A Convergent Adaptive Uzawa Finite Element Method for the Nonlinear Stokes Problem

Dissertation zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Augsburg

vorgelegt von Christian Kreuzer

April 2008

Erster Gutachter:	Prof. Dr. K. G. Siebert, Augsburg, Deutschland
Zweiter Gutachter:	Prof. Dr. R. H. Nochetto, College Park, USA
Dritter Gutachter:	Prof. Dr. A. Veeser, Mailand, Italien

Mündliche Prüfung:	23. Juli, 2008

Danksagung

Obgleich ich diese Arbeit selbst verfasst und mich keiner fremden Hilfe bedient habe, gibt es doch einige Menschen, die zur Entstehung der vorliegenden Seiten beigetragen haben.

In erster Linie möchte ich mich bei meinem Doktorvater Kunibert G. Siebert bedanken, der mir einerseits viele Freiräume gelassen hat, andererseits bei Problemen immer zur Stelle war. Auch bedanken möchte ich mich bei ihm für die tolle Zusammenarbeit und die vielen fruchtbaren Kontakte zu anderen Forschungsgruppen, die er mir ermöglicht hat.

Weiterhin danke ich allen Kollegen, vor allem Christian Möller, der ausdauernd als Korrektor fungiert hat, und falls nötig (und das war es oft), mit Kaffee zur Stelle war. Außerdem möchte ich noch Carina Lorenzen danken, die es auf sich genommen hat, das Englisch der Arbeit zu verbessern wo sie es verstanden hat.

Dank gebührt auch dem Projekt C.1 der DFG-Research-Unit "Nonlinear Partial Differential Equations" Generalized Newtonian fluids and electrorheological fluids, dem ich es zu verdanken habe, dass ich nicht Hunger leiden musste.

Abschließend möchte ich noch meiner Familie und meinen Freunden danken, die in der letzten Zeit doch sehr zurückstecken mussten und trotzdem immer wenn es nötig war — und sei es für ein Bier oder mehrere — zur Stelle waren. Yuri, danke für die Zigaretten.

Contents

Chapter 1

Introduction

Partial differential equations like the stationary Stokes problem arise in numerous physical models, particularly in the modeling of Quasi-Newtonian fluids; see section 1.1. We know the formulation of the stationary Stokes equations to be

$$(1)\qquad \begin{aligned} -\operatorname{div}\mathbf{A}(\nabla u) + \nabla p &= f && \text{in } \Omega,\\ \operatorname{div} u &= 0 && \text{in } \Omega,\\ u &= 0 && \text{on } \partial\Omega, \end{aligned}$$

with $\mathbf{A}$ being a vector-field, which in general is nonlinear.

The main objective of this dissertation is the formulation of a convergent adaptive Uzawa algorithm (AUA) for the numerical solution of the nonlinear stationary Stokes problem. For this purpose, we reformulate the system (1) into a saddle-point problem, which is equivalent to minimizing a functional $\mathcal{F}$ relative to the pressure. The basic idea behind AUA is the method of the steepest descent [18, 24], which is equivalent to the Uzawa method in the linear case [64, 6].

It turns out that the derivative of $\mathcal{F}$ for the pressure q is the divergence of the solution to the nonlinear elliptic equation

$$(2)\qquad \begin{aligned} -\operatorname{div}\mathbf{A}(\nabla u_q) &= f - \nabla q && \text{in } \Omega,\\ u_q &= 0 && \text{on } \partial\Omega. \end{aligned}$$

Hence, $\mathfrak{d}$ is a descent direction of $\mathcal{F}$ in q if and only if

$$D\mathcal{F}(q)(\mathfrak{d}) = \int_\Omega \mathfrak{d} \operatorname{div} u_q < 0,$$

where $D\mathcal{F}$ is the Fréchet derivative of $\mathcal{F}$. We compute a numerical solution of (2) using an adaptive finite element method (AFEM) proposed in [27]. Adaptive finite element methods are a powerful and efficient tool for solving elliptic partial differential equations. Usually they consist of the loop

(AFEM) $\qquad$ SOLVE → ESTIMATE → MARK → REFINE and their convergence has been analyzed in [57, 58, 74, 55, 28, 19, 61, 60]. In particular, our AFEM, based on the quasi-norm error concept introduced in [8], converges to the true solution in a linear fashion.

This motivates the use of the quasi-norm techniques in the AUA as well. As a consequence we define a so called quasi-steepest descent direction. Then starting from an initial guess Q_0 of the pressure p, the AUA consists of a loop

$$Q_{j+1} := Q_j + \mu\,\mathfrak{D}_j, \tag{AUA}$$

where $\mu \geq 0$ and we instrumentalize the AFEM to compute a reasonable approximation $\mathfrak{D}_j$ to the quasi-steepest descent direction in the jth step. The main result shows convergence of the AUA for a fixed step-size μ.

1.1 Quasi-Newtonian Flows

The viscosity ν of a fluid describes its resistance to flow. It is defined to be the proportionality constant between the shear stress τ and the shear rate, i.e., the symmetric part of the velocity gradient $\mathbf{E}(u) = \frac{1}{2}(\nabla u + \nabla u^t)$

$$\tau = \nu\,\mathbf{E}(u).$$

Newton's law of viscosity states that the viscosity ν does not change with the shear rate, i.e., ν is constant.

However, many fluids do not obey Newton's hypothesis, i.e., the viscosity depends on the shear rate: When paint is sheared with a brush, it flows comfortably, but when the shear stress is removed, its viscosity increases so that it no longer flows easily.

We speak of a *pseudo-plastic* or a *shear thinning* fluid, if the viscosity decreases with increasing shear rate. Examples of shear thinning fluids are polymer melts, polymer solutions and some paints. The opposite behavior called *dilatant* or *shear thickening* is found in corn starch, clay slurries, and some surfactants. Fluids of this kind are called *quasi-Newtonian* fluids.

The traditional engineering model for quasi-Newtonian fluids is the so-called *power law*

$$\nu(|\mathbf{E}(u)|) = \nu_0\,|\mathbf{E}(u)|^{r-2},$$

where $\nu_0 > 0$. Thereby pseudo-plastic fluids correspond to $r \in (1, 2)$ whereas dilatant fluids correspond to $r > 2$. It seems to work well for dilatant fluids, but seems to be rather inconvenient for pseudo plastic ones since the power $r-2$ becomes negative. Moreover, many shear-thinning fluids exhibit Newtonian behavior at extreme shear, both low and high. These difficulties can be overcome by the *Carreau law*

$$\nu(|\mathbf{E}(u)|) = \nu_\infty + (\nu_0 - \nu_\infty)(\kappa^2 + |\mathbf{E}(u)|^2)^{\frac{r-2}{2}},$$

where $\kappa > 0$ and $\nu_0 > \nu_\infty \geq 0$. In the case of pseudo-plastic fluids, i.e., when $r \in (1,2)$, for $|\mathbf{E}(u)| \ll \kappa$, the fluid is almost Newtonian with $\nu \approx \nu_\infty + (\nu_0 - \nu_\infty)\kappa^2$. And for $|\mathbf{E}(u)| \gg \kappa$ the fluid is again Newtonian with $\nu \approx \nu_\infty$. In most polymers ν_∞ is zero.

The steady state of a fluid can be modeled by the stationary Stokes equations

$$\begin{aligned} -\operatorname{div}\big(\nu\,\mathbf{E}(u)\big) + \nabla p &= f \qquad \text{in } \Omega,\\ \operatorname{div} u &= 0 \qquad \text{in } \Omega,\\ u &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{1}$$

where u is the velocity and p the kinematic pressure of a fluid inside a domain Ω due to an external body force f. Thereby the definition of the viscosity ν has to be chosen according to the Newtonian, pseudo-plastic, or dilatant behavior of the fluid.

For the ease of exposition we decided to formulate the thesis for the gradient of the velocity instead of its symmetric gradient; see (1). However, thanks to Korn's inequality all results transfer themselves to the formulation with the symmetric gradient; see Remarks 112 and 162.

1.2 Outline

This work starts from analytical fundamentals in Chapter 2 in which we introduce the necessary facts about Orlicz and Orlicz-Sobolev spaces. These spaces are the basis for the treatment of the partial differential equations in the subsequent chapters.

The following Chapter 3 is devoted to the finite element approximation of the analytical solution of nonlinear elliptic problems. It starts with some analytical results on existence and uniqueness of the the solution and then introduces the concept of quasi-norms, which is suitable for quantifying the error of the finite element solution. For this error concept we prove residual based reliable and efficient a posteriori estimators. The main result of this chapter establishes linear convergence of an adaptive finite element method based on the selection criterion of Dörfler for the estimators.

Chapter 4 addresses the numerical solution of the nonlinear stationary Stokes equations. By the use of the theory of saddle-points the weak formulation of the problem can be reformulated to a minimizing problem. A first infinite dimensional Uzawa algorithm, which adapts the idea of the method of steepest descent to quasi-norms, highlights the role of elliptic equations for determinating a reasonable descent direction. Substituting the analytical solutions of the elliptic pde by sufficient good approximations of the AFEM lead to an adaptive Uzawa algorithm (AUA). The main result of this chapter states convergence of AUA.

Chapter 2

Analytical Background

In this chapter we introduce the necessary analytical facts and fix the notation for this work. We start with basic notations and definitions in the first part and introduce Orlicz and Sobolev-Orlicz spaces, which may not be so familiar to the reader in the second part. For the reader's convenience we have provided a table of symbols in Appendix B.

2.1 Preliminaries

We denote by $\mathbb{R}$ the set of real numbers and by $\mathbb{R}_+$ its subset of nonnegative real numbers. The set of natural numbers is denoted by $\mathbb{N}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. The Euclidean scalar product on $\mathbb{R}^m$, $m \in \mathbb{N}$, will be denoted by $\xi \cdot \eta = \sum_{i=1}^m \xi_i \eta_i$ for all $\xi = (\xi_1, \dots, \xi_m)^T, \eta = (\eta_1, \dots, \eta_m)^T \in \mathbb{R}^m$. The corresponding Euclidean product on matrix spaces will be denoted by $\mathbf{P} : \mathbf{Q} = \sum_{i,j=1}^m p_{ij} q_{ij}$ for all $\mathbf{P} = (p_{ij})_{i,j=1,\dots,m}, \mathbf{Q} = (q_{ij})_{i,j=1,\dots,m} \in \mathbb{R}^{m\times m}$, $m \in \mathbb{N}$. Furthermore, we denote the absolute value of real numbers as well as the Euclidean norm on $\mathbb{R}^m$, $\mathbb{R}^{m\times m}$, $m \in \mathbb{N}$, as $|\cdot|$. For $A \subset X$ being a subset of a topological space X, let $\overline{A}$ be the closure of A and ∂A the boundary of A. If $A \subset \mathbb{R}^d$, $d \in \mathbb{N}$, and A is measurable, we denote by $|A|$ the d or $(d-1)$ dimensional Hausdorff measure of A. It will be always clear from the context, which kind of measure is meant.

In the following we will always denote by $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, a bounded polyhedral domain. Let $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$, $d \in \mathbb{N}$, a multi-index and $|\alpha| := \alpha_1 + \dots + \alpha_d$, then $D^\alpha = D_1^{\alpha_1} \dots D_d^{\alpha_d}$, where $D_i = \frac{\partial}{\partial x_i}$ denotes the partial derivative with respect to the i-th component of $\mathbb{R}^d$ and D_i^0 denotes the identity. The number $|\alpha|$ is called the order of the derivative D^α. Let $A \subset \mathbb{R}^m$ be a Lebesgue-measurable set and let $f : A \to \mathbb{R}$ be a measurable function. We denote the Lebesgue integral of f over A by $\int_A f\, dx$. Note, that we suppress the dependence of f on $x \in A$.

The following definitions and results are standard in the theory of partial differential equations. For more details consider, e.g., the books [2, 43, 41, 47, 48, 46]. We denote the space of test-functions as $\mathcal{D}(\Omega) = C_0^\infty(\Omega)$, i.e., as the space of infinitely differentiable functions f that have a compact support $\operatorname{supp}(f)$ in Ω.

Definition 1 (Lebesgue spaces)**.** *We define $L^1_{loc}(\Omega)$ to be the set of locally integrable functions, i.e., the set of all measurable functions $f : \Omega \to \mathbb{R}$ such that*

$$\int_K f\,dx < \infty$$

for all compact subsets $K \subset \Omega$. Let $r \in [1, \infty]$, we define

$$L^r(\Omega) := \{f : \Omega \to \mathbb{R} : f \textit{ is measurable and } \|f\|_{L^r(\Omega)} < \infty\},$$
$$\textit{where } \|f\|_{L^p(\Omega)} := \begin{cases} \left(\int_\Omega |f|^r \, dx \right)^{1/r}, & \textit{if } r < \infty, \\ \operatorname{ess\,sup}_{x\in\Omega} |f(x)|\,, & \textit{if } r = \infty. \end{cases}$$

The closed subspace of $L^r(\Omega)$ consisting of the functions with mean-value zero is denoted by $L^r_0(\Omega)$. Furthermore, we define the quotient space $L^r(\Omega)/\mathbb{R}$ by identifying functions in $L^r(\Omega)$, which only differ by a constant value. A norm on this space is given by

$$\|q\|_{L^r(\Omega)/\mathbb{R}} := \inf_{c\in\mathbb{R}} \|q - c\|_{L^r(\Omega)}\,.$$

As usual, the Lebesgue spaces are actually defined as equivalence classes of functions whose values only differ on a set of Lebesgue measure zero. With this identification, the Lebesgue spaces $\left(L^r(\Omega), \|\cdot\|_{L^r(\Omega)}\right)$ and $\left(L^r(\Omega)/\mathbb{R}, \|\cdot\|_{L^r(\Omega)/\mathbb{R}}\right)$ become Banach spaces. Lebesgue spaces are reflexive if and only if $r \in (1, \infty)$. In particular, for $r' \in (1, \infty)$ with $1/r + 1/r' = 1$, it holds $L^r(\Omega)^* = L^{r'}(\Omega)$ via the representation

$$\langle g, f\rangle_{L^r(\Omega)^*\times L^r(\Omega)} = \int_\Omega fg\,dx \quad \text{for all } f \in L^r(\Omega), g \in L^{r'}(\Omega),$$

where $\langle \cdot, \cdot\rangle_{X^*\times X}$ denotes the dual pairing of the space X. We shall skip the subscript at the duality braces in situations where this cannot give rise to any misunderstanding.

Definition 2 (weak derivatives)**.** *Let $\alpha \in \mathbb{N}_0^d$ and let $f \in L^1_{loc}(\Omega)$ be a locally integrable function. Then, f is said to have α-th weak derivative if there exists a locally integrable function $g \in L^1_{loc}(\Omega)$ such that*

$$\int_\Omega f D^\alpha v\,dx = (-1)^{|\alpha|} \int_\Omega gv\,dx \quad \textit{for all } v \in \mathcal{D}(\Omega).$$

We call $D^\alpha f := g$ the α-th weak derivative of f.

Definition 3 (Sobolev spaces)**.** *Let $r \in [1,\infty]$, and $k \in \mathbb{N}$. We define:*

i) The Sobolev space

$$W^{k,r}(\Omega) := \big\{ f \in L^r(\Omega) : D^\alpha f \in L^r(\Omega) \text{ for all } |\alpha| \le k \big\},$$

with the norm

$$\|f\|_{W^{k,r}(\Omega)} := \begin{cases} \Big(\sum_{|\alpha|\le k} \|D^\alpha f\|^r_{L^r(\Omega)} \Big)^{1/r} & \text{for } r < \infty, \\ \max_{|\alpha|\le k} \|D^\alpha f\|^r_{L^\infty(\Omega)} & \text{for } r = \infty, \end{cases}$$

as well as with the semi-norm

$$|f|_{W^{k,r}(\Omega)} := \begin{cases} \Big(\sum_{|\alpha|= k} \|D^\alpha f\|^r_{L^r(\Omega)} \Big)^{1/r} & \text{for } r < \infty, \\ \max_{|\alpha|= k} \|D^\alpha f\|^r_{L^\infty(\Omega)} & \text{for } r = \infty. \end{cases}$$

ii) The Sobolev space with zero boundary values $W_0^{k,r}(\Omega)$ to be the closure of $C_0^\infty(\Omega)$ in $W^{k,r}(\Omega)$.

iii) For $r' \in (0,\infty)$ with $\frac{1}{r} + \frac{1}{r'} = 1$ we define $W^{-k,r'}(\Omega)$ to be the dual space of $W_0^{k,r}(\Omega)$.

The spaces $(W^{k,r}(\Omega), \|\cdot\|_{W^{k,r}(\Omega)})$ are Banach spaces. Thanks to Poincaré-Friedrich's inequality, on $W_0^{r,k}(\Omega)$ the Sobolev norm is equivalent to the semi-norm, hence $(W_0^{k,r}(\Omega), |\cdot|_{W^{k,r}(\Omega)})$ is also a Banach space. Moreover, those spaces are reflexive if and only if $r \in (1,\infty)$.

All definitions can be generalized to vector-valued functions. A function f with values in $\mathbb{R}^m$, $m \in \mathbb{N}$, is said to be in $L^r(\Omega)^m$ if each of its component functions lies in $L^r(\Omega)$. Recalling that norms on $\mathbb{R}^m$ are denoted in the same way as the absolute value of real numbers, the spaces become Banach spaces with the same definition of norms as in Definition 1. In the same way Sobolev spaces generalize to vector valued functions.

Finally, we want to mention Jensen's inequality, which is fundamental in the analysis of convex functions; see, e.g., [49].

Lemma 4 (Jensen's inequality)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space with $\mu(X) = 1$, $I \subset \mathbb{R}$, be an interval, and $f : X \to I$ be μ-integrable. Then $\int_X f\, d\mu \in I$ and for each convex function $\phi : I \to \mathbb{R}$ it holds*

$$\phi\Big(\int_X f \, d\mu \Big) \le \int_X \phi \circ f \, dx.$$

2.2 Orlicz and Orlicz-Sobolev Spaces

In the theory of weak solutions the solution spaces are closely related to the problem. The Orlicz and Orlicz-Sobolev spaces are the appropriate solution spaces for the weak formulation of the nonlinear problems in Sections 3.1 and 4.1; compare Introduction 1. They are a generalization of the well-known Lebesgue and Sobolev spaces respectively. In fact, many properties of Orlicz-Sobolev spaces are obtained by very straightforward generalizations of the proofs for Sobolev spaces. A detailed presentation of Orlicz spaces can be found in [66, 63, 51]. A short overview of the topic of Orlicz-Sobolev spaces is given in [2, 66], for more detailed information see, e.g., [35].

2.2.1 N-functions

Orlicz spaces are closely connected to N-functions and we concentrate our presentation to properties of N-functions necessary in the subsequent analysis. As the reader may not that familiar with the theory of N-functions, we decided to provide some of the proofs in order to give insight into the techniques that are used in this area. For more detailed presentations we refer to the books of Rao and Ren [66], of Krasnosel'skij and Rutitskij [51].

Definition 5 (N-functions)**.** *A 'nice' Young function, termed an N-function, is a continuous, convex, and strictly monotone function* $\phi : \mathbb{R}^+ \mapsto \mathbb{R}^+$*, such that*

- $\phi(0) = 0$ *and* $\phi(t) > 0$*, if* $t > 0$*,*
- $\lim_{t\to 0} \frac{\phi(t)}{t} = 0$,
- $\lim_{t\to \infty} \frac{\phi(t)}{t} = \infty$.

The following proposition gives a different characterization of N-functions at hand.

Proposition 6 (right derivative)**.** *Let* ϕ *be an N-function. Then it can be represented as*

$$\phi(t) = \int_0^t \phi'(s)\, ds, \qquad t \in \mathbb{R}^+,$$

where $\phi' : \mathbb{R}^+ \mapsto \mathbb{R}^+$ *is a nondecreasing, right continuous function with* $\phi'(0) = 0$ *and* $\lim_{t\to\infty} \phi'(t) = \infty$.

Proof. [66, Corollary 1.3.2] □

N-functions come in mutually complementary pairs. In fact, for an N-function ϕ we can define a right inverse function $(\phi')^{-1}$ of its right derivative via

$$(\phi')^{-1}(t) := \inf\{s : \phi'(s) > t\}, \qquad t > 0.$$

If ϕ' is strictly increasing, then $(\phi')^{-1}$ is the inverse function of ϕ. The function $(\phi')^{-1} : \mathbb{R}^+ \mapsto \mathbb{R}^+$ itself defines an N-function

$$\phi^*(t) := \int_0^t (\phi')^{-1}(s)\,ds, \qquad t > 0, \tag{2.1}$$

called the dual or complementary N-function of ϕ. Obviously it holds $(\phi^*)' = (\phi')^{-1}$ and $(\phi^*)^* = \phi$. Since $(\phi')^{-1}$ is the right inverse for all $t \geq 0$ and all sufficiently small $\epsilon > 0$ there holds

$$\begin{aligned} (\phi^*)'\big(\phi'(t) - \epsilon\big) \leq t \leq (\phi^*)'\big(\phi'(t)\big),\\ \phi'\big((\phi^*)'(t) - \epsilon\big) \leq t \leq \phi'\big((\phi^*)'(t)\big). \end{aligned} \tag{2.2}$$

It is geometrically clear that the pair of N-functions ϕ, ϕ^* forms a pair of Young functions, i.e., it holds

$$st \leq \phi(s) + \phi^*(t) \qquad \text{for all } s, t > 0; \tag{2.3}$$

see Figure 2.1 and [51]. Moreover, if we choose $s = \phi'(t)$ or $t = (\phi^*)'(s)$ it holds equality, i.e.,

$$st = \phi(s) + \phi^*(t). \tag{2.4}$$

Consequently, this implies an alternative definition of ϕ^*

$$\phi^*(t) = \max\{st - \phi(s) : s \geq 0\}. \tag{2.5}$$

The following proposition collects some basic properties of N-functions.

Proposition 7. *Let ϕ, ψ be N-functions. Then for all $t \geq 0$*

$$\phi(\alpha\, t) \leq \alpha\, \phi(t) \qquad \textit{for all } \alpha \in [0, 1], \tag{2.6a}$$

$$\frac{t}{2}\, \phi'\Big(\frac{t}{2}\Big) \leq \phi(t) \leq t\, \phi'(t), \tag{2.6b}$$

$$t \leq (\phi^*)^{-1}(t)\, \phi^{-1}(t) \leq 2\, t, \tag{2.6c}$$

$$\phi\Big(\frac{\phi^*(t)}{t}\Big) \leq \phi^*(t) \leq \phi\Big(2\, \frac{\phi^*(t)}{t}\Big), \tag{2.6d}$$

$$\phi(t) \leq \psi(t) \;\Rightarrow\; \psi^*(t) \leq \phi^*(t). \tag{2.6e}$$

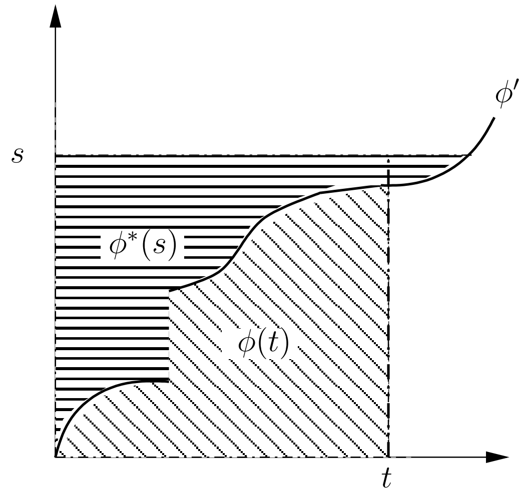

Figure 2.1: A geometric interpretation of Young's inequality.

Proof. Assertion (2.6a) follows immediately from $\phi(0) = 0$ and the convexity of ϕ, since

$$\phi(\alpha\, t) = \phi\big((1-\alpha)\,0 + \alpha\, t\big) \le (1-\alpha)\,\phi(0) + \alpha\,\phi(t) = 0 + \alpha\,\phi(t).$$

To prove assertion (2.6b) we employ the monotonicity of ϕ' to obtain

$$\frac{t}{2}\,\phi'\big(\frac{t}{2}\big) = \int_{t/2}^{t} \phi'\big(\frac{t}{2}\big)\,ds \le \int_{t/2}^{t} \phi'(s)\,ds \le \int_{0}^{t} \phi'(s)\,ds = \phi(t),$$

and

$$\phi(t) = \int_0^t \phi'(s)\,ds \le \int_0^t \phi'(t)\,ds = t\,\phi'(t).$$

For the proof of assertion (2.6c) note that ϕ as well as ϕ^* are strictly monotone functions and thus their inverse functions exist. The right-hand inequality is an immediate consequence of the Young inequality (2.3). In particular,

$$\phi^{-1}(t)\,(\phi^*)^{-1}(t) \le \phi\big((\phi^{-1}(t)\big) + \phi^*\big((\phi^*)^{-1}(t)\big) = 2t.$$

To prove the left inequality of (2.6c), we obtain by the mean value theorem for any $a > 0$, that $\frac{\phi(a)}{a} \le \phi'(\theta)$ for some $\theta \in (0, a)$. Analogously $\phi^*\big(\frac{\phi(a)}{a}\big) \le \frac{\phi(a)}{a}\,(\phi^*)'(\tilde\theta)$ for some $\tilde\theta \in (0, \frac{\phi(a)}{a})$. Combining these estimates, we get by the monotonicity of $(\phi^*)'$ and the definition of the generalized inverse $(\phi^*)' = (\phi')^{-1}$, that

$$\begin{aligned}\phi^*\Big(\frac{\phi(a)}{a}\Big) &\le \frac{\phi(a)}{a}\,(\phi^*)'(\tilde\theta) \le \frac{\phi(a)}{a}\,(\phi^*)'\Big(\frac{\phi(a)}{a}\Big) \le \frac{\phi(a)}{a}\,(\phi^*)'\big(\phi'(\theta)\big)\\ &\le \frac{\phi(a)}{a}\,(\phi^*)'\big(\phi'(a)\big) \le \frac{\phi(a)}{a}\,a = \phi(a).\end{aligned}$$

Now, the assertion follows by taking $a = \phi^{-1}(t)$ and applying $(\phi^*)^{-1}$ to the whole inequality.
Note that the left hand side of (2.6d) is already proven by the last display interchanging the roles of ϕ and ϕ^*. The inequality at the right-hand side is a consequence of (2.6c): In fact, taking $t = \phi^*(s)$ in (2.6c) we get

$$\phi^{-1}\big(\phi^*(s)\big)\, s \le 2\,\phi^*(s).$$

Dividing by s and applying ϕ on each side yield the assertion.

The statement (2.6e) is an easy consequence of (2.5). □

For our purpose one class of N-functions is essential, namely the class of N-functions that satisfies the Δ_2-condition.

Definition 8 (Δ_2-condition)**.** *An N-function ϕ is said to satisfy the Δ_2-condition, if there exists a constant $C > 0$ such that*

$$\phi(2t) \le C\,\phi(t) \qquad \text{for all } t \ge 0.$$

Furthermore, we define $\Delta_2(\phi)$ to be the minimum of the possible constants C. For a family $\{\phi_\lambda\}$ of N-functions for which each member satisfies the Δ_2-condition we define $\Delta_2(\{\phi_\lambda\}) := \sup_\lambda\{\Delta_2(\phi_\lambda)\}$.

Remark 9. *Observe that $\Delta_2(\phi) < \infty$ does not necessarily imply $\Delta_2(\phi^*) < \infty$. In particular, the N-function*

$$\phi^*(t) := e^t - t - 1$$

does not satisfy the Δ_2-condition inasmuch as it increases more rapidly than any polynomial function. The fact that the function ϕ complementary to ϕ^* satisfies the Δ_2-condition can be verified directly from

$$\phi(t) = (1+t)\ln(1+t) - t;$$

for more details consider, e.g., [51].

For the rest of this chapter we use the notation $f \preccurlyeq g$ to indicate $f \le Cg$, with a generic constant C solely depending on some fixed parameters like the Δ_2-constants of given N-functions. We denote $f \preccurlyeq g \preccurlyeq f$ by $f \approx g$.

Based on the Δ_2-property lots of fundamental relations can be derived. First of all we observe that those N-functions satisfy quasi-norm properties.

Corollary 10. *Let $\Delta_2(\phi) < \infty$, then for each constant $\alpha > 0$, there exists a constant $C = C(\alpha, \Delta_2(\phi)) > 0$ such that*

$$\phi(\alpha\, t) \le C\,\phi(t) \qquad \text{for all } t \ge 0.$$

Furthermore,

$$\phi(s+t) \le \frac{\Delta_2(\phi)}{2}\,\phi(s) + \frac{\Delta_2(\phi)}{2}\,\phi(t) \qquad \text{for all } t \ge 0.$$

Proof. The first assertion can be shown in a similar way to the proof of Proposition 11. In particular, let $k \in \mathbb{N}_0$ with $\alpha \leq 2^k$, then taking $C = \Delta_2(\phi)^k$ yields

$$\phi(\alpha\, t) \leq \phi(2^k\, t) \leq \Delta_2(\phi)^k\, \phi(t) = C\, \phi(t).$$

The second assertion is a consequence of the convexity of ϕ. In particular,

$$\phi(s+t) = \phi\big(\frac{1}{2}(2s) + \frac{1}{2}(2t)\big) \leq \frac{1}{2}\phi(2s) + \frac{1}{2}\phi(2t) \leq \frac{\Delta_2(\phi)}{2}\phi(s) + \frac{\Delta_2(\phi)}{2}\phi(t).$$

□

Moreover, we get a generalized Young inequality.

Proposition 11. *Let ϕ be an N-function with $\Delta_2(\phi) < \infty$. Then, for all $\delta > 0$, there exists a constant $C_\delta > 0$, depending on $\Delta_2(\phi)$ and δ, such that*

$$st \leq \delta\, \phi^*(s) + C_\delta\, \phi(t).$$

Proof. It holds by Young's inequality (2.3)

$$st = \delta\, s\, \frac{1}{\delta}\, t \leq \delta\, \phi^*(s) + \delta\, \phi(\frac{1}{\delta}\, t).$$

Let $k \in \mathbb{N}$ such that $\frac{1}{\delta} \leq 2^k$, then we get by the monotonicity of ϕ and the Δ_2-condition

$$\delta\, \phi^*(s) + \delta\, \phi(\frac{1}{\delta}\, t) \leq \delta\, \phi^*(s) + \delta\, \phi(2^k\, t) \leq \delta\, \phi^*(s) + \delta\, \Delta(\phi)^k\, \phi(t).$$

Setting $C_\delta := \delta\, \Delta_2(\phi)^k$ proves the assertion. □

Remark 12. *By duality also it holds*

$$st \leq \delta\, \phi(s) + C_\delta^*\, \phi^*(t)$$

if $\Delta_2(\phi^*) < \infty$. For the ease of simplicity, if $\Delta_2(\{\phi, \phi^*\}) < \infty$, we will not distinguish between the two constants C_δ, C_δ^* and take the maximum of both. We will then say that C_δ depends on $\Delta_2(\{\phi, \phi^*\})$.

Remark 13. *For $r \in (1, \infty)$ and $\kappa \geq 0$, $\nu_0 > \nu_\infty \geq 0$ the N-functions $t \mapsto \frac{1}{r}\, t^r$ and $t \mapsto \int_0^t \big(\nu_\infty + (\nu_0 - \nu_\infty)(\kappa^2 + s^2)^{(r-2)/2}\big) s\, ds$ as well as their dual functions satisfy the Δ_2-condition. In particular, for $\phi(t) = \frac{1}{r}\, t^r$ we have $\Delta_2(\phi) = 2^r$. Moreover, it holds $\phi'(t) = t^{r-1}$, i.e., $(\phi^*)'(t) = (\phi')^{-1}(t) = t^{\frac{1}{r-1}}$. Therefore, we get*

$$\phi^*(t) = \int_0^t s^{\frac{1}{r-1}}\, ds = \frac{1}{r'}\, t^{r'},$$

with $\frac{1}{r} + \frac{1}{r'} = 1$. Hence Young's inequality (11) coincides with the well known classical Young inequality

$$st \leq \delta\, \frac{1}{r}\, t^r + \delta^{\frac{-1}{r-1}}\, \frac{1}{r'}\, t^{r'} \quad \textit{for all } s, t \geq 0.$$

The next proposition sheds light on the nature of pairs of complementary N-functions that satisfy the Δ_2-condition.

Proposition 14. *Let ϕ be an N-function, then the following properties are each equivalent to $\Delta_2(\phi) < \infty$:*

i) There exists $C > 0$ such that

$$\phi'(t)\, t \le C\, \phi(t) \qquad \text{for all } t \ge 0.$$

In particular, $C = \Delta_2(\phi)$.

ii) It holds

$$\nabla_2\,(\phi^*)\, \phi^*(t) \le (\phi^*)'(t)\, t \qquad \text{for all } t \ge 0,$$

for some $\nabla_2(\phi^*) > 1$ depending only on $\Delta_2(\phi)$.

iii) There exists $\alpha > 1$ such that

$$\phi(t) \precsim t^\alpha \quad \text{or equivalently} \quad t^{\alpha^*} \precsim \phi^*(t) \qquad \text{for all } t \ge 0,$$

where $\frac{1}{\alpha} + \frac{1}{\alpha^*} = 1$. The constant α depends solely on $\Delta_2(\phi)$.

Proof. See for instance [66, Theorem 2.3.3, Corollary 2.3.5]. The claim $\frac{1}{\alpha}+\frac{1}{\alpha^*} = 1$ in iii) is a consequence of (2.6e) and the fact that the two functions $t \mapsto \frac{1}{\alpha} t^\alpha$ and $t \mapsto \frac{1}{\alpha^*} t^{\alpha^*}$ are dual; see Remark 13. □

The next Corollary is a direct consequence of Proposition 14.

Corollary 15. *Let ϕ be an N-function. Then $\Delta_2(\{\phi, \phi^*\}) < \infty$ is equivalent to the existence of a constant $\nabla_2(\phi) > 1$, such that*

$$\nabla_2(\phi)\, \phi(t) \le \phi'(t)\, t \le \Delta_2(\phi)\, \phi(t).$$

In particular,

$$\phi'(t)\, t \approx \phi(t). \tag{2.7}$$

Remark 16. *In the literature an N-function ϕ^* satisfying property ii) of Proposition 14 is said to satisfy the ∇_2-condition. This condition in turn is equivalent to $\Delta_2(\phi) < \infty$, thereby recalling that $\phi = (\phi^*)^*$ is the dual function of ϕ^*.*

Proposition 14 iii) further implies that there exist constants $C, c > 0$, $\alpha, \beta \in (1, \infty)$ depending only on $\Delta_2(\{\phi, \phi^*\})$ such that for all $t \ge 0$

$$c\, t^\beta \le \phi(t) \le C\, t^\alpha \quad \text{and} \quad c\, t^{\alpha^*} \le \phi^*(t) \le C\, t^{\beta^*},$$

where $\frac{1}{\alpha} + \frac{1}{\alpha^*} = 1 = \frac{1}{\beta} + \frac{1}{\beta^*}$.

As an immediate consequence of (2.6d) and Corollary 15, we get for N-functions ϕ with $\Delta_2(\{\phi,\phi^*\}) < \infty$ that

$$c\,\phi^*(t) \le \phi\big((\phi^*)'(t)\big) \le C\,\phi^*(t), \tag{2.8}$$

for some constants $c, C > 0$ solely depending on $\Delta_2(\{\phi,\phi^*\})$. Moreover, ϕ' also satisfies a Δ_2-condition.

Corollary 17. *Let ϕ be an N-function with $\Delta_2(\phi) < \infty$, then*

$$\phi'(2t) \le \frac{\Delta_2(\phi)^2}{2}\,\phi'(t).$$

Moreover, for each constant $\alpha > 0$ there exists a constant $C = C(\alpha, \Delta_2(\phi)) > 0$ such that

$$\phi'(\alpha\, t) \le C\,\phi'(t)$$

for all $t \ge 0$.

Proof. It follows from Proposition 14 for N-functions ϕ with $\Delta_2(\phi) < \infty$ that

$$\phi'(2t) = \frac{\phi'(2t)2t}{2t} \le \Delta_2(\phi)\,\frac{\phi(2t)}{2t} \le \Delta_2(\phi)^2\,\frac{\phi(t)}{2t} \le \frac{\Delta_2(\phi)^2}{2}\,\phi'(t). \tag{2.9}$$

The second claim can be deduced as in the proof of Corollary 10. In fact, let $k \in \mathbb{N}_0$ with $\alpha \le 2^k$, then taking $C = \frac{\Delta_2(\phi)^{2k}}{2^k}$ and the monotonicity of ϕ' yield

$$\phi'(\alpha\, t) \le \phi'(2^k\, t) \le \frac{\Delta_2(\phi)^{2k}}{2^k}\,\phi'(t) = C\,\phi'(t).$$

This proves the assertion. □

Remark 16 suggests that an N-function raised to the power of some $\theta \in (0,1)$ close to one, stays similar to an N-function.

Lemma 18. *Let ϕ be a given N-function with $\Delta_2(\{\phi,\phi^*\}) < \infty$. Then, there exists $\theta \in (0,1)$ and an N-function ρ with $\Delta_2(\{\rho,\rho^*\}) < \infty$ such that*

$$\rho(t) \approx (\phi(t))^\theta$$

for all $t \ge 0$. Thereby θ, $\Delta_2(\{\rho,\rho^*\})$, and the constants hidden in $\approx$ depend only on $\Delta_2(\{\phi,\phi^*\})$.

Proof. The proof of this statement for even more general functions can be found in [50, Lemma 1.2.2 and Lemma 1.2.3]. We present here an alternative proof where we explicitely track the dependence on the Δ_2-constant. First we observe that, thanks to Proposition 14 ii) applied to ϕ instead of ϕ^*,

$$\log\left(\frac{\phi(lt)}{\phi(t)}\right) = \int_t^{lt} \frac{\phi'(s)}{\phi(s)}\, ds \geq \nabla_2(\phi) \int_t^{lt} \frac{1}{s}\, ds = \nabla_2(\phi) \log(l),$$

where $\nabla_2(\phi)$ depends only on $\Delta_2(\phi^*)$. Recalling from Proposition 14 *ii)* that $\nabla_2(\phi) > 1$ we can choose $l > 1$ such that $l^{\nabla_2(\phi)-1} > 2$ to obtain

$$\phi(t) \leq \frac{1}{2l}\, \phi(lt) \qquad \text{for all } t \geq 0.$$

Since $\nabla_2(\phi)$ depends only on $\Delta_2(\phi^*)$, l depends only on $\Delta_2(\phi^*)$, too. Let $\theta \in (0,1)$ be chosen later. Direct calculations yield for any $t \geq 0$

$$\big(\phi(t)\big)^\theta \leq \frac{1}{(2l)^\theta}\, \big(\phi(lt)\big)^\theta.$$

We take $\log_{2l}(\frac{3l}{2}) < \theta < 1$ and set $\psi = \phi^\theta$ and $\lambda = l^2 > 1$, hence

$$\psi(t) \leq \frac{2}{3l}\, \psi(lt) \leq \frac{2}{3l}\frac{2}{3l}\, \psi(l^2 t) \leq \frac{1}{2l^2}\, \psi(l^2 t) = \frac{1}{2\lambda}\psi(\lambda t). \tag{2.10}$$

The next step is to prove that

$$\frac{\psi(t_1)}{t_1} \leq \frac{\lambda\psi(\lambda t_2)}{t_2} \tag{2.11}$$

whenever $0 < t_1 < t_2$; see [50, Lemma 1.2.3]. Let $0 < t_1 < t_2 \leq \lambda t_1$, then as ψ is increasing in $[0,\infty)$ it is

$$\frac{\psi(\lambda t_2)}{t_2} \geq \frac{\psi(t_2)}{t_2} \geq \frac{\psi(t_1)}{t_2} \geq \frac{\psi(t_1)}{\lambda t_1}.$$

Conversely let $0 < t_1 < t_2$ and $t_2 > \lambda t_1$. For $r \in \mathbb{R}$ we denote the greatest integer less or equal than r by $\lfloor r \rfloor$. We deduce from a repeatedly application of (2.10)

$$\begin{aligned}\psi(t_2) &= \psi\Big(\frac{t_2}{t_1} t_1\Big) \geq \psi\big(\lambda^{\lfloor \log_\lambda (t_2/t_1)\rfloor} t_1\big) \geq (2\lambda)^{\lfloor \log_\lambda (t_2/t_1)\rfloor}\, \psi(t_1) \\ &\geq (2\lambda)^{\log_\lambda (t_2/t_1)-1}\, \psi(t_1) \geq 2^{\log_\lambda (t_2/t_1)-1} \lambda^{\log_\lambda (t_2/t_1)} \lambda^{-1}\, \psi(t_1) \geq \frac{t_2}{t_1}\lambda^{-1}\psi(t_1).\end{aligned}$$

Recalling the definition of $\lambda = l^2 > 1$, it follows

$$\psi(\lambda t_2) \geq \psi(t_2) \geq \frac{t_2}{t_1}\lambda^{-1}\psi(t_1).$$

and hence (2.11) is established. We observe by basic calculations that the function

$$\rho(t) := \frac{1}{\lambda}\int_0^{t/\lambda} \sup_{0<\tau<s} \frac{\psi(\tau)}{\tau}\,ds$$

is convex with $\rho(t) \le \psi(t)$ and $2\lambda\,\rho(2\lambda t) \ge \psi(t)$. Furthermore, it follows from $\Delta_2(\phi) < \infty$ that

$$\begin{aligned}\rho(2t) \le \psi(2t) = (\phi(2t))^\theta &= \Big(\phi\Big(4\lambda\frac{t}{2\lambda}\Big)\Big)^\theta \le \Big(\phi\Big(2^{\lfloor\log_2(4\lambda)\rfloor+1}\frac{t}{2\lambda}\Big)\Big)^\theta \\ &\le \Delta_2(\phi)^{\theta(\lfloor\log_2(4\lambda)\rfloor+1)}\Big(\phi\Big(\frac{t}{2\lambda}\Big)\Big)^\theta = \Delta_2(\phi)^{\theta(\lfloor\log_2(4\lambda)\rfloor+1)}\,\psi\Big(\frac{t}{2\lambda}\Big) \\ &\le \Delta_2(\phi)^{\theta(\lfloor\log_2(4\lambda)\rfloor+1)}\,2\lambda\,\rho(t).\end{aligned}$$

Thus $\Delta_2(\rho) \le \Delta_2(\phi)^{\theta(\log_2(4\lambda)+1)} < \infty$.

It remains to prove that ρ is an N-function for some $\theta \in (0,1)$ and that $\Delta_2(\rho^*) < \infty$. Let $1 < \beta < \alpha$ as in Remark 16; depending only on $\Delta_2(\{\phi,\phi^*\})$; i.e.,

$$t^\beta \preccurlyeq \phi(t) \preccurlyeq t^\alpha$$

Choosing θ such that $\frac{1}{\beta} < \theta < 1$ yields

$$\frac{\rho(t)}{t} \approx \frac{\big(\phi(t)\big)^\theta}{t} \preccurlyeq \frac{t^{\theta\alpha}}{t} \to 0,$$

as $t \to 0$. On the other hand

$$\frac{\rho(t)}{t} \approx \frac{\big(\phi(t)\big)^\theta}{t} \succcurlyeq \frac{t^{\theta\beta}}{t} \to \infty,$$

as $t \to \infty$. Furthermore, thanks to Proposition 14 iii), the estimate

$$\rho(t) \succcurlyeq t^{\theta\beta} \qquad \text{for all } t \ge 0$$

with $\theta\beta > 1$ implies $\Delta_2(\rho^*) < \infty$ depending only on $\Delta_2(\{\phi,\phi^*\})$. □

Corollary 19. *Let ϕ be an N-function that satisfies $\Delta_2(\{\phi,\phi^*\}) < \infty$. Then there exist constants $C > 0$, $s > 1$, such that*

$$\phi(\alpha t) \le \alpha^s C\,\phi(t) \qquad \textit{for all } t \ge 0.$$

The constants s, C depend solely on $\Delta_2(\{\phi,\phi^*\})$.

Proof. Due to Lemma 18, there exist $\theta \in (0,1)$ and an N-function ρ such that $\rho(t) \approx (\phi(t))^\theta$. Hence, it holds by (2.6a)

$$\phi(\alpha t) \approx (\rho(\alpha t))^{\frac{1}{\theta}} \le \alpha^{\frac{1}{\theta}}(\rho(t))^{\frac{1}{\theta}} \approx \alpha^{\frac{1}{\theta}}\phi(t).$$

Taking $s = 1/\theta$ proves the assertion. □

2.2.2 Orlicz Spaces

Based on the N-functions we can generalize the concept of Lebesgue spaces.

Definition 20 (Orlicz space). *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, $d \in \mathbb{N}$ and let ϕ be an N-function. Then the Orlicz class $\tilde{L}^\phi(\Omega)$ consists of all measurable functions $u : \Omega \to \mathbb{R}$, such that*

$$\int_\Omega \phi(|u|)\, dx < \infty.$$

The quantity $\int_\Omega \phi(|\cdot|)\, dx$ is called the modular induced by ϕ. The Orlicz space is defined as

$$L^\phi(\Omega) := \{u : \Omega \mapsto \mathbb{R} \text{ measurable} : \int_\Omega uv\, dx < \infty \quad \text{for all } v \in \tilde{L}^{\phi^*}(\Omega)\},$$

where we again identify functions that differ on a set of Lebesgue measure zero.

The subspace $L_0^\phi(\Omega)$ as well as the fraction space $L^\phi(\Omega)/\mathbb{R}$ can be defined analogously to the case of Lebesgue functions Definition 1.

Proposition 21. *For an N-function ϕ the Orlicz space $L^\phi(\Omega)$ becomes a Banach space together with the norm*

$$\|u\|_\phi := \sup_{\int_\Omega \phi^*(|v|)\, dx \le 1} \Big| \int_\Omega uv\, dx \Big|. \tag{2.12}$$

Remark 22. *Obviously $L^\phi(\Omega)$ is a linear space and it holds with Young's inequality that $\tilde{L}^\phi(\Omega) \subset L^\phi(\Omega)$. However, in general those two spaces are not equal and $\tilde{L}^\phi(\Omega)$ even does not define a linear space. In fact, this is the case if and only if $\Delta_2(\phi) < \infty$. Then it holds $\tilde{L}^\phi(\Omega) = L^\phi(\Omega)$ (see [51, §8]). Furthermore, in the case $\Delta_2(\{\phi, \phi^*\}) < \infty$ Orlicz functions can be continuously embedded into Lebesgue spaces and vice versa. In particular, it holds with $1 < \beta < \alpha < \infty$ from Remark 16*

$$L^\alpha(\Omega) \subset L^\phi(\Omega) \subset L^\beta(\Omega).$$

One can define another norm on $L^\phi(\Omega)$. In fact, for $v \in L^\phi(\Omega)$ take the Minkowski functional (or Luxemburg norm)

$$\|v\|_{(\phi)} := \inf \big\{ \lambda \in (0, \infty) : \int_\Omega \phi\Big(\frac{|v|}{\lambda}\Big)\, dx \le 1 \big\}. \tag{2.13}$$

It turns out that both norms are equivalent, in particular it holds for all $v \in L^\phi(\Omega)$

$$\|v\|_{(\phi)} \le \|v\|_\phi \le 2\, \|v\|_{(\phi)}\, ; \tag{2.14}$$

see [66, Proposition 3.3.4].

Remark 23. *For an N-function ϕ the Δ_2-condition $\Delta_2(\phi) < \infty$ implies*

$$\int_\Omega \phi\Big(\frac{v}{\|v\|_{(\phi)}}\Big)\, dx = 1.$$

But if this condition is not satisfied, then functions $v \in L^\phi(\Omega)$ can be found such that $\int_\Omega \phi\big(v/\,\|v\|_{(\phi)}\big)\, dx < 1$. Moreover, the equality

$$\int_\Omega \phi\Big(\frac{v}{\lambda_0}\Big)\, dx = 1$$

always implies $\lambda_0 = \|v\|_{(\phi)}$; see [51].

The two norms $\|\cdot\|_\phi$ and $\|\cdot\|_{(\phi^*)}$ are dual in that there holds a Hölder inequality; see, e.g., [66, 51] and Proposition 25.

Proposition 24. *Let ϕ be an N-function. Then for every $v \in L^\phi(\Omega)$, $w \in L^{\phi^*}(\Omega)$ we have*

$$\Big|\int_\Omega v\, w\, dx\Big| \le \|v\|_{(\phi)}\, \|w\|_{\phi^*}$$

and

$$\Big|\int_\Omega v\, w\, dx\Big| \le \|v\|_{\phi}\, \|w\|_{(\phi^*)}\,.$$

We introduce the space E^ϕ to be the closure of the space of bounded functions $L^\infty(\Omega)$ in $L^\phi(\Omega)$. With this definition E^ϕ is a separable Banach space. The following proposition states among other facts that even equality in the Hölder inequality Proposition 24 can be obtained; see [51, Chapter II, §14] or [66, Chapter VI, Theorems 6 and 7].

Proposition 25. *Let ϕ be an N-function and ϕ^* its complementary function. Then*

$$\Big(E^\phi(\Omega), \|\cdot\|_\phi\Big)^* = \Big(L^{\phi^*}(\Omega), \|\cdot\|_{(\phi^*)}\Big)$$

and

$$\Big(E^\phi(\Omega), \|\cdot\|_{(\phi)}\Big)^* = \Big(L^{\phi^*}(\Omega), \|\cdot\|_{\phi^*}\Big).$$

In particular, it holds for $w \in L^{\phi^*}(\Omega)$

$$\sup_{v \in E^\phi(\Omega), \|v\|_\phi = 1} \int_\Omega w\, v\, dx = \|w\|_{(\phi^*)}$$

and

$$\sup_{v \in E^\phi(\Omega), \|v\|_{(\phi)} = 1} \int_\Omega w\, v\, dx = \|w\|_{\phi^*}\,.$$

The following proposition underlines the role, which the Δ_2-condition plays in the theory of Orlicz spaces; see, e.g., [51, 66].

Proposition 26. *The following assertions are equivalent for an N-function ϕ:*

i) $L^\phi(\Omega)$ *is separable;*

ii) $L^\phi(\Omega) = E^\phi(\Omega)$*;*

iii) $\tilde{L}^\phi(\Omega) = L^\phi(\Omega)$*;*

iv) $\big(L^\phi(\Omega), \|\cdot\|_\phi\big)^* = \big(L^{\phi^*}(\Omega), \|\cdot\|_{(\phi^*)}\big)$*;*

v) $\big(L^\phi(\Omega), \|\cdot\|_{(\phi)}\big)^* = \big(L^{\phi^*}(\Omega), \|\cdot\|_{\phi^*}\big)$*;*

vi) $\Delta_2(\phi) < \infty$*.*

Remark 27. *As a consequence of Proposition 26, for an N-function ϕ, $L^\phi(\Omega)$ is reflexive if and only if $\Delta(\{\phi, \phi^*\}) < \infty$.*

Remark 28. *When we revisit Remark 13, i.e., taking $\phi(t) = \frac{1}{r} t^r$, $r \in (1, \infty)$ we get for $u \in L^\phi(\Omega)$*

$$\|u\|_{(p)} = \inf \big\{\lambda \geq 0 : \int_\Omega \frac{1}{r} \Big|\frac{u}{\lambda}\Big|^r \, dx \leq 1\big\},$$

and thus $\|\cdot\|_{(\phi)} = \frac{1}{r^{1/r}} \|\cdot\|_{L^r(\Omega)}$, i.e., $L^\phi(\Omega) = L^r(\Omega)$. Therefore, the Orlicz spaces are a generalization of the well known Lebesgue spaces.

In Remark 73 we show that also for $\phi(t) = \int_0^t (\kappa + s)^{r-2} s \, ds$ and $\phi(t) = \int_0^t (\kappa^2 + s^2)^{\frac{r-2}{2}} s \, ds$ with $\kappa \geq 0$, it holds $\|\cdot\|_{(\phi)} \approx \|\cdot\|_{L^r(\Omega)}$.

The next result sheds light on the relation between the defining N-functions of different Orlicz spaces.

Proposition 29. *Let ϕ, ψ be to N-functions with $\Delta_2(\{\phi, \psi\}) < \infty$, then*

$$L^\phi(\Omega) \subset L^\psi(\Omega)$$

if and only if there exists $t_0 > 0$, such that

$$\psi(t) \preccurlyeq \phi(t) \qquad \text{for all } t \geq t_0.$$

Proof. From [51, Chapter II, Theorem 13.1] we have that for general N-functions a necessary and sufficient condition that $L^\phi(\Omega) \subset L^\psi(\Omega)$ is that there exists $t_0, k > 0$, such that

$$\psi(t) \leq \phi(k\,t) \qquad \text{for all } t \geq t_0. \tag{2.15}$$

Hence, it suffices to prove that this condition is equivalent to

$$\psi(t) \preccurlyeq \phi(t) \qquad \text{for all } t \geq t_0. \tag{2.16}$$

Since ϕ satisfies the Δ_2-condition it follows from Corollary 10 that $\phi(k\,t) \preccurlyeq \phi(t)$ and therefore (2.15) implies (2.16). On the other hand, it holds for $C \geq 1$ by the monotonicity of ϕ' and (2.6b)

$$C\,\phi(t) \leq C\,\phi'(t)t = \phi'(t)(C\,t) \leq \phi'(C\,t)(C\,t) \leq \phi(2C\,t)$$

for all $t \geq 0$. Hence, (2.16) also implies (2.15). □

Finally we introduce another convergence concept on Orlicz spaces.

Definition 30 (mean convergence)**.** *For an N-function ϕ, we say that a sequence of functions $(v_n)_{n\in\mathbb{N}} \subset L^\phi(\Omega)$ is mean (or modular) convergent to a function $v \in L^\phi(\Omega)$, if*

$$\int_\Omega \phi(|v - v_n|)\,dx \to 0 \quad \text{as } n \to \infty.$$

Proposition 31. *Let ϕ be an N-function, then norm convergence implies mean convergence. If additionally ϕ satisfies the Δ_2-condition then mean-convergence also implies norm convergence.*

Proof. The proof can be found in [51, Theorem II.9.4] □

Remark 32. *Proposition 31 further implies, that if the N-function ϕ satisfies a Δ_2-condition, a sequence $(v_n)_{n\in\mathbb{N}} \subset L^\phi(\Omega)$ stays bounded in mean if and only if it stays bounded in $L^\phi(\Omega)$. In fact, let $(v_n)_{n\in\mathbb{N}} \subset L^\phi(\Omega)$ be a bounded sequence in the norm sense, i.e., $\|v_n\|_{(\phi)} \leq \alpha$ for an $\alpha > 0$. It holds by the monotonicity of ϕ and Corollary 10*

$$\begin{aligned}\int_\Omega \phi(|v_{n_k}|)\,dx &= \int_\Omega \phi\Big(\frac{\|v_{n_k}\|_{(\phi)}}{\|v_{n_k}\|_{(\phi)}}\,|v_{n_k}|\Big)\,dx \leq \int_\Omega \phi\Big(\frac{\alpha}{\|v_{n_k}\|_{(\phi)}}\,|v_{n_k}|\Big)\,dx \\ &\leq \int_\Omega C\,\phi\Big(\frac{|v_{n_k}|}{\|v_{n_k}\|_{(\phi)}}\Big)\,dx \leq C,\end{aligned}$$

for a constant $C > 0$ depending on α and $\Delta_2(\phi)$.

On the other hand assume that $(v_n)_{n\in\mathbb{N}} \subset L^\phi(\Omega)$ diverge in the norm sense, i.e.,

$$\|v_n\|_{(\phi)} \to \infty$$

as $n \to \infty$. Thus, we may assume w.l.o.g. that $\|v_n\|_{(\phi)} \geq 1$ for all $n \in \mathbb{N}$ and hence with (2.6a)

$$1 = \int_\Omega \phi\Big(\frac{|v_n|}{\|v_n\|_{(\phi)}}\Big)\,dx \leq \frac{1}{\|v_n\|_{(\phi)}} \int_\Omega \phi(|v_n|)\,dx,$$

where the left equality is due to $\Delta_2(\phi) < \infty$*; see Remark 23 and [51]. Hence, the sequence* $(v_n)_{n\in\mathbb{N}} \subset L^\phi(\Omega)$ *is divergent in the modular sense, too. Note that the equality* $1 = \int_\Omega \phi(v/\|v\|_{(\phi)})\,dx$, $v \in L^\phi(\Omega)$ *is a consequence of the* Δ_2*-condition and the definition of the norm* $\|\cdot\|_{(\phi)}$ *and does not hold for general N-functions* ϕ*, see [51, 66].*

2.2.3 Orlicz-Sobolev Spaces

In order to establish the nonlinear partial differential equations in Sections 3.1 and 4.1 we need to have weak derivatives of Orlicz functions. This leads to the so called Orlicz-Sobolev spaces. A detailed presentation can, e.g., be found in [2, 66, 35].

Definition 33 (Orlicz-Sobolev spaces)**.** *Let* ϕ *be an N-function,* $k \in \mathbb{N}$*. We define:*

i) The space $W^{k,\phi}(\Omega)$ *consists of all functions* f *in the Orlicz space* $L^\phi(\Omega)$ *with weak derivatives* $D^\alpha f \in L^\phi(\Omega)$*, where* $\alpha \in \mathbb{N}^d$*,* $|\alpha| \leq k$*. We equip* $W^{k,\phi}(\Omega)$ *with a norm*

$$\|f\|_{W^{k,\phi}(\Omega)} := \sum_{|\alpha|\leq k} \|D^\alpha f\|_\phi\,,$$

and a semi-norm

$$|f|_{W^{k,\phi}(\Omega)} := \sum_{|\alpha|=k} \|D^\alpha f\|_\phi\,.$$

ii) The space $W_0^{k,\phi}(\Omega)$ *is defined to be the closure of* $C_0^\infty(\Omega)$ *in* $W^{k,\phi}(\Omega)$*.*

iii) We denote $WE^{k,\phi}(\Omega)$ *to be the closure of* $W^{k,\infty}(\Omega)$ *in* $W^{k,\phi}(\Omega)$*.*

iv) If $\Delta_2(\{\phi,\phi^*\}) < \infty$*, we denote* $W^{-k,\phi^*}(\Omega)$ *to be the dual space of* $W_0^{k,\phi}(\Omega)$*.*

v) We say that a sequence $(f_n)_{n\in\mathbb{N}} \subset W^{k,\phi}(\Omega)$ *converges in mean if each of the sequences* $(D^\alpha f_n)_{n\in\mathbb{N}}$*,* $\alpha \in \mathbb{N}^d$*,* $|\alpha| \leq k$ *converges in mean in* $L^\phi(\Omega)$*.*

The definitions and results above extend to functions with values in $\mathbb{R}^m$, $m \in \mathbb{N}$ in the same way as Lebesgue spaces and Sobolev spaces do. We shall denote the resulting spaces as $L^\phi(\Omega)^m$, $W^{k,\phi}(\Omega)^m$, $W_0^{k,\phi}(\Omega)^m$, and $W^{-k,\phi^*}(\Omega)^m$ respectively.

Lemma 34 (Poincaré-Friedrich's inequality)**.** *Let* ϕ *be a given N-function with* $\Delta_2(\phi) < \infty$ *and* $f \in W_0^{1,\phi}(\Omega)$*, then*

$$\int_\Omega \phi(|f|)\,dx \preccurlyeq \int_\Omega \phi(|\nabla f|).$$

The constant hidden in $\preccurlyeq$ *solely depends on* $\Delta_2(\phi) < \infty$ *and* Ω*.*

Proof. Since $C_0^\infty(\Omega)$ is dense in $W_0^{1,\phi}(\Omega)$ and norm convergence implies mean-convergence (see Proposition 31), it suffices to establish the inequality for $f \in C_0^\infty(\Omega)$. We may assume that $\Omega \subset W = \{(x_1,\dots,x_d) : -s < x_i < s\}$ for some $s > 0$, and set $f \equiv 0$ in $W \setminus \Omega$. By the fundamental theorem of calculus, we then get for $x = (x_1,\dots,x_d)$

$$\begin{aligned}|f(x)| &= |f(x) - f(-s, x_2, \dots, x_d)| \\ &\le \int_{-s}^{x_1} |D_1 f(t, x_2, \dots, x_d)|\, dt \le \int_{-s}^{s} |D_1 f(t, x_2, \dots, x_d)|\, dt;\end{aligned}$$

see, e.g., [13]. Now, we apply ϕ on both sides and obtain with the monotonicity of ϕ, that

$$\phi(|f(x)|) \le \phi\Big(\int_{-s}^{s} |D_1 f(t, x_2, \dots, x_d)|\, dt\Big).$$

Since ϕ is convex, we can apply Jensen's inequality (Lemma 4) to get

$$\phi(|f(x)|) \le \frac{1}{2s}\int_{-s}^{s} \phi\big(2s\,|D_1 f(t, x_2, \dots, x_d)|\,\big)\, dt.$$

Observe that the right hand side is independent of x_1, hence

$$\int_{-s}^{s} \phi(|f(x)|)\, dx_1 \le \int_{-s}^{s} \phi\big(2s\,|D_1 f(t, x_2, \dots, x_d)|\,\big)\, dt.$$

Then integrating with respect to the other coordinates yields

$$\int_W \phi(|f(x)|)\, dx \le \int_W \phi\big(2s\,|D_1 f(x)|\,\big)\, dx \le \int_W \phi\big(2s\,|\nabla f(x)|\,\big)\, dx.$$

Now, $2s$ can be dragged out by Corollary 10 and hence the assertion is proved. □

Lemma 35. *Let X be a space with norms $\|\cdot\|_1, \|\cdot\|_2$ that define the same convergence, i.e., a sequence $(x_n)_{n\in\mathbb{N}} \subset X$ converges with respect to $\|\cdot\|_1$ if and only if it converges with respect to $\|\cdot\|_2$. Then, the two norms are equivalent.*

Proof. Assume contrary. Then, w.l.o.g, there exists a sequence $(x_n)_{n\in\mathbb{N}} \subset X$, $x_n \ne 0$, $n \in \mathbb{N}$, such that $\|x_n\|_1 = C_n \|x_n\|_2$ with $C_n \to 0$ as $n \to \infty$. Dividing x_n by $\|x_n\|_2$ yields

$$\left\|\frac{x_n}{\|x_n\|_2}\right\|_1 = C_n \to 0$$

as $n \to \infty$. Since $\|\cdot\|_1$ and $\|\cdot\|_2$ define the same convergence it follows

$$1 = \left\|\frac{x_n}{\|x_n\|_2}\right\|_2 \to 0$$

as $n \to \infty$. This is a contradiction. □

Corollary 36. *Let ϕ be as in Lemma 34, then it holds for $f \in W_0^{1,\phi}(\Omega)$*

$$\|f\|_{W_0^{1,\phi}(\Omega)} \approx |f|_{W_0^{1,\phi}(\Omega)} \approx \|\nabla f\|_\phi .$$

Furthermore, if $(f_n)_{n\in\mathbb{N}} \subset W_0^{1,\phi}(\Omega)$ converges in mean, then $(\nabla f_n)_{n\in\mathbb{N}} \subset L^\phi(\Omega)^d$ converges in mean.

Proof. To prove the second statement, we observe by Corollary 10 that

$$\phi(|\nabla f|) \leq \max_{i=1,\dots,d} \phi\big(\sqrt{d}\,|D_i f|\big) \preccurlyeq \sum_{i=1}^d \phi(|D_i f|).$$

On the other hand,

$$\sum_{i=1}^d \phi(|D_i f|) \leq d \max_{i=1,\dots,d} \phi\big(\,|D_i f|\,\big) \leq d\,\phi(|\nabla f|).$$

Integrating over Ω the claim follows with Lemma 34.

Now, Lemma 34, Proposition 31, and the above observations imply that the three expressions

$$\|\cdot\|_{W_0^{1,\phi}(\Omega)}, \quad |\cdot|_{W_0^{1,\phi}(\Omega)}, \text{ and } \quad \|\nabla\cdot\|_\phi$$

are norms, which define the same convergence. Hence, the assertion follows by Lemma 35. ☐

We summarize some properties of Orlicz-Sobolev spaces in the next proposition; see [2]. We refer the reader to the corresponding results for Sobolev spaces for method of proof. The details can, e.g., be found in [35].

Proposition 37. *Let ϕ be an N-function and $k \in \mathbb{N}$.*

i) The spaces $W^{k,\phi}(\Omega)$, $WE^{k,\phi}(\Omega)$, and $W_0^{k,\phi}(\Omega)$ are Banach spaces equipped with the norm $\|\cdot\|_{W^{k,\phi}(\Omega)}$.

ii) The spaces $WE^{k,\phi}(\Omega)$, $W_0^{k,\phi}(\Omega)$ are separable.

iii) The spaces $W^{k,\phi}(\Omega)$ and $W_0^{k,\phi}(\Omega)$ are reflexive if and only if $\Delta_2(\{\phi,\phi^*\}) < \infty$. Moreover, this is equivalent to $W^{k,\phi}(\Omega) = WE^{k,\phi}(\Omega)$.

iv) Each element v of the dual space $(WE^{k,\phi}(\Omega))^*$ is given by

$$v(u) = \sum_{|\alpha|\leq k} \int_\Omega (D^\alpha u)\, v_\alpha \, dx$$

for some functions $v_\alpha \in L^{\phi^*}(\Omega)$, $\alpha \in \mathbb{N}_0^d$, $0 \leq |\alpha| \leq k$.

Chapter 3

Adaptive Finite Elements for the Nonlinear Poisson Problem

After a short overview on existence and uniqueness of a solution for the nonlinear Poisson equation we introduce in Section 3.2 an error concept based on the so called *quasi-norm*, introduced by Barrett and Liu; cf. [8, 9]. The next section, Section 3.3 is concerned with the finite element framework for the discrete nonlinear Poisson problem. Based on the error bounds of Section 3.4, the last section, Section 3.5, contains the convergence analysis of an adaptive finite element method AFEM based on [28, 27, 19].

Note that we consider the problem for d-dimensional vector valued functions, i.e., for a d-dimensional system of Poisson equations.

3.1 Nonlinear Poisson Equation

In this section we discuss the analytical aspects of the nonlinear Poisson equation with homogeneous Dirichlet boundary values. Since the nonlinearity of the problem is defined by an N-function, the natural space for weak solutions turns out to be an Orlicz-Sobolev space. We restrict ourselves to the case of N-functions satisfying $\Delta_2(\{\phi, \phi^*\}) < \infty$. Therefore, Orlicz-Sobolev spaces become separable and reflexive Banach spaces and thus the well established theory of monotone operators provides existence and uniqueness of a solution; see for instance [69, 81]. Finally, we introduce an energy functional whose minimal function coincides with the solution of the nonlinear Poisson equation.

3.1.1 Stating the Problem

Let ϕ be an N-function with $\Delta_2(\{\phi, \phi^*\}) < \infty$. In the sequel we discuss vector valued partial differential equations of the form: Find $u : \Omega \to \mathbb{R}^d$ such that for given $g : \Omega \to \mathbb{R}^d$

$$\begin{aligned} -\operatorname{div} \mathbf{A}(\nabla u) &= g \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{3.1}$$

where $\mathbf{A} : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ is defined as

$$\mathbf{A}(\mathbf{Q}) = \phi'(|\mathbf{Q}|)\frac{\mathbf{Q}}{|\mathbf{Q}|}.$$

Hereafter we assume that $g \in W^{-1,\phi^*}(\Omega)^d$. The weak form of (3.1) reads as follows: For an N-function ϕ with $\Delta_2(\{\phi,\phi^*\}) < \infty$ find $u \in W_0^{1,\phi}(\Omega)^d$ such that

$$\int_\Omega \mathbf{A}(\nabla u) : \nabla v \, dx = \langle g, v\rangle \qquad \text{for all} \quad v \in W_0^{1,\phi}(\Omega)^d. \tag{3.2}$$

Remark 38. *Note that, in face of the Stokes problem in Chapter 4, we formulated problem* (3.1) *for functions with d-dimensional values. However, this restriction is only for the ease of presentation. All statements of this chapter carry over to problems where $u \in W_0^{1,\phi}(\Omega)^m$ and $g \in L^{\phi^*}(\Omega)^m$ for any $m \in \mathbb{N}$.*

Remark 39. *The expressions in problem* (3.2) *are well-defined. In fact, it follows from* (2.8) *that $\mathbf{A}(\nabla u) \in L^{\phi^*}(\Omega)^{d\times d}$. Furthermore, it holds with Proposition 26 that $L^{\phi^*}(\Omega)^{d\times d} = (L^{\phi}(\Omega)^{d\times d})^*$ and thus the left hand side is well-defined since $\nabla v \in L^{\phi}(\Omega)^{d\times d}$ for all $v \in W_0^{1,\phi}(\Omega)^d$. The right hand side is well-defined by the choice of g.*

We can interprete equation (3.1) as an operator-equation in the dual space $W^{-1,\phi^*}(\Omega)^d$, defining the non-linear operator $-\operatorname{div} \mathbf{A}(\nabla\cdot) \in W^{-1,\phi^*}(\Omega)^d$ by

$$\langle -\operatorname{div} \mathbf{A}(\nabla u), v\rangle := \int_\Omega \mathbf{A}(\nabla u) : \nabla v \, dx.$$

Hence, (3.1) is equivalent to

$$-\operatorname{div} \mathbf{A}(\nabla u) = g \qquad \text{in } W^{-1,\phi^*}(\Omega).$$

For the numerical analysis the following assumption is crucial. It is the key ingredient to proof continuity and ellipticity of (3.1).

Assumption 40. Let ϕ be an N-function with $\Delta_2(\{\phi,\phi^*\}) < \infty$ and let $\phi \in C^2((0,\infty))$ such that there exist constants $c, C > 0$ with

$$c\, t\, \phi''(t) \le \phi'(t) \le C\, t\, \phi''(t) \qquad \text{for all } t \ge 0,$$

where we extend $t\,\phi''(t)$ continuously to zero by setting $t\,\phi''(t) := 0$ for $t = 0$.

The next theorem is from [26] and states that Assumption 40 carries over to dual functions.

Proposition 41. *Let ϕ be an N-function with $\Delta_2(\{\phi,\phi^*\}) < \infty$. Then ϕ satisfies Assumption 40 if and only if ϕ^* satisfies Assumption 40.*

Proof. We just have to prove one direction, the other direction follows by duality. Assume that ϕ satisfies Assumption 40. From $(\phi^*)'(t) = (\phi')^{-1}$ we find by the inverse function theorem, Assumption 40, (2.7), (2.8), and Proposition 14 (ϕ^* replaced by ϕ) that for $t > 0$

$$(\phi^*)''(t) = \frac{1}{\phi''\big((\phi^*)'(t)\big)} \approx \frac{\big((\phi^*)'(t)\big)^2}{\phi\big((\phi^*)'(t)\big)} \approx \frac{\big((\phi^*)'(t)\big)^2}{\phi^*(t)} \approx \frac{\big(\phi^*(t)\big)^2}{\phi^*(t)\, t^2} = \frac{\phi^*(t)}{t^2}.$$

This proves the assertion. □

Remark 42. *Assumption 40 implies that ϕ is strictly convex since $\phi'(t) > 0$ for $t > 0$ and hence $\phi''(t) \approx \frac{\phi'(t)}{t} > 0$ on $(0,\infty)$. Moreover, ϕ' is strictly monotone increasing and thus the inverse function of ϕ' exists.*

Recalling Remark 13, the N-functions $t \mapsto \frac{1}{r}t^r$ and $t \mapsto \int_0^t \big(\nu_\infty + (\nu_0 - \nu_\infty)(\kappa^2 + s^2)^{(r-2)/2}\big) s\, ds$ for $r \in (1,\infty)$, $\kappa \geq 0$, and $\nu_0 > \nu_\infty \geq 0$ satisfy Assumption 40. In particular, for $\phi(t) = \frac{1}{r}t^r$ it holds $\big(\frac{1}{r}\, t^r\big)'' = (r-1)\, t^{r-2}$. Therefore, the constants in Assumption 40 can be determinated exactly as $c = C = r - 1$. This means that the PDE (3.1) covers the well-known nonlinear Poisson equation

$$\begin{aligned} -\operatorname{div} |\nabla u|^{r-2}\, \nabla u &= g && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned}$$

as well as the variants, which are widely used in the modeling of quasi-Newtonian flow; see Section 1.1.

3.1.2 Existence and Uniqueness of Solutions

To establish the existence and uniqueness of solutions of (3.2) we have to analyze the vector field $\mathbf{A}$. The proof of the next proposition can be found in [26], but since it is one of the key estimates in the subsequent analysis we decided to prove it in detail.

Proposition 43. *Let ϕ be an N-function satisfying Assumption 40, then there exist constants $c, C > 0$ such that for all $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{d\times d}$*

$$\begin{aligned} \big(\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})\big) : (\mathbf{P} - \mathbf{Q}) &\geq c\, \phi''(|\mathbf{P}| + |\mathbf{Q}|)\, |\mathbf{P} - \mathbf{Q}|^2\,, \\ |\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})| &\leq C\, \phi''(|\mathbf{P}| + |\mathbf{Q}|)\, |\mathbf{P} - \mathbf{Q}|\,. \end{aligned}$$

The constants c, C depend solely on $\Delta_2(\{\phi,\phi^*\})$ and the constants of Assumption 40. For $\mathbf{P}, \mathbf{Q} = 0$ extend the right hand sides continuously to zero; cf., Assumption 40.

Remark 44. *The estimates of Proposition 43 are a generalization of those of Barret and Liu in [9, 8]. In fact, for* $\phi(t) = \frac{1}{r}t^r$*, with* $r \in (1, \infty)$*, we have* $\phi''(t) = (r-1)\,t^{r-2}$ *for* $t > 0$ *and thus Proposition 43 becomes*

$$\begin{aligned}\big(|\mathbf{P}|^{r-2}\,\mathbf{P} - |\mathbf{Q}|^{r-2}\,\mathbf{Q}\big) : (\mathbf{P}-\mathbf{Q}) &\geq c\,(|\mathbf{P}| + |\mathbf{Q}|)^{r-2}\,|\mathbf{P}-\mathbf{Q}|^2\,,\\ \big||\mathbf{P}|^{r-2}\,\mathbf{P} - |\mathbf{Q}|^{r-2}\,\mathbf{Q}\big| &\leq C\,(|\mathbf{P}| + |\mathbf{Q}|)^{r-2}\,|\mathbf{P}-\mathbf{Q}|\,.\end{aligned}$$

To prove Proposition 43 we need some basic inequalities. The first lemma is essentially contained in [1] and proved with sharp constants in [25].

Lemma 45. *Let* $\alpha > -1$*, then for all* $\mathbf{P}_0, \mathbf{P}_1 \in \mathbb{R}^{d\times d}$ *with* $|\mathbf{P}_0| + |\mathbf{P}_1| > 0$

$$c(\alpha)\,(|\mathbf{P}_0| + |\mathbf{P}_1|)^\alpha \leq \int_0^1 |\mathbf{P}_\theta|^\alpha\,d\theta \leq C(\alpha)\,(|\mathbf{P}_0| + |\mathbf{P}_1|)^\alpha$$

with

$$c(\alpha) = \min\{\frac{1}{\alpha+1}, \frac{2^{-\alpha}}{\alpha+1}, 2^{-\alpha}\}, \qquad C(\alpha) = \max\{\frac{1}{\alpha+1}, \frac{2^{-\alpha}}{\alpha+1}, 2^{-\alpha}\}$$

where $\mathbf{P}_\theta = (1-\theta)\,\mathbf{P}_0 + \theta\,\mathbf{P}_1$*. The constants* $c(\alpha), C(\alpha)$ *are optimal.*

The next lemma from [26] contains a generalization of the above lemma to the case of N-functions.

Lemma 46. *Let* ϕ *be an N-function with* $\Delta_2(\{\phi, \phi^*\}) < \infty$*. Then, for all* $\mathbf{P}_1, \mathbf{P}_2 \in \mathbb{R}^{d\times d}$ *with* $|\mathbf{P}_1| + |\mathbf{P}_2| > 0$ *it holds*

$$\frac{\phi'(|\mathbf{P}_1| + |\mathbf{P}_2|)}{|\mathbf{P}_1| + |\mathbf{P}_2|} \approx \int_0^1 \frac{\phi'(|\mathbf{P}_\theta|)}{|\mathbf{P}_\theta|}\,d\theta,$$

where $\mathbf{P}_\theta = (1-\theta)\,\mathbf{P}_1 + \theta\,\mathbf{P}_2$*. The constants hidden in* $\approx$ *solely depend on* $\Delta_2(\{\phi, \phi^*\})$*.*

Proof. From Proposition 14 and Jensen's inequality (Lemma 4) we derive

$$\int_0^1 \frac{\phi'(|\mathbf{P}_\theta|)}{|\mathbf{P}_\theta|}\,d\theta \succeq \int_0^1 \frac{\phi(|\mathbf{P}_\theta|)}{|\mathbf{P}_\theta|^2}\,d\theta \geq \int_0^1 \frac{\phi(|\mathbf{P}_\theta|)}{(|\mathbf{P}_1| + |\mathbf{P}_2|)^2}\,d\theta \geq \frac{\phi(\int_0^1 |\mathbf{P}_\theta|\,d\theta)}{(|\mathbf{P}_1| + |\mathbf{P}_2|)^2}.$$

Since by Lemma 45 $\int_0^1 |\mathbf{P}_\theta| \geq \frac{1}{4}\,|\mathbf{P}_1| + |\mathbf{P}_2|$, we obtain by means of Corollary 15

$$\frac{\phi(\int_0^1 |\mathbf{P}_\theta|\,d\theta)}{(|\mathbf{P}_1| + |\mathbf{P}_2|)^2} \geq \frac{\phi(\frac{1}{4}(|\mathbf{P}_1| + |\mathbf{P}_2|))}{(|\mathbf{P}_1| + |\mathbf{P}_2|)^2} \geq \frac{1}{\Delta_2(\phi)^2}\frac{\phi(|\mathbf{P}_1| + |\mathbf{P}_2|)}{(|\mathbf{P}_1| + |\mathbf{P}_2|)^2} \approx \frac{\phi'(|\mathbf{P}_1| + |\mathbf{P}_2|)}{(|\mathbf{P}_1| + |\mathbf{P}_2|)}.$$

This proves the first part. For the second part we recall from Lemma 18 that there exists some $\gamma \in (0,1)$ and some N-function ρ with $\Delta_2(\{\rho, \rho^*\}) < \infty$ such that $\phi^\gamma \approx \rho$, where $\Delta_2(\{\rho,\rho^*\})$ as well as the constants hidden in $\approx$ solely depend on $\Delta_2(\{\phi,\phi^*\})$. Again involving Corollary 15, i.e., $\phi(t)\approx\phi'(t)t$ and $\rho(t)\approx\rho'(t)t$, we deduce

$$\begin{aligned}\int_0^1 \frac{\phi'(|\mathbf{P}_\theta|)}{|\mathbf{P}_\theta|}\,d\theta &\approx \int_0^1 \frac{\phi(|\mathbf{P}_\theta|)}{|\mathbf{P}_\theta|^2}\,d\theta \approx \int_0^1 \frac{\big(\rho(|\mathbf{P}_\theta|)\big)^{\frac{1}{\gamma}}}{|\mathbf{P}_\theta|^2}\,d\theta\\ &\approx \int_0^1 \big(\rho'(|\mathbf{P}_\theta|)\big)^{\frac{1}{\gamma}}\,|\mathbf{P}_\theta|^{\frac{1}{\gamma}-2}\,d\theta.\end{aligned}$$

The monotonicity of ρ' and Lemma 45 with $\alpha=\frac{1}{\gamma}-2>-1$ imply

$$\begin{aligned}\int_0^1 \frac{\phi'(|\mathbf{P}_\theta|)}{|\mathbf{P}_\theta|}\,d\theta &\preccurlyeq \int_0^1 \big(\rho'(|\mathbf{P}_1|+|\mathbf{P}_2|)\big)^{\frac{1}{\gamma}}\,|\mathbf{P}_\theta|^{\frac{1}{\gamma}-2}\,d\theta\\ &= \big(\rho'(|\mathbf{P}_1|+|\mathbf{P}_2|)\big)^{\frac{1}{\gamma}}\int_0^1 |\mathbf{P}_\theta|^{\frac{1}{\gamma}-2}\,d\theta\\ &\preccurlyeq \big(\rho'(|\mathbf{P}_1|+|\mathbf{P}_2|)\big)^{\frac{1}{\gamma}}(|\mathbf{P}_1|+|\mathbf{P}_2|)^{\frac{1}{\gamma}-2}\\ &\preccurlyeq \frac{\phi'(|\mathbf{P}_1|+|\mathbf{P}_2|)}{(|\mathbf{P}_1|+|\mathbf{P}_2|)}.\end{aligned}$$

This completes the proof. □

We are now prepared to prove Proposition 43.

Proof of Proposition 43. We define $\Phi(\mathbf{Q}) := \phi(|\mathbf{Q}|)$, $\mathbf{Q}\in\mathbb{R}^{d\times d}$. Recall from Definition 5 that $\phi'(0)=0$. We denote $\mathbf{Q}=(Q_{ij})_{i,j=1,\dots,d}, \mathbf{P}=(P_{ij})_{i,j=1,\dots,d}\in\mathbb{R}^{d\times d}$, as well as $\mathbf{A}(\mathbf{Q})=(A_{ij}(\mathbf{Q}))_{i,j=1,\dots,d}\in\mathbb{R}^{d\times d}$. Let further D_{ij} be the partial derivative in direction of the ij-th matrix component and $\mathbf{D}=(D_{ij})_{i,j=1,\dots,d}$. Observe that

$$(D_{ij}\Phi)(\mathbf{Q})=\phi'(|\mathbf{Q}|)\frac{Q_{ij}}{|\mathbf{Q}|},$$

and

$$(D_{ij}D_{kl}\Phi)(\mathbf{Q})=\phi'(|\mathbf{Q}|)\Big(\frac{\delta_{ik}\delta_{jl}}{|\mathbf{Q}|}-\frac{Q_{ij}Q_{kl}}{|\mathbf{Q}|^3}\Big)+\phi''(|\mathbf{Q}|)\frac{Q_{ij}}{|\mathbf{Q}|}\frac{Q_{kl}}{|\mathbf{Q}|}. \tag{3.3}$$

We assume $[\mathbf{Q},\mathbf{P}]_t=(1-t)\,\mathbf{Q}+t\,\mathbf{P}\neq 0$ for all $t\in[0,1]$. Since $\phi\in C^2((0,\infty))$, according to Assumption 40, it holds

$$\begin{aligned}A_{ij}(\mathbf{P})-A_{ij}(\mathbf{Q}) &= (D_{ij}\Phi)(\mathbf{P})-(D_{ij}\Phi)(\mathbf{Q})\\ &= \sum_{k,l=1}^d \int_0^1 (D_{ij}D_{kl}\Phi)([\mathbf{Q},\mathbf{P}]_t)(P_{kl}-Q_{kl})\,dt.\end{aligned} \tag{3.4}$$

Lemma 46 and Assumption 40 yield

$$\begin{aligned}|\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})| &\preccurlyeq \int_0^1 \frac{\phi'(|[\mathbf{Q},\mathbf{P}]_t|)}{|[\mathbf{Q},\mathbf{P}]_t|}\,dt\ |\mathbf{P}-\mathbf{Q}| \\ &\preccurlyeq \frac{\phi'(|\mathbf{P}|+|\mathbf{Q}|)}{|\mathbf{P}|+|\mathbf{Q}|}\,|\mathbf{P}-\mathbf{Q}| \preccurlyeq \phi''(|\mathbf{P}|+|\mathbf{Q}|)\,|\mathbf{P}-\mathbf{Q}|\,.\end{aligned}$$

This proves the second assertion. On the other hand due to Assumption 40 there exists $c\in(0,1)$ such that $\phi'(t)\geq c\,\phi''(t)t$. Therefore, (3.4) and (3.3) imply

$$\begin{aligned}(\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})):(\mathbf{P}-\mathbf{Q}) &= \int_0^1 \frac{\phi'(|[\mathbf{Q},\mathbf{P}]_t|)}{|[\mathbf{Q},\mathbf{P}]_t|}\Big(\,|\mathbf{P}-\mathbf{Q}|^2-\frac{|(\mathbf{P}-\mathbf{Q}):[\mathbf{Q},\mathbf{P}]_t|^2}{|[\mathbf{Q},\mathbf{P}]_t|^2}\Big) \\ &\quad + \phi''(|[\mathbf{Q},\mathbf{P}]_t|)^2\frac{|(\mathbf{P}-\mathbf{Q}):[\mathbf{Q},\mathbf{P}]_t|^2}{|[\mathbf{Q},\mathbf{P}]_t|^2}\,dt \\ &\geq \int_0^1 c\,\phi''(|[\mathbf{Q},\mathbf{P}]_t|)\Big(\,|\mathbf{P}-\mathbf{Q}|^2-\frac{|(\mathbf{P}-\mathbf{Q}):[\mathbf{Q},\mathbf{P}]_t|^2}{|[\mathbf{Q},\mathbf{P}]_t|^2}\Big) \\ &\quad + \phi''(|[\mathbf{Q},\mathbf{P}]_t|)\frac{|(\mathbf{P}-\mathbf{Q}):[\mathbf{Q},\mathbf{P}]_t|^2}{|[\mathbf{Q},\mathbf{P}]_t|^2}\,dt \\ &\geq c\int_0^1 \phi''(|[\mathbf{Q},\mathbf{P}]_t|)\,|\mathbf{P}-\mathbf{Q}|^2\ dt.\end{aligned}$$

Note that we made use of the Cauchy-Schwartz inequality to obtain $|\mathbf{R}|^2-\frac{|\mathbf{RS}|^2}{|\mathbf{S}|^2}\geq 0$ for $\mathbf{R},\mathbf{S}\in\mathbb{R}^{d\times d}$ in the above estimate. Assumption 40 and Lemma 46 yield again that

$$\begin{aligned}(\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})):(\mathbf{P}-\mathbf{Q}) &\succcurlyeq \int_0^1 \frac{\phi'(|[\mathbf{Q},\mathbf{P}]_t|)}{|[\mathbf{Q},\mathbf{P}]_t|}\,|\mathbf{P}-\mathbf{Q}|^2\ dt \\ &\approx \frac{\phi'(|\mathbf{P}|+|\mathbf{Q}|)}{|\mathbf{Q}|+|\mathbf{P}|}\,|\mathbf{P}-\mathbf{Q}|^2 \\ &\approx \phi''(|\mathbf{P}|+|\mathbf{Q}|)\,|\mathbf{P}-\mathbf{Q}|^2\,.\end{aligned}\tag{3.5}$$

Hence, the assertion is established in the case $[\mathbf{Q},\mathbf{P}]_t\neq 0$ for all $t\in[0,1]$. We observe that both sides are continuous in $\mathbf{P}$ and $\mathbf{Q}$. For $\mathbf{P}=\mathbf{Q}=0$ the assertion is obvious, hence for arbitrary $\mathbf{P},\mathbf{Q}$ we may assume, w.l.o.g., that $\mathbf{P}\neq 0$. Then there exists a sequence $(\mathbf{Q}_n)_{n\in\mathbb{N}}\subset\mathbb{R}^{d\times d}$ that converges to $\mathbf{Q}$ such that $[\mathbf{Q}_n,\mathbf{P}]_t\neq 0$ for all $t\in[0,1]$ and $n\in\mathbb{N}$. Therefore, it holds (3.5) and hence

$$\begin{array}{ccc}(\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q}_n)):(\mathbf{P}-\mathbf{Q}_n) & \succcurlyeq & \phi''(|\mathbf{P}|+|\mathbf{Q}_n|)\,|\mathbf{P}-\mathbf{Q}_n|^2 \\ \Big\downarrow n\to\infty & & \Big\downarrow n\to\infty \\ (\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})):(\mathbf{P}-\mathbf{Q}) & \succcurlyeq & \phi''(|\mathbf{P}|+|\mathbf{Q}|)\,|\mathbf{P}-\mathbf{Q}|^2\,.\end{array}$$

Hence, the assertion is proved for all $\mathbf{P},\mathbf{Q}\in\mathbb{R}^{d\times d}$. □

Remark 47. *Note that in the case* $\phi(t) = \frac{1}{r}t^r$ *with* $r \in (1,\infty)$ *Lemma 45 leads to the sharp estimates*

$$\begin{aligned}(\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})) : (\mathbf{P} - \mathbf{Q}) &\geq c(r)\,(|\mathbf{P}| + |\mathbf{Q}|)^{r-2}\,|\mathbf{P} - \mathbf{Q}|^2\,,\\ |\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})| &\leq C(r)\,(|\mathbf{P}| + |\mathbf{Q}|)^{r-2}\,|\mathbf{P} - \mathbf{Q}|\,,\end{aligned}$$

with $c = \min\{2^{2-r}, (r-1)\,2^{2-r}\}$ *and* $C = \max\{1, 2^{2-r}, (r-1)\,2^{2-r}\}$*; see also [25, 17].*

As a consequence of Proposition 43 we get the following result.

Lemma 48. *Let* ϕ *be an N-function satisfying Assumption 40. Then the Operator*

$$-\operatorname{div}\mathbf{A}(\nabla\cdot) : W_0^{1,\phi}(\Omega)^d \to W^{-1,\phi^*}(\Omega)^d$$

is continuous, strictly monotone, and coercive.

Proof. We start with proving the continuity. Let $(v_n)_{n\in\mathbb{N}} \subset W_0^{1,\phi}(\Omega)^d$ such that $v_n \to v \in W_0^{1,\phi}(\Omega)^d$ as $n \to \infty$. It follows from Assumption 40 and (2.8) that

$$\phi''(|\nabla v_n| + |\nabla v|)\,|\nabla v_n - \nabla v| \preccurlyeq \phi'(|\nabla v_n| + |\nabla v|) \in L^{\phi^*}(\Omega). \tag{3.6}$$

Thus, Proposition 43, Proposition 25 and Corollary 36 imply that it suffices to prove

$$\phi''(|\nabla v_n| + |\nabla v|)\,|\nabla v_n - \nabla v| \to_{n\to\infty} 0 \qquad \text{in } L^{\phi^*}(\Omega).$$

Lebesgue measure theory yields the existence of a subsequence $(v_{n_k})_{k\in\mathbb{N}} \subset (v_n)_{n\in\mathbb{N}}$ such that $\nabla v_{n_k} \to_{k\to\infty} \nabla v$ a.e. in Ω; see, e.g., [23, Propositions 3.1.4 and 3.1.2]. Since $\phi'' : (0,\infty) \to (0,\infty)$ is continuous, it follows that

$$\phi''(|\nabla v_{n_k}| + |\nabla v|)\,|\nabla v_{n_k} - \nabla v| \to_{k\to\infty} 0 \qquad \text{a.e. in } \Omega.$$

Note that for $\nabla v = 0$, the statement follows with the continuous extension $t\,\phi''(t) = 0$ for $t = 0$; see Assumption 40. We have by (3.6) that $\phi'(|\nabla v_{n_k}| + |\nabla v|)$ is up to a constant a majorizing sequence of $\phi''(|\nabla v_{n_k}| + |\nabla v|)\,|\nabla v_{n_k} - \nabla v|$ and therefore it holds with (2.8) and mean-convergence

$$\int_\Omega \phi^*\big(\phi'(|\nabla v_{n_k}| + |\nabla v|)\big)\,dx \eqsim \int_\Omega \phi(|\nabla v_{n_k}| + |\nabla v|)\,dx \to \int_\Omega \phi(2\,|\nabla v|)\,dx,$$

as $k \to \infty$. Now, a generalized version of Lebesgue's majorized convergence theorem [81, Appendix (19a)] implies that

$$\phi''(|\nabla v_{n_k}| + |\nabla v|)\,|\nabla v_{n_k} - \nabla v| \to_{k\to\infty} 0 \qquad \text{in } L^{\phi^*}(\Omega). \tag{3.7}$$

The assertion for the whole sequence follows by assuming that there exists a subsequence $(v_{n_l})_{l\in\mathbb{N}} \subset (v_n)_{n\in\mathbb{N}}$ such that $\phi''(|\nabla v_{n_l}| + |\nabla v|)\,|\nabla v_{n_l} - \nabla v|$ is bounded away from zero in $L^{\phi^*}(\Omega)$. Recalling that $v_{n_l} \to v$ in $W_0^{1,\phi}(\Omega)^d$ as $l \to \infty$, the above calculations prove that a subsequence of $(v_{n_l})_{l\in\mathbb{N}}$ satisfies (3.7), which is a contradiction.

It is clear from Proposition 43 that $-\operatorname{div} \mathbf{A}\nabla$ is a monotone operator. However, in order to prove strict monotonicity we notice that Proposition 43 yields

$$\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla v)\big) : \big(\nabla u - \nabla v\big)\, dx \succcurlyeq \int_\Omega \phi''(|\nabla u| + |\nabla v|)\, |\nabla u - \nabla v|^2\, dx,$$

for $u, v \in W_0^{1,\phi}(\Omega)^d$. If we now assume the left hand side to be zero, we obtain that

$$\phi''(|\nabla u| + |\nabla v|)\, |\nabla u - \nabla v|^2 = 0 \qquad \text{a.e. in } \Omega,$$

which in turn implies $\nabla u = \nabla v$ a.e. in Ω. Hence, with Corollary 36 it follows $u = v$ in $W_0^{1,\phi}(\Omega)^d$.

It remains to prove the coercivity of $-\operatorname{div} \mathbf{A}(\nabla\cdot)$. Due to Lemma 18 there exists $\gamma \in (0,1)$ and an N-function ρ with $\Delta_2(\{\rho,\rho^*\}) < \infty$ such that $\phi^\gamma \approx \rho$. Recalling the definition of $\|\cdot\|_{(\phi)}$ we get from [51]

$$1 = \int_\Omega \phi\Big(\frac{|\nabla v|}{\|\nabla v\|_{(\phi)}}\Big)\, dx \approx \int_\Omega \rho\Big(\frac{|\nabla v|}{\|\nabla v\|_{(\phi)}}\Big)^{\frac{1}{\gamma}}\, dx;$$

see Remark 23. Since we want to consider the limit $\|\nabla v\|_{(\phi)} \to \infty$, we may assume that $\|\nabla v\|_{(\phi)} > 1$. Then it follows from (2.6a) that

$$1 \preccurlyeq \int_\Omega \Big(\frac{\rho(|\nabla v|)}{\|\nabla v\|_{(\phi)}}\Big)^{\frac{1}{\gamma}}\, dx \approx \int_\Omega \frac{\phi(|\nabla v|)}{\|\nabla v\|_{(\phi)}^{\frac{1}{\gamma}}}\, dx.$$

Thus, with the definition of $\mathbf{A}$ and Proposition 14 we have

$$\int_\Omega \frac{\mathbf{A}(\nabla v) : \nabla v}{\|\nabla v\|_{(\phi)}}\, dx = \int_\Omega \frac{\phi'(|\nabla v|)\, |\nabla v|}{\|\nabla v\|_{(\phi)}}\, dx \approx \int_\Omega \frac{\phi(|\nabla v|)}{\|\nabla v\|_{(\phi)}}\, dx \succcurlyeq \|\nabla v\|_{(\phi)}^{\frac{1}{\gamma}-1} \to \infty,$$

as $\|\nabla v\|_{(\phi)} \to \infty$. This proves coercivity and thus the Lemma. □

Now, the well established theory of monotone operators yields the existence and uniqueness of a solution.

Theorem 49. *Let ϕ be an N-function that satisfies Assumption 40. Then there exists a unique solution $u \in W_0^{1,\phi}(\Omega)^d$ of (3.2).*

Proof. The assertion follows from the theory of monotone operators. In particular, as $-\operatorname{div} \mathbf{A}(\nabla\cdot) : W_0^{1,\phi}(\Omega)^d \to W^{-1,\phi^*}(\Omega)^d$ is continuous, strictly monotone and coercive (see Lemma 48), the existence of a solution follows from the Minty-Browder Theorem; see e.g. [69, Theorem II.2.2] or [81, Theorem 26.A]. The uniqueness is a consequence of the strict monotonicity: Suppose that there exists a second solution $u \neq v \in W_0^{1,\phi}(\Omega)^d$ of (3.2), then

$$0 = \langle g - g, u - v\rangle = \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla v)\big) : (\nabla u - \nabla v)\, dx > 0. \tag{3.8}$$

This is a contradiction. □

Let $X \subset W_0^{1,\phi}(\Omega)^d$ be a (not necessarily finite dimensional) closed sub-space. Note that by $\langle y, x\rangle_{X^*\times X} := \langle y, x\rangle_{W^{-1,\phi^*}(\Omega)^d \times W_0^{1,\phi}(\Omega)^d}$ for $y \in W^{-1,\phi^*}(\Omega)^d$, $x \in X$, each linear functional on $W_0^{1,\phi}(\Omega)^d$ defines a linear functional on X. Thus for $g \in W^{-1,\phi^*}(\Omega)$ we can define the weak sub-problem of (3.2):

Find $U \in X$ such that

$$\int_\Omega \mathbf{A}(\nabla U) : \nabla V\, dx = \langle g, V\rangle \qquad \text{for all} \quad V \in X. \tag{3.9}$$

Since the properties of the nonlinear operator $-\operatorname{div}\mathbf{A}(\nabla\cdot)$ of Lemma 48 carry over to any closed sub-space $X \subset W_0^{1,\phi}(\Omega)^d$ and $W^{-1,\phi^*}(\Omega)^d \subset X^*$ we get the following corollary analogously to Theorem 49.

Corollary 50. *Let $X \subset W_0^{1,\phi}(\Omega)^d$ be a closed sub-space. Then problem* (3.9) *possesses a unique solution $U \in X$.*

Remark 51. *Note that existence and uniqueness results for more general nonlinearities are available; see, e.g., [16, 34]. In both works nonlinearities are considered that in general lead to non-reflexive Orlicz-Sobolev spaces, which is equivalent to $\Delta_2(\{\phi, \phi^*\}) = \infty$; see Proposition 37. In the sequel we will see that the Δ_2-condition however is crucial for lots of estimates that are important for numerical analysis.*

3.1.3 The Energy Functional

We establish an energy functional whose unique extremal point is the weak solution of (3.2). In particular, let ϕ be an N-function that satisfies Assumption 40 and let $g \in W^{-1,\phi^*}(\Omega)^d$. We define the functional $\mathcal{J} : W_0^{1,\phi}(\Omega)^d \to \mathbb{R}$ by

$$\mathcal{J}(v) := \int_\Omega \phi(|\nabla v|)\, dx - \langle g, v\rangle, \qquad v \in W_0^{1,\phi}(\Omega)^d. \tag{3.10}$$

From the definition of Orlicz-Sobolev spaces and Remark 22 it is clear that the energy functional is well-defined. In the following we are concerned in finding a minimizer $u \in W_0^{1,\phi}(\Omega)^d$ of $\mathcal{J}$, i.e.,

$$\inf_{v \in W_0^{1,\phi}(\Omega)^d} \mathcal{J}(v) = \mathcal{J}(u). \tag{3.11}$$

First we state the connection of the minimizing problem (3.11) to the PDE (3.1).

Proposition 52. *Let ϕ be an N-function that satisfies Assumption 40, then the energy functional defined in* (3.11) *is Fréchet differentiable with derivative*

$$\mathcal{J}'(v) = -\operatorname{div} \mathbf{A}(\nabla v) - g \in W^{-1,\phi^*}(\Omega),$$

in $v \in W_0^{1,\phi}(\Omega)$.

Proof. Since the proof is standard, we just list its basic ideas. We know from Lemma 48 that the functional $\mathcal{J}' : W_0^{1,\phi}(\Omega) \to W^{-1,\phi^*}(\Omega)$ is continuous. Hence it suffices to prove that $\mathcal{J}$ is Gâteaux differentiable with derivative $\mathcal{J}'(v)$ in $v \in W_0^{1,\phi}(\Omega)$; see [80, Chapter 4]. We restrict ourselves to the nonlinear part of $\mathcal{J}$ since the assertion for the linear part g is obvious. First, we observe that for $h \in W_0^{1,\phi}(\Omega)$

$$\frac{\phi(|\nabla(v + th)|) - \phi(|\nabla v|)}{t} \longrightarrow \mathbf{A}(\nabla v) : \nabla h \qquad \text{a.e. in } \Omega,$$

as $t \to 0$. In order to find an integrable majorant for this difference quotient, we observe that by the monotonicity of ϕ' it holds

$$\begin{aligned}
|\phi(|\nabla(v + th)|) - \phi(|\nabla v|)| &\leq \int_0^t \phi'(|\nabla(v + sh)|)\, |\nabla h| \, ds \\
&\leq \int_0^t \phi'(|\nabla v| + s\, |\nabla h|)\, |\nabla h| \, ds \\
&\leq t\, \phi'(|\nabla v| + |\nabla h|)\, |\nabla h| \,,
\end{aligned}$$

for $t \leq 1$. Therefore an integrable majorant for the above difference quotient is given by $\phi'(|\nabla v| + |\nabla h|)\, |\nabla h|$. Hence by Lebesgue's majorized convergence theorem

$$\frac{\phi(|\nabla(v + th)|) - \phi(|\nabla v|)}{t} \longrightarrow \mathbf{A}(\nabla v) : \nabla h \qquad \text{in } L^1(\Omega),$$

as $t \to 0$, which is the desired assertion. □

Knowing about the derivative of $\mathcal{J}$ we can at once deduce the next corollary from Lemma 48; see also [79, Proposition 42.6].

Corollary 53. *Under the assumptions of Proposition 52 the energy functional $\mathcal{J}$ is continuous, strictly convex and coercive.*

This in turn implies the existence of a minimizer of $\mathcal{J}$ as well as its uniqueness.

Theorem 54. *Let ϕ be an N-function that satisfies Assumption 40. Then, the minimizing problem* (3.11) *possesses a unique solution. Moreover, the minimizer is the solution of* (3.2).

Proof. Since direct methods for variational problems are somehow standard in nonlinear analysis we only sketch the proof providing precise information where to find the used assertions in literature. The convexity and continuity of $\mathcal{J}$ imply that $\mathcal{J}$ is weak sequentially lower semi-continuous; see [79, Proposition 38.7] or [45, Theorem 4.3]. Together with the coercivity of $\mathcal{J}$ this implies the existence of a solution; cf. [79, Proposition 38.15] or [45, Theorem 4.6]. The uniqueness follows from the strict convexity of $\mathcal{J}$; see [79, Theorem 38C].

By Proposition 52 the minimal function is the solution of (3.2) since a minimal point of a potential is a critical point of its linearization. The one to one correspondence follows from the uniqueness of the solution of (3.2); see Proposition 49. □

Since continuity, convexity, and coercivity are inherited by any closed subspace of $W_0^{1,\phi}(\Omega)$ there exists a unique minimizer of $\mathcal{J}$ in those spaces as well.

Corollary 55. *Under the conditions of Theorem 54 let $X \subset W_0^{1,\phi}(\Omega)$ be a closed sub-space. Let $\mathcal{J}_X : X \to \mathbb{R}$ be the restriction of $\mathcal{J}$ to X. Then there exists a unique minimizer $U \in X$ of $\mathcal{J}_X$. Moreover, the minimizer is the solution of* (3.9).

3.2 Concept of Distance

In 1993 Barrett and Liu introduced an new error concept for the nonlinear Laplacian; see [8, 9]. In particular, they introduced an error notion called *quasi-norm*, which is directly related to the residual of the problem; see, e.g., Remark 79. The concept of distance presented in this section is a generalization of the quasi-norm from [26, 31], and [32].

3.2.1 Shifted N-functions

A modified N-function called *shifted N-function* turned out to be very useful for a generalization of the quasi-norm concept to the case of N-functions.

Definition 56 (Shifted N-functions)**.** *Let ϕ be an N-function with $\Delta_2(\phi) < \infty$. For given $a \geq 0$ we define*

$$\phi_a'(t) := \frac{\phi'(a+t)}{a+t}\, t \quad \textit{and} \quad \phi_a(t) := \int_0^t \phi_a'(s)\, ds.$$

In the following we state some properties of shifted N-functions, which are crucial in the subsequent analysis.

Lemma 57. *Let ϕ be an N-function with $\Delta_2(\phi) < \infty$. The function ϕ_a is an N-function for all $a \geq 0$ and it holds $\Delta_2(\{\phi_a\}_{a\geq 0}) \leq 2\,\Delta_2(\phi)^2$, i.e., the family $(\phi_a)_{a\geq 0}$ satisfies a Δ_2-condition uniformly in $a \geq 0$.*

Proof. We fix $a \geq 0$. Since ϕ is an N-function, $\phi'(a + \cdot)$ is non decreasing and right continuous with $\phi'(a+t) \to \infty$ as $t \to \infty$. Moreover, $\frac{t}{a+t}$ is increasing and continuous and obviously $\phi_a'(0) = 0$. Thus, ϕ_a is an N-function. It remains to prove the Δ_2-condition. Together with Corollary 17 we get

$$\begin{aligned}\phi_a(2t) &= \int_0^t \frac{\phi'(a+2s)}{a+2s}\, 4s\, ds \leq \int_0^t \frac{\phi'(2a+2s)}{(a+s)}\, 4s\, ds \\ &\leq \frac{\Delta_2(\phi)^2}{2} \int_0^t \frac{\phi'(a+s)}{(a+s)}\, 4s\, ds = 2\, \Delta_2(\phi)^2\, \phi_a(t),\end{aligned}$$

which is the desired assertion. □

Lemma 58. *Let ϕ be an N-function with $\Delta_2(\phi) < \infty$. Then for any $a, b \geq 0$ it holds*

$$(\phi_a')_b(t) = \phi_{a+b}'(t) \qquad \text{for all } t \geq 0.$$

Proof. With Definition 56 we have $\Delta_2(\phi_a) < \infty$ and thus the left hand side is well defined. Furthermore,

$$(\phi_a)_b'(t) = \frac{\phi_a'(b+t)}{b+t}\, t = \frac{\phi'(a+b+t)}{a+b+t}\, t = \phi_{a+b}'(t),$$

which yields the assertion. □

Lemma 59. *Let ϕ be an N-function with $\Delta_2(\phi) < \infty$. Assume further that $0 \leq t \leq \Lambda\, a$ for $a, \Lambda > 0$. Then there exists $C > 0$ depending solely on Λ and $\Delta_2(\phi)$ such that for all $\alpha \leq 1$*

$$\phi_a(\alpha t) \leq \alpha^2\, C\, \phi_a(t).$$

Proof. By the definition of shifted N-functions Definition 56 it holds with $\frac{t+a}{1+\Lambda} \leq a$ that

$$\begin{aligned}\phi_a'(\alpha t) &= \frac{\phi'(a+\alpha\, t)}{a+\alpha\, t}\, \alpha\, t \leq \frac{\phi'(a+t)}{a}\, \alpha\, t \\ &\leq \alpha\, (1+\Lambda)\, \frac{\phi'(a+t)}{a+t} t = \alpha\, (1+\Lambda)\, \phi_a'(t).\end{aligned}$$

Now, the assertion follows with Corollary 15. □

The next lemma gives some information about what the dual function of a shifted N-function looks like.

Lemma 60. *Let ϕ be an N-function with $\Delta_2(\{\phi, \phi^*\}) < \infty$. Then there exist constants $c, C > 0$ depending solely on $\Delta_2(\{\phi, \phi^*\})$ such that for all $a \geq 0$*

$$c\, (\phi^*)_{\phi'(a)}(t) \leq (\phi_a)^*(t) \leq C\, (\phi^*)_{\phi'(a)}(t) \qquad \text{for all } t \geq 0.$$

Proof. We assume that ϕ satisfies Assumption 40 in order to avoid some technical complications. The proof for the general case can be found in [32]. Therefore, ϕ is continuous and its inverse function exists; see Remark 42. The case $a = 0$ is obvious, therefore we concentrate on $a > 0$. We start with estimating $(\phi^*)_{\phi'(a)}(\phi'_a(t))$ by distinguishing two cases, namely $t \leq a$ and $t > a$. In the first case we have $a \leq a + t \leq 2a$ and hence the monotonicity of ϕ' and Corollary 17 imply

$$\frac{\phi'(a+t)}{a+t} \leq \frac{\phi'(2a)}{a} \leq \Delta_2(\phi)^2 \, \frac{\phi'(a)}{2a}.$$

This, the definition of shifted N-functions, and Corollary 15 imply

$$\phi'(a) + \phi'_a(t) = \phi'(a) + \frac{\phi'(a+t)}{a+t}\, t \leq \phi'(a) + \frac{\phi'(a+t)}{a+t}\, a \lesssim \phi'(a).$$

Hence, with the obvious estimate $\phi'(a) + \phi'_a(t) \geq \phi'(a)$

$$\phi'(a) + \phi'_a(t) \approx \phi'(a)$$

Furthermore,

$$\phi'_a(t) = \frac{\phi'(a+t)}{a+t}\, t \lesssim \frac{\phi'(a)}{a}\, t$$

and

$$\frac{\phi'(a)}{a}\, t \lesssim \frac{\phi'(a)}{2a}\, t \leq \frac{\phi'(a+t)}{a+t}\, t = \phi'_a(t).$$

Using the definition of shifted N-functions, we get with Corollary 17

$$\begin{aligned}(\phi^*)'_{\phi'(a)}(\phi'_a(t)) &= \frac{(\phi^*)'\big(\phi'(a) + \phi'_a(t)\big)}{\phi'(a) + \phi'_a(t)}\, \phi'_a(t) \approx \frac{(\phi^*)'\big(\phi'(a)\big)}{\phi'(a)}\, \phi'_a(t) \\ &\approx \frac{(\phi^*)'\big(\phi'(a)\big)}{\phi'(a)}\, \frac{\phi'(a)}{a}\, t = \frac{(\phi^*)'\big(\phi'(a)\big)}{a}\, t.\end{aligned}$$

Recalling (2.1), i.e., $(\phi^*)' = (\phi')^{-1}$ yields

$$(\phi^*)'_{\phi'(a)}(\phi'_a(t)) \approx t. \tag{3.12}$$

In the second case, i.e., for $a < t$ it holds $t < a + t < 2t$, i.e., $t \approx a + t$. Therefore, we get with the monotonicity of ϕ' and Corollary 17

$$\phi'_a(t) = \frac{\phi'(a+t)}{a+t} t \leq \frac{\Delta_2(\phi)^2}{2} \frac{\phi'(t)}{t}\, t \lesssim \phi'(t).$$

On the other hand it holds

$$\phi'(t) = 2\frac{\phi'(t)}{2t}\,t \leq 2\frac{\phi'(t)}{a+t}\,t \leq 2\frac{\phi'(a+t)}{a+t}\,t = 2\phi_a'(t)$$

and hence

$$\phi'(t) \approx \phi_a'(t).$$

Now, the monotonicity of ϕ' yields $\phi'(a) \leq \phi'(t)$ and therefore

$$\phi'(a) + \phi_a'(t) \approx \phi'(a) + \phi'(t) \approx \phi'(t) \approx \phi_a'(t) \preccurlyeq \phi'(a) + \phi_a'(t).$$

With similar arguments as in the above case, this gives with Corollary 17

$$(\phi^*)_{\phi'(a)}'(\phi_a'(t)) = \frac{(\phi^*)'\big(\phi'(a) + \phi_a'(t)\big)}{\phi'(a) + \phi_a'(t)}\,\phi_a'(t) \approx \frac{(\phi^*)'\big(\phi'(t)\big)}{\phi'(t)}\,\phi'(t) = t.$$

Thus (3.12) holds for all $t \geq 0$ and hence with Corollary 15 we have for all $t \geq 0$

$$(\phi^*)_{\phi'(a)}\big(\phi_a'(t)\big) \approx (\phi^*)_{\phi'(a)}'\big(\phi_a'(t)\big)\,\phi_a'(t) \approx t\,\phi_a'(t) \approx \phi_a(t) \approx (\phi_a)^*\big(\phi_a'(t)\big),$$

where the last $\approx$ follows from (2.8). Since ϕ_a' is continuous, $\phi_a'(0) = 0$, and $\lim_{t\to\infty} \phi_a'(t) = \infty$, it follows that $\phi_a' : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is surjective and hence substituting $s = \phi_a'(t)$ completes the proof. □

Remark 61. *Let ϕ be an N-function with $\Delta_2(\{\phi, \phi^*\}) < \infty$. We observe from Lemma 57 and Lemma 60 that $(\phi^*)_{\phi'(a)}$ as well as $(\phi_a)^*$ are N-functions with $\Delta_2(\{(\phi^*)_{\phi'(a)}, \big((\phi^*)_{\phi'(a)}\big)^*, (\phi_a)^*, \phi_a\}) < \infty$ depending only on $\Delta_2(\{\phi, \phi^*\})$. Therefore, Corollary 15 holds for all those functions and thus Lemma 60 implies*

$$\big((\phi^*)_{\phi'(a)}\big)'(t) \approx \frac{(\phi^*)_{\phi'(a)}(t)}{t} \approx \frac{(\phi_a)^*(t)}{t} \approx \big((\phi_a)^*\big)'(t). \tag{3.13}$$

We will now introduce some quantities related to shifted N-functions. In particular, we introduce a vector field $\mathbf{F} : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ defined by

$$\mathbf{F}(\mathbf{Q}) := \sqrt{|\mathbf{A}(\mathbf{Q})|\,|\mathbf{Q}|}\frac{\mathbf{Q}}{|\mathbf{Q}|} \approx \sqrt{\phi(|\mathbf{Q}|)}\frac{\mathbf{Q}}{|\mathbf{Q}|} \tag{3.14}$$

for $\mathbf{Q} \in \mathbb{R}^{d\times d}$. The vector-field $\mathbf{F} : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ is bijective since ϕ is strictly monotone increasing. Furthermore, it is related to an N-function ψ defined by $\psi(t) := \sqrt{\phi'(t)t}$ as $\mathbf{A}$ is related to ϕ; see [26, 31]. The vector field $\mathbf{F}$ transforms L^ϕ-functions into L^2-functions. The connection between $\mathbf{A}$, $\mathbf{F}$, and $\{\phi_a\}_{a\geq 0}$ is best reflected in the following result from [26].

Proposition 62. *Let ϕ be an N-function that satisfies Assumption 40. Then, for all $\mathbf{Q}, \mathbf{P} \in \mathbb{R}^{d\times d}$ it holds*

$$\begin{aligned}\big(\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})\big) : \big(\mathbf{P} - \mathbf{Q}\big) &\approx \phi_{|\mathbf{P}|}(|\mathbf{P} - \mathbf{Q}|) \approx |\mathbf{F}(\mathbf{P}) - \mathbf{F}(\mathbf{Q})|^2 \\ &\approx \phi''(|\mathbf{P}| + |\mathbf{Q}|)\, |\mathbf{P} - \mathbf{Q}|^2 .\end{aligned}$$

The constants hidden in $\approx$ depend solely on $\Delta_2(\{\phi, \phi^*\})$ and the constants in Assumption 40.

Proof. To prove the first estimate we recall from Proposition 43 that

$$\big(\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})\big) : \big(\mathbf{P} - \mathbf{Q}\big) \approx \phi''(|\mathbf{P}| + |\mathbf{Q}|)\, |\mathbf{P} - \mathbf{Q}|^2 .$$

Assumption 40, the fact that $\frac{1}{2}(|\mathbf{P}| + |\mathbf{P} - \mathbf{Q}|) \leq |\mathbf{P}| + |\mathbf{Q}| \leq 2(|\mathbf{P}| + |\mathbf{P} - \mathbf{Q}|)$, and $\Delta_2(\phi) < \infty$ give

$$\begin{aligned}\big(\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})\big) : \big(\mathbf{P} - \mathbf{Q}\big) &\approx \phi''(|\mathbf{P}| + |\mathbf{Q}|)\, |\mathbf{P} - \mathbf{Q}|^2 \\ &\approx \frac{\phi'(|\mathbf{P}| + |\mathbf{Q}|)}{|\mathbf{P}| + |\mathbf{Q}|}\, |\mathbf{P} - \mathbf{Q}|^2 \\ &\approx \frac{\phi'(|\mathbf{P}| + |\mathbf{P} - \mathbf{Q}|)}{|\mathbf{P}| + |\mathbf{P} - \mathbf{Q}|}\, |\mathbf{P} - \mathbf{Q}|^2 \\ &= \phi'_{|\mathbf{P}|}(|\mathbf{P} - \mathbf{Q}|)\, |\mathbf{P} - \mathbf{Q}| \approx \phi_{|\mathbf{P}|}(|\mathbf{P} - \mathbf{Q}|).\end{aligned}$$

To prove the second estimate we observe that $\psi'(t) := \sqrt{\phi'(t)t}$ defines an N-function with $\Delta_2(\{\psi, \psi^*\}) < \infty$ solely depending on $\Delta_2(\{\phi, \phi^*\})$. Furthermore, ψ satisfies Assumption 40 with the constants therein depending only on the respective constants for ϕ; c.f. also [26, 31]. By the definition of $\mathbf{F}$ we have for $\mathbf{Q} \in \mathbb{R}^{d\times d}$ that $\mathbf{F}(\mathbf{Q}) = \psi(|\mathbf{Q}|)\frac{\mathbf{Q}}{|\mathbf{Q}|}$ and therefore Proposition 43 holds for $\mathbf{A}$ and ϕ replaced by $\mathbf{F}$ and ψ. Moreover, observe that $\psi''(t) \approx \sqrt{\phi''(t)}$ for all $t \geq 0$ and thus

$$|\mathbf{F}(\mathbf{P}) - \mathbf{F}(\mathbf{Q})| \approx \psi''(|\mathbf{P}| + |\mathbf{Q}|)\, |\mathbf{P} - \mathbf{Q}| \approx \sqrt{\phi''(|\mathbf{P}| + |\mathbf{Q}|)}\, |\mathbf{P} - \mathbf{Q}| .$$

Applying Proposition 43 proves the lemma. □

Remark 63. *Recalling our standard example $\phi'(t) = \frac{1}{r}\, t^r$, $r > 1$, then*

$$\phi'_{|\mathbf{P}|}(t) = \frac{(|\mathbf{P}| + t)^{r-1}}{|\mathbf{P}| + t}\, t = (|\mathbf{P}| + t)^{r-2}\, t = \frac{1}{r-1}\, \phi''(|\mathbf{P}| + t)\, t.$$

Therefore, the estimates of Proposition 62 correspond to the basic estimates of Barrett and Liu [8, 9]; see also Remark 44.

Corollary 64. *Under the assumptions of Proposition 62 it holds*

$$|\mathbf{A}(\mathbf{P}) - \mathbf{A}(\mathbf{Q})| \approx \phi'_{|\mathbf{P}|}(|\mathbf{P} - \mathbf{Q}|).$$

Proof. The estimate

$$|\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})| \succcurlyeq \phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|)$$

follows immediately from Corollary 15 and Proposition 62. For the converse estimate the second estimate of Proposition 43 states

$$|\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})| \preccurlyeq \phi''(|\mathbf{Q}|+|\mathbf{P}|)\,|\mathbf{P}-\mathbf{Q}|\,.$$

Recalling Assumption 40, then

$$\phi''(|\mathbf{P}|+|\mathbf{Q}|)\,|\mathbf{P}-\mathbf{Q}| \approx \frac{\phi'(|\mathbf{P}|+|\mathbf{Q}|)}{|\mathbf{P}|+|\mathbf{Q}|}\,|\mathbf{P}-\mathbf{Q}|\,.$$

Observing by the triangle inequality that $\frac{1}{2}(|\mathbf{Q}|+|\mathbf{P}|) \leq |\mathbf{P}-\mathbf{Q}|+|\mathbf{P}| \leq 2(|\mathbf{Q}|+|\mathbf{P}|)$, the assertion follows from Corollary 10 and the definition of shifted N-functions, in particular

$$\frac{\phi'(|\mathbf{P}|+|\mathbf{Q}|)}{|\mathbf{P}|+|\mathbf{Q}|}\,|\mathbf{P}-\mathbf{Q}| \approx \frac{\phi'(|\mathbf{P}|+|\mathbf{P}-\mathbf{Q}|)}{|\mathbf{P}|+|\mathbf{P}-\mathbf{Q}|}\,|\mathbf{P}-\mathbf{Q}| = \phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|).$$

Hence, the Corollary is proved. □

Corollary 65. *Supposing the assumptions of Proposition 62 then*

$$\big(\phi_{|\mathbf{P}|}\big)^*(|\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})|) \approx |\mathbf{F}(\mathbf{Q})-\mathbf{F}(\mathbf{P})|^2\,,$$

for all $\mathbf{P},\mathbf{Q} \in \mathbb{R}^{d\times d}$.

Proof. Corollary 64 and Corollary 10 yield

$$\big(\phi_{|\mathbf{P}|}\big)^*(|\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})|) \approx \big(\phi_{|\mathbf{P}|}\big)^*(\phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|)).$$

Now, by (2.8) it follows

$$\big(\phi_{|\mathbf{P}|}\big)^*(\phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|)) \approx \phi_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|).$$

Recalling Proposition 62, this proves the assertion. □

The following results deal with the change of the shift of a shifted N-function.

Lemma 66. *Let ϕ be an N-function that satisfies Assumption 40. We then have for all* $\mathbf{P},\mathbf{Q} \in \mathbb{R}^{d\times d}$

$$\phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|) \approx \phi'_{|\mathbf{Q}|}(|\mathbf{P}-\mathbf{Q}|)$$

and

$$\phi_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|) \approx \phi_{|\mathbf{Q}|}(|\mathbf{P}-\mathbf{Q}|),$$

for all $\mathbf{P},\mathbf{Q} \in \mathbb{R}^{d\times d}$. *The constants hidden in $\approx$ depend solely on $\Delta_2(\{\phi,\phi^*\})$.*

Proof. Observing that $\frac{1}{2}(|\mathbf{P}|+|\mathbf{P}-\mathbf{Q}|) \leq |\mathbf{P}|+|\mathbf{Q}| \leq 2(|\mathbf{Q}|+|\mathbf{P}-\mathbf{Q}|)$, the first assertion follows by the definition of shifted N-functions and Corollary 17

$$\frac{\phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|)}{|\mathbf{P}-\mathbf{Q}|} = \frac{\phi'(|\mathbf{P}|+|\mathbf{P}-\mathbf{Q}|)}{|\mathbf{P}|+|\mathbf{P}-\mathbf{Q}|} \approx \frac{\phi'(|\mathbf{Q}|+|\mathbf{P}-\mathbf{Q}|)}{|\mathbf{Q}|+|\mathbf{P}-\mathbf{Q}|} = \frac{\phi'_{|\mathbf{Q}|}(|\mathbf{P}-\mathbf{Q}|)}{|\mathbf{P}-\mathbf{Q}|}.$$

The second assertion follows by Proposition 15. □

Remark 67. *The assertion of Lemma 66 could also be deduced from Proposition 62 since the expression in terms of* $\mathbf{F}$ *is symmetric in* $\mathbf{P}$ *and* $\mathbf{Q}$ *there. In this case additionally the constants of Assumption 40 would be involved, which is avoided in the proof above.*

Lemma 68. *Let* ϕ *be an N-function with* $\Delta_2(\phi) < \infty$*, then for all* $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{d\times d}$ *and* $t \geq 0$ *it holds*

$$\phi'_{|\mathbf{P}|}(t) \preccurlyeq \phi'_{|\mathbf{Q}|}(t) + \phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|). \tag{3.15}$$

The constant hidden in $\preccurlyeq$ *depends only on* $\Delta_2(\phi)$*.*

Proof. Since $\phi'_{|\mathbf{P}|}(t) \approx \phi_{|\mathbf{P}|}(t)/t$ and $\phi_{|\mathbf{P}|}(2t) \approx \phi_{|\mathbf{P}|}(t)$, we have $\phi'_{|\mathbf{P}|}(2t) \approx \phi'_{|\mathbf{P}|}(t)$. All constants depend only on $\Delta_2(\phi_{|\mathbf{P}|})$, hence by Lemma 57 the constants depend only on $\Delta_2(\phi)$. We split the considerations into two cases:

Case $|\mathbf{P}-\mathbf{Q}| \leq \frac{1}{2}t$: From $|\mathbf{P}-\mathbf{Q}| \leq \frac{1}{2}t$ follows $0 \leq \frac{1}{2}(|\mathbf{Q}|+t) \leq |\mathbf{P}|+t \leq 2(|\mathbf{Q}|+t)$. Hence,

$$\phi'_{|\mathbf{P}|}(t) = \frac{\phi'(|\mathbf{P}|+t)}{|\mathbf{P}|+t}\,t \leq \frac{\phi'(2\,(|\mathbf{Q}|+t))}{\frac{1}{2}(|\mathbf{Q}|+t)}\,t \leq 2\,C\,\frac{\phi'(|\mathbf{Q}|+t)}{|\mathbf{Q}|+t}\,t = 2\,C\,\phi'_{|\mathbf{Q}|}(t).$$

Case $|\mathbf{P}-\mathbf{Q}| \geq \frac{1}{2}t$: We estimate

$$\phi'_{|\mathbf{P}|}(t) \leq \phi'_{|\mathbf{P}|}(2\,|\mathbf{P}-\mathbf{Q}|) \leq C\,\phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|).$$

Combining the two cases proves the lemma. □

Corollary 69. *Let* ϕ *be an N-function with* $\Delta_2(\{\phi,\phi^*\}) < \infty$*. Then for* $\delta > 0$ *there exists* $C_\delta > 0$ *depending solely on* δ *and* $\Delta_2(\{\phi,\phi^*\}) < \infty$ *such that for all* $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{d\times d}$ *and* $t \geq 0$

$$\phi_{|\mathbf{P}|}(t) \preccurlyeq (1+C_\delta)\,\phi_{|\mathbf{Q}|}(t) + \delta\,\phi_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|).$$

The constant hidden in $\preccurlyeq$ *depends only on* $\Delta_2(\{\phi,\phi^*\})$*.*

Let ϕ *additionally satisfy Assumption 40. Then for all* $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{d\times d}$ *and* $t \geq 0$

$$\phi_{|\mathbf{P}|}(t) \preccurlyeq (1+C_\delta)\,\phi_{|\mathbf{Q}|}(t) + \delta\,\big|\mathbf{F}(\mathbf{P}) - \mathbf{F}(\mathbf{Q})\big|^2.$$

The constant hidden in $\preccurlyeq$ *depends only on* $\Delta_2(\{\phi,\phi^*\})$ *and the constants in Assumption* (40)*.*

Proof. Due to Corollary 15 it holds $\phi_{|\mathbf{P}|}(t) \approx \phi'_{|\mathbf{P}|}(t)\, t$. With (3.15) and Young's inequality (Proposition 11) we obtain

$$\begin{aligned}\phi_{|\mathbf{P}|}(t) &\preccurlyeq \phi'_{|\mathbf{P}|}(t)\, t \preccurlyeq \phi'_{|\mathbf{Q}|}(t)\, t + \phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|)\, t \\ &\preccurlyeq \phi_{|\mathbf{Q}|}(t) + \delta\, \phi^*_{|\mathbf{Q}|}\big(\phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|)\big) + C_\delta\, \phi_{|\mathbf{Q}|}(t)\end{aligned}$$

for all $\delta > 0$. The constant C_δ depends on δ and $\Delta_2(\phi_{|\mathbf{Q}|})$ and thus on $\Delta_2(\phi)$; see Lemma 57. Now, it follows from Lemma 66, Corollary 17, and (2.8) that

$$\phi^*_{|\mathbf{Q}|}\big(\phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|)\big) \approx \phi^*_{|\mathbf{Q}|}\big(\phi'_{|\mathbf{Q}|}(|\mathbf{P}-\mathbf{Q}|)\big) \approx \phi_{|\mathbf{Q}|}(|\mathbf{P}-\mathbf{Q}|).$$

The second assertion follows with the help of Lemma 62. □

Lemma 70. *Let ϕ be an N-function with $\Delta_2(\{\phi,\phi^*\}) < \infty$, then for all $\mathbf{P},\mathbf{Q} \in \mathbb{R}^{d\times d}$ and $t \geq 0$ it holds*

$$\big((\phi_{|\mathbf{P}|})^*\big)'(t) \preccurlyeq \big((\phi_{|\mathbf{Q}|})^*\big)'(t) + |\mathbf{P}-\mathbf{Q}|\,.$$

The constant hidden in $\preccurlyeq$ depends solely on $\Delta_2(\{\phi,\phi^*\})$.

Proof. Observe that $\phi'(|\mathbf{P}|) = |\mathbf{A}(\mathbf{P})|$. This, in combination with Remark 61, yields

$$\big((\phi_{|\mathbf{P}|})^*\big)'(t) \approx \big((\phi^*)_{|\mathbf{A}(\mathbf{P})|}\big)'(t).$$

Applying Lemma 68 to $\big((\phi^*)_{|\mathbf{A}(\mathbf{P})|}\big)'(t)$, we have

$$\big((\phi^*)_{|\mathbf{A}(\mathbf{P})|}\big)'(t) \preccurlyeq \big((\phi^*)_{|\mathbf{A}(\mathbf{Q})|}\big)'(t) + \big((\phi^*)_{|\mathbf{A}(\mathbf{P})|}\big)'(|\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})|). \tag{3.16}$$

Recalling Corollary 64, we get for the last term

$$\big((\phi^*)_{|\mathbf{A}(\mathbf{P})|}\big)'(|\mathbf{A}(\mathbf{P})-\mathbf{A}(\mathbf{Q})|) \approx \big((\phi^*)_{|\mathbf{A}(\mathbf{P})|}\big)'\big(\phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|)\big).$$

Inserting this in (3.16), a re-transformation via Remark 61 yields

$$\begin{aligned}\big((\phi_{|\mathbf{P}|})^*\big)'(t) &\preccurlyeq \big((\phi_{|\mathbf{Q}|})^*\big)'(t) + \big((\phi_{|\mathbf{P}|})^*\big)'\big(\phi'_{|\mathbf{P}|}(|\mathbf{P}-\mathbf{Q}|)\big) \\ &= \big((\phi_{|\mathbf{Q}|})^*\big)'(t) + |\mathbf{P}-\mathbf{Q}|\,,\end{aligned}$$

where the last equality follows from the definition of dual functions (2.1). □

Corollary 71. *Let ϕ be an N-function with $\Delta_2(\{\phi,\phi^*\}) < \infty$. Then for $\delta > 0$ there exists $C_\delta > 0$ depending solely on δ and $\Delta_2(\{\phi,\phi^*\}) < \infty$ such that for all $\mathbf{P},\mathbf{Q} \in \mathbb{R}^{d\times d}$ and $t \geq 0$*

$$(\phi_{|\mathbf{P}|})^*(t) \preccurlyeq (1 + C_\delta)\,(\phi_{|\mathbf{Q}|})^*(t) + \delta\, \phi_{|\mathbf{Q}|}(|\mathbf{P}-\mathbf{Q}|).$$

The constant hidden in $\preccurlyeq$ depends only on $\Delta_2(\{\phi,\phi^*\})$.

If ϕ additionally satisfies Assumption 40 then for all $\mathbf{P},\mathbf{Q} \in \mathbb{R}^{d\times d}$ and $t \geq 0$

$$(\phi_{|\mathbf{P}|})^*(t) \preccurlyeq (1 + C_\delta)\,(\phi_{|\mathbf{Q}|})^*(t) + \delta\, \big|\mathbf{F}(\mathbf{P}) - \mathbf{F}(\mathbf{Q})\big|^2.$$

The constant hidden in $\preccurlyeq$ depends only on $\Delta_2(\{\phi,\phi^*\})$ and the constants in Assumption (40).

Proof. Due to Corollary 15 it holds $(\phi_{|\mathbf{P}|})^*(t) \approx \big((\phi_{|\mathbf{P}|})^*\big)'(t)\,t$. Thus, multiplying the estimate of Lemma 70 by t yields

$$(\phi_{|\mathbf{P}|})^*(t) \lesssim (\phi_{|\mathbf{Q}|})^*(t) + |\mathbf{P}-\mathbf{Q}|\;t.$$

Now, applying Young's inequality (Proposition 11), we get with Lemma 57

$$(\phi_{|\mathbf{P}|})^*(t) \lesssim (1+C_\delta)\,(\phi_{|\mathbf{Q}|})^*(t) + \delta\,\phi_{|\mathbf{Q}|}(|\mathbf{P}-\mathbf{Q}|),$$

where C_δ depends on $\Delta_2(\phi^*_{|\mathbf{Q}|})$ and thus on $\Delta_2(\phi^*)$; see Remark 61. The second assertion follows with Proposition 62. □

Remark 72. *Note that the constant C_δ in Corollary 69 depends on $\Delta_2(\phi_{|\mathbf{Q}|})$. In particular, $C_\delta \leq \delta\Delta_2(\phi_{|\mathbf{Q}|})^{\lfloor \log_2(1/\delta)\rfloor+1}$, where $\lfloor x\rfloor$, $x\in\mathbb{R}$, denotes the greatest integer less or equal x; see the proof of Proposition 11. The dependence on $\Delta_2(\{\phi,\phi^*\})$ then follows from Lemma 57. The same holds for the constant C_δ in Corollary 71 with $\Delta_2(\phi_{|\mathbf{Q}|})$ substituted by $\Delta_2(\phi^*_{|\mathbf{Q}|})$.*

Remark 73. *Note that $W_0^{1,\phi}(\Omega) = W_0^{1,\phi_a}(\Omega)$ for any $a \geq 0$ since mean convergence with ϕ implies mean convergence with ϕ_a and vice versa: Assume that $(v_n)_{n\in\mathbb{N}} \subset C_0^\infty(\Omega)$ is a Cauchy sequence in $W_0^{1,\phi}(\Omega)$ but not in $W_0^{1,\phi_a}(\Omega)$. Hence, there exists $v \in W_0^{1,\phi}(\Omega)$ such that $v_n \to v$ in $W_0^{1,\phi}(\Omega)$ as $n\to\infty$. Since norm convergence is equivalent to mean convergence and Corollary 36, there exist a subsequence $(v_{n_l})_{l\in\mathbb{N}}, \subset (v_n)_{n\in\mathbb{N}}$ such that $\int_\Omega \phi_a(|\nabla(v_{n_l}-v)|)\,dx > c > 0$ for all $l\in\mathbb{N}$. Therefore, Corollary 69 yields*

$$\begin{aligned}0 < c &< \int_\Omega \phi_a(|\nabla(v_{n_l}-v)|)\,dx \\ &\lesssim (1+C_\delta)\int_\Omega \phi(|\nabla(v_{n_l}-v)|)\,dx + \delta\int_\Omega \phi(a)\,dx.\end{aligned}$$

If we now choose δ small enough, we end up with

$$0 < c \lesssim \int_\Omega \phi(|\nabla(v_{n_l}-v)|)\,dx,$$

which is a contradiction since the right hand side converges to zero as $l\to\infty$. Recalling that $C_0^\infty(\Omega)$ is dense in $W_0^{1,\phi}(\Omega)$ and $W_0^{1,\phi_a}(\Omega)$ we get with Corollary 36 $W_0^{1,\phi}(\Omega) \subset W_0^{1,\phi_a}(\Omega)$. The other inclusion follows analogously with interchanged roles of ϕ and ϕ_a.

We consider $\phi(t) = \frac{1}{r}t^r$ with $r > 1$. Recalling the definition of shifted N-functions, we have for $\kappa \geq 0$

$$\phi_\kappa(t) = \int_0^t \phi_\kappa'(s)\,ds = \int_0^t \frac{\phi'(\kappa+s)}{\kappa+s}s\,ds = \int_0^t (\kappa+s)^{r-2}s\,ds.$$

Hence, Remark 73 and Corollary 36 imply $W_0^{1,r}(\Omega) = W_0^{1,\phi_\kappa}(\Omega)$. *The same assertion holds for* $\varphi(t) := \int_0^t (\kappa^2 + s^2)^{\frac{r-2}{2}} s\,ds$ *observing that* $a^2 + b^2 \approx (a+b)^2$ *for all* $a, b \geq 0$ *and therefore* $\phi_\kappa(t) \approx \varphi(t)$, *for all* $t \geq 0$. *Hence, all these families of N-functions lead to the same space* $W_0^{1,r}(\Omega) = W_0^{1,\phi}(\Omega) = W_0^{1,\varphi}(\Omega) = W_0^{1,\phi_\kappa}(\Omega)$.

Moreover, let us consider for $r \in (1, \infty)$ *and* $\kappa \geq 0$, $\nu_0 > \nu_\infty > 0$ *the N-function* $\hat{\phi}(t) := \int_0^t \big(\nu_\infty + (\nu_0 - \nu_\infty)(\kappa^2 + s^2)^{(r-2)/2}\big) s\,ds$. *Then*

$$\hat{\phi}(t) = \nu_\infty \frac{1}{2} t^2 + (\nu_0 - \nu_\infty)\, \varphi(t),$$

which in turn implies $W_0^{1,\hat{\phi}}(\Omega) = W_0^{1,\max\{2,r\}}(\Omega) = W_0^{1,2}(\Omega) \cap W_0^{1,\varphi}(\Omega)$.

3.2.2 Quasi-Norm

Once the shifted N-functions have been established we can use them to define error quantities, which generalize the classical quasi-norm.

Lemma 74. *Let* ϕ *be an N-function that satisfies Assumption 40, then for each* $v, w \in W_0^{1,\phi}(\Omega)$

$$\begin{aligned}
\langle -\operatorname{div} \mathbf{A}(\nabla v) + \operatorname{div} \mathbf{A}(\nabla w), v - w\rangle &= \int_\Omega \big(\mathbf{A}(\nabla v) - \mathbf{A}(\nabla w)\big) : \big(\nabla v - \nabla w\big)\, dx \\
&\approx \int_\Omega \phi''(|\nabla v| + |\nabla w|)\, |\nabla v - \nabla w|^2\, dx \\
&\approx \int_\Omega \phi_{|\nabla v|}(|\nabla v - \nabla w|)\, dx \\
&\approx \|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla w)\|_{L^2(\Omega)}^2 .
\end{aligned}$$

The constants hidden in $\approx$ *depend solely on* $\Delta_2(\{\phi, \phi^*\})$ *and the constants of Assumption 40.*

Proof. The assertion is a direct consequence of Proposition 62. □

Remark 75. *We will extensively use each of the proportional expressions in Lemma 74 since each of them exhibits different advantages. The first expression utilizes the properties of the partial differential equation.*

In the case of $\phi(t) = \frac{1}{r} t^r$, $r > 1$ *the classical* quasi-norm *of Barrett and Liu reads*

$$\|v - w\|_{(r)}^2 = \int_\Omega (|\nabla v| + |\nabla w|)^{r-2}\, |\nabla v - \nabla w|^2\, dx,$$

for $v, w \in W_0^{1,r}(\Omega)$. *Recalling Remark 63 we get*

$$\|v - w\|_{(r)}^2 = \frac{1}{r-1} \int_\Omega \phi''(|\nabla v| + |\nabla w|)\, |\nabla v - \nabla w|^2\, dx.$$

Thus, the expression defined via the second derivative of ϕ is closest to the classical quasi-norm and in the case $\phi(t) = \frac{1}{r}t^r$ all quantities in Lemma 74 are indeed proportional to the classical quasi-norm.

The expression $\int_\Omega \phi_{|\nabla v|}(|\nabla v - \nabla w|)\,dx$, based on the shifted N-function, enables us to apply Young's inequality as well as techniques for convex functions. With the calculations of Remark 63 we obtain for $\phi(t) = \frac{1}{r}t^r$

$$\int_\Omega \phi_{|\nabla v|}(|\nabla v - \nabla w|)\,dx = \int_\Omega \int_0^{|\nabla v - \nabla w|} \phi'_{|\nabla v|}(s)\,ds\,dx$$
$$= \int_\Omega \int_0^{|\nabla v - \nabla w|} (|\nabla v| + s)^{r-2} s\,ds\,dx.$$

The expression in terms of $\mathbf{F}$ is important for stating the results since it is convenient to have a symmetric error quantity. Moreover, it also plays an important role in the a priori analysis, since it seems to be the natural quantity to express regularity; see [39, 38, 26]. In fact, convergence of order h can be obtained if $\nabla\mathbf{F}(\nabla u)$ is square integrable. Particularly, let $\mathring{\mathbb{V}}(\mathcal{T}) \subset \mathbb{V}$ be a conforming finite element space. Then, for a suitable interpolation operator $\Pi_h : W_0^{1,\phi}(\Omega) \to \mathbb{V}(\mathcal{T})$

$$\|\mathbf{F}(\nabla u) - \mathbf{F}(\Pi_h u)\|_{L^2(\Omega)} \le C\, h_{\max}(\mathcal{T})\, \|\nabla \mathbf{F}(\nabla u)\|_{L^2(\Omega)},$$

where $h_{\max}(\mathcal{T})$ is the maximal mesh-size of the underlying mesh $\mathcal{T}$. For $\phi(t) = \frac{1}{r}t^r$ the error expression in terms of $\mathbf{F}$ becomes

$$\|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla w)\|^2_{L^2(\Omega)} = \int_\Omega \left| |\nabla v|^{\frac{r-2}{2}} \nabla v - |\nabla w|^{\frac{r-2}{2}} \nabla w \right|^2 dx.$$

In the case $\phi(t) = \frac{1}{2}t^2$, i.e., in the case when $\operatorname{div}\mathbf{A}(\nabla \cdot)$ coincides with the linear Laplacian, then $\phi'' \equiv 0$, $\mathbf{F} = id$, and $\phi_a(t) = \phi(t)$. Therefore the quasi-norm is equivalent to the usual Sobolev semi-norm $|\cdot|_{W_0^{1,2}(\Omega)}$.

These error quantities, which might seem dubious at first glance, are actually reasonable, since convergence in the quasi norm implies convergence in $W_0^{1,\phi}(\Omega)$ and vice versa.

Lemma 76. *Let ϕ be an N-function that satisfies Assumption 40. Let further $v, w \in L^\phi(\Omega)$ and $(v_n)_{n\in\mathbb{N}} \subset L^\phi(\Omega)$. Then*

$$\int_\Omega \phi_{|w|}(|v_n - v|)\,dx \to 0 \quad \text{as } n \to \infty$$

is equivalent to the convergence in $L^\phi(\Omega)$

$$v_n \to v \quad \text{in} \quad L^\phi(\Omega) \quad \text{as } n \to \infty.$$

Moreover, it holds $L^\phi(\Omega) = L^{\phi_{|w|}}(\Omega)$.

Proof. Starting from the quasi-norm convergence, we assume that $(v_n)_{n\in\mathbb{N}}$ does not converge to v in $L^\phi(\Omega)$. Hence according to Proposition 31 there exists a subsequence $(v_{n_l})_{l\in\mathbb{N}}$ such that

$$0 < c < \int_\Omega \phi(|v - v_{n_l}|)\,dx$$

for all $l \in \mathbb{N}$ and some $c > 0$. Corollary 69 implies for $\delta > 0$

$$\int_\Omega \phi(|v - v_{n_l}|)\,dx \preccurlyeq (1 + C_\delta) \int_\Omega \phi_{|w|}(|v - v_{n_l}|)\,dx + \delta \int_\Omega \phi(|w|)\,dx.$$

Since the left hand side is bounded away from zero and $\int_\Omega \phi(|w|)\,dx$ is bounded we get for δ small enough

$$c < \int_\Omega \phi(|v - v_{n_l}|)\,dx \preccurlyeq \int_\Omega \phi_{|w|}(|v - v_{n_l}|)\,dx \to 0,$$

as $l \to \infty$. This is a contradiction. The converse assertion can be proved in the same way by interchanging the roles of ϕ and $\phi_{|v|}$.

The assertion $L^\phi(\Omega) = L^{\phi_{|w|}}(\Omega)$ follows from the fact that mean convergence implies convergence (see Proposition 31) and from the density of $C_0^\infty(\Omega)$ in $L^\phi(\Omega)$ and $L^{\phi_{|w|}}(\Omega)$. □

Corollary 77. *Let ϕ be an N-function that satisfies Assumption 40. Let further $v \in W_0^{1,\phi}(\Omega)^d$ and $(v_n)_{n\in\mathbb{N}} \subset W_0^{1,\phi}(\Omega)^d$. Then the quasi-norm convergence*

$$\|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla v_n)\|_{L^2(\Omega)}\,dx \to 0 \quad \text{as } n \to \infty$$

is equivalent to the convergence in $W_0^{1,\phi}(\Omega)^d$

$$v_n \to v \quad \text{in} \quad W_0^{1,\phi}(\Omega)^d \quad \text{as } n \to \infty.$$

Proof. Lemma 74 implies that

$$\int_\Omega \phi_{|\nabla v|}(|\nabla v - \nabla v_n|)\,dx \approx \|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla v_n)\|^2_{L^2(\Omega)} \to 0,$$

as $n \to \infty$. Hence, the assertion follows with Lemma 76 by means of Corollary 36. □

The above results yields that the quasi-norm expression in terms of $\mathbf{F}$ is a metric.

Corollary 78. *Let ϕ be an N-function that satisfies Assumption 40. Then $(W_0^{1,\phi}(\Omega), \mathbf{d})$ is a closed metric space with*

$$\mathbf{d}(v, w) := \|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla w)\|_{L^2(\Omega)}\,.$$

Proof. The assertion is an easy consequence of Corollary 77 and the properties of the L^2-norm. □

Remark 79. *The quasi-norm approach naturally arises from the fundamental principle of estimating the error by a residual expression: For the ease of exposition we stick to the case $\phi(t) = \frac{1}{r} t^r$ with $r \in (1, \infty)$. Let $v \in W_0^{1,\phi}(\Omega)$ be an approximation to the solution u of (3.2). The residual $D\mathcal{J}(v)$ is a functional in the dual space $W^{-1,r'}(\Omega)$ with $\frac{1}{r} + \frac{1}{r'} = 1$. Quantifying it in the dual energy norm leads necessarily to a gap in the power of the upper and the lower bound. In particular, with $\|v\|_r := |v|_{W^{1,r}(\Omega)^d}$ for $v \in W_0^{1,r}(\Omega)^d$ and $\|D\mathcal{J}(v)\|_{r',*} := \sup_{v \in \mathbb{V}, \|v\|_r = 1} \int_\Omega g \cdot v - \mathbf{A}(\nabla v) : \nabla v \, dx$, it holds*

$$\|u - v\|_r^{r-1} \preccurlyeq \|D\mathcal{J}(v)\|_{r',*} \preccurlyeq (\|u\|_r + \|v\|_r)^{r-2} \|u - v\|_r,$$

if $r \geq 2$, and

$$\|u - v\|_r \preccurlyeq (\|u\| + \|v\|)^{2-r} \|D\mathcal{J}(v)\|_{r',*} \preccurlyeq (\|u\|_r + \|v\|_r)^{2-r} \|u - v\|_r^{r-1}$$

if $r \in (1, 2)$. The reason for this gap is that energy error and the dual energy norm of the residual are somehow not in 'balance'. The idea is now to find a primal measure of distance that is 'balanced' with the resulting dual measure for the residual: We shall consider a different formulation of the dual energy norm, namely

$$\frac{1}{r'} \|D\mathcal{J}(v)\|_{r',*}^{r'} = \sup_{w \in W_0^{1,r}(\Omega)} \langle D\mathcal{J}(v), w \rangle - \frac{1}{r} \|w\|_r^r$$

or in a more abstract equivalent formulation with N-functions

$$\|D\mathcal{J}(v)\|_{\phi^*,*}^{r'} = \sup_{w \in W_0^{1,r}(\Omega)} \langle D\mathcal{J}(v), w \rangle - \int_\Omega \phi(|\nabla w|) \, dx. \tag{3.17}$$

Roughly spoken, the dual norm is getting weaker as the primal norm is getting stronger and vice versa. In the quasi-norm concept, dual and primal error measure are balanced: Recall the equivalent quasi-norm quantities of Lemma 74. Then, defining

$$\|D\mathcal{J}(v)\|_{(\nabla u),*}^2 = \sup_{w \in W_0^{1,\phi}(\Omega)} \langle D\mathcal{J}(v), w \rangle - \int_\Omega \phi_{|\nabla u|}(|\nabla w|) \, dx$$

yields with Young's inequality (2.3)

$$\begin{aligned}\|D\mathcal{J}(v)\|^2_{(\nabla u),*} &= \sup_{w\in W^{1,\phi}_0(\Omega)} \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla v)\big) : \nabla w \, dx - \int_\Omega \phi_{|\nabla u|}(|\nabla w|)\, dx \\ &\le \sup_{w\in W^{1,\phi}_0(\Omega)} \int_\Omega \big(\phi_{|\nabla u|}\big)^*\big(\,|\mathbf{A}(\nabla u) - \mathbf{A}(\nabla v)|\,\big)\, dx \\ &\quad + \int_\Omega \phi_{|\nabla u|}(|\nabla w|)\, dx - \int_\Omega \phi_{|\nabla u|}(|\nabla w|)\, dx \\ &= \int_\Omega \big(\phi_{|\nabla u|}\big)^*\big(\,|\mathbf{A}(\nabla u) - \mathbf{A}(\nabla v)|\,\big)\, dx.\end{aligned}$$

On the other hand, testing the residual with $\alpha(u-v)$ *yields*

$$\begin{aligned}\|D\mathcal{J}(v)\|^2_{(\nabla u),*} \ge &\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla v)\big) : \nabla \alpha(u-v)\, dx \\ &- \int_\Omega \phi_{|\nabla u|}(|\nabla \alpha(u-v)|)\, dx.\end{aligned}$$

Hence, with Corollary 19 there exist $s > 1$, $C > 0$, *such that*

$$\begin{aligned}&\ge \alpha \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla v)\big) : \nabla(u-v)\, dx \\ &\quad - \alpha^s C \int_\Omega \phi_{|\nabla u|}(|\nabla(u-v)|)\, dx\end{aligned}$$

and thus with Lemma 74

$$\ge (\alpha - \alpha^s \tilde{C}) \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla v)\big) : \nabla(u-v)\, dx.$$

Now, choosing $\alpha > 0$ *small enough yields that the* dual quasi-norm *of the residual is equivalent to the quasi-norm of the error.*

The quasi-norm was first introduced by Barrett and Liu in [8, 9, 10]. In particular, they considered the case $\phi(t) = \frac{1}{r}t^r$ *with* $r \in (1,\infty)$. *As is shown in Remark 75, the approach from [31, 26, 32] and [28], which we present in this work, is a generalization of this concept. In fact, this generalization covers most common nonlinearities in the modeling of quasi-Newtonian flows; see Remark 42 and Section 1.1. Moreover, in the concept of shifted there is no need to treat different cases like* $r \in (1,2)$ *and* $r \ge 2$ *for* $\phi(t) = \frac{1}{r}t^r$ *separately.*

Remark 80. *The quasi-norm approach leads amongst other assertions to a Cea's Lemma, i.e., let* $U \in X$ *be the solution of* (3.9) *in a closed subspace* $X \subset \mathbb{V}$, *then*

$$\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|_{L^2(\Omega)} \preccurlyeq \inf_{V\in X} \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla V)\|_{L^2(\Omega)}\,;$$

see, e.g., [8, 31]. This is the starting point of the a priori analysis.

3.3 Finite Element Approach

This section provides the finite element framework. The subsequent definitions and concepts of triangulations and finite element spaces are taken from [13, 5, 14, 67]. The interpolation estimates of Section 3.3.3 are taken from [31].

3.3.1 Triangulation and Refinement Framework

This section fixes the notation regarding triangulations of Ω.

Definition 81 (simplex)**.** *For $s \in \mathbb{N}$, $0 \le s \le d$, let $a_0, \dots, a_s \in \mathbb{R}^d$. The s vectors $a_0 - a_1, \dots, a_0 - a_s$ are assumed to be linear independent.*

1. *The set*

$$\begin{aligned} T &:= \text{conv hull } \{a_0, \dots, a_s\} \\ &= \left\{ \sum_{i=0}^{s} \lambda_i : \lambda_i \ge 0 \textit{ and } \sum_{i=1}^{s} \lambda_i = 1 \right\} \end{aligned}$$

   is known as the s simplex spanned by $a_0, \dots, a_s$. The coefficients λ_i describing a point $x \in T$ are unique and known as the barycentric coordinates of x relative to the simplex T. Note that the simplex T is closed.

2. *Let T' be a k simplex spanned by $a'_0, \dots, a'_k \in \{a_0, \dots, a_s\}$. Then T' is called a k sub-simplex of T. The $d-1$ sub-simplices of T are called faces (sides) of T, whereas we denote the 1 sub-simplices of T as its vertices.*

3. *For an s simplex T we define the following characteristic quantities*

$$\begin{aligned} h_T &:= |T|^{1/s}, \\ \text{diam}(T) &:= \max\{|x-y| : x, y \in T\}, \\ \rho(T) &:= \max\{2r : B_r \subset T \textit{ is an } s\textit{-sphere of radius } r\}, \\ \sigma(T) &:= \frac{\text{diam}(T)}{\rho(T)}. \end{aligned}$$

4. *The reference d simplex $\hat{T} \subset \mathbb{R}^d$ is defined as*

$$\hat{T} := \text{conv hull}\{0, e_1, \dots, e_d\},$$

   where e_i are the standard unit vectors in $\mathbb{R}^d$.

For every d simplex T spanned by $\{a_0, \dots, a_d\}$, there exists a bijective affine linear mapping $F_T : \hat{T} \to T$. In particular,

$$F_T \hat{x} := \mathbf{C}_T\, \hat{x} + a_0, \qquad \text{with} \qquad \mathbf{C}_T := \begin{pmatrix} \vdots & & \vdots \\ a_1 - a_0 & \cdots & a_d - a_0 \\ \vdots & & \vdots \end{pmatrix} \in \mathbb{R}^{d \times d}.$$

Note that

$$\|\mathbf{C}_T\|_2 \leq \frac{\operatorname{diam}(T)}{\rho(\hat{T})}, \qquad \left\|\mathbf{C}_T^{-1}\right\| \leq \frac{\operatorname{diam}(\hat{T})}{\rho(T)}, \qquad |\det \mathbf{C}_T| = \frac{|T|}{|\hat{T}|},$$

where $\|\cdot\|_2$ is the matrix norm associated with the Euclidean norm on $\mathbb{R}^{d\times d}$; see, e.g., [21, 67, 13, 14]. We will often use scaling arguments, where we transform functions v defined on an d-simplex T to the standard d-simplex $\hat{T}$. We denote the scaled function by $\hat{v} = v \circ F_T$.

Definition 82 (conforming triangulation)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with polygonal boundary. A finite set $\mathcal{T}$ of d simplices is said to be a conforming triangulation of Ω if*

1. *the domain Ω is the interior of the set $\bigcup_{T\in\mathcal{T}} T$.*
2. *the intersection $T_1 \cap T_2$ of two d simplices $T_1, T_2 \in \mathcal{T}$ is either empty or a common sub-simplex of both T_1 and T_2.*

Let $\mathcal{T}$ be a conforming triangulation of Ω. Then the set of vertices (nodes) of all T, $T \in \mathcal{T}$ is denoted by $\mathcal{N}$, whereas $\mathring{\mathcal{N}}$ denotes the set of interior vertices (nodes), i.e., $\mathring{\mathcal{N}} = \mathcal{N} \cap \Omega$. The set of faces (sides) of T, $T \in \mathcal{T}$ is denoted by $\mathcal{S}$ and the set of interior sides is denoted by $\mathring{\mathcal{S}}$.

For $\sigma \in \mathcal{S}$ we denote by ω_σ the union of the adjacent elements sharing σ, i.e.,

$$\omega_\sigma := \text{interior}\left(\bigcup\{T \in \mathcal{T} \mid \sigma \subset T\}\right).$$

For $T \in \mathcal{T}$ we define

$$\omega_T := \text{interior}\left(\bigcup\{\omega_\sigma \mid \sigma \in \mathcal{S}, \sigma \subset T\}\right),$$

and

$$S_T := \text{interior}\left(\bigcup\{T' \in \mathcal{T} \mid T' \cap T \neq \emptyset\}\right).$$

Let $z \in \mathcal{N}$ be a node of the triangulation $\mathcal{T}$. The corresponding finite element star is then denoted by

$$\omega_z := \text{interior}\left(\bigcup\{T \in \mathcal{T} \mid z \in T\}\right)$$

and its interior sides by

$$\sigma_z := \bigcup\{\sigma \in \mathcal{S} \mid \sigma \cap \omega_z \neq \emptyset\}.$$

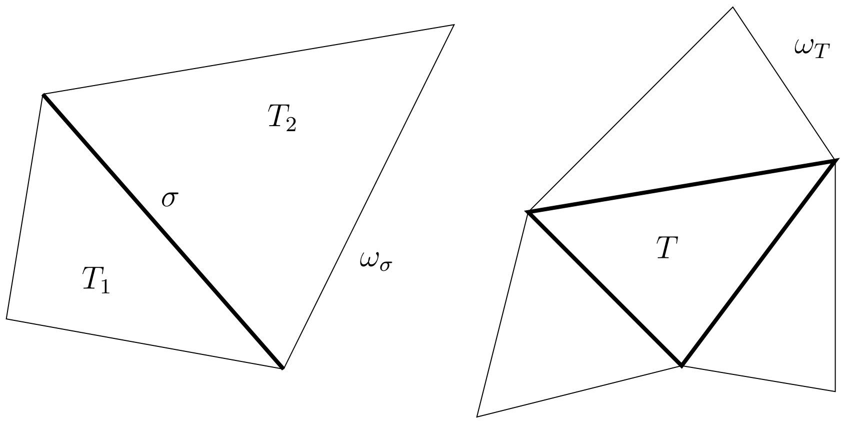

Figure 3.1: Neighborhood of σ and T in 2 dimensions.

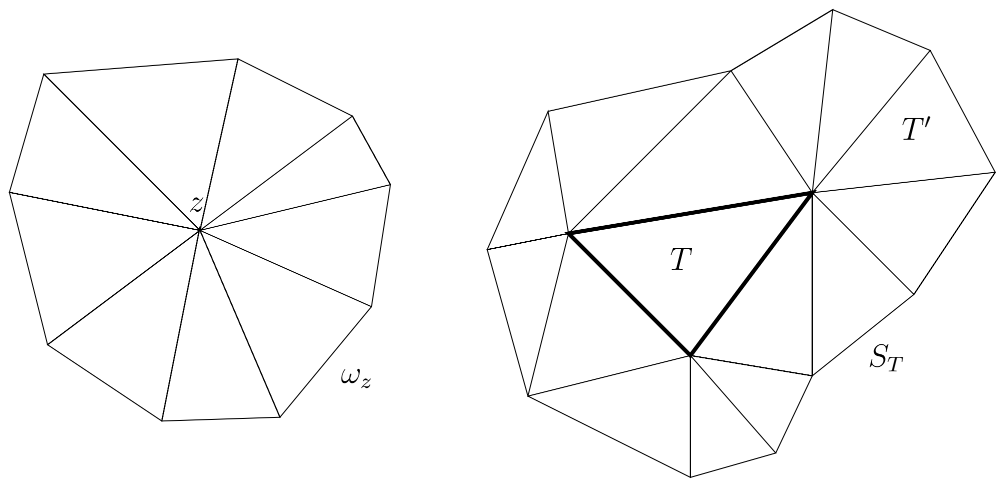

Figure 3.2: Finite element star ω_z for $z \in \mathring{\mathcal{N}}$ and Patch S_T of an interior element $T \in \mathcal{T}$ in 2 dimensions.

For $A \subset \Omega$ we define a sub-triangulation $\mathcal{T}(A) \subset \mathcal{T}$ by

$$\mathcal{T}(A) := \{T \in \mathcal{T} : T \subset \overline{A}\}.$$

We further define the *shape-regularity* of a conforming triangulation $\mathcal{T}$ by

$$\sigma(\mathcal{T}) := \max_{T \in \mathcal{T}} \sigma(T).$$

For $T \in \mathcal{T}$ the quantities h_T, $\operatorname{diam}(T)$, and $\rho(T)$ are mutually equivalent depending solely on the shape-regularity of $\sigma(T)$. The mesh-size of two neighboring elements is comparable, i.e., for $T_1, T_2 \in \mathcal{T}$, $T_1 \cap T_2 \in \mathcal{S}$ there exist $C, c > 0$ depending solely on $\sigma(\mathcal{T})$ such that

$$c\, h(T_1) \leq h(T_2) \leq C\, h(T_1).$$

Moreover, the minimum angle of $T \in \mathcal{T}$ is bounded depending on $\sigma(\mathcal{T})$, and hence the number of elements that are contained in the closure of S_T is bounded depending on the shape-regularity $\sigma(\mathcal{T})$.

A sequence $(\mathcal{T}_k)_{k \in \mathbb{N}}$ of conforming triangulations of Ω is called *shape-regular* if the parameter $\sigma(\mathcal{T}_k)$ remains bounded, i.e.,

$$\sup_{k \in \mathbb{N}} \sigma(\mathcal{T}_k) < \infty.$$

Let $\mathcal{T}$, $\mathcal{T}_*$ be two conforming triangulations of Ω, then we call $\mathcal{T}_*$ a refinement of $\mathcal{T}$ if for any $T \in \mathcal{T}$ the subset $\mathcal{T}_*(T) \subset \mathcal{T}_*$ is a conforming triangulation of T, i.e.,

$$T = \bigcup_{T' \in \mathcal{T}_*(T)} T'.$$

This defines a partial ordering on all conforming triangulations of Ω, i.e., we denote

$$\mathcal{T}_* \geq \mathcal{T}, \text{ if } \mathcal{T}_* \text{ is a refinement of } \mathcal{T}.$$

3.3.2 Finite Element Space and Discrete Problem

For the remainder of the chapter we denote $\mathbb{V} := W_0^{1,\phi}(\Omega)^d$ as the solution space of (3.2). Assume that $\mathcal{T}$ is a conforming triangulation of Ω. We specify $\mathcal{P}^s(T)$, $s \in \mathbb{N}$, to be the space of polynomials of degree s on $T \in \mathcal{T}$. The conforming finite element space of continuous, piecewise linear functions over $\mathcal{T}$ is then defined by

$$\mathbb{V}(\mathcal{T}) := \left\{ V \in C(\bar{\Omega}) : V|_T \in \mathcal{P}^1(T)^d,\ T \in \mathcal{T} \right\}.$$

Its subspace with homogenous boundary values is given by

$$\mathring{\mathbb{V}}(\mathcal{T}) := \{v \in \mathbb{V}(\mathcal{T}) : V = 0 \text{ on } \partial\Omega\}\,.$$

Note that a function $V \in \mathring{\mathbb{V}}(\mathcal{T})$ is uniquely defined by its values at the interior nodes of $\mathcal{T}$. Let $\mathring{\mathcal{N}} = \{z_1, \ldots, z_{N(\mathcal{T})}\}$ be the set of interior nodes of $\mathcal{T}$. Then the set of functions $\{\Phi_1^1, \ldots, \Phi_{N(\mathcal{T})}^1, \ldots, \Phi_1^d, \ldots, \Phi_{N(\mathcal{T})}^d\} \subset \mathring{\mathbb{V}}(\mathcal{T})$ with

$$\Phi_i^k(z_j) = \delta_{ij}\, e_k, \qquad i, j = 1, \ldots, N(\mathcal{T}),\; k = 1, \ldots, d$$

form a basis of $\mathbb{V}(\mathcal{T})$ called the Lagrange basis of $\mathbb{V}(\mathcal{T})$. Thereby δ_{ij} is the Kronecker delta and e_k is the k-th vector of the standard normal basis of $\mathbb{R}^d$. As an immediate consequence we have $\overline{\omega}_{z_i} = \operatorname{supp}(\Phi_i^k)$, $k = 1, \ldots, d$.

We observe further, that for a conforming triangulation $\mathcal{T}$ and a conforming refinement $\mathcal{T}_*$ of $\mathcal{T}$ the functions $V \in \mathring{\mathbb{V}}(\mathcal{T})$ are continuous and piecewise linear over $\mathcal{T}_*$. Hence, it holds $V \in \mathring{\mathbb{V}}(\mathcal{T}_*)$, i.e., the finite element spaces are nested;

$$\mathring{\mathbb{V}}(\mathcal{T}) \subset \mathring{\mathbb{V}}(\mathcal{T}_*).$$

Since $\mathring{\mathbb{V}}(\mathcal{T}) \subset W_0^{1,\infty}(\Omega)^d$, we obviously have $\mathring{\mathbb{V}}(\mathcal{T}) \subset W_0^{1,\phi}(\Omega)^d$; recall Definition 33 and Proposition 37.

Having the finite element space $\mathring{\mathbb{V}}(\mathcal{T})$ at hand, we can introduce the Ritz Galerkin solution (3.2). In particular, for $g \in W_0^{-1,\phi^*}(\Omega)$ we look for $U \in \mathring{\mathbb{V}}(\mathcal{T})$ such that

$$\int_\Omega \mathbf{A}(\nabla U) : \nabla V\, dx = \langle g, V\rangle, \quad \text{for all } V \in \mathring{\mathbb{V}}(\mathcal{T}), \tag{3.18}$$

where $\mathbf{A}(\mathbf{Q}) := \phi'(|\mathbf{Q}|)\frac{\mathbf{Q}}{|\mathbf{Q}|}$ for $\mathbf{Q} \in \mathbb{R}^{d\times d}$.

Proposition 83. *Let ϕ be an N-function that satisfies Assumption 40 and let $\mathcal{T}$ be a conforming triangulation of Ω. Then there exists a unique solution $U \in \mathring{\mathbb{V}}(\mathcal{T})$ of* (3.18)*. Moreover, U is the unique minimizer of the energy functional $\mathcal{J}(\cdot) = \int_\Omega \phi(|\nabla\cdot|)\, dx - \langle g, \cdot\rangle$ in $\mathring{\mathbb{V}}(\mathcal{T})$.*

Proof. Since $\mathring{\mathbb{V}}(\mathcal{T}) \subset \mathbb{V}$ is a finite dimensional subspace, it is closed. Hence, Corollary 50 yields the first assertion. The second follows analogously by Corollary 55. □

3.3.3 Modular Interpolation Estimates

In what follows, we assume that we have a suitable interpolation operator at hand. Note that the Scott-Zhang interpolation operator satisfies all our requirements; see [68].

Hereafter we use the notation $f \preccurlyeq k$ to indicate $f \leq C\,k$, with a generic constant C solely depending on the Δ_2-constants of some given N-functions, the dimension d, or the shape-regularity of some given triangulations. We denote $f \preccurlyeq k \preccurlyeq f$ as $f \approx k$.

Assumption 84 (interpolation operator)**.** Let $\mathcal{T}$ be a conforming triangulation of the polygonal domain $\Omega \subset \mathbb{R}^d$ and let $\mathring{\mathbb{V}}(\mathcal{T})$ be the finite element space according to Section 3.3. We assume that $\Pi_{\mathcal{T}} : W^{1,1}(\Omega)^d \to \mathbb{V}(\mathcal{T})$ has the following properties:

i) For $T \in \mathcal{T}$ it holds for all $v \in W^{1,1}(\Omega)^d$

$$\sum_{j=0}^{1} \int_T \left| h_T^j \nabla^j \Pi_{\mathcal{T}} v \right| \, dx \leq C \sum_{j=0}^{1} \int_{S_T} \left| h_T^j \nabla^j v \right| \, dx,$$

where the constant $C > 0$ depends only on d and $\sigma(\mathcal{T})$.

ii) The operator Π_h is invariant on $\mathcal{P}^1(\Omega)^d$, i.e., it holds for any linear polynomial $p \in \mathcal{P}^1(\Omega)^d$ that

$$\Pi_{\mathcal{T}} p = p.$$

Remark 85. *Assumption 84 is satisfied by many common interpolation operators as, e.g., the Clément [22] and the Scott-Zhang [68] interpolation operators. The Scott-Zhang operator additionally preserves homogeneous boundary values, i.e.,*

$$\Pi V = V \in \mathbb{V}(\mathcal{T}) \subset \mathbb{V} \qquad \text{for all } V \in \mathbb{V}(\mathcal{T}).$$

Remark 86. *Note that Assumption 84 is sufficient to get interpolation estimates in $W_0^{1,r}(T)^d$, $r \geq 1$; see e.g. [21, 14, 68]. In particular, it holds for all $v \in W^{1,r}(\Omega)$, $T \in \mathcal{T}$*

$$\sum_{i=0}^{1} h_T^i \left\| v - \Pi_{\mathcal{T}} v \right\|_{L^r(T)} \leq C \, h_T \, \left\| \nabla v \right\|_{L^r(S_T)}, \tag{3.19}$$

where C depends only on d, r, and the shape-regularity of $\mathcal{T}$.

Lemma 87. *Let $\mathcal{T}$ be a conforming triangulation of the polygonal domain Ω and let $\Pi_{\mathcal{T}} : W^{1,1}(\Omega)^d \to \mathring{\mathbb{V}}(\mathcal{T})$ satisfy Assumption 84. Then there exists a constant $C > 0$ such that for all $\sigma \in \mathcal{S}$, $v \in W^{1,1}(\Omega)$*

$$\left\| v - \Pi_{\mathcal{T}} v \right\|_{L^1(\sigma)} \leq C \left\| \nabla v \right\|_{L^1(S_T)},$$

where $T \in \mathcal{T}$ with $\sigma \subset \partial T$. The constant C depends only on d and $\sigma(\mathcal{T})$.

Proof. The proof is standard in the context of finite elements; see [21, 22]. In particular, one first maps $v - \Pi_{\mathcal{T}} v$ onto the reference simplex $\hat{T}$, then applies the trace theorem $W^{1,1}(\hat{T}) \hookrightarrow L^1(\hat{\sigma})$, where $\hat{\sigma} = \mathbf{F}_T^{-1}(\sigma)$. Now, back transformation from $\hat{T}$ to T and the interpolation estimate (3.19) yields the desired assertion. □

The proof of the following lemma can be found in [31]. For some of the main ideas consider also Remark 89.

Lemma 88 (stability and approximability)**.** *Let $\mathcal{T}$ be a conforming triangulation of Ω. Let ϕ be an N-function with $\Delta_2(\phi) < \infty$ and let $\Pi_{\mathcal{T}} : \mathbb{V} \to \mathring{\mathbb{V}}(\mathcal{T})$ satisfy Assumption 84. Then, for any $a \geq 0$, $T \in \mathcal{T}$*

$$\sum_{j=0}^{1} \int_T \phi_a \left(\left| h_T^j \nabla^j \Pi_{\mathcal{T}} v \right| \right) dx \leq C \sum_{j=0}^{1} \int_{S_T} \phi_a \left(\left| h_T^j \nabla^j v \right| \right) dx$$

and

$$\sum_{j=0}^{1} \int_T \phi_a \left(h_T^j \left| \nabla^j (v - \Pi_{\mathcal{T}} v) \right| \right) dx \leq C \int_{S_T} \phi_a \left(h_T \left| \nabla v \right| \right) dx,$$

where the constant $C > 0$ depends only on $\sigma(\mathcal{T})$, d, and $\Delta_2(\phi)$.

Remark 89. *The interpolation estimate of Lemma 88 is proved similar to the interpolation estimate in Sobolev spaces using approximability of functions by polynoms [14, 21]. In fact, it can be proven that there exists a polynomial $p \in \mathcal{P}^1(S_T)^d$ such that*

$$\sum_{j=0}^{1} \int_{S_T} \phi_a \left(h_T^j \left| \nabla^j (v - p) \right| \right) dx \leq C \int_{S_T} \phi_a \left(h_T \left| \nabla v \right| \right) dx, \tag{3.20}$$

where the constant $C > 0$ depends only on $\sigma(\mathcal{T})$ and $\Delta_2(\phi)$; see [31]. Therefore, the interpolation estimate of Lemma 88 can be obtained recalling the triangle like inequality of Corollary 10

$$\begin{aligned} \sum_{j=0}^{1} \int_T \phi_a \left(h_T^j \left| \nabla^j (v - \Pi_{\mathcal{T}} v) \right| \right) dx &\preccurlyeq \sum_{j=0}^{1} \int_T \phi_a \left(h_T^j \left| \nabla^j (v - p) \right| \right) dx \\ &\quad + \sum_{j=0}^{1} \int_T \phi_a \left(h_T^j \left| \nabla^j \Pi_{\mathcal{T}} (p - v) \right| \right) dx \\ &\preccurlyeq \sum_{j=0}^{1} \int_{S_T} \phi_a \left(h_T^j \left| \nabla^j (v - p) \right| \right) dx. \end{aligned}$$

3.4 A Posteriori Error Estimators

There have been made many efforts for proving a posteriori error estimators for the nonlinear Dirichlet problem. In particular, Baranger and El Amri proposed in [7] a posteriori error estimators for the error in the $\|\cdot\|_{W^{1,\phi}(\Omega)}$ norm for the case $\phi(t) = \frac{1}{r} t^r$; see also [77]. These estimates naturally lack in that there is a gap between the power of the upper and the lower bound; compare with Remark 97. Recently, Liu and Yan [53, 52] proved a posteriori estimates for the error measured in the quasi-norm. In this section we shall establish the estimators of Diening and Kreuzer [28, 27], which generalize the ones of Liu and Yan; see Remark 98.

We assume that ϕ is a fixed N-function that satisfies Assumption 40. Let $\mathcal{T}$ be a conforming triangulation of the polygonal domain $\Omega \subset \mathbb{R}^d$ and $\mathring{\mathbb{V}}(\mathcal{T})$ be the corresponding finite element space.

We want to estimate the error between the Ritz-Galerkin solution $U \in \mathring{\mathbb{V}}(\mathcal{T})$ (3.18) and the true solution $u \in \mathbb{V}$ of (3.2). Existence and uniqueness of u and U is established in Theorem 49 and Proposition 83; see also (3.18). Hereafter we assume $g \in L^{\phi^*}(\Omega)^d \subset W_0^{-1,\phi}(\Omega)$. Hence,

$$\int_\Omega \mathbf{A}(\nabla u) : \nabla v \, dx = \int_\Omega g \, v \, dx \qquad \text{for all} \quad v \in W_0^{1,\phi}(\Omega)^d, \tag{3.21}$$

and

$$\int_\Omega \mathbf{A}(\nabla U) : \nabla V \, dx = \int_\Omega g \, V \, dx \qquad \text{for all} \quad V \in \mathring{\mathbb{V}}(\mathcal{T}). \tag{3.22}$$

We start from the residual $D\mathcal{J}(U)$ and use the fact that it is orthogonal on $\mathring{\mathbb{V}}(\mathcal{T})$. Hence, we have for $v \in \mathbb{V}$ and $V \in \mathring{\mathbb{V}}(\mathcal{T})$

$$\begin{aligned}
\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)\big) : \nabla v \, dx
&= \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)\big) : \big(\nabla v - \nabla V\big) \, dx \\
&= \int_\Omega g \cdot (v - V) \, dx - \int_\Omega \mathbf{A}(\nabla U) : \big(\nabla v - \nabla V\big) \, dx \\
&= \sum_{T \in \mathcal{T}} \int_T g \cdot (v - V) \, dx - \sum_{T \in \mathcal{T}} \int_{\partial T} \mathbf{A}(\nabla U) \, n_T \cdot (v - V) \, d\sigma,
\end{aligned}$$

where we used integration by parts to obtain the last equality. Observing that each interior side is shared by two triangles, we have

$$\begin{aligned}
\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)\big) : \nabla v \, dx
&= \sum_{T \in \mathcal{T}} \int_T g \cdot (v - V) \, dx - \frac{1}{2} \sum_{T \in \mathcal{T}} \int_{\partial T} [\![\mathbf{A}(\nabla U)]\!] \, n \cdot (v - V) \, d\sigma \\
&= \sum_{T \in \mathcal{T}} \int_T g \cdot (v - V) \, dx - \sum_{\sigma \in \mathcal{S}} \int_\sigma [\![\mathbf{A}(\nabla U)]\!] \, n \cdot (v - V) \, d\sigma,
\end{aligned} \tag{3.23}$$

where the *jump* $[\![\mathbf{G}]\!]$ across inter-element sides $\sigma = T \cap T' \in \mathcal{S}$ is defined as

$$[\![\mathbf{G}]\!]\, n|_\sigma := \big[\mathbf{G}|_T - \mathbf{G}|_{T'}\big] n_T|_\sigma$$

for piecewise constant functions $\mathbf{G}$ with values in $\mathbb{R}^{d\times d}$ and n_T being the outer unit normal on $\sigma \subset \partial T$. Note that the jump is well defined, i.e., for $\sigma \in \mathcal{S}$ the definition of the jump does not depend on the choice of $T \in \mathcal{T}$, $\sigma \subset T$. Since there is no jump tangential to σ, taking the norm of the jump, we can omit the outer normal. We define $|[\![\mathbf{G}]\!]\,|_\sigma| := \big|\big[\mathbf{G}|_T - \mathbf{G}|_{T'}\big]\big| = \big|\big[\mathbf{G}|_T - \mathbf{G}|_{T'}\big] n_T\big|$.

We define the local error indicator for $v \in \mathbb{V}$, $W \in \mathring{\mathbb{V}}(\mathcal{T})$ on $T \in \mathcal{T}$ by

$$\eta^2(v, W, T, g) := \int_T \big(\phi_{|\nabla v|}\big)^*(h_T\,|g|)\,dx + \int_{\partial T\cap\Omega} h_T\,|[\![\mathbf{F}(\nabla W)]\!]|^2\,d\sigma. \tag{3.24}$$

The first term in (3.24) usually is called the element-estimator, whereas the second part is called the jump-estimator. Furthermore, we define for any subset $\hat{\mathcal{T}} \subset \mathcal{T}$

$$\eta^2(v, W, \hat{\mathcal{T}}, g) := \sum_{T\in\hat{\mathcal{T}}} \eta^2(v, W, T, g).$$

Finally, we denote

$$\eta(W, \hat{\mathcal{T}}, g) := \eta(W, W, \hat{\mathcal{T}}, g).$$

3.4.1 Upper Bound

Similar to [28] we show that the error estimator is an upper bound for the error measured in the quasi-norm.

Theorem 90 (upper bound)**.** *Let u, U be the solutions of* (3.21) *and* (3.22)*, respectively. Then there exists a constant $C_1 > 0$ such that*

$$\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|_{L^2(\Omega)} \le C_1\,\eta(U, \mathcal{T}, g). \tag{3.25}$$

The constant C_1 depends solely on d, $\Delta_2(\{\phi, \phi^*\})$ and the shape-regularity of $\mathcal{T}$.

To prove Theorem 90 we need a technical auxiliary result.

Lemma 91. *Suppose the assumptions of Theorem 90. Then for arbitrary $V \in \mathbb{V}(\mathcal{T})$, $T \in S_T$, it holds*

$$\sum_{T'\in\mathcal{T}(S_T)} |\mathbf{F}(\nabla V|_T) - \mathbf{F}(\nabla V|_{T'})|^2 \preccurlyeq \sum_{\sigma\in\Sigma_T} |[\![\mathbf{F}(\nabla V)]\!]\,|_\sigma|^2\,,$$

where $\Sigma_T := \{\sigma \in \mathcal{S} : \sigma \cap S_T \neq \emptyset\}$ is the set of sides inside S_T. The constant hidden in $\preccurlyeq$ depends only on the shape regularity of $\mathcal{T}$.

Proof. We observe that for $T \in \mathcal{T}$, $T' \in \mathcal{T}(S_T)$ one can reach T' from T by passing through a finite number of faces, bounded by the shape-regularity of $\mathcal{T}$; see Figure 3.3 for an example in $d = 2$. In particular, there exist $T_1, \ldots, T_N \in \mathcal{T}$, with $T \cap T_1 = \sigma_0, \ldots, T_i \cap T_{i+1} = \sigma_i, \ldots, T_N \cap T' = \sigma_N$, $\sigma_0, \ldots, \sigma_N \in \mathcal{S}$. We set $T_0 := T$ and $T_{N+1} := T'$. Then, by the triangle inequality

$$
\begin{aligned}
|\mathbf{F}(\nabla U|_T) - \mathbf{F}(\nabla U|_{T'})| &\le \sum_{i=0}^{N} \left|\mathbf{F}(\nabla U|_{T_i}) - \mathbf{F}(\nabla U|_{T_{i+1}})\right| \\
&= \sum_{i=0}^{N} \left|[\![\mathbf{F}(\nabla U)]\!]_{\sigma_i}\right| \\
&\le \sum_{\sigma \in \Sigma_T} |[\![\mathbf{F}(\nabla U)]\!]_{\sigma}| .
\end{aligned} \tag{3.26}
$$

Therefore,

$$
\sum_{T' \in \mathcal{T}(S_T)} |\mathbf{F}(\nabla U|_T) - \mathbf{F}(\nabla U|_{T'})|^2 \preccurlyeq \sum_{T' \in \mathcal{T}(S_T)} \sum_{\sigma \in \Sigma_T} \int_{S_T} |[\![\mathbf{F}(\nabla U)]\!]_{\sigma}|^2 .
$$

We observe that the addends of the right hand side are independent of $T' \in \mathcal{T}(S_T)$. Recall further that the number of elements in S_T and hence the number of sides in Σ_T are bounded with respect to the shape-regularity of $\mathcal{T}$. This yields the assertion. □

Proof of Theorem 90. Let $\Pi_{\mathcal{T}} : \mathbb{V} \to \mathring{\mathbb{V}}(\mathcal{T})$ be the Scott-Zhang interpolation operator. Recall, that it satisfies all requirements of Assumption 84. Moreover, it preserves homogeneous boundary values, i.e., $\Pi V \in \mathring{\mathbb{V}}(\mathcal{T})$ for all $V \in \mathbb{V}$. We choose $v = e := u - U$ and $V = \Pi_{\mathcal{T}} e \in \mathring{\mathbb{V}}(\mathcal{T})$ in (3.23), i.e.,

$$
\begin{aligned}
&\int_{\Omega} \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)\big) : \nabla e \, dx \\
&\qquad = \sum_{T \in \mathcal{T}} \int_T g \cdot (e - \Pi_{\mathcal{T}} e)\, dx - \frac{1}{2} \sum_{T \in \mathcal{T}} \int_{\partial T} [\![\mathbf{A}(\nabla U)]\!] \; n \cdot (e - \Pi_{\mathcal{T}} e)\, d\sigma \\
&\qquad =: (\mathtt{Upper}_1) + (\mathtt{Upper}_2).
\end{aligned}
$$

We handle the two terms $(\mathtt{Upper}_1)$ and $(\mathtt{Upper}_2)$ separately. To estimate $(\mathtt{Upper}_1)$ let $T \in \mathcal{T}$. Then with Young's inequality (Proposition 11) for $\delta > 0$

$$
\begin{aligned}
\int_T g \cdot (e - \Pi_{\mathcal{T}} e)\, dx &\le \int_T |g|\, |e - \Pi_{\mathcal{T}} e| \; dx \\
&\le \int_T C_\delta \left(\phi_{|\nabla U|}\right)^*(h_T\, |g|) + \delta\, \phi_{|\nabla U|} \left(\frac{|e - \Pi_{\mathcal{T}} e|}{h_T}\right) dx \\
&= \int_T C_\delta \left(\phi_{|\nabla U|}\right)^*(h_T\, |g|) + \delta\, \phi_{|\nabla U|} \left(\left|\frac{e}{h_T} - \Pi_{\mathcal{T}} \frac{e}{h_T}\right|\right) dx.
\end{aligned}
$$

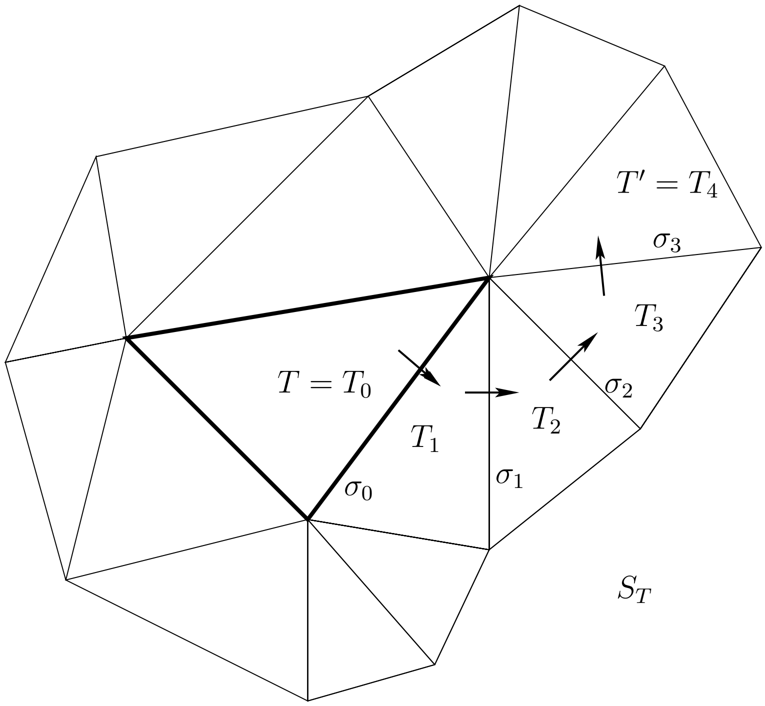

Figure 3.3: Element sides passed through from T to T'.

The constant C_δ depends on $\Delta_2(\{\phi_a\}_{a\geq 0})$ and hence on $\Delta_2(\phi)$; see Lemma 57. Now, the interpolation estimate Lemma 88 yields

$$\preccurlyeq \int_T C_\delta \left(\phi_{|\nabla U|}\right)^*(h_T\,|g|)\,dx + \delta \int_{S_T} \phi_{|\nabla U|_T|}\left(|\nabla e|\right)\,dx.$$

Note that for the last term the shift $|\nabla U|_T|$ is constant on S_T. Hence, in order to get this term compatible with the quasi-norm we shall change it on each $T' \in \mathcal{T}(S_T)$ with $T' \neq T$ to the shift $|\nabla U|_{T'}|$. We obtain according to Corollary 71 and Lemma 91

$$\begin{aligned}\int_{S_T} \phi_{|\nabla U|_T|}\left(|\nabla e|\right)\,dx &\preccurlyeq \int_{S_T} \phi_{|\nabla U|}\left(|\nabla e|\right)\,dx \\ &\quad + \sum_{T'\in\mathcal{T}(S_T)} \int_{S_T} \left|\mathbf{F}(\nabla U|_T) - \mathbf{F}(\nabla U_{T'})\right|^2\,dx. \\ &\preccurlyeq \int_{S_T} \phi_{|\nabla U|}\left(|\nabla e|\right)\,dx + \sum_{\sigma\in\Sigma_T} \left|[\![\mathbf{F}(\nabla U)]\!]\,|_\sigma\right|^2,\end{aligned} \tag{3.27}$$

where Σ_T is the set of interior sides of S_T defined in Lemma 91. Therefore,

$$\begin{aligned}(\texttt{Upper}_1) \preccurlyeq \sum_{T\in\mathcal{T}} \int_T C_\delta \left(\phi_{|\nabla U|}\right)^*(h_T\,|g|)\,dx + \delta \int_{S_T} \phi_{|\nabla U|}\left(|\nabla e|\right)\,dx \\ \delta \sum_{T\in\mathcal{T}} \sum_{\sigma\in\Sigma_T} \int_{S_T} |[\![\mathbf{F}(\nabla U)]\!]\,|_\sigma|^2\,dx.\end{aligned}$$

Observe that $|S_T| \approx |T| \approx h_\sigma\,|\sigma|$ for all $\sigma\in\Sigma_T$, where the constants hidden in $\approx$ solely depend on the shape-regularity of $\mathcal{T}$. Hence, it holds for $\sigma\in\Sigma_T$

$$\int_{S_T} |[\![\mathbf{F}(\nabla U)]\!]\,|_\sigma|^2\,dx = |S_T|\,|[\![\mathbf{F}(\nabla U)]\!]\,|_\sigma|^2 \approx \int_\sigma h_\sigma\,|[\![\mathbf{F}(\nabla U)]\!]\,|_\sigma|^2\,d\sigma.$$

Recall that the number of sides in Σ_T is bounded with respect to the shape-regularity of $\mathcal{T}$. Therefore, the finite overlapping of the S_T, $T\in\mathcal{T}$, implies

$$\begin{aligned}(\texttt{Upper}_1) \preccurlyeq C_\delta \sum_{T\in\mathcal{T}} \int_T \left(\phi_{|\nabla U|}\right)^*(h_T\,|g|)\,dx + \delta \int_\Omega \phi_{|\nabla U|}\left(|\nabla e|\right)\,dx \\ + \delta \sum_{\sigma\in\mathcal{S}} \int_\sigma h_\sigma\,|[\![\mathbf{F}(\nabla U)]\!]|^2\,d\sigma.\end{aligned} \tag{3.28}$$

To estimate the term $(\texttt{Upper}_2)$ we recall that ∇U is piecewise constant and thus $\mathbf{A}(\nabla U)$ is piecewise constant, too. By Lemma 87, then

$$\begin{aligned}(\texttt{Upper}_2) &\le \sum_{T\in\mathcal{T}} \sum_{\sigma\subset\partial T} |[\![\mathbf{A}(\nabla U)]\!]\,|_\sigma| \int_\sigma |e - \Pi_{\mathcal{T}} e|\,d\sigma \\ &\preccurlyeq \sum_{T\in\mathcal{T}} \sum_{\sigma\subset\partial T} |[\![\mathbf{A}(\nabla U)]\!]\,|_\sigma| \int_{S_T} |\nabla e|\,dx.\end{aligned}$$

Estimating the right hand side element-wise, Young's inequality (Proposition 11) yields for for $\delta > 0$

$$\begin{aligned}\sum_{\sigma\subset\partial T} &|[\![\mathbf{A}(\nabla U)]\!]\,|_\sigma| \int_{S_T} |\nabla e|\,dx \\ &\le \sum_{\sigma\subset\partial T} \Big\{ \int_{S_T} C_\delta \big(\phi_{|\nabla U|_T|}\big)^*(|[\![\mathbf{A}(\nabla U)]\!]\,|_\sigma|)\,dx \\ &\qquad + \delta \int_{S_T} \phi_{|\nabla U|_T|}(|\nabla e|)\,dx \Big\} \\ &\le \sum_{\sigma\subset\partial T} \int_{S_T} C_\delta \big(\phi_{|\nabla U|_T|}\big)^*(|[\![\mathbf{A}(\nabla U)]\!]\,|_\sigma|)\,dx \\ &\qquad + (d+1)\,\delta \int_{S_T} \phi_{|\nabla U|_T|}(|\nabla e|)\,dx.\end{aligned} \tag{3.29}$$

The constant C_δ depends on $\Delta_2(\{\phi_a\}_{a\geq 0})$ and hence on $\Delta_2(\phi)$; see Lemma 57. For the last inequality we used the fact that each element has at most $(d+1)$ sides. Recalling that $|[\![\mathbf{A}(\nabla U)]\!]|_\sigma|$ and $|\nabla U|_T|$ are constant, then by Corollary 65 for $\sigma \in \mathring{\mathcal{S}}$ and $\sigma \subset T, T' \in \mathcal{T}$

$$\begin{aligned}\big(\phi_{|\nabla U|_T|}\big)^*(|[\![\mathbf{A}(\nabla U)]\!]|_\sigma|) &= \big(\phi_{|\nabla U|_T|}\big)^*\big(\,|\mathbf{A}(\nabla U|_T) - \mathbf{A}(\nabla U|_{T'})|\,\big)\\ &\approx |[\![\mathbf{F}(\nabla U)]\!]|_\sigma|^2\,.\end{aligned}$$

Hence, by $|S_T| \approx h_\sigma\,|\sigma|$, depending on the shape regularity of $\mathcal{T}$, we have for $\sigma \in \mathring{\mathcal{S}}$, $\sigma \subset T \in \mathcal{T}$

$$\begin{aligned}\int_{S_T} \big(\phi_{|\nabla U|_T|}\big)^*(|[\![\mathbf{A}(\nabla U)]\!]|_\sigma|)\,dx &\approx \int_{S_T} |[\![\mathbf{F}(\nabla U)]\!]|_\sigma|^2\,dx\\ &\approx \int_\sigma h_\sigma\,|[\![\mathbf{F}(\nabla U)]\!]|^2\,d\sigma.\end{aligned} \tag{3.30}$$

The last term in (3.29) can be estimated as in (3.27). Altogether, this yields

$$\begin{aligned}(\texttt{Upper}_2) &\preccurlyeq \sum_{T\in\mathcal{T}} \Big\{C_\delta \sum_{\sigma\subset\partial T}\int_\sigma h_\sigma\,|[\![\mathbf{F}(\nabla U)]\!]|\,d\sigma + \delta \sum_{\sigma\in\Sigma_T}\int_\sigma h_\sigma\,|[\![\mathbf{F}(\nabla U)]\!]|\,d\sigma\\ &\qquad + \delta\int_{S_T}\phi_{|\nabla U|}(|\nabla e|)\,dx\Big\}\\ &\leq \sum_{T\in\mathcal{T}}\Big\{(C_\delta+\delta)\sum_{\sigma\in\Sigma_T}\int_\sigma h_\sigma\,|[\![\mathbf{F}(\nabla U)]\!]|\,d\sigma + \delta\int_{S_T}\phi_{|\nabla U|}(|\nabla e|)\,dx\Big\}\end{aligned}$$

The number of overlaps of S_T, $T \in \mathcal{T}$ as well as the number of sides $\sigma \in \Sigma_T$ are bounded with respect to the shape regularity of $\mathcal{T}$. Hence, we get

$$(\texttt{Upper}_2) \preccurlyeq (\delta + C_\delta)\sum_{\sigma\in\mathcal{S}}\int_\sigma h_\sigma\,|[\![\mathbf{F}(\nabla U)]\!]|\;d\sigma + \delta\int_\Omega \phi_{|\nabla U|}(|\nabla e|)\,dx. \tag{3.31}$$

Thus, combining (3.28) and (3.31) yields

$$\begin{aligned}\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)\big) : \nabla e\,dx &\preccurlyeq C_\delta \sum_{T\in\mathcal{T}}\int_T \big(\phi_{|\nabla U|}\big)^*(h_T\,|g|)\,dx\\ &\quad + (\delta + C_\delta)\sum_{\sigma\in\mathcal{S}}\int_\sigma h_\sigma\,|[\![\mathbf{F}(\nabla U)]\!]|^2\,d\sigma\\ &\quad + \delta\int_\Omega \phi_{|\nabla U|}(|\nabla e|)\,dx.\end{aligned}$$

Recalling Lemma 74, we have

$$\int_\Omega \phi_{|\nabla U|}(|\nabla e|)\,dx \approx \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|^2_{L^2(\Omega)} \approx \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)\big) : \nabla e\,dx.$$

Therefore, it follows

$$\begin{aligned}\|\mathbf{F}(\nabla u)-\mathbf{F}(\nabla U)\|_{L^2(\Omega)}^2 \preccurlyeq C_\delta \sum_{T\in\mathcal{T}} \int_T \left(\phi_{|\nabla U|}\right)^*(h_T\,|g|)\,dx \\ + (\delta + C_\delta)\sum_{\sigma\in\mathcal{S}}\int_\sigma h_\sigma\,|[\![\mathbf{F}(\nabla U)]\!]|^2\,d\sigma \\ + \delta\,\|\mathbf{F}(\nabla u)-\mathbf{F}(\nabla U)\|_{L^2(\Omega)}^2\,.\end{aligned}$$

Now, we can subtract the last term at the left hand side. Choosing δ small enough yields the desired estimate. □

Remark 92. *Note that in Lemma 88 it is crucial that $a \geq 0$ is constant; see also [31]. For this reason, our finite element spaces are restricted to piecewise linear polynomials, since this implies that the gradient is piecewise constant and thus can be used as shift.*

Moreover, for $T \in \mathcal{T}$ we need $a \geq 0$ to be constant on the whole patch S_T. We take $|\nabla V|$ as shift for some functions $V \in \mathring{\mathbb{V}}(\mathcal{T})$. This causes problems, since ∇V is piecewise constant on T, but may jump across inter-element sides. Hence, by Lemma 88, we find

$$\sum_{j=0}^{1}\int_T \phi_{|\nabla V|}\left(h_T^j\,|\nabla^j(v-\Pi_{\mathcal{T}}v)|\right)\,dx \leq C\int_{S_T}\phi_{|\nabla V|_T|}\left(h_T\,|\nabla v|\right)\,dx.$$

Recalling the proof of Theorem 90, this drawback can be overcome by a change of the shift and estimating the perturbation term by the jump of $\mathbf{F}(\nabla V)$ over inter-element sides; compare also Lemma 91. This term is proportional to the jump estimator.

3.4.2 Lower Bound

The proof of efficiency is based on the idea of Verfürth [75] of testing the residual by certain locally supported, nonnegative *bubble* functions; see also [77, 76]. We consider two types of bubble functions. *Interior* bubble functions, supported on a single element and *side* bubble functions supported on a pair of elements; see [77, 3].

Let $\hat\lambda_0,\ldots,\hat\lambda_d$ be the barycentric coordinates of the reference triangle $\hat T$. We define the interior bubble function on $\hat T$ by

$$\hat\psi := \frac{1}{d!}\,\frac{\hat\lambda_0\cdots\hat\lambda_d}{\int_{\hat T}\hat\lambda_0\cdots\hat\lambda_d\,d\hat x}.$$

For $i=0,\ldots,d$, let $\hat\sigma_i :=$ conv hull$\{e_0,\ldots e_{i-1},e_{i+1},\ldots e_d\}$ be the $d-1$ sub-simplex of $\hat T$ opposite to the node e_i. The side bubble function corresponding to

$\hat{\sigma}_i$ is then given by

$$\hat{\chi}_i := \frac{1}{(d-1)!} \frac{\hat{\lambda}_0 \cdots \hat{\lambda}_{i-1}\hat{\lambda}_{i+1} \cdots \hat{\lambda}_d}{\int_{\hat{\sigma}_i} \hat{\lambda}_0 \cdots \hat{\lambda}_{i-1}\hat{\lambda}_{i+1} \cdots \hat{\lambda}_d \, d\hat{\sigma}_i}.$$

The next step is to construct bubble functions on the physical elements. For a conforming triangulation $\mathcal{T}$ of Ω let for each $T \in \mathcal{T}$ the mapping $F_T : \hat{T} \to T$ as described in Section 3.3.1. We define the interior bubble function of $T \in \mathcal{T}$ by

$$\psi_T := \begin{cases} \hat{\psi} \circ F_T^{-1}, & \text{in } T, \\ 0, & \text{elsewhere.} \end{cases}$$

For the side bubble function let $\sigma \in \mathring{\mathcal{S}}$ and $T_1, T_2 \in \mathcal{T}$ be the elements with $T_1 \cap T_2 = \sigma$. Let further $i, j \in \{0, \ldots, d\}$ such that $\sigma = F_{T_1}(\sigma_i) = F_{T_2}(\sigma_j)$. Then we define the side bubble function χ_σ by

$$\chi_\sigma := \begin{cases} \hat{\chi}_i \circ F_{T_1}^{-1}, & \text{in } T_1, \\ \hat{\chi}_j \circ F_{T_2}^{-1}, & \text{in } T_2, \\ 0, & \text{elsewhere.} \end{cases}$$

Note that ψ_T and χ_σ are continuous piece-wise polynomials with zero boundary values on T, ω_σ, respectively. Hence, we obtain $\psi_T \in W_0^{1,\phi}(T)$ and $\chi_\sigma \in W_0^{1,\phi}(\omega_\sigma)$. The following lemma collects some properties of the bubble functions that can easily be deduced from their definition; see also [77, 3].

Lemma 93. *Let $\mathcal{T}$ be a conforming triangulation of Ω. Then there exists a constant $C > 0$ depending solely on the shape-regularity of $\mathcal{T}$, such that for all $T \in \mathcal{T}$, $\sigma \in \mathcal{S}$, $\psi_T \in W_0^{1,\phi}(T)$, $\chi_\sigma \in W_0^{1,\phi}(\omega_\sigma)$ and*

$$\int_T \psi_T \, dx = |T|, \qquad \|\psi_T\|_{L^\infty(T)} \le C, \qquad \|\nabla \psi_T\|_{L^\infty(T)} \le \frac{C}{h_T},$$
$$\int_\sigma \chi_\sigma \, d\sigma = |\sigma|, \qquad \|\chi_\sigma\|_{L^\infty(\omega_\sigma)} \le C, \qquad \|\nabla \chi_\sigma\|_{L^\infty(\omega_\sigma)} \le \frac{C}{h_\sigma}.$$

Proof. We prove only the assertions for the element bubble function, since the proofs for the side bubble function work in the same fashion. The first claim follows from transforming the bubble function onto the standard simplex $\hat{T}$

$$\int_T \psi_T \, dx = \int_{\hat{T}} \psi_T \circ F_T \, |\det DF_T| \, d\hat{x} = |\det DF_T| \int_{\hat{T}} \hat{\psi} \, d\hat{x} = \frac{|\det DF_T|}{d!}.$$

Observing that $|\det DF_T| = d!\,|T|$, yields the assertion. The second claim follows from $\|\psi_T\|_{L^\infty(T)} = \|\hat{\psi}\|_{L^\infty(\hat{T})}$ for all $T \in \mathcal{T}$ and the third claim follows by an inverse estimate. □

The concept of oscillation plays a fundamental role in the efficiency of the estimator. Since it is not possible to numerically evaluate the dual quasi-norm of the residual on an infinite dimensional space we estimate it by the computable quantity $\eta(U,\mathcal{T},g)$; see Remark 79 for the concept of the dual quasi-norm. In particular, the estimator uses the L^{ϕ^*}-regularity of the residual, which induces a stronger topology than the topology on $W_0^{-1,\phi}(\Omega)$; recall that $g \in L^{\phi^*}(\Omega)$ is assumed. This defect conditions the oscillation as a correction term in the lower bound Lemma 95.

For $v \in \mathbb{V}, T \in \mathcal{T}$, and $g \in L^{\phi^*}(\Omega)$, we define the oscillation by

$$\operatorname{osc}^2(v,T,g) := \int_T \left(\phi_{|\nabla v|}\right)^*(h_T\,|g-g_T|)\,dx,$$

where $g_T \in \mathbb{R}$ such that the expression becomes minimal. Observe that $g_T \in \mathbb{R}$ is uniquely defined, since the function $\int_T \left(\phi_{|\nabla v|}\right)^*(h_T\,|g-c|)\,dx \in \mathbb{R}$ is strictly convex in $c \in \mathbb{R}$ and tends to infinity as $|c|$ tends to infinity. We define for any subset $\hat{\mathcal{T}} \subset \mathcal{T}$

$$\operatorname{osc}^2(v,\hat{\mathcal{T}},g) := \sum_{T\in\hat{\mathcal{T}}} \operatorname{osc}^2(v,T,g).$$

Remark 94. *Note that oscillation is dominated by the estimator, since*

$$\operatorname{osc}^2(v,T,g) = \inf_{c\in\mathbb{R}} \int_T \left(\phi_{|\nabla v|}\right)^*(h_T\,|g-c|)\,dx \le \int_T \left(\phi_{|\nabla v|}\right)^*(h_T\,|g-0|)\,dx.$$

The last term corresponds to the element-estimator and is therefore dominated by $\eta^2(v,V,T,g)$ for any $V \in \mathbb{V}(\mathcal{T})$.

Now, we are prepared to state the lower estimate for the residual.

Theorem 95 (lower bound)**.** *Let u,U be the solutions of* (3.21) *and* (3.18)*, respectively. Then there exists constants $C_2,\tilde{C}_2 > 0$ such that for all $T \in \mathcal{T}$*

$$C_2\,\eta(U,T,g) \le \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|_{L^2(\omega_T)} + \operatorname{osc}(U,\mathcal{T}(\omega_T),g)$$

and

$$\tilde{C}_2\,\eta(U,T,g) \le \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|_{L^2(\omega_T)} + \operatorname{osc}(u,\mathcal{T}(\omega_T),g).$$

The constants $C_2,\tilde{C}_2$ depend solely on d, $\Delta_2(\{\phi,\phi^*\})$, and the shape-regularity of $\mathcal{T}$.

Proof. We start with estimating the element-estimator. Let $g_T \in \mathbb{R}^d$ be arbitrary. We observe that (2.4) also holds in the d-dimensional case, i.e., there exists $s_T \in \mathbb{R}^d$ such that

$$h_T\,g_T \cdot s_T = \left(\phi_{|\nabla U|_T|}\right)^*(h_T\,|g_T|) + \phi_{|\nabla U|_T|}(|s_T|),$$

Again we used that $\nabla U|_T = \nabla U|_T$ is constant. Recalling that $\psi_T \in W_0^{1,\phi}(T) \subset W_0^{1,\phi}(\Omega)$, we have $s_T\psi_T \in W_0^{1,\phi}(T)^d \subset W_0^{1,\phi}(\Omega)^d$. Hence, with the help of Lemma 93 and (3.23)

$$
\begin{aligned}
|T| \big(\phi_{|\nabla U|_T|}\big)^*(h_T\,|g_T|) + |T|\,\phi_{|\nabla U|_T|}(|s_T|) &= |T|\,s_T \cdot (h_T g_T) \\
&= \int_T h_T\, g_T \cdot s_T\, \psi_T\, dx = \int_T g \cdot h_T\, s_T\, \psi_T\, dx + \int_T h_T\,(g_T - g)\cdot s_T\,\psi_T\, dx \\
&= \int_T \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)\big) : \nabla(h_T\, s_T\, \psi_T)\, dx + \int_T h_T\,(g_T - g)\cdot s_T\,\psi_T\, dx \\
&\le \int_T |\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)|\,|s_T|\; h_T\; \|\nabla \psi_T\|_{L^\infty(T)}\; dx \\
&\qquad + \int_T h_T\; |g_T - g|\,|s_T|\,\|\psi_T\|_{L^\infty(T)}\; dx \\
&\le C \int_T |\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)|\,|\, s_T|\; dx + C \int_T h_T\; |g_T - g|\,|s_T|\; dx \\
&=: (\mathtt{Lower}_1) + (\mathtt{Lower}_2).
\end{aligned}
$$

Now, applying Young's inequality (Proposition 11) we get for $\delta > 0$

$$
(\mathtt{Lower}_1) \preccurlyeq \int_T C_\delta\,\big(\phi_{|\nabla U|}\big)^*\big(\,|\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)|\,\big) + \delta\,\phi_{|\nabla U|}(|(s_T\,\psi_T)|)\, dx.
$$

The first term can be estimated with Corollary 65

$$
\begin{aligned}
\int_T \big(\phi_{|\nabla U|}\big)^*\big(\,|\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)|\,\big)\, dx &\approx \int_T \big(\phi_{|\nabla U|}\big)^*\big(\phi'_{|\nabla U|}(|\nabla u - \nabla U|)\big)\, dx \\
&\approx \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|^2_{L^2(T)}\,.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
(\mathtt{Lower}_1) &\preccurlyeq C_\delta\,\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|^2_{L^2(T)} + \delta \int_T \phi_{|\nabla U|}(|s_T|)\, dx \\
&= C_\delta\,\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|^2_{L^2(T)} + \delta\,|T|\,\phi_{|\nabla U|_T|}(|s_T|).
\end{aligned}
\tag{3.32}
$$

Similarly, Young's inequality (Proposition 11) and Lemma 93 yield for the second term $(\mathtt{Lower}_2)$

$$
\begin{aligned}
(\mathtt{Lower}_2) &\preccurlyeq \int_T h_T\; |g_T - g|\,|s_T|\; dx \\
&\le C_\delta \int_T \big(\phi_{|\nabla U|}\big)^*(h_T\,|g_T - g|)\, dx + \delta\; |T|\,\phi_{|\nabla U|_T|}(|s_T|).
\end{aligned}
\tag{3.33}
$$

The constant C_δ depends on $\Delta_2(\{\phi_a\}_{a\geq 0})$ and hence on $\Delta_2(\phi)$; see Lemma 57. Combining (3.32) and (3.33) we get

$$\begin{aligned}|T|\,\big(\phi_{|\nabla U|_T|}\big)^*(h_T\,|g_T|) + |T|\,\phi_{|\nabla U|_T|}(|s_T|) &\\ \preceq C_\delta\,\|\mathbf{F}(\nabla u)-\mathbf{F}(\nabla U)\|^2_{L^2(T)} + C_\delta\int_T \big(\phi_{|\nabla U|}\big)^*(h_T\,|g_T-g|)\,dx &\\ +\,\delta\,|T|\,\phi_{|\nabla U|_T|}(|s_T|),&\end{aligned}$$

hence, choosing $\delta>0$ small enough, this yields

$$\int_T \big(\phi_{|\nabla U|}\big)^*(h_T\,|g_T|) \preceq \|\mathbf{F}(\nabla u)-\mathbf{F}(\nabla U)\|^2_{L^2(T)} + \int_T \big(\phi_{|\nabla U|}\big)^*(h_T\,|g_T-g|)\,dx.$$

The triangle like inequality of Corollary 10 implies

$$\int_T \big(\phi_{|\nabla U|}\big)^*(h_T\,|g|)\,dx \preceq \int_T \big(\phi_{|\nabla U|}\big)^*(h_T\,|g_T|) + \big(\phi_{|\nabla U|}\big)^*(h_T\,|g_T-g|)\,dx$$

Recalling that $g_T\in\mathbb{R}$ was arbitrary, we obtain

$$\int_T \big(\phi_{|\nabla U|}\big)^*(h_T\,|g|)\,dx \preceq \|\mathbf{F}(\nabla u)-\mathbf{F}(\nabla U)\|^2_{L^2(T)} + \operatorname{osc}^2(U,T,g). \tag{3.34}$$

It remains to estimate the jump-estimator. Let $\sigma\in\mathcal{S}$, $\sigma\subset T$ and recall from Corollary 65

$$\begin{aligned}\big(\phi_{|\nabla U|_T|}\big)^*(|[\![\mathbf{A}(\nabla U)]\!]_\sigma|) &\approx \big(\phi_{|\nabla U|_T|}\big)^*\big(\phi'_{|\nabla U|_T|}(|[\![\nabla U]\!]\,|_\sigma|)\big)\\ &\approx \phi_{|\nabla U|_T|}(|[\![\nabla U]\!]\,|_\sigma|)\\ &\approx |[\![\mathbf{F}(\nabla U)]\!]\,|_\sigma|\,.\end{aligned} \tag{3.35}$$

As in the estimate of the element-estimator, there exists $s_\sigma\in\mathbb{R}^d$ such that Young's inequality is sharp (see also (2.4)), i.e.,

$$[\![\mathbf{A}(\nabla U)]\!]\,n|_\sigma\cdot s_\sigma = \big(\phi_{|\nabla U|_T|}\big)^*\big(\,|[\![\mathbf{A}(\nabla U)]\!]\,|_\sigma|\,\big) + \phi_{|\nabla U|_T|}(|s_\sigma|).$$

Recalling that $\chi_\sigma\in W^{1,\phi}_0(\omega_\sigma)$ we have from Lemma 93 and (3.23)

$$\begin{aligned}h_\sigma\,|\sigma|\,\big(\phi_{|\nabla U|_T|}\big)^*\big(\,|[\![\mathbf{A}(\nabla U)]\!]\,|_\sigma|\,\big) + h_\sigma\,|\sigma|\,\phi_{|\nabla U|_T|}(|s_\sigma|) &= h_\sigma\,|\sigma|\,[\![\mathbf{A}(\nabla U)]\!]\,n|_\sigma\cdot s_\sigma\\ &= h_\sigma\int_\sigma [\![\mathbf{A}(\nabla U)]\!]\,n|_\sigma\cdot s_\sigma\chi_\sigma\,d\sigma\\ &= -\int_{\omega_\sigma}\big(\mathbf{A}(\nabla u)-\mathbf{A}(\nabla U)\big):\nabla(s_\sigma h_\sigma\chi_\sigma)\,dx + \int_{\omega_\sigma} h_\sigma g\cdot s_\sigma\chi_\sigma\,dx\\ &\leq \int_{\omega_\sigma}|\mathbf{A}(\nabla u)-\mathbf{A}(\nabla U)|\,|s_\sigma\nabla(h_\sigma\chi_\sigma)|\,dx + \int_{\omega_\sigma} h_\sigma\,|g|\,|s_\sigma\chi_\sigma|\,dx\\ &\leq C\int_{\omega_\sigma}|\mathbf{A}(\nabla u)-\mathbf{A}(\nabla U)|\,|s_\sigma|\,dx + C\int_{\omega_\sigma} h_\sigma\,|g|\,|s_\sigma|\,dx\\ &= (\mathtt{Lower}_3) + (\mathtt{Lower}_4).\end{aligned}$$

We estimate the two terms separately. For the first one we have with Young's inequality (Proposition 11) for $\delta > 0$

$$(\texttt{Lower}_3) \preccurlyeq \int_{\omega_\sigma} C_\delta \left(\phi_{|\nabla U|}\right)^*(|\mathbf{A}(\nabla u) - \nabla \mathbf{A}(\nabla U)|) + \delta\, \phi_{|\nabla U|}(|s_\sigma|)\, dx.$$

The constant C_δ depends on $\Delta_2(\{\phi_a\}_{a\geq 0})$ and hence on $\Delta_2(\phi)$; see Lemma 57. Corollary 65 then yields

$$(\texttt{Lower}_3) \preccurlyeq C_\delta \int_{\omega_\sigma} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla(U)|^2\, dx + \delta \int_{\omega_\sigma} \phi_{|\nabla U|}(|s_\sigma|)\, dx. \tag{3.36}$$

Similarly, for the second term $(\texttt{Lower}_4)$

$$(\texttt{Lower}_4) \preccurlyeq \int_{\omega_\sigma} C_\delta \left(\phi_{|\nabla U|}\right)^*(h_\sigma\, |g|) + \delta\, \phi_{|\nabla U|}(|s_\sigma|)\, dx. \tag{3.37}$$

Now, (3.36) and (3.37) imply

$$\begin{aligned} h_\sigma\, |\sigma| \left(\phi_{|\nabla U|_T|}\right)^*\big(\, |[\![\mathbf{A}(\nabla U)]\!]_\sigma|\,\big) &+ h_\sigma\, |\sigma|\, \phi_{|\nabla U|_T|}(|s_\sigma|) \\ &\preccurlyeq C_\delta \int_{\omega_\sigma} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla(U)|^2\, dx + C_\delta \int_{\omega_\sigma} \left(\phi_{|\nabla U|}\right)^*(h_\sigma\, |g|)\, dx \\ &\quad + \delta \int_{\omega_\sigma} \phi_{|\nabla U|}(|s_\sigma|)\, dx. \end{aligned}$$

To absorb the last term at the right hand side we need the constant shift $|\nabla U|_T|$ on ω_σ. Let $\tilde{T} \in \mathcal{T}$ be the other element adjacent to σ, i.e., $T \cap \tilde{T} = \sigma$ and $T \cup \tilde{T} = \omega_\sigma$. Then, $|\mathbf{F}(\nabla U|_{\tilde{T}}) - \mathbf{F}(\nabla U|_T)| = |[\![\mathbf{F}(\nabla U)]\!]\, |_\sigma|$ and hence we get with Corollary 69

$$\begin{aligned} h_\sigma\, |\sigma| \left(\phi_{|\nabla U|_T|}\right)^*\big(\, |[\![\mathbf{A}(\nabla U)]\!]_\sigma|\,\big) &+ h_\sigma\, |\sigma|\, \phi_{|\nabla U|_T|}(|s_\sigma|) \\ &\preccurlyeq C_\delta \int_{\omega_\sigma} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla(U)|^2\, dx + C_\delta \int_{\omega_\sigma} \left(\phi_{|\nabla U|}\right)^*(h_\sigma\, |g|)\, dx \\ &\quad + \delta \Big\{ \int_{\omega_\sigma} \phi_{|\nabla U|_T|}(|s_\sigma|) + |[\![\mathbf{F}(\nabla U)]\!]\, |_\sigma|\, dx \Big\}. \end{aligned}$$

Recall (3.35) and that $|\omega_\sigma| \approx h_\sigma\, |\sigma|$, with the constants hidden in $\approx$ solely depending on the shape-regularity of $\mathcal{T}$. Therfore, we get

$$\begin{aligned} h_\sigma\, |\sigma| \left(\phi_{|\nabla U|_T|}\right)^*\big(\, |[\![\mathbf{A}(\nabla U)]\!]_\sigma|\,\big) &+ h_\sigma\, |\sigma|\, \phi_{|\nabla U|_T|}(|s_\sigma|) \\ &\preccurlyeq C_\delta \int_{\omega_\sigma} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla(U)|^2\, dx + C_\delta \int_{\omega_\sigma} \left(\phi_{|\nabla U|}\right)^*(h_\sigma\, |g|)\, dx \\ &\quad + \delta\, h_\sigma\, |\sigma|\, \phi_{|\nabla U|_T|}(|s_\sigma|) + \delta\, h_\sigma\, |\sigma| \left(\phi_{|\nabla U|_T|}\right)^*\big(\, |[\![\mathbf{A}(\nabla U)]\!]_\sigma|\,\big). \end{aligned}$$

Now, choosing δ small enough, we obtain

$$h_\sigma\,|\sigma|\,\big(\phi_{|\nabla U|_T|}\big)^*\big(\,|[\![\mathbf{A}(\nabla U)]\!]\,|_\sigma|\,\big) \preccurlyeq \int_{\omega_\sigma} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla(U)|^2\,dx + \int_{\omega_\sigma} \big(\phi_{|\nabla U|}\big)^*(h_\sigma\,|g|)\,dx.$$

Since $h_\sigma \approx h_T \approx h_{\tilde{T}}$ for each of the two triangles $T, \tilde{T}$ adjacent to σ, the last term is equivalent to the element residual. Therefore, we can apply (3.34) element-wise to get

$$\int_\sigma \big(\phi_{|\nabla U|_T|}\big)^*\big(\,|[\![\mathbf{A}(\nabla U)]\!]\,|_\sigma|\,\big)\,d\sigma \preccurlyeq \int_{\omega_\sigma} |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla(U)|^2\,dx + \operatorname{osc}^2(U, \mathcal{T}(\omega_\sigma), g).$$

Now, summing this estimate over all $\sigma \in \mathcal{S}$, $\sigma \subset T$ together with (3.34) proves the first assertion.

To prove the second claim, we observe with Corollary 71 that

$$\int_T \big(\phi_{|\nabla U|}\big)^*(|g - g_T|)\,dx \preccurlyeq \int_T |\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)|^2\,dx + \int_T \big(\phi_{|\nabla u|}\big)^*(|g - g_T|)\,dx$$

for all $g_T \in \mathbb{R}$ and all $T \in \mathcal{T}$. Taking the infimum over all $g_T \in \mathbb{R}$ and substituting this into the first estimate yields the desired assertion. □

The lower estimates above are local. Summing over all $T \in \mathcal{T}$ and taking into account the finite overlapping of the ω_T immediately yield global versions.

Corollary 96. *Let u, U be the solutions of* (3.21) *and* (3.18)*, respectively. Then, it holds with the same constants $C_2, \tilde{C}_2 > 0$ as in Theorem 95*

$$C_2\,\eta(U, \mathcal{T}, g) \le \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|_{L^2(\Omega)} + \operatorname{osc}(U, \mathcal{T}, g)$$

and

$$\tilde{C}_2\,\eta(U, \mathcal{T}, g) \le \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|_{L^2(\Omega)} + \operatorname{osc}(u, \mathcal{T}, g).$$

Remark 97. *Former a posteriori estimates for the error in the energy norm lack in a gap in the power of the upper and the lower bound; see [7, 77, 74]. This gap is induced from the gap between the dual norm of the residual and the energy norm (see Remark 79 and [74]) and therefore cannot be avoided.*

Remark 98. *Liu and Yan proved in [52, 53] similar estimates for the case $\phi(t) = \frac{1}{r}t^r$, $r \in (1, \infty)$. In particular, they show*

$$\|u - U\|_{(r)}^2 \preccurlyeq (\eta_1^2 + \eta_2^2) + \eta^2$$
$$\eta_1^2 + \eta_2^2 \preccurlyeq \|u - U\|_{(r)}^2 + \epsilon^2,$$

where $\frac{1}{r} + \frac{1}{r'} = 1$ *and*

$$\begin{aligned}
\|u - U\|_{(r)}^2 &= \int_\Omega (|\nabla U| + |\nabla(u-U)|)^{r-2} |\nabla(u-U)|^2 \, dx, \\
\eta_1^2 &= \sum_{T \in \mathcal{T}} \int_T (|\nabla U|^{r-1} + h_T |g|)^{r'-2} h_T^2 |g|^2 \, dx, \\
\eta_2^2 &= \sum_{\sigma \in \mathcal{S}} \int_{\omega_\sigma} (|\nabla U|^{r-1} + |[\![\mathbf{A}(\nabla U)]\!]|_\sigma|)^{r'-2} |[\![\mathbf{A}(\nabla U)]\!]|_\sigma|^2 \, d\sigma, \\
\eta^2 &= \sum_{\sigma \in \mathcal{S}} \int_{\omega_\sigma} (|\nabla U| + |[\![\nabla U]\!]|_\sigma|)^{r-2} |[\![\nabla U]\!]|_\sigma|^2 \, d\sigma,
\end{aligned}$$

and

$$\epsilon^2 = \sum_{T \in \mathcal{T}} \int_T (|\nabla U|^{r-1} + h_T |g - g_T|)^{r'-2} h_T^2 |g - g_T|^2 \, dx.$$

In [53, 52] the contributions η_2 *and* η *are defined by integrating over a particularly chosen simplex in* $\mathcal{T}(\omega_\sigma)$, $\sigma \in \mathcal{S}$. *We neglected this special choice, since it is just a matter of constants: For fixed* $\sigma \in \mathcal{S}$ *let* $\{T_1, T_2\} = \mathcal{T}(\omega_\sigma)$. *Then, the triangle inequality yields*

$$\begin{aligned}
|\nabla U(T_1)| + |[\![\nabla U]\!]|_\sigma| &= |\nabla U(T_1)| + |\nabla U(T_1) - \nabla U(T_2)| \\
&\approx |\nabla U(T_2)| + |\nabla U(T_1) - \nabla U(T_2)| \\
&= |\nabla U(T_2)| + |[\![\nabla U]\!]|_\sigma| .
\end{aligned}$$

For η_2 *a similar argument applies. Thus, the above estimators are equivalent to the ones of Liu and Yan.*

We will now show that that our estimates generalize those of Liu and Yan. As we observed in Remark 75 and Lemma 74, it holds

$$\|u - U\|_{(r)}^2 \approx \int_\Omega \phi_{|\nabla U|}(|\nabla u - \nabla U)|) \, dx = \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|_{L^2(\Omega)}^2 .$$

Furthermore, we have by

$$\phi'(t) = t^{r-1} \qquad \text{and} \qquad (\phi^*)''(t) = (r'-1)\, t^{r'-2},$$

that

$$\begin{aligned}
\eta_1^2 &\approx \sum_{T \in \mathcal{T}} \int_T (\phi^*)''(\phi'(|\nabla U|) + h_T |g|)\, h_T^2 |g|^2 \, dx \\
&\approx \sum_{T \in \mathcal{T}} \int_T (\phi^*)_{\phi'(|\nabla U|}(h_T |g|) \, dx \\
&\approx \sum_{T \in \mathcal{T}} \int_T \left(\phi_{|\nabla U|}\right)^*(h_T |g|) \, dx,
\end{aligned}$$

where we used the estimates of Proposition 62 and Lemma 60. Hence η_1 is equivalent to the element-estimator. In the same way it can be shown that ϵ is equivalent to $\operatorname{osc}(U,\mathcal{T},g)$.

To handle the last two terms, η_2 and η, we observe by similar estimates as for η_1 that

$$\begin{aligned}(|\mathbf{Q}|^{r-1} + |\mathbf{A}(\mathbf{Q}) - \mathbf{A}(\mathbf{P})|)^{r'-2}\,|\mathbf{A}(\mathbf{Q}) - \mathbf{A}(\mathbf{P})|^2 \\ &\approx \phi''\big(\phi'(|\mathbf{Q}|) + |\mathbf{A}(\mathbf{Q}) - \mathbf{A}(\mathbf{P})|\big)\,|\mathbf{A}(\mathbf{Q}) - \mathbf{A}(\mathbf{P})|^2 \\ &\approx \big(\phi_{|\mathbf{Q}|}\big)^*(|\mathbf{A}(\mathbf{Q}) - \mathbf{A}(\mathbf{P})|) \\ &\approx |\mathbf{F}(\mathbf{P}) - \mathbf{F}(\mathbf{Q})|^2\,,\end{aligned}$$

for all $\mathbf{P},\mathbf{Q}\in\mathbb{R}^{d\times d}$, where the last estimate is shown in Corollary 65. Furthermore, Proposition 62 yields

$$|\mathbf{F}(\mathbf{P}) - \mathbf{F}(\mathbf{Q})|^2 \approx \phi_{|\mathbf{Q}|}(|\mathbf{P}-\mathbf{Q}|) \approx (|\mathbf{Q}| + |\mathbf{P}-\mathbf{Q}|)^{r-2}\,|\mathbf{P}-\mathbf{Q}|^2$$

for all $\mathbf{P},\mathbf{Q}\in\mathbb{R}^{d\times d}$. Hence, η_2 as well as η are equivalent to the jump estimator.

Summarizing, Theorems 90 and 95 generalize the estimates of Liu and Yan to more general N-functions; see [53, 52]. Moreover, they avoid unnecessary terms and clarify the presentation.

3.5 Adaptive Finite Elements

Although adaptive finite elements have been a powerful tool of engineering and scientific computing for about three decades, the convergence analysis is rather recent. It started with Dörfler [36], who introduced a crucial marking, from now on called Dörflers marking. Later Morin, Nochetto, and Siebert [57, 58] established linear convergence for linear elliptic problems. The first plain convergence result for the nonlinear Poisson equation is due to Veeser [74]. Further convergence results can be found in [20, 55, 19, 61, 60, 70]; see also Remark 111.

In Section 3.5.1 we introduce an adaptive finite element method (AFEM) for the nonlinear Poisson equation. Then, after some auxiliary results in Section 3.5.2 the main result in Section 3.5.3, which is basically from [28, 27], states linear convergence of AFEM. Finally, the section is closed by a result on the quasi-optimal convergence rate of AFEM based on the results in [71, 19, 27].

For the remainder of this chapter we assume that the polygonal domain $\Omega \subset \mathbb{R}^d$ is triangulated by a conforming initial triangulation $\mathcal{T}_0$.

3.5.1 Adaptive Finite Element Method (AFEM)

The adaptive finite element method AFEM for the nonlinear Poisson equation (3.21) consists of a loop

$$\mathsf{SOLVE} \to \mathsf{ESTIMATE} \to \mathsf{MARK} \to \mathsf{REFINE}.$$

The procedure SOLVE calculates the Ritz Galerkin solution. For any conforming triangulation $\mathcal{T}$ of Ω we suppose that the routine SOLVE outputs the exact Ritz-Galerkin solution $U \in \mathring{\mathbb{V}}(\mathcal{T})$ of (3.22) with right hand side $g \in L^{\phi^*}(\Omega)$

$$U = \mathsf{SOLVE}(\mathcal{T}, g).$$

Next, the error between the discrete solution U and the continuous solution u of (3.21) is estimated by ESTIMATE. We assume that, given a conforming triangulation $\mathcal{T}$ of Ω, the finite element solution $U \in \mathring{\mathbb{V}}(\mathcal{T})$, and the right hand side $g \in L^{\phi^*}(\Omega)$ of (3.21), the procedure ESTIMATE outputs the error indicators (3.24)

$$\{\eta(U, T, g)\}_{T \in \mathcal{T}} = \mathsf{ESTIMATE}(U, \mathcal{T}, g).$$

In the selection of elements for refinement we rely on Dörfler marking. Given a grid $\mathcal{T}$, the set of indicators $\{\eta(U, T, g)\}_{T \in \mathcal{T}}$, and a marking parameter $\theta \in (0, 1]$, we suppose that MARK outputs a subset $\mathcal{M} \subset \mathcal{T}$ of marked elements, i.e.,

$$\mathcal{M} = \mathsf{MARK}(\{\eta(U, T, g)\}_{T \in \mathcal{T}}, \mathcal{T}, \theta),$$

such that $\mathcal{M}$ satisfies the Dörfler property

$$\eta(U, \mathcal{M}, g) \geq \theta\, \eta(U, \mathcal{T}, g).$$

Refinement is based on shape-regular bisection of single elements. Any given d simplex is subdivided into two sub-simplices of the same size such that the minimal angle is uniformly bounded from below. We do not go too much into detail of refining routines and just assume that there exists a procedure REFINE, that produces a conforming refinement of a given triangulation $\mathcal{T}$ based on a certain subset $\mathcal{M} \subset \mathcal{T}$ of marked elements and an integer b. In particular, let

$$\mathcal{T}_* = \mathsf{REFINE}(\mathcal{T}, \mathcal{M}, b),$$

then $\mathcal{T}_*$ is a conforming triangulation of Ω such that for $T \in \mathcal{M}$ the set $\mathcal{T}_*(T)$ has at least 2^b elements, i.e., T is at least bisected b times. Moreover, bisection implies the mesh-size reduction of the refined elements $T' \in \mathcal{T}_*(T)$, $T \in \mathcal{M}$,

$$|T'| \leq 2^{-b}\, |T| \quad \text{or equivalently} \quad h_{T'} \leq 2^{-b/d} h_T. \tag{3.38}$$

Note that due to conformity of meshes additional refinements may be mandatory and therefore we do not have equality in the above display.

We call $\mathbb{T}$ the set of conforming triangulations of Ω that can be produced from $\mathcal{T}_0$ by finite many calls of REFINE. Furthermore, we suppose that the shape-regularity $\sigma(\mathbb{T})$ is bounded. For the existence of such a procedure REFINE we refer to [5, 54, 56, 67, 71, 72].

Let ϕ be an N-function that satisfies Assumption 40, we assume that $g \in L^{\phi^*}(\Omega)$ in (3.21). The precise formulation of AFEM is as follows.

Algorithm 99 (AFEM)**.** Given a conforming initial triangulation $\mathcal{T}_0$ of Ω, $b \in \mathbb{N}$ and a marking parameter $\theta \in (0, 1]$, let $k = 0$

1. $U_k = \mathsf{SOLVE}(\mathcal{T}_k, g)$;

2. $\{\eta(U_k, T, g)\}_{T\in\mathcal{T}_k} = \mathsf{ESTIMATE}(U_k, \mathcal{T}_k, g)$;

3. $\mathcal{M}_k = \mathsf{MARK}(\{\eta(U_k, T, g)\}_{T\in\mathcal{T}_k}, \mathcal{T}_k, \theta)$;

4. $\mathcal{T}_{k+1} = \mathsf{REFINE}(\mathcal{T}_k, \mathcal{M}_k, b)$; increment k and go to step (1).

3.5.2 Auxiliary Results

One of the basic ideas in proving linear convergence of Algorithm 99 (AFEM) in the linear case is the so called error reduction property; see [58, 57, 19] as well as Remark 101. This property can be generalized to the nonlinear case by the energy reduction property (see also [74]): Let $\mathbb{V}_1 \subset \mathbb{V}_2 \subset \mathbb{V}$ be closed subspaces and $u_1 \in \mathbb{V}_1$, $u_2 \in \mathbb{V}_2$, and $u \in \mathbb{V}$ be the unique minimizers of the energy functional $\mathcal{J}$ (3.10) in their respective spaces; compare with Corollary 55. Then, we have

$$\mathcal{J}(u_2) - \mathcal{J}(u) = \mathcal{J}(u_1) - \mathcal{J}(u) - \big(\mathcal{J}(u_1) - \mathcal{J}(u_2)\big). \tag{3.39}$$

Note that since $\mathbb{V}_1 \subset \mathbb{V}_2 \subset \mathbb{V}$, we have

$$\mathcal{J}(u) \leq \mathcal{J}(u_2) \leq \mathcal{J}(u_1).$$

Thus, (3.39) yields an energy reduction and it remains to find a link between the energy differences and the error. This is the content of the following proposition from [28].

Proposition 100 (energy reduction in nested spaces)**.** *Let $u_1 \in \mathbb{V}_1$ and $u_2 \in \mathbb{V}_2$ be the minimizers of the energy functional $\mathcal{J}$ with respect to the closed subspaces $\mathbb{V}_1 \subset \mathbb{V}_2 \subset \mathbb{V}$. Then there exist constants $C_3, c_3 > 0$ such that*

$$c_3 \left\|\mathbf{F}(\nabla u_1) - \mathbf{F}(\nabla u_2)\right\|^2_{L^2(\Omega)} \leq \mathcal{J}(u_1) - \mathcal{J}(u_2) \leq C_3 \left\|\mathbf{F}(\nabla u_1) - \mathbf{F}(\nabla u_2)\right\|^2_{L^2(\Omega)}.$$

The constants c_3, C_3 depend only on $\Delta(\{\phi, \phi^*\})$ and the constants of Assumption 40.

Proof. For the sake of completeness we sketch the proof. We define $\Phi(\mathbf{Q}) := \phi(|\mathbf{Q}|)$ for $\mathbf{Q} \in \mathbb{R}^{d\times d}$, hence $\mathcal{J}(v) = \int_\Omega \Phi(\nabla v) - g \cdot v\, dx$. Let $h(t) := \mathcal{J}([u_1, u_2]_t)$ for $t \in \mathbb{R}$, where $[u_2, u_1]_t := (1 - t)\, u_2 + t\, u_1$. Since u_2 is the minimal function of $\mathcal{J}$ in $\mathbb{V}_2 \supset \mathbb{V}_1$, we have $h'(0) = 0$. We denote as D_{ij} the partial derivative in direction of the ij-th matrix component and as $D_i v^j$ the i-th partial derivative of the j-th component of $v \in \mathbb{V}$. We get by Taylors formula

(3.40)
$$
\begin{aligned}
\mathcal{J}(u_1) - \mathcal{J}(u_2) &= h(1) - h(0) = \frac{1}{2}\int_0^1 h''(t)(1-t)\,dt \\
&= \frac{1}{2}\sum_{i,j,k,l}\int_0^1 \int_\Omega (D_{ij}D_{kl}\Phi)([u_2,u_1]_t)\,(D_i u_1^j - D_i u_2^j)(D_k u_1^l - D_k u_2^l)\,dx\,(1-t)\,dt.
\end{aligned}
$$

Note that the expression above is well defined if we extend $\phi''(t)t$ continuously to zero for $t = 0$; see Assumption 40. Recalling (3.3), then for $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{d\times d}$ with $\mathbf{Q} = (Q_{ij})_{i,j=1,\dots,d}$, it holds

$$
\sum_{i,j,k,l} D_{ij}D_{kl}\Phi(\mathbf{P})\,Q_{ij}Q_{kl} = \frac{\phi'(|\mathbf{P}|)}{|\mathbf{P}|}\Big(\,|\mathbf{Q}|^2 - \frac{|\mathbf{P}:\mathbf{Q}|}{|\mathbf{P}|^2}\Big) + \phi''(|\mathbf{P}|)\frac{|\mathbf{P}:\mathbf{Q}|^2}{|\mathbf{P}|^2}.
$$

By Assumption 40 there are constants $C, c > 0$ such that $c\,\phi'(t) \le t\,\phi''(t) \le C\,\phi'(t)$ for all $t \in [0,\infty)$. Therefore,

$$
\begin{aligned}
\sum_{i,j,k,l} D_{ij}D_{kl}\Phi(\mathbf{P})\,Q_{ij}Q_{kl} &\le \frac{\phi'(|\mathbf{P}|)}{|\mathbf{P}|}\,|\mathbf{Q}|^2 + C\,\frac{\phi'(|\mathbf{P}|)}{|\mathbf{P}|^3}\,|\mathbf{P}|^2\,|\mathbf{Q}|^2 \\
&\le (1+C)\,\frac{\phi'(|\mathbf{P}|)}{|\mathbf{P}|}\,|\mathbf{Q}|^2
\end{aligned}
$$

and on the other hand

$$
\begin{aligned}
\sum_{i,j,k,l} D_{ij}D_{kl}\Phi(\mathbf{P})\,Q_{ij}Q_{kl} &\ge \frac{\phi'(|\mathbf{P}|)}{|\mathbf{P}|}\,|\mathbf{Q}|^2 + (c-1)\frac{\phi'(|\mathbf{P}|)}{|\mathbf{P}|}\frac{|\mathbf{P}:\mathbf{Q}|^2}{|\mathbf{P}|^2} \\
&\ge c\,\frac{\phi'(|\mathbf{P}|)}{|\mathbf{P}|}\,|\mathbf{Q}|^2\,,
\end{aligned}
$$

i.e.,

$$
\sum_{i,j,k,l} D_{ij}D_{kl}\Phi(\mathbf{P})\,Q_{ij}Q_{kl} \approx \frac{\phi'(|\mathbf{P}|)}{|\mathbf{P}|}\,|\mathbf{Q}|^2\,,
$$

uniformly in $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{d\times d}$. Combining the last estimate with (3.40), we obtain

$$
\mathcal{J}(u_1) - \mathcal{J}(u_2) \approx \int_0^1 \int_\Omega \frac{\phi'(|[\nabla u_2, \nabla u_1]_t|)}{|[\nabla u_2, \nabla u_1]_t|}\,|\nabla u_1 - \nabla u_2|^2\,dx\,(1-t)\,dt. \tag{3.41}
$$

Since $(1-t) \leq 1$ we can estimate

$$
\begin{aligned}
\mathcal{J}(u_1) - \mathcal{J}(u_2) &\preccurlyeq \int_0^1 \int_\Omega \frac{\phi'(|[\nabla u_2, \nabla u_1]_t|)}{|[\nabla u_2, \nabla u_1]_t|} \left|\nabla u_1 - \nabla u_2\right|^2 \, dx \, dt \\
&= \int_\Omega \int_0^1 \frac{\phi'(|[\nabla u_2, \nabla u_1]_t|)}{|[\nabla u_2, \nabla u_1]_t|} \, dt \left|\nabla u_1 - \nabla u_2\right|^2 \, dx.
\end{aligned}
$$

Now, an application of Lemma 46, Assumption 40, and Lemma 74 yields

$$
\begin{aligned}
\mathcal{J}(u_1) - \mathcal{J}(u_2) &\preccurlyeq \int_\Omega \frac{\phi'(|\nabla u_2| + |\nabla u_1|)}{|\nabla u_2| + |\nabla u_1|} \left|\nabla u_1 - \nabla u_2\right|^2 \, dx \\
&\approx \int_\Omega \phi''(|\nabla u_2| + |\nabla u_1|) \left|\nabla u_1 - \nabla u_2\right|^2 \, dx \\
&\approx \|\mathbf{F}(\nabla u_1) - \mathbf{F}(\nabla u_2)\|_{L^2(\Omega)}^2 .
\end{aligned}
$$

On the other hand observe that $2(1-t)$ is a density of a probability measure on the Borel σ-algebra over $(0,1)$. Therefore, since ϕ is convex we can estimate (3.41) with Jensen's inequality (Lemma 4)

$$
\begin{aligned}
\mathcal{J}(u_1) - \mathcal{J}(u_2) &\succcurlyeq \int_\Omega \int_0^1 \frac{\phi'(|[\nabla u_2, \nabla u_1]_t|)}{|[\nabla u_2, \nabla u_1]_t|} (1-t) \, dt \, \left|\nabla u_1 - \nabla u_2\right|^2 \, dx \\
&\succcurlyeq \int_\Omega \int_0^1 \frac{\phi(|[\nabla u_2, \nabla u_1]_t|)}{(|\nabla u_2| + |\nabla u_1|)^2} (1-t) \, dt \, \left|\nabla u_1 - \nabla u_2\right|^2 \, dx \\
&\succcurlyeq \int_\Omega \frac{\phi\big(\int_0^1 |[\nabla u_2, \nabla u_1]_t| \, 2(1-t) \, dt\big)}{(|\nabla u_2| + |\nabla u_1|)^2} \left|\nabla u_1 - \nabla u_2\right|^2 \, dx.
\end{aligned}
$$

Both $\int_0^1 |[\mathbf{P}, \mathbf{Q}]_t| \, 2(1-t) \, dt$ and $|\mathbf{P}| + |\mathbf{Q}|$ define a norm on the space $\mathbb{R}^{d\times d} \times \mathbb{R}^{d\times d}$. Thus, they are equivalent, i.e.,

$$
\int_0^1 |[\mathbf{P}, \mathbf{Q}]_t| \, 2(1-t) \, dt \approx |\mathbf{P}| + |\mathbf{Q}| ,
$$

uniformly in $\mathbf{P}, \mathbf{Q}$. This, together with Assumption 40 and Lemma 74 yields

$$
\begin{aligned}
\mathcal{J}(u_1) - \mathcal{J}(u_2) &\succcurlyeq \int_\Omega \frac{\phi(|\nabla u_2| + |\nabla u_1|)}{(|\nabla u_2| + |\nabla u_1|)^2} \left|\nabla u_1 - \nabla u_2\right|^2 \, dx \\
&\succcurlyeq \int_\Omega \phi''(|\nabla u_2| + |\nabla u_1|) \left|\nabla u_1 - \nabla u_2\right|^2 \, dx \\
&\approx \|\mathbf{F}(\nabla u_1) - \mathbf{F}(\nabla u_2)\|_{L^2(\Omega)}^2 .
\end{aligned}
$$

Hence, the lemma is proven. □

Remark 101. *In the linear case, i.e., for $\phi(t) = \frac{1}{2}t^2$ we have with the notation of Proposition 100*

$$\mathcal{J}(u_1) - \mathcal{J}(u_2) = \int_\Omega \frac{1}{2}\left|\nabla u_1\right|^2 - \frac{1}{2}\left|\nabla u_2\right|^2 dx - \int_\Omega g \cdot (u_1 - u_2).$$

Since u_1, u_2 are minimal functions of $\mathcal{J}$ in their respective spaces $\mathbb{V}_1 \subset \mathbb{V}_2$, it holds

$$\langle DJ(u_i), v\rangle = \int_\Omega \nabla u_i : \nabla v - g \cdot v\, dx = 0 \qquad \text{for all } v \in \mathbb{V}_i,$$

$i = 1, 2$. Therefore, $\mathbb{V}_1 \subset \mathbb{V}_2$ implies

$$\int_\Omega g \cdot (u_1 - u_2)\, dx = \int_\Omega \left|\nabla u_1\right|^2 - \left|\nabla u_2\right|^2 dx$$

and

$$\int_\Omega g \cdot u_1\, dx = \int_\Omega \left|\nabla u_1\right|^2 dx = \int_\Omega \nabla u_2 : \nabla u_1\, dx.$$

Altogether this yields

$$\begin{aligned}\mathcal{J}(u_1) - \mathcal{J}(u_2) &= \int_\Omega \frac{1}{2}\left|\nabla u_2\right|^2 - \frac{1}{2}\left|\nabla u_1\right|^2 dx \\ &= \int_\Omega \frac{1}{2}\left|\nabla u_2\right|^2 - \nabla u_2 : \nabla u_1 + \frac{1}{2}\left|\nabla u_1\right|^2 dx \\ &= \int_\Omega \frac{1}{2}\left|\nabla u_1 - \nabla u_2\right|^2 dx.\end{aligned}$$

Thus, in the linear case the energy reduction property (3.39) *is equivalent to the error reduction property*

$$\begin{aligned}\frac{1}{2}\left\|u_2 - u\right\|^2_{L^2(\Omega)} &= \mathcal{J}(u_2) - \mathcal{J}(u) = \mathcal{J}(u_1) - \mathcal{J}(u) - \big(\mathcal{J}(u_1) - \mathcal{J}(u_2)\big) \\ &= \frac{1}{2}\left\|u_1 - u\right\|^2_{L^2(\Omega)} - \frac{1}{2}\left\|u_1 - u_2\right\|^2_{L^2(\Omega)};\end{aligned}$$

see [74].

Convergence of Algorithm 99 AFEM) is naturally based on properties of the estimator, since it contains the only available information on the error. The following technical results reveal the behavior of the estimator on perturbations.

Lemma 102. *Let $\mathcal{T}$ be a conforming triangulation of Ω, $v, w \in \mathbb{V}$, $V \in \mathring{\mathbb{V}}(\mathcal{T})$, then there exists $\Lambda_1 > 0$ solely dependent on $\Delta_2(\{\phi, \phi^*\})$, such that for all $T \in \mathcal{T}$, $\delta > 0$*

$$\eta^2(v, V, T, g) \le (1 + C_\delta)\, \Lambda_1\, \eta^2(w, V, T, g) + \delta\, \Lambda_1\, \left\|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla w)\right\|^2_{L^2(T)}.$$

The constant C_δ stems from Young's inequality (Proposition 11) and depends only on δ and $\Delta_2(\{\phi, \phi^*\})$.

Proof. Applying Corollary 71 to the element-estimator yields for $\delta > 0$

$$\int_\Omega \left(\phi_{|\nabla v|}\right)^*(h_T\,|g|)\,dx \preccurlyeq (1+C_\delta)\int_\Omega \left(\phi_{|\nabla w|}\right)^*(h_T\,|g|)\,dx \\ + \delta\,\|\mathbf{F}(\nabla v)-\mathbf{F}(\nabla w)\|^2_{L^2(T)}\,.$$

Therefore, we obtain

$$\begin{aligned}\eta^2(v,V,T,g) &= \int_\Omega \left(\phi_{|\nabla v|}\right)^*(h_T\,|g|)\,dx + \int_{\partial T} h_T\,\left|[\![\mathbf{F}(\nabla V)]\!]\right|^2\,d\sigma \\ &\preccurlyeq (1+C_\delta)\int_\Omega \left(\phi_{|\nabla w|}\right)^*(h_T\,|g|)\,dx + \delta\,\|\mathbf{F}(\nabla v)-\mathbf{F}(\nabla w)\|^2_{L^2(T)} \\ &\quad + \int_{\partial T} h_T\,\left|[\![\mathbf{F}(\nabla V)]\!]\right|^2\,d\sigma \\ &\le (1+C_\delta)\,\eta^2(w,V,T,g) + \delta\,\|\mathbf{F}(\nabla v)-\mathbf{F}(\nabla w)\|^2_{L^2(T)}\,.\end{aligned}$$

Choosing δ small enough yields the assertion. □

The following corollary is a direct consequence of Lemma 102 and the upper bound Theorem 90.

Corollary 103. *Let $\mathcal{T}$ be a conforming triangulation of Ω, let $u \in \mathbb{V}$ be the solution of (3.21) and $U \in \mathring{\mathbb{V}}(\mathcal{T})$ its Ritz-Galerkin approximation. Then there exist constants $c_4, C_4 > 0$ such that*

$$c_4\,\eta(u,U,\mathcal{T},g) \le \eta(U,\mathcal{T},g) \le C_4\,\eta(u,U,\mathcal{T},g),$$

where the constants c_4, C_4 depend solely on $\Delta_2(\{\phi,\phi^*\})$ and the shape-regularity of $\mathcal{T}$.

Proof. Summing over all $T \in \mathcal{T}$, Lemma 102 yields for $v, w \in \mathbb{V}$

$$\eta^2(v,U,\mathcal{T},g) \le (1+C_\delta)\,\Lambda_1\,\eta^2(w,U,\mathcal{T},g) + \delta\,\Lambda_1\,\|\mathbf{F}(\nabla v)-\mathbf{F}(\nabla w)\|^2_{L^2(\Omega)}\,.$$

If now $v = u$ and $w = U$ or $v = U$ and $w = u$, the last term can be estimated by the upper bound Theorem 90. This yields the assertion. □

Lemma 104. *Let $\mathcal{T}$ be a conforming triangulation, $v \in \mathbb{V}$ and $V, W \in \mathring{\mathbb{V}}(\mathcal{T})$. Then there exists a constant Λ_2 solely depending on the shape regularity of $\mathcal{T}$ and d such that*

$$\eta^2(v,V,T,g) \le (1+\delta)\,\eta^2(v,W,T,g) + (1+\delta^{-1})\,\Lambda_2\,\|\mathbf{F}(\nabla V)-\mathbf{F}(\nabla W)\|^2_{L^2(\omega_T)}\,,$$

for all $T \in \mathcal{T}$.

Proof. Since the element-residual does not depend on the discrete solution in the second argument of the estimator, it suffices to prove the assertion for the jump estimator. It holds for $\sigma \in \mathcal{S} \cap \partial T$ with the triangle inequality and Young's inequality $st \leq \frac{\delta}{2}s^2 + \frac{1}{2\delta}t^2$ for $\delta > 0$

$$\begin{aligned} h_T \, \|[\![\mathbf{F}(\nabla V)]\!]\|^2_{L^2(\sigma)} &= h_T \, \|[\![\mathbf{F}(\nabla V) - \mathbf{F}(\nabla W) + \mathbf{F}(\nabla W)]\!]\|^2_{L^2(\sigma)} \\ &\leq (1+\delta)\, h_T \, \|[\![\mathbf{F}(\nabla W)]\!]\|^2_{L^2(\sigma)} \\ &\quad + (1+\delta^{-1})\, h_T \, \|[\![\mathbf{F}(\nabla V) - \mathbf{F}(\nabla W)]\!]\|^2_{L^2(\sigma)} \, . \end{aligned}$$

Let now $T' \in \mathcal{T}$ such that $\sigma = T \cap T'$ and recall that $\nabla V, \nabla W$ are piecewise constant. Shape regularity yields $|T| \approx h_T \, |\sigma| \approx h_{T'} \, |\sigma| \approx |T'|$ and thus the second term can be estimated by

$$\begin{aligned} h_T \, \|[\![\mathbf{F}(\nabla V) - \mathbf{F}(\nabla W)]\!]\|^2_{L^2(\sigma)} &\leq 2\, h_T \, \|\mathbf{F}(\nabla V|_T) - \mathbf{F}(\nabla W|_T)\|^2_{L^2(\sigma)} \\ &\quad + 2\, h_T \, \|\mathbf{F}(\nabla V|_{T'}) - \mathbf{F}(\nabla W|_{T'})\|^2_{L^2(\sigma)} \\ &= 2\, h_T \, |\sigma| \, |\mathbf{F}(\nabla V|_T) - \mathbf{F}(\nabla W|_T)|^2 \\ &\quad + 2\, h_T \, |\sigma| \, |\mathbf{F}(\nabla V|_{T'}) - \mathbf{F}(\nabla W|_{T'})|^2_{L^2(\sigma)} \\ &\approx \int_T |\mathbf{F}(\nabla V|_T) - \mathbf{F}(\nabla W|_T)|^2 \, dx \\ &\quad + \int_{T'} |\mathbf{F}(\nabla V|_{T'}) - \mathbf{F}(\nabla W|_{T'})|^2_{L^2(\sigma)} \, dx \\ &= \|\mathbf{F}(\nabla V) - \mathbf{F}(\nabla W)\|^2_{L^2(\omega_\sigma)} \, . \end{aligned}$$

Since it holds $\omega_T = \text{interior} \bigcup \{\omega_\sigma \mid \sigma \in \mathcal{S} : \sigma \subset \partial T\}$ and T has at most $d+1$ sides, the assertion follows. □

A key observation of the subsequent convergence analysis is the following perturbed estimator reduction that stems from the mesh-size reduction of the refined elements in Algorithm 99 (AFEM).

Lemma 105 (perturbed estimator reduction)**.** *Let $u \in \mathbb{V}$ be the unique solution of* (3.21) *and let $(\mathcal{T}_k, \mathbb{V}(\mathcal{T}_k), U_k)_{k \in \mathbb{N}_0}$ be the sequence of meshes, finite element spaces, and discrete solutions produced by* AFEM*. Then, with $\lambda := 1 - 2^{-\frac{b}{d}} \in (0,1)$,*

$$\begin{aligned} \eta^2(u, U_{k+1}, \mathcal{T}_{k+1}, g) &\leq (1+\delta)\{\eta^2(u, U_k, \mathcal{T}_k, g) - \lambda\, \eta^2(u, U_k, \mathcal{M}_k, g)\} \\ &\quad + (1+\delta^{-1})\, \Lambda_3 \, \|\mathbf{F}(\nabla U_k) - \mathbf{F}(\nabla U_{k+1})\|^2_{L^2(\Omega)} \, , \end{aligned}$$

where the constant $\Lambda_3 > 0$ depends solely on the shape regularity of $\sigma(\{\mathcal{T}_k\}_{k \in \mathbb{N}})$ and d.

Proof. We observe from Lemma 104 and $U_k \in \mathbb{V}(\mathcal{T}_k) \subset \mathbb{V}(\mathcal{T}_{k+1})$ that

$$
\begin{aligned}
\eta^2(u, U_{k+1}, \mathcal{T}_{k+1}, g) &\le (1+\delta)\,\eta^2(u, U_k, \mathcal{T}_{k+1}, g) \\
&\quad + (1+\delta^{-1})\,\Lambda_2 \sum_{T\in\mathcal{T}_{k+1}} \|\mathbf{F}(\nabla U_{k+1}) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\omega_T)} \\
&\le (1+\delta)\,\eta^2(u, U_k, \mathcal{T}_{k+1}, g) \\
&\quad + (1+\delta^{-1})\,\Lambda_2\,(d+2)\,\|\mathbf{F}(\nabla U_{k+1}) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\Omega)}\,,
\end{aligned}
\tag{3.42}
$$

where we used that ω_T consists of at most $d+2$ elements. The error estimator can be splitted according to marked and non-marked elements, i.e.,

$$
\begin{aligned}
\eta^2(u, U_k, \mathcal{T}_{k+1}, g) &= \sum_{T'\in\mathcal{T}_{k+1}} \eta^2(u, U_k, T', g) \\
&= \sum_{T\in\mathcal{T}_k} \sum_{T'\in\mathcal{T}_{k+1}(T)} \eta^2(u, U_k, T', g) \\
&= \sum_{T\in\mathcal{T}_k\setminus\mathcal{M}_k} \sum_{T'\in\mathcal{T}_{k+1}(T)} \eta^2(u, U_k, T', g) \\
&\quad + \sum_{T\in\mathcal{M}_k} \sum_{T'\in\mathcal{T}_{k+1}(T)} \eta^2(u, U_k, T', g).
\end{aligned}
$$

Let $T \in \mathcal{M}_k$, recalling (3.38), we have for all all $T' \in \mathcal{T}_{k+1}(T)$, $T \in \mathcal{M}_k$, the mesh-size reduction $h_{T'} = |T'|^{1/d} \le (2^{-b}\,|T|)^{1/d} = 2^{-b/d} h_T$. Note further, that $U_k \in \mathbb{V}(\mathcal{T}_k) \subset \mathbb{V}(\mathcal{T}_{k+1})$. Therefore, ∇U_k jumps only across inter element sides of $\mathcal{T}_k$, i.e., $|[\![\nabla U_k]\!]| = 0$ and therefore $|[\![\mathbf{F}(\nabla U_k)]\!]| = 0$ on interior sides of $\mathcal{T}_{k+1}(T)$. With (2.6a) we have

$$
\begin{aligned}
\eta^2(u, U_k, \mathcal{T}_{k+1}(T), g) &= \sum_{T'\in\mathcal{T}_{k+1}(T)} \Big\{ \int_{T'} \big(\phi_{|\nabla u|}\big)^*(h_{T'}\,|g|)\,dx \\
&\qquad\qquad + h_{T'}\,\|[\![\mathbf{F}(\nabla U_k)]\!]\|^2_{L^2(\partial T)} \Big\} \\
&= \sum_{T'\in\mathcal{T}_{k+1}(T)} \Big\{ \int_{T'} \big(\phi_{|\nabla u|}\big)^*(2^{-b/d} h_T\,|g|)\,dx \\
&\qquad\qquad + 2^{-b/d} h_T\,\|[\![\mathbf{F}(\nabla U_k)]\!]\|^2_{L^2(\partial T)} \Big\} \\
&\le 2^{-b/d} \sum_{T'\in\mathcal{T}_{k+1}(T)} \Big\{ \int_{T'} \big(\phi_{|\nabla u|}\big)^*(h_T\,|g|)\,dx \\
&\qquad\qquad + h_T\,\|[\![\mathbf{F}(\nabla U_k)]\!]\|^2_{L^2(\partial T)} \Big\} \\
&= 2^{-b/d}\,\eta^2(u, U_k, T, g).
\end{aligned}
$$

For all other elements $T \in \mathcal{T}_k \setminus \mathcal{M}_k$ it follows from the monotonicity of the mesh-size and similar arguments

$$\eta^2(u, U_k, \mathcal{T}_{k+1}(T), g) \le \eta^2(u, U_k, T, g).$$

Hence, summing over all $T \in \mathcal{T}_k$ implies

$$\begin{aligned}\eta^2(u, U_k, \mathcal{T}_{k+1}, g) &\le \sum_{T \in \mathcal{T}_k \setminus \mathcal{M}_k} \eta^2(u, U_k, T, g) + 2^{-b/d} \sum_{T \in \mathcal{M}_k} \eta^2(u, U_k, T, g)\\ &= \eta^2(u, U_k, \mathcal{T}_k \setminus \mathcal{M}_k, g) + 2^{-b/d}\, \eta^2(u, U_k, \mathcal{M}_k, g)\\ &= \eta^2(u, U_k, \mathcal{T}_k, g) - \lambda\, \eta^2(u, U_k, \mathcal{M}_k, g).\end{aligned}$$

Inserting this in (3.42) yields the assertion. □

3.5.3 Contraction of AFEM

In this section we prove linear convergence of AFEM. The result is taken from [27] and improves the result in [28]. In particular, it combines the results of [28] with ideas of the linear case [19]; see also Remark 111.

Theorem 106 (Contraction of AFEM)**.** *Let $u \in \mathbb{V}$ be the solution of* (3.21) *and let $(\mathcal{T}_k, \mathbb{V}_k, U_k)_{k \in \mathbb{N}}$ be the sequence of meshes, finite element spaces, and discrete solutions produced by Algorithm 99* (AFEM)*. Then, there exists $\gamma > 0$, $\alpha \in (0, 1)$, depending solely on the shape-regularity of $\mathcal{T}_0$, b, $\Delta_2(\{\phi, \phi^*\})$, and the marking parameter $0 < \theta \le 1$, such that*

$$\begin{aligned}&\mathcal{J}(U_{k+1}) - \mathcal{J}(u) + \gamma\, \eta^2(u, U_{k+1}, \mathcal{T}_{k+1}, g)\\ &\qquad\qquad \le \alpha \big\{\mathcal{J}(U_k) - \mathcal{J}(u) + \gamma\, \eta^2(u, U_k, \mathcal{T}_k, g)\big\}.\end{aligned}$$

Proof. For the sake of convenience, we use the notation

$$\begin{aligned}\epsilon_k^2 &:= \mathcal{J}(U_k) - \mathcal{J}(u), \quad e_k^2 := \mathcal{J}(U_{k+1}) - \mathcal{J}(u),\\ \eta_k &:= \eta(u, U_k, \mathcal{T}_k, g), \quad \eta_k(\mathcal{M}_k) := \eta(u, U_k, \mathcal{M}_k, g).\end{aligned}$$

We combine the energy reduction (3.39) with the estimator reduction Corollary 105 and thus get for $\gamma > 0$

$$\epsilon_{k+1}^2 + \gamma\, \eta_{k+1} \le \epsilon_k^2 - e_k^2 + (1+\delta)\, \gamma\, (\eta_k^2 - \lambda\, \eta_k^2(\mathcal{M}_k)) + (1+\delta^{-1})\, \gamma\, \Lambda_2\, e_k^2.$$

Choose $\gamma := \frac{1}{(1+\delta^{-1})\,\Lambda_2}$ to obtain

$$\epsilon_{k+1}^2 + \gamma\, \eta_{k+1} \le \epsilon_k^2 + (1+\delta)\, \gamma\, (\eta_k^2 - \lambda\, \eta_k^2(\mathcal{M}_k)).$$

We take a closer look to the term $\eta_k^2(\mathcal{M}_k) = \eta^2(u, U_k, \mathcal{M}_k)$. In particular, we want to apply Dörfler's marking property and thus we have to substitute its first argument with the help of Proposition 102 to get for all $\rho > 0$

$$\begin{aligned}\eta_k^2(\mathcal{M}_k) &\geq \frac{1}{(1+C_\rho)\Lambda_1}\,\eta^2(U_k, U_k, \mathcal{M}_k, g) - \frac{\rho}{1+C_\rho}\,\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_k)\|^2 \\ &\geq \frac{\theta^2}{(1+C_\rho)\,\Lambda_1}\,\eta^2(U_k, \mathcal{T}_k, g) - \frac{\rho}{1+C_\rho}\,\epsilon_k^2.\end{aligned}$$

Therefore,

$$\begin{aligned}\epsilon_{k+1}^2 + \gamma\,\eta_{k+1} \leq & \left\{1 + (1+\delta)\,\gamma\,\lambda\,\frac{\rho}{1+C_\rho}\right\}\epsilon_k^2 \\ & + (1+\delta)\,\gamma\left(\eta_k^2 - \lambda\,\frac{\theta^2}{(1+C_\rho)\,\Lambda_1}\,\eta^2(U_k, \mathcal{T}_k, g)\right).\end{aligned}$$

We split the estimator $\eta_k^2 = \frac{1}{2}\,\eta_k^2 + \frac{1}{2}\,\eta_k^2$ into two parts and apply the upper bound Theorem 90 and Proposition 100 to the first part to get

$$\begin{aligned}\leq & \left\{1 + \frac{(1+\delta)\,\gamma\,\lambda}{1+C_\rho}\left(\rho - \frac{c_3}{2\,C_1}\frac{\theta^2}{\Lambda_1}\right)\right\}\epsilon_k^2 \\ & + (1+\delta)\,\gamma\left(\eta_k^2 - \lambda\,\frac{\theta^2}{2\,(1+C_\rho)\,\Lambda_1}\,\eta^2(U_k, \mathcal{T}_k, g)\right).\end{aligned}$$

Finally, Corollary 103 yields

$$\begin{aligned}\leq & \left\{1 + \frac{(1+\delta)\,\gamma\,\lambda}{1+C_\rho}\left(\rho - \frac{c_3}{2\,C_1}\frac{\theta^2}{\Lambda_1}\right)\right\}\epsilon_k^2 \\ & + (1+\delta)\left\{1 - \lambda\,\frac{c_4^2}{2\,(1+C_\rho)}\frac{\theta^2}{\Lambda_1}\right\}\gamma\,\eta_k^2.\end{aligned}$$

We set

$$\alpha := \max\left\{1 + \frac{(1+\delta)\,\gamma\,\lambda}{1+C_\rho}\left(\rho - \frac{c_3}{2\,C_1}\frac{\theta^2}{\Lambda_1}\right),\ (1+\delta)\left(1 - \lambda\,\frac{c_4^2}{2\,(1+C_\rho)}\frac{\theta^2}{\Lambda_1}\right)\right\}.$$

Now, choose $\rho \in (0, \frac{c_3}{2\,C_1}\frac{\theta^2}{\Lambda_1})$. Hence, the first term is less than 1 for all $\delta > 0$. For δ small enough, the second term becomes less than 1, too. This yields the desired estimate. □

The next result follows from Theorem 106 with induction over $k \in \mathbb{N}$.

Corollary 107. *Assume the conditions of Theorem 106, then for all $k \in \mathbb{N}$*

$$\mathcal{J}(U_k) - \mathcal{J}(u) + \gamma\,\eta^2(u, U_k, \mathcal{T}_k, g) \leq \alpha^k\left\{\mathcal{J}(U_0) - \mathcal{J}(u) + \gamma\,\eta^2(u, U_0, \mathcal{T}_k, g)\right\}.$$

Corollary 108. *Under the conditions of Theorem 106, there exists $C > 0$ depending on the shape-regularity of $\mathcal{T}_0$ and $\Delta_2(\{\phi, \phi^*\})$, such that for all $k \in \mathbb{N}$*

$$\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\Omega)} \le C\, \alpha^k$$

and

$$\eta^2(U_k, \mathcal{T}_k, g) \le C\, \alpha^k.$$

Proof. The first assertion is an immediate consequence of Corollary 107 and Proposition 100. The second assertion follows from Theorem 106 with the help of Corollary 103. □

It is shown in [27] that Algorithm 99 leads to quasi-optimal meshes. The proof of this result relies amongst others on the linear convergence rate of Algorithm 99 (AFEM) and is a generalization of the results in [71, 19] to the nonlinear case. To state the result, we need to introduce a suitable error quantity being controlled by AFEM and its associated approximation class $\mathbb{A}_s$. On the one hand, oscillation is dominated by the estimator according to Remark 94, thereby yielding with Corollary 103

$$\begin{aligned}\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\Omega)} &+ \operatorname{osc}^2(u, \mathcal{T}_k, g)\\ &\le \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\Omega)} + \eta^2(U_k, \mathcal{T}_k, g).\end{aligned}$$

On the other hand, the global lower bound (Corollary 96) implies

$$\begin{aligned}\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\Omega)} &+ \eta^2(U_k, \mathcal{T}_k, g)\\ &\le (1 + \frac{1}{\tilde{C}_2^{-2}})\big\{ \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\Omega)} + \operatorname{osc}^2(u, \mathcal{T}_k, g)\big\}.\end{aligned}$$

We thus realize that

$$\begin{aligned}\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\Omega)} &+ \operatorname{osc}^2(u, \mathcal{T}_k, g)\\ &\approx \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\Omega)} + \eta^2(U_k, \mathcal{T}_k, g),\end{aligned} \tag{3.43}$$

and call the square root of the right-hand side the *total error*. This is equivalent to the quantity being reduced by AFEM and motivates the following definition of the approximation class $\mathbb{A}_s$. The quality of the best approximation to the total error with at most N elements more than $\mathcal{T}_0$ is given by

$$\Sigma(N; u, g) := \inf_{\{\mathcal{T} \in \mathbb{T} : \#\mathcal{T} - \#\mathcal{T}_0 \le N\}} \inf_{V \in \mathbb{V}(\mathcal{T})} \Big(\|\mathbf{F}(\nabla V) - \mathbf{F}(\nabla u)\|^2_{L^2(\Omega)} + \operatorname{osc}^2(u, \mathcal{T}, g)\Big)^{1/2}.$$

Now, for $s > 0$ we define the nonlinear approximation class $\mathbb{A}_s$ to be

$$\mathbb{A}_s := \Big\{(u, g) \,:\, \sup_{N > 0} \big(N^s\, \Sigma(N; u, g)\big) < \infty\Big\}.$$

Now, we are prepared to state the result on quasi-optimal convergence rate of AFEM from [27].

Theorem 109. *Let $u \in \mathbb{V}$ be the solution of* (3.21)*, let the initial triangulation $\mathcal{T}_0$ of Ω satisfy condition* (*b*) *of §4 in [72], and let the routine* **REFINE** *be based on the conforming local refinement routine in [72]. Assume $(u, g) \in \mathbb{A}_s$ for some $s > 0$, then there exists $\theta_* \in (0, 1)$, such that the sequence $(\mathcal{T}_k, \mathbb{V}_k, U_k)_{k\in\mathbb{N}}$ of meshes, finite element spaces, and discrete solutions, produced by Algorithm 99* (AFEM) *with marking parameter $\theta \in (0, \theta_*)$, satisfies*

$$\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_k)\|^2_{L^2(\Omega)} + \operatorname{osc}^2(u, \mathcal{T}_k, g) \le C\big(\#\mathcal{T}_k - \#\mathcal{T}_0\big)^{-2s}$$

for all $k \in \mathbb{N}$. The constant $\theta^* \in (0, 1)$ depends only on $\Delta_2(\{\phi, \phi^*\})$, the constants in Assumption 40 and the shape regularity of $\mathcal{T}_0$. The constant $C > 0$ depends only on $\Delta_2(\{\phi, \phi^*\})$, the constants in Assumptions 40, the shape-regularity of $\mathcal{T}_0$, the refinement depth b, and the marking parameter θ.

Remark 110. *Note that due to the global lower bound (Corollary 96)*

$$\tilde{C}_2\, \eta^2(U_k, \mathcal{T}_k, g) \le \|\mathbf{F}(\nabla U_k) - \mathbf{F}(\nabla u)\|^2_{L^2(\Omega)} + \operatorname{osc}^2(u, \mathcal{T}_k, g).$$

On the other hand, we have by the fact that $\operatorname{osc}^2(u, \mathcal{T}_k) \le \eta^2(u, U_k, \mathcal{T}_k, g) \approx \eta^2(U_k, \mathcal{T}_k, g)$ (Remark 94 and Corollary 103) and the upper bound (Theorem 90) that

$$\begin{aligned}\|\mathbf{F}(\nabla U_k) - \mathbf{F}(\nabla u)\|^2_{L^2(\Omega)} + \operatorname{osc}^2(u, \mathcal{T}_k, g) &\le C_1\, \eta^2(U_k, \mathcal{T}_k, g) + \operatorname{osc}^2(u, \mathcal{T}_k, g)\\ &\preccurlyeq \eta^2(U_k, \mathcal{T}_k, g).\end{aligned}$$

Hence, it follows

$$\eta^2(U_k, \mathcal{T}_k, g) \approx \|\mathbf{F}(\nabla U_k) - \mathbf{F}(\nabla u)\|^2_{L^2(\Omega)} + \operatorname{osc}^2(u, \mathcal{T}_k, g). \tag{3.44}$$

Therefore, the total error and the estimator are equivalent and thus the approximation class could be equivalently defined substituting the total error by the estimator. This reflects the fact that AFEM *takes all its decisions depending on the indicators $\eta(U_k, T, g), T \in \mathcal{T}_k$, and therefore optimal meshes can only be expected with respect to this quantity.*

Remark 111. *Based on the crucial Dörfler marking [36], Morin, Nochetto, and Siebert established in [57, 58, 59] the first convergence result for an adaptive finite element method. Later these results have been extended to more general elliptic operators by Chen and Feng [20] and Mekchay and Nochetto [55]. What all these results have in common is that they incorporate a separate marking according to oscillation. In particular, in step* **MARK** *of* (AFEM) *the set $\mathcal{M}_k$ is further enlarged to satisfy additionally*

$$\operatorname{osc}^2(U_k, \mathcal{M}_k, g) \ge \theta^2 \operatorname{osc}^2(U_k, \mathcal{T}_k, g).$$

The first result on convergence of an adaptive finite element algorithm for the nonlinear Poisson problem was proved by Veeser in [74] using hierarchical estimators. The result is based on the error notion in the energy norm and thus the a posteriori error estimators are not optimal; see also Remark 97. This prevents proving linear convergence.

Diening and Kreuzer proved linear convergence in the quasi-norm of an adaptive finite element method for the nonlinear Poisson equation in [28]. There, the marking according to oscillation is completely avoided for the first time.

Binev, Dahmen, and DeVore showed in [12] a quasi-optimal convergence rate for an adaptive method using coarsening. Stevenson improved this result in [71] showing that an algorithm based on the method in [57] leads to quasi-optimal meshes.

Up to this point, all mentioned results rely on a so-called discrete lower bound, which estimates the distance of discrete solutions in nested spaces. For this reason it was crucial to have a discrete substitute of the bubble functions in Section 3.4.2. Thus, an interior node condition was mandatory on the marked elements. This condition can, e.g., be ensured by taking $b = 3$ *in* 2d *or* $b = 6$ *in* 3d *as refinement depth of* REFINE. *This condition could be completely avoided in recent works of Recently Morin, Siebert, and Veeser [61, 60]. They proved convergence of* (AFEM) *for general marking strategies, including maximum and equidistribution strategy besides Dörfler strategy. The main result is a plain convergence result. They do not provide a strict error reduction between two successive iterations, which is currently crucial for proving complexity results like in Theorem 109. Siebert extended these results to estimators without lower bound [70].*

Recently the interior node condition could be avoided in [19] for the linear case and in [27] for the nonlinear case, nevertheless providing linear convergence results for Dörfler marking. These works additionally established quasi-optimal convergence rates for the considered adaptive finite element methods.

Remark 112 (symmetric gradient)**.** *In the modeling of quasi-Newtonian fluids, the symmetric gradient appears rather then the gradient; see Section 1.1. In particular, models often lead to equations of the form*

$$\int_\Omega \mathbf{A}(\mathbf{E}(u)) : \mathbf{E}(v)\,dx = \langle g, v\rangle \qquad \text{for all } v \in W_0^{1,\phi}(\Omega)^d, \tag{3.45}$$

where the symmetric gradient is defined as $\mathbf{E}(v) := \frac{1}{2}(\nabla v + \nabla v^t)$. *Note, that for this equation the corresponding energy becomes*

$$\mathcal{J}_{\mathbf{E}}(v) := \int_\Omega \phi(|\mathbf{E}(v)|)\,dx - \langle g, v\rangle.$$

In order to handle this kind of equations, we need a so called Korn inequality, i.e.,

$$\int_\Omega \phi(|\nabla v|)\,dx \preccurlyeq \int_\Omega \phi(|\mathbf{E}(v)|)\,dx \qquad \text{for all } v \in W_0^{1,\phi}(\Omega)^d. \tag{3.46}$$

In the case $W_0^{1,\phi}(\Omega)^d = W_0^{1,r}(\Omega)^d$ *for some* $r \in (0,1)$ *a Korn inequality is proved, e.g., in [62, 29, 30]. For more general N-functions, a Korn inequality can be found in [33].*

Since the pointwise estimate $|\mathbf{E}(u)| \leq |\nabla u|$ *immediately implies the inverse inequality of* (3.46)*, we can deduce by Corollary 36 that* $\|\mathbf{E}(\cdot)\|_\phi$ *is equivalent to* $\|\cdot\|_{W_0^{1,\phi}(\Omega)}$*. This is the key observation for proving existence like in Section 3.1.2.*

Most estimates are based on the pointwise estimates of Sections 3.1.2 and 3.2.1. Hence, in these estimates we can easily insert $\mathbf{E}(v)$ *instead of* ∇v *in order to get the corresponding estimates to the ones in Section 3.2.2. With the same techniques as in Section 3.4 we get upper and lower bounds for the error. In particular, let* $\mathcal{T}$ *be a conforming triangulation of* Ω *and* $U \in \mathring{\mathbb{V}}(\mathcal{T})$ *be the finite element solution of* (3.45) *with* $g \in L^{\phi^*}(\Omega)^d$*, i.e.,*

$$\int_\Omega \mathbf{A}(\mathbf{E}(U)) : \mathbf{E}(V)\,dx = \langle g, V\rangle \qquad \text{for all } V \in \mathring{\mathbb{V}}(\mathcal{T}). \tag{3.47}$$

Then,

$$\|\mathbf{F}(\mathbf{E}(u)) - \mathbf{F}(\mathbf{E}(U))\|_{L^2(\Omega)} \preccurlyeq \eta_{\mathbf{E}}(U, \mathcal{T}, g)$$

and

$$\eta_{\mathbf{E}}(U, \mathcal{T}, g) \preccurlyeq \|\mathbf{F}(\mathbf{E}(u)) - \mathbf{F}(\mathbf{E}(U))\|_{L^2(\Omega)} + \operatorname{osc}_{\mathbf{E}}(U, \mathcal{T}, g),$$

where for $v \in \mathbb{V}$, $V \in \mathbb{V}(\mathcal{T})$

$$\eta^2_{\mathbf{E}}(v, V, \mathcal{T}, g) := \sum_{T\in\mathcal{T}} \Big\{ \int_T (\phi_{|\mathbf{E}(v)|})^*(h_T\,|g|)\,dx + \int_{\partial T} h_T \,|[\![\mathbf{F}(\mathbf{E}(V))]\!]|^2\,d\sigma \Big\},$$
$$\eta^2_{\mathbf{E}}(V, \mathcal{T}, g) := \eta^2_{\mathbf{E}}(V, V, \mathcal{T}, g),$$

and

$$\operatorname{osc}^2_{\mathbf{E}}(v, \mathcal{T}, g) := \inf_{g_T\in\mathbb{R}} \int_T (\phi_{|\mathbf{E}(v)|})^*(h_T\,|g - g_T|)\,dx.$$

In order to get a convergent adaptive finite element method ($\mathrm{AFEM}_{\mathbf{E}}$) *for* (3.45) *we have to modify Algorithm 99* (AFEM)*. In particular, the procedure* ***SOLVE*** *has to be substituted by a procedure* $U = \textbf{\textsf{SOLVE}}_{\mathbf{E}}(\mathcal{T}, g)$*, that, given a conforming triangulation* $\mathcal{T}$ *of* Ω *and a right-hand side* $g \in L^{\phi^*}(\Omega)$*, outputs the finite element solution* $U \in \mathring{\mathbb{V}}(\mathcal{T})$ *of* (3.47)*. Moreover, the routine* ***ESTIMATE*** *has to be modified into a routine* $\textbf{\textsf{ESTIMATE}}_{\mathbf{E}}$ *that outputs the estimators* $\{\eta_{\mathbf{E}}(U, T, g)\}_{T\in\mathcal{T}}$ *instead of* $\{\eta(U, T, g)\}_{T\in\mathcal{T}}$*. Now, we are able to define* ($\mathrm{AFEM}_{\mathbf{E}}$)*:*

Algorithm 113 ($\mathrm{AFEM}_{\mathbf{E}}$)**.** Given a conforming initial triangulation $\mathcal{T}_0$ of Ω, $b \in \mathbb{N}$ and a marking parameter $\theta \in (0, 1]$, let $k = 0$,

1. $U_k = \mathsf{SOLVE}_{\mathbf{E}}(\mathcal{T}_k, g)$;
2. $\{\eta_{\mathbf{E}}(U_k, T, g)\}_{T\in\mathcal{T}_k} = \mathsf{ESTIMATE}_{\mathbf{E}}(U_k, \mathcal{T}_k, g)$;
3. $\mathcal{M}_k = \mathsf{MARK}(\{\eta_{\mathbf{E}}(U_k, T, g)\}_{T\in\mathcal{T}_k}, \mathcal{T}_k, \theta)$;
4. $\mathcal{T}_{k+1} = \mathsf{REFINE}(\mathcal{T}_k, \mathcal{M}_k, b)$; increment k and go to step (1).

Using the same techniques as in the proof of Theorem 106, the $\mathrm{AFEM}_{\mathbf{E}}$ *yields a reduction of the energies, i.e., there exists* $\alpha \in (0,1)$*,* $\gamma > 0$*, such that*

$$\begin{aligned}\mathcal{J}_{\mathbf{E}}(U_{k+1}) - \mathcal{J}_{\mathbf{E}}(u) + \gamma\, \eta^2_{\mathbf{E}}(u, U_{k+1}, \mathcal{T}_{k+1}, g)&\\ \leq \alpha \left\{\mathcal{J}_{\mathbf{E}}(U_k) - \mathcal{J}_{\mathbf{E}}(u) + \gamma\, \eta^2_{\mathbf{E}}(u, U_k, \mathcal{T}_k, g)\right\};&\end{aligned}$$

Then analogously to the proof of Proposition 100 we get that the energy reduction is equivalent to error reduction and hence for all $k \in \mathbb{N}$

$$\|\mathbf{F}(\mathbf{E}(u)) - \mathbf{F}(\mathbf{E}(U_k))\|_{L^2(\Omega)} \leq \alpha^k C;$$

see Corollary 108. It remains the question if this result implies $U_k \to u$ *as* $k \to \infty$*. For this reason we need Korn's inequality. In fact, Lemma 76 together with* $\|\mathbf{F}(\mathbf{E}(u)) - \mathbf{F}(\mathbf{E}(U_k))\|^2_{L^2(\Omega)} \approx \int_\Omega \phi_{|\mathbf{E}(u)|}(|\mathbf{E}(u) - \mathbf{E}(U_k)|)\, dx$ *(see Lemma 74) implies*

$$\mathbf{E}(U_k) \to_{k\to\infty} \mathbf{E}(u) \qquad \text{in } L^\phi(\Omega)^{d\times d}.$$

Hence, the equivalence of the norms $\|\cdot\|_{W_0^{1,\phi}(\Omega)}$ *and* $\|\mathbf{E}(\cdot)\|_{L^\phi(\Omega)}$ *yields*

$$U_k \to_{k\to\infty} u \qquad \text{in } W_0^{1,\phi}(\Omega)^d.$$

Therfore, we can handle problems of the form (3.45)*, too.*

Chapter 4

Adaptive Uzawa Finite Element Method for the Nonlinear Stationary Stokes Problem

The nonlinear stationary Stokes equations are a well established physical model of, e.g., steady, viscous, incompressible quasi-Newtonian fluids; see Section 1.1. This chapter is concerned with the numerical solution of this problem. In the first part, we state the problem and proof existence and uniqueness of a solution. The second section §4.2 is concerned with a convergent *quasi-steepest descent* algorithm, which is a generalization of the Uzawa algorithm for the linear case. In the last part we proof convergence of a practicable adaptive Uzawa algorithm (AUA) using finite elements.

4.1 Nonlinear Stationary Stokes Equations

In this Section we introduce the nonlinear stationary stokes equation for a certain class of N-functions. We give a short overview on existence and uniqueness of solutions and finally, we introduce an equivalent minimizing problem that is crucial for the convergent adaptive algorithm in Sections 4.2 and 4.3.

4.1.1 Stating the Problem

In the following, let ϕ be a fixed N-function that satisfies Assumption 40. We discuss problems of the form: Find functions $u : \Omega \to \mathbb{R}^d$, $p : \Omega \to \mathbb{R}$, such that for a given right-hand side $f : \Omega \to \mathbb{R}^d$

$$\begin{aligned} -\operatorname{div} \mathbf{A}(\nabla u) + \nabla p &= f && \text{in } \Omega, \\ \operatorname{div} u &= 0 && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{4.1}$$

Thereby, the vector-field $\mathbf{A} : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ is defined as

$$\mathbf{A}(\mathbf{Q}) := \phi'(|\mathbf{Q}|)\frac{\mathbf{Q}}{|\mathbf{Q}|}.$$

For the weak formulation of (4.1) we suppose that $f \in L^{\phi^*}(\Omega)$. We are looking for $u \in W_0^{1,\phi}(\Omega)^d$, $p \in L^{\phi^*}(\Omega)/\mathbb{R}$, such that

$$\begin{aligned} \int_\Omega \mathbf{A}(\nabla u) : \nabla v\,dx - \int_\Omega p \operatorname{div} v\,dx &= \int_\Omega f\cdot v\,dx \quad &&\text{for all } v \in W_0^{1,\phi}(\Omega)^d,\\ \int_\Omega q \operatorname{div} u\,dx &= 0 &&\text{for all } q \in L^{\phi^*}(\Omega)/\mathbb{R}. \end{aligned} \tag{4.2}$$

Remark 114. *We recall the definition of the viscosity of quasi Newtonian fluids in Section 1.1. If we take*

$$\nu(t) = \frac{\phi'(t)}{t}, \qquad \text{for } t \geq 0,$$

then, for $r \in (1,\infty)$,

$$\phi(t) = \frac{1}{r}t^r, \qquad \mathbf{A}(\mathbf{Q}) = (|\mathbf{Q}|)^{r-2}\mathbf{Q}$$

correspond to the power law, whereas for $r \in (1,\infty)$, $\kappa \geq 0$ and $\nu_0 > \nu_\infty \geq 0$, the N-function and vector-field

$$\begin{aligned} \phi(t) &= \int_0^t \big(\nu_\infty + (\nu_0-\nu_\infty)(\kappa^2+s^2)^{\frac{r-2}{2}}\big)s\,ds,\\ \mathbf{A}(\mathbf{Q}) &= \big(\nu_\infty + (\nu_0-\nu_\infty)(\kappa^2+|\mathbf{Q}|^2)^{\frac{r-2}{2}}\big)\mathbf{Q} \end{aligned}$$

correspond to the Carreau law. Due to this physical interpretation of the nonlinear stationary Stokes problem, we call u velocity and p the pressure. Consequently, we call $W_0^{1,\phi}(\Omega)$ the velocity space and $L^{\phi^*}(\Omega)/\mathbb{R}$ the pressure space.

Remark 115. *Recalling Remark 39, we observe that the above problem is well posed, since $\operatorname{div} v \in L^{\phi}(\Omega) = (L^{\phi^*}(\Omega))^*$. Furthermore, the choice of the pressure space $L^{\phi^*}(\Omega)/\mathbb{R}$ is reasonable, since the pressure is only determined up to a constant. In particular, it holds for $q \in L^{\phi^*}(\Omega)$ and $v \in W_0^{1,\phi}(\Omega)$ that*

$$\int_\Omega (q+c) \operatorname{div} v\,dx = \int_\Omega q \operatorname{div} v\,dx - \int_\Omega (\nabla c)\cdot v\,dx = \int_\Omega q \operatorname{div} v\,dx + 0$$

for all $c \in \mathbb{R}$.

4.1.2 Existence and Uniqueness of Solutions

Existence and uniqueness of a solution of (4.2) are closely connected to the so called inf-sup condition. As is shown in [4] for $r > 1$, $\frac{1}{r} + \frac{1}{r'} = 1$, there exists a constant $\beta_0 > 0$ such that

$$\inf_{q \in L^{r'}(\Omega)/\mathbb{R}} \sup_{v \in W_0^{1,r}(\Omega)^d} \frac{\int_\Omega q \operatorname{div} v \, dx}{\|v\|_{W_0^{1,r}(\Omega)^d} \|q\|_{L^{r'}(\Omega)/\mathbb{R}}} > \beta_0. \tag{4.3}$$

In particular, the inf-sup condition asserts that

$$\|\nabla q\|_{W^{-1,r'}(\Omega)} \geq \beta_0 \|q\|_{L^{r'}(\Omega)/\mathbb{R}}$$

for all $q \in L^{r'}(\Omega)/\mathbb{R}$, where $\langle \nabla q, v \rangle := -\int_\Omega q \operatorname{div} v \, dx$ for $v \in W_0^{1,r}(\Omega)$. For this reason we restrict ourselves to a certain class of N-functions; compare also Remark 123.

Assumption 116. Let ϕ be an N-function that satisfies Assumption 40. We suppose that there exists $r > 1$ and $t_0 \geq 0$, such that

$$\phi(t) \approx t^r \qquad \text{for all } t \geq t_0.$$

Corollary 117. *Let ϕ be an N-function that satisfies Assumption 116 for an $r > 1$. Then,*

$$L^\phi(\Omega) = L^r(\Omega), \quad L^{\phi*}(\Omega) = L^{r'}(\Omega), \quad \textit{and} \quad W_0^{1,\phi}(\Omega) = W_0^{1,r}(\Omega),$$

with $\frac{1}{r} + \frac{1}{r'} = 1$. Moreover, the norms of each pair of function spaces are equivalent and therefore there exists $\beta > 0$, such that

$$\inf_{q \in L^{\phi*}(\Omega)/\mathbb{R}} \sup_{v \in W_0^{1,\phi}(\Omega)^d} \frac{\int_\Omega q \operatorname{div} v \, dx}{\|v\|_{W_0^{1,\phi}(\Omega)^d} \|q\|_{L^{\phi*}(\Omega)/\mathbb{R}}} > \beta. \tag{4.4}$$

Proof. The claim $L^\phi(\Omega) = L^r(\Omega)$ follows from Proposition 29 and the equivalence of their norms follows from Lemma 35. Thanks to (2.6e), the claim for the second pair of function spaces follows analogously. The assertion for the last pair of spaces, $W_0^{1,\phi}(\Omega) = W_0^{1,r}(\Omega)$, follows by the definition of their particular norms: In fact, their norms are defined via the $L^r(\Omega)$ and $L^\phi(\Omega)$ norms, respectively. As shown above, the norms of $L^r(\Omega)$ and $L^\phi(\Omega)$ are equivalent and hence the norms of $W_0^{1,\phi}(\Omega)$ and $W_0^{1,r}(\Omega)$ are also equivalent. Finally, $C_0^\infty(\Omega)$ is dense in each of the spaces, and therefore $W_0^{1,\phi}(\Omega) = W_0^{1,r}(\Omega)$.

The inf-sup condition (4.4) follows from (4.3) and the equivalence of the particular norms. □

Remark 118. *Basic calculations yield for all* $t \geq \kappa$

$$\frac{1}{r} t^r \leq \int_0^t (\kappa + s)^{r-2} s \, ds \leq 2^{r-2} \frac{1}{r} t^r,$$

if $r > 2$*. In the case* $r \in (1,2)$*, the inverse estimates hold true. Similar estimates can be shown for* $t \mapsto \int_0^t (\kappa^2 + s^2)^{\frac{r-2}{2}} s \, ds$*. In the case of the Carreau law it holds for all* $t \geq \kappa$

$$\int_0^t (\nu_\infty + (\nu_0 - \nu_\infty)(\kappa^2 + s^2)^{\frac{r-2}{2}}) s \, ds \approx t^{\max\{2,r\}},$$

where $\nu_0 > \nu_\infty > 0$ *and* $\kappa \geq 0$*. Hence, among many others, the class of N-functions satisfying Assumption 116 covers the most common nonlinearities appearing in the modeling of quasi-Newtonian flow like the power law and the Carreau law; see Section 1.1.*

However, we want to emphasize that we only miss an inf-sup condition for general N-functions and that beyond the inf-sup condition there is no need for any restriction to r-integrable functions; see also Remarks 123 and 142. To indicate that we do not use techniques particularly related to r-integrability we decided to keep the notation of the spaces via N-functions, i.e., we write $W_0^{1,\phi}(\Omega)$ instead of $W_0^{1,r}(\Omega)$, $L_0^\phi(\Omega)$ instead of $L_0^r(\Omega)$, and $L^{\phi^*}(\Omega)/\mathbb{R}$ instead of $L^{r'}(\Omega)/\mathbb{R}$; see also Corollary 117.

We start with two abstract results about Lagrange multipliers; see [79, Proposition 43.1] and [79, Corollary 43.2].

Proposition 119. *Assume that the following two conditions hold:*

i) X *and* Y *are real Banach-spaces.*

ii) The operators $\mathcal{A} : X \to \mathbb{R}$ *and* $\mathcal{B} : X \to Y$ *are continuous linear operators and* $\mathbf{R}(\mathcal{B}) := \{\mathcal{B}x : x \in X\}$ *is closed.*

Then if $\mathcal{A}h = 0$ *for all* $h \in X$ *such that* $\mathcal{B}h = 0$ *holds, there exists a* $\Lambda \in Y^*$ *such that*

$$\mathcal{A}k + \Lambda(\mathcal{B}k) = 0 \qquad \text{for all } k \in X.$$

For $\mathbf{R}(\mathcal{B}) = Y$*,* Λ *is unique.*

Corollary 120. *Suppose the assumptions of Proposition 119. If* $\mathbf{R}(\mathcal{B}) \neq Y$*, then, by the assumptions i) and ii), there exists a* $\Lambda \in Y^*$*,* $\Lambda \neq 0$*, such that*

$$\Lambda(\mathcal{B}k) = 0 \qquad \text{for all } k \in X.$$

In the following we discuss how Proposition 119 can be applied to problem (4.2) in order to obtain its unique solvability. In particular, we take $\mathcal{B} := \operatorname{div}$, $X = W_0^{1,\phi}(\Omega)^d$, and $Y = L_0^\phi(\Omega)$. Thus, $\mathcal{B}$ is a continuous linear operator on Banach spaces and we have that the subspace

$$Z := \{v \in W_0^{1,\phi}(\Omega)^d : \operatorname{div} v = 0\} \subset W^{1,\phi}(\Omega)^d$$

is closed. Therefore, with Corollaries 55 and 50, there exists a unique $u \in Z$ such that

$$\int_\Omega \mathbf{A}(\nabla u) : \nabla v \, dx = \int_\Omega f \cdot v \, dx \qquad \text{for all } v \in Z,$$

where we use the notation of (4.2). Now, we define the linear operator

$$\mathcal{A}v := \int_\Omega \mathbf{A}(\nabla u) : \nabla v \, dx - \int_\Omega f \cdot v \, dx \qquad \text{for } v \in W_0^{1,\phi}(\Omega),$$

which is continuous from $W_0^{1,\phi}(\Omega)^d$ to $\mathbb{R}$; see Lemma 48. The next lemma specifies the space of the Lagrange multiplicator Λ of Proposition 119.

Lemma 121. *Let ϕ be an N-function that satisfies Assumption 40, then*

$$\Big(L_0^\phi(\Omega), \|\cdot\|_\phi\Big)^* = \Big(L^{\phi^*}(\Omega)/\mathbb{R}, \inf_{c\in\mathbb{R}} \|\cdot - c\|_{(\phi^*)}\Big)$$

and

$$\Big(L_0^\phi(\Omega), \|\cdot\|_\phi\Big) = \Big(L^{\phi^*}(\Omega)/\mathbb{R}, \inf_{c\in\mathbb{R}} \|\cdot - c\|_{(\phi^*)}\Big)^*.$$

Proof. By the Hahn-Banach theorem we have $(L_0^\phi(\Omega))^* = L^{\phi^*}(\Omega)|_{L_0^\phi(\Omega))}$ and since $\int_\Omega ch\, dx = 0$ for all $c \in \mathbb{R}$, $h \in L_0^\phi(\Omega)$, it follows $(L_0^\phi(\Omega))^* \subset L^{\phi^*}(\Omega)/\mathbb{R}$. Let $q \in L^{\phi^*}(\Omega)$ such that $\langle q, h\rangle = 0$ for all $h \in L_0^\phi(\Omega)$. Then for all $\psi \in L^\phi(\Omega)$

$$0 = \int_\Omega q\,(\psi - \langle\psi\rangle)\, dx = \int_\Omega (q - \langle q\rangle)(\psi - \langle\psi\rangle)\, dx = \int_\Omega (q - \langle q\rangle)\,\psi\, dx, \tag{4.5}$$

where $\langle q\rangle := \frac{1}{|\Omega|}\int_\Omega q\, dx$. Therefore, we proved that any linear functional $\ell \in (L_0^\phi(\Omega))^*$ is representable in the form

$$\ell(h) = \int_\Omega q\, h\, dx, \qquad h \in L_0^\phi(\Omega),$$

with a $q \in L^{\phi^*}(\Omega)/\mathbb{R}$ and vice versa. It remains to prove that the norms on $(L_0^\phi(\Omega))^*$ and $L^{\phi^*}(\Omega)/\mathbb{R}$ are equal. We observe that Propositions 25 and 26 imply for $q \in L^{\phi^*}(\Omega)$

$$\begin{aligned}
\|q\|_{(L_0^\phi(\Omega))^*} &= \sup_{h \in L_0^\phi(\Omega), \|h\|_\phi = 1} \int_\Omega q\, h\, dx \\
&= \inf_{c \in \mathbb{R}} \sup_{h \in L_0^\phi(\Omega), \|h\|_\phi = 1} \int_\Omega (q-c)\, h\, dx \\
&\leq \inf_{c \in \mathbb{R}} \sup_{h \in L_0^\phi(\Omega), \|h\|_\phi = 1} \|q-c\|_{(\phi^*)} \|h\|_\phi = \inf_{c \in \mathbb{R}} \|q-c\|_{(\phi^*)} .
\end{aligned}$$

Thus, it suffices to show that

$$\sup_{h \in L_0^\phi(\Omega), \|h\|_\phi = 1} \int_\Omega q\, h\, dx \geq \inf_{c \in \mathbb{R}} \|q-c\|_{(\phi^*)} . \tag{4.6}$$

Let $q_0 \in L^{\phi^*}(\Omega)/\mathbb{R}$ be fixed, then by the considerations above, q_0 defines a linear functional on $L_0^\phi(\Omega)$. Since $L_0^\phi(\Omega)$ is a closed subspace of $L^\phi(\Omega)$, we know — by the Hahn-Banach extension theorem (cf. [78]) — that there exists $\bar{q}_0 \in \left(L^{\phi^*}(\Omega), \|\cdot\|_{(\phi^*)}\right) = \left(L^\phi(\Omega), \|\cdot\|_\phi\right)^*$, such that $\bar{q}_0$ is an extension of q_0, i.e.,

$$\int_\Omega q_0\, h\, dx = \int_\Omega \bar{q}_0\, h\, dx \qquad \text{for all } h \in L_0^\phi(\Omega)$$

and $\bar{q}_0$ and q_0 have equal operator norms

$$\sup_{h \in L_0^\phi(\omega), \|h\|_{\phi^*} = 1} \int_\Omega q_0\, h\, dx = \sup_{k \in L^\phi(\omega), \|k\|_{\phi^*} = 1} \int_\Omega \bar{q}_0\, k\, dx = \|\bar{q}_0\|_{(\phi^*)} ;$$

see also Propositions 25 and 26. Since, by (4.5), $\bar{q}_0$ and any other representative of q_0 only differ up to a constant, we have

$$\|\bar{q}_0\|_{(\phi^*)} \geq \inf_{c \in \mathbb{R}} \|q_0 - c\|_{(\phi^*)} .$$

Hence, (4.6) is established. The second claim states the reflexivity of $L_0^\phi(\Omega) \subset L^\phi(\Omega)$. Since closed subspaces of reflexive Banach spaces are reflexive [37, II.3.23] the assertion follows from the reflexivity of $L^\phi(\Omega)$; see Remark 27. □

Note with the help of Lemma 121, that $\nabla : L^{\phi^*}(\Omega)/\mathbb{R} \to W_0^{-1,\phi^*}(\Omega)^d$ is the dual operator of $\operatorname{div} : W_0^{1,\phi}(\Omega)^d \to L_0^\phi(\Omega)$ and observe that $\operatorname{div} : W_0^{1,\phi}(\Omega)^d \to L_0^\phi(\Omega)$ is a closed operator. Recalling the closed range theorem (see, e.g., [78] and [15]) the inf-sup condition (4.4) is equivalent to

$$\mathbf{R}(\operatorname{div}) = \mathbf{N}(\nabla)^\perp,$$

where $\mathbf{N}(\nabla)$ denotes the kernel of ∇ in $L^{\phi^*}(\Omega)/\mathbb{R}$. Moreover, by the inf-sup condition (4.4) we have that $\nabla : L^{\phi^*}(\Omega)/\mathbb{R} \to W^{-1,\phi^*}(\Omega)^d$ is injective, i.e., $\mathbf{N}(\nabla) = \{0\}$ and therefore

$$\mathbf{R}(\mathrm{div}) = \{0\}^\perp = L_0^\phi(\Omega).$$

Hence, we proved $\mathbf{R}(\mathcal{B}) = Y$ and therefore Proposition 119 yields the following existence and uniqueness result.

Theorem 122. *Let ϕ be an N-function that satisfies Assumption 116. Then there exists a unique solution $(u,p) \in W_0^{1,\phi}(\Omega)^d \times L^{\phi^*}(\Omega)/\mathbb{R}$ of (4.2).*

Remark 123. *For general N-functions no* inf-sup *condition is known so far. The above considerations and Corollary 120 show that the existence and uniqueness of $p \in L^{\phi^*}(\Omega)/\mathbb{R}$ in (4.2) is equivalent to the* inf-sup *condition*

$$\inf_{q \in L^{\phi^*}(\Omega)/\mathbb{R}} \sup_{v \in W_0^{1,\phi}(\Omega)^d} \frac{\int_\Omega q \,\mathrm{div}\, v \, dx}{\|v\|_{W_0^{1,\phi}(\Omega)^d} \|q\|_{L^{\phi^*}(\Omega)/\mathbb{R}}} > \beta$$

for some $\beta > 0$. We want to emphasize that all subsequent analysis is applicable to N-functions that satisfy Assumption 40 and for which such a inf-sup *condition holds.*

4.1.3 The Lagrangian Function

Following the approach in [40]. For a given N-function ϕ, we define the Lagrangian function $\mathcal{L} : W_0^{1,\phi}(\Omega)^d \times L^{\phi^*}(\Omega)/\mathbb{R} \to \mathbb{R}$ of (4.2) by

$$\mathcal{L}(v,q) := \int_\Omega \phi(|\nabla v|) - q \,\mathrm{div}\, v - f \cdot v \, dx.$$

For the ease of exposition, we will use the abbreviations

$$\mathbb{V} := W_0^{1,\phi}(\Omega)^d \qquad \text{and} \qquad \mathbb{Q} := L^{\phi^*}(\Omega)/\mathbb{R}$$

in the remainder of this chapter.

Proposition 124. *Let ϕ be an N-function that satisfies Assumption 116. Then the nonlinear Stokes problem (4.2) is equivalent to the saddle-point problem: Find functions $u \in \mathbb{V}$, $p \in \mathbb{Q}$, such that*

$$\inf_{v \in \mathbb{V}} \mathcal{L}(v,p) = \mathcal{L}(u,p) = \sup_{q \in \mathbb{Q}} \mathcal{L}(u,q), \tag{4.7}$$

i.e., the unique solution $(u,p) \in \mathbb{V} \times \mathbb{Q}$ of (4.2) is the unique saddle-point of $\mathcal{L}$.

Proof. Let (u, p) be the solution of (4.2). From

$$\int_\Omega q \operatorname{div} u \, dx = 0 \qquad \text{for all } q \in \mathbb{Q},$$

we get

$$\mathcal{L}(u, q) = \mathcal{L}(u, p), \qquad \text{for all } q \in \mathbb{Q}.$$

Hence, the second equality of (4.7) is established. We observe further, that u is the unique solution of the nonlinear Poisson equation (3.2) with right hand side $g = f - \nabla p \in W^{-1,\phi^*}(\Omega)^d$; see Theorem 49. Recalling Theorem 54, u is the unique minimizer of $\mathcal{L}(\cdot, p)$, which implies the left equality in (4.7).

On the other hand, let $(u, p) \in \mathbb{V} \times \mathbb{Q}$ be a saddle-point of $\mathcal{L}$, then we have that u is a minimizer of $\mathcal{L}(\cdot, p)$ and thus Theorem 54 yields

$$\int_\Omega \mathbf{A}(\nabla u) : \nabla v \, dx = \int_\Omega p \operatorname{div} v + f \cdot v \, dx \qquad \text{for all } v \in \mathbb{V}$$

Finally, the right equality of (4.7) implies

$$\mathcal{L}(u, q) - \mathcal{L}(u, p) = \int_\Omega (p - q) \operatorname{div} u \, dx \leq 0 \qquad \text{for all } q \in \mathbb{Q}.$$

Since $p \in \mathbb{Q}$ is arbitrary, this yields

$$\int_\Omega q \operatorname{div} u \, dx = 0 \qquad \text{for all } q \in \mathbb{Q}.$$

Therefore, we have proved that the solution of the saddle-point problem (4.7) is a solution of (4.2). Hence, the uniqueness of the saddle-point problem then follows by the uniqueness of solutions of (4.2); see Theorem 122. □

The following proposition is a general property of saddle-points; see e.g. [40, VI, Proposition 1.2].

Proposition 125. *Suppose the conditions of Proposition 124, then*

$$\sup_{q \in \mathbb{Q}} \inf_{v \in \mathbb{V}} \mathcal{L}(v, q) = \mathcal{L}(u, p) = \inf_{v \in \mathbb{V}} \sup_{q \in \mathbb{Q}} \mathcal{L}(v, q).$$

Based on the above results we define the nonlinear functional $\mathcal{F} : \mathbb{Q} \to \mathbb{R}$ by

$$\mathcal{F}(q) := - \inf_{v \in \mathbb{V}} \mathcal{L}(v, q) \qquad \text{for all } q \in \mathbb{Q}. \tag{4.8}$$

According to Proposition 125 our aim is to minimize $\mathcal{F}$.

Corollary 126. *Under the conditions of this section, the functional $\mathcal{F} : \mathbb{Q} \to \mathbb{R}$ possesses a unique minimizer $p \in \mathbb{Q}$.*

Proof. The assertion is an immediate consequence of Propositions 124 and 125. □

Note from the definition of the Lagrangian function, that evaluating $\mathcal{F}$ at $q \in \mathbb{Q}$ is a minimizing problem of the form (3.11), with $g = f - \nabla q \in W^{-1\phi^*}(\Omega)$. Hence, by Theorem 54, the unique minimizer $u_q \in \mathbb{V}$ of

$$\mathcal{F}(q) = -\mathcal{L}(u_q, q) = -\inf_{v \in \mathbb{V}} \mathcal{L}(v, q). \tag{4.9}$$

is the unique solution of the elliptic equation

$$\int_\Omega \mathbf{A}(\nabla u_q) : \nabla v \, dx = \int_\Omega f \cdot v + q \operatorname{div} v \, dx \quad \text{for all } v \in \mathbb{V}. \tag{4.10}$$

In the following, we will analyze the functional $\mathcal{F}$.

Proposition 127. *Under the conditions of Proposition 124 let $\mathcal{F} : \mathbb{Q} \to \mathbb{R}$ be defined as in* (4.8). *Then the mapping*

$$q \mapsto u_q,$$

defined by (4.9)*, is continuous from $\mathbb{Q}$ to $\mathbb{V}$. Moreover, $\mathcal{F} : \mathbb{Q} \to \mathbb{R}$ is continuous.*

In order to prove Proposition 127 we need some technical Lemmas. We start with a basic observation that will be used frequently in the following.

Lemma 128. *For an N-function ϕ with $\Delta_2(\phi) < \infty$ holds*

$$\phi_a(|\mathrm{tr}(\mathbf{Q})|) \preccurlyeq \phi_a(|\mathbf{Q}|)$$

for all $a \geq 0$ and $\mathbf{Q} = (Q_{ij})_{i,j} \in \mathbb{R}^{d \times d}$, where $\mathrm{tr}(\mathbf{Q}) = \sum_{i=1}^d Q_{ii}$. The constant hidden in $\preccurlyeq$ depends solely on $\Delta_2(\phi)$ and d.

Proof. First, we observe that $|\mathrm{tr}(\mathbf{Q})| \leq \sqrt{d}\,|\mathbf{Q}|$ for all $\mathbf{Q} \in \mathbb{R}^{d \times d}$. Therefore, the monotonicity of ϕ_a implies

$$\phi_a(|\mathrm{tr}(\mathbf{Q})|) \leq \phi_a(\sqrt{d}\,|\mathbf{Q}|).$$

Now, the assertion follows by Corollary 10, recalling that the Δ_2-constant of ϕ_a is bounded uniformly in $a \geq 0$; see Lemma 57. □

The next Lemma states that we can use $L_0^{\phi^*}(\Omega)$ as a representation space for $L^{\phi^*}(\Omega)/\mathbb{R}$.

Lemma 129. *Let ϕ be an N-function that satisfies Assumption 116. Then it holds*

$$\|q - \langle q \rangle\|_{(\phi^*)} \leq 2 \inf_{c \in \mathbb{R}} \|q - c\|_{(\phi^*)} \leq 2\, \|q - \langle q \rangle\|_{(\phi^*)}$$

for all $q \in L^{\phi^*}(\Omega)$, where $\langle q \rangle := \frac{1}{|\Omega|} \int_\Omega q \, dx$.

Proof. We have to show the equivalence of norms of $L^{\phi^*}(\Omega)/\mathbb{R}$ and $L_0^{\phi^*}(\Omega)$. It is clear that $\inf_{c\in\mathbb{R}} \|q-c\|_{(\phi^*)} \le \|q-\langle q\rangle\|_{(\phi^*)}$. On the other hand we have for any $c \in \mathbb{R}$

$$\|q-\langle q\rangle\|_{(\phi^*)} \le \|q-c\|_{(\phi^*)} + \|c-\langle q\rangle\|_{(\phi^*)} . \tag{4.11}$$

For the second summand of the right hand side, we obtain by Jensen's inequality (Lemma 4)

$$\begin{aligned}\int_\Omega \phi^*(|c-\langle q\rangle|)\,dx &\le \int_\Omega \phi^*\Big(\frac{1}{|\Omega|}\int_\Omega |c-q|\;dy\Big)\,dx \\ &\le \int_\Omega \frac{1}{|\Omega|}\int_\Omega \phi^*(|c-q|)\,dy\,dx = \int_\Omega \phi^*(|c-q|)\,dy.\end{aligned}$$

Therfore, by the definition of the Minkowski functional (2.13) we have for all $c \in \mathbb{R}$

$$\|c-\langle q\rangle\|_{(\phi^*)} \le \|c-q\|_{(\phi^*)} .$$

Applying this to (4.11) we get

$$\|q-\langle q\rangle\|_{(\phi^*)} \le 2\,\|q-c\|_{(\phi^*)} ,$$

which is the desired estimate since $c \in \mathbb{R}$ is arbitrary. □

Corollary 130. *Let $w \in W_0^{1,\phi}(\Omega)$ and $(q_n)_{n\in\mathbb{N}} \subset L^{\phi^*}(\Omega)$. Under the conditions of Lemma 129 the following assertions are equivalent:*

i)

$$\int_\Omega \big(\phi_{|\nabla w|}\big)^*(|q_n-\langle q_n\rangle|)\,dx \to 0, \qquad \textit{as } n\to\infty;$$

ii)

$$\inf_{c\in\mathbb{R}} \int_\Omega \big(\phi_{|\nabla w|}\big)^*(|q_n-c|)\,dx \to 0, \qquad \textit{as } n\to\infty;$$

iii)

$$\inf_{c\in\mathbb{R}} \|q_n-c\|_{(\phi^*)} \to 0, \qquad \textit{as } n\to\infty.$$

Proof. It holds

$$\inf_{c\in\mathbb{R}} \int_\Omega \big(\phi_{|\nabla w|}\big)^*(|q_n-c|)\,dx \le \int_\Omega \big(\phi_{|\nabla w|}\big)^*(|q_n-\langle q_n\rangle|)\,dx$$

for all $n \in \mathbb{N}$. Thus i) implies ii).

Now, assuming ii) we observe for fixed $n \in \mathbb{N}$ that the real function $c \mapsto \int_\Omega \big(\phi_{|\nabla w|}\big)^*(|q_n - c|)\, dx$ is continuous and tends to infinity as $|c|$ tends to infinity. Thus, it attains its minimum. Denoting a minimizer by $c_n \in \mathbb{R}$, it follows by ii) that

$$\int_\Omega \big(\phi_{|\nabla w|}\big)^*(|q_n - c_n|)\, dx \to 0 \quad \text{as } n \to \infty.$$

Hence, Lemma 76 implies $\|q_n - c_n\|_{(\phi^*)} \to 0$ as $n \to \infty$. The estimate

$$\inf_{c \in \mathbb{R}} \|q_n - c\|_{L^{\phi^*}(\Omega)} \le \|q_n - c_n\|_{L^{\phi^*}(\Omega)}$$

yields that ii) implies iii).

The fact that iii) implies i) can be deduced from Lemma 76 and the equivalence of norms in Lemma 129. $\square$

Lemma 131. *Let ϕ be an N-function that satisfies $\Delta_2(\{\phi, \phi^*\}) < \infty$. Then the functional $\mathcal{L} : \mathbb{V} \times \mathbb{Q} \to \mathbb{R}$ is continuous.*

Proof. Let $v, w \in \mathbb{V}$ and $q, h \in \mathbb{Q}$. Then, by the triangle inequality we have

$$\begin{aligned} |\mathcal{L}(v,q) - \mathcal{L}(w,h)| &\le \Big| \int_\Omega \phi(|\nabla v|) - \phi(|\nabla w|)\, dx \Big| \\ &\quad + \Big| \int_\Omega q \operatorname{div} v - h \operatorname{div} w\, dx \Big| + \Big| \int_\Omega f \cdot (v - w)\, dx \Big|. \end{aligned} \tag{4.12}$$

The first addend at the right hand side can be estimated by the quasi triangle inequality (Corollary 10)

$$\Big| \int_\Omega \phi(|\nabla v|) - \phi(|\nabla w|)\, dx \Big| \preccurlyeq \Big| \int_\Omega \phi(|\nabla v - \nabla w|)\, dx \Big|.$$

Thanks to the equivalence of norm-convergence and mean convergence (Proposition 31) this term becomes small as $\|v - w\|_{\mathbb{V}}$ becomes small. For the second addend we estimate

$$\begin{aligned} \Big| \int_\Omega q \operatorname{div} v - h \operatorname{div} w\, dx \Big| &= \Big| \int_\Omega q \operatorname{div} v - q \operatorname{div} w + q \operatorname{div} w - h \operatorname{div} w\, dx \Big| \\ &\le \Big| \int_\Omega q \operatorname{div}(v - w)\, dx \Big| + \Big| \int_\Omega (q - h) \operatorname{div} w\, dx \Big|. \end{aligned}$$

Recalling that the pressure is determinated up to a constant we obtain by Proposition 24

$$\Big| \int_\Omega q \operatorname{div} v - h \operatorname{div} w\, dx \Big| \le \|q - \tilde{c}\|_{\phi^*} \|\operatorname{div}(v - w)\|_{(\phi)} + \|q - h - \hat{c}\|_{\phi^*} \|\operatorname{div} v\|_{(\phi)}$$

for all $\tilde{c}, \hat{c} \in \mathbb{R}$. Taking the infimum over all $\tilde{c}, \hat{c}$, applying the point-wise estimate of Lemma 128, and (2.14), we can further deduce

$$\Big| \int_\Omega q \operatorname{div} v - h \operatorname{div} w \, dx \Big| \preceq \|q\|_{\mathbb{Q}} \|v - w\|_{\mathbb{V}} + \|q - h\|_{\mathbb{Q}} \|v\|_{\mathbb{V}} ,$$

which becomes small as $\|v - w\|_{\mathbb{V}}$ and $\|q - h\|_{\mathbb{Q}}$ becomes small — provided $\|q\|_{\mathbb{Q}}$ and $\|v\|_{\mathbb{V}}$ stay bounded. The last term of the right hand side of (4.12) can be estimated by Proposition 24

$$\Big| \int_\Omega f \cdot (v - w) \, dx \Big| \leq \|f\|_{(\phi^*)} \|v - w\|_\phi \leq \|f\|_{(\phi^*)} \|v - w\|_{\mathbb{V}} .$$

Hence, this term also becomes small as $\|v - w\|_{\mathbb{V}}$ becomes small. Applying these estimates to (4.12) yields the assertion. □

Proof of Proposition 127. From the preceeding considerations we know that u_q solves (4.10). According to Lemma 129 we can choose $q, h \in L_0^{\phi^*}(\Omega)$ as representatives of functions in $\mathbb{Q}$. It holds

$$\int_\Omega \big(\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u_{q+h})\big) : \nabla v \, dx = \int_\Omega h \operatorname{div} v \, dx \quad \text{for all } v \in \mathbb{V}.$$

Taking $v = u_q - u_{q+h}$ and applying Young's inequality (Proposition 11) we get

$$\int_\Omega \big(\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u_{q+h})\big) : \nabla(u_q - u_{q+h}) \, dx \\ \leq \int_\Omega C_\delta \, \big(\phi_{|\nabla u_q|}\big)^*(|h|) + \delta \, \phi_{|\nabla u_q|}(|\operatorname{div}(u_q - u_{q+h})|) \, dx.$$

Lemma 128 then implies

$$\int_\Omega \big(\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u_{q+h})\big) : \nabla(u_q - u_{q+h}) \, dx \\ \preceq \int_\Omega C_\delta \, \big(\phi_{|\nabla u_q|}\big)^*(|h|) + \delta \, \phi_{|\nabla u_q|}(|\nabla(u_q - u_{q+h})|) \, dx.$$

According to Lemma 74, for δ small enough, we obtain

$$\int_\Omega \phi_{|\nabla u_q|}(|\nabla(u_q - u_{q+h})|) \, dx \preceq \int_\Omega \big(\phi_{|\nabla u_q|}\big)^*(|h|) \, dx \approx \int_\Omega \phi^*_{\phi'(|\nabla u_q|)}(|h|) \, dx.$$

Now, Lemmas 76 and 129 imply the desired result.

The continuity of $\mathcal{F}$ follows from the continuity of $\mathcal{L}$ on $\mathbb{V} \times \mathbb{Q}$ (Lemma 131) and the continuity of $q \mapsto u_q$. □

We will now conclude our analytical considerations proving some properties of $\mathcal{F}$, which will be crucial in the convergence analysis of Sections 4.2 and 4.3.

Proposition 132. *Let ϕ be an N-function that satisfies Assumption 116. Then, the functional $\mathcal{F} : \mathbb{Q} \to \mathbb{R}$ defined in (4.8) is strictly convex.*

Proof. Let $q_1, q_2 \in \mathbb{Q}$ with $q_1 \neq q_2$, then for $t \in (0,1)$

$$\mathcal{L}(v, t\, q_1 + (1-t)\, q_2) = t\, \mathcal{L}(v, q_1) + (1-t)\, \mathcal{L}(v, q_2), \qquad \text{for all } v \in \mathbb{V},$$

since $\mathcal{L}$ is linear in its second argument. The strict convexity follows from recalling that u_q is the unique minimizer of $\mathcal{L}(\cdot, q)$. In particular,

$$\mathcal{L}(u_{t\, q_1 + (1-t)\, q_2}, q_1) < \mathcal{L}(u_{q_1}, q_1)$$

and

$$\mathcal{L}(u_{t\, q_1 + (1-t)\, q_2}, q_2) < \mathcal{L}(u_{q_2}, q_2)$$

for all $t \in (0,1)$. Hence,

$$\begin{aligned} \mathcal{F}(t\, q_1 + (1-t)\, q_2) &= -\mathcal{L}(u_{t\, q_1 + (1-t)\, q_2}, t\, q_1 + (1-t)\, q_2) \\ &= -t\, \mathcal{L}(u_{t\, q_1 + (1-t)\, q_2}, q_1) - (1-t)\, \mathcal{L}(u_{t\, q_1 + (1-t)\, q_2}, q_2) \\ &< -t\, \mathcal{L}(u_{q_1}, q_1) - (1-t)\, \mathcal{L}(u_{q_2}, q_2) \\ &= t\, \mathcal{F}(q_1) + (1-t)\, \mathcal{F}(q_2). \end{aligned}$$

This finishes the proof. □

Proposition 133. *Let ϕ be an N-function that satisfies Assumption 116. For $q \in \mathbb{Q}$, let u_q be the uniquely determined function from (4.10). The functional $\mathcal{F}$ is Fréchet differentiable in q with derivative $D\mathcal{F}(q) = \operatorname{div} u_q \in \mathbb{Q}^*$, i.e.,*

$$\langle D\mathcal{F}(q), h \rangle = \int_\Omega h \operatorname{div} u_q \, dx, \qquad \textit{for } h \in \mathbb{Q}.$$

Proof. To prove the assertion, we have to show that

$$\mathcal{F}(q+h) - \mathcal{F}(q) - \int_\Omega h \operatorname{div} u_q \, dx = o(\|h\|_{\mathbb{Q}}).$$

According to (4.9) and the definition of the Lagrangian function we have

$$\begin{aligned}
&\mathcal{F}(q+h)-\mathcal{F}(q)-\int_\Omega h\ \operatorname{div} u_q\,dx \\
&\qquad\qquad = -\mathcal{L}(u_{q+h},q+h)+\mathcal{L}(u_q,q)-\int_\Omega h\ \operatorname{div} u_q\,dx \\
&\qquad\qquad = \int_\Omega \phi(|\nabla u_q|)-q\ \operatorname{div} u_q - f\cdot u_q\,dx \\
&\qquad\qquad\quad - \int_\Omega \phi(|\nabla u_{q+h}|)-(q+h)\ \operatorname{div} u_{q+h} - f\cdot u_{q+h}\,dx \\
&\qquad\qquad\quad - \int_\Omega h\ \operatorname{div} u_q\,dx \\
&\qquad\qquad = \int_\Omega \phi(|\nabla u_q|)-q\ \operatorname{div} u_q - f\cdot u_q\,dx \\
&\qquad\qquad\quad - \int_\Omega \phi(|\nabla u_{q+h}|)-q\ \operatorname{div} u_{q+h} - f\cdot u_{q+h}\,dx \\
&\qquad\qquad\quad - \int_\Omega h\ \operatorname{div}(u_q-u_{q+h})\,dx.
\end{aligned}$$

Defining $\mathcal{J}_q$ as

$$\mathcal{J}_q(v):=\int_\Omega \phi(|\nabla v|)-f\cdot v-q\operatorname{div} v\,dx$$

yields

$$\begin{aligned}
&\mathcal{F}(q+h)-\mathcal{F}(q)-\int_\Omega h\ \operatorname{div} u_q\,dx \\
&\qquad\qquad = \mathcal{J}_q(u_q)-\mathcal{J}_q(u_{q+h})-\int_\Omega h\ \operatorname{div}(u_q-u_{q+h})\,dx.
\end{aligned}$$

Note that the definition of $\mathcal{J}_q$ corresponds to the definition of $\mathcal{J}$ in (3.10) with $g=f-\nabla q\in\mathbb{V}^*$. Therefore, since u_q is the minimizer of $\mathcal{J}_q$, we get from Proposition 100 and (4.2)

$$\begin{aligned}
|\mathcal{J}_q(u_q)-\mathcal{J}_q(u_{q+h})| &\approx \int_\Omega \big(\mathbf{A}(\nabla u_q)-\mathbf{A}(\nabla u_{q+h})\big):\big(\nabla u_q-\nabla u_{q+h}\big)\,dx \\
&= -\int_\Omega h\ \operatorname{div}(u_q-u_{q+h})\,dx,
\end{aligned}$$

where the constants hidden in $\approx$ solely depend on $\Delta_2(\{\phi,\phi^*\})$. Hence, it follows that

$$\begin{aligned}
\left|\mathcal{F}(q+h)-\mathcal{F}(q)-\int_\Omega h\ \operatorname{div} u_q\,dx\right| &\preceq \left|\int_\Omega h\ \operatorname{div}(u_q-u_{q+h})\,dx\right| \\
&\le \|h\|_{\mathbb{Q}}\,\|\operatorname{div}(u_q-u_{q+h})\|_\phi\,,
\end{aligned}$$

where we used that $\|\cdot\|_{\mathbb{Q}}$ equals to the operator-norm of $(L_0^\phi(\Omega))^*$; see Lemma 121. Thus, Lemma 128 implies

$$\left|\mathcal{F}(q+h) - \mathcal{F}(q) - \int_\Omega h \operatorname{div} u_q \, dx\right| \preccurlyeq \|h\|_{\mathbb{Q}} \|\nabla(u_q - u_{q+h})\|_\phi .$$

Now, the continuity of $q \mapsto u_q$ (Proposition 127), implies that $\|\nabla(u_q - u_{q+h})\|_\phi \to 0$ as $h \to 0$ in $\mathbb{Q}$. This proves the assertion. □

Corollary 134. *Assume the conditions of Proposition 133. Then* $D\mathcal{F} : \mathbb{Q} \to \mathbb{Q}^*$ *is strictly monotone.*

Proof. Proposition 133 asserts that $\mathcal{F}$ is Fréchet differentiable. The strict convexity of $\mathcal{F}$ (Proposition 132) implies the strict monotonicity of $D\mathcal{F}$; see [79, Proposition 42.6]. □

4.2 Generalized Uzawa Algorithm

This section contains an infinite-dimensional convergent steepest descent algorithm, which is the motivation for the convergent adaptive method for the nonlinear stationary Stokes equation in Section 4.3. It is a generalization of the well known Uzawa method (see, e.g., [73, 15]) to the nonlinear case. In the linear case the method is a contraction for certain relaxation parameters [64, 65]; compare with Remark 141. Due to the lack of an inf-sup condition for the quasi-norm it is currently not possible to show contraction for our nonlinear problem; see Remark 142.

The idea of the algorithm is to approximate the unique minimizer $p \in \mathbb{Q} = L^{\phi^*}(\Omega)/\mathbb{R}$ of

$$q \in \mathbb{Q} : \qquad \mathcal{F}(q) \to \min, \tag{4.13}$$

where $\mathcal{F}$ is defined as in (4.8); see also Corollary 126. Since we know from Proposition 133, that $\mathcal{F}$ is Fréchet differentiable with derivative

$$\langle D\mathcal{F}(q), h\rangle = \int_\Omega h \operatorname{div} u_q \, dx, \qquad \text{for } h \in \mathbb{Q},$$

we may think of using the method of steepest descent; cf. [24].

4.2.1 Quasi-Steepest Descent Direction

For norms, a steepest descent direction $\mathfrak{d} \in \mathbb{Q}$ of $D\mathcal{F}$ in $q \in \mathbb{Q}$ is defined by

$$\|D\mathcal{F}(q)\|_{\mathbb{Q}^*} = \sup_{h\in\mathbb{Q},\, \|h\|_{\phi^*}=1} \langle D\mathcal{F}(q), h\rangle = -\left\langle D\mathcal{F}(q), \frac{\mathfrak{d}}{\|\mathfrak{d}\|_\phi}\right\rangle.$$

However, the experience of Chapter 3 indicates that for nonlinear problems, like (4.13), norms might not be the appropriate concept of distance. Using the concept of quasi-norms the question arises what is the 'steepest' descent in this context. To generalize this principle to the case of quasi-norms, we have to generalize the dual or operator norm. In the case of $\phi(t) = \frac{1}{2}t^2$, i.e., the case when quasi-norm and norm coincide, we know for $l \in L^2_0(\Omega) = (L^2_0(\Omega))^*$, that

$$\frac{1}{2}\|l\|^2_{L^2(\Omega),*} = \sup_{h \in L^2(\Omega)} \Big\{\langle l, h\rangle - \frac{1}{2}\|h\|^2_{L^2(\Omega)}\Big\} = \sup_{h \in L^2(\Omega)} \Big\{\langle l, h\rangle - \int_\Omega \phi(|h|)\,dx\Big\}.$$

This motivates the following definition of the dual quasi-norm; see also Remark 79. For $l \in L^\phi_0(\Omega) = (L^{\phi^*}(\Omega)/\mathbb{R})^*$ (see Lemma 121), $w \in W^{1,\phi}_0(\Omega)$, we define

$$\|l\|^2_{(\nabla w),\mathbb{Q}^*} := \sup_{h \in L^{\phi^*}(\Omega)} \Big\{\langle l, h\rangle - \inf_{c \in \mathbb{R}} \int_\Omega \phi^*_{|\nabla w|}(|h - c|)\,dx\Big\}. \tag{4.14}$$

Recall, that $\langle l, h\rangle = \int_\Omega l\,h\,dx = \int_\Omega l\,(h - \hat{c})\,dx$ for all $\hat{c} \in \mathbb{R}$. We have according to Young's inequality (2.3)

$$\begin{aligned}\langle l, h\rangle - \int_\Omega \big(\phi_{|\nabla w|}\big)^*(|h - c|)\,dx \le \int_\Omega \phi_{|\nabla w|}(|l|) + \big(\phi_{|\nabla w|}\big)^*(|h - c|)\,dx \\ - \int_\Omega \big(\phi_{|\nabla w|}\big)^*(|h - c|)\,dx,\end{aligned}$$

and hence

$$\langle l, h\rangle - \int_\Omega \phi_{|\nabla w|}(|h - c|)\,dx \le \int_\Omega \phi_{|\nabla w|}(|l|)\,dx \tag{4.15}$$

for all $h \in L^\phi(\Omega)$, $c \in \mathbb{R}$.

On the other hand note, that by the properties of the N-function ϕ, for $h \in L^{\phi^*}(\Omega)$ there exists a unique $c_h \in \mathbb{R}$ that minimizes $\int_\Omega \phi^*(|h - c|)\,dx : c \in \mathbb{R}$. Moreover, by the strict convexity of ϕ, c_h is the unique solution of

$$\frac{\partial}{\partial c} \int_\Omega \big(\phi_{|\nabla w|}\big)^*(|h - c|)\,dx\Big|_{c = c_h} = \int_\Omega \big(\phi_{|\nabla w|}\big)^{*\prime}(|h - c_h|)\frac{h - c_h}{|h - c_h|}\,dx = 0.$$

Hence, taking $h = \phi'_{|\nabla w|}(|l|)\frac{l}{|l|} \in L^{\phi^*}(\Omega)$ it turns out that $\int_\Omega \big(\phi_{|\nabla w|}\big)^{*\prime}(|h|)\frac{h}{|h|}\,dx = \int_\Omega l\,dx = 0$. Therefore, $c_h = 0$ and we obtain by (2.4)

$$\|l\|^2_{(\nabla w),\mathbb{Q}^*} \ge \langle l, h\rangle - \int_\Omega \phi_{|\nabla w|}(|h|)\,dx = \int_\Omega \phi_{|\nabla w|}(|l|)\,dx.$$

Together with (4.15) this yields

$$\|l\|^2_{(\nabla w),\mathbb{Q}^*} = \int_\Omega \phi_{|\nabla w|}(|l|)\,dx, \tag{4.16}$$

which is exactly what we expect from a reasonable dual quasi-norm on $L_0^\phi(\Omega)$.

The next question is how to choose the shift $|\nabla w|$. Recalling Lemma 74, the quasi-norm is a quantity, which is equivalent to the residual tested with the error. Carrying over these ideas to the functional $\mathcal{F}$ suggests to test the residual $D\mathcal{F}(q)$, $q \in \mathbb{Q}$, with the error $q - p$:

$$\begin{aligned}\langle D\mathcal{F}(q), q-p\rangle &= \langle D\mathcal{F}(q) - D\mathcal{F}(p), q-p\rangle \\ &= \int_\Omega (q-p)\operatorname{div}(u_q - u)\,dx \\ &= \int_\Omega q\,\operatorname{div}(u_q-u) + f\,(u_q-u) \\ &\quad - p\,\operatorname{div}(u_q-u) - f\,(u_q-u)\,dx.\end{aligned}$$

According to (4.10) and Lemma 74 this leads to

$$\begin{aligned}\langle D\mathcal{F}(q) - D\mathcal{F}(p), q-p\rangle &= \int_\Omega \big(\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u)\big) : (\nabla u_q - \nabla u)\,dx \\ &\approx \int_\Omega \phi_{|\nabla u|}(|\nabla u - \nabla u_q|)\,dx. \\ &\approx \int_\Omega \phi_{|\nabla u_q|}(|\nabla u - \nabla u_q|)\,dx.\end{aligned} \tag{4.17}$$

Therefore, the residual of $\mathcal{F}$ is closely connected to the error $u - u_q$ in the quasi-norm with shift $|\nabla u|$ or $|\nabla u_q|$. Since the solution u is not at our disposal we decide for the later one in the following definition of the quasi-steepest descent direction; compare also Remark 143.

Definition 135 (quasi-steepest descent)**.** *Let ϕ be an N-function that satisfies Assumption 40 and assume the notation of this chapter. Then, the quasi-steepest descent direction with respect to $\mathcal{F}$ in $q \in \mathbb{Q}$ is defined as*

$$\mathfrak{d}_q := -\phi'_{|\nabla u_q|}(|\operatorname{div} u_q|)\frac{\operatorname{div} u_q}{|\operatorname{div} u_q|}. \tag{4.18}$$

4.2.2 Convergent Generalized Uzawa Algorithm (GUA)

Now, we are prepared to state the infinite-dimensional quasi-steepest descent algorithm.

Algorithm 136 (GUA)**.** Let $\mu > 0$ and $q_0 \in \mathbb{Q} = L^{\phi^*}(\Omega)/\mathbb{R}$ be an initial guess for the exact solution $p \in \mathbb{Q}$. Let $j = 0$;

1. (DERIVATIVE)

$$u_j \in \mathbb{V}: \quad \int_\Omega \mathbf{A}(\nabla u_j) : \nabla v\,dx = \int_\Omega f \cdot v\,dx + \int_\Omega q_j\,\operatorname{div} v\,dx$$

for all $v \in \mathbb{V}$;

2. (QUASI-STEEPEST DESCENT DIRECTION)

$$\mathfrak{d}_j := -\phi'_{|\nabla u_j|}(|\operatorname{div} u_j|)\frac{\operatorname{div} u_j}{|\operatorname{div} u_j|};$$

3. (UPDATE)

$$q_{j+1} := q_j + \mu\,\mathfrak{d}_j;$$

increment j and go to step (1);

Remark 137. *In step* (DERIVATIVE) *of Algorithm 136 the function $u_j = u_{q_j} \in \mathbb{V}$ is determined. This leads immediately to the derivative $D\mathcal{F}(q_j) = \operatorname{div} u_j$. Hence, in step* (QUASI-STEEPEST DESCENT DIRECTION) *the quasi-steepest descent direction, with respect to $D\mathcal{F}(q_j) = \operatorname{div} u_j$, is determined according to* (4.18). *Finally, in step* (UPDATE)*, the approximation q_j to the solution $p \in \mathbb{Q}$ is updated with the quasi-steepest descent direction scaled by a step-size parameter μ.*

Note, that Algorithm 136 (GUA) is driven by $\operatorname{div} u_j = D\mathcal{F}(q_j)$, $j \in \mathbb{N}$. Hence, the question arises what it means to $(q_j)_{j\in\mathbb{N}} \subset \mathbb{Q}$ if the sequence $(\operatorname{div} u_j) \subset L_0^\phi(\Omega)$ vanishes.

Lemma 138. *Let ϕ be an N-function that satisfy Assumption 40. For a sequence $(q_j)_{j\in\mathbb{N}} \subset \mathbb{Q}$, we define the sequence $(u_j)_{j\in\mathbb{N}} \subset \mathbb{V}$ by $u_j := u_{q_j}$ as in* (4.9)*. Then*

$$\operatorname{div} u_j \to_{j\to\infty} 0 \qquad \text{in } L_0^\phi(\Omega)$$

implies

$$q_j \to_{j\to\infty} p \qquad \text{in } \mathbb{Q},$$

where p is the unique minimizer of $\mathcal{F}$.

Proof. We assume the contrary. In particular, w.l.o.g., there exists a constant $c > 0$ such that $\|p - q_j\|_{\mathbb{Q}} > c$ — otherwise we pass to a subsequence. By the inf-sup condition (4.4) and Corollary 36, there exists a $\tilde{\beta} > 0$ such that

$$\begin{aligned}
\tilde{\beta}\,\|p - q_j\|_{\mathbb{Q}} &\leq \sup_{v\in W_0^{1,\phi}(\Omega)} \frac{\int_\Omega (p - q_j)\,\operatorname{div} v\,dx}{\|\nabla v\|_{(\phi)}}\\
&= \sup_{v\in W_0^{1,\phi}(\Omega)} \frac{\int_\Omega (p - q_j)\,\operatorname{div} v\,dx + \int_\Omega (f - f)\,v\,dx}{\|\nabla v\|_{(\phi)}}\\
&= \sup_{v\in W_0^{1,\phi}(\Omega)} \frac{\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla u_j)\big) : \nabla v\,dx}{\|\nabla v\|_{(\phi)}},
\end{aligned}$$

where we used (4.10) in the last equality. By means of Young's inequality (Proposition 11), it follows for $\delta > 0$

$$\begin{aligned}\tilde{\beta}\, \|p - q_j\|_{\mathbb{Q}} \leq C_\delta \int_\Omega \big(\phi_{|\nabla u|}\big)^*(|\mathbf{A}(\nabla u) - \mathbf{A}(\nabla u_j)|)\, dx \\ + \delta \sup_{v \in W_0^{1,\phi}(\Omega)} \int_\Omega \phi_{|\nabla u|}\Big(\frac{|\nabla v|}{\|\nabla v\|_{(\phi)}}\Big)\, dx,\end{aligned}$$

where the constant C_δ depends on δ and $\Delta_2(\{\phi_a\}_{a\geq 0})$ and thus on $\Delta_2(\{\phi, \phi^*\})$; see Lemma 57. The second term is bounded according to

$$\int_\Omega \phi_{|\nabla u|}\Big(\frac{|\nabla v|}{\|\nabla v\|_{(\phi)}}\Big)\, dx \lesssim \int_\Omega \phi\Big(\frac{|\nabla v|}{\|\nabla v\|_{(\phi)}}\Big) + \phi(|\nabla u|)\, dx \leq 1 + \int_\Omega \phi(|\nabla u|)\, dx;$$

see Corollary 69. Hence, for δ small enough, we have by the assumption $0 < c < \|p - q_j\|_{\mathbb{Q}}$ that

$$\tilde{\beta}\, \|p - q_j\|_{\mathbb{Q}} \lesssim C \int_\Omega \big(\phi_{|\nabla u|}\big)^*(|\mathbf{A}(\nabla u) - \mathbf{A}(\nabla u_j)|)\, dx,$$

For a constant $C > 0$ not depending on $j \in \mathbb{N}$. Furthermore, Corollary 65 and Lemma 74 imply

$$\begin{aligned}\tilde{\beta}\, \|p - q_j\|_{\mathbb{Q}} &\lesssim C \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla u_j)\big) : (\nabla u - \nabla u_j)\, dx \\ &= C \int_\Omega (p - q_j) \operatorname{div}(u - u_j)\, dx = C \int_\Omega (p - q_j) \operatorname{div} u_j\, dx \\ &\leq C\, \|p - q_j\|_{\mathbb{Q}}\, \|\operatorname{div} u_j\|_\phi\,,\end{aligned}$$

where we used (4.10) and the fact that $\operatorname{div} u = 0$; see (4.2). Since $\|\operatorname{div} u_j\|_\phi \to 0$ as $j \to \infty$, this is a contradiction and hence $q_j \to p$ in $\mathbb{Q}$ as $j \to \infty$. □

The next theorem asserts that for some fixed $\mu > 0$ the sequence $(q_j)_{j\in\mathbb{N}} \subset \mathbb{Q}$ produced by Algorithm 136 (GUA) converges to the real solution.

Theorem 139. *Let ϕ be an N-function that satisfies Assumption 116. There exists $\mu_0 > 0$ depending only on $\Delta_2(\{\phi, \phi^*\})$ and d, such that for all step-sizes $\mu \in (0, \mu_0)$, it holds for the sequence $(q_j)_{j\in\mathbb{N}} \subset \mathbb{Q}$ produced by Algorithm 136* (GUA) *that*

$$q_j \to p \quad \text{in } \mathbb{Q}, \text{ as } j \to \infty,$$

where $p \in \mathbb{Q}$ is the solution of (4.13).

Proof. Recall that $\Delta_2(\{\phi_a,(\phi_a)^*\})$ is bounded with respect to $\Delta_2(\{\phi,\phi^*\})$; see Lemma 57. For $q_j \in \mathbb{Q}$ we define an auxiliary function $\mathcal{H}_j : \mathbb{R} \to \mathbb{R}$ by

$$\mathcal{H}_j(\mu) := \mathcal{F}(q_j) - \mathcal{F}(q_j + \mu\,\mathfrak{d}_j).$$

By means of the mean value theorem and Proposition 133, for $\mu > 0$, there exists $\theta \in (0,\mu)$ such that

$$\begin{aligned}\mathcal{H}_j(\mu) &= \mu\,\mathcal{H}_j'(\theta) = -\mu\langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j), \mathfrak{d}_j\rangle \\ &= -\mu\langle D\mathcal{F}(q_j), \mathfrak{d}_j\rangle - \frac{\mu}{\theta}\langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\,\mathfrak{d}_j\rangle.\end{aligned} \tag{4.19}$$

Considering the first term, the definition of $\mathfrak{d}_j$ and (2.6b) imply

$$\begin{aligned}-\mu\langle D\mathcal{F}(q_j), \mathfrak{d}_j\rangle &= \mu \int_\Omega \phi'_{|\nabla u_j|}(|\operatorname{div} u_j|)\,|\operatorname{div} u_j|\; dx \\ &\geq \mu \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx.\end{aligned} \tag{4.20}$$

For the second term holds

$$\langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\,\mathfrak{d}_j\rangle = \int_\Omega \theta\,\mathfrak{d}_j\; \operatorname{div}(u_{q_j+\theta\mathfrak{d}_j} - u_j)\,dx,$$

where $u_{q_j+\theta\,\mathfrak{d}_j}$ is defined as in (4.10). For convenience we shall denote $u_\theta := u_{q_j+\theta\mathfrak{d}_j}$ in the sequel. Applying Young's inequality (Proposition 11) it follows for $\delta > 0$

$$\begin{aligned}&\langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\,\mathfrak{d}_j\rangle \\ &\qquad \leq \int_\Omega \delta\,\phi_{|\nabla u_j|}(|\operatorname{div}(u_\theta - u_j)|) + C_\delta\,\big(\phi_{|\nabla u_j|}\big)^*(|\theta\,\mathfrak{d}_j|)\,dx,\end{aligned}$$

where the constant C_δ solely depends on $\Delta_2(\{\phi,\phi^*\})$ and δ. By Lemma 128, then

$$\begin{aligned}&\langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\,\mathfrak{d}_j\rangle \\ &\qquad \preccurlyeq \int_\Omega \delta\,\phi_{|\nabla u_j|}(|\nabla(u_\theta - u_j)|) + C_\delta\,\big(\phi_{|\nabla u_j|}\big)^*(|\theta\,\mathfrak{d}_j|)\,dx,\end{aligned} \tag{4.21}$$

where the constant hidden in $\preccurlyeq$ depends only on $\Delta_2(\phi)$ and d. On the other hand we get, as in (4.17), with $(q_j + \theta\,\mathfrak{d}_j) - q_j = \theta\,\mathfrak{d}_j$ that

$$\begin{aligned}&\langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\,\mathfrak{d}_j\rangle \\ &\qquad = \int_\Omega \big(\mathbf{A}(\nabla u_\theta) - \mathbf{A}(\nabla u_j)\big) : \nabla(u_\theta - u_j)\,dx \\ &\qquad \approx \int_\Omega \phi_{|\nabla u_j|}(|\nabla(u_\theta - u_j)|)\,dx.\end{aligned} \tag{4.22}$$

Therefore, choosing $\delta > 0$ small enough in (4.21) yields

$$\langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\,\mathfrak{d}_j\rangle \preccurlyeq \int_\Omega \big(\phi_{|\nabla u_j|}\big)^*(|\theta\,\mathfrak{d}_j|)\,dx, \tag{4.23}$$

where the constant hidden in $\preccurlyeq$ depends only on $\Delta_2(\{\phi,\phi^*\})$. We continue to estimate the right hand side of (4.23). Lemma 60 implies

$$\big(\phi_{|\nabla u_j|}\big)^*(|\theta\,\mathfrak{d}_j|) \approx \phi^*_{\phi'(|\nabla u_j|)}(|\theta\,\mathfrak{d}_j|).$$

We may assume that $\mu < \mu_0 \leq 2$. Hence, Lemma 128, the definition of shifted N-functions (Definition 56), and Corollary 17 yield

$$2\,|\mathfrak{d}_j| = 2\,\phi'_{|\nabla u_j|}(|\operatorname{div} u_j|) \preccurlyeq 2\,\phi'_{|\nabla u_j|}(|\nabla u_j|) = 2\,\frac{\phi'(2\,|\nabla u_j|)}{2\,|\nabla u_j|}\,|\nabla u_j| \preccurlyeq \phi'(|\nabla u_j|),$$

where the constant hidden in $\preccurlyeq$ depends on $\Delta_2(\{\phi,\phi^*\})$ and d. Therefore, we can apply Lemma 59 with $\alpha = \frac{\theta}{2} \leq 1$ to obtain

$$\big(\phi_{|\nabla u_j|}\big)^*(|\theta\,\mathfrak{d}_j|) \approx \phi^*_{\phi'(|\nabla u_j|)}\Big(\frac{\theta}{2}\,2\,|\mathfrak{d}_j|\Big) \preccurlyeq \theta^2\phi^*_{\phi'(|\nabla u_j|)}(2\,|\mathfrak{d}_j|) \approx \theta^2\big(\phi_{|\nabla u_j|}\big)^*(|\mathfrak{d}_j|).$$

Note that the hidden constants of the last display solely depend on $\Delta_2(\{\phi,\phi^*\})$ and d. Recalling the definition of $\mathfrak{d}$ we get from (2.8) that

$$\big(\phi_{|\nabla u_j|}\big)^*(|\theta\,\mathfrak{d}_j|) \preccurlyeq \theta^2\phi_{|\nabla u_j|}(|\operatorname{div} u_j|). \tag{4.24}$$

Applying this to (4.23) yields

$$\begin{aligned}\langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\,\mathfrak{d}_j\rangle &\leq \tilde{C}\,\theta^2 \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx \\ &\leq \tilde{C}\,\mu^2 \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx\end{aligned} \tag{4.25}$$

with constant $\tilde{C} > 0$ depending only on $\Delta_2(\{\phi,\phi^*\})$ and d. Inserting this, together with (4.20) into (4.19), implies the estimate

$$\mathcal{H}_j(\mu) = \mu\,\mathcal{H}'_j(\theta) \geq \mu\,(1 - \tilde{C}\,\mu) \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx. \tag{4.26}$$

We can now choose $0 < \mu_0 \leq 2$ such that $\mu\,(1 - \tilde{C}\mu) > 0$ for all $\mu \in (0,\mu_0)$. For fixed $\mu \in (0,\mu_0)$ this implies that $\operatorname{div} u_j \to 0$ in $L^\phi_0(\Omega)$ as $j \to \infty$: In fact, observing that $q_j + \mu\,\mathfrak{d}_j = q_{j+1}$ and summing over j yield for any $J \in \mathbb{N}$

$$\begin{aligned}\mathcal{F}(q_0) - \mathcal{F}(q_j) &= \sum_{j=0}^{J-1} \mathcal{F}(q_j) - \mathcal{F}(q_{j+1}) \\ &\geq \mu\,(1 - \tilde{C}\,\mu) \sum_{j=0}^{J-1} \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx.\end{aligned}$$

Recalling Corollary 126, the left hand side can be estimated by $\mathcal{F}(q_0) - \mathcal{F}(p)$ and thus is independent of J. Hence, the series

$$\sum_{j=0}^{J-1} \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx \le \frac{1}{\mu\,(1-\tilde{C}\,\mu)} \big(\mathcal{F}(q_0) - \mathcal{F}(p)\big)$$

is bounded for all $J \in \mathbb{N}$, which implies

$$\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx \to 0,$$

as $j \to \infty$. Due to (4.26) and the choice of μ the sequence $(\mathcal{F}(q_j))_{j\in\mathbb{N}}$ is bounded. Combining (4.9) with (4.10) yields

$$\begin{aligned}
\mathcal{F}(q_0) \ge \mathcal{F}(q_j) = -\mathcal{L}(u_j, q_j) &= \int_\Omega -\phi(|\nabla u_j|) + q_j \operatorname{div} u_j + f u_j\, dx \\
&= \int_\Omega -\phi(|\nabla u_j|) + \mathbf{A}(\nabla u_j) : \nabla u_j\, dx \\
&= \int_\Omega -\phi(|\nabla u_j|) + \phi'(|\nabla u_j|)\, |\nabla u_j|\ dx \\
&\ge (\nabla(\phi) - 1) \int_\Omega \phi(|\nabla u_j|)\, dx \ge 0,
\end{aligned}$$

where the constant $\nabla(\phi) > 1$ depends only on $\Delta_2(\phi^*)$; see Proposition 14 ii). Therefore, the sequence $(\int_\Omega \phi(|\nabla u_j|)\, dx)_{j\in\mathbb{N}} \subset \mathbb{R}$ is bounded. Assume that $(\operatorname{div} u_j)_{j\in\mathbb{N}}$ does not converge to zero in $\mathbb{Q}$. Then, Proposition 31 implies w.l.o.g. that

$$0 < c < \int_\Omega \phi(|\operatorname{div} u_j|)\, dx \quad \text{for all } j \in \mathbb{N},$$

for a constant $c > 0$ — otherwise we pass to a subsequence. Hence, we get by Corollary 69 for $\delta > 0$

$$c < \int_\Omega \phi(|\operatorname{div} u_j|)\, dx \preccurlyeq (1 + C_\delta) \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx + \delta \int_\Omega \phi(|\nabla u_j|)\, dx$$

for all $j \in \mathbb{N}$. Since $(\int_\Omega \phi(|\nabla u_j|)\, dx)_{j\in\mathbb{N}}$ is bounded, we can choose $\delta > 0$ small enough to obtain

$$0 < c \preccurlyeq C \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx,$$

with a constant $C > 0$ not depending on $j \in \mathbb{N}$. This is a contradiction, since $\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx \to 0$, as $j \to \infty$. Thus, $\operatorname{div} u_j \to 0$ in $\mathbb{Q}$ as $j \to \infty$ and the assertion follows with Lemma 138. □

Corollary 140. *Suppose the assumptions of Theorem 139. Then for $\mu \in (0, \mu_0)$ there exists constants $C, c > 0$, such that for the reduction of $\mathcal{F}$*

$$c \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx \leq \mathcal{F}(q_j) - \mathcal{F}(q_{j+1}) \leq C \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx.$$

The constants C, c depend solely on $\Delta_2(\{\phi, \phi^*\})$, d, and the step-size μ.

Proof. The left inequality is proven by (4.26). For the right inequality we recall the prove of Theorem 139. In particular, we estimate the first term of the right hand side of (4.19) by the definition of $\mathfrak{d}_j$ and Corollary 15

$$\begin{aligned} -\mu \langle D\mathcal{F}(q_j), \mathfrak{d}_j \rangle &= \mu \int_\Omega \phi'_{|\nabla u_j|}(|\operatorname{div} u_j|)\, |\operatorname{div} u_j|\ dx \\ &\approx \mu \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx. \end{aligned}$$

Moreover, from (4.22) it holds for the second term of the right hand side of (4.19)

$$\frac{\mu}{\theta} \langle D\mathcal{F}(q_j + \theta\, \mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\, \mathfrak{d}_j \rangle > 0$$

Hence, neglecting this term in (4.19) yields

$$\mathcal{F}(q_j) - \mathcal{F}(q_{j+1}) \preccurlyeq \mu \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx$$

and the assertion is proved. □

Remark 141 (linear case $(r = 2)$)**.** *In the linear case, i.e., when $\phi(t) = \frac{1}{2} t^2$, the above algorithm corresponds to the Uzawa method, which is known to converge for appropriate values of the parameter μ; see, e.g., [15, 73, 6, 64]. In particular, it converges linearly for $\mu \in (0, 2)$ and the contraction factor seems to be optimal for $\mu = 1$ [64, 65]. We shall show that the convergence proof of Theorem 139 leads to the same result in the linear case.*

We use the notation of the proof of Theorem 139. Observe that in the linear case $\mathcal{F}(q) = \int_\Omega -\frac{1}{2} |\nabla u_q|^2 + f \cdot u_q + q \operatorname{div} u_q\, dx = \frac{1}{2} |\nabla u_q|_{W^{1,2}(\Omega)}$ and $\mathfrak{d}_j = \operatorname{div} u_j$. Moreover, we obtain by straight forward calculations that

$$\mathcal{F}(q_j) - \mathcal{F}(q_j + \mu \mathfrak{d}_j) = \mathcal{H}_j(\mu) = \mu\, \mathcal{H}'_j\big(\frac{\mu}{2}\big),$$

i.e., the mean value Theorem holds with $\theta = \frac{\mu}{2}$. As in the proof of Theorem 139 we get

$$\mathcal{H}_j(\mu) = \mu \mathcal{H}'_j\big(\frac{\mu}{2}\big) = -\mu \langle D\mathcal{F}(q_j), \mathfrak{d}_j \rangle - \frac{\mu}{\theta} \langle D\mathcal{F}(q_j + \theta\, \mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\, \mathfrak{d}_j \rangle. \tag{4.27}$$

Noting that $\|\operatorname{div} v\|_{L^2(\Omega)} \leq \|\nabla v\|_{L^2(\Omega)}$ *for* $v \in W_0^{1,2}(\Omega)$ *(see [64]), we get for the second term*

$$\begin{aligned}\|\nabla u_\theta - \nabla u_j\|^2_{L^2(\Omega)} &= \langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\,\mathfrak{d}_j\rangle \\ &= \int_\Omega \theta\,\mathfrak{d}_j \operatorname{div}(u_\theta - u_j)\,dx \\ &\leq \|\operatorname{div}(u_\theta - u_j)\|_{L^2(\Omega)} \|\theta\,\mathfrak{d}_j\|_{L^2(\Omega)} \\ &\leq \|\nabla u_\theta - \nabla u_j\|_{L^2(\Omega)} \|\theta\,\mathfrak{d}_j\|_{L^2(\Omega)}\,.\end{aligned}$$

Therefore, with $\mathfrak{d}_j = -\operatorname{div} u_j$

$$\|\nabla u_\theta - \nabla u_j\|^2_{L^2(\Omega)} = \langle D\mathcal{F}(q_j + \theta\,\mathfrak{d}_j) - D\mathcal{F}(q_j), \theta\,\mathfrak{d}_j\rangle \leq \|\theta \operatorname{div} u_j\|^2_{L^2(\Omega)}\,.$$

Thus, inserting this in (4.27) *we get with* $-\langle D\mathcal{F}(q_j), \mathfrak{d}_j\rangle = \|\operatorname{div} u_j\|^2_{L^2(\Omega)}$ *and* $\theta = \frac{\mu}{2}$

$$\begin{aligned}\mathcal{F}(q_j) - \mathcal{F}(q_j + \mu\,\mathfrak{d}_j) = \mathcal{H}_j(\mu) &\geq (\mu - \mu\,\theta)\,\|\operatorname{div} u_j\|^2_{L^2(\Omega)} \\ &= \mu\big(1 - \frac{\mu}{2}\big)\,\|\operatorname{div} u_j\|^2_{L^2(\Omega)}\,.\end{aligned} \tag{4.28}$$

Moreover, we observe by the inf-sup *condition,* $\operatorname{div} u = 0$, *and* (4.10) *that*

$$\begin{aligned}\|\operatorname{div} u_j\|^2_{L^2(\Omega)} &= \|\operatorname{div}(u_j - u)\|^2_{L^2(\Omega)} \\ &\geq \beta^2\,\|\nabla(u_j - u)\|^2_{L^2(\Omega)} \\ &= \beta^2 \int_\Omega (\nabla(u_j - u)) : (\nabla(u_j - u))\,dx \\ &= \beta^2 \int_\Omega (p - q_j)(\operatorname{div}(u_j - u))\,dx \\ &= \beta^2 \int_\Omega (p - q_j)(\operatorname{div}(u_j + u))\,dx \\ &= \beta^2 \int_\Omega (\nabla(u_j - u)) : (\nabla(u_j + u))\,dx \\ &= \beta^2\,(\|\nabla u_j\|^2_{L^2(\Omega)} - \|\nabla u\|^2_{L^2(\Omega)}) = 2\,\beta^2\,(\mathcal{F}(q_j) - \mathcal{F}(p)),\end{aligned} \tag{4.29}$$

as $\mathcal{F}(q) = \frac{1}{2}\,\|\nabla u_q\|^2_{L^2(\Omega)}$ *for* $q \in \mathbb{Q}$. *Altogether, we have with* $q_{j+1} = q_j + \mu\,\mathfrak{d}_j$ *for* $\mu \in (0, 2)$

$$\begin{aligned}\mathcal{F}(q_{j+1}) - \mathcal{F}(p) &= \mathcal{F}(q_j) - \mathcal{F}(p) - \big(\mathcal{F}(q_j) - \mathcal{F}(q_{j+1})\big) \\ &\leq \mathcal{F}(q_j) - \mathcal{F}(p) - \mu\big(1 - \frac{\mu}{2}\big)\,\|\operatorname{div} u_j\|^2_{L^2(\Omega)} \\ &\leq \mathcal{F}(q_j) - \mathcal{F}(p) - \mu\big(1 - \frac{\mu}{2}\big)\,2\,\beta^2\big(\mathcal{F}(q_j) - \mathcal{F}(p)\big) \\ &= \big(1 - \mu(2 - \mu)\,\beta^2\big)\big(\mathcal{F}(q_j) - \mathcal{F}(p)\big).\end{aligned}$$

Furthermore, we can deduce from (4.29) *that* $\|\nabla(u_j - u)\|^2_{L^2(\Omega)} = 2\,(\mathcal{F}(q_j) - \mathcal{F}(p))$ *and hence,*

$$\|\nabla(u_{j+1} - u)\|^2_{L^2(\Omega)} \leq \big(1 - \mu(2-\mu)\,\beta^2\big)\,\|\nabla(u_j - u)\|^2_{L^2(\Omega)}\,. \tag{4.30}$$

As $\beta < 1$ *(see [64]), this yields a contraction for* $\mu \in (0,2)$. *The contraction factor becomes minimal for* $\mu = 1$ *and is the same factor obtained in [64] for this case.*

Remark 142 (contraction)**.** *We observed in Remark 141 that, for some step-size* μ, *the Uzawa algorithm is a contraction for the linear case; see* (4.30) *and [64, 65]. Therefore, the question arises, if Algorithm 136* (GUA) *is also a contraction in the nonlinear case.*

We assume the conditions of Theorem 139. Recall that

$$\mathcal{F}(q) = -\mathcal{L}(u_q, q) = \sup_{v \in \mathbb{V}} -\mathcal{L}(v,q) \quad \textit{for } q \in \mathbb{Q}$$

and

$$\mathcal{F}(p) = \inf_{q \in \mathbb{Q}} \sup_{v \in \mathbb{V}} -\mathcal{L}(v,q),$$

i.e., u_q *is the minimizer of the functional* $\mathcal{J}_q(\cdot) := \mathcal{L}(\cdot, q)$.

By Corollary 140, there exists a $c > 0$ *solely depending on* $\Delta_2(\{\phi, \phi^*\})$, d, *and* μ, *such that*

$$\begin{aligned}\mathcal{F}(q_{j+1}) - \mathcal{F}(p) &= \mathcal{F}(q_j) - \mathcal{F}(p) - \big(\mathcal{F}(q_j) - \mathcal{F}(q_{j+1})\big)\\ &\leq \mathcal{F}(q_j) - \mathcal{F}(p) - c\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx.\end{aligned}$$

Thanks to Corollary 140, this estimate is optimal up to a constant. Hence, a fixed contraction for differences of the functional $\mathcal{F}$ *in each step is equivalent to*

$$\mathcal{F}(q_j) - \mathcal{F}(p) \preccurlyeq \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx. \tag{4.31}$$

Therefore, we shall analyze the term $\mathcal{F}(q_j) - \mathcal{F}(p)$. *On the one hand we obtain with* (4.10) *and Proposition 100*

$$\begin{aligned}\langle D\mathcal{F}(q) - D\mathcal{F}(p), q - p\rangle &= \int_\Omega \big(\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u)\big) : (\nabla u_q - \nabla u)\,dx\\ &\approx \int_\Omega \phi_{|\nabla u|}(|\nabla u_q - \nabla u|)\,dx\\ &\approx \mathcal{J}_q(u) - \mathcal{J}_q(u_q) = \mathcal{L}(u,q) - \mathcal{L}(u_q,q)\\ &\leq \mathcal{L}(u,p) - \mathcal{L}(u_q,q) = \mathcal{F}(q) - \mathcal{F}(p).\end{aligned} \tag{4.32}$$

Note that the involved constants solely depend on $\Delta_2(\{\phi,\phi^*\})$*, but not on* q*. On the other hand, the mean value theorem for some* $\theta \in (0,1)$ *implies*

$$\begin{aligned}\mathcal{F}(q)-\mathcal{F}(p) &= \langle D\mathcal{F}(p+\theta\,(q-p)), q-p\rangle\\ &= \langle D\mathcal{F}(q), q-p\rangle + \langle D\mathcal{F}(p+\theta(q-p)) - D\mathcal{F}(q), q-p\rangle\\ &= \langle D\mathcal{F}(q)-D\mathcal{F}(p), q-p\rangle + \langle D\mathcal{F}(p+\theta(q-p)) - D\mathcal{F}(q), q-p\rangle,\end{aligned}$$

where we use that $D\mathcal{F}(p) = \operatorname{div} u = 0$*; see Proposition 133. By the monotonicity of* $D\mathcal{F}$ *(Corollary 134) we have for the last term*

$$\begin{aligned}&\langle D\mathcal{F}(p+\theta\,(q-p)) - D\mathcal{F}(q), q-p\rangle\\ &\qquad = \frac{1}{\theta-1}\,\langle D\mathcal{F}(q+(\theta-1)\,(q-p)) - D\mathcal{F}(q), (\theta-1)\,(q-p)\rangle \le 0.\end{aligned}$$

Hence,

$$\mathcal{F}(q)-\mathcal{F}(p) \le \langle D\mathcal{F}(q)-D\mathcal{F}(p), q-p\rangle.$$

Thus, with (4.32)*, it holds for all* $q \in \mathbb{Q}$

$$\mathcal{F}(q)-\mathcal{F}(p) \eqsim \langle D\mathcal{F}(q)-D\mathcal{F}(p), q-p\rangle \eqsim \|\mathbf{F}(\nabla u_q)-\mathbf{F}(\nabla u)\|^2_{L^2(\Omega)}\,; \tag{4.33}$$

see also Lemma 74 Hence, by (4.17) *it follows that* (4.31) *is equivalent to*

$$\int_\Omega \phi_{|\nabla u_j|}(|\nabla u - \nabla u_j|)\,dx \preccurlyeq \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx.$$

The contraction should not depend on the specific sequence $(q_j)_{j\in\mathbb{N}}$*, which strongly depends on the initial guess* q_0 *and the step-size* μ*. Hence, the above observations lead to the question if it holds*

$$\int_\Omega \phi_{|\nabla u_q|}(|\nabla u - \nabla u_q|)\,dx \preccurlyeq \int_\Omega \phi_{|\nabla u_q|}(|\operatorname{div} u_q|)\,dx \tag{4.34}$$

for all $q \in \mathbb{Q}$*. In the linear case, the analog estimate is a consequence of the* inf-sup *condition; see* (4.29)*. Since we are dealing with quasi-norms, we have to look for an analog of the norm-*inf-sup *condition for quasi-norms; see* (4.4)*. For one possible generalization assume that there exists* $\beta > 0$ *such that for all* $q \in \mathbb{Q}$

$$\begin{aligned}\|\nabla(q-p)\|^2_{(\nabla u),*} &:= \sup_{v\in\mathbb{V}}\Big\{\int_\Omega (q-p)\operatorname{div} v\,dx - \int_\Omega \phi_{|\nabla u|}(|\nabla v|)\,dx\Big\}\\ &\ge \beta\,\inf_{c\in\mathbb{R}} \int_\Omega \big(\phi_{|\nabla u|}\big)^*(|q-p-c|)\,dx;\end{aligned} \tag{4.35}$$

compare also (4.14)*,* (4.3)*,* (4.4)*, and Corollary 130. Note, that this estimate is very meaningful according to the question whether we have an adequate error* *concept or not; see Remark 143. We want to show, that* (4.35) *implies* (4.34). *By* (4.10)*, Young's inequality* (2.3)*, Corollary 65, Lemma 74, and Proposition 11 it holds for all* $q \in \mathbb{Q}$*,* $v \in \mathbb{V}$*,* $\hat{c} \in \mathbb{R}$*, and* $\delta > 0$

$$\begin{aligned}
&\int_\Omega (q-p)\,\operatorname{div} v\,dx - \int_\Omega \phi_{|\nabla u|}(|\nabla v|)\,dx \\
&\qquad = \int_\Omega \big(\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u)\big) : \nabla v\,dx - \int_\Omega \phi_{|\nabla u|}(|\nabla v|)\,dx \\
&\qquad \le \int_\Omega \big(\phi_{|\nabla u|}\big)^*(|\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u)|)\,dx \\
&\qquad \eqsim \int_\Omega \big(\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u)\big) : \nabla(u_q - u)\,dx \\
&\qquad = \int_\Omega (q-p)\,\operatorname{div} u_q\,dx = \int_\Omega (q-p-\hat{c})\,\operatorname{div} u_q\,dx \\
&\qquad \le \delta \inf_{c\in\mathbb{R}} \int_\Omega \big(\phi_{|\nabla u|}\big)^*(|q-p-c|)\,dx + C_\delta \int_\Omega \phi_{|\nabla u|}(|\operatorname{div} u_q|)\,dx.
\end{aligned} \tag{4.36}$$

Taking the supremum over all $v \in \mathbb{V}$*,* (4.35) *implies for* $\delta > 0$ *small enough*

$$\int_\Omega \big(\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u)\big) : \nabla(u_q - u)\,dx \preccurlyeq \int_\Omega \phi_{|\nabla u|}(|\operatorname{div} u_q|)\,dx.$$

Now, the shift can be changed to $|\nabla u_q|$ *with Corollary 69. Hence,*

$$\begin{aligned}
&\int_\Omega \big(\mathbf{A}(\nabla u_q) - \mathbf{A}(\nabla u)\big) : \nabla(u_q - u)\,dx \\
&\qquad\qquad \preccurlyeq (1 + C_\delta) \int_\Omega \phi_{|\nabla u_q|}(|\operatorname{div} u_q|)\,dx + \delta \int_\Omega \phi_{|\nabla u_q|}(|\nabla(u - u_q)|)\,dx.
\end{aligned}$$

Choosing $\delta > 0$ *small enough, the last term can be hidden in the left hand side; compare also Lemma 74. Therefore,* (4.35) *implies* (4.34) *and hence contraction of* (GUA).

Remark 143 (concept of distance)**.** *In Remark 79 we proposed that it is important to use error concepts for which the dual error and the primal error are balanced with respect to the problem. In this chapter we implicitly introduced a concept of distance for the nonlinear Stokes problem. In particular, by* (4.16) *and the later choice of the shift, on* $L^{\phi^*}(\Omega)/\mathbb{R}$ *a measure of distance is defined by*

$$\|q - p\|^2_{(\nabla u_q),\mathbb{Q}} = \inf_{c\in\mathbb{R}} \int_\Omega \big(\phi_{|\nabla u_q|}\big)^*(|q-p-c|)\,dx;$$

see Corollary 130. The dual measure of distance on $L^\phi_0(\Omega)$ *for the residual of*

$q \in L^{\phi^*}(\Omega)/\mathbb{R}$ reads as

$$\|D\mathcal{F}(q)\|^2_{(\nabla u_q),\mathbb{Q}^*} := \sup_{\hat{q}\in\mathbb{Q}} \Big\{ \langle D\mathcal{F}(q), \hat{q}\rangle - \inf_{c\in\mathbb{R}} \int_\Omega \big(\phi_{|\nabla u_q|}\big)^*(|\hat{q}-c|)\,dx \Big\}$$
$$= \int_\Omega \phi_{|\nabla u_q|}(|\operatorname{div} u_q|)\,dx;$$

cf. (4.14) *and* (4.16). *Now, the question arises if these two quantities are balanced; see Remark 79. In fact, by* $\operatorname{div} u = 0$, *Lemma 128, and Lemma 74 it holds*

$$\int_\Omega \phi_{|\nabla u_q|}(|\operatorname{div} u_q|)\,dx = \int_\Omega \phi_{|\nabla u_q|}(|\operatorname{div} u_q - \operatorname{div} u|)\,dx$$
$$\preccurlyeq \int_\Omega \phi_{|\nabla u_q|}(|\nabla u_q - \nabla u|)\,dx$$
$$\eqsim \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla u_q)\big) : (\nabla u - \nabla u_q)\,dx.$$

Now, recalling (4.2) *and* (4.10) *we get by means of Young's inequality (Proposition 11) for all* $\delta > 0$ *and* $c \in \mathbb{R}$

$$\int_\Omega \phi_{|\nabla u_q|}(|\operatorname{div} u_q|)\,dx \preccurlyeq \int_\Omega (p-q)\ \operatorname{div}(u-u_q)\,dx + \int_\Omega (f-f)(u-u_q)\,dx$$
$$= \int_\Omega (p-q-c)\ \operatorname{div}(u-u_q)\,dx$$
$$\leq \int_\Omega C_\delta\,\big(\phi_{|\nabla u_q|}\big)^*(|q-p-c|) + \delta\,\phi_{|\nabla u_q|}(|\operatorname{div}(u-u_q)|)\,dx.$$

Recalling once again $\operatorname{div} u = 0$, *we get for* $\delta > 0$ *small enough*

$$\|D\mathcal{F}(q)\|^2_{(\nabla u_q),\mathbb{Q}^*} \preccurlyeq \|q-p\|^2_{(\nabla u_q),\mathbb{Q}}\,,$$

where we took the infimum over all $c \in \mathbb{R}$.

We want to prove that the converse estimate is equivalent to the suggested quasi-norm inf-sup-*condition* (4.35) *of Remark 142. Recalling* (4.36) *we observe that choosing* δ *small enough,* (4.35) *implies*

$$\|D\mathcal{F}(q)\|^2_{(\nabla u_q),\mathbb{Q}^*} \succcurlyeq \|q-p\|^2_{(\nabla u_q),\mathbb{Q}}\,, \tag{4.37}$$

where we additionally used Corollaries 69 and 71 to change the shift from $|\nabla u|$ *to* $|\nabla u_q|$. *On the other hand assuming* (4.37), *Lemma 128, Lemma 74, and* $\operatorname{div} u = 0$ *yield*

$$\|q-p\|^2_{(\nabla u_q),\mathbb{Q}} \preccurlyeq \|D\mathcal{F}(q)\|^2_{(\nabla u_q),\mathbb{Q}^*} = \int_\Omega \phi_{|\nabla u_q|}(|\operatorname{div} u_q|)\,dx$$
$$\preccurlyeq \int_\Omega \phi_{|\nabla u_q|}(|\nabla u_q - \nabla u|)\,dx$$
$$\eqsim \sup_{v\in\mathbb{V}} \Big\{ \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla u_q)\big) : \nabla v\,dx - \int_\Omega \phi_{|\nabla u|}(|\nabla v|)\,dx \Big\};$$

see Remark 79 for the last estimate. Hence, an application of Corollary 71, (4.2), *and* (4.10) *yields*

$$\|q-p\|^2_{(\nabla u),\mathbb{Q}} \preccurlyeq \sup_{v\in\mathbb{V}} \Big\{ \int_\Omega (p-q)\operatorname{div} v\,dx - \int_\Omega \phi_{|\nabla u|}(|\nabla v|)\,dx \Big\},$$

which is (4.35).

Therefore we proved that the error concept is balanced if and only if the quasi inf-sup *condition* (4.35) *holds. Moreover, if the error concept is balanced then Algorithm 136* (GUA) *yields linear convergence; see Remark 142.*

4.3 Adaptive Uzawa Finite Element Method

As in [6] for the linear case, we shall now bring together algorithms 136 (GUA) and 99 (AFEM) to formulate an adaptive Uzawa algorithm (AUA). Recall that in the GUA, in each iteration $j \in \mathbb{N}$, the quasi-steepest descent direction $\mathfrak{d}_j$ is used for the update. To determine $\mathfrak{d}_j$, a nonlinear Poisson equation has to be solved. Now, the idea is to use Algorithm 99 to approximate the quasi-steepest descent direction.

In Section 4.3.1 an adaptive finite element method based on Algorithm 99 (AFEM) is presented to calculate an approximation to the quasi-descent direction. Section 4.3.2 collects some technical results on interpolation of discrete functions that are needed to prove convergence of the adaptive algorithm in section 4.3.3. Some possible modifications of the algorithm like, e.g., an a posteriori error estimator of [11] are discussed in the proximate remarks.

4.3.1 Approximation of the Quasi-Steepest Descent Direction

As we know from Section 4.2, we have to solve a nonlinear elliptic system (4.10) for the quasi-steepest direction. Recalling Theorem 106, Algorithm 99 yields linear convergence for a right hand side $g \in L^{\phi^*}(\Omega)^d$. Therefore, due to the right hand side of (4.10) it is convenient that the gradient of the pressure is in $L^{\phi}(\Omega)^d$. In particular, for $\mathcal{T}$ being a conforming triangulation of Ω, we define

$$\mathbb{Q}(\mathcal{T}) := \{Q \in C(\Omega) : Q|_T \in \mathcal{P}^1(T) \quad \text{for all } T \in \mathcal{T}\}.$$

Since $\mathbb{Q}(\mathcal{T}) \subset W^{1,\infty}(\Omega)$, we obviously have $\mathbb{Q}(\mathcal{T}) \subset W^{1,\phi^*}(\Omega)$ for all N-functions ϕ; see Definition 33. Hence $\nabla Q \in L^{\phi^*}(\Omega)^d$ for $Q \in \mathbb{Q}(\mathcal{T})$. Note that $\mathbb{Q}(\mathcal{T})$ is not a subspace of $\mathbb{Q}$, but $\mathbb{Q}(\mathcal{T})/\mathbb{R} \subset \mathbb{Q}$. For convenience, we use the functions in $\mathbb{Q}(\mathcal{T})$ as representants of those in $\mathbb{Q}(\mathcal{T})/\mathbb{R}$ and say that two of them are equal if they differ by a constant value.

Let ϕ be an N-function that satisfies Assumption 40. Then, for $Q \in \mathbb{Q}(\mathcal{T})$ let $u_Q \in \mathbb{V}$ be defined according to (4.9). Since $Q \in \mathbb{Q}(\mathcal{T}) \subset W_0^{1,\phi^*}(\Omega)$, we have

$f - \nabla Q \in L^{\phi^*}(\Omega)^d$. Hence, we can reformulate the nonlinear system (4.10) — using integration by parts — into

$$\int_\Omega \mathbf{A}(\nabla u_Q) : \nabla v \, dx = \int_\Omega (f - \nabla Q) \cdot v \, dx \qquad \text{for all } v \in \mathbb{V}. \tag{4.38}$$

According to Definition 135 the quasi-steepest descent direction of $\mathcal{F}$ in Q is given by

$$\mathfrak{d}_Q = \phi'_{|\nabla u_Q|}(|\operatorname{div} u_Q|) \frac{\operatorname{div} u_Q}{|\operatorname{div} u_Q|}. \tag{4.39}$$

Now, the aim is to calculate an approximation $\mathfrak{D}_Q$ of $\mathfrak{d}_Q$. For this purpose, we modify Algorithm 99 (AFEM) to obtain a method

$$(U_Q, \mathcal{T}^*) = \mathsf{ELLIPT}(Q, \mathcal{T}, \epsilon, \theta)$$

that, given a conforming triangulation $\mathcal{T}$ of Ω, $\epsilon > 0$, $\theta \in (0,1)$, and $Q \in \mathbb{Q}(\mathcal{T})$, outputs an approximation U_Q of u_Q and a refinement $\mathcal{T}^*$ of $\mathcal{T}$. Since the method is based on Algorithm 99 (AFEM), for its precise formulation, we assume that we have the subroutines of Algorithm 99 (AFEM) at hand; see Section 3.5.1.

Algorithm 144 ($\mathsf{ELLIPT}(Q, \mathcal{T}, \epsilon, \theta)$)**.** Let $k = 0$, $\mathcal{T}_0 = \mathcal{T}$;

1. $U_k = \mathsf{SOLVE}(\mathcal{T}_k, f - \nabla Q)$;
2. $\{\eta(U_k, T, f - \nabla Q)\}_{T \in \mathcal{T}_k} = \mathsf{ESTIMATE}(U_k, \mathcal{T}_k, f - \nabla Q)$;
3. if $\eta(U_k, \mathcal{T}_k, f - \nabla Q) < \epsilon$, then
$$U_Q := U_k; \qquad \mathcal{T}^* := \mathcal{T}_k; \qquad \text{RETURN};$$
4. $\mathcal{M}_k = \mathsf{MARK}(\{\eta(U_k, T, f - \nabla Q)\}_{T \in \mathcal{T}_k}, \mathcal{T}_k, \theta)$;
5. $\mathcal{T}_{k+1} = \mathsf{REFINE}(\mathcal{T}_k, \mathcal{M}_k, b)$; increment k and go to step (1);

An approximation to the quasi-steepest descent direction in Q, based on $(U_Q, \mathcal{T}^*) = \mathsf{ELLIPT}(Q, \mathcal{T}, \epsilon, \theta)$, is then given by

$$\phi'_{|\nabla U_Q|}(|\operatorname{div} U_Q|) \frac{\operatorname{div} U_Q}{|\operatorname{div} U_Q|}. \tag{4.40}$$

Remark 145. *Note that the method* ELLIPT *is a modification of Algorithm 99* (AFEM) *for the right hand side* $g = f - \nabla Q \in L^{\phi^*}(\Omega)$ *in* (3.2)*. The only difference is step* (3)*, where a stopping criterion is added. Hence,* ELLIPT *terminates for any* $(Q, \mathcal{T}, \epsilon, \theta) \in W^{1,\phi^*}(\Omega) \times \mathbb{T} \times (0, \infty) \times (0,1)$*, since Corollary 108 states linear convergence of the estimator.*

In the adaptive algorithm the quasi-steepest descent direction in Q will be substituted by the approximation (4.40). To control the resulting error, we need the following Lemma that estimates the distance between descent directions.

Lemma 146. *Let ϕ be an N-function that satisfies Assumption 40. For $v, w \in \mathbb{V}$ we set*

$$\mathfrak{d}(v) := \phi'_{|\nabla v|}(|\operatorname{div} v|)\frac{\operatorname{div} v}{|\operatorname{div} v|} \quad \text{and} \quad \mathfrak{d}(w) := \phi'_{|\nabla w|}(|\operatorname{div} w|)\frac{\operatorname{div} w}{|\operatorname{div} w|}.$$

Then, for all $v, w \in \mathbb{V}$ it holds

$$\int_\Omega \big(\phi_{|\nabla v|}\big)^*(|\mathfrak{d}(v) - \mathfrak{d}(w)|)\,dx \preccurlyeq \|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla w)\|^2_{L^2(\Omega)},$$

where the constant hidden in $\preccurlyeq$ solely depends on $\Delta_2(\{\phi, \phi^*\})$ and d.

Proof. By Lemma 68, Lemma 66, and Corollary 10, it holds

$$\begin{aligned}
&\int_\Omega \big(\phi_{|\nabla v|}\big)^*\big(\,|\mathfrak{d}(v) - \mathfrak{d}(w)|\,\big)\,dx \\
&\qquad = \int_\Omega \big(\phi_{|\nabla v|}\big)^*\Big(\Big|\phi'_{|\nabla v|}(|\operatorname{div} v|)\frac{\operatorname{div} v}{|\operatorname{div} v|} - \phi'_{|\nabla w|}(|\operatorname{div} w|)\frac{\operatorname{div} w}{|\operatorname{div} w|}\Big|\Big)\,dx \\
&\qquad \preccurlyeq \int_\Omega \big(\phi_{|\nabla v|}\big)^*\Big(\Big|\phi'_{|\nabla v|}(|\operatorname{div} v|)\frac{\operatorname{div} v}{|\operatorname{div} v|} - \phi'_{|\nabla v|}(|\operatorname{div} w|)\frac{\operatorname{div} w}{|\operatorname{div} w|}\Big| \\
&\qquad\qquad\qquad + \phi'_{|\nabla v|}(|\nabla v - \nabla w|)\Big)\,dx \\
&\qquad \preccurlyeq \int_\Omega \big(\phi_{|\nabla v|}\big)^*\Big(\Big|\phi'_{|\nabla v|}(|\operatorname{div} v|)\frac{\operatorname{div} v}{|\operatorname{div} v|} - \phi'_{|\nabla v|}(|\operatorname{div} w|)\frac{\operatorname{div} w}{|\operatorname{div} w|}\Big|\Big)\,dx \\
&\qquad\quad + \int_\Omega \big(\phi_{|\nabla v|}\big)^*\big(\phi'_{|\nabla v|}(|\nabla v - \nabla w|)\big)\,dx,
\end{aligned}$$

where the constant hidden in $\preccurlyeq$ solely depends on $\Delta_2(\{\phi_{|\nabla v|}, (\phi_{|\nabla v|})^*\})$ and thus on $\Delta_2(\{\phi, \phi^*\})$; see Lemma 57. Applying Corollary 65 in 1-dimension for the N-function $\phi_{|\nabla v|}$ to the first addend and (2.8) to the second yields

$$\begin{aligned}
&\int_\Omega \big(\phi_{|\nabla v|}\big)^*\big(\,|\mathfrak{d}(v) - \mathfrak{d}(w)|\,\big)\,dx \\
&\qquad \preccurlyeq \int_\Omega \big(\phi_{|\nabla v|}\big)^*\Big(\big(\phi'_{|\nabla v|}\big)_{|\operatorname{div} v|}(|\operatorname{div} v - \operatorname{div} w|)\Big)\,dx \\
&\qquad\quad + \int_\Omega \phi_{|\nabla v|}(|\nabla v - \nabla w|)\,dx.
\end{aligned}$$

Lemma 128 yields the pointwise estimate $|\operatorname{div} v| \leq \sqrt{d}\,|\nabla v|$. Hence, Lemma 58, the monotonicity of ϕ', and Corollary 17 imply

$$\begin{aligned}\Big(\phi'_{|\nabla v|}\Big)_{|\operatorname{div} v|}(t) = \phi'_{|\nabla v|+|\operatorname{div} v|}(t) &= \frac{\phi'(|\nabla v| + |\operatorname{div} v| + t)}{|\nabla v| + |\operatorname{div} v| + t}\, t \\ &\leq \frac{\phi'\big((1+\sqrt{d})(|\nabla v| + t)\big)}{|\nabla v| + t}\, t \\ &\preccurlyeq \frac{\phi'(|\nabla v| + t)}{|\nabla v| + t}\, t = \phi'_{|\nabla v|}(t)\end{aligned} \tag{4.41}$$

for all $t \geq 0$. Therefore, by Corollary 10 and (2.8)

$$\begin{aligned}&\int_\Omega \big(\phi_{|\nabla v|}\big)^* \big(\,|\mathfrak{d}(v) - \mathfrak{d}(w)|\,\big)\, dx \\ &\qquad \preccurlyeq \int_\Omega \big(\phi_{|\nabla v|}\big)^* \big(\phi'_{|\nabla v|}(|\operatorname{div} v - \operatorname{div} w|)\big)\, dx + \int_\Omega \phi_{|\nabla v|}(|\nabla v - \nabla w|)\, dx \\ &\qquad \preccurlyeq \int_\Omega \phi_{|\nabla v|}(|\operatorname{div} v - \operatorname{div} w|)\, dx + \int_\Omega \phi_{|\nabla v|}(|\nabla v - \nabla w|)\, dx.\end{aligned}$$

Hence, applying Lemma 128 and Corollary 10 once more yields

$$\preccurlyeq \int_\Omega \phi_{|\nabla v|}(|\nabla v - \nabla w|)\, dx,$$

where the constant hidden in $\preccurlyeq$ depends only on $\Delta_2(\{\phi, \phi^*\})$ and d. Applying Lemma 74 yields the assertion. □

4.3.2 Interpolation of Discrete Functions

The approximation (4.40) is not suitable for updating the pressure, since it does not belong to the discrete pressure space $\mathbb{Q}(\mathcal{T})$ but to the space of piecewise constant functions

$$\mathbb{Q}_D(\mathcal{T}) := \{Q : Q|_T \in \mathcal{P}^0(T) \quad \text{for all } T \in \mathcal{T}\},$$

on a conforming conforming triangulation $\mathcal{T}$ of Ω — recall that the procedure ELLIPT requires a certain regularity of the pressure; see Section 4.3.1. To overcome this drawback we interpolate the approximation of the quasi-steepest direction (4.40) into the space of continuous, piecewise affine functions. The interpolation estimates presented in this section for discrete functions are a generalization of the ones in [6] to the quasi-norm case.

We use an interpolation operator $\Pi_{\mathcal{T}}^{\mathbb{Q}} : \mathbb{Q}_D(\mathcal{T}) \subset L^1(\Omega) \to \mathbb{Q}(\mathcal{T})$, which is closely related to the Clément operator [22]: Let $z \in \mathcal{N}$ be a node of the triangulation $\mathcal{T}$ and ω_z the corresponding finite element star; see Section 3.3.1.

For $q \in L^1(\Omega)$ let $\Pi_z^2 : L^1(\Omega) \to \mathcal{P}^1(\omega_z)$ be the L^2-projection into the space of continuous piecewise linear polynomials, i.e.,

$$\int_{\omega_z} (q - \Pi_z^2 q)\, Q\, dx = 0 \qquad \text{for all } Q \in \mathcal{P}^1(\omega_z). \tag{4.42}$$

We then set $\Pi_{\mathcal{T}}^{\mathbb{Q}} q(z) := \Pi_z^2 q(z)$; hence, $\Pi_{\mathcal{T}}^{\mathbb{Q}} q = \sum_{z\in\mathcal{N}} \Pi_z^2 q(z)\, \Phi_z \in \mathbb{Q}(\mathcal{T})$, where $\{\Phi_z : z \in \mathcal{N}\}$ denotes the Lagrange-basis of $\mathbb{Q}(\mathcal{T})$. Note that $\Pi_{\mathcal{T}}^{\mathbb{Q}} : L^1(\Omega) \to \mathbb{Q}(\mathcal{T})$ is a projection; see [22].

With this interpolation operator we can modify (4.40): For $Q \in \mathbb{Q}(\mathcal{T})$ let $U_Q \in \mathring{\mathbb{V}}(\mathcal{T})$ be the finite element approximation of u_Q, i.e.,

$$\int_\Omega \mathbf{A}(\nabla U_Q) : \nabla V\, dx = \int_\Omega (f - \nabla Q) \cdot V\, dx \qquad \text{for all } V \in \mathring{\mathbb{V}}(\mathcal{T}); \tag{4.43}$$

c.f. also Section 3.3. Then

$$\mathfrak{D}_Q := \Pi_{\mathcal{T}}^{\mathbb{Q}} \phi'_{|\nabla U_Q|}(|\operatorname{div} U_Q|) \frac{\operatorname{div} U_Q}{|\operatorname{div} U_Q|} \in \mathbb{Q}(\mathcal{T}),$$

is an *approximated steepest descent direction* in $\mathbb{Q}(\mathcal{T})$.

The aim of this section is to estimate the distance between $\mathfrak{d}_Q$ and $\mathfrak{D}_Q$. The following lemmas are an adaption of the $L^2(\Omega)$ estimates from [6] to the $L^1(\Omega)$-case and are the starting point for the quasi-norm estimates.

Lemma 147. *Let $\mathcal{T}$ be a conforming triangulation of Ω. Then, we have with the notation above that for any $q \in L^1(\Omega)$*

$$\int_T \left|q - \Pi_{\mathcal{T}}^{\mathbb{Q}} q\right| dx \preccurlyeq \sum_{z\in\mathcal{N}\cap T} \int_T \left|q - \Pi_z^2 q\right| dx,$$

where the constant hidden in $\preccurlyeq$ depends only on the shape regularity of $\mathcal{T}$ and d.

Proof. Let $\Pi_z^2 q$, $z \in \mathcal{N}$ be defined as in (4.42). Thus, we have for a fixed $z_0 \in$

$\mathcal{N} \cap T$ by the triangle inequality

$$
\begin{aligned}
\int_T \left|q - \Pi_T^{\mathbb{Q}} q\right| \, dx &= \int_T \Big|q - \sum_{z\in\mathcal{N}} \Pi_z^2 q(z)\, \Phi_z\Big| \, dx \\
&= \int_T \Big|q - \sum_{z\in\mathcal{N}\cap T} \Pi_z^2 q(z)\, \Phi_z\Big| \, dx \\
&\le \int_T \Big|q - \sum_{z\in\mathcal{N}\cap T} \Pi_{z_0}^2 q(z)\, \Phi_z\Big| \, dx \\
&\quad + \int_T \Big|\sum_{z\in\mathcal{N}\cap T: z\neq z_0} (\Pi_z^2 q(z) - \Pi_{z_0}^2 q(z))\Phi_z\Big| \, dx \\
&\le \int_T \Big|q - \Pi_{z_0}^2 q\Big| \, dx \\
&\quad + \sum_{z\in\mathcal{N}\cap T: z\neq z_0} \int_T \Big|\Pi_z^2 q(z) - \Pi_{z_0}^2 q(z)\Big| \, dx,
\end{aligned} \tag{4.44}
$$

where we used that the Lagrange basis is a partition of unity and that the basis functions have values in $[0, 1]$. Since for the first term nothing has to be done, we continue estimating the second term. Recall that $\Pi_z^2 q \in \mathcal{P}^1(T)$ is a polynomial. Hence, scaling it to the reference situation all its norms are equivalent. Thus, recalling Section 3.3.1, we have for fixed $z \in \mathcal{N} \cap T$ with $z \neq z_0$

$$
\begin{aligned}
\Big|\Pi_z^2 q(z) - \Pi_{z_0}^2 q(z)\Big| &\le \sup_T \left|\Pi_z^2 q - \Pi_{z_0}^2 q\right| \\
&= \sup_{\hat{T}} \left|\Pi_z^2 q \circ F_T - \Pi_{z_0}^2 q \circ F_T\right| \\
&\le \tilde{C} \int_{\hat{T}} \left|\Pi_z^2 q \circ F_T - \Pi_{z_0}^2 q \circ F_T\right| \, d\hat{x} \\
&= \tilde{C} \int_T \left|\Pi_z^2 q - \Pi_{z_0}^2 q\right| \left|\det \mathbf{C}_T^{-1}\right| \, dx \\
&= \tilde{C} \, \frac{|\hat{T}|}{|T|} \int_T \left|\Pi_z^2 q - \Pi_{z_0}^2 q\right| \, dx.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sum_{z\in\mathcal{N}\cap T: z\neq z_0} \int_T \Big|\Pi_z^2 q(z) - \Pi_{z_0}^2 q(z)\Big| \, dx &= \sum_{z\in\mathcal{N}\cap T: z\neq z_0} |T| \, \Big|\Pi_z^2 q(z) - \Pi_{z_0}^2 q(z)\Big| \\
&\le \tilde{C} \, |\hat{T}| \sum_{z\in\mathcal{N}\cap T: z\neq z_0} \int_T \left|\Pi_z^2 q - \Pi_{z_0}^2 q\right| \, dx \\
&\le \tilde{C} \, |\hat{T}| \, (d+1) \sum_{z\in\mathcal{N}\cap T} \int_T \left|q - \Pi_z^2 q\right| \, dx,
\end{aligned}
$$

where we used the triangle inequality and that the number of nodes in T is bounded by $d+1$ in the last step. Inserting this in (4.44) yields the desired estimate. $\square$

Corollary 148. *Suppose the assumptions of Lemma 147, then*

$$\int_\Omega \left|q - \Pi_{\mathcal{T}}^{\mathbb{Q}} q\right| \, dx \preccurlyeq \sum_{z\in\mathcal{N}} \int_{\omega_z} \left|q - \Pi_z^2 q\right| \, dx,$$

where the constant hidden in $\preccurlyeq$ depends only on the shape regularity of $\mathcal{T}$ and d.

Proof. The assertion follows from Lemma 147 by summing the estimates therein over all $T \in \mathcal{T}$. $\square$

Next, we make use of the fact that the functions we focus on, lie in $\mathbb{Q}_D(\mathcal{T}) \subset L^1(\Omega)$, which in turn is finite-dimensional.

Lemma 149. *In addition to the assumptions of Lemma 147, let $Q \in \mathbb{Q}_D(\mathcal{T})$. Then, for all $z \in \mathcal{N}$*

$$\int_{\omega_z} \left|Q - \Pi_z^2 Q\right| \, dx \preccurlyeq \operatorname{diam}(\omega_z) \int_{\sigma_z} |[\![Q]\!]| \, d\sigma,$$

where $[\![\cdot]\!]$ denotes the jump across inter-element sides; see Section 3.4. The constant hidden in $\preccurlyeq$ depends only on the shape regularity of $\mathcal{T}$ and d.

Proof. Clearly, $(id - \Pi_z^2)\,\mathbb{Q}_D(\mathcal{T}(\omega_z))$ is a finite dimensional linear space and hence all of its norms are equivalent. We have to prove that $\int_{\sigma_z} |[\![\cdot]\!]| \, d\sigma$ is a norm on $(id - \Pi_z^2)\,\mathbb{Q}_D(\mathcal{T}(\omega_z))$. Let $Q \in \mathbb{Q}_D(\mathcal{T}(\omega_z))$ with

$$\int_{\sigma_z} |[\![Q]\!]| \, d\sigma = 0,$$

i.e., Q does not jump across σ_z, thus $Q \in \mathcal{P}^1(\omega_z)$. Since Π_z^2 is the local L^2-projection onto $\mathcal{P}^1(\omega_z)$, we have that $Q - \Pi_z^2 Q = 0$. All other norm-properties follow by the properties of the L^1-norm on σ_z. Now, the assertion follows by scaling to the reference situation, applying equivalence of norms on finite dimensional spaces and scaling back to the physical finite element star. In particular, let $\hat{\omega}_z$ be the reference finite element star corresponding to ω_z and $\hat{\sigma}_z$ the union of its interior sides; see also [3]. Then, we have with $Q_z = \Pi_z^2 Q$

$$\begin{aligned} \int_{\omega_z} |Q - Q_z| \, dx &\le \operatorname{diam}(\omega_z)^d \int_{\hat{\omega}_z} \left|\hat{Q} - \hat{Q}_z\right| d\hat{x} \\ &\preccurlyeq \operatorname{diam}(\omega_z)^d \int_{\hat{\sigma}_z} \left|[\![\hat{Q}]\!]\right| d\hat{\sigma} \preccurlyeq \operatorname{diam}(\omega_z) \int_{\sigma_z} |[\![Q]\!]| \, d\sigma, \end{aligned}$$

where $\hat{Q}, \hat{Q}_z$ denote the functions Q, Q_z after scaling to the reference finite element star $\hat{\omega}_z$. This proves the Lemma. $\square$

In the next Lemma we generalize Lemma 149 to the quasi-norm case. The result is crucial for estimating the error that occurs during interpolation.

Lemma 150. *Let $\mathcal{T}$ be a conforming triangulation of Ω and ϕ be an N-function that satisfies Assumption 40. For $V \in \mathbb{V}(\mathcal{T})$ let $\mathfrak{d} := \phi_{|\nabla V|}(|\operatorname{div} V|)\frac{\operatorname{div} V}{|\operatorname{div} V|}$. Then, for all $T \in \mathcal{T}$*

$$\int_T (\phi_{|\nabla V|})^*\big(\,|\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}|\,\big)\,dx \preccurlyeq \sum_{T' \in \mathcal{T}(S_T)} \int_{\partial T'} h_{T'}\, |[\![\mathbf{F}(\nabla V)]\!]|^2\, d\sigma.$$

The constant hidden in $\preccurlyeq$ depends only on $\Delta_2(\{\phi, \phi^*\})$, the shape regularity of $\mathcal{T}$ and d. The nonlinear vector-field $\mathbf{F} : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ is defined as in (3.14).

Proof. We observe that $\mathfrak{d} \in \mathbb{Q}_D(\mathcal{T})$. Therfore, scaling $\mathfrak{d}$ to the reference element $\hat{T}$, applying equivalence of norms on finite dimensional spaces, and scaling back to the physical element T, we obtain

$$\sup_T |\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}| = \sup_{\hat{T}} \Big|\hat{\mathfrak{d}} - \widehat{\Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}}\Big| \preccurlyeq \int_{\hat{T}} \Big|\hat{\mathfrak{d}} - \widehat{\Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}}\Big|\,dx \preccurlyeq \frac{1}{|T|}\int_T |\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}|\,dx.$$

Thus, we can apply Lemmas 147 and 149 to get

$$\begin{aligned}
\sup_T |\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}| &\preccurlyeq \frac{1}{|T|}\sum_{z\in\mathcal{N}\cap T}\int_T |\mathfrak{d} - \Pi_z^2\mathfrak{d}|\,dx \\
&\leq \frac{1}{|T|}\sum_{z\in\mathcal{N}\cap T}\int_{\omega_z} |\mathfrak{d} - \Pi_z^2\mathfrak{d}|\,dx \\
&\preccurlyeq \frac{1}{|T|}\sum_{z\in\mathcal{N}\cap T}\operatorname{diam}(\omega_z)\int_{\sigma_z} |[\![\mathfrak{d}]\!]|\,d\sigma.
\end{aligned}$$

Depending on the shape-regularity of $\mathcal{T}$ we have that $\frac{\operatorname{diam}(\omega_z)}{|T|} \approx \frac{1}{|\sigma_z|}$. Therefore, there holds

$$\sup_T |\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}| \preccurlyeq \sum_{z\in\mathcal{N}\cap T}\frac{1}{|\sigma_z|}\int_{\sigma_z} |[\![\mathfrak{d}]\!]|\,d\sigma.$$

Since $\#(\mathcal{N}\cap T)$ is bounded by $d+1$, this estimate yields with Corollary 10

$$\int_T (\phi_{|\nabla V|})^*\big(\,|\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}|\,\big)\,dx \preccurlyeq \int_T \sum_{z\in\mathcal{N}\cap T} (\phi_{|\nabla V|})^*\Big(\frac{1}{|\sigma_z|}\int_{\sigma_z} |[\![\mathfrak{d}]\!]|\,d\sigma\Big)\,dx.$$

Now, Jensen's inequality (Lemma 4) implies for the fixed shift $|\nabla V|_T|$

$$\begin{aligned}
\int_T (\phi_{|\nabla V|})^*\big(\,|\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}|\,\big)\,dx &\preccurlyeq \int_T \sum_{z\in\mathcal{N}\cap T}\frac{1}{|\sigma_z|}\int_{\sigma_z} (\phi_{|\nabla V|_T|})^*(|[\![\mathfrak{d}]\!]|)\,d\sigma\,dx \\
&\preccurlyeq \sum_{z\in\mathcal{N}\cap T}\int_{\sigma_z} h_T\,(\phi_{|\nabla V|_T|})^*(|[\![\mathfrak{d}]\!]|)\,d\sigma \\
&\preccurlyeq \sum_{T'\in\mathcal{T}(S_T)}\int_{\partial T'} h_T\,(\phi_{|\nabla V|_T|})^*(|[\![\mathfrak{d}]\!]|)\,d\sigma.
\end{aligned}$$

Similar to (3.27), we obtain with the help of Corollary 71

$$\int_T \big(\phi_{|\nabla V|}\big)^*\big(\,\big|\mathfrak{d}-\Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}\big|\,\big)\,dx \preccurlyeq \sum_{T'\in\mathcal{T}(S_T)}\int_{\partial T'} h_T\,\big(\phi_{|\nabla V|_{T'}|}\big)^*\big(\,|[\![\mathfrak{d}]\!]|\,\big)\,d\sigma + \sum_{T'\in\mathcal{T}(S_T)}\int_{\partial T'} h_T\,|\mathbf{F}(\nabla V|_{T'})-\mathbf{F}(\nabla V|_T)|^2\,d\sigma.$$

Note that the integrand of the last term is constant. By Lemma 91 we derive

$$|\mathbf{F}(\nabla V|_{T'})-\mathbf{F}(\nabla V|_T)| \le \sum_{\sigma\in\Sigma_T} |[\![\mathbf{F}(\nabla V)]\!]_\sigma|\,,$$

where $\Sigma_T=\{\sigma\in\mathcal{S}:\sigma\cap S_T\neq\emptyset\}$ is the set of sides in the interior of S_T; see also Figure 3.3. We recall that the amount of sides in Σ_T as well as the amount of elements in S_T is bounded with respect to the shape-regularity of $\mathcal{T}$. Hence, we get with the fact that $|\sigma'|\approx|\sigma|$ for all $\sigma,\sigma'\in\Sigma_T$ and $h_{T'}\approx h_T$ for all $T'\in\mathcal{T}(S_T)$ that

$$\sum_{T'\in\mathcal{T}(S_T)}\int_{\partial T'} h_T\,|\mathbf{F}(\nabla V|_{T'})-\mathbf{F}(\nabla V|_T)|^2\,d\sigma \preccurlyeq \sum_{T'\in\mathcal{T}(S_T)}\int_{\partial T'} h_{T'}\,|[\![\mathbf{F}(\nabla V)]\!]|^2\,d\sigma$$

and thus,

$$\int_T \big(\phi_{|\nabla V|}\big)^*\big(\big|\mathfrak{d}-\Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}\big|\big)\,dx \preccurlyeq \sum_{T'\in\mathcal{T}(S_T)}\int_{\partial T'} h_T\,\big(\phi_{|\nabla V|}\big)^*\big(\,|[\![\mathfrak{d}]\!]|\,\big)\,d\sigma + \sum_{T'\in\mathcal{T}(S_T)}\int_{\partial T'} h_{T'}\,|[\![\mathbf{F}(\nabla V)]\!]|^2\,d\sigma. \tag{4.45}$$

It remains to estimate the first term of the right-hand side of (4.45). For $\sigma\in\mathcal{S}$, let $T_1,T_2\in\mathcal{T}$ be the adjacent simplices, i.e., $\sigma=T_1\cap T_2$. Applying the definition of $\mathfrak{d}=\phi'_{|\nabla V|}(|\operatorname{div}V|)\frac{\operatorname{div}V}{|\operatorname{div}V|}$ Corollary 69 implies

$$\begin{aligned}
|[\![\mathfrak{d}]\!]_\sigma| &= \left|\phi'_{|\nabla V|}(|\operatorname{div}V|)\frac{\operatorname{div}V}{|\operatorname{div}V|}\Big|_{T_1} - \phi'_{|\nabla V|}(|\operatorname{div}V|)\frac{\operatorname{div}V}{|\operatorname{div}V|}\Big|_{T_2}\right| \\
&\preccurlyeq \left|\phi'_{|\nabla V|_{T_1}|}(|\operatorname{div}V|)\frac{\operatorname{div}V}{|\operatorname{div}V|}\Big|_{T_1} - \phi'_{|\nabla V|_{T_1}|}(|\operatorname{div}V|)\frac{\operatorname{div}V}{|\operatorname{div}V|}\Big|_{T_2}\right| \\
&\quad + \big|\phi'_{|\nabla V|_{T_1}|}\big(\,|\nabla V|_{T_1}-\nabla V|_{T_2}|\,\big)\big| \\
&= \left|\phi'_{|\nabla V|_{T_1}|}(|\operatorname{div}V|)\frac{\operatorname{div}V}{|\operatorname{div}V|}\Big|_{T_1} - \phi'_{|\nabla V|_{T_1}|}(|\operatorname{div}V|)\frac{\operatorname{div}V}{|\operatorname{div}V|}\Big|_{T_2}\right| \\
&\quad + \big|\phi'_{|\nabla V|_{T_1}|}\big(\,|[\![\nabla V]\!]_\sigma|\,\big)\big|,
\end{aligned} \tag{4.46}$$

Now, we can estimate the first addend, with the help of Corollary 64 by

$$\begin{aligned}\Big|\phi'_{|\nabla V|_{T_1}|}(|\operatorname{div} V|)\frac{\operatorname{div} V}{|\operatorname{div} V|}\Big|_{T_1} - \phi'_{|\nabla V|_{T_1}|}(|\operatorname{div} V|)\frac{\operatorname{div} V}{|\operatorname{div} V|}\Big|_{T_2}\Big| \\ \approx \Big(\phi'_{|\nabla V_{T_1}|}\Big)_{|\operatorname{div} V|_{T_1}|}(|[\![\nabla V]\!]_\sigma|),\end{aligned} \tag{4.47}$$

where the constants hidden in $\approx$ depend only on $\Delta_2(\{\phi_{|\nabla V|_{T_1}|}, (\phi_{|\nabla V|_{T_1}|})^*\})$ and thus on $\Delta_2(\{\phi, \phi^*\})$; see Lemma 57. Recalling Lemma 58, we have with $|\operatorname{div} V| \le \sqrt{d}\,|\nabla V|$ and the monotonicity of ϕ'

$$\begin{aligned}\Big(\phi'_{|\nabla V|_{T_1}|}\Big)_{|\operatorname{div} V|_{T_1}|}(t) &= \phi'_{|\nabla V|_{T_1}|+|\operatorname{div} V|_{T_1}|}(t) \\ &= \frac{\phi'(|\nabla V|_{T_1}| + |\operatorname{div} V|_{T_1}| + t)}{|\nabla V|_{T_1}| + |\operatorname{div} V|_{T_1}| + t}\, t \\ &\le \frac{\phi'\big((1+\sqrt{d})(|\nabla V|_{T_1}| + t)\big)}{|\nabla V|_{T_1}| + t}\, t \\ &\preccurlyeq \frac{\phi'(|\nabla V|_{T_1}| + t)}{|\nabla V|_{T_1}| + t}\, t = \phi'_{|\nabla V|}(t)\end{aligned}$$

for all $t \ge 0$. Thereby the last inequality follows from $\Delta_2(\{\phi, \phi^*\}) < \infty$ with Corollary 10. Applying this to (4.47) gives

$$\Big|\phi'_{|\nabla V|_{T_1}|}(|\operatorname{div} V|)\frac{\operatorname{div} V}{|\operatorname{div} V|}\Big|_{T_1} - \phi'_{|\nabla V|_{T_1}|}(|\operatorname{div} V|)\frac{\operatorname{div} V}{|\operatorname{div} V|}\Big|_{T_2}\Big| \preccurlyeq \phi'_{|\nabla V|_{T_1}|}(|[\![\nabla V]\!]_\sigma|),$$

where the constant hidden in $\preccurlyeq$ depends only on $\Delta_2(\{\phi, \phi^*\}) < \infty$ and d. Inserting this in (4.46) implies

$$|[\![\mathfrak{d}]\!]_\sigma| \preccurlyeq \phi'_{|\nabla V|_{T_1}|}(|[\![\nabla V]\!]_\sigma|).$$

Choosing $T_1 = T'$ for every addend of the right-hand side of (4.45), we have by $\Delta_2(\{\phi, \phi^*\}) < \infty$ and Corollary 10

$$\begin{aligned}\int_T \big(\phi_{|\nabla V|}\big)^*\big(\big|\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}\big|\big)\,dx \preccurlyeq \sum_{T'\in\mathcal{T}(S_T)} \int_{\partial T'} h_T \big(\phi_{|\nabla V|}\big)^*\big(\phi'_{|\nabla V|}(|[\![\nabla V]\!]|)\big)\,d\sigma \\ + \sum_{T'\in\mathcal{T}(S_T)} \int_{\partial T'} h_{T'}\, |[\![\mathbf{F}(\nabla V)]\!]|^2\,d\sigma.\end{aligned}$$

Now, (2.8) and Proposition 62 imply

$$\begin{aligned}\int_T (\phi_{|\nabla V|})^*(|\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}|)\,dx &\preccurlyeq \sum_{T'\in\mathcal{T}(S_T)} \int_{\partial T'} h_T\,\phi_{|\nabla V|}(|[\![\nabla V]\!]|)\,d\sigma \\ &\quad + \sum_{T'\in\mathcal{T}(S_T)} \int_{\partial T'} h_{T'}\,|[\![\mathbf{F}(\nabla V)]\!]|^2\,d\sigma \\ &\preccurlyeq \sum_{T'\in\mathcal{T}(S_T)} \int_{\partial T'} h_{T'}\,|[\![\mathbf{F}(\nabla V)]\!]|^2\,d\sigma,\end{aligned}$$

where we additionally used that $h_T \approx h_{T'}$ for all $T' \in \mathcal{T}(S_T)$ depending on the shape-regularity of $\mathcal{T}$. This is the asserted estimate. □

Using the finite overlapping of the S_T, $T \in \mathcal{T}$, we can immediately deduce the following global version of Lemma 150.

Corollary 151. *Assuming the conditions of Lemma 150 it holds*

$$\int_\Omega (\phi_{|\nabla V|})^*\big(|\mathfrak{d} - \Pi_{\mathcal{T}}^{\mathbb{Q}}\,\mathfrak{d}|\big)\,dx \preccurlyeq \sum_{T\in\mathcal{T}} \int_{\partial T} h_T\,|[\![\mathbf{F}(\nabla V)]\!]|^2\,d\sigma.$$

Where the constant hidden in $\preccurlyeq$ depends only on $\Delta_2(\{\phi,\phi^*\})$ and the shape regularity of $\mathcal{T}$.

The next corollary combines the above results to the particular case of the finite element approximation of the quasi-steepest descent direction. In particular, it estimates the error between $\mathfrak{d}_Q$ and $\mathfrak{D}_Q$ by the quantity that is controlled by ELLIPT, namely by the estimator of the error between u_Q and U_Q.

Corollary 152. *Let ϕ be an N-function that satisfies Assumption 40 and let $\mathcal{T}$ be a conforming triangulation of the domain $\Omega \subset \mathbb{R}^d$. Then, with the notation of this section*

$$\int_\Omega (\phi_{|\nabla U_Q|})^*(|\mathfrak{d}_Q - \mathfrak{D}_Q|) \preccurlyeq \eta^2(U_Q, \mathcal{T}, f - \nabla Q),$$

where η denotes the error estimator defined in (3.24). *Thereby the constant hidden in $\preccurlyeq$ depends only on $\Delta_2(\{\phi,\phi^*\})$, the shape regularity of $\mathcal{T}$, and d.*

Proof. We start with the triangle like inequality of Corollary 10 and thus obtain

$$\begin{aligned}\int_\Omega (\phi_{|\nabla U_Q|})^*(|\mathfrak{d}_Q - \mathfrak{D}_Q|) &\preccurlyeq \int_\Omega (\phi_{|\nabla U_Q|})^*\Big(\Big|\mathfrak{d}_Q - \phi'_{|\nabla U_Q|}(|\operatorname{div} U_Q|)\frac{\operatorname{div} U_Q}{|\operatorname{div} U_Q|}\Big|\Big) \\ &\quad + (\phi_{|\nabla U_Q|})^*\Big(\Big|\phi'_{|\nabla U_Q|}(|\operatorname{div} U_Q|)\frac{\operatorname{div} U_Q}{|\operatorname{div} U_Q|} - \mathfrak{D}_Q\Big|\Big)\,dx,\end{aligned}$$

where we used that the Δ_2-constant of $(\phi_{|\nabla U_Q|})^*$ depends only on $\Delta_2(\{\phi, \phi^*\})$; see Lemma 57. Now, the first term can be estimated by Lemma 146. In particular,

$$\int_\Omega (\phi_{|\nabla U_Q|})^* \Big(\Big| \mathfrak{d}_Q - \phi'_{|\nabla U_Q|}(|\operatorname{div} U_Q|) \frac{\operatorname{div} U_Q}{|\operatorname{div} U_Q|} \Big| \Big) \, dx \preccurlyeq \|\mathbf{F}(\nabla u_Q) - \mathbf{F}(\nabla U_Q)\|^2_{L^2(\Omega)} .$$

This term can be estimated by the upper bound (Theorem 90). Furthermore, by Corollary 151 then

$$\int_\Omega (\phi_{|\nabla U_Q|})^* \Big(\Big| \phi'_{|\nabla U_Q|}(|\operatorname{div} U_Q|) \frac{\operatorname{div} U_Q}{|\operatorname{div} U_Q|} - \mathfrak{D}_Q \Big| \Big) \, dx \\ \preccurlyeq \sum_{T \in \mathcal{T}} \int_{\partial T} h_T \, |[\![\mathbf{F}(\nabla U_Q)]\!]|^2 \, d\sigma.$$

Recalling (3.24), this is a part of the estimator and thus obviously can be estimated by $\eta^2(U_Q, \mathcal{T}, f - \nabla Q)$. Hence the proposition is proved. □

Remark 153. *In our case, it is crucial to have the approximation of the quasi-steepest descent direction inside the pressure space $\mathbb{Q}(\mathcal{T})$ — recall that the procedure* ELLIPT *requires sufficient regular functions in its first argument. This requires interpolation estimates for a suitable interpolation operator from $\mathbb{Q}_D(\mathcal{T})$ into $\mathbb{Q}(\mathcal{T})$, since the divergence of the discrete velocity is not sufficiently regular.*

Similar estimates may be mandatory if one deals with stable pairs of discrete function spaces; see [6]. In particular, often the divergence of the discrete velocity is not contained in the discrete pressure space and hence has to be projected into it. For example consider the popular Taylor Hood elements $P_2 - P_1$, i.e., continuous piecewise second order polynomials for the discrete velocity and continuous piecewise linear elements for the discrete pressure. Thus the divergence of the velocity is piecewise linear but may jump over inter-element sides and therefore is not contained in the pressure space.

Another example is the so called Mini*-element, which is close to our case. In fact, piecewise linear continuous elements are used for the discretization of the pressure space. The discrete velocity space also contains piecewise linear continuous elements, but is additionally enriched by element bubble functions in order to obtain stability. However, the divergence of the discrete velocity is again not contained in the discrete pressure space and hence a projection-estimate is required.*

4.3.3 Convergent Adaptive Uzawa Algorithm (AUA)

Thanks to the above results on the approximated steepest descent direction, we are now able to state the adaptive finite element algorithm for the stationary Stokes problem. We suppose that ϕ is an N-function that satisfies Assumption 116.

Algorithm 154 (AUA)**.** Let $\mathcal{T}_0$ be a conforming initial triangulation of Ω and let $Q_0 \in \mathbb{Q}(\mathcal{T}_0)$ be an initial guess for $p \in \mathbb{Q}$. Fix $\theta, \rho \in (0,1)$, and $\mu > 0$ and let $j = 0$;

1. (APPROXIMATED DERIVATIVE)

$$(U_{Q_j}, \mathcal{T}_{j+1}) := \mathsf{ELLIPT}(\mathcal{T}_j, \rho^j, Q_j, \theta);$$

2. (APPROXIMATED QUASI-STEEPEST DESCENT DIRECTION)

$$\mathfrak{D}_j := \Pi^{\mathbb{Q}}_{\mathcal{T}_{j+1}} \phi'_{|\nabla U_{Q_j}|}\big(\,|\operatorname{div} U_{Q_j}|\,\big)\frac{\operatorname{div} U_{Q_j}}{|\operatorname{div} U_{Q_j}|};$$

3. (UPDATE)

$$Q_{j+1} := Q_j + \mu\,\mathfrak{D}_j;$$

increment j and go to step (1);

Remark 155. *For the reason of numerical cancellations it may be convenient to try to avoid extreme values of Q_j. For this purpose one may consider functions with mean value zero since the pressure is only determinated up to a constant value. Hence, starting Algorithm 154* (AUA) *with an initial guess $Q_0 \in \mathbb{Q}(\mathcal{T}_0)$, which has mean value zero we can substitute step 3* (UPDATE) *of* (AUA) *by*

3'. (UPDATE')

$$Q_{j+1} := Q_j + \mu\,\mathfrak{D}_j - \frac{1}{|\Omega|}\int_\Omega \mu\,\mathfrak{D}_j\,dx;$$

increment j and go to step (1)*.*

Therefore, by induction $(Q_j)_{j\in\mathbb{N}} \subset L_0^{\phi^*}(\Omega)$. Note that the modifications do not affect the theoretical behavior of (AUA)*, since the pressure is only defined up to a constant, i.e., $\mathbb{Q} = L^{\phi^*}(\Omega)/\mathbb{R}$. Hence, subtracting the mean-value has no theoretical effect. Moreover, recall from Lemma 129 and Corollary 130 that the convergence of the sequence $(Q_j)_{j\in\mathbb{N}} \subset \mathbb{Q}$ is equivalent to the convergence of its representants in $L_0^{\phi^*}(\Omega)$. Thus, for numerical evaluation it is rather convenient to consider error quantities related to $L_0^{\phi^*}(\Omega)$ instead of the corresponding quantities in $\mathbb{Q}$, which require a minimization over $\mathbb{R}$; cf. Lemma 129 and Corollary 130.*

Theorem 156. *Let ϕ be an N-function that satisfies Assumption 116. Then there exists $\mu_0 > 0$ depending only on $\Delta_2(\{\phi, \phi^*\})$ and d, such that for all step-sizes $\mu \in (0, \mu_0)$, it holds for the sequence $(Q_j)_{j\in\mathbb{N}} \subset \mathbb{Q}$ produced by Algorithm 154* (AUA) *that*

$$Q_j \to p \quad \text{in } \mathbb{Q}, \text{ as } j \to \infty.$$

Proof. For convenience, we use the abbreviations

$$\mathfrak{d}_j = \mathfrak{d}_{Q_j} = -\phi'_{|\nabla u_j|}(|\operatorname{div} u_j|)\frac{\operatorname{div} u_j}{|\operatorname{div} u_j|} \qquad \text{and} \qquad u_j = u_{Q_j};$$

see also (4.39). Recall that $\Delta_2(\{\phi_a, (\phi_a)^*\})$ depends only on $\Delta_2(\{\phi, \phi^*\}$; cf. Lemma 57. As in the proof of Theorem 139 let for $Q_j \in \mathbb{Q}(\mathcal{T}_j)$, $j \in \mathbb{N}$,

$$\mathcal{H}_j(\mu) := \mathcal{F}(Q_j) - \mathcal{F}(Q_j + \mu\,\mathfrak{D}_j)$$

By means of the mean value theorem and Proposition 133, for $\mu > 0$, there exists $\theta \in (0, \mu)$, such that

$$\begin{aligned}\mathcal{H}_j(\mu) &= \mu\,\mathcal{H}'_j(\theta) = -\mu\,\langle D\mathcal{F}(Q_j + \theta\,\mathfrak{D}_j), \mathfrak{D}_j\rangle \\ &= -\mu\,\langle D\mathcal{F}(Q_j), \mathfrak{D}_j\rangle - \frac{\mu}{\theta}\langle D\mathcal{F}(Q_j + \theta\,\mathfrak{D}_j) - D\mathcal{F}(Q_j), \theta\,\mathfrak{D}_j\rangle \\ &= -\mu\,\langle D\mathcal{F}(Q_j), \mathfrak{d}_j\rangle + \mu\,\langle D\mathcal{F}(Q_j), \mathfrak{d}_j - \mathfrak{D}_j\rangle \\ &\quad - \frac{\mu}{\theta}\langle D\mathcal{F}(Q_j + \theta\,\mathfrak{D}_j) - D\mathcal{F}(Q_j), \theta\,\mathfrak{D}_j\rangle.\end{aligned} \tag{4.48}$$

We handle the terms at the right hand side separately. First, we have from (2.6b)

$$-\langle DF(Q_j), \mathfrak{d}_j\rangle = \int_\Omega \phi'_{|\nabla u_j|}(|\operatorname{div} u_j|)\,|\operatorname{div} u_j|\;dx \geq \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx. \tag{4.49}$$

The next term can be estimated with the help of Young's inequality (Proposition 11) for $\delta > 0$

$$\begin{aligned}|\langle D\mathcal{F}(Q_j), \mathfrak{d}_j - \mathfrak{D}_j\rangle| &\leq \int_\Omega |(\mathfrak{d}_j - \mathfrak{D}_j)\;\operatorname{div} u_j|\;dx \\ &\leq \delta \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx \\ &\quad + C_\delta \int_\Omega \big(\phi_{|\nabla u_j|}\big)^*(|\mathfrak{d}_j - \mathfrak{D}_j|)\,dx.\end{aligned}$$

The constant C_δ depends only on $\Delta_2(\{\phi_a\}_{a\geq 0})$ and thus on $\Delta_2(\{\phi, \phi^*\})$; see Lemma 57. Now, applying Lemma 146, there exists a constant $\hat{C} > 0$ depending only on $\Delta_2(\{\phi, \phi^*\})$ and d, such that

$$\begin{aligned}|\langle D\mathcal{F}(Q_j), \mathfrak{d}_j - \mathfrak{D}_j\rangle| &\leq \delta \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx \\ &\quad + C_\delta\,\hat{C}\,\big\|\mathbf{F}(\nabla u_j) - \mathbf{F}(\nabla U_{Q_j})\big\|^2_{L^2(\Omega)}.\end{aligned} \tag{4.50}$$

The last term in (4.48) can be estimated as in the proof of Theorem 139; see (4.23). In particular,

$$\langle D\mathcal{F}(Q_j + \theta\,\mathfrak{D}_j) - D\mathcal{F}(Q_j), \theta\,\mathfrak{D}_j\rangle \preccurlyeq \int_\Omega \big(\phi_{|\nabla u_j|}\big)^*(|\theta\,\mathfrak{D}_j|)\,dx,$$

where the constant in $\preccurlyeq$ depends only on $\Delta_2(\{\phi,\phi^*\})$.

Next, we change the shift with the help of Lemma 71 to $|\nabla U_{Q_j}|$, therefore obtaining

$$\langle D\mathcal{F}(Q_j+\theta\,\mathfrak{D}_j)-D\mathcal{F}(Q_j),\theta\,\mathfrak{D}_j\rangle \preccurlyeq \int_\Omega \big(\phi_{|\nabla U_{Q_j}|}\big)^*(|\theta\,\mathfrak{D}_j|)\,dx \\ + \big\|\mathbf{F}(\nabla u_j)-\mathbf{F}(\nabla U_{Q_j})\big\|^2_{L^2(\Omega)}.$$

Assuming $\mu_0\le 2$, we get similar to (4.24)

$$\big(\phi_{|\nabla U_{Q_j}|}\big)^*(|\theta\,\mathfrak{D}_j|)\preccurlyeq \theta^2\,\phi_{|\nabla U_{Q_j}|}(|\operatorname{div} U_{Q_j}|).$$

Where the constants of the last two displays, that are hidden in $\preccurlyeq$ solely depend on $\Delta_2(\{\phi,\phi^*\})$. Hence, there exists a constant $\tilde{C}$ solely depending on $\Delta_2(\{\phi,\phi^*\})$ and d, such that

$$\langle D\mathcal{F}(Q_j+\theta\,\mathfrak{D}_j)-D\mathcal{F}(Q_j),\theta\,\mathfrak{D}_j\rangle \preccurlyeq \tilde{C}\int_\Omega \theta^2\,\phi_{|\nabla U_{Q_j}|}(|\operatorname{div}\mathfrak{D}_j|)\,dx \\ + \tilde{C}\,\big\|\mathbf{F}(\nabla u_j)-\mathbf{F}(\nabla U_{Q_j})\big\|^2_{L^2(\Omega)}.$$

This, (4.49), and (4.50), applied to (4.48) yields

$$\begin{aligned}
\mathcal{H}_j(\mu) &= \mathcal{F}(Q_j)-\mathcal{F}(Q_j+\mu\,\mathfrak{D}_j)\\
&\ge \mu\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx\\
&\quad -\mu\Big\{\delta\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx + C_\delta\,\hat{C}\,\big\|\mathbf{F}(\nabla u_j)-\mathbf{F}(\nabla U_{Q_j})\big\|^2_{L^2(\Omega)}\Big\}\\
&\quad -\frac{\mu}{\theta}\Big\{\tilde{C}\int_\Omega \theta^2\,\phi_{|\nabla U_{Q_j}|}(|\operatorname{div}\mathfrak{D}_j|)\,dx - \tilde{C}\,\big\|\mathbf{F}(\nabla u_j)-\mathbf{F}(\nabla U_{Q_j})\big\|^2_{L^2(\Omega)}\Big\}\\
&= \mu(1-\delta-\tilde{C}\,\theta)\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx\\
&\quad -(\mu\,C_\delta\,\hat{C}+\frac{\mu}{\theta}\,\tilde{C})\,\big\|\mathbf{F}(\nabla u_j)-\mathbf{F}(\nabla U_{Q_j})\big\|^2_{L^2(\Omega)}.
\end{aligned}$$

Recall that $\theta\le\mu$, hence

$$\mathcal{H}_j(\mu) = \mu\,(1-\delta-\tilde{C}\,\mu)\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx \\ -(\mu\,C_\delta\,\hat{C}+\tilde{C})\,\big\|\mathbf{F}(\nabla u_j)-\mathbf{F}(\nabla U_{Q_j})\big\|^2_{L^2(\Omega)}.$$

Observe that for $\mu_0\in(0,1/\tilde{C})$, $\delta:=(1-\tilde{C}\mu)/2>0$, we have for all $\mu\in(0,\mu_0)$ that $c_\mu:=\mu\,(1-\delta-\tilde{C}\,\mu)>0$. Take $C_\mu:=(\mu\,C_\delta\,\hat{C}+\tilde{C})$, then

$$\begin{aligned}
\mathcal{H}_j(\mu) &= \mathcal{F}(Q_j)-\mathcal{F}(Q_j+\mu\,Dc_j)\\
&\ge c_\mu\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx - C_\mu\,\big\|\mathbf{F}(\nabla u_j)-\mathbf{F}(\nabla U_{Q_j})\big\|^2_{L^2(\Omega)}.
\end{aligned}\tag{4.51}$$

The constants $c_\mu, C_\mu > 0$ depend only on $\Delta_2(\{\phi, \phi^*\})$, the step-size μ and d. Note that due to Algorithm 154 (AUA) — step 1 (APPROXIMATED DERIVATIVE) — and the upper bound (Theorem 90), U_{Q_j} is an approximation of u_j with accuracy at least $C_1 \rho^j$, i.e.,

$$\left\|\mathbf{F}(\nabla u_j) - \mathbf{F}(\nabla U_{Q_j})\right\|_{L^2(\Omega)} \leq C_1\, \eta(U_{Q_j}, \mathcal{T}_j, f - \nabla Q_j) \leq C_1\, \rho^j.$$

Therfore, we have

$$\begin{aligned}\mathcal{H}_j(\mu) &= \mathcal{F}(Q_j) - \mathcal{F}(Q_j + \mu\, \mathfrak{D}_j)\\ &\geq c_\mu \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx - C_\mu\, C_1\, \rho^{2j}.\end{aligned}$$

Now, the aim is to prove that $\operatorname{div} u_j \to_{j\to\infty} 0$ in $L_0^\phi(\Omega)$, since this implies $Q_j \to_{j\to\infty} p$ in $\mathbb{Q}$; see Lemma 138. Recalling that $Q_{j+1} = Q_j + \mu\, \mathfrak{D}_j$, we have for all $J \in \mathbb{N}$ the telescopic sum

$$\begin{aligned}\mathcal{F}(Q_0) - \mathcal{F}(Q_J) &= \sum_{j=0}^{J-1} \mathcal{F}(Q_j) - \mathcal{F}(Q_{j+1})\\ &\geq c_\mu \sum_{j=0}^{J-1} \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx - C_\mu\, C_1 \sum_{j=0}^{J-1} \rho^{2j}.\end{aligned}$$

The last term can be estimated by a geometric series and thus by $1/(1-\rho^2)$. On the other hand we can estimate $\mathcal{F}(Q_0) - \mathcal{F}(p) \geq \mathcal{F}(Q_0) - \mathcal{F}(Q_J)$, since $p \in \mathbb{Q}$ is the minimizer of $\mathcal{F}$. Therefore,

$$\begin{aligned}\mathcal{F}(Q_0) - \mathcal{F}(p) &\geq \mathcal{F}(Q_0) - \mathcal{F}(Q_J)\\ &\geq c_\mu \sum_{j=0}^{J-1} \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx - C_\mu\, C_1 \frac{1}{1-\rho^2}\end{aligned} \tag{4.52}$$

for all $J \in \mathbb{N}$. In other words, the series $\sum_{j=0}^{J-1} \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx$ is bounded. Since all its addends are positive, we get that

$$\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\, dx \to 0, \qquad \text{as } j \to 0.$$

As in the proof of Theorem 139 it remains to show that this implies $\operatorname{div} u_j \to 0$ in $\mathbb{Q}$ as $j \to \infty$. Then, the assertion follows by Lemma 138. In particular, we obtain by (4.52)

$$\mathcal{F}(Q_0) + C_\mu\, C_1 \frac{1}{1-\rho^2} \geq \mathcal{F}(Q_j)$$

for all $j \in \mathbb{N}$, i.e., $(\mathcal{F}(Q_j))_{j\in\mathbb{N}}$ is bounded. Combining (4.9) with (4.10) gives

$$\begin{aligned}\mathcal{F}(Q_0) \geq \mathcal{F}(Q_j) = -\mathcal{L}(u_j, Q_j) &= \int_\Omega -\phi(|\nabla u_j|) + Q_j \operatorname{div} u_j + f u_j \, dx \\ &= \int_\Omega -\phi(|\nabla u_j|) + \mathbf{A}(\nabla u_j) : \nabla u_j \, dx \\ &= \int_\Omega -\phi(|\nabla u_j|) + \phi'(|\nabla u_j|)\, |\nabla u_j| \, dx \\ &\geq (\nabla(\phi) - 1) \int_\Omega \phi(|\nabla u_j|) \, dx \geq 0,\end{aligned}$$

where the constant $\nabla(\phi) > 1$ depends only on $\Delta_2(\phi^*)$; see Proposition 14 ii). Therefore, the sequence $(\int_\Omega \phi(|\nabla u_j|)\,dx)_{j\in\mathbb{N}} \subset \mathbb{R}$ is bounded. Assume that $(\operatorname{div} u_j)_{j\in\mathbb{N}}$ does not converge to zero in $\mathbb{Q}$. Then, Proposition 31 implies, w.l.o.g., that there exists $c > 0$ such that

$$0 < c < \int_\Omega \phi(|\operatorname{div} u_j|) \, dx \quad \text{for all } j \in \mathbb{N}$$

— otherwise we pass to a subsequence. Hence, we get by Corollary 69 for $\delta > 0$

$$c < \int_\Omega \phi(|\operatorname{div} u_j|) \, dx \preccurlyeq (1 + C_\delta) \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|) \, dx + \delta \int_\Omega \phi(|\nabla u_j|) \, dx$$

for all $j \in \mathbb{N}$. Since $(\int_\Omega \phi(|\nabla u_j|)\,dx)_{j\in\mathbb{N}}$ is bounded, we can choose $\delta > 0$ small enough to obtain

$$0 < c \preccurlyeq C \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|) \, dx,$$

with a constant $C > 0$ not depending on $j \in \mathbb{N}$. This is a contradiction. Thus, $\operatorname{div} u_j \to 0$ in $\mathbb{Q}$, as $j \to \infty$ and the assertion follows with Lemma 138. □

Remark 157 (stopping criterion)**.** *Finding a stopping criterion for Algorithm 154* (AUA) *for an adequate distance quantity turns out to be no easy task. In fact, proving reasonable a posteriori estimates usually requires a continuous* inf-sup *condition; see [3, Section 9.2]. To have a reasonable estimator for a quasi-norm error notion, we need a* inf-sup *condition, which is somehow related to the quasi-norm; see* (4.35)*. Since such a condition is not available so far, we have to settle for non-optimal estimates like in [11]. They prove an upper bound for mixed finite element approximations. In our case* $(U_j, Q_j) \in \mathbb{V} \times \mathbb{Q}$*,* $j \in \mathbb{N}$*, is not a solution of the discrete Stokes problem. This makes our error analysis a bit unusual. However, since the same techniques as reported in [11] apply in our context, we only sketch the proof for completeness. We assume that*

$$\phi(t) = \int_0^t \big(\nu_\infty + (\nu_0 - \nu_\infty)(\kappa^2 + s^2)^{(r-2)/2}\big) s \, ds,$$

for fixed $\kappa \geq 0$, $\nu_0 > \nu_\infty \geq 0$. *This corresponds to the power law for* $\kappa = \nu_\infty = 0$, *and for* $\kappa > 0$ *to the Carreau law; see Section 1.1 and Remark 114. Note that* ϕ *satisfies Assumption 116; see Remark 118. Let* $\mathcal{T}$ *be a conforming triangulation of* Ω. *For* $Q \in \mathbb{Q}_D(\mathcal{T})$ *let* $U_Q \in \mathbb{V}(\mathcal{T})$ *be the finite element solution of* (4.43). *Then for* $u \in \mathbb{V}$ *and* $p \in \mathbb{Q}$ *being the unique solution of* (4.2) *we have like in [11], for any* $v \in \mathbb{V}$, $q \in \mathbb{Q}$, *and* $V \in \mathring{\mathbb{V}}(\mathcal{T})$

$$\begin{aligned}\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U_Q)\big) &: \nabla v - (p-Q)\,\operatorname{div} v - q\,\operatorname{div}(u - U_Q)\,dx \\ &= \int_\Omega f \cdot v - \mathbf{A}(\nabla U_Q) : \nabla v + Q\,\operatorname{div} v - q\,\operatorname{div} U_Q\,dx \\ &= \int_\Omega f \cdot (v - V) - \mathbf{A}(\nabla U_Q) : \nabla(v - V) + Q\,\operatorname{div}(v - V) - q\,\operatorname{div} U_Q\,dx.\end{aligned}$$

Element-wise integration by parts yields

$$\begin{aligned}\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U_Q)\big) &: \nabla v - (p-Q)\,\operatorname{div} v - q\,\operatorname{div}(u - U_Q)\,dx = \\ &= \sum_{T\in\mathcal{T}} \int_T (f - \nabla Q)\cdot(v - V)\,dx - \sum_{T\in\mathcal{T}} \int_{\partial T} [\![\mathbf{A}(\nabla U_Q)]\!]\,(v - V)\,d\sigma \\ &\quad + \sum_{T\in\mathcal{T}} \int_T q\,\operatorname{div} U_Q\,dx.\end{aligned}$$

Now, choosing $V = \Pi_{\mathcal{T}} v$ *the Scott-Zhang interpolant ([68]), we can estimate as in [11]*

$$\begin{aligned}\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U_Q)\big) &: \nabla v - (p-Q)\,\operatorname{div} v - q\,\operatorname{div}(u - U_Q)\,dx \\ &\lesssim \Big(\sum_{T\in\mathcal{T}} \Big\{ h_T^{r'} \|\mathbf{R}_1\|_{L^{r'}(T)}^{r'} + h_T \|\mathbf{R}_2\|_{L^{r'}(\partial T)}^{r'} \Big\}\Big)^{1/r'} |v|_{W^{1,r}(\Omega)} \\ &\quad + \Big(\sum_{T\in\mathcal{T}} \|R_3\|_{L^r(T)}^r\Big)^{1/r} \inf_{c\in\mathbb{R}} \|q - c\|_{L^{r'}(\Omega)},\end{aligned}$$

where $\frac{1}{r} + \frac{1}{r'} = 1$ *and*

$$\begin{aligned}\mathbf{R}_1|_T &:= f - \nabla Q|_T, && \text{for } T \in \mathcal{T}, \\ \mathbf{R}_2|_\sigma &:= [\![\mathbf{A}]\!]\,n|_\sigma, && \text{for } \sigma \in \mathcal{S},\end{aligned}$$

and

$$R_3|_T := \operatorname{div} U_Q|_T, \qquad \text{for } T \in \mathcal{T}.$$

Since q, v are arbitrary, taking $q = 0$ and then the supremum over all $v \in \mathbb{V}$, we get

$$
\begin{aligned}
\|\mathbf{S}_1\|_{\mathbb{V}^*} &:= \sup_{v\in\mathbb{V}} \frac{\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U_Q)\big) : \nabla v - (p-Q)\,\operatorname{div} v\, dx}{|v|_{W^{1,r}(\Omega)}} \\
&\preccurlyeq \Big(\sum_{T\in\mathcal{T}} \Big\{ h_T^{r'} \|\mathbf{R}_1\|_{L^{r'}(T)}^{r'} \Big)^{1/r'} + h_T \|\mathbf{R}_2\|_{L^{r'}(\partial T)}^{r'} \Big\} \Big)^{1/r'}.
\end{aligned} \tag{4.53}
$$

On the other hand, taking $w = 0$ and then the supremum over $q \in \mathbb{Q}$ yields

$$
\|S_2\|_{\mathbb{Q}^*} := \sup_{q\in\mathbb{Q}} \frac{\int_\Omega q\ \operatorname{div}(u - U_Q)\, dx}{\|q\|_{\mathbb{Q}}} \preccurlyeq \Big(\sum_{T\in\mathcal{T}} \|R_3\|_{L^r(T)}^r \Big)^{1/r}. \tag{4.54}
$$

To continue, we cite two estimates from [11] (see also [9, 8]), which connect the quasi-norm to the $W^{1,r}$-norm. In particular, for $v, w \in \mathbb{V}$ then

$$
\begin{aligned}
\|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla w)\|_{L^2(\Omega)}^{2/r} &\preccurlyeq |v - w|_{W^{1,r}(\Omega)} \\
|v - w|_{W^{1,r}(\Omega)} &\preccurlyeq [\phi(|\nabla v|_{W^{1,r}(\Omega)} + |\nabla w|_{W^{1,r}(\Omega)})]^{(2-r)/2} \|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla w)\|_{L^2(\Omega)}
\end{aligned}
$$

if $r \in (1, 2]$ and

$$
\begin{aligned}
|v - w|_{W^{1,r}(\Omega)}^{r/2} &\preccurlyeq \|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla w)\|_{L^2(\Omega)} \\
\|\mathbf{F}(\nabla v) - \mathbf{F}(\nabla w)\|_{L^2(\Omega)} &\preccurlyeq [\phi(|\nabla v|_{W^{1,r}(\Omega)} + |\nabla w|_{W^{1,r}(\Omega)})]^{(r-2)/2} |v - w|_{W^{1,r}(\Omega)}
\end{aligned}
$$

if $r \in (2, \infty)$. Furthermore, it holds

$$
\left| \int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U_Q)\big) : \nabla w\, dx \right| \preccurlyeq \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_Q)\|_{L^2(\Omega)}^{\min\{1,\frac{2}{r'}\}} |w|_{W^{1,r}(\Omega)}. \tag{4.55}
$$

With these estimates at hand, we can deduce from the inf-sup *condition* (4.4) *and* (4.53) *that*

$$
\|p - Q\|_{\mathbb{Q}} \preccurlyeq \|\mathbf{S}_1\|_{\mathbb{V}^*} + \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_Q)\|_{L^2(\Omega)}^{\min\{1,\frac{2}{r'}\}}. \tag{4.56}
$$

Again from (4.53) *and then using* (4.54), *we find that*

$$
\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_Q)\|_{L^2(\Omega)}^2 \preccurlyeq \|\mathbf{S}_1\|_{V^*} |u - U_Q|_{W^{1,r}(\Omega)} + \|S_2\|_{\mathbb{Q}^*} \|p - Q\|_{\mathbb{Q}}.
$$

Now, we can apply (4.56) *to obtain by the above estimates and the classical Young inequality (see Remark 13) like in [11] that*

$$
\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U_Q)\|_{L^2(\Omega)}^2 \preccurlyeq \|\mathbf{S}_1\|_{\mathbb{V}^*}^{\mathcal{R}'} + \|\mathbf{S}_1\|_{\mathbb{V}^*} \|S_2\|_{\mathbb{Q}^*} + \|S_2\|_{\mathbb{Q}^*}^{\mathfrak{R}'},
$$

where $\mathcal{R} = \max\{r, 2\}$, $\mathfrak{R} = \max\{r', 2\}$, $\frac{1}{\mathcal{R}} + \frac{1}{\mathcal{R}'} = 1$, *and* $\frac{1}{\mathfrak{R}} + \frac{1}{\mathfrak{R}'} = 1$. *Hence,*

$$\|u - U_Q\|_{\mathbb{V}}^{\mathcal{R}} \preccurlyeq \|\mathbf{S}_1\|_{\mathbb{V}^*}^{\mathcal{R}'} + \|\mathbf{S}_1\|_{\mathbb{V}^*} \|S_2\|_{\mathbb{Q}^*} + \|S_2\|_{\mathbb{Q}^*}^{\mathfrak{R}'}$$

and

$$\|p - Q\|_{\mathbb{Q}}^{\mathfrak{R}} \preccurlyeq \|\mathbf{S}_1\|_{\mathbb{V}^*}^{\mathfrak{R}} + \|\mathbf{S}_1\|_{\mathbb{V}^*}^{\mathcal{R}'} + \|\mathbf{S}_1\|_{\mathbb{V}^*} \|S_2\|_{\mathbb{Q}^*} + \|S_2\|_{\mathbb{Q}^*}^{\mathfrak{R}'}.$$

Thus, by (4.53) *and* (4.54) *we have computable a posteriori error bounds.*

Remark 158 (coarsening)**.** *Since the right hand side* $f - \nabla Q_j$ *of* (4.38) *in Algorithm 154* (AUA) *is changing in each iteration, it might be reasonable to apply a coarsening step in order to obtain optimal meshes. Recall, that for the proof of the convergence of* AUA *we only used that* $\eta(U_{Q_j}, \mathcal{T}_k, f - \nabla Q_j) \le \rho^k$. *In fact, the procedure* ELLIPT *can be substituted by any procedure that approximates* u_{Q_j} *up to this accuracy. Hence, it is possible to apply a coarsening routine, e.g., after step* (3) (UPDATE) *of the* AUA. *Note, that* Q_j *is defined on the common refinement of all triangulations* $\mathcal{T}_i$ $i = 1, \ldots, k$. *Therefore, it may be necessary to handle two grids, namely one grid for calculating* U_{Q_j} *in step* (1) *and then the common refinement of all triangulations* $\mathcal{T}_i$, $i = 1, \ldots, k$, *in order to store* Q_j.

Remark 159. *In [18] an algorithm for optimization of general convex functionals is proposed. As in our case, their algorithm is based on approximating the quasi-steepest descent direction. Actually, they ensure that the approximation of the quasi-steepest descent direction is still a descent direction. For our problem, this means that step* (1) *of Algorithm 154* (AUA) *is substituted by a method, which yields an approximation* U_{Q_j} *of the true solution* u_j *of* (4.38), *such that*

$$c_\mu \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx \ge \gamma\, C_\mu\, \left\|\mathbf{F}(\nabla u_j) - \mathbf{F}(\nabla U_{Q_j})\right\|_{L^2(\Omega)}^2,$$

for $\gamma \in (0,1)$, *where the constants* c_μ, C_μ *are those of* (4.51). *If we assume* $\operatorname{div} u_j \ne 0$ *— otherwise it holds* $u_j = u$ *and we are finished —, this goal is achievable: In fact, we can estimate by the generalized triangle inequality (Corollary 10), Corollary 69 and Lemma 128*

$$\int_\Omega \phi_{|\nabla U_{Q_j}|}(|\operatorname{div} U_{Q_j} - \operatorname{div} u_j|)\,dx \le \hat{C}\, \left\|\mathbf{F}(\nabla u_j) - \mathbf{F}(\nabla U_{Q_j})\right\|_{L^2(\Omega)}^2$$

and

$$\int_\Omega \phi_{|\nabla U_{Q_j}|}(|\operatorname{div} U_{Q_j}|)\,dx \preccurlyeq \hat{C} \left\{ \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|) + \left\|\mathbf{F}(\nabla u_j) - \mathbf{F}(\nabla U_{Q_j})\right\|_{L^2(\Omega)}^2 \right\},$$

with $\hat{C} > 0$ *depending only on* $\Delta_2(\{\phi, \phi^*\})$ *and* d. *Note that by Corollary 108 we can modify step* (1) *of* AUA, *such that the error* $\left\|\mathbf{F}(\nabla u_j) - \mathbf{F}(\nabla U_{Q_j})\right\|_{L^2(\Omega)}^2$

is sufficiently small. In particular, by the above estimates and the assumption $\operatorname{div} u_j \neq 0$*, we have for* $A > 0$ *that*

$$\begin{aligned} A \left\|\mathbf{F}(\nabla u_j) - \mathbf{F}(\nabla U_{Q_j})\right\|_{L^2(\Omega)}^2 &\leq \int_\Omega \phi_{|\nabla U_{Q_j}|}(|\operatorname{div} U_{Q_j}|) \\ &\leq \hat{C}\left\{\int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|)\,dx + \left\|\mathbf{F}(\nabla u_j) - \mathbf{F}(\nabla U_{Q_j})\right\|_{L^2(\Omega)}^2\right\}, \end{aligned}$$

i.e.,

$$\frac{A - \hat{C}}{\hat{C}} \left\|\mathbf{F}(\nabla u_j) - \mathbf{F}(\nabla U_{Q_j})\right\|_{L^2(\Omega)}^2 \leq \int_\Omega \phi_{|\nabla u_j|}(|\operatorname{div} u_j|).$$

Hence, for $A > 0$ *such that*

$$\frac{A - \hat{C}}{\hat{C}} \geq \gamma \frac{C_\mu}{c_\mu}.$$

we get the desired estimate. In view of (4.51)*, this yields a descent for* $\mathcal{F}$ *in each iteration* k*. The drawback of this method is that we need to know the constants* $c_\mu, C_\mu, \hat{C}$ *in order to calculate an approximation with sufficient accuracy. Furthermore, the accuracy may be much too high for an reasonable descent direction. For these reasons, we decided not to use a descent of* $\mathcal{F}$ *in each step.*

In [18] a new step-size is chosen in each iteration by a line-search algorithm, such that an adapted Wolfe's condition is satisfied; see also [24]. This line-search algorithm may require several approximate evaluations of the functional $\mathcal{F}$ *at different points. Since evaluating* $\mathcal{F}$ *is equivalent to solving a nonlinear Poisson equation, line search is expensive. For the benefit that* AUA *converges for a fixed step-size* μ *the special structure of our problem and in particular the quasi-norm techniques seem to be crucial.*

Remark 160. *Note that the spaces* $\mathring{\mathbb{V}}(\mathcal{T}), \mathbb{Q}(\mathcal{T})$ *are not stable in the sense, that they satisfy a discrete* inf-sup *condition*

$$\inf_{Q \in \mathbb{Q}(\mathcal{T})} \sup_{V \in \mathring{\mathbb{V}}(\mathcal{T})} \frac{\int_\Omega Q \operatorname{div} v\,dx}{\|Q\|_{\mathbb{Q}} \|V\|_{\mathbb{V}}} \geq \beta_{\mathcal{T}} > 0,$$

with $\beta_{\mathcal{T}}$ *independent of the triangulation* $\mathcal{T}$*; for pairs of stable function spaces cf., e.g., [9, 15, 44, 42]. However, Algorithm 154* (AUA) *is an generalized inexact Uzawa iteration at an infinite dimensional level. The convergence of our algorithm does not require a discrete* inf*-sup condition but rather the continuous* inf-sup *condition* (4.4).

Remark 161. *In Algorithm 154* (AUA) *we use an approximation to the quasi-steepest descent direction that is continuous and piecewise linear. This is due to* *the fact that the procedure* ESTIMATE *of* ELLIPT *requires a* $L^{\phi^*}(\Omega)^d$ *right hand side in* (4.38). *The reason for this is that for* $T \in \mathcal{T}$ *the interpolation estimate of Lemma 88 requires a constant shift on the whole patch* S_T. *According to Remark 92 this leads to a perturbation term of the form*

$$\sum_{T\in\mathcal{T}} \int_{\partial T} h_T \, |[\![\mathbf{F}(\nabla U)]\!]|^2 \, d\sigma, \tag{4.57}$$

where $U \in \mathring{\mathbb{V}}(\mathcal{T})$ *is the discrete Galerkin solution of the respective problem. Consider problem* (3.2) *with right hand side* $f - \nabla Q \in W^{-1,\phi^*}(\Omega)$ *for* $Q \in \mathbb{Q}_D(\mathcal{T})$, *i.e.,* $u \in \mathbb{V}$ *such that*

$$\int_\Omega \mathbf{A}(\nabla u) : \nabla v \, dx = \int_\Omega f \cdot v + Q \operatorname{div} v \, dx \quad \textit{for all } v \in \mathbb{V} \tag{4.58}$$

Furthermore let $U \in \mathring{\mathbb{V}}(\mathcal{T})$ *be its Ritz-Galerkin solution*

$$\int_\Omega \mathbf{A}(\nabla U) : \nabla V \, dx = \int_\Omega f\, V + Q \, \operatorname{div} V \, dx \quad \textit{for all } V \in \mathbb{V}(\mathcal{T}). \tag{4.59}$$

Then, similarly as in (3.23), *we obtain by integration by parts*

$$\begin{aligned}\int_\Omega \big(\mathbf{A}(\nabla u) - \mathbf{A}(\nabla U)\big) : \nabla v \, dx \\ = \sum_{T\in\mathcal{T}} \int_T f \cdot (v - V) \, dx - \sum_{\sigma\in\mathcal{S}} \int_\sigma [\![\mathbf{A}(\nabla U) - \nabla Q]\!] \; n \cdot (v - V) \, d\sigma;\end{aligned}$$

see [6] for the linear case. Therefore, the jump part of the estimator is not only determined by the jumps of ∇U, *but also by the jumps of* P. *Thus, the estimator becomes*

$$\begin{aligned}\eta_D^2(U, T, f - \nabla Q) = \int_T \big(\phi_{|\nabla U|}\big)^*(h_T \; |f|) \, dx \\ + \int_{\partial T} h_T \, \big(\phi_{|\nabla U|}\big)^*(|[\![\mathbf{A}(\nabla U) - Q \, id]\!]|) \, dx.\end{aligned}$$

The second part of the expression reflects the fact, that the jumps of ∇u *are related to the jumps of* Q. *Note that the jump estimator is essentially different from the terms in* (4.57). *Hence, the term* (4.57) *appears additionally in the upper bound*

$$\begin{aligned}\|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|_{L^2(\Omega)} \preccurlyeq \eta_D(U, \mathcal{T}, g) \\ + \Big(\sum_{T\in\mathcal{T}} \int_{\partial T} h_T \; |[\![\mathbf{F}(\nabla U)]\!]|^2 \; d\sigma\Big)^{1/2}.\end{aligned} \tag{4.60}$$

Similarly, we obtain with the techniques of the proof of Theorem 95

$$\eta_D(U,T,g) \preccurlyeq \|\mathbf{F}(\nabla u) - \mathbf{F}(\nabla U)\|_{L^2(\omega_T)} + \operatorname{osc}(U,\mathcal{T}(\omega_T)) + \Big(\int_{\partial T} h_T\, |[\![\mathbf{F}(\nabla U)]\!]|^2\, d\sigma\Big)^{1/2}. \tag{4.61}$$

For $Q \in \mathbb{Q}_D(\mathcal{T}_0)$, i.e., Q only jumps across interior sides of the initial triangulation, Algorithm 99 (AFEM) *still yields a contraction for the energy differences plus the estimator. This is due to the fact that only the upper bound is involved in the proof of Theorem 106. In particular, a perturbed estimator reduction (Lemma 105) is still valid, since all terms in $\eta_D(U,\mathcal{T},g)$ are scaled by the mesh-size. It seems that the estimator overestimates the error and thus we get an error reduction for the estimator that may not necessarily be close to the reduction of the error. This can be observed by the fact that by* (4.61) *and* (4.60) *we do not get an reasonable total error concept as in* (3.43)*. In particular, from* (4.58) *the jumps of ∇u are related to the jumps of Q. Therefore, we cannot expect that the jumps of ∇U across interior sides of the initial triangulations vanish and hence* (4.57) *can be of lower order.*

Remark 162 (symmetric gradient)**.** *Recall from Section 1.1 that physical models of quasi-Newtonian flow involve the symmetric gradient rather than the gradient in the formulation of the nonlinear Stokes equations, i.e., $u \in \mathbb{V}$, $p \in \mathbb{Q}$, such that*

$$\begin{aligned} \int_\Omega \mathbf{A}(\mathbf{E}(u)) : \mathbf{E}(v)\, dx - \int_\Omega p \operatorname{div} v\, dx &= \int_\Omega f\cdot v\, dx && \text{for all } v \in W_0^{1,\phi}(\Omega)^d \\ \int_\Omega q \operatorname{div} u\, dx &= 0 && \text{for all } q \in L^{\phi^*}(\Omega)/\mathbb{R}, \end{aligned} \tag{4.62}$$

where $\mathbf{E}(u) := \frac{1}{2}(\nabla u + \nabla u^t)$. Thanks to Korn's inequality (3.46)*, the norms $\|\nabla\cdot\|_\phi$ and $\|\mathbf{E}(\cdot)\|_\phi$ are equivalent norms on $W_0^{1\phi}(\Omega)$ and thus an* inf-sup *condition is valid, if ϕ satisfies Assumption 116; see* (4.4)*. Therefore, existence and uniqueness of a solution can be obtained as in Section 4.1.2.*

All definitions and results of Section 4.1.3 carry over to the case of (4.62) *substituting the gradient by the symmetric gradient — note that Lemma 128 remains valid, since $\operatorname{tr}(\mathbf{Q}) = \operatorname{tr}(\frac{1}{2}(\mathbf{Q}+\mathbf{Q}^t))$. In particular, this leads to a functional $\mathcal{F}_\mathbf{E} : \mathbb{Q} \to \mathbb{R}$, which is minimal in $p \in \mathbb{Q}$. Then for $\mathcal{F}_\mathbf{E}$ the quasi steepest descent direction* (4.18) *in $q \in \mathbb{Q}$ becomes*

$$\mathfrak{d}_q := -\phi'_{|\mathbf{E}(u_q)|}(|\operatorname{div} u_q|)\frac{\operatorname{div} u_q}{|\operatorname{div} u_q|},$$

where $u_q \in \mathbb{V}$ is the unique solution of (3.45) *with right-hand side $g = f - \nabla q$. Adapting Algorithm 136 according to the above considerations for the symmetric* *gradient, it produces a sequence* $(q_j)_{j\in\mathbb{N}} \subset \mathbb{Q}$ *that converges to the solution* p *of* (4.62).

Finally, recalling Remark 112, we can modify the procedure **ELLIPT** *(Algorithm 144) to get a method* $\mathsf{ELLIPT}_{\mathbf{E}}$ *in the same fashion as we modified the* AFEM *in Remark 112, ie, substituting* **SOLVE** *by* $\mathsf{SOLVE}_{\mathbf{E}}$ *and* **ESTIMATE** *by* $\mathsf{ESTIMATE}_{\mathbf{E}}$*. Hence, substituting* **ELLIPT** *by* $\mathsf{ELLIPT}_{\mathbf{E}}$ *in Algorithm 154* (AUA) *and changing step 2 of* AUA *into*

1. *APPROXIMATED QUASI-STEEPEST DESCENT DIRECTION*

$$\mathfrak{D}_j := \Pi^{\mathbb{Q}}_{\mathcal{T}_{j+1}} \phi'_{|\mathbf{E}(U_{Q_j})|}\big(\,|\operatorname{div} U_{Q_j}|\big)\frac{\operatorname{div} U_{Q_j}}{|\operatorname{div} U_{Q_j}|},$$

yields a convergent adaptive Uzawa finite element method for the pressure of the nonlinear stationary Stokes problem with symmetric gradient (4.62). *The proof of convergence works in the same fashion as the proof of Theorem 156.*

4.4 Conclusions and Outlook

We have presented algorithms for the nonlinear Poisson equation as well as for the nonlinear stationary Stokes problem with guaranteed convergence to the true solution.

For the nonlinear Poisson equation a posteriori analysis yields estimates for the error quantified in the so-called quasi-norm without a gap in the power of the upper and the lower bound. Moreover, a standard adaptive finite element method based upon these estimates features linear convergence.

For the nonlinear stationary Stokes equations we proposed an infinite dimensional steepest descent algorithm, which also makes use of the quasi-norm techniques.

Combining those two methods yields a practicable convergent adaptive algorithm for the nonlinear stationary Stokes equations.

Future work might concentrate on the following points:

- Numerical experiments for the adaptive algorithm for the nonlinear stationary Stokes problem. This is of great interest in confirming the obtained results as well as numerically validating some educated guesses.

- Improvement of quasi-norm interpolation estimates in order to use piecewise constant pressure in Algorithm 154 (AUA); compare with Remark 161.

- Generalization of the quasi-norm techniques to higher order elements. This is important for reducing the numerical complexity of (AFEM) as well as to allow for inf-sup stable function spaces in Algorithm 154.

- Prove an inf-sup condition for more general N-functions; see Remark 123. Such a condition would allow Assumption 116 to be weakened.

- Checking the quasi inf-sup condition (4.35). For this reason it is helpful to verify whether numerical experiments for Algorithm 154 show linear convergence or not; see Remark 142. The task of proving the quasi inf-sup condition may be passed forward to some pure analysts.

- Having a quasi-norm inf-sup at hand, efforts should be made to prove new a posteriori error estimates for the Stokes problem, making use of the quasi-norm techniques.

Appendix A

Bibliography

[1] E. Acerbi and N. Fusco. Regularity for minimizers of non-quadratic functionals: The case $1 < p < 2$. *J. Math. Anal. Appl.*, 140(1):115–135, 1989.

[2] R. A. Adams. *Sobolev spaces.* Pure and Applied Mathematics, 65. A Series of Monographs and Textbooks. New York-San Francisco-London: Academic Press, Inc., a subsidiary of Harcourt Brace Jovanovich, Publishers, 1975.

[3] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis.* Pure and Applied Mathematics. A Wiley-Interscience Series of Texts, Monographs, and Tracts. Chichester: Wiley, 2000.

[4] C. Amrouche and V. Girault. Decomposition of vector spaces and application to the Stokes problem in arbitrary dimension. *Czech. Math. J.*, 44(1):109–140, 1994.

[5] E. Bänsch. Local mesh refinement in 2 and 3 dimensions. *IMPACT Comput. Sci. Eng.*, 3(3):181–191, 1991.

[6] E. Bänsch, P. Morin, and R. H. Nochetto. An adaptive Uzawa FEM for the Stokes problem: Convergence without the Inf-sup condition. *SIAM J. Numer. Anal.*, 40(4):1207–1229, 2002.

[7] J. Baranger and H. El Amri. Estimateurs a posteriori d'erreur pour le calcul adaptatif d'écoulements quasi-Newtoniens. (A posteriori error estimators for adaptive calculation of quasi-Newtonian flows). *RAIRO, Anal. NumÃ©r.*, 25:31–48.

[8] J. W. Barrett and W. B. Liu. Finite element approximation of the p-Laplacian. *Math. Comput.*, 61(204):523–537, 1993.

[9] J. W. Barrett and W. B. Liu. Finite element error analysis of a quasi-Newtonian flow obeying the Carreau or power law. *Numer. Math.*, 64(4):433–453, 1993.

[10] J. W. Barrett and W. B. Liu. Quasi-norm error bounds for the finite element approximation of a non-Newtonian flow. *Numer. Math.*, 68(4):437–456, 1994.

[11] J. W. Barrett, J. A. Robson, and E. Süli. A posteriori error analysis of mixed finite element approximations to quasi-Newtonian incompressible flows. *Research Reports from the Numerical Analysis Group of the computing laboratory at Oxford University, UK*, NA-04/13:1–16, 2004.

[12] P. Binev, W. Dahmen, and R. DeVore. Adaptive finite element methods with convergence rates. *Numer. Math.*, 97(2):219–268, 2004.

[13] D. Braess. *Finite elements. Theory, fast solvers, and applications in solid mechanics. Transl. from the German by Larry L. Schumaker. 2nd ed.* Cambridge: Cambridge University Press, 2001.

[14] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods. 2nd ed.* Texts in Applied Mathematics. 15. Berlin: Springer, 2002.

[15] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods.* Springer Series in Computational Mathematics. 15. New York etc.: Springer-Verlag, 1991.

[16] F.E. Browder. Nonlinear elliptic functional equations in nonreflexive Banach spaces. *Bull. Am. Math. Soc.*, 72:89–95, 1966.

[17] J. Byström. Sharp constants for some inequalities connected to the p-Laplace operator. *JIPAM, J. Inequal. Pure Appl. Math.*, 6(2, paper no. 56):8p., electronic only, 2005.

[18] C. Canuto and K. Urban. Adaptive optimization of convex functionals in Banach spaces. *SIAM J. Numer. Anal.*, 42(5):2043–2075, 2005.

[19] M. Cascon, C. Kreuzer, R. H. Nochetto, and K. G. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *accepted for publication in SIAM J. Numer. Anal.*, 2008.

[20] Z. Chen and J. Feng. An adaptive finite element algorithm with reliable and efficient error control for linear parabolic problems. *Math. Comput.*, 73(247):1167–1193, 2004.

[21] P. G. Ciarlet. *The finite element method for elliptic problems.* Studies in Mathematics and its Applications. Vol. 4. Amsterdam - New York - Oxford: North-Holland Publishing Company, 1978.

[22] Ph. Clément. Approximation by finite element functions using local regularization. *RAIRO Anal. Numér.*, 9(R-2):77–84, 1975.

[23] Donald L. Cohn. *Measure theory.* Boston, Basel, Stuttgart: Birkhäuser, 1980.

[24] J. E. jun. Dennis and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations.* Prentice-Hall Series in Computational Mathematics. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1983.

[25] L. Diening, C. Ebmeyer, and M. Růžička. Optimal convergence for the implicite space-time discretization of parabolic systems with p structure. *SIAM J. Numer. Anal.*, 45:457–472, 2007.

[26] L. Diening and F. Ettwein. Fractional estimates for non-differentiable elliptic systems with general growth. *Forum Mathematicum*, 3:523–556, 2008.

[27] L. Diening and C. Kreuzer. Quasi-optimal convergence rate for an adaptive finite element method for the nonlinear Laplace equation. *In preparation.*

[28] L. Diening and C. Kreuzer. Linear convergence of an adaptive finite element method for the p-Laplacian equation. *SIAM J. Numer. Anal.*, 46(2):614–638, 2008.

[29] L. Diening and M. Růžička. Integral operators on the halfspace in generalized Lebesgue spaces $L^{p(\cdot)}$, part I. *J. Math. Anal. Appl.*, 298(2):559–571, 2004.

[30] L. Diening and M. Růžička. Integral operators on the halfspace in generalized Lebesgue spaces $L^{p(\cdot)}$, part II. *J. Math. Anal. Appl.*, 298(2):572–588, 2004.

[31] L. Diening and M. Růžička. Interpolation operators in Orlicz-Sobolev spaces. *Numer. Math.*, 107(1):107–129, 2007.

[32] L. Diening and M. Růžička. Non-Newtonian Fluids and Function Spaces. *Nonlinear Analysis, Function Spaces and Applications*, 8:95–143, 2007.

[33] L. Diening, M. Růžička, and K. Schumacher. A decomposition technique for John domains. *in preparation.*

[34] T. K. Donaldson. Nonlinear elliptic boundary-value problems in Orlicz-Sobolev spaces. *J. Differ. Equations*, 10:507–528, 1971.

[35] T. K. Donaldson and N. S. Trudinger. Orlicz-Sobolev spaces and imbedding theorems. *J. Funct. Anal.*, 8:52–75, 1971.

[36] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996.

[37] Nelson Dunford and Jacob T. Schwartz. *Linear Operators. I. General theory.* (Pure and Applied Mathematics. Vol. 6) New York and London: Interscience Publishers, 1958.

[38] C. Ebmeyer. Global regularity in Sobolev spaces for elliptic problems with p-structure on bounded domains. Rodrigues, José F. (ed.) et al., Trends in partial differential equations of mathematical physics. Selected papers of the international conference held on the occasion of the 70th birthday of V. A. Solonnikov, Óbidos, Portugal, June 7–10, 2003. Basel: Birkhäuser. Progress in Nonlinear Differential Equations and their Applications 61, 81-89 (2005)., 2005.

[39] C. Ebmeyer and W. B. Liu. Quasi-norm interpolation error estimates for the piecewise linear finite element approximation of p-Laplacian problems. *Numer. Math.*, 100(2):233–258, 2005.

[40] I. Ekeland and R. Temam. *Convex analysis and variational problems. Translated by Minerva Translations, Ltd., London.* Studies in Mathematics and its Applications. Vol. 1. Amsterdam - Oxford: North-Holland Publishing Company; New York: American Elsevier Publishing Company, Inc., 1976.

[41] L. C. Evans. *Partial differential equations.* Graduate Studies in Mathematics. 19. Providence, RI: American Mathematical Society (AMS), 1998.

[42] M. Fortin. Old and new finite elements for incompressible flows. *Int. J. Numer. Methods Fluids*, 1:347–364, 1981.

[43] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order. Reprint of the 1998 ed.* Classics in Mathematics. Berlin: Springer, 2001.

[44] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations. Theory and algorithms. (Extended version of the 1979 publ.).* Springer Series in Computational Mathematics, 5. Berlin etc.: Springer-Verlag, 1986.

[45] E. Giusti. *Direct methods in the calculus of variations.* Singapore: World Scientific, 2003.

[46] P. Grisvard. *Elliptic problems in nonsmooth domains.* Monographs and Studies in Mathematics, 24. Pitman Advanced Publishing Program. Boston-London-Melbourne: Pitman Publishing Inc., 1985.

[47] W. Hackbusch. *Elliptic differential equations: theory and numerical treatment. Transl. from the German by Regine Fadiman and Patrick D. F. Ion.* Springer Series in Computational Mathematics. 18. Berlin: Springer- Verlag, 1992.

[48] J. Jost. *Partial differential equations. Expanded translation of the original German version.* Graduate Texts in Mathematics 214. New York, NY: Springer, 2002.

[49] A. F. Karr. *Probability.* Springer Texts in Statistics. New York, NY: Springer-Verlag, 1993.

[50] V. Kokilashvili and M. Krbec. *Weighted inequalities in Lorentz and Orlicz spaces.* Singapore etc.: World Scientific Publishing Co. Pte. Ltd., 1991.

[51] M. A. Krasnosel'skij and Ya. B. Rutitskij. *Convex functions and Orlicz spaces.* Groningen-The Netherlands: P. Noordhoff Ltd., 1961.

[52] W. Liu and N. Yan. Quasi-norm local error estimators for p-Laplacian. *SIAM J. Numer. Anal.*, 39(1):100–127, 2001.

[53] W. Liu and N. Yan. On quasi-norm interpolation error estimation And A posteriori error estimates for p-Laplacian. *SIAM J. Numer. Anal.*, 40(5):1870–1895, 2002.

[54] J. M. Maubach. Local bisection refinement for n-simplicial grids generated by reflection. *SIAM J. Sci. Comput.*, 16(1):210–227, 1995.

[55] K. Mekchay and R. H. Nochetto. Convergence of adaptive finite element methods for general second order linear elliptic PDEs. *SIAM J. Numer. Anal.*, 43(5):1803–1827, 2005.

[56] W. F. Mitchell. A comparison of adaptive refinement techniques for elliptic problems. *ACM Trans. Math. Softw.*, 15(4):326–347, 1989.

[57] P. Morin, R. H. Nochetto, and K. G. Siebert. Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.*, 38(2):466–488, 2000.

[58] P. Morin, R. H. Nochetto, and K. G. Siebert. Convergence of adaptive finite element methods. *SIAM Rev.*, 44(4):631–658, 2002.

[59] P. Morin, R. H. Nochetto, and K. G. Siebert. Local problems on stars: A posteriori error estimators, convergence, and performance. *Math. Comput.*, 72(243):1067–1097, 2003.

[60] P. Morin, K. G. Siebert, and A. Veeser. Convergence of finite elements adapted for weaker norms. *V. Cutello, G. Fotia, and L. Puccio (Eds.): Applied and Industrial Matematics in Italy - II, Selected Contributions from the 8th SIMAI Conference*, 08:468–479.

[61] P. Morin, K. G. Siebert, and A. Veeser. A basic convergence result for conforming adaptive finite elements. *Math. Models Methods Applications, to appear*, 18, 2008.

[62] P. P. Mosolov and V. P. Myasnikov. A proof of Korn's inequality. *Sov. Math., Dokl.*, 12:1618–1622, 1971.

[63] J. Musielak. *Orlicz spaces and modular spaces.* Lecture Notes in Mathematics. 1034. Berlin etc.: Springer-Verlag, 1983.

[64] R. H. Nochetto and J.-H. Pyo. Optimal relaxation parameter for the Uzawa method. *Numer. Math.*, 98(4):695–702, 2004.

[65] J.-H. Pyo. *The Gauge Uzawa and Related Projection Finite Element Methods for the Evolution Navier Stokes Equation.* Dissertation, University of Maryland, College Park, 2002.

[66] M. M. Rao and Z. D. Ren. *Theory of Orlicz spaces.* Pure and Applied Mathematics, 146. New York etc.: Marcel Dekker, Inc., 1991.

[67] A. Schmidt and K. G. Siebert. *Design of adaptive finite element software. The finite element toolbox ALBERTA. With CD-ROM.* Lecture Notes in Computational Science and Engineering 42. Berlin: Springer, 2005.

[68] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comput.*, 54(190):483–493, 1990.

[69] R. E. Showalter. *Monotone operators in Banach space and nonlinear partial differential equations.* Mathematical Surveys and Monographs. 49. Providence, RI: American Mathematical Society, 1997.

[70] K. G. Siebert. A Convergence Proof for Adaptive Finite Elements without Lower Bound. *In preparation.*

[71] R. Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007.

[72] R. Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comput.*, 77(261):227–241, 2008.

[73] R. Temam. *Navier-Stokes equations. Theory and numerical analysis. 3rd (rev.) ed.* Studies in Mathematics and its Applications, Vol. 2. Amsterdam-New York- Oxford: North-Holland, 1984.

[74] A. Veeser. Convergent adaptive finite elements for the nonlinear Laplacian. *Numer. Math.*, 92(4):743–770, 2002.

[75] R. Verfürth. A posteriori error estimators for the Stokes equations. *Numer. Math.*, 55(3):309–325, 1989.

[76] R. Verfürth. A posteriori error estimation and adaptive mesh-refinement techniques. *J. Comput. Appl. Math.*, 50(1-3):67–83, 1994.

[77] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques.* Wiley-Teubner Series Advances in Numerical Mathematics. Chichester: John Wiley; Sons. Stuttgart: B. G. Teubner, 1996.

[78] K. Yosida. *Functional analysis. Repr. of the 6th ed.* Berlin: Springer-Verlag, 1994.

[79] E. Zeidler. *Nonlinear functional analysis and its applications. III: Variational methods and optimization.* New York etc.: Springer-Verlag, 1985.

[80] E. Zeidler. *Nonlinear functional analysis and its applications. I: Fixed-point theorems.* New York etc.: Springer-Verlag, 1986.

[81] E. Zeidler. *Nonlinear functional analysis and its applications. II/B: Nonlinear monotone operators.* New York etc.: Springer-Verlag, 1990.

Appendix B

Notation Index

$\mathbb{N}$	set of nonzero natural numbers
$\mathbb{N}_0$	set of natural numbers with zero: $\mathbb{N} \cup \{0\}$
$\mathbb{R}$	set of real numbers
$\overline{A}$	closure of the set $A \subset \mathbb{R}^m$, $m \in \mathbb{N}$
∂A	boundary set of the set $A \subset \mathbb{R}^m$, $m \in \mathbb{N}$
$B \subset\subset A$	set $B \subset \mathbb{R}^m$ is a compact subset of the set $A \subset \mathbb{R}^m$, $m \in \mathbb{N}$
$\|A\|$	m-dimensional Hausdorff measure of the set $A \subset \mathbb{R}^m$, $m \in \mathbb{N}$
$\|\xi\|, \|\mathbf{Q}\|$	Euclidean norm of $\xi \in \mathbb{R}^m$ and $\mathbf{Q} \in \mathbb{R}^{m \times m}$, $m \in \mathbb{N}$, respectively
$\xi^t, \mathbf{Q}^t$	transposed $\xi \in \mathbb{R}^m$ and $\mathbf{Q} \in \mathbb{R}^{m \times m}$, $m \in \mathbb{N}$, respectively
$\Delta_2(\phi)$	Δ_2-constant of the N-function ϕ
ϕ_a	N-function ϕ with shift $a \geq 0$
$\operatorname{supp}(f)$	support of a function f
D_i	partial derivative with respect to the i-th variable
∇v	gradient of a function v
$\operatorname{div} v$	divergence of a function v
$\mathbf{E}(v)$	symmetric gradient of a function v defined as $\mathbf{E}(v) = \frac{1}{2}(\nabla v + \nabla v^t)$
$(X, \\|\cdot\\|_X)$	pair of Banach space X and corresponding norm $\\|\cdot\\|_X$
$\langle f, g \rangle_{X^* \times X}$	dual pairing of $f \in X^*$ with $g \in X$ defined by $f(g)$
$C_0^\infty(\Omega)$	set of test-functions on a set $\Omega \subset \mathbb{R}^d$
$L^r(\Omega)$	space of r-integrable Lebesgue functions over $\Omega \subset \mathbb{R}^d$
$W_0^{k,r}(\Omega)$	Sobolev space of functions with zero boundary values and weak derivatives up to order k in $L^r(\Omega)$
$(L^\phi(\Omega), \\|\cdot\\|_\phi)$	Orlicz space corresponding to the N-function ϕ with norm $\\|\cdot\\|_\phi$
$\\|\cdot\\|_{(\phi)}$	Luxemburg norm on $L^\phi(\Omega)$

$W_0^{k,\phi}(\Omega)$	Orlicz Sobolev space of functions with zero boundary values and weak derivatives up to order k in $L^\phi(\Omega)$
$W_0^{k,\phi}(\Omega)^d$	Orlicz Sobolev space of d-dimensional vector valued functions with each component function in $W_0^{k,\phi}(\Omega)$
$\mathbb{V}$	velocity space defined as $W_0^{1,\phi}(\Omega)^d$
$\mathcal{J}$	energy functional of the nonlinear Poisson equation, $\mathcal{J} : \mathbb{V} \to \mathbb{R}$
$\mathbb{Q}$	pressure space defined as $L^{\phi^*}(\Omega)/\mathbb{R}$
$\mathcal{L}$	Lagrange function of the nonlinear stationary Stokes problem, $\mathcal{L} : \mathbb{V} \times \mathbb{Q} \to \mathbb{R}$
$\mathcal{F}$	functional defined as $\mathcal{F}(q) := -\inf_{v\in\mathbb{V}} \mathcal{L}(v, q)$, $q \in \mathbb{Q}$
$D\mathcal{J}, D\mathcal{F}$	Fréchet derivative of the functional $\mathcal{J}$ and $\mathcal{F}$ respectively
$\mathcal{T}, \mathcal{N}, \mathring{\mathcal{N}}, \mathcal{S}, \mathring{\mathcal{S}}$	conforming triangulation of the polyhedral domain $\Omega \subset \mathbb{R}^d$ and corresponding sets of nodes $\mathcal{N}$, interior nodes $\mathring{\mathcal{N}}$, sides $\mathcal{S}$, and interior sides $\mathring{\mathcal{S}}$
$\sigma(\mathcal{T})$	shape-regularity of $\mathcal{T}$
$\mathcal{T}^* \geq \mathcal{T}$	the conforming triangulation $\mathcal{T}^*$ is a refinement of $\mathcal{T}$
$\mathcal{T}(A)$	sub-triangulation of elements $T \in \mathcal{T}$ with $T \subset \overline{A}$, $A \subset \mathbb{R}^d$
h_T	mesh-size of a simplex $T \in \mathcal{T}$
$\hat{T}$	reference simplex
S_T	patch of a simplex $T \in \mathcal{T}$
ω_σ	union of simplices adjacent to $\sigma \in \mathcal{S}$
ω_T	union of simplices adjacent to $T \in \mathcal{T}$
ω_z	finite element star of the node $z \in \mathcal{N}$
$\mathbb{Q}_D(\mathcal{T})$	space of piecewise constant functions over $\mathcal{T}$
$\mathbb{Q}(\mathcal{T})$	discrete pressure space defined as space of piecewise linear continuous functions over $\mathcal{T}$
$\mathbb{V}(\mathcal{T})$	space of d-dimensional vector-valued piecewise linear continuous functions over $\mathcal{T}$
$\mathring{\mathbb{V}}(\mathcal{T})$	discrete velocity space defined as the subspace of $\mathbb{V}(\mathcal{T})$ of the functions with zero boundary values
$[\![\mathbf{G}]\!]$	jump of a function $\mathbf{G}$ across inter-element sides $\sigma \in \mathcal{S}$
$\eta(v, W, T, g)$	residual based a posteriori error estimator for the nonlinear Poisson equation
$\operatorname{osc}(v, T, g)$	oscillation related to the estimator $\eta(v, W, T, g)$
$\mathfrak{d}_q$	quasi-steepest descent direction of $\mathcal{F}$ in $q \in \mathbb{Q}$
$\mathfrak{D}_Q$	approximation of the quasi-steepest direction $\mathfrak{d}_Q$ of $\mathcal{F}$ in $Q \in \mathbb{Q}(\mathcal{T})$

Lebenslauf

Persönliche Daten

Name:	Christian Kreuzer
Geburtsort:	Augsburg
Geburtstag:	06.04.1978
Nationalität:	Deutsch
Familienstand:	Ledig

Ausbildung

10/1999 - 04/2002	Grundstudium Mathematik an der Universität Augsburg
05/2002 - 02/2005	Hauptstudium Mathematik an der Universität Augsburg, Abschluss mit dem Diplom am 25.02.2005 Thema der Diplomarbeit: “Globale Zweige schwacher Lösungen elliptischer Systeme über Gebieten mit nichtglattem Rand”
03/2005 - 07/2008	Promotion am Lehrstuhl für angewandte Analysis mit Schwerpunkt Numerische Mathematik der Universität Augsburg bei Prof. Dr. K. G. Siebert

Wissenschaftliche Arbeiten

1. C. Kreuzer, *Globale Zweige schwacher Lösungen elliptischer Systeme über Gebieten mit nichtglattem Rand*, Diplomarbeit, Institut für Mathematik, Universität Augsburg, 2005
2. L. Diening und C. Kreuzer, *Linear convergence of an adaptive finite element method for the p-Laplacian equation*, SIAM J. Numer. Anal., 46(2): 614-638, 2008
3. J. M. Cascon, C. Kreuzer, R.H. Nochetto und K. G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, Preprint Universität Augsburg 2007/9 - erscheint in SIAM J. Numer. Anal.
4. L. Diening und C. Kreuzer, *Quasi-optimal convergence rate of an adaptive finite element method for the nonlinear Laplacian*, in Vorbereitung