3-2022

Existence and Uniqueness of Minimizers for A Nonlocal Variational Problem

Michael Pieper
University of Nebraska - Lincoln

EXISTENCE AND UNIQUENESS OF MINIMIZERS FOR A NONLOCAL VARIATIONAL PROBLEM

by

Michael Pieper

AN UNDERGRADUATE HONORS THESIS

Submitted in Partial Fulfillment of

The University Honors Program Requirements

University of Nebraska-Lincoln

Major: Mathematics, Philosophy

Under the Supervision of Professor Mikil Foss

Lincoln, Nebraska

May, 2022

EXISTENCE AND UNIQUENESS OF MINIMIZERS FOR A NONLOCAL VARIATIONAL PROBLEM

Michael Pieper, B.S.

University of Nebraska, 2022

Advisor: Mikil Foss

Nonlocal modeling is a rapidly growing field, with a vast array of applications and connections to questions in pure math. One goal of this work is to present an approachable introduction to the field and an invitation to the reader to explore it more deeply. In particular, we explore connections between nonlocal operators and classical problems in the calculus of variations. Using a well-known approach, known simply as The Direct Method, we establish well-posedness for a class of variational problems involving a nonlocal first-order differential operator. Some simple numerical experiments demonstrate the behavior of these problems for specific choices of kernel and boundary conditions.

GRANT INFORMATION

This work was supported by an Undergraduate Creative Activities and Research Experience at Nebraska Grant.

Table of Contents

List of Figures

Chapter 1

Background and Motivation

The goal of this chapter is to introduce the fundamental concepts of nonlocal variational problems. The hope is that the introduction in first two subsections will be understandable for anyone with knowledge of multivariable calculus and linear algebra. A brief history is presented to place the questions in their proper context, and key concepts are given some intuitive motivation. The discussion is somewhat informal and technical details are mostly saved for later.

1.1 Calculus of Variations

A vast array of problems across the mathematical sciences can be viewed as optimization problems: minimizing error, maximizing profits, minimizing energy or material used in a given process, maximizing the strength or stability of a structure, and so on. Some of these can be handled using elementary techniques from calculus. If you can write down a sufficiently smooth function modeling the quantity to be optimized, just take the derivative and set it equal to zero. This generally leads to an algebraic equation, which can either be solved exactly or with well-established numerical techniques. Even in cases with many variables, the same basic procedure generally holds: take the gradient of the function, find where it is 0, and solve the corresponding system of equations.

From a mathematical standpoint, it is natural to wonder if these techniques can be extended to a more general setting. Can we solve optimization problems in an abstract vector space? Is there an analogous procedure? Additionally, the process mentioned above for the finite dimensional case only gives *necessary* conditions for local extrema. For example, $f(x) = x^3$ satisfies $f'(0) = 0$, but there is no local extremum at the origin. So we would also like to know when optimal solutions

even exist: what are the *sufficient* conditions for the presence of an extreme value? These are the sorts of questions investigated by the calculus of variations.

Studying these more general problems is not merely mathematical abstraction. As we will see, problems in the calculus of variations have found all sorts of applications throughout science and engineering.

1.1.1 Origins

The *Isoperimetric Problem* is perhaps the first problem ever solved in the calculus of variations. Though a (non-rigorous) solution was known to the Ancient Greeks, the problem is now most strongly associated with Dido, Queen of Carthage. One version of the story goes like this: Upon arriving at the north coast of Africa, Dido negotiated to buy as much land as could be enclosed enclosed by an oxskin. She cut the skin into thin strips and arranged them into a circle, thus solving the problem of maximizing enclosed area given a fixed perimeter (Blasjö 2005).

The key difference between the isoperimetric problem and the optimization problems of elementary calculus is that Dido was looking for a *curve* rather than a point. To fully specify a continuous path through space, one needs to specify the functional values for an infinite list of points. Thus, the problem cannot be solved using the basic techniques of multivariable calculus, which only handle finite dimensional spaces. To find necessary conditions for minimimization problems in infinite dimensional spaces, we need the techniques of the calculus of variations.

The story of the isoperimetric problem also highlights another key component of the calculus of variations. Beginning in 1838, Jakob Steiner published a series of five proofs that the circle does in fact maximize area given a fixed perimeter. He relied primarily on geometric arguments (which are presented in Blasjö 2005), but made a crucial mistake. He never gave any justification for the fact that the problem had a solution. As Weierstrass noted, Steiner's arguments "do not prove that there is an actual maximum, and not just an upper bound" (Blasjö 2005). Thus, it was not until 1879, when Weierstrass showed the problem was well-posed, that the isoperimetric problem was fully solved.

We've seen that the types of problems we will be concerned with thus involve two considerations:

1. Sufficient conditions: when does a solution even exist?

2. Necessary conditions: what properties must the solution satisfy?

For much of the history of the calculus of variations, the focus was on this second question. To better explore the ideas involved, we turn to another classic example from the history of mathematics.

In June of 1696, John Bernoulli posed the following problem:

> Given points A and B in a vertical plane to find the path AMB down which a movable point M must, by virtue of its weight, proceed from A to B in the shortest possible time (Goldstine 1980).

Note that this does not ask for the shortest possible distance, which is of course just a strraight line connecting A and B. Rather, the problem requires minimizing the time of travel. The problem is illustrated in Figure 1.1, and has become known as the brachistochrone problem (from the Greek for 'shortest time').[1]

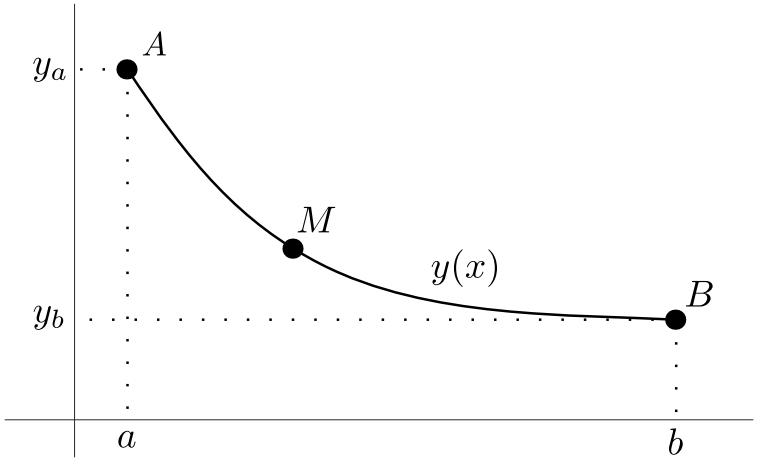


Figure 1.1: Bernoulli's Brachistochrone problem: find the path connecting A and B which minimizes the travel time of the falling object M. Image adapted from Kot 2014.

Many of the top mathematicians in Europe submitted solutions, including Leibniz, l'Hôpital, Newton and John's brother James Bernoulli.[2] The solution to the problem is a curve called the

[1] Some very nice animations and interactive explorations of the problem are hosted by the MAA at https://www.maa.org/press/periodicals/convergence/historical-activities-for-calculus-module-3-optimization-galileo-and-the-brachistochrone-problem.

[2] The Bernoulli family included a shockingly large number of hugely influential mathematicians. In fact, in his

cycloid. This is the path a point on a circle traces out as it rolls. Supposedly, Newton wrote his solution the very same day he received the problem. For whatever reason, he decided to submit his solution anonymously. But upon receiving the work, John Bernoulli famously said he knew immediately that it was Newton, because he could "recognize the lion by his claw" (Goldstine 1980).

These initial solutions, like Steiner's solution to the isoperimetric problem, primarily relied on geometrical reasoning. The Bernoullis and other mathematicians of the time went on to study many related problems, including determining the shape of a hanging wire, supported only at each end (known as a catenary curve). This work, and the brachistrocrone problem in particular, led Euler to publish *The Method of Finding Curves that Show Some Property of Maximum or Minimum* in 1744. Euler considered 100 optimization problems and began formulating the analytic framework that we use today (Goldstine 1980).[3] This work inspired Lagrange, who was the first to introduce the notion of variations and solidify the foundations of the field in more modern terms.[4]

To illustrate the basic method, let's set up the brachistochrone in more detail.[5] Suppose that the path connecting A and B is given by a function $y(x)$. Let s denote the arc length and $\dot{s}$ denote its time derivative, the velocity of the object M moving along the curve. If L is the total length of

well-known text *History of Mathematics*, Carl Boyer refers to the whole period between Newton and Euler as simply "The Bernoulli Era" (Boyer 1968). Both Euler and his father were pupils of a Bernoulli.

The relationship between James and John is particularly notable. Although both made many contributions to extending infintesmial calculus to new problems, there had been a longstanding rivalry between the brothers. Interestingly, one of their public feuds involved the isoperimetric problem mentioned earlier. James found a solution in 1697, recognizing it as a calculus of variations problem. But for some reason it was not made public until 1706. John published an incorrect solution, and refused to admit both that his approach was flawed and that his brother had given a correct analysis (Stillwell 2010, p. 271).

[3]For some additional discussion of Euler's method, see Section 2.2 of Kot 2014.

[4]Lagrange was only 19 when he wrote to Euler with his general solution to problems in the calculus of variations. Euler, who had independently come to the same conclusion (what is now called the Euler-Lagrange Equation, which will be discussed later) intentionally withheld his paper so that the young Lagrange could publish first. Euler continued to support Lagrange throughout his career, as it was Euler's recommendation that secured Lagrange a spot as a foreign member in Berlin's Academy. In fact, Lagrange's work across math and physics was so impressive that when Lagrange moved to Prussia in 1766, King Frederick the Great welcomed him as "Europe's Greatest Mathematician." Lagrange led a very interesting life, getting tangled up with royals and nobles during the French revolution. (Motz and Weaver 1993, p. 159). Even Napoleon referred to Lagrange as "the lofty pyramid of the mathematical sciences" (Motz and Weaver 1993, p. 161).

[5]The discussion below is based on Chapter 1 of Kot 2014.

the curve, then the total time T of travel from A to B is given by

$$\int_0^L \frac{1}{\dot{s}}\,\mathrm{d}s.$$

We can use the arc length formula from calculus (essentially the infinitesimal version of the Pythagorean Theorem) to find

$$T = \int_a^b \frac{1}{\dot{s}}\sqrt{1+y'^2}\,\mathrm{d}x.$$

Note that $\dot{s}$ is the derivative with respect to time, while y' is the derivative of height or vertical position with respect to x. Recall that kinetic energy is given by $\frac{1}{2}m\dot{s}^2$ and gravitational potential energy is mgy. Assuming that M starts at rest and energy is conserved, we have

$$\frac{1}{2}mg\dot{s}^2 + mgy = mgy_a,$$

where y_a is the initial height of the point A (as in Figure 1.1). We can then solve for the velocity:

$$\dot{s} = \sqrt{2g(y_a - y)}.$$

Thus, total time is given by

$$T = \frac{1}{\sqrt{2g}}\int_a^b \sqrt{\frac{1+y'^2}{y_a - y}}\,\mathrm{d}x. \tag{1.1}$$

There are several things to note about Equation 1.1. First, the quantity T depends not on the choice of a single input, but on the entire path taken. Thus, we think of T as a *functional*: a mapping that takes a path or function as an input (or, more generally, an element of some abstract vector space), and outputs a real number. Also, both y and its first spatial derivative, y' appear inside the integral. This turns out to be fairly typical: functionals studied in the calculus of variations can be represented as something like

$$J[y] = \int_D f\left(x, y(x), y'(x)\right)\mathrm{d}x, \tag{1.2}$$

where J is the functional and D is some domain of integration.[6] To simplify notation, we write "y" instead of "$y(x)$," since it is usually clear from context what the argument of u should be. To ensure the problem is well-formed, we need to specify the set of possible solutions. For example, Bernoulli probably implicitly assumed that the path of least time should be fairly smooth, without any jumps. Thus, when fully describing a minimization problem in the calculus of variations, we always provide the functional J, as well as its *admissible class*. For example, the admissible class for the original brachistochrone problem might be something like

$$\mathcal{A} = \left\{ y \in C^2\left([a,b]\right) : y(a) = y_a, y(b) = y_b \right\},$$

where $C^2\left([a,b]\right)$ denotes the set of all functions from $[a,b]$ into $\mathbb{R}$ that have continuous second derivative.

Now that the problem is set up, we turn to Lagrange's notion of variations.[7] One way to think of the derivative is as measuring the change in the functional output when we nudge or perturb the input by a small amount. In one dimensional problems, there's really only one direction to nudge the input. But in higher dimensions, it doesn't really make sense to simply take a derivative. You must first specify the *direction* in which you are perturbing the input. Once a point and a vector are chosen, we can take the derivative of the function in the direction of that vector. The basic idea of Lagrange's "first variation" is to extend this notion of a directional derivative to the space of functions. Suppose we have a path y in the admissible class $\mathcal{A}$. Fix a "point", y, and a "direction," v. Then we nudge the input y a tiny bit in the direction of v, and measure the corresponding change in the functional J. We call v the *variation*. To make sure that everything is still well-defined, we need to choose v carefully: $y(x) + \varepsilon v(x)$ should still be in the admissible class, as long as ε is

[6] It turns out that focusing on integration doesn't limit the scope of inquiry much, since the Riesz Representation Theorem (Fonseca and Leoni 2007, p. 159) implies that most functionals we care about in this context can be represented as some kind of integration. See also Buttazzo 1989 for some additional facts about integral representation of functionals.

[7] This should not be viewed as a historical account: we are combining Lagrange's original ideas with some more modern concepts and terminology here.

sufficiently small. For example, in the brachistochrone problem, the set of admissible variations is

$$\mathcal{V} = \left\{ v \in C^2\left([a,b]\right) : v(a) = 0,\ v(b) = 0 \right\}.$$

This ensures that the boundary conditions are still satisfied whenever we add some $v \in V$ to any $y \in \mathcal{A}$. Figure 1.2 depicts one potential choice of a function y and an admissible variation v.

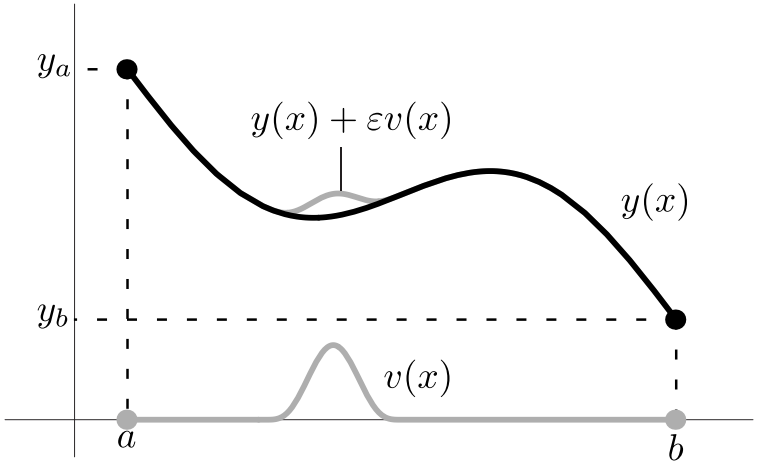


Figure 1.2: We think of a variation as determining the direction in which we perturb the function. Image adapted from Hrusa and Foss 2002.

Now we can define a notion of directional derivative that is completely analogous to the familiar definition of a derivative:

$$\delta J[y; v] \coloneqq \lim_{\varepsilon \to 0} \frac{J[y + \varepsilon v] - J[y]}{\varepsilon} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon} J[y + \varepsilon v]\Big|_{\varepsilon = 0}.$$

When this limit exists, we say $\delta J[y; v]$ is the *Gâteaux derivative* of J at y, in the direction of v. Now suppose that the Gâteaux derivative in the direction of v is nonzero for some y. Then we could modify y by pushing it slightly in the direction of v and either increase or decrease the value of J (see Figure 1.3). Thus, y could not be a local extremum of J. So, like in elementary calculus, there is a simple necessary condition for local extrema in terms of directional derivatives: If y minimizes (or maximizes) the functional J, then $\delta J[y, v] = 0$ for every admissible variation $v \in \mathcal{V}$.

The central result of Lagrange's variational approach was deriving a differential equation from the necessary condition listed above. This argument leads to the *Euler-Lagrange Equation:* If $\hat{y}$ is

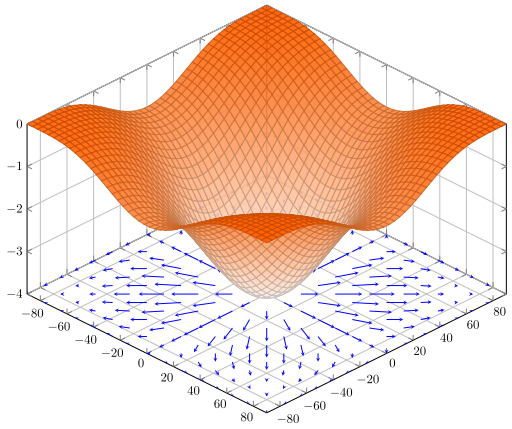


Figure 1.3: In the finite dimensional case, extreme values of smooth functions must occur at points where the directional derivative in every possible direction is 0. Image credit: MartinThoma, WikiMedia Commons.

a minimizer of J over the set $\mathcal{A}$, then it must satisfy

$$\frac{\partial f}{\partial \hat{y}} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial \hat{y}'} = 0. \tag{1.3}$$

Thus, instead of getting an algebraic equation and solving for a point line in the one dimensional problem, we have a differential equation and solve for a function. We skip the details of the proof since they are not needed for the main results of this work. But the derivation is not very difficult to follow, especially in the particularly nice case considered above, where all our functions live in $C^2\left([a,b]\right)$. Many books contain the relevant details; see for example Logan 1977. It is worth noting, however, that one of the key tools used in the proof is integration by parts. Roughly speaking, as long as we are working with functions where integration by parts makes sense, the argument goes through. Thus, it is more natural to study variational problems in a space of functions much larger than $C^2\left([a,b]\right)$, called *Sobolev space*. Again, we skip those details here. See Evans 1998 or Leoni 2009 for more information.

Now that the basic concepts have been introduced, we look at some applications of variational problems to mechanics to help motivate further study.

1.1.2 Connections with Mechanics

The development of mechanics is deeply interlinked with the development of variational methods. Indeed, as Cornelius Lanczos correctly noted: "There is hardly any other branch of the mathematical sciences in which abstract mathematical speculation and concrete physical evidence go so fully together and complement each other so beautifully" (Lanczos 1949, p. vii).

Intuitively, we know that nature often 'takes the path of least resistance.' This concept is reflected across many areas of physics: soap bubbles generally for a shape minimizing surface area, water flows down the hill rather than up, elastic bodies minimize internal energy, and so on. Even before the brachistochrone problem, minimization problems of the sort studied by the calculus of variations had found their way into the physical sciences. In 1662, Fermat gave a derivation of Snell's Law, which models the refraction of light as it enters a new medium, based on the belief that "nature operates by means and ways that are 'easiest and fastest'" (Goldstine 1980, p. 1).[8] This is now known as *Fermat's Principle*: light travels along paths that minimize total time. This principle was fundamental to understanding optics at the time, and it serves as the first example in the European tradition of a variational principle forming the basis for a physical theory. Over the next few centuries, variational principles of this sort massively influenced advances in physics.

The history here is complicated but fascinating; only a very rough outline is presented here. For historical detail, Dugas 1955 is a classic text on mechanics, while Goldstine 1980 focuses specifically on the development of the calculus of variations. For a wonderful overview that includes historical and philosophical insight in addition to the details of the physical theories, see the classic text Lanczos 1949. For a more mathematical approach, Levi 2014 presents the intuition and numerous examples, while Arnold 1989 gives a more rigorous account of the usual physical theories.

In Newtonian mechanics, the fundamental objects of study are vectors, like momentum or force. The vectorial approach to mechanics surely has it strengths, but it is not the only way to classical mechanics. Leibniz introduced a notion of *vis viva*, or 'living force,' which is what we know today

[8] His method was developed as far back as 1629, long before Newton or Leibniz formulated modern calculus. However, Fermat's technique is not so different from more recent methods; Goldstine gives the full argument, and notes "Essentially what Fermat does when he wishes to maximize or minimize a function f of E is to calculate $f'(0)$ and set this value to zero" (Goldstine 1980, p. 3). It was Fermat's method that inspired John Bernoulli's solution to the brachristochrone.

as kinetic energy (only missing the factor of $\frac{1}{2}$, which turns out to be significant). Around the same time, people were studying *work*, which is a line integral of force along the path of motion, and is essentially equivalent to potential energy. These two types of energy formed the foundation of an alternative approach to Newton's force and momentum vectors.[9] Let's look at some of the most important developments in this tradition.

Building on the work of John Bernoulli and others,[10] in 1743 the French mathematician d'Alembert gave a formulation of the *Principle of Virtual Work* that applied to both static and dynamic problems (Lanczos 1949, Parts III and IV). We skip the details, but here's the central idea: if a particle moves along a trajectory y, we consider *virtual displacements*, which correspond to the variations $y(t) + \varepsilon v(t)$ introduced in the previous section. The Principle of Virtual Work states that the net work done along any of these perturbed paths must be 0. Since work is an integral of force along a path, this principle takes a form like Equation 1.2. Over the next several decades, Maupertuis, Euler, Lagrange, Hamilton, Jacobi and Gauss all formulated versions of essentially this same idea. (Many of these turn out to be equivalent in many contexts; see for example page 93 of Arnold 1989.) The most common version involves a quantity known as *action*. There is no simple intuition for what exactly action measures, but in most cases it can computed by integrating the difference between kinetic and potential energy:

$$S[y] = \int_a^b L(t, y, y')\,\mathrm{d}t = \int_a^b T - U\,\mathrm{d}t, \tag{1.4}$$

where T is kinetic energy, generally $\frac{1}{2}my'^2$, U is potential energy, which usually only depends on y, and the integrand L is called the *Lagrangian*. The central variational principle of classical mechan-

[9] Although it seems so familiar now, there was actually a great deal of controversy around the notion of energy. See Part Three, Chapter Two of Dugas 1955.

[10] As usual in the history of the sciences, the standard stories focus on the contributions of men. Thus, the significance of Émilie du Châtelet has often been overlooked. She was the first to translate Newton's *Principia* into French in 1759, and her version is still the standard used today. It is hard to overstate the significance of bringing Newton's work to France. Additionally, she gave three compelling arguments in favor of using kinetic and potential energy, one philosophical, one empirical and one mathematical. These developments were crucial for the progress made later by d'Alembert and Maupertuis (Reichenberger 2018). In particular, she helped develop the concept of conservation of energy, which is absolutely essential to our modern understanding of mechanics. Independent of her contributions to physics, du Châtelet is quite an interesting character in her own right. She corresponded with John (the second one) Bernoulli and Euler, was taught by two of John (the first one) Bernoulli's most notable students, and was the mistress of Voltaire. Her work in philosophy is significant as well, see her entry in the *Stanford Encyclopedia of Philosophy*.

ics is the *Principle of Stationary Action*: a particle will always move along a trajectory y satisfying $\delta S[y; v]$ for all admissible variations v. There are various theories built on this fundamental principle. We focus on *Lagrangian Mechanics*, developed around 1789, about 100 years after Newton's *Principia*.

Since the preceding discussion has been rather vague, let's look at the Principle of Stationary Action in a simple special case. Suppose that the Lagrangian for a given particle is

$$L(t, y, y') = \frac{1}{2}my'^2 - U(y),$$

for some potential energy function U and a constant $m > 0$. Since the particle moves on a path of stationary action, its trajectory must satisfy the Euler-Lagrange Equation:

$$\frac{\partial L}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial y'} = 0.$$

Recall from elementary physics that for conservative forces,

$$F = -\frac{\partial U}{\partial y}\,.$$

Thus, the partial derivative of L with respect to y is just the force F acting on the particle. Computing the other derivatives, the Euler-Lagrange Equation implies

$$F - \frac{\mathrm{d}}{\mathrm{d}t}(my') = 0 \implies F = my''.$$

Thus, we have recovered Newton's Second Law: force is mass times acceleration.

This may seem like a lot of work just to end up at Newton's Second Law, which was of course well-known by the time these variational principles were formulated. So it is natural to wonder why all the additional work of introducing the calculus of variations is worthwhile. In fact, there are a number of advantages to Lagrangian mechanics. First, note that the fundamental object of the theory, the Lagrangian L, is a scalar. Contrast this with the standard Newtonian approach, where the basic concepts are necessarily vectors. Scalars are usually easier to work with, so for some

problems the Lagrangian approach is computationally easier. Similarly, Lagrangian mechanics is usually set up using *generalized coordinates*. This means that constraints can be viewed as reducing the number of dimensions of the problem. For example, consider a bead moving along a wire in three-dimensional space. Newton's formulation requires computing a normal force at each instant, which pushes on the bead and keeps it on the wire. But Lagrangian mechanics can describe the same situation quite simply as a one-dimensional problem, viewing the constraints as removing degrees of freedom instead of adding new forces. Additionally, Newtonian mechanics is best suited for individual particles, while the variational methods extend quite naturally to continuous problems as well (fields, fluids, elastic bodies, etc).

The Lagrangian approach to coordinates turns out to be especially significant. The central ideas are not committed to the same sort of Euclidean space needed for Newtonian mechanics. Thus, the techniques of Lagrangian mechanics carry over to modern physics in a way Newtonian mechanics does not. For example, it is possible to formulate the Lagrangian function in a *covariant* fashion, which is one of the requirements of Einstein's relativity (Lanczos 1949). Even in quantum field theory, physicists often specify a particular model by simply writing down its Lagrangian. This can make the symmetry of the problem more apparent, as well. This leads to another significant advantage of the variational approach to mechanics: *Noether's Theorem*.

Emmy Noether was one of the sharpest mathematicians of the early 20th Century. She worked closely with Klein, Hilbert and Einstein, making significant contributions across multiple fields of mathematics. In 1918, she published *Invariante Variationsprobleme*, proving two theorems on the certain symmetries of variational problems (in particular, regarding the invariance of the variational problem under the action of a Lie group with a finite number of infinitesimal generators). Essentially, her work established a one-to-one correspondence between the symmetries and the conservation laws of a given problem. For example, Euclidean space is isotropic (the same in all directions), so Lagrangians in this setting are often invariant under rotations. This not so surprising: the outcome of the experiment doesn't usually depend on whether the lab is facing East or West. Noether's Theorem shows that the law of convservation of angular momentum follows as an immediate consequence of invariance under rotations. It is difficult to overestimate the impact of

this work on the progress of theoretical physics in the following years. As relativity was developed and Klein's *Erlangen Program* advanced, the role of symmetry became absolutely essential to our understanding of space and time. And conservation laws are powerful tools, both mathematically and philosophically, for understanding problems in dynamics.

Formulating the fundamental principles of mechanics in terms of variational problems was also crucial to Noether's insight. David Hilbert was especially enthusiastic about the power of this formulation. In 1921 Felix Klein wrote to Wolfgang Pauli, a leading physicist, accusing Hilbert of "a fanatical belief in the variational principles, that they can explain the reality of nature by means purely of mathematical considerations" (Kosmann-Schwarzbach 2010, p. 27). This commitment to variational principles as the true, fundamental laws of nature can also be seen in Hilbert's work on general relativity: he gave a formulation of Einstein's theory in terms of a variational principle, now known as the Einstein-Hilbert Action.[11] And the Principle of Stationary Action, along with Noether's results, led Richard Feynman to give his path integral formulation of quantum mechanics, part of what earned him a Nobel Prize in 1965. The insights gained from the calculus of variations have been absolutely essential to nearly all of the major advances of modern theoretical physics.

For more information on the deep connections between modern physics and Noether's work, see Kosmann-Schwarzbach 2010. Logan 1977 discusses some of the relevant mathematical theory involved. For a physics perspective on the importance of symmetry, see Schwichtenberg 2015. The path integral formalism is an especially fascinating extension of the Principle of Stationary Action and has deeply influenced math as well as physics. See Johnson and Lapidus 2002 for a detailed overview.

Although we have focused on applications of variational principles to problems in classical mechanics, this is not the only reason to care about the calculus of variations. Applications have

[11] Although it is usually credited to Einstein, there are thorny historical debates about who exactly should get credit for the various parts of general relativity. Poincare, Riemann, Minkowski, Hilbert, Weyl and others all made contributions in one way or another. In fact, in 1923 Hilbert essentially claimed that he found the fundamental field equations of general relativity, in the form of a variational principle, prior to Einstein: "the final result of Einstein's latest work amounts to a Hamiltonian principle that is similar to the one that I had originally proposed. Indeed, it might be the case that the content of this latest Einsteinian theory is completely equivalent to the theory originally advanced by myself" (Majer and Sauer 2005). He accused Einstein of taking a "colossal detour" for 8 years before finally arriving at the same equation Hilbert had suggested, which Hilbert saw as a "a beautiful confirmation" of his own work. It is not totally clear if Hilbert is correct about this or if he was making self-serving comments with the benefit of hindsight and a bit of revisionist history. See Majer and Sauer 2005 for discussion.

been found in economics, fluid dynamics aerodynamics, traffic flow, image processing, machine learning and thin-film physics (Hardt 2004; Troutman 1996; Calder 2022; Evans 1998). In the 1960s and on, variational methods were applied to problems in nonlinear elastic bodies with great success (see, for example, Ball 1976). Perhaps the most extensive usage of variational techniques is found in the modern field of *optimal control theory*. This is a vast subject in engineering involving solving differential equations coming from control problems while simultaneously optimizing some objective function. For example, programming a self-balancing robotic arm while minimizing the amount of battery power used. For details and examples connecting optimal control and the calculus of variations, see Troutman 1996 and Levi 2014. A classic text with numerous examples at a slightly more advanced level is Cesari 1983.

There is plenty of motivation for studying variational problems from within mathematics, independent of the many applications mentioned above. The tools of variational calculus are essential to modern approaches to partial differential equations, through Lax-Milgram theory and Dirichlet's Principle for example (Evans 1998). There are also connections to problems in differential geometry, where *geodesics* (paths of minimal distance) play a fundamental role. For a lighthearted and brief discussion of connections to differential geometry, dynamical systems and Morse theory, see Hardt 2004, p.13-36. A much more technical discussion of connecting variational methods to other branches of mathematics can be found in Struwe 2008. Extending the techniques of calculus and analysis to functionals like 1.2 also motivated the work of Weierstrass, Hilbert, Fréchet, Lebesgue and others to study abstract metric spaces and function spaces, laying the groundwork for modern day functional analysis.

Analytical mechanics mostly relies on the tools of the calculus of variations to find a minimizer of some integral. As we noted in the previous section, questions of existence and uniqueness also arise in the context of variational problems. In the next section we explore more recent work in finding sufficient conditions for the existence of a minimizer.

1.1.3 Modern Developments and The Direct Method

Throughout most of the history discussed in the previous section, it was simply assumed that solutions to the variational problems always existed. Especially for nonnegative functionals applied to problems in nature, there was essentially no question that a unique minimizer would always exist. That is, it was assumed that functionals always attained their absolute minimum. The first counterexample was found by Weierstrass. (This discussion is based on Goldstine 1980, p. 391, and a related case is discussed in Example 4.6 of Dacorogna 2008.) He showed that the functional

$$W[y] = \int_{-1}^{1} x^2 y'^2 \, \mathrm{d}x \tag{1.5}$$

does not attain its minimum over the admissible class

$$\mathcal{A} = \left\{ y \in C^1\left([-1,1]\right) : y(-1) = 0, y(1) = 1 \right\}.$$

Since $y(-1) \neq y(1)$, note that y cannot be constant and so y' cannot be identically zero. However, Weierstrass showed that the functional J can be made arbitrarily close to 0 using the following functions:

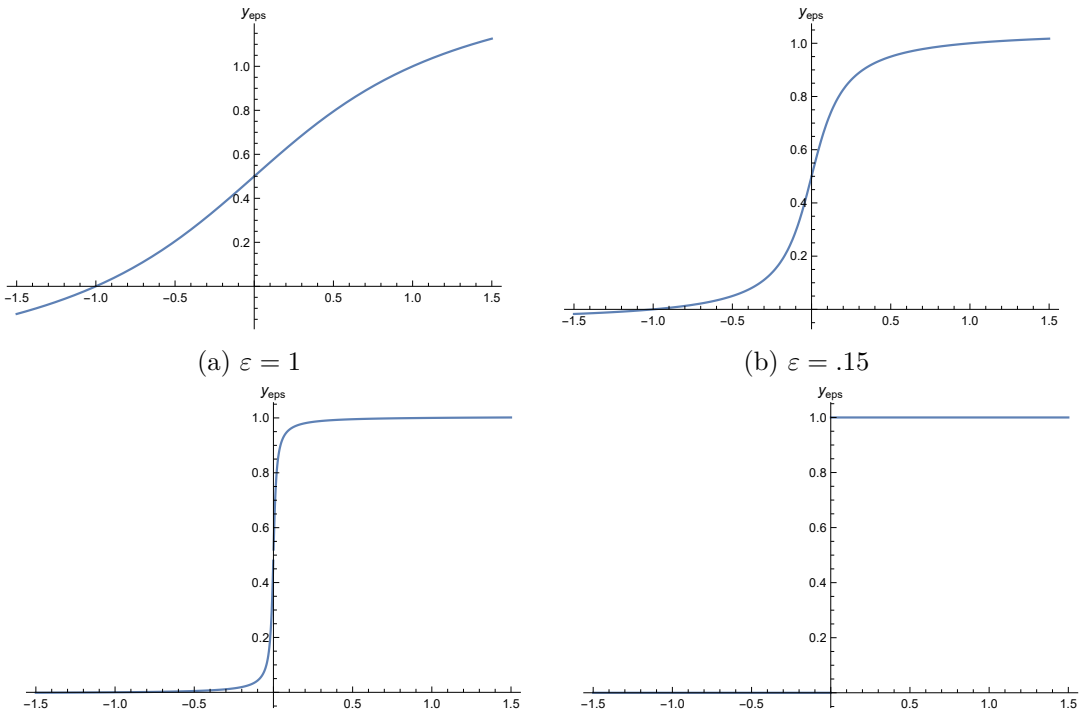
$$y_\varepsilon(x) = \frac{1}{2} + \frac{1}{2} \cdot \frac{\arctan\left(x/\varepsilon\right)}{\arctan\left(1/\varepsilon\right)}.$$

For every $\varepsilon > 0$, the function y_ε is indeed in the admissible class. But as $\varepsilon \to 0$, $J[y_\varepsilon] \to 0$. Figure 1.4 shows plots for two choices of ε.[12]

As this was the era of increased rigor throughout mathematics, this example motivated the community to investigate questions like existence and uniqueness of minimizers more deeply. This is reflected in Hilbert's famous lecture, "Mathematical Problems," delivered to the International Congress of Mathematicians in 1900. Several of his problems concern the calculus of variations; his final 23rd problem simply calls for "further development of the methods of the calculus of variations" (Hilbert 1902).

Perhaps the first major result in establishing the existence of a minimizer is also due to Weier-

[12] To better see the limiting behavior as ε approaches 0, an interactive plot is hosted at https://www.wolframcloud.com/obj/michael.pieper/Published/WeierstrassCounterexample.nb.



(a) $\varepsilon = 1$ (b) $\varepsilon = .15$

(c) $\varepsilon = 0.015$ (d) $\varepsilon = 1 \times 10^{-9}$

Figure 1.4: For any $\varepsilon > 0$, y_ε is smooth and satisfies the boundary conditions. As $\varepsilon \to 0$, this family of functions approaches a step function with a discontinuity at 0. This step function makes the functional W equal to 0, but is not in the admissible class.

strass. His theorem illustrates some of the crucial elements of the direct method needed later, so we present the theorem and a proof in full. But first, we need to introduce a few definitions.

Recall that in the familiar setting of functions from the real line to itself, one way to ensure that a function attains its minimum is to require (1) that the domain is compact and (2) that the function is continuous. These assumptions can be weakened if we introduce the following concept.

Definition 1.1.1: Let X be any topological space, and f be some function from X into $\mathbb{R} \cup \{\pm\infty\}$. We say that f is **lower semi-continuous** if the following holds for all $x_0 \in X$:

$$f(x_0) \leq \liminf_{x \to x_0} f(x).$$

Clearly, if f is continuous it is also lower semi-continuous. Intuitively, we think of the above

condition as stating that any time f has a discontinuity, any jumps can only decrease the function's value. This is illustrated in Figure 1.5.

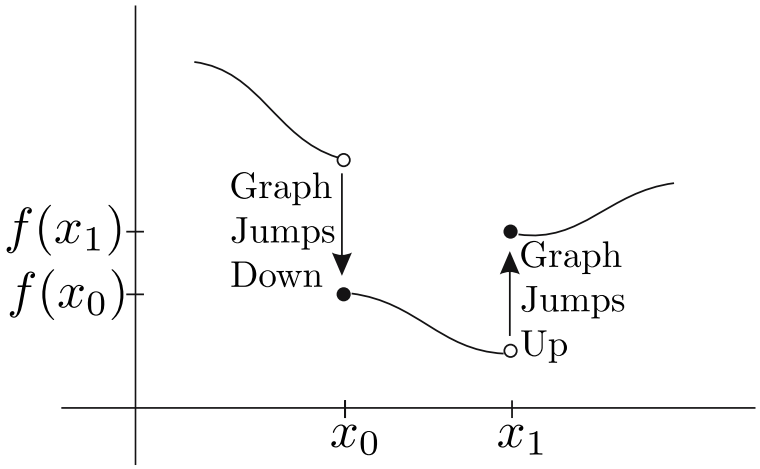


Figure 1.5: Although f is discontinuous at x_0, it still satisfies $f(x) < \liminf_{x \to x_0} f(x)$. Thus, f is lower semi-continuous at x_0. However, at x_1 the function value suddenly increases, so f is not lower semi-continuous there. Image adapted from Zabarankin and Kurdila 2005.

An equivalent characterization will be useful in proving Weierstrass' theorem:

Lemma 1.1.2: A function $f : X \to \mathbb{R} \cup \{\pm\infty\}$ is lower semi-continuous if and only if, for every $t \in \mathbb{R}$, the set

$$\{x \in X : f(x) \leq t\}$$

is closed in X.

For a proof of this fact, see Proposition 3.4 of Fonseca and Leoni 2007. The significance of lower semi-continuity is demonstrated by the following theorem of Weierstrass which can be found as Theorem 3.6 in Fonseca and Leoni 2007 or 7.3.1 of Zabarankin and Kurdila 2005.

Theorem 1.1.3 (Weierstrass)**:** Let X be a topological space, K be some compact subspace of X, and $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be lower semi-continuous. Then there exists some $x_0 \in K$ such that

$$f(x_0) = \inf_{x \in K} f(x) = \min_{x \in K} f(x) = m.$$

Proof. Suppose for the sake of contradiction that the infimum is *not* attained. Then for every $x \in K$, we can find some $m_x \in \mathbb{R}$ such that $m < m_x < f(x)$. For each $x \in K$, define the following set:

$$U_x \coloneqq \left\{ y \in L : f(y) > m_y \right\}.$$

First, note that our choice of m_x and definition of U_x immediately imply that, for any $x \in K$, x is in U_x. Lemma 1.1.2 implies that the set $\left\{ y \in K : f(y) \leq r \right\}$ is closed for any $r \in \mathbb{R}$. Hence, the set $\left\{ y \in K : f(y) > r \right\}$ is open, since it is the complement of a closed set. Each of our sets U_x are precisely of this form, so the collection $\mathcal{U} \coloneqq \{U_x : x \in K\}$ forms an open cover of K. Since K is compact, there must be a finite subcover of $\{U_x\}_{x \in K}$. That is, there exist some $U_{x_1}, U_{x_2}, \ldots, U_{x_n}$ such that

$$K \subseteq \bigcup_{i=1}^{n} U_{x_i}.$$

Each set U_{x_i} corresponds to some real number m_{x_i}. Since there are only finitely many, we can take a minimum:

$$m_0 \coloneqq \min_{i \leq n} m_{x_i}.$$

Now choose any arbitrary $y \in K$. Since the collection $U_{x_1}, \ldots, U_{x_n}$ covers K, there is some set U_{x_k} that contains y. Then, by definition,

$$f(y) > m_{x_k} > m_0.$$

Therefore, m_0 is a lower bound for $\left\{ f(y) : y \in K \right\}$. But m_0 is one of the numbers we chose satisfying $m_0 > m = \inf_{x \in K} f(x)$. This means we have found a lower bound greater than the infimum, which is a contradiction. □

The result above shows that a compact domain and a lower semi-continuous functional are sufficient for the existence of a minimizer. However, in the contexts we are interested in, these two conditions are too restrictive. This motivates moving to a weaker topology, where the sufficient conditions are more easily met. Here is the general strategy, as described in Section 3.2 of Fonseca and Leoni 2007.

1. Start with a minimizing sequence $\{x_n\}_{n=1}^{\infty}$ in a space X:

$$\lim_{n\to\infty} f(x_n) = \inf_{y\in X} f(y).$$

As long as degenerate cases are ruled out, such a sequence will always exist.

2. By moving to a weaker topology on X, establish sequential compactness of the domain. Sequential compactness then implies the existence of a subsequence $\left\{x_{n_k}\right\}_{k=1}^{\infty}$, converging (in the weaker topology) to some $x_0 \in X$. Also, since this is still a subsequence of a minimizing sequence, we have

$$\lim_{k\to\infty} f(x_{n_k}) = \inf_{y\in X} f(y).$$

3. Prove that f is lower semi-continuous in the weaker topology. This implies

$$\lim_{k\to\infty} f(x_{n_k}) \geq f(x_0).$$

4. Since $x \in X$, the definition of infimum immediately implies

$$f(x_0) \geq \inf_{y\in X} f(y).$$

Combining the above equations, we have

$$\inf_{y\in X} f(y) = \lim_{k\to\infty} f(x_{n_k}) \geq f(x_0) \geq \inf_{y\in X} f(y).$$

Thus, x_0 minimizes f over X.

This technique is called *The Direct Method*, and it was built up by Tonelli, Morrey, and many others in the decades following Weierstrass and Hilbert.[13] The two essential ingredients in modern formulations of the direct method are a *convexity* condition, and a growth condition called *coercivity*. An intuitive sketch of these concepts is below. For a very readable introduction to convex analysis

[13] See Remarks 3.16 and 3.25 of Dacorogna 2008 and the bibliographic notes at the end of Chapter 2 in Cesari 1983 for a list of contributors.

and the tools needed for the direct method, see Chapter 7 of Zabarankin and Kurdila 2005 or Chapter 4 of Calder 2022. Section 8.2 of Evans 1998 introduces the relevant concepts and goes on to connect the direct method to problems in partial differential equations. A more thorough discussion, especially concerning different versions of the convexity condition, can be found in Chapters 3 and 4 of Dacorogna 2008. The text Cesari 1983 is a classic, and Fonseca and Leoni 2007 gives a more modern treatment. Sometimes counterexamples can be more illuminating than cases where everything goes smoothly. The clear and concise work by Serovaĭskiĭ 2004 gives several natural problems where the convexity or coercivity conditions are not met and shows that the problems are ill-posed. Also, section 6 of this book discusses a problem lacking continuous dependence on parameters (sometimes called Hadamard well-posedness).

Convexity is the key tool needed to ensure lower semi-continuity in the weak topology. In an introductory calculus course, convexity is usually described in terms of the second derivative test: f is convex if $f'' \geq 0$. A similar idea can be captured in terms of the first derivative: f is convex if the graph of f lies above its tangent line ($f(x) \geq f(x_0) + f'(x_0)(x - x_0)$). However, we will be working in spaces of functions where the derivative may not be defined. Thus, we use a more general notion of convexity. First, we need to introduce the notion of a convex set.

Definition 1.1.4: Let V be a vector space. We say that $X \subseteq V$ is a **convex set** if for all $x, y \in X$ and all $t \in [0, 1]$,

$$tx + (1 - t)y \in X.$$

The expression $tx + (1 - t)y$ in the previous definition should be thought of as a parametric equation for a line, connecting x and y. This is illustrated in Figure 1.6.

We apply a similar idea to functions: f is convex if every secant line connecting two points on the curve lies below the curve itself. Note that this definition only makes sense if the domain of f is a convex set.

Definition 1.1.5: Let X be a convex set and let $f : X \to \mathbb{R}$ be given. We say f is **convex** if for

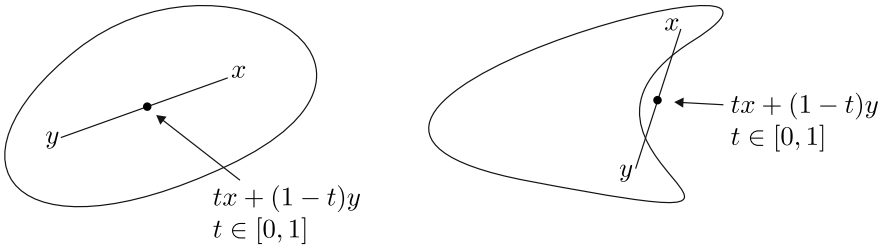


Figure 1.6: The set on the left is convex. However, the set on the right is not, since there is a point on the line connecting x and y that outside the set. Image adapted from Zabarankin and Kurdila 2005.

every $x, y \in X$ and every $t \in [0,1]$,

$$f\left(tx + (1-t)y\right) \leq tf(x) + (1-t)f(y).$$

This concept is illustrated in Figure 1.7. This figure also illustrates why it is not so surprising that convexity would be associated with existence of a minimizer, and why *strict* convexity is related to uniqueness of minimizers.[14]

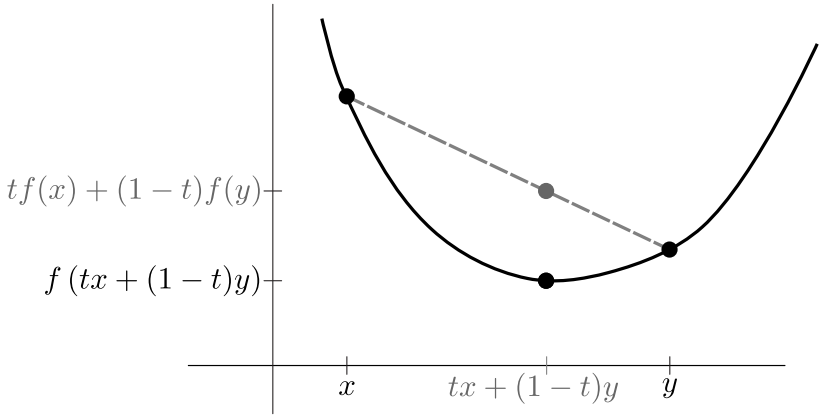


Figure 1.7: This function is convex, since its graph lies below any secant line.

In many of the relevant cases, convexity implies lower semi-continuity in both the strong and weak topologies.[15] (In some cases convexity and lower semi-continuity even turn out to be equiva-

[14]There is a lot more to say about convexity. In addition to the citations a few paragraphs above, Rockafellar 1970 is a standard reference. We focus on one dimensional problems, but different convexity conditions (quasiconvexity, polyconvexity) are more natural choices for problems in high dimensions.

[15]See Sections 2.18 and 2.19 of Cesari 1983, Proposition 4.26 of Fonseca and Leoni 2007, Theorems 7.2.4-7.2.6 of Zabarankin and Kurdila 2005, Section 3.2 of Dacorogna 2008, and Theorem 4.22 of Calder 2022. Also, we are leaving out relevant details from functional analysis here. Related results on the relevant Banach spaces can be found in, for

lent.) Through methods involving *relaxation*, the direct method can be extended even to functions failing to meet our convexity condition, but we do not consider such cases here (see, for example, Buttazzo 1989).

Now that we have a simple condition to ensure lower semi-continuity, we turn to the other key component in the direct method: coercivity. This is a growth condition that ensures a certain kind of boundedness. The term "coercive" is often used to refer to a certain type of property, as opposed to referring to a single fixed condition. The exact meaning varies based on context; here are a few examples of coercivity conditions:

- A functional f defined on a normed vector space $(X, \|\cdot\|)$ is **coercive** if

$$\|x\| \to \infty \implies \big|f(x)\big| \to \infty.$$

- A functional $f : X \to [0, \infty)$ is **coercive** if there is some $\lambda > \inf_{x \in X} f(x)$ such that $\big\{x \in X : f(x) < \lambda\big\}$ is sequentially precompact.

- A functional $f : \mathbb{R}^3 \to \mathbb{R}$ is **coercive** if it satisfies

$$f(x, y, z) \geq c_1 |z|^p - c_2,$$

  for some $c_1, c_2 > 0$ and $p > 1$.

- A functional $f : \mathbb{R}^3 \to \mathbb{R}$ is **coercive** if it grows superlinearly with respect to its third argument:

$$\lim_{z \to \infty} \frac{f(x, y, z)}{|z|} = \infty, \forall x, y.$$

These are all slightly different ways of capturing the same idea. One way of recognizing the significance of coercivity is through the following theorem. Intuitively, it states that when looking for minimizers of f we don't need to search the entire domain. It is sufficient to restrict out attention to some relatively closed and bounded subset.

example, V.4.2 and V.13.1 of Conway 2019 or Chapter V of Yosida 1980. We cite the relevant details as needed in Sections 2.3 and 3.2.

Theorem 1.1.6: Let X be a normed vector space and $M \subseteq X$ be closed in the norm topology. If $f : M \to \mathbb{R}$ is coercive, there exist some $z \in M$ and $R \in \mathbb{R}$ such that

$$\inf_{x \in M} f(x) = \inf \left\{ f(x) : x \in M \cap \overline{B_R(z)} \right\},$$

where $\overline{B_R(z)}$ is the closure of a ball centered at z with radius R.

For a proof, see Theorem 7.3.1 of Zabarankin and Kurdila 2005. For some intuition, consider a simple function from $\mathbb{R}$ to $\mathbb{R}$. If we require that $f(x)$ grows rapidly as $|x| \to \infty$, we end up with a picture like in Figure 1.8. As the graph suggests (and the theorem confirms), any minimizing sequence of inputs $\{x_j\}_{j=1}^{\infty}$ will eventually enter a bounded set and stay there.

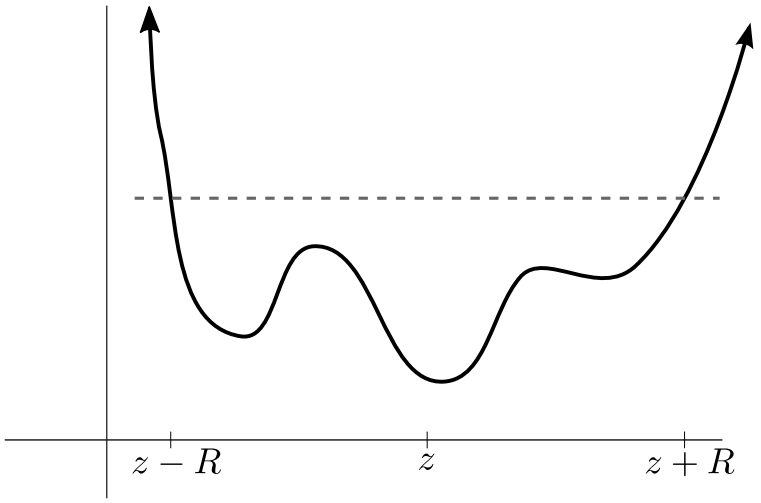


Figure 1.8: Because f eventually grows very quickly, only finitely many terms of any minimizing sequence lie outside $[z - R, z + R]$.

The third and fourth characterizations of coercivity give an alternate intuition. Suppose our function f depends on x, u, and u'. If there is a point where u changes suddenly, like the point $x = 0$ in Weierstrass' Examples (see Figure 1.4), then the derivative u' will be very large at that point. Coercivity implies that the function f will penalize these jumps harshly; if the slope u' is very large, the functional f is even larger. We see that Weierstrass' functional fails this condition: at $x = 0$, the derivative y' can grow arbitrarily large without increasing the functional W (Equation 1.5). Thus, the Direct Method does not apply to W and there is no guarantee the minimization

problem is well-posed.

We give our specific coercivity condition later, and show its precise implications in Section 3.2. The key idea is that it helps establish a uniform bound on the minimizing sequence. We can then apply some results from functional analysis to see that the minimizing sequence in contained in a weakly sequentially compact set, and thus there is a weakly convergent subsequence. Since convexity gives lower semi-continuity even in this weak topology, the direct method described above can go through.

Aside 1.1.7: There is another subtlety here that would be easily overlooked. The choice of admissible class $\mathcal{A}$ can change the behavior of the problem in ways that may not be expected. First, consider the function $q(x) = (x - \sqrt{2})^2$. The absolute minimum value is 0, and this minimum occurs at $x = \sqrt{2}$. Although there is no rational number $r \in \mathbb{Q}$ such that $q(r) = 0 = \min_{x\in\mathbb{Q}} q(x)$, we still have

$$\inf_{x\in\mathbb{Q}} q(x) = \inf_{x\in\mathbb{R}} q(x).$$

Because the rationals are dense in the reals, we can approximate $\sqrt{2}$ with arbitrarily high precision, and thus make $q(r)$ as close to 0 as we like. Intuitively, this behavior should carry over to nice looking functionals. It turns out this does not always hold; sometimes looking at a dense subclass of the admissible class is not enough to determine the true infimum of the functional. This is called *Lavrentiev's Phenomenon*.

Although it was discovered in the early 1900s, not too long after Tonelli gave his famous theorem, Lavrentiev's Phenomenon did not receive much attention until around the 1980s. In this era, Mizel, Ball, Heinricher and others found fairly natural examples of the phenomenon, renewing interest in the problem. The literature on the topic is now quite extensive. See Section 4.7 of Dacorogna 2008, the survey Buttazzo and Belloni 1995, or the dissertation Foss 2005 for more information. We present a single example, from Heinricher and Mizel 1988.

Let $\mathcal{A}^1 = \left\{u \in W^{1,1}\left([0,1]\right) : u(0) = 0,\ u(1) = 1\right\}$, which we think of as the absolutely continuous functions on $[0,1]$. And let$\mathcal{A}^\infty = \left\{u \in W^{1,\infty}\left([0,1]\right) : u(0) = 0,\ u(1) = 1\right\}$, the Lipschitz

functions on that interval. It is well known that $\mathcal{A}^\infty$ is dense in $\mathcal{A}^1$. Then consider the functional

$$J[u] = \int_0^1 (u^2 - x)^2 (u')^6 \, \mathrm{d}x.$$

The function $\hat{u} = \sqrt{x}$ is in $\mathcal{A}^1$, and satisfies $J[\hat{u}] = 0$. It is not too hard to see this is the smallest possible value of J, so $\hat{u}$ is a minimizer. Since we can approach $\hat{u}$ with functions in $\mathcal{A}^\infty$, we might expect that the infimum of J over the set $\mathcal{A}^\infty$ would be 0. However, it can be shown that

$$\inf_{u \in \mathcal{A}^\infty} J[u] \geq \frac{9}{2^{10}} \left(\frac{3}{5} \right)^6 > 0.$$

$\Diamond$

To summarize the main points of this section, coercivity and convexity together give precisely the conditions need to extend Weierstrass' theorem to a wider class of functionals. Because most of the argument in the direct method can be formulated in terms of abstract normed vector spaces, the same ideas can be applied to a wide variety of cases. As we saw in the previous sections, most functions of interest are of the form

$$J[u] = \int_\Omega f(x, u, u') \, \mathrm{d}x.$$

Thus, most of the historical work was done in the setting of Sobolev space. Only a few adjustments are need to apply the direct method to functionals of the form

$$J[u] = \int_\Omega f(x, u, Du) \, \mathrm{d}x,$$

where D is some other operator. For example, Bourdin et al. 2013 applies the direct method where D is a Caputo derivative and Cueto 2021 applies it to the case of Riesz fractional gradients, in the context of fractional-order Sobolev space. As we see in the next sections, recent advances in nonlocal modeling give us a reason to move away from Sobolev space and to the larger Lebesgue spaces, L^p. Most of the work goes into setting up an operator that acts on L^p functions (a very

general setting), while maintaining the required structural properties as the classical derivative so that the direct method applies. We establish the relevant results in Chapters 2 and 3.

1.2 Nonlocal Models

In this section, nonlocal models are introduced. This is a very active area of research, and the literature is vast. The discussion is kept relatively brief, with the main goal of motivating the class of problems considered in Chapter 2.

1.2.1 Historical Introduction

Once the notion of an nth order derivative was introduced, it is natural to wonder if non-integer-order derivatives can be defined. Is there an operator T such that

$$T(Tu) = T^2 u = u'?$$

Is there a sensible way to interpolate between the space of continuous functions and the space of differentiable functions? Such questions were already discussed in the era of Newton and Leibniz (Odzijewicz et al. 2012).

Motivated by the Fundamental Theorem of Calculus, one approach to defining fractional order derivatives is to view them as inverse operators to fractional order integrals. Consider Cauchy's iterated integral formula:

$$\int_a^x \int_a^{t_1} \int_a^{t_2} \cdots \int_a^{t_{n-1}} u(t_n)\,\mathrm{d}t_n\,\mathrm{d}t_{n-1}\ldots\mathrm{d}t_1 = \frac{1}{(n-1)!}\int_a^x (x-t)^{n-1}u(t)\,\mathrm{d}t.$$

One might hope this can be extended to a non-integer number of applications of the integral by simply replacing the factorial with the Gamma function, so that the expression makes sense for non-integers. It turns out this works quite well, and leads to the Riemann-Liouville integral of order $\alpha \in (0,1)$:

$$I^\alpha u(x) \coloneqq \frac{1}{\Gamma(\alpha)}\int_a^x (x-t)^{\alpha-1}u(t)\,\mathrm{d}t. \tag{1.6}$$

From a purely abstract point of view, this is quite a simple generalization of the classic theory. However, just like in the calculus of variations, the problem was also motivated by a problem in physics: a generalization of the *Tautochrone Problem*. The tautochrone (from the Greek for "same time") is a curve such that the time for a heavy particle to fall to the bottom of this curve is independent of the upper starting point (Kot 2014, p. 4). Coincidentally, the solution to this problem is the cycloid; the exact same curve that solves Bernoulli's brachistochrone problem![16]

In 1823, Abel came across a version of Equation 1.6 in a context completely unrelated to iterated integrals. He was investigating variant of the tautochrone problem (Wazwaz 2011, p. 238): Given a function $f(y)$ of vertical position, Abel wanted to find a curve u in the vertical plane such that the particle takes exactly time $f(y)$ to move from the height y to position 0. He derived the following integral equation:

$$f(y) = \int_0^y \frac{u(t)}{\sqrt{y-t}}\,\mathrm{d}t.$$

If scaled by a constant $1/\Gamma(.5)$, this is just a special case of Equation 1.6. In modern terms, Abel's key idea was to solve for u by viewing the right side as an operator applied to u:

$$f(y) = I^{1/2}u(y)$$

By applying something like $I^{1/2}$ to both sides, he found

$$I^{1/2}f(y) = I^{1/2}I^{1/2}u(y) = I^1u(y) = \int_0^y u(t)\,\mathrm{d}t.$$

He could then isolate u using the Fundamental Theorem of Calculus:

$$u(y) = \frac{\mathrm{d}}{\mathrm{d}y}I^{1/2}f(y) = \frac{\mathrm{d}}{\mathrm{d}y}\frac{1}{\Gamma(1/2)}\int_0^y \frac{f(t)}{\sqrt{x-t}}\,\mathrm{d}t.$$

Now u is given in terms of an integral equation that is easier to work with.[17] Of course, the argument

[16] Kot 2014 also has a decent bibliography for more reading on the history here. Levi 2014 discusses the problem on page 55.

[17] This discussion is based on Luchko 2020. See Cartwright and McMullen 1978 for a short proof that this fractional integral is the "unique" operator satisfying certain desirable properties.

takes for granted the fact that $I^{1/2}I^{1/2} = I^1$, which is not immediately obvious. Developing such algebraic rules to formalize the idea of a fractional derivative (by defining it in terms of a fractional integral) took time, and a variety of approaches were found throughout the 1800s (Pooseh 2013). One important generalization of the Riemann-Liouville conception introduced above is the *Riesz potential*; for example, see Cueto 2021 or Malinowska et al. 2015 for more information.

1.2.2 Recent Developments

Although the concept of fractional calculus is quite old, there has been a recent spike in interest in the analysis of these fractional operators. This is partially due to progress in analytical tools generally, improvements in numerical methods and computational power, and new found applications to problems in engineering. For example, some recent connections to partial differential equations include Shieh and D. E. Spector 2018; Brasco and Lindgren 2017; L . Caffarelli and L. 2007; L. Caffarelli and Stinga 2016; L. Caffarelli and Silvestre 2009; Di Castro et al. 2016; Kuusi et al. 2014; Kreisbeck and Schönberger 2022; Ros-Oton 2016 and Šilhavỳ 2020. Studying Sobolev spaces in terms of these fractional operators has also become common; see the influential work Bourgain et al. 2001. For a nice overview of some recent progress on fractional Sobolev space, see Di Nezza et al. 2012, and Daoud and Laamri 2022 has a survey on fractional Laplacians.

In the classical case, the derivative of u at x depends only on the behavior of u on some infinitesimal neighborhood of x. For example, changing the value of $u(x+1)$ generally has no impact on the value of $u'(x)$. Thus, we call the classical derivative a local operator. Fractional derivatives and integrals are examples of *nonlocal operators*. For many of the formulations of a fractional derivative D^s, for $s \in (0,1)$, the value of $D^s u(x)$ depends on the behavior of u at some finite distance away from x. Generalizing from the fractional case, we can consider operators that transform the value of u at x in a way that is sensitive to points "far" from x. The most studied nonlocal operators are defined in terms of integrals, since this is a natural tool for collecting information about a function over some region.

Many laws of nature are formulated locally, often in terms of a differential equation (examples are abundant; some sort of differential equation appears in nearly every branch of science). Obviously

the differential operator has many strengths. However, there are some noticeable drawbacks:

1. the derivative cannot easily capture long-range interactions that occur throughout the natural world

2. derivatives are only defined for relatively nice functions, so they are not well-suited for contexts involving discontinuities

3. a more general theory, allowing for fractional order operators and a larger domain of functions, would be more satisfying.

These reasons are partially what motivated Silling to formulate *peridynamics*, a nonlocal framework for continuum mechanics (Silling 2000). In the study of continuous bodies, the local notions of stress and strain do not always adequately describe the dynamics of the body. The most obvious example is cases of fracture: if a sharp discontinuity develops in the body, the classical, local theories need to be modified to accurately model the break. Silling's idea was to replace the spatial derivatives in the classical formulation with integral operators. This idea turned out to be quite successful, and helped drive a flurry of research activity in nonlocal modeling.[18]

In addition to the success found by applying nonlocal models to continuum mechanics, a staggering array of other applications have been found. None of these applications will be discussed in detail, but here is a sampling of applications involving fractional derivatives and other nonlocal operators done in recent years.

- Nonlocal Diffusion (Andreu-Vaillo et al. 2010; Mellet et al. 2011; Vázquez 2012; Metzler and Klafter 2000; Du, Gunzburger, et al. 2012; S. Jafarzadeh et al. 2020; Vázquez 2017)
- Conservation Laws (Biler, Karch, and Woyczyński 2001)

[18] It's worth noting that Silling was not the first to introduce nonlocal elements into a theory of deformation and elastic bodies. See the introduction of Madenci and Oterkus 2013 for some discussion of the relevant history, as well as Evgrafov and Bellido 2019. Peridynamics is not necessary for the work developed later in this thesis, so we skip the details. Bobaru, Foster, et al. 2016 is an excellent resource on the topic; Chapters 1 through 4 are especially relevant to this project. The editorial Silling and Madenci 2019 is a nice overview of the field, and the dissertation Cueto 2021 also has some relevant discussion and many great references.

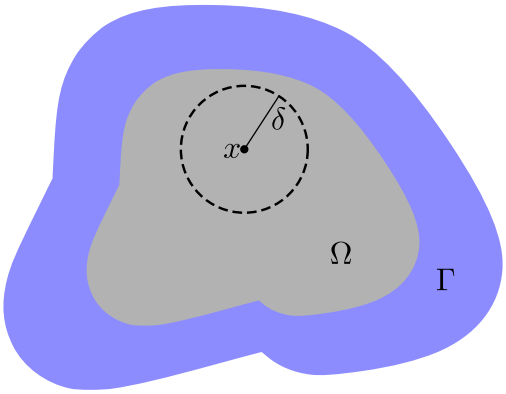


Figure 1.9: For the nonlocal models we consider, the behavior at x typically depends on the values at every point inside a ball of radius $\delta > 0$, centered at x. Thus, we usually decompose the relevant domain into the interior, Ω, and the collar, Γ. This sort of nonlocal operator will only be defined on Ω.

- Image Processing (Aubert and Kornprobst 2009; Gilboa and Osher 2009a; Gilboa and Osher 2007; Gilboa and Osher 2009b; Buades et al. 2010; Kindermann et al. 2005; Tadmor and Athavale 2009; Lou et al. 2010; Brezis and Nguyen 2018)
- Materials Science (Silvestre 2007; Weckner and Abeyaratne 2005)
- Stochastic Jump Processes (Barlow et al. 2009; Bass et al. 2010; Applebaum 2009; Du, Huang, et al. 2014; Burch et al. 2014; D'Elia et al. 2017; Meerschaert and Sikorskii 2019; Metzler and Klafter 2004)
- Machine Learning (Rosasco et al. 2010; Antil et al. 2020; Pang et al. 2020; Wei et al. 2020)
- Self-Organized Dynamics (Mengesha and D. Spector 2015)
- Particle Systems (Bodnar and Velázquez 2006)
- Multiscale Systems (Askari et al. 2008; Du, Engquist, et al. 2020)
- Traffic Models (Colombo et al. 2021)
- Biology (Carrillo and P. Fife 2005)
- Population Dispersal (Cortazar et al. 2007; Kao et al. 2010)
- Coagulation Models (Fournier and Laurençot 2006)
- Flocking Models (Mogilner and Edelstein-Keshet 1999)
- Prous Media Flow (Cushman and Ginn 1993; Dagan 1994)
- Water Waves (L. Caffarelli, Mellet, et al. 2012; Craig and Groves 1994; Craig and Nicholls 2000; Craig, Schanz, et al. 1997; de la Llave and Panayotaros 1996; Gächter and Grote 2003; Hu and Nicholls 2005; Uspenskii 1960)

- Subsurface Transport (Benson et al. 2000; Katiyar et al. 2020)
- Finance (Cont and Tankov 2004; Jakobsen and Karlsen 2005; Sabatelli et al. 2002; Scalas et al. 2000)
- Magnetihydrodynamics (Schekochihin et al. 2008)
- Phase Transitions (Cabré and Solà-Morales 2005; Farina and Valdinoci 2011; Garroni and Palatucci 2006; Sire and Valdinoci 2009; Bates and Chmaj 1999; P. Fife 2003; Alberti and Bellettini 1998; Chen and P. C. Fife 2000; Dayal and Bhattacharya 2006; Cozzi et al. 2017; Savin and Valdinoci 2012)
- Stratified Materials (Chermisi and Valdinoci 2010)
- Crystal Dislocation (Biler, Karch, and Monneau 2008; Gonzalez and Monneau 2010; Toland 1997)
- Soft Thin Films (Kurzke 2006)
- Corrosion (Siavash Jafarzadeh et al. 2018)
- Nonlocal Heat Conduction (Bobaru and Duangpanya 2010)
- Thin Obstacle Problem (Silvestre 2007)
- Gradient Potential Theory (Mingione 2010)
- Minimal Surfaces (L. Caffarelli, Roquejoffre, et al. 2009; L. Caffarelli and Valdinoci 2011)
- Quasi-geostrophic Flow (L. Caffarelli and Vasseur 2010; Cordoba 1998; Majda and Tabak 1996)
- Scattering Theory (Colton et al. 1998; Duistermaat 1975; Grote and Kirsch 2004)
- Optimization (Duvant and Lions 2012)
- Quantum Mechanics (Fefferman and de la Llave 1986; Laskin 2000)

These numerous examples show the potential strengths of developing an analytic theory that does not lean so heavily on classical derivatives. In the next chapter, we consider one particular nonlocal operator, and show that it forms a natural extension of the classical theory.

Chapter 2

A Nonlocal Operator and Nonlocal Variational Problems

This chapter presents the main notation and assumptions for the central class of variational problems studied in the next chapter. A nonlocal operator is introduced, and some of its basic properties are explored. Additionally, some well-known theorems are included for easy reference later on.

2.1 Connecting Nonlocal Models and the Calculus of Variations

In the previous chapter, we saw that variational problems involving functionals of the form

$$J[u] = \int_\Omega f(x, u, u') \, \mathrm{d}x$$

arise throughout physics and engineering. However, we also saw that there are advantages to formulating problems in terms of nonlocal operators. It is natural to ask what happens when these two lines of thought are combined: what can we say about minimizing functionals of the form

$$J_\delta[u] = \int_\Omega f(x, u, \mathcal{D}_\delta u) \, \mathrm{d}x, \tag{2.1}$$

for some nonlocal operator $\mathcal{D}_\delta$? Because some nonlocal operators do not even require weak differentiability, functionals of the form J_δ can be defined on function spaces much larger than the classical Sobolev spaces $W^{1,p}$. From a purely mathematical standpoint, then, we would like to know how much of the classical theory carries over to this new setting. For example, Dirichlet's Principle provides a way of translating a partial differential equation into a variational problem. Does some-

thing similar occur when the classical Laplacian is replaced with its nonlocal counterpart? Is there an analogous form of the Euler-Lagrange equations providing a necessary condition on minimizers of J_δ?

Apart from this mathematical motivation, functionals like J_δ have already been shown to have a wide variety of applications. As mentioned at the end of Section 1.1.2, one recent motivation for investigating variational methods has come from material science, in the study of nonlinear elasticity. Peridynamics, the primary motivation of studying nonlocal models, also deals with elastic bodies. So it should not be surprising that the calculus of variations and nonlocal modeling have been combined in many recent studies. Section 1.1.2 also discussed the significance of Noether's theorem and variational principles like the Principle of Virtual Work (Oterkus et al. 2012; Huang 2019). Analogs to both of these have been found in the nonlocal setting. Some other recent work connecting nonlocal models and variational problems includes Foss, Radu, and Wright 2018; Bellido et al. 2020; Cueto 2021; Mengesha and Du 2013 and Hinds and Radu 2012.

2.2 Introducing the Problem

The goal is to find an operator $\mathcal{D}_\delta$ that behaves like the classical derivative in some ways, but has the advantages of the nonlocal operators discussed in Section 1.2. We would like $\mathcal{D}_\delta$ to be a linear operator, like the derivative, but defined on a larger space of functions. We also want $\mathcal{D}_\delta$ to be capable of capturing long-range interactions, of the sort that arise in peridynamics or other nonlocal models. Finally, as the horizon δ goes to 0, in cases where the classical derivative is well defined, we should have $\mathcal{D}_\delta \to \dfrac{\mathrm{d}}{\mathrm{d}x}$ (in some sense). In this chapter, we defined an operator $\mathcal{D}_\delta$ and show it has precisely these properties.

The basic idea is to think of the integral as a sort of averaging operator. Integrating collects information about the behavior of a function over a whole region, while the classical derivative (defined in terms of a limit) describes the deeper behavior fo the function at a single point. When δ is very small, we might expect:

$$⨍_\delta^\delta \frac{u(x+z)-u(x)}{z}\,\mathrm{d}z \sim \lim_{z\to x}\frac{u(x+z)-u(x)}{z} = u'(x), \tag{2.2}$$

where the integral sign $⨍$ denotes an average:

$$⨍_\Omega g(t)\,\mathrm{d}t = \frac{1}{\mathrm{meas}(\Omega)}\int_\Omega g(t)\,\mathrm{d}t.$$

We can already see one of the features mentioned in the first paragraph coming into play. The left side of Equation 2.2 requires only that the difference quotient $\frac{u(x+z)-u(x)}{z}$ be integrable. This is a much weaker requirement than differentiability of u. As $\delta \to 0$, we can approximate $z \in [-\delta, \delta]$ simply by $z \sim \delta$. So the left side of Equation 2.2 behaves roughly like

$$⨍_{-\delta}^{\delta} \frac{u(x+z)-u(x)}{z}\,\mathrm{d}z \sim \int_{\delta}^{\delta} u(x+z)-u(x)\cdot\frac{1}{\delta^2}\,\mathrm{d}z.$$

This gives the simplest case of the operator defined in Definition 2.2.1, and Example 2.5.1 illustrates how this simple idea captures the behavior of the derivative.

Generalizing from this example, we might hope that an operator given by

$$\mathcal{D}_\delta u(x) = \int_{-\delta}^{\delta}[u(x+z)-u(x)]\mu_\delta(z)\,\mathrm{d}z,$$

for some appropriately chosen kernel μ_δ, satisfies the properties from the beginning of this subsection. Different choices of μ_δ may be motivated by specific modeling considerations. For example, the kernel may be designed to capture specific long-range interactions within an elastic body. It turns out that the central existence theorem, Theorem 3.2.1, and the convergence result, 2.4.2, require very little of the kernel μ_δ. The theory is surprisingly flexible in this sense.

Motivated by the discussion above, we define the nonlocal operator $\mathcal{D}_\delta$.

Definition 2.2.1: Let Ω be some interval in $\mathbb{R}$ and fix a $\delta > 0$. Let Γ denote the collar of Ω, with width δ. Then let $p > 1$ and $u \in L^p(\Omega \cup \Gamma)$ be given. For any $\mu_\delta \in L^1\big((-\delta,\delta)\big)$, we define the first order nonlocal differential operator $\mathcal{D}_\delta$ by

$$\mathcal{D}_\delta u(x) \coloneqq \int_{-\delta}^{\delta}[u(x+z)-u(x)]\mu_\delta(z)\,\mathrm{d}z. \tag{2.3}$$

In cases where it does not create ambiguity, we drop the subscript δ and simply write $\mathcal{D}$ and μ.

Remark 2.2.2: We focus on the one-dimensional case, but the hope is that these concepts and results extend easily to higher dimensions.

We can finally state the central question of this thesis: when do functionals like the one given in Equation 2.1 have minimimzers? When are they unique? Theorems 3.2.1 and 3.3.2 answer these questions.

In the rest of this chapter, we show that this choice of operator does in fact have the desired properties from the beginning of this section. Proposition 2.4.1 shows that $\mathcal{D}_\delta$ is a bounded linear operator between L^p spaces. Theorem 2.4.2 and its corollary show that, under some additional hypotheses, $\mathcal{D}_\delta$ converges to the classical derivative as δ goes to 0. And Section 2.5 presents a few examples of the behavior of $\mathcal{D}_\delta$ for different choices of μ_δ and u.

2.3 Some Helpful Facts

We collect some useful theorems needed for later discussion. Since these are all classic results, we provide references instead of complete proofs. Throughout, we assume $p > 1$.

Lemma 2.3.1 (Minkowski's Integral Inequality)**:** If A and B are measurable subsets of $\mathbb{R}$ and $g : A \times B \to \mathbb{R}$ is a measurable function,

$$\left(\int_A \left(\int_B g(x,y)\,\mathrm{d}y \right)^p \mathrm{d}x \right)^{1/p} \leq \int_B \left(\int_A \left(g(x,y) \right)^p \mathrm{d}x \right)^{1/p} \mathrm{d}y$$

(Hardy et al. 1952, Theorem 202).

Lemma 2.3.2 (Hölder's Inequality)**:** Let $u \in L^p(X)$ and $v \in L^q(X)$, where $q \coloneqq \frac{p}{p-1}$ denotes the

Hölder conjugate of p. If X is a measurable subset of $\mathbb{R}$, then

$$\int_X \left|u(x)v(x)\right| \mathrm{d}x \leq \left(\int_X \left|u(x)\right|^p \mathrm{d}x\right)^{\frac{1}{p}} \left(\int_X \left|v(x)\right|^q \mathrm{d}x\right)^{\frac{1}{q}}$$

(Hardy et al. 1952, Theorem 189)

Lemma 2.3.3 (A Special Case of Jensen's Inequality (integral form))**:** For $u \in L^p(-\delta, \delta)$, we have

$$\left(\int_{-\delta}^{\delta} \left|u(x)\right| \mathrm{d}x\right)^p \leq (2\delta)^{p-1} \int_{-\delta}^{\delta} \left|u(x)\right|^p \mathrm{d}x$$

(Dacorogna 2008, Theorem 2.28).

Proof. Set $g = 1$ in Hölder's Inequality. For an alternative proof of a much more general result, see (Fonseca and Leoni 2007, Theorem 4.80) □

Lemma 2.3.4 (Lebesgue-Fatou Lemma)**:** Let u_n be a sequence of real-valued, integrable functions on a measurable subset $X \subseteq \mathbb{R}$. If there exists a real-valued integrable function g such that, for all $n \in \mathbb{N}$, $g(x) \leq u_n(x)$ for almost all $x \in X$, then

$$\int_X \liminf_{n\to\infty} u_n(x) \, \mathrm{d}x \leq \liminf_{n\to\infty} \int_X u_n(x) \, \mathrm{d}x$$

(Yosida 1980, p. 17).

Lemma 2.3.5 (Fubini-Tonelli Theorem)**:** Let $g : X \times Y \to \mathbb{R}$ be a measurable function over $X \times Y$, a measurable subset of $\mathbb{R}^2$. Then g is integrable over $X \times Y$ if and only if at least of of the iterated integrals

$$\int_X \int_Y \left|g(x,y)\right| \mathrm{d}y \, \mathrm{d}x, \text{ or } \int_Y \int_X \left|g(x,y)\right| \mathrm{d}x \, \mathrm{d}y$$

is finite. In that case, we have

$$\int_X \int_Y \left|g(x,y)\right| \,\mathrm{d}y \,\mathrm{d}x = \int_Y \int_X \left|g(x,y)\right| \,\mathrm{d}x \,\mathrm{d}y$$

(Yosida 1980, p. 18).

Lemma 2.3.6: For any $1 \leq p < \infty$ and $h \in \mathbb{R}$, define the translation operator $\tau_h : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ by

$$\tau_h f(x) \coloneqq f(x-h).$$

Then translation is continuous in h with respect to the L^p norm: for any $f \in L^p(\mathbb{R})$,

$$\lim_{h \to 0} \|\tau_h f - f\|_{L^p(\mathbb{R})} = 0$$

(Folland 1999, Proposition 8.5).

Proof. The trick is to use density of continuous functions with compact support and the triangle inequality. See □

Lemma 2.3.7: A bounded sequence in a reflexive Banach space contains a weakly convergent subsequence (Zabarankin and Kurdila 2005, Theorem 7.3.3).

Lemma 2.3.8: For a measurable set $X \subseteq \mathbb{R}$ and for $p > 1$, $L^p(X)$ is reflexive.

Proof. This follows from the Riesz Representation Theorem. See for example Conway 2019, p. III.11.3 or Fonseca and Leoni 2007, Theorem 2.37. □

Lemma 2.3.9 (Mazur's Lemma)**:** Let $(X, \|\cdot\|)$ be a normed space and $\{x_n\}_{n=1}^{\infty}$ be a sequence in X converging weakly to $x_0 \in X$. Then there is a sequence of convex combinations $\{y_n\}_{n=1}^{\infty}$ converging strongly to x_0 such that $y_n = \sum_{k=n}^{N_k} \lambda_k x_k$ where $\sum_{k=n}^{N_k} \lambda_k = 1$ and $\lambda_k \geq 0$ for $n \leq k \leq N_k$, (Ekeland

and Temam 1976, p. 6)

Lemma 2.3.10 (Jensen's Inequality (finite form))**:** Suppose that $X \subseteq \mathbb{R}^n$ is convex and f is a convex function on X. If $\lambda_1, \ldots, \lambda_r$ are non-negative scalars such that $\sum_{i=1}^{r} \lambda_i = 1$ and $x_1, \ldots, x_r$ are points in X, then

$$f(\lambda_1 x_1 + \lambda_2 x_2 + \ldots \lambda_r x_r) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots + \lambda_r f(x_r)$$

(Hardy et al. 1952, Theorems 86 and 90).

2.4 Continuity of $\mathcal{D}$ and Convergence to the Classical Derivative

The following result establishes some central structural properties of the operator $\mathcal{D}$.

Proposition 2.4.1: The operator $\mathcal{D}$ is linear, and it continuously sends elements of $L^p(\Omega \cup \Gamma)$ into the space $L^p(\Omega)$.

Proof. First, note that the domain of the function $\mathcal{D}u$ must be Ω. The function u is defined on $(a - \delta, b + \delta)$, so the term $u(x + z)$ is only well-defined when $x \in \Omega$ (since z ranges from $-\delta$ to δ). We need to establish that $\mathcal{D}u$ is in fact in $L^p(\Omega)$, but first we show that $\mathcal{D}$ is linear. Let $u, v \in L^p(\Omega \cup \Gamma)$ and $\alpha \in \mathbb{R}$ be given. Then for all $x \in \Omega$,

$$\begin{aligned}
\mathcal{D}(u+v)(x) &= \int_{-\delta}^{\delta} [(u+v)(x+z) - (u+v)(x)]\mu(z)\,\mathrm{d}z \\
&= \int_{-\delta}^{\delta} [u(x+z) - u(x) + v(x+z) - v(x)]\mu(z)\,\mathrm{d}z \\
&= \int_{-\delta}^{\delta} [u(x+z) - u(x)]\mu(z) + [v(x+z) - v(x)]\mu(z)\,\mathrm{d}z \\
&= \int_{-\delta}^{\delta} [u(x+z) - u(x)]\mu(z)\,\mathrm{d}z + \int_{-\delta}^{\delta} [v(x+z) - v(x)]\mu(z)\,\mathrm{d}z \\
&= \mathcal{D}u(x) + \mathcal{D}v(x).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathcal{D}(\alpha u)(x) &= \int_{-\delta}^{\delta} [(\alpha u)(x+z) - (\alpha u)(x)]\mu(z)\,\mathrm{d}z \\
&= \alpha \int_{-\delta}^{\delta} [u(x+z) - u(x)]\mu(z)\,\mathrm{d}z \\
&= \alpha \mathcal{D}u(x).
\end{aligned}$$

These equalities hold for all $x \in \Omega$, so we see $\mathcal{D}(u+v) = \mathcal{D}u + \mathcal{D}v$ and $\mathcal{D}(\alpha u) = \alpha \mathcal{D}u$. Hence, $\mathcal{D}$ is a linear operator.

Now we show that $\|\mathcal{D}u\|_{L^p(\Omega)} \leq 2\,\|u\|_{L^p}\,\|\mu\|_{L^1}$. This means that $\mathcal{D}u \in L^p$, but is also establishes that the operator norm $\|\mathcal{D}\| = \sup\left\{\|\mathcal{D}u\|_{L^p(\Omega)} : \|u\|_{L^p(\Omega\cup\Gamma)} = 1\right\}$ is finite. Hence, $\mathcal{D}$ is continuous.

Fix any $u \in L^p(\Omega \cup \Gamma)$. Applying Minkowski's Integral Inequality (Lemma 2.3.1) to the defini-

tion of $\mathcal{D}$, we have

$$\begin{aligned}
\|\mathcal{D}u\|_{L^p(\Omega)} &= \left(\int_\Omega \left| \int_{-\delta}^{\delta} [u(x+z) - u(x)]\mu(z)\,\mathrm{d}z \right|^p \mathrm{d}x \right)^{1/p} \\
&\leq \int_{-\delta}^{\delta} \left(\int_a^b \left| [u(x+z) - u(x)]\mu(z) \right|^p \mathrm{d}x \right)^{1/p} \mathrm{d}z \\
&= \int_{-\delta}^{\delta} \left|\mu(z)\right|^{p/p} \left(\int_a^b \left| u(x+z) - u(x) \right|^p \mathrm{d}x \right)^{1/p} \mathrm{d}z \\
&= \int_{-\delta}^{\delta} \left|\mu(z)\right| \left\| u(\cdot + z) - u(\cdot) \right\|_{L^p(\Omega)} \mathrm{d}z
\end{aligned}$$

Now we apply the triangle inequality for the L^p norm (which is sometimes also called the Minkowski inequality) to find

$$\begin{aligned}
\|\mathcal{D}u\|_{L^p(\Omega)} &\leq \int_{-\delta}^{\delta} \left|\mu(z)\right| \left(\left\| u(\cdot + z) \right\|_{L^p(\Omega)} + \|u\|_{L^p(\Omega)} \right) \mathrm{d}z \\
&= \int_{-\delta}^{\delta} \left|\mu(z)\right| \left\| u(\cdot + z) \right\|_{L^p(\Omega)} \mathrm{d}z + \int_{-\delta}^{\delta} \left|\mu(z)\right| \|u\|_{L^p(\Omega)} \mathrm{d}z \\
&= \int_{-\delta}^{\delta} \left|\mu(z)\right| \left(\int_\Omega \left| u(x+z) \right|^p \mathrm{d}x \right)^{1/p} \mathrm{d}z + \|u\|_{L^p(\Omega)} \|\mu\|_{L^1\left((-\delta,\delta)\right)}
\end{aligned}$$

Consider the inner integral of this first term. Changing variables to $y = x + z$, we have

$$\int_a^b \left| u(x+z) \right|^p \mathrm{d}x = \int_{a+z}^{b+z} \left| u(y) \right|^p \mathrm{d}y$$

Clearly $\left|u(y)\right|$ is never negative, so increasing the size of the domain of integration will never make the integral smaller. Since $z \in [-\delta, \delta]$, this implies

$$\int_\Omega \left| u(x+z) \right|^p \mathrm{d}x \leq \int_{\Omega \cup \Gamma} \left| u(y) \right|^p \mathrm{d}y.$$

Combining this with what we had above produces

$$\begin{aligned}
\|\mathcal{D}u\|_{L^p(\Omega)} &\leq \int_{-\delta}^{\delta} \left|\mu(z)\right| \left(\int_{\Omega\cup\Gamma} \left|u(y)\right|^p \mathrm{d}y\right)^{1/p} \mathrm{d}z + \|u\|_{L^p(\Omega)} \|\mu\|_{L^1\left((-\delta,\delta)\right)} \\
&= \int_{-\delta}^{\delta} \left|\mu(z)\right| \|u\|_{L^1(\Omega\cup\Gamma)} \,\mathrm{d}z + \|u\|_{L^p(\Omega)} \|\mu\|_{L^1\left((-\delta,\delta)\right)} \\
&= \|\mu\|_{L^1\left((-\delta,\delta)\right)} \|u\|_{L^p(\Omega\cup\Gamma)} + \|u\|_{L^p(\Omega)} \|\mu\|_{L^1\left((-\delta,\delta)\right)} \\
&= 2\,\|\mu\|_{L^1\left((-\delta,\delta)\right)} \|u\|_{L^p(\Omega\cup\Gamma)}
\end{aligned}$$

Each operator $\mathcal{D}$ is determined by fixing a kernel μ and a horizon δ, so $\|\mu\|_{L^1\left((-\delta,\delta)\right)}$ is some constant that depends on $\mathcal{D}$ but not on u. Thus, $\mathcal{D}$ is bounded (and therefore continuous), with operator norm $\|\mathcal{D}\| \leq 2\,\|\mu\|_{L^1\left((-\delta,\delta)\right)}$. □

For an interval $\Omega = (a, b)$, the operator $\mathcal{D}_\delta$ discussed above is defined on all of $L^p\left((a-\delta, b+\delta)\right)$, which includes many functions which are not even weakly differentiable. To motivate the definition of $\mathcal{D}_\delta$, we show that in cases where u' is well defined, $\mathcal{D}_\delta u$ converges to u' as the horizon δ shrinks to 0. This justifies the intuition given above for thinking of $\mathcal{D}_\delta$ as a kind of differential operator.

Theorem 2.4.2: Suppose that $u \in W^{1,1}(\mathbb{R})$, and that $\{\mu_\delta\}_{\delta>0} \subseteq L^1(\mathbb{R})$ is a family of integrable kernels satisfying the following:

1. For each δ, the kernel μ_δ is zero outside the interval $(-\delta, \delta)$.

2. The first moments are uniformly bounded in the L^1 norm: there is some $M > 0$ such that, for all $\delta > 0$,
$$\int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \mathrm{d}z < M. \tag{2.4}$$

3. As δ approaches 0 from the right,
$$\int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z \to 1. \tag{2.5}$$

Then $\mathcal{D}_\delta u$ converges to u' in the L^1 norm:

$$\lim_{\delta \to 0^+} \left\| \mathcal{D}_\delta u - u' \right\|_{L^1(\mathbb{R})} = 0.$$

Proof. Fix any $\varepsilon > 0$. Since $u \in W^{1,1}(\mathbb{R})$, both u and its weak derivative u' are in $L^1(\mathbb{R})$. If the L^1 norm of u' is 0, then for almost all $x \in \mathbb{R}$,

$$\left| u(x+z) - u(x) \right| = \left| \int_x^{x+z} u'(s)\, \mathrm{d}s \right| \leq \int_x^{x+z} \left| u'(s) \right| \mathrm{d}s \leq \int_{\mathbb{R}} \left| u'(s) \right| \mathrm{d}s = 0.$$

Hence, $[u(x+z) - u(x)] = 0$ almost everywhere and so $\mathcal{D}_\delta u(x) = 0$ for all δ. Then convergence in L^1 is immediate. So we suppose that $\left\| u' \right\|_{L^1(\mathbb{R})} \neq 0$. From Equation 2.5, there is some $\delta_1 > 0$ such that

$$\delta \leq \delta_1 \implies \left| \int_{-\delta}^{\delta} z\mu(z)\, \mathrm{d}z - 1 \right| < \frac{\varepsilon}{2 \left\| u' \right\|_{L^1(\mathbb{R})}}. \tag{2.6}$$

Additionally, Lemma 2.3.6, implies there is some $\delta_2 > 0$ such that

$$|\delta| \leq \delta_2 \implies \left\| \tau_\delta u' - u' \right\|_{L^1(\mathbb{R})} < \frac{\varepsilon}{2M}. \tag{2.7}$$

Now select any positive $\delta \leq \min\{\delta_1, \delta_2\}$. We show that $\left\| D_\delta u - u' \right\|_{L^1(\mathbb{R})} < \varepsilon$.

The idea is to express $\mathcal{D}_\delta u$ in terms of an integral involving u', so that we can more easily compute the distance between the two functions. We do this by applying the tal Theorem of Calculus. If we introduce the parameter s, note that

$$\frac{\mathrm{d}}{\mathrm{d}s} u(x+sz) = u'(x+sz)z, \text{ so } \int_0^1 u'(x+sz)z\, \mathrm{d}s = \left[u(x+sz) \right]_{s=0}^{s=1} = u(x+z) - u(x),$$

which is exactly the difference occurring inside the integrand of $\mathcal{D}_\delta$ (Equation 2.3). Thus, for any $x \in \mathbb{R}$,

$$\mathcal{D}_\delta u(x) = \int_{-\delta}^{\delta} [u(x+z) - u(x)] \mu_\delta(z)\, \mathrm{d}z = \int_{-\delta}^{\delta} \left[\int_0^1 u'(x+sz) z\, \mathrm{d}s \right] \mu_\delta(z)\, \mathrm{d}z.$$

Next, add and subtract $u'(x)z$ inside the innermost integral:

$$\begin{aligned}\mathcal{D}_\delta u(x) &= \int_{-\delta}^{\delta} \left[\int_0^1 u'(x+sz)z - u'(x)z + u'(x)z \,\mathrm{d}s \right] \mu_\delta(z)\,\mathrm{d}z \\ &= \int_{-\delta}^{\delta} \left[\int_0^1 \left(u'(x+sz) - u'(x)\right) z \,\mathrm{d}s + \int_0^1 u'(x)z \,\mathrm{d}s \right] \mu_\delta(z)\,\mathrm{d}z\end{aligned}$$

Since $\int_0^1 u'(x)z\,\mathrm{d}s = u'(x)z$, we have

$$\mathcal{D}_\delta u(x) = \int_{-\delta}^{\delta} \left(z\mu_\delta(z) \cdot \int_0^1 u'(x+sz) - u'(x)\,\mathrm{d}s \right) \mathrm{d}z + u'(x) \int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z. \tag{2.8}$$

We are now ready to compute the L^1 distance. Using the triangle inequality repeatedly yields

$$\begin{aligned}\left\|\mathcal{D}_\delta u - u'\right\|_{L^1(\mathbb{R})} &= \int_{\mathbb{R}} \left| \int_{-\delta}^{\delta} \left(z\mu_\delta(z) \cdot \int_0^1 u'(x+sz) - u'(x)\,\mathrm{d}s \right) \mathrm{d}z + u'(x) \left(\int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z - 1 \right) \right| \mathrm{d}x \\ &\le \int_{\mathbb{R}} \left| \int_{-\delta}^{\delta} \left(z\mu_\delta(z) \cdot \int_0^1 u'(x+sz) - u'(x)\,\mathrm{d}s \right) \mathrm{d}z \right| + \left|u'(x)\right| \left| \int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z - 1 \right| \mathrm{d}x \\ &\le \int_{\mathbb{R}} \int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \int_0^1 \left|u'(x+sz) - u'(x)\right| \mathrm{d}s\,\mathrm{d}z + \left|u'(x)\right| \left| \int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z - 1 \right| \mathrm{d}x\end{aligned}$$

Hence,

$$\left\|\mathcal{D}_\delta u - u'\right\|_{L^1(\mathbb{R})} = \int_{\mathbb{R}} \int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \int_0^1 \left|u'(x+sz) - u'(x)\right| \mathrm{d}s\,\mathrm{d}z\,\mathrm{d}x + \int_{\mathbb{R}} \left|u'(x)\right| \left| \int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z - 1 \right| \mathrm{d}x \tag{2.9}$$

We handle each of the two terms separately.

Applying Fubini's Theorem (Lemma 2.3.5) to the first term of Equation 2.9, we obtain

$$\int_{\mathbb{R}} \int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \int_0^1 \left|u'(x+sz) - u'(x)\right| \mathrm{d}s\,\mathrm{d}z\,\mathrm{d}x = \int_{-\delta}^{\delta} \int_0^1 \int_{\mathbb{R}} \left|u'(x+sz) - u'(x)\right| \mathrm{d}x\,\mathrm{d}s \left|z\mu_\delta(z)\right| \mathrm{d}z. \tag{2.10}$$

From the definition of the translation operator and the L^1 norm,

$$\int_{\mathbb{R}} \left|u'(x+sz) - u'(x)\right| \mathrm{d}x = \left\|\tau_{-sz}u' - u'\right\|_{L^1(\mathbb{R})}.$$

Since $z \in [-\delta, \delta]$ and $s \in [0,1], |sz| \leq \delta$. We chose $\delta \leq \delta_2$, so Equation 2.7 implies

$$\left\|\tau_{-sz}u' - u'\right\|_{L^1(\mathbb{R})} < \frac{\varepsilon}{2M}.$$

This bound is independent of s, so integrating from $s = 0$ to $s = 1$ doesn't change anything. Therefore, Equation 2.10 becomes

$$\int_{\mathbb{R}} \int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \int_0^1 \left|u'(x+sz) - u'(x)\right| \mathrm{d}s \, \mathrm{d}z \, \mathrm{d}x < \frac{\varepsilon}{2M} \int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \mathrm{d}z.$$

Finally, apply Equation 2.4 to conclude

$$\int_{\mathbb{R}} \int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \int_0^1 \left|u'(x+sz) - u'(x)\right| \mathrm{d}s \, \mathrm{d}z \, \mathrm{d}x < \frac{\varepsilon}{2M} \int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \mathrm{d}z < \frac{\varepsilon}{2}.$$

Now for the second term in Equation 2.9. Since $\delta \leq \delta_1$, Equation 2.6 implies that

$$\begin{aligned}
\int_1^b \left|u'(x)\right| \left| \int_{-\delta}^{\delta} z\mu_\delta(z) \, \mathrm{d}z - 1 \right| \mathrm{d}x &< \int_{\mathbb{R}} \left|u'(x)\right| \frac{\varepsilon}{2\left\|u'\right\|_{L^1(\mathbb{R})}} \mathrm{d}x \\
&= \frac{\varepsilon}{2\left\|u'\right\|_{L^1(\mathbb{R})}} \int_{\mathbb{R}} \left|u'(x)\right| \mathrm{d}x \\
&= \frac{\varepsilon}{2}.
\end{aligned}$$

Putting things back together, we have shown

$$\begin{aligned}
\left\|\mathcal{D}_\delta u - u'\right\|_{L^1(\mathbb{R})} &\leq \int_{\mathbb{R}} \int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \int_0^1 \left|u'(x+sz) - u'(x)\right| \mathrm{d}s \, \mathrm{d}z \, \mathrm{d}x + \int_{\mathbb{R}} \left|u'(x)\right| \left| \int_{-\delta}^{\delta} z\mu_\delta(z) \, \mathrm{d}z - 1 \right| \mathrm{d}x \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

Thus, whenever $\delta \leq \min\{\delta_1, \delta_2\}$, $\left\|\mathcal{D}_\delta u - u'\right\| < \varepsilon$. This proves the theorem. □

Remark 2.4.3: In the previous theorem, the choice of $\mathbb{R}$ for the domain was primarily for conve-

nience. For example, the same argument holds if $u \in W^{1,1}\left((a-C, b+C)\right)$ (for any $C > 0$) and we only consider kernels $\{\mu_\delta\}_{\delta \in (0,C)}$. Alternatively, it is possible to extend $u : \Omega \to \mathbb{R}$ to a function $\bar{u} : \mathbb{R} \to \mathbb{R}$ such that $\bar{u}(x) = u(x)$ for all $x \in \Omega$ and $\|\bar{u}\|_{W^{1,1}(\mathbb{R})} \leq C_u \|u\|_{W^{1,1}(\mathbb{R})}$ for some constant C_u (Evans 1998, p. 254). So there is no real loss of generality by assuming $u \in W^{1,1}(\mathbb{R})$.

With a stronger assumption on u and a slight modification on the behavior of μ_δ, we get pointwise convergence as well.

Corollory 2.4.4: Suppose that $u \in C^{1,\alpha}(\mathbb{R})$, and that $\{\mu_\delta\}_{\delta>0} \subseteq L^1(\mathbb{R})$ is a family of integrable kernels satisfying the following:

1. For each δ, the kernel μ_δ is zero outside the interval $(-\delta, \delta)$.
2. The first moments are uniformly bounded in the L^1 norm: there is some $M > 0$ such that, for all $\delta > 0$,
$$\int_{-\delta}^{\delta} \left|z\mu_\delta(z)\right| \mathrm{d}z < M. \tag{2.11}$$
3. For all $\delta > 0$,
$$\int_{-\delta}^{\delta} z\mu_\delta(z) \, \mathrm{d}z = 1. \tag{2.12}$$

Then there is a constant $C \in \mathbb{R}$ such that, for all $x \in \mathbb{R}$ and $\delta > 0$,

$$\left|\mathcal{D}_\delta u(x) - u'(x)\right| < C\delta^\alpha.$$

In particular, for any $x \in \mathbb{R}$,

$$\lim_{\delta \to 0^+} \mathcal{D}_\delta u(x) = u'(x).$$

Proof. Select any $\delta > 0$ and $x \in \mathbb{R}$. Since $u \in C^{1,\alpha}$, there is some C_0 such that $\left|u'(x) - u'(y)\right| \leq C_0|x-y|^\alpha$ for any $y \in \mathbb{R}$. This is a special case of the previous theorem, so Equation 2.8 still holds:

$$\mathcal{D}_\delta u(x) = \int_{-\delta}^{\delta} \left(z\mu_\delta(z) \cdot \int_0^1 u'(x+sz) - u'(x) \, \mathrm{d}s \right) \mathrm{d}z + u'(x) \int_{-\delta}^{\delta} z\mu_\delta(z) \, \mathrm{d}z. \tag{2.13}$$

Thus,

$$\begin{aligned}\left|\mathcal{D}_\delta u(x) - u'(x)\right| &= \left|\int_{-\delta}^{\delta}\left(z\mu_\delta(z)\cdot\int_0^1 u'(x+sz)-u'(x)\,\mathrm{d}s\right)\mathrm{d}z + u'(x)\int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z - u'(x)\right| \\ &\leq \int_{-\delta}^{\delta}\left|z\mu_\delta(z)\right|\cdot\int_0^1\left|u'(x+sz)-u'(x)\right|\mathrm{d}s\,\mathrm{d}z + \left|u'(x)\right|\left|\int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z - 1\right| \\ &\leq C_0\int_{-\delta}^{\delta}\left|z\mu_\delta(z)\right|\cdot\int_0^1\left|(x+sz)-x\right|^\alpha\mathrm{d}s\,\mathrm{d}z\end{aligned}$$

Note that we used both assumption 3 in the statement of the Corollary and the Hölder continuity of u' to get this last line. Since $s\in[0,1]$ and $z\in[-\delta,\delta]$, $|sz|\leq|\delta|$. Thus,

$$\left|\mathcal{D}_\delta u(x) - u'(x)\right| \leq C_0\int_{-\delta}^{\delta}\left|z\mu_\delta(z)\right|\cdot\int_0^1\delta^\alpha\,\mathrm{d}s\,\mathrm{d}z = C_0\int_{-\delta}^{\delta}\left|z\mu_\delta(z)\right|\delta^\alpha\,\mathrm{d}z$$

By assumption 4, it follows that

$$\left|\mathcal{D}_\delta u(x) - u'(x)\right| < C_0 M\delta^\alpha =: C\delta^\alpha,$$

as desired. □

2.5 Examples

The following examples demonstrate how different choices of kernel can approximate the behavior of the classical derivative. Note that if each μ_δ is symmetric about the origin, Equation 2.5 will never be satisfied. One way to ensure the correct limit is to appropriately scale each kernel so that each first moment is exactly one. Also, there is no need to consider all $\delta > 0$ since we are primarily interested in cases where δ is very small. Thus, the condition given in Equation 2.4 really only needs to apply once δ is sufficiently close to 0. (See also Remark 2.4.3.)

Example 2.5.1: This simple case illustrates the procedure for properly scaling the kernels. Let $u(x) = ax^2 + bx + c$ for some $a, b, c \in \mathbb{R}$. Suppose that each kernel μ_δ is an antisymmetric, piecewise

constant function:

$$\mu_\delta(z) \coloneqq \begin{cases} -C_\delta, & z < 0, \\ C_\delta, & z \geq 0, \end{cases}$$

where C_δ is a scaling constant to be determined. We find the scaling using Equation 2.5:

$$\int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z = 2C_\delta \int_0^\delta z\,\mathrm{d}z = 2C_\delta\left(\frac{\delta^2}{2} - 0\right) = C_\delta\delta^2.$$

Setting C_δ equal to $\frac{1}{\delta^2}$ for each δ, ensures that Equation 2.5 is satisfied.

Now we compute the nonlocal derivative. Using the definition of $\mathcal{D}_\delta$, we have

$$\begin{aligned} \mathcal{D}_\delta u(x) &= \int_{-\delta}^{\delta} \Big[\Big(a(x+z)^2 + b(x+z) + c\Big) - \Big(ax^2 + bx + c\Big)\Big]\,\mu_\delta(z)\,\mathrm{d}z \\ &= \int_{-\delta}^{\delta} \Big[az^2 + (2ax+b)z\Big]\,\mu_\delta(z)\,\mathrm{d}z \end{aligned}$$

Then plugging in our choice of μ_δ yields

$$\begin{aligned} \mathcal{D}_\delta u(x) &= \int_{-\delta}^{0} \Big[az^2 + (2ax+b)z\Big]\left(-\frac{1}{\delta^2}\right)\mathrm{d}z + \int_0^\delta \Big[az^2 + (2ax+b)z\Big]\left(\frac{1}{\delta^2}\right)\mathrm{d}z \\ &= \left(-\frac{1}{\delta^2}\right)\left[\frac{a}{3}z^3 + \frac{2ax+b}{2}z^2\Big|_{-\delta}^{0} + \left(\frac{1}{\delta^2}\right)\left[\frac{a}{3}z^3 + \frac{2ax+b}{2}z^2\Big|_0^{\delta}\right.\right. \\ &= \left(-\frac{1}{\delta^2}\right)\left(\frac{a}{3}\delta^3 - \frac{2ax+b}{2}\delta^2\right) + \left(\frac{1}{\delta^2}\right)\left(\frac{a}{3}\delta^3 + \frac{2ax+b}{2}\delta^2\right) \\ &= \left(-\frac{a}{3}\delta + \frac{2ax+b}{2}\right) + \left(\frac{a}{3}\delta + \frac{2ax+b}{2}\right) \\ &= 2ax + bx \end{aligned}$$

Thus, we see that antisymmetric kernels behaving like $\sim \frac{1}{\delta^2}$ successfully recover the classical derivative. △

Remark 2.5.2: The basic idea of proof of Theorem 2.4.2 was to use the Fundamental Theorem of Calculus to relate the difference inside the definition of $\mathcal{D}$ to the classical derivative u'. One

advantage of this process is that it provides an expression depending on s, a scalar quantity whose values we can control. When working in higher dimensions, this turns out to be a handy trick for establishing bounds. Another point about the procedure: there was no reason we needed to stop at the first derivative. We could continue the steps above, adding and subtracting a term and then applying the Fundamental Theorem of Calculus. This can be used to ensure $\mathcal{D}$ captures the higher order behavior of u. For example, if $u \in C^6(\mathbb{R})$, say, we can add the following constraints on the family $\{\mu_\delta\}_{\delta>0}$:

$$\forall \delta > 0,\ \forall z \in (-\delta, \delta), \quad \mu_\delta(-z) = -\mu_\delta(z) \tag{2.14}$$

$$\lim_{\delta \to 0^+} \int_{-\delta}^{\delta} \mu_\delta(z) z^3 \, \mathrm{d}z = 0 \tag{2.15}$$

$$\lim_{\delta \to 0^+} \int_{-\delta}^{\delta} \mu_\delta(z) z^5 \, \mathrm{d}z = 0. \tag{2.16}$$

Then our nonlocal operator does an even better job of approximating the behavior of the classical derivative.

The same idea can be captured (perhaps more intuitively) in terms of Taylor Series. Suppose that μ_δ is odd and u is sufficiently smooth in the relevant domain. Then we replace u with a Taylor polynomial centered at x:

$$\begin{aligned}
\mathcal{D}_\delta u(x) &= \int_{-\delta}^{\delta} [u(x+z) - u(z)] \mu_\delta(z) \, \mathrm{d}z \\
&= \int_{-\delta}^{\delta} \left(u'(x) z + \frac{u''(x) z^2}{2} + \frac{u'''(x) z^3}{6} + \frac{u^{(4)}(x) z^4}{4!} + \frac{u^{(5)}(x) z^5}{5!} + \dots \right) \mu_\delta(z) \, \mathrm{d}z
\end{aligned}$$

And so

$$\begin{aligned}
\mathcal{D}_\delta u(x) = u'(x) \int_{-\delta}^{\delta} z \mu_\delta(z) \, \mathrm{d}z + \frac{u''(x)}{2} \int_{-\delta}^{\delta} z^2 \mu_\delta(z) \, \mathrm{d}z + \frac{u'''(x)}{6} \int_{-\delta}^{\delta} z^3 \mu_\delta(z) \, \mathrm{d}z \\
+ \frac{u^{(4)}(x)}{4!} \int_{-\delta}^{\delta} z^4 \mu_\delta(z) \, \mathrm{d}z + \frac{u^{(5)}(x)}{5!} \int_{-\delta}^{\delta} z^5 \mu_\delta(z) \, \mathrm{d}z + \dots
\end{aligned} \tag{2.17}$$

Since μ_δ is odd, and we are integrating over the interval $[-\delta, \delta]$, the terms with even powers of z

cancel out. This leaves

$$\mathcal{D}_\delta u(x) = u'(x)\int_{-\delta}^{\delta} z\mu_\delta(z)\,\mathrm{d}z + \frac{u'''(x)}{6}\int_{-\delta}^{\delta} z^3\mu_\delta(z)\,\mathrm{d}z + \frac{u^{(5)}(x)}{5!}\int_{-\delta}^{\delta} z^5\mu_\delta(z)\,\mathrm{d}z + \dots.$$

If Equations 2.15 and 2.16 hold, then in the limit $\delta \to 0^+$ we get

$$\mathcal{D}_\delta u(x) \to u'(x) + 0 + 0 + \text{ (higher order terms).}$$

Thus, by choosing the kernels μ_δ appropriately, we can approximate the classical derivative to any degree of accuracy. This computation also demonstrates why the choice of anti-symmetric kernels is natural for approximating first-order behavior. (We also saw this in Example 2.5.1.) Notice, for example, that if μ is symmetric about the origin $\int z\mu(z)\,\mathrm{d}x = 0$, so Equation 2.5 is never satisfied. Radially symmetric kernels would be better suited for second-order operators like the Laplacian, since the odd-degree terms in the above Taylor expansion all drop out.

Example 2.5.3: A classic set of examples in analysis is given by

$$f(x) = \begin{cases} x^a \sin\left(\frac{1}{x^b}\right), & x \neq 0, \\ 0, & x = 0, \end{cases}$$

for some $a, b > 0$. As $x \to 0$, x^a goes to 0 and $\frac{1}{x^b}$ blows up to infinity. So the behavior near 0 depends on the relative sizes of a and b. If $a \gg b > 1$, we might expect the amplitude of the oscillations to decay much quicker than the increase in frequency, making the function differentiable. Consider, for example, setting $a = 2$ and $b = 1$. Then it can be shown f is continuous at 0, but not differentiable: for any $x \neq 0$,

$$f'(x) = 2x\sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right).$$

As x goes to zero, the cosine term oscillates too rapidly, so that $\lim_{x\to 0} f'(x)$ does not exist. The

functions f and f' are illustrated in Figure 2.1.

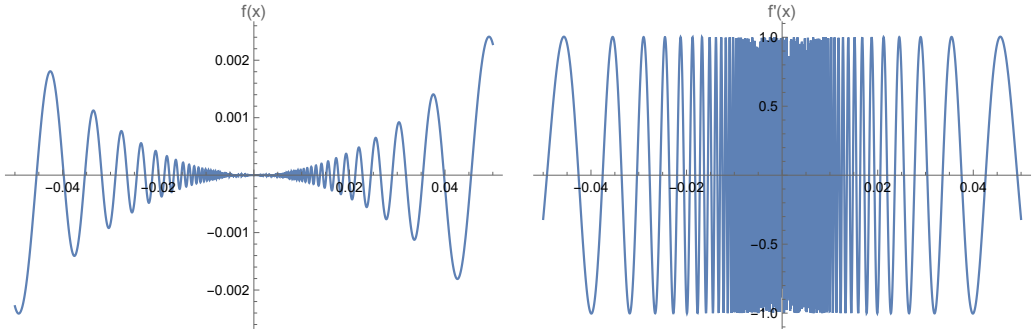


Figure 2.1: Near $x = 0$, the oscillations in f decay (left), but not fast enough for the derivative (right) to converge.

But how will the nonlocal derivative $\mathcal{D}_\delta f$ behave? Let μ be the same piecewise constant kernel from tExample 2.5.1. Note that

$$\mathcal{D}_\delta f(0) = \int_{-\delta}^{\delta} [f(0+z) - f(0)]\mu_\delta(z)\,\mathrm{d}z = \int_{-\delta}^{\delta} f(z)\mu_\delta(z)\,\mathrm{d}z$$

Substituting in our choice of μ_δ, then changing variables, this becomes

$$\begin{aligned}\mathcal{D}_\delta f(0) &= -\frac{1}{\delta^2}\int_{-\delta}^{0} z^2 \sin\left(\frac{1}{z}\right)\mathrm{d}z + \frac{1}{\delta^2}\int_{0}^{\delta} z^2 \sin\left(\frac{1}{z}\right)\mathrm{d}z \\ &= \frac{2}{\delta^2}\int_{0}^{\delta} z^2 \sin\left(\frac{1}{z}\right)\mathrm{d}z.\end{aligned}$$

We can then get an upper bound on $\mathcal{D}_\delta f(x)$ in terms of δ:

$$\left|\mathcal{D}_\delta f(0)\right| \leq \frac{2}{\delta^2}\int_{0}^{\delta} \left| z^2 \sin\left(\frac{1}{z}\right)\right| \mathrm{d}z.$$

Since $z \in (0, \delta)$, $\left|\sin(y)\right| \leq 1$ for all $y \in \mathbb{R}$, and $\int_0^\delta \mathrm{d}z = 1$,

$$\left|\mathcal{D}_\delta f(0)\right| \leq \frac{2}{\delta^2} \cdot \delta^2 \cdot 1 \cdot \delta = 2\delta.$$

Thus, as $\delta \to 0$, $\mathcal{D}_\delta f(x) \to 0$.

Let

$$g(\delta) = \mathcal{D}_\delta f(0) = \frac{2}{\delta^2} \int_0^\delta z^2 \sin\left(\frac{1}{z}\right) \mathrm{d}z.$$

Then we can plot g as a function of δ to see how the nonlocal derivative behaves when δ is near 0. The numerical approximation is shown in Figure 2.2.

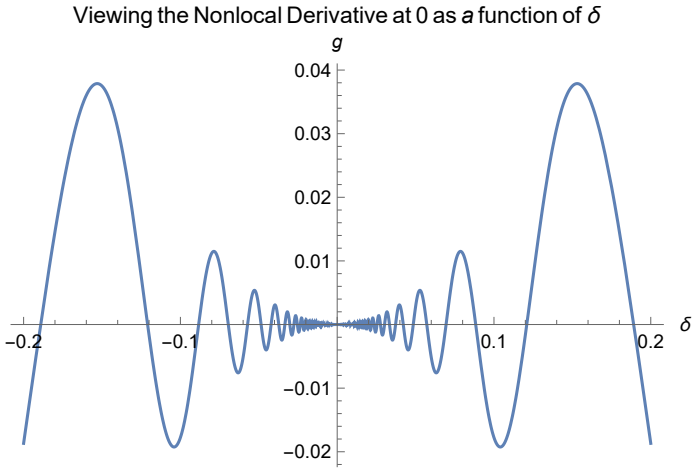


Figure 2.2: When δ is small, $g(\delta) = \mathcal{D}_\delta f(0)$ is small.

The preceding discussion has focused on the strengths of using the operator $\mathcal{D}_\delta$. But this example illustrates a potential drawback, depending on the point of view. Our choice of kernel leads to a lot of cancellation when computing $\mathcal{D}_\delta u(0)$. There is a sense in which, compared with the classical derivative, we lose a lot of information about the behavior of the function near the origin. △

This example illustrates one case where $\mathcal{D}_\delta u$ is well-defined even though the classical does not exist. The next two examples demonstrate what happens in other cases where u' does not exist. In each of the following plots, $\delta = 0.05$, and we consider the interval $[0, 1]$. Also, we use an appropriately scaled and truncated version of $z^{-1/3}$ for our kernel, shown below in Figure 2.3. Note that the figures below use a discretized version of u and the trapezoid rule to compute the nonlocal derivative, with a spacing of $h = 0.005$ between points. This dramatically reduces computation time, since the integrals involved in computing $\mathcal{D}_\delta$ can sometimes be challenging for systems like *Mathematica* to handle efficiently. Any loss in precision should not be enough to obscure the main

point of these examples. More details on numerical approximation can be found in Chapter 4 and Appendix A.

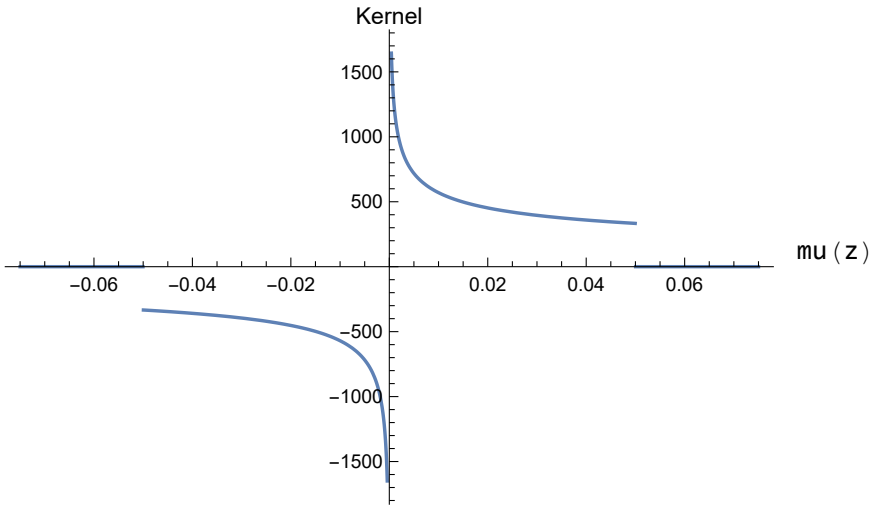


Figure 2.3: Inside the interval $(-0.05, 0.05)$, the kernel is given by $\mu(z) = \frac{122.801}{z^{1/3}}$. This kernel is in L^1 and has been scaled to meet the conditions of Theorem 2.4.2.

Example 2.5.4: For functions with cusps, the classical derivative $u'(x)$ blows up. In some cases, however, the nonlocal derivative $\mathcal{D}_\delta u(x)$ stays well-behaved. This is illustrated by the following plots. The function $u(x) = (x - \frac{1}{2})^{2/3}$, shown in Figure 2.4, has a cusp at $x = \frac{1}{2}$.

The nonlocal derivative collects up information about the function over a whole interval, and acts as a sort of average of the slopes in that region. Thus, $\mathcal{D}_\delta u$ smoothly interpolates between the values of $u'(x)$ at points where u is not differentiable. △

Example 2.5.5: Using the same kernel as in Figure 2.3, consider the function $u(x) = \left|x - \frac{1}{2}\right|$. Once again, we see that far from the corner, the nonlocal and classical deriavtives are nearly indistinguishable. But at the corner, the nonlocal derivative continuously joins the two branches of u', while the classical derivative jumps.

△

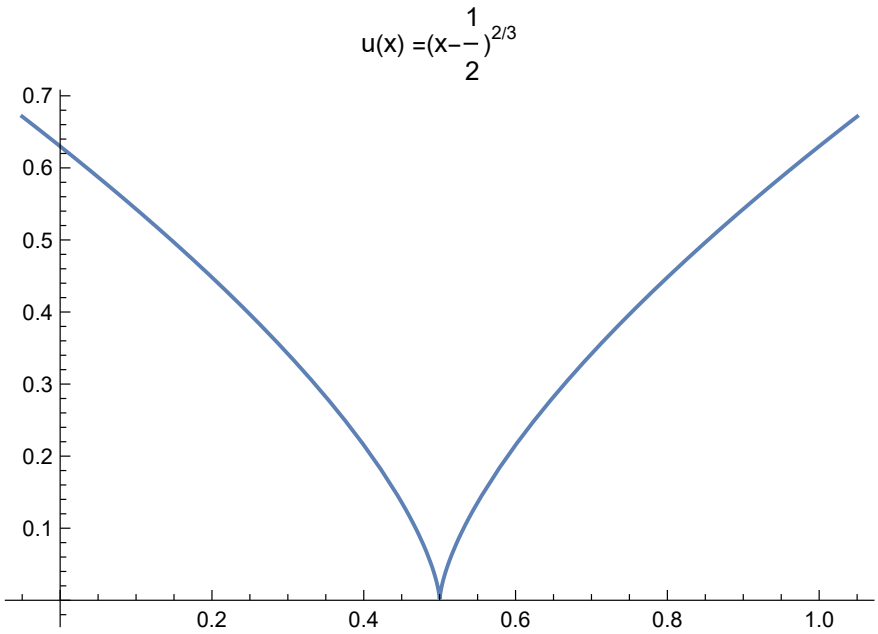


Figure 2.4: A function with a cusp.

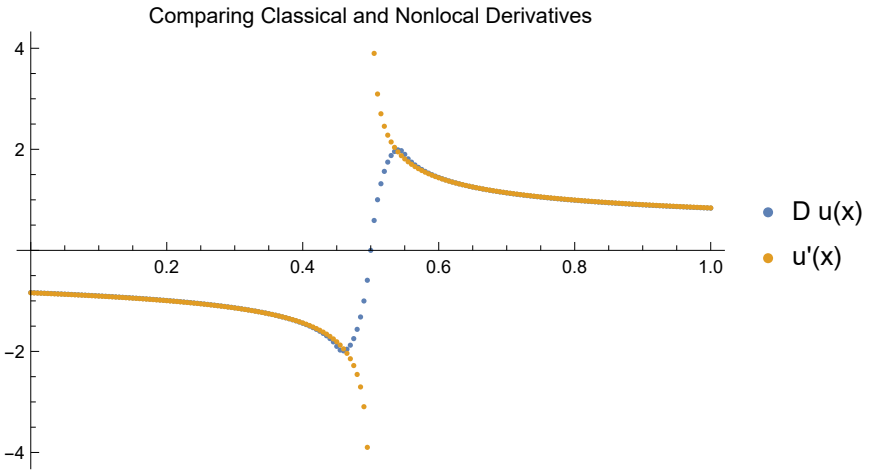


Figure 2.5: Away from the cusp, the nonlocal and classical derivatives agree. Near $x = \frac{1}{2}$, however, u' blows up, while $\mathcal{D}_\delta u$ connects the two branches relatively smoothly.

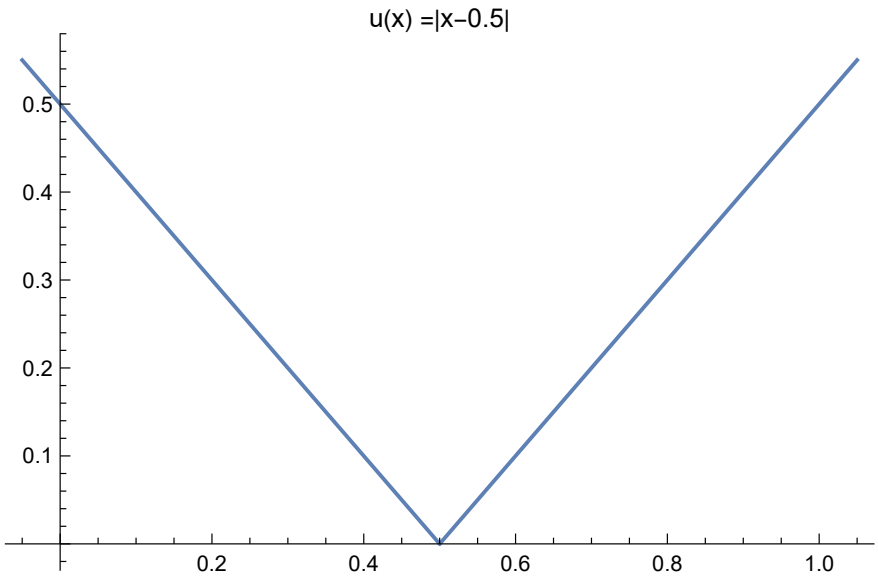


Figure 2.6: A function with a corner.

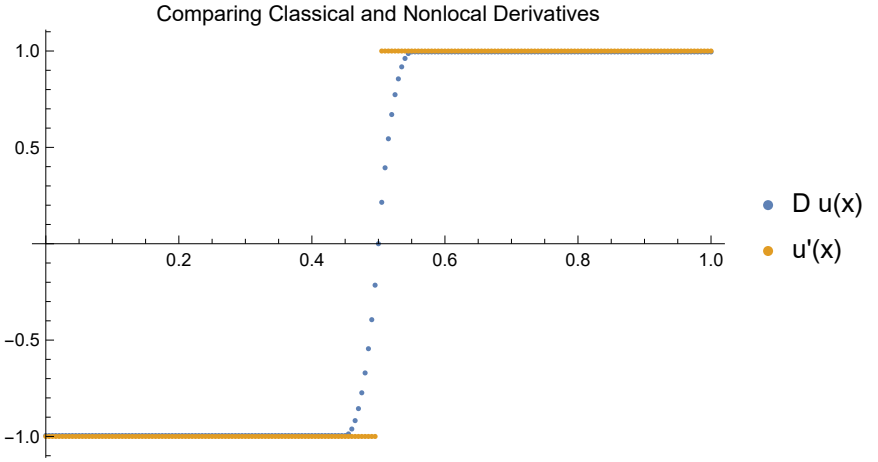


Figure 2.7: Just like with the cusp, the nonlocal derivative doesn't have the same sort of discontinuity that we see in u'.

Chapter 3

Existence and Uniqueness of Minimizers

Now we set up the minimization problem introduced in Section 2.2 in more detail. Immediate consequences of the initial assumptions are established in Section 3.1 for convenient reference later. The main results concerning existence and uniqueness are given in Theorems 3.2.1 and 3.3.2.

3.1 Notation and Assumptions

Fix any $p \in (1, \infty)$ and $\delta > 0$. Let $\Omega = (a, b)$ and $\Gamma = (a - \delta, a] \cup [b, b + \delta)$. Suppose that $\mu : (-\delta, \delta) \to \mathbb{R}$ is a function in $L^1\left((-\delta, \delta)\right)$. Let $\mathcal{D}$ denote the nonlocal differential operator given by Equation 2.3. We drop the subscript μ in this section since the kernel is fixed.

We consider a functional of the form

$$J[u] = \int_\Omega f(x, u, \mathcal{D}u)\, \mathrm{d}x,$$

where $f : \Omega \cup \Gamma \times \mathbb{R} \times \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ is in $L^1\left(\Omega \cup \Gamma \times \mathbb{R} \times \mathbb{R}\right)$. We are interested in problems where behavior on the collar is fixed, so let $u_0 : \Omega \cup \Gamma \to \mathbb{R}$ be a given function in $L^p\left(\Omega \cup \Gamma\right)$ and let $\mathcal{A}$ denote the admissible class of the functional J:

$$\mathcal{A} \coloneqq \left\{u \in L^p\left(\Omega \cup \Gamma\right) : u(x) = u_0(x) \text{ a.e. on } \Gamma\right\}.$$

Remark 3.1.1: There are other possible choices for the boundary conditions. For example, in numerical experiments (see Chapter 4), we found cases where boundary conditions of the type

above did not admit minimizers converging to the classical minimizer. Instead, we fixed the values just at the endpoints a and b, then extended the function u onto the collar via linear reflection. More on this later.

Before moving on to the main theorem of this section, we note a few properties that follow immediately from the assumptions made so far.

Lemma 3.1.2: Recall that f is defined on an open set, so every point in its domain is also in the interior of that set. By assumption 3, f is convex in its second and third arguments. It is therefore continuous in those arguments as well. (See Dacorogna 2008, Theorem 2.31 or Fonseca and Leoni 2007, Theorem 4.36.)

Lemma 3.1.3: The admissible class $\mathcal{A}$ is a convex set and the functional $J : \mathcal{A} \to \mathbb{R}$ is convex: for all $t \in [0, 1]$,

$$J[tu_1 + (1-t)u_2] \leq tJ[u_1] + (1-t)J[u_2].$$

Proof. To see that $\mathcal{A}$ is convex, select any $u, v \in \mathcal{A}$ and choose $t \in [0, 1]$. Since $u, v \in L^p(\Omega \cup \Gamma)$, and L^p is a vector space, any convex combination of u and v will stay in L^p. The only thing to check is that they agree with u_0 on the collar. By assumption, $u(x) = v(x) = u_0(x)$ for almost every $x \in \Gamma$. Thus, on the set Γ,

$$tu(x) + (1-t)v(x) = tu_0(x) + (1-t)u_0(x) = u_0(x) \text{ a.e.}$$

So any convex combination of u and v agrees with the boundary values and is therefore an element of $\mathcal{A}$.

Now we show that J is convex. Choose $u, v \in \mathcal{A}$. Using Assumption 3 and the linearity of $\mathcal{D}$

(see Proposition 2.4.1), we compute

$$\begin{aligned} J[tu+(1-t)v] &= \int_a^b f\left(x, tu+(1-t)v, \mathcal{D}\left(tu+(1-t)v\right)\right) \mathrm{d}x \\ &= \int_a^b f\left(x, tu+(1-t)v, t\mathcal{D}u+(1-t)\mathcal{D}v\right) \mathrm{d}x \\ &\leq \int_a^b t\cdot f\left(x,u,\mathcal{D}u\right)+(1-t)\cdot f\left(x,v,\mathcal{D}v\right) \mathrm{d}x \\ &= t\int_a^b f\left(x,u,\mathcal{D}u\right) \mathrm{d}x + (1-t)\int_a^b f\left(x,v,\mathcal{D}v\right) \mathrm{d}x \\ &= tJ[u]+(1-t)J[v]. \end{aligned}$$

□

Lemma 3.1.4: If there is at least one function $v \in \mathcal{A}$ such that $J[v] < \infty$, then there exists a minimizing sequence $\left\{u_j\right\}_{j=1}^{\infty}$ in $\mathcal{A}$ such that

$$\lim_{j\to\infty} J[u_j] = \inf\left\{J[u] : u \in \mathcal{A}\right\}.$$

Proof. This is immediate from the definition of a minimizing sequence. □

3.2 Existence of a Minimizer: The Direct Method

Now we are ready to apply the direct method, described in Section 1.1.3 to the problem set up in Sections 2.2 and 3.1.

Theorem 3.2.1: In addition to the definitions and assumptions given in 3.1, suppose that the following conditions hold:

1. There is a Poincaré Inequality for the operator $\mathcal{D}$. That is, there exists a constant C_p, depending only on p, such that

$$\int_\Omega |u|^p \,\mathrm{d}x \leq C_p \int_\Omega |\mathcal{D}u|^p \,\mathrm{d}x, \text{ for all } u \in L^p\left(\Omega\cup\Gamma\right) \text{ such that } u = 0 \text{ a.e. on } \Gamma. \tag{3.1}$$

2. There exists a function $v \in \mathcal{A}$ such that $J[v] < \infty$.

3. For every $x \in \Omega \cup \Gamma$, the map $(u, \xi) \mapsto f(x, u, \xi)$ is jointly convex on all of $\mathbb{R} \times \mathbb{R}$: for all $t \in [0, 1]$,

$$f\left(x, tu_1 + (1-t)u_2, t\xi_1 + (1-t)\xi_2\right) \leq tf\left(x, u_1, \xi_1\right) + (1-t)f\left(x, u_2, \xi_2\right). \tag{3.2}$$

4. The function f is coercive: there exists a constant $M > 0$ and a function $P \in L^1\left(\Omega \cup \Gamma\right)$ such that

$$f(x, u, \xi) \geq M|\xi|^p + P(x), \text{ for all } (x, u, \xi) \in \Omega \cup \Gamma \times \mathbb{R} \times \mathbb{R}. \tag{3.3}$$

Then there exists a minimizer of J over the set $\mathcal{A}$. That is, there is some $\hat{u} \in \mathcal{A}$ such that $J[\hat{u}] = \inf_{u \in \mathcal{A}} J[u]$.

Before proving the theorem, a few comments on the assumptions made above.

Remark 3.2.2: We chose $p > 1$ to ensure that L^p is a reflexive Banach space. Since the dual of L^1 is L^∞, but the dual of L^∞ is not generally L^1, the argument given below does not go through for $p = 1$. However, using the Dunford-Pettis theorem (Fonseca and Leoni 2007, Theorem 2.54), equi-integrability, and our coercivity condition, it should be possible to make the argument work for $p = 1$ as well.

Remark 3.2.3: The joint convexity condition is needed to ensure lower semi-continuity of the functional J. In higher dimensions, this assumption can be weakened. For further discussion of convexity in the classical case, see Dacorogna 2008 or Sections 2.19 and 2.20 of Cesari 1983. More recent work on lower semi-continuity in the nonlocal setting includes Elbau 2011, Kreisbeck and Schönberger 2022, Bellido et al. 2020, Cueto 2021, and Kreisbeck and Zappale 2020.

Remark 3.2.4: Assuming that there is a Poincaré-type Inequality for this particular nonlocal setting is a nontrivial assumption. This type of bound is essential for the standard arguments, but

it is unclear whether any existing results would apply to this context (without further assumptions, for example, on the kernel μ). For some examples of classical Poincaré-tpe inequalities, see Leoni 2009, Theorems 7.21. 7.23, and Section 12.2 or Ziemer 2012, Chapter 4. For some examples of nonlocal analogs, see Bobaru, Foster, et al. 2016, p. 78 and the associated references, Mengesha and Du 2014, Proposition 2.7, Mengesha and Du 2013, Proposition 2, Cueto 2021, Theorem 6.4.2, and Andreu-Vaillo et al. 2010, Propositions 5.3, 5.5, 6.19, and 6.25.

Now onto the proof of Theorem 3.2.1. The goal is to construct an element $\hat{u} \in \mathcal{A}$ so that

$$J[\hat{u}] = \inf_{u \in \mathcal{A}} J[u]. \tag{3.4}$$

The general idea is to establish uniform bounds on our minimizing sequence $\{u_j\}_{j=1}^{\infty}$ so we can extract a weakly convergent subsequence. Then Mazur's Lemma provides a related (strongly) convergent sequence. Applying convexity, we show that the limit of that new sequence is in fact a minimizer of J over $\mathcal{A}$.

Proof. Let $\{u_j\}_{j=1}^{\infty}$ be any minimizing sequence given by Lemma 3.1.4. Since f is coercive, for each $j \in \mathbb{N}$ and all $x \in \Omega$ we have

$$P(x) + M\left|\mathcal{D}u_j(x)\right|^p \leq f(x, u_j, \mathcal{D}u_j).$$

Since this holds for the whole interval, we can integrate both sides to obtain

$$\int_\Omega P(x)\,\mathrm{d}x + M\int_\Omega \left|\mathcal{D}u_j(x)\right|^p \mathrm{d}x \leq \int_\Omega f(x, u, \mathcal{D}u)\,\mathrm{d}x = J[u_j].$$

Since $P \in L^1(\Omega)$, we can find $K \in \mathbb{R}$ such that $\int_\Omega P(x)\,\mathrm{d}x < K$. Thus, we can rearrange to find the following inequality for each $j \in \mathbb{N}$:

$$\int_\Omega \left|\mathcal{D}u_j(x)\right|^p \mathrm{d}x \leq \frac{J[u_j] - K}{M}.$$

Since $\{u_j\}$ is a minimizing sequence and Assumption 2 guarantees $J[u_j]$ is not always infinite, the $J[u_j]$ can be bounded independently of j. That is, there exists a constant $R < \infty$ such that, for all j, $\frac{J[u_j]-K}{M} < R$. Thus, we see that coercivity ensures

$$\int_\Omega \left|\mathcal{D}u_j(x)\right|^p \mathrm{d}x < R, \ \forall j \in \mathbb{N}. \tag{3.5}$$

We'd like to use this to get a uniform bound on our minimizing sequence. Recall that our Poincaré Inequality only applied to functions that are 0 on the collar. Thus, for each $j \in \mathbb{N}$ we introduce a new function w_j defined by

$$w_j \coloneqq u_j - u_0. \tag{3.6}$$

Since each $u_j \in \mathcal{A}$, we know $u_j(x) = u_0(x)$ for any $x \in \Gamma$. Thus, $w_j = 0$ on the collar, and we can apply Poincaré's Inequality (Assumption 1):

$$\int_\Omega \left|w_j(x)\right|^p \mathrm{d}x \leq C_p \int_\Omega \left|\mathcal{D}w_j(x)\right|^p \mathrm{d}x. \tag{3.7}$$

We'd like to express the right side in terms of u_j so that we can apply Equation 3.5. In Proposition 2.4.1, we saw that the operator $\mathcal{D}$ is linear, so

$$\int_\Omega \left|\mathcal{D}w_j(x)\right|^p \mathrm{d}x = \int_\Omega \left|\mathcal{D}u_j(x) - \mathcal{D}u_0(x)\right|^p \mathrm{d}x. \tag{3.8}$$

Since $p > 1$, the map $x \mapsto |x|^p$ is convex. Thus, viewing the right hand side of the above equation as a convex combination, we can further break up the integral:

$$\begin{aligned}
\left|\mathcal{D}u_j(x) - \mathcal{D}u_0(x)\right|^p &= \left|\frac{1}{2}\left(2\mathcal{D}u_j(x)\right) + \frac{1}{2}\left(-2\mathcal{D}u_0(x)\right)\right|^p \\
&\leq \frac{1}{2}\left|2\mathcal{D}u_j(x)\right|^p + \frac{1}{2}\left|-2\mathcal{D}u_0(x)\right|^p \\
&= 2^{p-1}\left|\mathcal{D}u_j(x)\right|^p + 2^{p-1}\left|\mathcal{D}u_0(x)\right|^p
\end{aligned}$$

(This is a standard trick; see for example Fonseca and Leoni 2007, p. 259.) Plugging this back into Equation 3.8 yields

$$\int_\Omega \left|\mathcal{D}w_j(x)\right|^p \,\mathrm{d}x \leq 2^{p-1} \int_\Omega \left|\mathcal{D}u_j(x)\right|^p \,\mathrm{d}x + 2^{p-1} \int_\Omega \left|\mathcal{D}u_0(x)\right|^p \,\mathrm{d}x.$$

The rightmost term, involving $\int_\Omega \left|\mathcal{D}u_0(x)\right|^p \,\mathrm{d}x$, is some finite constant determined by the boundary conditions, call it C_0. Putting this back together with the inequality given in Equation 3.7, we have

$$\int_\Omega \left|w_j(x)\right|^p \,\mathrm{d}x \leq C_p 2^{p-1} \left(\int_\Omega \left|\mathcal{D}u_j(x)\right|^p \,\mathrm{d}x + C_0 \right). \tag{3.9}$$

Combining this with Equation 3.5, we obtain a bound independent of j:

$$\int_\Omega \left|w_j(x)\right|^p \,\mathrm{d}x < C_p 2^{p-1} \left(R + C_0\right) < \infty, \ \forall j \in \mathbb{N}. \tag{3.10}$$

Now that we have a uniformly bounded sequence $\left\{w_j\right\}_{j=1}^\infty$ in $L^p(\Omega \cup \Gamma)$, we apply Lemmas 2.3.7 and 2.3.8 to find a weakly convergent subsequence. Now we can apply Mazur's Lemma (2.3.9) to find a sequence $\{y_n\}_{n=1}^\infty$ of convex combinations of the elements of $\left\{w_{j_k}\right\}_{k=1}^\infty$ such that y_n converges strongly to $\hat{y}$. More precisely, for every n, there exists a set of nonnegative real numbers $\left\{\lambda_k^{(n)}\right\}_{k=n}^{M_n}$ such that

1. $y_n = \sum_{k=n}^{M_n} \lambda_k^{(n)} w_{j_k}$,
2. $\sum_{k=n}^{M_n} \lambda_k^{(n)} = 1$, and
3. $y_n \to \hat{y}$ in the L^p norm.

Note that applying the operator $\mathcal{D}$ to this sequence yields a strongly convergent sequence in $L^p(\Omega)$: Proposition 2.4.1 shows $\mathcal{D}$ is linear and bounded, so

$$\left\|\mathcal{D}y_n - \mathcal{D}\hat{y}\right\|_{L^p(\Omega)} = \left\|\mathcal{D}(y_n - \hat{y})\right\|_{L^p(\Omega)} \leq \|\mathcal{D}\| \left\|y_n - \hat{y}\right\|_{L^p(\Omega)},$$

which can be made arbitrarily small by Property 3 above.

Recall that we defined each w_j by subtracting off u_0 from the corresponding u_j, which was a member of the original minimizing sequence. We can rewrite Property 1 of the y_n above as follows,

$$y_n = \sum_{k=n}^{M_n} \lambda_k^{(n)} w_{j_k} = \sum_{k=n}^{M_n} \lambda_k^{(n)} (u_{j_k} - u_0) = \sum_{k=n}^{M_n} \left(\lambda_k^{(n)} u_{j_k}\right) - \sum_{k=n}^{M_n} \left(\lambda_k^{(n)} u_0\right)$$

But u_0 is independent of the indices we are summing over, so we can pull it out of the last summation and use the fact that $\sum_{k=n}^{M_n} \lambda_k^{(n)} = 1$ to obtain

$$y_n = \sum_{k=n}^{M_n} \left(\lambda_k^{(n)} u_{j_k}\right) - u_0. \tag{3.11}$$

Next, for each n define

$$\hat{u}_n \coloneqq y_n + u_0 = \sum_{k=n}^{M_n} \lambda_k^{(n)} u_{j_k} \tag{3.12}$$

These $\hat{u}_n$ are all in the admissible class $\mathcal{A}$, since they are finite linear combinations of elements of $\mathcal{A}$. Next, define $\hat{u} = \hat{y} + u_0$, so that

$$\hat{u} = \hat{y} + u_0 = \lim_{n\to\infty} y_n + \lim_{n\to\infty} u_0 = \lim_{n\to\infty} (y_n + u_0) = \lim_{n\to\infty} \hat{u}_n.$$

Note that $\mathcal{A}$ is a closed subset of a complete vector space, and $\hat{u}$ is defined as the limit of elements of $\mathcal{A}$. Thus, $\hat{u} \in \mathcal{A}$. Our goal now is to show this $\hat{u}$ we've constructed is the minimizer of J: $J[\hat{u}] = \inf\left\{J[u] : u \in \mathcal{A}\right\}$.

Evaluating J using Equation 3.12 and the linearity of $\mathcal{D}$, we have the following for every $n \in \mathbb{N}$:

$$\begin{aligned}
J[\hat{u}_n] &= \int_\Omega f(x, \hat{u}_n, \mathcal{D}\hat{u}_n)\,\mathrm{d}x \\
&= \int_\Omega f\left(x, \sum_{k=n}^{M_n} \lambda_k^{(n)} u_{j_k}, \mathcal{D}\left(\sum_{k=n}^{M_n} \lambda_k^{(n)} u_{j_k}\right)\right)\mathrm{d}x \\
&= \int_\Omega f\left(x, \sum_{k=n}^{M_n} \lambda_k^{(n)} u_{j_k}, \sum_{k=n}^{M_n} \lambda_k^{(n)} \mathcal{D} u_{j_k}\right)\mathrm{d}x
\end{aligned}$$

From the convexity of J (Lemmas 2.3.10 and 3.1.3), we can pull the sums out to get an upper

bound:

$$J[\hat{u}_n] \leq \sum_{k=n}^{M_n} \lambda_k^{(n)} \int_\Omega f\left(x, u_{j_k}, \mathcal{D}u_{j_k}\right) \mathrm{d}x = \sum_{k=n}^{M_n} \lambda_k^{(n)} J[u_{j_k}]. \tag{3.13}$$

Now fix an $\varepsilon > 0$. Since $\left\{u_{j_k}\right\}_{k=1}^{\infty}$ is a subsequence of a minimizing sequence, we know there is some $N = N(\varepsilon) \in \mathbb{N}$ such that $J[u_{j_k}] < \inf_{u \in \mathcal{A}} J[u] + \varepsilon$ whenever $k \geq N$. Setting $n = N$ in Equation 3.13, we see that

$$\begin{aligned} J[\hat{u}_N] &\leq \sum_{k=N}^{M_n} \lambda_k^{(n)} J[u_{j_k}] \\ &< \sum_{k=N}^{M_n} \lambda_k^{(n)} \left(\inf_{u \in \mathcal{A}} J[u] + \varepsilon \right) \\ &= \left(\inf_{u \in \mathcal{A}} J[u] + \varepsilon \right) \sum_{k=N}^{M_n} \lambda_k^N \\ &= \inf_{u \in \mathcal{A}} J[u] + \varepsilon, \end{aligned}$$

where we used Property 2 of the coefficients λ_k^N. Since this holds for all $\varepsilon > 0$, we conclude

$$\lim_{n \to \infty} J[\hat{u}_n] \leq \inf_{u \in \mathcal{A}} J[u].$$

From the definition of infimum, we also have $\lim_{n\to\infty} J[\hat{u}_n] \leq \inf_{u\in\mathcal{A}} J[u]$. Hence,

$$\lim_{n \to \infty} J[\hat{u}_n] = \inf_{u \in \mathcal{A}} J[u]. \tag{3.14}$$

Next we use the fact that $\mathcal{D}$ is a continuous linear operator and that f is continuous in its second

and third arguments (Lemmas 2.4.1 and 3.1.2, respectively) to obtain the following

$$\begin{aligned} J[\hat{u}] &= \int_\Omega f(x, \hat{u}, \mathcal{D}\hat{u})\,\mathrm{d}x \\ &= \int_\Omega f(x, \lim_{n\to\infty} \hat{u}_n, \mathcal{D}\left(\lim_{n\to\infty} \hat{u}_n\right))\,\mathrm{d}x \\ &= \int_\Omega f(x, \lim_{n\to\infty} \hat{u}_n, \lim_{n\to\infty} \mathcal{D}\hat{u}_n)\,\mathrm{d}x \\ &= \int_\Omega \lim_{n\to\infty} f(x, \hat{u}_n, \mathcal{D}\hat{u}_n)\,\mathrm{d}x \end{aligned}$$

The growth condition of assumption 4 implies that $f(x, \hat{u}_n(x), \mathcal{D}\hat{u}_n(x))$ is bounded below by the measurable function $M\left|\mathcal{D}\hat{u}_n(x)\right|^p + P(x)$. Thus, for each n, $P(x) \leq f(x, \hat{u}_n(x), \mathcal{D}\hat{u}_n(x))$. Then we apply a variant of Fatou's Lemma (2.3.4) and Equation 3.14 to obtain

$$J[\hat{u}] = \int_\Omega \lim_{n\to\infty} f(x, \hat{u}_n, \mathcal{D}\hat{u}_n)\,\mathrm{d}x \leq \lim_{n\to\infty} \int_\Omega f(x, \hat{u}_n, \mathcal{D}\hat{u}_n)\,\mathrm{d}x = \lim_{n\to\infty} J[\hat{u}_n] = \inf_{u\in\mathcal{A}} J[u].$$

Therefore, $\hat{u} \in \mathcal{A}$ minimizes the functional J over our chosen admissible class. □

3.3 Uniqueness

There is no reason to think that a convex functional will have a unique minimum. To ensure uniqueness, a stronger convexity assumption is needed.

Definition 3.3.1: We say a function $f : \Omega \cup \Gamma \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is **strictly convex** in its second and third arguments if, for all $x \in \Omega \cup \Gamma$ and all $u_1, u_2, \xi_1, \xi_2 \in \mathbb{R}$ such that $u_1 \neq u_2$ and $\xi_1 \neq \xi_2$, the function f satisfies

$$f(x, tu_1 + (1-t)u_2, t\xi_1 + (1-t)\xi_2) < tf(x, u_1, \xi_1) + (1-t)f(x, u_2, \xi_2). \tag{3.15}$$

Theorem 3.3.2: Suppose that the conditions of Theorem 3.2.1 hold. Further assume that f is strictly convex in its second and third arguments. Then there exists a unique (up to equivalence almost everywhere) $\hat{u}$ minimizing J over the set $\mathcal{A}$.

Remark 3.3.3: The proof given below is no different than what happens in the classical case. See for example page 108 of Dacorogna 2008 or page 449 of Evans 1998.

Proof. Suppose that $\hat{u}$ and $\hat{v}$ are two minimizers of the functional J over the set $\mathcal{A}$:

$$J[\hat{u}] = J[\hat{v}] = \inf_{u \in \mathcal{A}} J[u].$$

Then let $w = \frac{1}{2}(\hat{u} + \hat{v})$. From Lemma 3.1.3, $\mathcal{A}$ is a convex set (so $w \in \mathcal{A}$) and J is a convex functional, so

$$J[w] = J\left[\frac{1}{2}\hat{u} + \frac{1}{2}\hat{v}\right] \leq \frac{1}{2}J[\hat{u}] + \frac{1}{2}J[\hat{v}] = \inf_{u \in \mathcal{A}} J[u].$$

Thus, w is a minimizer of J. It immediately follows that

$$\int_\Omega \frac{1}{2} f(x, \hat{u}, \mathcal{D}\hat{u})\, \mathrm{d}x + \int_\Omega \frac{1}{2} f(x, \hat{v}, \mathcal{D}\hat{v})\, \mathrm{d}x - \int_\Omega f(x, \hat{w}, \mathcal{D}\hat{w})\, \mathrm{d}x = 0. \tag{3.16}$$

Now, for every x, the map $(u, \xi) \mapsto f(x, u, \xi)$ is convex. With the linearity of $\mathcal{D}$, this implies

$$f(x, w, \mathcal{D}w) = f\left(x, \frac{\hat{u} + \hat{v}}{2}, \mathcal{D}\left(\frac{\hat{u} + \hat{v}}{2}\right)\right) \leq \frac{1}{2} f(x, \hat{u}, \mathcal{D}\hat{u}) + \frac{1}{2} f(x, \hat{v}, \mathcal{D}\hat{v}).$$

So

$$\frac{1}{2} f(x, \hat{u}, \mathcal{D}\hat{u}) + \frac{1}{2} f(x, \hat{v}, \mathcal{D}\hat{v}) - f(x, \hat{w}, \mathcal{D}\hat{w}) \geq 0,$$

and Equation 3.16 implies that the integral of the left side quantity is 0. Therefore,

$$\frac{1}{2} f(x, \hat{u}, \mathcal{D}\hat{u}) + \frac{1}{2} f(x, \hat{v}, \mathcal{D}\hat{v}) - f(x, \hat{w}, \mathcal{D}\hat{w}) = 0 \text{ a.e. in } \Omega. \tag{3.17}$$

Now suppose for the sake of contradiction that there is a set of strictly positive measure, S, such that $\hat{u}(x) \neq \hat{v}(x)$ for all $x \in S$. Then the strict convexity of f (Equation 3.15) implies

$$f(x, w, \mathcal{D}w) = f\left(x, \frac{\hat{u} + \hat{v}}{2}, \mathcal{D}\left(\frac{\hat{u} + \hat{v}}{2}\right)\right) < \frac{1}{2} f(x, \hat{u}, \mathcal{D}\hat{u}) + \frac{1}{2} f(x, \hat{v}, \mathcal{D}\hat{v}),$$

for all $x \in S$. In particular, $\frac{1}{2}f(x, \hat{u}, \mathcal{D}\hat{u}) + \frac{1}{2}f(x, \hat{v}, \mathcal{D}\hat{v}) - f(x, w, \mathcal{D}w) > 0$ on a set of positive measure, contradicting Equation 3.17. Thus, if $\hat{u}$ and $\hat{v}$ differ at all, it cannot be on a set of positive measure. □

Chapter 4

Numerical Examples and Discussion

In Theorems 3.2.1 and 3.3.2, we showed the minimization problems of 2.2 are well-posed in a fairly general setting: the kernel was only required to be in L^1 and the values on the collar were required to match some function $u_0 \in L^p$. Now we would like to explore the consequences of different choices of μ and different boundary conditions to get a better grasp of the problems at hand. For this, we turn to numerical simulations performed in *Mathematica*. The details of the numerical scheme along with the complete code can be found in Appendix A. Rather than optimizing for numerical precision or computational speed, the algorithm used was chosen due to its simplicity and ease of presentation. Additionally, we will only discuss relatively nice problems, where additional precision is unlikely to grant any additional insight. Thus, these results should be seen as suggestive at most; the investigation was not sufficiently thorough to come to any firm conclusions.

When comparing classical and nonlocal variational problems, there is a question of how to extend the classical boundary conditions to the new problem. We consider two possibilities: constant extensions and linear reflections. Details of this implementation can be found in the code itself; Figure 4.1 should be sufficiently clear to illustrate the basic concepts.

In this section, the notation $\|u - v\|$ without a subscript denotes the Euclidean norm, measuring distance between finite dimensional vectors $u, v \in \mathbb{R}^n$. The symbol $\|u - v\|_\infty$ measures the maximum difference between the entries of u and v. Throughout, $\hat{u}_c$ refers to the minimizer of the classical problem and $\hat{u}_n$ is the approximate minimizer of the nonlocal problem. The plots labeled "Final" show the conclusion of the gradient descent algorithm, with the blue function labeled "u" denoting the approximate minimizer to the nonlocal problem. In the plots labeled "Nonlocal

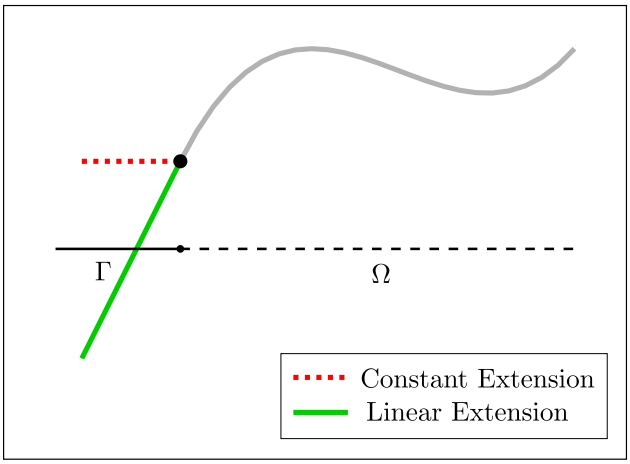


Figure 4.1: Linear versus constant extensions. Image adapted from Foss, Radu, and Yu 2022.

Derivative," the blue data, "w", denotes the approximate nonlocal derivative.

4.1 Constant Extensions

Consider the classical functional

$$J[u] = \int_0^1 u^2 + u'^2 \, \mathrm{d}x,$$

with admissible class

$$\mathcal{A} = \left\{ u \in W^{1,1}\left([0,1]\right) : u(0) = 0, \ u(1) = 1 \right\}.$$

From the Euler-Lagrange Equation, it can be shown that the minimizer is given by

$$\hat{u}_c(x) = \frac{e^x - e^{-x}}{e - e^{-1}}. \tag{4.1}$$

To convert this to a nonlocal problem, we replace u' with $\mathcal{D}u$ and adjust the admissible class. Instead of only fixing the endpoints, Theorem 3.2.1 requires specifying the values of u on the entire collar, $(-\delta, 0] \cup [1, 1+\delta)$. Perhaps the simplest way to convert the classical boundary conditions to the nonlocal setting is by constant extension. So we first consider the following nonlocal variation problem.

Example 4.1.1: Select $\delta \in (0, 1]$ and consider the domain $\Omega = [0, 1]$ with collar $\Gamma = (-\delta, 0] \cup$

$[1, 1+\delta)$. Let $\mu : (-\delta, \delta) \to \mathbb{R}$ be given by

$$\mu_\delta(z) = \begin{cases} 0, & x \leq 0 \\ \frac{2}{\delta^2}, & x > 0. \end{cases} \tag{4.2}$$

No matter which δ is chosen, this kernel satisfies

1. the corresponding $\mathcal{D}_\delta$ given by Equation 2.3 satisfies the Poincaré Inequality in Equation 3.1 (See Example 3.1 in Foss 2019),
2. $\int_{-\delta}^{\delta} \left| z\mu_\delta(z) \right| \mathrm{d}z < \infty$, and
3. $\int_{-\delta}^{\delta} z\mu_\delta(z) \, \mathrm{d}z = 1$.

Thus, the conditions of Theorem 2.4.2, its corollary, and Theorem 3.2.1 are all satisfied. Then define the functional

$$J_\delta[u] = \int_0^1 u^2 + (\mathcal{D}_\delta u)^2 \, \mathrm{d}x$$

with admissible class

$$\mathcal{A} \coloneqq \left\{ u \in L^p\left([-\delta, 1+\delta]\right) : u(x) = 0 \text{ for all } x \leq 0, \text{ and } u(x) = 1 \text{ for all } x \geq 1 \right\}.$$

What is the minimizer of J_δ over the set $\mathcal{A}$? How does it compare to the classical minimizer as δ approaches 0? △

The functional J_δ satisfies the conditions of Theorems 3.2.1 and 3.3.2, so there should be a unique minimizer. Using the numerical methods described in Appendix A, we can approximate this minimizer by converting the problem into a finite dimensional optimization problem.

Remark 4.1.2: Besides the many possible sources of error within the specific algorithm used, one should be skeptical of even this basic concept. For general functions in L^p, we have very little reason to think that a finite sampling of points will do a very good job of capturing the behavior of

the function. The hope is that, for small δ, the nonlocal problem behaves like the classical problem. (This is suggested, but not guaranteed, by Corollary 2.4.4.) Since the minimizer to the classical problem is quite smooth, it can be approximated reasonably well via finite sampling.

Remark 4.1.3: Extremely limited initial testing suggests that changing the initial function for the gradient descent algorithm does not change the outcome much. For simplicity, and to control this potential variable, all of the figures shown in this chapter use the following function as the initial data:

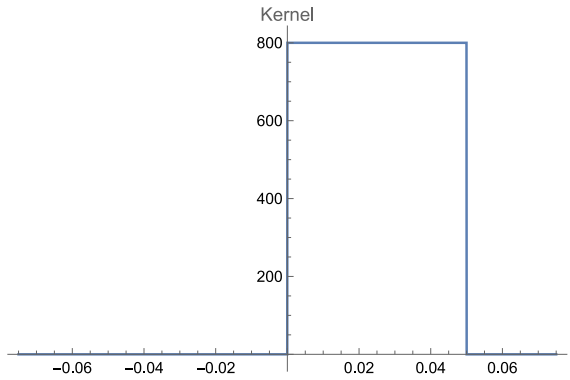
$$u(x) = \begin{cases} 0, & x \leq 0 \\ x + \sin(2\pi x), & x \in (0,1) \\ 1, & x \geq 1. \end{cases}$$

All parameters, like δ and the step size, are also fixed. These values can be found in Appendix A.

Example 4.1.1 fixes the value of u on the collar independent of what the function u is doing inside the interval $[0,1]$. Thus, in general, the nonlocal derivative does not agree well with the classical derivative for points near 0 or 1. Also, the kernel chosen in Equation 4.2 is only non-zero to the right of $z = 0$. Hence, the operator $\mathcal{D}_\delta$ can only "see" what happens to its right; points to the left of x make no contribution to the integral in the definition of $\mathcal{D}_\delta$. This can be see in Figure 4.2(c): the values of $\mathcal{D}_\delta u$ are shifted and the collar values for $x < 0$ make no impact at all.

Because of this last observation, there is nothing enforcing the boundary conditions near 0: u can be discontinuous at 0 with a large jump without being "penalized." Figure 4.2(d) shows the result of the gradient descent method applied to Example 4.1.1. The nonlocal minimizer behaves, in an extremely rough sense, qualitatively similarly to the classical minimizer. But it is not clear that the nonlocal minimizer converges to the classical as $\delta \to 0$. More testing is needed.

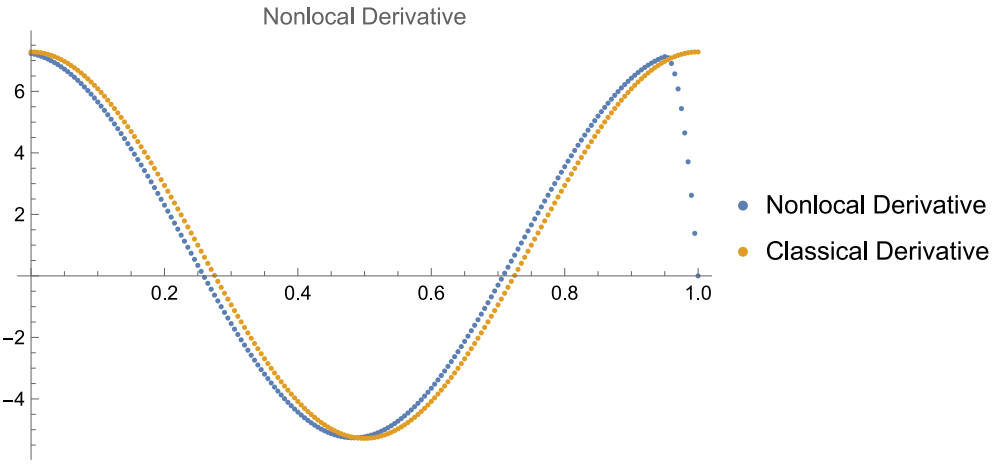
Keeping the constant collar values and the same kernel, let's look at the behavior of a different (but similar) functional. The action functional (recall Equation 1.4) for a simple harmonic oscillator



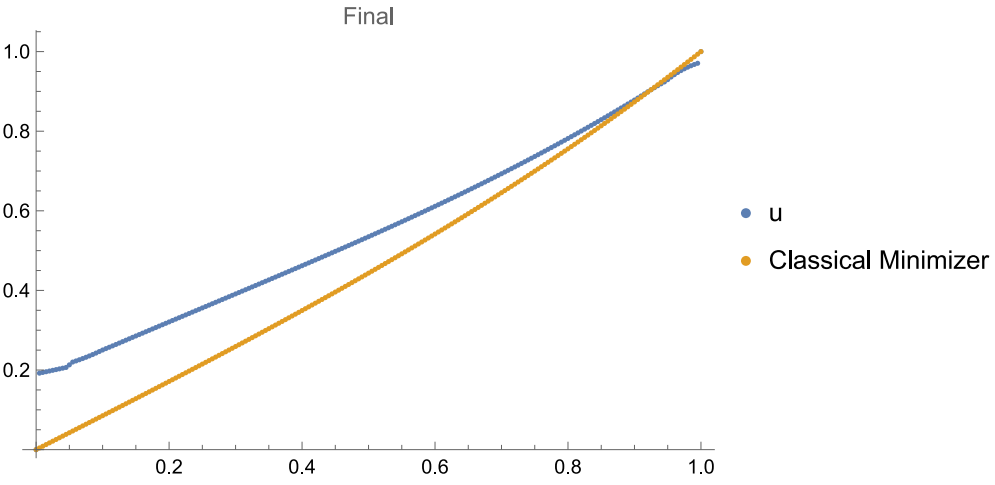
(a)

$\|\mathcal{D}u - u'\| \approx 13.2991$
$\|\mathcal{D}u - u'\|_\infty \approx 7.28319$
$J[\hat{u}_c] \approx 1.31305$
$J_\delta[\hat{u}_n] \approx 1.16373$
$\|\hat{u}_c - \hat{u}_n\| \approx 1.49413$
$\|\hat{u}_c - \hat{u}_n\|_\infty \approx 0.187401$

(b)



(c)



(d)

Figure 4.2: An approximate minimizer for the problem in Example 4.1.1. Near $x = 1$, the nonlocal derivative does not accurately capture the classical derivative. The gradient descent method did converge to something, but it doesn't look like the nonlocal minimizer is converging to the classical minimizer.

is given by

$$H[u] = \int_0^1 2u'^2 - \frac{\pi^2}{2}u^2 \, \mathrm{d}x. \tag{4.3}$$

The corresponding Euler-Lagrange Equation (which is just Hooke's Law) gives the minimizer as

$$\hat{u}_c(x) = \sin\left(\frac{\pi x}{2}\right).$$

Now we convert this to a nonlocal problem as in the previous example

Example 4.1.4: Select $\delta \in (0, 1]$ and consider the domain $\Omega = [0, 1]$ with collar $\Gamma = (-\delta, 0] \cup [1, 1 + \delta)$. Let $\mu : (-\delta, \delta) \to \mathbb{R}$ be given by

$$\mu_\delta(z) = \begin{cases} 0, & x \leq 0 \\ \frac{2}{\delta^2}, & x > 0. \end{cases} \tag{4.4}$$

Define the functional

$$H_\delta[u] = \int_0^1 2(\mathcal{D}_\delta u)^2 - \frac{\pi^2}{2}u^2 \, \mathrm{d}x$$

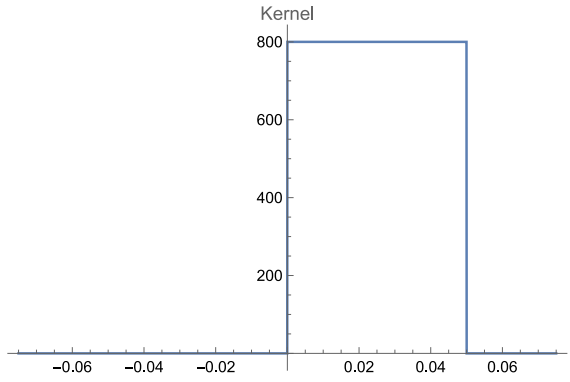
with admissible class

$$\mathcal{A} \coloneqq \left\{ u \in L^p\left([-\delta, 1 + \delta]\right) : u(x) = 0 \text{ for all } x \leq 0, \text{ and } u(x) = 1 \text{ for all } x \geq 1 \right\}.$$

How does the minimizer of H_δ compare to the classical minimizer of H? △

The result of the numerical approximation is shown in Figure 4.3. Again, the nonlocal minmizer does not satisfy $\lim_{x \to 0^+} u(x) = 0$, because $\mathcal{D}_\delta$ does not pick up this discontinuity. Apart from this difference, the nonlocal and classical minimizers are qualitatively similar.

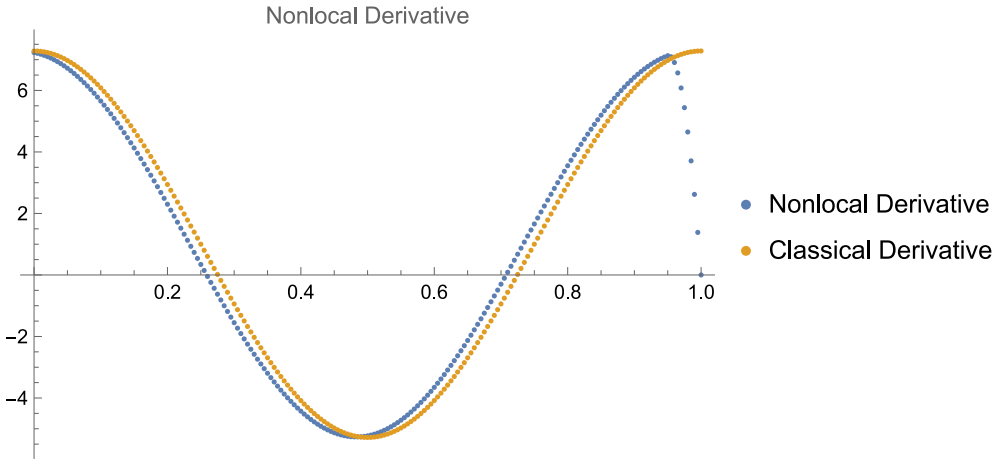
For the same constant collar values, it is interesting to see the effects of different choices of kernel. As noted in Example 2.5.1, an anti-symmetric kernel likely does a better job of behaving like the classical derivative. Thus, return to Example 4.1.1 but replace the kernel with the anti-symmetric,



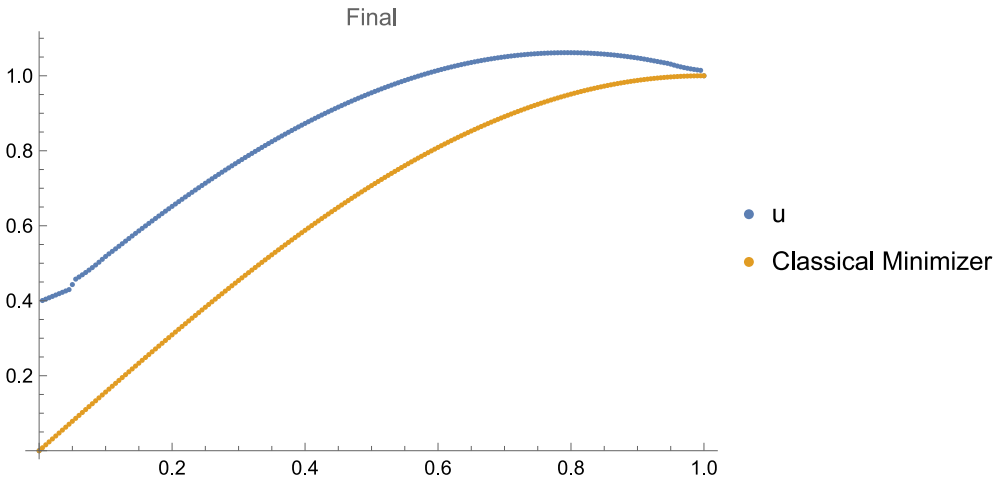
(a)

$\left\|\mathcal{D}u - u'\right\| \approx 13.2991$
$\left\|\mathcal{D}u - u'\right\|_\infty \approx 7.28319$
$J[\hat{u}_c] \approx 0.0$
$J_\delta[\hat{u}_n] \approx -1.22377$
$\|\hat{u}_c - \hat{u}_n\| \approx 3.59452$
$\|\hat{u}_c - \hat{u}_n\|_\infty \approx 0.393176$

(b)



(c)



(d)

Figure 4.3: For the harmonic oscillator functional of Example 4.1.4, with constant extensions, the nonlocal variational problem has a solution roughly similar to the local problem. Again, the graph is qualitatively similar but not particularly close to the classical minimizer, especially near $x = 0$.

piece-wise constant kernel shown in Figure 4.4.

Note that this kernel satisfies the hypotheses of Theorem 2.4.2, so it should behave like u' for small δ. In fact, the examples of Section 2.5 suggest that an anti-symmetric kernel should do an even better job of approximating the classical problem than the asymmetric kernl of the previous figures. However, the corrsponding nonlcoal operator may not have a Poincaré Inequality, so the existence and uniqueness of a minimizer is not guaranteed by Theorem 3.2.1. Either a more general Poincaré Inequality or some other further justification is needed.

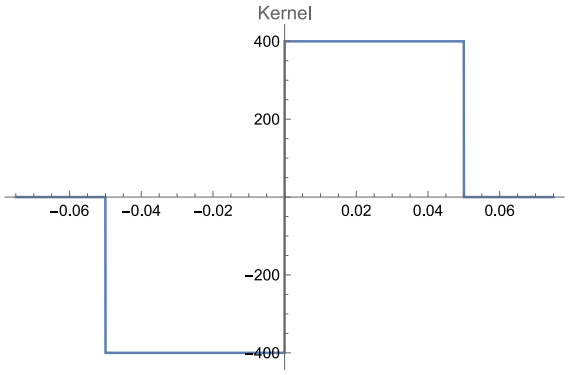
As Figure 4.4(c) shows, this kernel does a better job of approximating the derivative than the asymmetric kernel of Example 4.1.1. However, the nonlocal minimizer found by the iterative process behaves surprisingly differently from the classical minimizer. This is shown in Figure 4.4(d).

More research is needed to determine what exactly is going on here. The oscillations shown do not seem to depend on the size of δ or the mesh size h. A similar phenomenon occurs for other choices of anti-symmetric kernel as well; it is unclear why. Additionally, when the same anti-symmetric, piecewise-constant kernel is applied to the simple harmonic oscillator problem, the algorithm does not converge at all. It runs indefinitely, producing sharply discontinuous values of u. Curiously, similar oscillations do not occur when the function is linearly reflected onto the collar, as we see in the next section.

4.2 Linear Reflections

As seen in the figures from the last section, extending u by a constant generally causes sharp corners in the graph of u. This means $\mathcal{D}u(x)$ is not a good approximation to $u'(x)$ for x near the end points. One way to mitigate this effect is by reflecting the values of u across the endpoint. This reduces the spike in $\mathcal{D}u$ near the endpoints, but it also comes at a cost. Since the values on the collar now depend on the values of u inside the interval, Theorem 3.2.1 no longer applies. However, the numerical results are somewhat promising in this case. Fewer strange cases have been found, motivating the following conjecture.

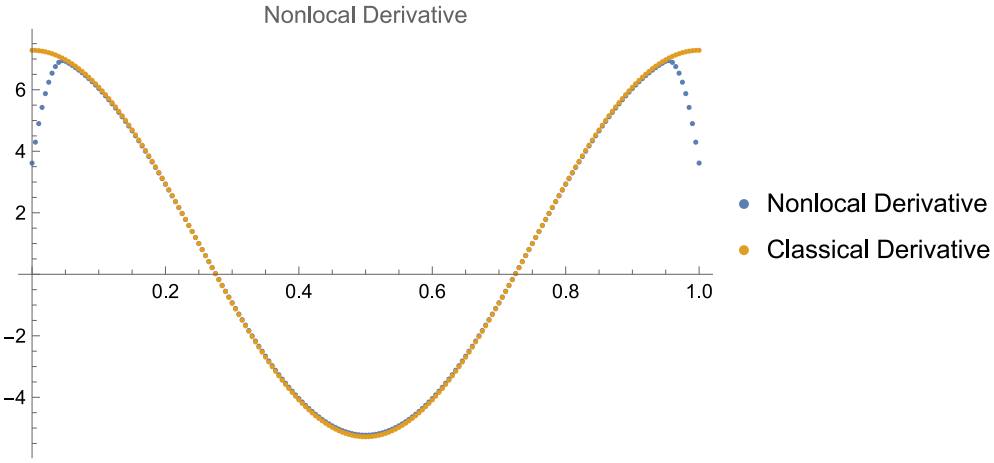
Conjecture 4.2.1: Theorem 3.2.1 can be extended to variational problems where the function



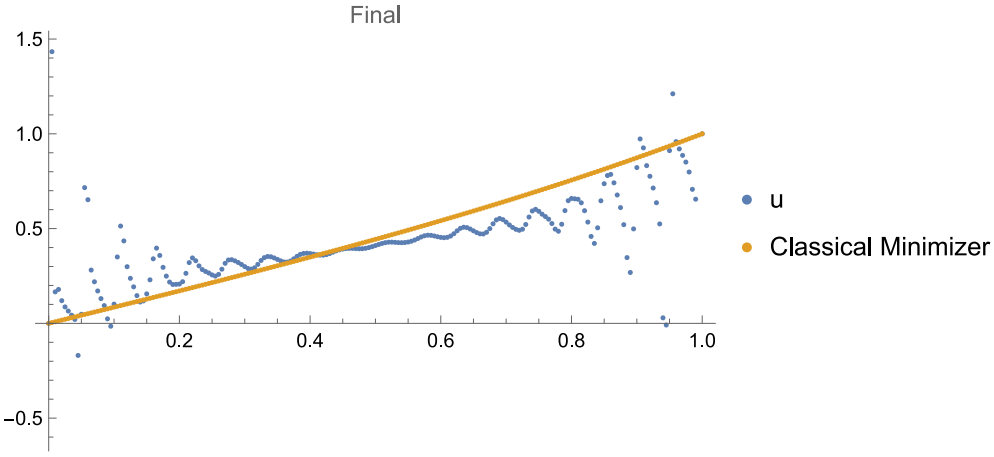
(a)

$\left\|\mathcal{D}u - u'\right\| \approx 8.34289$
$\left\|\mathcal{D}u - u'\right\|_\infty \approx 3.6676$
$J[\hat{u}_c] \approx 1.31305$
$J_\delta[\hat{u}_n] \approx 0.754286$
$\|\hat{u}_c - \hat{u}_n\| \approx 3.59502$
$\|\hat{u}_c - \hat{u}_n\|_\infty \approx 2.11912$

(b)



(c)



(d)

Figure 4.4: Applying an anti-symmetric kernel to Example 4.1.1. This kernel better approximates the derivative. But for some reason does not approach the classical minmizer like the asymmetric kernel does. Note that this behavior appears only for the constant extensions.

u is extended to the collar via linear reflection and the kernel μ is anti-symmetric. Additionally, in cases where the classical problem is well-defined and the kernel of $\mathcal{D}$ satisfies Theorem 2.4.2, the minimizer of the nonlocal variational problem converges to the minimizer of the corresponding classical problem. $\lozenge$

To see some potential evidence of this claim, we return to the functional from Example 4.1.1. Now the admissible class has changed; the values at 0 and 1 are fixed, and the values of u on the collar must be linear reflections of the values of u inside the interval. We consider 6 different kernels and corresponding operators. All of these functions are set to zero outside $(-\delta, \delta)$, and the scaling constants C are selected so that $\int_{-\delta}^{\delta} z\mu(z)\,\mathrm{d}z = 1$.

1. $$\mu_1(z) \coloneqq \begin{cases} C_1, & x \in (0, \delta), \\ 0, & x \notin (0, \delta) \end{cases}$$

2. $$\mu_2(z) \coloneqq \begin{cases} -C_2, & x \in (-\delta, 0), \\ 0, & x = 0, \\ C_2, & x \in (0, \delta) \end{cases}$$

3. $$\mu_3(z) \coloneqq C_3 z$$

4. $$\mu_4(z) \coloneqq \frac{C_4}{z^{1/3}}$$

5. $$\mu_5(z) \coloneqq C_5 z^5$$

6.

$$\mu_6(z) \coloneqq \frac{C_6}{z}$$

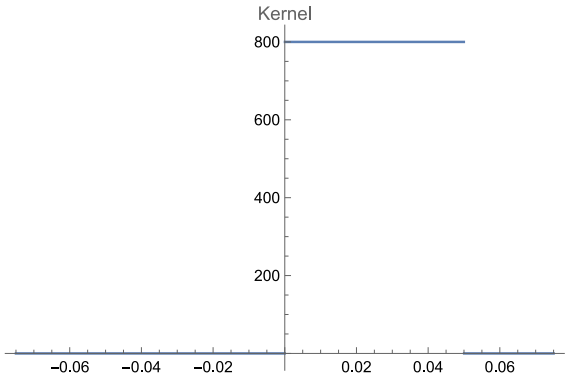
Note that μ_1 and μ_2 are the same kernels used in the previous section. The fourth kernel was also mentioned previously, in Section 2.5. Intuitively, it gives points near x a nearly infinite amount of weight, while points further away contribute less to the value of $\mathcal{D}u(x)$. Compare this to μ_5, which has nearly the opposite effect: points close to x contribute almost nothing. One might expect that μ_4 gets much close to the classical derivative and therefore the classical minimizer. This turns out not to be the case; the two behave somewhat similarly. Finally, note that μ_6 is not in L^1, so none of the results from the previous sections apply. However, it is still interesting to see how the functional behaves for this sort of singular kernel.

A summary of results is given in Table 4.1, though the plots are likely more informative. Many of the nonlocal minimizer graphs develop small ripples near $x = 1$. The cause of this effect is not yet clear. Unlike the oscillations in Figure 4.4, they stay contained near the right endpoint and the amplitudes remain relatively small. Also, note that the plots of the nonlocal derivatives stay closer to the classical derivative near $x = 0$ and $x = 1$ compared to the plots of the previous section.

Kernels μ_2 through μ_6 are all anti-symmetric. Despite the kernels behaving quite differently from one another, the resulting minimizers are fairly close to each other and fairly close to the classical minimzer. This suggests the possibility of a Poincaré type inequality for these anti-symmetric kernels, though we have no proof of this fact so far.

	$\|\mathcal{D}u - u'\|$	$\|\mathcal{D}u - u'\|_\infty$	$J_\delta[\hat{u}_n]$	$\|\hat{u}_c - \hat{u}_n\|$	$\|\hat{u}_c - \hat{u}_n\|_\infty$
μ_1	6.60044	0.659946	1.19439	1.71542	0.191849
μ_2	0.522794	0.05202	1.31451	0.195205	0.00554286
μ_3	0.322977	0.0364033	1.32474	0.190166	0.0531383
μ_4	0.771154	0.081053	1.30554	0.167338	0.0260365
μ_5	1.44124	0.16924	1.38671	0.194979	0.0310755
μ_6	3.58004	0.398679	1.21586	0.139082	0.0168013

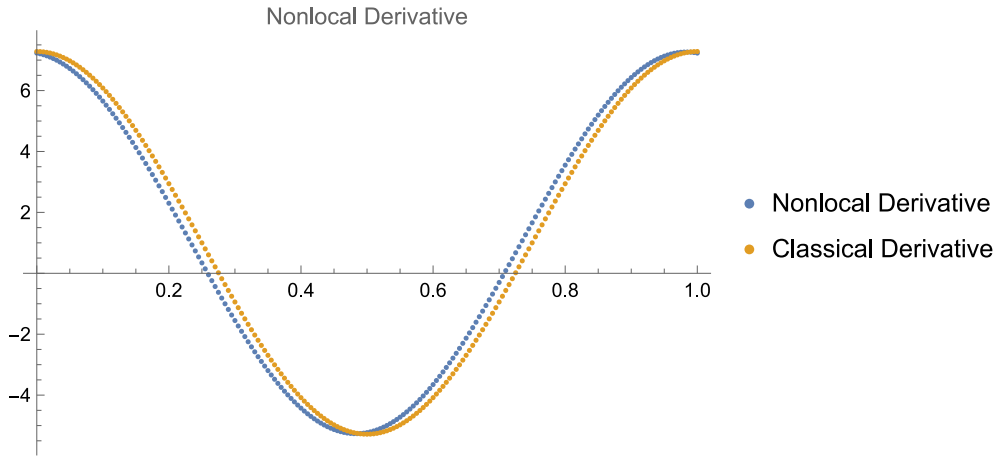
Table 4.1: Summary of results for different choices of the kernel μ.



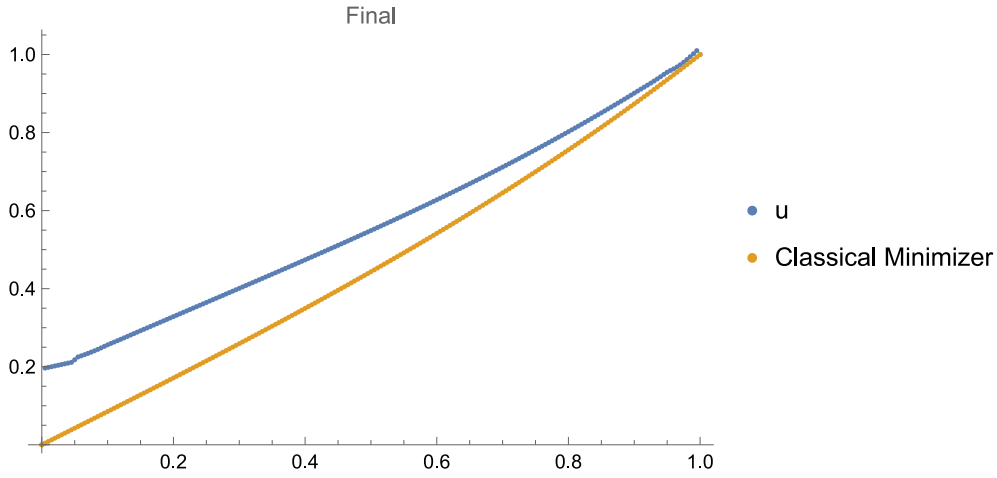
(a)

$\left\|\mathcal{D}u - u'\right\| \approx 6.60044$
$\left\|\mathcal{D}u - u'\right\|_\infty \approx 0.659946$
$J[\hat{u}_c] \approx 1.31305$
$J_\delta[\hat{u}_n] \approx 1.19439$
$\|\hat{u}_c - \hat{u}_n\| \approx 1.71542$
$\|\hat{u}_c - \hat{u}_n\|_\infty \approx 0.191849$

(b)

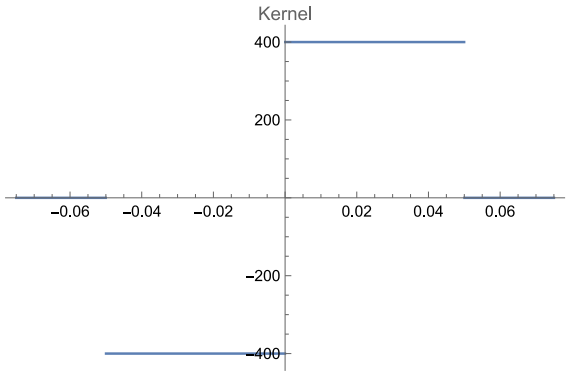


(c)



(d)

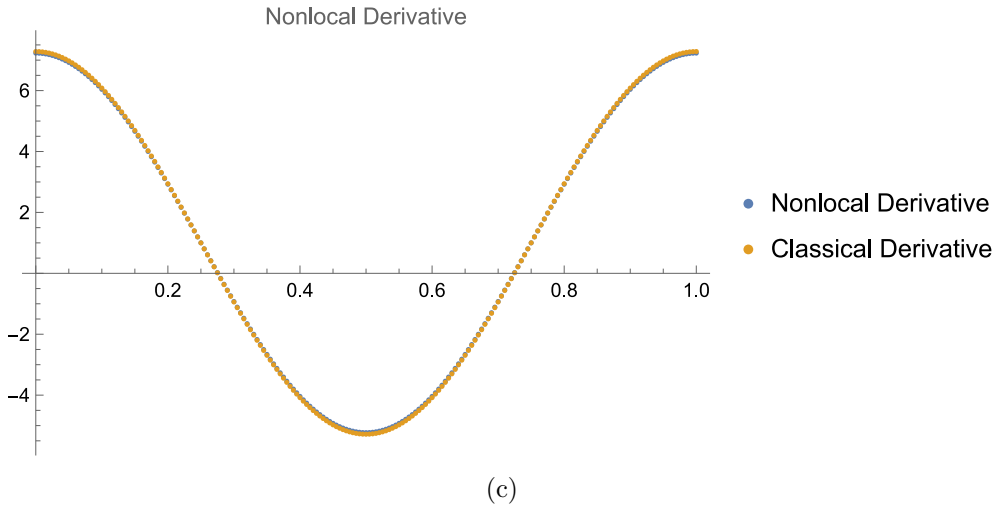
Figure 4.5: (μ_1) A constant, asymmetric kernel. Compare with Figure 4.4 which applied the same kernel to the same functional, with different boundary conditions. Note that the nonlocal functional value is lower than the classical.



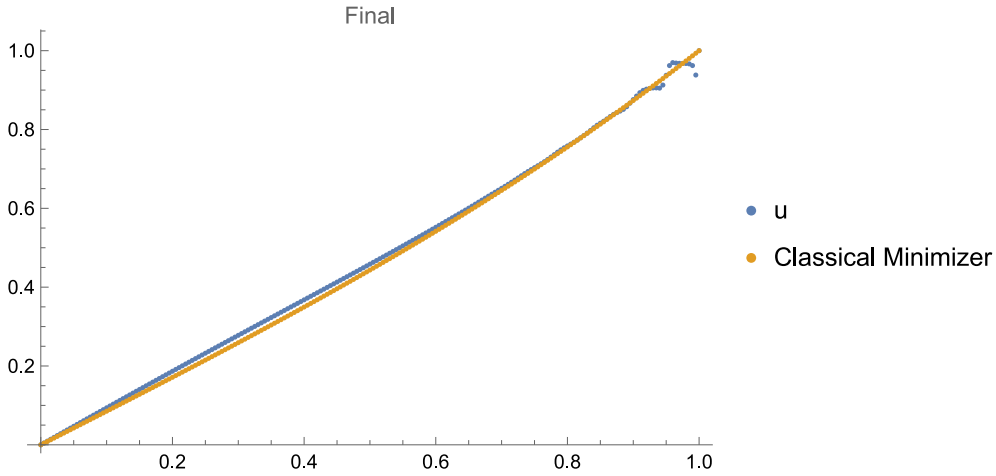
(a)

$\left\|\mathcal{D}u - u'\right\| \approx 0.522794$
$\left\|\mathcal{D}u - u'\right\|_\infty \approx 0.05202$
$J[\hat{u}_c] \approx 1.31305$
$J_\delta[\hat{u}_n] \approx 1.31451$
$\|\hat{u}_c - \hat{u}_n\| \approx 0.195205$
$\|\hat{u}_c - \hat{u}_n\|_\infty \approx 0.0554286$

(b)

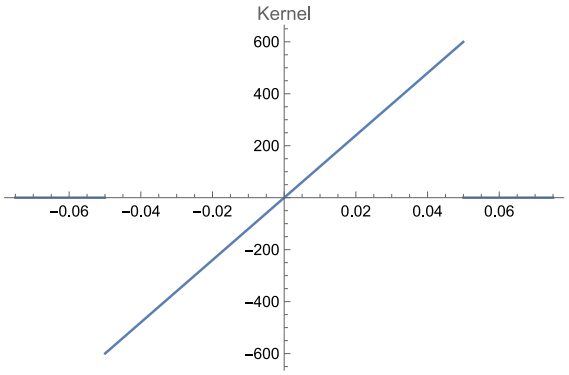


(c)



(d)

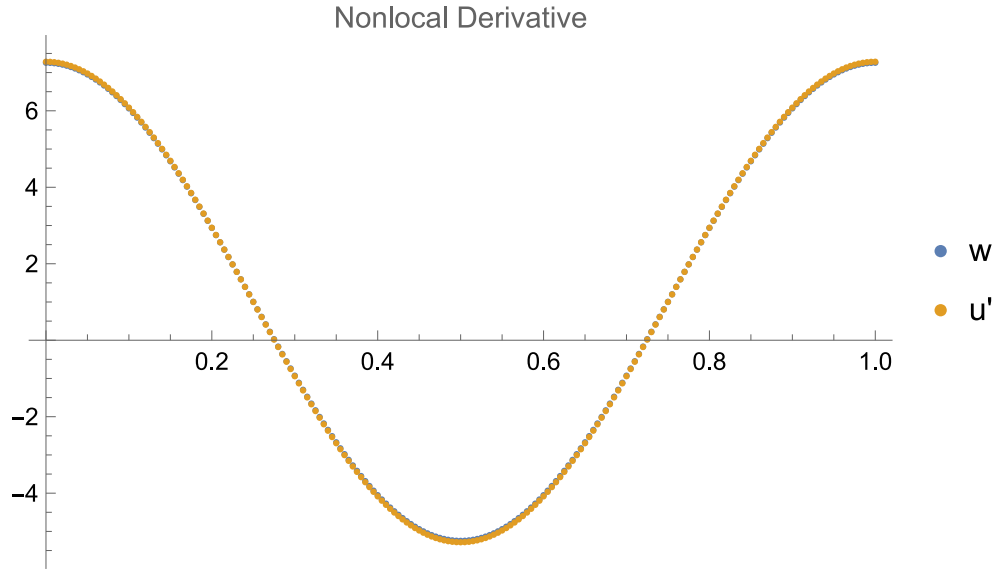
Figure 4.6: (μ_2) An anti-symmetric, piece-wise constant kernel, which gets closer to both the classical derivative and the classical minimizer than μ_1. Compare with Figure 4.4 which applied the same kernel to the same functional, with different boundary conditions.



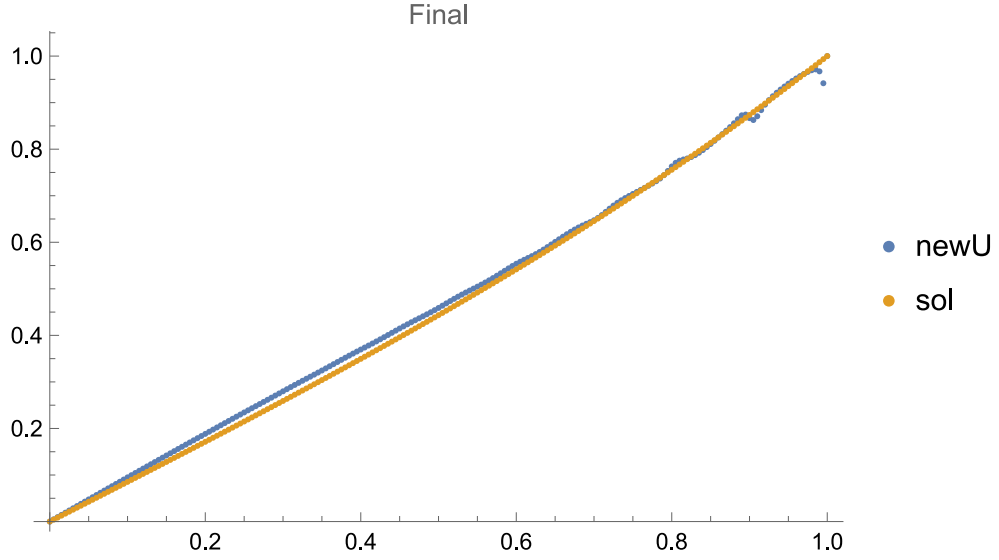
(a)

$\left\|\mathcal{D}u - u'\right\| \approx 0.322977$
$\left\|\mathcal{D}u - u'\right\|_\infty \approx 0.0364033$
$J[\hat{u}_c] \approx 1.31305$
$J_\delta[\hat{u}_n] \approx 1.32474$
$\|\hat{u}_c - \hat{u}_n\| \approx 0.190166$
$\|\hat{u}_c - \hat{u}_n\|_\infty \approx 0.0531383$

(b)

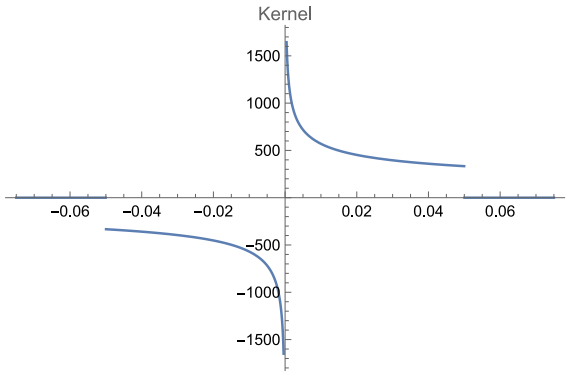


(c)



(d)

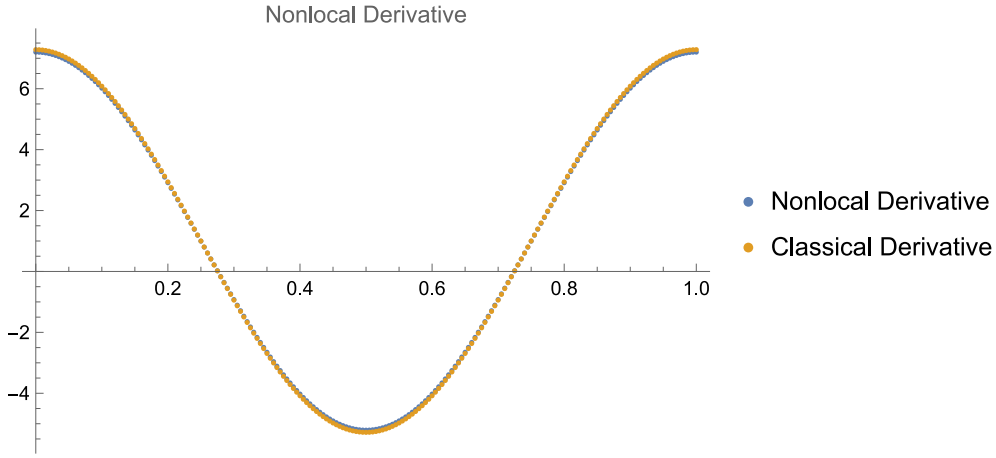
Figure 4.7: (μ_3) A simple linear kernel.



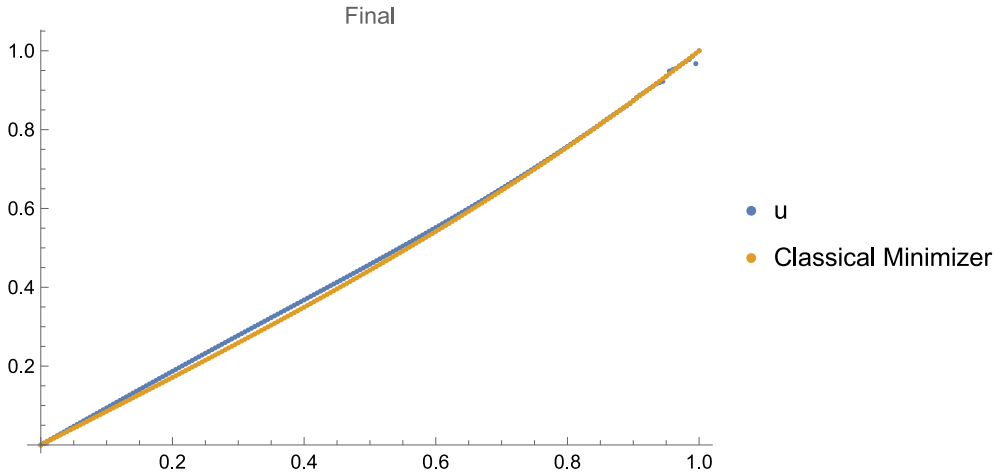
(a)

$\left\|\mathcal{D}u - u'\right\| \approx 0.771154$
$\left\|\mathcal{D}u - u'\right\|_\infty \approx 0.081053$
$J[\hat{u}_c] \approx 1.31305$
$J_\delta[\hat{u}_n] \approx 1.30554$
$\|\hat{u}_c - \hat{u}_n\| \approx 0.167338$
$\|\hat{u}_c - \hat{u}_n\|_\infty \approx 0.0260365$

(b)

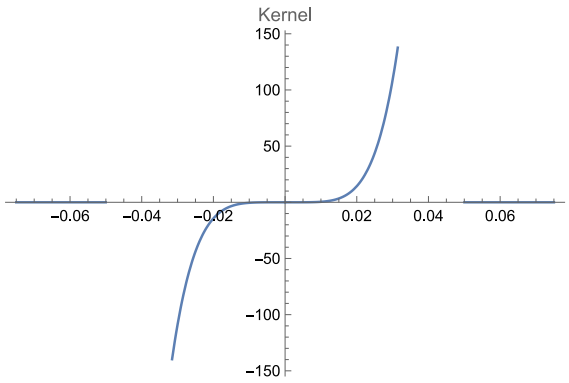


(c)



(d)

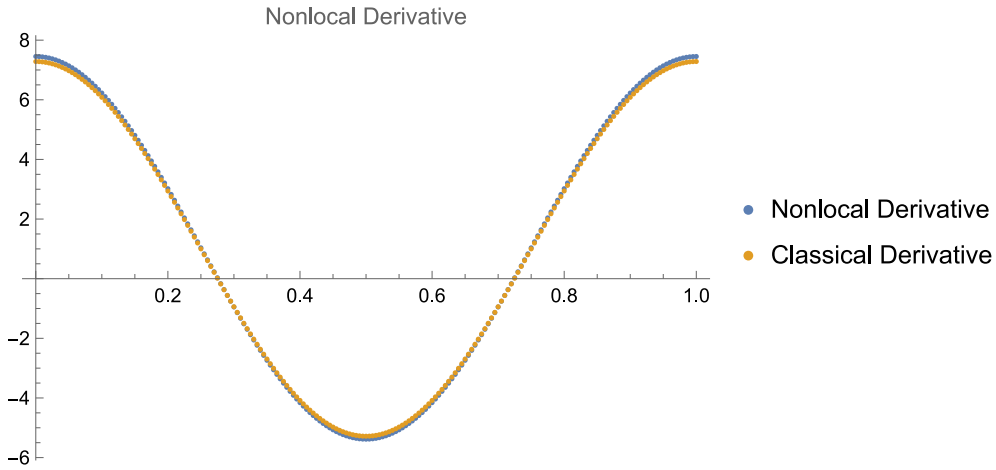
Figure 4.8: (μ_4) The cuberoot kernel used in Section 2.5.



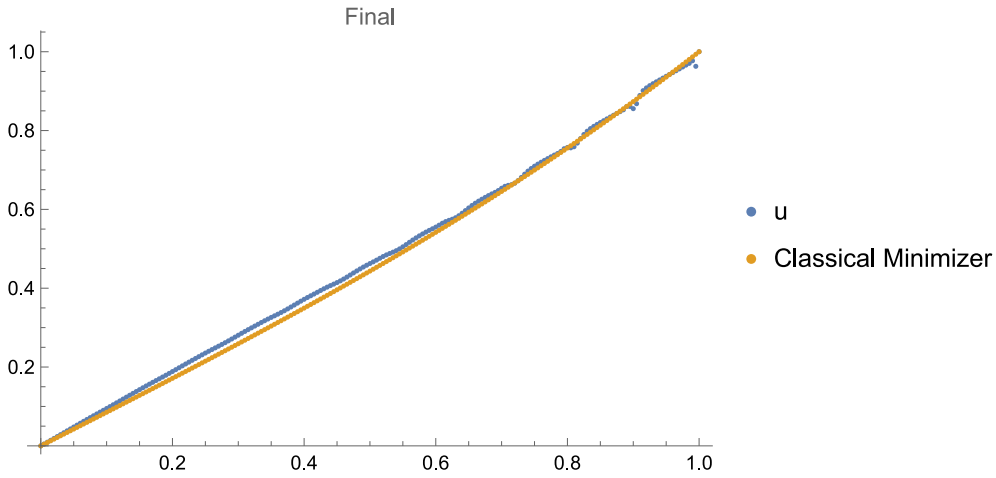
(a)

$\left\|\mathcal{D}u - u'\right\| \approx 1.44124$
$\left\|\mathcal{D}u - u'\right\|_\infty \approx 0.16924$
$J[\hat{u}_c] \approx 1.31305$
$J_\delta[\hat{u}_n] \approx 1.38671$
$\|\hat{u}_c - \hat{u}_n\| \approx 0.194979$
$\|\hat{u}_c - \hat{u}_n\|_\infty \approx 0.0310755$

(b)

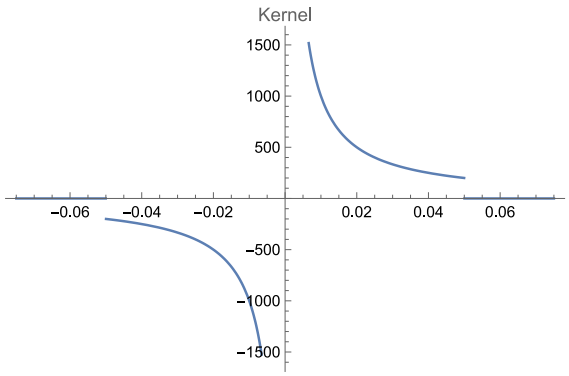


(c)



(d)

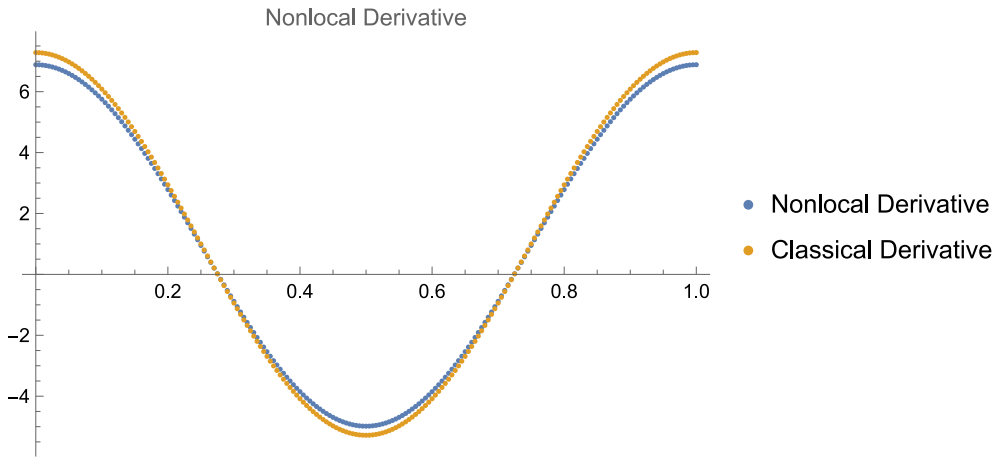
Figure 4.9: (μ_5) Surprisingly, the quintic kernel doesn't produce results all that different from the inverse cuberoot kernel, despite having essentially opposite structures.



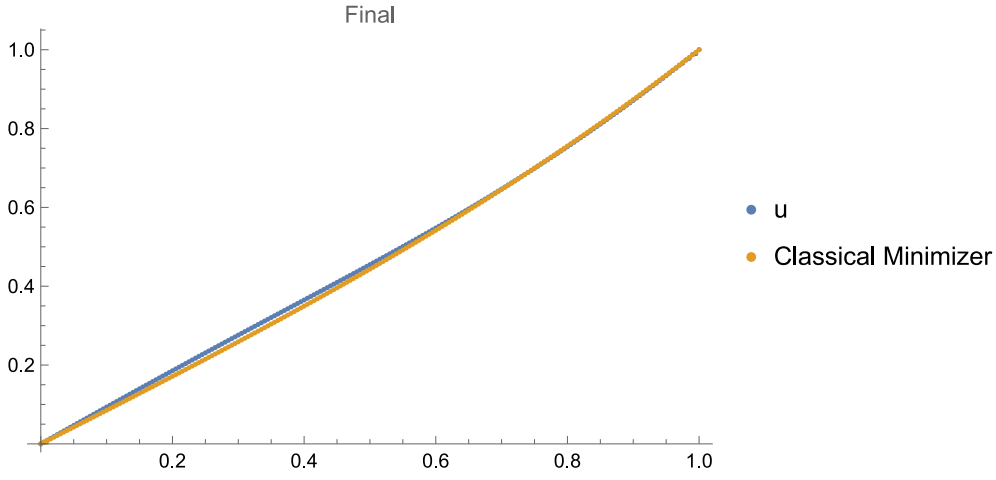
(a)

$\left\|\mathcal{D}u - u'\right\| \approx 3.58004$
$\left\|\mathcal{D}u - u'\right\|_\infty \approx 0.398679$
$J[\hat{u}_c] \approx 1.31305$
$J_\delta[\hat{u}_n] \approx 1.21586$
$\|\hat{u}_c - \hat{u}_n\| \approx 0.139082$
$\|\hat{u}_c - \hat{u}_n\|_\infty \approx 0.0168013$

(b)



(c)



(d)

Figure 4.10: (μ_6) This kernel doesn't reproduce the derivative as well as some of the other anti-symmetric kernels. Yet it looks like the nonlocal minimizer still approaches the classical minimizer.

Chapter 5

Concluding Remarks and Further Questions

5.1 Summary of Main Results

We introduced a class of nonlocal operator, and showed that it is a natural extension of the classical derivative. It is linear, bounded, and preserves the integrability of functions. The flexibility of μ makes the operator $\mathcal{D}$ potentially useful in a wide variety of settings. In cases where both are defined, we showed that D acts as a good approximation to the classical derivative.

Because of the structural similarities between $\mathcal{D}u$ and u', it is natural to investigate variational problems involving this nonlocal operator. One goal is to extend as much as possible of the classical theory on Sobolev spaces to the more general L^p setting, replacing $\dfrac{\mathrm{d}}{\mathrm{d}x}$ with $\mathcal{D}$. Establishing existence and uniqueness of minimization problems is a good first step in this direction. However, many questions are left open. We present some of these problems here.

5.2 Further Research

As noted, a crucial assumption made for the proof of Theorem 3.2.1 was the existence of a Poincaré type inequality. As far as I am aware, no general result has been found for this particular setting. Establishing such a bound would be an important step in advancing the theory of nonlocal variational problems of the sort discussed above.

We showed that $\mathcal{D}_\delta u$ converges to u', but this alone is not enough to show that the minimizer of the nonlocal problem

$$J_\delta[u] = \int_a^b f(x, u, \mathcal{D}_\delta u)\,\mathrm{d}x$$

converges to the corresponding classical minimizer of the local problem

$$J[u] = \int_a^b f(x, u, u') \, \mathrm{d}x.$$

The numerical results of Chapter 4 suggest that establishing this convergence may be tricky, but it does hold in some special cases.

A related notion is that of Γ-convergence, which concerns the convergence of the functionals themselves. (See Braides 2002 or Dal Maso 2012 for an introduction. Bellido et al. 2021 applies these techniques to a problem very similar to the work done in this thesis.) It would be interesting to know when J_δ Γ-converges to J.

A deeper numerical study of the relevant class of nonlocal variational problems is needed. The results above are based on a rough, inefficient algorithm. In addition to giving deeper insight into the nature of the problem, more advanced numerical techniques would help in applying the theoretical framework to real-world problems.

An essential property of the classical derivative is its integration by parts rule. It is not clear if the operator $\mathcal{D}$ satisfies something analogous. If so, this would provide a useful tool in establishing other key results about the operator and related variational problems.

As noted in Section 1.1, the Euler-Lagrange Equations form large part in the usefulness of variational problems. Additionally, nonlocal analogs of the Euler-Lagrange Equations have been found in numerous other settings; see Foss, Radu, and Wright 2018 and Bellido et al. 2020 for example. Do analogous equations hold in our context?

Finally, recall the Lavrentiev Phenomenon introduced in Aside 1.1.7. Can the same phenomenon occur for the nonlocal functionals we discussed? What are some conditions to rule out that possibility?

Appendix A

Mathematica Code and Additional Data

We lay out the general process of the numerical approximations presented in Chapter 4 and record the complete *Mathematica* implementation.

A.1 Description of Methods

We used a simple gradient descent algorithm to approximate minimizers to the nonlocal variational problem. See https://blog.skz.dev/gradient-descent for an intuitive explanation of the general gradient descent method, including some very nice interactive visuals.

First, the setup. To translate the infinite dimensional problem into something a computer can handle, we sample the functions in question at discrete points x_i. For simplicity, we choose a regular step size $x_{i+1} - x_i = h$, for some fixed $h > 0$. To appropriately capture the behavior of the operator $\mathcal{D}$, we need several data points inside the interval $(x_i - \delta, x_i + \delta)$ for each i. For simplicity, we assume δ is an integer multiple of h and call this number $r \coloneqq \frac{\delta}{h}$. This way the same number of sampled points fall within each interval of length 2δ. The problems below only consider the interval $[0, 1]$, with collar $(-\delta, 0] \cup [1, 1 + \delta)$; in theory the code is set up so this can be changed easily.

For a given map $u : [0, 1] \to \mathbb{R}$, let $u_i = u(x_i)$ whenever $x_i \in [0, 1]$. Of course, as $h \to 0$, we sample more and more points, the vector $U = (u_1, \ldots, u_N)$ does a better and better job of approximating the function u. As mentioned in Chapter 4, we apply a linear reflection to extend u onto the collar. For example, suppose that $u_L = u(x_L) = u(1) = 1$, where L denotes the last index corresponding to a point in $[0, 1]$. Then we compute the value of u_{L+1}, the first point on the collar, by drawing a line connecting u_{L-1} and u_L. We set u_{L+1} to whatever the value of this line

is at the point x_{L+1}. In general, the value of u_j for $j > L$ (thus, x_j is a point to the right of the interval $[0, 1]$) is given by

$$u_j = 2 \cdot u_L - u_{2L-j}.$$

To specify the operator $\mathcal{D}_\delta$, we need to know the kernel. Since we want it to behave like the classical derivative, the kernel should satisfy the conditions of Theorem 2.4.2. It is sufficient to enter the basic structure; the code will automatically re-scale it the kernel so that $\int z\mu(z)\,\mathrm{d}z = 1$.

The nonlocal derivative $\mathcal{D}u$ is approximated by the finite-dimensional vector W, which we compute using the trapezoidal rule. A change of variables $y = x_i + z$ centers the integral at x_i and makes the computation clearer:

$$\mathcal{D}u(x_i) = \int_{-\delta}^{\delta} [u(x_i + z) - u(x_i)]\mu(z)\,\mathrm{d}z = \int_{x_i-\delta}^{x_i+\delta} [u(y) - u(x_i)]\mu(y - x_i)\,\mathrm{d}y.$$

Using the fact that $x_i - \delta = x_{i-r}$ and $x_i + \delta = x_{i+r}$, the trapezoidal rule gives

$$\mathcal{D}u(x_i) \approx w_i \coloneqq \frac{h}{2}\left([u_{i-r} - u_i]\mu(x_{i-r} - x_i) + [u_{i+r} - u_i]\mu(x_{i+r} + x_i)\right) + h \cdot \sum_{l=1-r}^{r-1} [u_{i+l} - u_i]\mu(x_{i+l} - x_i) =: w_i$$

Now that we have a way of computing the nonlocal derivative at each point, we need to set up for the gradient descent algorithm. Thus, we need to find the gradient of the functional. This part needs to be done by hand. We illustrate the process by considering

$$J[u] = \int_0^1 u^2 + u'^2\,\mathrm{d}x.$$

We convert this integral into a trapezoidal approximation and replace the classical derivative with the approximate nonlocal derivative:

$$J_n[U] = \frac{h}{2}\left(u_F^2 + w_F^2 + u_L^2 + w_L^2\right) + h \cdot \sum_{l=F+1}^{L-1} \left(u_l^2 + w_l^2\right),$$

where F is the first index such that x_j is in $[0,1]$, and L is the last such index. The gradient is a vector whose jth entry is $\dfrac{\partial J}{\partial u_j}$. So we need to find the partial derivatives using the chain rule. Two things to keep in mind here: usually u_i will not depend on u_j unless $i = j$. But the linear extension to the collar means that the points outside $[0,1]$ *do* depend on points inside $[0,1]$. Thus, the values of $\dfrac{\partial u_i}{\partial u_j}$ (and therefore the gradient of J) depend on the type of boundary conditions. Also, each w_i is a function of many of the u_i, so we need to compute the partial derivatives $\dfrac{\partial w_i}{\partial u_j}$ as well. The partial derivatives of U and W are stored as matrices, so that when we compute the gradient of J we can just call those values. This means that nothing in the code computing these Jacobian matrices needs to change if the functional J is changed, although the formula for the gradient of J will of course need to be updated. As an example, the jth entry of the gradient vector with linear extensions is given by

$$\frac{\partial J_n}{\partial u_j} = \frac{h}{2}\left(2u_F\frac{\partial u_F}{\partial u_j} + 2w_F\frac{\partial w_F}{\partial u_j} + 2u_L\frac{\partial u_L}{\partial u_j} + 2w_L\frac{\partial w_L}{\partial u_j}\right) + h\cdot\sum_{l=F+1}^{L-1}\left(2u_l\frac{\partial u_l}{\partial +}2w_l\frac{\partial w_l}{\partial u_j}\right).$$

Once we have a way of computing the gradient of the functional J, we can proceed with the gradient descent method. This is an iterative process, continuing until the difference between successive values falls under the user-inputted error value ε. Here is the process of the While loop:

1. Using the current value of U on the interval $[0,1]$, extend U to the collar.
2. The extended U is then used to compute the nonlocal derivative w_j at each point x_j in $[0,1]$.
3. The gradient of J_n is computed using these updated values of U and W as inputs.
4. The gradient is scaled by the user-inputted parameter "step", then added to the current value of U. This gives the new values of U inside the interval.
5. Return to step 1.

The remaining code is used to save the output to a given directory, so that the process can be analyzed. The loop saves a plot of the current U values once every m steps. Then, once the loop terminates, the final plots and various data are saved.

A.2 The Code

The parameters used in the examples of Chapter 4 are:

- horizon: $\delta = .05$
- sample rate or mesh size: $h = .0015$
- step size in the gradient descent algorithm: .001
- cutoff distance between two consecutive values: .000001

A.2.1 The Gradient Descent Algorithm

Here is the full code for computing the nonlocal functional and approximating the minimizer with the gradient descent algorithm described above.

```
(*Set file directory to save the outputs, and title the current dataset*)
dir = "ENTER DIRECTORY HERE";
title="TITLE THE RUN\n";
initial=AbsoluteTime[];
ClearAll[i, j, k, l, firstIndex, lastIndex, X, L, a, b, ua, ub, \
delta, h, step, eps, r, int, c, mu, trapJ, gradJ, gradient, u, U, W, \
diff, eps]

(*Set parameters. Delta and h need to be integer multiples of each other for this to
   work*)
a=0;
b=1;
ua=0;
ub=1;
delta=.05;
h=.0015;
step=.001;
eps=.000001;
r=(delta)/(h);

```

```
param = "a = "<>ToString[DecimalForm[a]]<>"\nb= " <>ToString[DecimalForm[b]] <>"\nua
  = " <>ToString[DecimalForm[ua]] <>"\nub= " <>ToString[DecimalForm[ub]] <>"\ndelta=
  " <>ToString[DecimalForm[delta]] <>"\nh= " <>ToString[DecimalForm[h]] <>"\nr= " <>
  ToString[DecimalForm[r]] <>"\nstep= " <>ToString[DecimalForm[step]] <>"\neps= " <>
  ToString[DecimalForm[eps]];


(*Create the mesh,and store some helpful index values for later*)
X = Range[a - delta,b +delta,h];
L = Length[X];
firstIndex = Part[Flatten[Position[X,_?(#>=a&)]],1]; (*This is the index of the
  first x value that falls inside [a,b]*)
lastIndex = Part[Flatten[Position[X,_?(#<= b&)]],-1]; (*This is the index of the
  last x value that falls inside [a,b]*)
trimX=Take[X,{firstIndex,lastIndex}];


(*Input the kernel. Use ctrl enter to add a new line.:*)
unscaledMu[z_]:=\[Piecewise] 0  z<-delta
-1  -delta<=z<0
0 z==0
1 0<z<=delta
0 z>delta

;


(*Need to find appropriate scaling*)
int=NIntegrate[unscaledMu[z]*z,{z,-delta,delta}];
c=1/int;
mu[z_]:=c*unscaledMu[z];

(*Use this if you want to see the kernel to make sure it's working*)
Plot[mu[x],{x,-delta-.5*delta,delta+.5delta},PlotLabel->"Kernel"]
```

```
48
49  kernel="unscaledMu[x]:=" <>ToString[Definition[unscaledMu]];
50
51  (*Enter the functional (as a trapezoidal approximation), the classical solution, and
      the initial data. Maybe later I will do this symbolically but for now: IF YOU
    CHANGE THE FUNCTIONAL YOU MUST ALSO UPDATE THE GRADIENT*)
52
53  (*This is the functional discussed in the document
54  (*trapJ[x_,y_]:=h*(.5*(x[[firstIndex]]^2+y[[firstIndex]]^2+x[[lastIndex]]^2+y[[
      lastIndex]]^2)+Sum[x[[i]]^2+y[[i]]^2,{i,firstIndex+1, lastIndex-1}]);
55  gradient[x_,dx_,y_,dy_,i_]:=h*(x[[firstIndex]]*dx[[firstIndex,i]]+y[[firstIndex]]*dy
      [[firstIndex,i]]+x[[lastIndex]]*dx[[lastIndex,i]]+y[[lastIndex]]*dy[[lastIndex,i]])
       + 2*h*(Sum[x[[l]]*dx[[l,i]]+y[[l]]*dy[[l,i]],{l,firstIndex+1,lastIndex-1}]);
56  sol[x_]:=(\[ExponentialE]^(1-x) (-1+\[ExponentialE]^(2 x)))/(-1+\[ExponentialE]^2);
      *)
57
58  (* This is the functional, gradient and classical minimizer for the simple harmonic
      oscillator with m=4,k=(Pi^2). *)
59  trapJ[x_,y_]:=h/2 (2*y[[firstIndex]]^2-Pi^2/2*x[[firstIndex]]^2+2y[[lastIndex]]^2-Pi
      ^2/2*x[[lastIndex]]^2) + h * Sum[2y[[i]]^2-Pi^2/2*x[[i]]^2,{i,firstIndex+1,
      lastIndex-1}];
60  gradient[x_,dx_,y_,dy_,i_]:=h/2 (4*y[[firstIndex]]*dy[[firstIndex,i]]-Pi^2*x[[
      firstIndex]]*dx[[firstIndex,i]]+4*y[[lastIndex]]*dy[[lastIndex,i]]-Pi^2*x[[
      lastIndex]]*dx[[lastIndex,i]]) + h*Sum[4*y[[l]]*dy[[l,i]]-Pi^2*x[[l]]*dx[[l,i]],{l,
      firstIndex+1,lastIndex-1}];
61  sol[x_]:=Sin[(\[Pi] x)/2];
62
63  (* Enter the inital function u.*)
64  u[x_]:=x;
65
66  functional ="trapJ[x_,y_]:="<> ToString[Definition[trapJ]];
67  gradString="gradient[x_,dx_,y_,dy_,i_]:= "<>ToString[Definition[gradient]];
68  solString="sol[x_]:= "<>ToString[Definition[sol]];
69  uString="u[x_]:= "<>ToString[Definition[u]];
```

```
70
71
72  (*This produces the discretized version of the initial function u, using linear
      extensions at the endpoints :*)
73  U=ConstantArray[0,L];
74  For[
75  i=1,i<=L,i++,
76  Which[
77  i< firstIndex , U[[i]]= 2*ua-u[X[[2*firstIndex-i]]] ,
78  i>=firstIndex && i<lastIndex, U[[i]]=u[X[[i]]] ,
79  i>=lastIndex, U[[i]]=2*ub-u[X[[2*lastIndex-i]]]
80  ];
81  ];
82
83
84  (*Now we can produce the list of  approximate nonlocal derivatives, using the
      trapezoidal rule:*)
85  W=ConstantArray[0,L];
86  nonlocalDeriv[U_,i_]:=h/2 ((U[[i-r]]-U[[i]])*mu[X[[i-r]]-X[[i]]]+(U[[i+r]]-U[[i]])*
      mu[X[[i+r]]-X[[i]]])+h*Sum[(U[[i+l]]-U[[i]])*mu[X[[i+l]]-X[[i]]],{l,1-r,r-1}];
87
88  For[
89  i=firstIndex,i<=lastIndex,i++,
90  W[[i]]=nonlocalDeriv[U,i]
91  ];
92  trimW=Take[W,{firstIndex,lastIndex}];
93
94  (*Make sure the nonlocal deriv. approximation is behaving correctly:*)
95  ListPlot[Transpose @ {X,U} , PlotLabel->"u"]
96  trimW=Take[W,{firstIndex,lastIndex}];
97  ListPlot[{Transpose @ {trimX,trimW} , Transpose @ {trimX,Map[u',trimX] }}, PlotLabel
      ->"w in blue,u' in orange"]
98  ListPlot[Transpose @ {trimX,Map[u',trimX]/trimW }, PlotLabel->"u'/w" ]
99
```

```
100
101 (*To compute the gradient of the functional J, we  need to know the partial
      derivatives of all the u_i's and w_i's with respect to each u)j. We store these
      values in the matrices below.
102
103 Due to the linear reflection/extension, note that the values of U on the collar will
       depend on the values of U just inside the interval. Thus, the matrix will look
      like the identity, except for some symmetric behavior at the corners. This matrix
      will only change if we implement different boundary conditions.
104 *)
105 JacobU=SparseArray[{
106   {i_,j_}/;i==j ->1, (*the partial of u_i with respect to u_i is always 1*)
107   {i_,j_}/; (i<firstIndex && j==firstIndex ) || (i>lastIndex && j==lastIndex)->2, (*
      values of u on the collar depend on the endpoint*)
108   {i_,j_}/; (i<firstIndex && j==(2*firstIndex-i))||(i>lastIndex && j==(2*lastIndex-i
      ))->-1} , (*from the reflection: as U[[firstIndex+1]] moves up, U[[firstIndex-1]]
      moves down*)
109 {L,L}];
110
111 JacobW = ConstantArray[0,{L,L}];
112 For[
113 i=firstIndex,i<=lastIndex,i++,
114 For[
115 j= firstIndex, j<=lastIndex, j++,
116 Which[
117 j<i-r, JacobW[[i,j]] =0,
118 j>=i-r&& j<= i+r, JacobW[[i,j]] =h/2*((JacobU[[i-r,j]]-JacobU[[i,j]])*mu[X[[i-r]]-X
      [[i]]]+(JacobU[[i+r,j]]-JacobU[[i,j]])*mu[X[[i+r]]-X[[i]]])+h*Sum[mu[X[[i+l]]-X[[i
      ]]]*(JacobU[[i+l,j]]-JacobU[[i,j]]),{l,1-r,r-1}],
119 j>i+r, JacobW[[i,j]] =0
120 ];
121 ];
122 ];
123
```

```
rules=ArrayRules[JacobW];
JacobW=SparseArray[rules,{L,L}];


(*With these partial derivatives, we can compute the gradient of the functional.*)
gradJ= ConstantArray[0,L];
For[
i=firstIndex+1,i< lastIndex,i++, (*Only store values of the gradient for points
  strictly inside the interval, so that the endpoints cannot move.*)
gradJ[[i]] = gradient[U,JacobU,W,JacobW,i]
];


(*Now we have everything we need for the gradient descent algorithm.*)

curU = U; (*initialize variables*)
newW=ConstantArray[0,L];
curW=W;
newGrad = gradJ;
(* The size of the difference between functional values is what determines when the
  loop stops. I initialize it at something higher than the error, so that the loop
  runs at least once. *)
diff=2*eps;
loopTracker=0;


(*This is one massive While loop. It iterates until diff <=eps and implements a
  basic gradient descent algorithm.*)
While[diff >eps,

(*using the gradient, we can find the next U*)
newU = curU - (step)(newGrad);

```

```
153 (*Since the gradient vector is zero at the endpoints and on the collar, we need to
      update the values of U on the collar manually, using linear extensions as above.*)
154 For[
155 i=1,i<=L,i++,
156 Which[
157 i< firstIndex , newU[[i]]= 2*ua-newU[[2*firstIndex-i]] ,
158 i>=lastIndex, newU[[i]]=2*ub-newU[[2*lastIndex-i]]
159 ];
160 ];
161
162 (*Now we use the new U to update the nonlocal derivative W.*)
163 For[
164 i=firstIndex,i<=lastIndex,i++,
165 newW[[i]]=nonlocalDeriv[newU,i]
166 ];
167
168 (*Then we can update the gradient using the new U and W values.*)
169 For[
170 i=firstIndex+1,i< lastIndex,i++,
171 newGrad[[i]] = gradient[newU,JacobU,newW,JacobW,i]
172 ];
173
174 (*Now we have the previous values of U and W stored in the "cur" variables, and
      these updated ones in the "new" variables. So we can check the difference in
      corresponding functional values, then update the variables before the next loop. *)
175 diff=Abs[trapJ[curU,curW]-trapJ[newU,newW]];
176 curU=newU;
177 curW=newW;
178 loopTracker=loopTracker+1;
179
180 (*For debugging, this block produces graphs once every m cycles.*)
181 m=500;
182 If[Element[loopTracker/m,Integers] || loopTracker == 1,
183
```

```
trimNewU=Take[newU,{firstIndex,lastIndex}];
trimNewW=Take[newW,{firstIndex,lastIndex}];
trimNewGrad=Take[newGrad,{firstIndex,lastIndex}];

Export[dir<>"Plot "<>ToString[loopTracker]<>".svg",ListPlot[{Transpose @ {trimX,
  trimNewU} ,Transpose @ {trimX,Map[sol,trimX] }} , PlotLabel->"LoopTracker="<>
  ToString[loopTracker],PlotLegends->{"u","Classical Minimizer"}]];

(*ListPlot[{Transpose @ {trimX,trimNewU} ,Transpose @ {trimX,Map[sol,trimX] }} ,
  PlotLabel\[Rule]"LoopTracker="<>ToString[loopTracker],PlotLegends->{"newU","sol"}]

ListPlot[Transpose @ {trimX,trimNewGrad} , PlotLabel\[Rule]"New Gradient "]//Print;
Print[trapJ[newU,newW]]*)
];
];
Export[dir<>"Final.svg",ListPlot[{Transpose @ {trimX,trimNewU} ,Transpose @ {trimX,
  Map[sol,trimX] }} , PlotLabel->"Final",PlotLegends->{"u","Classical Minimizer"}]];
Export[dir<>"Initial U.svg",ListPlot[Transpose @ {X,U} , PlotLabel->"Initial Values"
  ]];
Export[dir<>"Kernel.svg", Plot[mu[x],{x,-delta-.5*delta,delta+.5delta},PlotLabel->"
  Kernel"]];
Export[dir<>"Nonlocal Derivative.svg",ListPlot[{Transpose @ {trimX,trimW} ,
  Transpose @ {trimX,Map[u',trimX] }}, PlotLabel->"Nonlocal Derivative",PlotLegends
  ->{"Nonlocal Derivative","Classical Derivative"}]];
Export[dir<>"Ratio.svg",ListPlot[Transpose @ {trimX,Map[u',trimX]/trimW }, PlotLabel
  ->"Classical to Nonlocal Ratio", PlotLegends->{"u'/D"} ]];
trapJ[newU,newW];
trapJ[sol[X],sol'[X]];
Norm[newU-sol[X]];
Norm[newU-sol[X],Infinity];
loopTracker;
final=AbsoluteTime[];

```

```
output="trapJ[newU,newW]= "<>ToString[DecimalForm[trapJ[newU,newW]]]<>"\ntrapJ[sol[X
  ],sol'[X]]= "<>ToString[DecimalForm[trapJ[sol[X],sol'[X]]]]<>"\nNorm[trimW-Map[u',
  trimX]]= "<>ToString[DecimalForm[Norm[trimW-Map[u',trimX]]]]<>"\nNorm[trimW-Map[u',
  trimX],Infinity]= "<>ToString[DecimalForm[Norm[trimW-Map[u',trimX],Infinity]]]<>"\
  nNorm[newU-sol[X]]= "<>ToString[DecimalForm[Norm[newU-sol[X]]]]<>"\nNorm[newU-sol[X
  ],Infinity]= "<>ToString[DecimalForm[Norm[newU-sol[X],Infinity]]]<>"\nloopTracker=
  "<>ToString[loopTracker];
summary=title<>"\n"<>param<>"\n"<>kernel<>"\n"<>functional<>"\n"<>gradString<>"\n"<>
  solString<>"\n"<>uString<>"\n"<>output<>"\nTotal computation time= "<>ToString[
  DecimalForm[(final-initial)/60]]<>" minutes";
Export[dir<>"Data and Parameters.txt",summary];
```

A.2.2 Nonlocal Derivative Approximator

Here is the code just for approximating and plotting nonlocal derivatives. This was used to produce the plots of $\mathcal{D}_\delta$ for functions like absolute value or the map with a cusp.

```
ClearAll[i, j, k, l, firstIndex, lastIndex, X, L, a, b, ua, ub, \
delta, h, step, eps, r, int, c, mu, trapJ, gradJ, gradient, u, U, W, \
diff, eps]

(*Set parameters. Note that delta and h need to be integer multiples of each other
  for this to work*)
a=0;
b=1;
ua=0;
ub=1;
delta=.05;
h=.005;
step=.001;
eps=.000001;
r=(delta)/(h);

(*This creates the mesh and stores some helpful index values for later. *)
X = Range[a - delta,b +delta,h];
```

```
L = Length[X];
firstIndex = Part[Flatten[Position[X,_?(#>=a&)]],1]; (*This is the index of the
  first x value that falls inside [a,b]*)
lastIndex = Part[Flatten[Position[X,_?(#<= b&)]],-1]; (*This is the index of the
  last x value that falls inside [a,b]*)
trimX=Take[X,{firstIndex,lastIndex}];

(*Input the kernel. It must satisfy mu[z] = 0 for z outside [-delta, delta], and the
   integral of mu cannot be 0 (so mu cannot be symmetric).*)

unscaledMu[z_]:=z^5;


(*This automatically scales the kernel for you, to ensure convergence to the
  classical derivative.*)
int=NIntegrate[unscaledMu[z]*z,{z,-delta,delta}];
c=1/int;
mu[z_]:=c*unscaledMu[z];

(*Use this if you want to see the kernel to make sure it's working*)
Plot[mu[x],{x,-delta-.5*delta,delta+.5delta},PlotLabel->"Kernel",PlotLegends->"mu(z)
  "]


(*Enter the function you want to differentiate.*)
u[x_]:=CubeRoot[(x-1/2)^2];

(*Now we can produce the list of  approximate nonlocal derivatives, using the
  trapezoidal rule:*)
U=Map[u,X];
W=ConstantArray[0,L];
nonlocalDeriv[U_,i_]:=h/2 ((U[[i-r]]-U[[i]])*mu[X[[i-r]]-X[[i]]]+(U[[i+r]]-U[[i]])*
  mu[X[[i+r]]-X[[i]]])+h*Sum[(U[[i+l]]-U[[i]])*mu[X[[i+l]]-X[[i]]],{l,1-r,r-1}];

```

```
45  For[
46  i=firstIndex,i<=lastIndex,i++,
47  W[[i]]=nonlocalDeriv[U,i]
48  ];
49  trimW=Take[W,{firstIndex,lastIndex}];
50
51  (*Make sure the nonlocal deriv. approximation is behaving correctly:*)
52  Plot[u[x],{x,-delta,1+delta}, PlotLabel->"u(x)"]
53  trimW=Take[W,{firstIndex,lastIndex}];
54  ListPlot[{Transpose @ {trimX,trimW} , Transpose @ {trimX,Map[u',trimX] }}, PlotLabel
      ->"Comparing Classical and Nonlocal Derivatives", PlotLegends->{"D u(x)", "u'(x)"}]
```

Bibliography

Alberti, G. and G. Bellettini (1998), "A non-local anisotropic model for phase transitions: asymptotic behaviour of rescaled energies", *European Journal of Applied Mathematics*, 9, 3, pp. 261-284.

Andreu-Vaillo, F., J. M. Mazón, J. D. Rossi, and J. J. Toledo-Melero (2010), *Nonlocal diffusion problems*, 165, American Mathematical Soc.

Antil, H., Z. W. Di, and R. Khatri (2020), "Bilevel optimization, deep learning and fractional Laplacian regularization with applications in tomography", *Inverse Problems*, 36, 6, p. 064001.

Applebaum, D. (2009), *Lévy processes and stochastic calculus*, Cambridge university press.

Arnold, V. I. (1989), *Mathematical Methods of Classical Mechanics*, 2nd ed. Graduate texts in Mathematics ; 60, Springer-Verlag, New York.

Askari, E., F. Bobaru, R. B. Lehoucq, M. L. Parks, S. A. Silling, and O. Weckner (2008), "Peridynamics for multiscale materials modeling", in *Journal of Physics: Conference Series*, 1, IOP Publishing, vol. 125, p. 012078.

Aubert, G. and P. Kornprobst (2009), "Can the nonlocal characterization of Sobolev spaces by Bourgain et al. be useful for solving variational problems?", *SIAM Journal on Numerical Analysis*, 47, 2, pp. 844-860.

Ball, J. M. (1976), "Convexity Conditions and Existence Theorems in Nonlinear Elasticity", *Archive for Rational Mechanics and Analysis*, 63, 4, pp. 337-403, ISSN: 0003-9527, https://link.springer.com/article/10.1007/BF00279992.

Barlow, M., R. Bass, Zhen-Qing Chen, and M. Kassmann (2009), "Non-local Dirichlet forms and symmetric jump processes", *Transactions of the American Mathematical Society*, 361, 4, pp. 1963-1999.

Bass, R. F., M. Kassmann, and T. Kumagai (2010), "Symmetric jump processes: localization, heat kernels and convergence", in *Annales de l'IHP Probabilités et statistiques*, 1, vol. 46, pp. 59-71.

Bates, P. W. and A. Chmaj (1999), "An integrodifferential model for phase transitions: stationary solutions in higher space dimensions", *Journal of statistical physics*, 95, 5, pp. 1119-1139.

Bellido, J. C., J. Cueto, and C. Mora-Corral (2020), "Fractional Piola identity and Polyconvexity in Fractional Spaces", in *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 4, Elsevier, vol. 37, pp. 955-981.

— (2021), "Γ-convergence of polyconvex functionals involving s-fractional gradients to their local counterparts", *Calculus of Variations and Partial Differential Equations*, 60, 1, pp. 1-29.

Benson, D. A., S. W. Wheatcraft, and M. M. Meerschaert (2000), "Application of a fractional advection-dispersion equation", *Water resources research*, 36, 6, pp. 1403-1412.

Biler, P., G. Karch, and R. Monneau (2008), "Nonlinear diffusion of dislocation density and self-similar solutions", *arXiv preprint arXiv:0812.4979*.

Biler, P., G. Karch, and W. A. Woyczyński (2001), "Critical nonlinearity exponent and self-similar asymptotics for Lévy conservation laws", in *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 5, Elsevier, vol. 18, pp. 613-637.

Blasjö, V. (2005), "The Isoperimetric Problem", *American Mathematical Monthly*, 112, 6, p. 526, ISSN: 0002-9890, DOI: 10.2307/30037526, https://www.maa.org/sites/default/files/pdf/upload_library/22/Ford/blasjo526.pdf.

Bobaru, F. and M. Duangpanya (2010), "The peridynamic formulation for transient heat conduction", *International Journal of Heat and Mass Transfer*, 53, 19-20, pp. 4047-4059.

Bobaru, F., J. T. Foster, P. H. Geubelle, and S. A. Silling (2016), *Handbook of peridynamic modeling*, CRC press.

Bodnar, M. and Juan José Lopéz Velázquez (2006), "An integro-differential equation arising as a limit of individual cell-based models", *Journal of Differential Equations*, 222, 2, pp. 341-380.

Bourdin, L., T. Odzijewicz, and D. F. M. Torres (2013), "Existence of Minimizers for Fractional Variational Problems Containing Caputo Derivatives", *Advances In Dynamical Systems And Applications*, 8, 1, pp. 3-12, ISSN: 0973-5321, https://arxiv.org/abs/1208.2363v1.

Bourgain, J., H. Brezis, and P. Mironescu (2001), "Another Look at Sobolev Spaces", Original research article appeared at in Optimal Control and Partial Differential Equations IOS Press ISBN 1 58603 096 5, pp. 439-455, https://hal.archives-ouvertes.fr/hal-00747692.

Boyer, C. B. (1968), *A History of Mathematics*, eng, Wiley, New York.

Braides, A. (2002), *Gamma-convergence for beginners*, eng, Oxford lecture series in mathematics and its applications ; 22, Oxford University Press, New York, ISBN: 0198507844.

Brasco, L. and E. Lindgren (2017), "Higher Sobolev regularity for the fractional p-Laplace equation in the superquadratic case", *Advances in Mathematics*, 304, pp. 300-354, ISSN: 0001-8708, DOI: https://doi.org/10.1016/j.aim.2016.03.039, https://www.sciencedirect.com/science/article/pii/S0001870815300876.

Brezis, H. and H. Nguyen (2018), "Non-local functionals related to the total variation and connections with image processing", *Annals of PDE*, 4, 1, pp. 1-77.

Buades, A., B. Coll, and J. Morel (2010), "Image denoising methods. A new nonlocal principle", *SIAM review*, 52, 1, pp. 113-147.

Burch, N., M. D'Elia, and R. B. Lehoucq (2014), "The exit-time problem for a Markov jump process", *The European Physical Journal Special Topics*, 223, 14, pp. 3257-3271.

Buttazzo, G. (1989), *Semicontinuity, Relaxation, and Integral Representation in the Calculus of Variations*, eng, Pitman research notes in mathematics series, 207, Longman Scientific and Technical, Harlow, Essex, England, ISBN: 0470213434.

Buttazzo, G. and M. Belloni (1995), "A Survey on Old and Recent Results about the Gap Phenomenon in the Calculus of Variations", in, *Recent Developments in Well-Posed Variational Problems*, Mathematics and Its Applications, Springer, vol. 331, chap. 1, https://doi.org/10.1007/978-94-015-8472-2_1.

Cabré, X. and J. Solà-Morales (2005), "Layer solutions in a half-space for boundary reactions", *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 58, 12, pp. 1678-1732.

Caffarelli, L . and Silvestre L. (2007), "An Extension Problem Related to the Fractional Laplacian", *Communications in Partial Differential Equations*, 32, 8, pp. 1245-1260, DOI: 10.1080/03605300600987306, https://doi.org/10.1080/03605300600987306.

Caffarelli, L., A. Mellet, and Y. Sire (2012), "Traveling waves for a boundary reaction–diffusion equation", *Advances in Mathematics*, 230, 2, pp. 433-457.

Caffarelli, L., J. M. Roquejoffre, and O. Savin (2009), "Nonlocal minimal surfaces", *arXiv preprint arXiv:0905.1183*.

Caffarelli, L. and L. Silvestre (2009), "Regularity theory for fully nonlinear integro-differential equations", *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 62, 5, pp. 597-638.

Caffarelli, L. and P. R. Stinga (2016), "Fractional Elliptic Equations, Caccioppoli Estimates and Regularity", *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 33, 3, pp. 767-807, ISSN: 0294-1449, DOI: https://doi.org/10.1016/j.anihpc.2015.01.004, https://www.sciencedirect.com/science/article/pii/S0294144915000153.

Caffarelli, L. and E. Valdinoci (2011), "Uniform estimates and limiting arguments for nonlocal minimal surfaces", *Calculus of Variations and Partial Differential Equations*, 41, 1, pp. 203-240.

Caffarelli, L. and A. Vasseur (2010), "Drift diffusion equations with fractional diffusion and the quasi-geostrophic equation", *Annals of Mathematics*, pp. 1903-1930.

Calder, J. (2022), *The Calculus of Variations*, Online Lecture Notes, https://www-users.cse.umn.edu/~jwcalder/8385F19/CalculusOfVariations.pdf.

Carrillo, C. and P. Fife (2005), "Spatial effects in discrete generation population models", *Journal of Mathematical Biology*, 50, 2, pp. 161-188.

Cartwright, D. and J. McMullen (1978), "A Note on the Fractional Calculus", *Proceedings of the Edinburgh Mathematical Society.*, 21, 1, pp. 79-80, ISSN: 0013-0915, DOI: 10.1017/S0013091500015911, https://www.cambridge.org/core/journals/proceedings-of-the-edinburgh-mathematical-society/article/note-on-the-fractional-calculus/D7F7BE1A09DEDC7C0FDB58AF8B48ADBC.

Cesari, L. (1983), *Optimization—Theory and Applications: Problems with Ordinary Differential Equations*, eng, Applications of Mathematics, Springer New York, New York, NY, vol. 17, ISBN: 1461381673, https://link.springer.com/book/10.1007/978-1-4613-8165-5.

Chen, C. K. and P. C. Fife (2000), "Nonlocal models of phase transitions in solids", *Advances in Mathematical Sciences and Applications*, 10, 2, pp. 821-849.

Chermisi, M. and E. Valdinoci (2010), "A symmetry result for a general class of divergence form PDEs in fibered media", *Nonlinear Analysis: Theory, Methods & Applications*, 73, 3, pp. 695-703.

Colombo, M., G. Crippa, E. Marconi, and L. V. Spinolo (2021), "Local limit of nonlocal traffic models: Convergence results and total variation blow-up", *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 38, 5, pp. 1653-1666, ISSN: 0294-1449, DOI: https://doi.org/10.1016/j.anihpc.2020.12.002, https://www.sciencedirect.com/science/article/pii/S0294144920301220.

Colton, D., R. Kress, and R. Kress (1998), *Inverse acoustic and electromagnetic scattering theory*, Springer, vol. 93.

Cont, R. and P. Tankov (2004), *Financial modelling with jump processes*, 2003, https://hal.archives-ouvertes.fr/hal-00002693.

Conway, J.B. (2019), *A Course in Functional Analysis*, eng, 2nd ed, Graduate Texts in Mathematics, Springer, vol. 96, ISBN: 9780387972459, https://link.springer.com/book/10.1007/978-1-4757-4383-8.

Cordoba, D. (1998), "Nonexistence of simple hyperbolic blow-up for the quasi-geostrophic equation", *Annals of Mathematics*, 148, 3, pp. 1135-1152.

Cortazar, C., J. Coville, M. Elgueta, and S. Martinez (2007), "A nonlocal inhomogeneous dispersal process", *Journal of Differential Equations*, 241, 2, pp. 332-358.

Cozzi, M., S. Dipierro, and E. Valdinoci (2017), "Nonlocal phase transitions in homogeneous and periodic media", *Journal of Fixed Point Theory and Applications*, 19, 1, pp. 387-405.

Craig, W. and M. D. Groves (1994), "Hamiltonian long-wave approximations to the water-wave problem", *Wave motion*, 19, 4, pp. 367-389.

Craig, W. and D. P. Nicholls (2000), "Traveling two and three dimensional capillary gravity water waves", *SIAM Journal on Mathematical Analysis*, 32, 2, pp. 323-359.

Craig, W., U. Schanz, and C. Sulem (1997), "The modulational regime of three-dimensional water waves and the Davey-Stewartson system", in *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 5, Elsevier, vol. 14, pp. 615-667.

Cueto, J. (2021), *Mathematical Analysis of Fractional and Nonlocal Models from Nonlinear Solid Mechanics*, PhD thesis, University of Castilla-La Mancha, https://www.researchgate.net/profile/Javier-Cueto/publication/355145923_Mathematical_analysis_of_fractional_and_nonlocal_models_from_nonlinear_Solid_Mechanics/links/6194eb9b3068c54fa5f4aa64/Mathematical-analysis-of-fractional-and-nonlocal-models-from-nonlinear-Solid-Mechanics.pdf.

Cushman, J. H. and T. R. Ginn (1993), "Nonlocal dispersion in media with continuously evolving scales of heterogeneity", *Transport in porous media*, 13, 1, pp. 123-138.

D'Elia, M., Q. Du, M. Gunzburger, and R. Lehoucq (2017), "Nonlocal convection-diffusion problems on bounded domains and finite-range jump processes", *Computational Methods in Applied Mathematics*, 17, 4, pp. 707-722.

Dacorogna, B. (2008), *Direct Methods in the Calculus of Variations*, eng, Applied mathematical sciences ; v. 78, Springer, New York, ISBN: 9780387357799, https://link.springer.com/book/10.1007/978-0-387-55249-1.

Dagan, G. (1994), "The significance of heterogeneity of evolving scales to transport in porous formations", *Water Resources Research*, 30, 12, pp. 3327-3336.

Dal Maso, G. (2012), *An introduction to* Γ*-convergence*, Springer Science & Business Media, vol. 8.

Daoud, M. and E. H. Laamri (2022), "Fractional Laplacians : A short survey", *Discrete & Continuous Dynamical Systems - S*, 15, 1, pp. 95-116.

Dayal, K. and K. Bhattacharya (2006), "Kinetics of phase transformations in the peridynamic formulation of continuum mechanics", *Journal of the Mechanics and Physics of Solids*, 54, 9, pp. 1811-1842.

De la Llave, R. and P. Panayotaros (1996), "Gravity waves on the surface of the sphere", *Journal of Nonlinear Science*, 6, 2, pp. 147-167.

Di Castro, A., T. Kuusi, and G. Palatucci (2016), "Local Behavior of Fractional p-Minimizers", *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 33, 5, pp. 1279-1299, ISSN: 0294-1449, DOI: https://doi.org/10.1016/j.anihpc.2015.04.003, https://www.sciencedirect.com/science/article/pii/S0294144915000451.

Di Nezza, E., G. Palatucci, and E. Valdinoci (2012), "Hitchhiker s Guide to the Fractional Sobolev Spaces", *Bulletin des Sciences Mathématiques*, 136, 5, pp. 521-573, ISSN: 0007-4497, DOI: https://doi.org/10.1016/j.bulsci.2011.12.004, https://www.sciencedirect.com/science/article/pii/S0007449711001254.

Du, Q., B. Engquist, and X. Tian (2020), "Multiscale modeling, homogenization and nonlocal effects: Mathematical and computational issues", *Contemporary Mathematics*, 754, pp. 115-140.

Du, Q., M. Gunzburger, R. B. Lehoucq, and K. Zhou (2012), "Analysis and approximation of nonlocal diffusion problems with volume constraints", *SIAM review*, 54, 4, pp. 667-696.

Du, Q., Z. Huang, and R. B. Lehoucq (2014), "Nonlocal convection-diffusion volume-constrained problems and jump processes", *Discrete & Continuous Dynamical Systems-B*, 19, 2, p. 373.

Dugas, R. (1955), *A History of Mechanics*, eng, Editions du Griffon, Neuchatel.

Duistermaat, J. J. (1975), "On the spectrum of positive elliptic operators and periodic bicharacteristics", in *Fourier Integral Operators and Partial Differential Equations*, Springer, pp. 15-22.

Duvant, Ge. and J. L. Lions (2012), *Inequalities in mechanics and physics*, Springer Science & Business Media, vol. 219.

Ekeland, I. and R. Temam (1976), *Convex analysis and variational problems*, eng, Studies in mathematics and its applications ; v. 1, North-Holland Pub. Co., Amsterdam, ISBN: 044410898X.

Elbau, P. (2011), "Sequential lower semi-continuity of non-local functionals", *arXiv preprint arXiv:1104.2686*.

Evans, L. C. (1998), *Partial Differential Equations*, eng, Graduate Studies in Mathematics, v. 19, American Mathematical Society, Providence, R.I, ISBN: 0821807722.

Evgrafov, A. and J. C. Bellido (2019), "From non-local Eringen's model to fractional elasticity", *Mathematics and Mechanics of Solids*, 24, 6, pp. 1935-1953.

Farina, A. and E. Valdinoci (2011), "Rigidity results for elliptic PDEs with uniform limits: an abstract framework with applications", *Indiana University Mathematics Journal*, pp. 121-141.

Fefferman, C. and R. de la Llave (1986), "Relativistic stability of matter–I", *I. Rev. Mat. Iberoamericana*, 2, 1-2, pp. 119-213.

Fife, P. (2003), "Some nonclassical trends in parabolic and parabolic-like evolutions", *Trends in nonlinear analysis*, pp. 153-191.

Folland, G. B. (1999), *Real Analysis: Modern Techniques and their Applications*, eng, Second edition. Pure and applied mathematics, Wiley, New York, ISBN: 0471317160.

Fonseca, I. and G. Leoni (2007), *Modern Methods in the Calculus of Variations: L^p Spaces*, eng, Springer New York, New York, NY, ISBN: 038735784X, https://link.springer.com/book/10.1007/978-0-387-69006-3.

Foss, M. (2005), *On Lavrentiev's Phenomenon*, PhD thesis, Carnegie Mellon University.

— (2019), "Nonlocal Poincaré Inequalities for Integral Operators with Integrable Nonhomogeneous Kernels", *arXiv preprint arXiv:1911.10292*.

Foss, M., P. Radu, and C. Wright (2018), "Existence and regularity of minimizers for nonlocal energy functionals", *Differential and Integral Equations*, 31, 11/12, pp. 807-832.

Foss, M., P. Radu, and Y. Yu (2022), "Convergence analysis and numerical studies for linearly elastic peridynamics with dirichlet-type boundary conditions", *Journal of Peridynamics and Nonlocal Modeling*, pp. 1-36.

Fournier, N. and P. Laurençot (2006), "Well-posedness of Smoluchowski's coagulation equation for a class of homogeneous kernels", *Journal of functional Analysis*, 233, 2, pp. 351-379.

Gächter, G. K. and M. J. Grote (2003), "Dirichlet-to-Neumann map for three-dimensional elastic waves", *Wave Motion*, 37, 3, pp. 293-311.

Garroni, A. and G. Palatucci (2006), "A singular perturbation result with a fractional norm", in *Variational problems in materials science*, Springer, pp. 111-126.

Gilboa, G. and S. Osher (2007), "Nonlocal linear image regularization and supervised segmentation", *Multiscale Modeling & Simulation*, 6, 2, pp. 595-630.

— (2009a), "Nonlocal operators with applications to image processing", *Multiscale Modeling & Simulation*, 7, 3, pp. 1005-1028.

— (2009b), "Nonlocal operators with applications to image processing", *Multiscale Modeling & Simulation*, 7, 3, pp. 1005-1028.

Goldstine, H. (1980), *A History of the Calculus of Variations from the Seventeenth Through the Nineteenth Century*, Studies in the history of mathematics and physical sciences ; 5, Springer-Verlag, New York.

Gonzalez, Maria del Mar and R. Monneau (2010), "Slow motion of particle systems as a limit of a reaction-diffusion equation with half-Laplacian in dimension one", *arXiv preprint arXiv:1007.0740*.

Grote, M. J. and C. Kirsch (2004), "Dirichlet-to-Neumann boundary conditions for multiple scattering problems", *Journal of Computational Physics*, 201, 2, pp. 630-650.

Hardt, R. (ed.) (2004), *Six Themes on Variation*, eng, Student mathematical library ; v. 26, American Mathematical Society, Providence, RI, ISBN: 0821837206.

Hardy, G. H., J.E. Littlewood, and G. Polya (1952), *Inequalities*, eng, 2d ed. University Press, Cambridge Eng.

Heinricher, A. C. and V. J. Mizel (1988), "The Lavrentiev phenomenon for invariant variational problems", *Archive for Rational Mechanics and Analysis*, 102, pp. 57-96, DOI: 10.1007/BF00250924.

Hilbert, D. (1902), "Mathematical Problems", eng, *Bulletin of the American Mathematical Society*, 8, 10, pp. 437-479, https://www.ams.org/journals/bull/1902-08-10/S0002-9904-1902-00923-3/S0002-9904-1902-00923-3.pdf.

Hinds, B. and P. Radu (2012), "Dirichlet's principle and wellposedness of solutions for a nonlocal p-Laplacian system", *Applied Mathematics and Computation*, 219, 4, pp. 1411-1419, ISSN: 0096-3003, DOI: https://doi.org/10.1016/j.amc.2012.07.045, https://www.sciencedirect.com/science/article/pii/S0096300312007618.

Hrusa, W. and M. Foss (2002), "Lecture Notes: Calculus of Variations".

Hu, B. and D. P. Nicholls (2005), "Analyticity of Dirichlet–Neumann Operators on Hölder and Lipschitz Domains", *SIAM journal on mathematical analysis*, 37, 1, pp. 302-320.

Huang, Z. (2019), "Noether's theorem in peridynamics", *Mathematics and Mechanics of Solids*, 24, 11, pp. 3394-3402.

Jafarzadeh, S., A. Larios, and F. Bobaru (2020), "Efficient solutions for nonlocal diffusion problems via boundary-adapted spectral methods", *Journal of Peridynamics and Nonlocal Modeling*, 2, 1, pp. 85-110.

Jafarzadeh, Siavash, Ziguang Chen, and Florin Bobaru (2018), "Peridynamic modeling of repassivation in pitting corrosion of stainless steel", *Corrosion*, 74, 4, pp. 393-414.

Jakobsen, E. R. and K. H. Karlsen (2005), "Continuous dependence estimates for viscosity solutions of integro-PDEs", *Journal of Differential Equations*, 212, 2, pp. 278-318.

Johnson, G. W. and M. L. Lapidus (2002), *The Feynman Integral and Feynman's Operational Calculus*, eng, New Edition, Oxford Mathematical Monographs, Oxford University Press Incorporated, ISBN: 1-280-37517-5.

Kao, Chiu-Yen, Y. Lou, and W. Shen (2010), "Random dispersal vs. non-local dispersal", *Discrete & Continuous Dynamical Systems*, 26, 2, p. 551.

Katiyar, A., S. Agrawal, H. Ouchi, P. Seleson, J. T. Foster, and M. M. Sharma (2020), "A general peridynamics model for multiphase transport of non-Newtonian compressible fluids in porous media", *Journal of Computational Physics*, 402, p. 109075.

Kindermann, S., S. Osher, and P. W. Jones (2005), "Deblurring and denoising of images by nonlocal functionals", *Multiscale Modeling & Simulation*, 4, 4, pp. 1091-1115.

Kosmann-Schwarzbach, Y. (2010), *The Noether Theorems: Invariance and Conservation Laws in the Twentieth Century*, eng, trans. by B. E. Schwarzbach, Sources and Studies in the History of Mathematics and Physical Sciences, Springer New York, New York, NY, pp. 205-, ISBN: 038787867X, https://link.springer.com/book/10.1007/978-0-387-87868-3.

Kot, M. (2014), *A First Course in the Calculus of Variations*, Student Mathematical Library, American Mathematical Society, Rhode Island.

Kreisbeck, C. and H. Schönberger (2022), "Quasiconvexity in the Fractional Calculus of Variations: Characterization of Lower Semicontinuity and Relaxation", *Nonlinear Analysis*, 215, p. 112625.

Kreisbeck, C. and E. Zappale (2020), "Lower semicontinuity and relaxation of nonlocal L^∞-functionals", *Calculus of variations and partial differential equations*, 59, 4, pp. 1-36.

Kurzke, M. (2006), "A nonlocal singular perturbation problem with periodic well potential", *ESAIM: Control, Optimisation and Calculus of Variations*, 12, 1, pp. 52-63.

Kuusi, T., G. Mingione, and Y. Sire (2014), "A fractional Gehring lemma, with applications to nonlocal equations", *Rendiconti Lincei-Matematica e Applicazioni*, 25, 4, pp. 345-358.

Lanczos, Cornelius (1949), *The Variational Principles of Mechanics*, eng, Heritage, University of Toronto Press, ISBN: 1487581777.

Laskin, N. (2000), "Fractional quantum mechanics and Lévy path integrals", *Physics Letters A*, 268, 4-6, pp. 298-305.

Leoni, G. (2009), *A First Course in Sobolev Spaces*, eng, Graduate studies in mathematics ; v. 105, American Mathematical Society, Providence, R.I, ISBN: 9780821847688.

Levi, M. (2014), *Classical Mechanics with Calculus of Variations and Optimal Control: an Intuitive Introduction*, The Student Mathematical Library, American Mathematical Society, vol. 69.

Logan, J. D. (1977), *Invariant Variational Principles*, Mathematics in science and engineering ; v. 138, Academic Press, New York.

Lou, Y., X. Zhang, S. Osher, and A. Bertozzi (2010), "Image recovery via nonlocal operators", *Journal of Scientific Computing*, 42, 2, pp. 185-197.

Luchko, Y. (2020), "Fractional derivatives and the fundamental theorem of fractional calculus", *Fractional Calculus and Applied Analysis*, 23, 4, pp. 939-966.

Madenci, E. and E. Oterkus (2013), *Peridynamic Theory and Its Applications*, eng, 2014th ed., Springer New York, New York, NY, ISBN: 9781461484646.

Majda, A. J. and E. G. Tabak (1996), "A two-dimensional model for quasigeostrophic flow: comparison with the two-dimensional Euler flow", *Physica D: Nonlinear Phenomena*, 98, 2-4, pp. 515-522.

Majer, U. and T. Sauer (2005), "Hilbert's 'World Equations' and His Vision of a Unified Science", in *The Universe of General Relativity*, ed. by Jean Kox A. J. and Eisenstaedt, Birkhäuser Boston, Boston, MA, pp. 259-276, ISBN: 978-0-8176-4454-3, https://link.springer.com/chapter/10.1007/0-8176-4454-7_14.

Malinowska, A. B., T. Odzijewicz, and D. F. M. Torres (2015), *Advanced methods in the fractional calculus of variations*, Springer.

Meerschaert, M. M. and A. Sikorskii (2019), *Stochastic models for fractional calculus*, de Gruyter.

Mellet, A., S. Mischler, and C. Mouhot (2011), "Fractional diffusion limit for collisional kinetic equations", *Archive for Rational Mechanics and Analysis*, 199, 2, pp. 493-525.

Mengesha, T. and Q. Du (2013), "Analysis of a scalar nonlocal peridynamic model with a sign changing kernel", *Discrete & Continuous Dynamical Systems-B*, 18, 5, p. 1415.

— (2014), "Nonlocal constrained value problems for a linear peridynamic Navier equation", *Journal of Elasticity*, 116, 1, pp. 27-51.

Mengesha, T. and D. Spector (2015), "Localization of nonlocal gradients in various topologies", *Calculus of Variations and Partial Differential Equations*, 52, 1, pp. 253-279.

Metzler, R. and J. Klafter (2000), "The random walk's guide to anomalous diffusion: a fractional dynamics approach", *Physics reports*, 339, 1, pp. 1-77.

— (2004), "The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics", *Journal of Physics A: Mathematical and General*, 37, 31, R161.

Mingione, G. (2010), "Gradient potential estimates", *Journal of the European Mathematical Society*, 13, 2, pp. 459-486.

Mogilner, A. and L. Edelstein-Keshet (1999), "A non-local model for a swarm", *Journal of mathematical biology*, 38, 6, pp. 534-570.

Motz, L. and J. H. Weaver (1993), *The Story of Mathematics*, eng, Plenum Press, New York, ISBN: 0306445085.

Odzijewicz, T., A. B. Malinowska, and D. F. M. Torres (2012), "Generalized fractional calculus with applications to the calculus of variations", *Computers & Mathematics with Applications*, 64, 10, pp. 3351-3366.

Oterkus, E., A. Barut, and E. Madenci (2012), "Peridynamics based on the principle of virtual work", in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*, p. 1540.

Pang, G., M. D'Elia, M. Parks, and G. E. Karniadakis (2020), "nPINNs: nonlocal Physics-Informed Neural Networks for a parametrized nonlocal universal Laplacian operator. Algorithms and Applications", *Journal of Computational Physics*, 422, p. 109760.

Pooseh, S. (2013), *Computational methods in the fractional calculus of variations and optimal control*, PhD thesis, Universidade de Aveiro (Portugal).

Reichenberger, A. (2018), "Émilie Du Châtelet's interpretation of the laws of motion in the light of 18th century mechanics", *Studies in History and Philosophy of Science Part A*, 69, pp. 1-11, ISSN: 0039-3681, DOI: https://doi.org/10.1016/j.shpsa.2018.01.006, https://www.sciencedirect.com/science/article/pii/S0039368118300177.

Rockafellar, R. T. (1970), *Convex Analysis*, Princeton University Press, ISBN: 9781400873173, DOI: doi:10.1515/9781400873173, https://www.degruyter.com/document/doi/10.1515/9781400873173/html.

Ros-Oton, X. (2016), "Nonlocal elliptic equations in bounded domains: a survey", *Publicacions matematiques*, pp. 3-26.

Rosasco, L., M. Belkin, and E. De Vito (2010), "On learning with integral operators." *Journal of Machine Learning Research*, 11, 2.

Sabatelli, L., S. Keating, J. Dudley, and P. Richmond (2002), "Waiting time distributions in financial markets", *The European Physical Journal B-Condensed Matter and Complex Systems*, 27, 2, pp. 273-275.

Savin, O. and E. Valdinoci (2012), "Γ-convergence for nonlocal phase transitions", in *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 4, Elsevier, vol. 29, pp. 479-500.

Scalas, E., R. Gorenflo, and F. Mainardi (2000), "Fractional calculus and continuous-time finance", *Physica A: Statistical Mechanics and its Applications*, 284, 1-4, pp. 376-384.

Schekochihin, A. A., S. C. Cowley, and T. A. Yousef (2008), "MHD turbulence: Nonlocal, anisotropic, nonuniversal?", in *IUTAM Symposium on computational physics and new perspectives in turbulence*, Springer, pp. 347-354.

Schwichtenberg, J. (2015), *Physics from Symmetry*, eng, 1st ed. 2015. Undergraduate Lecture Notes in Physics, Springer International Publishing, Cham, ISBN: 3-319-19201-9, https://link.springer.com/book/10.1007/978-3-319-19201-7.

Serovaĭskiĭ, Y. Ya. (2004), *Counterexamples in Optimal Control Theory*, eng, Inverse and Ill-Posed Problems Series, VSP, Utrecht ; ISBN: 9067644005.

Shieh, T. and D. E. Spector (2018), "On a New Class of Fractional Partial Differential Equations II", *Advances in Calculus of Variations*, 11, 3, pp. 289-307, DOI: doi:10.1515/acv-2016-0056, https://doi.org/10.1515/acv-2016-0056.

Šilhavỳ, M. (2020), "Fractional Vector Analysis Based on Invariance Requirements (Critique of Coordinate Approaches)", *Continuum Mechanics and Thermodynamics*, 32, 1, pp. 207-228.

Silling, S. A. (2000), "Reformulation of Elasticity Theory for Discontinuities and Long-Range Forces", *Journal of the Mechanics and Physics of Solids*, 48, 1, pp. 175-209, ISSN: 0022-5096, DOI: https://doi.org/10.1016/S0022-5096(99)00029-0, https://www.sciencedirect.com/science/article/pii/S0022509699000290.

Silling, S. A. and E. Madenci (2019), "The World is Nonlocal", *Journal of Peridynamics and Nonlocal Modeling*, 1, 1, pp. 1-2.

Silvestre, L. (2007), "Regularity of the obstacle problem for a fractional power of the Laplace operator", *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 60, 1, pp. 67-112.

Sire, Y. and E. Valdinoci (2009), "Fractional Laplacian phase transitions and boundary reactions: a geometric inequality and a symmetry result", *Journal of Functional Analysis*, 256, 6, pp. 1842-1864.

Stillwell, J. (2010), *Mathematics and its History*, eng, 3rd ed, Undergraduate texts in mathematics, Springer, New York, ISBN: 9781441960528, https://link.springer.com/book/10.1007/978-1-4419-6053-5.

Struwe, M. (2008), *Variational Methods Applications to Nonlinear Partial Differential Equations and Hamiltonian Systems*, eng, 4th ed. A Series of Modern Surveys in Mathematics, 34, Springer, Berlin, Heidelberg, https://link.springer.com/book/10.1007/978-3-540-74013-1.

Tadmor, E. and P. Athavale (2009), "Multiscale image representation using novel integro-differential equations", *Inverse Problems & Imaging*, 3, 4, p. 693.

Toland, J. F. (1997), "The Peierls–Nabarro and Benjamin–Ono equations", *Journal of functional analysis*, 145, 1, pp. 136-150.

Troutman, J. (1996), *Variational Calculus and Optimal Control: Optimization with Elementary Convexity*, 2nd ed, Undergraduate texts in mathematics, Springer, New York, https://link.springer.com/book/10.1007/978-1-4612-0737-5.

Uspenskii, S. V. (1960), "Properties of W_p^r classes with a fractional derivative on differentiable manifolds", in *Doklady Akademii Nauk*, 1, Russian Academy of Sciences, vol. 132, pp. 60-62.

Vázquez, J. L. (2012), "Nonlinear diffusion with fractional Laplacian operators", in *Nonlinear partial differential equations*, Springer, pp. 271-298.

— (2017), "The mathematical theories of diffusion: Nonlinear and fractional diffusion", in *Nonlocal and nonlinear diffusions and interactions: new methods and directions*, Springer, pp. 205-278.

Wazwaz, A. (2011), *Linear and Nonlinear Integral Equations Methods and Applications*, eng, 1st ed. 2011. Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN: 1-283-47623-1.

Weckner, O. and R. Abeyaratne (2005), "The effect of long-range forces on the dynamics of a bar", *Journal of the Mechanics and Physics of Solids*, 53, 3, pp. 705-728.

Wei, Y., Y. Kang, W. Yin, and Y. Wang (2020), "Generalization of the gradient method with fractional order gradient direction", *Journal of the Franklin Institute*, 357, 4, pp. 2514-2532.

Yosida, K. (1980), *Functional Analysis*, eng, 6th ed, Grundlehren der mathematischen Wissenschaften ; v.123, Springer-Verlag, Berlin ; ISBN: 0387102108.

Zabarankin, M. and A. Kurdila (2005), *Convex Functional Analysis*, eng, Systems and Control: Foundations and Applications, Birkhauser, ISBN: 9783764321987, https://link.springer.com/book/10.1007/3-7643-7357-1.

Ziemer, W. P. (2012), *Weakly differentiable functions: Sobolev spaces and functions of bounded variation*, Springer Science & Business Media, vol. 120.