Euler operators for mis-specified physics-informed neural networks

Charlie Cowen-Breen [1 2] **Yongji Wang** [3 4] **Stephen Bates** [1] **Ching-Yao Lai** [4 5]

Abstract

Integration between equations and data is ubiquitously useful in the physical sciences: for example, data assimilation plays an important role in generating seamless data for climate and geophysics applications. However, in many practical scenarios, we have only partial or imprecise knowledge of these equations, leading to mis-specified problems, i.e. problems for which there is no solution to the given equations when coupled with real-world data. Physics-informed neural networks (PINNs) have been increasingly used for data assimilation, yet their efficacy has primarily been demonstrated for well-specified problems where solutions for the given equation and data exist. Moreover, in applications involving real-world noisy data and complex PDEs, it is often difficult to know whether the given problem is mis-specified. We introduce a framework for deriving well-specified problems from mis-specified problems—based on the Euler operator (related to the Euler-Lagrange equation) and other tools from mathematical analysis—closing the gap between problems which may be mis-specified in practice and the already-proven theory for the well-specified case. Based on this framework, we propose a diagnostic tool for determining whether given PDE problems are mis-specified, and prove generalization bounds for mis-specified PINNs. Our contributions enhance PINN reliability in the presence of partial knowledge of governing equations or imperfect data.

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA [2]Department of Mathematics, Princeton University, Princeton, NJ, USA [3]Courant Institute of Mathematical Sciences, New York University, New York, NY, USA [4]Department of Geosciences, Princeton University, Princeton, NJ, USA [5]Department of Geophysics, Stanford University, Stanford, CA, USA. Correspondence to: Charlie Cowen-Breen <ccbreen@mit.edu>, Ching-Yao Lai <cyaolai@stanford.edu>.

5th AI for Science Workshop at the 41st International Conference on Machine Learning (ICML), Vienna, Austria, 2024.

Understanding the equations that govern natural phenomena and integrating them with data to model systems of interest is a core aim of the physical sciences. In some domains, the governing equations are well-understood. In others, however, we may have only partial knowledge of the governing equations, yet this partial knowledge can still aid in better modeling in the presence of potentially noisy observations. In such settings, integrating data with governing equations is challenged by *possible mis-specification* of the equations with the given data, by which we mean that there may be no solution to the hypothesized governing equations which also fits the data precisely (Zou et al., 2024; Duffin et al., 2021; Muchiri et al., 2024; Browning et al., 2020; Murphy et al., 2023).

In practice, problems can be mis-specified in several ways. Mis-specification can occur due to (i) noisy data, (ii) inaccurate governing equations, or (iii) both. In PDE-constrained optimization, for instance, observations may be so noisy that there is no solution to the governing equation which fits them (e.g. Figure 2) (c.f. Lu et al. (2021)). In complex systems, the given governing equations may not be fully understood and thus inaccurate. Mis-specification can also occur due to non-existence of solutions to governing equations (as we see in subsection 5.2).

Physics-informed neural networks (PINNs) are an emerging tool for numerically solving PDEs, and are especially well-suited for inverse problems (Raissi et al., 2019; Karnakov et al., 2023). However, much of the PINN literature has focused on establishing the performance of PINNs for almost exclusively well-specified problems.

In this work, we introduce a set of tools for diagnosing and addressing mis-specified problems by reducing them to well-specified alternatives. This is accomplished using the Euler-Lagrange equations and other tools from mathematical analysis such as generalized functions. We harness this toolkit to prove bounds on the error of mis-specified PINNs, and to propose a diagnostic tool for determining whether given problems are mis-specified.

More formally, we consider problems of the following type:

$$\begin{aligned} \mathcal{N}[u] &= 0 \quad \text{on } \Omega \\ u &= f \quad \text{on } \mathcal{D} \end{aligned} \tag{1}$$

where $\mathcal{N}[u] = 0$ represents a PDE to be solved on a do-

main Ω, and f represents data collected and/or boundary conditions enforced on a subset $\mathcal{D}$ of the domain of interest. We say that Problem 1 is *well-specified* if there exists a piecewise-smooth function u which satisfies Equation 1; otherwise, the problem is *mis-specified*.

PINNs approach solving problems in the form of Equation 1 by minimizing a loss of the form

$$\mathcal{L}[u] := \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \lambda\|u - f\|^2_{L^2(\mathcal{D})}$$

where $\lambda > 0$ is the hyper-parameter that balances the contribution of two losses in the cost function (Raissi et al., 2019). Here and throughout, we will assume that the minimal loss value $\inf_u \mathcal{L}[u]$ is achieved by some piecewise smooth function u. With this said, the condition that Problem 1 is mis-specified is equivalent to the condition that $\min_u \mathcal{L}[u] > 0$.

Our contributions: We introduce a framework for transforming mis-specified problems of type 1 into well-specified problems, with the property that every minimizer of $\mathcal{L}[u]$ is a solution of the well-specified problem we define.[1] This allows us to

- Derive so-called *generalization bounds* of the form

  $$\|u - u^*\|^2 \leq C\tilde{\mathcal{L}}[u]$$

  based on preceding theory of Mishra & Molinaro (2022), where $u^* = \arg\min_u \mathcal{L}[u]$, and $\tilde{\mathcal{L}}[u]$ is the PINN loss corresponding to the well-specified problem we construct, whose rigorous definition we postpone to subsection 4.1.

- Propose a diagnostic tool for determining whether a given problem is mis-specified, according to the behavior of $\mathcal{L}$ and $\tilde{\mathcal{L}}$. This is based on theory showing that, with regularity assumptions,
  - If the original PDE problem is well-specified, then $\mathcal{L}[u] = O(\tilde{\mathcal{L}}[u])$.
  - If the original PDE problem is mis-specified, then $\mathcal{L}[u] = \Omega(1)$ for all u, while $\tilde{\mathcal{L}}[u] = o(1)$ for a certain sequence of u's.

1. Setting

Let $\mathcal{N}[u] = 0$ denote a PDE defined on a domain Ω, and $\mathcal{B}[u] = 0$ denote a set of additional conditions (typically data-based conditions enforced at the $\mathcal{B}$oundary), defined on a dataset $\mathcal{D}$ such as $\partial\Omega$. In Equation 1, we took $\mathcal{B}[u] = u - f$.

Formally, a *PDE problem* is a pair $(\mathcal{N}, \mathcal{B})$ of continuous differential operators defined on sets $\Omega, \mathcal{D} \subset \mathbb{R}^n$, respectively.

[1]Code to be made available at https://anonymous.4open.science/r/EulerPINNs-42E6/README.md.

Such a PDE problem is *well-specified* if the system

$$\begin{aligned} \mathcal{N}[u] &= 0 \quad \text{on } \Omega \\ \mathcal{B}[u] &= 0 \quad \text{on } \mathcal{D} \end{aligned} \tag{2}$$

has a piecewise-smooth solution u, and *mis-specified* otherwise.

Every PDE problem has an associated *PINN loss* $\mathcal{L}^{(\mathcal{N},\mathcal{B})}$, which is the functional defined by

$$\mathcal{L}^{(\mathcal{N},\mathcal{B})}[u] := \|\mathcal{N}[u]\|_{L^2(\Omega)} + \lambda\|\mathcal{B}[u]\|_{L^2(\mathcal{D})} \tag{3}$$

where $\lambda > 0$ is a hyperparameter. Where it is clear from context, we will simply denote $\mathcal{L} = \mathcal{L}^{(\mathcal{N},\mathcal{B})}$.

The *minimizer* of the PDE problem $(\mathcal{N}, \mathcal{B})$ is defined to be the minimizer $u^* \in \tilde{\mathcal{C}}^\infty$ of $\mathcal{L}^{(\mathcal{N},\mathcal{B})}$, if it exists, where $\tilde{\mathcal{C}}^\infty$ denotes the class of piecewise-smooth functions on Ω. In what follows, we will always assume that u^* exists.[2] When we write $\inf_u$ or $\min_u$, we will intend the optima to be taken over piecewise-smooth u, unless otherwise specified.

2. Overview of results

We introduce a set of tools for transforming mis-specified PDE problems into well-specified alternatives. More specifically, we define a transformation $\mathscr{T}$ associating to every PDE problem $(\mathcal{N}, \mathcal{B})$ a necessarily well-specified PDE problem $\mathscr{T}(\mathcal{N}, \mathcal{B})$, whose definition we postpone for simplicity to subsection 4.1. The $\mathscr{T}$-transformed problem has the property that every minimizer of $(\mathcal{N}, \mathcal{B})$ is a solution to $\mathscr{T}(\mathcal{N}, \mathcal{B})$, as summarized by the schematic in Figure 1. We will often denote $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}}) = \mathscr{T}(\mathcal{N}, \mathcal{B})$ and $\tilde{\mathcal{L}} = \mathcal{L}^{(\tilde{\mathcal{N}},\tilde{\mathcal{B}})}$. The $\mathscr{T}$-transformation involves the action of the *Euler operator* $\mathscr{EL}$, which is related to the Euler-Lagrange equations. Intuitively, $\mathscr{EL}$ can be thought of as a first derivative in function space; for more intuition on the $\mathscr{T}$-transformation, see Appendix A.

We demonstrate two applications of this transformation of mis-specified problems into well-posed problems:

2.1. Bounds on generalization error of mis-specified PINNs

Let u^* be the minimizer of a given PDE problem, i.e. the minimizer of a PINN loss $\mathcal{L}[u]$. Mishra & Molinaro (2022) answered the question "when does $\mathcal{L}[u] \to 0$ imply that $u \to u^*$?" for certain well-specified problems. For mis-specified problems, the premise fails, since by definition $\mathcal{L}[u^*] > 0$, so we never have $\mathcal{L}[u] \to 0$.[3]

[2]For sufficient conditions on the existence of a minimizer, we defer the reader to section 8.2 of Evans (2010).

[3]One may ask why we cannot simply define $\tilde{\mathcal{L}}[u] = \mathcal{L}[u] - \min_u \mathcal{L}$. The issue is that $\min_u \mathcal{L}$ is generally unknown for prob-

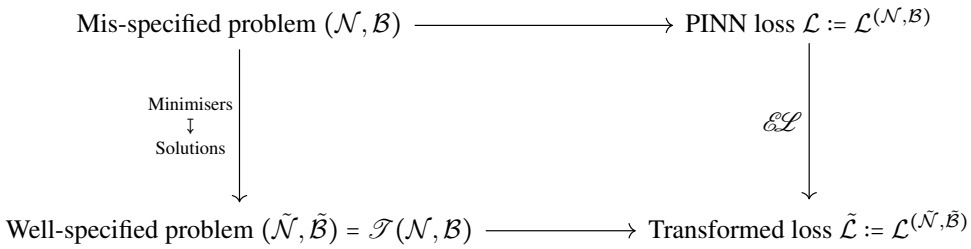


Figure 1: Outline of proposed method: mis-specified problems with associate loss $\mathcal{L}$ are transformed into well-specified problems with associated loss $\tilde{\mathcal{L}}$ through a procedure described in subsection 4.1 involving the Euler operator $\mathscr{EL}$.

How can we derive bounds on the rate at which $u \to u^*$ in the mis-specified setting? The $\mathscr{T}$-transformation provides one answer to this question: since $\tilde{\mathcal{L}}[u^*] = 0$ by construction, we can indeed have $\tilde{\mathcal{L}}[u] \to 0$ despite the fact that $\mathcal{L}[u] \nrightarrow 0$.

Provided sufficient regularity of the Euler operator, we demonstrate bounds of the form

$$\|u - u^*\|^2_{L^2(\Omega)} \leq C\tilde{\mathcal{L}}[u]$$

on the generalization error of PINNs (well-specified or not).

2.2. Diagnosing when PDE problems are mis-specified

Many PDE problems are mis-specified when combined with real-world data. On the other hand, some applied modeling problems may simply appear mis-specified on account of the limitations of the modeling method, such as poor optimization or expressivity.

How can practitioners distinguish between the cases of true PDE problem mis-specification and that of methodological limitations? We propose an application of the Euler operator as a diagnostic tool for qualitatively distinguishing these cases.

Namely, with regularity assumptions, we show that

- If the original PDE problem is well-specified, then $\mathcal{L}[u] = O(\tilde{\mathcal{L}}[u])$.
- If the original PDE problem is mis-specified, then $\mathcal{L}[u] = \Omega(1)$ for all u, while $\tilde{\mathcal{L}}[u] = o(1)$ for a certain sequence of u's.

Given these facts, consider minimizing $\tilde{\mathcal{L}}[u]$ and recording the values of $\mathcal{L}[u]$. If $\tilde{\mathcal{L}}[u]$ tends to zero while $\mathcal{L}[u]$ plateaus, this may provide some indication that the original PDE problem is mis-specified. On the other hand, if $\tilde{\mathcal{L}}[u]$ and $\mathcal{L}[u]$ plateau simultaneously, the lack of improvement may be instead due to the modeling method, since by construction $\min_u \tilde{\mathcal{L}}[u] = 0$.

lems which are potentially mis-specified. Indeed, if one knew e.g. whether or not $\min_u \mathcal{L} = 0$, one would immediately know whether the problem was well-specified or not.

3. Related work

The theoretical work that exists has largely focused on problems that are *well-specified* and have additional regularity such as uniqueness and stability of solutions. Mishra & Molinaro (2022; 2021); De Ryck & Mishra (2022); De Ryck et al. (2024) all show, in a variety of well-specified settings, that if u^* is the unique solution to Equation 2, then $\mathcal{L}[u] \to 0$ implies $\|u - u^*\| \to 0$ in various constructed norms. Separately, Shin (2020) show that, in well-specified settings, PINNs are asymptotically consistent, in the sense that $\|u - u^*\| \to 0$ as the number of boundary datapoints provided grows. Toward mis-specified settings, Doumèche et al. (2023) observe that in mis-specified settings, PINNs are *not* necessarily asymptotically consistent, and show that a regularization term is necessary to ensure consistency. Given the focus of the theoretical literature on well-specified settings, we aim to provide a general framework for reducing mis-specified problems to well-specified alternatives, thereby inheriting the preceding theory for the well-specified case.

PINNs are increasingly deployed in real world settings (Wang et al.; Eusebi et al., 2024; Teisberg et al., 2021), in which context one expects most models to be mis-specified (Lv & Liu, 2014). In Lu et al. (2021), a mis-specified PINN is employed by a practitioner who does not realize that the data specified does not satisfy the governing Helmholtz equation; in Wang et al., PINN mis-specification led to an observation that the equation previously believed to be governing the system of interest may not accurately model ice-sheet dynamics.

Making such observations, however, is complicated by the difficulty in distinguishing between the case of PDE mis-specification and the case of methodological (e.g. optimization, architectural, etc) shortcomings. Both cases will result in $\mathcal{L}[u]$ being bounded away from zero, in which domain the preceding theory of Mishra & Molinaro (2022) does not apply. To illustrate this challenge, consider that, across optimization schemes and architecture, it is difficult for PINNs to achieve loss at the level of machine precision even for well-specified problems, with loss smaller than 10^{-5} being unachievable in many practical settings (Wang & Lai, 2024).

On the other hand, in the deliberately mis-specified setting of Lu et al. (2021), PINN loss was of order 10^{-6}. In such contexts, it seems that the PINN loss alone is not sufficient to differentiate between well-specified and mis-specified problems. In this work, we circumvent this difficulty by considering the loss function of a *necessarily well-specified* problem constructed according to the Euler operator, allowing us to attribute large loss values universally to methodological shortcomings, rather than PDE mis-specification.

Recently, Zou et al. (2024) considers applications of physics-informed deep learning to mis-specified problems. However, their approach is to learn the governing dynamics automatically, in a manner similar to Chen et al. (2021) and Brunton et al. (2016).

The calculus of variations allows us to characterize minimizers of the loss by using variations in function space. Relatedly, Müller & Zeinhofer (2023) and Jnini et al. (2024) also consider variations in function space, although they do so using energy natural gradients instead of the calculus of variations, and consider their proposed scheme as a optimization method for solving well-specified problems, rather than examining mis-specified problems.

Lastly, there is a long history within mathematical PDEs of defining loss functionals with the property that they are minimized precisely when a PDE of interest is solved (Evans, 2010). This is not the approach taken by PINNs, but it is the approach taken by certain computational methods such as E & Yu (2018), which design auxiliary well-behaved loss functionals that are minimized precisely when a PDE of interest is solved. PINN loss functionals are typically not as well behaved as those designed by methods such as E & Yu (2018), and in this work we study the properties of their minimizers.

4. Formal results

4.1. Transforming mis-specified problems into well-posed ones with the $\mathscr{T}$-transformation

In this section we describe the general procedure for mapping mis-specified problems $(\mathcal{N}, \mathcal{B})$ to well-specified ones $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}}) = \mathscr{T}(\mathcal{N}, \mathcal{B})$.

This procedure involves computing a functional derivative of the PINN loss $\mathcal{L}$. This is accomplished using the Euler operator from the calculus of variations, which we define below in accordance with the Euler-Lagrange equation E-O; Dacorogna (2024). We state the definition only for first order here for readability, and postpone the general definition to Definition E.1 in the appendix.

Definition 4.1 (Euler operator)**.** The *Euler operator* is the operator which accepts a function $F(x, u, \nabla u)$ of x and the derivatives of u up to first-order and acts according to

$$\mathscr{EL}[F[u]](x) = \left[\frac{\partial}{\partial u} - \nabla \cdot \frac{\partial}{\partial \nabla u}\right] F(x, u, \nabla u) \quad (4)$$

The Euler operator can be thought of as a functional derivative of $\int_\Omega F$: the critical points u of $\int_\Omega F$ are those such that Equation 4 equals zero for all $x \in \Omega$. When set to zero for all x, Equation 4 is referred to as the Euler-Lagrange equation.

With this said, recall that the PINN loss $\mathcal{L} = \mathcal{L}^{(\mathcal{N},\mathcal{B})}$ typically consists of two terms, one an integral over Ω, and one an integral over $\mathcal{D}$. The $\mathscr{T}$-transformation consists of two steps:

1. Transforming the integral over $\mathcal{D}$ to one over Ω. This is typically done using either Stoke's theorem (the case $\mathcal{D} = \partial\Omega$), the theory of generalized functions such as the Dirac delta (the case when $\mathcal{D} \subset \Omega^o$), or Poincaré duality more generally.
2. Once the loss is expressed as a single integral over Ω, e.g.
$$\mathcal{L}[u] = \int_\Omega F[u]$$
we apply the Euler operator to F and define the transformed problem $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}})$ accordingly:
$$\tilde{\mathcal{N}}[u] = \mathscr{EL}[F[u]]$$

Fact 1. *Every minimizer of $\mathcal{L}[u] = 0$ is a solution to $\tilde{\mathcal{N}}[u] = 0$, as a consequence of Theorem D.2.*

The explicit form of $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}})$ is made precise by the following theorems.

4.1.1. $\mathscr{T}$-TRANSFORMATION FOR BOUNDARY DATA

The following theorem characterizes minimizers of mis-specified problems when boundary data is employed. For the general statement for arbitrary nonlinear PDEs, see Theorem B.1. We state the special case for first-order below:

Theorem 4.2. *Let $\mathcal{N}[u] = 0$ be an arbitrary first-order PDE, $\Omega \subset \mathbb{R}^n$ a domain with piece-wise smooth boundary $\partial\Omega$, and f smooth on $\partial\Omega$. Additionally, let $\mathscr{EL}$ denote the Euler operator. Consider the PINN loss function*

$$\mathcal{L}[u] = \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \lambda \|u - f\|^2_{L^2(\partial\Omega)}$$

Then every minimizer of $\mathcal{L}$ over $u \in \mathcal{C}^\infty$ is a solution to the PDE

$$\tilde{\mathcal{N}}[u] = \mathscr{EL}\left[\mathcal{N}[u]^2\right] = 0$$

on Ω, subject to the boundary conditions

$$\tilde{\mathcal{B}}[u] = 2\lambda(u - f) - \frac{\partial \mathcal{N}[u]^2}{\partial u_n} = 0$$

on $\partial\Omega$, where u_n denotes the component of ∇u normal to $\partial\Omega$. Thus, the loss function of the $\mathscr{T}$-transformed problem,

$$\tilde{\mathcal{L}}[u] = \| \mathscr{EL}\left[\mathcal{N}[u]^2\right] \|^2_{L^2(\Omega)} + \left\| 2\lambda(u-f) - \frac{\partial \mathcal{N}[u]^2}{\partial u_n} \right\|^2_{L^2(\partial\Omega)}$$

satisfies $\min \tilde{\mathcal{L}}[u] = 0$ *as long as* $\min \mathcal{L}[u]$ *exists.*

4.1.2. $\mathscr{T}$-TRANSFORMED FOR INTERIOR DATA

The following theorem characterizes minimizers of the mis-specified problem when interior data is employed. For the general statement for arbitrary nonlinear PDEs, see Theorem B.2. We state the special case of first-order PDEs below.

Theorem 4.3. *Let $\mathcal{N}[u]$ be a smooth first-order differential operator of order k, and $\mathcal{D} \subset \Omega$ be a finite set. Let $f : \mathcal{D} \longrightarrow \mathbb{R}$ be an arbitrary function and consider the PINN loss*

$$\mathcal{L}[u] = \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \lambda \|u - f\|^2_{L^2(\mathcal{D})}$$

Then every minimizer of over $u \in \tilde{\mathcal{C}}^\infty$ is a solution to the PDE

$$\tilde{\mathcal{N}}[u] = \mathscr{EL}\left[\mathcal{N}[u]^2\right] + 2\lambda \sum_{x\in\mathcal{D}} (u-f)\delta_x = 0$$

on Ω, subject to the boundary conditions

$$\tilde{\mathcal{B}}[u] = \frac{\partial \mathcal{N}[u]^2}{\partial u_n} = 0$$

on $\partial\Omega$, where u_n denotes the component of ∇u normal to $\partial\Omega$, and δ_x denotes the Dirac delta distribution centered at x.

Moreover, if $(\tilde{\mathcal{N}}, \mathcal{B})$ is mis-specified, then every solution of $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}})$ must develop singularities of order ≤ 2 precisely at the points $x \in \mathcal{D}$ such that $u(x) \neq f(x)$. Namely, ***there are no smooth minimizers*** *u* ***of*** *$\mathcal{L}$ in this case.*

4.2. Bounds on generalization error of mis-specified PINNs

In this section, we demonstrate generalization bounds on the error of PINNs, whether mis-specified or not, by employing the loss of the $\mathscr{T}$-transformed problem.

We recall the standard stability assumption of van der Meer et al. (2022), Mishra & Molinaro (2022), and Mishra & Molinaro (2021):

Assumption 4.4 (Stability of PDE problem $(\mathcal{N}, \mathcal{B})$)**.** Suppose that there exists a constant C such that

$$\|u - v\| \leq C\left(\|\mathcal{N}[u] - \mathcal{N}[v]\| + \|\mathcal{B}[u] - \mathcal{B}[v]\|\right)$$

where each norm is $\|\cdot\|_{L^2(\Omega)}$.

This assumption holds for many linear PDEs (see e.g. Chapter 6 of Evans (2010) or Dautray & Lions (2000)) and for some nonlinear PDEs whose linearized operator has a bounded inverse (see Mishra & Molinaro (2022)).

Consider any PDE problem $(\mathcal{N}, \mathcal{B})$, and the corresponding well-specified problem induced by the $\mathscr{T}$-transformation. Then we have the following result:

Theorem 4.5. *Suppose that Assumption 4.4 holds for the transformed problem $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}})$, and suppose that the original problem $(\mathcal{N}, \mathcal{B})$ has the minimizer $u^* = \arg\min_u \mathcal{L}[u]$. Then for $c = C\sqrt{2}\max\{1, 1/\sqrt{\lambda}\}$,*

$$\|u - u^*\|_{L^2(\Omega)} \leq c\tilde{\mathcal{L}}[u]^{1/2}$$

4.3. Diagnosing when PDE problems are mis-specified

Let $\mathcal{N}[u] = 0$ denote a first-order PDE on a domain Ω, and $f : \partial\Omega \longrightarrow \mathbb{R}$ denote a continuous function on $\partial\Omega$. Define

$$\mathcal{L}[u] = \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \lambda \|u - f\|^2_{L^2(\partial\Omega)}$$

and by $\mathscr{EL}$ the Euler operator. Let

$$\tilde{\mathcal{L}}[u] := \left\|\mathscr{EL}\left[\mathcal{N}[u]^2\right]\right\|^2_{L^2(\Omega)} + \lambda \left\| 2(u-f) - \frac{\partial \mathcal{N}[u]^2}{\partial u_n} \right\|^2_{L^2(\partial\Omega)}$$

denote the loss of the $\mathscr{T}$-transformed problem (c.f. Theorem 4.2).

Assumption 4.6 (Stability of $\mathscr{EL}$)**.** Suppose that there is a constant C such that

$$\|\mathcal{N}[u] - \mathcal{N}[v]\|_{W^{1,2}(\Omega)} \leq C \left\|\mathscr{EL}\left[\mathcal{N}[u]^2 - \mathcal{N}[v]^2\right]\right\|_{L^2(\Omega)}$$

for all smooth u, v on Ω. Moreover, suppose that there is a constant D such that

$$\|\partial_{u_\alpha} \mathcal{N}[u]^2\|_{L^2(\partial\Omega)} \leq D \|\mathcal{N}[u]\|_{L^2(\partial\Omega)}$$

where u_α denotes a partial derivative of u.

For sufficient conditions for the second assumption, note that this holds for all linear PDEs with bounded coefficients: writing $\mathcal{N}[u] = \sum_\alpha a(x;\alpha) u_\alpha$ we see that $\partial_{u_\alpha} \mathcal{N}[u]^2 = 2\mathcal{N}[u] a(x;\alpha)$. By Hölder's inequality, we have $\|\partial_{u_\alpha} \mathcal{N}[u]^2\|_{L^2} \leq 4 \sup_{x\in\Omega} a(x;\alpha) \|\mathcal{N}[u]\|_{L^2}$.

Theorem 4.7. *Suppose that $\mathcal{N}$ satisfies Assumption 4.6. Suppose*

$$u^* = \arg\min_u \mathcal{L}[u]$$

exists. Then there exists a constant c depending on $\mathcal{N}$, Ω and λ such that

$$\mathcal{L}[u] \leq c\,\tilde{\mathcal{L}}[u] \qquad \forall\, u$$

if and only if

$$\mathcal{N}[u^*] = 0$$

i.e., if and only if the original problem is well-specified.

4.4. When is a solution to the $\mathscr{T}$-transformed problem a minimizer of $\mathcal{L}$?

So far, we have shown that solving the $\mathscr{T}$-transformed problem is a *necessary* condition for minimizing $\mathcal{L}$. When is it also sufficient? The strongest result we can obtain in this direction is for linear PDEs. As usual, we prove the general result for arbitrary linear PDEs in Theorem B.3, and state the special case of first-order linear PDEs below.

Theorem 4.8. *Let $\mathcal{N}[u] = 0$ be a first-order linear PDE, and let $\mathcal{B}[u] = u - f$ provide data for u on any subdomain $\mathcal{D} \subset \Omega$. Consider the PINN loss function*

$$\mathcal{L}[u] = \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \|u - f\|^2_{L^2(\mathcal{D})}$$

Then the minimizers of $\mathcal{L}[u]$ are precisely the solutions to the $\mathscr{T}$-transformed problem specified by

$$\tilde{\mathcal{N}}[u] = \mathscr{EL}\left[\mathcal{N}[u]^2\right] + 2\lambda(u - f) = 0$$

subject to the boundary conditions

$$\tilde{\mathcal{B}}[u] = \frac{\partial \mathcal{N}[u]^2}{\partial u_n} = 0.$$

Even in the most general case of highly nonlinear PDEs, we show that the loss $\tilde{\mathcal{L}}[u]$ of the $\mathscr{T}$-transformed problem is not trivial, in the sense that not every function minimizes it unless the underlying PDE is trivial. Note that we state the following theorem for PDEs of arbitrary order, not just first.

Theorem 4.9. *Suppose that $\mathcal{N}$ is a differential operator which is not identically zero on every subdomain of Ω, and that $(\mathcal{N}, \mathcal{B})$ is well-specified. Then not every smooth u is a minimizer of the loss of the transformed problem, $\tilde{\mathcal{L}}[u]$.*

5. Examples

The purpose of this section is to demonstrate the $\mathscr{T}$-transformation through two simple example problems. In each case, we provide simple arguments that the problems are mis-specified, but assume that *the practitioner may be ignorant* of this mis-specification. In these settings, we observe how the standard PINN problem setup behaves, before comparing it to the proposed $\mathscr{T}$-transformed problem.

5.1. ODE example with interior data

Consider an ODE of the form

$$\mathcal{N}[u] = u' - G(u) = 0$$

on the interval $[0, 1] \subset \mathbb{R}$. Provided G is continuous and bounded, there is a global solution to the governing equation which passes through any given point (x_0, y_0). Provided G is globally Lipschitz, there is a unique such global solution.[4]

[4] Theorem B, Section 70, Chapter 13 of Simmons (2022).

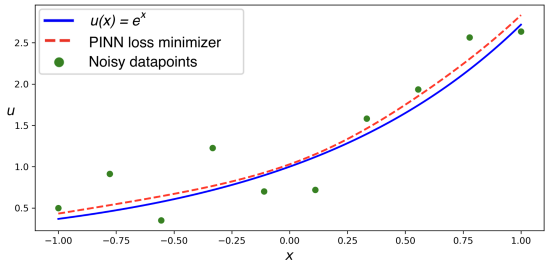


Figure 2: 10 datapoints are drawn from the function $u(x) = e^x$, and corrupted with Gaussian noise of variance $1/3$. The corresponding PINN loss $\mathcal{L}[u]$ consists of one term which penalizes deviation from these 10 noisy datapoints, and one term which penalizes deviation from the theorized governing equation $u = u'$. The theoretical minimizer of $\mathcal{L}$ is plotted in red, after being derived in Appendix F.

For example, when G is the identity, this ODE has the solutions $u(x) = \alpha e^x$ for some $\alpha \in \mathbb{R}$, and hence is uniquely determined by specifying the value of u to be $u(x_0) = y_0$ at a single point (x_0, y_0) (in which case we have $\alpha = y_0/e^{x_0}$). This particular example is showcased in Figure 2.

Therefore, when G is globally Lipschitz, the PDE problem

$$\begin{array}{rcll} \mathcal{N}[u] & = & u' - G(u) = 0 & \text{on } [0,1] \\ u(x_i) & = & y_i & \text{for } 1 \le i \le n \end{array} \tag{5}$$

is well-specified when $n = 1$, and mis-specified when $n > 1$ for almost every choice of data.[5]

Suppose that this fact is not known to a practitioner, who accidentally prescribes data which specifies the value of u at a set of multiple points $\mathcal{D} = \{x_i\}_{i \le n} \subset (0, 1)$. The corresponding PINN loss function is then

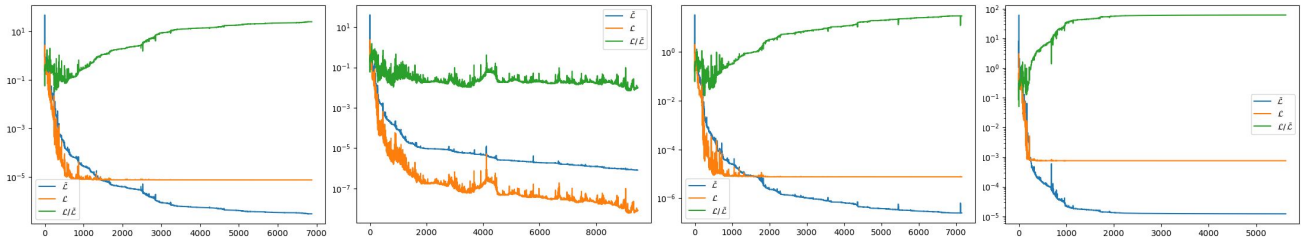
$$\mathcal{L}[u] = \int_0^1 (u'(x) - G(u(x)))^2 dx + \lambda \sum_{i=1}^n (u(x_i) - y_i)^2$$

We will now construct the $\mathscr{T}$-transformed problem corresponding to this mis-specified problem, and in the case when G is the identity, we will show that the minimum of value of $\mathcal{L}[u]$ is positive, and that it is uniquely obtained by some u^* which is also the unique solution to the transformed problem we construct.

The first step is to transform $\mathcal{L}[u]$ such that it is a single integral over $[0, 1]$: this will allow us to subsequently apply the Euler operator. This is accomplished by using Dirac delta distributions:

$$\mathcal{L}[u] = \int_0^1 (u' - G(u))^2 + \lambda \sum_{i=1}^n (u - y)^2 \delta_{x_i} dx$$

[5] I.e., with probability 1, assuming any model in which the datapoints are drawn i.i.d. from any distribution which is absolutely continuous with respect to the Lebesgue measure.



(a) Mis-specified: $\alpha = 0.99$. The ratio $\mathcal{L}/\tilde{\mathcal{L}}$ *increases*.
(b) Well-specified: $\alpha = 1.00$. The ratio $\mathcal{L}/\tilde{\mathcal{L}}$ *is bounded*.
(c) Mis-specified: $\alpha = 1.01$. The ratio $\mathcal{L}/\tilde{\mathcal{L}}$ *increases*.
(d) Mis-specified: $\alpha = 1.1$. The ratio $\mathcal{L}/\tilde{\mathcal{L}}$ *increases*.

Figure 3: Example with fitting $u_x = t$, $u_t = \alpha x$ by minimizing $\tilde{\mathcal{L}}$. This is well-specified only when $\alpha = 1.00$. Throughout training, we plot $\tilde{\mathcal{L}}$ (blue), $\mathcal{L}$ (orange), and $\mathcal{L}/\tilde{\mathcal{L}}$ (green). We optimize using L-BFGS until convergence and a PINN with width 20 and depth 8.

where $y : [0,1] \longrightarrow \mathbb{R}$ is any smooth function satisfying $y(x_i) = y_i$. Now that $\mathcal{L}[u]$ has the form $\int_0^1 F[u]dx$, we can apply the Euler operator to the integrand,[6] resulting in

$$\begin{aligned}&\mathscr{EL}[(u' - G(u))^2 + \lambda \sum_{i=1}^n (u-y)^2 \delta_{x_i}] \\ &= 2(G(u)G'(u) - u'') + 2\lambda \sum_{i=1}^n (u-y)\delta_{x_i}\end{aligned}$$

Applying the natural boundary conditions of Theorem D.2, which enforce that $\partial F/\partial u' = 0$ at the boundary $\{0,1\}$, we recover that the governing equation $u' = u$ must be satisfied at the boundaries. The $\mathscr{T}$-transformed problem is therefore determined by

$$\begin{aligned}\tilde{\mathcal{N}}[u] &= (G(u)G'(u) - u'') \\ &\quad + \lambda \sum_{i=1}^n (G(u) - y)\delta_{x_i} = 0 \qquad \text{on } [0,1] \\ \tilde{\mathcal{B}}[u] &= u - u' = 0 \qquad \text{on } \{0,1\}\end{aligned}$$

In the appendix, we solve this problem exactly and demonstrate that it has a unique solution $u^* \in \tilde{\mathcal{C}}^\infty$ when G is the identity. Coupled with the fact that $F[u]$ is convex in (u, u') in this case, we see that $\mathcal{L}[u]$ must have u^* as a unique minimizer via Theorem D.2.

Taken together, we have shown that the minimum $\min_u \mathcal{L} > 0$ is achieved by a unique piecewise smooth u^*, which is also the unique solution to the $\mathscr{T}$-transformed problem.

5.2. PDE example with boundary data

In the previous section, problem mis-specification was due to a mismatch between the governing equation $\mathcal{N}[u] = 0$ and the provided data. In this section, we will see how problem mis-specification can occur as a result of incompatibility within a system of governing equations.

[6]Note that Theorem D.2 does not apply here, because the integrand $F[u]$ is not $\mathcal{C}^2$. Nevertheless, the Euler-Lagrange equation still holds in this context, per Konjik et al. (2008) and Derr & Kinzebulatov (2007)

Consider the function $u(x,y) = xy$ on the unit square $[0,1]^2 \subset \mathbb{R}$. This function satisfies the system of PDEs

$$\begin{aligned}\frac{\partial u}{\partial x} &= y \\ \frac{\partial u}{\partial y} &= x\end{aligned} \tag{6}$$

but clearly not the system

$$\begin{aligned}\frac{\partial u}{\partial x} &= y \\ \frac{\partial u}{\partial y} &= \alpha x\end{aligned} \tag{7}$$

where $\alpha \neq 1$ is arbitrary. In fact, for $\alpha \neq 1$, Equation 7 has *no* $\mathcal{C}^2$ *solutions* (not even locally) on any open set $U \subset \mathbb{R}$. To see this, note that if $u \in \mathcal{C}^2(U)$, then

$$1 = \frac{\partial^2 u}{\partial y \partial x} = \frac{\partial^2 u}{\partial x \partial y} = \alpha$$

In this case, the argument above was sufficient to prove that Equation 7 is a mis-specified system.[7] In more general cases, however, it may be much less clear if a problem is mis-specified.

We consider a practitioner who uses a PINN to fit Equation 7 for general α. So that the problem is uniquely determined, as usual, we enforce the boundary conditions that $u = f$ on the boundary $\partial[0,1]^2$ of the square, for arbitrary smooth f. This yields the PINN loss function

$$\begin{aligned}\mathcal{L}[u] &= \int_0^1 \int_0^1 (u_x - y)^2 + (u_y - \alpha x)^2 dxdy \\ &\quad + \lambda \Bigg[\int_0^1 (u-f)^2(0,y)dy + \int_0^1 (u-f)^2(x,0)dx \\ &\quad + \int_0^1 (u-f)^2(1,y)dy + \int_0^1 (u-f)^2(x,1)dx \Bigg]\end{aligned}$$

Again, the first step is to transform $\mathcal{L}[u]$ such that it is a single integral over the entire domain $\Omega = [0,1]^2$: this will

[7]In fact, that this system is mis-specified can be seen to be an example of the much more general Cauchy–Kovalevskaya–Kashiwara theorem.

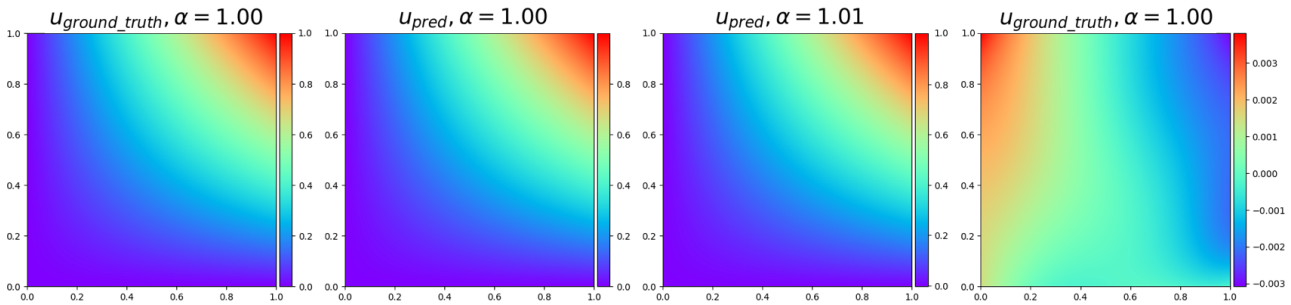


Figure 4: Predicted u as a result of minimizing the PINN loss $\mathcal{L}$ considered in subsection 5.2, with $\alpha \in \{1.00, 1.01\}$. From left to right: (i) ground truth for $\alpha = 1.00$; (ii) PINN prediction for $\alpha = 1.00$; (iii) PINN prediction for $\alpha = 1.01$; (iv) difference between PINN predictions for $\alpha = 1.00, 1.01$. There is little difference between the learned u fields: the relative error is less than 0.5%. Additionally, each prediction has less than 0.5% relative error from the ground truth for $\alpha = 1.00$.

subsequently allow us to apply the Euler operator. Here, this is accomplished using Green's theorem: one can see that the sum of integrals multiplied by λ has the alternate form

$$\lambda \int_C (2y-1)(u-f)^2 dx + (2x-1)(u-f)^2 dy \quad (8)$$

where C is a curve traversing $\partial\Omega$ counter-clockwise. By Green's theorem, Equation 8 is equal to

$$\lambda \int_0^1 \int_0^1 4(u-f)(u_x+u_y) dx dy$$

And thus we may write $\mathcal{L}[u]$ with a single integral as

$$\mathcal{L}[u] = \int_\Omega (u_x - y)^2 + (u_y - \alpha x)^2 + 4\lambda(u-f)(\nabla \cdot u)$$

Finally, we can apply the Euler operator to the integrand, resulting in the Laplacian $-2\Delta u$. Applying the natural boundary conditions of Theorem D.2, which enforce that $\partial F / \partial \nabla u \cdot n$ at the boundary $\partial\Omega$, we recover that u must satisfy certain Robin boundary conditions. Taken together, the $\mathscr{T}$-transformed is the Poisson problem:

$$\begin{aligned}
\tilde{\mathcal{N}}[u] &= -2\Delta u = 0 && \text{for } x, y \in [0,1]^2 \\
\tilde{\mathcal{B}}_1[u] &= 2(u_x - y) + 4\lambda(u-f) = 0 && \text{for } x \in \{0,1\} \\
\tilde{\mathcal{B}}_2[u] &= 2(u_y - \alpha x) + 4\lambda(u-f) = 0 && \text{for } y \in \{0,1\}
\end{aligned}$$

This problem has a unique solution $u^* \in L^2([0,1]^2)$, per Lanzani & Méndez (2006). Again, coupled with the fact that the integrand of $\mathcal{L}[u]$ is convex in u and its derivatives, we see that $\mathcal{L}[u]$ has u^* as a unique minimizer via Theorem D.2. Moreover, we remark that this problem satisfies Assumption 4.4, per Theorem 5.2 of Lanzani & Méndez (2006).

In Figure 4, we plot the PINN predictions for u with system of governing equations given by Equation 7, with $\alpha = 1.00, 1.01$. As expected, the predictions are in close agreement with each other and with the ground truth for $\alpha = 1.00$ (which is $u(x,y) = xy$). However, when plotting $\mathcal{L}$ and $\tilde{\mathcal{L}}$ in each case, $\mathcal{L}$ is bounded away from zero in only the cases when $\alpha \neq 1.00$: the results are shown in Figure 3.

5.3. Redux

In each of the previous cases, we have presented mis-specified PDE problems $(\mathcal{N}, \mathcal{B})$ and shown how the well-specified $\mathscr{T}$-transformed problem $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}})$ is constructed so that every minimizer of $(\mathcal{N}, \mathcal{B})$ is a solution to $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}})$.

We emphasize that nearly everything shown in the previous examples about the construction of the transformed problem generalizes: in any dimension, a finite number of interior datapoints can always be handled via Dirac delta distributions, and boundary data can always be handled by Stokes' theorem. Precise versions of these statements are provided in Theorem 4.2 and Theorem 4.3.

Moreover, while the previous examples may seem to have involved tedious calculations, nearly all of these calculations are mechanistically computable with automatic differentiation in TensorFlow or PyTorch. Lastly, we remark that the $\mathscr{T}$-transformed problem is always well-specified, but its solutions may not always be unique. In these examples presented here, and for certain well-behaved PDE in general (see Theorem 4.8), uniqueness holds.

6. Conclusion

In this paper, we introduced a framework to derive well-specified PDE problems from mis-specified ones and used it to prove error bounds for mis-specified PINNs. We also proposed a diagnostic tool to detect mis-specified problems. As our work is primarily theoretical, future research should focus on experimental validation, particularly to test the robustness of our diagnostic tool in practical settings. Additionally, there is potential for theoretical advancements by relaxing the assumptions needed to prove error bounds and tool guarantees, with insights likely drawn from the *stability and regularity of the Euler operator* in the calculus of variations.

References

Euler operator - encyclopedia of mathematics. URL https://encyclopediaofmath.org/wiki/Euler_operator.

Euler–lagrange equation_2024, May 2024. URL https://en.wikipedia.org/w/index.php?title=Euler%E2%80%93Lagrange_equation&oldid=1224640696. Page Version ID: 1224640696.

Browning, A. P., Warne, D. J., Burrage, K., Baker, R. E., and Simpson, M. J. Identifiability analysis for stochastic differential equation models in systems biology. *Journal of The Royal Society Interface*, 17 (173):20200652, December 2020. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2020.0652. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2020.0652.

Brunton, S. L., Proctor, J. L., and Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15): 3932–3937, April 2016. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1517384113. URL https://pnas.org/doi/full/10.1073/pnas.1517384113.

Chen, Z., Liu, Y., and Sun, H. Physics-informed learning of governing equations from scarce data. *Nature Communications*, 12(1):6136, October 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-26434-1. URL https://www.nature.com/articles/s41467-021-26434-1.

Dacorogna, B. *Introduction to the Calculus of Variations*, volume a of *Essential Textbooks in Mathematics*. WORLD SCIENTIFIC (EUROPE), 4 edition, July 2024. ISBN 9781800615267. doi: 10.1142/q0451. URL https://www.worldscientific.com/worldscibooks/10.1142/q0451.

Dautray, R. and Lions, J.-L. *Mathematical Analysis and Numerical Methods for Science and Technology*. Springer Berlin Heidelberg, 2000. doi: 10.1007/978-3-642-61566-5. URL https://link.springer.com/deleted.

De Ryck, T. and Mishra, S. Error analysis for physics-informed neural networks (pinns) approximating kolmogorov pdes. *Advances in Computational Mathematics*, 48(6):79, December 2022. ISSN 1019-7168, 1572-9044. doi: 10.1007/s10444-022-09985-9. URL https://link.springer.com/10.1007/s10444-022-09985-9.

De Ryck, T., Jagtap, A. D., and Mishra, S. Error estimates for physics-informed neural networks approximating the navier–stokes equations. *IMA Journal of Numerical Analysis*, 44(1):83–119, February 2024. ISSN 0272-4979, 1464-3642. doi: 10.1093/imanum/drac085. URL https://academic.oup.com/imajna/article/44/1/83/6989836.

Derr, V. Y. and Kinzebulatov, D. M. Dynamical generalized functions and the multiplication problem. *Russian Mathematics*, 51(5):32–43, May 2007. ISSN 1066-369X, 1934-810X. doi: 10.3103/S1066369X07050040. URL http://link.springer.com/10.3103/S1066369X07050040.

Doumèche, N., Biau, G., and Boyer, C. Convergence and error analysis of pinns. 2023. doi: arXiv:2305.01240.

Duffin, C., Cripps, E., Stemler, T., and Girolami, M. Statistical finite elements for misspecified models. *Proceedings of the National Academy of Sciences*, 118(2):e2015006118, January 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2015006118. URL https://pnas.org/doi/full/10.1073/pnas.2015006118.

E, W. and Yu, B. The deep ritz method: A deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, March 2018. ISSN 2194-6701, 2194-671X. doi: 10.1007/s40304-018-0127-z. URL http://link.springer.com/10.1007/s40304-018-0127-z.

Eusebi, R., Vecchi, G. A., Lai, C.-Y., and Tong, M. Realistic tropical cyclone wind and pressure fields can be reconstructed from sparse data using deep learning. *Communications Earth Environment*, 5(1):8, January 2024. ISSN 2662-4435. doi: 10.1038/s43247-023-01144-2. URL https://www.nature.com/articles/s43247-023-01144-2.

Evans, L. C. *Partial differential equations*. American Mathematical Society, Providence, R.I., 2010. ISBN 9780821849743 0821849743.

Jnini, A., Vella, F., and Zeinhofer, M. Gauss-newton natural gradient descent for physics-informed computational fluid dynamics, 2024.

Karnakov, P., Litvinov, S., and Koumoutsakos, P. Solving inverse problems in physics by optimizing a discrete loss: Fast and accurate learning without neural networks. *PNAS Nexus*, 3(1):pgae005, December 2023. ISSN 2752-6542. doi: 10.1093/pnasnexus/pgae005. URL https://academic.oup.com/pnasnexus/article/doi/10.1093/pnasnexus/pgae005/7516080.

Konjik, S., Kunzinger, M., and Oberguggenberger, M. Foundations of the calculus of variations in generalized function algebras. *Acta Applicandae Mathematicae*, 103(2):169–199, September 2008. ISSN 0167-8019, 1572-9036. doi: 10.1007/s10440-008-9228-0. URL http://link.springer.com/10.1007/s10440-008-9228-0.

Lanzani, L. and Méndez, O. The poisson's problem for the laplacian with robin boundary condition in non-smooth domains. *Revista Matemática Iberoamericana*, 22(1): 181–204, April 2006. ISSN 0213-2230, 2235-0616. doi: 10.4171/rmi/453. URL https://ems.press/doi/10.4171/rmi/453.

Lu, L., Pestourie, R., Yao, W., Wang, Z., Verdugo, F., and Johnson, S. G. Physics-informed neural networks with hard constraints for inverse design. *SIAM Journal on Scientific Computing*, 43(6):B1105–B1132, January 2021. ISSN 1064-8275, 1095-7197. doi: 10.1137/21M1397908. URL https://epubs.siam.org/doi/10.1137/21M1397908.

Lv, J. and Liu, J. S. Model selection principles in misspecified models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):141–167, January 2014. ISSN 1369-7412, 1467-9868. doi: 10.1111/rssb.12023. URL https://academic.oup.com/jrsssb/article/76/1/141/7075936.

Mishra, S. and Molinaro, R. Estimates on the generalization error of physics-informed neural networks for approximating a class of inverse problems for PDEs. *IMA Journal of Numerical Analysis*, 42(2): 981–1022, 06 2021. ISSN 0272-4979. doi: 10.1093/imanum/drab032. URL https://doi.org/10.1093/imanum/drab032.

Mishra, S. and Molinaro, R. Estimates on the generalization error of physics-informed neural networks for approximating PDEs. *IMA Journal of Numerical Analysis*, 43(1):1–43, 01 2022. ISSN 0272-4979. doi: 10.1093/imanum/drab093. URL https://doi.org/10.1093/imanum/drab093.

Muchiri, D. K., Hewett, J. N., Sellier, M., Moyers-Gonzalez, M., and Monnier, J. Numerical simulations of dam-break flows of viscoplastic fluids via shallow water equations. *Theoretical and Computational Fluid Dynamics*, May 2024. ISSN 0935-4964, 1432-2250. doi: 10.1007/s00162-024-00696-z. URL https://link.springer.com/10.1007/s00162-024-00696-z.

Müller, J. and Zeinhofer, M. Achieving high accuracy with pinns via energy natural gradient descent. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Murphy, R. J., Maclaren, O. J., and Simpson, M. J. Implementing measurement error models with mechanistic mathematical models in a likelihood-based framework for estimation, identifiability analysis, and prediction in the life sciences, 2023.

Olver, P. J. and Sivaloganathan, J. The structure of null lagrangians. *Nonlinearity*, 1(2):389–398, May 1988. ISSN 0951-7715, 1361-6544. doi: 10.1088/0951-7715/1/2/005. URL https://iopscience.iop.org/article/10.1088/0951-7715/1/2/005.

Raissi, M., Perdikaris, P., and Karniadakis, G. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, February 2019. ISSN 00219991. doi: 10.1016/j.jcp.2018.10.045. URL https://linkinghub.elsevier.com/retrieve/pii/S0021999118307125.

Shin, Y. On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type pdes. *Communications in Computational Physics*, 28(5):2042–2074, June 2020. ISSN 1815-2406, 1991-7120. doi: 10.4208/cicp.OA-2020-0193. URL http://global-sci.org/intro/article_detail/cicp/18404.html.

Simmons, G. F. *Differential equations with applications and historical notes*. Textbooks in mathematics. CRC Press, Taylor Francis Group, Boca Raton London New York, third edition, first edition in paperback edition, 2022. ISBN 9781498702591.

Teisberg, T. O., Schroeder, D. M., and MacKie, E. J. A machine learning approach to mass-conserving ice thickness interpolation. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp. 8664–8667, Brussels, Belgium, July 2021. IEEE. ISBN 9781665403696. doi: 10.1109/IGARSS47720.2021.9555002. URL https://ieeexplore.ieee.org/document/9555002/.

van der Meer, R., Oosterlee, C. W., and Borovykh, A. Optimally weighted loss functions for solving pdes with neural networks. *Journal of Computational and Applied Mathematics*, 405:113887, 2022. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2021.113887. URL https://www.sciencedirect.com/science/article/pii/S0377042721005100.

Wang, Y. and Lai, C.-Y. Multi-stage neural networks: Function approximator of machine precision.

Journal of Computational Physics, 504:112865, May 2024. ISSN 00219991. doi: 10.1016/j.jcp.2024.112865. URL https://linkinghub.elsevier.com/retrieve/pii/S0021999124001141.

Wang, Y., Lai, C.-Y., Prior, D., and Cowen-Breen, C. Discovering the rheology of antarctic ice shelves via physics-informed deep learning.

Zou, Z., Meng, X., and Karniadakis, G. E. Correcting model misspecification in physics-informed neural networks (pinns). *Journal of Computational Physics*, 505:112918, May 2024. ISSN 00219991. doi: 10.1016/j.jcp.2024.112918. URL https://linkinghub.elsevier.com/retrieve/pii/S0021999124001670.

A. Intuition for $\mathscr{T}$-transformation

Intuitively, the $\mathscr{T}$-transformation amounts to taking a first derivative of the PINN loss $\mathcal{L}$ in function space, so that the solutions to $\mathscr{T}(\mathcal{N}, \mathcal{B})$ are precisely the critical points of $\mathcal{L}$. Importantly, this is sensible because critical points of $\mathcal{L}$ in function space are often genuine minima (see e.g. Theorem 4.8), in contrast to the typically non-convex loss landscapes encountered in parameter space within machine learning.

The $\mathscr{T}$-transformation is made precise by Theorem 4.2 (the case $\mathcal{D}$ is the boundary of Ω) and Theorem 4.3 (the case when $\mathcal{D}$ contains interior points of Ω). In each case, we are able to characterize the critical points of $\mathcal{L}$ as solutions to the *Euler-Lagrange equations* for a suitable Lagrangian, and the "first-derivative in function space" is computed with the *Euler operator*, which we denote by $\mathscr{EL}$.

By transforming mis-specified PDE problems into well-specified ones, the $\mathscr{T}$-transformation provides one avenue for proving bounds on generalization error of mis-specified PINNs, as well as guarantees about the diagnostic tool we propose. At present, such statements require assumptions (namely, 4.4 and 4.6) about the *stability and regularity of the $\mathscr{EL}$ operator*, which is a well-studied area with the field of the calculus of variations. In future work, we expect that many insights can be drawn from this area in order to relax these assumptions.

B. Generalized theorem statements

Here we state versions of Theorem 4.2, Theorem 4.3 and Theorem 4.8 generalized to the case of higher-order derivatives.

Theorem B.1. *Let $\mathcal{N}[u] = 0$ be an arbitrary first-order PDE, $\Omega \subset \mathbb{R}^n$ a domain with piece-wise smooth boundary $\partial\Omega$, and f smooth on $\partial\Omega$. Additionally, let $\mathscr{EL}$ denote the Euler operator. Consider the PINN loss function*

$$\mathcal{L}[u] = \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \lambda \|u - f\|^2_{L^2(\partial\Omega)}$$

Then every minimizer of $\mathcal{L}$ over $u \in \mathcal{C}^\infty$ is a solution to the PDE

$$\tilde{\mathcal{N}}[u] = \mathscr{EL}\left[\mathcal{N}[u]^2\right] = 0$$

on Ω, subject to the boundary conditions the system of boundary conditions

$$\tilde{\mathcal{B}}^{(\gamma)}[u] = \begin{cases} \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] & \gamma \neq \varnothing \\ \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] + 2\lambda(u - f) & \gamma = \varnothing \end{cases} = 0$$

on $\partial\Omega$ for every multi-subset γ of $[n]$ of size k, where $\mathscr{EL}^{(\gamma)}$ is defined in Definition E.1. Thus, the loss of the $\mathscr{T}$-transformed problem,

$$\tilde{\mathcal{L}}[u] = \| \mathscr{EL}\left[\mathcal{N}[u]^2\right] \|^2_{L^2(\Omega)} + \left\| 2\lambda(u - f) + \mathscr{EL}^{(\varnothing)}\left[\mathcal{N}[u]^2\right] \right\|^2_{L^2(\partial\Omega)} + \sum_{\gamma \neq \varnothing} \left\| \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] \right\|^2_{L^2(\partial\Omega)}$$

satisfies $\min \tilde{\mathcal{L}}[u] = 0$ as long as $\min \mathcal{L}[u]$ exists.

Theorem B.2. *Let $\mathcal{N}[u]$ be a smooth first-order differential operator of order k, and $\mathcal{D} \subset \Omega \subset \mathbb{R}^n$ be a finite set. Let $f : \mathcal{D} \longrightarrow \mathbb{R}$ be an arbitrary function and consider the PINN loss*

$$\mathcal{L}[u] = \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \lambda \|u - f\|^2_{L^2(\mathcal{D})}$$

Then every minimizer of over $u \in \tilde{\mathcal{C}}^\infty$ is a solution to the PDE

$$\tilde{\mathcal{N}}[u] = \mathscr{EL}\left[\mathcal{N}[u]^2\right] + 2\lambda \sum_{x \in \mathcal{D}} (u - f)\delta_x = 0$$

on Ω, subject to the system of boundary conditions

$$\tilde{\mathcal{B}}^{(\gamma)}[u] = \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] = 0$$

on $\partial\Omega$ for every multi-subset γ of $[n]$ of size k, where $\mathscr{EL}^{(\gamma)}$ is defined in Definition E.1 and δ_x denotes the Dirac delta distribution centered at x.

Theorem B.3. *Let $\mathcal{N}[u] = 0$ be a linear PDE, and let $\mathcal{B}[u] = u - f$ provide data for u on the entire domain $\mathcal{D} = \Omega$. Consider the PINN loss function*

$$\mathcal{L}[u] = \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \lambda\|u - f\|^2_{L^2(\Omega)}$$

The minimizers of $\mathcal{L}[u]$ are precisely the solutions to the $\mathscr{T}$-transformed problem

$$\tilde{\mathcal{N}}[u] = \mathscr{EL}\left[\mathcal{N}[u]^2\right] + 2\lambda(u - f) = 0$$

on Ω, subject to the system of boundary conditions

$$\tilde{\mathcal{B}}^{(\gamma)}[u] = \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] = 0$$

on $\partial\Omega$ for every multi-subset γ of $[n]$ of size k, where $\mathscr{EL}^{(\gamma)}$ is defined in Definition E.1.

C. Proofs

Proof of Theorem B.1. Denote $\Sigma = \partial\Omega$, so that Ω is an orientable n-manifold and Σ is an $n-1$-manifold. Extend f to a smooth function on Ω (this can be done via the tubular neighborhood theorem and bump functions). Via this extension, $\omega = (u - f)^2 d\Sigma$ is an $n-1$-form on Ω. By Stokes' theorem,

$$\int_\Sigma \omega = \int_\Omega d\omega$$

First of all, note that G is a total derivative, where $d\omega = Gd\Omega$, so it is a null Lagrangian by Lemma 5 of Olver & Sivaloganathan (1988). This is for us to prove what $\tilde{\mathcal{N}}$ must be: note that then

$$\mathcal{L}[u] = \int_\Omega \mathcal{N}[u]^2 d\Omega + \lambda d\omega$$

and so the Euler-Lagrange equation reads

$$\mathscr{EL}\left[\mathcal{N}[u]^2 + \lambda G\right] = \mathscr{EL}\left[\mathcal{N}[u]^2\right] + \lambda\,\mathscr{EL}[G] = \mathscr{EL}\left[\mathcal{N}[u]^2\right]$$

by linearity of $\mathscr{EL}$ and the fact that G is a null Lagrangian. And so we have

$$\tilde{\mathcal{N}}[u] = \mathscr{EL}\left[\mathcal{N}[u]^2\right]$$

In order to compute $\tilde{\mathcal{B}}$, we must know the normal derivative of $d\omega$ at $\partial\Omega$. To do so, we change coordinate if necessary such that in a local chart Ω is given by the half-space $\{x \in \mathbb{R}^n : x_1 \geq 0\}$, and the boundary Σ is given by the plane $\{x \in \mathbb{R}^n : x_1 = 0\}$.

In this case, we have $\omega = (u - f)^2 dx_2 \wedge \cdots \wedge dx_n$, and so

$$d\omega = 2(u - f)(u_1 - f_1)dx_1 \wedge \cdots \wedge dx_n = Gd\Omega$$

where $G = 2(u-f)(u_1-f_1)$. Then we have $\frac{\partial G}{\partial u_1} = 2(u-f)$. Additionally, $\frac{\partial G}{\partial u_\alpha} = 0$ for any α other than $\{x_1\}$. Consequently, we have

$$\frac{\partial G}{\partial u_n} = 2(u - f)$$

where u_n denotes the partial derivative with respect to the component of u normal to Σ.

Observe from Definition E.1 that $\mathscr{EL}^{(\gamma)}$ depends on u_α for $|\alpha| = 1$ only when $\gamma = \varnothing$, and that $\mathscr{EL}^{(\gamma)}$ acts linearly for every γ. We therefore derive that

$$\mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2 + \lambda G\right] = \begin{cases} \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] & \gamma \neq \varnothing \\ \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] + 2\lambda(u - f) & \gamma = \varnothing \end{cases}$$

and thus that the $\mathscr{T}$-transformed problem is specified by the $\tilde{\mathcal{N}}[u] = 0$ and the system of boundary conditions

$$\tilde{\mathcal{B}}^{(\gamma)}[u] = \begin{cases} \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] & \gamma \neq \varnothing \\ \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] + 2\lambda(u - f) & \gamma = \varnothing \end{cases} = 0$$

for every multi-subset γ of $[n]$ of size k. □

Proof of Theorem B.2. We write

$$\|u - f\|^2_{L^2(\mathcal{D})} = \sum_{x \in \mathcal{D}} ((u-f)(x))^2 = \int_\Omega \sum_{x \in \mathcal{D}} (u-f)^2 \delta_x$$

The conclusion then follows from linearity of $\mathscr{EL}$ and $\mathscr{EL}^{(\gamma)}$ according to Definition E.1 and the generalizations of Theorem D.2 to generalized functions provided in (Konjik et al., 2008; Simmons, 2022; Derr & Kinzebulatov, 2007). □

Proof of Theorem B.3. We have

$$\mathcal{L}[u] = \int_\Omega \mathcal{N}[u]^2 + \lambda(u-f)^2 =: \int_\Omega F[u]$$

By linearity of $\mathscr{EL}$ and $\mathscr{EL}^{(\gamma)}$ and the fact that $(u - f)$ does not depend on u_α for any $\alpha \neq \varnothing$, we immediately have that u^* solves

$$\tilde{\mathcal{N}}[u] = \mathscr{EL}\left[\mathcal{N}[u]^2\right] + 2\lambda(u-f) = 0$$

on Ω, subject to the system of boundary conditions

$$\tilde{\mathcal{B}}^{(\gamma)}[u] = \mathscr{EL}^{(\gamma)}\left[\mathcal{N}[u]^2\right] = 0$$

on $\partial\Omega$ for every multi-subset γ of $[n]$ of size k. To show uniqueness, we appeal to Theorem D.2 and its generalizations in Dacorogna (2024) (see the exercises following Chapter 1), which state that every solution to $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}})$ is unique when the map $(u_\alpha)_{|\alpha| \leq k} \mapsto F[u]$ is strictly convex for all $x \in \Omega$. To check this, write

$$\mathcal{N}[u] = \sum_{|\alpha| \leq k} a(x;\alpha) u_\alpha$$

where $a(x;\alpha)$ denotes the coefficient function of u_α in the linear PDE $\mathcal{N}[u] = 0$. It follows that $\mathcal{N}[u]^2$ is convex as a function of $(u_\alpha)_{|\alpha| \leq k}$, while $\lambda(u-f)^2$ is convex; hence their sum is also convex as desired. □

Proof of Theorem 4.9. On the contrary, suppose every smooth u minimizes $\tilde{\mathcal{L}}[u]$. Since the original problem is well-specified, there exists a loss minimizer u^* such that $\mathcal{L}[u^*] = 0$, and hence $\tilde{\mathcal{L}}[u^*] = 0$. Since then $\mathscr{EL}[\mathcal{N}[u]^2] = 0$ for every u, by definition, $\mathcal{N}[u]^2$ is a null Lagrangian.

Define f to be the trace of u^*, i.e. its value on $\partial\Omega$. By Theorem 1 of Section 8.1.4. of Evans (2010), we know that every smooth u with trace f satisfies $\mathcal{L}[u] = \mathcal{L}[u^*] = 0$, so in particular $\mathcal{N}[u] = 0$ for all u with trace f. Now let U be an arbitrary proper open subset of Ω. For any smooth $\tilde{u}$ defined on U, we can smoothly extend $\tilde{u}$ to a function u on Ω with trace f (e.g. using bump functions). But then $\mathcal{N}[\tilde{u}] = 0$ on U for all smooth $\tilde{u}$, and so $\mathcal{N} \equiv 0$ on U. □

Proof of Theorem 4.7. First of all, if $\mathcal{N}[u^*] \neq 0$, it follows from continuity of $\mathcal{N}$ that $\|\mathcal{N}[u^*]\|_{L^2(\Omega)} > 0$, and so $\mathcal{L}[u] > 0$. Since u^* is the minimizer of $\mathcal{L}$, we have that $\mathcal{L}[u]$ is lower-bounded by a constant over all u. On the other hand, the fact that $\mathcal{L}$ has a minimum at u^* implies that $\tilde{\mathcal{L}}[u^*] = 0$ by construction. Thus, there can be no constant C' with the desired property. This suffices for one direction of the proof.

If $\mathcal{N}[u^*] = 0$, it follows that u^* is a global minimizer of $\|\mathcal{N}[u]\|^2_{L^2(\Omega)}$, and so we have $\mathscr{EL}\left[\mathcal{N}[u^*]\right] = 0$. Applying the first inequality of Assumption 4.6, we then have

$$\|\mathcal{N}[u]\|_{L^2(\Omega)} \leq \|\mathcal{N}[u]\|_{W^{1,2}(\Omega)} \leq C \left\|\mathscr{EL}\left[\mathcal{N}[u]^2\right]\right\|_{L^2(\Omega)}$$

for all u. Applying the second inequality of Assumption 4.6 and the trace theorem, we have

$$\left\|\frac{\partial \mathcal{N}[u]^2}{\partial u_n}\right\|_{L^2(\partial\Omega)} \leq D\|\mathcal{N}[u]\|_{L^2(\partial\Omega)} \leq c' D \|\mathcal{N}[u]\|_{W^{1,2}(\Omega)}$$

where c' is a constant depending only on Ω. Finally, we apply each of the above to conclude that

$$\begin{aligned}
\mathcal{L}[u] &= \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \lambda\|u-f\|^2_{L^2(\partial\Omega)} \\
&\le \|\mathcal{N}[u]\|^2_{W^{1,2}(\Omega)} + \lambda\|u-f\|^2_{L^2(\partial\Omega)} \\
&\le \sqrt{2}\|\mathcal{N}[u]\|_{W^{1,2}(\Omega)} + \lambda\|2(u-f)\|^2_{L^2(\partial\Omega)} \\
&\le (\sqrt{2}+\lambda c'^2 D^2)\|\mathcal{N}[u]\|^2_{W^{1,2}(\Omega)} + \lambda\|2(u-f)\|^2_{L^2(\partial\Omega)} - \lambda\left\|\frac{\partial\mathcal{N}[u]^2}{\partial u_n}\right\|^2_{L^2(\partial\Omega)} \\
&\le (\sqrt{2}+\lambda c'^2 D^2)\|\mathcal{N}[u]\|^2_{W^{1,2}(\Omega)} + \lambda\left\|2(u-f) - \frac{\partial\mathcal{N}[u]^2}{\partial u_n}\right\|^2_{L^2(\partial\Omega)} \\
&\le C^2(\sqrt{2}+\lambda c'^2 D^2)\left\|\mathscr{L}\left[\mathcal{N}[u]^2\right]\right\|^2_{L^2(\Omega)} + \lambda\left\|2(u-f) - \frac{\partial\mathcal{N}[u]^2}{\partial u_n}\right\|^2_{L^2(\partial\Omega)} \\
&\le c\,\tilde{\mathcal{L}}[u]
\end{aligned}$$

where, in order, we applied the inequality $\|\cdot\|_{L^2} \le \|\cdot\|_{W^{1,2}}$, multiplication by $\sqrt{2}$, adding the non-negative quantity

$$c'^2 D^2\|\mathcal{N}[u]\|^2_{W^{1,2}(\Omega)} - \left\|\frac{\partial\mathcal{N}[u]^2}{\partial u_n}\right\|^2_{L^2(\partial\Omega)} \ge 0$$

the reverse triangle inequality, the first inequality of Assumption 4.6, and lastly the definition

$$c := \max\left\{1, C^2(\sqrt{2}+\lambda c'^2 D^2)\right\}$$

□

Proof of Theorem 4.5. By the stability of $(\tilde{\mathcal{N}}, \tilde{\mathcal{B}})$, there exists a constant $C > 0$ such that

$$\|u-v\| \le C(\|\tilde{\mathcal{N}}[u]-\tilde{\mathcal{N}}[v]\| + \|\tilde{\mathcal{B}}[u]-\tilde{\mathcal{B}}[v]\|)$$

Since u^* minimizes $\mathcal{L}$, it also solves the $\mathscr{T}$-transformed problem, i.e. $\tilde{\mathcal{N}}[u^*], \tilde{\mathcal{B}} = 0$. We then have

$$\|u-u^*\| \le C(\|\mathcal{N}[u]\| + \|\mathcal{B}[u]\|)$$

and so

$$\|u-u^*\|^2 \le C^2(\|\mathcal{N}[u]\| + \|\mathcal{B}[u]\|)^2 \le 2C^2\left(\|\mathcal{N}[u]\|^2 + \|\mathcal{B}[u]\|^2\right) \le 2C^2\max\{1, 1/\lambda\}\tilde{\mathcal{L}}[u]$$

□

D. Background on PDEs and calculus of variations

D.1. Partial differential equations

Partial differential equations (PDEs) are ubiquitous in science and engineering, and are widely used to model physical phenomena. In the most general setup, $\Omega \subset \mathbb{R}^n$ denotes a domain of interest (e.g. a subset of space-time), and we seek a smooth function

$$u : \Omega \longrightarrow \mathbb{R}^m$$

such that a PDE

$$\mathcal{N}[u] = 0 \tag{9}$$

is satisfied at all points $x \in \Omega$, where $\mathcal{N}[u]$ is a differential operator[8] applied to u. To fix notation, if u is a function of several variables—say, x and t—we will let u_x and u_t denote the respective partial derivatives. More generally, to differentiate multiple times—say, twice with respect to x and once with respect to t—we will denote u_{xxt}. Most generally, we will denote

[8] E.g. $\frac{\partial}{\partial t} + \frac{\partial^2}{\partial x^2}$

u_α to be the partial derivative with respect to an arbitrary multiset of variables: in the case above, α was the string xxt. We will let $|\alpha|$ denote the total number of derivatives taken, which in this case is 3. Lastly, we will denote by ∂_α the operation of partial differentiation with respect to the collection of variables specified by α. With this said, a differential operator $\mathcal{N}$ of order k is a functional which accepts a $\mathcal{C}^k$ function u and a point $x \in \Omega$ as input, and outputs a value depending only on x and the derivatives of u up to order k, i.e. $\{u_\alpha\}_{|\alpha|\leq k}$.

As currently stated, Equation 9 tends to have many solutions; for this reason, u is often additionally constrained by a requirement specifying its value to be f on a subset $\mathcal{D} \subset \Omega$ of the domain:

$$u|_{\mathcal{D}} = f|_{\mathcal{D}}$$

This constraint is the general methodology by which data is incorporated into PDE-based modeling problems: $\mathcal{D}$ represents the subdomain on which data is available, and f represents the observed data.

In the general case, exactly what data should be specified in order for u to be uniquely specified is a major open question in mathematics: for example, when $\mathcal{N} = \nabla^2$, it is known that $\mathcal{D}$ should be $\partial\Omega$, and when $\mathcal{N}[u] = 0$ denotes the Navier-Stokes equations, it is an open question what $\mathcal{D}$ and f must be.

D.2. Integration of data with PDEs

While well-posedness in the general case remains an open question inside mathematics, the issue becomes further complicated by the specifics of the setup which practitioners may be confronted with. For instance, practitioners may have access to a set of data (or otherwise) constraints $\mathcal{D}$ beyond what would be required to uniquely specify a solution to the underlying problem. Or, practitioners may hope to encode situation-specific terms into their PDEs which do appear in the versions which are well studied in the mathematical literature.

One hope of PINNs is that they will be capable of addressing such integration issues in the general case. But if this hope is to be realized, PINNs must be able to handle *ill-posed* as well as *well-posed* cases.

D.3. Physics-informed neural networks

PINNs are neural networks which optimize an objective intended to enforce **??**. Specifically, PINNs minimize a loss function which is the Lagrangian relaxation

$$\mathcal{L}[u] = \|\mathcal{N}[u]\|^2_{L^2(\Omega)} + \lambda\|u - f\|^2_{L^2(\mathcal{D})} \tag{10}$$

where $\lambda > 0$ is a hyperparameter.

The PINN loss is convenient in part because it is able to handle mis-specified problems: in the mis-specified case, one may still solve the optimization problem $\min_u \mathcal{L}[u] > 0$, provided that the minimum exists.

D.4. The calculus of variations

The calculus of variations is a field of analysis used to find minima u of functionals of the form

$$\mathcal{I}[u] = \int_\Omega F(x, \{u_\alpha\}_{|\alpha|\leq k})dx$$

By analogy to the first derivative test in calculus, a necessary condition for u to minimize $\mathcal{I}[u]$ is that $\mathcal{I}'[u] = 0$, where $\mathcal{I}'[u]$ is some suitable notion of a "derivative" of $\mathcal{I}$ with respect to u. This notion is made precise by the following definition: (E-O) (E-L, 2024)

Definition D.1 (Euler operator)**.** The *Euler operator* is the operator which accepts a function $F(x, \{u_\alpha\}_{|\alpha|\leq k})$ of x and the derivatives of u and acts according to

$$\mathscr{EL} = \sum_{|\alpha|\leq k} (-1)^{|\alpha|}\partial_\alpha \frac{\partial}{\partial u_\alpha}$$

Concretely, when $k = 1$ and thus $F = F(x, u, \nabla u)$, we have

$$\mathscr{EL}[F](x) = \left[\frac{\partial}{\partial u} - \nabla \cdot \frac{\partial}{\partial \nabla u}\right] F(x, u, \nabla u)$$

$\mathscr{EL}$ generalizes the notion of a derivative with respect to u. With this said, the intuition from calculus that every minimizer of u should satisfy "$\mathcal{I}'[u] = 0$" will be replaced with the condition that $\mathscr{EL}[F[u]] = 0$.

The following theorem makes this precise by providing a set of necessary and sufficient conditions for u to minimize $\mathcal{I}[u]$. For simplicity, we state the theorem only for $k = 1$, and include the more general version in the appendix.

Theorem D.2 (Theorem 2.1 of (Dacorogna, 2024))**.** *Let $F(x, u, \nabla u)$ be twice-continuously differentiable on $\Omega \times \mathbb{R} \times \mathbb{R}^n$, and consider the minimization problem*

$$\inf_{u \in \mathcal{C}^2(\Omega, \mathbb{R})} \mathcal{I}[u] \tag{11}$$

If Equation 11 has a minimizer $u^* \in \mathcal{C}^2(\Omega, \mathbb{R})$ then u^* satisfies the Euler-Lagrange equation

$$\mathscr{EL}[F[u^*]] = 0 \tag{12}$$

on the domain Ω, subject to the boundary conditions

$$\frac{\partial F}{\partial \nabla u} \cdot n = 0$$

on the boundary $\partial\Omega$, where n denotes a vector normal to the boundary. Conversely, if u^* satisfies Equation 12 and the map $(\alpha, \beta) \mapsto F(x, \alpha, \beta)$ is convex for all $x \in \Omega$, then u^* is a minimizer of Equation 11 Moreover, if the map is instead strictly convex, then u^* is the unique minimizer of Equation 11.

E. Statement and derivation of Euler-Lagrange equations for higher derivatives

The authors were unable to find statements of the Euler-Lagrange equations with natural boundary conditions in full generality in the literature. For this reason, we derive the equations and appropriate boundary conditions here, remarking that they are references without statement in texts such as Dacorogna (2024); Evans (2010).

E.1. Statement

Let $\Omega \subset \mathbb{R}^n$ be an oriented $n-$manifold with boundary $\partial\Omega$, and let $\hat{n}$ denote a vector normal to $\partial\Omega$. Let F be a differential operator of order k, and u a smooth function on Ω.

Definition E.1 (The general Euler operator and natural boundary conditions)**.** The Euler-operator $\mathscr{EL}$ is defined by the Euler-Lagrange equation

$$\mathscr{EL}[F] = \sum_{|\alpha| \leq k} (-1)^{|\alpha|} \partial_\alpha F_{u_\alpha}$$

and the natural boundary conditions are defined by the set of equations

$$\mathscr{EL}^{(\gamma)}[F] = \sum_{\mu \in [n]} \hat{n}_\mu \sum_{|\alpha| + |\gamma| + 1 \leq k} (-1)^{|\alpha|} \partial_\alpha F_{u_{\alpha\gamma\mu}}$$

for every γ a multi-subset of $[n]$ of size k.

E.2. Derivation

Let

$$\mathcal{I}[u] = \int_\Omega F\left(x, \{u_\alpha\}_{|\alpha| \leq k}\right)$$

where Ω is a domain with smooth boundary $\partial\Omega$, F is a smooth function of x and the derivatives of u up to order k, for some $k \geq 0$. We recall the definition of the *first-variation* of $\mathcal{I}$, as it appears in Dacorogna (2024):

$$\delta\mathcal{I}[u] = \int_\Omega \sum_{|\alpha| \leq k} F_{u_\alpha} \delta u_\alpha$$

By construction, $\delta\mathcal{I}[u^*] = 0$ for all smooth δu whenever u^* extremizes $\mathcal{I}[u]$. We will now derive the Euler-Lagrange equations and the appropriate boundary conditions from the statement that $\delta\mathcal{I}[u] = 0$.

We will generally use $\mu_1, \mu_2, \cdots$ to denote dummy indices, and employ the Einstein summation convention. Let us define recursively the k vector fields

$$G^{(k)}_{\mu_1 \cdots \mu_k} = F_{u_{\mu_1 \cdots \mu_k}} \qquad G^{(i)}_{\mu_1 \cdots \mu_i} = F_{u_{\mu_1 \cdots \mu_i}} - \partial_{\mu_{i+1}} G^{(i+1)}_{\mu_1 \cdots \mu_{i+1}} \quad \forall 0 \leq i < k$$

Claim: we have $\delta \mathcal{I}[u^*]$ for all smooth δu if and only if u^* solves the system

$$\begin{aligned} G^{(0)} &= 0 \quad \text{on } \Omega \\ G^{(i)}_{\mu_1 \cdots \mu_i} n_{\mu_i} &= 0 \quad \text{on } \partial\Omega, \quad \forall \mu_1 \cdots \mu_{i-1}, \quad \forall 0 < i \leq k \end{aligned}$$

where n is a vector normal to $\partial\Omega$.

Proof of claim: Recalling Green's first identity, we have for any smooth vector field $G : \Omega \longrightarrow \mathbb{R}^{i+1}$ that

$$\int_\Omega G_{\mu_1 \cdots \mu_{i+1}} \delta u_{\mu_1 \cdots \mu_{i+1}} = \int_{\partial\Omega} G_{\mu_1 \cdots \mu_{i+1}} n_{\mu_{i+1}} \delta u_{\mu_1 \cdots \mu_i} - \int_\Omega \partial_{\mu_{i+1}} G_{\mu_1 \cdots \mu_{i+1}} \delta u_{\mu_1 \cdots \mu_i}$$

We will assume the inductive hypothesis that

$$\int_\Omega \sum_{i=\ell}^{k} F_{u_{\mu_1 \cdots \mu_i}} \delta u_{\mu_1 \cdots \mu_i} = - \int_\Omega G^{(\ell)}_{\mu_1 \cdots \mu_\ell} \delta u_{\mu_1 \cdots \mu_\ell} + \sum_{i=\ell}^{k} (-1)^{i-\ell} \int_{\partial\Omega} G^{(i)}_{\mu_1 \cdots \mu_i} n_{\mu_i} \delta u_{\mu_1 \cdots \mu_i}$$

Note that the left hand side is equal to $\delta \mathcal{I}[u]$ when $\ell = 0$. On the other hand, when $\ell = k$, the equality is trivially true by the definition of $G^{(k)}$.

We will proceed by assuming the inductive hypothesis for ℓ, and prove it for $\ell - 1$. We have

$$\begin{aligned} \int_\Omega \sum_{i=\ell-1}^{k} F_{u_{\mu_1 \cdots \mu_i}} \delta u_{\mu_1 \cdots \mu_i} &= \int_\Omega \sum_{i=\ell}^{k} F_{u_{\mu_1 \cdots \mu_i}} \delta u_{\mu_1 \cdots \mu_i} - \int_\Omega F_{u_{\mu_1 \cdots \mu_{\ell-1}}} \delta u_{\mu_1 \cdots \mu_{\ell-1}} \\ &= - \int_\Omega G^{(\ell)}_{\mu_1 \cdots \mu_\ell} \delta u_{\mu_1 \cdots \mu_\ell} - \int_\Omega F_{u_{\mu_1 \cdots \mu_{\ell-1}}} \delta u_{\mu_1 \cdots \mu_{\ell-1}} \\ &\quad + \sum_{i=\ell}^{k} (-1)^{i-\ell} \int_{\partial\Omega} G^{(i)}_{\mu_1 \cdots \mu_i} n_{\mu_i} \delta u_{\mu_1 \cdots \mu_i} \\ &= - \int_{\partial\Omega} G^{(\ell)}_{\mu_1 \cdots \mu_\ell} n_{\mu_\ell} \delta u_{\mu_1 \cdots \mu_{\ell-1}} + \int_\Omega \partial_{\mu_\ell} G^{(\ell)}_{\mu_1 \cdots \mu_\ell} \delta u_{\mu_1 \cdots \mu_{\ell-1}} \\ &\quad - \int_\Omega F_{u_{\mu_1 \cdots \mu_{\ell-1}}} \delta u_{\mu_1 \cdots \mu_{\ell-1}} + \sum_{i=\ell}^{k} (-1)^{i-\ell} \int_{\partial\Omega} G^{(i)}_{\mu_1 \cdots \mu_i} n_{\mu_i} \delta u_{\mu_1 \cdots \mu_i} \\ &= - \int_\Omega G^{(\ell-1)}_{\mu_1 \cdots \mu_{\ell-1}} \delta u_{\mu_1 \cdots \mu_{\ell-1}} + \sum_{i=\ell-1}^{k} (-1)^{i-\ell} \int_{\partial\Omega} G^{i}_{\mu_1 \cdots \mu_i} n_{\mu_i} \delta u_{\mu_1 \cdots \mu_i} \end{aligned}$$

as desired.

With this claim proven, consider the case $\ell = 0$. We then have

$$\delta \mathcal{I}[u] = - \int_\Omega G^{(0)} \delta u + \sum_{i=0}^{k} (-1)^i \int_{\partial\Omega} G^{(i)}_{\mu_1 \cdots \mu_i} n_{\mu_i} \delta u_{\mu_1 \cdots \mu_i}$$

Thus, we have $\delta \mathcal{I}[u^*] = 0$ for all smooth δu if and only if u^* solves the system

$$\begin{aligned} G^{(0)} &= 0 \quad \text{on } \Omega \\ G^{(i)}_{\mu_1 \cdots \mu_i} n_{\mu_i} &= 0 \quad \text{on } \partial\Omega, \quad \forall \mu_1 \cdots \mu_{i-1}, \quad \forall 0 < i \leq k \end{aligned}$$

as desired. This final implication requires an argument which is more technical than is appropriate here, so we defer the reader to the proofs of either Theorem 2.1 of Dacorogna (2024) or Chapter 8 of Evans (2010). Intuitively, one can prove that $G^{(0)}$ must vanish on any arbitary open subset U of Ω by considering δu to be a bump function on U; this will not change the

values of δu and its derivatives at the boundary $\partial\Omega$. Similarly, one can adjust the derivatives of δu at the boundary arbitrarily so conclude that the specified boundary conditions must hold.

Finally, we specify the form of the general Euler operator and the natural boundary conditions. The quantity $G^{(0)}$ derived in the previous section is exactly the action of the Euler operator on $F[u]$. In general, we have

$$G^{(0)} = \sum_{i=0}^{k} (-1)^i \partial_{\mu_1 \cdots \mu_i} F_{u_{\mu_1 \cdots \mu_i}}$$
$$G^{(\ell)}_{\nu_1 \cdots \nu_\ell} = \sum_{i=0}^{k-\ell} (-1)^i \partial_{\mu_1 \cdots \mu_i} F_{u_{\mu_1 \cdots \mu_i \nu_1 \cdots \nu_\ell}}$$

i.e.

$$G^{(0)} = \sum_{|\alpha| \le k} (-1)^{|\alpha|} \partial_\alpha F_{u_\alpha}$$
$$G^{(|\beta|)}_{\beta} = \sum_{|\alpha|+|\beta| \le k} (-1)^{|\alpha|} \partial_\alpha F_{u_{\alpha\beta}}$$

Finally, we define

Fact E.2 (The general Euler operator and natural boundary conditions)**.**

$$\mathscr{EL}[F] = G^{(0)}$$
$$\mathscr{EL}^{(\gamma)}[F] = G^{(|\gamma|+1)}_{\gamma\mu} n_\mu$$

The first line is the Euler-Lagrange equation, and the second line represents the natural boundary conditions, of which there are as many as multi-subsets γ of $[n]$ of size at most $k-1$. For instance, when $n = 1$, there are k boundary conditions.

F. Exact solution to the ODE example with interior data

Consider the equation

$$\ddot{u} - u = (u - y) \frac{1-\gamma}{\gamma m} \sum_{i=1}^{m} \delta_{x_i^d} \tag{13}$$

and recall that the solutions to the equation $\ddot{u} - u = 0$ are

$$u(x) = c_1 e^x + c_2 e^{-x} \tag{14}$$

for some $c_1, c_2 \in \mathbb{R}$.

Now we seek to solve Equation 13. Note that except at the finitely many datapoints $\{x_i^d\}_{i \le m}$, the equation is equivalent to $\ddot{u} = u$. Therefore, on the intervals between datapoints, $u(x)$ takes the form of Equation 14, where the constants c_1, c_2 may be different on each interval.

How can we determine these constants? Notice that in Equation 13, $\ddot{u}$ may blow up at the datapoints x_i^d. This is because the defining property of $\delta_{x_i^d}$ is that it is infinite at x_i^d. However, it will still integrate to a finite quantity. This means that we can expect $\dot{u}$ to be finite everywhere, although it might unfortunately be discontinuous (because it has an infinite rate of change at the points x_i^d).

Here's the good news: the fact that $\dot{u}$ is finite means that it will integrate to a continuous function, and so we expect u to be continuous, although maybe not $\mathcal{C}^1$. So what does u look like? Piece-wise on each interval between the datapoints, it will take the form of Equation 14, and at the datapoints it will have a corner such that Equation 13 is satisfied.

To determine the shape of these corners, we may extract some information from Equation 13 in the following way: Pick a datapoint $x_0 \in \{x_i^d\}_{i \le m}$, and fixing some small $\epsilon > 0$, integrate Equation 13 to achieve

$$\int_{x_0-\epsilon}^{x_0+\epsilon} \ddot{u} = \int_{x_0-\epsilon}^{x_0+\epsilon} u + (u - y) \frac{1-\gamma}{\gamma m} \sum_{i=1}^{m} \delta_{x_i^d}$$

We now compute this integral term-wise. Applying the fundamental theorem of calculus, the left hand side becomes

$$\dot{u}(x_0+\epsilon)-\dot{u}(x_0-\epsilon)$$

Asserting the existence of a continuous solution u, we know that

$$\lim_{\epsilon\to 0}\int_{x_0-\epsilon}^{x_0+\epsilon} u = 0$$

When ϵ is small enough that the interval $(x_0-\epsilon, x_0+\epsilon)$ does not contain any datapoints $\{x_i^d\}_{i\le m}$ other than x_0, we have[9]

$$\int_{x_0-\epsilon}^{x_0+\epsilon}(u-y)\frac{1-\gamma}{\gamma m}\sum_{i=1}^{m}\delta_{x_i^d} = (u(x_0)-y(x_0))\frac{1-\gamma}{\gamma m}$$

Taking the limit as $\epsilon\to 0$, we conclude that

$$\lim_{\epsilon\to 0}\dot{u}(x_0+\epsilon)-\dot{u}(x_0-\epsilon) = (u(x_0)-y(x_0))\frac{1-\gamma}{\gamma m} \tag{15}$$

This formula contains many interested facts. First, it demonstrates that **the slope of the corner at each datapoint is directly proportional to the error in the prediction of that datapoint** (i.e. the quantity $u(x_0)-y(x_0)$). The prediction will contain a corner at a datapoint if and only if it does not fit it precisely.

Moreover, because the change in slope has the same sign as $u(x_0)-y(x_0)$, the formula shows that **at each datapoint x_0, the prediction bends away from the given datapoint** $(x_0, f(x_0))$ (to see this, note that if $u(x_0) > f(x_0)$, then the left-hand side shows that the change in slope of u at x_0 is positive). Lastly, notice that the change in slope decreases with γ, and is inversely proportional to m. This means that the development of corners becomes less pronounced as the equation loss is weighted more heavily, and disappears as $m\to\infty$. It is perhaps for this reason that these corners are not observed in large-scale PINN experiments with many datapoints.

Returning to solving Equation 13, recall that the optimal u will take the form of Equation 14 on each of the intervals between the datapoints. There are $N+1$ such intervals, where $N\le m$ is the number of distinct interior datapoints among $\{x_i^d\}_{i\le m}$. Since Equation 14 is determined by the two constants c_1, c_2 which are fixed on each interval, there are $2(N+1)$ total degrees of freedom for u of this form.

At each of the N distinct interior datapoints, continuity of u enforces one constraint, and the corner equation (Equation 15) describing the jump in $\dot{u}$ enforces another. Thus we have $2N$ total constraints, and $2(N+1)$ degrees of freedom, meaning that a solution to Equation 13 is uniquely determined by two constants α, β. This is consistent with the fact that Equation 13 is a second-order linear ODE, albeit one containing Dirac delta functions.

In order to minimize the loss, we therefore fix some constants α, β for which Equation 13 has a unique solution, and then optimize the loss over all choices of α, β.

Returning to the problem at hand:

In our case, we consider only a single interior datapoint $\left(\frac{1}{2}, 1\right)$. This interior datapoint partitions the domain into two halves, so that

$$u(x) = \begin{cases} c_1^{(1)}e^x + c_2^{(1)}e^{-x} & x\in\left[0,\frac{1}{2}\right] \\ c_1^{(2)}e^x + c_2^{(2)}e^{-x} & x\in\left(\frac{1}{2},1\right]\end{cases} \tag{16}$$

where at $x=\frac{1}{2}$ the constants satisfy the continuity constraint

$$c_1^{(1)}e^{1/2} + c_2^{(1)}e^{-1/2} = c_1^{(2)}e^{1/2} + c_2^{(2)}e^{-1/2}$$

and the corner constraint (see Equation 15)

$$\left\{c_1^{(1)}e^{1/2} - c_2^{(1)}e^{-1/2}\right\} - \left\{c_1^{(2)}e^{1/2} - c_2^{(2)}e^{-1/2}\right\} = \left(c_1^{(1)}e^{1/2} + c_2^{(1)}e^{-1/2} - 1\right)\frac{1-\gamma}{3\gamma}$$

[9] Note that if the datapoint x_0 is repeated among $\{x_i^d\}_{i\le m}$ then this is not possible. But in this case, the integral formula below will just multiply by however many distinct datapoints are equal to x_0.

These two constraints form a linear system

$$\begin{pmatrix} e^{1/2} & e^{-1/2} & e^{1/2} & e^{-1/2} \\ e^{1/2}\left(1-\frac{1-\gamma}{3\gamma}\right) & -e^{1/2}\left(1+\frac{1-\gamma}{3\gamma}\right) & -e^{1/2} & e^{-1/2} \end{pmatrix} \begin{pmatrix} c_1^{(1)} \\ c_2^{(1)} \\ c_1^{(2)} \\ c_2^{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{\gamma-1}{3\gamma} \end{pmatrix} \tag{17}$$

the solutions to which form a 2-dimensional vector space which we will parametrize by the constants α, β.

Now recall that the ideal loss is

$$\begin{aligned}
(1-\gamma)\mathcal{L}_{data}^{1,1,1} + \gamma\mathcal{L}_{eqn} &= (1-\gamma)\left\{\frac{1}{3}(u(0)-1)^2 + \frac{1}{3}(u(1/2)-1)^2\frac{1}{3}(u(1)-1)^2\right\} \\
&\quad + \gamma\int_0^1 (u-\dot{u})^2 dx + \\
&= (1-\gamma)\left\{\frac{1}{3}(u(0)-1)^2 + \frac{1}{3}(u(1/2)-1)^2\frac{1}{3}(u(1)-1)^2\right\} \\
&\quad + \gamma\int_0^{1/2} (2c_2^{(1)}e^{-x})^2 dx + \gamma\int_{1/2}^1 (2c_2^{(2)}e^{-x})^2 \\
&= (1-\gamma)\left\{\frac{1}{3}(c_1^{(1)}+c_2^{(1)}-1)^2 + \frac{1}{3}(c_1^{(1)}e^{1/2}+c_2^{(1)}e^{-1/2}-1)^2\frac{1}{3}(c_1^{(2)}e + c_2^{(2)}/e-1)^2\right\} \\
&\quad + 2\gamma\frac{e-1}{e}{c_2^{(1)}}^2 + 2\gamma\frac{e-1}{e^2}{c_2^{(2)}}^2
\end{aligned}$$

which is again a quadratic form in the constants α, β which parametrize the solutions to Equation 17. Using the computer algebra system SymPy, by optimizing over all α, β, we can compute the precise minimum ideal loss to be 0.0836. We remark that this would also be possible to do using a more traditional method such as Lagrange multipliers.