9.3, Rem. 7], where

$$d_\Omega(x,y) = \inf\left\{\int_0^1 |\dot\gamma(t)|\,dt : \gamma \in C^1([0,1],\Omega) \text{ with } \gamma(0)=x, \gamma(1)=y\right\}$$

denotes the *geodesic distance* on Ω. If Ω is convex, we have that $d_\Omega(x,y) = |x-y|$ for every $x, y \in \Omega$ and therefore Eq. (2.5) yields $\mathrm{Lip}(u) = \|\nabla u\|_\infty$. However, this situation changes for non-convex domains, see Example 2.11. Additionally it is often necessary to define a distance measure on the closure of $\Omega \subset \mathbb{R}^d$. In order to have a geodesic on $\overline{\Omega}$ one can simply consider $d_{\overline{\Omega}}$, see e.g. [**unif**], which then yields the length space $(\overline{\Omega}, d_{\overline{\Omega}})$. In the classical theory developed in [Jen93] one alternatively considers

$$\tilde d_{\overline{\Omega}}(x,y) := \liminf_{(\tilde x,\tilde y)\to(x,y)} d_\Omega(\tilde x,\tilde y).$$

The differences between these notation are demonstrated in the following example. Also note, that $\tilde d_{\overline{\Omega}}$ is only a semi-metric on $\overline{\Omega}$ since it is lacking a triangle inequality.

Example 2.11. For $I = [-\pi, c] \cup [c, \pi]$ with $c = \pi/6$ we consider the domain

$$\bigcup_{\theta\in I} B_1\left((\cos(\theta), \sin(\theta)\right)$$

which is visualized in Fig. 2.1 and the points $x = (2\,c, 1), y = (2\,c, -1)$. The line segment between x and y contains the point $z = (2\,c, 0)$, however $z \notin \Omega$. One can show that the geodesic has the length $d_\Omega(x,y) = 4\,\cos(\pi/6) + \pi \approx 6.606$ which is the length of the dotted path in Fig. 2.1. However, we observe that

$$\overline{\Omega} = \bigcup_{\theta\in I} \overline{B_1\left((\cos(\theta), \sin(\theta)\right)}$$

and in particular $z \in \overline{B_1\left((c,c)\right)}$, therefore $d_{\overline{\Omega}}(x,y) = 2$.

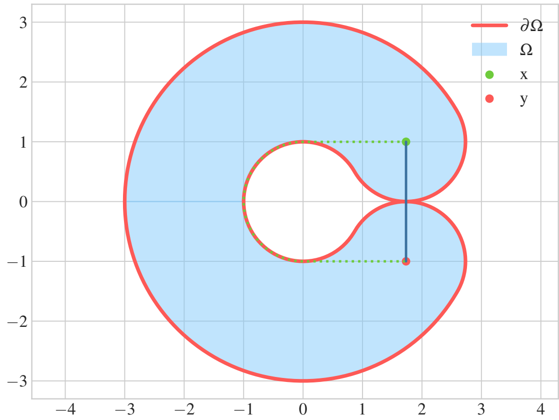


Figure 2.1.: The domain in Example 2.11.

Solutions to the gradient-sup problem Before generalizing the theory of Lipschitz extension to arbitrary metric spaces, we first note, that one can explicitly construct solutions. Namely, for given $g \in \operatorname{Lip}(\partial\Omega)$ the functions

$$\begin{aligned} \overline{g}(x) &:= \inf_{y\in\partial\Omega} g(y) + \operatorname{Lip}_{\tilde{d}_{\overline{\Omega}}}(g;\partial\Omega)\cdot \tilde{d}_{\overline{\Omega}}(x,y) \\ \underline{g}(x) &:= \sup_{y\in\partial\Omega} g(y) - \operatorname{Lip}_{\tilde{d}_{\overline{\Omega}}}(g;\partial\Omega)\cdot \tilde{d}_{\overline{\Omega}}(x,y) \end{aligned} \tag{2.6}$$

are solutions to the gradient-sup problem Problem 2.9 that coincide with g on $\partial\Omega$, see [Jen93, Th. 1.8].

Remark 2.12. The same concept of constructing solutions is applied in the following sections in a more abstract setting. These solutions are then called Whitney and McShane or respectively maximal and minimal extensions, see Lemma 2.15. △

One easily observes that there are cases where $\overline{g} \neq \underline{g}$ and therefore the sup-gradient problem does not admit a unique solution. A concrete example, to showcase this phenomena is given in [Jen93, p. 53].

Lipschitz extensions in metric spaces The problem considered in the last section was motivated by a variational problem for $\mathcal{E}_\infty(u) = \|\nabla u\|_\infty$. However, the theory of Lipschitz extensions provides a more general framework. Namely, here we do not assume that Ω is a subset of $\mathbb{R}^d$ and rather consider a metric space $(\tilde{\Omega}, d)$ with $\Omega \subset \tilde{\Omega}$.

Remark 2.13. For applications within this thesis we have that $\Omega \subset \mathbb{R}^d$ is an open bounded domain and then consider $\tilde{\Omega} := \overline{\Omega}$, i.e., the closure of Ω within the topology induced by the Euclidean distance. In this abstract setting however, we use an abstract space $\tilde{\Omega}$ while still being close notation wise. △

A result originally due to Kierszbraun [Kir34] states that for two Hilbert spaces H_1, H_2, a subset $\mathcal{O} \subset H_1$ and a function $g : \mathcal{O} \to H_1$ there exists a function $u : H_1 \to H_2$ such that

$$u = g \text{ on } U,$$
$$\mathrm{Lip}(u; H_1) = \mathrm{Lip}(g; \mathcal{O}).$$

Here, the metrics for the respective Lipschitz constants are induced by the inner products of the Hilbert spaces. We refer to [Kir34] for the original proof and to [Sch69, Th. 1.31] for a proof of the version as stated above. In this work we only consider the case $H_2 = \mathbb{R}$ which allows for more general assumption on the space H_1. We now formulate the Lipschitz extension problem in our setting.

Problem 2.14 (Lipschitz Extensions). Let $(\tilde{\Omega}, d)$ be a metric space and $\mathcal{O} \subset \tilde{\Omega}$ be a bounded subset. For a given Lipschitz function $g : \mathcal{O} \to \mathbb{R}$ find a Lipschitz function $u : \tilde{\Omega} \to \mathbb{R}$ such that

$$\mathrm{Lip}_d(u; \tilde{\Omega}) = \mathrm{Lip}_d(g; \mathcal{O}).$$

A function $u : \tilde{\Omega}$ with this property is called *Lipschitz extension* of g to $\tilde{\Omega}$.

In this setting one can explicitly construct solutions of the Lipschitz extension task. They are not unique, however one has an upper and a lower bound. In fact, conceptually these solutions are very similar to the functions in Eq. (2.6) and even coincide, whenever the sup-norm of the gradient is given as the Lipschitz constant.

Lemma 2.15. In the setting of Problem 2.14 we have that the

- **Whitney (or maximal) extension:** $\overline{g}(x) := \inf_{y \in \mathcal{O}} g(y) + \mathrm{Lip}_d(g; \mathcal{O}) \cdot d(x, y)$ and the
- **McShane (or minimal) extension:** $\underline{g}(x) := \sup_{y \in \mathcal{O}} g(y) - \mathrm{Lip}_d(g; \mathcal{O}) \cdot d(x, y)$

defined for $x \in \tilde{\Omega}$ are Lipschitz extensions of g to $\tilde{\Omega}$. Moreover, let $u : \tilde{\Omega} \to \mathbb{R}$ be any Lipschitz extension of g, the we have that

$$\underline{g} \leq u \leq \overline{g}.$$

Proof. We refer to [Whi92] and [McS34] for the proofs of the respective result. □

As demonstrated in Example 2.16, there are cases where $\overline{g} \neq \underline{g}$ and therefore, Lipschitz extensions are not unique in general. Furthermore, [ACJ04] points out that the Whitney and McShane extension do not allow for a comparison principle, which can also be observed in Example 2.16.

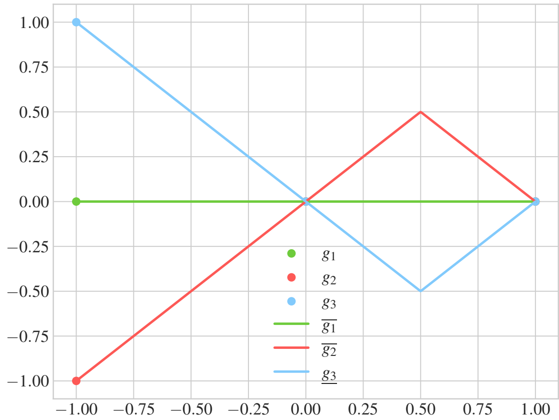


Figure 2.2.: The maximal extension does not admit a comparison principle, as demonstrated in Example 2.16.

Example 2.16. Consider the set $\tilde{\Omega} = [-1, 1]$ and $\mathcal{O} = \{-1, 0, 1\}$ with

$$\begin{aligned} g_1(x) &:= 0, \\ g_2(x) &:= 1/2(x - \operatorname{sgn}(x) \cdot x), \\ g_3(x) &:= -g_2. \end{aligned}$$

Then we have that $g_2 \leq g_1$ on $\mathcal{O}$ but

$$\overline{g_2} > \overline{g_1} \text{ in } (0, 1),$$

see Fig. 2.2 for a visualization. Analogously, we have that $g_3 \geq g_1$ on $\mathcal{O}$ but

$$\underline{g_3} < \underline{g_1} \text{ in } (0, 1).$$

Absolutely Minimizing Extension Sending $p \to \infty$ in the variational formulation of the p-Laplace equation yields the Lipschitz extension task, which however does not admit for unique solutions. So the question arises, which property is lost in the limit case. For $p < \infty$ one has the crucial local minimization property. Let $\Omega \subset \mathbb{R}^d$ be an open domain and denote by u_p the solution of the p-Laplace problem with boundary values $g \in W^{1,p}(\Omega)$. Then we know that u_p is also a minimizer of $\mathcal{E}_p$ on every subset $V \subset \Omega$, i.e.,

$$\int_V |\nabla u_p|^p \, dx \leq \int_V |\nabla v|^p \, dx$$

for any function v such that $(u_p - v) \in W_0^{1,p}(V)$, see e.g. [<**empty citation**>]. This lead Aronsson to introduce the concept of *absolutely minimizing Lipschitz extension* in

[Aro67]. A function $u_\infty \in W^{1,\infty}$ is called absolutely minimal, if

$$\operatorname*{ess\,sup}_{x\in V} |\nabla u| \leq \operatorname*{ess\,sup}_{x\in V} |\nabla v| \quad \text{for every open } V \subset \Omega \tag{2.7}$$

and every function v such that $(u-v) \in W_0^{1,\infty}$. I fact one can also show, that $u_p \xrightarrow{p\to\infty} u_\infty$ ([Aro67]), which seems to validate the notion of absolute minimizers. In [**tour**] it was shown, that one has an equivalent formulation involving the Lipschitz constant. For a given Lipschitz function $f : \overline{\Omega} \to \mathbb{R}$ we have that u_∞ with $(u_\infty - f) \in W_0^{1,\infty}(\Omega)$ fulfills Eq. (2.7) iff

$$\operatorname{Lip}(u_\infty; V) \leq \operatorname{Lip}(v; V) \text{ for every } V \subset \Omega$$

and every function v such that $(u-v) \in W_0^{1,\infty}(V)$, see [**tour**].

Maybe not Sobolev here

In this thesis we work with a notion of absolute minimizers, which is equivalent to the above formulation for convex domains in $\mathbb{R}^d$. However, ...

Problem 2.17 (AMLEs). Let $(\tilde{\Omega}, d)$ be a length space, $\mathcal{O} \subset \tilde{\Omega}$ a closed subset and $g : \mathcal{O} \to \mathbb{R}$ a Lipschitz function. Find an extension $u \in C(\tilde{\Omega})$ such that $u = g$ on $\mathcal{O}$ and

$$\operatorname{Lip}_d(u; \overline{V}) = \operatorname{Lip}_d(u, \partial V) \text{ for all open and connected sets } V \subset \tilde{\Omega} \setminus \mathcal{O}.$$

A function u fulfilling this property is called absolutely minimizing Lipschitz extension of g.

Remark 2.18. In our application Ω is an open subset of $\mathbb{R}^d$ and we then choose $\tilde{\Omega} = \overline{\Omega}$. Here, it is important to note that the topological notions like boundary and interior are to be understood relative to $\overline{\Omega}$. A visualization odf this concept can be found in Fig. 2.3. $\triangle$

Comparison with Cones

The infinity Laplacian

2.3.2. Graph Lipschitz Extensions

We now consider the limit $p \to \infty$ of Problem 2.5 in the graph case. Analogously to **??** we derive

$$\lim_{p\to\infty} \left(\mathbf{E}_p^{w_n}(\mathbf{u})\right)^{1/p} = \max_{x,y\in\Omega_n} w_n(x,y)\,|\mathbf{u}(y) - \mathbf{u}(x)| =: \mathbf{E}_\infty^{w_n}(\mathbf{u})$$

which extends the graph p-Laplacian energy to the case $p = \infty$. Again we notice structural similarities to the continuum version $\mathcal{E}_\infty$.

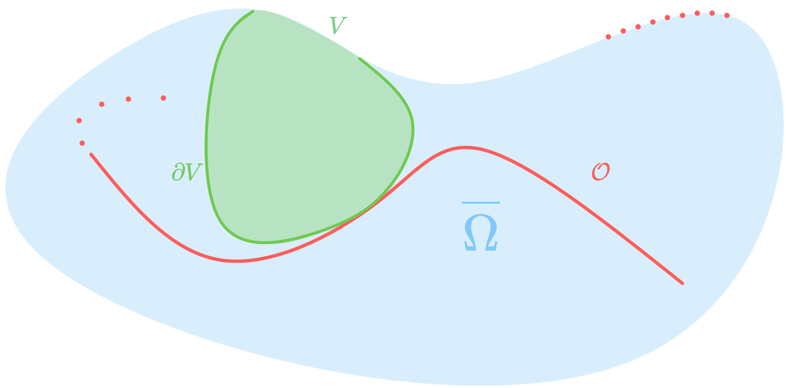


Figure 2.3.: A set $V \subset \overline{\Omega}$ can be relatively open w.r.t. the metric space $\overline{\Omega}$ although, $V \cap \partial\Omega \neq \emptyset$, where $\partial\Omega$ is the boundary within the standard topology on $\mathbb{R}^d$. The relative boundary of $\partial_{\overline{\Omega}} V$ does not include any parts of $\partial\Omega$.

Remark 2.19. Informally speaking the functional $\mathbf{E}$ combines elements of a gradient and Lipschitz constant. Assuming that $w_n(x,y)$ relates to $|x-y|$ we see that the finite difference approximation resembles a Lipschitz constant. However, typically $w_n(x,y)$ has also some localizing property which fits the interpretation of a gradient better. △

This functional now leads to the graph Lipschitz extension problem.

Problem 2.20 (Graph Energy Minimization). Given a weighted graph (Ω_n, w_n) and a labeling function $\mathbf{g} : \mathcal{O}_n \to \mathbb{R}$, for $\mathcal{O}_n \subset \Omega_n$ we consider the problem

$$\min_{\mathbf{u}:\Omega_n \to \mathbb{R}} \mathbf{E}_\infty^{w_n}(\mathbf{u}) \text{ subject to } \mathbf{u}(x) = \mathbf{g}(x) \text{ for all } x \in \mathcal{O}_n.$$

Since the weighting function $w_n : \Omega_n \times \Omega_n \to \mathbb{R}_0^+$ does not induce a metric, Problem 2.20 does not directly fit the framework of the abstract Lipschitz Extension in Problem 2.14. However, we can consider paths in (Ω_n, w_n), connecting arbitrary $x, y \in \Omega_n$ i.e. vectors $\gamma \in \Omega_n^{\times k}$ such that

$$\begin{aligned} w_n(\gamma_i, \gamma_{i+1}) &> 0 \quad \text{for all } i = 1, \dots, k-1, \\ \gamma_1 &= x, \\ \gamma_k &= y \end{aligned}$$

for which we define the length as

$$|\gamma| = \sum_{i=1}^{k-1} w(\gamma_i, \gamma_{i+1})^{-1}.$$

This yields the metric space (Ω_n, d_{w_n}), where $d_{w_n} : \Omega_n \times \Omega_n \to \mathbb{R}$ is defined as

$$d_{w_n}(x,y) := \min \{ |\gamma| : \gamma \text{ is a path in } (\Omega_n, w_n) \text{ from } x \text{ to } y \}. \tag{2.8}$$

Remark 2.21. We note that it is important to only consider non-negative weights, otherwise any loop with a negative "length" would decrease the length of the whole path arbitrarily. However, restricting ourselves to non-negative weights we can easily see, that the minimum in Eq. (2.8) is indeed attained. △

With this definition we can consider the Lipschitz extension task of $g : \mathcal{O}_n \to \mathbb{R}$ to Ω_n within the metric space (Ω_n, d_{w_n}), i.e. within the setting of Problem 2.14. Therefore the question arises, whether the minimization problem in Problem 2.20 is equivalent to the metric Lipschitz extension problem for which we have the following lemma.

Lemma 2.22. For a graph (Ω_n, w_n) with non-negative weights and a function $\mathbf{u} : \Omega_n \to \mathbb{R}$ we have that

$$\mathbf{E}_\infty^{w_n}(\mathbf{u}) = \mathrm{Lip}_{d_{w_n}}(\mathbf{u}).$$

Furthermore, for $\mathcal{O}_n \subset \Omega_n$ and a function $\mathbf{g} : \mathcal{O}_n \to \mathbb{R}$ and we have that

$$\mathbf{g} = \mathbf{u} \text{ on } \mathcal{O}_n \Rightarrow \mathrm{Lip}_{d_{w_n}}(\mathbf{g}; \mathcal{O}_n) \leq \mathrm{Lip}_{d_{w_n}}(\mathbf{u}).$$

Proof. **Step 1:** We show that $\mathrm{Lip}_{d_{w_n}}(\mathbf{u}) \leq \mathbf{E}_\infty^{w_n}(\mathbf{u})$.
We can choose a path $\gamma \in \Omega_n^{\times k}$ such that

$$\mathrm{Lip}_{d_{w_n}}(\mathbf{u}) = \frac{\mathbf{u}(\gamma_1) - \mathbf{u}(\gamma_k)}{|\gamma|}.$$

The path γ allows to compare vertices $\gamma_1, \gamma_k \in \Omega_n$ that aren't necessarily neighbors in the graph. However, each consecutive vertices in the path are neighbors in the graph and therefore we have

$$w_n(\gamma_i, \gamma_{i+1}) \left|\mathbf{u}(\gamma_{i+1}) - \mathbf{u}(\gamma_i)\right| \leq \mathbf{E}_\infty^{w_n}(\mathbf{u}) \quad \text{for all } i = 1, \dots, k-1. \tag{2.9}$$

We now employ an elementary result for numbers $a_i \in \mathbb{R}_0^+, b_i \in \mathbb{R}^+, i = 1, \dots m \in \mathbb{N}$, namely

$$[a_i \cdot b_i \leq c \in \mathbb{R} \quad \text{for } i = 1, \dots, m] \Rightarrow \frac{\sum_{i=1}^m a_i}{\sum_{i=1}^m b_i^{-1}} \leq c \tag{2.10}$$

which can be seen as follows

$$\begin{aligned} & a_i \cdot b_i \leq c \quad \text{for } i = 1, \dots, m \\ & \Rightarrow a_i \leq b_i^{-1} \cdot c \quad \text{for } i = 1, \dots, m \\ \Rightarrow & \sum_{i=1}^m a_i \leq \left(\sum_{i=1}^m b_i^{-1}\right) \cdot c \\ \Rightarrow & \frac{\sum_{i=1}^m a_i}{\sum_{i=1}^m b_i^{-1}} \leq c. \end{aligned}$$

This then yields

$$\frac{|\mathbf{u}(\gamma_1) - \mathbf{u}(\gamma_k)|}{|\gamma|} \leq \frac{\sum_{i=1}^{k-1} |\mathbf{u}(\gamma_i) - \mathbf{u}(\gamma_{i+1})|}{|\gamma|} = \frac{\sum_{i=1}^{k-1} |\mathbf{u}(\gamma_i) - \mathbf{u}(\gamma_{i+1})|}{\sum_{i=1}^{k-1} w_n(\gamma_i, \gamma_{i+1})^{-1}} \leq \mathbf{E}_\infty^{w_n}(\mathbf{u})$$

where in the last inequality we employed Eq. (2.10) together with Eq. (2.9).

Step 2: We show that $\mathrm{Lip}_{d_{w_n}}(\mathbf{u}) \geq \mathbf{E}_\infty^{w_n}(\mathbf{u})$.

Let $x, y \in \Omega_n$, then we know that $d_w(x,y) \leq w_n(x,y)^{-1}$ and therefore

$$|\mathbf{u}(x) - \mathbf{u}(y)|\, w_n(x,y) \leq \frac{|\mathbf{u}(x) - \mathbf{u}(y)|}{d_w(x,y)} \leq \max_{\bar{x},\bar{y} \in \Omega_n} \frac{|\mathbf{u}(\bar{x}) - \mathbf{u}(\bar{y})|}{d_w(\bar{x},\bar{y})} = \mathrm{Lip}_{d_{w_n}}(\mathbf{u}).$$

Since this holds for arbitrary $x, y \in \Omega_n$ we have that

$$\mathbf{E}_\infty^{w_n}(\mathbf{u}) = \max_{x,y \in \Omega_n} |\mathbf{u}(x) - \mathbf{u}(y)|\, w_n(x,y) \leq \mathrm{Lip}_{d_{w_n}}(\mathbf{u}).$$

Step 3: We show that $\mathrm{Lip}_{d_{w_n}}(\mathbf{g}; \mathcal{O}_n) \leq \mathrm{Lip}_{d_{w_n}}(\mathbf{u})$.

If $\mathbf{g} = \mathbf{u}$ on $\mathcal{O}$ this simply follows since the maximum for the Lipschitz constant of $\mathbf{u}$ is taken over a larger set. Indeed,

$$\begin{aligned}\mathrm{Lip}_{d_{w_n}}(\mathbf{u}) &= \max_{x,y \in \Omega_n} \frac{|\mathbf{u}(x) - \mathbf{u}(y)|}{d_w(x,y)} \geq \max_{x,y \in \mathcal{O}_n} \frac{|\mathbf{u}(x) - \mathbf{u}(y)|}{d_w(x,y)} = \max_{x,y \in \mathcal{O}_n} \frac{|\mathbf{g}(x) - \mathbf{g}(y)|}{d_w(x,y)} \\ &= \mathrm{Lip}_{d_{w_n}}(\mathbf{u}; \mathcal{O}_n).\end{aligned}$$

□

This lemma shows that the abstract Lipschitz extension task considered on the metric space (Ω_n, d_{w_n}) and the Graph ∞-Dirichlet minimization task are indeed equivalent. Therefore, we also have the Whitney and McShane extensions

$$\begin{aligned}\overline{\mathbf{g}}(x) &= \inf_{y \in \mathcal{O}_n} \mathbf{g}(y) + d_{w_n}(x,y) \\ \underline{\mathbf{g}}(x) &= \sup_{y \in \mathcal{O}_n} \mathbf{g}(y) - d_{w_n}(x,y)\end{aligned}$$

as solutions on the graph. Analogously, the problem does not admit for unique solutions.

Absolutely Minimizing Graph Extensions Similarly to Section 2.3.1 we can now consider absolutely minimizing extensions. However, the problem in Problem 2.17 uses a notion of a boundary and it is not directly clear how to infer this concept to the discrete set Ω_n. Therefore, we define the following what we mean by "boundary" on a graph.

> **Definition 2.23.** Let (Ω_n, w_n) be a weight graph and let $V \subset \Omega_n$ be a subset, then we define

- the **exterior** boundary as $\partial^{\text{ext}} := \{x \in \Omega_n \setminus V : w_n(x,y) > 0 \text{ for some } y \in V\}$,
- the **interior** boundary as $\partial^{\text{int}} := \{x \in V : w_n(x,y) > 0 \text{ for some } y \in \Omega_n \setminus V\}$.

The closure of V is then defined as $\overline{V}^{\text{ext}} V := V \cup \partial^{\text{ext}}$ and the interior as $\overset{\circ}{V}^{\text{int}} := V \setminus \partial^{\text{int}} V$.

We note that it is not possible to define a topology on Ω_n that would yield the above notions. Namely, the only admissible topology in our case would be the discrete topology, i.e., 2^{Ω_n}. However, in this topology the only closed sets are $\emptyset$ and Ω_n which is not useful for the applications in the following. Using the Kuratowski closure axioms [Kur22] we remark the following.

Lemma 2.24. The exterior closure on a weighted graph (Ω_n, w_n) is a preclosure or Čech closure.

Proof. Here, we use the notion of a preclosure in [ČFK66]. We first see that $\overline{\emptyset}^{\text{ext}} = \emptyset$ and that $V \subset \overline{V}^{\text{ext}}$ for every subset $V \subset \Omega_n$, i.e. the above defined closure preserves the empty set and is extensive. Furthermore, for two sets $V_1, V_2 \subset \Omega_n$ we have that

$$
\begin{gathered}
x \in \partial^{\text{ext}}(V_1 \cup V_2) \\
\Leftrightarrow [x \notin V_1 \cup V_2] \wedge [\exists y \in V_1 \cup V_2 : w_n(x,y)] \neq 0 \\
\Leftrightarrow [x \notin V_1 \cup V_2] \wedge \Big([\exists y \in V_1 : w_n(x,y) \neq 0] \vee [\exists y \in V_2 : w_n(x,y) \neq 0] \Big) \\
\Leftrightarrow \Big[x \in \partial^{\text{ext}} V_1 \setminus V_2\Big] \vee \Big[x \in \partial^{\text{ext}} V_2 \setminus V_1\Big] \\
\Leftrightarrow x \in (\partial^{\text{ext}} V_1 \cup \partial^{\text{ext}} V_2) \setminus (V_1 \cup V_2).
\end{gathered}
$$

We have shown that $\partial^{\text{ext}}(V_1 \cup V_2) = (\partial^{\text{ext}} V_1 \cup \partial^{\text{ext}} V_2) \setminus (V_1 \cup V_2)$. Therefore, we have that

$$
\begin{aligned}
\overline{V_1 \cup V_2}^{\text{ext}} &= V_1 \cup V_2 \cup \partial^{\text{ext}}(V_1 \cup V_2) \\
&= V_1 \cup V_2 \cup ((\partial^{\text{ext}} V_1 \cup \partial^{\text{ext}} V_2) \setminus (V_1 \cup V_2)) \\
&= V_1 \cup \partial^{\text{ext}} V_1 \cup V_2 \cup \partial^{\text{ext}} V_2 \\
&= \overline{V_1}^{\text{ext}} \cup \overline{V_2}^{\text{ext}}.
\end{aligned}
$$

This shows that the closure preserves binary unions and therefore we have shown, that it is indeed a Čech closure. □

The missing property, that inhibits the closure to induce a topology is the so-called idempotence. Namely, there are sets $V \subset \Omega_n$ such that

$$
\overline{V}^{\text{ext}} \neq \overline{\overline{V}^{\text{ext}}}^{\text{ext}}.
$$

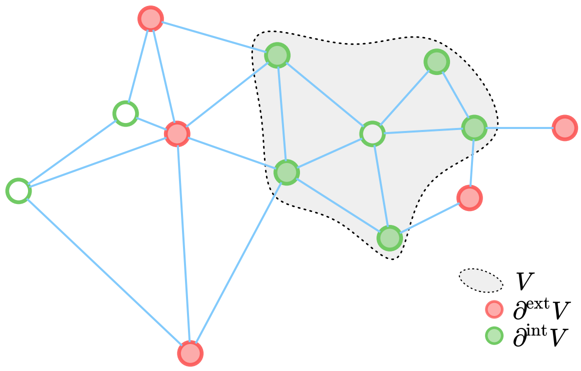


Figure 2.4.: Visualization of exterior and interior boundary on a graph.

E.g. in the example visualized in Fig. 2.4 we see that $\overline{\overline{V}^{\text{ext}}}^{\text{ext}} = \Omega_n \neq \overline{V}^{\text{ext}}$. Since the closure we employ here does not induce a topology, we have a slightly modified notion of absolutely minimizers.

Problem 2.25 (Graph AMLEs). Given a connected weighted graph (Ω_n, w_n), $\mathcal{O}_n \subset \Omega_n$) and a function $\mathbf{g} : \mathcal{O}_n \to \mathbb{R}$ find a function $\mathbf{u} : \Omega_n \to \mathbb{R}$ such that

$$\begin{aligned} \operatorname{Lip}_{d_{w_n}}(\mathbf{u}; \overline{V}^{\text{ext}}) &= \operatorname{Lip}_{d_{w_n}}(\mathbf{u}; \partial^{\text{ext}} V) \text{ for all connected } V \subset \Omega_n \setminus \mathcal{O}_n, \\ \mathbf{u} &= \mathbf{g} \text{ on } \mathcal{O}_n. \end{aligned}$$

Comparison with graph Distance functions Analogously to the continuum case **??** we can also consider comparison with distance functions on graphs. The main ingredients here, are the graph distance function d_{w_n} and the notion of closure on a graph as developed in the last section.

Definition 2.26. For a weighted graph (Ω_n, w_n) we say a function $\mathbf{u} : \Omega_n \to \mathbb{R}$ fulfills comparison with distance function from above (CDFA) on a subset $U \subset \Omega_n$ if for every $V \subset U$ we have

$$\max_{\overline{V}^{\text{ext}}} (u + a\, d_{w_n}(\cdot, z)) = \max_{\partial^{\text{ext}} V} (u + a\, d_{w_n}(\cdot, z)) \tag{CDFA}$$

for every $z \in \Omega_n \setminus V$ and every $a \in \mathbb{R}$. We say that $\mathbf{u}$ fulfills comparison with distance function from below (CDFB) on a subset $U \subset \Omega_n$ if for every $V \subset U$ we have

$$\min_{\overline{V}^{\text{ext}}} (u - a\, d_{w_n}(\cdot, z)) = \min_{\partial^{\text{ext}} V} (u - a\, d_{w_n}(\cdot, z)) \tag{CDFB}$$

for every $z \in \Omega_n \setminus V$ and every $a \in \mathbb{R}$.

Analogously to the continuum case, we say that a function fulfills comparison with distance functions, if it fulfills both, Eq. (CDFA) and Eq. (CDFB). Existence of such functions is established later, we are first interested in the question of uniqueness. Since the notion of graph boundaries is not directly compatible with the usual definitions on metric spaces, we prove it separately. Here, we adapt arguments from [smart] and [LeGruyer]. To do so we first consider the operators

$$\mathbf{S}^\varepsilon \mathbf{u}(x) := \max_{y\in\Omega_n : d_{w_n}(x,y)\leq\varepsilon} \mathbf{u}(y) \quad \mathbf{S}_\varepsilon \mathbf{u}(x) := \min_{y\in\Omega_n : d_{w_n}(x,y)\leq\varepsilon} \mathbf{u}(y)$$

and proof the following lemma, which is the analogue of

The Graph infinity Laplacian We can also obtain the limit of the Graph p-Laplace operator via the following formal calculation,

$$\begin{aligned} &\Delta_p^{w_n}\mathbf{u}(x) = 0 \\ \Leftrightarrow \sum_{y\in\Omega_n} &w_n(x,y)^p \left|\mathbf{u}(y)-\mathbf{u}(x)\right|^{p-2}(\mathbf{u}(y)-\mathbf{u}(x)) = 0 \\ &\Leftrightarrow \sum_{y:\mathbf{u}(x)\leq\mathbf{u}(y)} w_n(x,y)^p(\mathbf{u}(y)-\mathbf{u}(x))^{p-1} = \sum_{y:\mathbf{u}(x)>\mathbf{u}(y)} w_n(x,y)^p(\mathbf{u}(x)-\mathbf{u}(y))^{p-1}. \end{aligned}$$

Taking the terms on the left and right hand side to the power of $1/p$ and then formally sending $p\to\infty$ then yields

$$\begin{aligned} \max_{y:\mathbf{u}(x)\leq\mathbf{u}(y)} w_n(x,y)(\mathbf{u}(y)-\mathbf{u}(x)) &= \max_{y:\mathbf{u}(x)>\mathbf{u}(y)} w_n(x,y)(\mathbf{u}(x)-\mathbf{u}(y)) \\ \Leftrightarrow \max_{y\in\Omega_n} w_n(x,y)(\mathbf{u}(y)-\mathbf{u}(x)) &= -\min_{y\in\Omega_n} w_n(x,y)(\mathbf{u}(y)-\mathbf{u}(x)). \end{aligned}$$

This calculation motivates the definition of the graph infinity Laplacian

$$\Delta_\infty^{w_n}\mathbf{u}(x) := \max_{y\in\Omega_n} w_n(x,y)(\mathbf{u}(y)-\mathbf{u}(x)) + \min_{y\in\Omega_n} w_n(x,y)(\mathbf{u}(y)-\mathbf{u}(x)),$$

which then allows to formulate the associated problem as an extension of Problem 2.7.

Problem 2.27 (Graph ∞-Laplacian). Given a weighted graph (Ω_n, w_n) and a labeling function $\mathbf{g}:\mathcal{O}_n\to\mathbb{R}$ with $\mathcal{O}_n\subset\Omega_n$, find a function $\mathbf{u}:\Omega_n\to\mathbb{R}$ such that

$$\begin{aligned} \Delta_\infty^{w_n}\mathbf{u} &= 0, \text{ in } \Omega_n\setminus\mathcal{O}_n, \\ \mathbf{u} &= \mathbf{g} \text{ on } \mathcal{O}_n. \end{aligned}$$

This problem is again well-posed, which is formulated in the following lemma.

Lemma 2.28. There exists a unique solution for Problem 2.27.

Relation between the Graph Lipschitz Extensions We now establish the connection between the different notions of Lipschitz extensions. Compared to the continuum case we do not establish the full equivalences but only the necessary implications required for the convergence proofs in [LIP-II].

First we see, that graph AMLEs are indeed special solutions of the basic Lipschitz extension problem on the graph.

Lemma 2.29. A graph AMLE is also a Lipschitz extension.

Proof. □

We now state the main result concerning the relation between the graph infinity Laplacian, graph AMLEs and comparison cones.

Lemma 2.30. Let (Ω_n, w_n) be a weighted connected graph and $\mathbf{g} : \mathcal{O}_n \to \mathbb{R}$ be a given function for $\mathcal{O}_n \subset \Omega_n$. Furthermore, let $\mathbf{u} : \Omega_n \to \mathbb{R}$ be graph infinity harmonic on $\Omega_n \setminus \mathcal{O}_n$ with boundary conditions given by $\mathbf{g}$, i.e., $\mathbf{u}$ solves Problem 2.27 then we have that

- $\mathbf{u}$ is an graph AMLE, i.e., $\mathbf{u}$ solves Problem 2.25,
- $\mathbf{u}$ fulfills comparison with cones.

Proof. Both of the stament are proven in [LIP-II]. From [LIP-II, Prop. 3.8] we have that $\mathbf{u}$ is an graph AMLE. Furthermore, form [LIP-II, Th. 3.2] we have that $\mathbf{u}$ fulfills comparison with cones. In fact, [LIP-II, Th. 3.2], shows a more refined statement, namely that

$$-\Delta_\infty^{w_n}\mathbf{u} \leq 0 \Rightarrow \mathbf{u} \text{ fulfills CDFA},$$
$$-\Delta_\infty^{w_n}\mathbf{u} \geq 0 \Rightarrow \mathbf{u} \text{ fulfills CDFB}.$$

□

2.4. Gamma Convergence

2.5. Ratio Convergence

Chapter 3

Supervised Learning

[CLIP; BREG-II; FNO; BREG-I]

3.1. Setting

3.2. Sparsity

3.3. Adversarial Stability

3.4. Resolution Stability

Part II.

Prints

Bibliography

Books

[Bre+84] L. Breiman et al. *Clasificaion and regression trees.* CRC Press, New York, 1984.

[Zhu05] X. Zhu. *Semi-supervised learning with graphs.* Carnegie Mellon University, 2005.

[Lin17] P. Lindqvist. *Notes on the p-Laplace equation.* 161. University of Jyväskylä, 2017.

[BB11] H. Brezis and H. Brézis. *Functional analysis, Sobolev spaces and partial differential equations.* Vol. 2. 3. Springer, 2011.

[Sch69] J. T. Schwartz. *Nonlinear Functional Analysis -.* Boca Raton, Fla: CRC Press, 1969.

Articles

[LIP-I] T. Roith and L. Bungert. “Continuum limit of Lipschitz learning on graphs.” In: *Foundations of Computational Mathematics* (2022), pp. 1–39.

[LIP-II] L. Bungert, J. Calder, and T. Roith. “Uniform convergence rates for Lipschitz learning on graphs.” In: *IMA Journal of Numerical Analysis* (Sept. 2022). DOI: 10.1093/imanum/drac048.

[BREG-I] L. Bungert et al. “A bregman learning framework for sparse neural networks.” In: *Journal of Machine Learning Research* 23.192 (2022), pp. 1–43.

[ST14] A. Subramanya and P. P. Talukdar. “Graph-based semi-supervised learning.” In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 8.4 (2014), pp. 1–125.

[Ste+56] H. Steinhaus et al. “Sur la division des corps matériels en parties.” In: *Bull. Acad. Polon. Sci* 1.804 (1956), p. 801.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum likelihood from incomplete data via the EM algorithm.” In: *Journal of the royal statistical society: series B (methodological)* 39.1 (1977), pp. 1–22.

[GS15] N. García Trillos and D. Slepčev. "Continuum Limit of Total Variation on Point Clouds." In: *Archive for Rational Mechanics and Analysis* 220.1 (2015), pp. 193–241. DOI: 10.1007/s00205-015-0929-z.

[SB09] A. Szlam and X. Bresson. "A total variation-based graph clustering algorithm for cheeger ratio cuts." In: *UCLA Cam report* (2009), pp. 09–68.

[Gar+16] N. García Trillos et al. "Consistency of Cheeger and ratio graph cuts." In: *The Journal of Machine Learning Research* 17.1 (2016), pp. 6268–6313.

[GMT22] N. García Trillos, R. Murray, and M. Thorpe. "From graph cuts to isoperimetric inequalities: Convergence rates of Cheeger cuts on data clouds." In: *Archive for Rational Mechanics and Analysis* 244.3 (2022), pp. 541–598.

[GS18] N. García Trillos and D. Slepčev. "A variational approach to the consistency of spectral clustering." In: *Applied and Computational Harmonic Analysis* 45.2 (2018), pp. 239–281.

[CV95] C. Cortes and V. Vapnik. "Support-vector networks." In: *Machine learning* 20 (1995), pp. 273–297.

[MS63] J. N. Morgan and J. A. Sonquist. "Problems in the analysis of survey data, and a proposal." In: *Journal of the American statistical association* 58.302 (1963), pp. 415–434.

[Ros58] F. Rosenblatt. "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6 (1958), p. 386.

[MP69] M. Minsky and S. Papert. "An introduction to computational geometry." In: *Cambridge tiass., HIT* 479 (1969), p. 480.

[Sch15] J. Schmidhuber. "Deep learning in neural networks: An overview." In: *Neural Networks* 61 (2015), pp. 85–117. DOI: https://doi.org/10.1016/j.neunet.2014.09.003.

[Cho+20] Y. Chong et al. "Graph-based semi-supervised learning: A review." In: *Neurocomputing* 408 (2020), pp. 216–230.

[Jen93] R. Jensen. "Uniqueness of Lipschitz extensions: minimizing the sup norm of the gradient." In: *Archive for Rational Mechanics and Analysis* 123 (1993), pp. 51–74.

[Kir34] M. Kirszbraun. "Über die zusammenziehende und Lipschitzsche Transformationen." In: *Fundamenta Mathematicae* 22 (1934), pp. 77–108. DOI: 10.4064/fm-22-1-77-108.

[Whi92] H. Whitney. "Analytic extensions of differentiable functions defined in closed sets." In: *Hassler Whitney Collected Papers* (1992), pp. 228–254.

[McS34] E. J. McShane. "Extension of range of functions." In: (1934).

[ACJ04] G. Aronsson, M. Crandall, and P. Juutinen. "A tour of the theory of absolutely minimizing functions." In: *Bulletin of the American mathematical society* 41.4 (2004), pp. 439–505.

[Aro67] G. Aronsson. “Extension of functions satisfying Lipschitz conditions.” In: *Arkiv för Matematik* 6.6 (1967), pp. 551–561.

[Kur22] C. Kuratowski. “Sur l’opération A de l’analysis situs.” In: *Fundamenta Mathematicae* 3.1 (1922), pp. 182–199.

[ČFK66] E. Čech, Z. Frolík, and M. Katětov. “Topological spaces.” In: (1966).

[BREG-II] L. Bungert et al. “Neural Architecture Search via Bregman Iterations.” In: (2021). arXiv: `2106.02479 [cs.LG]`.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht sind und dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt wurde.

Erlangen, den 14.Juni 2023 ....

Tim Roith