

NUMERICAL METHODS FOR POWER-LAW DIFFUSION PROBLEMS *

IOANNIS TOULOPOULOS*[†] AND THOMAS WICK*[‡]

Abstract. In this paper, we consider numerical methods for nonlinear diffusion problems where the diffusion term follows a power law, e.g., p -Laplace type problems. In the first part, we present continuous higher order finite element discretizations for the model problem and we derive error estimates. In the second part, we discuss Newton iterative methods based on residual-based line search and error-oriented globalization, which are employed for the numerical solution of the produced nonlinear algebraic system. Thirdly, we formulate the original problem as a saddle point problem in the frame of augmented Lagrangian techniques and present two iterative methods for its solution. We conduct a systematic investigation of all solution algorithms. These algorithms are compared with respect to computational cost and their efficiency. Numerical results demonstrating the theoretical error estimates are also presented in five examples.

Key words. power-law diffusion problems, p -Laplace type problems, high order finite element discretizations, Newton iterative methods, augmented Lagrangian techniques

AMS subject classifications. 65N12, 65N15, 65L20, 65L60, 49M15

1. Introduction. Second order elliptic problems with p -type gradient nonlinearities in the diffusion coefficient, that is $(\varepsilon^2 + |\nabla u|^2)^{\frac{p-2}{2}}$, with $\varepsilon \geq 0$ and $p \in (1, \infty)$, are of special interest and are relevant in many practical applications, e.g., in aerodynamics, porous media, non-Newtonian flows, plasticity and glaciology, and fluid-structure interaction [1, 2, 3, 4]. Therefore, the numerical solution of these problems is of great importance. In the application of numerical methods for solving these type of problems, we meet several difficulties associated mainly with the presence of the gradient in the p -nonlinear structure of the diffusion term.

The most classical p -type model problem that has been studied in the literature is the p -Laplace problem, i.e., $\varepsilon = 0$. The first analysis of finite element methods (FE) for p -Laplace was undertaken in [5] and in [6] (chapter 5), where (sup-optimal) error estimates have been shown in the $W^{1,p}$ -norm. These results were further improved in [7] and recently, were used in [8, 9] for developing k and hk finite element methods, where h denotes the spatial discretization parameter and k the degree of the polynomial space¹. Recently, optimal error bounds have been shown for p -Laplace and more general p -power law elliptic problems by using quasi-norm interpolation estimates. The main idea in this approach is first to establish interpolation estimates in a quasi-norm, which is predetermined by the p -nature of the problem, and then to derive error estimates by exploiting the relations between the quasi- and Sobolev norms. Among others we quote [10, 11, 12, 13]. In [14], interpolation operators in Orlicz-Sobolev spaces were studied and were utilized for approximating solutions of more general p -type problems, the so-called p -structure problems.

Over the last two decades, there has been an increasing interest on devising discontinuous Galerkin (DG) methods for the numerical solution of p -type problems, see [15, 16, 17, 18]. In all these DG methods the numerical fluxes were developed by following the p -nonlinear nature of the problem. Particularly, in [16], local DG methods were studied for p -structure problems. The quasi-norm interpolation estimates presented in [14], were applied in the frame of broken paces and optimal error estimates were shown for linear elements. The same Local DG methods have been used later in [17] for solving more realistic p -type problems and in [19] for solving non-Newtonian flow problems.

We point out that the degenerate nature of the p -type problems makes the study of their regularity properties quite difficult. In general we cannot guarantee high regularity for the solution u even with smooth problem data, [20], and hence, the FE methods that have been applied for these problems usually consider first order polynomial spaces. In [21], global $W^{2,2}$ regularity properties were established under some assumptions on the right hand side of the p -type model under consideration. These results were used in the discretization error analysis in [21, 22]. Furthermore, it has been shown in [22] that under some further assumptions on the problem data, we can obtain $W^{1+2/p,p}$ regularity for the solutions of the p -Laplacian, and this regularity is sufficient to ensure optimal error bounds.

However, recently there has been much work on applying high-order tensor product spaces for solving elliptic problems, even in cases with reduced regularity, and many techniques have been developed, e.g., adaptive meshes, graded meshes, for recovering the optimal rates, see for example [8, 23, 24, 25, 26, 27]. In this paper, motivated by the results in [14], we discuss high order interpolation error estimates in the same quasi-norms for solving p -type diffusion problems, with $p \in (1, \infty)$ and $0 < \varepsilon \leq 1$.²

*This work was supported by Austrian Science Fund (FWF) under the grant NFN S117-03.

*Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences (ÖAW)

[†]ioannis.toulopoulos@oeaw.ac.at

[‡]thomas.wick@ricam.oeaw.ac.at

¹In the literature, p usually denotes the degree of the polynomial space. Due to the appearance of the parameter p in the p -Laplace operator, we denote the polynomial degree by k .

²We emphasize that in the numerical examples, we always use $\varepsilon^2 > 0$, which helps on relaxing the singular behavior of the

In particular, assuming appropriate regularity properties for u , see Assumptions 2.1 and 2.3, we combine the structure (monotonicity) conditions presented in [14], with known high order interpolation estimates in Sobolev spaces, [28]. This helps us to derive quasi-norm interpolation error estimates using high order spaces. Having these estimates, we proceed and show optimal error bounds for sufficiently smooth solutions. We mention that in [14] quasi-norm error estimates have been shown for linear spaces but using a different methodology.

After the finite element discretization of the p -type diffusion problem we must solve a nonlinear algebraic system. In the second part of the paper, we present two Newton methods for solving the nonlinear algebraic system, (i) a classical residual-based procedure and (ii) an error-oriented approach, [29]. We also mention that in [30], inexact Newton methods with adaptive stopping criteria for the nonlinear and the linear solver have been presented. Therein, the stopping criteria is based on a posteriori error estimates, which split the global error into three parts, the discretization, the linearization and the algebraic part.

Initially inspired by the results in [31], we show bounds for the condition number of the iterative Newton Jacobian matrix. Then we present two methods. The first method is simple to be materialized and uses a classical criterion in order to check at each step whether the residual has been decreased. The second method is based on a natural monotonicity criterion and rather than observing the residuals, two update steps are compared. This second method is more expensive since an additional simplified problem needs to be solved. Later, in the numerical examples, we investigate and compare the performance and the corresponding convergence rates of the two Newton methods with respect to the variation of the problem parameters ε and p . The comparisons are made in a wide range of problems with high and low regularity solutions.

Lastly, we treat the original problem by developing techniques which provide a separation of the gradient of the solution from the nonlinearity. We achieve this by means of augmented Lagrangian (ALG) techniques. The original problem is reformulated as a saddle point problem where the produced variational problem is discretized by finite elements. We propose two iterative methods for solving the resulting nonlinear algebraic system. The first iterative method is the classical ALG1 iterative method, which is usually used in the literature, see a detailed analysis in [32]. It can be interpreted as a variant of the Uzawa algorithm, where the Lagrange multiplier is separately updated after a sequence of nested iteration procedure and is denoted by sALG1. The second proposed iterative method can be characterized as a monolithic approach, mALG1, where all the unknown variables (state variable and Lagrange multiplier) are simultaneously computed in one step. All proposed methods are compared with respect to computational cost and to the convergence rates in several examples. We point out that both ALG1 methods yield the same numerical solutions (in terms of accuracy) as the Newton methods. The mALG1 method is of optimal complexity and gives rise to a more easy handling of the p -nonlinear nature of the model, but is not as fast as the Newton methods. To the best of our knowledge, a comparison of these iterative techniques including also higher order finite elements discretizations has not yet been presented in the existing literature.

The outline of the paper is as follows. In Section 2, the model problem is introduced. Furthermore, the FE discretization analysis to this problem is presented. In Section 3, residual-based and error-oriented Newton methods are formulated. Also, in the same section the solution of the problem under the augmented Lagrangian methodology is presented and the two ALG1 iterative methods are in details described. Finally, in Section 4 systematic comparisons of all solution algorithms in terms of convergence rates, Newton iterations, and CPU times are measured. We also present numerical results to confirm the theoretical error bounds. The paper closes with the conclusions in Section 5.

2. The model problem. Let Ω be a bounded polygonal domain in \mathbb{R}^d , with $d = 2$ and $\Gamma_D = \partial\Omega$. We consider the following scalar p -type problem

$$(2.1) \quad -\operatorname{div} \mathbf{A}(\nabla u) = f \quad \text{in } \Omega, \quad u = u_D \quad \text{on } \Gamma_D,$$

where $f : \Omega \rightarrow \mathbb{R}$ and $u_D : \Gamma_D \rightarrow \mathbb{R}$ are given smooth functions. The operator $\mathbf{A}(\nabla u) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ has the following p -power law form

$$(2.2) \quad \mathbf{A}(\nabla u) = (\varepsilon^2 + |\nabla u|^2)^{\frac{p-2}{2}} \nabla u,$$

where $p \in (1, \infty)$ and $\varepsilon > 0$ are model parameters and $|\cdot|^2 = (\cdot, \cdot)$. The function $a(\nabla u) = (\varepsilon^2 + |\nabla u|^2)^{\frac{p-2}{2}}$ is the diffusivity term of (2.1). As we can observe by (2.2), the nonlinear nature of the problem is due to the appearance of $|\nabla u|$ in the diffusivity function and this poses numerical challenges. We introduce the closely related function $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ to the operator \mathbf{A} by

$$(2.3) \quad \mathbf{F}(\mathbf{a}) = (\varepsilon^2 + |\mathbf{a}|^2)^{\frac{p-2}{4}} \mathbf{a}.$$

problem at points where $|\nabla u| = 0$. However, we perform numerical tests setting $\varepsilon^2 = 10^{-8}$, which can give us an idea about the behavior of the algorithm for the classical p -Laplace problem.

2.1. Preliminaries and notation. Let $1 \leq p \leq \infty$ be fixed and l be a non-negative integer. As usual, $L^p(\Omega)$ denotes Lebesgue spaces for which $\int_{\Omega} |u(x)|^p dx < \infty$, endowed with the norm $\|u\|_{L^p(\Omega)} = (\int_{\Omega} |u(x)|^p dx)^{\frac{1}{p}}$, and $W^{l,p}(\Omega)$ is the Sobolev space, which consists of the functions $\phi : \Omega \rightarrow \mathbb{R}$ such that their weak derivatives $D^\alpha \phi$ with $|\alpha| \leq l$ belong to $L^p(\Omega)$. If $\phi \in W^{l,p}(\Omega)$, then its norm is defined by

$$\|\phi\|_{W^{l,p}(\Omega)} = \left(\sum_{0 \leq |\alpha| \leq l} \|D^\alpha \phi\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}} \quad \text{and} \quad \|\phi\|_{W^{l,\infty}(\Omega)} = \max_{0 \leq |\alpha| \leq l} \|D^\alpha \phi\|_{L^\infty(\Omega)},$$

for $1 \leq p < \infty$ and $p = \infty$, respectively. We refer the reader to [33] for more details about Sobolev spaces. Furthermore, we define the spaces

$$(2.4) \quad W_D^{l,p} := \{u \in W^{l,p}(\Omega) : u|_{\partial\Omega} = u_D\}, \quad \text{and} \quad W_0^{l,p} := \{u \in W^{l,p}(\Omega) : u|_{\partial\Omega} = 0\}.$$

96 In what follows, positive constants c and C appearing in the inequalities are generic constants which do
97 not depend on the mesh-size h . We indicate on what may the constants depend for a better understanding
98 of the proofs. Frequently, we will write $a \sim b$ meaning that $ca \leq b \leq Ca$, with c and C independent of the
99 mesh size.

2.2. The weak problem . The weak formulation for (2.1) reads as follows: Find $u \in W_D^{1,p}$ such that

$$(2.5a) \quad B(u, \phi) = l_f(\phi), \quad \forall \phi \in W_0^{1,p}(\Omega), \quad \text{where} \quad B(u, \phi) = \int_{\Omega} \mathbf{A}(\nabla u) \cdot \nabla \phi dx, \quad \text{and} \quad l_f(\phi) = \int_{\Omega} f \phi dx.$$

100 Depending on the form of \mathbf{A} and on the range of p , the well-posedness has been examined by means of
101 monotone operators in several works, see e.g., [6, 7].

Problem (2.5) is equivalent to the minimization problem:

$$(2.6) \quad \text{Find } u \in W_D^{1,p} \text{ such that } J(u) \leq J(\phi), \quad \forall \phi \in W_D^{1,p},$$

where $J : W_D^{1,p} \rightarrow \mathbb{R}$ is defined by

$$(2.7) \quad J(\phi) = \frac{1}{p} \int_{\Omega} (\varepsilon^2 + |\nabla \phi|^2)^{\frac{p}{2}} dx - \int_{\Omega} f \phi dx.$$

Furthermore, one can show that the Gateaux derivative of B is given by

$$(2.8) \quad B'(u)(v, w) = \int_{\Omega} (\varepsilon^2 + |\nabla u|^2)^{\frac{p-2}{2}} \nabla v \cdot \nabla w dx \\ + (p-2) \int_{\Omega} (\varepsilon^2 + |\nabla u|^2)^{\frac{p-4}{2}} (\nabla u \cdot \nabla v)(\nabla u \cdot \nabla w) dx, \quad \text{for } u, v, w \in W_D^{1,p}.$$

102 **ASSUMPTION 2.1.** Let $l \geq 2$ be an integer and let $p \in (1, \infty)$ and $d = 2$. We assume that the solution u
103 of (2.5) belongs to $V := W_D^{l,p}(\Omega)$, where either $(l-1)p > d$ and $p > 1$ or $(l-1)p < d$ and $p > \frac{2d}{d+1}$. Further,
104 we assume that $\nabla \mathbf{F}(\nabla u) \in L^2(\Omega)$. For simplicity, we assume the following for the boundary data:

105 **ASSUMPTION 2.2.** Let $u_D \in \mathbb{P}^k(\Gamma_D)$, where \mathbb{P}^k is the space of polynomials degree less than or equal to k .
106 Further, we assume that $|\int_{\Gamma_D} u_D ds| = 0$.

2.3. Known inequalities. The following inequalities are going to be used in several places in the text. Hölder's and Young's inequalities read: For any ϵ , $0 < \epsilon < \infty$, and $1 < p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$, for $u \in L^p(\Omega)$ and $v \in L^q(\Omega)$, there holds

$$(2.9a) \quad \left| \int_{\Omega} uv dx \right| \leq \|u\|_{L^p(\Omega)} \|v\|_{L^q(\Omega)}, \quad \left| \int_{\Omega} uv dx \right| \leq \frac{\epsilon}{p} \|u\|_{L^p(\Omega)}^p + \frac{\epsilon^{-\frac{q}{p}}}{q} \|v\|_{L^q(\Omega)}^q.$$

Poincaré-Friedrichs inequality, see [28]: for any $u \in W^{1,2}(\Omega)$, it holds

$$(2.9b) \quad \|u\|_{L^2(\Omega)} \leq C(\Omega, \partial\Omega) \left(\|\nabla u\|_{L^2(\Omega)} + \left| \int_{\partial\Omega} u ds \right| \right).$$

Next, we introduce some functions which will be useful to the following. For $p > 1$ and $\varepsilon > 0$, we define the functions

$$(2.10) \quad \widehat{\varphi}(t) := (\varepsilon + t)^{p-2}$$

$$(2.11) \quad \varphi(t) := \int_0^t (\varepsilon^2 + s^2)^{\frac{p-2}{2}} s ds, \quad \text{and it follows } \varphi'(t) = (\varepsilon^2 + t^2)^{\frac{p-2}{2}} t.$$

107 By an easy computation we can show that $\varphi'(t) \sim t\varphi''(t)$ and $\varphi''(t) \sim \widehat{\varphi}(t)$.

LEMMA 2.1. *Let \mathbf{A} be given by (2.2) and let \mathbf{F} be defined by (2.3). Then the relations*

$$(2.12a) \quad (\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})) \cdot (\mathbf{P} - \mathbf{Q}) \sim |\mathbf{F}(\mathbf{P}) - \mathbf{F}(\mathbf{Q})|^2,$$

$$(2.12b) \quad \sim \widehat{\varphi}(|\mathbf{P}| + |\mathbf{Q}|)|\mathbf{P} - \mathbf{Q}|^2,$$

$$(2.12c) \quad |\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})| \sim \widehat{\varphi}(|\mathbf{P}| + |\mathbf{Q}|)|\mathbf{P} - \mathbf{Q}|,$$

108 holds for all $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^2$.

109 *Proof.* The proofs are given in [34]. \square

110 **2.4. Finite element notation.** Let $T_h = \{E_i\}_{i=1}^{N_E}$ be a quasi-uniform subdivision of Ω into quadrilateral
111 elements (without hanging nodes) with the following properties: (i) $\overline{\Omega} = \cup_{i=1}^{N_E} \overline{E}_i$, (ii) if $E_i, E_j \in T_h$ then
112 their closures are either disjoint, or have common vertex, or have common edge. As usual, we set h_{E_i} to be
113 the diameter of $E_i \in T_h$ and the mesh size h is considered to be the maximum diameter of $E_i \in T_h$, i.e.,
114 $h := \max_{E_i \in T_h} h_{E_i}$.

Let $E \in T_h$ and $k \in \mathbb{N}$, we denote $\mathbb{Q}_k(E)$ the space of tensor product polynomials on E of degree less than or equal to k in each variable. On T_h , we define the approximation spaces $V_{D,h}^{(k)} \subset V$, $V_{0,h}^{(k)}$ and $V_h^{(k)}$ as

$$(2.13a) \quad V_{D,h}^{(k)} := \{\phi_h \in C(\overline{\Omega}) : \phi_h|_E \in \mathbb{Q}_k(E), \forall E \in T_h, \text{ and } \phi_h|_{\partial\Omega} = u_D\},$$

$$(2.13b) \quad V_{0,h}^{(k)} := \{\phi_h \in C(\overline{\Omega}) : \phi_h|_E \in \mathbb{Q}_k(E), \forall E \in T_h, \text{ and } \phi_h|_{\partial\Omega} = 0\},$$

$$(2.13c) \quad V_h^{(k)} := \{\phi_h \in C(\overline{\Omega}) : \phi_h|_E \in \mathbb{Q}_k(E), \forall E \in T_h\}.$$

LEMMA 2.2. *Let \mathcal{I}_h^k be the corresponding global interpolation operator and let $u \in W^{l,p}(\Omega)$ with $k+1 \geq l \geq 2$ and $p > 1$. Then, there exists a constant $C_{intp} > 0$ independent of h such that, the interpolation estimate*

$$(2.14) \quad |u - \mathcal{I}_h^k u|_{W^{s,p}(\Omega)} \leq C_{intp} h^{l-s} |u|_{W^{l,p}(\Omega)},$$

115 holds for $0 \leq s \leq l$.

116 *Proof.* The proof can be found in [28]. \square

117 REMARK 2.1. *We present the error analysis for the spaces in (2.13) defined on quadrilateral elements, because these spaces are available in the software package deal.II [35, 36], which we use for performing the numerical tests. The same analysis can be applied for other spaces defined on other mesh elements, e.g. triangular elements.*

121 We proceed to the discretization analysis under the following assumption

122 ASSUMPTION 2.3. *Let $1 < p < 2$, $q = \frac{p}{p-1}$ and let $\tau = \varepsilon + |\nabla u| + |\nabla \mathcal{I}_h^k u|$. We assume $\frac{1}{\tau} \in L^q(\Omega)$.*

123 REMARK 2.2. *Note that, if $\tau \geq 1$, i.e., $\varepsilon = 1$, we have $\|\frac{1}{\tau}\|_{L^q(\Omega)} \leq \|\tau\|_{L^q(\Omega)}$. Now, we can mention two cases. If $u \in W^{l,p}$ with $l = 2$ and $2 > p > \frac{2d}{d+1}$, (for the $d = 2$ case that we study here, this means
124 $2 > p > \frac{4}{3}$), then using the Sobolev inequality $\|u\|_{L^q(\Omega)} \leq \|u\|_{W^{1,p}(\Omega)}$, it is easy to see that the Assumption 2.3
125 is fulfilled. Furthermore, if $u \in W^{l,p}$ with $l > 2$ and $(l-1)p - d > 0$, then again using the Sobolev inequality
126 $\|u\|_{L^q(\Omega)} \leq \|u\|_{W^{1,p}(\Omega)}$, we can see that the Assumption 2.3 is fulfilled.*

In the analysis below, we will use some special cases of Sobolev embeddings theorem, see [33]. Let \hat{E} be the reference element, for example the unit square, and let $\hat{u} \in W^{1+m,p}(\hat{E})$ with the integer $m \geq 1$ and $2 > p > 1$ such that: either $mp > d$ and $p > 1$ or $mp < d$ and $p > \frac{2d}{d+1}$, (i.e., for the $d = 2$, we have $mp > 2$ and $p > 1$ or $mp < 2$ and $p > \frac{4}{3}$). Then for $q > p$, it holds that

$$(2.15) \quad \|\hat{u}\|_{W^{1,q}(\hat{E})} \leq C \|\hat{u}\|_{W^{1+m,p}(\hat{E})},$$

where the constant C depends on d, p, q, m and \hat{E} . By the Sobolev embedding (2.15) we can derive that

$$(2.16) \quad \|\hat{u}\|_{W^{1,q}(\hat{E})} \leq C \left(\|\hat{u}\|_{L^p(\hat{E})}^p + |\hat{u}|_{W^{1,p}(\hat{E})}^p + |\hat{u}|_{W^{2,p}(\hat{E})}^p \right)^{\frac{1}{p}}.$$

Let $E \in T_h$ and let T_E be the unique affine transformation that maps the reference element \hat{E} to $E \in T_h$,

$$(2.17) \quad T_E : \hat{E} \rightarrow E, \quad \text{with} \quad T_E(\hat{x}) = B\hat{x} + b,$$

where $|\det(B)| = |E|$. By applying a change of variables and using scaling arguments, we can show that

$$(2.18) \quad |\hat{u}|_{W^{j,p}(\hat{E})} = h^{j-\frac{d}{p}} |u|_{W^{j,p}(E)}, \quad \text{for } 0 \leq j \leq l.$$

Finally, using (2.18) into (2.16), a simple computation yields that

$$(2.19) \quad |u|_{W^{1,q}(E)}^2 \leq Ch^{2(\frac{d}{q}-\frac{d}{p}-1)} (\|u\|_{L^p(E)}^p + h^p |u|_{W^{1,p}(E)}^p + h^{2p} |u|_{W^{2,p}(E)}^p)^{\frac{2}{p}}.$$

COROLLARY 2.3. *Let $u \in W^{1,q}(\Omega) \cap W^{l,p}(\Omega)$ with $q > p$, $l \geq 2$, and let \mathcal{I}_h^k be the corresponding interpolation. Then by (2.14) and (2.19), it follows that*

$$(2.20) \quad |u - \mathcal{I}_h^k u|_{W^{1,q}(E)}^2 \leq Ch^{2(\frac{d}{q}-\frac{d}{p}-1)} h^{2l} \|u\|_{W^{2,p}(E)}^2.$$

128 **LEMMA 2.4.** *Let \mathbf{A} be given by (2.2) and let \mathbf{F} be defined by (2.3). For $\delta > 0$, there exist a constant $c(\delta)$, such that for $u, v, w \in V$ the following relation holds*

$$(2.21) \quad \int_{\Omega} (\mathbf{A}(\nabla u) - \mathbf{A}(\nabla v)) \cdot (\nabla w - \nabla v) dx \leq \delta \int_{\Omega} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla v)|^2 dx + c(\delta) \int_{\Omega} |\mathbf{F}(\nabla w) - \mathbf{F}(\nabla v)|^2 dx.$$

129

130 *Proof.* The proof can be found in [14]. \square

2.5. Finite element approximation and error bounds . In this paragraph, we present the finite element discretization of (2.5). We develop our analysis inspired by the results in [14]. The finite element approximation of (2.5) reads as follows: find $u_h \in V_{D,h}^{(k)}$ such that for all $\phi_h \in V_{0,h}^{(k)}$ holds

$$(2.22) \quad B(u_h, \phi_h) = l_f(\phi_h).$$

131 **PROPOSITION 2.5.**

(i) *Let $\{a_i\}_{i=1}^N$ be a sequence of nonnegative numbers. If $p < q < \infty$ then*

$$(2.23) \quad \left(\sum_{i=1}^N a_i^p \right)^{\frac{1}{p}} \leq N^{\frac{1}{p}-\frac{1}{q}} \left(\sum_{i=1}^N a_i^q \right)^{\frac{1}{q}}.$$

(ii) *Let $1 < p \leq q < \infty$ and let $E \in T_h$. If $u \in L^q(E)$ then there exist a C depending on the quasi-uniformity properties of T_h such that*

$$(2.24) \quad \|u\|_{L^p(E)} \leq Ch^{d(\frac{1}{p}-\frac{1}{q})} \|u\|_{L^q(E)}.$$

(iii) *Let $\{a_i\}_{i=1}^N$ be a sequence of nonnegative numbers. If $1 < p < q < \infty$ then*

$$(2.25) \quad \left(\sum_{i=1}^N a_i^q \right)^{\frac{1}{q}} \leq \left(\sum_{i=1}^N a_i^p \right)^{\frac{1}{p}}.$$

132 *Proof.*

(i) Let $p^* = \frac{q}{p} > 1$ and $q^* = \frac{p^*}{p^*-1} = \frac{q}{q-p}$ its conjugate exponent. Recalling (2.9a) for vectors, we have

$$(2.26) \quad \left(\sum_{i=1}^N a_i^p 1 \right) \leq \left(\sum_{i=1}^N a_i^q \right)^{\frac{p}{q}} \left(\sum_{i=1}^N 1 \right)^{\frac{q-p}{q}}.$$

133 Taking the $\frac{1}{p}$ th root in (2.26), we obtain (2.23).

134 (ii) Inequality (2.24) can be shown in the same way as inequality (2.23).

135 (iii) We observe that the function $f(x) = (a_1^x + a_2^x + \dots + a_N^x)^{\frac{1}{x}}$ is decreasing for $x \geq 1$. Then inequality
136 (2.25) follows.

\square

137 Below we prove the basic interpolation estimates in the related quasi-norms.

THEOREM 2.6. *Let u satisfies Assumptions 2.1 and 2.2 and let $\mathcal{I}_h^k u$ be the corresponding interpolant as defined above. Then, there exist $c \geq 0$ depending on ε , $|\nabla u|$, $|\nabla \mathcal{I}_h^k u|$, the constant in (2.14) and the constant in (2.19), but independent of the grid size h , such that the following estimate holds true*

$$(2.27) \quad \int_{\Omega} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla \mathcal{I}_h^k u)|^2 dx \leq ch^{2(l-1)} \|u\|_{W^{l,p}(\Omega)}^2.$$

138

Proof. For simplicity in the formulas below, we introduce the notation $\tau := \widehat{\varphi}(|\nabla u| + |\nabla \mathcal{I}_h^k u|)^{\frac{1}{p-2}} = \varepsilon + |\nabla u| + |\nabla \mathcal{I}_h^k u|$. By the relations (2.12), we conclude that

$$(2.28) \quad \int_{\Omega} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla \mathcal{I}_h^k u)|^2 dx \leq \int_{\Omega} \widehat{\varphi}(|\nabla u| + |\nabla \mathcal{I}_h^k u|) |\nabla u - \nabla \mathcal{I}_h^k u|^2 dx.$$

139 Let us first prove the required relation for $p > 2$. Applying (2.9a) in (2.28) and setting $q = \frac{p}{2}$, we obtain

$$(2.29) \quad \begin{aligned} \int_{\Omega} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla \mathcal{I}_h^k u)|^2 dx &\leq c \left(\int_{\Omega} |\nabla u - \nabla \mathcal{I}_h^k u|^{\frac{2p}{2}} dx \right)^{\frac{2}{p}} \left(\int_{\Omega} (\widehat{\varphi}(|\nabla u| + |\nabla \mathcal{I}_h^k u|)^{\frac{p}{p-2}} dx \right)^{\frac{p-2}{p}} \\ &\leq c |\nabla u - \nabla \mathcal{I}_h^k u|_{W^{1,p}(\Omega)}^2 \|\tau\|_{L^p(\Omega)}^{p-2} \leq ch^{2(l-1)} \|\tau\|_{L^p(\Omega)}^{p-2}, \end{aligned}$$

140 where the interpolation estimate (2.14) has been used.

For the case $p < 2$ we proceed as follows. Let $E \in \mathcal{T}_h(\Omega)$ and let $q^* = \frac{q}{2}$ and $p^* = \frac{p}{2-p}$. By relations (2.12), we can conclude that

$$(2.30) \quad \int_E |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla \mathcal{I}_h^k u)|^2 dx \leq \int_E \widehat{\varphi}(|\nabla u| + |\nabla \mathcal{I}_h^k u|) |\nabla u - \nabla \mathcal{I}_h^k u|^2 dx.$$

We apply (2.9a) in (2.30) and consequently (2.19), and obtain

$$(2.31) \quad \begin{aligned} \sum_E \int_E |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla \mathcal{I}_h^k u)|^2 dx &\leq c \sum_E \left(\int_E |\nabla u - \nabla \mathcal{I}_h^k u|^{\frac{2q}{2}} dx \right)^{\frac{2}{q}} \left(\int_E \tau^{-p} dx \right)^{\frac{2-p}{p}} \\ &\leq c \sum_E |u - \mathcal{I}_h^k u|_{W^{1,q}(E)}^2 \|\frac{1}{\tau}\|_{L^p(E)}^{2-p} \leq c \left(\sum_E |u - \mathcal{I}_h^k u|_{W^{1,q}(E)}^q \right)^{\frac{2}{q}} \left(\sum_E \|\frac{1}{\tau}\|_{L^p(E)}^p \right)^{\frac{2-p}{p}} \\ &\leq c \left(\sum_E |u - \mathcal{I}_h^k u|_{W^{1,q}(E)}^2 \right)^{\frac{2}{q}} \left(\sum_E \|\frac{1}{\tau}\|_{L^p(E)}^p \right)^{\frac{2}{p}} \left(\sum_E \|\frac{1}{\tau}\|_{L^p(E)}^p \right)^{-1} \\ &\leq ch^{2(\frac{d}{q} - \frac{d}{p} - 1)} h^{2l} \|u\|_{W^{2,p}(\Omega)}^2 h^{2d(\frac{1}{p} - \frac{1}{q})} \left(\sum_E \|\frac{1}{\tau}\|_{L^q(E)}^p \right)^{\frac{2}{p}} \left(\|\frac{1}{\tau}\|_{L^p(\Omega)}^p \right)^{-1} \\ &\leq ch^{2l-2} \|u\|_{W^{2,p}(\Omega)}^2 \left(\sum_E \|\frac{1}{\tau}\|_{L^q(E)}^p \right)^{\frac{2}{p}} \left(\|\frac{1}{\tau}\|_{L^p(\Omega)}^p \right)^{-1}, \end{aligned}$$

where the interpolation estimate (2.14) and inequalities in Proposition 2.5 have been used. Finally, putting together the results (2.29) and (2.31), we have that

$$(2.32) \quad \int_{\Omega} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla \mathcal{I}_h^k u)|^2 dx \leq ch^{2l-2} \|u\|_{W^{l,p}(\Omega)}^2.$$

141 This proves the theorem. \square

142 Let us now use the previous results for showing an estimate for the approximation error $u_h - u$.

THEOREM 2.7. *Let $u \in V$ be the solution of (2.5) under the Assumption 2.1, and let $u_h \in V_{D,h}^{(k)}$ be the solution of (2.22). Then, there exist $C \geq 0$ depending on the constant in (2.27) but independent of the grid size h , such that*

$$(2.33) \quad \int_{\Omega} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla u_h)|^2 dx \leq Ch^{2(l-1)} \|u\|_{W^{l,p}(\Omega)}^2.$$

143 *Proof.* Let $\phi_h \in V_{0,h}^{(k)}$ and let $\mathcal{I}_h^k u \in V_{D,h}^{(k)}$ be the interpolant of u . By forms (2.5) and (2.22), we can deduce that

$$(2.34) \quad \int_{\Omega} (\mathbf{A}(\nabla u) - \mathbf{A}(\nabla u_h)) \cdot \nabla \phi_h dx = 0,$$

where it follows that

$$(2.35) \quad \int_{\Omega} (\mathbf{A}(\nabla u_h) - \mathbf{A}(\nabla \mathcal{I}_h^k u)) \cdot \nabla \phi_h dx = \int_{\Omega} (\mathbf{A}(\nabla u) - \mathbf{A}(\nabla \mathcal{I}_h^k u)) \cdot \nabla \phi_h dx.$$

144 Now, choosing $\phi_h = u_h - \mathcal{I}_h^k u$ and $\delta > 0$ small enough, relations (2.12) and inequality (2.21) imply that

$$(2.36) \quad \int_{\Omega} |\mathbf{F}(\nabla u_h) - \mathbf{F}(\nabla \mathcal{I}_h^k u)|^2 dx \leq c(\delta) \int_{\Omega} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla \mathcal{I}_h^k u)|^2 dx,$$

145 where by triangle inequality it follows that

$$(2.37) \quad \int_{\Omega} |\mathbf{F}(\nabla u_h) - \mathbf{F}(\nabla u)|^2 dx \leq c(\delta) \int_{\Omega} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla \mathcal{I}_h^k u)|^2 dx.$$

146 The desired estimate (2.33) follows immediately by applying the interpolation estimate (2.27) on the right
147 hand side of (2.37). \square

148 **3. Nonlinear iterative processes.** In this section, we design several iterative methods for solving
149 problem (2.22). All proposed schemes are based on Newton-type approaches. We begin with an inspection
150 of the behavior of the eigenvalues of the Jacobian matrix of \mathbf{A} , see (2.2), in order to get a prefigure about the
151 behavior of the corresponding Newton iterative matrix around the critical points. Then, in Section 3.3, two
152 Newton-type iterative methods, e.g., residual-based descent and error-oriented techniques, are formulated. In
153 Section 3.4, augmented Lagrangian-type techniques are employed in order to reformulate the problem (2.1)
154 as a saddle point problem. We discretize this last problem using a classical finite element procedure and
155 solve the resulting nonlinear system by applying two iterative procedures. Our final goal is to investigate the
156 performance of all the aforementioned methods through the numerical examples.

3.1. Preliminary considerations. In view of the form of operator \mathbf{A} in (2.2), we introduce the function $\mathbf{A} \in C^1(\mathbb{R}^2, \mathbb{R}^2)$ defined by

$$(3.1) \quad \mathbf{A}(\eta) = (\varepsilon^2 + \eta_1^2 + \eta_2^2)^{\frac{p-2}{2}} (\eta_1, \eta_2).$$

By an easy computation, we can show that

$$(3.2) \quad J_{\mathbf{A}} = \begin{bmatrix} (\eta_1^2 + \eta_2^2 + \varepsilon^2)^{(p-4)/2} ((p-1)\eta_1^2 + \eta_2^2 + \varepsilon^2) & (p-2)\eta_1\eta_2(\eta_1^2 + \eta_2^2 + \varepsilon^2)^{(p-4)/2} \\ (p-2)\eta_1\eta_2(\eta_1^2 + \eta_2^2 + \varepsilon^2)^{(p-4)/2} & (\eta_1^2 + \eta_2^2 + \varepsilon^2)^{(p-4)/2} ((p-1)\eta_2^2 + \eta_1^2 + \varepsilon^2) \end{bmatrix}$$

with the eigenvalues

$$(3.3a) \quad \lambda_1 = (\varepsilon^2 + \eta_1^2 + \eta_2^2)^{(p-2)/2},$$

$$(3.3b) \quad \lambda_2 = (\varepsilon^2 + \eta_1^2 + \eta_2^2)^{(p-4)/2} ((p-1)(\eta_1^2 + \eta_2^2) + \varepsilon^2).$$

157 The behavior of the eigenvalues of the Jacobian $J_{\mathbf{A}} := \frac{\partial \mathbf{A}(\eta)}{\partial \eta}$ with respect to the parameters ε and p , outline
158 the behavior of the eigenvalues of the corresponding Jacobian matrix, which appears in the Newton iterative
159 process. Namely, for the cases where the eigenvalues of $J_{\mathbf{A}}$ are close to zero, we can expect that Jacobian
160 matrix of the Newton method to be singular or ill-conditioned. Note that the eigenvalues λ_i , $i = 1, 2$ in (3.3)
161 have a similar form as the function $\widehat{\varphi}$ defined in (2.10). For example, if we set $t = |\eta|$ then $\lambda_i(t) \sim \widehat{\varphi}(t)$, $i = 1, 2$.
162 Let $r = (\eta_1, \eta_2)$ be the radial function with $r^2 \leq 1$. In Fig. 1, we plot λ_1 and λ_2 with respect to $r^2 = \eta_1^2 + \eta_2^2$
163 for several values of ε and p . We observe that for $p < 2$ and $\varepsilon < 1$ the values of both eigenvalues are getting
164 very high as $r^2 \rightarrow 0$ and the values are reduced as $r^2 \rightarrow 1$, see Figs 1(a),(b),(d),(e). Conversely for the case
165 where $\varepsilon = 1$ the eigenvalues are close to one, see Figs. 1(c),(f). Concerning the graphs for $p > 2$, we can
166 see that for $\varepsilon < 1$ the eigenvalues are close to zero as $r^2 \rightarrow 0$, and are increasing rapidly as $r^2 \rightarrow 1$, see
167 Figs. 1(g),(h). On the other hand, for $\varepsilon = 1$ the eigenvalues are greater than one for all r^2 values, see
168 Fig. 1(i).

169 **REMARK 3.1.** For simplicity the analysis above has been presented in \mathbb{R}^2 . It can be easily extended to
170 \mathbb{R}^{N_h} , where N_h is the dimension of $V_{D,h}^{(k)}$.

3.2. The nonlinear algebraic system. Let N_h be the dimension of the space $V_{D,h}^{(k)}$ and $\{\phi_{h,i}\}_{i=1}^{N_h}$ are
the basis functions of $V_{D,h}^{(k)}$. The solution $u_h \in V_{D,h}^{(k)}$ of problem (2.22) is expressed as $u_h = \sum_i^{N_h} U_i \phi_{h,i}(x)$
where U_i are the degrees of freedom. When this expression is substituted into (2.22), we obtain the following
nonlinear algebraic problem: Find the vector $\mathbf{U} = [U_1, \dots, U_i, \dots, U_{N_h}]$ which satisfies the system of N_h
nonlinear equations

$$(3.4) \quad \mathbf{B}(\mathbf{U}) = \mathbf{f},$$

where the entries $B_i(\mathbf{U})$ of \mathbf{B} and f_i of \mathbf{f} are specified by (2.5a), i.e.,

$$(3.5) \quad B_i(\mathbf{U}) = B(u_h, \phi_{h,i}), \text{ and } f_i = l_f(\phi_{h,i}).$$

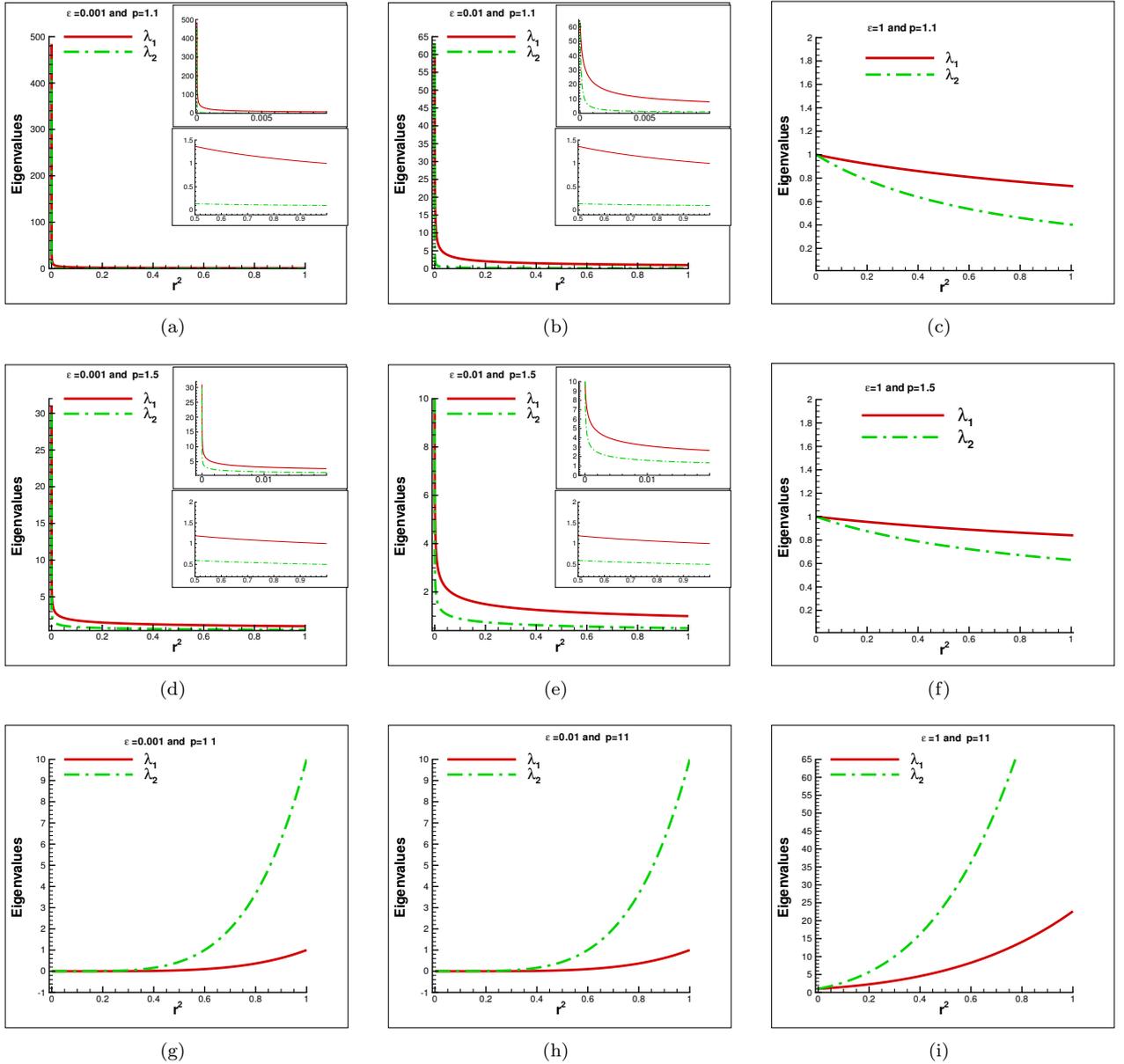


FIG. 1. Eigenvalues λ_1, λ_2 of $J_{\mathbf{A}}$: (a) plots for $\varepsilon = 0.001$ and $p = 1.1$, (b) plots for $\varepsilon = 0.01$ and $p = 1.1$, (c) plots for $\varepsilon = 1$ and $p = 1.1$, (d) plots for $\varepsilon = 0.001$ and $p = 1.5$, (e) plots for $\varepsilon = 0.01$ and $p = 1.5$, (f) plots for $\varepsilon = 1$ and $p = 1.5$, (g) plots for $\varepsilon = 0.001$ and $p = 11$, (h) plots for $\varepsilon = 0.01$ and $p = 11$, (i) plots for $\varepsilon = 1$ and $p = 11$.

171 The nonlinear system (3.4) is generally large-sized and hence it is important to develop efficient iterative
 172 methods for its numerical solution, [37]. Also, it is important to investigate the influence of the parameters
 173 p and ε with respect to the convergence speed of the nonlinear iterative procedure.

174 Throughout this section, \mathbf{U}^* will denote the exact solution of (3.4) and \mathbf{U}^n the solution derived at step n
 175 of the iterative procedure. Furthermore $\|\cdot\|$ will denote the l^2 -norm as well the induced matrix norm. Recall
 176 that the condition number of a matrix A relative to the norm $\|\cdot\|$ is given by $\kappa(A) = \|A\| \|A^{-1}\|$. We denote
 177 the set of eigenvalues of A by $\sigma(A)$ and $\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|$ is the the spectral radius of A .

178 3.3. Newton-like methods.

3.3.1. Definition and algorithms.. In this subsection, we present two Newton techniques for solving

(3.4). Let $v_h = \sum_{i=1}^{N_h} V_i \phi_{h,i}$ to be an element in $V_{D,h}^{(k)}$ with $\mathbf{V} = [V_1, \dots, V_i, \dots, V_{N_h}]$. Using (2.8) we define

$$(3.6) \quad \mathbf{B}'(\mathbf{V})_{j,i} := B'(v_h)(\phi_{h,i}, \phi_{h,j}), \quad i, j = 1, \dots, N_h,$$

179 to be the iterative matrix of Newton's method. Then the main steps of a Newton method (in a matrix-vector
 180 form) can be described as follows:

181 ALGORITHM 3.1 (Newton's method as defect-correction scheme). Choose an initial Newton guess \mathbf{U}^0 :

1. For $n = 0, 1, 2, \dots, n_{max}$ compute the solution \mathbf{Z}^n from

$$(3.7a) \quad \mathbf{B}'(\mathbf{U}^n)\mathbf{Z}^n = -(\mathbf{B}(\mathbf{U}^n) - \mathbf{f}),$$

Determine $\lambda_n \in (0, 1]$, (e.g., see Algorithm 3.2)

$$(3.7b) \quad \text{Update: } \mathbf{U}^{n+1} = \mathbf{U}^n + \lambda_n \mathbf{Z}^n.$$

182 2. If the stopping criterion (see below either (3.8) or (3.9)) is satisfied then exit and set $\mathbf{U}^* := \mathbf{U}^{n+1}$.

183 Otherwise repeat 1.

Usually the stopping criterion is related to a control of the magnitude of the residual of the new iteration \mathbf{U}^{n+1} , i.e.,

$$(3.8) \quad res_{n+1} := \|\mathbf{B}(\mathbf{U}^{n+1}) - \mathbf{f}\| \leq TOL,$$

where $TOL \sim 10^{-12}$. Based on that we determine the parameter λ_n for the updating step (3.7b). However, in the literature, see e.g. [29, 37], more sophisticated Newton methods have been proposed, where the stopping criterion, the damping parameter and the updating step are determined by estimating the magnitude of the norms

$$(3.9) \quad \|\mathbf{Z}^n\| \leq TOL, \quad \text{or} \quad \|\mathbf{Z}_{simp}^{n+1}\| \leq TOL,$$

184 where \mathbf{Z}_{simp}^{n+1} is a Newton update from solving a simplified problem, see below (3.12).

185 Thus, we formulate two different Newton procedures, (i) a residual-based line-search procedure, which
186 is related to the first case (3.8); see Algorithm 3.2. And (ii) an error-oriented Newton procedure, which is
187 related to second case (3.9); see Algorithm 3.3.

188 ALGORITHM 3.2 (Residual-based line-search). In this procedure, λ_n is specified by a direct decreasing of
189 the residual in each n -step:

190 1. Set $\lambda_{n,l=0} := \lambda_{initial} = 1$

191 2. For $l = 0, \dots, l_M$, compute \mathbf{Z}^n by (3.7a) and update $\mathbf{U}^{n+1,l} = \mathbf{U}^n + \lambda_{n,l} \mathbf{Z}^n$

192 3. Evaluate $res_{n,l+1}$,

193 4. If $res_{n,l+1} \leq res_{n,l}$ then solution found $\mathbf{U}^{n+1} := \mathbf{U}^{n+1,l}$,

194 otherwise go to step 2, setting $\lambda_{n,l+1} = \frac{\lambda_{n,l}}{2}$,

195 5. If $l+1 > l_M$ stop. Convergence failure.

196 ALGORITHM 3.3 (Error-oriented Newton's method). This procedure is based on a natural monotonicity
197 test, i.e., $\|\mathbf{Z}_{simp}^{n+1}\| < \|\mathbf{Z}^n\|$.

198 1. Choose an initial guess \mathbf{U}^0 . Set the minimal damping factor, e.g., $\lambda_{min} \sim 10^{-8}$ and $\lambda_0 < 1$. For
199 $n = 0, 1, 2, 3, \dots$:

2. Solve

$$\mathbf{B}'(\mathbf{U}^n)\mathbf{Z}^n = -(\mathbf{B}(\mathbf{U}^n) - \mathbf{f}), \quad (\text{see Newton step (3.7a)}).$$

If $\|\mathbf{Z}^n\| \leq TOL$ then the solution found and set

$$\mathbf{U}^* := \mathbf{U}^n + \mathbf{Z}^n.$$

3. If $\|\mathbf{Z}^n\| > TOL$, then take \mathbf{Z}_{simp}^n , which has been already computed in the previous Newton step $n-1$
via (3.12). Determine a new prediction value λ_n for the damping factor as follows:

$$\lambda_n := \min(1, \mu_n), \quad \text{where} \quad \mu_n := \frac{\|\mathbf{Z}^{n-1}\| \cdot \|\mathbf{Z}_{simp}^n\|}{\|\mathbf{Z}_{simp}^n - \mathbf{Z}^n\| \cdot \|\mathbf{Z}^n\|}.$$

If

$$(3.10) \quad \lambda_n < \lambda_{min}$$

200 then stop. Convergence failure and abort computation.

4. If $\lambda_n > \lambda_{min}$ continue and compute trial iterate

$$(3.11) \quad \mathbf{U}^{n+1} := \mathbf{U}^n + \lambda_n \mathbf{Z}^n$$

and evaluate the new residual $-(\mathbf{B}(\mathbf{U}^{n+1}) - \mathbf{f})$. Solve the simplified linear system using the old
Jacobian and the new residual

$$(3.12) \quad \mathbf{B}'(\mathbf{U}^n)\mathbf{Z}_{simp}^{n+1} = -(\mathbf{B}(\mathbf{U}^{n+1}) - \mathbf{f}).$$

5. Compute the monitoring functions:

$$\theta_n := \frac{\|\mathbf{Z}_{simp}^{n+1}\|}{\|\mathbf{Z}^n\|}, \quad \mu'_n := \frac{0.5\|\mathbf{Z}^n\|\lambda_n^2}{\|\mathbf{Z}_{simp}^{n+1} - (1 - \lambda_n)\mathbf{Z}^n\|}.$$

6. If $\theta_n \geq 1$ (no convergence of the updates), then set

$$\lambda_n := \lambda'_n := \min(\mu'_n, \frac{1}{2}\lambda_n)$$

and go to (3.10) and continue from there.

7. Else, if $\theta_n < 1$, we have convergence and continue with the next steps.

Set $\lambda'_n := \min(1, \mu'_n)$.

(a) If $\lambda'_n = \lambda_n = 1$ then: if $\|\mathbf{Z}_{simp}^{n+1}\| \leq TOL$ stop and the solution is found:

$$\mathbf{U}^* := \mathbf{U}^{n+1} + \mathbf{Z}_{simp}^{n+1},$$

otherwise go to the beginning to Step 2 and increment $n \rightarrow n + 1$.

(b) If $\lambda'_n \neq \lambda_n < 1$ go to the beginning to Step 2 and increment $n \rightarrow n + 1$.

REMARK 3.2 (Motivation of Step 3 and Step 5). In the above algorithm, the important steps are No. 3 and 5. Both are based on adaptive trust region strategies as derived in [29], Section 3.3.3. Specifically, Step 3 is a prediction strategy that is based on a estimate of the simplified update and the full Newton update, while using Lipschitz continuity of the Jacobian. Step 5, is a correction strategy that is based on a damped Newton iteration.

We underline that for finding the numerical solution \mathbf{U}^* the generated global error includes two parts, one coming from the FE discretization and the second coming by the Newton iterative procedure. Recalling (2.33) and the stopping criterion in the Newton algorithms, we can practically express the global error as

$$\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla u_{h,n})\| \lesssim \mathcal{O}(h^{l-1}) + res_n,$$

where res_n is the residual at the last n -th Newton iteration step. Thus, the Newton iterations must be continued until the discretization error component dominates.

REMARK 3.3. In view of Jacobian $J_{\mathbf{A}}$ defined in (3.2), the entries of iterative matrix in (3.6) are equal to $\mathbf{B}'(\mathbf{V})_{j,i} = \int_{\Omega} J_{\mathbf{A}}(\nabla v_h) \nabla \phi_{h,i} \nabla \phi_{h,j} dx$ and consequently Newton's step in (3.7a) can be expressed as: Find $z_h \in V_{D,h}^{(k)}$ such that

$$\int_{\Omega} J_{\mathbf{A}}(\nabla u_h^n) \nabla z_h \cdot \nabla \phi_h dx = - \left(\int_{\Omega} \mathbf{A}(\nabla u_h^n) \cdot \nabla \phi_h dx - \int_{\Omega} f \phi_h dx \right), \quad \phi_h \in V_{0,h}^{(k)},$$

where $u_h^n \in V_{D,h}^{(k)}$ is the updated numerical solution of the previous step.

3.3.2. Remarks on the properties for the condition number of \mathbf{B}' . In practice we can check the well-conditioned status of system (3.4) (and also (3.7a)) through the condition number. The condition number $\kappa(\mathbf{B}')$ is a measure of sensitivity of the solution to relative small perturbations of the system, e.g., finite precision arithmetic, as well as a measure of the distance from a singular matrix. Roughly speaking an almost singular matrix is expected to have a large condition number and thus we expect to meet problems during the solution of the system. In our studies this could be the case where $p > 2$ and $\varepsilon \rightarrow 0$, see Figs. 1(g),(h). Next, we quote a Lemma related to inverse estimates for $\phi_h \in V_{D,h}^{(k)}$.

LEMMA 3.1. Let $\phi_h = \sum_{i=1}^{N_h} v_i \phi_{h,i}$ and $\mathbf{v} = (v_1, \dots, v_{N_h})$. There exist positive constants C_0 , C_m and C_M depending only on the quasi-uniform properties of the meshes and on k , such that

$$\|\nabla \phi_h\|_{L^2(\Omega)} \leq C_0 h^{-1} \|\phi_h\|_{L^2(\Omega)},$$

$$C_m h \|\mathbf{v}\| \leq \|\phi_h\|_{L^2(\Omega)} \leq C_M h \|\mathbf{v}\|,$$

where $\|\mathbf{v}\|^2 = v_1^2 + \dots + v_{N_h}^2$.

Proof. Both relations can be shown by applying scaling arguments, see details in [28]. \square

Based on (2.8), we can clearly see that the matrix \mathbf{B}' defined in (3.6) is symmetric and is positive definite. Indeed by (3.6) and (2.8), we have

$$\begin{aligned} B'(u_h)(w_h, w_h) &= \int_{\Omega} (\varepsilon^2 + |\nabla u_h|^2)^{\frac{p-2}{2}} |\nabla w_h|^2 dx \\ &+ (p-2) \int_{\Omega} (\varepsilon^2 + |\nabla u_h|^2)^{\frac{p-4}{2}} (\nabla u_h, \nabla w_h)^2 dx. \end{aligned}$$

Now, let us try to derive bounds for $\frac{B'(u_h)(w_h, w_h)}{\|\mathbf{w}\|^2}$. As a result, using that $(a^2 + \beta^2) \leq (a + \beta)^2 \leq 2(a^2 + \beta^2)$ for $a, \beta > 0$, we have that

$$\begin{aligned}
 (3.17) \quad B'(u_h)(w_h, w_h) &\leq \int_{\Omega} (\varepsilon^2 + |\nabla u_h|^2)^{\frac{p-2}{2}} |\nabla w_h|^2 dx \\
 &+ |p-2| \int_{\Omega} (\varepsilon^2 + |\nabla u_h|^2)^{\frac{p-4}{2}} (\varepsilon^2 + |\nabla u_h|^2) |\nabla w_h|^2 dx \\
 &\leq C(p) \int_{\Omega} (\varepsilon^2 + |\nabla u_h|^2)^{\frac{p-2}{2}} |\nabla w_h|^2 dx.
 \end{aligned}$$

If $p > 2$ we get immediately that

$$(3.18) \quad B'(u_h)(w_h, w_h) \geq \int_{\Omega} (\varepsilon^2 + |\nabla u_h|^2)^{\frac{p-2}{2}} |\nabla w_h|^2 dx.$$

For the case $1 < p < 2$, a simple modification of the last term in (3.16) yields

$$\begin{aligned}
 (3.19) \quad B'(u_h)(w_h, w_h) &\geq \int_{\Omega} (\varepsilon^2 + |\nabla u_h|^2)^{\frac{p-2}{2}} |\nabla w_h|^2 dx \\
 &+ (p-2) \int_{\Omega} (\varepsilon^2 + |\nabla u_h|^2)^{\frac{p-4}{2}} |\nabla u_h|^2 |\nabla w_h|^2 dx \\
 &\geq (p-1) \int_{\Omega} (\varepsilon^2 + |\nabla u_h|^2)^{\frac{p-2}{2}} |\nabla w_h|^2 dx \\
 &> 0.
 \end{aligned}$$

222 In the following, inspired by [31], we try to derive estimates for the condition number $\kappa(\mathbf{B}')$.

PROPOSITION 3.2. *The condition number $\kappa(\mathbf{B}')$ can be estimated by*

$$(3.20) \quad \kappa(\mathbf{B}') = \frac{\lambda_{Max}}{\lambda_{min}} \leq \begin{cases} C(\varepsilon + |\nabla u_h|_{\infty})^{p-2} \varepsilon^{2-p} h^{-2} & \text{for } p > 2, \\ C(\varepsilon + |\nabla u_h|_{\infty})^{2-p} \varepsilon^{p-2} h^{-2} & \text{for } 1 < p < 2, \end{cases}$$

223 where λ_{Max} and λ_{min} are the largest and the smallest eigenvalue of \mathbf{B}' respectively.

Proof. For $p > 2$, it follows from (3.17) and (3.15) that

$$\begin{aligned}
 (3.21) \quad \frac{B'(u_h)(w_h, w_h)}{\|\mathbf{w}\|^2} &\leq C(\varepsilon + \|\nabla u_h\|_{\infty})^{p-2} \frac{|\nabla w_h|_{L^2(\Omega)}^2}{h^{-2} \|w_h\|_{L^2(\Omega)}^2} \leq C(\varepsilon + \|\nabla u_h\|_{\infty})^{p-2} \frac{h^{-2} \|w_h\|_{L^2(\Omega)}^2}{h^{-2} \|w_h\|_{L^2(\Omega)}^2} \\
 &\leq C(\varepsilon + |\nabla u_h|_{\infty})^{p-2}.
 \end{aligned}$$

For $p > 2$ using (3.18) and (2.9b), we obtain

$$(3.22) \quad \frac{B'(u_h)(w_h, w_h)}{\|\mathbf{w}\|^2} \geq C\varepsilon^{p-2} \frac{|\nabla w_h|_{L^2(\Omega)}^2}{h^{-2} \|w_h\|_{L^2(\Omega)}^2} \geq C\varepsilon^{p-2} \frac{\|w_h\|_{L^2(\Omega)}^2}{h^{-2} \|w_h\|_{L^2(\Omega)}^2} \geq C\varepsilon^{p-2} h^2.$$

On the other hand for $1 < p < 2$ we have

$$(3.23) \quad \frac{B'(u_h)(w_h, w_h)}{\|\mathbf{w}\|^2} \leq C\varepsilon^{p-2} \frac{|\nabla w_h|_{L^2(\Omega)}^2}{h^{-2} \|w_h\|_{L^2(\Omega)}^2} \leq C\varepsilon^{p-2} \frac{h^{-2} \|w_h\|_{L^2(\Omega)}^2}{h^{-2} \|w_h\|_{L^2(\Omega)}^2} \leq C\varepsilon^{p-2}.$$

Also, for $1 < p < 2$ using (3.19) and (2.9b), we obtain

$$\begin{aligned}
 (3.24) \quad \frac{B'(u_h)(w_h, w_h)}{\|\mathbf{w}\|^2} &\geq C(\varepsilon + \|\nabla u_h\|_{\infty})^{p-2} \frac{|\nabla w_h|_{L^2(\Omega)}^2}{h^{-2} \|w_h\|_{L^2(\Omega)}^2} \geq C(\varepsilon + \|\nabla u_h\|_{\infty})^{p-2} \frac{\|w_h\|_{L^2(\Omega)}^2}{h^{-2} \|w_h\|_{L^2(\Omega)}^2} \\
 &\geq C(\varepsilon + |\nabla u_h|_{\infty})^{p-2} h^2.
 \end{aligned}$$

224 Gathering together all the previous inequalities we can show (3.20). \square

225 REMARK 3.4. *We note that a necessary condition for obtaining the solution \mathbf{Z}^n during the Newton*
 226 *process, is that the iterative matrix \mathbf{B}' is invertible, see Algorithm 3.1. In the proofs of the convergence of*
 227 *Newton's methods that have been presented in the literature, the existence of \mathbf{B}'^{-1} (at least in a neighborhood*
 228 *of \mathbf{U}^*) is usually supposed. Additional assumptions are the existence of $\beta_m > 0$ and $\beta_M > 0$ such that*
 229 *$\beta_m \leq \|\mathbf{B}'(\mathbf{V})^{-1}\| \leq \beta_M$. Usually, the bound β_M is used to define the radius of the area around \mathbf{U}^* for*
 230 *choosing the initial \mathbf{U}^0 , in order to be possible to show analytically the convergence of the method, see [29],*
 231 *[37]. From (3.20) and (3.3), one sees that for our case it is difficult to find uniform bounds with respect to*
 232 *ε and p . In the numerical examples, we observed that, indeed the choice of \mathbf{U}^0 influences the performance*
 233 *of the Newton solver. Choosing as initial guess to be the projection of the previous mesh solution, we have a*
 234 *fast convergence of the iterative procedure, see for instance Table 2 of Example 1.*

Algorithm sALG1:for $\lambda^0 \in (L^q(\Omega))^2$ given**For** each ALG iterative step $i > 0, \in \mathbb{N}$ find $u^i := u^{i,n_M}$, $\mathbf{q}^i := \mathbf{q}^{i,n_M}$ where n_M isthe final index n for which the

inner Newton loop is converged.

For each Newton step $n = 0, 1, 2, \dots$ solve for fixed i :

$$\begin{cases} -\widehat{r}\Delta u^{i,n} + \widehat{r}\nabla \cdot \mathbf{q}^{i,n} = \nabla \cdot \lambda^i + f, \\ (\varepsilon^2 + |\mathbf{q}^{i,n}|^2)^{\frac{p-2}{2}} \mathbf{q}^{i,n} + \widehat{r}\mathbf{q}^{i,n} - \widehat{r}\nabla u^{i,n} = \lambda^i, \end{cases}$$

EndFor Newton loop**update:** $\lambda^{i+1} = \lambda^i + \widehat{\rho}(\nabla u^{i,n_M} - \mathbf{q}^{i,n_M})$ **EndFor** ALG loop**Algorithm mALG1:**for $\lambda^0 \in (L^q(\Omega))^2$ given**For** each iterative Newton step $n > 0 \in \mathbb{N}$ find u^n , \mathbf{q}^n and λ^n as follows

$$\begin{cases} -\widehat{r}\Delta u^n - \nabla \cdot \lambda^n + \widehat{r}\nabla \cdot \mathbf{q}^n = f, \\ ((\varepsilon^2 + |\mathbf{q}^n|^2)^{\frac{p-2}{2}} \mathbf{q}^n + \widehat{r}\mathbf{q}^n - \widehat{r}\nabla u^n - \lambda^n = 0, \\ \lambda^n - \widehat{\rho}(\nabla u^n - \mathbf{q}^n) = \lambda^{n-1}, \end{cases}$$

EndFor

TABLE 1

The steps of the two ALG1 algorithms.

3.4. Augmented Lagrangian techniques. In the following two paragraphs, we transform the original problem (2.1) into a saddle-point problem using augmented Lagrangian techniques. Let q be the conjugate exponent of p , that is $\frac{1}{p} + \frac{1}{q} = 1$, and let us define the space $W \subset W_D^{1,p} \times (L^p(\Omega))^2$ by

$$(3.25) \quad W = \{(v, \mathbf{q}) | (v, \mathbf{q}) \in W_D^{1,p} \times (L^p(\Omega))^2 : \nabla v - \mathbf{q} = 0\}.$$

Following [38], we introduce the augmented Lagrangian $\mathcal{L}_{\widehat{r}}$ defined, for $\widehat{r} > 0$, by

$$(3.26) \quad \mathcal{L}_{\widehat{r}}(v, \mathbf{q}, \lambda) = \frac{1}{p} \int_{\Omega} (\varepsilon^2 + |\mathbf{q}|^2)^{\frac{p}{2}} dx - \int_{\Omega} f v dx + \frac{\widehat{r}}{2} \int_{\Omega} |\nabla v - \mathbf{q}|^2 dx + \int_{\Omega} \lambda \cdot (\nabla v - \mathbf{q}) dx,$$

and the following associated to (3.26) saddle-point problem:

$$(3.27a) \quad \text{Find } \{u, \mathbf{q}, \lambda\} \in W_D^{1,p} \times (L^p(\Omega))^2 \times (L^q(\Omega))^2 \text{ such that}$$

$$(3.27b) \quad \mathcal{L}_{\widehat{r}}(u, \mathbf{q}, \mu) \leq \mathcal{L}_{\widehat{r}}(u, \mathbf{q}, \lambda) \leq \mathcal{L}_{\widehat{r}}(v, \mathbf{w}, \lambda),$$

$$(3.27c) \quad \forall \{v, \mathbf{w}, \mu\} \in W_D^{1,p} \times (L^p(\Omega))^2 \times (L^q(\Omega))^2.$$

235 In that way, we reduce the solution of (2.1) to finding the saddle-point solutions of $\mathcal{L}_{\widehat{r}}$. We apply a variational
 236 analysis to (3.27) and then for computing the saddle point solutions, we employ two variants of the first
 237 augmented Lagrangian iterative algorithm, (ALG1), described in [32]. The two algorithms are listed in Table
 238 1, where $0 < \widehat{\rho} < 2\widehat{r}$. In the first case the Lagrangian λ is updated separately by applying a post-processing
 239 procedure at the end of every iterative circle, which has been previously performed for computing u, \mathbf{q} . We
 240 call the first version as splitting ALG1, (sALG1), and we present the algorithm on the left column in Table
 241 1. In the second iterative ALG1 procedure, all the unknowns are simultaneously computed at every step,
 242 and we call this procedure as monolithic ALG1, (mALG1). The algorithm of the mALG1 iterative method
 243 is given on the right column in Table 1. The analysis for showing the existence and uniqueness of the saddle
 244 point problems seems to be quite hard, and as well the study of the convergence properties of the resulting
 245 ALG methods. Such results are discussed in [32] for several problems set in $W^{2,2}$.

At first glance the ALG1 technique seems not to be so attractive due to the introduction of the two variables \mathbf{q} and λ . But as we can see from Table 1, the ALG1 method helps us to simplify the nonlinear structure of the original problem. In particular we are able to uncouple the strong nonlinearity of the model and the derivatives of u . The convergence properties of algorithm sALG1 have been studied in [32] for $\widehat{r} > 0$. Recalling the finite dimensional spaces defined in (2.13), the discrete variational form of the sALG1 method has as follows: for u_h^0 given, find (u_h^i, \mathbf{q}_h^i) such that

$$(3.28a) \quad \widehat{r}(\nabla u_h^{i,n}, \nabla v_h) + (\widehat{r}\nabla \cdot \mathbf{q}_h^{i,n}, v_h) = (\nabla \cdot \lambda_h^i + f, v_h), \quad \forall v_h \in V_{0,h}^{(k)},$$

$$(3.28b) \quad ((\varepsilon^2 + |\mathbf{q}_h^{i,n}|^2)^{\frac{p-2}{2}} \mathbf{q}_h^{i,n}, \phi_h) - \widehat{r}(\nabla u_h^{i,n} - \mathbf{q}_h^{i,n}, \phi_h) = (\lambda_h^i, \phi_h), \quad \forall \phi_h \in (V_h^{(k)})^2,$$

$$(3.28c) \quad \text{Update: } \lambda_h^{i+1} = \lambda_h^i + \widehat{\rho}(\nabla u_h^{i,n_M} - \mathbf{q}_h^{i,n_M}).$$

Following a standard FE methodology, as in Section 3.2, we obtain the following nonlinear algebraic system

$$(3.29) \quad \mathbf{B}_{sALG1} \begin{pmatrix} \mathbf{U} \\ \mathbf{Q} \end{pmatrix} = \mathbf{f}_{sALG1},$$

where \mathbf{U} and \mathbf{Q} are the vectors with the degrees of freedom of $u_h^{i,n}$ and $\mathbf{q}_h^{i,n}$ respectively. A similar discrete variational form can be derived for the mALG1 method and then we get the following nonlinear algebraic system

$$(3.30) \quad \mathbf{B}_{mALG1} \begin{pmatrix} \mathbf{U} \\ \mathbf{Q} \\ \mathbf{\Lambda} \end{pmatrix} = \mathbf{f}_{mALG1},$$

246 where \mathbf{U} , \mathbf{Q} and $\mathbf{\Lambda}$ are the vectors with the degrees of freedom of u_h^n and \mathbf{q}_h^n and λ_h^n respectively.

REMARK 3.5. *The resulting nonlinear algebraic systems (3.29) and (3.30) can be solved by utilizing any of the two Newton variants, see Algorithm 3.2 and Algorithm 3.3. In the numerical examples, we use Algorithm 3.2 where the stopping criterion is defined in an analogous way as in (3.8). In particular, the stopping criterion for the augmented Lagrangian iterative loop in sALG1 is defined as*

$$\max\{\|\mathbf{U}^{i+1} - \mathbf{U}^i\|, \|\mathbf{Q}^{i+1} - \mathbf{Q}^i\|, \|\mathbf{\Lambda}^{i+1} - \mathbf{\Lambda}^i\|\} \leq 10^{-6}.$$

247
248 REMARK 3.6. *We notice that all linear equation systems in this paper are solved with the direct solver*
249 *UMPACK [39]. We are currently working on developing iterative solvers based on geometric multigrid.*

250 REMARK 3.7. *In the numerical examples, the parameters \hat{r} and $\hat{\rho}$ have been specified after carrying out*
251 *few computations using different values. As outlined in [32], the optimal choice for \hat{r} is an open question.*
252 *Through the performance of the numerical tests, we have observed that the method sALG1 is more sensitive*
253 *to the choice of \hat{r} than the mALG1. The mALG1 method appears to have greater robustness on the choice*
254 *of \hat{r} and $\hat{\rho}$ with respect to p . In the numerical examples, we have used $\hat{\rho} = \hat{r} = 1$ for sALG1 and $\hat{r} = 100$,*
255 *$\hat{\rho} = 0.1$ for mALG1. Choosing this high value for \hat{r} , in fact we increase the strength of the penalization in*
256 *the mALG1 method.*

257 **4. Numerical examples.** In this section, we perform several numerical tests by selecting different
258 values for the exponent p and investigate the order of accuracy of the proposed FE method. We mainly
259 focus on cases where $1 < p < 2$ but do also consider a few tests with $p > 2$. The examples have been
260 performed using first ($k = 1$) and second order ($k = 2$) polynomial spaces. We compare the error convergence
261 rates versus the grid size for several values of ε ranging between 10^{-4} and 1. Each example has been
262 solve applying several mesh refinement steps with h_i, h_{i+1}, \dots , satisfying Assumption 2.3. The numerical
263 convergence rates r have been computed by the ratio $r = \frac{\ln(e_i/e_{i+1})}{\ln(h_i/h_{i+1})}$, $i = 1, 2, \dots$, where the corresponding
264 error is $e_i := \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla u_{h,n})\|_{L^2}$, see (2.33), and $u_{h,n}$ is the final solution computed by the iterative
265 procedure. We mention that, in the test cases with smooth solutions, i.e., $k + 1 \leq l$, the approximation
266 order in (2.33) is expected to be equal to k . The column in the tables below which are related to N indicate
267 the maximum number of the nonlinear iterative steps. All tests have been performed with the C++ library
268 deal.II [35, 36] using an Intel(R) Core(TM) i5-3320M CPU 2.60GHz computer.

4.1. Newton like methods.

269 **4.1. Newton like methods.**
270 *Example 1, ($|\nabla u| \neq 0, \forall x \in \Omega$).* The first numerical example is a simple test case for validating the error
271 estimates given in Theorem 2.7 and demonstrating the performance of the Newton method for several values
272 of ε . We take $\Omega \subset \mathbb{R}^2$ to be the square domain $(0, \frac{\pi}{2})^2$. The boundary data u_D and the function f in (2.1),
273 are chosen so that the exact solution in (2.1) to be $u(x, y) = \sin(x)$. The exact solution is shown in Fig. 2
274 and it holds $\{x \in \Omega : \nabla u = 0\} = \emptyset$.

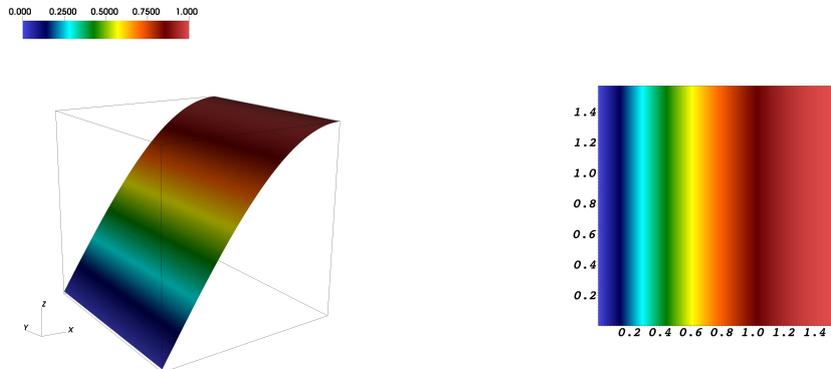


FIG. 2. *Example 1: Display of the numerical solution for the case $p = 1.01$ and $\varepsilon = 1.0e - 4$.*

275 We compute numerically the convergence rates of the error on a sequence of successively refined meshes
276 and also we investigate the performance of the residual-based Newton method for different choices of ε . Here

we consider two different choices for the starting solution u_h^0 for Newton methods, (i) we set $u_h^0 = 0$, (ii) $u_h^0 = Proj(u_{2h}^n)$, where $Proj(u_{2h}^n)$ is the projection of the solution u_{2h}^n , which has been computed on the previous space $V_{D,2h}^{(k)}$, to the current space $V_{D,h}^{(k)}$. Here the index n denotes the last (i.e., converged) solution on the coarser space $V_{D,2h}^{(k)}$. The projection is carried out as nodal interpolation from the coarse mesh on the finer mesh.

In Table 2, we display the results for polynomial degree $k = 1$, for $p = 1.01$ and for $\varepsilon = 10^{-3}$ (right columns) and $\varepsilon = 10^{-4}$ (left columns). For both ε cases, we observe that the rates r related to the first coarse meshes are little less than the expected rates, but progressively as the mesh is refined, the rates tend to the optimal value $r = 1$, which is in agreement with Theorem 2.7. For the case $u_h^0 = 0$, we observe that the Newton method is not mesh independent in the sense that the number of the steps is not stable as the mesh is further refined. On the other hand for the case $u_h^0 = Proj(u_{2h}^n)$, we observe mesh independence and the further number of steps is significantly reduced, see also comments in Remark 3.4. For theoretical results on mesh independence for general nonlinear problems we refer the reader to [40]. As expected, this behavior of the Newton method affects the corresponding CPU times, which are provided on the last row of the Table 2. If not further indicated in all remaining tests, we always use as initial guess $u_h^0 = Proj(u_{2h}^n)$.

		$k = 1$ and $p = 1.01$							
		$\varepsilon = 10^{-4}$				$\varepsilon = 10^{-3}$			
DoFs	$\ F - F_h\ $	r	$N, u_h^0 = 0$	$N, u_h^0 = P(u_{2h}^n)$	$\ F - F_h\ $	r	$N, u_h^0 = 0$	$N, u_h^0 = P(u_{2h}^n)$	
289	3.64207e-02	-	16	16	3.66426e-02	-	13	13	
1089	1.96867e-02	0.88	18	7	2.15775e-02	0.76	16	8	
4225	1.05728e-02	0.9	22	8	1.28714e-02	0.75	23	9	
16641	5.69064e-03	0.9	26	7	6.77354e-03	0.93	28	9	
66049	3.17429e-03	0.85	32	8	3.31612e-03	1.03	32	8	
263169	1.78568e-03	0.83	43	10	1.65509e-03	1.0	34	7	
1050625	9.36320e-04	0.94	47	8	8.54432e-04	0.96	36	6	
CPU	-	-	1.33e+04s	2.13e+03s	-	-	9.87e+03s	1.80e+03s	

TABLE 2

Example 1: The results for the different values of ε .

We repeat the previous computations choosing $\varepsilon = 10^{-2}$ and $\varepsilon = 1$. The results are shown in Table 3. We can clearly see that the rates r are in fully agreement with Theorem 2.7. The Newton method is mesh independent and in particular for the case $\varepsilon = 1$ only few steps (less than 3) are needed. In comparison to the corresponding data in Table 2 the CPU times are reduced in Table 3 as a result of the fewer iterative steps. By the results of the two previous numerical tests, we conclude that the rates r are very close to the expected value for all ε but the Newton steps are increased as the values of ε become smaller.

		$k = 1$ and $p = 1.01$					
		$\varepsilon = 10^{-2}$			$\varepsilon = 1$		
DoFs	$\ F - F_h\ $	r	N	$\ F - F_h\ $	r	N	
289	4.64339e-02	-	12	2.77038e-02	-	12	
1089	2.27756e-02	1.02	8	1.38533e-02	0.99	5	
4225	1.16848e-02	0.96	9	6.92684e-03	1.0	5	
16641	6.05204e-03	0.95	10	3.46344e-03	1.0	3	
66049	3.04030e-03	0.99	8	1.73172e-03	1.0	3	
263169	1.51975e-03	1.0	5	8.65862e-04	1.0	3	
1050625	7.59810e-04	1.0	4	4.32931e-04	1.0	2	
CPU	-	-	8.46e+02s	-	-	5.07e+02s	

TABLE 3

Example 1: The results for the different values of ε .

Next, we investigate the convergence rate of the error and the performance of the residual-based Newton method for $p = 1.1$ and for three different choices of ε . The results are shown in Table 4. Here, we can see that for each of the three values of ε , the error $\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla u_h^n)\|_{L^2}$ converges to zero with rate r close to one, which is in agreement with Theorem 2.7. The Newton performance appears to be mesh independent. The required total steps are increased as the value of ε is decreased.

As a last computation in this example, we solve the problem using $k = 2$ polynomials. The parameter p takes the limiting values $p = 1.01$ and $p = 1.1$. Again, as in the above tests, we investigate the behavior of

$k = 1$ and $p = 1.1$									
$\varepsilon = 10^{-4}$									
$\varepsilon = 10^{-3}$									
$\varepsilon = 10^{-1}$									
DoFs	$\ F - F_h\ $	r	N	$\ F - F_h\ $	r	N	$\ F - F_h\ $	r	N
289	3.61674e-02	-	11	3.62331e-02	-	10	3.90390e-02	-	9
1089	1.92142e-02	0.91	5	1.93753e-02	0.9	5	1.95622e-02	1.0	4
4225	1.00983e-02	0.93	5	1.04294e-02	0.9	6	9.78098e-03	1.0	4
16641	5.26816e-03	0.94	5	5.64053e-03	0.88	5	4.89046e-03	1.0	3
66049	2.73798e-03	0.95	5	2.89079e-03	0.96	6	2.44523e-03	1.0	3
263169	1.42804e-03	0.94	6	1.46773e-03	0.98	5	1.22261e-03	1.0	3
1050625	7.44909e-04	0.94	8	7.51022e-04	0.97	5	6.11307e-04	1.0	2
CPU	-	-	1.05e+03s	-	-	7.27e+02s	-	-	3.47e+02s

TABLE 4

Example 1: The results for the different values of ε .

305 the convergence of the error as well the behavior of the iterative solver for several values of ε . The results are
 306 displayed in Table 5 and in Table 6. Here, we observe for $\varepsilon = 10^{-4}$ that the error converges with the expected
 307 rate r for both p -cases. We repeat the computations setting $\varepsilon = 0.1$. Also, in this case the convergence rates
 308 of the error tend to obtain the optimal value $r = 2$ as the mesh is refined. For both p -cases, the rates are
 309 optimal for all mesh refinement levels and the iterative solver needs less than two steps for obtaining the
 310 solution. Again, this test illustrates the affection of the parameter ε to the performance of the iterative solver.

$k = 2$ and $p = 1.01$							
$\varepsilon = 10^{-4}$ $p = 1.01$							
$\varepsilon = 10^{-1}$ $p = 1.01$							
DoFs	$\ F - F_h\ $	r	N	$\ F - F_h\ $	r	N	
256	2.221e-04	-	20	2.316e-04	-	11	
1024	5.509e-05	2.01	4	5.787e-05	2.00	2	
4096	1.382e-05	1.99	4	1.446e-05	2.00	2	
16384	3.457e-06	2.00	5	3.616e-06	2.00	1	
65536	8.645e-07	2.00	5	9.041e-07	2.00	1	
262144	2.161e-07	2.00	5	2.260e-07	2.00	1	
CPU	-	-	6.999e+02s	-	-	1.658e+02s	

TABLE 5

Example 1: The results for $k = 2$ for different choices of ε .

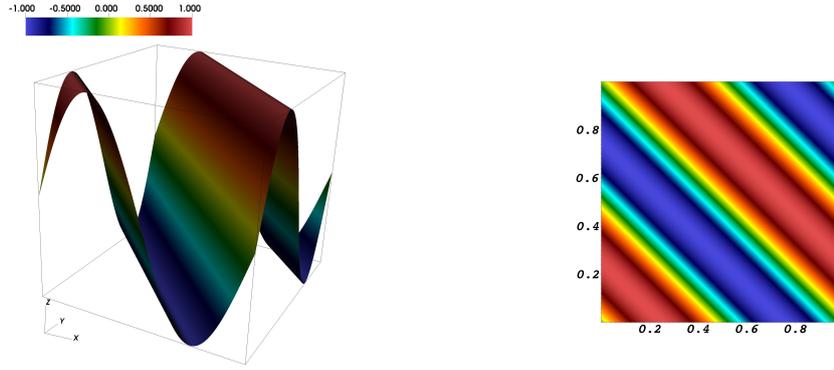
$k = 2$ and $p = 1.1$						
$\varepsilon = 10^{-4}$ $p = 1.1$			$\varepsilon = 10^{-1}$ $p = 1.1$			
$\ F - F_h\ $	r	N	$\ F - F_h\ $	r	N	
2.394e-04	-	14	2.476e-04	-	14	
5.939e-05	2.01	4	6.190e-05	2.00	2	
1.490e-05	1.99	4	1.547e-05	2.00	2	
3.726e-06	2.00	4	3.868e-06	2.00	1	
9.318e-07	2.00	4	9.672e-07	2.00	1	
2.329e-07	2.00	4	2.418e-07	2.00	1	
CPU	-	5.719e+02s	-	-	1.671e+02s	

TABLE 6

Example 1: The results for $k = 2$ for different choices of ε .

311 *Example 2, (existence of $x_0 \in \Omega : |\nabla u(x_0)| = 0$).*

312 In the second example, we want to investigate the convergence rates of the error and the corresponding
 313 behavior of the residual-based Newton method in case of the presence of critical points, i.e., $\{x_0 \in \Omega : |\nabla u(x_0)| = 0\}$. In this situation the iterative matrix \mathbf{B}' in (3.7a) tends to become singular or ill-conditioned
 314 as $\varepsilon \rightarrow 0$. To this end, the domain is $\Omega = (0, 1)^2$ and the data u_D and f are selected so that the exact solution
 315 to be given by $u(x, y) = \sin(2\pi(x + y))$, which is plotted in Fig. 3. Note that the solution contains critical
 316 points along the lines $y = \frac{1}{4} - x$ and $y = \frac{3}{4} - x$. We solved the problem using $k = 1$ polynomial space and
 317 initial guess $u_h^0 = Proj(u_{2h}^2)$. We compute the convergence rates on a sequence of seven successively finer
 318 meshes for several values of p and ε ranging from 10^{-4} to 1. In Table 7, we show the convergence rates r
 319

FIG. 3. Example 2: Display of the solution for the case $p = 1.01$ and $\varepsilon = 1.0$.

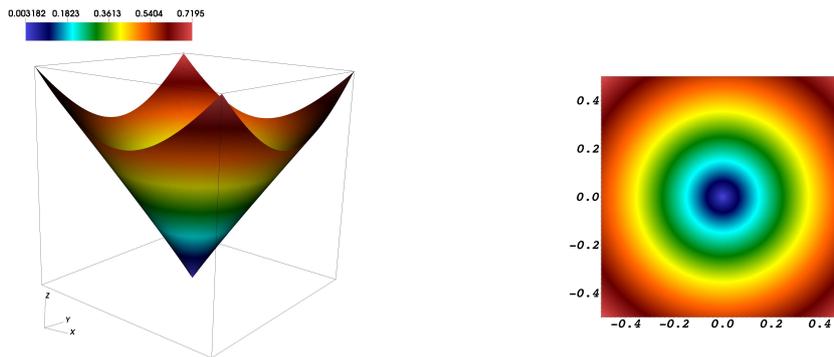
320 and the Newton iterations for each selected p and each ε . The reported rate r is the minimum rate computed
 321 on the last three finer meshes. The reported number of Newton iterations N , is the maximum number of
 322 the Newton iterations that have been used on the three finer meshes. We can clearly see, that for $p < 2$
 323 the performance of the Newton solver is very sensitive to the choice of ε . In particular the value of ε must
 324 be close to one in order to perform successfully the Newton iterative procedure. An exception is the case
 325 of $p = 1.8$ where the problem has been solved for all values of ε using a relatively small number of iterative
 326 steps N . On the other hand for the tests with $p > 2$, the problem has been successfully solved for all the
 327 selected values of ε . We can see that the rates have the expected values and the Newton method appears to
 328 have a mesh independent behavior. For the case $p = 11$, we observe an increase in the Newton steps without
 having a deterioration of the convergence rates.

ε	$p = 1.01$		$p = 1.1$		$p = 1.3$		$p = 1.5$		$p = 1.8$		$p = 2.25$		$p = 3$		$p = 4.3$		$p = 11$	
	r	N	r	N	r	N	r	N	r	N	r	N	r	N	r	N	r	N
10^{-4}							1.3	8	1.08	9	1	6	1	7	1	10	1	20
10^{-3}					1.8	11	1.27	8	1.1	9	1	6	1	7	1	10	1	20
10^{-2}					2.4	12	1.37	8	1.1	6	1	5	1	7	1	9	1	19
10^{-1}			1	7	1	9	1	6	1	5	1	5	1	7	1	6	1	18
1	1	4	1	4	1	5	1	5	1	5	1	5	1	4	1	5	1	9

TABLE 7

Example 2: The rates r and the Newton iteration steps with respect to ε for several values of p .

329

FIG. 4. Example 3: Display of the numerical solution for the case $p = 1.5$ and $\varepsilon = 1.0$ and $\gamma = 0.8$ and $k = 1$.

330 Example 3: (low-regularity solution, $u \in W^{l,p}(\Omega)$, $l = 2$ and $l = 3$). This example consists of a problem
 331 with low regularity solution, i.e., $u(x, y) = (x^2 + y^2)^{\frac{\gamma}{2}}$, for several values for the parameter γ , see also [17]. The
 332 computational domain is the rectangle $\Omega = (-0.5, 0.5)^2$. The source function f and u_D are manufactured by
 333 the exact solution. We point out that the regularity of the solution is specified by the value of the parameter
 334 γ and in this example, we investigate the ability of the whole method (FE discretization and Newton solver)
 335 to approximate low regularity solutions of (2.1) with the expected accuracy. Thus, this example seems to
 336 be quite interesting, since we can predefine the regularity of the proposed exact solution to be close to the
 337 required regularity of the solution, which has been assumed in the error analysis, see Assumptions 2.1, 2.3

$u \in W^{l,p}(\Omega)$ with $u = x ^\gamma$, $\varepsilon = 1$			
space \mathbb{Q}_k	$\gamma = 0.72, l = 2, p = 1.5$	$\gamma = 0.92, l = 2, p = 1.8$	$\gamma = 1.4, l = 2, p = 3$
	expected rates r		
$k = 1$ and $k = 2$	1	1	1
	$\gamma = 1.72, l = 3, p = 1.5$	$\gamma = 1.92, l = 3, p = 1.8$	$\gamma = 2.4, l = 3, p = 3$
	expected rates r		
$k = 1$	1	1	1
$k = 2$	2	2	2

TABLE 8

Example 3: the expected convergence rates r for the several values of the parameters γ and p .

338 and Remark 2.2. In Table 8, for several values of the parameters γ and p , we display the space $W^{l,p}$ where
 339 the solution u belongs to. Also, the expected convergence rates r are mentioned. For validation, we have
 340 computed the convergence rates in all cases setting $\varepsilon = 1$, see Table 8. Note that in this case we have that
 341 $\frac{1}{\tau} \in L^q(\Omega)$, $q = \frac{p}{p-1}$, see Assumption 2.3 and Remark 2.2. We solve the problem using $k = 1$ and $k = 2$
 342 polynomial spaces and utilizing both Newton iterative solvers, residual based-line, see Algorithm 3.2 and
 343 error-oriented, see Algorithm 3.3.

$u \in W^{l,p}(\Omega)$ with $u = x ^\gamma$, $\varepsilon = 1$																		
$\gamma = 0.72, l = 2, p = 1.5$						$\gamma = 0.92, l = 2, p = 1.8$						$\gamma = 1.4, l = 2, p = 3$						
$k = 1$			$k = 2$			$k = 1$			$k = 2$			$k = 1$			$k = 2$			
$\frac{h}{2^s}$	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}
$s=0$	-	7	4	-	7	5	-	5	3	-	5	4	-	6	4	-	7	6
$s=1$	0.69	4	2	0.81	4	2	0.84	3	2	0.99	3	2	0.98	4	2	1.43	3	2
$s=2$	0.73	4	2	0.82	4	2	0.87	3	2	0.99	3	2	0.99	4	2	1.42	3	1
$s=3$	0.77	4	2	0.82	4	2	0.89	3	2	0.99	3	2	0.99	3	2	1.42	3	1
$s=4$	0.78	4	2	0.88	4	2	0.90	3	2	0.99	3	2	0.99	3	2	1.41	3	1
CPU	-	27s	25s	-	130s	120s	-	21s	25s	-	100s	120s	-	22s	19s	-	110s	87s
$\gamma = 1.72, l = 3, p = 1.5$						$\gamma = 1.92, l = 3, p = 1.8$						$\gamma = 2.4, l = 3, p = 3$						
$k = 1$			$k = 2$			$k = 1$			$k = 2$			$k = 1$			$k = 2$			
$\frac{h}{2^s}$	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}	r	N_{RB}	N_{EO}
$s=0$	-	5	4	-	5	4	-	4	3	-	5	3	-	5	4	-	7	5
$s=1$	1	3	2	1.74	3	1	1	3	2	1.85	2	1	1	4	2	2	2	1
$s=2$	1	3	2	1.76	2	1	1	3	2	1.87	2	1	1	3	2	2	2	1
$s=3$	1	3	2	1.77	2	1	1	3	1	1.90	2	1	1	3	2	2	2	1
$s=4$	1	3	1	1.78	2	1	1	2	1	1.91	2	1	1	3	2	2	2	1
CPU	-	17s	19s	-	76s	46s	-	17s	20s	-	51s	45s	-	23s	21s	-	49s	54s

TABLE 9

Example 3: the numerical convergence rates r and the nonlinear iterative steps N for the several values of γ and p .

344 In Table 9, we present the convergence rates r , as well the maximum number of the nonlinear iterative
 345 steps. In particular, the column of N_{RB} corresponds to the steps of the residual-based method, and the
 346 column of N_{EO} corresponds to the error-oriented method. In Table 9, we report only one column with
 347 convergence rates r , because both Newton iterative methods yield the same result. We observe that the
 348 rates related to $\gamma = 0.72, l = 2, p = 1.5$ are slightly below the expected value $r = 1$, see Table 8, but
 349 are progressively increased as we use finer meshes. For the $p = 1.8$ test case, the rates r corresponding to
 350 $k = 1$ and $k = 2$ solutions are very close to the expected values. For the last $p = 3$ test case, the rates r
 351 corresponding to $k = 1$ are in good agreement with Table 8, but the rates r corresponding to $k = 2$ are little
 352 higher. Here, we would like to note that in several numerical experiments that we have performed using
 353 $p > 2$, the numerical rates have been found to be greater than those predicted by the theory, e.g., see Tables
 354 9, 10 and 12 below. Similar rate behavior has been observed in the corresponding tests presented in [17].³

355 Concerning the iterative methods, we observe that both have mesh independent and p independent
 356 behavior. However, the performance of the error-oriented method needs less steps. However, this does not
 357 entail into a remarkable reduction of the corresponding CPU times, see the corresponding row in Table 8.
 358 The reason is that the error-oriented method is more expensive than the residual-based method, because we
 359 solve the linearized system two times per step; see Algorithm 3.3. Continuing with the results for $l = 3$ that

³For a further investigation, we have performed more computations using different values for γ and p . Again, in some cases, we found little higher rates than the rates expected by the theory.

are displayed in last rows in Table 9, we can see that for every p -case the rates r related to $k = 1$ solutions are optimal. For $p = 1.5$, the rates related to $k = 2$ spaces are little below than the expected value. For the test case $p = 1.8$ the rates are close to the expected value and finally for the case $p = 3$ the rates become optimal. In conclusion, the results presented in Table 9 show that the approximation error either converges to the predicted rate, or in some cases related to $\gamma = 0.72$ tends to convergence. Both Newton iterative schemes can be utilized for solving the final nonlinear algebraic system.

4.2. Augmented Lagrangian methods. In this paragraph we present two numerical examples that have been solved using the two ALG1 variants, see Table 1 in Section 3.4. Concretely, we have solved the same problems as in Example 2 and in Example 3 in order to compare the residual-based Newton method with the two ALG1 approaches.

Example 4: (existence of $x_0 \in \Omega : |\nabla u(x_0)| = 0$). We set $\Omega = (0, 1)^2$ and the boundary data u_D and f are prescribed by the exact solution $u(x, y) = \sin(2\pi(x+y))$, see also Example 2 above. We consider six levels of mesh refinement and four values for $p \in \{1.01, 1.5, 3, 4.3\}$. The numerical tests has been performed using $k = 1$ polynomial space and $\varepsilon = 1$. We mention that we also performed computations using the remaining values of ε in Table 7. The derived results were similar to that which are described in Table 7 and thus we report only the case $\varepsilon = 1$. In Table 10, we show the numerical convergence rates and the corresponding iterative steps that we have found for the sALG1 method. In Table 11, we show the corresponding data of mALG1 method. In both tables the column associated with N indicates the number of nonlinear (Newtons) iterations necessary to achieve the convergence criteria. In general, we observe that both methods sALG1 and mALG1 converge quite fast, as few as 2 iterations on fine meshes. In some cases, the proposed mALG1 method requires less steps to reach the stopping criterion, see for example the $p = 1.5$ case. The number N of iterations remains stable and very low for the last mesh levels, thus we can say that both ALG1 methods exhibit a mesh independent behavior. Also we observe that the convergence rates of sALG1 appear to be little higher than the theory predicts. On the other hand the convergence rates r corresponding to mALG1 method are very close to the rates predicted by the theory.

-	sALG1, $\varepsilon = 1$ and $u_h^0 = P(u_{2h}^n)$ and $k = 1$							
grid size	$p = 1.01$		$p = 1.5$		$p = 3$		$p = 4.3$	
$\frac{h}{2^s}$	r	N	r	N	r	N	r	N
s=0	-	4	-	4	-	4	-	3
s=1	1.1	3	1.1	4	1.2	4	1.4	3
s=2	1.0	3	1.3	3	1.1	3	1.4	2
s=3	1.3	2	1.2	3	1.1	3	1.3	2
s=4	1.2	2	1.1	2	1.0	3	1.2	1

TABLE 10

Example 4: sALG1 method. The convergence rates and the iterative steps for several values of p .

-	mALG1, $\varepsilon = 1$ and $u_h^0 = P(u_{2h}^n)$ and $k = 1$							
grid size	$p = 1.01$		$p = 1.5$		$p = 3$		$p = 4.3$	
$\frac{h}{2^s}$	r	N	r	N	r	N	r	N
s=0	-	4	-	4	-	7	-	12
s=1	0.86	3	0.98	2	1.1	3	1.2	3
s=2	1.0	3	1.0	2	1.0	2	1.2	2
s=3	1.0	2	1.0	2	1.0	2	1.0	2
s=4	1.0	2	1.0	2	1.0	2	1.0	1

TABLE 11

Example 4: mALG1 method. The convergence rates and the iterative steps for several values of p .

Example 5: (low regularity solution). We consider $\Omega = (-0.5, 0.5)^2$ and the exact solution $u(x, y) = (x^2 + y^2)^{\frac{\gamma}{2}}$, see also Example 3, Table 8 and Table 9. We have solved the problem using linear elements $k = 1$ giving to γ and p the values of the second row in Table 8. We compare the two different ALG1 methods by choosing three different test cases associated with three different values for p , i.e., $p = \{1.5, 1.8, 3\}$. Table 12 corresponds to sALG1 method and it displays the convergence rates of the error and the iterative steps N for every mesh level. The results related to mALG1 method are displayed in Table 13. The initial guess is defined to be the projection of the previous coarse mesh solution. We begin with the results of sALG1 displayed in Table 12 and are associated with $l = 2$ regularity. We observe that for $p = 1.5$ the rates, for both $k = 1$ and $k = 2$ spaces, are slightly less than the expected optimal $r = 1$ rate, but they progressively increase towards the optimal value. For the case of $p = 1.8$ the rates r of both numerical solutions $k = 1$ and $k = 2$ are very close to the optimal value $r = 1$ which is determined by the regularity of u . On the other hand the rates for $p = 3$ case are little higher on the first meshes but are progressively reduced. In general, the convergence rates in Table 12, which are related to $l = 2$ regularity, are closer to the rates presented in Table 8, in comparison to the rates that we show in Table 9. Looking at Table 12 the column with the number of the nonlinear iterative steps, we observe that in general the sALG1 method has a mesh independent behavior. The iterative steps corresponding to $k = 1$ solutions are considerably more than the steps corresponding to $k = 2$ solutions. This does not have an impact on the total CPU time, since we observe that the CPU time of $k = 2$ computations is much higher than the CPU time of $k = 1$ computations, see the corresponding row in Table 12.

Also, Table 12 shows the results for the $l = 3$ tests (see last rows). We observe that the rates related to $k = 1$ for all p -cases are little higher than the $r = 1$ (expected value). On the other hand the rates r related

sALG1, $u \in W^{l,p}(\Omega)$ with $u = x ^\gamma$, $\varepsilon = 1$												
$\gamma = 0.72, l = 2, p = 1.5$				$\gamma = 0.92, l = 2, p = 1.8$				$\gamma = 1.4, l = 2, p = 3$				
$\frac{h}{2^s}$	$k = 1$	N	$k = 2$	N	$k = 1$	N	$k = 2$	N	$k = 1$	N	$k = 2$	N
-	Rates r and steps N for sALG1											
$s = 0$	-	22	-	14	-	17	-	11	-	24	-	16
$s = 1$	0.85	18	0.85	10	1.02	16	0.95	11	1.50	16	1.45	11
$s = 2$	0.79	17	0.86	9	0.94	15	0.95	8	1.30	12	1.43	10
$s = 3$	0.79	16	0.86	8	0.93	13	0.96	7	1.22	12	1.42	10
$s = 4$	0.80	14	0.87	7	0.94	11	0.95	6	1.17	9	1.41	7
CPU	-	94s	-	394s	-	120s	-	306s	-	101s	-	440s
$\gamma = 1.72, l = 3, p = 1.5$												
$\gamma = 1.72, l = 3, p = 1.5$				$\gamma = 1.92, l = 3, p = 1.8$				$\gamma = 2.4, l = 3, p = 3$				
$\frac{h}{2^s}$	$k = 1$	N	$k = 2$	N	$k = 1$	N	$k = 2$	N	$k = 1$	N	$k = 2$	N
-	Rates r and steps N for sALG1											
$s = 0$	-	20	-	10	-	18	-	8	-	24	-	14
$s = 1$	1.35	16	1.76	6	1.35	14	1.90	6	1.17	18	2.1	12
$s = 2$	1.46	11	1.77	6	1.39	12	1.91	6	1.28	12	2.07	10
$s = 3$	1.39	8	1.78	6	1.35	8	1.92	6	1.30	11	2.04	10
$s = 4$	1.30	7	1.79	6	1.29	7	1.90	6	1.23	10	2.02	9
CPU	-	64s	-	344s	-	61s	-	389s	-	117s	-	471s

TABLE 12

Example 5: sALG1 for solving problems with low regularity solution. The numerical convergence rates and the iterative steps for the several values of γ and p .

mALG1, $u \in W^{l,p}(\Omega)$ with $u = x ^\gamma$, $\varepsilon = 1$												
$\gamma = 0.72, l = 2, p = 1.5$				$\gamma = 0.92, l = 2, p = 1.8$				$\gamma = 1.4, l = 2, p = 3$				
$\frac{h}{2^s}$	$k = 1$	N	$k = 2$	N	$k = 1$	N	$k = 2$	N	$k = 1$	N	$k = 2$	N
-	Rates r and steps N for mALG1											
$s = 0$	-	9	-	9	-	6	-	6	-	9	-	10
$s = 1$	0.73	6	0.85	9	0.77	8	0.91	5	0.97	5	1.32	10
$s = 2$	0.85	6	0.90	5	0.90	8	0.95	7	1.02	7	1.40	6
$s = 3$	0.86	5	0.90	5	0.91	7	0.95	6	1.02	5	1.41	4
$s = 4$	0.87	5	0.90	4	0.92	7	0.95	6	1.01	4	1.41	3
CPU	-	133s	-	1174s	-	188s	-	1597s	-	98s	-	883s
$\gamma = 1.72, l = 3, p = 1.5$												
$\gamma = 1.72, l = 3, p = 1.5$				$\gamma = 1.92, l = 3, p = 1.8$				$\gamma = 2.4, l = 3, p = 3$				
$\frac{h}{2^s}$	$k = 1$	N	$k = 2$	N	$k = 1$	N	$k = 2$	N	$k = 1$	N	$k = 2$	N
-	Rates r and steps N for mALG1											
$s = 0$	-	8	-	8	-	5	-	6	-	9	-	8
$s = 1$	1.00	4	1.74	3	1.02	4	1.81	5	0.97	7	2.02	8
$s = 2$	1.00	3	1.76	2	1.01	3	1.87	2	1.02	5	2.07	3
$s = 3$	1.00	3	1.78	2	1.01	2	1.90	2	1.01	4	2.00	2
$s = 4$	1.00	2	1.79	1	1.00	2	1.91	1	1.01	4	2.00	1
CPU	-	76s	-	356s	-	61s	-	335s	-	76s	-	338s

TABLE 13

Example 5: mALG1 for solving problems with low regularity solution. The numerical convergence rates and the iterative steps for the several values of γ and p .

406 to $p = 1.5$, $k = 2$ solutions are little lower than expected. For the two other p -cases, the convergence rates of
 407 $k = 2$ numerical solutions are close to the expected $r = 2$ rate. Concerning the variations of the numbers N of
 408 the iterative steps, we observe similar behavior as in the previous $l = 2$ test case. More precisely, although the
 409 steps of $k = 1$ numerical solutions are again quite higher than the steps of $k = 2$ solutions, the corresponding
 410 CPU time of $k = 1$ solutions is smaller. This occurs due to the fact that ALG1 methods introduce two new
 411 variables, see Table 1. Thus, when we use $k = 2$ spaces, the resulting algebraic linear system at each Newton
 412 step, has a high number of unknowns, and therefore its solution requires much more computing time. Here,
 413 advanced techniques, e.g. efficient preconditioners such as for instance multigrid techniques must be applied.
 414 This is a topic that we will discuss in a forthcoming paper; see also Remark 3.6.

We continue with the results of mALG1 method for $l = 2$ regularity case, see first rows in Table 13. For the $p = 1.5$ test case, we observe that the rates r corresponding to $k = 1$ solutions are little lower than the optimal value but moving to finer meshes the rates are approaching the optimal value. For the other two p -cases the rates of $k = 1$ solutions are much closer to the expected $r = 1$ value. Concerning the rates of $k = 2$ solutions, we can see that are close to $r = 1$ which is not surprising since $u \in W^{l=2,p}$. We have to remark that the rates obtained for $p = 3$ test case are little bit more higher than expected. Finally, we mention that, for all p -cases, the required number of iterative steps is not increased as we use finer grids, even for $k = 2$ spaces. This indicates that the proposed mALG1 method follows a mesh-independent behavior. Here, we also observe that the number of the iterative steps for both cases $k = 1$ and $k = 2$ are very close, and such a behavior did not appear in sALG1 method. Concerning the CPU times displayed in the last row, similar comments as in previous tests of sALG1 method can be expressed.

The last rows in Table 13 show the results related to $l = 3$ regularity test case. We observe that for each of the three p -tests the error related to $k = 1$ spaces convergences at the optimal rate $r = 1$, as predicted by Theorem 2.33, see also Table 8. As in the previous examples, the error convergence rates related to $k = 2$ numerical solutions is little below the expected value for the case of $p = 1.5$, but the rates are in agreement with the predicted rates for the rest p -cases. In general, we can say that the rates behave in a similar manner as in Example 3, where the residual-based method applied, see last lines in Table 8. Also, Table 13 presents the actual number of the nonlinear iterative steps at every mesh. We see that the iterative steps of both $k = 1$ and $k = 2$ solutions are very close. For each choice of p , the solution is obtained in very few iterative steps even on the last mesh levels. Also we note that the number of the iterative steps is not increased as we increase the value of p . We did not observe the same in Table 12 which shows the results for the sALG1 method. This fact indicates the robustness of the new proposed mALG1 method with respect the choice of p . Finally, in the last row in Table 13, we can see the corresponding CPU times. As in the $l = 2$ case and as in sALG1 method, the CPU times of $k = 2$ solutions are quite larger than the CPU times of $k = 1$ solutions. This is not the case for the two Newton methods, see last row in Table 9.

As a final comment we can say that the two ALG1 methods appear to be quite appropriate for solving the nonlinear system in (3.4). This seems to be a result of the particular decomposition of the problem and the treatment of the nonlinear terms. On the other hand, this ALG1 decomposition introduces additional unknowns, which results in the creation of a large system.

5. Conclusions. In this article, we applied finite element methods for the numerical solution of p -power-law diffusion problems. Under some regularity assumptions for the exact solution we derived a priori error estimates. The theoretical error estimates were demonstrated by numerical examples using several values for the parameters of the problem. Next, we presented nonlinear iterative solvers. Initially, we analyzed two Newton-like iterative methods, a residual-based and an error-oriented technique. We discussed in detail the condition properties of the Newton iterative matrix. Even though the error-oriented procedure finds the solution in less iterations than the residual-based technique, the latter is easier to be materialized. However, in the error-oriented method the linear system is solved twice, which finally makes the two methods to be appear equivalent in terms of CPU time. In addition, we applied two augmented Lagrangian methodologies for solving the original p -type elliptic problem. The resulting iterative methods can be characterized by the simplicity of the algorithm and are able to provide a nice handling of the nonlinear terms via the decomposition of the original problem. Nevertheless, the introduction of the additional unknowns that accompanies the ALG1 framework, yields a large system at each Newton step. For an efficient solution of this system, it seems that we need to apply more advanced techniques, which we envisage to develop in future work.

Acknowledgments. The authors wish to thank Prof. J. W. Barrett for providing Ref. [22]. The first author was supported by the Austrian Science Fund (FWF) under the grant NFN S117-03.

REFERENCES

- [1] Glowinski R. and Rappaz J. Approximation of a nonlinear elliptic problem arising in a non-Newtonian fluid flow model in glaciology. *Math. Model. Numer. Anal.*, 37:175–186, 2003.
- [2] Lee S., Mikelic A., Wheeler S., and Wick T. Phase-field modeling of proppant-filled fracture in a poroelastic medium. accepted for publication in *Computer Methods in Applied Mechanics and Engineering*, doi: 10.1016/j.cma.2016.02.008, Feb 2016.
- [3] Janela J., Moura A., and Sequeira A. A 3d non-newtonian fluid-structure interaction model for blood flow in arteries. *Journal of Computational and Applied Mathematics*, 234(9):2783 – 2791, 2010.
- [4] Robertson A.M., Sequeira A., and Owens R.G. *Rheological models for blood*, volume 1 of *Cardiovascular Mathematics. Modeling and Simulation of the Cardiovascular System*, pages 211–241. Springer, Berlin Heidelberg.
- [5] Glowinski R. and Marrocco A. Sur l’approximation par elements finite d’ordre un, et la resolution, par penalization-dualite, d’une classe de Dirichlet non-lineaires. *RAIRO R-2*, pages 41–76, 1975.
- [6] Ciarlet P. G. *The Finite Element Method for Elliptic Problems*. Studies in Mathematics and its Applications. North Holland Publishing Company, 1978.

- 474 [7] Chow S. S. Finite Element Estimates for Non-Linear Elliptic Equations of Monotone Type. *Numer. Math.*, 54(4):373–393,
475 1989.
- 476 [8] Ainsworth M. and Kay D. The approximation theory for the p-version finite element method and application to non-linear
477 elliptic PDEs. *Numerische Mathematik*, 82(3):351–388, 1999.
- 478 [9] Ainsworth M. and Kay D. Approximation theory for the hp-version finite element method and application to the non-linear
479 Laplacian. *Appl. Numer. Math.*, 34:329–344, 2000.
- 480 [10] Barrett J. W. and Liu W. B. Finite element approximation of the p-Laplacian. *Math. Comp.*, 61(204):523–537, 1993.
- 481 [11] Barrett J. W. and Liu W. B. Finite element approximation of some degenerate monotone quasilinear elliptic systems.
482 *SIAM J. Numer. Anal.*, 33(1):88–106, 1996.
- 483 [12] Ebmeyer C. and Liu W. B. Quasi-norm interpolation error estimates for the piecewise linear finite element approximation
484 of p-laplacian problems. *Numer. Math*, 100:233–258, 2005.
- 485 [13] Liu W. B. and Yan N. On quasi-norm interpolation error estimation and a posteriori error estimates for p-Laplacian.
486 *SIAM J. Numer. Anal.*, 40(5):1870–1895, 2002.
- 487 [14] Diening L. and Růžička M. Interpolation operators in Orlicz-Sobolev spaces. *Numer. Math.*, 107:107–129, 2007.
- 488 [15] Burman E. and Ern A. Discontinuous Galerkin approximations with discrete variational principle for the nonlinear
489 Laplacian. *C. R. Acad. Sci. Paris, Ser I(346)*:1013–1016, 2008.
- 490 [16] Diening L., Kröner D., Růžička M., and Touloupoulos I. A local discontinuous Galerkin approximation for systems with
491 p-structure. *IMA J Numer Anal*, 34(4):1447–1488, 2014.
- 492 [17] Kröner D., Růžička M., and Touloupoulos I. Numerical solutions of systems with (p, δ) -structure using local discontinuous
493 Galerkin finite element methods. *Int. J. Numer. Methods Fluids*, 76:855–874, 2014.
- 494 [18] Touloupoulos I. An interior penalty discontinuous Galerkin finite element method for quasilinear parabolic problems. *Finite
495 Elements in Analysis and Design*, 95:42 – 50, 2015.
- 496 [19] Kröner D., Růžička M., and Touloupoulos I. Local discontinuous galerkin numerical solutions of non - Newtonian incom-
497 pressible flows modeled by p - Navier Stokes equations. *Journal of Computational Physics*, 270:182 – 202, 2014.
- 498 [20] Liu W.B. and Barrett J. W. Higher-order regularity for the solutions of some degenerate quasilinear elliptic equations in
499 the plane. *SIAM J. Math. Anal.*, 24(6):1522–1536, 1993.
- 500 [21] Liu W.B. and Barrett J. W. A remark on the regularity of the solutions of the p-Laplacian and its application to their
501 finite element approximation. *Journal of Mathematical Analysis and Applications*, 178(2):470 – 487, 1993.
- 502 [22] Liu W.B. and Barrett J. W. A further remark on the regularity of the solutions of the p-Laplacian and its applications to
503 their finite element approximation. *Nonlinear Analysis: Theory, Methods and Applications*, 21(5):379 – 387, 1993.
- 504 [23] Dobrev V. A., Kolev T. V., and Rieben R. N. High-order curvilinear finite element methods for Lagrangian hydrodynamics.
505 *SIAM J. Sci. Comput.*, 34(5):B606 – B641, 2012.
- 506 [24] Cotrell J. A., Hughes T. J. R., and Bazilevs Y. *Isogeometric Analysis, Toward Integration of CAD and FEA*. John Wiley
507 and Sons, United Kingdom, 2009.
- 508 [25] Langer U., Mantzaflaris A., Moore S. E., and Touloupoulos I. Mesh grading in Isogeometric Analysis. *Computers and
509 Mathematics with Applications*, 70(7):1685–1700, 2015.
- 510 [26] Richter T. and Wick T. Variational localizations of the dual weighted residual estimator. *Journal of Computational and
511 Applied Mathematics*, 279(0):192 – 208, 2015.
- 512 [27] Apel. T., Saendig A.-M., and Whiteman J. R. Graded mesh refinement and error estimates for finite element solutions of
513 elliptic boundary value problems in non-smooth domains. *Math. Method. Appl. Sci.*, 19(1):63–85, 1996.
- 514 [28] Brenner S.C. and Scott L.R. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied
515 Mathematics*. Springer-Verlag New York, New York, third edition, 1994.
- 516 [29] Deuffhard P. *Newton Methods for Nonlinear Problems, Affine Invariance and Adaptive Algorithms*, volume 35 of *Springer
517 Series in Computational Mathematics*. Springer Berlin Heidelberg, 2011.
- 518 [30] Ern A. and Vohralík M. Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs.
519 *SIAM J. Sci. Comput.*, 35(4):A1761– A1791, 2013.
- 520 [31] Hirn A. Finite element approximation of singular power-law systems. *Math. Comp*, 82(283):1247–1268, 2013.
- 521 [32] Glowinski R. *Numerical Methods for Nonlinear Variational Problems (Scientific Computation)*. Scientific Computation.
522 Springer-Verlag, Berlin-Heidelberg, 2 edition, 2008.
- 523 [33] Adams R. A. and Fournier J. J. F. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. ACADEMIC PRESS-
524 imprint Elsevier Science, second edition, 2003.
- 525 [34] Diening L. and Ettwein F. Fractional Estimates for Non-differentiable Elliptic Systems with General Growth. *Forum
526 Mathematicum*, 20(3):523–556, 2008.
- 527 [35] Bangerth W., Heister T., and Kanschhat G. *Differential Equations Analysis Library*, 2012.
- 528 [36] Bangerth W., Hartmann R., and Kanschhat G. deal.II – a general purpose object oriented finite element library. *ACM
529 Trans. Math. Softw.*, 33(4):24/1–24/27, 2007.
- 530 [37] Faragó I. and Karatson J. *Numerical Solution of Nonlinear Elliptic Problems Via Preconditioning operators: Theory and
531 Applications*, volume 11. Nova Science Publishers, Inc, New York, 2002.
- 532 [38] Glowinski R. and Tallec P. L. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, volume 9.
533 Society for Industrial and Applied Mathematics, Philadelphia, USA, 1989.
- 534 [39] T. A. Davis and I. S. Duff. An unsymmetric-pattern multifrontal method for sparse LU factorization. *SIAM J. Matrix
535 Anal. Appl.*, 18(1):140–158, 1997.
- 536 [40] Weiser M., Schiela A., and Deuffhard P. Asymptotic mesh independence of newton’s method revisited. *SIAM Journal on
537 Numerical Analysis*, 42(5):1830–1845, 2005.