Advances in Continuous and Discrete Models

RESEARCH Open Access

Uniform convergence guarantees for the deep Ritz method for nonlinear problems

Patrick Dondl[1*], Johannes Müller[2] and Marius Zeinhofer[3]

*Correspondence: patrick.dondl@mathematik.uni-freiburg.de
[1]Department of Applied Mathematics, University of Freiburg, Hermann-Herder-Straße 10, 79104 Freiburg i. Br., Germany
Full list of author information is available at the end of the article

Abstract

We provide convergence guarantees for the Deep Ritz Method for abstract variational energies. Our results cover nonlinear variational problems such as the p-Laplace equation or the Modica–Mortola energy with essential or natural boundary conditions. Under additional assumptions, we show that the convergence is uniform across bounded families of right-hand sides.

Keywords: Calculus of variations; Nonlinear problems; Ritz method; Boundary penalty method; Neural networks

1 Introduction

The idea of the Deep Ritz Method is to use variational energies as an objective function for neural network training to obtain a finite-dimensional optimization problem that allows solving the underlying partial differential equation approximately. The idea of deriving a finite-dimensional optimization problem from variational energies dates back to Ritz [28], was widely popularized in the context of finite element methods (see, e.g., Braess [4]), and was recently revived by E and Yu [13] using deep neural networks. In the following, we give a more thorough introduction to the Deep Ritz Method. Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain and consider the variational energy corresponding to the Lagrangian L and a force f, namely

$$E\colon X \to \mathbb{R}, \qquad E(u) = \int_\Omega L\big(\nabla u(x), u(x), x\big) - f(x)u(x)\,\mathrm{d}x, \tag{1}$$

defined on a suitable function space X, usually a Sobolev space $W^{1,p}(\Omega)$. One is typically interested in minimizers of E on subsets $U \subseteq X$ where U encodes further physical constraints, such as boundary conditions. Here, we consider either unconstrained problems or zero Dirichlet boundary conditions and use the notation $U = X_0$ for the latter case. In other words, for zero boundary conditions, one aims to find

$$u \in \operatorname*{argmin}_{v \in X_0} \int_\Omega L\big(\nabla v(x), v(x), x\big) - f(x)v(x)\,\mathrm{d}x. \tag{2}$$

To solve such a minimization problem numerically, the idea dating back to Ritz [28] is to use a parametric ansatz class

$$A := \left\{ u_\theta \in X \mid \theta \in \Theta \subseteq \mathbb{R}^P \right\} \subseteq U \tag{3}$$

and to consider the finite-dimensional minimization problem of finding

$$\theta^* \in \operatorname*{argmin}_{\theta\in\Theta} \int_\Omega L\big(\nabla v_\theta(x), v_\theta(x), x\big) - f(x)v_\theta(x)\,\mathrm{d}x$$

which can be approached by different strategies, depending on the class A. For instance, if A is chosen to be a finite element ansatz space or polynomials and the structure of E is simple enough, one uses optimality conditions to solve this problem.

In this manuscript, we focus on ansatz classes that are given through (deep) neural networks. When choosing such ansatz functions, the method is known as the Deep Ritz Method and was recently proposed by E and Yu [13]. Neural network type ansatz functions possess a parametric form as in (3), however, it is difficult to impose zero boundary conditions on the ansatz class A. To circumvent this problem, one can use a penalty approach, relaxing the energy to the full space, but penalizing the violation of zero boundary conditions, to include these. This means that for a penalization parameter $\lambda > 0$ one aims to find

$$\theta_\lambda^* \in \operatorname*{argmin}_{\theta\in\Theta} \int_\Omega L\big(\nabla v_\theta(x), v_\theta(x), x\big) - f(x)v_\theta(x)\,\mathrm{d}x + \lambda \int_{\partial\Omega} v_\theta^2\,\mathrm{d}s. \tag{4}$$

The idea of using neural networks for the approximate solution of PDEs can be traced back at least to the works of Lee and Kang [21], Dissanayake and Phan-Thien [10], Takeuchi and Kosugi [32], Lagaris et al. [20]. Since the recent successful application of neural network based methods to stationary and instationary PDEs by E et al. [12], E and Yu [13], Sirignano and Spiliopoulos [30], there is an ever growing body of theoretical works contributing to the understanding of these approaches. For a collection of the different methods, we refer to the overview articles by Beck et al. [3], Han et al. [15].

The error in the Deep Ritz Method, which decomposes into an approximation, optimization, and generalization terms, has been studied by Luo and Yang [25], Xu [34], Duan et al. [11], Hong et al. [17], Jiao et al. [18], Lu et al. [23], Lu et al. [24], Müller and Zeinhofer [26]. However, those works either consider non-essential boundary conditions or they require a term with a positive potential, apart from Müller and Zeinhofer [26]. This excludes the prototypical Poisson equation, which was originally treated by the Deep Ritz Method by E and Yu [13]. More importantly, those works only study linear problems, which excludes many important applications.

In this work, we thus study the convergence of the Deep Ritz Method when a sequence of growing ansatz classes $(A_n)_{n\in\mathbb{N}}$, given through parameter sets Θ_n and a penalization of growing strength $(\lambda_n)_{n\in\mathbb{N}}$ with $\lambda_n \nearrow \infty$, is used in the optimization problem (4) with more modest assumptions on L, f, and Ω.

Denote a sequence of (almost) minimizing parameters of problem (4) with parameter set Θ_n and penalization λ_n by θ_n. We then see that under mild assumptions on $(A_n)_{n\in\mathbb{N}}$ and E, the sequence $(u_{\theta_n})_{n\in\mathbb{N}}$ of (almost) minimizers converges weakly in X to the solution of

the continuous problem, see Theorem 7 in Sect. 3. We then strengthen this result in Sect. 4 where we show that the aforementioned convergence is uniform across certain bounded families of right-hand sides f, see Theorem 12. This means that a fixed number of degrees of freedom in the ansatz class can be used independently of the right-hand side to achieve a given accuracy. Alternatively, given a discretization of the space of right-hand sides, one may discretize the solution operator that maps f to the minimizer u of (2) and still obtain a convergence guarantee (although this is not necessarily a viable numerical approach).

To the best of our knowledge, our results currently comprise the only convergence guarantees for the Deep Ritz Method for nonlinear problems. However, since we prove these results using Γ-convergence methods, no rates of convergence are obtained – as mentioned above, for linear elliptic equations some error decay estimates are known. Our results also do not provide insight into the finite-dimensional optimization problem (4) which is a challenging problem in its own right, see, for instance, Wang et al. [33], Courte and Zeinhofer [8]. However, they guarantee that given one is able to solve (4) to a reasonable accuracy, one is approaching the solution of the continuous problem (2).

Our results are formulated for neural network type ansatz functions due to the current interest in using these in numerical simulations, yet other choices are possible. For instance, our results do apply directly to finite element functions.

The remainder of this work is organized as follows. Section 2 discusses some preliminaries and the used notation. The main results, namely Γ-convergence and uniformity of convergence are provided in Sects. 3 and 4, respectively. Finally, in Sect. 5 we discuss how the p-Laplace and a phase field model fit into our general framework.

2 Notation and preliminaries

We fix our notation and present the tools that our analysis relies on.

2.1 Notation of Sobolev spaces and Friedrich's inequality

We denote the space of functions on $\Omega \subseteq \mathbb{R}^d$ that are integrable in the pth power by $L^p(\Omega)$, where we assume that $p \in [1, \infty)$. Endowed with

$$\|u\|_{L^p(\Omega)}^p := \int_\Omega |u|^p \, \mathrm{d}x,$$

this is a Banach space, i.e., a complete normed space. If u is a multivariate function with values in $\mathbb{R}^m$, we interpret $|\cdot|$ as the Euclidean norm. We denote the subspace of $L^p(\Omega)$ of functions with weak derivatives up to order k in $L^p(\Omega)$ by $W^{k,p}(\Omega)$, which is a Banach space with the norm

$$\|u\|_{W^{k,p}(\Omega)}^p := \sum_{l=0}^{k} \left\|D^l u\right\|_{L^p(\Omega)}^p.$$

This space is called a *Sobolev space* and we denote its dual space, i.e., the space consisting of all bounded and linear functionals on $W^{k,p}(\Omega)$ by $W^{k,p}(\Omega)^*$. The closure of all compactly supported smooth functions $C_c^\infty(\Omega)$ in $W^{k,p}(\Omega)$ is denoted by $W_0^{k,p}(\Omega)$. It is well known that if Ω has a Lipschitz continuous boundary the operator that restricts a Lipschitz continuous function on $\overline{\Omega}$ to the boundary admits a linear and bounded extension $\mathrm{tr}\colon W^{1,p}(\Omega) \to L^p(\partial\Omega)$. This operator is called the *trace operator* and its kernel is precisely

$W_0^{1,p}(\Omega)$. Further, we write $\|u\|_{L^p(\partial\Omega)}$ whenever we mean $\|\mathrm{tr}(u)\|_{L^p(\partial\Omega)}$. In the following we mostly work with the case $p = 2$ and write $H^k_{(0)}(\Omega)$ instead of $W^{k,2}_{(0)}(\Omega)$.

In order to study the boundary penalty method, we use the Friedrich inequality which states that the $L^p(\Omega)$ norm of a function can be estimated by the norm of its gradient and boundary values. We refer to Gräser [14] for a proof.

Proposition 1 (Friedrich's inequality) *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded and open set with Lipschitz boundary $\partial\Omega$ and $p \in (1,\infty)$. Then there exists a constant $c > 0$ such that*

$$\|u\|^p_{W^{1,p}(\Omega)} \leq c^p \cdot \left(\|\nabla u\|^p_{L^p(\Omega)} + \|u\|^p_{L^p(\partial\Omega)}\right) \quad \text{for all } u \in W^{1,p}(\Omega). \tag{5}$$

2.2 Neural networks

Here we introduce our notation for the functions represented by a feedforward neural network. Consider natural numbers $d, m, L, N_0, \ldots, N_L \in \mathbb{N}$ and let

$$\theta = \big((A_1, b_1), \ldots, (A_L, b_L)\big)$$

be a tuple of matrix–vector pairs where $A_l \in \mathbb{R}^{N_l \times N_{l-1}}$, $b_l \in \mathbb{R}^{N_l}$ and $N_0 = d$, $N_L = m$. Every matrix–vector pair (A_l, b_l) induces an affine linear map $T_l \colon \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}$. The *neural network function with parameters* θ and with respect to some *activation function* $\rho \colon \mathbb{R} \to \mathbb{R}$ is the function

$$u^\rho_\theta \colon \mathbb{R}^d \to \mathbb{R}^m, \qquad x \mapsto T_L\big(\rho\big(T_{L-1}\big(\rho\big(\cdots \rho\big(T_1(x)\big)\big)\big)\big)\big).$$

The set of all neural network functions of a certain architecture is given by $\{u^\rho_\theta \mid \theta \in \Theta\}$, where Θ collects all parameters of the above form with respect to fixed natural numbers $d, m, L, N_0, \ldots, N_L$. If we have $f = u^\rho_\theta$ for some $\theta \in \Theta$ we say the function f can be *realized* by the neural network $\mathcal{F}^\rho_\Theta$. Note that we often drop the superscript ρ if it is clear from the context.

A particular activation function often used in practice and relevant for our results is the *rectified linear unit* or *ReLU activation function*, which is defined via $x \mapsto \max\{0, x\}$. Arora et al. [2] showed that the class of ReLU networks coincides with the class of continuous and piecewise linear functions. In particular, they are weakly differentiable. Since piecewise linear functions are dense in $H^1_0(\Omega)$, we obtain the following universal approximation result which we prove in detail in the appendix.

Theorem 2 (Universal approximation with zero boundary values) *Consider an open set $\Omega \subseteq \mathbb{R}^d$ and fix a function $u \in W^{1,p}_0(\Omega)$ with $p \in [1,\infty)$. Then for all $\varepsilon > 0$ there exists $u_\varepsilon \in W^{1,p}_0(\Omega)$ that can be realized by an ReLU network of depth $\lceil \log_2(d+1) \rceil + 1$ such that*

$$\|u - u_\varepsilon\|_{W^{1,p}(\Omega)} \leq \varepsilon.$$

To the best of our knowledge, this is the only available universal approximation result where the approximating neural network functions are guaranteed to have zero boundary values. This relies on the special properties of the ReLU activation function and it is unclear for which classes of activation functions universal approximation with zero boundary values hold.

2.3 Gamma convergence

We recall the definition of Γ-convergence with respect to the weak topology of reflexive Banach spaces. For further reading, we point the reader towards Dal Maso [9].

Definition 3 (Γ-convergence) Let X be a reflexive Banach space as well as $F_n, F\colon X \to (-\infty, \infty]$. Then $(F_n)_{n\in\mathbb{N}}$ is said to be Γ-*convergent* to F if the following two properties are satisfied:

(i) (*Liminf inequality*) For every $x \in X$ and $(x_n)_{n\in\mathbb{N}}$ with $x_n \rightharpoonup x$, we have

$$F(x) \le \liminf_{n\to\infty} F_n(x_n).$$

(ii) (*Recovery sequence*) For every $x \in X$, there is $(x_n)_{n\in\mathbb{N}}$ with $x_n \rightharpoonup x$ such that

$$F(x) = \lim_{n\to\infty} F_n(x_n).$$

The sequence $(F_n)_{n\in\mathbb{N}}$ is called *equicoercive* if the set

$$\bigcup_{n\in\mathbb{N}} \{x \in X \mid F_n(x) \le r\}$$

is bounded in X (or equivalently, relatively compact with respect to the weak topology) for all $r \in \mathbb{R}$. We say that a sequence $(x_n)_{n\in\mathbb{N}}$ are *quasiminimizers* of the functionals $(F_n)_{n\in\mathbb{N}}$ if we have

$$F_n(x_n) \le \inf_{x\in X} F_n(x) + \delta_n,$$

where $\delta_n \to 0$.

We need the following property of Γ-convergent sequences. We want to emphasize the fact that there are no requirements regarding the continuity of any of the functionals and that the functionals $(F_n)_{n\in\mathbb{N}}$ are not assumed to admit minimizers.

Theorem 4 (Convergence of quasiminimizers) *Let X be a reflexive Banach space and $(F_n)_{n\in\mathbb{N}}$ be an equicoercive sequence of functionals that Γ-converges to F. Then, any sequence $(x_n)_{n\in\mathbb{N}}$ of quasiminimizers of $(F_n)_{n\in\mathbb{N}}$ is relatively compact with respect to the weak topology of X and every weak accumulation point of $(x_n)_{n\in\mathbb{N}}$ is a global minimizer of F. Consequently, if F possesses a unique minimizer x, then $(x_n)_{n\in\mathbb{N}}$ converges weakly to x.*

3 Abstract Γ-convergence result for the deep Ritz method

For the abstract results, we work with an abstract energy $E\colon X \to \mathbb{R}$, instead of an integral functional of the form (1). This reduces technicalities in the proofs and separates abstract functional-analytic considerations from applications.

Setting 5 *Let $(X, \|\cdot\|_X)$ and $(B, \|\cdot\|_B)$ be reflexive Banach spaces and $\gamma \in \mathcal{L}(X, B)$ be a continuous linear map. We set X_0 to be the kernel of γ, i.e., $X_0 = \gamma^{-1}(\{0\})$. Let $\rho\colon \mathbb{R} \to \mathbb{R}$ be some activation function and denote by $(\Theta_n)_{n\in\mathbb{N}}$ a sequence of neural network parameters.*

We assume that any function represented by such a neural network is a member of X and define

$$A_n := \{x_\theta \mid \theta \in \Theta_n\} \subseteq X. \tag{6}$$

Here, x_θ denotes the function represented by the neural network with the parameters θ. Let $E\colon X \to (-\infty, \infty]$ be a functional and $(\lambda_n)_{n\in\mathbb{N}}$ a sequence of real numbers with $\lambda_n \to \infty$. Furthermore, let $p \in (1,\infty)$ and $f \in X^*$ be fixed and define the functional $F_n^f\colon X \to (-\infty,\infty]$ by

$$F_n^f(x) = \begin{cases} E(x) + \lambda_n \|\gamma(x)\|_B^p - f(x) & \text{for } x \in A_n, \\ \infty & \text{otherwise,} \end{cases}$$

as well as $F^f : X \to (-\infty,\infty]$ by

$$F^f(x) = \begin{cases} E(x) - f(x) & \text{for } x \in X_0, \\ \infty & \text{otherwise.} \end{cases}$$

Then assume the following holds:

(A1) *For every $x \in X_0$, there is $x_n \in A_n$ such that $x_n \to x$ and $\lambda_n \|\gamma(x_n)\|_B^p \to 0$ for $n \to \infty$.*

(A2) *The functional E is bounded from below, weakly lower semicontinuous with respect to the weak topology of $(X, \|\cdot\|_X)$ and continuous with respect to the norm topology of $(X, \|\cdot\|_X)$.*

(A3) *The sequence $(F_n^f)_{n\in\mathbb{N}}$ is equicoercive with respect to the norm $\|\cdot\|_X$.*

Remark 6 We discuss Assumptions (A1) to (A3) in view of their applicability to concrete problems.

(i) In applications, $(X, \|\cdot\|_X)$ will usually be a Sobolev space with its natural norm, the space B contains boundary values of functions in X and the operator γ is a boundary value operator, e.g., the trace map. However, if the energy E is coercive on all of X, i.e., without adding boundary terms to it, we might choose $\gamma = 0$ and obtain $X_0 = X$. This is the case for non-essential boundary value problems.

(ii) Assumption (A1) compensates that, in general, we cannot penalize with arbitrary strength. However, if we can approximate any member of X_0 by a sequence $x_{\theta_n} \in A_n \cap X_0$ then any divergent sequence $(\lambda_n)_{n\in\mathbb{N}}$ can be chosen. This is, for example, the case for the ReLU activation function and the space $X_0 = H_0^1(\Omega)$. More precisely, we can choose A_n to be the class of functions expressed by a (fully connected) ReLU network of depth $\lceil \log_2(d+1) \rceil + 1$ and width n, see Theorem 2.

Theorem 7 (Γ-convergence) *Assume we are in Setting 5. Then the sequence $(F_n^f)_{n\in\mathbb{N}}$ of functionals Γ-converges towards F^f. In particular, if $(\delta_n)_{n\in\mathbb{N}}$ is a sequence of nonnegative real numbers converging to zero, any sequence of δ_n-quasiminimizers of F_n^f is bounded and all its weak accumulation points are minimizers of F^f. If, additionally, F^f possesses a unique*

minimizer $x^f \in X_0$, any sequence of δ_n-quasiminimizers converges to x^f in the weak topology of X.

Proof We begin with the limes inferior inequality. Let $x_n \rightharpoonup x$ in X and assume that $x \notin X_0$. Then $f(x_n)$ converges to $f(x)$ as real numbers and $\gamma(x_n)$ converges weakly to $\gamma(x) \neq 0$ in B. Combining this with the weak lower semicontinuity of $\|\cdot\|_B^p$, we get, using the boundedness from below, that

$$\liminf_{n\to\infty} F_n^f(x_n) \geq \inf_{x\in X} E(x) + \liminf_{n\to\infty} \lambda_n \left\|\gamma(x_n)\right\|_B^p - \lim_{n\to\infty} f(x_n) = \infty.$$

Now let $x \in X_0$. Then by the weak lower semicontinuity of E, we find

$$\liminf_{n\to\infty} F_n^f(x_n) \geq \liminf_{n\to\infty} E(x_n) - f(x) \geq E(x) - f(x) = F^f(x).$$

Now let us have a look at the construction of the recovery sequence. For $x \notin X_0$, we can choose the constant sequence and estimate

$$F_n^f(x_n) \geq E(x) + \lambda_n \left\|\gamma(x)\right\|_B^p - f(x).$$

Hence we find that $F_n^{f_n}(x) \to \infty = F^f(x)$. If $x \in X_0$, we approximate it with a sequence $(x_n) \subseteq X$, according to Assumption (A1), such that $x_n \in A_n$ and $x_n \to x$ in $\|\cdot\|_X$ and $\lambda_n \|\gamma(x_n)\|_B^p \to 0$. It follows that

$$F_n^f(x_n) = E(x_n) + \lambda_n \|x_n\|_B^p - f(x_n) \to E(x) - f(x) = F^f(x). \qquad \square$$

A sufficient criterion for equicoercivity of the sequence $(F_n^f)_{n\in\mathbb{N}}$ from Assumption (A3) in terms of the functional E is given by the following lemma.

Lemma 8 (Criterion for equicoercivity) *Assume we are in Setting 5. If there is a constant $c > 0$ such that it holds for all $x \in X$ that*

$$E(x) + \left\|\gamma(x)\right\|_B^p \geq c \cdot \left(\|x\|_X^p - \|x\|_X - 1\right),$$

then the sequence $(F_n^f)_{n\in\mathbb{N}}$ is equicoercive.

Proof It suffices to show that the sequence

$$G_n^f \colon X \to \mathbb{R} \quad \text{with } G_n^f(x) = E(x) + \lambda_n \left\|\gamma(x)\right\|_B^p - f(x)$$

is equicoercive, as $G_n^f \leq F_n^f$. So let $r \in \mathbb{R}$ be given and assume that $r \geq G_n^f(x)$. We estimate, assuming without loss of generality that $\lambda_n \geq 1$,

$$\begin{aligned} r &\geq E(x) + \lambda_n \left\|\gamma(x)\right\|_B^p - f(x) \\ &\geq c \cdot \left(\|x\|_X^p - \|x\|_X - 1\right) - \|f\|_{X^*} \cdot \|x\|_X \\ &\geq \tilde{c} \cdot \left(\|x\|_X^p - \|x\|_X - 1\right). \end{aligned}$$

As $p > 1$, a scaled version of Young's inequality clearly implies a bound on the set

$$\bigcup_{n\in\mathbb{N}} \{x \in X \mid G_n^f(x) \leq r\}$$

and hence the sequence $(F_n^f)_{n\in\mathbb{N}}$ is seen to be equicoercive. □

4 Abstract uniform convergence result for the deep Ritz method

In this section we present an extension of Setting 5 that allows proving uniform convergence results over certain bounded families of right-hand sides.

Setting 9 *Assume we are in Setting 5. Furthermore, let there be an additional norm* $|\cdot|$ *on* X *such that the dual space* $(X, |\cdot|)^*$ *is reflexive. However, we do not require* $(X, |\cdot|)$ *to be complete. Then, let the following assumptions hold*:

(A4) *The identity* $\mathrm{Id}\colon (X, \|\cdot\|_X) \to (X, |\cdot|)$ *is completely continuous, i.e., maps weakly convergent sequences to strongly convergent ones.*

(A5) *For every* $f \in X^*$, *there is a unique minimizer* $x_f \in X_0$ *of* F^f *and the solution map*

$$S\colon X_0^* \to X_0 \quad \text{with } f \mapsto x^f$$

is demicontinuous, i.e., maps strongly convergent sequences to weakly convergent ones.

Remark 10 As mentioned earlier, $(X, \|\cdot\|_X)$ is usually a Sobolev space with its natural norm. The norm $|\cdot|$ may then chosen to be an $L^p(\Omega)$ or $W^{s,p}(\Omega)$ norm, where s is strictly smaller than the differentiability order of X. In this case, Rellich's compactness theorem provides Assumption (A4).

Lemma 11 (Compactness) *Assume we are in Setting 9. Then the solution operator* $S\colon (X, |\cdot|)^* \to (X_0, |\cdot|)$ *is completely continuous, i.e., maps weakly convergent sequences to strongly convergent ones.*

Proof We begin by clarifying what we mean by S being defined on $(X, |\cdot|)^*$. Denote by i the inclusion map $i\colon X_0 \to X$ and consider

$$\big(X, |\cdot|\big)^* \xrightarrow{\mathrm{Id}^*} \big(X, \|\cdot\|_X\big)^* \xrightarrow{i^*} \big(X_0, \|\cdot\|_X\big)^* \xrightarrow{S} \big(X_0, \|\cdot\|_X\big) \xrightarrow{\mathrm{Id}} \big(X_0, |\cdot|\big).$$

By abusing notation, always when we refer to S as defined on $(X, |\cdot|)^*$ we mean the above composition, i.e., $\mathrm{Id} \circ S \circ i^* \circ \mathrm{Id}^*$. Having explained this, it is clear that it suffices to show that Id^* maps weakly convergent sequences to strongly convergent ones since i^* is continuous, S demicontinuous, and Id strongly continuous. This, however, is a consequence of Schauder's theorem, see, for instance, Alt [1], which states that a linear map $L \in \mathcal{L}(X, Y)$ between Banach spaces is compact if and only if $L^* \in \mathcal{L}(Y^*, X^*)$ is. Here, compact means that L maps bounded sets to relatively compact ones. Let X_c denote the completion of $(X, |\cdot|)$. Then, using the reflexivity of $(X, \|\cdot\|_X)$ it is easily seen that $\mathrm{Id}\colon (X, \|\cdot\|_X) \to X_c$ is compact. Finally, using that $(X, |\cdot|)^* = X_c^*$ the desired compactness of Id^* is established. □

The following theorem is the main result of this section. It shows that the convergence of the Deep Ritz method is uniform on bounded sets in the space $(X, |\cdot|)^*$. The proof of the uniformity follows an idea from Cherednichenko et al. [7], where in a different setting a compactness result was used to amplify pointwise convergence to uniform convergence across bounded sets, compare to Theorem 4.1 and Corollary 4.2 in Cherednichenko et al. [7].

Theorem 12 (Uniform convergence of the Deep Ritz Method) *Assume that we are in Setting 9 and let $\delta_n \searrow 0$ be a sequence of real numbers. For $f \in X^*$, we set*

$$S_n(f) := \left\{ x \in X \;\middle|\; F_n^f(x) \leq \inf_{z \in X} F_n^f(z) + \delta_n \right\},$$

which is the approximate solution set corresponding to f and δ_n. Furthermore, denote the unique minimizer of F^f in X_0 by x^f and fix $R > 0$. Then we have

$$\sup\left\{ \left| x_n^f - x^f \right| \mid x_n^f \in S_n(f), \|f\|_{(X,|\cdot|)^*} \leq R \right\} \to 0 \quad \text{for } n \to \infty.$$

In the definition of this supremum, f is measured in the norm of the space $(X, |\cdot|)^*$. This means that $f : (X, |\cdot|) \to \mathbb{R}$ is continuous, which is a more restrictive requirement than the continuity with respect to $\|\cdot\|_X$. Also the computation of this norm takes place in the unit ball of $(X, |\cdot|)$, i.e.,

$$\|f\|_{(X,|\cdot|)^*} = \sup_{|x| \leq 1} f(x).$$

Before we prove Theorem 12, we need a Γ-convergence result similar to Theorem 7. The only difference is that now also the right-hand side may vary along the sequence.

Proposition 13 *Assume that we are in Setting 9, however, we do not need Assumption* (A5) *for this result. Let $f_n, f \in (X, |\cdot|)^*$ be such that $f_n \rightharpoonup f$ in the weak topology of the reflexive space $(X, |\cdot|)^*$. Then the sequence $(F_n^{f_n})_{n \in \mathbb{N}}$ of functionals Γ-converges to F^f in the weak topology of $(X, \|\cdot\|_X)$. Furthermore, the sequence $(F_n^{f_n})_{n \in \mathbb{N}}$ is equicoercive.*

Proof The proof is almost identical to that of Theorem 7 but, since it is brief, we include it for the reader's convenience. We begin with the limes inferior inequality. Let $x_n \rightharpoonup x$ in X and $x \notin X_0$. Then $x_n \to x$ with respect to $|\cdot|$ which implies that $f_n(x_n)$ converges to $f(x)$. Using that $\gamma(x_n) \rightharpoonup \gamma(x)$ in B, combined with the weak lower semicontinuity of $\|\cdot\|_B^p$, we get

$$\liminf_{n \to \infty} F_n^{f_n}(x_n) \geq \inf_{x \in X} E(x) + \liminf_{n \to \infty} \lambda_n \left\| \gamma(x_n) \right\|_B^p - \lim_{n \to \infty} f_n(x_n) = \infty.$$

Now let $x \in X_0$. Then by the weak lower semicontinuity of E, we find

$$\liminf_{n \to \infty} F_n^{f_n}(x_n) \geq \liminf_{n \to \infty} E(x_n) - f(x) \geq E(x) - f(x) = F^f(x).$$

Now let us have a look at the construction of the recovery sequence. For $x \notin X_0$, we can choose the constant sequence and estimate

$$F_n^{f_n}(x) \geq \inf_{x \in X} E(x) + \lambda_n \left\| \gamma(x) \right\|_B^p - \|f_n\|_{(X,|\cdot|)'} \cdot |x|.$$

As $\|f_n\|_{(X,|\cdot|)^*}$ is bounded we find $F_n^{f_n}(x) \to \infty = F^f(x)$. If $x \in X_0$, we approximate it with a sequence $(x_n) \subseteq X$, according to Assumption (A1), such that $x_n \in A_n$ and $x_n \to x$ in $\|\cdot\|_X$ and $\lambda_n \|\gamma(x_n)\|_B^p \to 0$. It follows that

$$F_n^{f_n}(x_n) = E(x_n) + \lambda_n \|x_n\|_B^p - f_n(x_n) \to E(x) - f(x) = F^f(x).$$

The equicoercivity was already assumed in (A3) so it does not need to be shown. □

Proof of Theorem 12 We can choose $(f_n) \subseteq (X, |\cdot|)^*$ and $\|f_n\|_{(X,|\cdot|)^*} \leq R$ and $x_n^{f_n} \in S_n(f_n)$ such that

$$\sup_{\|f\|_{(X,|\cdot|)^*} \leq R x_n^f \in S_n(f)} \left| x_n^f - x^f \right| \leq \left| x_n^{f_n} - x^{f_n} \right| + \frac{1}{n}.$$

Now it suffices to show that $|x_n^{f_n} - x^{f_n}|$ converges to zero. Since $(f_n)_{n\in\mathbb{N}}$ is bounded in $(X, |\cdot|)^*$ and this space is reflexive, we can, without loss of generality, assume that $f_n \rightharpoonup f$ in $(X, |\cdot|)^*$. This implies by Lemma 11 that $x^{f_n} \to x^f$ in $(X, |\cdot|)$. The Γ-convergence result of the previous proposition yields $x_n^{f_n} \rightharpoonup x^f$ in X and hence $x_n^{f_n} \to x^f$ with respect to $|\cdot|$ which concludes the proof. □

5 Examples

We discuss different concrete examples that allow the application of our abstract results and focus on nonlinear problems. In particular, we consider a phase field model illustrating the basic Γ-convergence result of Sect. 3 and the p-Laplacian as an example for the uniform results of Sect. 4.

5.1 General practical considerations

In practice, when solving the optimization problem (4), in order to obtain an approximate solution of the variational problem (2) there are a lot of choices to make. We give an overview over some of them here and report our specific choices in the individual examples.

Optimization One can use almost any kind of optimization algorithm for the approximation solution of the optimization problem (4), where gradient type algorithms and quasi-Newton methods are the most common choice. We use a combination of the Adam optimizer and L-BFGS. The former is a version of stochastic gradient descent with with adaptive moment estimation, see Kingma and Ba [19] and the latter is a quasi-Newton method, see Liu and Nocedal [22]. Typically, the optimization process begins with applying Adam until convergence slows down and the fast local convergence properties of Newton methods can be exploited through the application of the L-BFGS optimizer.

Quadrature In practice, one does not have access to the true gradient of the objective function. Hence, one usually uses estimates for the gradient as update directions. For example, E and Yu [13] used an online SGD estimator, i.e., a Monte Carlo approximation of the integral with fixed sample size. However, one can use any quadrature rule for the evaluation of the integrals in order to obtain approximations of the true gradient. We used a

uniform grid for the discretization of the integral, i.e., the integral is approximated by the sum of the functions values at the grid points divided by the number of grid points. On the boundary of the domains, equispaced integration points are used and the quadrature rule is analogue to the one previously described.

We choose the number of integration points such that no further improved accuracy of the method is observed when increasing their amount. As this was computationally tractable without problems, no more elaborate integration routines were deemed necessary.

Activation functions The only requirements on the activation function present in Setting 5 are that the associated neural network functions belong to the considered Banach space $A_n \subseteq X$ as well as that Condition (A1) holds. The first one is usually of no concern as in practice X is often a space of given smoothness and hence for sufficently smooth activations $A_n \subseteq X$ is satisfied. Note that this is in particular the case $X = H^1(\Omega)$ and for the ReLU activation function. Further, in order to Condition (A1) to hold, it is necessary that the neural network type ansatz classes (A_n) have the universal approximation property in X_0, i.e., that $X_0 \subseteq \overline{\bigcup_{n\in\mathbb{N}} A_n}$. Note that this is the case for shallow[1] networks of increasing width and $X = H^k(\Omega)$ as long as the activation function is k times continuously differentiable and nonpolynomial (Pinkus [27]).

Penalization strength Condition (A1) couples the penalization strength to the norm of the (generalized) boundary values required for approximation of a general element $x \in X$. Consider, for example, the case that for any $x \in X_0$ there are $x_n \in X_n$ such that $x_n \to x$ and $\|\gamma(x_n)\|_B \leq c(x)\delta_n$ for some $\delta_n \to 0$. Then any choice of penalization strengths $\lambda_n \to \infty$ with $\lambda_n\delta_n \to 0$ satisfies Condition (A1). Let us first consider the case with inessential boundary values, which corresponds to $\gamma \equiv 0$ in the notation of Setting 5. Then, as argued above smooth and nonpolynomial activations together with arbitrarily strong penalization strengths are allowed.

Let us consider the case $X = H^1(\Omega)$ and $\gamma = \mathrm{tr}$ and $B = L^2(\Omega)$. For the ReLU activation function, Theorem 2 guarantees the existence of the $u_n \to u$ with $\|u_n\|_{L^2(\partial\Omega)} = 0$, which allows for arbitrarily strong penalization. For other activation functions, which do not possess the universal approximation property with exact (generalized) zero boundary values, the proof of (A1) is more delicate and has to be established in specific cases.

5.2 A phase field model

Let $\varepsilon > 0$ be fixed, $\Omega \subseteq \mathbb{R}^d$ a bounded Lipschitz domain and consider the following energy:

$$E\colon H^1(\Omega) \cap L^4(\Omega) \to [0,\infty), \qquad E(u) = \frac{\varepsilon}{2}\int_\Omega |\nabla u|^2 \,\mathrm{d}x + \frac{1}{\varepsilon}\int_\Omega W(u)\,\mathrm{d}x,$$

where $W\colon \mathbb{R} \to \mathbb{R}$ is the nonlinear function given by

$$W(u) = \frac{1}{4}u^2(u-1)^2 = \frac{1}{4}u^4 - \frac{1}{2}u^3 + \frac{1}{4}u^2.$$

[1] and hence also for deep as can be seen by composing with almost identity layers

The functional E constitutes a way to approximately describe phase separation and the parameter ε encodes the length-scale of the phase transition, see Cahn and Hilliard [6]. We describe now how the Setting 5 is applicable to fully connected neural network ansatz functions with tanh activation function. For the Banach spaces in Setting 5, we choose

$$X = H^1(\Omega) \cap L^4(\Omega), \qquad \|\cdot\|_X = \|\cdot\|_{H^1(\Omega)} + \|\cdot\|_{L^4(\Omega)}.$$

We choose $\gamma \equiv 0$, hence the choice of the space B and its norm is irrelevant. The choice of $\gamma \equiv 0$ corresponds to the case of homogeneous Neumann boundary conditions. The space X is reflexive as it is an intersection of reflexive spaces. We define

$$A_n := \{u_\theta \mid \theta \in \Theta_n\} \subseteq H^1(\Omega) \cap L^4(\Omega),$$

where Θ_n implicitly encodes that we use scalar valued neural networks with input dimension d and arbitrary fixed depth larger or equal to two. The width of all layers (except the input and output) is set to n. With $\gamma \equiv 0$ and this definition of $(A_n)_{n\in\mathbb{N}}$ the requirements of Assumption (A1) are satisfied, as can be seen by well known universal approximation results, we refer to Pinkus [27].

To proceed, the continuity of E with respect to $\|\cdot\|_X$ is clear, hence we turn to the weak lower semicontinuity. To this end, we write E in the following form:

$$E(u) = \underbrace{\frac{\varepsilon}{2}\int_\Omega |\nabla u|^2 \,\mathrm{d}x + \frac{1}{4\varepsilon}\int_\Omega u^4 \,\mathrm{d}x}_{:=E_1(u)} + \underbrace{\frac{1}{\varepsilon}\int_\Omega \frac{1}{4}u^2 - \frac{1}{2}u^3 \,\mathrm{d}x}_{:=E_2(u)}$$

and treat E_1 and E_2 separately. The term E_1 is continuous with respect to $\|\cdot\|_X$ and convex, hence weakly lower semicontinuous. To treat E_2, note that we have the compact embedding

$$H^1(\Omega) \cap L^4(\Omega) \hookrightarrow\hookrightarrow L^3(\Omega).$$

This implies that a sequence that converges weakly in $H^1(\Omega) \cap L^4(\Omega)$ converges strongly in $L^3(\Omega)$ and consequently shows that the term E_2 is continuous with respect to weak convergence in X. Finally, for fixed $f \in X^*$, we need to show that the sequence $(F_n^f)_{n\in\mathbb{N}}$ defined in (6) is equicoercive with respect to $\|\cdot\|_X$. To this end, it suffices to show that the functional

$$G^f : X \to \mathbb{R}, \qquad G^f(u) = \frac{\varepsilon}{2}\int_\Omega |\nabla u|^2 \,\mathrm{d}x + \frac{1}{\varepsilon}\int_\Omega W(u) \,\mathrm{d}x - f(u)$$

is coercive as it holds $F_n^f \geq G^f$. Let $r \in \mathbb{R}$ be fixed and consider all $u \in X$ with $G^f(u) \geq r$. Then we estimate

$$\begin{aligned} r &\geq G^f(u) \\ &\geq \frac{\varepsilon}{2}\int_\Omega |\nabla u|^2 \,\mathrm{d}x + \frac{1}{\varepsilon}\int_\Omega W(u) \,\mathrm{d}x - f(u) \\ &\geq c\|u\|^2_{H^1(\Omega)} - \|f\|_{X^*}\big(\|u\|_{H^1(\Omega)} + \|u\|_{L^4(\Omega)}\big) + \frac{1}{4\varepsilon}\|u\|^4_{L^4(\Omega)} - \frac{1}{3\varepsilon}\|u\|^3_{L^3(\Omega)} \end{aligned}$$

$$\geq c\|u\|_{H^1(\Omega)}^2 - \|f\|_{X^*}\|u\|_{H^1(\Omega)} + \frac{1}{4\varepsilon}\|u\|_{L^4(\Omega)}^4 - \frac{|\Omega|^{1/4}}{3\varepsilon}\|u\|_{L^4(\Omega)}^{3/4} - \|f\|_{X^*}\|u\|_{L^4(\Omega)},$$

where we used the estimate

$$\|u\|_{L^3(\Omega)}^3 \leq |\Omega|^{1/4}\|u\|_{L^4(\Omega)}^{3/4}$$

due to Hölder's inequality. This clearly implies a bound on the set

$$\bigcup_{n\in\mathbb{N}}\{u \in H^1(\Omega) \cap L^4(\Omega) \mid G^f(u) \leq r\}$$

and hence $(F_n^f)_{n\in\mathbb{N}}$ is equicoercive.

Description of the experiment Figure 1 shows two exemplary numerical realizations of the Deep Ritz Method with the unit disk $\Omega = B_1(0)$ as a domain and with right-hand sides

$$f_i = \chi_{B_{r_i}(0,-1/2)} - \chi_{B_{r_i}(0,1/2)}$$

for $r_1 = 0.1$ and $r_2 = 0.4$ corresponding to the left and right picture, respectively. Further, we considered $\varepsilon = 0.01$ and used a fully connected network with tanh activation and three hidden layers of width 16. Note that by Theorem 2 ReLU networks of depth $\lceil \log_2(2+1) \rceil + 1 = 3$ satisfy the universal approximation property with exact zero boundary values. Hence, the number of trainable parameters is 609 in this case. As we were solving a homogeneous Neumann boundary value problem, no penalization was needed. For the discretization of the integral over the unit disk $B_1(0)$, we used an evenly spaced grid and gave equal weights in the numerical approximation of the integrals to the function values at every grid point. For the optimization of the networks parameters, we used Adam with full batch size until the optimization slowed down and then used L-BFGS in order to exploit the fast local convergence properties of quasi-Newton methods. Note that in the case of f_1, a phase transition around the ball $B_{r_1}(0, 1/2)$ is energetically more favorable than the configuration in the right figure, where the radius r_2 is much larger.

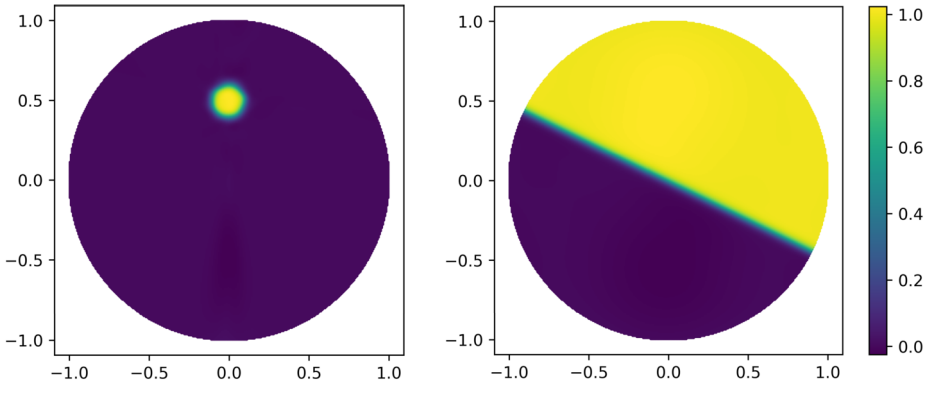

Figure 1 Exemplary numerical realization of the Deep Ritz Method for a Cahn–Hilliard functional with right-hand sides given through $f = \chi_{B_r(0,-1/2)} - \chi_{B_r(0,1/2)}$ with $r = 0.1$ for the left plot and $r = 0.4$ for the right plot. The value of ε is set to 0.01. We used zero Neumann boundary conditions and fully connected feed-forward networks with three hidden layers of width 16 and tanh activation. The number of trainable parameters is 609

Remark 14 (Stability under compact perturbations) With a similar – even simpler – approach, we may also show that energies of the form

$$\hat{E}(u) = E(u) + F(u)$$

fall in the Setting 5 provided E does and F is bounded from below and continuous with respect to weak convergence in X. Note also that in the space dimension $d = 2$ this includes the above example, however, the slightly more involved proof presented here works independently of the space dimension d.

5.3 The p-Laplacian

As an example for the uniform convergence of the Deep Ritz method, we discuss the p-Laplacian. To this end, consider the p-Dirichlet energy for $p \in (1, \infty)$ given by

$$E\colon W^{1,p}(\Omega) \to \mathbb{R}, \qquad u \mapsto \frac{1}{p} \int_\Omega |\nabla u|^p \,\mathrm{d}x.$$

Note that for $p \neq 2$ the associated Euler–Lagrange equation – the p-Laplace equation – is nonlinear. In strong formulation it is given by

$$\begin{aligned} -\operatorname{div}\left(|\nabla u|^{p-2}\nabla u\right) &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned}$$

see, for example, Struwe [31] or Růžička [29]. Choosing the ReLU activation function, the abstract setting is applicable as we will describe now. For the Banach spaces, we choose

$$X = W^{1,p}(\Omega), \qquad B = L^p(\partial\Omega), \qquad |u| = \|u\|_{L^p(\Omega)},$$

where the norms $\|\cdot\|_X$ and $\|\cdot\|_B$ are chosen to be the natural ones. Clearly, $W^{1,p}(\Omega)$ endowed with the norm $\|\cdot\|_{W^{1,p}(\Omega)}$ is reflexive by our assumption $p \in (1, \infty)$. Note that

$$\left(W^{1,p}(\Omega), \|\cdot\|_{L^p(\Omega)}\right)^* = L^p(\Omega)^* \cong L^{p'}(\Omega),$$

which is also reflexive. We set $\gamma = \mathrm{tr}$, i.e.,

$$\mathrm{tr}\colon W^{1,p}(\Omega) \to L^p(\partial\Omega) \quad \text{with } u \mapsto u|_{\partial\Omega}$$

We use the same ansatz sets $(A_n)_{n\in\mathbb{N}}$ as in the previous example, hence Assumption (A1) holds. Rellich's theorem provides the complete continuity of the embedding

$$\left(W^{1,p}(\Omega), \|\cdot\|_{W^{1,p}(\Omega)}\right) \to \left(W^{1,p}(\Omega), \|\cdot\|_{L^p(\Omega)}\right)$$

which shows Assumption (A4). As for Assumption (A3), Friedrich's inequality provides the assumptions of Lemma 8. Furthermore, E is continuous with respect to $\|\cdot\|_{W^{1,p}(\Omega)}$ and convex, hence also weakly lower semicontinuous. By Poincaré's and Young's inequalities, we find for all $u \in W_0^{1,p}(\Omega)$ that

$$F^f(u) = \frac{1}{p} \int_\Omega |\nabla u|^p \,\mathrm{d}x - f(u)$$

$$\geq C\|u\|^p_{W^{1,p}(\Omega)} - \|f\|_{W^{1,p}(\Omega)'}\|u\|_{W^{1,p}(\Omega)}$$
$$\geq C\|u\|^p_{W^{1,p}(\Omega)} - \tilde{C}.$$

Hence, a minimizing sequence in $W^{1,p}_0(\Omega)$ for F^f is bounded and as F^f is strictly convex on $W^{1,p}_0(\Omega)$ it possesses a unique minimizer. Finally, to provide the demicontinuity, we must consider the operator $S\colon W^{1,p}_0(\Omega)^* \to W^{1,p}_0(\Omega)$ mapping f to the unique minimizer u_f of $E - f$ on $W^{1,p}_0(\Omega)$. By the Euler–Lagrange formalism, u minimizes F^f if and only if

$$\int_\Omega |\nabla u|^{p-2}\nabla u \cdot \nabla v \,\mathrm{d}x = f(v) \quad \text{for all } v \in W^{1,p}_0(\Omega).$$

Hence, the solution map S is precisely the inverse of the mapping

$$W^{1,p}_0(\Omega) \to W^{1,p}_0(\Omega)^*, \qquad u \mapsto \left(v \mapsto \int_\Omega |\nabla u|^{p-2}\nabla u \cdot \nabla v \,\mathrm{d}x\right)$$

and this map is demicontinuous, see, for example, Růžička [29].

Description of the experiment Figure 2 shows two numerical realizations of the Deep Ritz Method for the p-Laplacian with right-hand side $f \equiv 1$ and $p_1 = 3/2$ in the left picture and $p_2 = 10$ in the right picture. The penalization value is set to $\lambda = 250$ in both simulations to approximately enforce zero boundary values. We used fully connected feed-forward networks with three hidden layers of width 16 and GELU activation (Hendrycks and Gimpel [16]) for the left plot and ReLU activation for the right plot. The quadrature follows the same strategy as in the previous example. Note that the exact solution to the homogeneous p-Laplace problem on the disk with $f \equiv 1$ is given by

$$u_p(x) = C \cdot \left(1 - |x|^{\frac{p}{p-1}}\right)$$

for a suitable constant C that depends on the spatial dimension and the value of p. We see that the solution u_p converges pointwise to zero for $p \searrow 0$ and for $p \nearrow \infty$ the function u_p tends to $x \mapsto C(1 - |x|)$. This asymptotic behavior is clearly visible in our simulations.

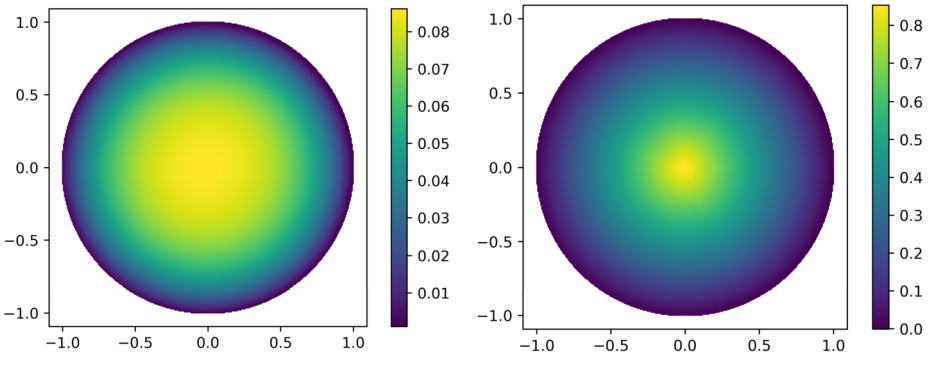

Figure 2 Exemplary numerical realization of the Deep Ritz Method for the p-Laplacian with right-hand $f = 1$ and $p = 1.5$ in the left plot and $p = 10$ in the right plot. Zero Dirichlet boundary conditions are enforced through a penalty parameter $\lambda = 250$. We used fully connected feed-forward networks with three hidden layers of width 16 and GELU activation for the left plot and ReLU activation for the right plot. The number of trainable parameters is 609. Note the difference in the scaling of the axis in the two plots

In case of the ReLU ansatz function, the architecture considered agrees with the analysis presented in the previous paragraph. For $p_1 = 3/2$, we found the GELU activation function to provide good performance. However, for the GELU activation function, establishing condition (A1) is not entirely obvious. The GELU activation is defined as $\mathrm{GELU}(x) := x\Phi(x)$, where Φ is the cumulative distribution function of the Gaussian normal. It is often interpreted as a smoothed version of the ReLU since $t^{-1}\,\mathrm{GELU}(tx) \to \mathrm{ReLU}(x)$ and $\partial_x(t^{-1}\,\mathrm{GELU}(tx)) \to \partial_x\,\mathrm{ReLU}(x)$ pointwise for $t \to \infty$. This gives some intuition why the GELU activation function could admit a universal approximation result with almost zero boundary values and hence satisfy (A1), however, we leave a rigorous statement for future research. This aligns very well with our numerical experiments, which do not indicate problems in resolving the zero boundary values in this practical example.

Appendix: Universal approximation with zero boundary values

Here we prove the universal approximation result which we stated as Theorem 2 in the main text. Our proof uses that every continuous, piecewise-linear function can be represented by a neural network with ReLU activation function and then shows how to approximate Sobolev functions with zero boundary conditions by such functions. The precise definition of a piecewise linear function is the following.

Definition 15 (Continuous piecewise-linear function) We say that a function $f\colon \mathbb{R}^d \to \mathbb{R}$ is *continuous piecewise linear,* or in short *piecewise linear,* if there exists a finite set of closed polyhedra whose union is $\mathbb{R}^d$, and f is affine linear over each polyhedron. Note every piecewise linear functions is continuous by definition since the polyhedra are closed and cover the whole space $\mathbb{R}^d$, and affine functions are continuous.

Theorem 16 (Universal expression) *Every ReLU neural network function $u_\theta\colon \mathbb{R}^d \to \mathbb{R}$ is a piecewise-linear function. Conversely, every piecewise-linear function $f\colon \mathbb{R}^d \to \mathbb{R}$ can be expressed by an ReLU network of depth at most $\lceil \log_2(d+1) \rceil + 1$.*

For the proof of this statement, we refer to Arora et al. [2]. We turn now to the approximation capabilities of piecewise linear functions.

Lemma 17 *Let $\varphi \in C_c^\infty(\mathbb{R}^d)$ be a smooth function with compact support. Then for every $\varepsilon > 0$ there is a piecewise-linear function s_ε such that for all $p \in [1,\infty]$ it holds*

$$\|s_\varepsilon - \varphi\|_{W^{1,p}(\mathbb{R}^d)} \le \varepsilon \quad \textit{and} \quad \operatorname{supp}(s_\varepsilon) \subseteq \operatorname{supp}(\varphi) + B_\varepsilon(0).$$

Here, we set $B_\varepsilon(0)$ to be the ε-ball around zero, i.e., $B_\varepsilon(0) = \{x \in \mathbb{R} \mid |x| < \varepsilon\}$.

Proof In the following we will denote by $\|\cdot\|_\infty$ the uniform norm on $\mathbb{R}^d$. To show the assertion, choose a triangulation $\mathcal{T}$ of $\mathbb{R}^d$ of width $\delta = \delta(\varepsilon) > 0$, consisting of rotations and translations of one nondegenerate simplex K. We choose s_ε to agree with φ on all vertices of elements in $\mathcal{T}$. Since φ is compactly supported, it is uniformly continuous, and hence it is clear that $\|\varphi - s_\varepsilon\|_\infty < \varepsilon$ if δ is chosen small enough.

To show convergence of the gradients, we show that also $\|\nabla\varphi - \nabla s_\varepsilon\|_\infty < \varepsilon$ which will be shown on one element $K \in \mathcal{T}$ and as the estimate is independent of K is understood

to hold on all of $\mathbb{R}^d$. So let $K \in \mathcal{T}$ be given and denote its vertices by $x_1, \dots, x_{d+1}$. We set $v_i = x_{i+1} - x_1$, $i = 1, \dots, d$ to be the vectors spanning K. By the one-dimensional mean value theorem, we find ξ_i on the line segment joining x_1 and x_i such that

$$\partial_{v_i} s_\varepsilon(v_1) = \partial_{v_i} \varphi(\xi_i).$$

Note that $\partial_{v_i} s_\varepsilon$ is constant on all of K where it is defined. Now for arbitrary $x \in K$, we compute with setting $w = \sum_{i=1}^d \alpha_i v_i$ for $w \in \mathbb{R}^d$ with $|w| \leq 1$. Note that the α_i are bounded uniformly in w, where we use that all elements are the same up to rotations and translations,

$$\begin{aligned} \left|\nabla\varphi(x) - \nabla s_\varepsilon(x)\right| &= \sup_{|w|\leq 1} \left|\nabla\varphi(x)w - \nabla s_\varepsilon(x)w\right| \\ &\leq \sup_{|w|\leq 1} \sum_{i=1}^d |\alpha_i| \cdot \underbrace{\left|\partial_{v_i}\varphi(x) - \partial_{v_i} s_\varepsilon(x)\right|}_{=(*)}, \end{aligned}$$

where again $(*)$ is uniformly small due to the uniform continuity of $\nabla\varphi$. Noting that the $W^{1,\infty}$-case implies the claim for all $p \in [1, \infty)$ finishes the proof. □

We turn to the proof of Theorem 2 which we state again for the convenience of the reader.

Theorem 18 (Universal approximation with zero boundary values) *Consider an open set $\Omega \subseteq \mathbb{R}^d$ and let $u \in W_0^{1,p}(\Omega)$ with $p \in [1, \infty)$. Then for all $\varepsilon > 0$ there exists a function $u_\varepsilon \in W_0^{1,p}(\Omega)$ that can be expressed by an ReLU network of depth $\lceil \log_2(d+1) \rceil + 1$ such that*

$$\|u - u_\varepsilon\|_{W^{1,p}(\Omega)} \leq \varepsilon.$$

Proof Let $u \in W_0^{1,p}(\Omega)$ and $\varepsilon > 0$. By the density of $C_c^\infty(\Omega)$ in $W_0^{1,p}(\Omega)$, see, for instance, Brezis [5], we choose a smooth function $\varphi_\varepsilon \in C_c^\infty(\Omega)$ such that $\|u - \varphi_\varepsilon\|_{W^{1,p}(\Omega)} \leq \varepsilon/2$. Furthermore, we use Lemma 17 and choose a piecewise-linear function u_ε such that $\|\varphi_\varepsilon - u_\varepsilon\|_{W^{1,p}(\Omega)} \leq \varepsilon/2$ and such that u_ε has compact support in Ω. This yields

$$\|u - u_\varepsilon\|_{W^{1,p}(\Omega)} \leq \|u - \varphi_\varepsilon\|_{W^{1,p}(\Omega)} + \|\varphi_\varepsilon - u_\varepsilon\|_{W^{1,p}(\Omega)} \leq \varepsilon$$

and, by Theorem 16, we know that u_ε is in fact a realization of a neural network with depth at most $\lceil \log_2(d+1) \rceil + 1$. □

Acknowledgements
None to declare.

Funding
PD and MZ acknowledge support by BMBF via the e:Med consortium SyMBoD (grant number 01ZX1910C). JM acknowledges support by the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement no 757983), the International Max Planck Research School for Mathematics in the Sciences and the Evangelisches Studienwerk Villigst e.V. Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials
The manuscript does not make use of data. Numerical code is not made available.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author contributions
All authors have contributed to all aspects of this manuscript and have reviewed its final draft. All authors read and approved the final manuscript.

Author details
[1]Department of Applied Mathematics, University of Freiburg, Hermann-Herder-Straße 10, 79104 Freiburg i. Br., Germany. [2]Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany. [3]Simula Research Laboratory, Department of Numerical Analysis and Scientific Computing, Kristian Augusts Gate 23, 0164 Oslo, Norway.

Publisher's Note

Received: 1 December 2021 Accepted: 20 June 2022 Published online: 15 July 2022

References
1. Alt, H.W.: Linear Functional Analysis. An Application Oriented Introduction (1992)
2. Arora, R., Basu, A., Mianjy, P., Mukherjee, A.: Understanding deep neural networks with rectified linear units (2016). arXiv:1611.01491. arXiv preprint
3. Beck, C., Hutzenthaler, M., Jentzen, A., Kuckuck, B.: An overview on deep learning-based approximation methods for partial differential equations (2020). arXiv preprint. arXiv:2012.12348
4. Braess, D.: Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics. Cambridge University Press, Cambridge (2007)
5. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer, Berlin (2010)
6. Cahn, J.W., Hilliard, J.E.: Free energy of a nonuniform system. I. interfacial free energy. J. Chem. Phys. **28**(2), 258–267 (1958)
7. Cherednichenko, K., Dondl, P., Rösler, F.: Norm-resolvent convergence in perforated domains. Asymptot. Anal. **110**(3–4), 163–184 (2018)
8. Courte, L., Zeinhofer, M.: Robin pre-training for the deep Ritz method (2021). arXiv:2106.06219. arXiv preprint
9. Dal Maso, G.: An Introduction to Γ-Convergence, vol. 8. Springer, Berlin (2012)
10. Dissanayake, M., Phan-Thien, N.: Neural-network-based approximations for solving partial differential equations. Commun. Numer. Methods Eng. **10**(3), 195–201 (1994)
11. Duan, C., Jiao, Y., Lai, Y., Lu, X., Yang, Z.: Convergence rate analysis for deep Ritz method (2021). arXiv:2103.13330. arXiv preprint
12. E, W., Han, J., Jentzen, A.: Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. Commun. Math. Stat. **5**(4), 349–380 (2017)
13. E, W., Yu, B.: The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. Commun. Math. Stat. **6**(1), 1–12 (2018)
14. Gräser, C.: A note on Poincaré- and Friedrichs-type inequalities (2015). arXiv preprint. arXiv:1512.02842
15. Han, J., Jentzen, A., et al.: Algorithms for solving high dimensional PDEs: from nonlinear Monte Carlo to machine learning. Nonlinearity **35**(1), 278–310 (2021)
16. Hendrycks, D., Gimpel, K.: Gaussian error linear units (GELUs) (2016). arXiv preprint. arXiv:1606.08415
17. Hong, Q., Siegel, J.W., Xu, J.: A priori analysis of stable neural network solutions to numerical PDEs (2021). arXiv preprint. arXiv:2104.02903
18. Jiao, Y., Lai, Y., Luo, Y., Wang, Y., Yang, Y.: Error analysis of deep Ritz methods for elliptic equations (2021). arXiv preprint. arXiv:2107.14478
19. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2014). arXiv preprint. arXiv:1412.6980
20. Lagaris, I.E., Likas, A., Fotiadis, D.I.: Artificial neural networks for solving ordinary and partial differential equations. IEEE Trans. Neural Netw. **9**(5), 987–1000 (1998)
21. Lee, H., Kang, I.S.: Neural algorithm for solving differential equations. J. Comput. Phys. **91**(1), 110–131 (1990)
22. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Math. Program. **45**(1), 503–528 (1989)
23. Lu, Y., Lu, J., Wang, M.: A priori generalization analysis of the deep Ritz method for solving high dimensional elliptic partial differential equations. In: Belkin, M., Kpotufe, S. (eds.) Proceedings of Thirty Fourth Conference on Learning Theory. Proceedings of Machine Learning Research, vol. 134, pp. 3196–3241. PMLR (2021)
24. Lu, Y., Chen, H., Lu, J., Ying, L., Blanchet, J.: Machine learning for elliptic PDEs: fast rate generalization bound, neural scaling law and minimax optimality (2021). arXiv preprint. arXiv:2110.06897
25. Luo, T., Yang, H.: Two-layer neural networks for partial differential equations: optimization and generalization theory (2020). arXiv preprint. arXiv:2006.15733
26. Müller, J., Zeinhofer, M.: Error estimates for the variational training of neural networks with boundary penalty (2021). arXiv preprint. arXiv:2103.01007
27. Pinkus, A.: Approximation theory of the MLP model in neural networks. Acta Numer. **8**, 143–195 (1999)
28. Ritz, W.: Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. J. Reine Angew. Math. **1909**(135), 1–61 (1909)
29. Růžička, M.: Nichtlineare Funktionalanalysis: Eine Einführung. Springer, Berlin (2006)
30. Sirignano, J., Spiliopoulos, K.: DGM: a deep learning algorithm for solving partial differential equations. J. Comput. Phys. **375**, 1339–1364 (2018)

31. Struwe, M.: Variational Methods, vol. 31999. Springer, Berlin (1990)
32. Takeuchi, J., Kosugi, Y.: Neural network representation of finite element method. Neural Netw. **7**(2), 389–395 (1994)
33. Wang, S., Teng, Y., Perdikaris, P.: Understanding and mitigating gradient flow pathologies in physics-informed neural networks. SIAM J. Sci. Comput. **43**(5), A3055–A3081 (2021)
34. Xu, J.: The finite neuron method and convergence analysis (2020). arXiv preprint. arXiv:2010.01458