Calcolo (2022) 59:4
https://doi.org/10.1007/s10092-021-00444-3

On the convergence rate of the Kačanov scheme for shear-thinning fluids

Pascal Heid[1] · **Endre Süli**[1]

Received: 5 January 2021 / Revised: 13 October 2021 / Accepted: 14 October 2021 /
Published online: 28 November 2021

Abstract
We explore the convergence rate of the Kačanov iteration scheme for different models of shear-thinning fluids, including Carreau and power-law type explicit quasi-Newtonian constitutive laws. It is shown that the energy difference contracts along the sequence generated by the iteration. In addition, an a posteriori computable contraction factor is proposed, which improves, on finite-dimensional Galerkin spaces, previously derived bounds on the contraction factor in the context of the power-law model. Significantly, this factor is shown to be independent of the choice of the cut-off parameters whose use was proposed in the literature for the Kačanov iteration applied to the power-law model. Our analytical findings are confirmed by a series of numerical experiments.

Keywords Non-Newtonian fluids · Kačanov's method · Energy contraction · Carreau model · Power-law model

Mathematics Subject Classification 65J15 · 35Q35 · 35J62

PH acknowledges the financial support of the Swiss National Science Foundation (SNF), Project No. P2BEP2_191760.

✉ Pascal Heid
pascal.heid@maths.ox.ac.uk

Endre Süli
endre.suli@maths.ox.ac.uk

[1] Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, UK

1 Introduction

In this work, we focus on the iterative solution of nonlinear partial differential equations that arise in models of steady flows of incompressible shear-thinning fluids, including models with explicit constitutive relations of Carreau and power-law type. In particular, we consider the following quasi-Newtonian fluid flow problem: find $(\mathbf{u}, p)$ such that

$$\begin{aligned} -\nabla \cdot \{\mu(\boldsymbol{x}, |\underline{e}(\mathbf{u})|^2)\underline{e}(\mathbf{u})\} + \nabla p &= \mathbf{f} \quad \text{in } \Omega, \\ \nabla \cdot \mathbf{u} &= 0 \quad \text{in } \Omega, \\ \mathbf{u} &= \mathbf{0} \quad \text{on } \partial\Omega, \end{aligned} \tag{1}$$

where $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, is a bounded Lipschitz domain, the source term $\mathbf{f} \in \mathrm{L}^2(\Omega)^d$ is a given external force, $\mathbf{u}$ is the velocity vector, p denotes the pressure, and $\underline{e}(\mathbf{u})$ is the $d \times d$ rate-of-strain tensor defined by

$$e_{ij}(\mathbf{u}) := \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \qquad i, j = 1, \ldots, d.$$

here $|\underline{e}(\mathbf{u})|$ denotes the Frobenius norm of $\underline{e}(\mathbf{u})$, and the (real-valued) viscosity coefficient μ is assumed to satisfy the following structural assumptions:

(A1) $\mu \in C(\overline{\Omega} \times \mathbb{R}_{\geq 0})$ and it is differentiable in the second variable;
(A2) There exist constants $m_\mu, M_\mu > 0$ such that

$$m_\mu(t - s) \leq \mu(\boldsymbol{x}, t^2)t - \mu(\boldsymbol{x}, s^2)s \leq M_\mu(t - s), \qquad t \geq s \geq 0, \quad \boldsymbol{x} \in \overline{\Omega}; \tag{2}$$

(A3) μ is decreasing in the second variable, i.e., $\mu'(\boldsymbol{x}, t) \leq 0$ for all $t \geq 0$ and all $\boldsymbol{x} \in \overline{\Omega}$, where μ' denotes the derivative of μ with respect to the variable t.

The assumption (A3) asserts that the viscosity decreases with increasing strain rate, in line with our assumption that the fluid under consideration is shear-thinning. Moreover, (A2) immediately implies that μ is bounded from above and below; indeed, by setting $s = 0$, we obtain

$$m_\mu \leq \mu(\boldsymbol{x}, t) \leq M_\mu \qquad \text{for all } \boldsymbol{x} \in \overline{\Omega},\ t \geq 0. \tag{3}$$

The bounds m_μ and M_μ are, in general, closely related to the infinite and zero shear viscosity plateau, respectively. In the sequel, the dependence of μ on $\boldsymbol{x} \in \Omega$ will be suppressed.

Upon defining $V := \{\mathbf{u} \in \mathrm{H}^1_0(\Omega)^d : \nabla \cdot \mathbf{u} = 0\}$, the weak formulation of (1) is as follows:

$$\text{find } \mathbf{u} \in V \text{ such that} \qquad \int_\Omega \mu(|\underline{e}(\mathbf{u})|^2)\underline{e}(\mathbf{u}) : \underline{e}(\mathbf{v})\,\mathrm{d}\boldsymbol{x} = \int_\Omega \mathbf{f} \cdot \mathbf{v}\,\mathrm{d}\boldsymbol{x} \qquad \text{for all } \mathbf{v} \in V, \tag{4}$$

where $\underline{e}(\mathbf{u}) : \underline{e}(\mathbf{v})$ denotes the Frobenius inner product of $\underline{e}(\mathbf{u})$ and $\underline{e}(\mathbf{v})$; we refer to [2] Sect. 2 for more details concerning the weak formulation (4). The space V is endowed with the inner product

$$(\mathbf{u}, \mathbf{v})_V = \int_\Omega \underline{e}(\mathbf{u}) : \underline{e}(\mathbf{v}) \, \mathrm{d}\boldsymbol{x}, \qquad \mathbf{u}, \mathbf{v} \in V, \tag{5}$$

and the induced norm $|||\mathbf{u}|||_\Omega^2 = (\mathbf{u}, \mathbf{u})_V$, $\mathbf{u} \in V$. We emphasize that

$$\frac{1}{2} \sum_{i=1}^{d} \int_\Omega |\nabla u_i|^2 \, \mathrm{d}\boldsymbol{x} \le \int_\Omega |\underline{e}(\mathbf{u})|^2 \, \mathrm{d}\boldsymbol{x} \le \sum_{i=1}^{d} \int_\Omega |\nabla u_i|^2 \, \mathrm{d}\boldsymbol{x} \qquad \text{for all } \mathbf{u} \in V,$$

i.e., the norm $|||\cdot|||_\Omega$ is equivalent to the standard norm on $\mathrm{W}^1_0(\Omega)^d$; the first inequality is a special case of Korn's inequality (see, e.g., inequality (1.7) in [17]), while the second can be easily verified by invoking the Cauchy–Schwarz inequality. In particular, V endowed with the inner product of (5) and induced norm $|||\cdot|||_\Omega$ is a Hilbert space.

The weak form (4) of the boundary-value problem under consideration is known to have a unique solution $\mathbf{u}^\star \in V$, which will be shown, nonetheless, in Sect. 2; moreover, this element $\mathbf{u}^\star \in V$ is the unique minimiser of the energy functional

$$\mathsf{E}(\mathbf{u}) := \int_\Omega \varphi\Big(\big|\underline{e}(\mathbf{u})\big|^2\Big) \, \mathrm{d}\boldsymbol{x} - \int_\Omega \mathbf{f} \cdot \mathbf{u} \, \mathrm{d}\boldsymbol{x}, \qquad \mathbf{u} \in V, \tag{6}$$

where

$$\varphi(s) := \frac{1}{2} \int_0^s \mu(t) \, \mathrm{d}t.$$

Indeed, a straightforward calculation reveals that, for a given $\mathbf{u} \in V$,

$$\mathsf{E}'(\mathbf{u})(\mathbf{v}) = \int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u})\big|^2\Big) \underline{e}(\mathbf{u}) : \underline{e}(\mathbf{v}) \, \mathrm{d}\boldsymbol{x} - \int_\Omega \mathbf{f} \cdot \mathbf{v} \, \mathrm{d}\boldsymbol{x}, \qquad \mathbf{v} \in V, \tag{7}$$

where E' denotes the Gâteaux derivative; we refer to [1, Prop. 2.1] for details. In particular, the weak formulation (4) is the Euler–Lagrange equation for the minimisation of E over V.

A prominent iterative solver for the nonlinear problem (4) is Kačanov's scheme, which, in simple terms, fixes the nonlinearity at the previous iterate: for a given $\mathbf{u}^n \in V$ find $\mathbf{u}^{n+1} \in V$ such that

$$\int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big) \underline{e}(\mathbf{u}^{n+1}) : \underline{e}(\mathbf{v}) \, \mathrm{d}\boldsymbol{x} = \int_\Omega \mathbf{f} \cdot \mathbf{v} \, \mathrm{d}\boldsymbol{x} \qquad \text{for all } \mathbf{v} \in V, \quad n = 0, 1, \dots, \tag{8}$$

where $\mathbf{u}^0 \in V$ is an arbitrary initial guess. Early references concerning this iterative method include [14], where it was used to compute a stationary magnetic field in nonlinear media, and [5], where the convergence of the Kačanov iteration was investigated in the context of Galerkin methods; Fučík, Kratochvíl and Nečas point in

their work [5] to pages 369–370 of Michlin's 1966 monograph [15] for a description of the iterative method introduced by Kačanov in [13], in the context of variational methods for plasticity problems. Kačanov's iteration scheme has been, by now, carefully examined; see, e.g., the monographs [16, Sect. 4.5] and [21, Sect. 25.14], or the papers [7, 8, 10]. More recently, it was shown in the articles [9] and [4] that the energy E from (6) contracts along the sequence generated by the Kačanov iteration (8). Indeed, the first of these two papers established the energy contraction for a more general iteration scheme, and the latter focuses on the Kačanov scheme for a 'relaxed p-Poisson problem' involving a truncation of the nonlinearity from below and from above using a pair of positive cut-off parameters ε_- and ε_+. The derived upper bound on the contraction factor depends on the quotient m_μ/M_μ involving ε_- and ε_+, and may be extremely close to 1 in certain situations; interestingly, this unfavourable predicted dependence of the contraction factor on the ratio m_μ/M_μ has not been observed in numerical experiments. It is this mismatch between the observed behaviour of the method and the rather more pessimistic results of the analysis reported in the literature that motivated the work outlined herein.

We will establish an improved upper bound on the contraction factor of the Kačanov iteration for a general class of shear-thinning fluids. The resulting bound will then be further examined for fluids obeying either the Carreau law or a relaxed power-law, to be specified in the lines below. It will be shown that for (finite-dimensional) Galerkin approximations of the relaxed power-law model it is the power-law exponent, rather than the ratio m_μ/M_μ, that is responsible for the rate of convergence of the iteration. Specifically, we will show that the contraction factor of the iteration on finite-dimensional spaces is independent of the choice of the lower and upper cut-off parameters featuring in the so-called relaxed Kačanov iteration, where a truncation of the power-law nonlinearity from below and above is carried out by means of these two positive truncation parameters. To the best of our knowledge the proof of such a result was an open question in the literature.

The paper is structured as follows. In Sect. 2 we will show that the weak formulation (4) of the problem under consideration has a unique solution, which, in turn, is the unique minimiser of E in V. The proof is based on auxiliary results, which will also be decisive for the derivation of the contraction factor in Sect. 3. In Sect. 4 we will perform a series of numerical experiments, which confirm our theoretical results. The paper closes with concluding remarks recorded in Sect. 5.

2 Existence and uniqueness of the solution

We will show in this section that the weak formulation (4) has a unique solution. To this end we define, for given $\mathbf{u} \in V$, the linear operator $\mathsf{A}[\mathbf{u}] : V \to V^\star$, where $V^\star$ denotes the dual space of V, by

$$\mathsf{A}[\mathbf{u}](\mathbf{v})(\mathbf{w}) := \int_\Omega \mu(|\underline{e}(\mathbf{u})|^2)\underline{e}(\mathbf{v}) : \underline{e}(\mathbf{w})\,\mathrm{d}\boldsymbol{x}, \qquad \mathbf{v}, \mathbf{w} \in V, \tag{9}$$

and the linear form $\ell \in V^\star$ by

$$\ell(\mathbf{w}) := \int_\Omega \mathbf{f} \cdot \mathbf{w} \, \mathrm{d}\boldsymbol{x}, \qquad \mathbf{w} \in V.$$

In terms of these, the weak formulation (4) can be restated in the following equivalent form:

$$\text{find } \mathbf{u} \in V \text{ such that} \qquad \mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{v}) = \ell(\mathbf{v}) \qquad \text{for all } \mathbf{v} \in V, \tag{10}$$

and the Kačanov iteration (8) takes the form: given $\mathbf{u}^0 \in V$,

$$\text{find } \mathbf{u}^{n+1} \in V \text{ such that} \quad \mathsf{A}[\mathbf{u}^n](\mathbf{u}^{n+1})(\mathbf{v}) = \ell(\mathbf{v}) \qquad \text{for all } \mathbf{v} \in V, \quad n = 0, 1, \dots. \tag{11}$$

By (7) and the definitions of A and ℓ we further have that

$$\mathsf{E}'(\mathbf{u}) = \mathsf{A}[\mathbf{u}](\mathbf{u}) - \ell. \tag{12}$$

We will now show that the operator $\mathbf{u} \mapsto \mathsf{A}[\mathbf{u}](\mathbf{u})$ is Lipschitz continuous and strongly monotone, since, in that case, the theory of strongly monotone operators implies that the weak equation (10) has a unique solution $\mathbf{u}^\star \in V$; see, e.g., [16, Sect. 3.3] or [21, Sect. 25.4]. For the proof of Lipschitz continuity and strong monotonicity of the operator $\mathbf{u} \mapsto \mathsf{A}[\mathbf{u}](\mathbf{u})$ we require the following result, which, as well as its proof, is largely borrowed from [2, Lemma 3.1]. However, we place emphasis on sharp bounds, since these will be crucial for our convergence analysis below, leading to the improved factors appearing in (15) and (16).

Lemma 2.1 *Let μ satisfy the assumptions (A1)–(A3) and define $\xi(t) := \mu(t^2)t$, $t \geq 0$. Then, for any $\underline{\kappa}, \underline{\tau} \in \mathbb{R}^{d\times d}$, the following inequalities hold*:

$$\left|\mu(|\underline{\kappa}|^2)\underline{\kappa} - \mu(|\underline{\tau}|^2)\underline{\tau}\right|^2 \leq C(\underline{\kappa}, \underline{\tau})|\underline{\kappa} - \underline{\tau}|^2 \leq 3M_\mu^2|\underline{\kappa} - \underline{\tau}|^2 \tag{13}$$

and

$$(\mu(|\underline{\kappa}|^2)\underline{\kappa} - \mu(|\underline{\tau}|^2)\underline{\tau}) : (\underline{\kappa} - \underline{\tau}) \geq c(\underline{\kappa}, \underline{\tau})|\underline{\kappa} - \underline{\tau}|^2 \geq m_\mu|\underline{\kappa} - \underline{\tau}|^2, \tag{14}$$

where

$$C(\underline{\kappa}, \underline{\tau}) := \left(\sup_{t\in(0,1)} \xi'(|\underline{\kappa}| + t(|\underline{\tau}| - |\underline{\kappa}|))\right)^2 + 2\mu(|\underline{\kappa}|^2)\mu(|\underline{\tau}|^2), \tag{15}$$

$$c(\underline{\kappa}, \underline{\tau}) := \inf_{t\in(0,1)} \xi'(|\underline{\kappa}| + t(|\underline{\tau}| - |\underline{\kappa}|)). \tag{16}$$

Proof We will only prove (14), since the contraction factor will strongly rely on this bound, but not on (13). For the proof of the latter, we refer to [2].

A simple and straightforward calculation reveals that

$$(\mu(|\underline{\kappa}|^2)\underline{\kappa} - \mu(|\underline{\tau}|^2)\underline{\tau}) : (\underline{\kappa} - \underline{\tau}) = \mu(|\underline{\kappa}|^2)|\underline{\kappa}|^2 + \mu(|\underline{\tau}|^2)|\underline{\tau}|^2 - \mu(|\underline{\kappa}|^2)\underline{\kappa} : \underline{\tau} - \mu(|\underline{\tau}|^2)\underline{\tau} : \underline{\kappa}$$
$$= \big(\mu(|\underline{\kappa}|^2)|\underline{\kappa}| - \mu(|\underline{\tau}|^2)|\underline{\tau}|\big)(|\underline{\kappa}| - |\underline{\tau}|) \tag{17}$$
$$+ (\mu(|\underline{\kappa}|^2) + \mu(|\underline{\tau}|^2))(|\underline{\kappa}||\underline{\tau}| - \underline{\kappa} : \underline{\tau}). \tag{18}$$

We note that the summand in (17) can be written as $(\xi(|\underline{\kappa}|) - \xi(|\underline{\tau}|))(|\underline{\kappa}| - |\underline{\tau}|)$, since $\xi(t) = \mu(t^2)t$ for $t \geq 0$. Then, the mean value theorem implies that

$$(\xi(|\underline{\kappa}|) - \xi(|\underline{\tau}|))(|\underline{\kappa}| - |\underline{\tau}|) \geq \inf_{t\in(0,1)} \xi'(|\underline{\kappa}| + t(|\underline{\tau}| - |\underline{\kappa}|))(|\underline{\kappa}| - |\underline{\tau}|)^2 = c(\underline{\kappa}, \underline{\tau})(|\underline{\kappa}| - |\underline{\tau}|)^2. \tag{19}$$

Furthermore, since $\mu'(t) \leq 0$ for all $t \geq 0$ by (A3), and, in turn, $\xi'(t) = \mu(t^2) + 2t^2\mu'(t^2) \leq \mu(t^2)$, we find that

$$c(\underline{\kappa}, \underline{\tau}) = \inf_{t\in(0,1)} \xi'(|\underline{\kappa}| + t(|\underline{\tau}| - |\underline{\kappa}|)) \leq \min\left\{\mu\Big(\big|\underline{\kappa}\big|^2\Big), \mu\Big(\big|\underline{\tau}\big|^2\Big)\right\}.$$

Consequently, the summand in (18) can be bounded from below as follows

$$\left(\mu\Big(\big|\underline{\kappa}\big|^2\Big) + \mu\Big(\big|\underline{\tau}\big|^2\Big)\right)(|\underline{\kappa}||\underline{\tau}| - \underline{\kappa} : \underline{\tau}) \geq 2c(\underline{\kappa}, \underline{\tau})(|\underline{\kappa}||\underline{\tau}| - \underline{\kappa} : \underline{\tau}), \tag{20}$$

since $|\underline{\kappa}||\underline{\tau}| - \underline{\kappa} : \underline{\tau} \geq 0$ by the Cauchy–Schwarz inequality. Hence, using the established bounds (19) and (20) for the summands in (17) and (18), respectively, yields

$$\left(\mu\Big(\big|\underline{\kappa}\big|^2\Big)\underline{\kappa} - \mu\Big(\big|\underline{\tau}\big|^2\Big)\underline{\tau}\right) : (\underline{\kappa} - \underline{\tau}) \geq c(\underline{\kappa}, \underline{\tau})\big((|\underline{\kappa}| - |\underline{\tau}|)^2 + 2\big(|\underline{\kappa}||\underline{\tau}| - \underline{\kappa} : \underline{\tau}\big)\big)$$
$$= c(\underline{\kappa}, \underline{\tau})\big(|\underline{\kappa}|^2 + |\underline{\tau}|^2 - 2\underline{\kappa} : \underline{\tau}\big)$$
$$= c(\underline{\kappa}, \underline{\tau})|\underline{\kappa} - \underline{\tau}|^2.$$

Finally we note that dividing (2) by $(t - s)$, for $t > s$, and taking the limit $s \to t$ yields that $m_\mu \leq \xi'(t) \leq M_\mu$ for all $t \geq 0$, i.e., $c(\underline{\kappa}, \underline{\tau}) \geq m_\mu$. □

Now we are ready to show that the operator A, cf. (9), is strongly monotone and Lipschitz continuous, which implies the unique solvability of (4).

Proposition 2.2 *Let* A *be defined as in* (9), *with* μ *satisfying (A1)–(A3).*

(a) *For given* $\mathbf{u} \in V$, $\mathsf{A}[\mathbf{u}](\cdot)(\cdot)$ *is a uniformly bounded and coercive, symmetric bilinear form on* $V \times V$. *In particular, the following inequalities hold*:

$$\mathsf{A}[\mathbf{u}](\mathbf{v})(\mathbf{w}) \leq M_\mu |||\mathbf{v}|||_\Omega |||\mathbf{w}|||_\Omega$$

and

$$\mathsf{A}[\mathbf{u}](\mathbf{v})(\mathbf{v}) \geq m_\mu |||\mathbf{v}|||_\Omega^2 \tag{21}$$

for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$.

(b) *The mapping* $\mathbf{u} \mapsto \mathsf{A}[\mathbf{u}](\mathbf{u})$ *is Lipschitz continuous with*

$$\mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{w}) - \mathsf{A}[\mathbf{v}](\mathbf{v})(\mathbf{w}) \le \sqrt{3} M_\mu |||\mathbf{u}-\mathbf{v}|||_\Omega |||\mathbf{w}|||_\Omega, \qquad \mathbf{u},\mathbf{v},\mathbf{w} \in V \tag{22}$$

and strongly monotone with

$$\mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{u}-\mathbf{v}) - \mathsf{A}[\mathbf{v}](\mathbf{v})(\mathbf{u}-\mathbf{v}) \ge m_\mu |||\mathbf{u}-\mathbf{v}|||_\Omega^2, \qquad \mathbf{u},\mathbf{v} \in V. \tag{23}$$

Consequently, the problem (4) *has a unique solution* $\mathbf{u}^\star \in V$.

Proof Ad (a): By invoking the definition of A, cf. (9), the boundedness of the viscosity coefficient μ, cf. (3), and applying the Cauchy–Schwarz inequality twice, first for the Frobenius inner product and subsequently for the $\mathrm{L}^2(\Omega)$-inner product, we obtain

$$\begin{aligned}
\mathsf{A}[\mathbf{u}](\mathbf{v})(\mathbf{w}) &= \int_\Omega \mu(|\underline{e}(\mathbf{u})|^2)\underline{e}(\mathbf{v}) : \underline{e}(\mathbf{w})\,\mathrm{d}\boldsymbol{x} \\
&\le M_\mu \int_\Omega |\underline{e}(\mathbf{v}) : \underline{e}(\mathbf{w})|\,\mathrm{d}\boldsymbol{x} \\
&\le M_\mu \left(\int_\Omega |\underline{e}(\mathbf{v})|^2\,\mathrm{d}\boldsymbol{x}\right)^{1/2} \left(\int_\Omega |\underline{e}(\mathbf{w})|^2\,\mathrm{d}\boldsymbol{x}\right)^{1/2} \\
&= M_\mu |||\mathbf{v}|||_\Omega |||\mathbf{w}|||_\Omega.
\end{aligned}$$

Similarly, the inequality (3) implies the uniform coercivity (21).

Ad (b): The definition of A, cf. (9), and the Cauchy–Schwarz inequality yield

$$\begin{aligned}
\mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{w}) - \mathsf{A}[\mathbf{v}](\mathbf{v})(\mathbf{w}) &= \int_\Omega \left[\mu(|\underline{e}(\mathbf{u})|^2)\underline{e}(\mathbf{u}) - \mu(|\underline{e}(\mathbf{v})|^2)\underline{e}(\mathbf{v})\right] : \underline{e}(\mathbf{w})\,\mathrm{d}\boldsymbol{x} \\
&\le \int_\Omega |\mu(|\underline{e}(\mathbf{u})|^2)\underline{e}(\mathbf{u}) - \mu(|\underline{e}(\mathbf{v})|^2)\underline{e}(\mathbf{v})||\underline{e}(\mathbf{w})|\,\mathrm{d}\boldsymbol{x}.
\end{aligned}$$

Hence, by (13) and the linearity of $\underline{e}(\cdot)$, this leads to

$$\mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{w}) - \mathsf{A}[\mathbf{v}](\mathbf{v})(\mathbf{w}) \le \sqrt{3} M_\mu \int_\Omega |\underline{e}(\mathbf{u}-\mathbf{v})||\underline{e}(\mathbf{w})|\,\mathrm{d}\boldsymbol{x}.$$

Applying once more the Cauchy–Schwarz inequality implies that

$$\begin{aligned}
\mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{w}) - \mathsf{A}[\mathbf{v}](\mathbf{v})(\mathbf{w}) &\le \sqrt{3} M_\mu \left(\int_\Omega |\underline{e}(\mathbf{u}-\mathbf{v})|^2\,\mathrm{d}\boldsymbol{x}\right)^{1/2} \left(\int_\Omega |\underline{e}(\mathbf{w})|^2\,\mathrm{d}\boldsymbol{x}\right)^{1/2} \\
&= \sqrt{3} M_\mu |||\mathbf{u}-\mathbf{v}|||_\Omega |||\mathbf{w}|||_\Omega.
\end{aligned}$$

Similarly, by the definition of A, cf. (9), and (14) we obtain

$$\begin{aligned}\mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{u}-\mathbf{v}) - \mathsf{A}[\mathbf{v}](\mathbf{v})(\mathbf{u}-\mathbf{v}) &= \int_\Omega \left(\mu(|\underline{e}(\mathbf{u})|^2)\underline{e}(\mathbf{u}) - \mu(|\underline{e}(\mathbf{v})|^2)\underline{e}(\mathbf{v})\right) : (\underline{e}(\mathbf{u}) - \underline{e}(\mathbf{v}))\,\mathrm{d}\boldsymbol{x} \\ &\geq m_\mu \int_\Omega |\underline{e}(\mathbf{u}-\mathbf{v})|^2\,\mathrm{d}\boldsymbol{x} \\ &= m_\mu |||\mathbf{u}-\mathbf{v}|||_\Omega^2.\end{aligned}$$

The existence and uniqueness of a solution to the Eq. (4) now follows from the theory of monotone operators, cf. [16, Sect. 3.3] or [21, Sect. 25.4]. □

Remark 2.3 Since (4) is the Euler–Lagrange equation of the minimisation problem

$$\min_{\mathbf{u}\in V} \mathsf{E}(\mathbf{u}),$$

the above proposition yields that $\mathbf{u}^\star \in V$ is the unique minimiser of the functional E; we emphasise that E is strictly convex thanks to the strong monotonicity (23), cf. [21, Prop. 25.10]. Moreover, the Kačanov scheme (11) is well defined thanks to Proposition 2.2 (a) and the Lax–Milgram theorem.

3 Energy contraction

In this section, we will show that the energy error, given by $\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star)$, contracts along the sequence $\{\mathbf{u}^n\}$ generated by the Kačanov iteration (11). To this end, we need an auxiliary result.

Lemma 3.1 *Let* A *and* E *be defined as in* (9) *and* (6), *respectively*, *with* μ *satisfying* (*A*1)–(*A*3). *Then*

$$\mathsf{E}(\mathbf{u}) - \mathsf{E}(\mathbf{v}) \geq \frac{1}{2}\mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{u}) - \frac{1}{2}\mathsf{A}[\mathbf{u}](\mathbf{v})(\mathbf{v}) - \ell(\mathbf{u}) + \ell(\mathbf{v}), \qquad \mathbf{u},\mathbf{v}\in V. \quad (24)$$

This result is well-known for the Kačanov iteration in the given setting, and the proof can be found, e.g., in [21, Sect. 25.12] or [16, Sect. 4.5]. However, as it is stated in a slightly different form in those references, and also for the sake of completeness, we will include the proof of this statement nonetheless.

Proof It can be shown that

$$\varphi(t) - \varphi(s) \geq \frac{1}{2}\mu(t)(t-s), \qquad t,s\geq 0,$$

see, e.g., [10, Sect. 5.1], and therefore

$$\int_\Omega \varphi\Big(\big|\underline{e}(\mathbf{u})\big|^2\Big) - \varphi\Big(\big|\underline{e}(\mathbf{v})\big|^2\Big)\,\mathrm{d}\boldsymbol{x} \geq \frac{1}{2}\int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u})\big|^2\Big)\Big(\big|\underline{e}(\mathbf{u})\big|^2 - \big|\underline{e}(\mathbf{v})\big|^2\Big)\,\mathrm{d}\boldsymbol{x}$$
$$= \frac{1}{2}(\mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{u}) - \mathsf{A}[\mathbf{u}](\mathbf{v})(\mathbf{v})),$$

for any $\mathbf{u}, \mathbf{v} \in V$. Hence, by the definition of E, cf. (6), we find that

$$\mathsf{E}(\mathbf{u}) - \mathsf{E}(\mathbf{v}) = \int_\Omega \varphi\Big(\big|\underline{e}(\mathbf{u})\big|^2\Big)\,\mathrm{d}\boldsymbol{x} - \int_\Omega \varphi\Big(\big|\underline{e}(\mathbf{v})\big|^2\Big)\,\mathrm{d}\boldsymbol{x} - \ell(\mathbf{u}) + \ell(\mathbf{v})$$
$$\geq \frac{1}{2}\mathsf{A}[\mathbf{u}](\mathbf{u})(\mathbf{u}) - \frac{1}{2}\mathsf{A}[\mathbf{u}](\mathbf{v})(\mathbf{v}) - \ell(\mathbf{u}) + \ell(\mathbf{v}),$$

which completes the proof of the claim. □

Now we are in a position to prove the contraction of the energy along the sequence generated by the Kačanov scheme (11). We note that similar results can be found, e.g., in [9, Thm. 2.1] or [4, Cor. 19].

Theorem 3.2 *Assume that (A1)–(A3) hold and let* E *be defined as in* (6)*. Then, the energy error contracts along the sequence* $\{\mathbf{u}^n\}$ *generated by the Kačanov iteration* (11) *in the sense that*

$$0 \leq \mathsf{E}(\mathbf{u}^{n+1}) - \mathsf{E}(\mathbf{u}^\star) \leq q(n)\big(\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star)\big),$$

where

$$q(n) := 1 - \frac{1}{4}\left\{\operatorname*{ess\,sup}_{x\in\Omega} \frac{\mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)}{\inf_{t\in(-1,1)} \xi'\big(\max\big\{0, \big|\underline{e}(\mathbf{u}^\star)\big| + t\big|\underline{e}(\mathbf{u}^n) - \underline{e}(\mathbf{u}^\star)\big|\big\}\big)}\right\}^{-1}, \tag{25}$$

and $\xi(t) = \mu(t^2)t$ *for* $t \geq 0$.

Proof We largely proceed along the lines of [4]. However, as we want to improve the contraction factor from that reference and, in addition, remove any unknown constants, some non-trivial modifications are necessary in the second part of the proof.

Let us define the real-valued function $\psi(t) := \mathsf{E}(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star))$, $t \in [0,1]$. Then, by invoking the fundamental theorem of calculus, we obtain

$$\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) = \int_0^1 \psi'(t)\,\mathrm{d}t.$$

We will first show that $\psi'(t)$, $t \in [0,1]$, is increasing. A straightforward calculation reveals that

$$\psi''(t) = \int_\Omega 2\mu'\Big(\big|\underline{e}(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star))\big|^2\Big)\big(\underline{e}(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star)) : \underline{e}(\mathbf{u}^n - \mathbf{u}^\star)\big)^2 \,\mathrm{d}\boldsymbol{x} + \int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star))\big|^2\Big)\big|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)\big|^2 \,\mathrm{d}\boldsymbol{x}.$$

By assumption (A3) we have that $\mu'(t) \le 0$ for $t \ge 0$, and thus, by the Cauchy–Schwarz inequality,

$$\begin{aligned}\psi''(t) &\ge \int_\Omega \Big(\mu\Big(\big|\underline{e}(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star))\big|^2\Big) + 2\mu'\Big(\big|\underline{e}(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star))\big|^2\Big)\big|\underline{e}(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star))\big|^2\Big) \\ &\quad \cdot |\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)|^2 \,\mathrm{d}\boldsymbol{x} \\ &= \int_\Omega \xi'(|\underline{e}(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star))|)|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)|^2 \,\mathrm{d}\boldsymbol{x},\end{aligned}$$

where $\xi(t) = \mu(t^2)t$ as before. Finally, since $\xi'(t) \ge m_\mu > 0$ for all $t \ge 0$, see the end of the proof of Lemma 2.1, we find that $\psi''(t) \ge 0$, i.e., $\psi'(t)$ is increasing. As a consequence, we immediately get that

$$\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) \le \psi'(1) = \mathsf{E}'(\mathbf{u}^n)(\mathbf{u}^n - \mathbf{u}^\star) = \mathsf{A}[\mathbf{u}^n](\mathbf{u}^n)(\mathbf{u}^n - \mathbf{u}^\star) - \ell(\mathbf{u}^n - \mathbf{u}^\star).$$

Moreover, by the definition of the Kačanov scheme (11), we have that

$$\mathsf{A}[\mathbf{u}^n](\mathbf{u}^{n+1})(\mathbf{u}^n - \mathbf{u}^\star) = \ell(\mathbf{u}^n - \mathbf{u}^\star).$$

Consequently, the above inequality becomes

$$\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) \le \mathsf{A}[\mathbf{u}^n](\mathbf{u}^n - \mathbf{u}^{n+1})(\mathbf{u}^n - \mathbf{u}^\star) = \int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)\underline{e}(\mathbf{u}^n - \mathbf{u}^{n+1}) : \underline{e}(\mathbf{u}^n - \mathbf{u}^\star)\,\mathrm{d}\boldsymbol{x}.$$

Let us recall that, for $a, b \ge 0$, $ab \le \frac{1}{2\gamma}a^2 + \frac{\gamma}{2}b^2$ for all $\gamma > 0$: Indeed, this holds true as the function $\gamma \mapsto \frac{1}{2\gamma}a^2 + \frac{\gamma}{2}b^2$ takes its minimum ab at $\gamma = a/b$. Consequently, we obtain

$$\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) \le \frac{1}{2\gamma}\int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)|\underline{e}(\mathbf{u}^n - \mathbf{u}^{n+1})|^2 \,\mathrm{d}\boldsymbol{x} + \frac{\gamma}{2}\int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)\big|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)\big|^2 \,\mathrm{d}\boldsymbol{x}. \tag{26}$$

We will now examine the two summands on the right-hand side above separately.

The first summand can be bounded from above in a similar manner as in the proof of [4, Thm. 18]. First, we note that

$$\begin{aligned}\int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)\big|\underline{e}(\mathbf{u}^n - \mathbf{u}^{n+1})\big|^2 \,\mathrm{d}\boldsymbol{x} &= \int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)\big|\underline{e}(\mathbf{u}^n)\big|^2 \,\mathrm{d}\boldsymbol{x} - \int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)\big|\underline{e}(\mathbf{u}^{n+1})\big|^2 \,\mathrm{d}\boldsymbol{x} \\ &\quad - 2\int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)\underline{e}(\mathbf{u}^{n+1}) : \underline{e}(\mathbf{u}^n)\,\mathrm{d}\boldsymbol{x} \\ &\quad + 2\int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)\underline{e}(\mathbf{u}^{n+1}) : \underline{e}(\mathbf{u}^{n+1})\,\mathrm{d}\boldsymbol{x},\end{aligned}$$

and thus, by the definition of the operator A, cf. (9),

$$\int_\Omega \mu(|\underline{e}(\mathbf{u}^n)|^2)|\underline{e}(\mathbf{u}^n-\mathbf{u}^{n+1})|^2\,\mathrm{d}\boldsymbol{x} = \mathsf{A}[\mathbf{u}^n](\mathbf{u}^n)(\mathbf{u}^n) - \mathsf{A}[\mathbf{u}^n](\mathbf{u}^{n+1})(\mathbf{u}^{n+1}) - 2\mathsf{A}[\mathbf{u}^n](\mathbf{u}^{n+1})(\mathbf{u}^n) + 2\mathsf{A}[\mathbf{u}^n](\mathbf{u}^{n+1})(\mathbf{u}^{n+1}).$$

Recall that $\mathsf{A}[\mathbf{u}^n](\mathbf{u}^{n+1})(\mathbf{v}) = \ell(\mathbf{v})$ for all $\mathbf{v}\in V$, cf. (11), which leads to

$$\int_\Omega \mu(|\underline{e}(\mathbf{u}^n)|^2)|\underline{e}(\mathbf{u}^n-\mathbf{u}^{n+1})|^2\,\mathrm{d}\boldsymbol{x} = \mathsf{A}[\mathbf{u}^n](\mathbf{u}^n)(\mathbf{u}^n) - \mathsf{A}[\mathbf{u}^n](\mathbf{u}^{n+1})(\mathbf{u}^{n+1}) - 2\ell(\mathbf{u}^n) + 2\ell(\mathbf{u}^{n+1}),$$

and hence, by (24),

$$\frac{1}{2}\int_\Omega \mu(|\underline{e}(\mathbf{u}^n)|^2)|\underline{e}(\mathbf{u}^n-\mathbf{u}^{n+1})|^2\,\mathrm{d}\boldsymbol{x} \le \mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^{n+1}). \tag{27}$$

Next, we will take care of the second summand in (26). As was done in [4], we want to bound this summand in terms of the energy difference $\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star)$. However, in order to improve the contraction factor whilst removing all unknown constants, some modifications to the argument presented in [4] are necessary. For $\psi(t) = \mathsf{E}(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star))$, the fundamental theorem of calculus implies that

$$\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) = \int_0^1 \psi'(t)\,\mathrm{d}t = \int_0^1 \mathsf{E}'(\mathbf{u}^\star + t(\mathbf{u}^n-\mathbf{u}^\star))(\mathbf{u}^n-\mathbf{u}^\star)\,\mathrm{d}t.$$

Recall that $\mathsf{E}'(\mathbf{u})(\mathbf{v}) = \mathsf{A}[\mathbf{u}][\mathbf{u}](\mathbf{v}) - \ell(\mathbf{v})$, cf. (12), and, since $\mathbf{u}^\star\in V$ is the unique solution of (10), $\ell(\mathbf{v}) = \mathsf{A}[\mathbf{u}^\star](\mathbf{u}^\star)(\mathbf{v})$ for any $\mathbf{v}\in V$. As a consequence, we have that

$$\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) = \int_0^1 \big(\mathsf{A}[\mathbf{u}^\star + t(\mathbf{u}^n-\mathbf{u}^\star)](\mathbf{u}^\star + t(\mathbf{u}^n-\mathbf{u}^\star)) - \mathsf{A}[\mathbf{u}^\star](\mathbf{u}^\star)\big)(\mathbf{u}^n-\mathbf{u}^\star)\,\mathrm{d}t.$$

Invoking the definition of A, cf. (9), and (14) further implies that

$$\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) \ge \int_0^1 t\int_\Omega c(\underline{e}(\mathbf{u}^\star) + t(\underline{e}(\mathbf{u}^n) - \underline{e}(\mathbf{u}^\star)), \underline{e}(\mathbf{u}^\star))|\underline{e}(\mathbf{u}^n-\mathbf{u}^\star)|^2\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t. \tag{28}$$

Moreover, by the definition of $c(\cdot,\cdot)$, cf. (16), and a brief argument based on reflection we get

$$c(\underline{e}(\mathbf{u}^\star) + s(\underline{e}(\mathbf{u}^n) - \underline{e}(\mathbf{u}^\star)), \underline{e}(\mathbf{u}^\star)) \ge \inf_{t\in(-1,1)} \xi'\big(\max\big\{0, |\underline{e}(\mathbf{u}^\star)| + t|\underline{e}(\mathbf{u}^n) - \underline{e}(\mathbf{u}^\star)|\big\}\big) \tag{29}$$

for all $s\in[0,1]$, where $\xi(t) = \mu(t^2)t$; indeed, it is easily verified that, for any $\underline{\kappa},\underline{\tau}\in\mathbb{R}^{d\times d}$, we have $c(\underline{\kappa},\underline{\tau}) = c(\underline{\tau},\underline{\kappa})$, and, in turn,

$$c(\underline{\kappa} + s(\underline{\tau}-\underline{\kappa}), \underline{\kappa}) = \inf_{t\in(0,1)} \xi'((1-t)|\underline{\kappa}| + t|\underline{\kappa} + s(\underline{\tau}-\underline{\kappa})|), \qquad s\in[0,1]. \tag{30}$$

The triangle inequality yields that

$$|\underline{\kappa}| - ts|\underline{\tau}-\underline{\kappa}| \le (1-t)|\underline{\kappa}| + t|\underline{\kappa} + s(\underline{\tau}-\underline{\kappa})| \le |\underline{\kappa}| + ts|\underline{\tau}-\underline{\kappa}| \qquad \text{for all } t\in(0,1),$$

and thus

$$(1-t)|\underline{\kappa}| + t|\underline{\kappa} + s(\underline{\tau} - \underline{\kappa})| = |\underline{\kappa}| + (2r-1)st|\underline{\kappa} - \underline{\tau}| \qquad \text{for some } r \in [0,1].$$

This further implies that, for any $s \in [0,1]$, we have

$$\{(1-t)|\underline{\kappa}| + t|\underline{\kappa} + s(\underline{\tau} - \underline{\kappa})| : t \in (0,1)\} \subseteq \{|\underline{\kappa}| + t|\underline{\kappa} - \underline{\tau}| : t \in (-1,1)\},$$

and consequently, in regard of (30),

$$c(\underline{\kappa} + s(\underline{\tau} - \underline{\kappa}), \underline{\kappa}) \geq \inf_{t\in(-1,1)} \xi'\big(\max\big\{0, |\underline{\kappa}| + ts|\underline{\tau} - \underline{\kappa}|\big\}\big),$$

which immediately implies (29). Combining the equalities (28) and (29) yields

$$\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) \geq \frac{1}{2}\int_\Omega \inf_{t\in(-1,1)} \xi'\big(\max\big\{0, |\underline{e}(\mathbf{u}^\star)| + t|\underline{e}(\mathbf{u}^n) - \underline{e}(\mathbf{u}^\star)|\big\}\big)|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)|^2 \,\mathrm{d}\boldsymbol{x}. \tag{31}$$

Since, in addition,

$$\begin{aligned}
&\frac{1}{2}\int_\Omega \mu\Big(|\underline{e}(\mathbf{u}^n)|^2\Big)|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)|^2 \,\mathrm{d}\boldsymbol{x} \\
&\quad = \frac{1}{2}\int_\Omega \frac{\mu\Big(|\underline{e}(\mathbf{u}^n)|^2\Big)}{\inf_{t\in(-1,1)} \xi'\big(\max\big\{0, |\underline{e}(\mathbf{u}^\star)| + t|\underline{e}(\mathbf{u}^n) - \underline{e}(\mathbf{u}^\star)|\big\}\big)} \\
&\qquad \cdot \inf_{t\in(-1,1)} \xi'\big(\max\big\{0, |\underline{e}(\mathbf{u}^\star)| + t|\underline{e}(\mathbf{u}^n) - \underline{e}(\mathbf{u}^\star)|\big\}\big)|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)|^2 \,\mathrm{d}\boldsymbol{x},
\end{aligned}$$

the lower bound (31) implies that

$$\frac{1}{2}\int_\Omega \mu\Big(|\underline{e}(\mathbf{u}^n)|^2\Big)|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)|^2 \,\mathrm{d}\boldsymbol{x} \leq Q(n)\big(\mathsf{E}(\mathbf{u}^n) - \mathsf{E}\big(\mathbf{u}^\star\big)\big), \tag{32}$$

where

$$Q(n) := \operatorname*{ess\,sup}_{x\in\Omega} \frac{\mu\Big(|\underline{e}(\mathbf{u}^n)|^2\Big)}{\inf_{t\in(-1,1)} \xi'\big(\max\big\{0, |\underline{e}(\mathbf{u}^\star)| + t|\underline{e}(\mathbf{u}^n) - \underline{e}(\mathbf{u}^\star)|\big\}\big)}.$$

Finally, combining (26), (27), and (32) yields

$$\gamma(1 - \gamma Q(n))\big(\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star)\big) \leq \mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^{n+1}),$$

and, in turn,

$$\mathsf{E}(\mathbf{u}^{n+1}) - \mathsf{E}(\mathbf{u}^\star) = \mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) - \big(\mathsf{E}(\mathbf{u}^n) - \mathsf{E}\big(\mathbf{u}^{n+1}\big)\big) \leq (1 - \gamma(1 - \gamma Q(n)))\big(\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star)\big).$$

It is straightforward to verify that the contraction factor is minimal for $\gamma = 1/2Q(n)$, and, in that case, one has that

$$0 \leq \mathsf{E}(\mathbf{u}^{n+1}) - \mathsf{E}(\mathbf{u}^\star) \leq \left(1 - \frac{1}{4Q(n)}\right)\big(\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star)\big),$$

which proves the claim. □

Remark 3.3 Since $m_\mu \leq \mu(t) \leq M_\mu$ as well as $m_\mu \leq \xi'(t) \leq M_\mu$ for all $t \geq 0$, we get the following crude uniform bound on the contraction factor:

$$q(n) \leq \left(1 - \frac{m_\mu}{4M_\mu}\right) \in [0.75, 1), \qquad n \geq 0.$$

We note that, in the context of the relaxed power-law model, cf. Sect. 3.2, this bound, in principle, coincides with the contraction factor from [4].

We note that the contraction factor (25) is not computable as it involves $\mathbf{u}^\star$, and the uniform upper bound from Remark 3.3 is rather pessimistic. In the following, we will establish an improved computable bound, up to higher order error terms, for the contraction factor on finite-dimensional subspaces.

Theorem 3.4 *Assume that (A1)–(A3) hold, let $W \subset V$ be a finite-dimensional subspace, and let E be defined as in (6) (restricted to W). Then, the energy error contracts along the sequence $\{\mathbf{u}^n\} \subset W$ generated by the Kačanov iteration (11) on W in the sense that*

$$0 \leq \mathsf{E}(\mathbf{u}^{n+1}) - \mathsf{E}(\mathbf{u}^\star) \leq q_A(n)\big(\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star)\big) + o_W\Big(|||\mathbf{u}^\star - \mathbf{u}^n|||_\Omega^2\Big),$$

where now $\mathbf{u}^\star$ denotes the unique minimiser of E in W,

$$q_A(n) := 1 - \frac{1}{4}\left\{\operatorname*{ess\,sup}_{x\in\Omega} \frac{\mu(|\underline{e}(\mathbf{u}^n)|^2)}{2\mu'(|\underline{e}(\mathbf{u}^n)|^2)|\underline{e}(\mathbf{u}^n)|^2 + \mu(|\underline{e}(\mathbf{u}^n)|^2)}\right\}^{-1}, \tag{33}$$

and $o_W(|||\mathbf{u} - \mathbf{u}^n|||_\Omega^2)$ denotes a remainder term depending on W.

Proof This result follows, in principle, from the proof of Theorem 3.2 with a modification of the bound from (32). Consider the map $\omega : W \to W^\star$ given by $\omega(\mathbf{u}) := \mathsf{A}[\mathbf{u}](\mathbf{u})$ for $\mathbf{u} \in W$. A lengthy, but straightforward calculation reveals that the Gâteaux derivative of ω exists and is given by

$$\begin{aligned}\omega'(\mathbf{u})(\mathbf{v})(\mathbf{w}) = \int_\Omega &2\mu'\Big(\big|\underline{e}(\mathbf{u})\big|^2\Big)(\underline{e}(\mathbf{u}) : \underline{e}(\mathbf{v}))(\underline{e}(\mathbf{u}) : \underline{e}(\mathbf{w})) \\ &+ \mu\Big(\big|\underline{e}(\mathbf{u})\big|^2\Big)\big(\underline{e}(\mathbf{v}) : \underline{e}(\mathbf{w})\big)\,\mathrm{d}\boldsymbol{x}, \quad \mathbf{v}, \mathbf{w} \in W.\end{aligned}$$

Since W is a finite-dimensional space and $\omega : W \to W^\star$ is Lipschitz continuous by Proposition 2.2, the Gâteaux derivative coincides with the Fréchet derivative, see [19, Prop. 3.5]. By definition of the Fréchet derivative, one has that

$$\omega(\mathbf{u}) = \omega(\mathbf{u}^n) + \omega'(\mathbf{u}^n)(\mathbf{u} - \mathbf{u}^n) + o_W\big(|||\mathbf{u} - \mathbf{u}^n|||_\Omega\big);$$

here, $o_W(|||\mathbf{u} - \mathbf{u}^n|||_\Omega)$ denotes a remainder in the dual space $W^\star$. Combining these two observations yields

$$\begin{aligned}
&\big(\mathsf{A}[\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star)](\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star)) - \mathsf{A}[\mathbf{u}^\star](\mathbf{u}^\star)\big)(\mathbf{u}^n - \mathbf{u}^\star)\\
&\quad = \big(\omega(\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star)) - \omega(\mathbf{u}^\star)\big)(\mathbf{u}^n - \mathbf{u}^\star)\\
&\quad = t\omega'(\mathbf{u}^n)(\mathbf{u}^n - \mathbf{u}^\star)(\mathbf{u}^n - \mathbf{u}^\star) + o_W\Big(|||\mathbf{u}^\star - \mathbf{u}^n|||_\Omega^2\Big)\\
&\quad = t\int_\Omega 2\mu'(|\underline{e}(\mathbf{u}^n)|^2)(\underline{e}(\mathbf{u}^n) : \underline{e}(\mathbf{u}^n - \mathbf{u}^\star))^2\\
&\qquad + \mu(|\underline{e}(\mathbf{u}^n)|^2)|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)|^2\,\mathrm{d}\boldsymbol{x} + o_W\Big(|||\mathbf{u}^\star - \mathbf{u}^n|||_\Omega^2\Big).
\end{aligned}$$

Recall that, by assumption (A3), $\mu'(t) \le 0$ for all $t \ge 0$. Therefore, the Cauchy–Schwarz inequality implies that

$$\begin{aligned}
&\big(\mathsf{A}[\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star)](\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star)) - \mathsf{A}[\mathbf{u}^\star](\mathbf{u}^\star)\big)(\mathbf{u}^n - \mathbf{u}^\star)\\
&\qquad \ge t\int_\Omega \Big\{2\mu'(|\underline{e}(\mathbf{u}^n)|^2)|\underline{e}(\mathbf{u}^n)|^2 + \mu(|\underline{e}(\mathbf{u}^n)|^2)\Big\}|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)|^2\,\mathrm{d}\boldsymbol{x}\\
&\qquad\quad + o_W\Big(|||\mathbf{u}^\star - \mathbf{u}^n|||_\Omega^2\Big).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathsf{E}(\mathbf{u}^n) - \mathsf{E}(\mathbf{u}^\star) &= \int_0^1 \big(\mathsf{A}[\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star)](\mathbf{u}^\star + t(\mathbf{u}^n - \mathbf{u}^\star)) - \mathsf{A}[\mathbf{u}^\star](\mathbf{u}^\star)\big)(\mathbf{u}^n - \mathbf{u}^\star)\,\mathrm{d}t\\
&\ge \frac{1}{2}\int_\Omega \big\{2\mu'\big(|\underline{e}(\mathbf{u}^n)|^2\big)|\underline{e}(\mathbf{u}^n)|^2 + \mu\big(|\underline{e}(\mathbf{u}^n)|^2\big)\big\}|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)|^2\,\mathrm{d}\boldsymbol{x} + o_W\Big(|||\mathbf{u}^\star - \mathbf{u}^n|||_\Omega^2\Big),
\end{aligned}$$

and thus

$$\begin{aligned}
\frac{1}{2}\int_\Omega \mu\Big(\big|\underline{e}(\mathbf{u}^n)\big|^2\Big)\big|\underline{e}(\mathbf{u}^n - \mathbf{u}^\star)\big|^2\,\mathrm{d}\boldsymbol{x} &\le \operatorname*{ess\,sup}_{x\in\Omega} \frac{\mu(|\underline{e}(\mathbf{u}^n)|^2)}{2\mu'(|\underline{e}(\mathbf{u}^n)|^2)\big|\underline{e}(\mathbf{u}^n)\big|^2 + \mu(|\underline{e}(\mathbf{u}^n)|^2)}\Big(\big(\mathsf{E}(\mathbf{u}^n) - \mathsf{E}\big(\mathbf{u}^\star\big)\big)\\
&\quad + o_W\Big(|||\mathbf{u}^\star - \mathbf{u}^n|||_\Omega^2\Big)\Big).
\end{aligned}$$

The rest follows as in the proof of Theorem 3.2. We note, however, that the factor of the remainder term $o_W\Big(|||\mathbf{u}^\star - \mathbf{u}^n|||_\Omega^2\Big)$ above cancels by the multiplication with γ in (26). □

Remark 3.5 We emphasize that the contraction factor q_A from (33) is independent of the finite-dimensional subspace $W \subset V$. However, the remainder term o_W may depend on the choice of the given discrete subspace, as indicated by the subscript.

Finally we remark that the energy error is equivalent to the norm error, i.e., the norm error contracts, up to some uniform constant, along the sequence generated by the Kačanov scheme as well. This equivalence was already established in a similar

setting, e.g., in [11, Lem. 2.3] and [6, Lem. 5.1]. The proof can also be found in those references.

Proposition 3.6 *Let* E *be defined as in* (6), *with* μ *satisfying (A1)–(A3), and let* $\mathbf{u}^\star$ *be the unique minimiser of* E *in* V; *then*,

$$\frac{m_\mu}{2}|||\mathbf{u}^\star - \mathbf{v}|||_\Omega^2 \le \mathsf{E}(\mathbf{v}) - \mathsf{E}(\mathbf{u}^\star) \le \frac{\sqrt{3}M_\mu}{2}|||\mathbf{u}^\star - \mathbf{v}|||_\Omega^2 \qquad \text{for all } \mathbf{v} \in V. \tag{34}$$

An analogous result holds on any finite-dimensional subspace $W \subset V$, *with* V *replaced by* W *in the assertion above*.

3.1 Application to the Carreau model

A widely used model for the flow of incompressible non-Newtonian fluids is the Carreau law, cf. [3]. In that case the viscosity coefficient μ in (1) is of the form

$$\mu(t) = \mu_\infty + (\mu_0 - \mu_\infty)(1 + \lambda t)^{\frac{r-2}{2}}, \tag{35}$$

where for shear-thinning fluids, $r \in (1,2)$, $\lambda > 0$ is the relaxation time, and $0 < \mu_\infty < \mu_0 < \infty$ denote the infinite and zero shear rate, respectively. The function μ from (35) is smooth, decreasing since $r \in (1,2)$, and satisfies the structural assumption (A2), cf. (2), thanks to the following lemma.

Lemma 3.7 *Let* $r \in (1,2)$, $\lambda > 0$, *and* $0 < \mu_\infty < \mu_0 < \infty$. *Then, the following inequalities hold*:

$$\mu_\infty(t-s) \le \mu(t^2)t - \mu(s^2)s \le \mu_0(t-s), \qquad t \ge s \ge 0.$$

Proof Define $\xi(t) := \mu(t^2)t$, $t \ge 0$. The mean value theorem yields

$$\inf_{\tau \ge 0} \xi'(\tau)(t-s) \le \xi(t) - \xi(s) \le \sup_{\tau \ge 0} \xi'(\tau)(t-s),$$

and thus we need to show that $\mu_\infty = \inf_{\tau \ge 0} \xi'(\tau)$ and $\mu_0 = \sup_{\tau \ge 0} \xi'(\tau)$. A straightforward calculation reveals that $\xi''(\tau) \ne 0$ for all $\tau \ge 0$, i.e., ξ' has no local extrema in the interval $(0, \infty)$. Since, in addition, $\lim_{\tau \to 0} \xi'(\tau) = \mu_0$ and $\lim_{\tau \to \infty} \xi'(\tau) = \mu_\infty$, the lemma is established. □

In particular, we may apply Theorems 3.2 and 3.4 to the Carreau model. In this case, the computable contraction factor from (33) reads as follows, with $\mathbf{u}^n \in W$:

$$q_A(n) := 1 - \frac{1}{4}\left(\operatorname*{ess\,sup}_{x\in\Omega} \frac{\mu_\infty + (\mu_0 - \mu_\infty)(1+\lambda|\underline{e}(\mathbf{u}^n)|^2)^{\frac{r-2}{2}}}{\mu_\infty + (\mu_0 - \mu_\infty)(1+\lambda|\underline{e}(\mathbf{u}^n)|^2)^{-1+\frac{r-2}{2}}(1+\lambda(r-1)|\underline{e}(\mathbf{u}^n)|^2)} \right)^{-1}$$
$$= 1 - \frac{1}{4} \operatorname*{ess\,inf}_{x\in\Omega} \frac{\mu_\infty + (\mu_0 - \mu_\infty)(1+\lambda|\underline{e}(\mathbf{u}^n)|^2)^{-1+\frac{r-2}{2}}(1+\lambda(r-1)|\underline{e}(\mathbf{u}^n)|^2)}{\mu_\infty + (\mu_0 - \mu_\infty)(1+\lambda|\underline{e}(\mathbf{u}^n)|^2)^{\frac{r-2}{2}}}.$$

Let us further examine this factor. First we note that

$$\frac{\mu_\infty + (\mu_0 - \mu_\infty)(1+\lambda t^2)^{-1+\frac{r-2}{2}}(1+\lambda(r-1)t^2)}{\mu_\infty + (\mu_0 - \mu_\infty)(1+\lambda t^2)^{\frac{r-2}{2}}} \to 1 \qquad \text{as } t \to \infty,$$

which is optimal from the point of view of contraction. Consequently, we do not expect a significant deterioration of the convergence rate if the rate-of-strain tensor of the solution, i.e., $\underline{e}(\mathbf{u}^\star)$, is unbounded, cf. Experiment 4.1.1.

Moreover, an elementary calculation reveals that

$$\frac{\mu_\infty + (\mu_0 - \mu_\infty)(1+\lambda t^2)^{-1+\frac{r-2}{2}}\left(1+\lambda(r-1)t^2\right)}{\mu_\infty + \left(\mu_0 - \mu_\infty\right)(1+\lambda t^2)^{\frac{r-2}{2}}} \geq \frac{1+\lambda(r-1)t^2}{1+\lambda t^2} \qquad \text{for all } t \geq 0,$$

and therefore

$$q_A(n) = 1 - \frac{1}{4} \operatorname*{ess\,inf}_{x\in\Omega} \frac{\mu_\infty + \left(\mu_0 - \mu_\infty\right)\left(1+\lambda\left|\underline{e}(\mathbf{u}^n)\right|^2\right)^{-1+\frac{r-2}{2}}\left(1+\lambda(r-1)\left|\underline{e}(\mathbf{u}^n)\right|^2\right)}{\mu_\infty + \left(\mu_0 - \mu_\infty\right)\left(1+\lambda\left|\underline{e}(\mathbf{u}^n)\right|^2\right)^{\frac{r-2}{2}}}$$
$$\leq 1 - \frac{1}{4} \operatorname*{ess\,inf}_{x\in\Omega} \frac{1+\lambda(r-1)\left|\underline{e}(\mathbf{u}^n)\right|^2}{1+\lambda\left|\underline{e}(\mathbf{u}^n)\right|^2}$$
$$\leq 1 - \frac{1}{4}(r-1).$$

In combination with Remark 3.3, we get

$$q_A(n) \leq \min\left\{ 1 - \frac{1}{4}\frac{\mu_\infty}{\mu_0}, 1 - \frac{1}{4}(r-1) \right\}, \tag{36}$$

i.e., the convergence rate may only deteriorate drastically if $r \to 1$ and, in addition, $\mu_\infty/\mu_0 \to 0$.

3.2 Application to the relaxed power-law model

Another prominent model for non-Newtonian fluids, e.g., in polymer processing, is the power-law model, see, e.g., [20, Ch. 3.3]. For this model, the weak formulation (4) of the boundary-value problem under consideration is as follows:

$$\text{find } \mathbf{u} \in X \text{ such that } \quad \int_\Omega |\underline{e}(\mathbf{u})|^{r-2}\underline{e}(\mathbf{u}) : \underline{e}(\mathbf{v})\,\mathrm{d}\boldsymbol{x} = \ell(\mathbf{v}) \qquad \text{for all } \mathbf{v} \in X; \tag{37}$$

here $X := \{\mathbf{u} \in \mathrm{W}^{1,r}_0(\Omega)^d : \nabla \cdot \mathbf{u} = 0\}$ and $\ell \in X^\star$, where, for shear-thinning fluids, $r \in (1,2)$. In particular, the viscosity coefficient is given by

$$\mu(t) = t^{\frac{r-2}{2}}.$$

Clearly, $\mu : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is neither bounded away from zero nor bounded from above, i.e., (A2) is not satisfied. Therefore, as was proposed in the work [4], we consider a relaxed version of μ: for $0 < \varepsilon_- < \varepsilon_+ < \infty$ we define the viscosity coefficient

$$\mu_\varepsilon(t) := \begin{cases} \varepsilon_-^{r-2} & \text{for } 0 \le t < \varepsilon_-^2, \\ t^{\frac{r-2}{2}} & \text{for } \varepsilon_-^2 \le t \le \varepsilon_+^2, \\ \varepsilon_+^{r-2} & \text{for } t \ge \varepsilon_+^2. \end{cases} \tag{38}$$

The function μ_ε is decreasing, strictly positive, bounded, globally Lipschitz continuous, and satisfies (A2) with

$$(r-1)\varepsilon_+^{r-2}(t-s) \le \mu(t^2)t - \mu(s^2)s \le \varepsilon_-^{r-2}(t-s), \qquad t \ge s \ge 0;$$

it is, furthermore, differentiable at all $t \in [0,\infty) \setminus \{\varepsilon_-^2, \varepsilon_+^2\}$ and has finite left- and right-derivatives at $t = \varepsilon_-^2$ and $t = \varepsilon_+^2$, respectively. Hence, even though μ_ε is not continuously differentiable on $[0,\infty)$, Theorem 3.2 can, nevertheless, be applied in the given setting. Moreover, in the generic case when the set $\Omega_S^n := \{\boldsymbol{x} \in \Omega : |\underline{e}(\mathbf{u}^n(\boldsymbol{x}))| \in \{\varepsilon_-, \varepsilon_+\}\}$, for every $n \ge 0$, has Lebesgue measure zero, the operator ω from the proof of Theorem 3.4 is Fréchet differentiable at $\mathbf{u}^n \in W$. Thus, in turn, Theorem 3.4 can then also be applied to the relaxed power-law model[1]. A simple calculation reveals that the computable contraction factor from (33) can again be bounded; indeed,

$$q_A(n) \le 1 - \frac{1}{4}(r-1). \tag{39}$$

Moreover, one even has that $q_A(n) = 1 - 4^{-1}(r-1)$ if the set $\{\boldsymbol{x} \in \Omega : \varepsilon_- \le |\underline{e}(\mathbf{u}^n(\boldsymbol{x}))| \le \varepsilon_+\}$ is of positive Lebesgue measure. We further remark that

$$\frac{m_\mu}{M_\mu} = \frac{(r-1)\varepsilon_+^{r-2}}{\varepsilon_-^{r-2}} < (r-1),$$

since $r \in (1,2)$. This shows that the bound (39) is, for every value $r \in (1,2)$, sharper than the bound from Remark 3.3. Furthermore, this bound predicts that it

[1] Nonetheless we will present a continuously differentiable version of the viscosity coefficient (38) in the Appendix.

is the physical parameter r that affects the convergence rate of the iteration, in the finite-dimensional setting at least, rather than the quotient $\varepsilon_+^{r-2}/\varepsilon_-^{r-2}$ implied by existing bounds on the contraction factor, cf. [4, Cor. 19]. Significantly, the upper bound $(r-1)$ on the contraction factor appearing of the right-hand side of (39) is independent of the relaxation parameters $\varepsilon_\pm$. This is of importance as we are interested in the power-law model (37) and we thus need to let $\varepsilon_- \to 0$ and $\varepsilon_+ \to \infty$. We note that the existence of a bound independent of $\varepsilon_\pm$ on the contraction factor of the relaxed Kačanov iteration applied to the power-law model with $r \in (1,2)$ was stated in the infinite-dimensional case as an open problem in [4, Ex. 20].

We further note that the energy functional E_ε corresponding to the viscosity from (38) coincides with the energy functional $\mathcal{J}_\epsilon$ from [4] up to a constant shift depending on ε_-. To be precise, one has that

$$\mathsf{E}_\varepsilon(\mathbf{u}) = \mathcal{J}_\epsilon(\mathbf{u}) + \left(\frac{1}{2} - \frac{1}{r}\right)\varepsilon_-^r, \qquad \mathbf{u} \in V,$$

and thus the results established in [4] may be directly applied in our setting. In particular, this implies that the sequence of unique minimisers $\mathbf{u}_\varepsilon^\star \in V$ of E_ε converges in $\mathrm{W}_0^{1,r}(\Omega)^d$ to the unique minimiser $\mathbf{u}^\star \in X$ of

$$\mathsf{E}(\mathbf{u}) = \frac{1}{r}\int_\Omega |\underline{e}(\mathbf{u})|^r \,\mathrm{d}\boldsymbol{x} - \ell(\mathbf{u}),$$

cf. [4, Cor. 10].

Remark 3.8 The relaxed power-law model could also be solved by using a (damped) Newton method, cf. [10, Prop. 5.3]. However, it is unclear whether and how the convergence rate will deteriorate as $\varepsilon_- \to 0$ and $\varepsilon_+ \to \infty$. For an application of Newton's method to the power-law model with a different regularisation approach we refer to [12]; however, the convergence rate in relation to the choice of the regularisation parameter ε is not examined in that work.

Remark 3.9 We emphasise that our analysis does also apply to a variable (measurable) exponent $r : \Omega \to (1,2)$ for both the relaxed power-law model as well as the Carreau model. Then, in (39) and (36), respectively, we need to replace $1 - 1/4(r-1)$ by $1 - 1/4(\operatorname{ess\,inf}_{\boldsymbol{x}\in\Omega} r(\boldsymbol{x}) - 1)$.

4 Experiments

In this section, we will perform some numerical tests to assess our findings. To this end, we consider the simplified problem

$$\text{find } u \in \mathrm{H}_0^1(\Omega) \text{ such that} \qquad \int_\Omega \mu(|\nabla u|^2)\nabla u \cdot \nabla v = \int_\Omega f v \,\mathrm{d}\boldsymbol{x} \qquad \text{for all } v \in \mathrm{H}_0^1(\Omega),$$

where $\Omega := (-1,1)^2 \setminus [0,1] \times [-1,0] \subset \mathbb{R}^2$ is an L-shaped domain, $f \in \mathrm{L}^2(\Omega)$, and the coefficient μ either obeys the Carreau law (35) or the relaxed power-law (38). We

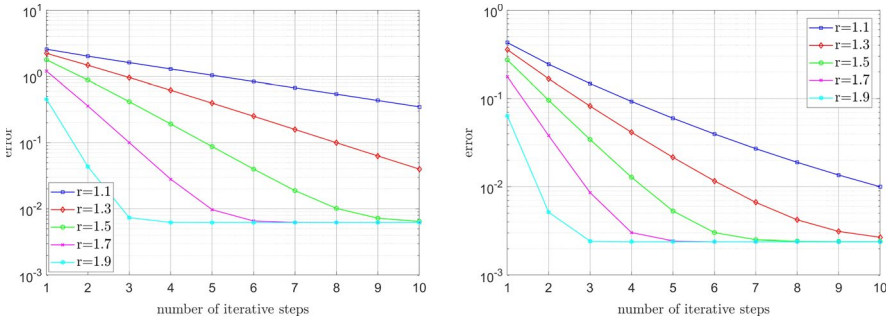

Fig. 1 Carreau model: Influence of the physical parameter r on the convergence rate in the smooth case (left) and irregular case (right), where $\mu_\infty = 1$, $\mu_0 = 100$, and $\lambda = 2$

remark that the theory derived before equally applies to this simpler case. In all our experiments below, we use a conforming P1-finite element discretisation, where the mesh consists of $\mathcal{O}(10^6)$ triangles, except where explicitly stated otherwise.

4.1 Error decay in dependence on *r*

First, we will examine how the convergence rate of the *error* depends on the exponent $\frac{r-2}{2}$; recall that the norm error is equivalent to the energy error, cf. Proposition 3.6. This will be done for both the Carreau and the relaxed power-law model, for smooth and irregular solutions.

4.1.1 Error decay for the Carreau model

Let the function μ obey the Carreau law (35), with $\mu_\infty = 1$, $\mu_0 = 100$, $\lambda = 2$, and varying values of $r \in (1, 2)$. The source term f is chosen so that the unique solution is given by

(a) The smooth function $u^\star(x, y) = \sin(\pi x)\sin(\pi y)$, where $(x, y) \in \mathbb{R}^2$ denote the Euclidean coordinates;
(b) The function

$$u^\star(R, \varphi) = R^{2/3}\sin(2\varphi/3)(1 - R\cos(\varphi))(1 + R\cos(\varphi))(1 - R\sin(\varphi))(1 + R\sin(\varphi))\cos(\varphi),$$

where R and φ are polar coordinates, which exhibits a singularity at the origin (0, 0).

In the smooth case (a) the mesh is uniform, and in the singular case (b) the mesh is increasingly refined in the vicinity of the singularity point (0,0). In Fig. 1 we plot the error $\|\nabla u^n - \nabla u^\star\|_{\mathrm{L}^2(\Omega)}$ against the number of iterative steps n. We can clearly see that the convergence rate deteriorates with decreasing r, as was predicted in Sect. 3. We further note that the irregularity of the solution in (b) does not affect the convergence rate, as was conjectured in Sect. 3.1.

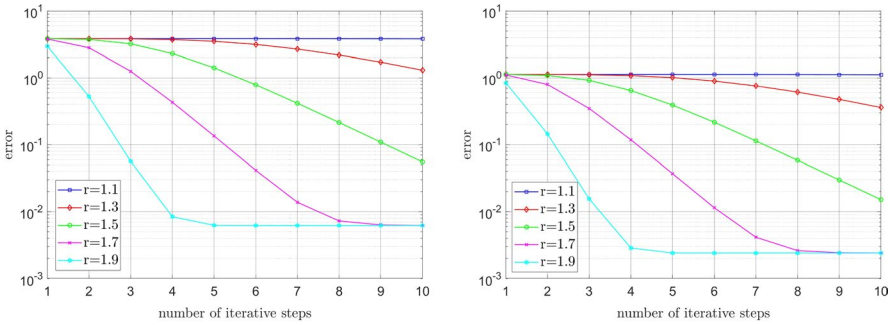

Fig. 2 Relaxed power-law model: Influence of the physical parameter r on the convergence rate in the smooth case (left) and irregular case (right)

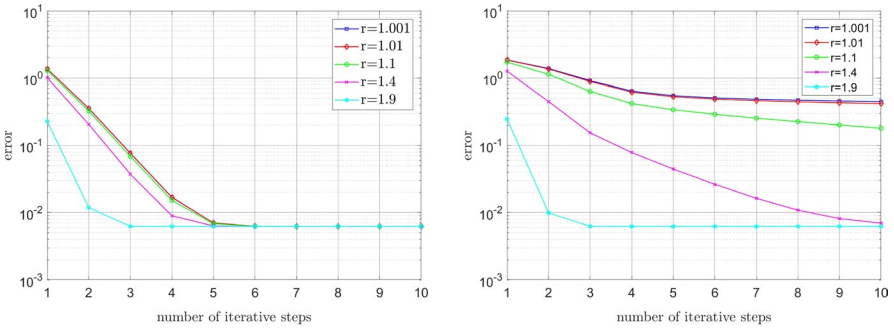

Fig. 3 Influence of the physical parameter r for almost constant viscosity. Left: Carreau model. Right: Relaxed power-law model

4.1.2 Error decay for the relaxed power-law model

Now consider the relaxed power-law model, cf. (38), with $\varepsilon_- = 10^{-6}$ and $\varepsilon_+ = 10^6$. As before, the source term f is chosen so that (a) $u^\star$ is smooth, and (b) $u^\star$ exhibits a singularity at the origin (0,0). In Fig. 2, the error $\left\|\nabla u^n - \nabla u^\star\right\|_{\mathrm{L}^2(\Omega)}$ is plotted against the number of iterative steps n. We observe that for the power-law model the dependence of the convergence rate on the exponent is even stronger than for the Carreau model.

4.1.3 Error decay for close to constant viscosity

In the experiments before we had that the ratio of the infinite and zero shear rates was much smaller than $(r-1)$, cf. (36). Now we choose the parameters so that the 'shear stress' depends almost linearly on the 'shear rate', and we further let the source term f be such that the unique solution of (4) is given by the smooth function $u^\star(x,y) = \sin(\pi x)\sin(\pi y)$. For the Carreau law we set $\mu_\infty = 1$, $\mu_0 = 2$, $\lambda = 2$, and take again varying values of $r \in (1,2)$; we emphasize that, in

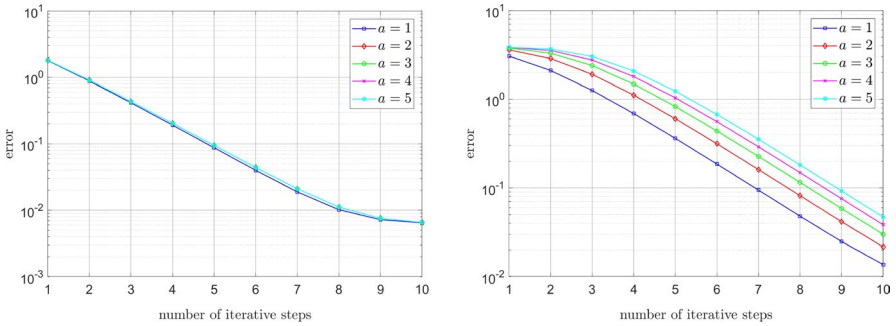

Fig. 4 Influence of the ratio of the infinite and zero shear rates on the convergence. Left: Carreau model with $\lambda = 2$, $r = 1.5$, $\mu_0 = 10^a$, and $\mu_\infty = 10^{-a}$. Right: Relaxed power-law model with $r = 1.5$, $\varepsilon_- = 10^{-a}$, and $\varepsilon_+ = 10^a$

this test, we consider even smaller values of r than in the experiments before. In view of the a posteriori computable contraction factor (36) we expect that the convergence rate will not deteriorate drastically for r close to one, which is confirmed by our numerical experiment, cf. Fig. 3 (left). In the case of the relaxed power-law model, we set $\varepsilon_- = 1$, $\varepsilon_+ = 2$, and test the same values $r \in (1, 2)$ as for the Carreau model. We note that these choices of the relaxation parameters $\varepsilon_\pm$ are in practice of no interest, as one is, rather, interested in $\varepsilon_- \to 0$ and $\varepsilon_+ \to \infty$. Nonetheless, we still presume that the convergence deteriorates for r close to one by our analysis in Sect. 3.2. This is indeed the case, as illustrated in Fig. 3 (right).

4.2 Error decay in dependence on the zero and infinite shear rates

Next, we will show that, for fixed $r \in (1, 2)$, the convergence rate does not essentially deteriorate when the ratio of the infinite and zero shear rates decreases. As in the experiment before, we choose the source term f so that the unique solution is given by the smooth function $u^\star(x, y) = \sin(\pi x)\sin(\pi y)$. For the Carreau model we set $\lambda = 2$, $r = 1.5$, $\mu_0 = 10^a$, and $\mu_\infty = 10^{-a}$ for $a \in \{1, 2, 3, 4, 5\}$. As we can see from Fig. 4 (left), the convergence rate is (almost) independent of a, i.e., the convergence does not deteriorate for a decreasing quotient μ_∞/μ_0. For the relaxed power-law model we set $r = 1.5$, $\varepsilon_- = 10^{-a}$, and $\varepsilon_+ = 10^a$ for $a \in \{1, 2, 3, 4, 5\}$. Even though the plots differ for the various values of a, the convergence *rate* is almost the same for all of them; indeed, no significant deterioration of the convergence rate can be observed in Fig. 4 (right) for increasing a.

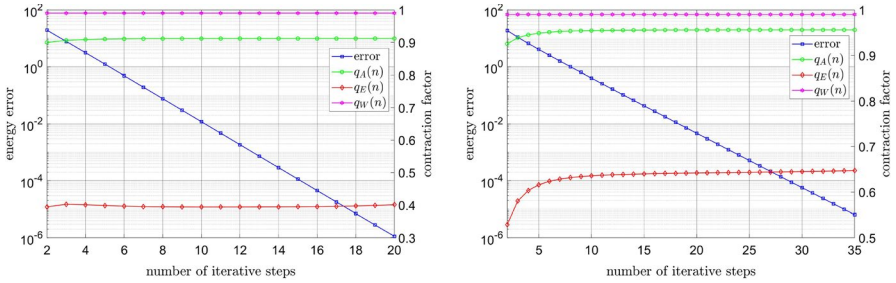

Fig. 5 Energy decay and the contraction factors for the Carreau model with $r = 1.3$ (left) and $r = 1.1$ (right)

4.3 Energy decay and the contraction factor

We now focus on the energy decay, and compare the exact contraction factor, cf. (40), the a posteriori computable factor (41), and the worst case factor from Remark 3.3, cf. (42). Again, this will be done for the Carreau and the relaxed power-law models. In our figures below, we plot the energy decay $\mathsf{E}(u^n) - \mathsf{E}(u^\star)$, as well as the aforementioned factors

$$q_E(n) = \frac{\mathsf{E}(u^n) - \mathsf{E}(u^\star)}{\mathsf{E}(u^{n-1}) - \mathsf{E}(u^\star)}, \tag{40}$$

$$q_A(n) = 1 - \frac{1}{4}\left\{ \operatorname*{ess\,sup}_{x\in\Omega} \frac{\mu\left(\left|\underline{e}(\mathbf{u}^n)\right|^2\right)}{2\mu'\left(\left|\underline{e}(\mathbf{u}^n)\right|^2\right)\left|\underline{e}(\mathbf{u}^n)\right|^2 + \mu\left(\left|\underline{e}(\mathbf{u}^n)\right|^2\right)} \right\}^{-1}, \tag{41}$$

$$q_W(n) = 1 - \frac{1}{4}\frac{m_\mu}{M_\mu}, \tag{42}$$

against the number of iteration steps n.

4.3.1 Energy contraction for the Carreau model

We consider the Carreau model, cf. (35), for $\mu_\infty = 1$, $\mu_0 = 100$, $\lambda = 2$, and $r = 1.3$, respectively $r = 1.1$. In both cases, we approximate the discrete solution for the source term f from case (a) before by performing seventy steps of the Kačanov iteration (11), and subsequently use this approximation for the determination of the reference energy $\mathsf{E}(u^\star)$; here, $u^\star$ denotes the unique minimiser in the finite element space. We can clearly observe in Fig. 5 that, on the one hand, the a posteriori computable factor $q_A(n)$, cf. (41), is much larger than the actual factor $q_E(n)$, cf. (40). On the other hand, however, the factor $q_A(n)$ clearly still improves the worst case factor $q_W(n)$ from Remark 3.3, cf. (42).

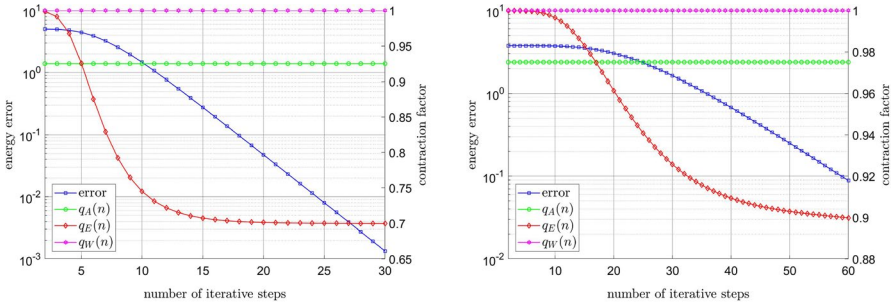

Fig. 6 Energy decay and the contraction factors for the power-law model with $r = 1.3$ (left) and $r = 1.1$ (right)

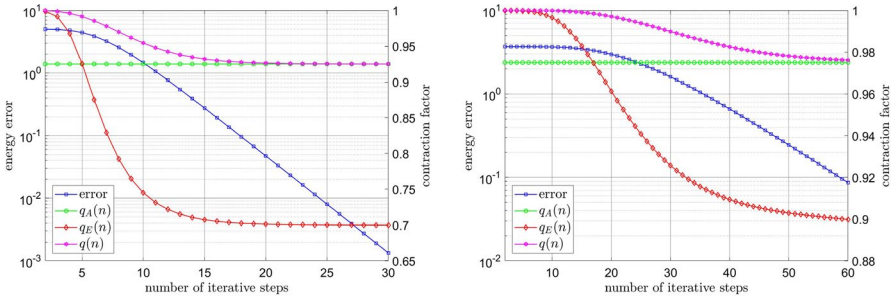

Fig. 7 Energy decay and the contraction factors for the power-law model with $r = 1.3$ (left) and $r = 1.1$ (right) in the coarser mesh

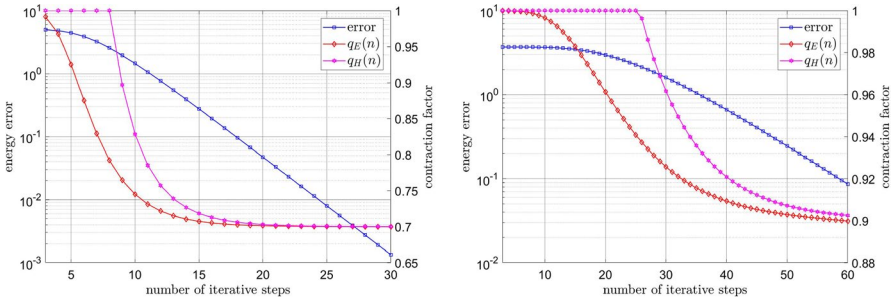

Fig. 8 Energy decay and the contraction factors for the power-law model with $r = 1.3$ (left) and $r = 1.1$ (right)

4.3.2 Energy contraction for the relaxed power-law model

Let the coefficient μ obey the relaxed power-law model with $\varepsilon_- = 10^{-6}$, $\varepsilon_+ = 10^{6}$, and $r = 1.3$, respectively $r = 1.1$. In this experiment, we approximate the discrete solution $u^\star$ by performing fifty and one hundred iteration steps for $r = 1.3$ and

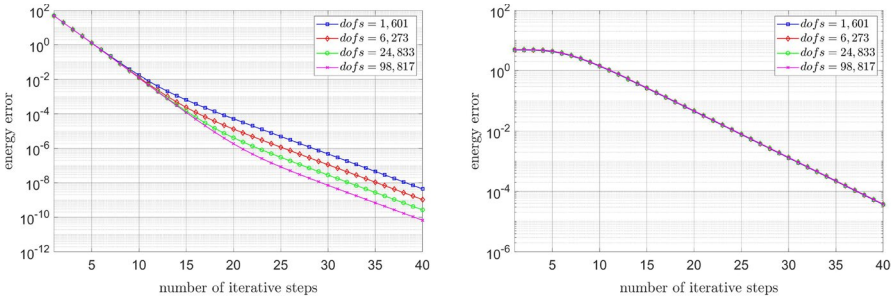

Fig. 9 Energy decay for the Carreau model (left) and relaxed power-law model (right) for different mesh sizes

$r = 1.1$, respectively. As before, the a posteriori computable contraction factor $q_A(n)$ is noticeably larger than the exact factor $q_E(n)$, however, this is less marked than before; see Fig. 6. Moreover, it considerably improves the worst case contraction factor $q_W(n) \approx 1 - 10^{-12}$.

We now repeat this experiment on a coarser mesh consisting of $\mathcal{O}(10^5)$ uniform triangles. In Fig. 7 we plot the factors $q_A(n)$, $q_E(n)$, as well as $q(n)$ from (25) against the number of iteration steps. We observe that the (non-computable) factor $q(n)$ from (25) has a similar trend as the exact factor $q_E(n)$, cf. (40), and approximates the computable factor $q_A(n)$, cf. (41), as the number of iteration steps increases.

Finally, we remark that, in the context of fixed point iterations, the contraction factor can be (heuristically) approximated by

$$q_H(n) := \min\left\{1, \frac{\mathsf{E}(u^n) - \mathsf{E}(u^{n-1})}{\mathsf{E}(u^{n-1}) - \mathsf{E}(u^{n-2})}\right\} \tag{43}$$

as $n \to \infty$, see, e.g., [18]; we emphasize that $q_H(n) \geq 0$, for $n \geq 2$, thanks to (27). As can be observed in Fig. 8, the factor $q_H(n)$, cf. (43), does indeed approximate the exact factor $q_E(n)$ from (40) well for sufficiently large n. However, in contrast with the bound $q_A(n)$ from Theorem 3.4, the computable factor $q_H(n)$ does not provide any guaranteed *a priori* information.

4.4 Energy decay for different mesh sizes

We conclude this section with a comparison of the energy decay for different mesh sizes. For the Carreau model, cf. (35), we set $\mu_\infty = 1$, $\mu_0 = 100$, $\lambda = 2$, and $r = 1.3$. In the case of the relaxed power-law model, let $\varepsilon_- = 10^{-6}$, $\varepsilon_+ = 10^6$, and $r = 1.3$. In each case we approximated the discrete solution, and, in turn, the corresponding energy by performing one hundred iteration steps. As we can see from Fig. 9, the asymptotic convergence rates (almost) coincide for the different mesh sizes.

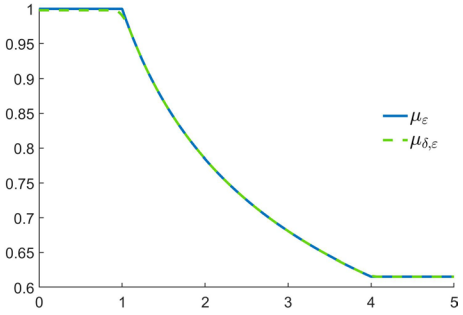

Fig. 10 Comparison of μ_ε and $\mu_{\delta,\varepsilon}$ for $r = 1.3$, $\varepsilon_- = 1$, $\varepsilon_+ = 2$, and $\delta = 0.1$

5 Conclusion

In this article, we established an a posteriori computable (energy) contraction factor for the Kačanov scheme (on finite-dimensional Galerkin spaces), motivated by applications to quasi-Newtonian fluid flow problems. For the relaxed power-law model, this factor is independent of the relaxation parameters $\varepsilon_\pm$; we also demonstrated that it is, instead, the power-law exponent that affects the convergence rate of the iteration. In contrast, existing bounds on the contraction factor of the relaxed Kačanov iteration depend on the relaxation parameters $\varepsilon_\pm$ in an unfavourable manner, in the sense that they tend to 1 as $\varepsilon_- \to 0$ and/or $\varepsilon_+ \to \infty$. A series of numerical tests have confirmed that our a posteriori computable contraction factor improves, on finite-dimensional Galerkin spaces, existing bounds, and that, as predicted by our analysis, for the power-law model it is in fact the closeness of the power-law exponent $r \in (1, 2)$ to 1 that influences the convergence rate of the iteration. However, our experiments revealed that the theoretically derived bound on the contraction factor of the Kačanov scheme is still too pessimistic.

Appendix

A Smoothly relaxed power-law model

In this appendix, we will introduce a continuously differentiable approximation of the relaxed power-law viscosity (38). In particular, for $0 < \delta < \varepsilon_-^2 < \varepsilon_+^2$, we want to define a function $\mu_{\delta,\varepsilon} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ which

(a) Is continuously differentiable;
(b) Coincides with μ_ε, cf. (38), in the domain $[\varepsilon_-^2 + \delta, \varepsilon_+^2 - \delta]$;
(c) Is constant on $[0, \varepsilon_-^2 - \delta] \cup [\varepsilon_+^2 + \delta, \infty)$;
(d) Converges pointwise to μ_ε for $\delta \to 0$ (Fig. 10).

The idea is to identify quadratic functions $g^{\pm}_{\delta,\varepsilon}$ on $[\varepsilon_{\pm}^2-\delta,\varepsilon_{\pm}^2+\delta]$, respectively, which smoothly connect the constant parts of μ_ε with the map $t\mapsto t^{\frac{r-2}{2}}=\mu(t)$ on $[\varepsilon_-^2+\delta,\varepsilon_+^2-\delta]$, i.e.,

$$\mu_{\delta,\varepsilon}(t)=\begin{cases} g^-_{\delta,\varepsilon}(\varepsilon_-^2-\delta) & 0\le t\le \varepsilon_-^2-\delta\\ g^-_{\delta,\varepsilon}(t) & \varepsilon_-^2-\delta<t\le\varepsilon_-^2+\delta\\ t^{\frac{r-2}{2}} & \varepsilon_-^2+\delta<t\le\varepsilon_+^2-\delta\\ g^+_{\delta,\varepsilon}(t) & \varepsilon_+^2-\delta<t\le\varepsilon_+^2+\delta\\ g^+_{\delta,\varepsilon}(\varepsilon_+^2+\delta) & t>\varepsilon_+^2+\delta. \end{cases}$$

A straightforward calculation reveals that the properties (a)–(d) are satisfied for

$$g^-_{\delta,\varepsilon}(t)=\frac{\left(\varepsilon_-^2+\delta\right)^{\frac{r-4}{2}}(r-2)}{8\delta}t^2-\frac{\left(\varepsilon_-^2+\delta\right)^{\frac{r-4}{2}}\left(\varepsilon_-^2-\delta\right)(r-2)}{4\delta}t\\ -\frac{\left(\varepsilon_-^2+\delta\right)^{\frac{r-2}{2}}\left(-14\delta+2\varepsilon_-^2+3\delta r-\varepsilon_-^2 r\right)}{8\delta}$$

and

$$g^+_{\delta,\varepsilon}(t)=-\frac{\left(\varepsilon_+^2-\delta\right)^{\frac{r-4}{2}}(r-2)}{8\delta}t^2+\frac{\left(\varepsilon_+^2-\delta\right)^{\frac{r-4}{2}}\left(\varepsilon_+^2+\delta\right)(r-2)}{4\delta}t\\ -\frac{\left(\varepsilon_+^2-\delta\right)^{\frac{r-2}{2}}\left(-14\delta-2\varepsilon_+^2+3\delta r+\varepsilon_+^2 r\right)}{8\delta}.$$

References

1. Baranger, J., Najib, K.: Analyse numérique des écoulements quasi-newtoniens dont la viscosité obéit à la loi puissance ou la loi de Carreau. Numer. Math. **58**(1), 35–49 (1990)
2. Barrett, J.W., Liu, W.B.: Finite element error analysis of a quasi-Newtonian flow obeying the Carreau or power law. Numer. Math. **64**(4), 433–453 (1993)
3. Carreau, P.J.: Rheological equations from molecular network theories. Trans. Soc. Rheol. **16**(1), 99–127 (1972)
4. Diening, L., Fornasier, M., Tomasi, R., Wank, M.: A relaxed Kačanov iteration for the *p*-poisson problem. Numer. Math. **145**(1), 1–34 (2020)

5. Fučík, S., Kratochvíl, A., Nečas, J.: Kačanov-Galerkin method. Comment. Math. Univ. Carolinae **14**, 651–659 (1973) (**MR 365300**)
6. Gantner, G., Haberl, A., Praetorius, D., Stiftner, B.: Rate optimal adaptive FEM with inexact solver for nonlinear operators. IMA J. Numer. Anal. **38**(4), 1797–1831 (2018)
7. Garau, E.M., Morin, P., Zuppa, C.: Convergence of an adaptive Kačanov FEM for quasi-linear problems. Appl. Numer. Math. **61**(4), 512–529 (2011)
8. Han, W., Jensen, S., Shimansky, I.: The Kačanov method for some nonlinear problems. Appl. Numer. Meth. **24**, 57–79 (1997)
9. Heid, P., Praetorius, D., Wihler, T.P.: Energy contraction and optimal convergence of adaptive iterative linearized finite element methods. Comput. Methods Appl. Math. **21**(2), 407–422 (2021) (**MR 4235817**)
10. Heid, P., Wihler, T.P.: Adaptive iterative linearization Galerkin methods for nonlinear problems. Math. Comput. **89**(326), 2707–2734 (2020)
11. Heid, P., Wihler, T.P.: On the convergence of adaptive iterative linearized Galerkin methods. Calcolo **57**(3), 24 (2020) (**MR 4131951**)
12. Hirn, A.: Finite element approximation of singular power-law systems. Math. Comput. **82**(283), 1247–1268 (2013)
13. Kačanov, L.M.: Variational methods of solution of plasticity problems. J. Appl. Math. Mech. **23**, 880–883 (1959)
14. Kačur, J., Nečas, J., Polák, J., Souček, J.: Convergence of a method for solving the magnetostatic field in nonlinear media. Apl. Mat. **13**(6), 456–465 (1968)
15. Michlin, S.G.: Čislennaja Realizacija Variacionnych Metodov. Nauka, Izd (1966)
16. Nečas, J.: Introduction to the Theory of Nonlinear Elliptic Equations. Wiley, New Jersey (1986)
17. Neff, P., Pauly, D., Witsch, K.-J.: Poincaré meets Korn via Maxwell: extending Korn''s first inequality to incompatible tensor fields. J. Differ. Equ. **258**(4), 1267–1302 (2015)
18. Senning, J.R.: *Computing and Estimating the Rate of Convergence*, 2007, online lecture notes available from: http://fourier.eng.hmc.edu/e176/lectures/rate.pdf
19. Shapiro, A.: On concepts of directional differentiability. J. Optim. Theory Appl. **66**(3), 477–487 (1990) (**MR 1080259**)
20. Tadmor, Z., Gogos, C.G.: Principles of Polymer Processing. Wiley, EngineeringPro collection, New Jersey (2006)
21. Zeidler, E.: Nonlinear Functional Analysis and its Applications. Springer, New York, II/B (1990)