Robust Variational Physics-Informed Neural Networks

Sergio Rojas [a]   Paweł Maczuga [b]   Judit Muñoz-Matute[c,d]   David Pardo[e,c,f]

Maciej Paszyński[b]

[a] Instituto de Matemáticas, Pontificia Universidad Católica de Valparaíso, Chile.
`sergio.rojas.h@pucv.cl`.
[b] AGH University of Krakow, Poland.
[c] Basque Center for Applied Mathematics (BCAM), Spain.
[d] Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, USA.
[e] University of the Basque Country (UPV/EHU), Spain.
[f] Ikerbasque, Spain.

May 21, 2024

Abstract

We introduce a Robust version of the Variational Physics-Informed Neural Networks method (RVPINNs). As in VPINNs, we define the quadratic loss functional in terms of a Petrov-Galerkin-type variational formulation of the PDE problem: the trial space is a (Deep) Neural Network (DNN) manifold, while the test space is a finite-dimensional vector space. Whereas the VPINN's loss depends upon the selected basis functions of a given test space, herein, we minimize a loss based on the discrete dual norm of the residual. The main advantage of such a loss definition is that it provides a reliable and efficient estimator of the true error in the energy norm under the assumption of the existence of a local Fortin operator. We test the performance and robustness of our algorithm in several advection-diffusion problems. These numerical results perfectly align with our theoretical findings, showing that our estimates are sharp.

Keywords: Robustness, Variational Physics-Informed Neural Networks, Petrov-Galerkin formulation, Riesz representation, Minimum Residual principle, a posteriori error estimation, quasi-minimizers.

1 Introduction

The remarkable success of Deep Learning (DL) algorithms across different scientific areas [28, 36, 27] in the last decade has recently led to explore the potential of this discipline to address classical problems in physics and mathematics. These problems include approximating the solutions of Partial Differential Equations (PDEs) employing (deep) Neural Networks (NN), which can accurately approximate continuous functions. The exponential growth of interest in these techniques started with the Physics Informed Neural Networks (PINNs) ([47]). This method incorporates the governing physical laws the PDE describes in the learning process. Then, the network is trained on a dataset that consists of a random selection of points in the physical domain and its boundary. PINNs have been successfully applied to solve a wide range of problems in scientific computing, including fluid mechanics [10, 41], wave propagation [3, 48], or inverse problems [16, 42], among many others. However, the loss function in PINNs is given by the strong form of the PDE residual, and it is known that in some problems with low regularity of the data, the solution only makes sense in a variational form. Therefore, PINNs fails to provide accurate solutions in those cases [35].

A natural continuation to PINNs addressing the aforementioned limitation is the so-called Variational PINNs (VPINNs) [31]. Here, the authors introduce a variational formulation of the underlying PDE within a Petrov-Galerkin framework. The solution is then approximated by a (deep) NN, whereas the test functions belong to linear vector spaces. Finally, the authors define a variational loss function to minimize during the training process. However, this methodology was not as popular as its predecessor. The main reason might be that the approach is sensible to the choice of the basis test functions. That is, given a discrete test space, the method's loss, stability, and robustness heavily depends upon the choice of the basis functions of the given test space. In [5], authors

1

present an a posteriori error analysis for discretizing elliptic boundary-value problems with VPINNs employing piecewise polynomials for the test space. As the loss function in VPINNs is, in general, not robust with respect to the true error (for example, the loss function can tend to zero even if the true error does not), they provide an error estimator employing classical techniques in finite element analysis to obtain practical information on the quality of the approximation. They also study in [6] the importance of employing appropriate quadrature rules and the degree of piecewise polynomial test functions to guarantee optimal convergence rates for VPINNs when performing mesh refinements. Finally, in [32], the same authors from VPINNs introduced *hp*-VPINNs. In the latter, they employ piecewise polynomials for testing and allow for *hp*-refinements via domain decomposition by selecting non-overlapping test functions. However, the aforementioned issue of adopting an appropriate loss function for each problem remains also in this approach.

The strategy we propose in this article overcomes the aforementioned limitation by following the core ideas introduced in Minimum Residual (MinRes) methods. The latter is a class of numerical methods for solving PDEs that guarantee stability. The spirit of MinRes methods is to minimize the dual norm of the residual of the PDE given in variational form. Many strategies have been developed based on this idea over the last five decades, including the families of Galerkin Least-Squares methods [7, 30, 29], First Order Least-squares [12, 13], residual minimization methods on dual discontinuous Galerkin norms [14, 18, 17, 50, 37], isogeometric residual minimization methods [15, 39, 38, 40], and Discontinuous Petrov-Galerkin (DPG) methods [20, 21, 22, 43, 49]. As the residual operator lives in the dual of the test space and the dual norm is difficult to compute, the use of the Riesz representation theorem [45] is natural in this context. The latter maps elements of the dual space into elements on the test space. Therefore, in many of the aforementioned methods, instead of minimizing the residual in the dual norm, they minimize its Riesz representative (which is an element of the test space) in a given test norm.

In this work, we revisit the initial work on VPINNs from Kharazmi et al. in [31], and we introduce a Robust version of Variational Physics-Informed Neural Networks (RVPINNs). As in VPINNs, we consider a Petrov-Galerkin formulation of the PDE, approximate the solution employing NNs, and introduce a finite-dimensional test space. Our goal is to minimize the residual in the discrete dual norm. Therefore, we introduce a single test function that is the Riesz representation of the weak residual over the discrete test space. We define an inner product and span the Riesz representative on a discrete basis so the resulting loss function will include the inverse of the Gram matrix corresponding to the selected inner product. We prove that the norm of such test function is efficient and reliable, i.e., the norm of the true error is bounded from below and above (up to an oscillation term) by the norm of the residual representative. Moreover, as we show in the numerical results, our strategy is insensitive (unlike VPINNs) to the choice of the discrete basis spanning the test space (assuming that the numerical integration and the inversion of the Gram matrix are sufficiently accurate). It is easy to see that minimizing the proposed loss functional is equivalent to minimizing the test norm of the Riesz representation of the residual, which is the idea behind classical MinRes methods. Summarizing, we provide a general mathematical framework to define robust loss functionals in VPINNs. Our strategy relies on two ideas: (a) the appropriate selection of the inner product in the test space, ensuring the stability of the variational formulation, and (b) the adoption of a single test function, that is the Riesz representation of the weak residual.

In particular, if we consider an orthonormal discrete basis with respect to the inner product in the test space, the Gram matrix becomes the identity, and we recover the original definition of the loss functional in VPINNs. The difference with classical VPINNs is that our strategy is robust and independent of the choice of the basis functions. Other works employ similar ideas closely related to minimum residual methods [11, 9]. In [2], the authors approximate the solution of symmetric and positive definite problems employing the discrete weak residual of the variational problem. For that, they define an adaptive construction of a sequence of finite-dimensional subspaces whose basis functions are realizations of a sequence of neural networks. In [53], the authors minimize the dual norm of the weak residual for symmetric variational formulations in H^1 and rectangular domains via a spectral decomposition. In [52], the same authors extend this method to time-harmonic Maxwell equations in the context of symmetric $H(curl)$-formulations. Also, in [54], in the context of Petrov-Galerkin formulations, the authors minimize the dual norm of the residual based on the concept of optimal testing form [22]. They approximate both the solution of the variational problem and the corresponding optimal test functions employing NNs by solving two nested deep Ritz problems. More recently, in [4], the authors interpolate the neural networks onto finite element spaces to represent the (partial) unknowns. This strategy overcomes, in particular, the challenges related to the imposition of boundary conditions and approximations of the NN derivatives during training.

Finally, apart from the usual limitations of DL-based technologies (optimizer, integration, etc.), the largest bottleneck of RVPINNs at the moment is that in certain configurations, we need to invert the Gram matrix corresponding to the basis and inner product of the test space. We will explore how to optimize this step in the future. However, there are particular cases where the inversion of the Gram matrix is trivial (as in the spectral case) or easy to compute. For example, if we consider a strong variational formulation, the test inner product is L^2, and

employing piecewise discontinuous polynomials for testing the Gram matrix becomes block diagonal. On the other hand, for parametric problems, the Gram matrix may remain the same for different values of the parameters. Therefore, the inversion can be done offline, being valid for a large class of problems. Parametric problems are essential to solve inverse problems [26], so RVPINNs could be of great interest in this area.

The article is organized as follows: In Section 2, we introduce the variational formulation of the model problem we consider in this article, the NN framework for approximation of the solution, and a brief overview of VPINNs. We also provide an alternative definition of VPINNs employing a single test function in the loss. Section 3 presents the methodology of RVPINNs and the connection with other methods based on VPINNs. Section 4 is devoted to the derivation of robust error estimates. We first prove that the residual representative is a reliable and efficient a posteriori error estimator of the true error in the sense of equivalent classes. Then, we demonstrate under the assumption of a local Fortin operator's existence that the true error is equivalent to the residual error estimator up to an oscillation term. Finally, we derive an *a priori* error estimate in the sense of quasi-optimizers. In Section 5 we test our method in several 1D and 2D advection-diffusion problems, showing the robustness of the approach. Finally, Section 6 summarizes the conclusions and future research lines.

2 Preliminaries

2.1 Abstract framework

Let U and V denote Hilbert spaces with norms $\|\cdot\|_U$ and $\|\cdot\|_V$. Assume that we are interested in obtaining an approximation of a PDE problem admitting a variational formulation of the form:

$$\text{Find } u \in U, \text{ such that: } r(u,v) := l(v) - a(u,v) = 0, \forall v \in V, \tag{1}$$

where $l(\cdot) \in V'$ is a bounded linear form, with V' denoting the dual space of V; $a(\cdot,\cdot)$ is a bounded inf-sup stable bilinear form in $U \times V$. That is, there exist constants $\mu, \alpha > 0$, respectively, such that:

$$a(w,v) \leq \mu \|w\|_U \|v\|_V, \quad \forall w \in U, v \in V, \tag{2}$$

and

$$\sup_{0 \neq v \in V} \frac{a(w,v)}{\|v\|_V} \geq \alpha \|w\|_U, \quad \forall w \in U. \tag{3}$$

We also assume that, for all $v \in V$, the operator $a(\cdot,v) \in U'$ satisfies:

$$a(w,v) = 0, \forall w \in U \implies v = 0. \tag{4}$$

Problem (1) admits a unique solution by means of the Banach-Nečas-Babŭska Theorem, and the following a priori estimate holds (see [23, Theorem 1.1]):

$$\|u\|_U \leq \frac{1}{\alpha} \|l(\cdot)\|_{V'}, \tag{5}$$

with

$$\|l(\cdot)\|_{V'} := \sup_{0 \neq v \in V} \frac{l(v)}{\|v\|_V}. \tag{6}$$

Additionally, it is well-known that (3) and (4) imply that the following adjoint inf-sup condition holds (see, e.g., [19, Theorem 1]):

$$\sup_{0 \neq w \in U} \frac{a(w,v)}{\|w\|_U} \geq \alpha \|v\|_V, \quad \forall v \in V. \tag{7}$$

Remark 1 (The weak residual). *For a given $w \in U$, we refer to $r(w,\cdot) \in V'$ (see (1)) as the weak residual.*

2.2 Neural Network framework

To numerically approximate (1), we consider a (Deep) Neural Network (DNN) function with input $\mathbf{x} = (x_1, \ldots, x_d)$ and output $u_\theta(\mathbf{x})$, where $\theta \in \mathbb{R}^S$ represents the trainable parameters. For simplicity, in this work, we employ a simple fully-connected feedforward Neural Network structure composed of L layers. Each layer l in the Neural Network consists of a set of neurons, which compute a weighted sum of their inputs plus a bias followed by a nonlinear activation function in the first $L-1$ layers and the identity function as the activation in the last layer. More precisely, the output of layer l, with $l = 1, \ldots L-1$, is given by:

$$\mathbf{z}^{(l)} = \sigma(\mathbf{w}^{(l)} \mathbf{z}^{(l-1)} + \mathbf{b}^{(l)}), \tag{8}$$

where σ is a nonlinear activation function (e.g., the tanh activation function), $\mathbf{w}^{(l)}$, $\mathbf{b}^{(l)}$ are the weights and biases respectively associated with the layer l, and $\mathbf{z}^{(0)} = \mathbf{x}$ is the input to the first layer. The final layer L is given by:

$$u_\theta = \mathbf{w}^{(L)}\mathbf{z}^{(L-1)} + \mathbf{b}^{(L)}. \tag{9}$$

Using an optimization algorithm, the Neural Network weights and biases are learned from a training set by minimizing a loss functional of input-output pairs.

We denote by U_{NN} the manifold consisting of all possible realizations for a given DNN architecture belonging to U, where U is defined in Section 2.1. That is,

$$U_{NN} := \{u_\theta, \forall \theta \in \mathbb{R}^S\} \cap U. \tag{10}$$

In the following, we assume $U_{NN} \neq \emptyset$. Finally, for a given discrete space $V_M \subseteq V$, we consider the following Petrov-Galerkin-type discretization of (1):

$$\text{Find } u_{NN} \in U_{NN}, \text{ such that: } r(u_{NN}, v_M) = 0, \forall v_M \in V_M. \tag{11}$$

Remark 2 (Existence of solutions for the Petrov-Galerkin-type discretization)**.** *The non-emptiness for U_{NN} is a necessary assumption to guarantee the existence of solutions to problem* (11)*, as an inadequate DNN structure setup could generate a manifold U_{NN} that is not contained in U. This is the case of a strong variational formulation set in $U \subset H^2(\Omega)$ (cf. Remark 11) with activation functions that are not sufficiently smooth (e.g., ReLU). However, we also notice that non-emptiness is not a sufficient condition to guarantee the existence of solutions, as they could be meaningful only in a dense sense (i.e., when $S \to +\infty$, with S denoting the number of trainable parameters). For instance, when considering smooth activation functions, but solutions are not smooth (cf. Section 5.5). The last can be circumvented by introducing the concept of quasi-minimizers, as we will see in Section 4.3.*

2.3 Variational Physics-Informed Neural Networks

Variational Physics Informed Neural Networks (VPINNs) approximate the solution u of problem (1) by minimizing a loss functional defined in terms of the weak residual (see Remark 1), and a discrete space $V_M \subseteq V$ with basis $\{\varphi_m\}_{m=1}^M$. That is, it approximates u by solving the following minimization problem[1]:

$$\text{Find } u_{\theta^*}, \text{ such that } \theta^* = \underset{\theta \in \mathbb{R}^S}{\text{argmin}} \, \mathscr{L}_r(u_\theta), \tag{12}$$

with $\mathscr{L}_r(u_\theta)$ being a loss functional defined in terms of the weak residual and the discrete basis. This approach allows to consider scenarios where the solution belongs to less regular spaces than in collocation PINNs. For example, elliptic problems with solutions in $H^{2-\varepsilon}(\Omega)$, with $\Omega \subset \mathbb{R}^2$ and $0 < \varepsilon \leq 1$.

A typical definition for the loss functional in (12) found in the literature is the following (see, e.g., [31, 33, 5, 6]):

$$\mathscr{L}_r(u_\theta) := \sum_{m=1}^M r(u_\theta, \varphi_m)^2 + C(u_\theta), \tag{13}$$

where $C(\cdot)$ is a V_M-independent non-negative-valued cost functional employed to impose data constraints, including boundary conditions, for u_{NN} that can be taken as $C(u_\theta) = 0$ if boundary conditions are strongly imposed in the DNN architecture, and no data for the interpolation is available.

Remark 3 (Classical approach)**.** *We will refer to problem* (12)*, with the loss functional defined by* (13)*, as the classical VPINNs approach.*

If we assume sufficient approximability of the DNN architecture and an ideal optimizer (solver), a minimizer for (13) for which the loss functional vanishes is also a solution to the Petrov-Galerkin-type problem (11). However, adopting such a loss functional may be computationally inefficient. Indeed, given the test space V_M, Eq. (13) depends heavily upon the choice of its basis. In particular, a simple re-scaling of one basis function (e.g., by making it sufficiently large) may easily lead to catastrophic results (cf. [53]). Consequently, there is a natural demand for robust loss definitions in practical VPINNs.

Remark 4 (Constraint assumption)**.** *We assume $C(\cdot)$ in* (13) *as a given cost functional since the present work focuses on the contribution in terms of the first term in* (13)*. The efficient implementation considering both contributions is a topic that requires further exploration.*

---

[1]Note that this problem may not admit solutions as a consequence of Remark 2

2.4 An alternative definition of VPINNs

To motivate how robust losses can be constructed, we first notice that, for a given set of trainable parameters θ, we can define the following function in V_M:

$$\widetilde{\varphi} = \sum_{m=1}^{M} r(u_\theta, \varphi_m)\,\varphi_m. \tag{14}$$

Then, as a consequence of the linearity of the weak residual, it holds (cf. [31]):

$$\mathscr{L}_r(u_\theta) = \sum_{m=1}^{M} r(u_\theta, \varphi_m)^2 + C(u_\theta) = r(u_\theta, \widetilde{\varphi}) + C(u_\theta). \tag{15}$$

The last equation suggests that VPINNs can be generalized in the following way. For a given $\widetilde{\varphi} \in V_M$, a VPINNs loss functional can be defined as:

$$\mathscr{L}_r(u_\theta) = r(u_\theta, \widetilde{\varphi}) + C(u_\theta). \tag{16}$$

The goal now is to define the loss functional as (16) in terms of a particular discrete test function $\widetilde{\varphi} \in V_M$, depending upon u_θ, such that VPINNs becomes robust. In the following Section, we present a strategy based on defining such a discrete function as the Riesz residual representative in V_M with respect to the norm inducing the inf-sup stability (3).

3 Robust Variational Physics-Informed Neural Networks

In Robust Variational Physics-Informed Neural Networks (RVPINNs), we first introduce an inner product $(\cdot, \cdot)_V$[2], whose associated norm satisfies the inf-sup stability condition (3). Next, for a given trainable parameter θ, we compute $\phi := \phi(\theta) \in V_M$ being the solution of the following Galerkin problem:

$$(\phi, \varphi_n)_V = r(u_\theta, \varphi_n), \text{ for } n = 1, \ldots, M, \tag{17}$$

and define the loss functional as:

$$\mathscr{L}_r^\phi(u_\theta) := r(u_\theta, \phi) + C(u_\theta), \tag{18}$$

where $C(\cdot)$ is a V_M-independent functional to impose constraints as in (13). Finally, we obtain the approximation of the solution of problem (1) by solving the following minimization problem (cf., [31]):

$$\text{Find } u_{\theta^*}, \text{ such that } \theta^* = \underset{\theta \in \mathbb{R}^S}{\operatorname{argmin}} \mathscr{L}_r^\phi(u_\theta), \tag{19}$$

when the DNN architecture allows for solutions for the associated Petrov-Galerkin-type problem (see Remark 2).

Remark 5 (Riesz representative of the weak residual). *The solution ϕ of problem* (17) *is the Riesz representative in V_M of the residual.*

As a consequence of the linearity of the weak residual and the linearity of the inner product with respect to the second variable, defining:

$$\phi := \sum_{m=1}^{M} \eta_m(\theta)\varphi_m, \tag{20}$$

problem (17) leads to the resolution of the following problem written in matrix form:

$$G\eta(\theta) = \mathscr{R}(\theta), \tag{21}$$

with $\eta(\theta)$ the vector of coefficients $\eta_m(\theta)$, G the (θ-independent) symmetric and positive definite Gram matrix of coefficients $G_{nm} = (\varphi_m, \varphi_n)_V$, and $\mathscr{R}(\theta)$ the vector of coefficients $\mathscr{R}_n(\theta) = r(u_\theta, \varphi_n)$, with $n = 1, \ldots, M$. Thus,

$$r(u_\theta, \phi) = \sum_{m=1}^{M} \eta_m(\theta) r(u_\theta, \varphi_m) = \sum_{m=1}^{M} \mathscr{R}_m(\theta)\eta_m(\theta) = \mathscr{R}(\theta)^T G^{-1} \mathscr{R}(\theta), \tag{22}$$

implying that the loss functional (18) is equivalently written as:

$$\mathscr{L}_r^\phi(u_\theta) = \mathscr{R}(\theta)^T G^{-1} \mathscr{R}(\theta) + C(u_\theta). \tag{23}$$

We also notice that, by definition, to minimize the loss functional (23) is equivalent, up to the constraint $C(u_\theta)$, to minimize the quantity $\|\phi\|_V^2$. Indeed, from (17), we deduce:

$$\|\phi\|_V^2 = (\phi, \phi)_V = r(u_\theta, \phi). \tag{24}$$

---

[2]i.e., $\|v\|_V^2 := (v, v)_V$, for all $v \in V$.

Remark 6 (Gram matrix inversion). *A practical implementation of RVPINNs requires the computation of $r(u_\theta, \phi)$ following (22). Computing explicitly G^{-1} could be computationally expensive in some scenarios, for instance, when considering standard FEM discrete test spaces defined in terms of piece-wise polynomials. Efficient inversions of the system (21) may require iterative solvers and preconditioners (see, e.g., [51]).*

In Section 4.2, we prove that $\|\phi\|_V$ is, up to the constraint term $C(u_\theta)$ and an oscillation term, a local robust error estimation for the error $\|u - u_\theta\|_U$ under the assumption of the existence of a local Fortin's operator.

3.1 Orthonormal discrete basis and relation with other VPINNs

When a test space V_M is the span of an orthonormal set $\{\varphi_m\}_{m=1}^M$ with respect to the $\|\cdot\|_V$-norm, the Gram matrix G coincides with the identity matrix; therefore, the corresponding residual representative has the form:

$$\phi = \sum_{m=1}^M r(u_\theta, \varphi_m) \varphi_m, \tag{25}$$

and the loss functional is explicitly written as

$$\mathscr{L}_r^\phi(u_\theta) = \sum_{m=1}^M r(u_\theta, \varphi_m)^2 + C(u_\theta). \tag{26}$$

Thus, the classical VPINNs loss functional definition (13) is recovered for this case. This is also the case of the recently proposed Deep Fourier Residual method (see [53, 52]), where authors first define the loss functional as the continuous dual norm of the weak residual, that later is approximated by considering a truncation of a series in terms of an orthonormal basis for the test space. In particular, in [53], authors consider diffusion problems and approximate the norm of the weak residual in H^{-1}. This turns out to be equivalent to (26) considering the standard H^1-norm for the space V, and the functions φ_m as orthonormalized sinusoidal (spectral) functions with respect to the H^1-norm.

4 Error estimates for RVPINNs

One of the main complexities for proving the robustness of the residual estimator $\|\phi\|_V$ is that the solution of the Petrov-Galerkin problem (11) may not have a solution or, if there exists, it may be non-unique since the space of all possible realizations of the NN architecture defines a manifold instead of a finite-dimensional space (see, e.g., [6, Section 6.3]). Thus, standard FEM arguments based on a discrete inf-sup condition cannot be applied in this context.

Nevertheless, we can derive a posteriori error estimates using a different strategy. First, we introduce in Section 4.1 an equivalence class that allows us to neglect the part of the error that is a-orthogonal to V_M. For that equivalence class, we prove that the residual representative always defines a reliable and efficient a posteriori estimator for the error. Then, in Section 4.2, we consider the case of the full error, for which we demonstrate its equivalence to the residual error estimator up to an oscillation term and under the assumption of the existence of a local Fortin operator.

4.1 A posteriori error estimates for RVPINNs in the sense of equivalence classes

We start by introducing the following Null space of the operator $A : U \mapsto V_M'$, defined in terms of the bilinear form $a(\cdot, \cdot)$ associated with problem (1) (cf. [19]):

$$U_M^0 := \{w \in U : \langle A(w), v_M \rangle := a(w, v_M) = 0, \forall v_M \in V_M\}, \tag{27}$$

and the following norm for the quotient space U/U_M^0:

$$\|[w]\|_{U/U_M^0} := \inf_{w_0 \in U_M^0} \|w + w_0\|_U. \tag{28}$$

We extend the definition of the bilinear form $a(\cdot, \cdot)$ to the product space $U/U_M^0 \times V$ as:

$$a([w], v) = a(w, v), \quad \text{with } w \in [w] \text{ being any arbitrary representative of the equivalence class.} \tag{29}$$

The following result follows as a consequence of the boundedness and inf-sup stability properties of the bilinear form $a(\cdot, \cdot)$ (see equations (2) and (3)), respectively:

Proposition 1. *The following boundedness and semi-discrete inf-sup conditions, respectively, are satisfied by the bilinear form* (29):

$$a([w], v_M) \leq \mu \|[w]\|_{U/U_M^0} \|v_M\|_V, \quad \forall [w] \in U/U_M^0, v_M \in V_M, \tag{30}$$

$$\sup_{0 \neq v_M \in V_M} \frac{a([w], v_M)}{\|v_M\|_V} \geq \alpha \|[w]\|_{U/U_M^0}, \quad \forall [w] \in U/U_M^0. \tag{31}$$

Proof. First, notice that (30) is direct from the definition (28) and the boundedness property of the bilinear form $a(\cdot, \cdot)$. Indeed, for all $w \in U$ and $w_0 \in U_M^0$, it holds:

$$a(w, v_M) = a(w + w_0, v_M) \leq \mu \|w + w_0\|_U \|v_M\|_V, \quad \forall v_M \in V_M.$$

Thus, (30) follows from taking the infimum. To prove (31), first notice that, for all $v_M \in V_M$, (7) implies:

$$\alpha \|v_M\|_V \leq \sup_{0 \neq w \in U} \frac{a(w, v_M)}{\|w\|_U} = \sup_{[0] \neq [w] \in U/U_M^0} \frac{a([w], v_M)}{\|w\|_U} \leq \sup_{[0] \neq [w] \in U/U_M^0} \frac{a([w], v_M)}{\|[w]\|_{U/U_M^0}},$$

where the last inequality follows from the fact that $\|[w]\|_{U/U_M^0} \leq \|w\|_U$, since $w_0 = 0 \in U_M^0$. Additionally, by construction, $\forall [w] \in U/U_M^0$,

$$a([w], v_M) = 0, \forall v_M \in V_M \implies [w] = [0]. \tag{32}$$

Therefore, (31) follows as a consequence of Theorem 1 in [19]. $\square$

Using the previous proposition, we can establish the main result of this section:

Theorem 2 (Error class bounds in terms of the residual representative). *Let $u \in U$ be the solution of the continuous problem* (1); *$u_\theta \in U_{NN}$ denote a DNN architecture with trainable parameters $\theta \in \mathbb{R}^S$; $V_M \subseteq V$ denote a finite dimensional space with norm $\|\cdot\|_V$; and $\phi \in V_M$ be the solution of problem* (17). *It holds:*

$$\frac{1}{\mu} \|\phi\|_V \leq \|[u - u_\theta]\|_{U/U_M^0} \leq \frac{1}{\alpha} \|\phi\|_V. \tag{33}$$

Proof. First notice that, by definition of ϕ and consistency of the analytical solution u, it holds:

$$\|\phi\|_V = \sup_{0 \neq v_M \in V_M} \frac{(\phi, v_M)_V}{\|v_M\|_V} = \sup_{0 \neq v_M \in V_M} \frac{r(u_\theta, v_M)}{\|v_M\|_V} = \sup_{0 \neq v_M \in V_M} \frac{a([u - u_\theta], v_M)}{\|v_M\|_V}.$$

Therefore, the left inequality in (33) is obtained by using the boundedness property (30). On another side, as a consequence of the inf-sup stability (31), it holds:

$$\alpha \|[u - u_\theta]\|_{U/U_M^0} \leq \sup_{0 \neq v_M \in V_M} \frac{a([u - u_\theta], v_M)}{\|v_M\|_V} = \sup_{0 \neq v_M \in V_M} \frac{a(u - u_\theta, v_M)}{\|v_M\|_V}$$

$$= \sup_{0 \neq v_M \in V_M} \frac{r(u_\theta, v_M)}{\|v_M\|_V}$$

$$= \sup_{0 \neq v_M \in V_M} \frac{(\phi, v_M)_V}{\|v_M\|_V}$$

$$= \|\phi\|_V,$$

proving the right inequality in (33). $\square$

Remark 7 (Robustness of the residual representative in the sense of classes). *Inequalities* (33) *show that, for any θ, $\|\phi\|_V$ always, up to some multiplicative constants, defines an efficient (lower) and a reliable (upper) bound (thus, robust) estimation for the error $\|[u - u_\theta]\|_{U/U_M^0}$. This, in practice, implies that to minimize $\|\phi\|_V^2 + C(u_\theta)$ is equivalent to minimize $\|[u - u_\theta]\|_{U/U_M^0}^2 + C(u_\theta)$. Moreover, if we assume that problem* (11) *admits a solution and consider $C(u_\theta) = 0$, $\|\phi\|_V \to 0^+$ implies that $u - u_\theta$ converges to a function that belongs to U_M^0 as a consequence of the robustness. Thus, since $u \notin U_M^0$, we conclude that u_θ converges to a function u_{θ^*} satisfying:*

$$0 = a(u - u_{\theta^*}, v_M) = r(u_{\theta^*}, v_M), \quad \forall v_M \in V_M.$$

Therefore, u_θ converges to a solution of the Petrov-Galerkin-type problem (11)*, as expected.*

4.2 Energy norm error estimates based on local semi-discrete inf-sup condition

Even if the a posteriori error estimates of the previous section only ensure that the RVPINNs definition of the loss functional is robust in the sense of equivalence classes, numerical experiments confirm that it can also be robust concerning the true error in the energy norm.

We first notice that, as a consequence of Theorem 2 and the quotient norm definition (28), the following result can be deduced:

Corollary 3 (Lower bound for the error in the energy norm). *Under the same hypothesis of Theorem 2, it holds:*

$$\frac{1}{\mu}\|\phi\|_V \le \|u - u_\theta\|_U. \tag{34}$$

Therefore, for a given θ, $\|\phi\|_V$ always defines an efficient bound for the error in the energy norm.

If we assume that the DNN architecture allows for solutions of the Petrov-Galerkin problem (11), we can also obtain a local reliable bound for $\|u - u_\theta\|_U$ through the following Assumption[3]:

Assumption 1 (Local Fortin's condition I). *There exists a parameter $\theta^* \in \mathbb{R}^S$, such that $u_{\theta^*} \in U_{NN}$ is a solution of the Petrov-Galerkin problem (11). Additionally, there exists $R > 0$ such that, for all $\theta \in B(\theta^*, R)$, there is an operator $\Pi_\theta : V \mapsto V_M$, and a θ-independent constant $C_\Pi > 0$, verifying:*

a) $a(u_\theta, v - \Pi_\theta v) = 0, \quad \forall v \in V$,

b) $\|\Pi_\theta v\|_V \le C_\Pi \|v\|_V, \quad \forall v \in V$,

where $B(\theta^*, R)$ denotes an open ball of center θ^* and radius R, with respect to a given norm of $\mathbb{R}^S$.

Proposition 4 (Upper bound of the error in the energy norm). *Under the same hypothesis of Theorem 2. If Assumption 1 is satisfied, it holds:*

$$\|u - u_\theta\|_U \le \frac{1}{\alpha}\operatorname{osc}(u) + \frac{1}{C_\Pi \alpha}\|\phi\|_V, \quad \forall \theta \in B(\theta^*, R), \tag{35}$$

with

$$\operatorname{osc}(u) := \sup_{0 \ne v \in V} \frac{a(u, v - \Pi_\theta v)}{\|v\|_V}.$$

Proof. If Assumption 1 is satisfied, it holds:

$$
\begin{aligned}
\alpha \|u - u_\theta\|_U &\le \sup_{0 \ne v \in V} \frac{a(u - u_\theta, v)}{\|v\|_V} = \sup_{0 \ne v \in V} \frac{a(u - u_\theta, v - \Pi_\theta v)}{\|v\|_V} + \sup_{0 \ne v \in V} \frac{a(u - u_\theta, \Pi_\theta v)}{\|v\|_V} \\
&\le \operatorname{osc}(u) + \frac{1}{C_\Pi} \sup_{0 \ne v \in V} \frac{a(u - u_\theta, \Pi_\theta v)}{\|\Pi_\theta v\|_V} \\
&\le \operatorname{osc}(u) + \frac{1}{C_\Pi} \sup_{0 \ne v_M \in V_M} \frac{a(u - u_\theta, v_M)}{\|v_M\|_V} \\
&= \operatorname{osc}(u) + \frac{1}{C_\Pi} \sup_{0 \ne v_M \in V_M} \frac{r(u_\theta, v_M)}{\|v_M\|_V} \\
&= \operatorname{osc}(u) + \frac{1}{C_\Pi} \sup_{0 \ne v_M \in V_M} \frac{(\phi, v_M)_V}{\|v_M\|_V} \\
&= \operatorname{osc}(u) + \frac{1}{C_\Pi}\|\phi\|_V,
\end{aligned}
$$

proving (35). $\qquad\square$

Remark 8 (Existence of a local Fortin's operator). *First, notice that, for a given θ, a Fortin operator's existence is only possible if $u_\theta \ne 0$ does not belong to the space U_M^0. Indeed, if $u_\theta \in U_M^0$, we have*

$$a(u_\theta, \Pi_\theta v) = 0, \text{ for all Fortin operator } \Pi_\theta,$$

---

[3]This Assumption is not standard; it is an adaptation of the classical Fortin's condition, commonly used in the discrete inf-sup stability analysis of mixed FEM formulations (see [8]) and, more recently, in the analysis of residual minimization-based FEMs, such as the DPG method (see [44]).

since $\Pi_\theta v \in V_M$. Thus, in such a case, the residual representative $\|\phi\|_V$ is only a reliable estimate in the spirit of Theorem 2. Nevertheless, it is expected that, in the minimization procedure, the NN solution approaches a solution of the Petrov-Galerkin problem (11) (see Remark 7), implying that the NN solution will not belong to the space U_M^0 in a further iteration. Therefore, it is meaningful to assume that, after a sufficient number of iterations, the nonlinear solver will reach a parameter θ belonging to a neighborhood of a parameter θ^*, in which θ^* is the local minimum, implying that a local semi-discrete inf-sup condition is satisfied, as stated in the following Lemma (cf. [25]).

Lemma 5 (Local Fortin's lemma). *If Assumption 1 is satisfied, then the following local semi-discrete inf-sup condition is satisfied:*

$$\frac{\alpha}{C_\Pi}\|u_\theta\|_U \leq \sup_{0\neq v_M\in V_M} \frac{a(u_\theta,v_M)}{\|v_M\|_V}, \quad \forall u_\theta \in U_{NN}^{\theta^*} := \text{span}\{u_\theta : \theta \in B(\theta^*,R)\}. \tag{36}$$

Proof. Using the properties of Fortin's operator and the linearity of the bilinear form $a(\cdot,\cdot)$ with respect to the first variable, it holds:

$$\sup_{0\neq v_M\in V_M} \frac{a(u_\theta,v_M)}{\|v_M\|_V} \geq \sup_{0\neq v\in V} \frac{a(u_\theta,\Pi_\theta v)}{\|\Pi_\theta v\|_V} = \sup_{0\neq v\in V} \frac{a(u_\theta,v)}{\|\Pi_\theta v\|_V} \geq \frac{1}{C_\Pi}\sup_{0\neq v\in V} \frac{a(u_\theta,v)}{\|v\|_V} \geq \frac{\alpha}{C_\Pi}\|u_\theta\|_U.$$

$\square$

Finally, with the help of the previous Lemma, we can prove the following local a priori error estimate:

Proposition 6 (Local a priori error estimate I). *If Assumption 1 is satisfied, it holds:*

$$\|u-u_{\theta^*}\|_U \leq \left(1+\frac{\mu C_\Pi}{\alpha}\right) \inf_{u_\theta\in U_{NN}^{\theta^*}} \|u-u_\theta\|_U. \tag{37}$$

Proof. First notice that, for all $u_\theta \in U_{NN}^{\theta^*}$, it holds

$$\frac{\alpha}{C_\Pi}\|u_{\theta^*}-u_\theta\|_U \leq \sup_{0\neq v_M\in V_M} \frac{a(u_{\theta^*}-u_\theta,v_M)}{\|v_M\|_V} = \sup_{0\neq v_M\in V_M} \frac{a(u-u_\theta,v_M)}{\|v_M\|_V} \leq \mu\|u-u_\theta\|_U.$$

Therefore, the result (37) follows from the triangular inequality

$$\|u-u_{\theta^*}\|_U \leq \|u-u_\theta\|_U + \|u_{\theta^*}-u_\theta\|_U.$$

$\square$

Remark 9 (Fortin's lemma converse). *Following [24], it can be proved that the converse of Lemma 5 is also valid. This could be useful in scenarios where proving the local semi-discrete inf-sup condition is simpler than proving the existence of a Fortin operator. Note that the continuity constant of the Fortin operator is a measure of the loss of stability. Ideally, we would like to find a Fortin operator which is indeed an orthogonal projection, i.e., $C_\Pi = 1$. However, the construction of a Fortin operator [44] depends upon the formulation, the spaces, and the discretization. The construction of local Fortin operators for formulations involving NNs are out of the scope of this article.*

Remark 10 (Norm for the discrete space). *For the sake of simplicity, we assume that the norm for the discrete space V_M coincides with the norm $\|\cdot\|_V$, of the space V. However, under certain conditions, it could be convenient to consider a different norm for the space V_M being equivalent to $\|\cdot\|_V$. For instance, when properties of the discrete test allow obtaining a better inf-sup constant for the local semi-discrete inf-sup condition. In such a case, previous results are straightforwardly extended to this scenario.*

4.3 Error estimates in the quasi-minimizer sense

In the previous section, we derived a priori and a posteriori error estimates near the convergence regime by assuming the existence of a local solution of the Petrov-Galerkin-type problem (11). However, as mentioned in Remark (2), this assumption cannot be satisfied in several practical scenarios. Additionally, we notice that our a priori error estimate does not take into consideration the contribution of the constraint $C(\cdot)$, since it is not necessarily true that a solution u_{θ^*} of the corresponding Petrov-Galerkin-type problem will also satisfy that $C(u_{\theta^*})$ will be bounded (up to a constant) by the infimum of $C(u_\theta)$. This is the case when $C(\cdot)$ is defined in terms of noisy data and the DNN architecture allows for functions u_θ satisfying $C(u_\theta) = 0$. Thus, a more general framework allowing for a less restrictive assumption must be considered. To this end, following [9], we start by considering the following definition:

Definition 1 (Quasi-minimizers). *Let $\mathscr{L} : U \mapsto \mathbb{R}$ be a cost functional. Let $\delta_S > 0$ and $U_{NN} \subset U$ be the manifold introduced in (10), represented by a set $\theta \in \mathbb{R}^S$ of trainable parameters. A function $u_{\theta_S} \in U_{NN}$ is said to be a quasi-minimizer of $\mathscr{L}$ if the following holds true:*

$$\mathscr{L}(u_{\theta_S}) \leq \inf_{\theta \in \mathbb{R}^S} \mathscr{L}(u_{\theta}) + \delta_S. \tag{38}$$

The following is a direct consequence of the RVPINN cost function being positive-valued (cf. Eq. (24)).

Proposition 7. *For a given DNN structure U_{NN}, there exists a sufficiently small $\delta^* \geq 0$ and a function $u_{\theta^*} \in U_{NN}$ such that, for all $\delta_S > \delta^*$, u_{θ^*} is a quasi-minimizer of the cost functional (18).*

The following Lemma is a direct consequence of the quasi-minimizer definition and the property of the Riesz representative (17) (cf. proof of Theorem 2):

Lemma 8. *Let u_{θ^*} be a quasi-minimizer and $\delta^* > 0$ be the smallest constant satisfying (38). It holds*

$$\left(\sup_{0 \neq v_M \in V_M} \frac{a(u - u_{\theta^*}, v_M)}{\|v_M\|_V} \right)^2 + C(u_{\theta^*}) \leq \left(\sup_{0 \neq v_M \in V_M} \frac{a(u - u_{\theta}, v_M)}{\|v_M\|_V} \right)^2 + C(u_{\theta}) + \delta^*, \quad \forall u_{\theta} \in U_{NN}. \tag{39}$$

We can now derive an a priori error estimate in the quasi-minimizer sense by considering the following relaxation of the Fortin's Assumption 1[4].

Assumption 2 (Local Fortin's condition II). *Given $\delta_S \geq 0$ such that $u_{\theta_S} \in U_{NN}$ is a quasi-minimizer of (18), there exists $R > 0$ such that, for all $\theta \in B(\theta_S, R)$, an operator $\Pi_{\theta} : V \mapsto V_M$, and a θ-independent constant $C_{\Pi} > 0$, verifying the same properties of Assumption 1.*

Proposition 9 (Local a priori error estimate II). *Let u_{θ^*} be a quasi-minimizer and $\delta^* > 0$ be the smallest constant satisfying (38). If Assumption 2 is satisfied, it holds:*

$$\|u - u_{\theta^*}\|_U^2 + \left(\frac{\sqrt{2}C_{\Pi}}{\alpha} \right)^2 C(u_{\theta^*}) \leq \inf_{u_{\theta} \in U_{NN}^{\theta^*}} \left\{ 2 \left(1 + \left(\frac{2\mu C_{\Pi}}{\alpha} \right)^2 \right) \|u - u_{\theta}\|_U^2 + \left(\frac{\sqrt{2}C_{\Pi}}{\alpha} \right)^2 \left(C(u_{\theta}) + \delta^* \right) \right\}. \tag{40}$$

Proof. The proof is similar as in Proposition 6 by noticing that, for all $u_{\theta} \in U_{NN}^{\theta^*}$, with $U_{NN}^{\theta^*}$ defined as in (36), it holds

$$\left(\frac{\alpha}{2C_{\Pi}} \right)^2 \|u_{\theta^*} - u_{\theta}\|_U^2 + \frac{1}{2} C(u_{\theta^*}) \leq \frac{1}{4} \left(\sup_{0 \neq v_M \in V_M} \frac{a(u_{\theta^*} - u_{\theta}, v_M)}{\|v_M\|_V} \right)^2 + \frac{1}{2} C(u_{\theta^*})$$

$$\leq \frac{1}{4} \left(\sup_{0 \neq v_M \in V_M} \frac{a(u - u_{\theta}, v_M)}{\|v_M\|_V} + \sup_{0 \neq v_M \in V_M} \frac{a(u - u_{\theta^*}, v_M)}{\|v_M\|_V} \right)^2 + \frac{1}{2} C(u_{\theta^*})$$

$$\leq \left(\sup_{0 \neq v_M \in V_M} \frac{a(u - u_{\theta}, v_M)}{\|v_M\|_V} \right)^2 + \frac{1}{2} \left(C(u_{\theta}) + \delta^* \right),$$

where the last inequality follows from the Young's inequality $2xy \leq x^2 + y^2, \forall x, y \in \mathbb{R}$, and Lemma 8. $\square$

5 Numerical examples

To illustrate the proposed strategy's performance, we consider several diffusion-advection problems subject to Dirichlet-type boundary conditions. We focus on problems that either are challenging for the standard finite element method (advection-dominated problems) or do not admit a solution belonging to $H^2(\Omega)$. We consider two distinct discrete test spaces to construct the loss functionals: one employing standard FE piece-wise polynomials and another employing spectral orthonormal test functions. Here, we do not intend to compare the performance between different test spaces, but rather, we aim to numerically validate the robustness of RVPINNs.

---

[4]This is indeed a relaxation of Assumption 1, since in the case where there exists a solution of the Petrov-Galerkin-type problem (11) and $C(u_{\theta}) = 0$, we can consider $\delta_S = 0$ and $\theta_S = \theta^*$.

5.1 Diffusion-advection model problem

For $d \geq 1$, given a bounded and open Lipschitz polyhedra $\Omega \in \mathbb{R}^d$ with boundary $\partial \Omega$, this work considers the following linear diffusion-advection problem:

Find u, such that:

$$
\begin{aligned}
-\nabla \cdot (\varepsilon \nabla u) + \beta \cdot \nabla u &= f, \quad \text{in } \Omega, \\
u &= 0, \quad \text{on } \partial \Omega,
\end{aligned}
\tag{41}
$$

where $\varepsilon > 0$ denotes a diffusive coefficient term, $\beta \in [L^{\infty}(\Omega)]^d$ is a divergence-free (i.e., $\text{div}\beta \equiv 0$) advective coefficient function. Finally, $f \in V'$ is a given source term, with $V' := H^{-1}(\Omega)$ being the dual space of $V := H_0^1(\Omega)$. Problem (41) admits the following continuous variational formulation:

$$
\text{Find } u \in U := V : r(u,v) := l(v) - a(u,v), \forall v \in V,
\tag{42}
$$

with

$$
a(u,v) := (\varepsilon \nabla u, \nabla v)_0 + (\beta \cdot \nabla u, v)_0, \quad \text{and} \quad l(v) = \langle f, v \rangle,
\tag{43}
$$

where $(\cdot, \cdot)_0$ denotes the L^2-inner product[5], and $\langle \cdot, \cdot \rangle$ denotes the duality map between V' and V, coinciding with $(f, v)_0$ when f belongs to $L^2(\Omega)$.

Let us equip the Hilbert spaces U, V with the norms

$$
\| \cdot \|_U^2 := \| \cdot \|_V^2 := \varepsilon (\nabla \cdot, \nabla \cdot)_0.
\tag{44}
$$

Using the Cauchy-Schwarz inequality, the following boundedness estimate is proved (cf. (2)):

$$
a(w,v) \leq \left(1 + \frac{C_{\Omega} \|\beta\|_{\infty}}{\varepsilon} \right) \|w\|_U \|v\|_V, \quad \forall w \in U, v \in V,
\tag{45}
$$

where C_{Ω} denotes the Poincaré constant (see [46, 1]). Moreover, as a consequence of the fact that (recall that $\text{div}\beta \equiv 0$):

$$
(\beta \cdot \nabla v, v)_0 = 0, \quad \forall v \in V,
\tag{46}
$$

a straightforward verification shows that the following coercive estimate holds:

$$
a(v,v) \geq \|v\|_V^2, \quad \forall v \in V.
\tag{47}
$$

Notice also that the coercive property (47) implies that the following inf-sup condition is satisfied (cf. (3)):

$$
\sup_{0 \neq v \in V} \frac{a(w,v)}{\|v\|_V} \geq \|w\|_U, \quad \forall w \in U.
\tag{48}
$$

Remark 11 (Alternative Variational Formulations). *With the spirit of considering a favorable scenario for spectral and FEM test spaces, we consider a standard weak variational formulation with the same test and trial spaces. However, we recall that the strategy is general, allowing for taking into account alternative formulations as a starting point. As an illustrative example, consider the pure diffusive particular case of problem (41) (i.e., with $\varepsilon = 1$, $\beta = \mathbf{0}$). Then, the following two alternative variational formulations of the form (42) are admissible:*

a) Strong VF: $U = H^2(\Omega) \cap H_0^1(\Omega)$, $V = L^2(\Omega)$, $a(u,v) = -(\Delta u, v)_0$, $l(v) = \langle f, v \rangle$.

b) Ultraweak VF: $U = L^2(\Omega)$, $V = H^2(\Omega) \cap H_0^1(\Omega)$, $a(u,v) = -(u, \Delta v)_0$, $l(v) = \langle f, v \rangle$.

Moreover, it can be proved that conditions (2)-(3) are satisfied by endowing the Hilbert space $H^2(\Omega) \cap H_0^1(\Omega)$ with the norm $(\Delta \cdot, \Delta \cdot)_0^{1/2}$, and $L^2(\Omega)$ with its standard norm.

5.2 A weak BCs imposition for 1D problems

Setting $\Omega = (-1,1)$, $U = H^1(\Omega)$, $V = H_0^1(\Omega)$, and $\widehat{u} = (u(-1), u(1))$, the variational formulation is defined as follows[6]

$$
\text{Find } u \in U : a(u,(v,\widehat{v})) := (\varepsilon u', v')_0 + (\beta u', v)_0 + \widehat{u} \cdot \widehat{v} = l((v,\widehat{v})) := \langle f, v \rangle, \forall (v,\widehat{v}) \in V \times \mathbb{R}^2.
\tag{49}
$$

---

[5]We adopt the same notation for the scalar and vectorial L^2-inner products for the sake of simplicity.

[6]Notice that this way of imposing BCs can also be applied to other alternative variational formulations. For instance, the Strong VF in the spirit of Remark 11.

Variational formulation (49) is well-posed. Indeed, defining $\|\cdot\|_V^2 := \varepsilon\,('\,,\,')_0$ and denoting by $\|\cdot\|_E$ the Euclidean norm of $\mathbb{R}^2$, we can consider the norms

$$\|u\|_U^2 := \|u\|_V^2 + \|\widehat{u}\|_E^2 \quad \text{and} \quad \|(v,\widehat{v})\|_{V\times\mathbb{R}^2}^2 := \|v\|_V^2 + \|\widehat{v}\|_E^2, \tag{50}$$

for the trial and test space, respectively. A straightforward verification proves that the boundedness condition (2) is satisfied with the upper bound constant μ of Eq. (45). It is also immediate to verify that the condition (4) is satisfied. Indeed, noticing that $(\beta v', v)_0 = 0$, for all $v \in V$ (cf. Eq. (46)), if $a(u,(v,\widehat{v})) = 0$ holds for all $u \in U$, then it holds for $u = v + u_{\widehat{v}}$, with the last being the (unique) solution of the problem:

$$\text{Find } u_{\widehat{v}} \in U : -\varepsilon u_{\widehat{v}}'' + \beta u_{\widehat{v}}' = 0, \quad u_{\widehat{v}}(-1) = \widehat{v}(1), \quad u_{\widehat{v}}(1) = \widehat{v}(2),$$

implying that $\|(v,\widehat{v})\|_{V\times\mathbb{R}^2}$ must be equal to zero. Finally, we can prove that the inf-sup condition (3) is satisfied with $\alpha = 1$ by first noticing that it is trivially satisfied for all $u \in V$. In addition, for a given u such that $\widehat{u} \neq 0$, given $v = u - u_d$ with u_d any lifting of the boundary data $\widehat{u}$, and $\widehat{v} = \widehat{u} - \eta \dfrac{\widehat{u}}{\|\widehat{u}\|_E^2}$, with $\eta = -(\varepsilon u', u_d')_0 + (\beta u', u - u_d)_0$, it holds

$$\frac{a(u,(v,\widehat{v}))}{\|(v,\widehat{v})\|_{V\times\mathbb{R}^2}} = \|u\|_U.$$

The result follows from noticing that the supreme over $V \times \mathbb{R}^2$ is always larger or equal to the left-hand side of the above equality.

With this alternative definition, it can be proved that the corresponding loss functional is given by

$$\mathcal{L}_r^\phi(u_\theta) = \|\phi\|_V^2 + |u_\theta(-1)|^2 + |u_\theta(1)|^2 + C(u_\theta), \tag{51}$$

with ϕ being the residual representative in $V_M \subseteq V$ with residual $r(u,\cdot)$ defined as in the variational formulation (42). Indeed, choosing the canonical basis for $\mathbb{R}^2$, the corresponding (block-diagonal) Gram matrix $\widehat{G}$ associated to the inner product inducing the norm for the space $V \times \mathbb{R}^2$, and the rhs $\widehat{\mathcal{R}}(\theta)$, respectively, have the following structure

$$\widehat{G} = \begin{pmatrix} G & \\ & I_2 \end{pmatrix} \quad \text{and} \quad \widehat{\mathcal{R}}(\theta) = \begin{pmatrix} \mathcal{R}(\theta) \\ -u_\theta(-1) \\ -u_\theta(1) \end{pmatrix}, \tag{52}$$

with I_2 denoting the 2×2 identity matrix, G the Gram matrix associated with the inner product of V, and $\mathcal{R}(\theta)$ the rhs associated with the residual $r(u_\theta,\cdot) := \langle f,\cdot\rangle - (\varepsilon u',\cdot')_0 - (\beta u',\cdot)_0$ in V'.

Remark 12 (Penalized weak BCs imposition). *We notice that it is also possible to consider a penalization for the weak imposition by multiplying the term $\widehat{u} \times \widehat{v}$ with a (possibly boundary-dependent) positive penalization term. This could serve as a further stabilization strategy in complex scenarios. For instance, in the advection-dominated regime.*

5.3 Discrete setting

For the sake of simplicity, in all the numerical examples we assume that there is no data available for interpolation; thus, we consider $C(u_\theta) = 0$. For one-dimensional problems, we set $\Omega := (-1,1)$ and define two kinds of test spaces. First, we consider V_M as FE space of the standard globally continuous and piece-wise linear functions that belong to $H_0^1(\Omega)$, defined over a uniform mesh partition of Ω into $M+1$ elements $[x_i, x_{i+1}]$, with $x_0 = -1 < x_1 < \cdots < x_M < x_{M+1} = 1$. Similarly, we consider a 2D case where $\Omega := (0,1)^2$. Here, we employ tensor-product piece-wise linear FEM test functions. The corresponding Gram matrix in both cases is constructed using the inner product, inducing the norm $\|\cdot\|_V$, defined in (44). Herein, we compute G^{-1} explicitly in both 1D and 2D examples employing FEM test spaces.

The second test space is intended to show examples involving an orthonormal basis for the discrete space V_M. Here, we normalize with respect to the inner product (44) the first M eigenfunctions of the Laplacian satisfying the Dirichlet homogeneous boundary conditions. The m-th eigenfunction and the corresponding orthonormal set, respectively, are given by

$$s_m := \sin\left(m\pi\frac{x+1}{2}\right), \qquad E_M = \left\{\varphi_m := \frac{2s_m}{\sqrt{\varepsilon}\pi m}\right\}_{m=1}^M. \tag{53}$$

Therefore, the loss functional (26) is explicitly given as:

$$\mathcal{L}_r^\phi(u_\theta) = \frac{4}{\varepsilon\pi^2} \sum_{m=1}^M \frac{1}{m^2} r(u_\theta, s_m)^2. \tag{54}$$

We also consider a two-dimensional numerical example involving an orthonormal basis for the discrete test space. In such a case, we set $\Omega := (0,1)^2$ and construct the first M and N eigenfunctions in the x and y coordinates, respectively. Here, the m,n-th eigenfunction and the corresponding orthonormal set, respectively, are given by

$$s_{m,n} := \sin(m\pi x)\sin(n\pi y), \qquad E_{MN} = \left\{ \varphi_{m,n} := \frac{s_{m,n}}{\sqrt{\varepsilon}\pi(m+n)} \right\}_{m=1,n=1}^{M,N}. \tag{55}$$

Thus, the loss function (26) is explicitly given as:

$$\mathscr{L}_r^\phi(u_\theta) = \frac{1}{\varepsilon\pi^2} \sum_{n=1}^{N} \sum_{m=1}^{M} \frac{1}{n^2+m^2} r(u_\theta, s_{m,n})^2. \tag{56}$$

For the architecture of the DNN approximator u_θ, we adopt the following structure for the 1D problems: a fixed feed-forward fully connected network comprising five layers, each with 25 neurons. Throughout all layers, the activation function employed is tanh. When strongly enforcing the homogeneous Dirichlet boundary conditions, we multiply the last layer's output by $(x+1)(x-1)$. Another version includes the weak imposition of homogeneous Dirichlet boundary conditions by adding to the functional loss the following quantity (see Section 5.2):

$$|u_\theta(-1)|^2 + |u_\theta(1)|^2. \tag{57}$$

As our choice for the nonlinear solver, we opt for the ADAM optimizer (refer to [34]) and initialize it with a learning rate of 0.0005. Our nonlinear solver undergoes up to 6000 iterations (epochs). To approximate the integral terms appearing in the loss functional definitions of the 1D problems, we employ (a) a fixed Gaussian quadrature rule with five nodes per element when considering FE functions, and (b) a trapezoidal rule with 4000 equally-distributed nodes when considering spectral functions. Lastly, the error $|u-u_\theta|_U$ is numerically estimated using a trapezoidal rule, with 10000 equally-distributed nodes. For the 2D example, we consider a fixed feed-forward fully connected network comprising four layers, each with 40 neurons and the same type of activation function as in the 1D cases. For the training in the spectral case, 500×500 equally-distributed integration points, while equally-distributed 1000×1000 integration points for estimating the error $|u-u_\theta|_U$. We strongly impose the Dirichlet boundary conditions by multiplying the last layer's output by $x(x-1)y(y-1)$.

5.4 Smooth diffusion problems

We first consider a simple diffusion problem with a smooth solution. More precisely, setting $\Omega = (-1,1)$, we consider the variational problem (42) with $\varepsilon = 1$, $\beta = 0$ and forcing term f defined in such a way that the analytical solution is given by

$$u(x) = x\sin(\pi(x+1)).$$

Figure 1 shows the numerical results obtained with RVPINNs with 50 sinusoidal test functions and strong imposition of the BCs. We observe a good performance of the RVPINNs strategy, obtaining an accurate approximation of the analytical solution, as shown in Figure 1a. Moreover, we observe a perfect match between the quantities $\|u-u_\theta\|_U$, $\|\phi\|_V$, and $\sqrt{\mathscr{L}_r^\phi(u_\theta)}$ as shown in figures 1b and 1c where, in particular, the last one shows a perfect correlation between the true error and the square root of the loss functional. Here, we compute $\|\phi\|_V$ employing the right-hand-side of (22). Note that $\|\phi\|_V$ and $\sqrt{\mathscr{L}_r^\phi(u_\theta)}$ coincide when strongly imposing BCs. Figure 2 shows an almost identical behavior when considering 100 FE test functions and strong imposition of BCs. We obtain similar results when considering the weak imposition of BCs, but we omit them here for brevity.

Note that in the case of pure diffusive problems, the continuity constant μ and the inf-sup stability constant α are equal to one when considering the H^1-seminorm (cf. (45) and (48)). Therefore, $\|\phi\|_V$ always defines a lower bound for the error (cf. (34)) and an upper bound up to a constant close to one if the oscillation term in (35) is sufficiently small. These theoretical considerations are confirmed in figures 1 and 2.

Finally, to show the effect of the oscillation term (35), Figure 3 shows the results obtained with only five test FE basis functions. Figure 3b reveals that the estimation is not robust after (approximately) 100 iterations and only becomes a lower bound, as the DNN has reached (up to implementation precision) a solution of the Petrov-Galerkin problem (11).

To conclude this section, we show the results obtained for a 2D pure diffusion example, set in the domain $\Omega = (0,1)^2$ analogously so the analytical solution coincides with
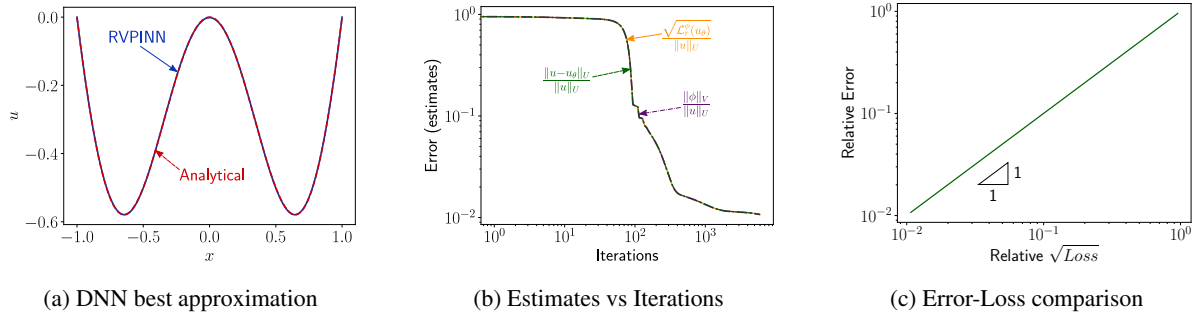
$$u(x,y) = \sin(\pi x)\sin(\pi y). \tag{58}$$

(a) DNN best approximation   (b) Estimates vs Iterations   (c) Error-Loss comparison

Figure 1: RVPINNs approximation of the smooth diffusion problem; strong BCs imposition, 50 spectral test functions.

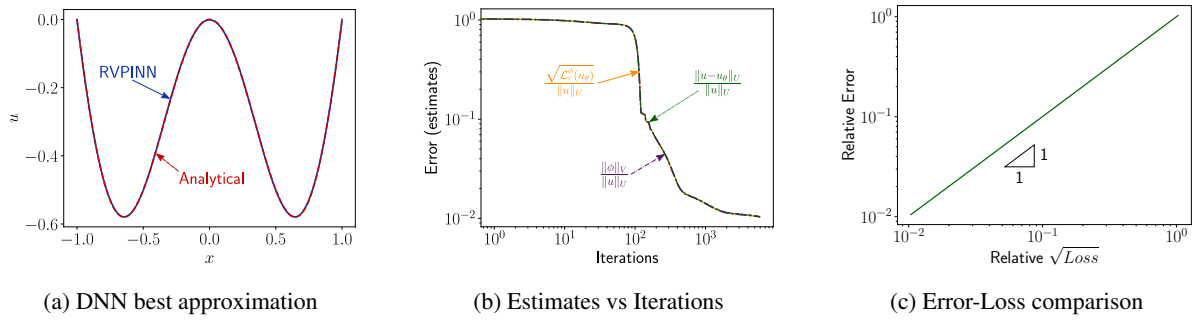(a) DNN best approximation   (b) Estimates vs Iterations   (c) Error-Loss comparison

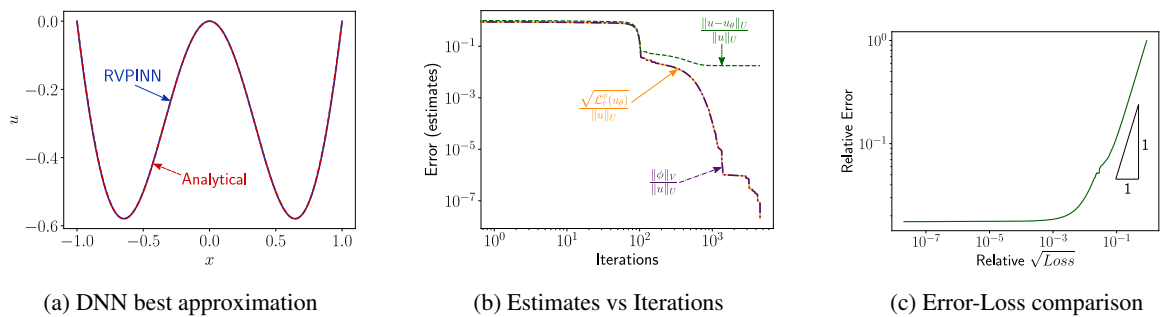Figure 2: RVPINNs approximation of the smooth diffusion problem; strong BCs imposition, 100 FE test functions.

(a) DNN best approximation   (b) Estimates vs Iterations   (c) Error-Loss comparison

Figure 3: RVPINNs approximation of the 1D smooth diffusion problem; strong BCs imposition, 5 FE test functions.

14

(a) $u - (\text{best}) \, u_\theta$ (b) Estimates vs Iterations   (c) Error-Loss comparison
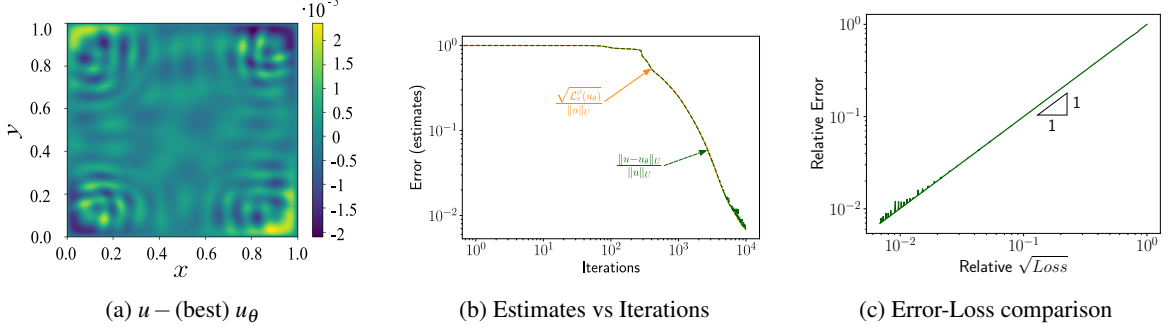
Figure 4: RVPINNs approximation of the 2D smooth problem; strong BCs imposition, 20×20 spectral test functions.

(a) $u - (\text{best}) \, u_\theta$ (b) Estimates vs Iterations   (c) Error-Loss comparison
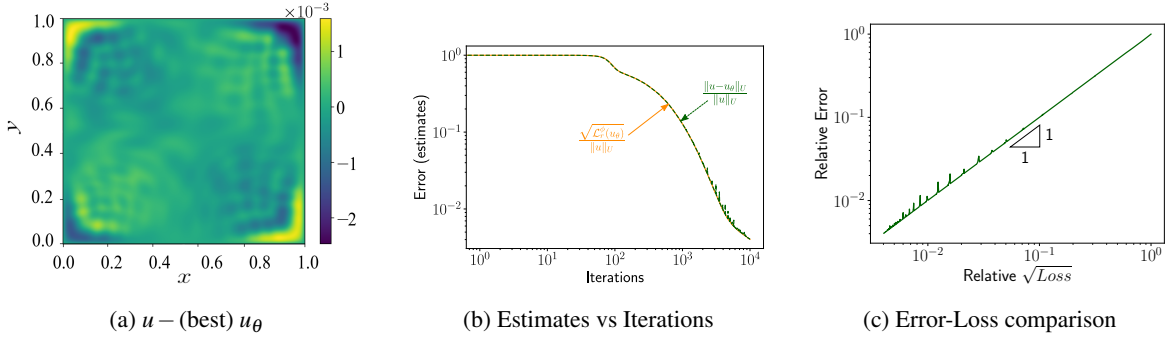
Figure 5: RVPINNs approximation of the 2D smooth problem; strong BCs imposition, 100×100 elements mesh with piece-wise linear FEM test functions.

Figure 4 displays the results when we consider the spectral loss functional (55) during training with $M = N = 20$ and strong imposition of the BCs. Similarly, Figure 5 displays the results when considering a 100×100-element mesh with piece-wise linear FEM test functions. We observe a similar behavior, up to some oscillation due to the inexact integration, to the 1D case, showing in particular that the strategy is dimension-independent, as expected.

To showcase the behavior of RVPINNs in extreme scenarios, in the following, we will consider 1D challenging problems for the particular setting described in Section 5.3.

5.5 Delta source problem

We consider again the pure 1D diffusion problem (i.e., eq. (42) with $\varepsilon = 1$ and $\beta = 0$), defined over $\Omega = (-1,1)$ and subject to homogeneous Dirichlet BCs, but now with a Dirac delta source $\delta_{1/2} \in H^{-1}(\Omega)$, that is explicitly defined through the action:

$$l(v) := \langle \delta_{1/2}, v \rangle := v(1/2), \quad \forall v \in H_0^1(\Omega). \tag{59}$$

The analytical solution to this problem is explicitly given as:

$$u(x) = \begin{cases} \dfrac{1}{4}(x+1), & -1 \leq x \leq \dfrac{1}{2}, \\ \dfrac{3}{4}(1-x), & \dfrac{1}{2} < x \leq 1, \end{cases}$$

which is only a $C^0(\Omega)$ solution, while the activation function tanh is smooth. Therefore, a major dominance of the oscillation term is expected, even when considering a sufficiently large number of test functions, as appreciated in Figure 6 where we have considered 100 FE test functions and strong imposition of the BCs. Figure 7 shows similar results for the same configuration but with the weak imposition of the BCs. In both cases, a good representation of the analytical solution is obtained despite being approximated with a smooth function. Additionally, figures 6 and 7 show that the loss functional defines a robust estimation of the error in the pre-asymptotic regime (before the dominance of the oscillation term). A similar behavior is obtained when considering the loss functional with spectral test functions– omitted here for brevity.
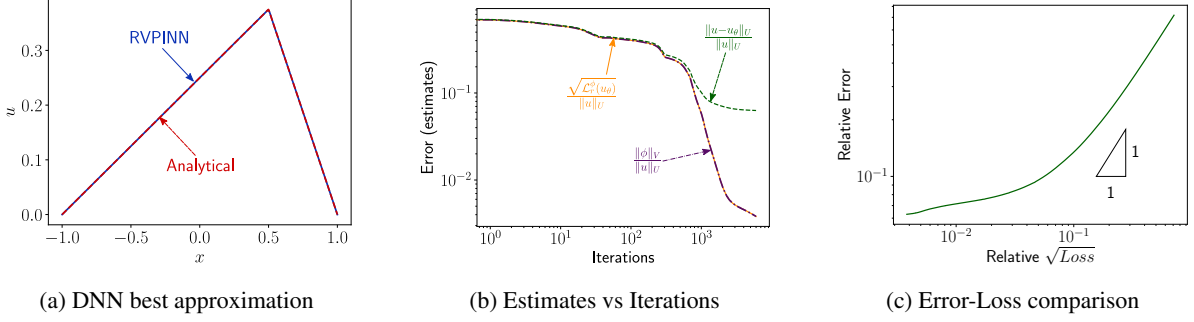
15

(a) DNN best approximation   (b) Estimates vs Iterations   (c) Error-Loss comparison

Figure 6: RVPINNs approximation of the delta source diffusion problem; strong BCs imposition, 100 FE test functions.

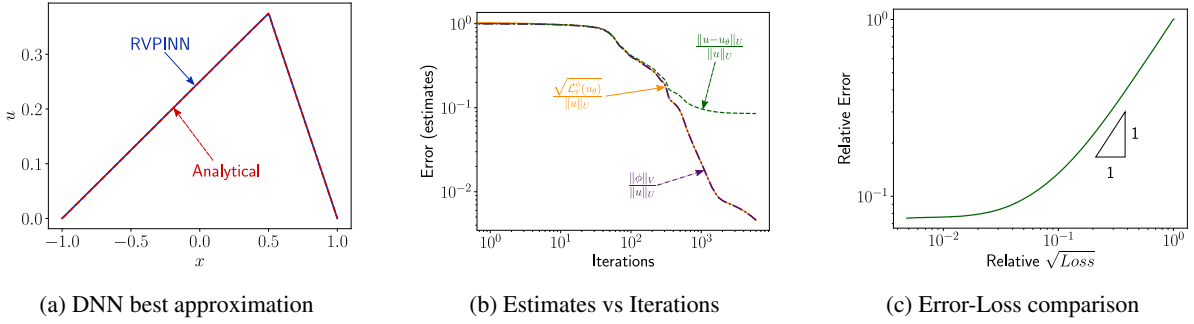(a) DNN best approximation   (b) Estimates vs Iterations   (c) Error-Loss comparison

Figure 7: RVPINNs approximation of the delta diffusion problem; weak BCs imposition, 100 FE test functions.

5.6   Advection-dominated diffusion problem

As our last example, we consider problem (42) in an advection-dominant regime, with the spirit of testing the performance and robustness of the estimates for RVPINNs in near unstable scenarios. For this, we set $\beta = 1$, the source term as $f = 1$, and consider small values of ε. For a given ε, the analytical solution of (42) in this case is given by:

$$u(x) = \frac{2(1 - e^{\frac{x-1}{\varepsilon}})}{1 - e^{-\frac{2}{\varepsilon}}} + x - 1,$$

exhibiting a strong gradient near the right boundary when $\varepsilon \to 0$ (cf. figures 8a and 10a). We restrict the discussion to spectral test functions (figures 8– 10), as we observe similar behaviors when considering FE test functions.
Figure 8 shows the results obtained for $\varepsilon = 0.1$, considering 50 test functions and strong imposition of the BCs. Contrary to the pure diffusive case, Figure 8b reflects that the lines representing the error and the square root of the loss functional (respectively, the norm of the discrete Riesz representative with respect to the norm of the test space $V \times \mathbb{R}^2$) do not overlap. This behavior is expected as, in this case, the continuity constant μ is not equal to one. Indeed, since the Poincaré constant is well-known in 1D, we can explicitly calculate it: $\mu = 1 + 2/(\varepsilon\pi)$. It deteriorates when $\varepsilon \to 0$, i.e., the estimations for this particular setting are not robust in terms of the physical parameter ε. Nevertheless, as we can see in Figure 8c, our theoretical a posteriori error bounds (34) and (35) are satisfied. We also observe a decay in the expected optimal correlation between the error and the square root of the loss starting near iteration 3000, apparently due to stability issues, that is in agreement with the oscilation term appearing in the a posteriori error estimate (35). This suboptimality is improved when considering the weak imposition of the boundary conditions, as shown in Figure 9. This weak imposition of BCs could thus serve as a stabilization technique. This is further confirmed in Figure 10, where we show the results obtained for $\varepsilon = 0.005$, with 200 test basis functions and weak BCs. When considering instead the strong imposition of the BCs and the same number of test functions, we do not observe convergence within the first 6000 iterations–omitted here for brevity. We observe that when considering the weak BCs, the loss functional exhibits some differences with the norm of the Riesz representative in V, as we intentionally plot the incomplete estimation to highlight the contribution of weak BC imposition in $\mathbb{R}^2$ of the full Riesz representative (cf. Eq. (52)). Nevertheless, in this case, we still observe a good correlation between the loss and the true error, as expected.

     To guarantee good estimations in RVPINNs, we need adequate variational formulations and norms for the trial and test spaces, so the resulting inf-sup and boundedness constants are independent of the physical parameters.
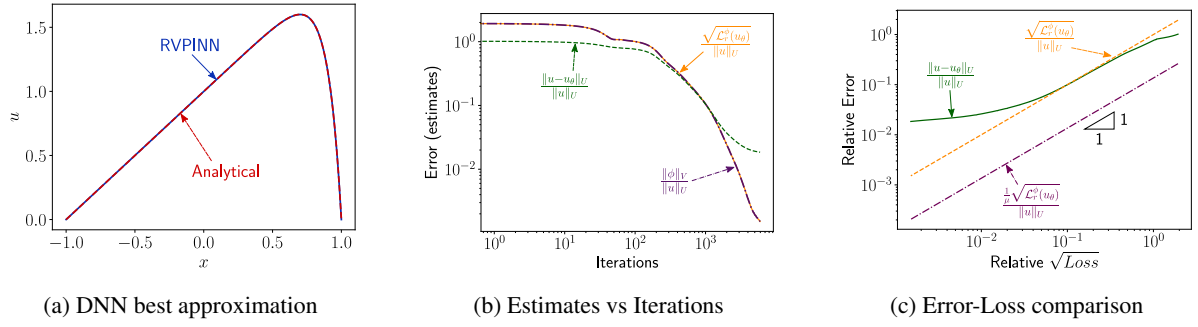
16

| (a) DNN best approximation | (b) Estimates vs Iterations | (c) Error-Loss comparison |

Figure 8: RVPINNs approximation of the diffusion-advection problem with $\varepsilon = 0.1$; strong BCs imposition, 50 spectral test functions.

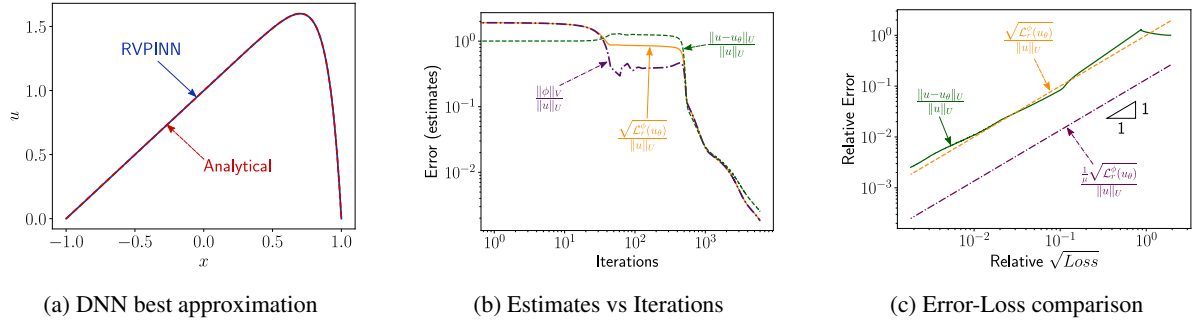| (a) DNN best approximation | (b) Estimates vs Iterations | (c) Error-Loss comparison |

Figure 9: RVPINNs approximation of the diffusion-advection problem with $\varepsilon = 0.1$; weak BCs imposition, 50 spectral test functions.

To exemplify this with a numerical example, consider the following (weighted) alternative variational formulation (with $\Omega := (-1, 1)$)

$$\text{Find } u \in U := H_0^1(\Omega) : (\varepsilon\, e^{-1/\varepsilon} u', v')_0 = (e^{-1/\varepsilon} f, v)_0, \quad \forall v \in V := U,$$

where the last can be deduced from (41), by first multiplying by $e^{-1/\varepsilon} v$ (assuming enough regularity for the solution u), and then performing integration by parts for the term involving the diffusion term. This alternative variational formulation is bounded and coercive, with constants $\mu = \alpha = 1$, with respect to the norms:

$$\| \cdot \|_U^2 := \| \cdot \|_V^2 := (\varepsilon\, e^{-1/\varepsilon}\cdot', \cdot')_0,$$

expecting a perfect correlation of the involved quantities as in the pure diffusive case. Indeed, Figure 11 shows the numerical results obtained for $\varepsilon = 0.5$, considering 200 piece-wise linear functions, confirming the robustness of RVPINNs.
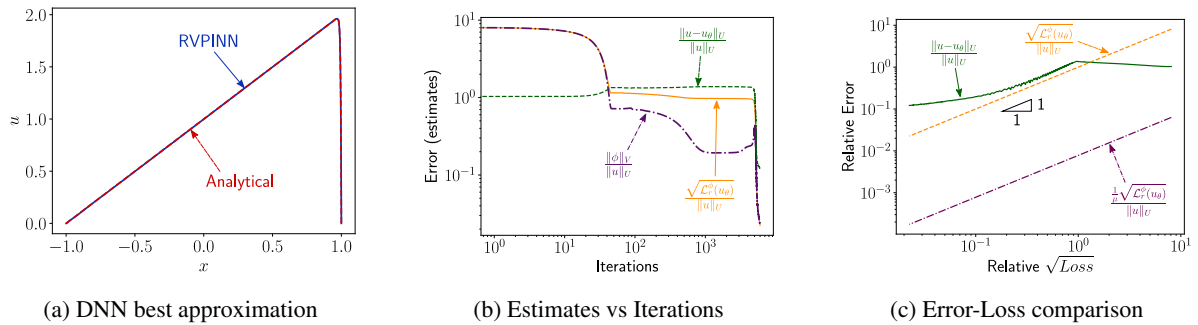

| (a) DNN best approximation | (b) Estimates vs Iterations | (c) Error-Loss comparison |

Figure 10: RVPINNs approximation of the diffusion-advection problem with $\varepsilon = 0.005$; weak BCs imposition, 200 spectral test functions.

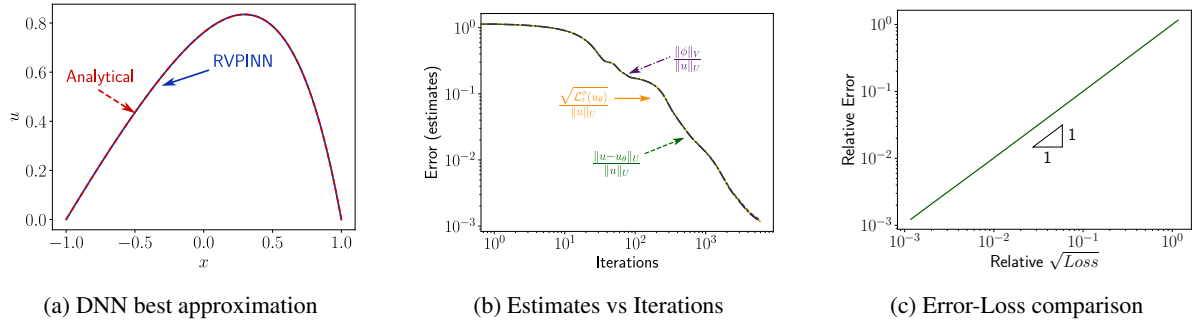(a) DNN best approximation   (b) Estimates vs Iterations   (c) Error-Loss comparison

Figure 11: RVPINNs approximation of the diffusion-advection problem with $\varepsilon = 0.5$; weighted variational formulation, strong BCs imposition, 200 linear FEM test functions.

6 Conclusions

In this article, we provide a general mathematical framework to construct robust loss functionals based on VPINNs. For that, we first generalize the definition of the loss functional in VPINNs to the choice of a single test function. Then, following a minimum residual principle, we define such test function as the Riesz representative of the weak residual over a given discrete test space. We expand the Riesz representation over the discrete test space so the loss functional includes the inversion of the Gram matrix corresponding to the inner product. We prove that the true error in the energy norm is equivalent to the test norm of the residual error estimator. To prove the robustness of the method, we need to introduce a test norm that induces the inf-sup stability at the continuous level. We numerically show the robustness of our a posteriori error estimator and also that our methodology is insensitive to the choice of the basis functions in the selected discrete test space.

Possible future research lines include: (a) the application of the framework to other families of PDEs like wave propagation, nonlinear or time-dependent problems, (b) the extension to nonconforming formulations, (c) the application of RVPINNs to parametric problems with an off-line inversion of the Gram matrix, and (d) an efficiency study of the method for a given variational formulation, NN configuration, and test space.

7 Acknowledgments

References

[1] G. ACOSTA AND R. DURÁN, *An optimal Poincaré inequality in L^1 for convex domains*, Proceedings of the american mathematical society, 132 (2004), pp. 195–202.

[2] M. AINSWORTH AND J. DONG, *Galerkin neural networks: A framework for approximating variational equations with error control*, SIAM Journal on Scientific Computing, 43 (2021), pp. A2474–A2501.

[3] Z. ALDIRANY, R. COTTEREAU, M. LAFOREST, AND S. PRUDHOMME, *Multi-level neural networks for accurate solutions of boundary-value problems*, arXiv preprint arXiv:2308.11503, (2023).

[4] S. BADIA, W. LI, AND A. F. MARTÍN, *Finite element interpolated neural networks for solving forward and inverse problems*, Computer Methods in Applied Mechanics and Engineering, 418 (2024), p. 116505.

[5] S. BERRONE, C. CANUTO, AND M. PINTORE, *Solving PDEs by variational physics-informed neural networks: an a posteriori error analysis*, ANNALI DELL'UNIVERSITA'DI FERRARA, 68 (2022), pp. 575–595.

[6] ———, *Variational physics informed neural networks: the role of quadratures and test functions*, Journal of Scientific Computing, 92 (2022), p. 100.

[7] P. B. BOCHEV AND M. D. GUNZBURGER, *Least-squares finite element methods*, vol. 166, Springer Science & Business Media, 2009.

[8] D. BOFFI, F. BREZZI, M. FORTIN, ET AL., *Mixed finite element methods and applications*, vol. 44, Springer, 2013.

[9] I. BREVIS, I. MUGA, AND K. G. VAN DER ZEE, *Neural control of discrete weak formulations: Galerkin, least squares & minimal-residual methods with quasi-optimal weights*, Computer Methods in Applied Mechanics and Engineering, 402 (2022), p. 115716.

[10] S. CAI, Z. MAO, Z. WANG, M. YIN, AND G. E. KARNIADAKIS, *Physics-informed neural networks (PINNs) for fluid mechanics: A review*, Acta Mechanica Sinica, 37 (2021), pp. 1727–1738.

[11] Z. CAI, J. CHEN, AND M. LIU, *Least-squares ReLU neural network (LSNN) method for scalar nonlinear hyperbolic conservation law*, Applied Numerical Mathematics, 174 (2022), pp. 163–176.

[12] Z. CAI, R. LAZAROV, T. A. MANTEUFFEL, AND S. F. MCCORMICK, *First-order system least squares for second-order partial differential equations: Part I*, SIAM Journal on Numerical Analysis, 31 (1994), pp. 1785–1799.

[13] Z. CAI, T. A. MANTEUFFEL, AND S. F. MCCORMICK, *First-order system least squares for second-order partial differential equations: Part II*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 425–454.

[14] V. M. CALO, A. ERN, I. MUGA, AND S. ROJAS, *An adaptive stabilized conforming finite element method via residual minimization on dual discontinuous Galerkin norms*, Computer Methods in Applied Mechanics and Engineering, 363 (2020), p. 112891.

[15] V. M. CALO, M. ŁOŚ, Q. DENG, I. MUGA, AND M. PASZYŃSKI, *Isogeometric residual minimization method (iGRM) with direction splitting preconditioner for stationary advection-dominated diffusion problems*, Computer Methods in Applied Mechanics and Engineering, 373 (2021), p. 113214.

[16] Y. CHEN, L. LU, G. E. KARNIADAKIS, AND L. DAL NEGRO, *Physics-informed neural networks for inverse problems in nano-optics and metamaterials*, Optics express, 28 (2020), pp. 11618–11633.

[17] R. J. CIER, S. ROJAS, AND V. M. CALO, *A nonlinear weak constraint enforcement method for advection-dominated diffusion problems*, Mechanics Research Communications, 112 (2021), p. 103602.

[18] ———, *Automatically adaptive, stabilized finite element method via residual minimization for heterogeneous, anisotropic advection–diffusion–reaction problems*, Computer Methods in Applied Mechanics and Engineering, 385 (2021), p. 114027.

[19] L. DEMKOWICZ, *Babuška $\iff$ Brezzi*, ICES Report, (2006), pp. 06–08.

[20] L. DEMKOWICZ AND J. GOPALAKRISHNAN, *A class of discontinuous Petrov-Galerkin methods. Part I: The transport equation*, Comput. Methods Appl. Mech. Eng., 199 (2010), pp. 1558 – 1572.

[21] L. DEMKOWICZ AND N. HEUER, *Robust DPG method for convection-dominated diffusion problems*, SIAM Journal on Numerical Analysis, 51 (2013), pp. 2514–2537.

[22] L. F. DEMKOWICZ AND J. GOPALAKRISHNAN, *An overview of the discontinuous Petrov Galerkin method*, Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations: 2012 John H Barrett Memorial Lectures, (2014), pp. 149–180.

[23] D. A. DI PIETRO AND A. ERN, *Mathematical aspects of discontinuous Galerkin methods*, vol. 69, Springer Science, 2012.

[24] A. ERN AND J.-L. GUERMOND, *A converse to Fortin's Lemma in Banach spaces*, Comptes Rendus Mathematique, 354 (2016), pp. 1092–1095.

[25] M. FORTIN, *An analysis of the convergence of mixed finite element methods*, RAIRO. Analyse numérique, 11 (1977), pp. 341–354.

[26] H. GAO, M. J. ZAHR, AND J.-X. WANG, *Physics-informed graph neural Galerkin networks: A unified framework for solving PDE-governed forward and inverse problems*, Computer Methods in Applied Mechanics and Engineering, 390 (2022), p. 114502.

[27] M. GHEISARI, G. WANG, AND M. Z. A. BHUIYAN, *A survey on deep learning in big data*, in 2017 IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC), vol. 2, IEEE, 2017, pp. 173–180.

[28] G. HINTON, L. DENG, D. YU, G. E. DAHL, A.-R. MOHAMED, N. JAITLY, A. SENIOR, V. VANHOUCKE, P. NGUYEN, T. N. SAINATH, ET AL., *Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups*, IEEE Signal processing magazine, 29 (2012), pp. 82–97.

[29] T. J. HUGHES, L. P. FRANCA, AND G. M. HULBERT, *A new finite element formulation for computational fluid dynamics: Viii. the galerkin/least-squares method for advective-diffusive equations*, Computer methods in applied mechanics and engineering, 73 (1989), pp. 173–189.

[30] B.-N. JIANG, *The least-squares finite element method: theory and applications in computational fluid dynamics and electromagnetics*, Springer Science & Business Media, 1998.

[31] E. KHARAZMI, Z. ZHANG, AND G. E. KARNIADAKIS, *Variational physics-informed neural networks for solving partial differential equations*, arXiv preprint arXiv:1912.00873, (2019).

[32] ———, *hp-vpinns: Variational physics-informed neural networks with domain decomposition*, Computer Methods in Applied Mechanics and Engineering, 374 (2021), p. 113547.

[33] E. KHARAZMI, Z. ZHANG, AND G. E. KARNIADAKIS, *hp-VPINNs: Variational physics-informed neural networks with domain decomposition*, Computer Methods in Applied Mechanics and Engineering, 374 (2021), p. 113547.

[34] D. P. KINGMA AND J. BA, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980, (2014).

[35] A. KRISHNAPRIYAN, A. GHOLAMI, S. ZHE, R. KIRBY, AND M. W. MAHONEY, *Characterizing possible failure modes in physics-informed neural networks*, Advances in Neural Information Processing Systems, 34 (2021), pp. 26548–26560.

[36] A. KRIZHEVSKY, I. SUTSKEVER, AND G. E. HINTON, *Imagenet classification with deep convolutional neural networks*, Communications of the ACM, 60 (2017), pp. 84–90.

[37] F. E. KYBURG, S. ROJAS, AND V. M. CALO, *Incompressible flow modeling using an adaptive stabilized finite element method based on residual minimization*, International Journal for Numerical Methods in Engineering, 123 (2022), pp. 1717–1735.

[38] M. ŁOŚ, I. MUGA, J. MUÑOZ-MATUTE, AND M. PASZYŃSKI, *Isogeometric residual minimization (iGRM) for non-stationary Stokes and Navier–Stokes problems*, Computers & Mathematics with Applications, 95 (2021), pp. 200–214.

[39] M. ŁOŚ, J. MUÑOZ-MATUTE, I. MUGA, AND M. PASZYŃSKI, *Isogeometric residual minimization method (iGRM) with direction splitting for non-stationary advection–diffusion problems*, Computers & Mathematics with Applications, 79 (2020), pp. 213–229.

[40] M. ŁOŚ, S. ROJAS, M. PASZYŃSKI, I. MUGA, AND V. M. CALO, *DGIRM: Discontinuous Galerkin based isogeometric residual minimization for the Stokes problem*, Journal of Computational Science, 50 (2021), p. 101306.

[41] Z. MAO, A. D. JAGTAP, AND G. E. KARNIADAKIS, *Physics-informed neural networks for high-speed flows*, Computer Methods in Applied Mechanics and Engineering, 360 (2020), p. 112789.

[42] S. MISHRA AND R. MOLINARO, *Estimates on the generalization error of physics-informed neural networks for approximating a class of inverse problems for PDEs*, IMA Journal of Numerical Analysis, 42 (2022), pp. 981–1022.

[43] J. MUÑOZ-MATUTE, D. PARDO, AND L. DEMKOWICZ, *A DPG-based time-marching scheme for linear hyperbolic problems*, Computer Methods in Applied Mechanics and Engineering, 373 (2021), p. 113539.

[44] S. NAGARAJ, S. PETRIDES, AND L. F. DEMKOWICZ, *Construction of DPG Fortin operators for second order problems*, Computers & Mathematics with Applications, 74 (2017), pp. 1964–1980.

[45] J. T. ODEN AND L. DEMKOWICZ, *Applied functional analysis*, CRC press, 2017.

[46] L. PAYNE AND H. WEINBERGER, *An optimal Poincaré inequality for convex domains*, Arch. Ration. Mech. Anal., 5 (1960), pp. 286–292.

[47] M. RAISSI, P. PERDIKARIS, AND G. 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 (2019), pp. 686–707.

[48] M. RASHT-BEHESHT, C. HUBER, K. SHUKLA, AND G. E. KARNIADAKIS, *Physics-informed neural networks (pinns) for wave propagation and full waveform inversions*, Journal of Geophysical Research: Solid Earth, 127 (2022), p. e2021JB023120.

[49] N. V. ROBERTS, T. BUI-THANH, AND L. DEMKOWICZ, *The DPG method for the Stokes problem*, Computers & Mathematics with Applications, 67 (2014), pp. 966–995.

[50] S. ROJAS, D. PARDO, P. BEHNOUDFAR, AND V. M. CALO, *Goal-oriented adaptivity for a conforming residual minimization method in a dual discontinuous Galerkin norm*, Computer Methods in Applied Mechanics and Engineering, 377 (2021), p. 113686.

[51] Y. SAAD, *Iterative methods for sparse linear systems*, SIAM, 2003.

[52] J. M. TAYLOR, M. BASTIDAS, D. PARDO, AND I. MUGA, *Deep Fourier Residual method for solving time-harmonic Maxwell's equations*, arXiv preprint arXiv:2305.09578, (2023).

[53] J. M. TAYLOR, D. PARDO, AND I. MUGA, *A deep Fourier residual method for solving PDEs using neural networks*, Computer Methods in Applied Mechanics and Engineering, 405 (2023), p. 115850.

[54] C. URIARTE, D. PARDO, I. MUGA, AND J. MUÑOZ-MATUTE, *A Deep Double Ritz Method (D2RM) for solving Partial Differential Equations using Neural Networks*, Computer Methods in Applied Mechanics and Engineering, 405 (2023), p. 115892.