ESAIM: M2AN 56 (2022) 433–450
https://doi.org/10.1051/m2an/2022008

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

A MODIFIED KAČANOV ITERATION SCHEME WITH APPLICATION TO QUASILINEAR DIFFUSION MODELS

Pascal Heid[1,*] and Thomas P. Wihler[2]

Abstract. The classical Kačanov scheme for the solution of nonlinear variational problems can be interpreted as a fixed point iteration method that updates a given approximation by solving a linear problem in each step. Based on this observation, we introduce a *modified Kačanov method*, which allows for (adaptive) damping, and, thereby, to derive a new convergence analysis under more general assumptions and for a wider range of applications. For instance, in the specific context of quasilinear diffusion models, our new approach does no longer require a standard monotonicity condition on the nonlinear diffusion coefficient to hold. Moreover, we propose two different adaptive strategies for the practical selection of the damping parameters involved.

Mathematics Subject Classification. 35J62, 47J25, 47H05, 47H10, 65J15, 65N12.

Received March 8, 2021. Accepted January 13, 2022.

1. Introduction

In this article we focus on a novel iterative Kačanov type procedure for the solution of quasilinear elliptic partial differential equations (PDE) of the form

$$-\operatorname{div}\Big\{\mu\Big(|\nabla u|^2\Big)\nabla u\Big\} = f \qquad \text{in } \Omega \tag{1.1a}$$

$$u = g \qquad \text{on } \Gamma_1 \tag{1.1b}$$

$$-\mu\Big(|\nabla u|^2\Big)\nabla u\cdot\mathbf{n} = h \qquad \text{on } \Gamma_2, \tag{1.1c}$$

where $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, is a non-empty, open, and bounded domain with a Lipschitz boundary $\partial\Omega$. We suppose that $\partial\Omega = \overline{\Gamma}_1 \cup \overline{\Gamma}_2$ is composed of a Dirichlet boundary part $\Gamma_1 \neq \emptyset$ (of non-zero surface measure) and a Neumann boundary part Γ_2, and $\mathbf{n}$ denotes the unit outward normal vector on Γ_2. Moreover, μ is a real-valued diffusion coefficient, and g and h are Dirichlet and Neumann boundary condition functions, respectively. Nonlinear equations of this type are widely used in mathematical models of physical applications including, for instance, hydro- and gas-dynamics, as well as elasticity and plasticity, see, *e.g.*, [9] and the references therein; we further refer to Sections 69.2 and 69.3 of [16] and Section 1.1 of [1] for a discussion of the physical meaning.

Keywords and phrases. Quasilinear elliptic PDE, strongly monotone problems, fixed point iterations, Kačanov method, quasi-Newtonian fluids, shear-thickening fluids.

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

[2] Mathematics Institute, University of Bern, CH-3012 Bern, Switzerland.

*Corresponding author: pascal.heid@maths.ox.ac.uk

For a given initial guess u^0 with $u^0 = g$ on Γ_1, the traditional Kačanov scheme for the solution of (1.1) is given by

$$-\operatorname{div}\Big\{\mu\Big(|\nabla u^n|^2\Big)\nabla u^{n+1}\Big\} = f \qquad \text{in } \Omega \tag{1.2a}$$

$$u^{n+1} = g \qquad \text{on } \Gamma_1 \tag{1.2b}$$

$$-\mu\Big(|\nabla u^n|^2\Big)\nabla u^{n+1}\cdot\mathbf{n} = h \qquad \text{on } \Gamma_2; \tag{1.2c}$$

this iteration scheme was originally introduced by Kačanov in [15] in the context of variational methods for plasticity problems. Observing that the above boundary value problem is a linear PDE for u^{n+1} (given u^n), we can see Kačanov's scheme as an *iterative linearization method*, *cf.*, the general abstract iterative linearization methodology in [12].

In the literature, standard assumptions on the nonlinearity μ, which guarantee the convergence of (1.2), are expressed as follows, see, *e.g.*, [7,9,12,17]:

(μ1) The diffusion function $\mu : [0,\infty) \to [0,\infty)$ is continuously differentiable;
(μ2) The diffusion function μ is decreasing, *i.e.*, $\mu'(t) \le 0$ for all $t \ge 0$;
(μ3) There are positive constants m_μ and M_μ such that $m_\mu \le \mu(t) \le M_\mu$ for all $t \ge 0$;
(μ4) There exists a positive constant $c_\mu > 0$ such that $2\mu'(t^2)t^2 + \mu(t^2) \ge c_\mu$ for all $t \ge 0$; if we let

$$\phi(t) := \int_0^t \mu(s^2)s\,\mathrm{d}s, \qquad t \ge 0, \tag{1.3}$$

then we note that this condition implies that ϕ is strictly convex.

Examples satisfying the assumptions (μ1)–(μ4) can be found, *e.g.*, in [9] and the references therein.

In this work we address the open question of whether the monotonicity assumption (μ2) is necessary or not for the convergence of the Kačanov scheme. For instance, this assumption can be omitted when the Picard (also termed Zarantonello) iteration is applied, *cf.*, Proposition 5.1 of [12]; in that context, we mention the recent work [8], which, for a closely related type of problem, proves convergence under quasi-optimal cost of a numerical scheme composed of an adaptive interplay of finite element discretizations, Picard iterations, and a contractive linear algebra solver. Since the Kačanov scheme outperforms the Picard iteration in general, however, it would be of practical relevance if the former could as well be applied without imposing assumption (μ2). Here, numerical experiments in [7,12] indicate that it may indeed be dropped. Based on this observation, in this work, we will introduce a *modified Kačanov iteration method that converges without imposing condition* (μ2), and, thereby, allows for the application to a considerably wider range of physical models. For instance, in the context of quasi-Newtonian fluids, the analysis of the traditional Kačanov method is limited to shear-thinning materials corresponding to a decreasing viscosity coefficient, whilst our new scheme can be applied, in addition, to shear-thickening substances, which have an increasing viscosity coefficient μ.

We note that the classical Kačanov method was already applied to incompressible generalized Newtonian fluid flow problems with a power-law like rheology of possibly shear-thickening fluids in [2]; in that work, however, very restrictive conditions needed to be imposed in order to show the convergence of the sequence of iterates to a solution of a regularized problem. Moreover, an advanced convergence analysis for the Kačanov scheme has been carried out for a relaxed p-Poisson problem in [5]; those results are restricted to decreasing diffusion coefficients μ, however, they apply for problems with nonlinear p-growth, namely for $p \in (1,2]$, *i.e.*, assumption (μ3), which essentially expresses linear growth, is not required there. We remark that the modified Kačanov scheme introduced in the present work could possibly be applied as a subroutine in place of the classical Kačanov scheme in [5] for the case $p > 2$. Finally, we point to [11], where the convergence rate of the Kačanov scheme was studied for a class of decreasing diffusion coefficients, which correspond to the viscosity function of shear-thinning fluids with zero and infinite shear plateaus, *i.e.*, (μ3) is satisfied. Indeed, the analysis in [11] can also

be applied to the generalized Stokes problem, *cf.*, Section 4.1 of [10], where the convergence is shown for the Bercovier–Engelman regularization of steady Bingham fluids.

We emphasize once more that a key prerequisite in the convergence analysis of the Kačanov scheme in [5,10, 11], as well as in the classical proof, is the monotonicity of the diffusion coefficient; in fact, in all the convergence proofs of the Kačanov scheme we are aware of, except for the one in [2], it is shown that an underlying (energy) functional decays along the sequence generated by the Kačanov scheme, for which, again, it is standard to assume that the diffusion coefficient is monotonically decreasing. The proof in the present work is also based on the decay of the energy functional, which, however, can be obtained without imposing the assumption ($\mu 2$) when a damping parameter is introduced.

The key idea in devising the modified scheme together with the proof of convergence is based on the fact that (1.2) can be cast into the unified iteration scheme introduced in [12], see also [13]. To sketch the idea, for $\Gamma_2 = \emptyset$, upon defining the PDE residual

$$\mathsf{F} := -\operatorname{div}\Big\{\mu\Big(|\nabla(\cdot)|^2\Big)\nabla(\cdot)\Big\} - f,$$

and, for given u, the *linear preconditioning operator*

$$\mathsf{A}(u)(\cdot) := -\operatorname{div}\Big\{\mu\Big(|\nabla u|^2\Big)\nabla(\cdot)\Big\},$$

the iterative procedure (1.2) can be written (formally) in terms of the *fixed point iteration*

$$u^{n+1} = u^n - \mathsf{A}(u^n)^{-1}\mathsf{F}(u^n), \qquad n \geq 0.$$

With the aim of obtaining an improved control of the updates in each step, we introduce a *step size parameter* $\delta(u^n) > 0$ in the iteration, *viz.*

$$u^{n+1} = u^n - \delta(u^n)\mathsf{A}(u^n)^{-1}\mathsf{F}(u^n), \qquad n \geq 0.$$

This yields the *modified Kačanov method* proposed and analyzed in this work.

Outline

We begin by deriving an appropriate framework for abstract nonlinear variational problems in Section 2. In particular, we introduce a modified version of the classical Kačanov iteration scheme, and prove a new convergence result under assumptions that are milder than in the classical setting. The purpose of Section 3 is to devise two different adaptive strategies for the selection of the damping parameters in the modified method. Subsequently, our general theory is applied to quasilinear diffusion models in Section 4, which also contains a numerical study within the framework of finite element discretizations. Finally, we add some concluding remarks in Section 5.

2. Abstract analysis

Throughout, Y is a reflexive real Banach space, equipped with a norm denoted by $\|\cdot\|_Y$, and $K \subset Y$ is a closed, convex subset.

2.1. Nonlinear variational problems

Consider a (nonlinear) Gâteaux continuously differentiable functional $\mathsf{H} : K \to \mathbb{R}$ that has a strongly monotone Gâteaux derivative, *i.e.*, there exists a constant $\nu > 0$ such that

$$\langle \mathsf{H}'(u) - \mathsf{H}'(v), u - v\rangle \geq \nu\|u - v\|_Y^2 \qquad \forall u, v \in K, \tag{2.1}$$

where $\langle\cdot,\cdot\rangle$ is the duality product on $Y^\star \times Y$, with $Y^\star$ signifying the dual space of Y.

Proposition 2.1. *Suppose that* $\mathsf{H} : K \to \mathbb{R}$ *is a (Gâteaux) continuously differentiable functional that satisfies the strong monotonicity condition* (2.1)*. Then, there exists a unique minimizer* $u^\star \in K$ *of* H *in* K*, i.e.,* $\mathsf{H}(u^\star) \leq \mathsf{H}(v)$ *for all* $v \in K$*. Furthermore,* $u^\star \in K$ *is the unique solution of the weak inequality*

$$\langle \mathsf{H}'(u^\star), v - u^\star \rangle \geq 0 \qquad \forall v \in K. \tag{2.2}$$

Proof. We follow along the lines of the proof of Theorem 25.L from [17].

(1) Since H is a Gâteaux continuously differentiable functional on K, it is, in particular, continuous on K. Moreover, from (2.1) we infer that H is strictly convex, see, *e.g.*, Proposition 25.10 of [17]. These two properties, in turn, imply that H is weakly sequentially lower semicontinuous, see Proposition 25.20 of [17].

(2) If the set K is bounded, then the functional H, being weakly sequentially lower semicontinuous, has a minimum on K, see Theorem 25.C of [17]. Otherwise, if K is unbounded, then we show that H is weakly coercive. To this end, take any $u, v \in K$, and define the scalar function

$$\varphi(t) := \mathsf{H}(u + t(v - u)), \qquad t \in [0, 1]; \tag{2.3}$$

since K is convex, note that $u + t(v - u) \in K$ for all $t \in [0, 1]$. Applying the chain rule reveals that

$$\varphi'(t) = \langle \mathsf{H}'(u + t(v - u)), v - u \rangle. \tag{2.4}$$

Thus, by virtue of the fundamental theorem of calculus, we deduce that

$$\begin{aligned} \mathsf{H}(v) - \mathsf{H}(u) &= \int_0^1 \langle \mathsf{H}'(u + t(v - u)), v - u \rangle \, \mathrm{d}t \\ &= \int_0^1 \langle \mathsf{H}'(u + t(v - u)) - \mathsf{H}'(u), v - u \rangle \, \mathrm{d}t + \langle \mathsf{H}'(u), v - u \rangle. \end{aligned}$$

Therefore, exploiting (2.1), it follows that

$$\mathsf{H}(v) - \mathsf{H}(u) \geq \frac{\nu}{2} \|v - u\|_Y^2 - \|\mathsf{H}'(u)\|_{Y^\star} \|v - u\|_Y.$$

Hence, we see that $\mathsf{H}(v) \to \infty$ for $\|v\|_Y \to \infty$, *i.e.*, H is weakly coercive on K. Then, owing to Theorem 25.D of [17], we conclude that H has a minimum $u^\star$ in K. We note that this minimum is unique since H is strictly convex.

(3) Finally, if $u^\star$ is the minimum of H in K, then the function $\varphi(t)$ from (2.3), with $u = u^\star$, has a minimum at $t = 0$. This implies that $\varphi'(0) \geq 0$. In turn, exploiting (2.4), this holds true if and only if $\langle \mathsf{H}'(u^\star), v - u^\star \rangle \geq 0$, which yields (2.2). Conversely, since H' is strongly monotone, *cf.*, (2.1), and $u^\star$ satisfies the weak inequality (2.2), the function $\varphi(t)$ is increasing on $[0, 1]$. In fact, for $t \in (0, 1]$, we have

$$\begin{aligned} \varphi'(t) &= \frac{1}{t} \langle \mathsf{H}'(u^\star + t(v - u^\star)) - \mathsf{H}'(u^\star), t(v - u^\star) \rangle + \langle \mathsf{H}'(u^\star), v - u^\star \rangle \\ &\geq \frac{1}{t} \langle \mathsf{H}'(u^\star + t(v - u^\star)) - \mathsf{H}'(u^\star), t(v - u^\star) \rangle \geq \nu t \|v - u^\star\|_Y^2 \geq 0. \end{aligned}$$

Therefore, we obtain $\mathsf{H}(v) = \varphi(1) \geq \varphi(0) = \mathsf{H}(u^\star)$, *i.e.*, $u^\star$ is the minimum of H in K since v was arbitrary.

This completes the proof. □

Remark 2.2. We emphasize that the application of Theorems 25.C and 25.D from [17] in the proof of Proposition 2.1 require the space Y to be reflexive. More precisely, these results make use of the fact that every bounded sequence in a reflexive Banach space has a weakly convergent subsequence.

Consider the following assumption on the closed and convex subset K:

(K) The set $X := \{u - v : u, v \in K\}$ is a linear closed subspace of Y, and $x + y \in K$ for all $x \in X$ and $y \in K$.

Corollary 2.3. *Suppose that the subset K has property* (K)*, and let the assumptions of Proposition 2.1 hold true. Then, the unique minimizer $u^\star \in K$ of H satisfies*

$$\langle \mathsf{H}'(u^\star), v\rangle = 0 \qquad \forall v \in X. \tag{2.5}$$

Proof. Let $u^\star \in K$ be the unique minimizer of H on K. Then, for any $v \in X$, owing to property (K), it holds that $v + u^\star \in K$. Thus, using (2.2), we have $0 \le \langle \mathsf{H}'(u^\star), (v + u^\star) - u^\star\rangle = \langle \mathsf{H}'(u^\star), v\rangle$. Similarly, upon replacing v by $-v$, we infer that $0 \le -\langle \mathsf{H}'(u^\star), v\rangle$, which concludes the argument. □

2.2. Modified Kačanov method

We consider mappings $a : K \times Y \times X \to \mathbb{R}$ and $b : K \times X \to \mathbb{R}$, with X being the linear subspace from property (K) above, which satisfy the following properties:

(A1) For any given $u \in K$, we suppose that $a(u; \cdot, \cdot)$ is a bilinear form on $Y \times X$, and $b(u, \cdot) \in X^\star$; in the sequel, we use the notation $b(u, \cdot) = \langle b(u), \cdot\rangle$, where the dual product is evaluated on the space $X^\star \times X$.

(A2) There exist positive constants $\alpha, \beta > 0$ such that, for any $u \in K$, the form $a(u; \cdot, \cdot)$ is uniformly bounded on $Y \times X$ and coercive on $X \times X$ in the sense that

$$a(u; v, w) \le \beta \|v\|_Y \|w\|_Y \qquad \forall v \in Y,\ \forall w \in X, \tag{2.6}$$

and

$$a(u; v, v) \ge \alpha \|v\|_Y^2 \qquad \forall v \in X, \tag{2.7}$$

respectively; in particular, if the set K satisfies property (K), then it follows that

$$a(u; v - w, v - w) \ge \alpha \|v - w\|_Y^2 \qquad \forall v, w \in K. \tag{2.8}$$

(A3) There are Gâteaux continuously differentiable functionals $\mathsf{G} : K \to \mathbb{R}$ and $\mathsf{B} : K \to \mathbb{R}$ such that, for all $u \in K$, it holds $\mathsf{G}'(u)|_X = a(u; u, \cdot)$ and $\mathsf{B}'(u)|_X = b(u)$ in $X^\star$.

(A4) The (continuously differentiable) functional $\mathsf{H} : K \to \mathbb{R}$ given by $\mathsf{H}(u) := \mathsf{G}(u) - \mathsf{B}(u)$, $u \in K$, with G and B from (A3), satisfies the strong monotonicity condition (2.1).

If the closed and convex subset $K \subset Y$ fulfils property (K), then the unique minimizer $u^\star \in K$ of the functional H from (A4) solves the weak formulation

$$0 = \langle \mathsf{H}'(u^\star), v\rangle = \langle \mathsf{G}'(u^\star) - \mathsf{B}'(u^\star), v\rangle = a(u^\star; u^\star, v) - \langle b(u^\star), v\rangle \qquad \forall v \in X, \tag{2.9}$$

cf., Corollary 2.3. Now, for given $u \in K$, define the *linear* operator $\mathsf{A}(u) : Y \to X^\star$, $v \mapsto \mathsf{A}(u)v$, by

$$\langle \mathsf{A}(u)v, w\rangle = a(u; v, w) \qquad \forall w \in X.$$

Then, the weak formulation (2.9) can be expressed by

$$\mathsf{A}(u^\star)u^\star = b(u^\star) \qquad \text{in } X^\star.$$

In light of (A2), for any $u \in K$, we notice that $a(u; \cdot, \cdot)$ is a bounded and coercive bilinear form on the closed subspace $X \times X$. In particular, thanks to the Lax–Milgram theorem, for any $u \in K$ and $\ell \in X^\star$, we conclude that there exists a unique $w_{u,\ell} \in X$ such that $\mathsf{A}(u)w_{u,\ell} = \ell$ in $X^\star$, *i.e.*, $\mathsf{A}(u)|_X : X \to X^\star$ is invertible for any $u \in K$. Hence, noticing that

$$\mathsf{F}(u) := \mathsf{H}'(u) = \mathsf{A}(u)u - b(u) \in X^\star, \tag{2.10}$$

the classical Kačanov method in abstract form, for given $u^n \in K$, reads as

$$u^{n+1} = u^n - \rho^n, \qquad n \geq 0, \tag{2.11a}$$

where $\rho^n \in X$ is uniquely defined through

$$\mathsf{A}(u^n)\rho^n = \mathsf{F}(u^n) \quad \text{in } X^\star. \tag{2.11b}$$

A modification of this procedure is obtained by invoking a parameter $\delta : K \to (0, \infty)$, thereby yielding the new scheme

$$u^{n+1} = u^n - \delta(u^n)\rho^n, \qquad n \geq 0, \tag{2.12}$$

with ρ^n as in (2.11b). Equivalently, upon introducing the forms

$$a_{\delta(u)}(u; v, w) := \frac{1}{\delta(u)} a(u; v, w), \qquad b_{\delta(u)}(u) := \frac{1}{\delta(u)} a(u; u, \cdot) - \mathsf{F}(u), \tag{2.13}$$

for $u \in K, v \in Y$, and $w \in X$, we derive the *modified Kačanov iteration* in weak form:

$$u^{n+1} \in K : \qquad a_{\delta^n}(u^n; u^{n+1}, v) = \langle b_{\delta^n}(u^n), v\rangle \qquad \forall v \in X, \qquad n \geq 0, \tag{2.14}$$

where we use the notation $\delta^n := \delta(u^n)$. Clearly, for $\delta \equiv 1$, the traditional Kačanov scheme (2.11) is recovered.

Proposition 2.4. *Suppose* (A1)–(A4)*, and that K satisfies property* (K)*. Then, for any initial guess $u^0 \in K$, the modified Kačanov iteration* (2.14) *remains well-defined for each $n \geq 0$, i.e., for given $u^n \in K$, the solution $u^{n+1} \in K$ of the weak formulation* (2.14) *exists and is unique.*

Proof. For fixed $u^n \in K$, the solution $\rho^n \in X$ of (2.11b) exists and is unique since $\mathsf{A}(u^n)|_X : X \to X^\star$ is invertible. Moreover, owing to property (K), we infer that $u^{n+1} \in K$ in (2.12). □

2.3. Convergence analysis

We are now in the position to state and prove the main result of our work.

Theorem 2.5. *Given* (A1)–(A4) *and* (K)*. We further assume the following conditions:*

(a) H' *is continuous with respect to the weak topology on $X^\star$ in the sense that, for any sequence $\{z^n\}_{n\geq 0} \subset K$ with a limit $z^\star \in K$, it holds that*

$$\lim_{n\to\infty} \langle \mathsf{H}'(z^n), w\rangle = \langle \mathsf{H}'(z^\star), w\rangle \qquad \forall w \in X; \tag{2.15}$$

(b) *there exists a damping strategy such that $\delta(u^n) \geq \delta_{\min} > 0$ and*

$$\mathsf{H}(u^n) - \mathsf{H}\big(u^{n+1}\big) \geq \gamma \big\|u^{n+1} - u^n\big\|_Y^2 \qquad \forall n \geq 0, \tag{2.16}$$

for some constants $\delta_{\min}, \gamma > 0$ independent of n.

Then, the damped Kačanov iteration (2.12) *converges to the unique solution $u^\star \in K$ of* (2.5) *for any initial guess $u^0 \in K$.*

Proof. We will proceed in three steps: First, we show that the difference of two consecutive iterates, *i.e.*, $\big\|u^{n+1} - u^n\big\|_Y$, tends to zero as $n \to \infty$. Subsequently, we will verify the convergence of $\{u^n\}_{n\geq 0}$, and finally that the limit equals to $u^\star$. For this purpose, we will essentially follow along the lines of the proof of Proposition 2.1 from [12]; see also the closely related argument in Theorem 25.L of [17].

(1) In light of Proposition 2.1 we recall that H is bounded from below by $\mathsf{H}(u^\star)$. Moreover, $\{\mathsf{H}(u^n)\}_{n\geq 0}$ is decreasing thanks to the assumption (2.16). Hence, this sequence converges, and we conclude that

$$0 \leq \gamma \big\|u^{n+1} - u^n\big\|_Y^2 \leq \mathsf{H}\big(u^{n+1}\big) - \mathsf{H}(u^n) \to 0 \qquad \text{as } n \to \infty;$$

i.e., $\lim_{n\to\infty} \big\|u^{n+1} - u^n\big\|_Y = 0$.

(2) Next, we shall verify the existence of the limit of the sequence $\{u^n\}_{n\geq 0}$. By the strong monotonicity (2.1), for any $m \geq n \geq 0$, it holds that

$$\nu \|u^m - u^n\|_Y^2 \leq \langle \mathsf{H}'(u^m) - \mathsf{H}'(u^n), \epsilon_{m,n}\rangle,$$

with $\epsilon_{m,n} := u^m - u^n \in X$. Combining (2.10) and (2.13), for $\delta = \delta(u) > 0$, we note that

$$\mathsf{H}'(u) = \mathsf{F}(u) = a_\delta(u; u, \cdot) - b_\delta(u) \qquad \forall u \in K \tag{2.17}$$

in $X^\star$, and thus,

$$\begin{aligned}\nu \|\epsilon_{m,n}\|_Y^2 \leq\ & a_{\delta^m}(u^m; u^m, \epsilon_{m,n}) - \langle b_{\delta^m}(u^m), \epsilon_{m,n}\rangle \\ & - a_{\delta^n}(u^n; u^n, \epsilon_{m,n}) + \langle b_{\delta^n}(u^n), \epsilon_{m,n}\rangle.\end{aligned}$$

Using (2.14), this further leads to

$$\nu \|\epsilon_{m,n}\|_Y^2 \leq a_{\delta^m}(u^m; u^m - u^{m+1}, \epsilon_{m,n}) - a_{\delta^n}(u^n; u^n - u^{n+1}, \epsilon_{m,n}).$$

Applying (2.6) yields

$$\nu \|\epsilon_{m,n}\|_Y^2 \leq \frac{\beta}{\delta_{\min}} \|\epsilon_{m,n}\|_Y \big(\big\|u^m - u^{m+1}\big\|_Y + \big\|u^n - u^{n+1}\big\|_Y\big),$$

and thus

$$\|\epsilon_{m,n}\|_Y \leq \frac{\beta}{\nu\delta_{\min}} \big(\big\|u^m - u^{m+1}\big\|_Y + \big\|u^n - u^{n+1}\big\|_Y\big).$$

From the first step of the proof, we conclude that $\{u^n\}_{n\geq 0}$ is a Cauchy sequence in K. Since K is a closed subset of a Banach space, the sequence $\{u^n\}_{n\geq 0}$ has a limit $\mathfrak{u} \in K$.

(3) It remains to verify that $\mathfrak{u}$ is a solution of (2.5). To this end, from (2.17) and (2.14) it follows that

$$a_{\delta^n}(u^n; u^{n+1} - u^n, w) + \langle \mathsf{H}'(u^n), w\rangle = a_{\delta^n}(u^n; u^{n+1}, w) - \langle b_{\delta^n}(u^n), w\rangle = 0,$$

for all $w \in X$. Recalling that $\{u^{n+1} - u^n\}_{n\geq 0}$ is a vanishing sequence in X, and exploiting (2.6), we have that $a_\delta(u^n; u^{n+1} - u^n, w) \to 0$ as $n \to \infty$, for any $w \in X$. Moreover, by the weak continuity property (2.15) of H', we obtain

$$\langle \mathsf{H}'(\mathfrak{u}), w\rangle = \lim_{n\to 0} \langle \mathsf{H}'(u^n), w\rangle = 0 \qquad \forall w \in X,$$

i.e., $\mathfrak{u}$ is a solution of (2.5). Since the solution is unique thanks to Proposition 2.1 and Corollary 2.3, we infer that $\mathfrak{u} = u^\star$. This completes the argument.

□

Remark 2.6. The classical convergence theory for the (standard) Kačanov method requires the following key inequality to hold:

$$\mathsf{G}(u) - \mathsf{G}(v) \geq \frac{1}{2}(a(u; u, u) - a(u; v, v)) \qquad \forall u, v \in K; \tag{2.18}$$

see, *e.g.*, Theorem 25.L and equation (106) of [17]. In order to verify (2.18) in the context of our model problem (1.1), the monotonicity assumption ($\mu 2$) is decisive. On the contrary, the analysis in our present work is based on the bound (2.16) which, in turn, allows to omit the monotonicity of the (nonlinear) diffusion coefficient μ in the application to quasilinear elliptic PDE (1.1); see Theorem 4.4 below. Furthermore, in contrast to the traditional framework, the operator B from assumption (A3) does not need to be linear in our analysis, and, in addition, the symmetry of $a(u,\cdot,\cdot)$ is no longer necessary. We remark that these improvements come at the expense of condition (K) as well as of the crucial estimate (2.16); in the context of quasilinear elliptic PDE (1.1), however, these assumptions do not implicate any drawback. Finally, we note that if B is linear, then (2.18) implies (2.16) with $\gamma = \alpha/2$.

The next result states that if H' is Lipschitz continuous in the sense that there exists a constant $L_{\mathsf{H}} > 0$ such that

$$\langle \mathsf{H}'(u) - \mathsf{H}'(v), u - v\rangle \le L_{\mathsf{H}} \|u - v\|_Y^2 \qquad \forall u, v \in K, \tag{2.19}$$

then our key assumption (2.16) from Theorem 2.5 is satisfied for sufficiently small damping parameters.

Corollary 2.7. *Assume* (A1)–(A4) *and* (K), *and suppose that* (2.19) *holds. Then, we have the estimate*

$$\mathsf{H}(u^n) - \mathsf{H}\big(u^{n+1}\big) \ge \left(\frac{\alpha}{\delta^n} - \frac{L_{\mathsf{H}}}{2}\right) \big\|u^{n+1} - u^n\big\|_Y^2,$$

where $\alpha > 0$ *is the constant from* (2.7). *In particular, if* $\delta^n \in [\delta_{\min}, \delta_{\max}]$, *with* $0 < \delta_{\min} \le \delta_{\max} < 2\alpha/L_{\mathsf{H}}$, *for all* $n \ge 0$, *and* H' *is continuous with respect to the weak topology on* $X^\star$, *cf., condition* (a) *from Theorem 2.5, then the damped Kačanov iteration* (2.12) *converges to the unique solution* $u^\star \in K$ *of* (2.5) *for any initial guess* $u^0 \in K$.

Proof. Our argument follows along the lines of the proof of Theorem 2.6 from [12], however, with a different bilinear form on account of the present iteration scheme (2.14). Similarly as in the proof of Proposition 2.1, we define the scalar function $\varphi(t) := \mathsf{H}(u^n + t(u^{n+1} - u^n))$, for $t \in [0,1]$ and $n \ge 0$. Then, we find that

$$\begin{aligned} \mathsf{H}\big(u^{n+1}\big) - \mathsf{H}(u^n) &= \int_0^1 \big\langle \mathsf{H}'(u^n + t(u^{n+1} - u^n)), u^{n+1} - u^n\big\rangle \,\mathrm{d}t \\ &= \int_0^1 \big\langle \mathsf{H}'(u^n + t(u^{n+1} - u^n)) - \mathsf{H}'(u^n), u^{n+1} - u^n\big\rangle \,\mathrm{d}t \\ &\quad + \big\langle \mathsf{H}'(u^n), u^{n+1} - u^n\big\rangle. \end{aligned} \tag{2.20}$$

Hence, by invoking the Lipschitz continuity (2.19), the identity (2.17), and the modified Kačanov scheme (2.14), we obtain

$$\begin{aligned} \mathsf{H}\big(u^{n+1}\big) - \mathsf{H}(u^n) &\le \frac{L_{\mathsf{H}}}{2}\big\|u^{n+1} - u^n\big\|_Y^2 + a_{\delta^n}(u^n; u^n, u^{n+1} - u^n) - \big\langle b_{\delta^n}(u^n), u^{n+1} - u^n\big\rangle \\ &= \frac{L_{\mathsf{H}}}{2}\big\|u^{n+1} - u^n\big\|_Y^2 - a_{\delta^n}(u^n; u^{n+1} - u^n, u^{n+1} - u^n). \end{aligned}$$

Furthermore, employing the coercivity assumption (2.8), it follows that

$$\mathsf{H}\big(u^{n+1}\big) - \mathsf{H}(u^n) \le \left(\frac{L_{\mathsf{H}}}{2} - \frac{\alpha}{\delta^n}\right) \big\|u^{n+1} - u^n\big\|_Y^2.$$

Moreover, if $\delta^n \le \delta_{\max}$ for all $n \ge 0$, then we further deduce the bound

$$\mathsf{H}(u^n) - \mathsf{H}\big(u^{n+1}\big) \ge \left(\frac{\alpha}{\delta_{\max}} - \frac{L_{\mathsf{H}}}{2}\right) \big\|u^{n+1} - u^n\big\|_Y^2. \tag{2.21}$$

If $\delta_{\max} < 2\alpha/L_{\mathsf{H}}$, then $(\alpha/\delta_{\max} - L_{\mathsf{H}}/2) > 0$, and, in turn, Theorem 2.5 implies the convergence of the sequence $\{u^n\}_{n\ge 0}$ to $u^\star$. □

Remark 2.8. Applying the abstract analysis in [14], given the assumptions of Theorem 2.5, it can be shown that the iterates generated by the modified Kačanov scheme (2.14) satisfy the contraction property

$$\mathsf{H}\big(u^{n+1}\big) - \mathsf{H}(u^\star) \leq q(\mathsf{H}(u^n) - \mathsf{H}(u^\star)) \qquad \forall n \geq 0,$$

for some constant $0 < q < 1$, where $u^\star$ is the solution of (2.5). In particular, in view of Corollary 2.7 with a constant damping parameter $\delta = \delta^n \in (0, {}^{2\alpha}/_{L_\mathsf{H}})$ for all $n \geq 0$, we have that

$$q(\delta) = \left(1 - \frac{2\delta\nu^2(\alpha - {}^{\delta L_\mathsf{H}}/_2)}{\beta^2 L_\mathsf{H}}\right),$$

cf., Theorem 2.1 of [14]. By taking the derivative with respect to δ, it follows immediately that the minimum is attained at $\delta = {}^{\alpha}/_{L_\mathsf{H}}$; noticing that the derivation of q involves some rough estimates, however, this choice is typically suboptimal with regards to the convergence rate.

3. Adaptive step size control

In this section, we will present two adaptive methods for selecting the damping parameter δ^n in the modified Kačanov iteration (2.14). To this end, recall the key inequality (2.16) from Theorem 2.5, and set $\delta_{\max} = {}^{\alpha}/_{L_\mathsf{H}}$ in (2.21) (which is a possibly pessimistic choice as mentioned in Rem. 2.8); then (2.16) holds for $\gamma = {}^{L_\mathsf{H}}/_2$. Alternatively, from Remark 2.6, we recall that within the setting of the classical Kačanov scheme, *i.e.*, for $\delta \equiv 1$, the bound (2.16) can be shown for the constant $\gamma = {}^{\alpha}/_2$ under more restrictive assumptions on the nonlinearity. This observation may suggest that a smaller choice of γ potentially relates to a larger size of the damping parameter. We thus propose that the sequence $\{u^n\}_{n\geq 0}$ is required to satisfy an estimate of the form

$$\mathsf{H}(u^n) - \mathsf{H}\big(u^{n+1}\big) \geq \theta \min\{\alpha, L_\mathsf{H}\}\big\|u^{n+1} - u^n\big\|_Y^2, \tag{3.1}$$

for a constant $0 < \theta \leq {}^1/_2$, which still guarantees the convergence of the modified Kačanov scheme (2.14) in regard to Theorem 2.5 without imposing an upper bound on the damping parameter. In our numerical experiments in Section 4.5, we let $\theta = 0.1$. Moreover, in order to prevent too small steps, we set $\delta_{\min} := {}^{\alpha}/_{4L_\mathsf{H}}$, which, in view of Remark 2.8, is a reasonable choice. We emphasize that the constants α and L_H must both be known *a priori*; in particular, we assume that H' is Lipschitz continuous as proposed in (2.19).

The two adaptive step size procedures to be presented below both pursue a similar strategy, namely, to maximize the difference $\mathsf{H}(u^n) - \mathsf{H}\big(u^{n+1}\big)$ in each step by choosing an appropriate step size $\delta^n = \delta(u^n) \geq \delta_{\min}$. Indeed, recalling that $u^\star$ is the unique minimizer of H in K, it seems obvious that a maximal decay of the functional H along the sequence $\{u^n\}_{n\geq 0}$ will potentially accelerate the convergence of $\{u^n\}_{n\geq 0}$ to $u^\star$.

3.1. Step size control *via* Taylor expansion

We begin by recalling (2.20), which in regard to (2.10), can be stated as

$$\mathsf{H}\big(u^{n+1}\big) - \mathsf{H}(u^n) = \int_0^1 \big\langle \mathsf{F}(u^n + t(u^{n+1} - u^n)), u^{n+1} - u^n \big\rangle \,\mathrm{d}t; \tag{3.2}$$

in particular, in view of the discussion above, we aim to maximize the integral on the right-hand side. For that purpose, we will employ a Taylor expansion of the integrand at $t = 0$, provided that $\mathsf{F} : K \to Y^\star$ from (2.10) is Fréchet differentiable. Specifically, let us first define the (continuously differentiable) function

$$\psi_n(t) := \big\langle \mathsf{F}(u^n + t(u^{n+1} - u^n)), u^{n+1} - u^n \big\rangle, \qquad t \in [0, 1].$$

Then, if the difference $u^{n+1} - u^n$ is sufficiently small, applying a Taylor expansion at $t = 0$ yields

$$\psi_n(t) \approx \psi_n(0) + \psi_n'(0)t = \big\langle \mathsf{F}(u^n), u^{n+1} - u^n \big\rangle + t\big\langle \mathsf{F}'(u^n)(u^{n+1} - u^n), u^{n+1} - u^n \big\rangle.$$

Since (2.12) implies that $u^{n+1} - u^n = -\delta(u^n)\rho^n$, for each $n \geq 0$, we obtain

$$\psi_n(t) \approx -\delta(u^n)\langle \mathsf{F}(u^n), \rho^n\rangle + t\delta(u^n)^2\langle \mathsf{F}'(u^n)\rho^n, \rho^n\rangle, \tag{3.3}$$

where we have exploited the fact that the Fréchet derivative $\mathsf{F}'(u^n)$ is a linear operator. Hence, by recalling (3.2) and integrating (3.3) from $t = 0$ to $t = 1$, we find that

$$\mathsf{H}(u^n) - \mathsf{H}\big(u^{n+1}\big) \approx \delta(u^n)\langle \mathsf{F}(u^n), \rho^n\rangle - \frac{\delta(u^n)^2}{2}\langle \mathsf{F}'(u^n)\rho^n, \rho^n\rangle.$$

Then, a simple calculation reveals that the right-hand side is maximized for the damping parameter

$$\delta^n = \delta(u^n) := \frac{\langle \mathsf{F}(u^n), \rho^n\rangle}{\langle \mathsf{F}'(u^n)\rho^n, \rho^n\rangle}. \tag{3.4}$$

In account of (3.1) and the lower bound $\delta_{\min} = \alpha/4L_{\mathsf{H}}$, this leads to the step size Algorithm 1. We note that the stopping criterion in line 6 will be satisfied once the damping parameter is small enough, *cf.*, Corollary 2.7, *i.e.*, the procedure terminates after finitely many steps; indeed, the stopping criterion is certainly satisfied once we reach $\delta^n = \delta_{\min}$. Moreover, we underline that the derivative $\mathsf{F}'(u^n)$ must be available in the step size Algorithm 1, *cf.*, (3.4).

Algorithm 1. Step size control *via* Taylor expansion.

Input: Given $u^n \in K$, a correction factor $\sigma \in (1/2, 1)$, and a parameter $\theta \in (0, 1/2]$.

1: Solve the linear problem $\mathsf{A}(u^n)\rho^n = \mathsf{F}(u^n)$ for $\rho^n \in X$, *cf.*, (2.11b).
2: Compute δ^n from (3.4) and set $\delta^n \leftarrow \max\{\delta_{\min}, \delta^n\}$.
3: **repeat**
4: Compute $u^{n+1} := u^n - \delta^n\rho^n$, *cf.*, (2.12).
5: Set $\delta^n \leftarrow \max\{\sigma\delta^n, \delta_{\min}\}$.
6: **until** $\mathsf{H}(u^n) - \mathsf{H}\big(u^{n+1}\big) \geq \theta \min\{\alpha, L_{\mathsf{H}}\}\big\|u^{n+1} - u^n\big\|_Y^2$
7: **return** u^{n+1}.

3.2. Step size control *via* a prediction-correction strategy

We will present a second adaptive damping parameter selection procedure that is partially based on ideas from Section 3.1 of [3]. This strategy is more "*ad hoc*" than the Taylor expansion approach above, however, it does not require the differentiability of the operator $\mathsf{F} = \mathsf{H}'$, *cf.*, (2.10). The idea is fairly straightforward: For a given correction factor $\sigma \in (1/2, 1)$ and damping factor $\delta > 0$, we compare the energy decay for the damping parameters δ and $\delta' = \sigma^p\delta$, where $p \in \{-1, 1\}$ depends on the previous step; we note that $p = -1$ yields an increased step size, whereas $p = 1$ decreases the damping parameter. If applying δ' results in a larger energy decay, then we choose the damping parameter to be δ' in the present and subsequent steps, with p unchanged; otherwise, if δ outperforms δ', then δ is retained, however, in the next step we replace p by $-p$. This leads to Algorithm 2.

4. Application to quasilinear diffusion models

In this section, we discuss the weak formulations of the boundary value problem (1.1) as well as of the Kačanov iteration scheme (1.2). In addition, an equivalent variational setting will be established. Furthermore, a series of numerical experiments in the framework of finite element discretizations will be presented.

Algorithm 2. Step size control *via* prediction-correction strategy.

Input: Given $u^n \in K$, a damping parameter $\delta \geq \delta_{\min}$, an exponent $p \in \{-1, 1\}$, a correction factor $\sigma \in (1/2, 1)$, and a parameter $\theta \in (0, 1/2]$.

Let $C := \theta \min\{\alpha, L_H\}$.
Solve the linear problem $\mathsf{A}(u^n)\rho^n = \mathsf{F}(u^n)$ for $\rho^n \in X$, *cf.*, (2.11b).
if $p = 1$ and $\delta < \sigma^{-1}\delta_{\min}$ **then**
    Set $p \leftarrow -1$.
end if
Set $\delta' := \sigma^p \delta$ and compute $\widetilde{u}^{n+1} := u^n - \delta'\rho^n$, *cf.*, (2.12).
if $\mathsf{H}(u^n) - \mathsf{H}\big(\widetilde{u}^{n+1}\big) \geq C\big\|u^n - \widetilde{u}^{n+1}\big\|_Y^2$ **then**
    Compute $u^{n+1} := u^n - \delta\rho^n$, *cf.*, (2.12).
    if $\mathsf{H}\big(\widetilde{u}^{n+1}\big) \leq \mathsf{H}\big(u^{n+1}\big)$ or $\mathsf{H}(u^n) - \mathsf{H}\big(u^{n+1}\big) < C\big\|u^n - u^{n+1}\big\|_Y^2$ **then**
        Set $\delta \leftarrow \delta'$ and $u^{n+1} \leftarrow \widetilde{u}^{n+1}$.
    else
        Set $p \leftarrow -p$.
    end if
else
    Set $p \leftarrow 1$, $\delta \leftarrow \delta'$, and $u^{n+1} := \widetilde{u}^{n+1}$.
    while $\mathsf{H}(u^n) - \mathsf{H}\big(u^{n+1}\big) < C\big\|u^{n+1} - u^n\big\|_Y^2$ **do**
        Set $\delta \leftarrow \sigma\delta$ and compute $u^{n+1} := u^n - \delta\rho^n$, *cf.*, (2.12).
    end while
end if
return δ, u^{n+1}, and p.

4.1. Sobolev spaces

Let $Y := \mathrm{H}^1(\Omega)$ be the standard Sobolev space of $\mathrm{L}^2(\Omega)$-functions with weak derivatives in $\mathrm{L}^2(\Omega)$. We endow Y with the inner product

$$(u, v)_Y := \int_\Omega \nabla u \cdot \nabla v \,\mathsf{dx} + \int_\Omega uv \,\mathsf{dx}, \qquad u, v \in Y,$$

and with the induced H^1-norm $\|u\|_Y := \sqrt{(u,u)_Y}$, $u \in Y$.

Moreover, consider the closed linear subspace $X := \{w \in Y : w|_{\Gamma_1} = 0\} \subset Y$, where $w|_{\Gamma_1}$ denotes the trace of $w \in Y$ on the (non-vanishing) Dirichlet boundary part $\Gamma_1 \subset \partial\Omega$. We equip X with the H^1-seminorm $\|u\|_X := \|\nabla u\|_{\mathrm{L}^2(\Omega)}$, for $u \in X$; owing to the Poincaré-Friedrichs inequality, we note that the norm $\|\cdot\|_X$ is equivalent to the norm $\|\cdot\|_Y$ on X, *i.e.*, there exists a constant $c > 0$ such that $c\|u\|_Y \leq \|u\|_X \leq \|u\|_Y$ for all $u \in X$.

Finally, we consider the closed and convex subset

$$K := \{w \in Y : w|_{\Gamma_1} = g \text{ on } \Gamma_1\} \subset Y, \tag{4.1}$$

with g the Dirichlet boundary data from (1.2b). Evidently, K has property (K). In particular, if $\Gamma_1 = \partial\Omega$ and $g \equiv 0$ on $\partial\Omega$, then we may consider $Y = X = K = \mathrm{H}^1_0(\Omega)$ with the norm $\|\cdot\|_X$.

4.2. Weak formulations

For any given $u \in K$, we define a (symmetric) bilinear form $a(u; \cdot, \cdot)$ on $Y \times Y$ by

$$a(u; v, w) := \int_\Omega \mu\Big(|\nabla u|^2\Big)\nabla v \cdot \nabla w \,\mathsf{dx}, \qquad v, w \in Y. \tag{4.2}$$

Moreover, we introduce the linear functional

$$\langle b, v\rangle := \int_\Omega f v \,\mathsf{dx} - \int_{\Gamma_2} h v \,\mathsf{dx}, \qquad v \in X, \tag{4.3}$$

where $\langle\cdot,\cdot\rangle$ denotes the duality pairing in $X^\star \times X$, with $X^\star$ signifying the dual space of X. If the source function $f \in \mathrm{L}^2(\Omega)$ and the Neumann boundary data $h \in \mathrm{L}^2(\partial\Omega)$, then we notice that $b \in X^\star$; incidentally, more general assumptions on the data can be made, see, *e.g.*, Remark 25.32 of [17].

In terms of the above forms, the weak formulation of (1.1) reads as follows:

$$\text{Find } u \in K: \qquad a(u; u, v) = \langle b, v\rangle \qquad \forall v \in X. \tag{4.4}$$

Furthermore, for given $u^n \in K$, $n \geq 0$, the weak form of the Kačanov scheme (1.2) is to find $u^{n+1} \in K$ such that

$$a(u^n; u^{n+1}, v) = \langle b, v\rangle \qquad \forall v \in X.$$

The ensuing result follows from standard arguments.

Proposition 4.1. *If the diffusion coefficient μ satisfies (μ3), then the form $a(\cdot;\cdot,\cdot)$ from (4.2) is bounded in the sense that*

$$|a(u; v, w)| \leq M_\mu \|v\|_Y \|w\|_Y \qquad \forall u \in K,\ \forall v, w \in Y.$$

Moreover, there exists a constant $\alpha > 0$ such that, for any $u \in K$, we have the coercivity property

$$a(u; v, v) \geq \alpha \|v\|_Y^2 \qquad \forall v \in X, \tag{4.5}$$

and, especially,

$$a(u; v - w, v - w) \geq \alpha \|v - w\|_Y^2 \qquad \forall v, w \in K. \tag{4.6}$$

4.3. Variational framework

We introduce the (nonlinear) functional $\mathsf{G}: K \to \mathbb{R}$ by

$$\mathsf{G}(u) := \int_\Omega \psi\big(|\nabla u|^2\big) \,\mathsf{dx}, \qquad \text{with} \quad \psi(s) := \frac{1}{2}\int_0^s \mu(t)\,\mathrm{d}t, \quad s \geq 0. \tag{4.7}$$

For $u \in K$, the Gâteaux derivative of G is given by

$$\langle \mathsf{G}'(u), v\rangle = \int_\Omega 2\psi'\big(|\nabla u|^2\big)\nabla u \cdot \nabla v \,\mathsf{dx} = \int_\Omega \mu\Big(|\nabla u|^2\Big)\nabla u \cdot \nabla v \,\mathsf{dx}, \tag{4.8}$$

for all $v \in X$, *i.e.*, $\mathsf{G}'(u) = a(u; u, \cdot)$ in $X^\star$.

Now, introduce the (energy) potential $\mathsf{H}: K \to \mathbb{R}$ by

$$\mathsf{H}(u) := \mathsf{G}(u) - \langle b, u\rangle, \tag{4.9}$$

with G and b from (4.7) and (4.3), respectively. If the diffusion coefficient μ satisfies the estimates

$$m_\mu(t - s) \leq \mu(t^2)t - \mu(s^2)s \leq M_\mu(t - s), \qquad t \geq s \geq 0, \tag{4.10}$$

then the Lipschitz condition (2.19) and the strong monotonicity property (2.1) can be deduced with $\nu = m_\mu$ and $L_{\mathsf{H}} = 3M_\mu$ (with respect to the norm $\|\cdot\|_X$), *cf.*, Proposition 25.26 of [17].

Remark 4.2. We comment on the assumption (4.10):

(a) It is easily shown that (μ1)–(μ4) implies (4.10), however, we emphasize that the latter assumption does not require μ to be decreasing nor differentiable. Yet, if μ is differentiable, then (4.10) implies (μ3) and (μ4).

(b) If the diffusion coefficient μ is continuously differentiable, then we can (easily) compute the bounds in (4.10) by taking into account the mean value theorem. In particular, we may set

$$m_\mu = \inf_{t\geq 0} \xi'(t) \qquad \text{and} \qquad M_\mu = \sup_{t\geq 0} \xi'(t),$$

where $\xi(t) = \mu(t^2)t$.

(c) Recall from (μ4) that the continuous function $\phi(t)$, *cf.*, (1.3), is strictly convex and increasing for $t \geq 0$ by (μ3) with $\phi(0) = 0$; we note that these properties relate to the class of Orlicz functions. In this aspect, our work links to the more general context of Orlicz type nonlinearities which have been studied, for instance, in [4].

The following result is a direct consequence of Corollary 2.3.

Proposition 4.3. *Suppose that the diffusion coefficient μ satisfies (4.10). Then, the functional H from (4.9) has a unique minimizer $u^\star \in K$, cf., (4.1), which satisfies the weak formulation*

$$\int_\Omega \mu\Big(|\nabla u|^2\Big)\nabla u \cdot \nabla v \,\mathsf{dx} = \int_\Omega f v \,\mathsf{dx} - \int_{\Gamma_2} h v \,\mathsf{dx} \qquad \forall v \in X; \tag{4.11}$$

i.e., $u^\star$ is the unique (weak) solution of (4.4).

4.4. Convergence of the modified Kačanov method

Recall that the properties (A1)–(A4) as well as (K) are satisfied if the diffusion coefficient obeys the bounds (4.10) by our analysis in the previous Sections 4.2 and 4.3, see, in particular, Proposition 4.1 and (4.8). Hence, the assumptions for the convergence results, *cf.*, Theorem 2.5 and Corollary 2.7, are fulfilled in the context of the quasilinear elliptic PDE (1.1), *without* assuming (μ2).

Theorem 4.4. *Assume that the diffusion coefficient satisfies the bounds (4.10). Then, the damped Kačanov method (2.14), with $\delta : K \to [\delta_{\min}, \delta_{\max}]$ and $0 < \delta_{\min} \leq \delta_{\max} < {}^{2\alpha}/_{3M_\mu}$, converges to the unique weak solution $u^\star \in K$ of (4.11).*

Remark 4.5. We emphasize that the assumptions on the damping function δ from Theorem 4.4 are sufficient for the key inequality (2.16) to hold, *cf.*, Corollary 2.7, however, they are not necessary. Indeed, as both step size methods from Section 3 guarantee this inequality, they yield the convergence in the setting of Theorem 4.4 without the restriction on δ.

4.5. Numerical experiments

We will now perform a number of numerical tests for the modified Kačanov method based on the different step size methods from Section 3 in the context of the quasilinear elliptic boundary value problem (1.1).

In all experiments, we let $\Omega := (-1,1)^2 \setminus ([0,1] \times [-1,0])$ be a standard L-shaped domain in $\mathbb{R}^2$. We focus on homogeneous Dirichlet boundary conditions, *i.e.*, $\Gamma_1 = \partial\Omega$ and $g \equiv 0$, and therefore we set $Y := X = K = \mathrm{H}^1_0(\Omega)$. Moreover, we consider the norm $\|\cdot\|_Y := \|\nabla(\cdot)\|_{\mathrm{L}^2(\Omega)}$, so that we obtain $\alpha = m_\mu$ in (4.5) and (4.6). The source function f in (1.1a), respectively the linear functional $b \equiv b(u)$ in the abstract analysis in Section 2, is chosen such that the analytical solution of (1.1a) and (1.1b), with $g \equiv 0$ on the Dirichlet boundary $\Gamma_1 = \partial\Omega$, is given by the smooth function $u^\star(x,y) = \sin(\pi x)\sin(\pi y)$. For the numerical approximation, we will use a conforming $\mathbb{P}_1$-finite element framework with a uniform mesh consisting of approximately $3 \cdot 10^6$ triangles. Throughout we set the correction factor in the adaptive step size algorithms to be $\sigma = 0.9$. We have implemented our algorithms in Matlab, and solved the linear equations by means of the backslash operator. Furthermore, the errors to be illustrated in the figures below are taken with respect to the underlying exact *discrete* solution, which, in each case, was determined with the aid of 1000 steps of the Zarantonello iteration with a suitable damping parameter, *cf.*, Proposition 5.1 of [12].

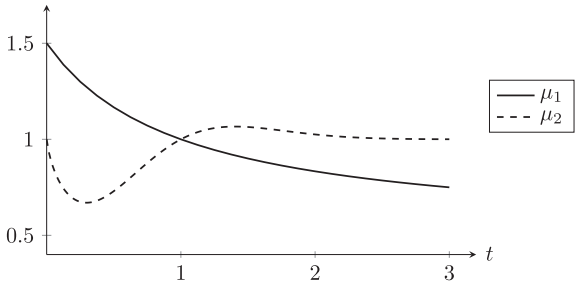

FIGURE 1. The diffusion coefficients $\mu(t) = \mu_1(t)$ and $\mu(t) = \mu_2(t)$ from Sections 4.5.1 and 4.5.2, respectively.

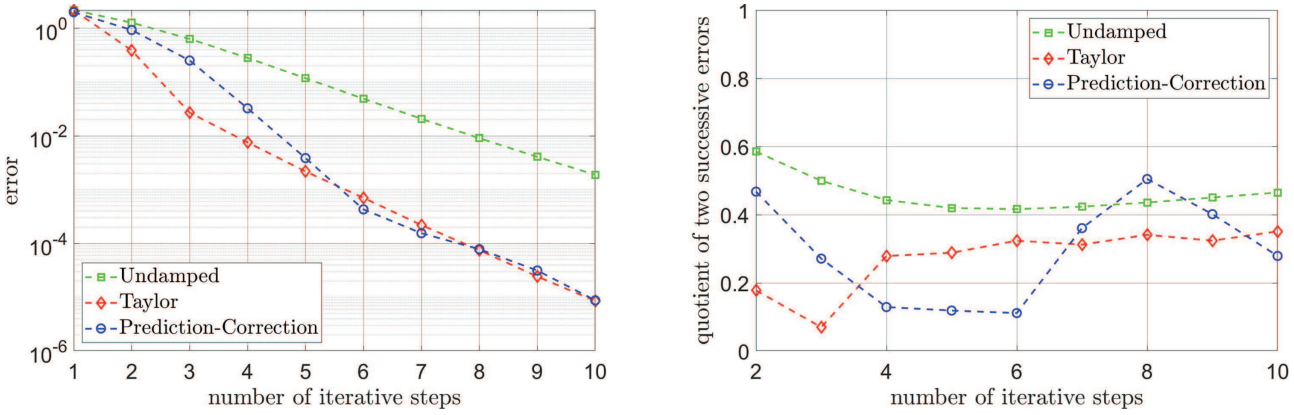

FIGURE 2. Experiment 4.5.1. Comparison of the performance of the classical Kačanov scheme ("Undamped") with the step size algorithms from Section 3.1 ("Taylor") and Section 3.2 ("Prediction-Correction"). *Left*: error decay. *Right*: ratio of two successive errors.

4.5.1. Monotonically decreasing diffusion

We consider the nonlinear diffusion coefficient $\mu(t) = \mu_1(t) = (t+1)^{-1} + 1/2$, for $t \geq 0$, see Figure 1. It is straightforward to verify that the diffusion parameter μ satisfies (4.10) as well as the properties (μ1)–(μ4). We compare the performance of the classical Kačanov scheme (2.11) with the damped Kačanov method (2.14) for both step size strategies from Section 3. For the application of the two step size methods, we recall that we need to know the values of the constants m_μ and M_μ *a priori*; in light of Remark 4.2 they are found to be $m_\mu = 3/8$ and $M_\mu = 3/2$. Moreover, here and in the two following experiments, we use the initial parameters $\delta = 1$ and $p = -1$ in Algorithm 2. Even though the diffusion parameter is monotonically decreasing and differentiable, which implies the convergence of the classical Kačanov scheme, we can see from Figure 2 that the damped Kačanov method with either the step size method from Section 3.1 or Section 3.2 performs (overall) better (in terms of error reduction per iteration step) than the undamped iteration. It is noteworthy that the damping parameters are larger than 1 in all steps for both approaches, see Figure 3.

4.5.2. Non-monotone diffusion

In our second experiment, we consider the nonlinear diffusion parameter $\mu(t) = \mu_2(t) = t\exp(-t^2)\log(t+\epsilon) + 1$, $t \geq 0$, for $\epsilon = 10^{-4}$, see Figure 1. It can be shown that μ satisfies (4.10) with $m_\mu \approx 0.483503$ and $M_\mu \approx 1.73565$, but is *not* monotonically decreasing (nor increasing). Even though Figure 4 indicates that the classical Kačanov scheme (2.11) may still converge, the convergence rate is really poor. In contrast, the modified

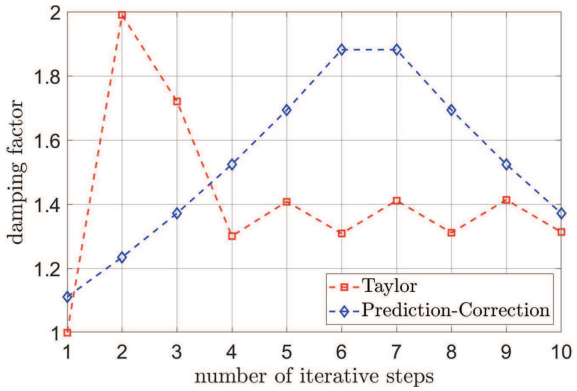

FIGURE 3. Experiment 4.5.1. Step sizes of each iterative step for the respective damped Kačanov scheme.

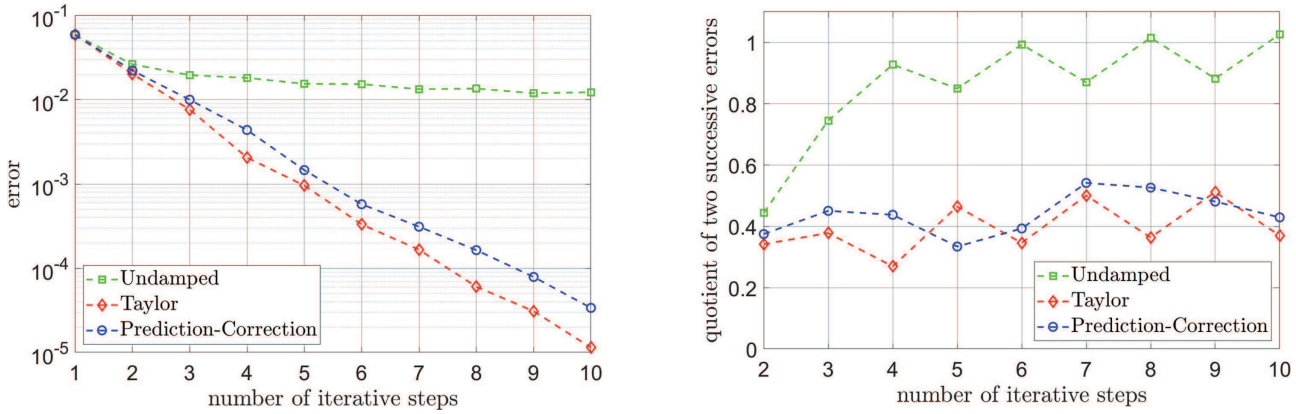

FIGURE 4. Experiment 4.5.2. Comparison of the performance of the classical Kačanov scheme ("Undamped") with the step size algorithms from Section 3.1 ("Taylor") and Section 3.2 ("Prediction-Correction"). *Left*: error decay. *Right*: ratio of two successive errors.

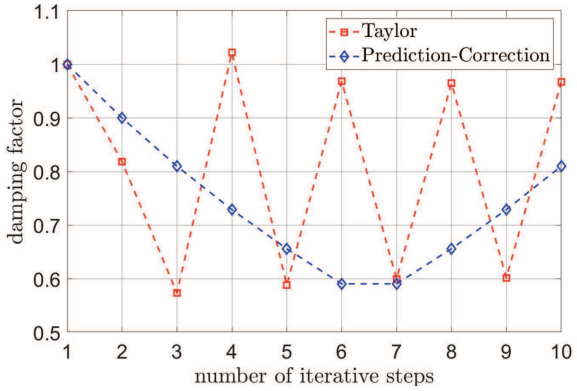

FIGURE 5. Experiment 4.5.2. Step sizes of each iterative step for the respective damped Kačanov scheme.

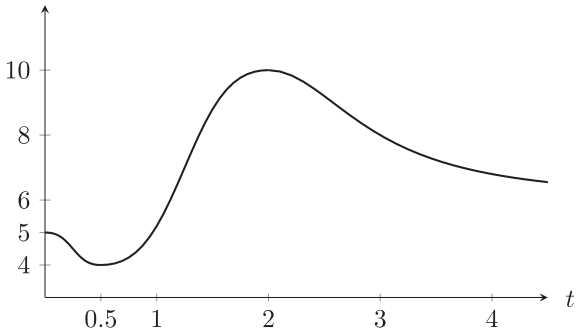

FIGURE 6. The diffusion coefficients μ_3 from Section 4.5.3.

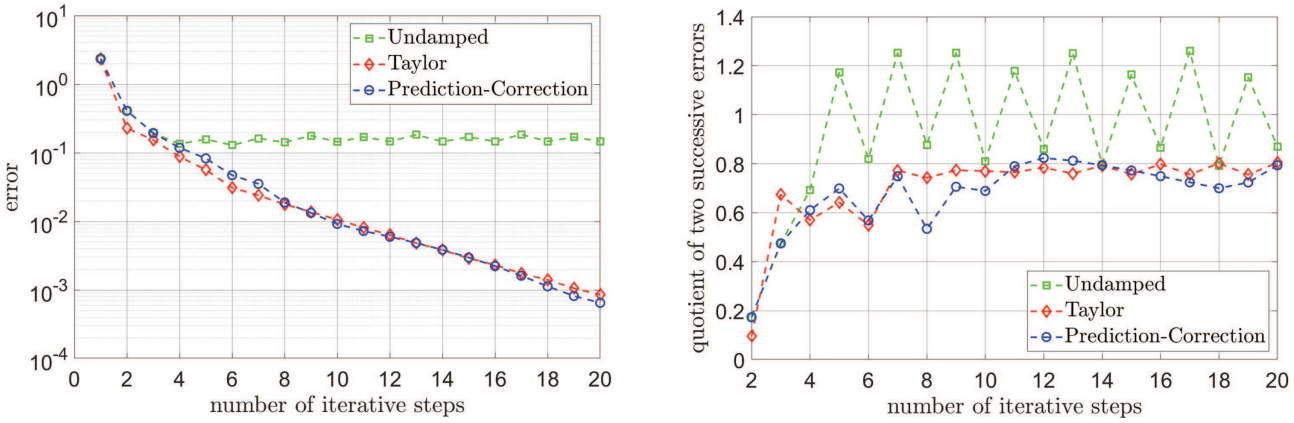

FIGURE 7. Experiment 4.5.3. Comparison of the performance of the classical Kačanov scheme ("Undamped") with the step size algorithms from Section 3.1 ("Taylor") and Section 3.2 ("Prediction-Correction"). *Left*: error decay. *Right*: ratio of two successive errors.

Kačanov scheme (2.14) with either damping strategy from Section 3 exhibits a considerably better performance. We can observe in Figure 5 that the damping parameters for both step size strategies from Section 3 are (mostly) smaller than 1 in this specific experiment.

4.5.3. Diffusion coefficient motivated by fluid viscosities with shear thinning and shear thickening zones

In our last experiment, we will consider a diffusion coefficient that is motivated by a model of a fluid viscosity with both shear thinning and shear thickening zones; we refer to the work [6] for the details about the rheological properties of the corresponding fluids. Specifically, let

$$\mu_3(t) = \begin{cases} \mu_c + \frac{\mu_0 - \mu_c}{1+\left(\frac{t^2}{t-t_c}\right)^2}, & 0 \le t \le t_c \\ \mu_{\max} + \frac{\mu_c - \mu_{\max}}{1+\left(\frac{t-t_c}{t-t_{\max}}\right)^2}, & t_c < t \le t_{\max} \\ \mu_\infty + \frac{\mu_{\max} - \mu_\infty}{1+(t-t_{\max})^2}, & t > t_{\max}, \end{cases}$$

whereby we set $t_c = 0.5$, $t_{\max} = 2$, $\mu_0 = 5$, $\mu_c = 4$, $\mu_{\max} = 10$, and $\mu_\infty = 6$; this diffusion coefficient is illustrated in Figure 6.

In [6] it is shown that the diffusion coefficient $\mu_3(t)$ is continuously differentiable. Once more, from Remark 4.2, it follows that (4.10) is satisfied with $m_\mu = 1.68$ and $M_\mu \approx 28.2696$. We clearly see in Figure 7 that the classical

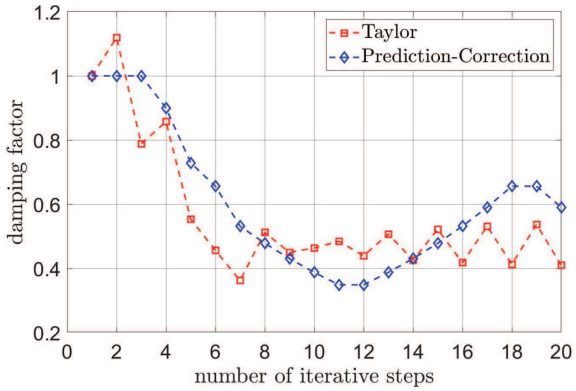

FIGURE 8. Experiment 4.5.3. Step sizes of each iterative step for the respective damped Kačanov scheme.

Kačanov scheme *does not converge* for this specific problem. In contrast, the modified Kačanov method with either of the two step size strategies from Section 3 converges perfectly, with the ratio of two successive errors being around 0.8. Here, the step sizes in the damped Kačanov scheme are, after an initial phase, between 0.3 and 0.7, and thus noticeably below 1, see Figure 8.

5. Conclusion

In this work, we have devised a modified version of the classical Kačanov iteration scheme. Exploiting the iterative linearization approach, *cf.*, [12], we have shown that the introduction of a damping parameter allows to derive a new convergence analysis, which applies to a wider class of problems. For instance, in the context of quasilinear elliptic PDE, a standard monotonicity condition on the diffusion coefficient can be dropped. Moreover, our numerical tests highlight that the modified Kačanov method, in combination with suitable damping strategies, outperforms the classical scheme for the examples under consideration. Especially, the final experiment in our work illustrates that the modified Kačanov scheme can effectively approximate nonlinear problems, for which the classical Kačanov method fails to generate a sequence converging to a solution. This underlines the relevance of our modified Kačanov scheme. We close by remarking that our work can be extended in a straightforward manner to quasilinear systems with applications to, *e.g.*, plasticity or quasi-Newtonian fluids.

Acknowledgements. The authors acknowledge the financial support of the Swiss National Science Foundation (SNF), Grant No. 200021_182524, and Project No. P2BEP2_191760.

References

[1] K. Astala, T. Iwaniec and G. Martin, Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane. *Princeton Mathematical Series*. Vol. 48. Princeton University Press, Princeton, NJ (2009).
[2] E. Carelli, J. Haehnle and A. Prohl, Convergence analysis for incompressible generalized newtonian fluid flows with nonstandard anisotropic growth conditions. *SIAM J. Numer. Anal.* **48** (2010) 164–190.
[3] P. Deuflhard, Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms. *Springer Series in Computational Mathematics*. Vol. 35. Springer-Verlag, Berlin (2004).
[4] L. Diening and M. Růžička, Interpolation operators in Orlicz–Sobolev spaces. *Numer. Math.* **107** (2007) 107–129.
[5] L. Diening, M. Fornasier, R. Tomasi and M. Wank, A relaxed Kačanov iteration for the p-Poisson problem. *Numer. Math.* **145** (2020) 1–34.
[6] F.J. Galindo-Rosales, F.J. Rubio-Hernández and A. Sevilla, An apparent viscosity function for shear thickening fluids. *J. Non-Newtonian Fluid Mech.* **166** (2011) 321–325.
[7] E.M. Garau, P. Morin and C. Zuppa, Convergence of an adaptive Kačanov FEM for quasi-linear problems. *Appl. Numer. Math.* **61** (2011) 512–529.

[8] A. Haberl, D. Praetorius, S. Schimanko and M. Vohralík, Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver. *Numer. Math.* **147** (2021) 679–725.
[9] W. Han, S. Jensen and I. Shimansky, The Kačanov method for some nonlinear problems. *Appl. Numer. Meth.* **24** (1997) 57–79.
[10] P. Heid and E. Süli, Adaptive iterative linearised finite element methods for implicitly constituted incompressible fluid flow problems and its application to Bingham fluids. Tech. Report arxiv:2109.05991 (2021).
[11] P. Heid and E. Süli, On the convergence rate of the Kačanov scheme for shear-thinning fluids. *Calcolo* **59** (2022) 27.
[12] P. Heid and T.P. Wihler, Adaptive iterative linearization Galerkin methods for nonlinear problems. *Math. Comp.* **89** (2020) 2707–2734.
[13] P. Heid and T.P. Wihler, On the convergence of adaptive iterative linearized Galerkin methods. *Calcolo* **57** (2020) 23.
[14] P. Heid, D. Praetorius and T.P. Wihler, Energy contraction and optimal convergence of adaptive iterative linearized finite element methods. *Comput. Methods Appl. Math.* **21** (2021) 407–422.
[15] L.M. Kačanov, Variational methods of solution of plasticity problems. *J. Appl. Math. Mech.* **23** (1959) 880–883.
[16] E. Zeidler, Nonlinear Functional Analysis and its Applications. IV. Applications to Mathematical Physics, Translated from the German and with a preface by Juergen Quandt. Springer-Verlag, New York (1988).
[17] E. Zeidler, Nonlinear Functional Analysis and its Applications. II/B. Springer-Verlag, New York (1990).