Notes on Exact Boundary Values in Residual Minimisation

Johannes Müller   JMUELLER@MIS.MPG.DE
Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

Marius Zeinhofer   MARIUSZ@SIMULA.NO
Simula Research Laboratory, Oslo, Norway

Editors: Bin Dong, Qianxiao Li, Lei Wang, Zhi-Qin John Xu

Abstract

We analyse the difference in convergence mode using exact versus penalised boundary values for the residual minimisation of PDEs with neural network type ansatz functions, as is commonly done in the context of physics informed neural networks. It is known that using an L^2 boundary penalty leads to a loss of regularity of $3/2$ meaning that approximation in H^2 yields a posteriori estimates in $H^{1/2}$. These notes demonstrate how this loss of regularity can be circumvented if the functions in the ansatz class satisfy the boundary values exactly. Furthermore, it is shown that in this case, the loss function provides a consistent a posteriori error estimator in H^2 norm made by the residual minimisation method. We provide analogue results for linear time dependent problems and discuss the implications of measuring the residual in Sobolev norms.

Keywords: Residual minimisation, Physics informed neural networks, Neural networks

1. Introduction

After their striking success in supervised learning tasks, neural network based methods have recently gained more attraction for problems from numerical analysis. Neural network based approaches for the approximate solutions of PDEs can be traced back to Dissanayake and Phan-Thien (1994) where they proposed to combine residual minimisation with a boundary penalty term in order to train the parameters of a neural network. This ansatz was recently revived by Sirignano and Spiliopoulos (2018), embracing increased computational power due to the heavy usage of GPUs and has received a growing amount of attention since then, see E and Yu (2018); Raissi and Karniadakis (2018) and subsequent work.

Due to the unconstrained nature of neural networks, resolving boundary conditions for PDEs is challenging as noted by E and Yu (2018); van der Meer et al. (2020); Cyr et al. (2020). The work of Lagaris et al. (1998) proposes to encode boundary conditions directly into the trial functions. This approach was refined and extended by Berg and Nyström (2018); Lyu et al. (2020) and it was observed that such an ansatz frequently helps in producing more accurate solutions.

As our main contribution we show that for residual minimisation H^2 approximation yields estimates in H^2 for ansatz classes with exact boundary values, see Theorem 5. In contrast, using an L^2 penalty on the boundary values leads to error estimates in norms not stronger than $H^{1/2}$. The extension to more general equations, including time dependant ones, is discussed as well.

1.1. Physics informed neural networks

Suppose, we want to approximately solve the Dirichlet problem

$$
\begin{aligned}
-\Delta u &= f \quad \text{in } \Omega \\
u &= g \quad \text{on } \partial\Omega,
\end{aligned}
\tag{1}
$$

where $f \in L^2(\Omega)$ and $g \in H^{3/2}(\partial\Omega)$. We assume that this problem is well posed and denote its solution by $u_f \in H^2(\Omega)$, compare also to Remark 2. There exist different approaches to this problem, which usually formulate the Dirichlet problem (1) as a minimisation problem over a function space and then minimise this energy over a suitable ansatz class. One of those approaches was proposed by Lagaris et al. (1998) is the method of *residual minimisation*. It has received growing attention in the recent advancement of neural network based methods where it is often referred to as *physics informed neural networks* or PINNs following the work of Raissi et al. (2019). The idea of residual minimisation is to perceive the Dirichlet problem as the problem of minimising the *residual energy*

$$
\mathcal{L} \colon \Theta \to \mathbb{R}, \quad \theta \mapsto \|\Delta u_\theta + f\|_{L^2(\Omega)}^2 + \tau \|u_\theta - g\|_{L^2(\partial\Omega)}^2
$$

for some $\tau \in (0, \infty)$.

As $\mathcal{L}(\theta)$ measures the deviation of $-\Delta u_\theta$ of the right hand side f and of the boundary values, it is a natural approach to minimise $\mathcal{L}$ over the parameter space Θ. In order to give a theoretical justification for this approach, one can exploit elliptic regularity theory in order to show the a posteriori estimate

$$
\|u_\theta - u_f\|_{H^s(\Omega)}^2 \le c \cdot \mathcal{L}(\theta).
\tag{2}
$$

If such an estimate holds, then successful minimisation of $\mathcal{L}$ implies approximation of the solution u_f in some norm. Note that an estimate of the form (2) provides an a posteriori error estimate which is readily accessible throughout the training process. In particular, if $\theta^* \in \Theta$ minimises $\mathcal{L}$ over Θ, then

$$
\|u_{\theta^*} - u_f\|_{H^s(\Omega)}^2 \le c \cdot \inf_{\theta \in \Theta} \mathcal{L}(\theta) \le \tilde{c} \cdot \inf_{\theta \in \Theta} \|u_\theta - u_f\|_{H^2(\Omega)}^2.
\tag{3}
$$

In particular, approximation capabilities in H^2 in combination with successful training imply error estimates in H^s. This directly yields that only $s \le 2$ are potentially achievable in the a posteriori estimate (2). For $s \le 1/2$ an according estimate has been established by Shin et al. (2020), i.e., H^2 approximation and successful training implies convergence in $H^{1/2}$. In this sense, the approach of residual minimisation leads to a loss in regularity of $3/2$. It is the purpose of these notes to show that it is not possible to obtain error estimates for $s > 1/2$ in general and that for function classes with exact boundary values, i.e. $u_\theta \in H^2(\Omega) \cap H^1_g(\Omega)$[1] the estimates can be improved to hold for all $s \le 2$. Hence, in this case the loss of regularity can be mitigated which shows a theoretical advantage of ansatz classes which exactly satisfy the boundary conditions. Our proofs rely on curvature based estimates. We discuss how measuring the residual in Sobolev norms implies estimates in higher order Sobolev norms and generalise our findings to a class of parabolic evolution equations.

In both the stationary and instationary case our estimates are with respect to stronger norms than for the case of penalised boundary values , see Shin et al. (2020); Mishra and Molinaro (2022). More precisely, Theorem 5 shows that H^2 approximation yields estimates in H^2 for exact boundary

---

1. here, $H^1_g(\Omega)$ denotes the affine subspace of $H^1(\Omega)$ of functions that agree with g on $\partial\Omega$

values, where it is only able to provide $H^{1/2}$ estimates for penalised boundary values. Where we work with the population loss, those works also include the quadrature of the appearing integrals. This routine work can be transferred to our analysis without problems.

Exact Dirichlet Boundary Conditions for Neural Networks Standard neural network architectures are usually designed for unconstrained optimization, e.g. it is hardly possible to encode boundary values into a standard neural network architecture. However, following Lagaris et al. (1998), we can transform any unconstrained neural network architecture into an ansatz space with the desired boundary conditions. Assume we want to solve the Poisson problem (1) on Ω with boundary values g. We then construct a smooth function $L\colon \Omega \to [0, \infty)$ that satisfies $L|_{\partial\Omega} = 0$ and $L|_{\Omega} \neq 0$. The function L is often referred to as a smooth approximation of the distance function to $\partial\Omega$. Furthermore, we denote by $G \in H^2(\Omega)$ a lift of the boundary conditions to all of Ω, i.e. G satisfies $G|_{\partial\Omega} = g$. For any neural network family $\{u_\theta \mid \theta \in \Theta\} \subseteq H^2(\Omega)$ we then consider the associated family

$$\{L \cdot u_\theta + G \mid \theta \in \Theta\} \subseteq H^2(\Omega) \cap H^1_g(\Omega) \tag{4}$$

and use these functions to approximate the solution of (1). For complex domains it is difficult to obtain L and G analytically and thus the approximation via neural networks was proposed by Berg and Nyström (2018). For time dependent problems, a similar construction to (4) using a smoothed distance function to the parabolic boundary of the space-time domain can be used. We refer the reader to Lyu et al. (2020) for an explicit example.

Using ansatz functions ~~of the above form~~ with exact boundary values has become increasingly popular as it has been observed to simplify the training process and produce more accurate solutions, see for instance Berg and Nyström (2018); Roth et al. (2021); Lyu et al. (2020); Chen et al. (2020). The works of Chen et al. (2020); Courte and Zeinhofer (2021) explicitly compare penalized boundary conditions to exactly enforced ones in numerical studies and found improved accuracy and a faster training process. This is in accordance with Krishnapriyan et al. (2021) that illustrates the difficulties in the training process stemming from soft penalties in residual minimisation. It is also possible to encode Neumann or Robin boundary conditions in a similar way, we refer the reader to Lyu et al. (2020). However, we mention that the approximation capabilities of such ansatz classes have not been studied so far.

1.2. Organisation

We state our main results in Section 2, where we begin with the improved mode of convergence in the presence of exact boundary conditions in Section 2.1. We discuss the failure of H^2 estimates with inexact boundary conditions in Section 2.2 and the implication of stronger penalisation of the residual in Section 2.3. We generalise our results to parabolic equations in Section 3.

1.3. Notation

On an open subset Ω of $\mathbb{R}^d$ with boundary $\partial\Omega$, we denote the Lebesgue spaces with integrability order $p \in [1, \infty]$ by $L^p(\Omega)$ and $L^p(\partial\Omega)$ respectively. The Sobolev spaces with order of integrability $p \in [1, \infty]$ and smoothness parameter $s \in \mathbb{R}$ are denoted by $W^{s,p}(\Omega)$ and $W^{s,p}(\partial\Omega)$. If $p = 2$ we instead write $H^s(\Omega)$ and $H^s(\partial\Omega)$. The Sobolev space with zero boundary values in the trace sense is denoted by $W_0^{1,p}(\Omega)$ or $H_0^1(\Omega)$ respectively. The reader is referred to Grisvard (2011) for the precise definitions and more information on Sobolev spaces. Furthermore, for a Banach space

V, we denote by $L^2(I, V)$ the Bochner space of Bochner measurable, square integrable functions defined on the time interval $I = [0, T]$ taking values in V. For the Sobolev space modelled on $L^2(I, V)$ we write $H^1(I, V)$ and refer the reader to Boyer and Fabrie (2012).

For multivariate, scalar valued functions, we use the symbol Δ for the Laplace operator, ∇ for the gradient and D^2 for the second derivative, i.e. the Hessian. By d_t we denote a time derivative, usually in the context of a vector-valued Sobolev space.

For the purpose of this work, it suffices to view neural networks as parametric families of functions. More precisely, let $\Theta \subseteq \mathbb{R}^N$ be a parameter set and V a (subset of a) function space. Suppose there is a map $\Theta \to V$ denoted by $\theta \mapsto u_\theta$. We then say u_θ is the (neural network) function represented by the parameters θ. Note that the usual neural network architectures fall in this category, however, this setting also includes general parametric function classes. We stick to the terminology of neural networks since their recent application is the main motivation of this work.

2. Implications of Exact Boundary Values in Residual Minimisation

In this section we show the theoretical benefits of using neural network type ansatz functions that satisfy Dirichlet boundary conditions exactly in the residual minimisation method for the Poisson problem. We see that the exact boundary conditions improve the mode of convergence from $H^{1/2}$ to H^2. Although being formulated for the Laplace operator, those results hold for any elliptic operator which is H^2 regular.

2.1. H^2 estimates for Residual Minimisation with Exact Boundary Values

We start by considering the case of exact boundary conditions and present two main results, one that allows to quantify the H^2 error using the value of the loss function and the other, an estimate based on Céa's Lemma that allows to link the approximation capabilities of the network class to the error made by residual minimisation.

Setting 1 *We consider again* (1)*, in particular, we assume that the problem is H^2 regular meaning that there is a constant $C_{reg} > 0$, satisfying*

$$\|u\|_{H^2(\Omega)} \leq C_{reg} \|\Delta u\|_{L^2(\Omega)} \quad \text{for all } u \in H^2(\Omega) \cap H^1_0(\Omega).$$

Furthermore, we assume that Θ is a parameter set of a neural network type ansatz class, such that for every $\theta \in \Theta$ we have $u_\theta \in H^2(\Omega)$ and $(u_\theta)|_{\partial\Omega} = g$. As our strategy is to minimise the residual we define the loss function

$$\mathcal{L} \colon \Theta \to \mathbb{R}, \quad \mathcal{L}(\theta) = \|\Delta u_\theta + f\|^2_{L^2(\Omega)}.$$

Remark 2 *Setting 1 is for example satisfied when $\partial\Omega \in C^{1,1}$, $f \in L^2(\Omega)$. Alternatively, one can replace the assumption $\partial\Omega \in C^{1,1}$ by requiring that the domain Ω is convex. We refer to Grisvard (2011) for a detailed discussion of the regularity properties of elliptic equations.*

The following result is a trivial corollary of the H^2 regularity we assumed and a similar result is due to van der Meer et al. (2020), although not exploiting the benefits of exact boundary conditions. Albeit being of simple nature, we believe it can be of practical relevance due to its easy and explicit error control.

4

Theorem 3 *Assume we are in the situation of Setting 1, then it holds for every $\theta \in \Theta$ that*

$$\|u_\theta - u_f\|_{H^2(\Omega)} \leq C_{reg}\sqrt{\mathcal{L}(\theta)}.$$

For convex domains, we may estimate the regularity constant explicitely. It holds

$$C_{reg} \leq \sqrt{1 + C_P} \leq \sqrt{1 + \left(\frac{|\Omega|}{\omega_d}\right)^{\frac{1}{d}}},$$

where d is the dimension of Ω, ω_d denotes the volume of the unit ball in $\mathbb{R}^d$ and C_P is the Poincaré constant for functions in $H_0^1(\Omega)$.

Proof The difference $u_\theta - u_f$ lies in $H^2(\Omega) \cap H_0^1(\Omega)$ and solves $-\Delta(u_f - u_\theta) = \Delta u_\theta + f$. The $H^2(\Omega)$ regularity theory then implies the desired estimate. Let us now assume that Ω is convex and derive the explicit estimate on C_{reg}. We expand the $H^2(\Omega)$ norm of $u_f \in H^2(\Omega) \cap H_0^1(\Omega)$

$$\|u_f\|_{H^2(\Omega)}^2 = \|u_f\|_{L^2(\Omega)}^2 + \|\nabla u_f\|_{L^2(\Omega)}^2 + \|D^2 u_f\|_{L^2(\Omega)}^2$$

Due to the zero boundary values and the convexity of Ω we have

$$\|D^2 u\|_{L^2(\Omega)}^2 = \|\Delta u\|_{L^2(\Omega)}^2 = \|f\|_{L^2(\Omega)}^2$$

and we refer the reader to Grisvard (2011) for details. The first two terms can be estimated jointly using the a priori estimates of the Lax-Milgram Theorem, this yields

$$\|u\|_{L^2(\Omega)}^2 + \|\nabla u\|_{L^2(\Omega)}^2 \leq C_P^2\|f\|_{H_0^1(\Omega)^*}^2 \leq C_P^2\|f\|_{L^2(\Omega)}^2.$$

This is due to the fact that C_P^{-1} is the coercivity constant of the Dirichlet Laplacian bilinear form, see Evans (1998). The explicit estimate of the Poincaré constant C_P can be found in Jost (2003). ∎

Remark 4 *Some remarks are in order.*

(i) *The zero boundary conditions are essential. If one instead resorts to a $L^2(\partial\Omega)$ penalty of the boundary values the best convergence one can hope for is $H^{1/2}(\Omega)$. We elaborate this in Section 2.2.*

(ii) *The theorem allows to compute an explicit upper bound on the error made by residual minimisation, once the training returns a parameter θ via computing the (continuous) loss. In particular, no access to the solution u_f is required. This means that if boundary conditions are encoded in the ansatz functions, the loss itself is a consistent a posteriori error estimator for the residual minimisation method.*

(iii) *The root in the estimate above does not indicate a slow convergence. In fact, the loss itself is a squared $L^2(\Omega)$ norm and the root accounts for that.*

The next theorem allows to quantify the error made by the residual minimisation method using the optimization quality and the expressiveness of the ansatz class. It is an application of the non-linear Céa Lemma as formulated by Müller and Zeinhofer (2021).

Theorem 5 *Assume we are in Setting 1, then for any $\theta \in \Theta$ it holds*

$$\|u_f - u_\theta\|_{H^2(\Omega)} \leq \sqrt{C_{reg}^2 \delta + C_{reg}^2 \inf_{\tilde{\theta} \in \Theta} \|\Delta(u_{\tilde{\theta}} - u_f)\|_{L^2(\Omega)}^2} \leq \sqrt{C_{reg}^2 \delta + C_{reg}^2 \inf_{\tilde{\theta} \in \Theta} \|u_{\tilde{\theta}} - u_f\|_{H^2(\Omega)}^2},$$

(5)

where $\delta = \mathcal{L}(\theta) - \inf_{\tilde{\theta} \in \Theta} \mathcal{L}(\tilde{\theta})$.

Proof We define the energy

$$E \colon H^2(\Omega) \to \mathbb{R}, \quad E(u) = \|\Delta u + f\|_{L^2(\Omega)}^2.$$

Note that E is defined on a different domain than the loss function $\mathcal{L}$, which is why we reserve an own symbol for it. The energy E is a quadratic energy

$$\|\Delta u + f\|_{L^2(\Omega)}^2 = \int_\Omega (\Delta u)^2 \mathrm{d}x + 2 \int_\Omega f \Delta u \mathrm{d}x + \int_\Omega f^2 \mathrm{d}x$$
$$= \frac{1}{2} a(u, u) - F(u) + c,$$

where the bilinear form $a : H^2(\Omega) \times H^2(\Omega) \to \mathbb{R}$, the functional $F \in H^2(\Omega)^*$ and the constant c are given by

$$a(u, v) = 2 \int_\Omega \Delta u \Delta v \mathrm{d}x, \quad F(u) = 2 \int_\Omega f(-\Delta u) \mathrm{d}x, \quad c = \int_\Omega f^2 \mathrm{d}x.$$

The unique minimiser of E in the affine subspace $H^2(\Omega) \cap H_g^1(\Omega)$ is precisely the solution u_f to the Poisson problem (1). The bilinear form a is coercive on the subspace $H^2(\Omega) \cap H_0^1(\Omega)$, which follows from elliptic regularity theory, see for instance Grisvard (2011). This allows to exploit a Céa Lemma for non-linear ansatz spaces, as described in Proposition 3.1 by Müller and Zeinhofer (2021). To transfer this to the affine space $H^2(\Omega) \cap H_g^1(\Omega)$ we choose $u_g \in H^2(\Omega)$ such that $-\Delta u_g = 0$ and $(u_g)|_{\partial\Omega} = g$. For an arbitrary u_θ we then expand

$$\|u_\theta - u_f\|_{H^2(\Omega)} = \|(u_\theta - u_g) - (u_f - u_g)\|_{H^2(\Omega)}.$$

Now note that $u_f - u_g$ solves $-\Delta(u_f - u_g) = f$ with zero boundary values, hence $u_f - u_g$ is the unique minimiser of E over the subspace $H^2(\Omega) \cap H_0^1(\Omega)$ and we can apply Céa's Lemma with the ansatz set $\{u_\theta - u_g \mid \theta \in \Theta\}$

$$\|u_\theta - u_f\|_{H^2(\Omega)} \leq \sqrt{\frac{2\delta}{\alpha} + \frac{1}{\alpha} \inf_{\tilde{\theta} \in \Theta} \|u_{\tilde{\theta}} - u_f\|_a^2},$$

where $\|\cdot\|_a$ denotes the norm induced by a. Using that the coercivity constant α of a is $2/C_{reg}^2$ and the norm $\|\cdot\|_a = 2\|\Delta\cdot\|_{L^2(\Omega)}$ we conclude. ∎

Remark 6 (Errors with Respect to the Complexity of the Ansatz Class) *Note that our results so far yield two possibilities to relate the complexity of the ansatz class to the smallest achievable*

error. First, as a consequence of the regularity estimate of Theorem 3 we can control the smallest achievable value of the loss function

$$\inf_{\theta \in \Theta} \mathcal{L}(\theta) \le c \cdot \inf_{\theta \in \Theta} \|u_\theta - u_f\|^2_{H^2(\Omega)}$$

as already discussed in (3). On the other hand, by the application of Theorem 5 we can estimate the smallest achievable error in the $H^2(\Omega)$ norm (setting $\delta = 0$, hence assuming perfect training)

$$\|u_f - u_\theta\|_{H^2(\Omega)} \le C_{reg} \inf_{\tilde{\theta} \in \Theta} \|u_{\tilde{\theta}} - u_f\|_{H^2(\Omega)}.$$

We stress however, that the complexity of the ansatz class is only implicitly present in the above estimates and that explicit estimates require a quantitative universal approximation result in $H^2(\Omega)$ norm with zero boundary values. To the best of our knowledge, no quantitative estimates on the H^2 approximation are available in the literature and there are no estimates for neural network approximation with zero boundary values at all.

Remark 7 (General Elliptic Equations) *The discussion of this chapter can be extended to more general elliptic equations. For coefficients $A \in C^{0,1}(\Omega, \mathbb{R}^{d \times d})$, a right-hand side $f \in L^2(\Omega)$ and boundary values $g \in H^{3/2}(\partial\Omega)$ consider the equation*

$$- \operatorname{div}(A\nabla u) = f \quad in \ \Omega,$$
$$u = g \quad on \ \partial\Omega.$$

If we assume that $\partial\Omega \in C^{1,1}$ (or that Ω is convex) and the coefficients are uniformly elliptic, i.e., for a constant $c_A > 0$ satisfy $A(x)\xi \cdot \xi \ge c_A |\xi|^2$ uniformly in $x \in \Omega$ and $\xi \in \mathbb{R}^d$, the problem admits a unique solution $u_f \in H^2(\Omega)$ and we can estimate

$$\|u_f\|_{H^2(\Omega)} \le c_{reg} \left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{3/2}(\partial\Omega)} \right).$$

Arguing as in the proof of Theorem 3 we obtain

$$\|u_\theta - u_f\|_{H^2(\Omega)} \le c_{reg} \sqrt{\mathcal{L}(\theta)}.$$

Similarly, Theorem 5 can be transferred to this setting.

2.2. Failure without Exact Boundary Values

In this section we show that *not* enforcing exact boundary values in the neural network ansatz functions leads to considerably weaker error estimates. Throughout this subsection, we work under the following assumptions.

Setting 8 *We consider again (1). We assume that Θ is a parameter set of a neural network type ansatz class, such that for every $\theta \in \Theta$ we have $u_\theta \in H^2(\Omega)$, but make no assumptions on its boundary values. As our strategy is to minimise the residual we define the loss function with boundary penalty*

$$\mathcal{L}_\tau \colon \Theta \to \mathbb{R}, \quad \mathcal{L}_\tau(\theta) = \|\Delta u_\theta + f\|^2_{L^2(\Omega)} + \tau \|u_\theta - g\|^2_{L^2(\partial\Omega)},$$

where $\tau \in (0, \infty)$ is a positive penalization parameter.

Without exact boundary values, the penalization of the deviations of the boundary values is required in order to enforce them approximately. Note that if u_θ has exact boundary values, it holds that $\mathcal{L}_\tau(\theta) = \mathcal{L}(\theta)$. With the penalization introduced above, we obtain a similar result to Theorem 3 but only with respect to the weaker $H^{1/2}$-norm, which is to the estimate by Shin et al. (2020) is sharp. However, we sharpen this result by showing that $1/2$ is the largest exponent for which such an estimate can hold in general.

Theorem 9 *Assume that we are in Setting 8 and that the domain $\Omega \subseteq \mathbb{R}^d$ has a smooth boundary $\partial\Omega \in C^\infty$. Then for $s \in \mathbb{R}$ there is a constant $c > 0$ such that*

$$\|u_\theta - u_f\|_{H^s(\Omega)} \leq c\sqrt{\mathcal{L}_\tau(\theta)} \quad \text{for all } \theta \in \Theta \tag{6}$$

and all parametric classes and data $f \in L^2(\Omega), g \in H^{3/2}(\partial\Omega)$ if and only if $s \leq 1/2$.

Proof First, we show that the estimate holds for $s \leq 1/2$, where it suffices to show it for $s = 1/2$. For this, we use the estimate

$$\|u\|_{H^s(\Omega)} \leq c\left(\|-\Delta u\|_{H^{s-2}(\Omega)} + \|u\|_{H^{s-1/2}(\partial\Omega)}\right), \tag{7}$$

for all $u \in C^\infty(\overline{\Omega})$ and $s \in \mathbb{R}$, see Theorem 2.1 in Schechter (1963) or Lemma 6.2 in Shin et al. (2020). Setting $s = 1/2$ and noting that it extends to functions $u \in H^2(\Omega)$ yields

$$\|u\|_{H^{1/2}(\Omega)} \leq c\left(\|\Delta u\|_{H^{-3/2}(\Omega)} + \|u\|_{L^2(\partial\Omega)}\right) \leq \tilde{c}\left(\|\Delta u\|_{L^2(\Omega)} + \|u\|_{L^2(\partial\Omega)}\right).$$

Setting $u := u_\theta - u_f$ yields

$$\|u_\theta - u_f\|_{H^{1/2}(\Omega)} \leq (1 + \tau^{-1/2})\tilde{c}\sqrt{\mathcal{L}_\tau(\theta)}.$$

To show that the estimate (6) can not in general be established for any stronger norms, we assume that it holds for some $s \in \mathbb{R}$. As in the proof of Theorem 5 we define the energy, this time penalising boundary values

$$E_\tau \colon H^2(\Omega) \to \mathbb{R}, \quad E_\tau(u) := \|\Delta u + f\|_{L^2(\Omega)}^2 + \tau\|u - g\|_{L^2(\partial\Omega)}^2.$$

If the estimate (6) holds for general parametric classes, this yields

$$\|v - u_f\|_{H^s(\Omega)}^2 \leq c \cdot E_\tau(v) \quad \text{for all } v \in H^2(\Omega), f \in L^2(\Omega), g \in H^{3/2}(\partial\Omega).$$

Choosing $f = 0$ and $g = 0$ yields

$$\|v\|_{H^s(\Omega)}^2 \leq c \cdot E_\tau(v) = c \cdot \left(\|\Delta v\|_{L^2(\Omega)}^2 + \tau\|v\|_{L^2(\partial\Omega)}^2\right) \quad \text{for all } v \in H^2(\Omega).$$

For $h \in H^{3/2}(\partial\Omega)$ let $u_h \in H^2(\Omega)$ denote the unique harmonic extension, i.e., the solution of

$$-\Delta u_h = 0 \quad \text{in } \Omega$$
$$u_h = h \quad \text{on } \partial\Omega.$$

8

Now we have

$$\|h\|_{H^{s-1/2}(\partial\Omega)}^2 \le c\|u_h\|_{H^s(\Omega)}^2 \le \tilde{c}\left(\|\Delta u_h\|_{L^2(\Omega)}^2 + \tau\|u_h\|_{L^2(\partial\Omega)}^2\right) = \tilde{c}\tau \cdot \|h\|_{L^2(\partial\Omega)}^2 \qquad (8)$$

for all $h \in H^{3/2}(\partial\Omega)$. In order to see that this implies $s \le 1/2$ we assume the contraty and set $\varepsilon := s - 1/2 > 0$. Then, the embedding $H^{3/2}(\partial\Omega) \hookrightarrow H^\varepsilon(\partial\Omega)$ is dense and hence (8) extends to $h \in H^\varepsilon(\partial\Omega)$. This yields that all norms $\|\cdot\|_{H^\delta(\partial\Omega)}$ for $\delta \in (0,\varepsilon)$ are equivalent to $\|\cdot\|_{L^2(\partial\Omega)}$, which implies that all spaces $H^\delta(\partial\Omega)$ agree which constitutes a contradiction. ■

Remark 10 (Stronger estimates through stronger penalty) *We have seen that the $L^2(\partial\Omega)$ penalisation can not lead to estimates in a stronger Sobolev norm than $H^{1/2}(\Omega)$. However, inspecting inequality (7) one could – at least in theory – penalise the boundary values in the $H^{3/2}(\partial\Omega)$ norm and would then obtain $H^2(\Omega)$ estimates. As the $H^{3/2}(\partial\Omega)$ norm is difficult to approximate in practice, this is no feasible numerical approach.*

Remark 11 (Stronger estimates through interpolation) *It is possible to bound the H^s error for $s \ge 1/2$ of residual minimisation with L^2 boundary penalty for the expense of worse rates and under the cost of an additional factor for which it is not clear whether it is bounded. Similar to Biswas et al. (2020) one can use an interpolation inequality for $s \in [1/2, 2]$ to obtain*

$$\|u\|_{H^s(\Omega)} \le \|u\|_{H^{1/2}(\Omega)}^{2(2-s)/3} \cdot \|u\|_{H^2(\Omega)}^{(2s-1)/3} \quad \text{for all } u \in H^2(\Omega).$$

Together with the a posteriori estimate on the $H^{1/2}$ norm, this yields

$$\|u_f - u_\theta\|_{H^s(\Omega)} \le \|u_f - u_\theta\|_{H^{1/2}(\Omega)}^{2(2-s)/3} \cdot \|u_f - u_\theta\|_{H^2(\Omega)}^{(2s-1)/3} \precsim \|u_f - u_\theta\|_{H^2(\Omega)}^{(2s-1)/3} \cdot L(\theta)^{(2-s)/3}$$

$$\le \left(\|u_f\|_{H^2(\Omega)} + \|u_\theta\|_{H^2(\Omega)}\right)^{(2s-1)/3} \cdot L(\theta)^{(2-s)/3}.$$

Hence, if it is possible to control the H^2 norm of the neural network functions, one obtains an a posteriori estimate on the H^s error. Note however, that the H^2 norm of the neural networks functions is not controlled through the loss function L and hence, this estimates requires an additional explicit or implicit control on the H^2 norm in order to be informative. Note, however, that the power of the a posteriori estimate decreases towards zero for $s \to 2$ and the estimate collapses to a trivial bound for $s = 2$.

2.3. Higher Order Sobolev Norms as a Residual Measurement

We discuss the potential benefit of using (higher order) Sobolev norms to measure the residual, as was already proposed by Son et al. (2021). We are again supposing the exact enforcement of boundary conditions. Our precise setting is the following.

Setting 12 *Let $p \in (1, \infty)$ and $k \ge 0$ be fixed. Assume that $\Omega \subseteq \mathbb{R}^d$ is a bounded, open domain with $C^{k+1,1}$ boundary and let $f \in W^{k,p}(\Omega)$ and $g \in W^{2+k-1/p,p}(\partial\Omega)$. Denote by u_f the solution to (1). Furthermore, let Θ be a parameter set of a neural network class, such that for every $\theta \in \Theta$ we have $u_\theta \in W^{k+2,p}(\Omega)$ and $u|_{\partial\Omega} = g$. We define the loss function*

$$\mathcal{L}: \Theta \to \mathbb{R}, \quad \mathcal{L}(\theta) = \|\Delta u_\theta + f\|_{W^{k,p}(\Omega)}^p. \qquad (9)$$

In total analogy to Theorem 3 we obtain the following result.

Theorem 13 *Assume we are in the situation of Setting 12, then it holds for every $\theta \in \Theta$ that*

$$\|u_\theta - u_f\|_{W^{k+2,p}(\Omega)} \le C_{reg}(p, k) \sqrt[p]{\mathcal{L}(\theta)}.$$

Proof The essential ingredient is the L^p regularity theory that holds under the assumptions made in Setting 12, see for instance chapter 2.5 in Grisvard (2011). The relevant result is that

$$-\Delta \colon W^{k+2,p}(\Omega) \cap W_0^{1,p}(\Omega) \to W^{k,p}(\Omega)$$

is a linear homeomorphism, where $C_{\text{reg}}(p, k)$ denotes the operator norm of its inverse. ∎

Remark 14 *The above result might be interesting if approximation of higher derivatives is desired. Furthermore, the empirical findings of Son et al. (2021) suggest that measuring the residual in a Sobolev norm might lead to fewer iterations in a gradient based optimization routine.*

3. Estimates for Parabolic Equations

The same observation made for the Poisson equation can be exploited for linear parabolic equations when both initial and boundary values are satisfied exactly by the ansatz class. Here, the key is maximal parabolic L^2 regularity theory. We begin by describing our setting.

Setting 15 *We consider again a domain $\Omega \subseteq \mathbb{R}^d$ that is H^2 regular for the Laplacian and a finite time interval $I = [0, T]$. For $f \in L^2(I, L^2(\Omega))$, $g \in H^{3/2}(\partial\Omega)$ and $u_0 \in H_0^1(\Omega)$ we consider the parabolic problem*

$$\begin{aligned}
d_t u - \Delta u &= f && \text{in } I \times \Omega \\
u(t)|_{\partial\Omega} &= g && \text{for all } t \in I \\
u(0) &= u_0.
\end{aligned} \tag{10}$$

Let Θ be a parameter set of a neural network class such that for every $\theta \in \Theta$ the function u_θ is a member of the space

$$\mathcal{X} = H^1(I, L^2(\Omega)) \cap L^2(I, H^2(\Omega) \cap H_g^1(\Omega)), \quad \|u\|_{\mathcal{X}} = \|d_t u\|_{L^2(I, L^2(\Omega))} + \|u\|_{L^2(I, H^2(\Omega))}$$

with $u_\theta(0) = u_0$. This means that both initial and boundary conditions are satisfied exactly. For an introduction to vector-valued Sobolev spaces we refer the reader to Boyer and Fabrie (2012). Then we define the loss function

$$\mathcal{L}(\theta) = \|d_t u_\theta - \Delta u_\theta - f\|_{L^2(I, L^2(\Omega))}^2$$

The following theorem is analogue to the case of the Laplacian and relies on a parabolic regularity result.

Theorem 16 *Assume we are in Setting 15. Then it holds for all $\theta \in \Theta$ that*

$$\|u_\theta - u_f\|_{\mathcal{X}} \le C\sqrt{\mathcal{L}(\theta)}$$

Proof We denote by $H_0^1(I, L^2(\Omega))$ the vector-valued Sobolev space with vanishing initial values. Maximal parabolic $L^2(\Omega)$ regularity theory tells us that

$$d_t - \Delta \colon H_0^1(I, L^2(\Omega)) \cap L^2(I, H^2(\Omega) \cap H_0^1(\Omega)) \longrightarrow L^2(I, L^2(\Omega))$$

is a linear homeomorphism and this implies the assertion, see for instance Arendt et al. (2017) for more information on maximal parabolic regularity. The constant C is then the operatornorm of $(d_t - \Delta)^{-1}$. ∎

Remark 17 *Of course this result is not limited to the heat equation. Indeed one can replace $-\Delta$ by a self-adjoint, coercive operator that satisfies $H^2(\Omega)$ regularity, we refer the reader again to Arendt et al. (2017) for the corresponding regularity theory. For information on the dependency of the constant C on data, we refer to Amann (1995), especially Theorem 4.10.8.*

Remark 18 *Mishra and Molinaro (2022) report error estimates for parabolic equations not enforcing initial and boundary conditions in the ansatz architecture. We stress that even though the solutions there are assumed to be classical, smooth solutions the error is only estimated in the $L^2(I \times \Omega)$ norm which is weaker than the estimates presented here. This is again due to advantage of exact boundary and initial conditions.*

Acknowledgments

References

Herbert Amann. *Linear and Quasilinear Parabolic Problems: Volume I: Abstract Linear Theory*, volume 1. Springer Science & Business Media, 1995.

Wolfgang Arendt, Dominik Dier, and Stephan Fackler. JL Lions' Problem on Maximal Regularity. *Archiv der Mathematik*, 109(1):59–72, 2017.

Jens Berg and Kaj Nyström. A unified deep artificial Neural Network Approach to Partial Differential Equations in complex Geometries. *Neurocomputing*, 317:28–41, 2018.

Animikh Biswas, Jing Tian, and Suleyman Ulusoy. Error Estimates for Deep Learning Methods in Fluid Dynamics. *arXiv preprint arXiv:2008.02844*, 2020.

Franck Boyer and Pierre Fabrie. *Mathematical Tools for the Study of the Incompressible Navier-Stokes Equations and related Models*, volume 183. Springer Science & Business Media, 2012.

Jingrun Chen, Rui Du, and Keke Wu. A Comparison Study of Deep Galerkin Method and Deep Ritz Method for Elliptic Problems with Different Boundary Conditions. *arXiv e-prints*, pages arXiv–2005, 2020.

Luca Courte and Marius Zeinhofer. Robin pre-training for the deep ritz method. *arXiv preprint arXiv:2106.06219*, 2021.

Eric C Cyr, Mamikon A Gulian, Ravi G Patel, Mauro Perego, and Nathaniel A Trask. Robust Training and Initialization of Deep Neural Networks: An adaptive Basis Viewpoint. In *Mathematical and Scientific Machine Learning*, pages 512–536. PMLR, 2020.

MWMG Dissanayake and Nhan Phan-Thien. Neural-Network-based Approximations for solving Partial Differential Equations. *Communications in Numerical Methods in Engineering*, 10(3): 195–201, 1994.

Weinan E and Bing Yu. The Deep Ritz method: a Deep Learning-based numerical Algorithm for solving Variational Problems. *Communications in Mathematics and Statistics*, 6(1):1–12, 2018.

Lawrence C Evans. *Partial Differential Equations*, volume 19. Rhode Island, USA, 1998.

Pierre Grisvard. *Elliptic Problems in nonsmooth Domains*. SIAM, 2011.

Jürgen Jost. *Partial Differential Equations*. Springer, 2003.

Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W Mahoney. Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems*, 34, 2021.

Isaac E Lagaris, Aristidis Likas, and Dimitrios I Fotiadis. Artificial Neural Networks for solving Ordinary and Partial Differential Equations. *IEEE transactions on neural networks*, 9(5):987–1000, 1998.

Liyao Lyu, Keke Wu, Rui Du, and Jingrun Chen. Enforcing exact Boundary and Initial Conditions in the deep mixed Residual Method. *arXiv preprint arXiv:2008.01491*, 2020.

Siddhartha Mishra and Roberto Molinaro. Estimates on the generalization error of physics-informed neural networks for approximating pdes. *IMA Journal of Numerical Analysis*, 2022.

Johannes Müller and Marius Zeinhofer. Error Estimates for the Variational Training of Neural Networks with Boundary Penalty. *arXiv preprint arXiv:2103.01007*, 2021.

Maziar Raissi and George Em Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, 2018.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. 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, 2019.

Julian Roth, Max Schröder, and Thomas Wick. Neural Network guided adjoint Computations in dual weighted residual Error Estimation. *arXiv preprint arXiv:2102.12450*, 2021.

Martin Schechter. On L^p Estimates and Regularity II. *Mathematica Scandinavica*, 13(1):47–69, 1963.

Yeonjong Shin, Zhongqiang Zhang, and George Em Karniadakis. Error Estimates of Residual Minimization using Neural Networks for linear PDEs. *arXiv preprint arXiv:2010.08019*, 2020.

Justin Sirignano and Konstantinos Spiliopoulos. DGM: A Deep Learning Algorithm for solving Partial Differential Equations. *Journal of computational physics*, 375:1339–1364, 2018.

Hwijae Son, Jin Woo Jang, Woo Jin Han, and Hyung Ju Hwang. Sobolev Training for the Neural Network Solutions of PDEs. *arXiv preprint arXiv:2101.08932*, 2021.

Remco van der Meer, Cornelis Oosterlee, and Anastasia Borovykh. Optimally weighted Loss Functions for solving PDEs with Neural Networks. *arXiv preprint arXiv:2002.06269*, 2020.