

ERROR BOUNDS FOR PDE-REGULARIZED LEARNING

CARSTEN GRÄSER¹ AND PREM ANAND ALATHUR SRINIVASAN²

ABSTRACT. In this work we consider the regularization of a supervised learning problem by partial differential equations (PDEs) and derive error bounds for the obtained approximation in terms of a PDE error term and a data error term. Assuming that the target function satisfies an unknown PDE, the PDE error term quantifies how well this PDE is approximated by the auxiliary PDE used for regularization. It is shown that this error term decreases if more data is provided. The data error term quantifies the accuracy of the given data. Furthermore, the PDE-regularized learning problem is discretized by generalized Galerkin discretizations solving the associated minimization problem in subsets of the infinite dimensional functions space, which are not necessarily subspaces. For such discretizations an error bound in terms of the PDE error, the data error, and a best approximation error is derived.

1. INTRODUCTION

The problem of learning an unknown function $u : \Omega \rightarrow \mathbb{R}$ is one of the central problems in the field of machine learning. A classical ansatz is to minimize the risk functional

$$\tilde{u} \mapsto |\Omega|^{-1} \int_{\Omega} L(\tilde{u}(x), u(x)) dx$$

in some ansatz set V_h for a loss functional $L : \mathbb{R}^2 \rightarrow \mathbb{R}$. In case of a quadratic loss functional this leads to an $L^2(\Omega)$ -best-approximation problem for u in $V_h \subset L^2(\Omega)$.

Solving this problem in general requires complete knowledge of $u \in L^2(\Omega)$. In practice the available data on u is often incomplete. A classical example of incomplete data is a finite set of known point data $(p_1, u(p_1)), \dots, (p_m, u(p_m))$ for a continuous function u . In this case, instead of minimizing the risk, one often considers the problem of *empirical risk minimization* given by

$$(1) \quad \tilde{u}_h = \operatorname{argmin}_{v \in V_h} \frac{1}{m} \sum_{i=1}^m |v(p_i) - u(p_i)|^2.$$

Here, the weighted sum can be viewed as Monte-Carlo approximation of the $L^2(\Omega)$ -norm using the sample set $\{p_1, \dots, p_m\}$.

Recently a lot of attention has been paid to learning with shallow or deep neural networks. In this case $V_h \subset L^2(\Omega)$ is given as the range $V_h = \Phi(\mathbb{R}^N)$ of a nonlinear map $\Phi : \mathbb{R}^N \rightarrow L^2(\Omega)$ over a finite dimensional parameter domain \mathbb{R}^N , such that the empirical risk minimization

¹INSTITUT FÜR MATHEMATIK, FREIE UNIVERSITÄT BERLIN, 14195 BERLIN, GERMANY

²INSTITUT FÜR MATHEMATIK, FREIE UNIVERSITÄT BERLIN, 14195 BERLIN, GERMANY

E-mail addresses: graeser@mi.fu-berlin.de, premanand@zedat.fu-berlin.de.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689). Project EF3-4 “Physics-regularized learning”.

problem in V_h can be written as

$$(2) \quad W^* = \operatorname{argmin}_{W \in \mathbb{R}^N} \frac{1}{m} \sum_{i=1}^m |\Phi(W)(p_i) - u(p_i)|^2$$

with $\tilde{u}_h = \Phi(W^*)$. The minimizer W^* is often approximated using stochastic gradient-type algorithms. While this approach has been used with reasonable success in practice, its theoretical understanding is still in its infancy. The central question is, if the total error, i.e. the difference $\tilde{u}_h - u$ of the computed approximation and the target function can be controlled. When analyzing this error, several aspects have to be taken into account.

Learning: Here the question is if we can control the error made by the algebraic solution method, e.g. stochastic gradient descent. If the algebraic error $W - W^*$ of a computed approximate parameter set W can be controlled, local Lipschitz continuity of the parametrization Φ allows to control the induced error $\Phi(W) - \Phi(W^*)$ of the associated functions. A major obstacle for an error analysis is, that the learning problem (2) is in general nonconvex and may have multiple global or local minimizers, saddle points, or plateaus. One direction to target this problem is to characterize network architectures, where local optimality implies global optimality or at least plateaus (see, e.g., [11], [27] and the references therein). Another direction is to interpret a (stochastic) gradient flow for shallow neural networks as evolution of an interacting particle system which allows to relate long time limits (i.e. stationary points) and the many particle limit [25]. In the present paper we do not consider this aspect and concentrate on the analysis of global minimizers.

Expressivity: This centers around the question of how well u can be approximated in V_h ? Starting from early results on universal approximation properties of neural networks (see, e.g. [5, 22]) there was significant recent progress on deriving best approximation error and expressivity bounds for deep neural networks [1, 2, 7, 9, 10, 20, 21, 26, 28]. An additional question is, to which extent it is possible to realize the theoretically derived best approximation error by solving the learning problem. Despite its importance this question is largely unexplored which is rooted in the fact that it can hardly be answered due to ill-posedness in case of incomplete data. In the present paper we will address this question by regularizing the problem and proving discretization error bounds for the minimizer \tilde{u}_h in terms of the best approximation error for u in V_h .

Generalization: This addresses the question if the trained function generalizes from the training data to other input data. In mathematical terms this leads to the question of how well the computed function \tilde{u}_h can approximate u on Ω given incomplete data at some points $\{p_1, \dots, p_m\} \not\subseteq \Omega$. This question is often discussed in a statistical setting considering the input data as randomly drawn samples. In the present paper we will take a deterministic perspective and are interested in error bounds for $\tilde{u}_h - u$ in terms of the amount of provided data. Again such bounds can in general not be derived due to ill-posedness, which we will address by a regularization of the problem.

The fact that it is hard to derive error bounds—even for global minimizers of the learning problem—is deeply related to the fact that problem (1) which only incorporates incomplete data is in general ill-posed leading to severe artifacts. For example, if the ansatz set V_h is 'large' in comparison to the available amount of data, one observes so called *overfitting*, where \tilde{u} matches u nicely in the points p_i but fails to provide a reasonable approximation in other

parts of Ω . This is in fact a consequence of ill-posedness ¹. A common technique to reduce such artifacts is to introduce regularization terms for W . However, it is unclear how the influence of such regularization on the error $\Phi(W^*) - u = \tilde{u}_h - u$ with respect to the $L^2(\Omega)$ -norm could be analyzed. Another obstacle in deriving error bounds for $\tilde{u}_h - u$ is, that it is unclear, how to understand (1) as discretization of a continuous problem and thus how such a discretization and the influence of incomplete data could be analyzed. In the present paper we target these questions by introducing regularizations of the learning problem by partial differential equations (PDEs). This is based on the assumption that additional knowledge on the process generating the data or at least on the smoothness of u is available. Using such regularizations we derive a framework which allows to prove error bounds for $\tilde{u}_h - u$ that quantify the effect of incomplete data and relate the discretization error in nonlinear ansatz spaces (like neural networks) to their approximation properties.

Due to their approximation power, neural networks have already been proposed for the solution of partial differential equations (PDEs). In [4] the authors introduce the *deep Ritz* method which approximates the solution of a variational formulation of a PDE using discretization by deep neural networks. In contrast to (2) this approach minimizes the Dirichlet energy associated to the PDE in the nonlinear neural network space using a simple penalty approach for essential boundary data. As a variant, it was proposed to minimize the consistent penalty formulation from Nitsche’s method using a heuristic penalty parameter [17]. While error bounds for both methods are unknown so far a convergence result based on Γ -convergence was recently derived [18]. A different approach was taken in [23], where the authors introduce so called *physics informed neural networks* (PINNs) which are trained using a collocation least squares functional. This ansatz is extended in [24] where, additionally to the PDE, point data of the target function is incorporated. Replacing collocation by a least squares Petrov–Galerkin ansatz leads to the *variational PINN* approach considered in [14]. It has also been highlighted that a neural network ansatz is especially promising for high-dimensional PDEs [4, 12, 9, 17]. Other uses of neural networks in the context of PDEs e.g. include reduced basis methods for parametrized problems [16].

Our contribution: In the present paper we will in general assume that u solves an elliptic PDE which is not known exactly. Since we cannot use the unknown exact PDE, an inexact auxiliary PDE is used to regularize the problem. Furthermore, to make the problem well-posed in a Lebesgue- and Sobolev-space setting, we first replace the point data $u(p_i)$ by local averages $f_{B_i} u \approx f(p_i)$ on sets $B_i \subset \Omega$ with $p_i \in B_i$ leading to the regularized learning problem

$$\tilde{u} = \operatorname{argmin}_{v \in V} \frac{1}{2} \sum_{i=1}^m |B_i| |f_{B_i} v - f_{B_i} u|^2 + \delta \left(\frac{1}{2} \tilde{a}(v, v) - \tilde{\ell}(v) \right).$$

Here, $\frac{1}{2} \tilde{a}(v, v) - \tilde{\ell}(v)$ is the Dirichlet energy associated to the elliptic auxiliary PDE and $\delta > 0$ is a regularization parameter that balances the data and PDE term. The main result of the paper is an error bound of the form

$$\|\tilde{u} - u\|_{L^2(\Omega)} \leq CR^2 \mathbb{E}_{\text{pde}}.$$

Here R is a constants that can be decreased by adding more data and \mathbb{E}_{pde} quantifies the error induced by using an inexact auxiliary PDE. Hence the accuracy of \tilde{u} can be improved by

¹This can be easily explained in the case of approximation by polynomials $V_h = \mathcal{P}_n$ of degree n . If $n = m - 1$, then the resulting \tilde{u} is exactly the interpolation polynomial which in general exhibits severe over- and undershoots. If $n \geq m$ then there is no unique solution and one can add an oscillatory polynomial of degree n with arbitrary amplitude to the interpolation polynomial leading to uncontrollably large errors.

either providing more data or by improving the exactness of the auxiliary PDE. In a second step we treat the case of given point values $u(p_i)$ by considering $u(p_i)$ as an inexact variant of $f_{B_i} u$. Assuming λ -Hölder-continuity of u we can control the additional error by an error bound

$$\|\tilde{u} - u\|_{L^2(\Omega)} \leq CR^2 \mathbb{E}_{\text{pde}} + Cr^\lambda$$

where r can again be decreased by adding more data.

Finally, using a nonlinear Céa-Lemma, we derive a bound

$$\|\tilde{u}_h - u\|_{L^2(\Omega)} \leq CR^2 \mathbb{E}_{\text{pde}} + Cr^\lambda + C \inf_{v \in V_h} \left(R \|\nabla(v - u)\|_{L^2(\Omega)} + \|v - u\|_{L^2(\Omega)} \right)$$

for the case of a generalized Galerkin discretization where \tilde{u}_h is computed by minimizing in a subset V_h of V . A variant for inexact minimizers is also presented. Since the result allows for non-subspace subsets $V_h \subset V$, it is in principle applicable to nonlinear approximation schemes like neural networks. Inserting known approximation error bounds (e.g. from [10]) on the right hand side, this leads to a discretization error bound for neural network discretizations under the assumption that a global minimizer can be computed, or, that the algebraic energy error can be controlled. While the latter can in general not be guaranteed, the derived results open a new perspective for an error analysis of neural networks. For example, the same arguments can be used to provide a-priori error bounds for the deep Ritz method [4] for Neumann problems with exact (or controlled inexact) global minimizers which allows to quantify recent convergence results [18].

The paper is organized as follows: First the PDE-regularized learning problem is introduced and its well-posedness is discussed in Section 2. Then an $L^2(\Omega)$ error bound for $\tilde{u} - u$ is derived in Section 3 in the infinite dimensional case. This section also discussed the quasi-optimality of the derived error bound for a pure data fitting problem without a-priori knowledge on the PDE. The generalized Galerkin discretization in subsets V_h is introduced and analyzed in Section 4. Finally, the theoretical findings are illustrated by numerical experiments with finite element and neural network discretizations in Section 5.

2. PROBLEM SETTING

2.1. Exact and inexact auxiliary PDEs. We are interested in approximating a function $u : \Omega \rightarrow \mathbb{R}$ on a bounded domain $\Omega \subset \mathbb{R}^d$ with Lipschitz boundary. Throughout the paper we make the assumption that u solves an elliptic partial differential equation (PDE) given in terms of a variational equation

$$(3) \quad u \in V : \quad a(u, v) = \ell(v) \quad \forall v \in V$$

for a closed subspace $V \subset H^1(\Omega)$ with $H_0^1(\Omega) \subset V$, a symmetric bilinear form

$$a(w, v) := \int_{\Omega} \alpha(x) \nabla w(x) \cdot \nabla v(x) + \sigma(x) w(x) v(x) dx$$

with uniformly bounded coefficient functions $\alpha : \Omega \rightarrow \mathbb{R}$ and $\sigma : \Omega \rightarrow \mathbb{R}$,

$$0 < \alpha_{\min} \leq \alpha(x) \leq \alpha_{\max} < \infty, \quad 0 \leq \sigma(x) \leq \|\sigma\|_{L^\infty(\Omega)} < \infty$$

and $\ell \in V^*$ given by

$$\ell(v) := \int_{\Omega} f(x) dx$$

for some $f \in L^2(\Omega)$. The subspace V is chosen such that $a(\cdot, \cdot)$ is coercive on V . Then (3) has a unique solution $u \in V$ by the Lax–Milgram theorem.

The basic assumption we will make is, that the PDE is not known exactly, that is, the exact functions α, σ, ℓ are unknown. Instead we will consider an inexact auxiliary PDE induced by a guess for α, σ , and f . The auxiliary PDE is given in terms a bilinear form

$$\tilde{a}(w, v) := \int_{\Omega} \tilde{\alpha}(x) \nabla w(x) \cdot \nabla v(x) + \tilde{\sigma}(x) w(x) v(x) dx$$

with uniformly bounded coefficient functions $\tilde{\alpha} : \Omega \rightarrow \mathbb{R}$ and $\tilde{\sigma} : \Omega \rightarrow \mathbb{R}$,

$$0 < \tilde{\alpha}_{\min} \leq \tilde{\alpha}(x) \leq \tilde{\alpha}_{\max} < \infty, \quad 0 \leq \tilde{\sigma}(x) \leq \|\tilde{\sigma}\|_{L^\infty(\Omega)} < \infty$$

and a functional $\tilde{\ell} \in V^*$ given by

$$\tilde{\ell}(v) := \int_{\Omega} \tilde{f}(x) v(x) dx$$

for some $\tilde{f} \in L^2(\Omega)$.

2.2. PDE-regularized learning problem. To compute \tilde{u} we will not just solve $\tilde{a}(\tilde{u}, \cdot) - \tilde{\ell} = 0$ but combine this PDE with the given (possibly inexact) local data on u . To this end we assume that the data is given in terms of local average values of u on open, nonempty subsets $B_i \subset \Omega$ for $i = 1, \dots, m$. We will also assume that the sets B_i can be extended to form a covering of Ω in the sense that for each B_i there is a convex set K_i with Lipschitz boundary and $B_i \subset K_i$ such that

$$\Omega \subset \bigcup_{i=1}^m \overline{K_i}.$$

The maximal overlap and the maximal diameter of the families (B_i) and (K_i) given by

$$M := \max_{x \in \Omega} |\{i \mid x \in K_i\}|, \quad r := \max_{i=1, \dots, m} \text{diam}(B_i), \quad R := \max_{i=1, \dots, m} \text{diam}(K_i)$$

will be used to quantify errors later on.

In the following we will use the notation $f_U v = |U|^{-1} \int_U v(x) dx$ for the average of v over a bounded set U . Using a regularization parameter $\delta > 0$ we define the augmented auxiliary forms

$$\tilde{c}(w, v) := \tilde{a}(w, v) + \delta^{-1} b(w, v), \quad \tilde{r}(v) := \tilde{\ell}(v) + \delta^{-1} \overline{b(u, v)}.$$

with the bilinear form $b(\cdot, \cdot)$ and the possibly inexact data term $\overline{b(u, \cdot)} \approx b(u, \cdot)$ given by

$$(4) \quad b(w, v) := \sum_{i=1}^m |B_i| f_{B_i} w f_{B_i} v, \quad \overline{b(u, v)} := \sum_{i=1}^m b_i |B_i| f_{B_i} v.$$

This data term can be viewed as an approximation of $b(u, v)$ which becomes exact for $b_i = f_{B_i} u$. Using this notation we introduce the PDE-regularized learning problem

$$(5) \quad \tilde{u} = \operatorname{argmin}_{v \in V} \tilde{J}(v),$$

for the functional

$$(6) \quad \tilde{J}(v) := \frac{1}{2} \tilde{c}(v, v) - \tilde{r}(v) = \frac{1}{2} \sum_{i=1}^m |B_i| \left| f_{B_i} v - b_i \right|^2 + \delta \left(\frac{1}{2} \tilde{a}(v, v) - \tilde{\ell}(v) \right) + \text{const}.$$

Using standard arguments we see, that this quadratic minimization problem is equivalent to

$$(7) \quad \tilde{u} \in V : \quad \tilde{c}(\tilde{u}, v) = \tilde{r}(v) \quad \forall v \in V.$$

Analogously to (7) we can define the augmented exact problem

$$(8) \quad u \in V : \quad c(u, v) = r(v) \quad \forall v \in V$$

for the exact augmented forms

$$c(w, v) := a(w, v) + \delta^{-1}b(w, v), \quad r(v) := \ell(v) + \delta^{-1}b(u, v).$$

It is straight forward to show that (8) is equivalent to (3).

2.3. Well-posedness by PDE-based regularization. We now discuss well-posedness of the PDE-regularized learning problem. For convenience we will denote the semi-norm $v \mapsto d(v, v)^{\frac{1}{2}}$ induced by a symmetric, positive semi-definite, bilinear form $d(\cdot, \cdot)$ by $\|\cdot\|_d = d(\cdot, \cdot)^{\frac{1}{2}}$. Using this notation it is clear that

$$\|\cdot\|_{\tilde{c}}^2 = \|\cdot\|_{\tilde{a}}^2 + \delta^{-1}\|\cdot\|_b^2, \quad \|\cdot\|_c^2 = \|\cdot\|_a^2 + \delta^{-1}\|\cdot\|_b^2.$$

The following lemma shows that the weighting of the terms in $b(\cdot, \cdot)$ is natural in the sense that it guarantee that $\|\cdot\|_b$ scales like $\|\cdot\|_{L^2(\Omega)}$.

Lemma 1. *The bilinear form $b(\cdot, \cdot)$ is $L^2(\Omega)$ -continuous with*

$$b(w, v) \leq M\|w\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}, \quad \|v\|_b^2 \leq M\|v\|_{L^2(\Omega)}^2, \quad \forall w, v \in L^2(\Omega).$$

Proof. Let $v, w \in L^2(\Omega)$. We first note that the Cauchy–Schwarz inequality gives

$$(9) \quad \int_{B_i} v \leq |B_i|^{-1}\|v\|_{L^2(B_i)}\|1\|_{L^2(B_i)} = |B_i|^{-1/2}\|v\|_{L^2(B_i)}.$$

Using this estimate and the Cauchy–Schwarz inequality in \mathbb{R}^m we get

$$b(w, v) \leq \sum_{i=1}^m \|w\|_{L^2(B_i)}\|v\|_{L^2(B_i)} \leq \left(\sum_{i=1}^m \|w\|_{L^2(B_i)}^2 \right)^{1/2} \left(\sum_{i=1}^m \|v\|_{L^2(B_i)}^2 \right)^{1/2} \leq M\|w\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}. \quad \square$$

Despite this upper bound, a pure data fitting problem

$$\min_{v \in L^2(\Omega)} \frac{1}{2} \sum_{i=1}^m |B_i| \left| \int_{B_i} v - b_i \right|^2 = \min_{v \in L^2(\Omega)} \frac{1}{2} \|v\|_b^2 - \widetilde{b(u, v)} + \text{const}$$

is in general ill-posed, since $b(\cdot, \cdot)$ is not coercive, or, equivalently, $\|\cdot\|_b$ cannot be bounded from below by $\|\cdot\|_{L^2(\Omega)}$. Due to the finite rank m of $b(\cdot, \cdot)$, the same is true whenever minimization is considered in a space with dimension larger than m . Thanks to the PDE-regularization, the situation is different for the PDE-regularized problem (5).

Proposition 1. *The PDE-regularized problem (5) or, equivalently, (7) has a unique solution $\tilde{u} \in V$ which depends Lipschitz-continuously on the provided data b_1, \dots, b_m .*

Proof. By the lower bound on $\tilde{\alpha}$ we obtain

$$\|\nabla v\|_{L^2(\Omega)}^2 \leq \tilde{\alpha}_{\min}^{-1} \|v\|_{\tilde{c}}^2 \quad \forall v \in V.$$

Furthermore $b(\cdot, \cdot)$ is coercive on the space of constant functions. Hence coercivity of \tilde{c} on $H^1(\Omega)$ follows from the Poincaré inequality (see [6, Proposition 2]). Furthermore the upper bounds on $\tilde{\alpha}$ and $\tilde{\sigma}$ and Lemma 1 imply continuity and thus ellipticity of $\tilde{c}(\cdot, \cdot)$ on V . Finally,

we get continuity of $\widetilde{b}(u, \cdot)$ and thus $\widetilde{r}(\cdot)$ similar to the proof of Lemma 1 such that the Lax–Milgram theorem guarantees existence of a unique solution $\tilde{u} \in V$ of (7). Lipschitz continuous dependency on b_1, \dots, b_m follows from standard arguments for linear elliptic problems. \square

3. ERROR BOUNDS FOR PDE-REGULARIZED LEARNING

3.1. Localized Poincaré inequality and improved $L^2(\Omega)$ -coercivity. A central ingredient in the error estimates shown later is the $L^2(\Omega)$ -coercivity of $\tilde{c}(\cdot, \cdot)$. In the proof of Proposition 1 we have seen, that $\tilde{c}(\cdot, \cdot)$ inherits coercivity with respect to the $H^1(\Omega)$ - and $L^2(\Omega)$ -norm from $\tilde{a}(\cdot, \cdot)$. However, the coercivity constants are independent of the data term. In this section we will show an improved $L^2(\Omega)$ -coercivity by applying localized Poincaré estimates. First we remind the classical Poincaré inequality on convex domains.

Lemma 2. *Let $U \subset \mathbb{R}^d$ be a convex, open, non-empty, bounded domain and $v \in H^1(U)$. Then*

$$\|v - f_U v\|_{L^2(U)} \leq C_U \|\nabla v\|_{L^2(U)}$$

with Poincaré constant $C_U = \frac{\text{diam}(U)}{\pi}$.

In fact the constant C_U given here is the best possible one depending only on the domain diameter. A proof can be found in [19]. As a direct consequence we get the following estimate.

Lemma 3. *Let $U \subset \mathbb{R}^d$ be a convex, open, non-empty, bounded domain and $v \in H^1(U)$. Then*

$$\|v\|_{L^2(U)}^2 \leq C_U^2 \|\nabla v\|_{L^2(U)}^2 + \|f_U v\|_{L^2(U)}^2 = C_U^2 \|\nabla v\|_{L^2(U)}^2 + |U| (f_U v)^2$$

with Poincaré constant $C_U = \frac{\text{diam}(U)}{\pi}$.

Proof. Using the $L^2(U)$ -orthogonality $v - f_U v$ and $f_U v$ and Lemma 2 we get

$$\|v\|_{L^2(U)}^2 = \|v - f_U v\|_{L^2(U)}^2 + \|f_U v\|_{L^2(U)}^2 \leq C_U^2 \|\nabla v\|_{L^2(U)}^2 + |U| (f_U v)^2.$$

\square

It is also possible to get bounds involving just averages over subsets, at the price of an additional constant. An abstract prove has been given in [6]. Since we are interested in the resulting constants we will give an explicit proof here.

Lemma 4. *Let $U \subset \mathbb{R}^d$ be a convex, open, non-empty, bounded domain, $W \subset U$ with $|W| > 0$, $t > 0$, and $v \in H^1(U)$. Then*

$$\|v\|_{L^2(U)}^2 \leq C_U^2 \left(1 + (1+t) \frac{|U|}{|W|}\right) \|\nabla v\|_{L^2(U)}^2 + (1+t^{-1})|U| (f_W v)^2$$

with Poincaré constant $C_U = \frac{\text{diam}(U)}{\pi}$ and especially (using $t = 1$)

$$\|v\|_{L^2(U)}^2 \leq 3 \frac{|U|}{|W|} \left(C_U^2 \|\nabla v\|_{L^2(U)}^2 + |W| (f_W v)^2\right).$$

Proof. Applying Lemma 3 on U we get

$$\|v\|_{L^2(U)}^2 \leq C_U^2 \|\nabla v\|_{L^2(U)}^2 + \|v_U\|_{L^2(U)}^2$$

where we used the notation $v_U = f_U v$. Utilizing the inequality

$$\langle a, a \rangle = \langle a - b, a - b \rangle + 2\langle b, a - b \rangle + \langle b, b \rangle \leq (1+t^{-1})\langle b, b \rangle + (1+t)\langle b - a, b - a \rangle$$

for the symmetric positive semi-definite bilinear form $\langle \cdot, \cdot \rangle = f_W(\cdot) f_W(\cdot)$ we get

$$\begin{aligned} \|v_U\|_{L^2(U)}^2 &= |U|(f_W v_U)^2 \leq |U|\left((1+t^{-1})(f_W v)^2 + (1+t)(f_W(v-v_U))^2\right) \\ &\leq |U|\left((1+t^{-1})(f_W v)^2 + (1+t)|W|^{-1}\|v-v_U\|_{L^2(W)}^2\right) \\ &\leq |U|(1+t^{-1})(f_W v)^2 + (1+t)\frac{|U|}{|W|}\|v-v_U\|_{L^2(U)}^2. \end{aligned}$$

Finally we get the assertion by using Lemma 2. \square

Note that using the optimal value for t the constant 3 can be slightly improved to $1 + \varphi < 3$ with the golden ratio $\varphi = \frac{1}{2}(1 + \sqrt{5})$.

Next we apply Lemma 3 and Lemma 4 locally to show a data dependent global Poincaré type estimate. To unify both estimates we encode the maximal mismatch of B_i and K_i in terms of the constant

$$\eta := \begin{cases} 1 & \text{if } B_i = K_i \quad \forall i, \\ 3 \max_{i=1, \dots, m} \frac{|K_i|}{|B_i|} & \text{else.} \end{cases}$$

Lemma 5. *Using the constants M, R, η defined above it holds for $v \in H^1(\Omega)$*

$$\|v\|_{L^2(\Omega)}^2 \leq \eta \left(\frac{R^2 M}{\pi^2} \|\nabla v\|_{L^2(\Omega)}^2 + b(v, v) \right).$$

Proof. Let $v \in H^1(\Omega)$. Then we also have $v \in H^1(B_i)$ for $i = 1, \dots, m$. Applying either Lemma 3 (if $\eta = 1$) or Lemma 4 (if $\eta > 1$) on each B_i we get

$$\begin{aligned} \|v\|_{L^2(\Omega)}^2 &\leq \sum_{i=1}^m \|v\|_{L^2(B_i)}^2 \leq \eta \sum_{i=1}^m \left(\frac{\text{diam}(K_i)^2}{\pi^2} \|\nabla v\|_{L^2(K_i)}^2 + |B_i| (f_{B_i} v)^2 \right) \\ &\leq \eta \left(\sum_{i=1}^m \frac{R^2}{\pi^2} \|\nabla v\|_{L^2(B_i)}^2 + b(v, v) \right). \end{aligned}$$

Using the fact that each part of Ω is a most covered M -times by the sets B_i we get the assertion. \square

The right hand side in the estimate of Lemma 5 almost coincides with $\tilde{c}(v, v)$. The following lemma balances the constants to finally derive an $L^2(\Omega)$ -coercivity of $\tilde{c}(\cdot, \cdot)$.

Lemma 6. *For any $v \in H^1(\Omega)$ it holds that*

$$\|v\|_{L^2(\Omega)}^2 \leq \Gamma \|v\|_{\tilde{c}}^2, \quad \Gamma := \eta \max \left\{ \frac{R^2 M}{\pi^2 \tilde{\alpha}_{\min}}, \delta \right\}.$$

Proof. Using Lemma 5, positive definiteness of $\|\sqrt{\tilde{\sigma}} \cdot\|_{L^2(\Omega)}^2$, and $\Gamma \delta^{-1} \geq \eta$ we get

$$\|v\|_{L^2(\Omega)}^2 \leq \eta \frac{R^2 M}{\pi^2 \tilde{\alpha}_{\min}} \|v\|_{\tilde{a}}^2 + \eta \|v\|_{\tilde{b}}^2 \leq \eta \frac{R^2 M}{\pi^2 \tilde{\alpha}_{\min}} \|v\|_{\tilde{a}}^2 + \Gamma \delta^{-1} \|v\|_{\tilde{b}}^2 \leq \Gamma \|v\|_{\tilde{c}}^2.$$

\square

In the following the relation of δ and the constant from Lemma 5 will play a crucial role. In the simplest case we would have

$$(10) \quad \delta = \frac{R^2 M}{\pi^2 \tilde{\alpha}_{\min}},$$

such that the constant in Lemma 6 reduces to $\Gamma = \eta\delta$. Since this constant may not be known exactly we quantify the relation by assuming that

$$(11) \quad \delta \in \left[\theta \frac{R^2 M}{\pi^2 \tilde{\alpha}_{\min}}, \theta^{-1} \frac{R^2 M}{\pi^2 \tilde{\alpha}_{\min}} \right]$$

for $\theta \in (0, 1]$. Note that by $\delta \in (0, \infty)$ such a $\theta = (0, 1)$ does always exist. Using this assumption we obtain the following bounds on Γ

$$(12) \quad \Gamma \leq \eta\theta^{-1} \frac{R^2 M}{\pi^2 \tilde{\alpha}_{\min}}, \quad \Gamma \leq \eta\theta^{-1} \delta.$$

3.2. Error analysis. Now we show an estimate for the error $u - \tilde{u}$ in the $L^2(\Omega)$ -norm.

Theorem 1. *Let u and \tilde{u} be the solutions of (3) and (7), respectively. Furthermore assume that $\tilde{\ell} - \tilde{a}(u, \cdot) \in L^2(\Omega)$. Then we have with Γ from Lemma 6*

$$(13) \quad \|\tilde{u} - u\|_{L^2(\Omega)} \leq \sqrt{\Gamma} \|\tilde{u} - u\|_{\tilde{c}} \leq \Gamma \left(\mathbb{E}_{\text{pde}} + \delta^{-1} \mathbb{E}_{\text{data}} \right)$$

and, using θ from (11),

$$(14) \quad \|\tilde{u} - u\|_{L^2(\Omega)} \leq \sqrt{\Gamma} \|\tilde{u} - u\|_{\tilde{c}} \leq \eta\theta^{-1} \left(\frac{R^2 M}{\pi^2 \tilde{\alpha}_{\min}} \mathbb{E}_{\text{pde}} + \mathbb{E}_{\text{data}} \right)$$

with the PDE and data error terms

$$\mathbb{E}_{\text{pde}} := \|\tilde{\ell} - \tilde{a}(u, \cdot)\|_{L^2(\Omega)}, \quad \mathbb{E}_{\text{data}} := \|\widetilde{b(u, \cdot)} - b(u, \cdot)\|_{L^2(\Omega)}.$$

Proof. Testing (7) with $\tilde{u} - u$ and subtracting $\tilde{c}(u, \tilde{u} - u)$ yields

$$\begin{aligned} \|\tilde{u} - u\|_{\tilde{c}}^2 &= \tilde{r}(\tilde{u} - u) - \tilde{c}(u, \tilde{u} - u) \\ &= (\tilde{\ell} - \tilde{a}(u, \cdot))(\tilde{u} - u) + \delta^{-1} (\widetilde{b(u, \cdot)} - b(u, \cdot))(\tilde{u} - u) \\ &\leq \left(\|\tilde{\ell} - \tilde{a}(u, \cdot)\|_{L^2(\Omega)} + \delta^{-1} \|\widetilde{b(u, \cdot)} - b(u, \cdot)\|_{L^2(\Omega)} \right) \|\tilde{u} - u\|_{L^2(\Omega)} \\ &\leq \left(\|\tilde{\ell} - \tilde{a}(u, \cdot)\|_{L^2(\Omega)} + \delta^{-1} \|\widetilde{b(u, \cdot)} - b(u, \cdot)\|_{L^2(\Omega)} \right) \sqrt{\Gamma} \|\tilde{u} - u\|_{\tilde{c}} \end{aligned}$$

where we used the $L^2(\Omega)$ -coercivity from Lemma 6 for the last estimate. Dividing by $\|\tilde{u} - u\|_{\tilde{c}}$ and using Lemma 6 again to bound the left hand side, we obtain (13). Estimate (14) is obtained by using the bounds on Γ from (12). \square

It should be noted that the residual $\tilde{\ell} - \tilde{a}(u, \cdot)$ is zero if the PDE is exact. Hence \mathbb{E}_{pde} quantifies the error induced by the inexact PDE. If the data points provide a more fine grained covering of Ω , then R is decreased. In this sense, the PDE error can be reduced by adding more data. On the other hand simply adding more data cannot cure the data error \mathbb{E}_{data} made by using inexact data.

The estimate also indicates that the PDE and data term should be balanced appropriately: In order minimize the constant Γ in front of the PDE error term, δ should be sufficiently small, while it should be sufficiently large in order minimize the constant $\Gamma\delta^{-1}$ in front of the

data error term. Since the estimate (14) bounds both constants in terms of θ^{-1} the optimal choice of δ is (10) which leads to $\theta = 1$. This motivates the following definition which allows to characterize the involved constants. It has to be understood in the sense that the data term and selected parameter are part of a sequence of problems.

Definition 1. *Problem (7) is called non-degenerate if the decomposition used in the data term is non-degenerate in the sense that $\eta, M \in O(1)$. It is called well-balanced if δ is selected such that there is a $\theta \in (0, 1)$ according to (11) with $\theta^{-1} \in O(1)$.*

The following corollary summarizes the result for a *non-degenerate* and *well-balanced* problem.

Corollary 1. *Let u and \tilde{u} be the solutions of (3) and (7), respectively. Furthermore assume that $\tilde{\ell} - \tilde{a}(u, \cdot) \in L^2(\Omega)$ and that the problem is non-degenerate and well-balanced. Then*

$$\|\tilde{u} - u\|_{L^2(\Omega)} \leq CR^2 \mathbb{E}_{\text{pde}} + C \mathbb{E}_{\text{data}}.$$

Next we will discuss the PDE error term.

Remark 1. *Under assumptions on the smoothness of $\partial\Omega$, u , and $\tilde{\alpha} - \alpha$ it can be shown, that $\tilde{\ell} - \tilde{a}(u, \cdot) \in L^2(\Omega)$ and furthermore that this term can be bounded with respect to $\tilde{\alpha} - \alpha$. To show this we first note that for all $v \in V$ we have*

$$\begin{aligned} |\tilde{\ell}(v) - \tilde{a}(u, v)| &= |\tilde{\ell}(v) - (\ell(v) - a(u, v)) - \tilde{a}(u, v)| \\ &\leq \left(\|\tilde{\ell} - \ell\|_{L^2(\Omega)} + \|\tilde{\sigma} - \sigma\|_{L^\infty(\Omega)} \|u\|_{L^2(\Omega)} \right) \|v\|_{L^2(\Omega)} + \left| \int_{\Omega} (\tilde{\alpha} - \alpha) \nabla u \cdot \nabla v \, dx \right|. \end{aligned}$$

To estimate the last term we note that for $H_0^1(\Omega) \subset V \subset H^1(\Omega)$, problem (3) corresponds to a mixed boundary value problem with

$$u = 0 \text{ on } (\partial\Omega)_D, \quad \frac{\partial u}{\partial \nu} = 0 \text{ on } (\partial\Omega)_N$$

for a decomposition $\partial\Omega = (\partial\Omega)_D \cup (\partial\Omega)_N$. For sufficiently smooth $\partial\Omega$ we then have

$$(15) \quad \left(\frac{\partial u}{\partial \nu} v \right) |_{\partial\Omega} = 0 \quad \forall v \in V.$$

Now, assuming that $\tilde{\alpha} - \alpha \in W^{1,\infty}(\Omega)$ and $u \in H^2(\Omega)$ we get

$$\begin{aligned} \left| \int_{\Omega} (\tilde{\alpha} - \alpha) \nabla u \cdot \nabla v \, dx \right| &= \left| - \int_{\Omega} \operatorname{div}((\tilde{\alpha} - \alpha) \nabla u) v \, dx + \int_{\partial\Omega} (\tilde{\alpha} - \alpha) \left(\frac{\partial u}{\partial \nu} v \right) ds \right| \\ &= \left| \int_{\Omega} \left((\tilde{\alpha} - \alpha) \Delta u + \nabla(\tilde{\alpha} - \alpha) \cdot \nabla u \right) v \, dx \right| \\ &\leq \left(\|\tilde{\alpha} - \alpha\|_{L^\infty(\Omega)} \|\Delta u\|_{L^2(\Omega)} + \|\nabla \tilde{\alpha} - \nabla \alpha\|_{L^\infty(\Omega)} \|\nabla u\|_{L^2(\Omega)} \right) \|v\|_{L^2(\Omega)} \\ &\leq \|\tilde{\alpha} - \alpha\|_{W^{1,\infty}(\Omega)} \|u\|_{H^2(\Omega)} \|v\|_{L^2(\Omega)} \quad \forall v \in V. \end{aligned}$$

Hence we have shown the PDE error bound

$$\mathbb{E}_{\text{pde}} = \|\tilde{\ell} - \tilde{a}(u, \cdot)\|_{L^2(\Omega)} \leq \|\tilde{\ell} - \ell\|_{L^2(\Omega)} + \|\tilde{\alpha} - \alpha\|_{W^{1,\infty}(\Omega)} \|u\|_{H^2(\Omega)} + \|\tilde{\sigma} - \sigma\|_{L^\infty(\Omega)} \|u\|_{L^2(\Omega)}.$$

Remark 2. *It can also be shown, that such a bound on $\|a(u, \cdot) - \tilde{a}(u, \cdot)\|_{L^2(\Omega)}$ does in general not exist if $\tilde{\alpha} - \alpha$ is not sufficiently smooth—regardless of the smoothness of u . To see this*

let $\Omega = (-1, 1)$, $(\tilde{\alpha} - \alpha)(x) = \beta(x) = -\text{sgn}(x) \in L^\infty(\Omega)$, and $u \in C^\infty([-1, 1])$ such that $u|_{[-0.5, 0.5]}(x) = x$. Now let $v_\varepsilon(x) = \max\{0, 1 - \varepsilon^{-1}|x|\}$ for $\varepsilon < 0.5$. Then

$$\int_{\Omega} \beta \nabla u \cdot \nabla v_\varepsilon \, dx = \int_{[-\varepsilon, \varepsilon]} (-\text{sgn}(x))(-\text{sgn}(x)\varepsilon^{-1}) \, dx = 2$$

while $\|v_\varepsilon\|_{L^2(\Omega)} \rightarrow 0$ for $\varepsilon \rightarrow 0$.

Finally, we will provide an error bound for the data error term \mathbb{E}_{data} in the special case of point-wise approximations

$$(16) \quad b_i = u(p_i) \approx f_{B_i} u$$

for points $p_i \in B_i$. To make sense of this expression, we need at least $u \in C(\Omega)$. However, to derive an error bound in terms of $\text{diam}(B_i)$ we will also need to relate $u(p_i) - u(x)$ to $p_i - x$ for points $x \in B_i$, independently of p_i . Hence we need to make additional assumptions on the regularity of u .

Proposition 2. *Assume $u \in C^{0,\lambda}(\overline{\Omega})$ for $\lambda \in (0, 1]$, i.e. u is λ -Hölder continuous or Lipschitz continuous (for $\lambda = 1$) and that the data $\widetilde{b(u, \cdot)}$ is given by (4) with (16). Then we can bound the data error term \mathbb{E}_{data} according to*

$$\mathbb{E}_{\text{data}} = \|\widetilde{b(u, \cdot)} - b(u, \cdot)\|_{L^2(\Omega)} \leq \|u\|_{C^{0,\lambda}(\overline{\Omega})} M |\Omega|^{1/2} r^\lambda.$$

Proof. Let $v \in L^2(\Omega)$ and $i = 1, \dots, m$. Then

$$|b_i - f_{B_i} u| = |B_i|^{-1} \left| \int_{B_i} u(p_i) - u(x) \, dx \right| \leq \|u\|_{C^{0,\lambda}(B_i)} \text{diam}(B_i)^\lambda \leq \|u\|_{C^{0,\lambda}(\overline{\Omega})} r^\lambda.$$

Using this estimate and proceeding as in the proof of Lemma 1 we get

$$\begin{aligned} |\widetilde{b(u, v)} - b(u, v)| &= \left| \sum_{i=1}^m (b_i - f_{B_i} u) |B_i| f_{B_i} v \right| \\ &\leq \|u\|_{C^{0,\lambda}(\overline{\Omega})} r^\lambda \sum_{i=1}^m |B_i| |f_{B_i} v| \leq \|u\|_{C^{0,\lambda}(\overline{\Omega})} r^\lambda M |\Omega|^{1/2} \|v\|_{L^1(\Omega)}. \end{aligned}$$

□

Now we summarize the results including the estimates for the PDE and data error terms.

Corollary 2. *Let u and \tilde{u} be the solutions of (3) and (7), respectively. Furthermore assume that*

- the problem $\widetilde{b(u, \cdot)}$ non-degenerate and well-balanced,
- the data $\widetilde{b(u, \cdot)}$ is given by (4) and (16),
- the exact and inexact coefficients satisfy $\tilde{\alpha} - \alpha \in W^{1,\infty}(\Omega)$,
- $\partial\Omega$ is smooth enough such that (15) holds true,
- the solution satisfies $u \in H^2(\Omega) \cap C^{0,\lambda}(\overline{\Omega})$ for $\lambda \in (0, 1]$.

Then

$$\|\tilde{u} - u\|_{L^2(\Omega)} \leq CR^2 (\|\tilde{\ell} - \ell\|_{L^2(\Omega)} + \|\tilde{\sigma} - \sigma\|_{L^\infty(\Omega)} + \|\tilde{\alpha} - \alpha\|_{W^{1,\infty}(\Omega)}) + Cr^\lambda.$$

3.3. Pure data fitting. As a special case of the setting outlined above, one can consider the pure data fitting problem, where nothing is known about the partial differential equation satisfied by u . In this case, one could consider using only the data in terms of a least squares ansatz

$$(17) \quad \tilde{u} \in V : \quad \|\tilde{u} - u\|_b^2 \leq \|v - u\|_b^2 \quad \forall v \in V,$$

or, equivalently,

$$\tilde{u} \in V : \quad b(\tilde{u}, v) = b(u, v) \quad \forall v \in V.$$

Due to the finite rank of $b(\cdot, \cdot)$ this problem is in general under determined and thus ill-posed. As a remedy, this can be considered in a finite dimensional subspace or submanifold V_h of V only, which may lead to a well-posed problem. However, this comes at the price, that the behavior of the fitted solution is largely determined by V_h .

As a simple example consider $V = H^1([0, 1])$ and $V_h = \mathcal{P}_{m-1}$. Then, for small sets B_i , the problem essentially reduces to interpolation problem in \mathcal{P}_{m-1} which may lead to uncontrollably large errors.

To overcome the ill-posedness in a controllable way, we introduce the regularized problem

$$(18) \quad \tilde{u} \in V : \quad \|\nabla \tilde{u}\|_{L^2(\Omega)}^2 + \delta^{-1} \|\tilde{u} - u\|_b^2 \leq \|\nabla v\|_{L^2(\Omega)}^2 + \delta^{-1} \|v - u\|_b^2 \quad \forall v \in V.$$

Noting that this takes the form of (7) with $\tilde{\alpha} = 1$ and $\tilde{f} = 0$ and exact data $\widetilde{b(u, \cdot)} = b(u, \cdot)$ we can utilize Theorem 1 to get an error estimate.

Theorem 2. *Assume that $V \subset H_0^1(\Omega)$, $u \in V \cap H^2(\Omega)$ and let \tilde{u} be the solution of (7) with $\tilde{\alpha} = 1$ and $\tilde{f} = 0$. Then*

$$(19) \quad \|\tilde{u} - u\|_{L^2(\Omega)} \leq \eta \theta^{-1} \frac{R^2 M}{\pi^2} \|\Delta u\|_{L^2(\Omega)} + \eta \theta^{-1} \|\widetilde{b(u, \cdot)} - b(u, \cdot)\|_{L^2(\Omega)}.$$

Proof. We only need to note that $u \in H^2(\Omega)$ solves the PDE (3) with $f = -\Delta u \in L^2(\Omega)$ and $\alpha = 1$. Then Theorem 1 provides the error estimate with $\tilde{\ell} - \tilde{a}(u, \cdot) = 0 - a(u, \cdot) = -f = \Delta u$. \square

In the case of exact data $\widetilde{b(u, \cdot)} = b(u, \cdot)$ in (18) the estimate reduces to

$$\|\tilde{u} - u\|_{L^2(\Omega)} \leq \eta \theta^{-1} \frac{R^2 M}{\pi^2} \|\Delta u\|_{L^2(\Omega)}.$$

It can also be shown that the result is quasi-optimal in a certain sense. To illustrate this, we investigate the question of how well we can approximate u only from the given data $f_{B_i} u$, $i = 1 \dots, m$ and the known regularity $u \in H_0^1(\Omega)$. Unfortunately, there is an infinite dimensional affine subspace

$$V_b = \left\{ v \in H_0^1(\Omega) \mid f_{B_i} v = f_{B_i} u, i = 1, \dots, b \right\}.$$

All functions in this space provide a perfect fit to the data and cannot be distinguished in terms of the available information. Hence the best we can afford in view of the regularity $u \in H_0^1(\Omega)$ is to compute a norm minimizing approximation in V_b , i.e.

$$u_{\text{opt}} = \operatorname{argmin}_{v \in V_b} \|\nabla v\|_{L^2(\Omega)}^2.$$

Using the orthogonality

$$(\nabla u_{\text{opt}}, \nabla(u_{\text{opt}} - v)) = 0 \quad \forall v \in V_b$$

and the fact that $\|u_{\text{opt}} - u\|_b = 0$ we obtain the identity

$$\tilde{c}(u_{\text{opt}} - u, u_{\text{opt}} - u) = (-\nabla u, \nabla(u_{\text{opt}} - u))$$

with $\tilde{c}(\cdot, \cdot)$ as in Theorem 2. Now assuming the additional regularity $u \in H^2(\Omega)$ and utilizing the coercivity from Lemma 6 we obtain

$$\|u_{\text{opt}} - u\|_{L^2(\Omega)}^2 \leq \eta \frac{R^2 M}{\pi^2} (\Delta u, u_{\text{opt}} - u) \leq \eta \frac{R^2 M}{\pi^2} \|\Delta u\|_{L^2(\Omega)} \|u_{\text{opt}} - u\|_{L^2(\Omega)}.$$

Thus—if we chose the optimal weighting parameter δ^{-1} such that $\theta = 1$ —the error estimate for \tilde{u} coincides with the one for u_{opt} which is the energy minimizing one among all functions that fit the data. It should be noted that the only reason for the inequality ‘ \leq ’ in the estimate for u_{opt} is the application of the coercivity estimate from Lemma 6 and the application of the Cauchy–Schwarz inequality in $L^2(\Omega)$. Hence we cannot expect to be able to derive a better estimate unless we sharpen the coercivity bound.

4. GENERALIZED GALERKIN DISCRETIZATION

4.1. Abstract a priori error estimate. Now we investigate the discretization of the PDE-regularized problem (7). To this end we consider an abstract generalized Galerkin discretization

$$(20) \quad \tilde{u}_h = \operatorname{argmin}_{v \in V_h} \tilde{J}(v),$$

of the minimization formulation (5) of (7) in a subset $V_h \subset V$. We call this problem a *generalized Galerkin discretization* since we do not require that V_h is a closed subspace of V as in a classical Galerkin discretization. As a consequence, existence and uniqueness of \tilde{u}_h is not guaranteed. However, if V_h is a closed subspace, then (20) is equivalent to the variational equation

$$(21) \quad \tilde{u}_h \in V_h : \quad \tilde{c}(\tilde{u}_h, v) = \tilde{r}(v) \quad \forall v \in V_h,$$

and uniqueness and existence is guaranteed by the Lax–Milgram theorem.

For several reasons we cannot directly apply the classical Céa-Lemma or Galerkin orthogonality to derive an error bound: The set V_h is in general not a subspace, we want to bound the error in the $L^2(\Omega)$ -norm, and the bilinear form incorporates data dependent weighting factors. Furthermore we are interested in directly bounding $\|\tilde{u}_h - u\|_{L^2(\Omega)}$ and not just $\|\tilde{u}_h - \tilde{u}\|_{L^2(\Omega)}$ for the auxiliary continuous solution \tilde{u} . As a remedy we will first use a nonlinear Céa-Lemma in terms of the weighted norm $\sqrt{\Gamma}\|\cdot\|_{\tilde{c}}$ and then go over to the $\|\cdot\|_{L^2(\Omega)}$ norm using Lemma 6. The idea of using generalized versions of Céa’s lemma for discretization in subsets goes back to [8, 13] where this was developed in a metric space setting. Since our setting is more special, we give a direct proof here.

Lemma 7. *Let $\tilde{u}_h \in V_h$ be a solution of (20) and $\tilde{u} \in V$ the solution of (5). Then we have*

$$\|\tilde{u}_h - \tilde{u}\|_{\tilde{c}} \leq \inf_{v \in V_h} \|v - \tilde{u}\|_{\tilde{c}}.$$

Proof. First we note that, in view of (7), we have for arbitrary $v \in V$

$$\|v - \tilde{u}\|_{\tilde{c}}^2 = \|v\|_{\tilde{c}}^2 - 2\tilde{c}(\tilde{u}, v) + \|\tilde{u}\|_{\tilde{c}}^2 = \|v\|_{\tilde{c}}^2 - 2\tilde{r}(v) + \|\tilde{u}\|_{\tilde{c}}^2 = 2\tilde{J}(v) + \|\tilde{u}\|_{\tilde{c}}^2.$$

Thus we get for the minimizer \tilde{u}_h of \tilde{J} in V_h

$$\|\tilde{u}_h - \tilde{u}\|_{\tilde{c}}^2 = 2\tilde{J}(\tilde{u}_h) + \|\tilde{u}\|_{\tilde{c}}^2 \leq 2\tilde{J}(v) + \|\tilde{u}\|_{\tilde{c}}^2 = \|v - \tilde{u}\|_{\tilde{c}}^2.$$

□

Theorem 3. *Let $\tilde{u}_h \in V_h$ be a solution of (20) and $u \in V$ the solution of (3). Furthermore assume that $\tilde{\ell} - \tilde{a}(u, \cdot) \in L^2(\Omega)$. Then \tilde{u}_h satisfies the error bound*

$$\begin{aligned} \|\tilde{u}_h - u\|_{L^2(\Omega)} &\leq \sqrt{\Gamma} \|\tilde{u}_h - u\|_{\tilde{c}} \leq 2\sqrt{\Gamma} \|\tilde{u} - u\|_{\tilde{c}} + \inf_{v \in V_h} \sqrt{\Gamma} \|v - u\|_{\tilde{c}} \\ &\leq 2\Gamma \left(\mathbb{E}_{\text{pde}} + \delta^{-1} \mathbb{E}_{\text{data}} \right) + \inf_{v \in V_h} \sqrt{\Gamma} \|v - u\|_{\tilde{c}} \end{aligned}$$

with the PDE- and data error terms \mathbb{E}_{pde} and \mathbb{E}_{data} as defined in Theorem 1.

Proof. Using Lemma 6, the triangle inequality, and the generalized Céa-Lemma 7 we get

$$\begin{aligned} \|\tilde{u}_h - u\|_{L^2(\Omega)} &\leq \sqrt{\Gamma} \|\tilde{u}_h - u\|_{\tilde{c}} \leq \sqrt{\Gamma} \|\tilde{u}_h - \tilde{u}\|_{\tilde{c}} + \sqrt{\Gamma} \|\tilde{u} - u\|_{\tilde{c}} \\ &\leq \sqrt{\Gamma} \|v - \tilde{u}\|_{\tilde{c}} + \sqrt{\Gamma} \|\tilde{u} - u\|_{\tilde{c}} \leq \sqrt{\Gamma} \|v - u\|_{\tilde{c}} + 2\sqrt{\Gamma} \|\tilde{u} - u\|_{\tilde{c}} \quad \forall v \in V_h. \end{aligned}$$

Now using Theorem 1 we get the assertion. □

While the first error term is the same as the one in the continuous case, we still need to take care for the best approximation error because it involves the weighted data dependent norm $\sqrt{\Gamma} \|\cdot\|_{\tilde{c}}$. The main ingredient is the following bound on this norm.

Lemma 8. *The weighted norm $\sqrt{\Gamma} \|\cdot\|_{\tilde{c}}$ can be bounded according to*

$$\sqrt{\Gamma} \|v\|_{\tilde{c}} \leq \sqrt{\eta\theta^{-1}M} \left(\frac{R}{\pi} \frac{\sqrt{\tilde{\alpha}_{\max}}}{\sqrt{\tilde{\alpha}_{\min}}} \|\nabla v\|_{L^2(\Omega)} + \|v\|_{L^2(\Omega)} \right).$$

Proof. As a direct consequence of the bounds in (12) and Lemma 1 we get

$$\Gamma \|v\|_{\tilde{c}}^2 \leq \eta\theta^{-1}M \left(\frac{R^2 \tilde{\alpha}_{\max}}{\pi^2 \tilde{\alpha}_{\min}} \|\nabla v\|_{L^2(\Omega)}^2 + \|v\|_{L^2(\Omega)}^2 \right).$$

□

To interpret this estimate we again consider a *non-degenerate, well-balanced* problem. In this case the estimate in Lemma 8 takes the form

$$\sqrt{\Gamma} \|v\|_{\tilde{c}} \leq C \left(R \|\nabla v\|_{L^2(\Omega)} + \|v\|_{L^2(\Omega)} \right)$$

where R is bounded and even decreasing if the data points cover the domain better and better. We summarize the result in the following corollary.

Corollary 3. *Let $\tilde{u}_h \in V_h$ be a solution of (20) and $u \in V$ the solution of (3). Furthermore assume that $\tilde{\ell} - \tilde{a}(u, \cdot) \in L^2(\Omega)$ and that the problem is non-degenerate and well-balanced. Then*

$$\|\tilde{u}_h - u\|_{L^2(\Omega)} \leq C \mathbb{E}_{\text{approx}} + CR^2 \mathbb{E}_{\text{pde}} + C \mathbb{E}_{\text{data}}$$

with the PDE and data error terms as in Theorem 1 and the approximation-error term

$$\mathbb{E}_{\text{approx}} := \inf_{v \in V_h} \left(R \|\nabla(v - u)\|_{L^2(\Omega)} + \|v - u\|_{L^2(\Omega)} \right).$$

Inserting the bounds on the PDE and data error we finally get:

Corollary 4. *Let $\tilde{u}_h \in V_h$ be a solution of (20) and $u \in V$ the solution of (3). Furthermore assume that the assumptions of Corollary 2 hold true. Then*

$$\begin{aligned} \|\tilde{u}_h - u\|_{L^2(\Omega)} &\leq CR^2(\|\tilde{\ell} - \ell\|_{L^2(\Omega)} + \|\tilde{\sigma} - \sigma\|_{L^\infty(\Omega)} + \|\tilde{\alpha} - \alpha\|_{W^{1,\infty}(\Omega)}) \\ &\quad + Cr^\lambda + C \inf_{v \in V_h} \left(R\|\nabla(v - u)\|_{L^2(\Omega)} + \|v - u\|_{L^2(\Omega)} \right). \end{aligned}$$

It is important to note that we neither require that V_h is a subspace, nor that the solution to (20) is unique. Hence the error estimate is in principle applicable to nonlinear approximation schemes. However, it will in general be hard to compute a global minimizer in V_h as required in (20) since practical optimization schemes often at most guarantee local optimality.

In case of inexact or local minimization we can at least bound the error in terms of the algebraic energy error. Let $\tilde{\tilde{u}}_h$ an approximation of \tilde{u}_h . Then using a straight forward modification of the proof of Lemma 7 we get

$$\|\tilde{\tilde{u}}_h - \tilde{u}\|_{\tilde{c}} \leq \inf_{v \in V_h} \|v - \tilde{u}\|_{\tilde{c}} + \sqrt{2(\tilde{J}(\tilde{\tilde{u}}_h) - \tilde{J}(\tilde{u}_h))}.$$

Thus, if we want to bound $\|\tilde{\tilde{u}}_h - u\|_{L^2(\Omega)}$ instead of $\|\tilde{u}_h - u\|_{L^2(\Omega)}$ we have to add the algebraic error term

$$\sqrt{2\Gamma(\tilde{J}(\tilde{\tilde{u}}_h) - \tilde{J}(\tilde{u}_h))} \leq CR\sqrt{\tilde{J}(\tilde{\tilde{u}}_h) - \tilde{J}(\tilde{u}_h)}$$

to the right hand sides of the estimates in Theorem 3, Corollary 3, and Corollary 4.

4.2. Finite element discretization. As an example for a linear Galerkin discretization we apply the abstract error bound to a finite element ansatz. To this end let $V_h \subset V$ be a conforming Lagrange finite element space of order k (piecewise polynomials in \mathcal{P}_k for simplex elements or piecewise tensor-polynomials in \mathcal{Q}_k for cubic elements) on a triangulation with mesh size h . Then we can apply the classical finite element interpolation error estimate which, in the present special case, provides:

Proposition 3. *Let $\Pi_h : C^0(\bar{\Omega}) \rightarrow V_h$ interpolation operator and $u \in H^{k+1}(\Omega) \cap C^0(\bar{\Omega})$. Then*

$$\|u - \Pi_h u\|_{L^2(\Omega)} \leq Ch^{k+1}|u|_{H^{k+1}(\Omega)}, \quad \|u - \Pi_h u\|_{H^1(\Omega)} \leq Ch^k|u|_{H^{k+1}(\Omega)}.$$

with a constant C depending only on the shape regularity of the triangulation. Here $|\cdot|_{H^{k+1}(\Omega)}$ denotes the $H^{k+1}(\Omega)$ -semi-norm containing only derivatives of order $k+1$.

For a proof we refer to [3, Theorem 3.2.1 and Remark 3.2.2]. Inserting the interpolation error estimate in the abstract error bound we get the following total error bound.

Corollary 5. *Let $\tilde{u}_h \in V_h$ be a solution of (20) and $u \in V$ the solution of (3). Furthermore assume that the assumptions of Corollary 2 and $u \in H^{k+1}(\Omega)$ hold true. Then*

$$\|\tilde{u}_h - u\|_{L^2(\Omega)} \leq CR^2(\|\tilde{\ell} - \ell\|_{L^2(\Omega)} + \|\tilde{\sigma} - \sigma\|_{L^\infty(\Omega)} + \|\tilde{\alpha} - \alpha\|_{W^{1,\infty}(\Omega)}) + Cr^\lambda + CRh^k + Ch^{k+1}.$$

4.3. A heuristic strategy for parameter selection. To guarantee *non-degenerate* and *well-balanced* problem in the sense of Definition 1 the size of the averaging sets B_i and the parameter δ should be carefully selected such that the quotient $|K_i|/|B_i|$ is bounded and such that δ scales like R^2M . In applications, where data points $p_1, \dots, p_m \in \Omega$ are given we can in general not assume that one is able to compute a decomposition into sets K_i and their maximal diameter R exactly. While this is in principle possible using a Voronoi decomposition, computing the latter is in general much to costly. Hence we propose a heuristic strategy to

determine the sets B_i and estimates for R and δ under the assumption of uniformly distributed data points p_i .

To this end we make the assumption that Ω fits into a cube $[0, D]^d$. Then we can expect that the points are in average uniformly spaced. For an ideal uniformly spaced distribution of m points in $[0, D]^d$ we can place m non-overlapping cubes \hat{K}_i with edge length $\hat{L} = Dm^{-1/d}$ on a uniform lattice into $\Omega = [0, D]^d$. The diameter of these boxes is

$$\hat{R} := \hat{L}\sqrt{d} = Dm^{-1/d}\sqrt{d}$$

which we will use as estimate for R . In order to fix $|\hat{K}_i|/|B_i| = Q$ for a parameter $Q > 1$, we will use cubes B_i centered at p_i with edge length \hat{l}

$$B_i := p_i + \hat{l}[-0.5, 0.5]^d, \quad \hat{l} := \hat{L}Q^{-1/d} = D(mQ)^{-1/d}.$$

Since we cannot control M explicitly, we assume $M \in O(1)$ and select

$$\delta := \frac{\hat{R}^2}{\pi^2 \tilde{\alpha}_{\min}} = \frac{D^2 m^{-2/d} d}{\pi^2 \tilde{\alpha}_{\min}}.$$

5. NUMERICAL EXPERIMENTS

5.1. A smooth test problem. Finally we illustrate the theoretical findings using numerical experiments for test problems. To this end we consider an example problem from [4] given by

$$-\Delta u + \sigma u = f \quad \text{in } \Omega, \quad \frac{\partial u}{\partial n} = 0 \quad \text{on } \Omega$$

with $\Omega = [0, 1]^d$. For the Eigenfunction $u(x) = \sum_{i=1}^d \cos(\pi x_i)$ of the Laplacian and $\sigma = \pi^2$ this equation is satisfied with $f(x) = (\pi^2 + \sigma)u(x) = 2\pi^2 u(x)$. We do not want to hide, that this example was selected, because—in contrast to Dirichlet boundary conditions—natural boundary conditions $\frac{\partial u}{\partial n} = 0$ do not require any approximation for neural network discretizations which was not covered in the error analysis.

As inexact auxiliary PDE we will use the same differential operator $-\Delta + \sigma$ but a scaled right hand side $\tilde{f} = (1 - \varepsilon)f$. Then we can explicitly compute the PDE error term

$$\mathbb{E}_{\text{pde}} = \|f - \tilde{f}\|_{L^2(\Omega)} = \varepsilon \|f\|_{L^2(\Omega)}.$$

In all experiments we generate data by creating m uniformly distributed random points p_1, \dots, p_m and use cubes B_i with edge length \hat{l} centered at the points p_i . The points are sampled in $[\hat{l}/2, 1 - \hat{l}/2]^d$ such that $B_i \subset \Omega$ is guaranteed. The edge length \hat{l} and the penalty parameter δ are selected according to the heuristic strategy proposed in Subsection 4.3 for different values of Q . Notice that the heuristic strategy does not guarantee a *non-degenerate* and *well-balanced* problem in the sense of Definition 1. Nevertheless, the method is still covered by the theory presented above, but the constants may blow up if the strategy fails, leading to different error rates.

5.2. Finite element discretization. In this section we discretize the problem with conforming finite elements for $d = 1, 2, 3$. Since we are not interested in illustrating well known finite element error bounds, we use a fixed finite element grid with uniformly spaced interval/rectangular/hexahedral elements and tensorial \mathcal{Q}_k Lagrange finite elements of order k throughout the experiments. For $d = 1$ we used 64 elements and order $k = 4$, for $d = 2$ we used $4069 = 64 \times 64$ elements and order $k = 4$, and for $d = 3$ we used $4069 = 16 \times 16 \times 16$ elements and order $k = 2$. In any case the discretization error can be neglected compared to the PDE

and data error terms reported in the following. All reported $L^2(\Omega)$ -errors are computed using a quadrature rule of order k^2 on the grid elements.

For the first experiment we fix $\tilde{f} = 0.5f$ and compute the solution for increasing number of data points m . The computations are done using exact data, i.e., $b_i = \mathcal{f}_{B_i} u$ and inexact data in the form of point values $b_i = f(p_i)$. Figure 1 depicts the error over m for dimensions $d = 1, 2, 3$ and parameters $Q = 4, 2$ in the heuristic strategy. Left and right picture show exact and inexact data, respectively. We observe that the error decays like $O(m^{-2/d}) = O(\hat{R}^2)$ as expected from the theoretical error bounds for exact data and uniformly distributed data points. We also find that the error increases by a constant if we increase $Q = |\hat{K}_i|/|B_i|$ which is again in accordance with the error estimate where this enters in the constant factor η . For inexact data, the situation is very similar. We observe the order $O(m^{-2/d}) = O(\hat{R}^2)$ but for very small errors obtained for $d = 1$ eventually the case $Q = 4$ becomes better compared to $Q = 2$. This can be explained by the fact that, while increasing Q increases the constant in the PDE error term, it also decreases r and thus the data error term. Hence, if the data error comes in the range of the PDE error, we expect that larger Q reduces the total error.

In the second experiment we fix $m = 512$ data points and compute the solution for inexact right hand sides $\tilde{f} = (1 - \varepsilon)f$ with varying ε and exact data $b_i = \mathcal{f}_{B_i} u$ as well as inexact data $b_i = f(p_i)$. Figure 2 depicts the error over the PDE error $\varepsilon = \mathbb{E}_{\text{pde}}/\|f\|_{L^2(\Omega)}$ for dimensions $d = 1, 2, 3$ and parameters $Q = 4, 2$ in the heuristic strategy. Left and right picture show exact and inexact data, respectively. Here we observe that the error scales like $O(\varepsilon) = O(\mathbb{E}_{\text{pde}})$ for a wide range of ε . This is expected from the theoretical error bound. Again, we observe that for small total error $Q = 4$ is better compared to $Q = 2$ where the error finally saturates for $d = 2, 3$. The latter indicates that the data error starts to dominate in this regime such that we no longer benefit from improving the PDE error.

5.3. Neural network discretization. Finally we consider the discretization of the test problem with neural networks. To this end we minimize the loss functional \tilde{J} defined in (6) over a set V_h of neural networks with a fixed architecture. In this case integrals are no longer evaluated exactly, but approximated using stochastic integration as proposed in [4]. More precisely, we approximate the local averages $\mathcal{f}_{B_i} v$ appearing in the functional by averaging v over $10d$ uniformly distributed sampling points in B_i . Since all sets B_i are the same up to translation, we also translated the sampling points. The integration over Ω was approximated using $100d$ uniformly distributed sampling points in Ω that are newly sampled in each step of the iterative algebraic solution method. Since we use natural boundary conditions, we only need to approximate the space $H^1(\Omega)$ without any hard boundary constraints. Hence no approximation of boundary conditions was needed.

The network architecture is as follows: A first layer inflates the input size to 16. This is followed by 3 blocks, each consisting of two densely connected layers with a residual connection. A final affine layer reduces the size from 16 to 1. In total the network takes the form

$$F = F_{\text{out}} \circ F_3 \circ F_2 \circ F_1 \circ F_{\text{in}}$$

where $F_{\text{in}} : \mathbb{R}^d \rightarrow \mathbb{R}^{16}$ is a linear map padding its input by $16 - d$ zeros, $F_{\text{out}} : \mathbb{R}^{16} \rightarrow \mathbb{R}$ is an affine map, and each block $F_i : \mathbb{R}^{16} \rightarrow \mathbb{R}^{16}$ has the form

$$F_i(x) = \psi(W_{i,2}\psi(W_{i,1}x - \theta_{i,1}) - \theta_{i,2}) + x$$

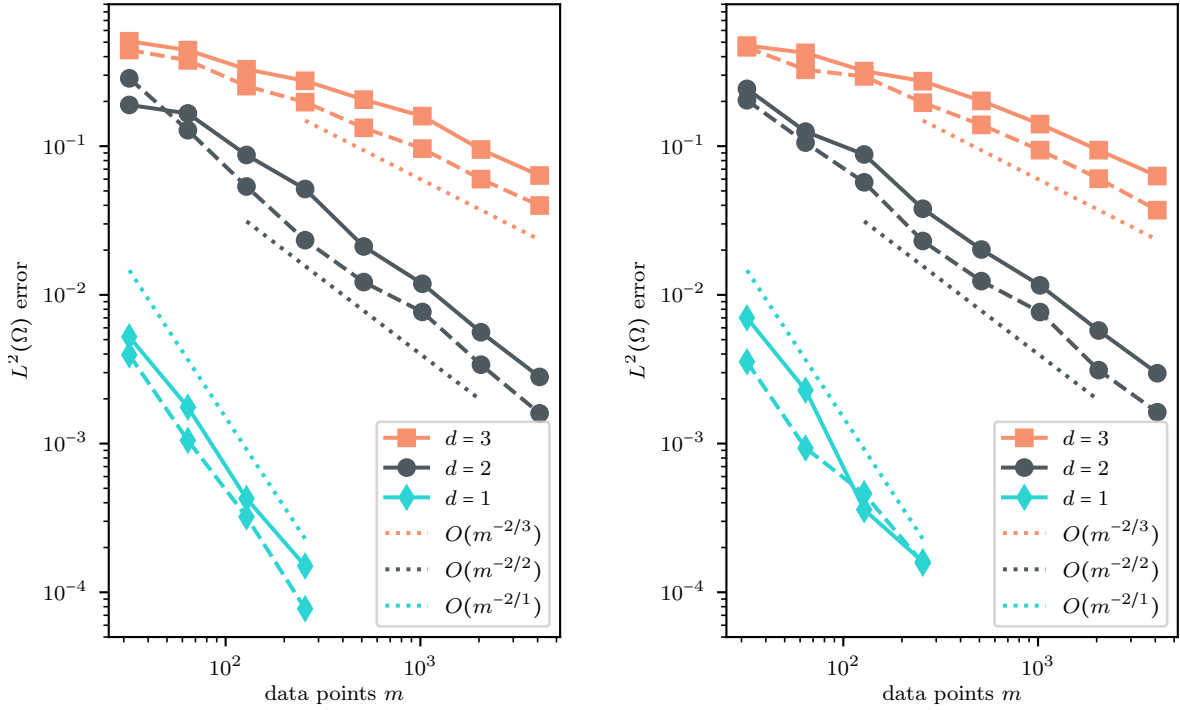


FIGURE 1. Finite element discretization. Total error over number of data points. Solid lines: $Q = 4$. Dashed lines: $Q = 2$. Dotted lines: Reference slope. Left: Exact local average data. Right: Point values as inexact data.

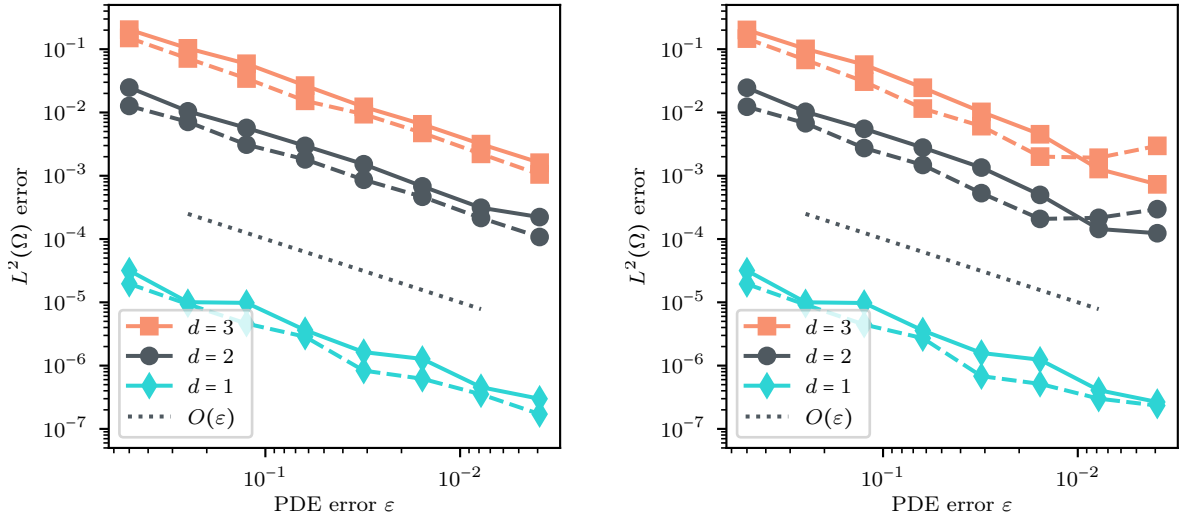


FIGURE 2. Finite element discretization. Total error over PDE error for exact local average data. Solid lines: $Q = 4$. Dashed lines: $Q = 2$. Dotted lines: Reference slope. Left: Exact local average data. Right: Point values as inexact data.

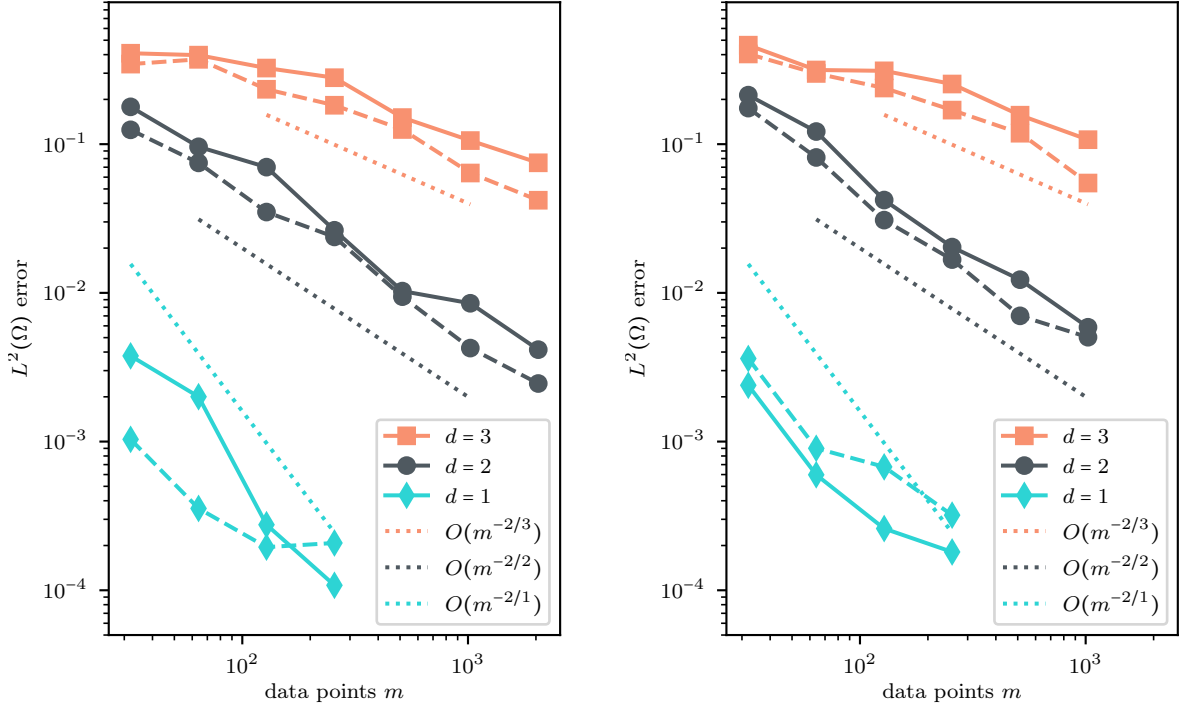


FIGURE 3. Neural network discretization. Total error over number of data points. Solid lines: $Q = 4$. Dashed lines: $Q = 2$. Dotted lines: Reference slope. Left: Exact local average data. Right: Point values as inexact data.

for dense weight matrices $W_{i,j} \in \mathbb{R}^{16 \times 16}$ and bias vectors $\theta_{i,j} \in \mathbb{R}^{16}$. In all layers we used the activation function $\psi(x) = \max\{x^3, 0\}$. All parameters involved in $F_1, F_2, F_3, F_{\text{out}}$ are determined in the training procedure.

The networks are trained using the Adam method [15] which is a variant of stochastic gradient descent. Training was stopped if the $L^2(\Omega)$ error $\|\tilde{u}_h - u\|_{L^2(\Omega)}$ did not decrease any more. While this is in general impractical due to the unknown solution u , we used this here to avoid effects of more heuristic stopping criteria.

Again we fix $\tilde{f} = 0.5f$ and compute the solution for increasing number of data points m . The computations are done using exact data, i.e., $b_i = f_{B_i} u$ and inexact data in the form of point values $b_i = f(p_i)$. Figure 3 depicts the error over m for dimensions $d = 1, 2, 3$ and parameters $Q = 4, 2$ in the heuristic strategy. Left and right picture show exact and inexact data, respectively. For $d = 2$ and $d = 3$ we again observe that the error decays like $O(m^{-2/d}) = O(\hat{R}^2)$. For $d = 1$ the situation is less clear. While we roughly observe the order $O(m^{-2/d}) = O(\hat{R}^2)$ again, there are some exceptions. Most importantly the error increases after adding more data in one case.

REFERENCES

- [1] H. Bölskei, P. Grohs, G. Kutyniok, and P. Petersen. Optimal approximation with sparsely connected deep neural networks. *SIAM J. Math. Data Sci.*, 1(1):8–45, 2019. arxiv:1705.01714.
- [2] M. Burger and A. Neubauer. Error bounds for approximation with neural networks. *J. Approx. Theory*, 112(2):235–250, 2001.

- [3] P. G. Ciarlet. *The finite element method for elliptic problems*, volume 4 of *Studies in Mathematics and its Applications*. North-Holland, Amsterdam New York Oxford, 1978.
- [4] W. E and B. Yu. The deep Ritz method: A deep learning-based numerical algorithm for solving variational problems. *Comm. Math. Stats.*, 6(1):1–12, 2018.
- [5] K.-I. Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural networks*, 2(3):183–192, 1989.
- [6] C. Gräser. A note on Poincaré- and Friedrichs-type inequalities, 2015. arxiv:1512.02842.
- [7] R. Griboval, G. Kutyniok, M. Nielsen, and F. Voigtlaender. Approximation spaces of deep neural networks, 2019. arxiv:1905.01208.
- [8] P. Grohs, H. Hardering, and O. Sander. Optimal a priori discretization error bounds for geodesic finite elements. *Found. Comput. Math.*, 15:1357–1411, 2015.
- [9] P. Grohs, F. Hornung, A. Jentzen, and P. von Wurstemberger. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations, 2018. arxiv:1809.02362.
- [10] I. Gühring, G. Kutyniok, and P. Petersen. Error bounds for approximations with deep ReLU neural networks in $W^{s,p}$ norms, 2019.
- [11] B. D. Haeffele and R. Vidal. Global optimality in neural network training. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7331–7339, 2017.
- [12] J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci.*, 115(34):8505–8510, 2018. arXiv:1707.02568.
- [13] H. Hardering. *Intrinsic Discretization Error Bounds for Geodesic Finite Elements*. PhD thesis, Freie Universität Berlin, 2015.
- [14] E. Kharazmi, Z. Zhang, and G. E. Karniadakis. Variational physics-informed neural networks for solving partial differential equations, 2019. arxiv:1912.00873.
- [15] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014. arxiv:1912.03937.
- [16] G. Kutyniok, P. Petersen, M. Raslan, and R. Schneider. A theoretical analysis of deep neural networks and parametric PDEs, 2019.
- [17] Y. Liao and P. Ming. Deep Nitsche method: Deep Ritz method with essential boundary conditions, 2019.
- [18] J. Müller and M. Zeinhofer. Deep Ritz revisited, 2019. arxiv:1912.03937.
- [19] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5(1):286–292, 1960.
- [20] D. Perekrestenko, P. Grohs, Elbrächter D., and H. Bölskei. The universal approximation power of finite-width deep ReLU networks, 2018.
- [21] P. Petersen and F. Voigtlaender. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *Neural Networks*, 108:296–330, 2018.
- [22] A. Pinkus. Approximation theory of the MLP model in neural networks. *Acta numerica*, 8:143–195, 1999.
- [23] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics informed deep learning (part I): Data-driven solutions of nonlinear partial differential equations, 2017. arxiv:1711.10561.
- [24] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics informed deep learning (part II): Data-driven discovery of nonlinear partial differential equations, 2017. arxiv:1711.10566.
- [25] G. M. Rotskoff and E. Vanden-Eijnden. Neural networks as interacting particle systems: Asymptotic convexity of the loss landscape and universal scaling of the approximation error, 2018. arxiv:1805.00915.
- [26] Z. Shen, H. Yang, and S. Zhang. Nonlinear approximation via compositions. *Neural Networks*, 119:74–84, Nov 2019.
- [27] R. Vidal, J. Bruna, R. Giryes, and S. Soatto. Mathematics of deep learning, 2017. arxiv:1712.04741.
- [28] D. Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Networks*, 94:103–114, 2017.