A RESIDUAL MINIMIZATION APPROACH FOR NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS SET IN BANACH SPACES*

IGNACIO MUGA†, JORGE PERERA‡, SERGIO ROJAS§, AND RICARDO RUIZ-BAIER¶

Abstract. In this work, we propose and analyze a residual-minimization strategy for the numerical solution of nonlinear PDEs posed in Banach spaces. Given a finite-dimensional trial space and a suitably enriched discrete test space (of higher dimension than the trial space), we approximate the solution by minimizing the variational residual in a discrete dual norm. This minimization is equivalent to a nonlinear saddle-point formulation for the discrete solution in the trial space together with a residual representative in the test space. The latter provides a natural a posteriori error estimator, enabling automatic mesh adaptivity. To solve the resulting nonlinear saddle-point problem, we propose a Newton iteration whose linearized saddle-point system is symmetric, thereby guaranteeing solvability at each step. We take the p-Laplacian as a model problem and support the theoretical developments with representative numerical experiments, using standard H^1-conforming piecewise linear functions for the trial space, and lowest-order Crouzeix–Raviart functions for the test space.

Key words. Residual minimization, nonlinear PDEs, Banach spaces, p-Laplacian, adaptivity.

MSC codes. 65N30, 65N12, 65N15.

1. Introduction.

1.1. Scope. Partial Differential Equations (PDEs) are commonly expressed in the abstract form:

$$\begin{cases} \text{Find } u \in \mathbb{U} : \\ A(u) = F \quad \text{in } \mathbb{V}^*, \end{cases} \tag{1.1}$$

where the mapping $A : \mathbb{U} \to \mathbb{V}^*$ represents the underlying PDE, $\mathbb{U}$ and $\mathbb{V}$ are infinite-dimensional Banach spaces and $F \in \mathbb{V}^*$ is a given functional. Although standard variational formulations form the basis of finite element discretizations, residual minimization (MinRes) methods have emerged as a robust alternative framework. The core idea of MinRes is to reformulate the PDE as an optimization problem by minimizing the norm of the residual $F - A(u)$ in a suitable dual space.

Different MinRes strategies based on finite elements have been proposed to numerically approximate the solution(s) of (1.1) in specific contexts. For instance, when the right-hand side functional F is in a Lebesgue space L^2, the least-squares method [2] can be applied, which reformulates PDE problems as the minimization of a functional that measures the residual error of the equation in the L^2 norm. In the context

*__Updated:__ April 2, 2026.

Funding: This work has been partially supported by ANID through FONDECYT projects 1240643 (SR) and 1230091 (IM), by the ANID National Doctoral Scholarship 21250504 (JP), by the National Center for Artificial Intelligence CENIA FB210017, Basal ANID, by the Australian Research Council through the Future Fellowship Grant FT220100496, and by the Center of Advanced Study (CAS) at the Norwegian Academy of Science and Letters under the program Mathematical Challenges in Brain Mechanics.

†Instituto de Matemáticas, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile (ignacio.muga@pucv.cl).

‡Corresponding author: Instituto de Matemáticas, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile (jorgeperera1597@gmail.com).

§School of Mathematics, Monash University, 9 Rainforest Walk, 3800 Melbourne VIC, Australia (sergio.rojas@monash.edu).

¶School of Mathematics, Monash University, 9 Rainforest Walk, Melbourne VIC 3800, Australia, and Universidad Adventista de Chile, Casilla 7-D Chillán, Chile (ricardo.ruizbaier@monash.edu).

of Hilbert spaces, the Discontinuous Petrov–Galerkin (DPG) method (see [14], and references therein for a recent overview) has established itself as a powerful methodology by minimizing the residual in dual norms, leading to optimal stability properties. More recently, these MinRes principles have been extended to the training of neural networks, such as in the Robust Variational Physics-Informed Neural Networks (RVPINNs) framework (see [26, 16]), which minimizes the discrete dual norm of the residual to guarantee stable and robust error estimation.

Other methods that minimize the residual in dual norms, or tackle negative and fractional Sobolev norms in the context of Hilbert spaces, can be found in [8, 9, 23]. Furthermore, the DPG framework has been extended to analyze nonlinear problems in Hilbert spaces, such as quasilinear elliptic PDEs [5], strongly monotone operators [4], and more recently, nonlinear evolution equations through multistage time-marching schemes [25].

Nevertheless, many physical phenomena and mathematical models are naturally governed by non-quadratic energies and, consequently, are best described in Banach spaces rather than Hilbert spaces. A classic example that arises in the modeling of non-Newtonian fluids, plasticity, glaciology, and nonlinear diffusion processes is the p-Laplacian problem: Given $p \in (1, \infty)$ and a source term $f \in (W_0^{1,p}(\Omega))^*$, find $u \in W_0^{1,p}(\Omega)$ such that:

$$\begin{cases} -\Delta_p u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1.2}$$

where $\Delta_p u := \operatorname{div}(|\nabla u|^{p-2}\nabla u)$ and $\Omega \subset \mathbb{R}^d$ is a bounded open set with a Lipschitz boundary $\partial\Omega$. For such problems, where the natural function spaces are Sobolev spaces $W^{1,p}(\Omega)$ with $p \neq 2$, restricting the numerical analysis to Hilbert structures is theoretically insufficient.

An approach generalizing residual minimization methods to Banach spaces was introduced in [24]. It is worth emphasizing that this framework, as well as the approaches in [8, 9], provides a posteriori error estimators that naturally drive adaptive mesh refinement. However, while the work in [24] successfully established the theory for continuous, bounded-below linear operators, extending it to nonlinear operators remains challenging.

Recent advancements have begun to address the highly nonlinear optimization hurdles that arise when minimizing residuals in Banach spaces, proposing strategies such as variable exponents [18] or regularized Kačanov iterations [28] for linear PDEs. Furthermore, as highlighted in [1], the inherent structural differences of the p-Laplacian necessitate analyzing the degenerate ($p \geq 2$) and singular ($1 < p < 2$) regimes separately to derive accurate stability and convergence bounds.

In this work, we aim to further bridge this gap by developing residual minimization methods for nonlinear PDEs that provide an a posteriori error estimator to automatically drive adaptivity, particularly when $\mathbb{V}$ and its dual space $\mathbb{V}^*$ are strictly convex, reflexive Banach spaces.

1.2. Outline. The remainder of this work is organized as follows. In Section 2, we present the necessary preliminaries, including definitions of functional spaces, properties of duality maps in strictly convex Banach spaces, and key results on best approximation. Section 3 is devoted to the continuous residual minimization method, establishing its well-posedness, equivalent characterizations, and a continuous a posteriori error estimator. In Section 4, we analyze the residual minimization method in discrete dual norms. We discuss its stability, provide a convergence analysis, and derive a discrete a posteriori error estimator. Finally, in Section 5, we detail the numerical implementation for the p-Laplacian problem, including the finite element spaces, the iterative algorithm, and numerical experiments validating our theoretical findings.

2. Preliminaries on functional setting. In this section, we present definitions of the function spaces we will consider in our model problem. Furthermore, we briefly explore essential results from the theory of residual minimization in Banach spaces.

Let $(\mathbb{X}, \| \ \|_{\mathbb{X}})$ be a normed space. We will denote by $\mathbb{X}^*$ the dual space of $\mathbb{X}$, and will refer to the action of $F \in \mathbb{X}^*$ over elements $x \in \mathbb{X}$, as a duality pairing between $\mathbb{X}^*$ and $\mathbb{X}$. This is,

$$\langle F, x\rangle_{\mathbb{X}^*,\mathbb{X}} := F(x). \tag{2.1}$$

Additionally, we denote the dual space $\mathbb{X}^*$ norm as

$$\| \bullet \|_{\mathbb{X}^*} := \sup_{x\in\mathbb{X}} \frac{\langle \bullet , x\rangle_{\mathbb{X}^*,\mathbb{X}}}{\|x\|_{\mathbb{X}}}. \tag{2.2}$$

Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain, where $d \in \mathbb{N}$ denotes its spatial dimension. For a given $p \geq 1$, consider the Sobolev spaces

$$W^{1,p}(\Omega) := \left\{ v \in L^p(\Omega) : \frac{\partial v}{\partial x_i} \in L^p(\Omega), \forall i \in \{1, \cdots , d\} \right\},$$

equipped with the norm

$$\| \bullet \|_{W^{1,p}(\Omega)} := \left(\| \bullet \|^p_{L^p(\Omega)} + \sum_{i=1}^{d} \left\| \frac{\partial(\bullet)}{\partial x_i} \right\|^p_{L^p(\Omega)} \right)^{1/p},$$

where $L^p(\Omega)$ stands for the standard Lebesgue space. In addition, consider the subspace $W^{1,p}_0(\Omega) \subset W^{1,p}(\Omega)$ being the closure of $\mathscr{C}^\infty_0(\Omega)$ (the space of smooth functions with compact support on Ω) in the normed space $W^{1,p}(\Omega)$. This is,

$$W^{1,p}_0(\Omega) := \overline{\mathscr{C}^\infty_0(\Omega)}^{\| \ \|_{W^{1,p}(\Omega)}}.$$

We recall that, as a consequence of the Poincaré inequality, $\| \ \|_{W^{1,p}(\Omega)}$ and

$$| \bullet |_{W^{1,p}(\Omega)} := \left(\int_\Omega |\nabla(\bullet)|^p \right)^{1/p} \tag{2.3}$$

are equivalent norms in the space $W^{1,p}_0(\Omega)$.

Since these Sobolev spaces, and their dual spaces, are strictly convex and reflexive for $1 < p < \infty$, they possess favorable geometric properties. A key tool that exploits these properties is the duality map. Although, in general, the duality map on a Banach space $\mathbb{X}$ is defined as a multivalued mapping from $\mathbb{X}$ to $2^{\mathbb{X}^*}$, when $\mathbb{X}^*$ is strictly convex it becomes a single-valued map (see [13, Prop. 12.3]). This property allows us to use the following definition of the duality map, which we will consider in this paper.

DEFINITION 2.1. *Let $\mathbb{X}$ be a Banach space such that $\mathbb{X}^*$ is strictly convex. For $p > 1$, the **duality map** $J_{p,\mathbb{X}} : \mathbb{X} \to \mathbb{X}^*$ is the unique operator that satisfies:*

i. $\langle J_{p,\mathbb{X}}(x), x\rangle_{\mathbb{X}^*,\mathbb{X}} = \|J_{p,\mathbb{X}}(x)\|_{\mathbb{X}^*} \|x\|_{\mathbb{X}}$.

ii. $\|J_{p,\mathbb{X}}(x)\|_{\mathbb{X}^*} = \|x\|^{p-1}_{\mathbb{X}}$.

The existence of $J_{p,\mathbb{X}}$ is guaranteed by the Hahn–Banach extension theorem, and the uniqueness is due to the strict convexity of $\mathbb{X}^*$. The following result considers a distinctive feature of the duality map.

PROPOSITION 2.1. *Let $\mathbb{X}$ be a strictly convex Banach space and consider $p > 1$. Let $\phi : \mathbb{X} \to \mathbb{R}$ be defined as $\phi(\bullet) := \frac{1}{p}\|\bullet\|_{\mathbb{X}}^p$. Then, ϕ is a Gâteaux differentiable functional for all $x \in \mathbb{X}$, and its derivative at $x \in \mathbb{X}$ satisfies the following identity:*

$$\nabla\phi(x) = J_{p,\mathbb{X}}(x). \tag{2.4}$$

Proposition 2.1 is very useful for determining the duality map $J_{p,\mathbb{X}}$ of a Banach space $\mathbb{X}$ from its norm. For a detailed proof, we refer the reader to [10, Th. 4.4], which establishes this relation for generalized duality maps by considering the weight function $\varphi(t) = t^{p-1}$.

The next result provides a particular case for which the bijectivity of the duality map is ensured. For a proof, see, for example, [10, Prop. 4.7].

PROPOSITION 2.2. *Let $\mathbb{X}$ and $\mathbb{X}^*$ be strictly convex and reflexive Banach spaces, and let $J_{p,\mathbb{X}} : \mathbb{X} \to \mathbb{X}^*$ be the duality map on $\mathbb{X}$ for $p > 1$. Then the following statements hold:*
(a) The duality map $J_{p,\mathbb{X}}$ is bijective.
(b) If $\mathcal{I}_{\mathbb{X}} : \mathbb{X} \to \mathbb{X}^{**}$ denotes the canonical injection and $p^* = p/(p-1)$, then

$$J_{p^*,\mathbb{X}^*} = \mathcal{I}_{\mathbb{X}} \circ J_{p,\mathbb{X}}^{-1}. \tag{2.5}$$

A last ingredient we require for the subsequent analysis is the duality map of a linear subspace.

PROPOSITION 2.3. *Let $\mathbb{X}$ be a Banach space such that $\mathbb{X}^*$ is strictly convex, and let $J_{p,\mathbb{X}} : \mathbb{X} \to \mathbb{X}^*$ be the duality map on $\mathbb{X}$. Let $\mathbb{M} \subset \mathbb{X}$ be a linear subspace of $\mathbb{X}$ and $p > 1$. Then, the duality map $J_{p,\mathbb{M}} : \mathbb{M} \to \mathbb{M}^*$ verify the following identity:*

$$I_{\mathbb{M}}^* \circ J_{p,\mathbb{X}} \circ I_{\mathbb{M}} = J_{p,\mathbb{M}}, \tag{2.6}$$

where $I_{\mathbb{M}} : \mathbb{M} \to \mathbb{X}$ is the natural injection.

Proof. The proof for $p = 2$ can be found in [24, Lem. 2.3]. For $p > 1$, it follows similarly. ∎

We now present a proposition establishing conditions for the existence and uniqueness of the best approximation from a Banach space to a nonempty, closed, and convex subset, along with its corresponding a priori bounds.

PROPOSITION 2.4 (Best approximation and a priori bounds). *Let $\mathbb{X}$ be a reflexive Banach space and let $\mathbb{M} \subset \mathbb{X}$ be a nonempty closed convex subset. Then, for every $x \in \mathbb{X}$, there exists $\hat{x} \in \mathbb{M}$ such that*

$$\|x - \hat{x}\|_{\mathbb{X}} = \inf_{y\in\mathbb{M}} \|x - y\|_{\mathbb{X}}.$$

Furthermore, if $\mathbb{X}$ is strictly convex, then $\hat{x}$ is unique. Additionally, if $0_{\mathbb{X}} \in \mathbb{M}$, the best approximation $\hat{x}$ satisfies the a priori bound:

$$\|\hat{x}\|_{\mathbb{X}} \leq 2\|x\|_{\mathbb{X}}.$$

Moreover, if $\mathbb{M}$ is a closed linear subspace, this bound can be sharpened to:

$$\|\hat{x}\|_{\mathbb{X}} \leq C_{\mathtt{BM}}(\mathbb{X})\|x\|_{\mathbb{X}},$$

where $C_{\mathtt{BM}}(\mathbb{X}) \in [1, 2]$ is the Banach–Mazur constant of $\mathbb{X}$.

Remark 2.1. *For the precise statement regarding the conditions for the existence and uniqueness of the best approximation, we refer to [6, Problem 5.14-4]. The derivation of the standard a priori bounds can be found in [24, Prop. 3.5]. It is worth noting that standard bounds in general Banach spaces can be pessimistic. To obtain the sharpened estimate involving the Banach–Mazur constant $C_{\mathtt{BM}}(\mathbb{X})$, it is necessary to exploit specific geometric properties of $\mathbb{X}$, particularly its proximity to a Euclidean space. For a formal definition of this constant and further geometric insights, see [27, Def. 2].*

3. Residual minimization in dual norms.

Let us consider a finite-dimensional subspace $\mathbb{U}_h \subset \mathbb{U}$. The residual minimization problem applied to problem (1.1) is as follows: given a $F \in \mathbb{V}^*$ and $p > 1$, find $u_h \in \mathbb{U}_h$ such that

$$u_h = \arg\min_{w_h \in \mathbb{U}_h} \frac{1}{p^*} \|F - A(w_h)\|_{\mathbb{V}^*}^{p^*}, \tag{3.1}$$

where p^* is such that $p^* p = p^* + p$. When analyzing problem (3.1), we assume the existence and uniqueness of a solution to the abstract problem (1.1).

For the problem (3.1) to be well-posed, the following assumptions are established.

Assumption 3.1. *In problem (3.1), we assume the following statements:*
1. ***Existence:*** *$\mathbb{V}^*$ is reflexive and $A(\mathbb{U}_h)$ is a nonempty closed convex subset of $\mathbb{V}^*$.*
2. ***Uniqueness:*** *$\mathbb{V}^*$ is strictly convex and $A : \mathbb{U} \to \mathbb{V}^*$ is injective.*
3. ***Stability:*** *There exists $\alpha > 0$ such that $\|w\|_{\mathbb{U}} \lesssim \|A(w)\|_{\mathbb{V}^*}^{\alpha}$ for all $w \in \mathbb{U}$.*

Indeed, assuming the above statements, from Proposition 2.4 it follows that there exists a unique $\hat{F}_h \in A(\mathbb{U}_h)$ such that

$$\left\| F - \hat{F}_h \right\|_{\mathbb{V}^*} = \min_{G_h \in A(\mathbb{U}_h)} \|F - G_h\|_{\mathbb{V}^*}. \tag{3.2}$$

In addition, since $A : \mathbb{U} \to \mathbb{V}^*$ is injective, then there exists a unique $u_h \in \mathbb{U}_h$ such that $A(u_h) = \hat{F}_h$. Thus, the existence and uniqueness of a solution to problem (3.1) is guaranteed. Hypothesis (3) is assumed to ensure the stability of the method, since Proposition 2.4 shows that

$$\|u_h\|_{\mathbb{U}} \leq \|A(u_h)\|_{\mathbb{V}^*}^{\alpha} \leq 2^{\alpha} \|F\|_{\mathbb{V}^*}^{\alpha}.$$

3.1. Useful characterizations.

Herein, we provide a characterization of the solution obtained by the MinRes method, which is a central result of this work.

Theorem 3.1 (Problem equivalence). *Let $A : \mathbb{U} \to \mathbb{V}^*$ be a Fréchet differentiable operator where $\mathbb{V}$ and $\mathbb{V}^*$ are strictly convex and reflexive Banach spaces. Let $\mathbb{U}_h$ be a finite-dimensional subspace of $\mathbb{U}$. Given $F \in \mathbb{V}^*$, the statements listed below are equivalent:*

1. *u_h is a solution to the minimization problem (3.1).*
2. *For all $p > 1$, $u_h \in \mathbb{U}_h$ is a solution of the problem:*

$$\left\langle \nabla A(u_h) w_h, J_{p,\mathbb{V}}^{-1}(F - A(u_h)) \right\rangle_{\mathbb{V}^*,\mathbb{V}} = 0, \quad \forall w_h \in \mathbb{U}_h. \tag{3.3}$$

3. *For all $p > 1$, there exists a residual representative $r \in \mathbb{V}$, such that $(r, u_h) \in \mathbb{V} \times \mathbb{U}_h$ solves the semi-infinite mixed formulation:*

$$\begin{cases} \langle J_{p,\mathbb{V}}(r), v \rangle_{\mathbb{V}} + \langle A(u_h), v \rangle_{\mathbb{V}^*,\mathbb{V}} & = \langle F, v \rangle_{\mathbb{V}^*,\mathbb{V}}, \quad \forall v \in \mathbb{V}, \\ \langle \nabla A(u_h) w_h, r \rangle_{\mathbb{V}^*,\mathbb{V}} & = 0, \quad \forall w_h \in \mathbb{U}_h. \end{cases} \tag{3.4}$$

Proof. Let us consider the functional $\mathcal{F} : \mathbb{U}_h \to \mathbb{R}$ such that

$$\mathcal{F}(w_h) := \frac{1}{p^*} \|F - A(w_h)\|_{\mathbb{V}^*}^{p^*}.$$

Thus, we can rewrite the problem (3.1) as: find $u_h \in \mathbb{U}_h$ such that

$$u_h = \arg\min_{w_h \in \mathbb{U}_h} \mathcal{F}(w_h).$$

From (2.4), we arrive at the following expression:

$$\begin{aligned}
\nabla\mathcal{F}(u_h)w_h &= \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u_h + tw_h)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{p^*}\|F - A(u_h + tw_h)\|_{\mathbb{V}^*}^{p^*}\Big|_{t=0}\\
&= -J_{p^*,\mathbb{V}^*}(F - A(u_h + tw_h))\nabla A(u_h + tw_h)w_h\Big|_{t=0}\\
&= -J_{p^*,\mathbb{V}^*}(F - A(u_h))\nabla A(u_h)w_h.
\end{aligned}$$

Writing the above with a duality product and taking into account the identity (2.5), we obtain that

$$\begin{aligned}
\langle\nabla\mathcal{F}(u_h), w_h\rangle_{\mathbb{U}^*,\mathbb{U}} &= -\langle J_{p^*,\mathbb{V}^*}(F - A(u_h)), \nabla A(u_h)w_h\rangle_{\mathbb{V}^{**},\mathbb{V}^*}\\
&= -\left\langle\left(\mathcal{I}_{\mathbb{V}} \circ J_{p,\mathbb{V}}^{-1}\right)(F - A(u_h)), \nabla A(u_h)w_h\right\rangle_{\mathbb{V}^{**},\mathbb{V}^*}\\
&= -\left\langle\nabla A(u_h)w_h, J_{p,\mathbb{V}}^{-1}(F - A(u_h))\right\rangle_{\mathbb{V}^*,\mathbb{V}}.
\end{aligned}$$

So, after setting $r := J_{p,\mathbb{V}}^{-1}(F - A(u_h)) \in \mathbb{V}$, it follows that (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3). ◻

Note that, from the definition of the duality map $J_{p,\mathbb{V}}$ and taking into account that $J_{p,\mathbb{V}}(r) := F - A(u_h)$, it follows that:

$$\|F - A(u_h)\|_{\mathbb{V}^*} = \|r\|_{\mathbb{V}}^{p-1}.$$

The following subsection presents an a posteriori error estimate based on this residual representative.

3.2. A continuous a posteriori error estimator. In this section, we derive reliability and efficiency estimates for the residual minimization method applied to the p-Laplacian problem for $p \in (1, \infty)$. We begin by stating some fundamental properties of the p-Laplacian operator, which are proven in [17, Sect. 5].

LEMMA 3.1 (Properties of the p-Laplacian). *Let $\mathbb{U} = \mathbb{V} = W_0^{1,p}(\Omega)$ and let $A : \mathbb{V} \to \mathbb{V}^*$ be the p-Laplacian operator defined as*

$$\langle A(u), v\rangle_{\mathbb{V}^*,\mathbb{V}} := \int_\Omega |\nabla u|^{p-2}\nabla u \cdot \nabla v \,\mathrm{d}\boldsymbol{x}, \quad \forall u, v \in \mathbb{V}. \tag{3.5}$$

Then, for $p > 1$, there exist positive constants $C_{\mathtt{L}}, c_{\mathtt{L}}, C_{\mathtt{M}}, c_{\mathtt{M}}$ that depend on p, such that for all $u, w \in \mathbb{V}$, the operator A satisfies:

(i) ***Continuity:***

$$\|A(u) - A(w)\|_{\mathbb{V}^*} \le \begin{cases} C_{\mathtt{L}}\left(\|u\|_{\mathbb{V}} + \|w\|_{\mathbb{V}}\right)^{p-2}\|u - w\|_{\mathbb{V}} & \text{if } p \ge 2,\\ c_{\mathtt{L}}\|u - w\|_{\mathbb{V}}^{p-1} & \text{if } 1 < p < 2.\end{cases} \tag{3.6}$$

(ii) ***Monotonicity:***

$$\langle A(u) - A(w), u - w\rangle_{\mathbb{V}^*,\mathbb{V}} \geq \begin{cases} C_{\mathtt{M}} \|u - w\|_{\mathbb{V}}^p & \text{if } p \geq 2, \\ c_{\mathtt{M}} (\|u\|_{\mathbb{V}} + \|w\|_{\mathbb{V}})^{p-2} \|u - w\|_{\mathbb{V}}^2 & \text{if } 1 < p < 2. \end{cases} \tag{3.7}$$

Additionally, the following lemma will be used to obtain several results presented in this paper applied to the p-Laplacian problem.

LEMMA 3.2 (Auxiliary bounds). *Let $\mathbb{U} = \mathbb{V} = W_0^{1,p}(\Omega)$ with $p \in (1, \infty)$ and let u be the solution to the abstract problem (1.1), where $A : \mathbb{U} \to \mathbb{V}^*$ is the p-Laplacian operator defined in (3.5). Let $u_h \in \mathbb{U}_h$ be the solution to the MinRes problem (3.1), and let $\hat{u}_h \in \mathbb{U}_h$ be the best approximation to u. Then, the following bounds hold:*

(i) $\|u\|_{\mathbb{V}} \leq \|F\|_{\mathbb{V}^*}^{\frac{1}{p-1}}$.

(ii) $\|u_h\|_{\mathbb{V}} \leq 2^{\frac{1}{p-1}} \|F\|_{\mathbb{V}^*}^{\frac{1}{p-1}}$.

(iii) $\|\hat{u}_h\|_{\mathbb{V}} \leq C_{\mathtt{BM}}(\mathbb{V}) \|F\|_{\mathbb{V}^*}^{\frac{1}{p-1}}$, *where $C_{\mathtt{BM}}(\mathbb{V})$ is the Banach–Mazur constant of $\mathbb{V}$ introduced in Remark 2.1.*

Proof. Directly from the definition of the operator in (3.5), we have the identity

$$\langle A(v), v\rangle_{\mathbb{V}^*,\mathbb{V}} = \|v\|_{\mathbb{V}}^p \quad \forall v \in \mathbb{V}. \tag{3.8}$$

Item (i) follows from applying (3.8) to the exact solution: $\|u\|_{\mathbb{V}}^p = \langle A(u), u\rangle_{\mathbb{V}^*,\mathbb{V}} \leq \|F\|_{\mathbb{V}^*} \|u\|_{\mathbb{V}}$.

For (ii), since $0 \in \mathbb{U}_h$ and $A(0) = 0$, the minimization property guarantees $\|F - A(u_h)\|_{\mathbb{V}^*} \leq \|F - A(0)\|_{\mathbb{V}^*} = \|F\|_{\mathbb{V}^*}$. By the triangle inequality, $\|A(u_h)\|_{\mathbb{V}^*} \leq 2\|F\|_{\mathbb{V}^*}$. Using (3.8) for u_h, we have $\|u_h\|_{\mathbb{V}}^p = \langle A(u_h), u_h\rangle_{\mathbb{V}^*,\mathbb{V}} \leq \|A(u_h)\|_{\mathbb{V}^*} \|u_h\|_{\mathbb{V}}$. Dividing by $\|u_h\|_{\mathbb{V}}$ gives $\|u_h\|_{\mathbb{V}}^{p-1} \leq \|A(u_h)\|_{\mathbb{V}^*} \leq 2\|F\|_{\mathbb{V}^*}$, which directly implies (ii).

Finally, from Proposition 2.4, we know that $\|\hat{u}_h\|_{\mathbb{V}} \leq C_{\mathtt{BM}}(\mathbb{V}) \|u\|_{\mathbb{V}}$. Combining this with (i) implies (iii). ∎

Based on these properties, we establish the following error estimator.

THEOREM 3.2 (Continuous a posteriori error estimator). *Let u and u_h be the solutions of the abstract problem (1.1) and the MinRes problem (3.1), respectively. Let $r \in \mathbb{V}$ be the residual representative defined in Theorem 3.1. Then, the following global error estimates hold:*

$$\begin{cases} C_1 \|r\|_{\mathbb{V}}^{p-1} \leq \|u - u_h\|_{\mathbb{V}} \leq C_2 \|r\|_{\mathbb{V}} & \text{if } p \geq 2, \\ c_1 \|r\|_{\mathbb{V}} \leq \|u - u_h\|_{\mathbb{V}} \leq c_2 \|r\|_{\mathbb{V}}^{p-1} & \text{if } 1 < p < 2, \end{cases} \tag{3.9}$$

where

$$C_1 = C_{\mathtt{L}}^{-1}(1 + C_{\mathtt{BM}}(\mathbb{V}))^{2-p} \|F\|_{\mathbb{V}^*}^{-\frac{p-2}{p-1}}; \qquad C_2 = C_{\mathtt{M}}^{-\frac{1}{p-1}};$$

$$c_1 = c_{\mathtt{L}}^{-\frac{1}{p-1}}; \qquad c_2 = c_{\mathtt{M}}^{-1}\left(1 + 2^{\frac{1}{p-1}}\right)^{2-p} \|F\|_{\mathbb{V}^*}^{\frac{2-p}{p-1}}.$$

Proof. For $p \geq 2$, we first prove the lower bound. Let $\hat{u}_h$ be the best aproximation of u in $\mathbb{U}_h$. Using the optimality of u_h in (3.1) and the Lipschitz continuity property (3.6), we obtain:

$$\|r\|_{\mathbb{V}}^{p-1} = \|F - A(u_h)\|_{\mathbb{V}^*} \leq \|F - A(\hat{u}_h)\|_{\mathbb{V}^*} \leq C_{\mathtt{L}} (\|u\|_{\mathbb{V}} + \|\hat{u}_h\|_{\mathbb{V}})^{p-2} \|u - \hat{u}_h\|_{\mathbb{V}}.$$

Invoking Lemma 3.2 and noting $\|u-\hat{u}_h\|_{\mathbb{V}} \leq \|u-u_h\|_{\mathbb{V}}$, we get the lower bound. For the upper bound, using the strong monotonicity (3.7) and the fact that $J_{p,\mathbb{V}}(r) = F - A(u_h)$, we get:

$$C_{\mathtt{M}}\|u-u_h\|_{\mathbb{V}}^p \leq \langle A(u)-A(u_h), u-u_h\rangle_{\mathbb{V}^*,\mathbb{V}} \leq \|J_{p,\mathbb{V}}(r)\|_{\mathbb{V}^*}\|u-u_h\|_{\mathbb{V}} = \|r\|_{\mathbb{V}}^{p-1}\|u-u_h\|_{\mathbb{V}}.$$

Simplifying gives the upper bound. For $1<p<2$, the lower bound follows directly from the definition of the residual and the Hölder continuity (3.6):

$$\|r\|_{\mathbb{V}}^{p-1} = \|A(u)-A(u_h)\|_{\mathbb{V}^*} \leq c_{\mathtt{L}}\|u-u_h\|_{\mathbb{V}}^{p-1} \implies c_{\mathtt{L}}^{-\frac{1}{p-1}}\|r\|_{\mathbb{V}} \leq \|u-u_h\|_{\mathbb{V}}.$$

For the upper bound, using the degenerate strong monotonicity (3.7), we get:

$$c_{\mathtt{M}}(\|u\|_{\mathbb{V}}+\|u_h\|_{\mathbb{V}})^{p-2}\|u-u_h\|_{\mathbb{V}}^2 \leq \langle A(u)-A(u_h), u-u_h\rangle_{\mathbb{V}^*,\mathbb{V}} \leq \|r\|_{\mathbb{V}}^{p-1}\|u-u_h\|_{\mathbb{V}}.$$

Dividing by $\|u-u_h\|_{\mathbb{V}}$ and bounding the weight term using Lemma 3.2 (i) and (ii) yields the result. ∎

4. Residual minimization in discrete dual norms.

A complication of the formulation (3.1) is that the dual norm in practice may not be computable, which happens when the space $\mathbb{V}$ is intractable (e.g., when $\mathbb{V} = W_0^{1,p}(\Omega)$). One way to solve this complication is to restrict the supremum to a discrete normed space $(\mathbb{V}_h, \|\ \|_{\mathbb{V}_h})$ which may be non-conforming (i.e. $\mathbb{V}_h \not\subset \mathbb{V}$). For this modification to make sense with a non-conforming space $\mathbb{V}_h$, we will assume the following:

ASSUMPTION 4.1. *Let $\mathbb{V}_h$ be a finite-dimensional normed space and let $\mathbb{V}(h) := \mathbb{V}+\mathbb{V}_h$ be endowed with a norm denoted by $\|\ \|_h$. We assume the following:*

1. $\|v\|_h \cong \|v\|_{\mathbb{V}}$ *and* $\|v_h\|_h \cong \|v_h\|_{\mathbb{V}_h}$ *for all* $v\in\mathbb{V}$ *and* $v_h\in\mathbb{V}_h$.
2. *There exists* $A_h : \mathbb{U}\to\mathbb{V}(h)^*$ *such that* $A_h(w)|_{\mathbb{V}} = A(w)$ *for all* $w\in\mathbb{U}$.

In this way, we can define the following discrete MinRes problem: given an $F_h \in \mathbb{V}(h)^*$ and $p>1$, find $u_h\in\mathbb{U}_h$ such that

$$u_h = \arg\min_{w_h\in\mathbb{U}_h} \frac{1}{p^*}\|I_h^*(F_h - A_h(w_h))\|_{(\mathbb{V}_h)^*}^{p^*}, \tag{4.1}$$

where $I_h : \mathbb{V}_h \to \mathbb{V}(h)$ is the natural injection and

$$\|\bullet\|_{(\mathbb{V}_h)^*} := \sup_{v_h\in\mathbb{V}_h} \frac{\langle \bullet, v_h\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}}{\|v_h\|_h}.$$

Here, the notation $(\mathbb{V}_h)^*$ is used to avoid confusion with a discrete subspace $\mathbb{V}_h^* \subset \mathbb{V}^*$.

For the well-posedness of problem (4.1), we establish the same assumptions as in the exact method (Assumption 3.1), but replacing the space $\mathbb{V}$ by $\mathbb{V}(h)$ and the operator A by A_h. However, in this case, we need an additional hypothesis to ensure the solution is unique. Thus, we have the following assumptions.

ASSUMPTION 4.2. *In problem (4.1), we assume the following statements:*

1. $\mathbb{U}$ *and* $\mathbb{U}_h$ *are subspaces of* $\mathbb{V}(h)$ *and* $\mathbb{V}_h$, *respectively.*

2. A_h *is a strictly monotone operator, i.e.,*

$$\langle A_h(u) - A_h(w), u - w\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} > 0, \quad \forall u, w \in \mathbb{U},\ u \neq w.$$

Assumption 4.2 is necessary since there might exist some $\tilde{u}_h \neq u_h$ in $\mathbb{U}_h$ such that

$$\langle A_h(u_h) - A_h(\tilde{u}_h), v_h\rangle_{(\mathbb{V}_h)^*,\mathbb{V}_h} = 0, \quad \forall v_h \in \mathbb{V}_h, \tag{4.2}$$

but $A_h(u_h) \neq A_h(\tilde{u}_h)$ in $\mathbb{V}(h)^*$. But, if A_h is strictly monotone, setting $v_h := u_h - \tilde{u}_h$ in (4.2), it follows that $u_h = \tilde{u}_h$. Furthermore, note that $\mathbb{U} \subset \mathbb{V}(h)$ and $\mathbb{U}_h \subset \mathbb{V}_h$ are necessary to make sense of the duality products in (4.2).

4.1. Discrete characterizations. Similarly to the exact method, a result that characterizes the solution to problem (4.1) is presented in the following theorem.

THEOREM 4.1 (Discrete problem equivalence). *Let $A_h : \mathbb{U} \to \mathbb{V}(h)^*$ be a Fréchet differentiable operator. Assume that $\mathbb{V}$ and $\mathbb{V}^*$ are strictly convex and reflexive Banach spaces, and that Assumption 4.1 holds true. Let $\mathbb{U}_h$ be a discrete subspace of $\mathbb{U}$ and let $\mathbb{V}_h$ be a finite-dimensional normed space. Given $F_h \in \mathbb{V}(h)^*$, the statements listed below are equivalent:*

i. u_h is a solution to the minimization problem (4.1).

ii. For all $p > 1$, $u_h \in \mathbb{U}_h$ is a solution of the problem:

$$\Big\langle \nabla A_h(u_h)w_h\, ,\, I_h J^{-1}_{p,\mathbb{V}_h} \circ I_h^*(F_h - A_h(u_h))\Big\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} = 0, \quad \forall w_h \in \mathbb{U}_h. \tag{4.3}$$

iii. For all $p > 1$, there exists a residual representative $r_h \in \mathbb{V}_h$, such that $(r_h, u_h) \in \mathbb{V}_h \times \mathbb{U}_h$ solves the finite-dimensional mixed formulation:

$$\big\langle J_{p,\mathbb{V}(h)}(r_h), v_h\big\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} + \langle A_h(u_h), v_h\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} = \langle F_h, v_h\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}\,, \tag{4.4a}$$

$$\langle \nabla A_h(u_h)w_h, r_h\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} = 0, \tag{4.4b}$$

for all $(v_h, w_h) \in \mathbb{V}_h \times \mathbb{U}_h$.

Proof. Analogously to the proof of Theorem 3.1, we consider the functional $\mathcal{F}_h : \mathbb{U}_h \to \mathbb{R}$ defined by

$$\mathcal{F}_h(w_h) := \frac{1}{p^*}\,\big\|I_h^*\big(F_h - A_h(w_h)\big)\big\|^{p^*}_{(\mathbb{V}_h)^*}\,.$$

Using (2.4), we derive its gradient as follows:

$$\langle \nabla \mathcal{F}_h(u_h), w_h\rangle_{\mathbb{U}_h^*,\mathbb{U}_h} = -\Big\langle \nabla A_h(u_h)w_h\, ,\, I_h J^{-1}_{p,\mathbb{V}_h} \circ I_h^*\big(F_h - A_h(u_h)\big)\Big\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}. \tag{4.5}$$

Since u_h minimizes $\mathcal{F}_h$ if and only if the gradient vanishes for all $w_h \in \mathbb{U}_h$, this establishes the equivalence (i) $\Leftrightarrow$ (ii).

Finally, we define the discrete residual representative $r_h := J^{-1}_{p,\mathbb{V}_h} \circ I_h^*\big(F_h - A_h(u_h)\big) \in \mathbb{V}_h$. By applying the duality map $J_{p,\mathbb{V}_h}$ to both sides and invoking Proposition 2.3 (which states that $J_{p,\mathbb{V}_h} = I_h^* J_{p,\mathbb{V}(h)} I_h$), we obtain the following identity

$$I_h^* \circ J_{p,\mathbb{V}(h)}(I_h r_h) = I_h^*\big(F_h - A_h(u_h)\big),$$

in $(\mathbb{V}_h)^*$. This equality is equivalent to the variational statement:

$$\big\langle J_{p,\mathbb{V}(h)}(r_h), v_h\big\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} = \langle F_h - A_h(u_h), v_h\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}\,, \quad \forall v_h \in \mathbb{V}_h,$$

which corresponds exactly to (4.4a). Combining this with the definition of r_h substituted into (4.5), it follows that (ii) $\Leftrightarrow$ (iii). ∎

4.2. Convergence analysis. In this section, we present a convergence analysis of the discrete MinRes method applied to the p-Laplacian problem with $p \in (1, \infty)$, with the configuration described below. Let $\Omega \subset \mathbb{R}^d$ (with $d \in \{2, 3\}$) be a polygonal or polyhedral domain, and let $\mathcal{T}^h$ be a conforming simplicial mesh of Ω. Let $\mathbb{V}_h$ be the lowest-order Crouzeix–Raviart finite element space [11] that vanishes at the barycenters of the boundary faces, and let $\mathbb{U}_h \subset \mathbb{V}_h$ be the H^1-conforming subspace. Let $\mathbb{U} = \mathbb{V} = W_0^{1,p}(\Omega)$ and let $\mathbb{V}(h) := \mathbb{V} + \mathbb{V}_h$ endowed with the broken norm

$$\| \bullet \|_h^p := \sum_{T \in \mathcal{T}^h} | \bullet |_{W^{1,p}(T)}. \tag{4.6}$$

As an extension of the p-Laplacian operator, let us consider $A_h : \mathbb{U} \to \mathbb{V}(h)^*$ defined as

$$\langle A_h(u), v \rangle_{\mathbb{V}(h)^*, \mathbb{V}(h)} := \sum_{T \in \mathcal{T}^h} \int_T |\nabla u|^{p-2} \nabla u \cdot \nabla v \, \mathrm{d}\boldsymbol{x}, \quad \forall (u, v) \in \mathbb{U} \times \mathbb{V}(h). \tag{4.7}$$

As shown below, with this configuration, the extended operator A_h satisfies a property similar to one of those established in Lemma 3.1 for the p-Laplacian operator. The convergence of the proposed method follows from the properties presented below.

LEMMA 4.1 (Local Lipschitz continuity). *Let $\mathbb{U} = \mathbb{V} = W_0^{1,p}(\Omega)$ and let $\mathbb{V}_h$ be the Crouzeix–Raviart finite element space that vanishes at the midpoint of the boundary. Let $\mathbb{V}(h) := \mathbb{V} + \mathbb{V}_h$ endowed with the broken norm (4.6) and let $A_h : \mathbb{U} \to \mathbb{V}(h)^*$ be the extended p-Laplacian operator defined in (4.7). Then, the operator A_h satisfies the continuity properties with the same constants C_{L} and c_{L} defined in Lemma 3.1:*

$$\|A_h(u) - A_h(w)\|_{\mathbb{V}(h)^*} \leq \begin{cases} C_{\mathrm{L}} \Big(\|u\|_{\mathbb{V}} + \|w\|_{\mathbb{V}} \Big)^{p-2} \|u - w\|_{\mathbb{V}} & \text{if } p \geq 2, \\ c_{\mathrm{L}} \, \|u - w\|_{\mathbb{V}}^{p-1} & \text{if } 1 < p < 2. \end{cases} \tag{4.8}$$

Proof. For $p = 2$, the result is trivial. We proceed with the proof for $p > 2$. We rely on the following vector inequality (see [20, Lem. 2.1]): For any $\boldsymbol{\xi}, \boldsymbol{\zeta} \in \mathbb{R}^d$, there exists a constant $C > 0$ such that

$$\big| |\boldsymbol{\xi}|^{p-2} \boldsymbol{\xi} - |\boldsymbol{\zeta}|^{p-2} \boldsymbol{\zeta} \big| \leq C \big(|\boldsymbol{\xi}| + |\boldsymbol{\zeta}| \big)^{p-2} |\boldsymbol{\xi} - \boldsymbol{\zeta}|. \tag{4.9}$$

Let $u, w \in \mathbb{V}$. Using the definition of A_h and the triangle inequality, we have

$$\begin{aligned} \big| \langle A_h(u) - A_h(w), z \rangle \big| &= \left| \sum_{T \in \mathcal{T}^h} \int_T \Big(|\nabla u|^{p-2} \nabla u - |\nabla w|^{p-2} \nabla w \Big) \cdot \nabla z \, \mathrm{d}\boldsymbol{x} \right| \\ &\leq \sum_{T \in \mathcal{T}^h} \int_T \big| |\nabla u|^{p-2} \nabla u - |\nabla w|^{p-2} \nabla w \big| \, |\nabla z| \, \mathrm{d}\boldsymbol{x} \qquad \forall z \in \mathbb{V}(h). \end{aligned}$$

Applying the inequality (4.9) pointwise inside the integrals, we obtain

$$\big| \langle A_h(u) - A_h(w), z \rangle \big| \leq C \sum_{T \in \mathcal{T}^h} \int_T \big(|\nabla u| + |\nabla w| \big)^{p-2} |\nabla (u - w)| |\nabla z| \, \mathrm{d}\boldsymbol{x}.$$

Now, we apply the generalized Hölder inequality with exponents $r = \frac{p}{p-2}$, $s = p$, and $t = p$ (satisfying $1/r + 1/s + 1/t = 1$). Since the sum of integrals over disjoint elements is equivalent to the integral over the whole domain, we can apply Hölder directly to the summation

$$\begin{aligned}\left|\langle A_h(u) - A_h(w), z\rangle\right| &\leq C \left(\sum_{T\in\mathcal{T}^h} \left\|(|\nabla u| + |\nabla w|)^{\frac{p}{r}}\right\|^r_{L^r(T)}\right)^{\frac{1}{r}} \\ &\qquad\times \left(\sum_{T\in\mathcal{T}^h} \|\nabla(u-w)\|^p_{L^p(T)}\right)^{\frac{1}{p}} \left(\sum_{T\in\mathcal{T}^h} \|\nabla z\|^p_{L^p(T)}\right)^{\frac{1}{p}} \\ &= C\left(\sum_{T\in\mathcal{T}^h}\int_T (|\nabla u| + |\nabla w|)^p \,\mathrm{d}\boldsymbol{x}\right)^{\frac{p-2}{p}} \|u-w\|_{\mathbb{V}} \|z\|_h .\end{aligned}$$

Recognizing the first term as the global L^p-norm raised to the power $p-2$, we conclude

$$\begin{aligned}\left|\langle A_h(u) - A_h(w), z\rangle\right| &\leq C\, \||\nabla u| + |\nabla w|\|^{p-2}_{L^p(\Omega)} \|u-w\|_{\mathbb{V}} \|z\|_h \\ &\leq C\big(\|u\|_{\mathbb{V}} + \|w\|_{\mathbb{V}}\big)^{p-2} \|u-w\|_{\mathbb{V}} \|z\|_h .\end{aligned}$$

The result follows by dividing by $\|z\|_h$ and taking the supremum over $z \in \mathbb{V}(h)\setminus\{0\}$.

For $1 < p < 2$, the proof follows an analogous structure. Instead of (4.9), we employ the vector inequality $\left||\boldsymbol{\xi}|^{p-2}\boldsymbol{\xi} - |\boldsymbol{\zeta}|^{p-2}\boldsymbol{\zeta}\right| \leq 2^{2-p}|\boldsymbol{\xi}-\boldsymbol{\zeta}|^{p-1}$ valid for $p<2$ (see, e.g., [19, Chapter 12]). Applying this inequality pointwise within the integrals over each element $T \in \mathcal{T}^h$, the result is directly obtained by using the standard Hölder's inequality with conjugate exponents $p/(p-1)$ and p over the broken domain.

The constants C_{L} and c_{L} arise exclusively from the algebraic vector inequalities applied pointwise within each element $T \in \mathcal{T}^h$. Consequently, they are independent of the global continuity of the functions across the mesh interfaces, and they coincide exactly with the constants of the continuous operator A established in Lemma 3.1. ◻

For the convergence analysis, we define the consistency error functional $\mathcal{E}_h(u) \in \mathbb{V}(h)^*$ as

$$\mathcal{E}_h(u) := F_h - A_h(u), \tag{4.10}$$

for a given $F_h \in \mathbb{V}(h)^*$. We then have the following Céa-type estimate.

THEOREM 4.2 (Céa estimate). *Let u and u_h be the solutions of the abstract problem (1.1) and the discrete MinRes problem (4.1), respectively, with the spaces, norm, and operator described in Lemma 4.1. Let $\mathcal{E}_h(u) \in \mathbb{V}(h)^*$ be the consistency error functional defined in (4.10). Assume that the best approximation error is small enough such that $\inf_{w_h\in\mathbb{U}_h} \|u-w_h\|_{\mathbb{V}} < 1$. Then, the following error estimate holds:*

$$\|u-u_h\|_{\mathbb{V}} \leq C_{\mathsf{app}} \inf_{w_h\in\mathbb{U}_h} \|u-w_h\|^{\alpha}_{\mathbb{V}} + C_{\mathsf{cons}} \|\mathcal{E}_h(u)\|^{\beta}_{(\mathbb{V}_h)^*} ,$$

where the exponents are given by:

$$\alpha := \begin{cases} \frac{1}{p-1} & \text{if } p \geq 2, \\ p-1 & \text{if } 1 < p < 2. \end{cases} \qquad \text{and} \qquad \beta := \begin{cases} \frac{1}{p-1} & \text{if } p \geq 2, \\ 1 & \text{if } 1 < p < 2. \end{cases}$$

The stability constants are defined as:

$$C_{\mathsf{app}} := \begin{cases} 1 + \left(\frac{2C_{\mathrm{L}}(1+C_{\mathtt{BM}}(\mathbb{V}))^{p-2}}{C_{\mathtt{M}}}\right)^{\frac{1}{p-1}} \|F\|^{\frac{p-2}{(p-1)^2}}_{\mathbb{V}^*} & \text{if } p \geq 2, \\ 1 + c_{\mathrm{L}} C_{\mathsf{cons}} & \text{if } 1 < p < 2. \end{cases}$$

$$C_{\mathtt{cons}} := \begin{cases} \left(\frac{2}{C_{\mathtt{M}}}\right)^{\frac{1}{p-1}} & \text{if } p \geq 2, \\ \frac{2}{c_{\mathtt{M}}}\left(2^{\frac{1}{p-1}} + C_{\mathtt{BM}}(\mathbb{V})\right)^{2-p} \|F\|_{\mathbb{V}^*}^{\frac{2-p}{p-1}} & \text{if } 1 < p < 2. \end{cases}$$

Proof. Let us start considering $p \geq 2$. Let $w_h \in \mathbb{U}_h$ be arbitrary. By the triangle inequality, we have

$$\|u - u_h\|_{\mathbb{V}} \leq \|u - w_h\|_{\mathbb{V}} + \|u_h - w_h\|_{\mathbb{V}}. \tag{4.11}$$

We focus on bounding the second term, $\|u_h - w_h\|_{\mathbb{V}}$. Since $\mathbb{U}_h \subset \mathbb{V}_h$, we can identify norms in $\mathbb{V}$ and $\mathbb{V}(h)$ restricted to $\mathbb{U}_h$.

Using the strong monotonicity (3.7) and the fact that u_h minimizes the residual, we derive

$$\begin{aligned} C_{\mathtt{M}} \|u_h - w_h\|_{\mathbb{V}}^{p-1} &\leq \frac{\langle A(u_h) - A(w_h), u_h - w_h\rangle_{\mathbb{V}^*,\mathbb{V}}}{\|u_h - w_h\|_{\mathbb{V}}} \\ &= \frac{\langle A_h(u_h) - A_h(w_h), u_h - w_h\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}}{\|u_h - w_h\|_h} \\ &\leq \|A_h(u_h) - A_h(w_h)\|_{(\mathbb{V}_h)^*} \\ &\leq \|F_h - A_h(u_h)\|_{(\mathbb{V}_h)^*} + \|F_h - A_h(w_h)\|_{(\mathbb{V}_h)^*} \\ &\leq 2\|F_h - A_h(w_h)\|_{(\mathbb{V}_h)^*}. \end{aligned}$$

Since $A_h(u) = F_h + \mathcal{E}_h(u)$ in $(\mathbb{V}_h)^*$, we apply the triangle inequality to bound the discrete residual:

$$\begin{aligned} \|F_h - A_h(w_h)\|_{(\mathbb{V}_h)^*} &\leq \|F_h - A_h(u)\|_{(\mathbb{V}_h)^*} + \|A_h(u) - A_h(w_h)\|_{(\mathbb{V}_h)^*} \\ &= \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*} + \|A_h(u) - A_h(w_h)\|_{(\mathbb{V}_h)^*}. \end{aligned}$$

Bounding the discrete dual norm by the full dual norm and applying Lemma 4.1, we obtain

$$C_{\mathtt{M}} \|u_h - w_h\|_{\mathbb{V}}^{p-1} \leq 2\|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*} + 2C_{\mathtt{L}} \left(\|u\|_{\mathbb{V}} + \|w_h\|_{\mathbb{V}}\right)^{p-2} \|u - w_h\|_{\mathbb{V}}.$$

Thus, since $(a+b)^q \leq a^q + b^q$ for $0 < q < 1$,

$$\|u_h - w_h\|_{\mathbb{V}} \leq \left(\frac{2}{C_{\mathtt{M}}}\right)^{\frac{1}{p-1}} \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*}^{\frac{1}{p-1}} + \left(\frac{2C_{\mathtt{L}}}{C_{\mathtt{M}}}\right)^{\frac{1}{p-1}} \left(\|u\|_{\mathbb{V}} + \|w_h\|_{\mathbb{V}}\right)^{\frac{p-2}{p-1}} \|u - w_h\|_{\mathbb{V}}^{\frac{1}{p-1}}.$$

Substituting this back into (4.11), taking w_h as the best approximation $\hat{u}_h$ of u in $\mathbb{U}_h$ and applying Lemma 3.2, we get

$$\begin{aligned} \|u - u_h\|_{\mathbb{V}} \leq \|u - \hat{u}_h\|_{\mathbb{V}} &+ \left(\frac{2C_{\mathtt{L}}(1 + C_{\mathtt{BM}}(\mathbb{V}))^{p-2}}{C_{\mathtt{M}}}\right)^{\frac{1}{p-1}} \|F\|_{\mathbb{V}^*}^{\frac{p-2}{(p-1)^2}} \|u - \hat{u}_h\|_{\mathbb{V}}^{\frac{1}{p-1}} \\ &+ \left(\frac{2}{C_{\mathtt{M}}}\right)^{\frac{1}{p-1}} \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*}^{\frac{1}{p-1}}. \end{aligned}$$

Consequently, assuming $\|u - \hat{u}_h\|_{\mathbb{V}} < 1$ (so that $\|u - \hat{u}_h\|_{\mathbb{V}} \leq \|u - \hat{u}_h\|_{\mathbb{V}}^{1/(p-1)}$ for $p \geq 2$), the proof is completed for $p \geq 2$.

The proof for $1 < p < 2$ proceeds analogously. It relies on applying Lemmas 3.1 and 4.1, and controlling the resulting weight term $\left(\|u_h\|_{\mathbb{V}} + \|w_h\|_{\mathbb{V}}\right)^{2-p}$ by means of the a priori bounds provided in Lemma 3.2. ∎

REMARK 4.1. *It is important to note that Theorem 4.2 only provides a theoretical lower bound for the convergence rate. In practice, our numerical experiments exhibit convergence rates that exceed those predicted by this theorem. The suboptimal theoretical rate stems from the fact that the proof strictly relies on global vector inequalities to control the monotonicity and continuity of the operator, as in Lemma 4.1. Consequently, the analysis yields pessimistic a priori bounds because it does not exploit the local regularity of the exact solution. Such local regularity is typically essential to derive sharp and optimal error estimates for nonlinear PDEs.*

4.3. Discrete a posteriori error estimator. In this section, we derive an a posteriori error estimator for the discrete MinRes method. We define the discrete residual representative $r_h \in \mathbb{V}_h$ as in Theorem 4.1.

First, we establish the estimator's efficiency (lower bound). This result relies solely on the properties of the operator A_h and the approximation properties of the space $\mathbb{U}_h$, without requiring additional stability assumptions.

PROPOSITION 4.1 (Efficiency). *Let u and u_h be the solutions of problems (1.1) and (4.1), respectively. Let $\mathcal{E}_h(u) \in \mathbb{V}(h)^*$ be the consistency error functional defined in (4.10). Then, the following lower bounds hold:*

$$\begin{cases} \|r_h\|_h^{p-1} \le C \|u - u_h\|_{\mathbb{V}} + \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*} & \text{if } p \ge 2, \\ \|r_h\|_h^{p-1} \le c_{\mathtt{L}} \|u - u_h\|_{\mathbb{V}}^{p-1} + \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*} & \text{if } 1 < p < 2, \end{cases} \tag{4.12}$$

where $C := C_{\mathtt{L}}(1 + C_{\mathtt{BM}}(\mathbb{V}))^{p-2} \|F\|_{\mathbb{V}^*}^{\frac{p-2}{p-1}}$.

Proof. Recall that $\|r_h\|_h^{p-1} = \|F_h - A_h(u_h)\|_{(\mathbb{V}_h)^*}$. Since u_h is the minimizer in $\mathbb{U}_h$, for any $w_h \in \mathbb{U}_h$, we have

$$\begin{aligned} \|r_h\|_h^{p-1} &\le \|F_h - A_h(w_h)\|_{(\mathbb{V}_h)^*} \\ &\le \|A_h(u) - A_h(w_h)\|_{(\mathbb{V}_h)^*} + \|F_h - A_h(u)\|_{(\mathbb{V}_h)^*} \\ &= \|A_h(u) - A_h(w_h)\|_{(\mathbb{V}_h)^*} + \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*}. \end{aligned}$$

For $p \ge 2$: Applying Lemma 4.1 for the specific approximation $w_h := \hat{u}_h$ defined in Lemma 3.2, and invoking the a priori bounds provided therein, gives

$$\|r_h\|_h^{p-1} \le C_{\mathtt{L}}(1 + C_{\mathtt{BM}}(\mathbb{V}))^{p-2} \|F\|_{\mathbb{V}^*}^{\frac{p-2}{p-1}} \|u - \hat{u}_h\|_{\mathbb{V}} + \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*}.$$

The result follows directly since $\|u - \hat{u}_h\|_{\mathbb{V}} \le \|u - u_h\|_{\mathbb{V}}$.

For $1 < p < 2$: We evaluate the error bound directly by choosing $w_h := u_h$. Applying the Hölder continuity from Lemma 4.1, it immediately yields:

$$\begin{aligned} \|r_h\|_h^{p-1} &\le \|A_h(u) - A_h(u_h)\|_{(\mathbb{V}_h)^*} + \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*} \\ &\le c_{\mathtt{L}} \|u - u_h\|_{\mathbb{V}}^{p-1} + \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*}, \end{aligned}$$

which completes the proof. ∎

On the other hand, establishing the estimator's reliability (upper bound) requires the following additional assumption.

ASSUMPTION 4.3 (Fortin operator). *There exists a linear operator $\Pi_h : \mathbb{V}(h) \to \mathbb{V}_h$ and a constant $C_\Pi > 0$ such that:*

$$\begin{cases} \|\Pi_h v\|_h \le C_\Pi \|v\|_h, & \forall v \in \mathbb{V}(h), \\ \langle A_h(w_h), v - \Pi_h v \rangle_{\mathbb{V}(h)^*, \mathbb{V}(h)} = 0, & \forall w_h \in \mathbb{U}_h,\ \forall v \in \mathbb{V}(h). \end{cases}$$

Under this assumption, we define the data oscillation term as:

$$\operatorname{osc}(F_h) := \sup_{v\in\mathbb{V}(h)} \frac{\langle F_h, v - \Pi_h v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}}{\|v\|_h}.$$

We are now in a position to state the reliability estimate.

THEOREM 4.3 (Reliability). *Let u and u_h be the solutions of problems (1.1) and (4.1), respectively. Let $\mathcal{E}_h(u) \in \mathbb{V}(h)^*$ be the consistency error functional, defined such that $F_h - A_h(u) = \mathcal{E}_h(u)$ in $\mathbb{V}(h)^*$. Under Assumption 4.3, the following upper bound holds:*

$$\|u-u_h\|_{\mathbb{V}} \lesssim \begin{cases} \|\mathcal{E}_h(u)\|_{\mathbb{V}(h)^*}^{\frac{1}{p-1}} + \operatorname{osc}(F_h)^{\frac{1}{p-1}} + \|r_h\|_h & \text{if } p \geq 2,\\ \|\mathcal{E}_h(u)\|_{\mathbb{V}(h)^*} + \operatorname{osc}(F_h) + \|r_h\|_h^{p-1} & \text{if } 1<p<2. \end{cases} \tag{4.13}$$

Here, the hidden constants depend on the monotonicity constants, C_Π, p, and, for $1 < p < 2$, on the a priori bounds of u and u_h.

Proof. For any $v \in \mathbb{V}(h)$, we decompose the action of the error using the identity $A_h(u) + \mathcal{E}_h(u) = F_h$ in $\mathbb{V}(h)^*$ and the splitting $v = (v - \Pi_h v) + \Pi_h v$:

$$\begin{aligned}\langle A_h(u) - A_h(u_h), v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} &= \langle A_h(u) - F_h, v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} + \langle F_h - A_h(u_h), v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}\\ &= -\langle \mathcal{E}_h(u), v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} + \langle F_h, v-\Pi_h v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}\\ &\quad - \langle A_h(u_h), v-\Pi_h v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}\\ &\quad + \langle F_h - A_h(u_h), \Pi_h v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}.\end{aligned}$$

Using the orthogonality property of the Fortin operator Π_h (since $u_h \in \mathbb{U}_h$), considering that $J_{p,\mathbb{V}(h)}(r_h) = F_h - A_h(u_h)$ in $(\mathbb{V}_h)^*$, we derive:

$$\begin{aligned}\langle A_h(u) - A_h(u_h), v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} &\leq \|\mathcal{E}_h(u)\|_{\mathbb{V}(h)^*}\|v\|_h + \operatorname{osc}(F_h)\|v\|_h\\ &\quad + \left\langle J_{p,\mathbb{V}(h)}(r_h), \Pi_h v\right\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)}\\ &\leq \left(\|\mathcal{E}_h(u)\|_{\mathbb{V}(h)^*} + \operatorname{osc}(F_h) + C_\Pi \|r_h\|_h^{p-1}\right)\|v\|_h.\end{aligned}$$

Dividing by $\|v\|_h$ and taking the supremum over $v \in \mathbb{V}(h)$, we obtain the bound for the dual norm:

$$\|A_h(u) - A_h(u_h)\|_{\mathbb{V}(h)^*} \leq \|\mathcal{E}_h(u)\|_{\mathbb{V}(h)^*} + \operatorname{osc}(F_h) + C_\Pi \|r_h\|_h^{p-1}. \tag{4.14}$$

For $p \geq 2$: Using the strong monotonicity (3.7), we have $C_{\mathtt{M}}\|u-u_h\|_{\mathbb{V}}^{p-1} \leq \|A_h(u) - A_h(u_h)\|_{\mathbb{V}(h)^*}$. Combining this with (4.14), raising both sides to the power $\frac{1}{p-1}$, and applying the inequality $(a+b+c)^q \leq a^q + b^q + c^q$ (valid for $0 < q \leq 1$), we obtain the desired result.

For $1 < p < 2$: Using the degenerate strong monotonicity (3.7), we derive

$$c_{\mathtt{M}}\left(\|u\|_h + \|u_h\|_h\right)^{p-2}\|u-u_h\|_{\mathbb{V}} \leq \|A_h(u) - A_h(u_h)\|_{\mathbb{V}(h)^*}.$$

Utilizing (4.14) and bounding the weight term $\left(\|u\|_h + \|u_h\|_h\right)^{2-p}$ with the a priori bounds from Lemma 3.2, the upper estimate follows. ∎

The reliability estimate in Theorem 4.3 involves the data oscillation term $\operatorname{osc}(F_h)$, which depends on the smoothness of the data relative to the discrete space. The following proposition establishes that $\operatorname{osc}(F_h)$ is indeed bounded by the best approximation error of the exact solution.

PROPOSITION 4.2 (Bound on data oscillation). *Let u be the solution of (1.1). Let $\mathcal{E}_h(u) \in \mathbb{V}(h)^*$ be the consistency error functional defined in (4.10). Under Assumption 4.3, for all $p > 1$, the data oscillation term satisfies:*

$$\operatorname{osc}(F_h) \le C \inf_{w_h \in \mathbb{U}_h} \|u - w_h\|_{\mathbb{V}}^{\alpha} + (1 + C_\Pi) \|\mathcal{E}_h(u)\|_{(\mathbb{V}_h)^*}, \tag{4.15}$$

where $\alpha = \min\{1, p-1\}$ and

$$C = \begin{cases} C_{\mathrm{L}}(1 + C_\Pi)(1 + C_{\mathtt{BM}}(\mathbb{V}))^{p-2} \|F\|_{\mathbb{V}^*}^{\frac{p-2}{p-1}} & \text{if } p \ge 2, \\ c_{\mathrm{L}}(1 + C_\Pi) & \text{if } 1 < p < 2. \end{cases}$$

Proof. Let $v \in \mathbb{V}(h)$ and let $w_h \in \mathbb{U}_h$ be arbitrary. Using the algebraic identity $F_h = A_h(u) + \mathcal{E}_h(u)$ in $\mathbb{V}(h)^*$, and the orthogonality property of the Fortin operator $\langle A_h(w_h), v - \Pi_h v\rangle = 0$ (Assumption 4.3), we derive

$$\begin{aligned} \langle F_h, v - \Pi_h v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} &= \langle A_h(u) + \mathcal{E}_h(u) - A_h(w_h), v - \Pi_h v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} \\ &\le \left(\|A_h(u) - A_h(w_h)\|_{\mathbb{V}(h)^*} + \|\mathcal{E}_h(u)\|_{\mathbb{V}(h)^*} \right) \|v - \Pi_h v\|_h \\ &\le (1 + C_\Pi) \left(\|A_h(u) - A_h(w_h)\|_{\mathbb{V}(h)^*} + \|\mathcal{E}_h(u)\|_{\mathbb{V}(h)^*} \right) \|v\|_h . \end{aligned}$$

Dividing by $\|v\|_h$ and taking the supremum over $v \in \mathbb{V}(h) \setminus \{0\}$, it follows that

$$\operatorname{osc}(F_h) \le (1 + C_\Pi) \|A_h(u) - A_h(w_h)\|_{\mathbb{V}(h)^*} + (1 + C_\Pi) \|\mathcal{E}_h(u)\|_{\mathbb{V}(h)^*} . \tag{4.16}$$

For $p \ge 2$: We apply the Lipschitz continuity of A_h from Lemma 4.1 to bound the first term. By choosing w_h as the best approximation $\hat{u}_h$ and using the a priori bounds from Lemma 3.2 to control the weight $(\|u\|_{\mathbb{V}} + \|\hat{u}_h\|_{\mathbb{V}})^{p-2}$, we obtain the constant C and the linear dependence on $\|u - \hat{u}_h\|_{\mathbb{V}}$.

For $1 < p < 2$: We apply the Hölder continuity of A_h from Lemma 4.1 directly, which immediately yields the bound in (4.16) with exponent $p - 1$, completing the proof. ◻

5. Numerical implementation. In this section, we present results obtained by applying the mixed formulation in Theorem 4.1 to a nonlinear PDE using a finite element discretization. The numerical realization of the algorithms has been carried out using the open-source library `Netgen/NGSolve` https://www.ngsolve.org.

5.1. Discretization of the model problem. Let $\Omega^h \subset \Omega$ be a polygonal (or polyhedral) approximation to Ω. Let $\mathcal{T}^h$ be a partition of Ω^h into a finite number of disjoint open regular simplices T (triangles for $d = 2$ or tetrahedra for $d = 3$) such that $\bigcup_{T \in \mathcal{T}^h} \overline{T} = \overline{\Omega^h}$, where $h := \max_{T \in \mathcal{T}^h} h_T$ with h_T being the maximum diameter of T. Suppose that each vertex associated with the simplicial mesh $\mathcal{T}^h$ that is in $\partial\Omega^h$ is also in $\partial\Omega$. Assume further that each element has at most one $(d-1)$-dimensional face on $\partial\Omega^h$ and that two elements T and T' such that $\overline{T} \cap \overline{T'} \neq \varnothing$ share either a common vertex, a common edge, or a common whole $(d-1)$-dimensional face. The following two finite element spaces are defined on the partition described above (for the foundational formulation and its 3D extensions, we refer to [11, 7]):

$$\mathbb{U}_h := \left\{ u_h \in W_0^{1,p}(\Omega^h) : u_h|_T \in \mathbb{P}_1(T), \forall T \in \mathcal{T}^h \right\},$$
$$\mathbb{V}_h := \Big\{ v_h \in L^p(\Omega^h) : v_h|_T \in \mathbb{P}_1(T),\ \forall T \in \mathcal{T}^h, v_h \text{ is continuous at the interior face barycenters of } \mathcal{T}^h, \text{ and vanishes at the face barycenters of } \partial\Omega^h \Big\},$$

where $\mathbb{P}_1(T)$ is the space of polynomials of degree up to 1 defined on the element T. The finite element spaces $\mathbb{U}_h$ and $\mathbb{V}_h$ are endowed with the norms

$$| \bullet |^p_{\mathbb{U}_h} := \sum_{i=1}^d \left\| \frac{\partial(\bullet)}{\partial x_i} \right\|^p_{L^p(\Omega^h)}, \qquad | \bullet |^p_{\mathbb{V}_h} := \sum_{T\in\mathcal{T}^h} \sum_{i=1}^d \left\| \frac{\partial(\bullet)}{\partial x_i} \right\|^p_{L^p(T)}, \tag{5.1}$$

respectively. Note that $\mathbb{U}_h$ is a subspace of $W_0^{1,p}(\Omega^h)$, but $\mathbb{V}_h \not\subset W_0^{1,p}(\Omega^h)$. Furthermore, it follows directly that $\mathbb{U}_h \subset \mathbb{V}_h$.

The pair of spaces $(\mathbb{U}_h, \mathbb{V}_h)$ defined above satisfies the stability condition necessary for the well-posedness of the discrete method.

Proposition 5.1 (Verification of Fortin condition). *Let $\mathbb{V}(h) = \mathbb{V} + \mathbb{V}_h$ endowed with the broken norm $\| \bullet \|_h$. There exists a Fortin operator $\Pi_h : \mathbb{V}(h) \to \mathbb{V}_h$ satisfying Assumption 4.3, i.e.,*

$$\begin{cases} \|\Pi_h v\|_h \le C_\Pi \|v\|_h, & \forall v \in \mathbb{V}(h), \\ \langle A_h(w_h), v - \Pi_h v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} = 0, & \forall w_h \in \mathbb{U}_h,\ \forall v \in \mathbb{V}(h). \end{cases}$$

Proof. Let us define $\Pi_h : \mathbb{V}(h) \to \mathbb{V}_h$ as the Crouzeix–Raviart interpolation operator. For any $v \in \mathbb{V}(h)$, its projection $\Pi_h v$ is defined element-wise such that, for all $T \in \mathcal{T}^h$, $\Pi_h v|_T \in \mathbb{P}_1(T)$ is the unique linear polynomial satisfying:

$$\int_F \Pi_h v \,\mathrm{d}s = \int_F v \,\mathrm{d}s, \quad \text{for all } (d-1)\text{-dimensional faces } F \subset \partial T. \tag{5.2}$$

Note that this operator is well-defined on $\mathbb{V}(h)$ since functions in $\mathbb{V}$ have traces on the faces, and functions in $\mathbb{V}_h$ are continuous at the barycenters of the faces, ensuring the integrals are single-valued. Moreover, classical interpolation theory ensures that Π_h is bounded in the broken norm [3, Th. 4.4.20].

To prove the orthogonality condition, we first analyze the gradient of the projected function. Using the divergence theorem on an element T, for any smooth function ϕ, we have $\int_T \nabla\phi \,\mathrm{d}\boldsymbol{x} = \int_{\partial T} \phi \mathbf{n}_T \,\mathrm{d}s$. Since $\Pi_h v$ is linear, $\nabla(\Pi_h v)$ is constant on T, and $\mathbf{n}_T$ is piecewise constant on the faces $F \subset \partial T$. Thus:

$$\begin{aligned} \int_T \nabla(\Pi_h v) \,\mathrm{d}\boldsymbol{x} &= \sum_{F\subset\partial T} \mathbf{n}_F \int_F \Pi_h v \,\mathrm{d}s \\ &= \sum_{F\subset\partial T} \mathbf{n}_F \int_F v \,\mathrm{d}s \quad \text{(by definition (5.2))} \\ &= \int_{\partial T} v \mathbf{n}_T \,\mathrm{d}s = \int_T \nabla v \,\mathrm{d}\boldsymbol{x}. \end{aligned}$$

This implies that $\int_T \nabla(v - \Pi_h v) \,\mathrm{d}\boldsymbol{x} = 0$ for all $T \in \mathcal{T}^h$.

Finally, considering the discrete operator A_h (the extended p-Laplacian), and noting that for any $w_h \in \mathbb{U}_h$, the gradient $\nabla w_h|_T$ is a constant vector (since $w_h \in \mathbb{P}_1(T)$), the term $|\nabla w_h|^{p-2}\nabla w_h$ is also constant on each element. Therefore,

$$\begin{aligned}\langle A_h(w_h), v - \Pi_h v\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} &= \sum_{T\in\mathcal{T}^h}\int_T |\nabla w_h|^{p-2}\nabla w_h \cdot \nabla(v-\Pi_h v)\,\mathrm{d}\boldsymbol{x}\\ &= \sum_{T\in\mathcal{T}^h} |\nabla w_h|^{p-2}\nabla w_h \cdot \int_T \nabla(v-\Pi_h v)\,\mathrm{d}\boldsymbol{x} = 0.\end{aligned}$$

This confirms that Π_h satisfies the Fortin condition. ◻

The model problem we consider is the p-Laplacian equation with homogeneous Dirichlet boundary conditions described in (1.2). To simplify the problem, assume that $\Omega^h = \Omega$ and $f \in L^{p^*}(\Omega)$. The numerical approximation in $\mathbb{U}_h$ is sought through the mixed framework established in Theorem 4.1, featuring the discrete operator A_h from (4.7) and the source functional $F_h \in \mathbb{V}(h)^*$ defined as follows:

$$\langle F_h, v_h\rangle_{\mathbb{V}(h)^*,\mathbb{V}(h)} := \int_\Omega f v_h \,\mathrm{d}\boldsymbol{x}.$$

5.2. The proposed algorithm. To implement the theoretical results from the preceding sections, we have developed a finite element algorithm derived from the mixed formulation in Theorem 4.1. Since this formulation involves the duality map (which is a nonlinear operator) and A_h is also generally nonlinear, our algorithm employs a damped Newton's method to solve the nonlinear system arising from the mixed formulation in Theorem 4.1 as follows: Given the discrete solution pair $(r_h^j, u_h^j) \in \mathbb{V}_h \times \mathbb{U}_h$ in an iterative step j, we seek for the increment $(\delta r_h, \delta u_h)$ in the next iteration step, solving the following linearized problem at the iteration $j+1$:

$$\begin{cases}\nabla J_{s,\mathbb{V}_h}(r_h^j;\delta r_h, v_h) + \nabla A_h(u_h^j;\delta u_h, v_h) & = F_h(v_h) - J_{s,\mathbb{V}_h}(r_h^j; v_h) - A_h(u_h^j; v_h),\\ \nabla A_h(u_h^j; w_h, \delta r_h) & = -\nabla A_h(u_h^j; w_h, r_h^j),\end{cases} \tag{5.3}$$

for all $(v_h, w_h) \in \mathbb{V}_h \times \mathbb{U}_h$. Then, update $r_h^{j+1} := r_h^j + \delta r_h$ and $u_h^{j+1} := u_h^j + \delta u_h$. Note that, for a more compact presentation of the iterative scheme (5.3), we have avoided the use of notation (2.1).

It is well known that the convergence of Newton's method is sensitive to the initial guess, particularly for values of p far from 2 where the nonlinearity becomes severe. For this reason, we employ a parameter continuation strategy on the exponent p. The process begins by solving the problem for $p = 2$ (the linear case), which guarantees a unique solution and requires no initial guess. This solution serves as the initial guess for a slightly perturbed problem. We proceed iteratively, updating the exponent p by a small step δp (denoted as `step` in Algorithm 5.1). Specifically, the update direction depends on the target exponent: we increment p if the target is greater than 2, and decrement it if the target is less than 2. This logic is represented by the notation $2 \pm$ `step` in the algorithm's loop. We continue this process, using the solution from the previous step as the initial guess for the current Newton solver, until the target value of p is reached. This procedure is outlined in Algorithm 5.1.

5.3. Numerical results. In this section, we present numerical experiments to validate the performance of the proposed discrete MinRes method and the associated a posteriori error estimator. We consider the model problem (1.2) on the domain

Algorithm 5.1 MinRes continuation algorithm for the p-Laplacian.

```
1: input p, step:=0.10, MaxNewton, MaxRef, tol.
2: for i ∈ {1, ··· , MaxRef} do
3:    Define test and trial spaces.
4:    Compute the number of DOFs.
5:    Compute (r_h^0, u_h^0) for P = 2 (linear case).
6:    for P = 2 ± step to p do
7:       while ||(δr_h, δu_h)|| < tol do
8:          Solve (5.3) to compute (δr_h, δu_h).
9:          Update (r_h^{j+1}, u_h^{j+1}) ← (r_h^j + δr_h, u_h^j + δu_h).
10:      end while
11:      if MaxNewton is reached then
12:         Refine the step and go to 7.
13:      end if
14:   end for
15:   Compute the error estimate and the actual error.
16:   Refine the mesh uniformly.
17: end for
```

$\Omega = (0,1)^d$ with $d \in \{2,3\}$. Following the benchmark problems proposed in [21], the exact solution is prescribed as the radially symmetric function:

$$u(\boldsymbol{x}) = \frac{p-1}{p-\sigma}\left(\frac{1}{d-\sigma}\right)^{\frac{1}{p-1}}\left(1 - r^{\frac{p-\sigma}{p-1}}\right), \tag{5.4}$$

where $r = |\boldsymbol{x} - \boldsymbol{x}_0|$ and $\sigma < d$. The source term is computed analytically as $f(r) = r^{-\sigma}$. In all experiments, we set $\sigma = 0.97$. The domain is discretized using a sequence of uniform simplicial meshes.

We report the error measured in the norm (5.1) and the computed estimator $\eta := \|r_h\|_{\mathbb{V}_h}^{p-1}$, which corresponds to the discrete dual norm of the residual (see Theorem 4.1). According to our theoretical findings, this quantity is expected to scale with the error. In all our experiments, unless otherwise stated, the initial mesh is generated by a standard diagonal split of the unit square followed by a single uniform refinement step. We consider four test cases to analyze the behavior of the proposed method across different regimes of p and solution regularity.

Case 1: Smooth solutions in 2D and 3D. To evaluate the performance of the method under high regularity conditions, we set $\boldsymbol{x}_0 = (-1,-1)^T$ for the two-dimensional domain and $\boldsymbol{x}_0 = (-1,-1,-1)^T$ for the three-dimensional domain, placing the singularity of the radial function outside the computational domain. With this configuration, the exact solution is highly regular, $u \in C^\infty(\overline{\Omega})$. We test both the singular ($p = 1.5$) and degenerate ($p = 3.0$) regimes under uniform mesh refinement.

Figure 5.1 illustrates the convergence history under uniform mesh refinement for all four configurations. In both dimensions and for both exponents, we observe that the actual error $\|u - u_h\|_{\mathbb{U}_h}$ and the computed estimator η converge at the optimal rate of $\mathcal{O}(h)$ (which corresponds to $\mathcal{O}(\text{NDOFs}^{-1/2})$ in 2D and $\mathcal{O}(\text{NDOFs}^{-1/3})$ in 3D) for linear elements. The estimator closely tracks the true error from the earliest refinement steps without artificial stabilization, thereby validating its robustness for both $p < 2$ and $p > 2$ under strong boundary conditions.

In addition, to evaluate the computational cost and scalability of our numerical approach, we track the total number of Newton iterations accumulated over the entire

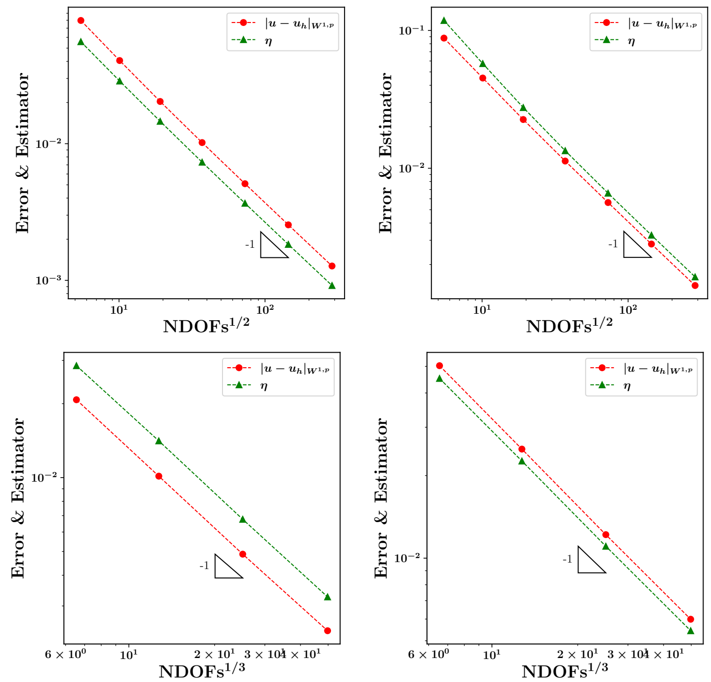

FIG. 5.1. *Convergence rates for smooth solutions. Top row: 2D results for $p = 1.5$ (left) and $p = 3$ (right). Bottom row: 3D results for $p = 1.5$ (left) and $p = 3$ (right).*

2D Smooth Cases			**3D Smooth Cases**		
NDOFs	$p = 1.5$	$p = 3.0$	**NDOFs**	$p = 1.5$	$p = 3.0$
30	26	55	281	44	57
102	35	56	2067	56	58
366	39	49	15791	70	60
1374	44	48	123399	117	60
5310	63	47	-	-	-
20862	79	44	-	-	-
82686	159	40	-	-	-

TABLE 5.1
Total accumulated Newton iterations required to reach the target exponent p during the continuation process for smooth solutions under uniform refinement.

parameter continuation path. Table 5.1 summarizes these counts. The nonlinear solver demonstrates remarkable robustness; notably, in the degenerate regime ($p = 3$), the number of iterations remains practically constant despite significant increases in the degrees of freedom across both dimensions.

Case 2: Singular right-hand side with $p = 1.5$. We now consider a problem with lower regularity by setting the source term singularity at the corner of the domain ($\boldsymbol{x}_0 = (0, 0)^T$). We set the exponent to $p = 1.5$. To fully understand the behavior of our method under these conditions, and specifically how the resolution of the right-hand side affects the error estimator, we analyze three mesh-refinement strategies.

First, we apply the standard uniform refinement starting from the previously

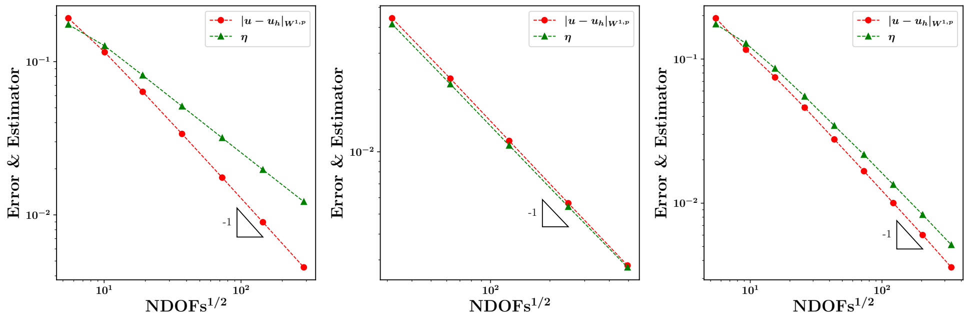

FIG. 5.2. *Convergence rates for a singular right-hand side with $p = 1.5$ in 2D. Left: Uniform refinement from a standard coarse mesh. Center: Uniform refinement from a pre-adapted initial mesh. Right: Adaptive mesh refinement starting from a standard coarse mesh.*

described coarse initial mesh. The results are shown in Figure 5.2 (left). In this scenario, the estimator's convergence rate is slower than the true error rate. We attribute this discrepancy to the right-hand side singularity, which is not adequately represented by the standard initial mesh and remains poorly resolved under purely uniform refinement.

To verify our hypothesis regarding the impact of the right-hand-side representation, we introduce an initial mesh refinement inspired by the pre-adaptation strategies in [22]. While that reference proposes starting with a regularized right-hand side, we leverage only the mesh-adaptation aspect of their procedure to isolate its geometric benefits. This provides us with a pre-adapted initial mesh that more accurately captures the singularity. Starting our main MinRes procedure from this adapted mesh and proceeding with uniform refinement, the estimator's convergence rate successfully realigns with the true error rate, as depicted in Figure 5.2 (center).

To establish a standalone procedure independent of the auxiliary initial MinRes step, we employ the main MinRes method directly with an adaptive mesh refinement strategy driven by our a posteriori error estimator η, starting from the original coarse mesh. For the adaptive process, we employ Dörfler marking [15] with a parameter $\theta = 0.5$. As illustrated in Figure 5.2 (right), the adaptive refinement dynamically resolves the singularity of the right-hand side as the iterative process advances. Consequently, the estimator accurately tracks the actual error rate without requiring a manually pre-adapted mesh.

To visually demonstrate the effectiveness of our a posteriori error estimator in capturing local features, Figure 5.3 shows the evolution of the mesh during the adaptive refinement process. The Dörfler marking strategy strongly localizes the refinement around this singular point, maintaining a coarser resolution in the rest of the domain where the solution is smoother. This targeted adaptivity is precisely what allows the method to recover the previously observed optimal convergence rates.

Finally, Table 5.2 presents the performance of the nonlinear solver for this singular problem using both the pre-adapted and the fully adaptive mesh strategies. Consistent with the observations in the smooth scenarios, the solver retains its remarkable robustness. The accumulated iteration counts grow very slowly and remain stable, even as the adaptive procedure significantly increases the local resolution and the total number of degrees of freedom.

6. Concluding remarks. In this work, we develop and analyze a residual minimization method for solving nonlinear PDEs in Banach spaces, with a particular

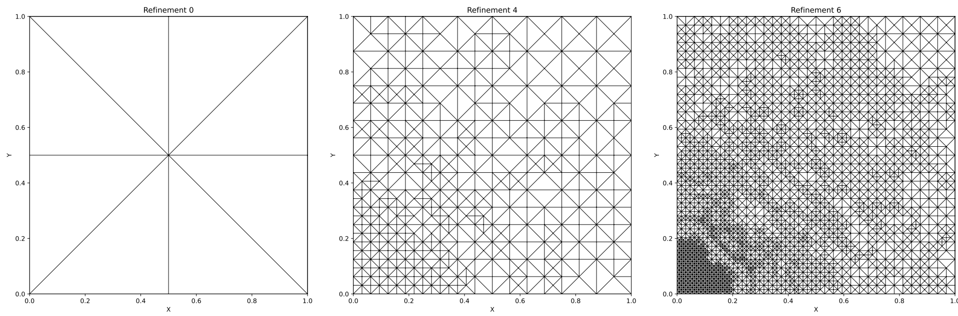

FIG. 5.3. *Snapshots of the adaptively refined mesh in 2D for the singular problem ($p = 1.5$). From left to right: the initial coarse mesh, an early adapted mesh, and the mesh at the sixth refinement step.*

Pre-adapted initial mesh		Adaptive mesh refinement	
NDOFs	Total Newton iterations	NDOFs	Total Newton iterations
1006	52	30	33
3919	80	86	40
15460	101	238	50
61402	176	671	85
244726	160	1891	88
-	-	5330	83
-	-	14861	92
-	-	41457	134
-	-	112921	176

TABLE 5.2
Case 2. Total accumulated Newton iterations required to reach the target exponent $p = 1.5$ for the singular problem in 2D.

focus on the p-Laplacian problem. By formulating the problem as a minimization of the residual in a dual norm, we derived a mixed formulation that involves the duality map of the test space. This approach naturally extends the framework proposed in [24] to nonlinear operators.

From a theoretical standpoint, we established the well-posedness of both the continuous and discrete minimization problems, relying on the strict convexity and reflexivity of the underlying Banach spaces. A key contribution of this study is the derivation of an a posteriori error estimator. We proved that the norm of the residual representative provides both upper and lower bounds for the approximation error, thereby guaranteeing reliability and efficiency. In the discrete setting, using a nonconforming Discontinuous Petrov–Galerkin (DPG) framework with broken Sobolev norms, we showed that the method remains stable and convergent.

The numerical experiments validated our theoretical findings. For smooth solutions, the method exhibits optimal convergence rates for both $p < 2$ and $p \geq 2$. In the presence of singularities, the proposed a posteriori error estimator drives an adaptive mesh refinement strategy. This adaptivity enables the method to recover optimal algebraic convergence rates, significantly outperforming uniform refinement, particularly for degenerate diffusion coefficients ($p = 3$).

Apart from the evident benefits of straightforwardly designing continuous a posteriori error estimators, we believe that the mechanisms developed in this work can be used to construct operator preconditioners for linear and nonlinear problems in L^p spaces (see, e.g., [12]). Moreover, these ideas may be useful for MinRes-based neural-network approximations of nonlinear PDEs, in the spirit of [16].

REFERENCES

[1] A. K. Balci, L. Diening, and J. Storn, *Relaxed Kačanov scheme for the p-Laplacian with large exponent*, SIAM Journal on Numerical Analysis, 61 (2023), pp. 2775–2794.

[2] P. B. Bochev and M. Gunzburger, *Least-squares finite element methods*, vol. 166, Springer Science & Business Media, 2009.

[3] S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, Springer, 2008.

[4] P. Cantin and N. Heuer, *A DPG framework for strongly monotone operators*, SIAM Journal on Numerical Analysis, 56 (2018), pp. 2731–2750.

[5] C. Carstensen, P. Bringmann, F. Hellwig, and P. Wriggers, *Nonlinear Discontinuous Petrov–Galerkin methods*, Numerische Mathematik, 139 (2018), pp. 529–561.

[6] P. G. Ciarlet, *Linear and Nonlinear Functional Analysis with Applications*, SIAM-Society for Industrial and Applied Mathematics, 2013.

[7] P. Ciarlet Jr, C. F. Dunkl, and S. A. Sauter, *A family of crouzeix–raviart finite elements in 3d*, Analysis and Applications, 16 (2018), pp. 649–691.

[8] R. J. Cier, T. Poulet, S. Rojas, M. Veveakis, and V. M. Calo, *Automatically adaptive stabilized finite elements and continuation analysis for compaction banding in geomaterials*, International Journal for Numerical Methods in Engineering, 122 (2021), pp. 6234–6252.

[9] 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. Special issue honoring G.I. Taylor Medalist Prof. Arif Masud.

[10] I. Cioranescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, Mathematics and Its Applications : Main Series №62, Springer, 1 ed., 1990.

[11] M. Crouzeix and P.-A. Raviart, *Conforming and nonconforming finite element methods for solving the stationary Stokes equations I*, Revue Française d'Automatique Informatique Recherche Opérationnelle. Mathématique, 7 (1973), pp. 33–75.

[12] R. Das, H. Hutridurga, A. K. Pani, and R. Ruiz-Baier, *Robust stability and preconditioning of Darcy–Forchheimer equations*, SIAM Journal on Numerical Analysis, (2026), p. under review.

[13] K. Deimling, *Nonlinear functional analysis*, Springer Science & Business Media, 2013.

[14] L. Demkowicz and J. Gopalakrishnan, *The discontinuous Petrov–Galerkin method*, Acta Numerica, 34 (2025), pp. 293–384.

[15] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM Journal on Numerical Analysis, 33 (1996), pp. 1106–1124.

[16] T. Führer and S. Rojas, *A posteriori analysis of neural network approximations*, arXiv preprint arXiv:2507.06017, (2025).

[17] R. Glowinski and A. Marroco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires*, Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique, 9 (1975), pp. 41–76.

[18] J. Li and L. Demkowicz, *An L^p-DPG method with application to 2D convection-diffusion problems*, Computational Methods in Applied Mathematics, 22 (2022), pp. 649–662.

[19] P. Lindqvist, *Notes on the p–Laplace equation*, no. 161, University of Jyväskylä, 2017.

[20] W. Liu and J. W. Barrett, *Quasi-norm error bounds for the finite element approximation of some degenerate quasilinear elliptic equations and variational inequalities*, ESAIM: Mathematical Modelling and Numerical Analysis, 28 (1994), pp. 725–744.

[21] W. Liu and N. Yan, *Some a posteriori error estimators for p–Laplacian based on residual estimation or gradient recovery*, Journal of Scientific Computing, 16 (2001), pp. 435–477.

[22] F. Millar, I. Muga, S. Rojas, and K. G. Van der Zee, *Projection in negative norms and the regularization of rough linear functionals*, Numerische Mathematik, 150 (2022), pp. 1087–1121.

[23] H. Monsuur, R. Stevenson, and J. Storn, *Minimal residual methods in negative or fractional Sobolev norms*, Mathematics of Computation, 93 (2024), pp. 1027–1052.

[24] I. Muga and K. G. Van Der Zee, *Discretization of linear problems in Banach spaces: Residual minimization, nonlinear Petrov–Galerkin, and monotone mixed methods*, SIAM

Journal on Numerical Analysis, 58 (2020), pp. 3406–3426.

[25] J. Muñoz-Matute and L. Demkowicz, *Multistage DPG time-marching scheme for nonlinear problems*, arXiv preprint arXiv:2309.00069, (2023).

[26] S. Rojas, P. Maczuga, J. Muñoz-Matute, D. Pardo, and M. Paszyński, *Robust Variational Physics-Informed Neural Networks*, Computer Methods in Applied Mechanics and Engineering, 425 (2024), p. 116904.

[27] A. Stern, *Banach space projections and Petrov-Galerkin estimates*, Numerische Mathematik, 130 (2015), pp. 125–133.

[28] J. Storn, *Solving minimal residual methods in* $W^{-1,p'}$ *with large exponents* p, Journal of Scientific Computing, 99 (2024), p. 35.