BIT Numer Math (2017) 57:787–810
DOI 10.1007/s10543-017-0653-1

BIT

Finite element convergence analysis for the thermoviscoelastic Joule heating problem

Axel Målqvist[1] · **Tony Stillfjord**[1]

Received: 2 November 2016 / Accepted: 2 March 2017 / Published online: 15 March 2017

Abstract We consider a system of equations that model the temperature, electric potential and deformation of a thermoviscoelastic body. A typical application is a thermistor; an electrical component that can be used e.g. as a surge protector, temperature sensor or for very precise positioning. We introduce a full discretization based on standard finite elements in space and a semi-implicit Euler-type method in time. For this method we prove optimal convergence orders, i.e. second-order in space and first-order in time. The theoretical results are verified by several numerical experiments in two and three dimensions.

Keywords Partial differential equations · Thermoviscoelastic · Joule heating · Thermistor · Convergence analysis · Finite elements

Mathematics Subject Classification 65M12 · 65M60 · 74D05 · 74H15

Communicated by Rolf Stenberg.

Both authors were supported by the Swedish Research Council under Grant 2015-04964.

✉ Tony Stillfjord
tony.stillfjord@gu.se

Axel Målqvist
axel@chalmers.se

[1] Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg, 412 96 Göteborg, Sweden

1 Introduction

Consider the following system of coupled equations:

$$\dot{\theta} = \Delta\theta + \sigma(\theta)|\nabla\phi|^2 - \mathbf{M} : \varepsilon(\dot{u}), \tag{1.1}$$

$$0 = \nabla \cdot \big(\sigma(\theta)\nabla\phi\big), \tag{1.2}$$

$$\ddot{u} = \nabla \cdot \big(\mathbf{A}\varepsilon(\dot{u}) + \mathbf{B}\varepsilon(u) - \mathbf{M}\theta\big) + f, \tag{1.3}$$

with initial conditions

$$\theta(0, x) = \theta_0(x), \quad u(0, x) = u_0(x) \quad \text{and} \quad \dot{u}(0, x) = v_0(x),$$

over the convex polygonal or polyhedral domain $\Omega \subset \mathbb{R}^d$ with $d \leq 3$. Together with appropriate boundary conditions, to be specified later, these equations describe the evolution of the temperature θ, electric potential ϕ and deformation u of a conducting body. Here $\mathbf{A}$, $\mathbf{B}$ are constant tensors of order 4, describing the viscosity and elasticity of the body, and $\mathbf{M}$ is a constant matrix describing the thermal expansion of the body. The vector f consists of external forces and $\sigma(\theta)$ denotes the electrical conductivity, which here depends on the temperature. In addition, we have used the notation

$$\varepsilon(u) = \frac{1}{2}\big(\nabla u + (\nabla u)^T\big)$$

for the linearized strain tensor and : for the Frobenius inner product.

The coupling of electricity and temperature through (1.1)–(1.2) is commonly known as *Joule heating* and is typically used to model thermistors, see e.g. [5,9]. These are electrical components used for example as surge protectors or temperature sensors. The inclusion of thermoviscoelastic effects through (1.3) allows us to also model their use as actuators on the micro-scale, cf. [16].

We note that the Joule heating problem, both stationary and time-dependent, has been considered extensively in different contexts. For discussions on existence and uniqueness, see e.g. [2,5,6,8,9,17–19,23,31] and the references therein. For the fully coupled, deformable problem the literature is less extensive. We refer mainly to [20] for the non-degenerate case that we consider here, with $\sigma \geq \sigma_{\min} > 0$. See also [30] for the degenerate case where $\sigma = 0$ is allowed; this requires a more generalized solution concept.

However, to our knowledge there exists no numerical analysis for methods applied to the fully coupled case. Many authors have analyzed methods for similar problems. For example, [12] considers the quasi-static version where the $\ddot{u}$-term is ignored, [1, 11,24] considers the non-deformable case, [13,14] treat the purely thermoviscoelastic case (no ϕ) with nonlinear constituent law, etc. Additionally, in the deformable case a common theme seems to be suboptimal convergence orders, i.e. errors of the form $\mathrm{O}(h + k)$ instead of $\mathrm{O}(h^2 + k)$.

The main contribution of this article is therefore an error analysis for a fully discrete discretization applied to the problem (1.1)–(1.3), which shows optimal convergence

orders in both time and space. For the spatial discretization we consider standard finite elements, and for the temporal discretization a semi-implicit Euler-type method. Our approach also allows us to analyze e.g. the implicit Euler method, but the semi-implicit method benefits from a greatly decreased computational cost while the errors are comparable.

The central idea of our proof is to bound the errors in ϕ and $\dot{u}$ in terms of the error in θ, in the spirit of [11,22]. The latter error then fulfills an equation similar to (1.1), to which we may apply a Grönwall inequality after properly handling the quadratic potential term. We note that we avoid any time step restrictions of the form $k \leq h^{d/r}$ by performing the analysis in two steps, where the first considers only the discretization in time, cf. [22]. Finally, in order to produce the $\dot{u}$ error bound, we extend the concept of Ritz–Volterra projections for damped wave equations (see [25]) to the discrete and vector-valued viscoelasticity case.

For simplicity, we consider Dirichlet boundary conditions,

$$\theta(t,x) = 0, \quad \phi(t,x) = \phi_b(t,x) \quad \text{and} \quad u(t,x) = 0$$

for $t \in [0,T]$ and $x \in \partial\Omega$. This is a simplified case of the ideal situation with an arbitrary polygon and mixed boundary conditions, corresponding to where the body is clamped and insulated. As is well known (see e.g. [15]) the solutions to such a problem would typically suffer from a lack of regularity in the vicinity of re-entrant corners and boundary condition transitions, which leads to suboptimal convergence orders for finite-element based numerical methods. We therefore restrict ourselves to the simplified model, and will indicate possible generalizations by our numerical experiments.

A brief outline of the article is as follows. In Sect. 2 we write the problem on weak form and discretize it in both time and space. The assumptions on the data and solutions to the continuous problem are given in Sect. 3, where we also perform the error analysis. In Sect. 3.1, the time-discrete system is shown to be first-order convergent, and then the full discretization is shown to be second-order convergent to the time-discrete system in Sect. 3.2. These results are confirmed by the numerical experiments presented in Sect. 4, and conclusions and future work is summarized in Sect. 5.

2 Weak formulation and discretization

In order to present a weak formulation of the problem, we introduce the spaces

$$V := H_0^1(\Omega) \subset L^2(\Omega), \quad \text{and} \quad \boldsymbol{V} := H_0^1(\Omega)^d \subset L^2(\Omega)^d =: \mathbf{L}^2(\Omega),$$

as well as the space of symmetric matrices,

$$Q = \left\{ \xi = (\xi_{ij})_{i,j=1}^d \subset L^2(\Omega)^{d\times d} \; ; \; \xi_{ji} = \xi_{ij}, 1 \leq i, j \leq d \right\}.$$

The idea here is that θ and $\phi - \phi_b$ belong to V, $u \in \boldsymbol{V}$ and $\varepsilon(u) \in Q$. On Q, we have the inner product

$$(\xi, \zeta)_Q := \int_\Omega \xi(x) : \zeta(x)\,\mathrm{d}x = \sum_{i,j=1}^{d} \left(\xi_{ij}, \zeta_{ij}\right)_{L^2(\Omega)}.$$

which gives rise to the norm $\|\cdot\|_Q$. To simplify some notation, we use the inner product

$$(u, v)_{\boldsymbol{V}} = (\varepsilon(u), \varepsilon(v))_Q$$

on $\boldsymbol{V}$ instead of the usual one. The norm $\|\cdot\|_{\boldsymbol{V}}$ induced by this inner product is equivalent to $\|\cdot\|_{H^1(\Omega)^d}$ by Korn's inequality, see e.g. [10, Chapter III, Theorems 3.1, 3.3] and [27]. We will on several occasions make use also of the norm $\|\cdot\|_{\mathbf{B}}$, which arises from the elasticity operator through

$$\|u\|_{\mathbf{B}}^2 = (\mathbf{B}\varepsilon(u), \varepsilon(u))_Q\,,$$

as well as the norm $\|\cdot\|_{\mathbf{A}+k\mathbf{B}}$ defined analogously for a small positive constant k. Under Assumption 3.1 in the next section, both of these norms are equivalent to the $\boldsymbol{V}$-norm. In the following, we will omit the specification of Ω and simply write L^2 or $\mathbf{L}^2$. Additionally, the L^2- and $\mathbf{L}^2$-norms will both simply be denoted by $\|\cdot\|$ and the corresponding inner products by $(\cdot, \cdot)$, where no confusion can arise.

By multiplying the Eqs. (1.1), (1.2) with the test function $\chi \in V$, Eq. (1.3) with $\boldsymbol{\chi} \in \boldsymbol{V}$ and then using Green's formula we get

$$\left(\dot{\theta}, \chi\right) + (\nabla\theta, \nabla\chi) = \left(\sigma(\theta)|\nabla\phi|^2, \chi\right) - (\mathbf{M} : \varepsilon(\dot{u}), \chi)\,, \tag{2.1}$$

$$(\sigma(\theta)\nabla\phi, \nabla\chi) = 0, \tag{2.2}$$

$$(\ddot{u}, \boldsymbol{\chi}) + (\mathbf{A}\varepsilon(\dot{u}) + \mathbf{B}\varepsilon(u), \varepsilon(\boldsymbol{\chi}))_Q = (\mathbf{M}\theta, \varepsilon(\chi))_Q + (f, \boldsymbol{\chi})\,, \tag{2.3}$$

for all $\chi \in V$ and $\boldsymbol{\chi} \in \boldsymbol{V}$, respectively. In (2.3), we have made use of the identity $(\varepsilon(u), \nabla v) = (\varepsilon(u), \varepsilon(v))$ as well as the similar identities $(\mathbf{A}\varepsilon(u), \nabla v) = (\mathbf{A}\varepsilon(u), \varepsilon(v))$ and $(\mathbf{B}\varepsilon(u), \nabla v) = (\mathbf{B}\varepsilon(u), \varepsilon(v))$. The latter two hold because we assume $\mathbf{A}$ and $\mathbf{B}$ to be symmetric; see Assumption 3.1 in the next section. Note also that we have omitted the time parameter here and in the original equation; both are supposed to hold for all times $t \in (0, T]$ for a given T.

We now discretize the time interval $[0, T]$ using a constant temporal step size k, which results in the grid $t_n = nk$ with $n = 1, 2, \ldots, N$ and $Nk = T$. We will abbreviate function evaluations at these times by sub-scripts, so that

$$\theta_n = \theta(t_n), \quad \phi_n = \phi(t_n), \quad u_n = u(t_n) \quad \text{and} \quad f_n = f(t_n).$$

The approximations of these solution values should belong to the same spaces as in the continuous case, and we will denote them by capital letters and superscripts:

$$\Theta^n \approx \theta_n, \quad \Phi^n \approx \phi_n \quad \text{and} \quad U^n \approx u_n.$$

Additionally, we denote by $\mathrm{D_t}$ the first-order backward difference quotient, i.e.

$$\mathrm{D_t}\,\Theta^n = \frac{\Theta^n - \Theta^{n-1}}{k}.$$

With this notation given, we now consider the following semi-implicit temporal discretization of Eqs. (1.1)–(1.3),

$$\mathrm{D_t}\,\Theta^n = \Delta\Theta^n + \sigma\left(\Theta^{n-1}\right)\left|\nabla\Phi^{n-1}\right|^2 - \mathbf{M} : \varepsilon\left(\mathrm{D_t}\,U^{n-1}\right), \tag{2.4}$$

$$0 = \nabla\cdot\left(\sigma(\Theta^n)\nabla\Phi^n\right), \tag{2.5}$$

$$\mathrm{D_t^2}\,U^n = \nabla\cdot\left(\mathbf{A}\varepsilon(\mathrm{D_t}\,U^n) + \mathbf{B}\varepsilon(U^n) - \mathbf{M}\Theta^n\right) + f_n, \tag{2.6}$$

where $\mathrm{D_t^2} = \mathrm{D_t}\,\mathrm{D_t}$, and its corresponding weak form,

$$\left(\mathrm{D_t}\,\Theta^n, \chi\right) + \left(\nabla\Theta^n, \nabla\chi\right) = \left(\sigma\left(\Theta^{n-1}\right)\left|\nabla\Phi^{n-1}\right|^2, \chi\right) - \left(\mathbf{M}{:}\varepsilon\left(\mathrm{D_t}\,U^{n-1}\right), \chi\right), \tag{2.7}$$

$$\left(\sigma(\Theta^n)\nabla\Phi^n, \nabla\chi\right) = 0, \tag{2.8}$$

$$\left(\mathrm{D_t^2}\,U^n, \boldsymbol{\chi}\right) + \left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\,U^n\right) + \mathbf{B}\varepsilon(U^n), \varepsilon(\boldsymbol{\chi})\right)_Q = \left(\mathbf{M}\Theta^n, \varepsilon(\chi)\right)_Q + \left(f_n, \boldsymbol{\chi}\right), \tag{2.9}$$

for $n = 1, \ldots, N$ and for all $\chi \in S_h$ and $\boldsymbol{\chi} \in \boldsymbol{S}_h$, respectively. The initial conditions are the same as in the continuous case: $\Theta^0 = \theta_0$, $U^0 = u_0$ and $\mathrm{D_t}\,U^0 = v_0$. (We use a fictitious point U^{-1} to define $\mathrm{D_t}\,U^0$.) Note that this discretization results in a decoupling of the equations; we solve first for Θ^n using (2.4) then use this to find Φ^n from (2.5) and U^n from (2.6). This implies a significant decrease in computational effort compared to the fully coupled case arising from e.g. the implicit Euler discretization.

For the spatial discretization, we introduce the finite element spaces $S_h \subset V$ and $\boldsymbol{S}_h \subset \boldsymbol{V}$. These consist of continuous, piecewise linear functions with zero trace on $\partial\Omega$, defined on a quasi-uniform mesh with mesh-width h. Then the fully discrete problem we are interested in is given by

$$\left(\mathrm{D_t}\,\Theta_h^n, \chi\right) + \left(\nabla\Theta_h^n, \nabla\chi\right) = \left(\sigma\left(\Theta_h^{n-1}\right)\left|\nabla\Phi_h^{n-1}\right|^2, \chi\right) - \left(\mathbf{M}{:}\varepsilon\left(\mathrm{D_t}\,U_h^{n-1}\right), \chi\right), \tag{2.10}$$

$$\left(\sigma\left(\Theta_h^n\right)\nabla\Phi_h^n, \nabla\chi\right) = 0, \tag{2.11}$$

$$\left(\mathrm{D_t^2}\,U_h^n, \boldsymbol{\chi}\right) + \left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\,U_h^n\right) + \mathbf{B}\varepsilon\left(U_h^n\right), \varepsilon(\boldsymbol{\chi})\right)_Q = \left(\mathbf{M}\Theta_h^n, \varepsilon(\chi)\right)_Q + \left(f_n, \boldsymbol{\chi}\right), \tag{2.12}$$

for $n = 1, \ldots, N$ and for all $\chi \in S_h$ and $\boldsymbol{\chi} \in \boldsymbol{S}_h$, respectively. Here, the approximations satisfy $\Theta_h^n \in S_h$, $\Phi_h^n - \phi_b(t_n) \in S_h$ and $U_h^n \in \boldsymbol{S}_h$. (We assume that $\phi_b(t_n)$ is

defined on all of Ω.) As initial conditions, we take $U_h^0 = 0$, $\mathrm{D_t}\, U_h^0 = 0$ and $\Theta_h^0 = I_h\theta_0$, the Lagrangian interpolant of the exact initial condition.

Remark 2.1 We assume the domain to be a convex polygon or polyhedron in order that the standard interpolation and regularity estimates for linear elliptic problems are satisfied, see [7, Section 3.2]. Similarly, the quasi-uniformity of the mesh guarantees that the standard inverse inequalities are satisfied. These are needed to handle the nonlinear potential term in (1.1), see [11,22].

3 Error analysis

Our main goal is to estimate the errors $\|\Theta_h^n - \theta_n\|$, $\|\Phi_h^n - \phi_n\|$ and $\|U_h^n - u_n\|$. In order to do this, we will generalize the analysis of [22] (cf. also [11]) for the case with no deformation. This consists of first showing that the time-discrete approximations are $\mathrm{O}(k)$-close to the solutions of the continuous system, and also proving that these approximations exhibit a certain regularity. The key part here is to express the error in the potential in terms of the error in the temperature, and then only working with the temperature equation. With the given regularity, the time-discrete and fully discrete approximations can then be compared and shown to be $\mathrm{O}(h^2)$-close. The main problem here is the nonlinear term $\sigma(\theta)|\nabla\phi|^2$, which is handled in a two-step fashion: first using that $\|\nabla(\Phi_h^n - \Phi^n)\| \leq C(h + \|\Theta_h^n - \Theta^n\|)$ to show that in fact $\|\nabla(\Phi_h^n - \Phi^n)\| \leq Ch$ and then using this to estimate $\nabla(\Phi_h^n - \Phi^n)$ in a stronger norm.

In our case, the temperature Eq. (1.1) contains the extra term $\mathbf{M} : \varepsilon(\dot{u})$, so our idea is to also bound the error in $\dot{u}$ by the error in the temperature. Then we show that the approximations U^n possess certain regularity, which may be used to also express the fully discrete deformation errors in terms of the fully discrete temperature errors. The key part in the latter step is to utilize the concept of Ritz–Volterra projections [25], which we here generalize to the vector-valued viscoelasticity case, as well as to discrete time.

Before we perform this extended analysis, we state the general assumptions on the given data. In these, as well as throughout the rest of the paper, C denotes a generic constant independent of k, h and n but possibly depending on T, that may differ from line to line.

Assumption 3.1 The viscosity and elasticity tensors $\mathbf{A} = (a_{ijkl})$ and $\mathbf{B} = (b_{ijkl})$ are symmetric, and both yield Lipschitz continuous and strongly coercive bilinear forms. That is,

$$a_{ijkl} = a_{jikl} = a_{klij}, \qquad b_{ijkl} = b_{jikl} = b_{klij},$$

and there are positive constants C_1, C_2 such that for all $u, v \in \boldsymbol{V}$ we have

$$\max\left((\mathbf{A}\varepsilon(u), \varepsilon(v))_Q , (\mathbf{B}\varepsilon(u), \varepsilon(v))_Q \right) \leq C_1 \|u\|_{\boldsymbol{V}} \|v\|_{\boldsymbol{V}} \quad \text{and}$$
$$\min\left((\mathbf{A}\varepsilon(u), \varepsilon(u))_Q , (\mathbf{B}\varepsilon(u), \varepsilon(u))_Q \right) \geq C_2 \|u\|_{\boldsymbol{V}}^2.$$

Assumption 3.2 The electrical conductivity σ belongs to $C^1(\mathbb{R})$ and there are positive constants $\sigma_{\min}$, $\sigma_{\max}$ and $\sigma'_{\max}$ such that for all $\theta \geq 0$ we have

$$0 < \sigma_{\min} \leq \sigma(\theta) \leq \sigma_{\max} \quad \text{and} \quad |\sigma'(\theta)| \leq \sigma'_{\max}.$$

Assumption 3.3 The function $f \in C(0, T; \mathbf{L}^2)$, $\theta_0 \in H^2 \cap H_0^1$ and $\phi_b \in L^\infty(0, T; L^2)$ is regular enough that

$$\|\phi_b\|_{L^\infty(0,T;\, W^{2,12/5})} + \|\dot{\phi}_b\|_{L^2(0,T;\, H^1)} + \|\nabla\phi_b\|_{L^\infty(0,T;\, L^\infty)} \leq C.$$

By [20], these assumptions guarantee the existence of a weak solution to the problem, i.e functions (θ, ϕ, u) satisfying (2.1)–(2.3) with the time derivatives interpreted in a weak sense. Thus for example $\theta \in L^2(0, T; V)$ and $\dot{\theta} \in L^2(0, T; V)'$. For optimal convergence orders more regularity is required, and explicit conditions on the data that guarantees such regularity is currently unknown. We therefore also make the following regularity assumption, where $\mathbf{H}^2 = H^2(\Omega)^d$:

Assumption 3.4 There exist solutions (θ, ϕ, u) to (2.1)–(2.3) over the time interval $[0, T]$ which are regular enough that

$$\begin{aligned}
\|\theta\|_{L^\infty(0,T;\, H^2)} + \|\dot{\theta}\|_{L^\infty(0,T;\, L^2)} + \|\dot{\theta}\|_{L^2(0,T;\, H^2)} + \|\ddot{\theta}\|_{L^1(0,T;\, L^2)} &\leq C,\\
\|\phi\|_{L^\infty(0,T;\, W^{2,12/5})} + \|\dot{\phi}\|_{L^2(0,T;\, H^1)} + \|\phi\|_{L^\infty(0,T;\, W^{1,\infty})} &\leq C,\\
\|\dot{u}\|_{L^\infty(0,T;\, \mathbf{H}^2)} + \|\ddot{u}\|_{L^\infty(0,T;\, \mathbf{H}^2)} + \|u^{(3)}\|_{L^1(0,T;\, \mathbf{L}^2)} &\leq C
\end{aligned}$$

The assumptions on θ and ϕ are essentially the same as in the non-deformable situation given in [22], while the assumptions on u and f are new. We note that for the non-deformable case, the existence of solutions with similar regularity properties was shown in [11] when $d \leq 2$, with weak requirements on the initial values. In the general elliptic/parabolic case, the absence of reentrant corners in the convex domain makes such regularity plausible, see e.g. [15, Chapters 3, 4] and [28, Chapter 19]. In the displacement equation the viscosity term acts as damping, and we expect regular solutions to be present also there, see e.g. [21]. We are not aware of any regularity results for the fully coupled system, but we note that our numerical experiments with smooth data suggest that Assumption 3.4 is satisfied in practice.

The following main theorem will be proved in the next two subsections:

Theorem 3.1 *Let Assumptions 3.1–3.4 be satisfied and let (θ, ϕ, u) and $(\Theta_h^n, \Phi_h^n, U_h^n)$ be solutions to the Eqs. (2.1)–(2.3) and (2.10)–(2.12), respectively. Then there are positive constants k_0 and h_0 such that if $k < k_0$ and $h < h_0$ we have for $n = 1, \ldots, N$ that*

$$\|\Theta_h^n - \theta_n\| + \|\Phi_h^n - \phi_n\| + \|\mathrm{D_t}\, U_h^n - \dot{u}_n\| \leq C(h^2 + k),$$

and

$$\|\Theta_h^n - \theta_n\|_{H^1} + \|\Phi_h^n - \phi_n\|_{H^1} + \|\mathrm{D_t}\, U_h^n - \dot{u}_n\|_V \leq C(h + k).$$

The constant C is independent of k, h and n, but may depend on the final time $T = Nk$ and the problem data.

To abbreviate expressions like the above in the following, we introduce

$$e_\theta^n = \Theta^n - \theta_n, \quad e_\phi^n = \Phi^n - \phi_n \quad \text{and} \quad e_u^n = U^n - u_n$$

as well as

$$e_{\theta,h}^n = \Theta_h^n - \Theta^n, \quad e_{\phi,h}^n = \Phi_h^n - \Phi^n \quad \text{and} \quad e_{u,h}^n = U_h^n - U^n.$$

3.1 The time-discrete case

We start by considering the semi-discrete case, and first provide a bound for $\mathrm{D_t}\, e_u^n$ in terms of e_θ^n.

Lemma 3.1 *Let Assumptions 3.1–3.4 be satisfied and let (θ, ϕ, u) and (Θ^n, Φ^n, U^n) be solutions to the Eqs. (2.1)–(2.3) and (2.7)–(2.9), respectively. Then we have*

$$\|\mathrm{D_t}\, e_u^n\|^2 + \|e_u^n\|_V^2 + k\sum_{j=1}^n \|\mathrm{D_t}\, e_u^j\|_V^2 \le Ck^2 + Ck\sum_{j=1}^n \|e_\theta^j\|^2,$$

for $n = 1, \dots, N$, with the constant C independent of k and n.

Proof By Eqs. (2.3) and (2.9), we see that the error e_u^n satisfies

$$\begin{aligned}\left(\mathrm{D_t^2}\, e_u^n, \boldsymbol{\chi}\right) + \left(\mathbf{A}\varepsilon(\mathrm{D_t}\, e_u^n) + \mathbf{B}\varepsilon(e_u^n), \varepsilon(\boldsymbol{\chi})\right) &= \left(\mathbf{M}e_\theta^n, \varepsilon(\boldsymbol{\chi})\right) + \left(\ddot{u}(t_n) - \mathrm{D_t^2}\, u(t_n), \boldsymbol{\chi}\right) \\ &\quad + \left(\mathbf{A}\varepsilon(\dot{u}(t_n) - \mathrm{D_t}\, u(t_n)), \varepsilon(\boldsymbol{\chi})\right) \\ &\le C\|e_\theta^n\|\|\boldsymbol{\chi}\|_V + Ck\|\boldsymbol{\chi}\| + Ck\|\boldsymbol{\chi}\|_V\end{aligned}$$

due to the regularity assumptions on u. We note that for any sequence $\{g^n\}$ we have

$$2\left(\mathrm{D_t}^2\, g^n, \mathrm{D_t}\, g^n\right) \ge \mathrm{D_t}\|\mathrm{D_t}\, g^n\|^2 \quad \text{and} \quad 2\left(\mathbf{B}\varepsilon(g^n), \varepsilon\left(\mathrm{D_t}\, g^n\right)\right) \ge \mathrm{D_t}\|g^n\|_{\mathbf{B}}^2,$$

where $\|\cdot\|_{\mathbf{B}}$ is the norm induced by the inner product $\left(\mathbf{B}\varepsilon(\cdot), \varepsilon(\cdot)\right)$. Thus by choosing $\boldsymbol{\chi} = \mathrm{D_t}\, e_u^n$ and using the Cauchy–Schwarz inequality as well as Young's inequality, $ab \le \frac{1}{2c}a^2 + \frac{c}{2}b^2$, we get

$$\mathrm{D_t}\|\mathrm{D_t}\, e_u^n\|^2 + 2C_2\|\mathrm{D_t}\, e_u^n\|_V + \mathrm{D_t}\|e_u^n\|_{\mathbf{B}}^2 \le Ck^2 + C\|e_\theta^n\|^2 + C_2\|\mathrm{D_t}\, e_u^n\|_V^2.$$

Canceling the final term, summing over n and modifying the constants then yields

$$\|\mathrm{D_t}\, e_u^n\|^2 + k\sum_{j=1}^{n}\|\mathrm{D_t}\, e_u^j\|_{\boldsymbol{V}} + \|e_u^n\|_{\mathbf{B}}^2 \le Ck^2 + Ck\sum_{j=1}^{n}\|e_\theta^j\|^2,$$

and the Lemma follows from the equivalence between the $\mathbf{B}$- and $\boldsymbol{V}$-norms. □

Theorem 3.2 *Let Assumptions 3.1–3.4 be satisfied and let* (θ, ϕ, u) *and* (Θ^n, Φ^n, U^n) *be solutions to the Eqs. (1.1)–(1.3) and (2.4)–(2.6), respectively. Then there is a positive constant* k_0 *such that if* $k < k_0$ *then*

$$\|e_\theta^n\|_{H^1}^2 + \|e_\phi^n\|_{H^1}^2 + \|\mathrm{D_t}\, e_u^n\|_{\boldsymbol{V}}^2 \le Ck^2,$$

for $n = 1, \ldots, N$*, with the constant* C *independent of* k *and* n*. In addition, the approximations have the following regularity:*

$$\begin{aligned}
\|\Theta^n\|_{H^2}^2 + \|\mathrm{D_t}\,\Theta^n\|^2 + k\sum_{j=1}^{n}\|\mathrm{D_t}\,\Theta^j\|_{H^2}^2 &\le C,\\
\|\Phi^n\|_{W^{2,12/5}} + \|\Phi^n\|_{W^{1,\infty}} &\le C,\\
\|\mathrm{D_t}\, U^n\|_{\mathbf{H}^2}^2 + \|\mathrm{D_t^2}\, U^n\|_{\boldsymbol{V}}^2 + k\sum_{j=1}^{n}\|\mathrm{D_t^2}\, U^j\|_{\mathbf{H}^2}^2 &\le C.
\end{aligned}$$

Proof To begin with, we see that the error e_ϕ^n satisfies

$$-\nabla\cdot\big(\sigma(\Theta^n)\nabla e_\phi^n\big) = \nabla\cdot\big((\sigma(\Theta^n) - \sigma(\theta_n))\nabla\phi_n\big).$$

Multiplying this equation by e_ϕ^n and integrating directly yields

$$\|\nabla e_\phi^n\|^2 \le C\|\nabla\phi_n\|_{L^\infty}\|e_\theta^n\|\|\nabla e_\phi^n\|,$$

so that

$$\|\nabla e_\phi^n\| \le C\|e_\theta^n\| \tag{3.1}$$

by the regularity assumptions. This inequality for e_ϕ^n corresponds to Lemma 3.1 for e_u^n. Further, we see that the error e_θ^n satisfies

$$\begin{aligned}
\mathrm{D_t}\, e_\theta^n - \Delta e_\theta^n = {}& \Big(\sigma(\Theta^{n-1}) - \sigma(\theta_{n-1})\Big)|\nabla\phi_{n-1}|^2 + \sigma(\Theta^{n-1})\Big(\nabla\Phi^{n-1} + \nabla\phi_{n-1}\Big)\cdot\nabla e_\phi^{n-1}\\
& - M : \varepsilon\left(\mathrm{D_t}\, e_u^{n-1}\right) + R_\theta^n,
\end{aligned} \tag{3.2}$$

where

$$\begin{aligned}
R_\theta^n = {}& \big(\sigma(\theta_{n-1}) - \sigma(\theta_n)\big)|\nabla\phi_{n-1}|^2 + \sigma(\theta_n)\big(\nabla\phi_{n-1} + \nabla\phi_n\big)\cdot\big(\nabla\phi_{n-1} - \nabla\phi_n\big)\\
& + M : \varepsilon(\dot{u}_n - \dot{u}_{n-1}) + M : \varepsilon(\dot{u}_{n-1} - \mathrm{D_t}\, u_{n-1}).
\end{aligned}$$

is bounded by $\|R_\theta^n\| \leq Ck$, again by the regularity assumptions. After multiplying by e_θ^n and integrating, we therefore get

$$\begin{aligned} \mathrm{D_t}\|e_\theta^n\|^2 + 2\|\nabla e_\theta^n\|^2 \leq C\|e_\theta^{n-1}\|\|e_\theta^n\|\|\nabla\phi_{n-1}\|_{L^\infty} + \left(M : \varepsilon\left(\mathrm{D_t}\, e_u^{n-1}\right), e_\theta^n\right) \\ + Ck\|e_\theta^n\| + \left(\sigma(\Theta^{n-1})\left(\nabla\Phi^{n-1} + \nabla\phi_{n-1}\right)e_\theta^n, \nabla e_\phi^{n-1}\right). \end{aligned} \tag{3.3}$$

The last term of this expression can be shown to be bounded by $C(\|e_\theta^n\|^2 + \|e_\phi\|_{H^1}^2)$, see [22, p. 627], and for the second term we observe that for a generic $u \in \boldsymbol{V}$,

$$(\mathbf{M} : (\nabla u), \chi)_{L^2} = (\nabla u, \mathbf{M}\chi)_Q = -(u, \nabla\cdot(\mathbf{M}\chi))_{\mathbf{L}^2} = -(u, \mathbf{M}\nabla\chi)_{\mathbf{L}^2}.$$

As a completely analogous calculation holds also for $(\nabla u)^T$ and $\mathbf{M}$ is symmetric, we thus have

$$(\mathbf{M} : \varepsilon(u), \chi) = -(u, \mathbf{M}\nabla\chi) \leq C\|u\|\|\nabla\chi\|. \tag{3.4}$$

This implies that (3.3) reduces to

$$\mathrm{D_t}\|e_\theta^n\|^2 + 2\|\nabla e_\theta^n\|^2 \leq C\left(k^2 + \|e_\theta^{n-1}\|^2 + \|e_\theta^n\|^2 + \|e_\phi^{n-1}\|_{H^1}^2 + \|\mathrm{D_t}\, e_u^{n-1}\|^2\right) + \|\nabla e_\theta^n\|^2.$$

Canceling the last term, summing up and using Eq. (3.1) and Lemma 3.1 thus yields

$$\|e_\theta^n\|^2 + k\sum_{j=1}^n \|\nabla e_\theta^j\|^2 \leq Ck^2 + Ck\sum_{j=1}^n \|e_\theta^j\|^2.$$

Under the step size restriction $Ck < 1$, we can eliminate the last term of the sum. An application of Grönwall's lemma then shows that the left-hand side is bounded by Ck^2. Using Eq. (3.1) and Lemma 3.1 again, we see that in fact

$$\|e_\theta^n\|^2 + k\sum_{j=1}^n \|\nabla e_\theta^j\|^2 + \|\nabla e_\phi^n\|^2 + \|\mathrm{D_t}\, e_u^n\|^2 + \|e_u^n\|_{\boldsymbol{V}}^2 + k\sum_{j=1}^n \|\mathrm{D_t}\, e_u^j\|_{\boldsymbol{V}}^2 \leq Ck^2$$

From these preliminary bounds, we may deduce the desired regularity of Θ^n and Φ^n and then test (3.2) with $-\Delta e_\theta^n$ to acquire

$$\|e_\theta^n\|_{H^1}^2 + k\sum_{j=1}^n \|\Delta e_\theta^j\|^2 \leq Ck^2.$$

For details, we refer to [22, Theorem 3.1]. Let us instead investigate the remaining questions of the regularity of U^n and the pointwise bound for $\mathrm{D_t}\, e_u^n$ in the $\boldsymbol{V}$-norm. By the defining equation, we have that

$$\begin{aligned} \nabla\cdot\left(\mathbf{A}\varepsilon(\mathrm{D_t}\, e_u^n) + \mathbf{B}\varepsilon(e_u^n)\right) = \mathrm{D_t}^2\, e_u^n + \nabla\cdot\left(\mathbf{M}\Theta^n\right) + \mathrm{D_t}^2\, u(t_n) - \ddot{u}(t_n) \\ + \nabla\cdot\left(\mathbf{A}\varepsilon(\mathrm{D_t}\, u(t_n) - \dot{u}(t_n))\right), \end{aligned} \tag{3.5}$$

where the right-hand side is in $\mathbf{L}^2$ since $\|\mathrm{D}_{\mathrm{t}}^2 e_u^n\| \le k^{-1}(\|\mathrm{D}_{\mathrm{t}} e_u^n\| + \|\mathrm{D}_{\mathrm{t}} e_u^{n-1}\|) \le C$. Let us denote it by g_n. Then we can rewrite the previous equation as

$$\nabla \cdot \left(\mathbf{A}\varepsilon(\mathrm{D}_{\mathrm{t}} e_u^n) + k\mathbf{B}\varepsilon(\mathrm{D}_{\mathrm{t}} e_u^n)\right) = g_n + \nabla \cdot \left(\mathbf{B}\varepsilon(e_u^{n-1})\right).$$

Now since both $\mathbf{B}$ and $\mathbf{A} + k\mathbf{B}$ induce bounded and coercive inner products on $\boldsymbol{V}$, we see that

$$\begin{aligned}\|\mathrm{D}_{\mathrm{t}} e_u^n\|_{\mathbf{H}^2}^2 &\le C\|\nabla \cdot \left(\mathbf{A}\varepsilon(\mathrm{D}_{\mathrm{t}} e_u^n) + k\mathbf{B}\varepsilon(\mathrm{D}_{\mathrm{t}} e_u^n)\right)\|^2 \\ &\le C\|g_n\|^2 + C\|e_u^{n-1}\|_{\mathbf{H}^2}^2\end{aligned}$$

But since $e_u^{n-1} = k\sum_{j=1}^{n-1} \mathrm{D}_{\mathrm{t}} e_u^j$, we can estimate the second term by Cauchy–Schwarz as

$$\|e_u^{n-1}\|_{\mathbf{H}^2}^2 \le k\sum_{j=1}^{n-1} \|\mathrm{D}_{\mathrm{t}} e_u^j\|_{\mathbf{H}^2}^2.$$

An application of Grönwall's lemma thus shows that

$$\|\mathrm{D}_{\mathrm{t}} e_u^n\|_{\mathbf{H}^2} \le C,$$

which also implies that e_u^n, U^n and $\mathrm{D}_{\mathrm{t}} U^n$ are all in $\mathbf{H}^2$. We may now multiply (3.5) by $\nabla \cdot \left((\mathbf{A} + k\mathbf{B})\varepsilon(\mathrm{D}_{\mathrm{t}} e_u^n)\right)$ and integrate to get

$$\begin{aligned}&\left(\mathrm{D}_{\mathrm{t}}\,\varepsilon\left(\mathrm{D}_{\mathrm{t}} e_u^n\right), (\mathbf{A} + k\mathbf{B})\varepsilon\left(\mathrm{D}_{\mathrm{t}} e_u^n\right)\right) \\ &\quad + \|\nabla \cdot \left((\mathbf{A} + k\mathbf{B})\varepsilon(\mathrm{D}_{\mathrm{t}} e_u^n)\right)\|^2 \le C\|e_\theta^n\|_{\mathbf{H}^1}^2 + C\|e_\theta^{n-1}\|_{\mathbf{H}^2}^2,\end{aligned}$$

where we have used the Cauchy–Schwarz and Young inequalities and canceled a term $\frac{1}{2}\|\nabla \cdot \left((\mathbf{A}+k\mathbf{B})\varepsilon(\mathrm{D}_{\mathrm{t}} e_u^n)\right)\|^2$. The first term on the left-hand side can be estimated from below by $\mathrm{D}_{\mathrm{t}}\|\mathrm{D}_{\mathrm{t}} e_u^n\|_{\mathbf{A}+k\mathbf{B}}$, so summing up and using the equivalence of the $(\mathbf{A} + k\mathbf{B})$- and $\boldsymbol{V}$-norms, we get

$$\|\mathrm{D}_{\mathrm{t}} e_u^n\|_{\boldsymbol{V}}^2 + k\sum_{j=1}^{n} \|\mathrm{D}_{\mathrm{t}} e_u^j\|_{\mathbf{H}^2}^2 \le Ck\sum_{j=1}^{n-1} \|e_\theta^j\|_{H^1}^2 + Ck\sum_{j=1}^{n-1} \|\mathrm{D}_{\mathrm{t}} e_u^j\|_{\mathbf{H}^2}^2.$$

But the first term in the right-hand side is bounded by Ck^2 and in the second term we may again use that $\|\mathrm{D}_{\mathrm{t}} e_u^j\|_{\mathbf{H}^2}^2 \le k\sum_{i=1}^{j} \|\mathrm{D}_{\mathrm{t}} e_u^i\|_{\mathbf{H}^2}^2$. Defining

$$w_n = \|\mathrm{D}_{\mathrm{t}} e_u^n\|_{\boldsymbol{V}}^2 + k\sum_{j=1}^{n} \|\mathrm{D}_{\mathrm{t}} e_u^j\|_{\mathbf{H}^2}^2,$$

we thus have

$$w_n \le Ck^2 + Ck\sum_{j=1}^{n-1} w_j,$$

and an application of Grönwall's lemma shows that $w_n \leq Ck^2$. This yields the final desired error bound, and additionally shows that $\|\mathrm{D_t}^2 e_u^n\|_{\boldsymbol{V}}^2 + k\sum_{j=1}^n \|\mathrm{D_t}^2 e_u^j\|_{\mathbf{H}^2}^2 \leq C$, which implies the stated regularity for U^n. □

3.2 The fully discrete case

We now turn to the fully discretized case and first prove an analogue to Lemma 3.1.

Lemma 3.2 *Let Assumptions 3.1–3.4 be satisfied and* (Θ^n, Φ^n, U^n) *and* $(\Theta_h^n, \Phi_h^n, U_h^n)$ *be solutions to Eqs. (2.7)–(2.9) and (2.10)–(2.12), respectively. Then there is a positive constant* k_0 *such that if* $k < k_0$ *we have for* $n = 1, \ldots, N$ *that*

$$\|e_{u,h}^n\|^2 + \|\mathrm{D_t}\, e_{u,h}^n\|^2 \leq Ch^4 + Ck\sum_{j=1}^n \|e_{\theta,h}^j\|^2 \quad \textit{and}$$

$$\|e_{u,h}^n\|_{\boldsymbol{V}}^2 + k\sum_{j=1}^n \|\mathrm{D_t}\, e_{u,h}^j\|_{\boldsymbol{V}}^2 \leq Ch^2 + Ck\sum_{j=1}^n \|e_{\theta,h}^j\|^2,$$

with the constant C *independent of* k, h *and* n.

Remark 3.1 In the case of a first-order equation, one would typically first add and subtract the Ritz projection of e_u^n in order to work only in the finite element space. This approach is viable also in the second-order case, if one defines the Ritz projection using the $(\mathbf{A}\varepsilon(\cdot), \varepsilon(\cdot))$ inner product. We refer to [29] for the scalar-valued case. However, we choose to instead work with a Ritz–Volterra projection, see [25] for the scalar-valued case. Such a projection takes both the $\mathbf{A}$- and $\mathbf{B}$-terms into account simultaneously, i.e. it is a projection of $C^1(0, T; \boldsymbol{V})$-functions rather than of elements in $\boldsymbol{V}$. In the present situation, we need of course to consider a discretized version, but it nevertheless simplifies matters.

Proof Subtracting (2.9) from (2.12), we see that

$$\left(\mathrm{D_t}^2 e_{u,h}^n, \boldsymbol{\chi}\right) + \left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\, e_{u,h}^n\right) + \mathbf{B}\varepsilon\left(e_{u,h}^n\right), \varepsilon(\boldsymbol{\chi})\right) = \left(\mathbf{M}e_{\theta,h}^n, \varepsilon(\boldsymbol{\chi})\right)$$

for all $\boldsymbol{\chi} \in \boldsymbol{S}_h$. Now let $e_{u,h}^n = \eta^n + \rho^n$, where

$$\eta^n = U_h^n - W^n \in \boldsymbol{S}_h \quad \text{and} \quad \rho^n = W^n - U^n,$$

with the discrete Ritz–Volterra projection W^n of U^n satisfying $W^0 = U^0 = 0$ and

$$\left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\, W^n - \mathrm{D_t}\, U^n\right) + \mathbf{B}\varepsilon\left(W^n - U^n\right), \varepsilon(\boldsymbol{\chi})\right) = 0 \tag{3.6}$$

for all $\boldsymbol{\chi} \in \boldsymbol{S}_h$. We note that Eq. (3.6) may also be stated as

$$\left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\, \rho^n\right) + \mathbf{B}\varepsilon(\rho^n), \varepsilon(\boldsymbol{\chi})\right) = 0,$$

and that since $\mathrm{D_t}\, U^0 = 0$, also $\mathrm{D_t}\, W^0 = 0$. Additionally, we need the Ritz projection $\mathbf{R}_h$ given by the viscosity term. For a generic $u \in \boldsymbol{V}$, this is defined by

$$(\mathbf{A}\varepsilon(\mathbf{R}_h u - u), \varepsilon(\boldsymbol{\chi})) = 0$$

for all $\boldsymbol{\chi} \in \boldsymbol{S}_h$, and we have the inequality

$$\|\mathbf{R}_h u - u\| + h\|\mathbf{R}_h u - u\|_{\boldsymbol{V}} \leq Ch^2 \|u\|_{\mathbf{H}^2}.$$

We start by estimating the $\boldsymbol{V}$-norms of $\mathrm{D_t}\, \rho^n$ and ρ^n. To this end, we observe that for a generic u, we have

$$\|u\|_{\boldsymbol{V}}^2 = \|\varepsilon(u)\|_Q^2 \leq \|\nabla u\|_Q^2 = \sum_{j=1}^{d} \left\| \frac{\partial u}{\partial x_j} \right\|^2$$

and that

$$\left\| \frac{\partial u}{\partial x_j} \right\| = \sup_{\varphi \in C_0^\infty(\Omega)^d, \|\varphi\|=1} \left(\frac{\partial u}{\partial x_j}, \varphi \right).$$

We therefore take $\varphi \in C_0^\infty(\Omega)^d$ with $\|\varphi\| = 1$ and let $\Psi \in \boldsymbol{V}$ be the solution to

$$(\mathbf{A}\varepsilon(\Psi), \varepsilon(\boldsymbol{\chi}))_Q = -\left(\frac{\partial \varphi}{\partial x_j}, \boldsymbol{\chi} \right).$$

Then

$$\begin{aligned}
\left(\frac{\partial\, \mathrm{D_t}\, \rho^n}{\partial x_j}, \varphi \right) &= -\left(\mathrm{D_t}\, \rho^n, \frac{\partial \varphi}{\partial x_j} \right) = \left(\mathbf{A}\varepsilon(\Psi), \varepsilon\left(\mathrm{D_t}\, \rho^n \right) \right) = \left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\, \rho^n \right), \varepsilon(\Psi) \right) \\
&= \left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\, \rho^n \right), \varepsilon\left(\Psi - \mathbf{R}_h \Psi \right) \right) + \left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\, \rho^n \right), \varepsilon\left(\mathbf{R}_h \Psi \right) \right) \\
&= \left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\, \rho^n \right), \varepsilon\left(\Psi - \mathbf{R}_h \Psi \right) \right) - \left(\mathbf{B}\varepsilon(\rho^n), \varepsilon(\mathbf{R}_h \Psi) \right) =: R_1 + R_2,
\end{aligned}$$

where the last term is bounded by

$$R_2 \leq C\|\rho^n\|_{\boldsymbol{V}} \|\mathbf{R}_h \Psi\|_{\boldsymbol{V}} \leq C\|\rho^n\|_{\boldsymbol{V}} (\|\mathbf{R}_h \Psi - \Psi\|_{\boldsymbol{V}} + \|\Psi\|_{\boldsymbol{V}}) \leq C\|\rho^n\|_{\boldsymbol{V}}.$$

Moreover, since $\mathrm{D_t}\, W^n \in \boldsymbol{S}_h$, the first term is bounded by

$$\begin{aligned}
R_1 = -\left(\mathbf{A}\varepsilon(\mathrm{D_t}\, U^n), \varepsilon(\Psi - \mathbf{R}_h \Psi) \right) &= \left(\mathbf{A}\varepsilon(\mathbf{R}_h\, \mathrm{D_t}\, U^n - \mathrm{D_t}\, U^n), \varepsilon(\Psi - \mathbf{R}_h \Psi) \right) \\
&= \left(\mathbf{A}\varepsilon(\mathbf{R}_h\, \mathrm{D_t}\, U^n - \mathrm{D_t}\, U^n), \varepsilon(\Psi) \right) \\
&\leq C\|\mathbf{R}_h\, \mathrm{D_t}\, U^n - \mathrm{D_t}\, U^n\|_{\boldsymbol{V}} \|\Psi\|_{\boldsymbol{V}} \\
&\leq Ch\|\mathrm{D_t}\, U^n\|_{\mathbf{H}^2}.
\end{aligned}$$

By expressing ρ^n in terms of $\mathrm{D_t}\,\rho^j$ and noting that $\rho^0 = 0$, we thus have

$$\|\mathrm{D_t}\,\rho^n\|_{\boldsymbol{V}} \leq Ch\|\mathrm{D_t}\,U^n\|_{\mathbf{H}^2} + Ck\sum_{j=1}^{n}\|\mathrm{D_t}\,\rho^j\|_{\boldsymbol{V}},$$

and under the step size restriction $Ck < 1$ we can eliminate the last term of the sum and apply Grönwall's lemma. This shows that

$$\|\mathrm{D_t}\,\rho^n\|_{\boldsymbol{V}} \leq Ch\Big(\|\mathrm{D_t}\,U^n\|_{\mathbf{H}^2} + Ck\sum_{j=1}^{n-1}\|\mathrm{D_t}\,U^j\|_{\mathbf{H}^2}\Big).$$

By using the regularity shown in Theorem 3.2 and then summing over n, we see that

$$\|\rho^n\|_{\boldsymbol{V}} + \|\mathrm{D_t}\,\rho^n\|_{\boldsymbol{V}} \leq Ch.$$

Using these bounds we may now estimate ρ also in the $\mathbf{L}^2$-norm, by instead letting $\Psi \in \boldsymbol{V}$ be the solution to

$$(\mathbf{A}\varepsilon(\Psi), \varepsilon(\boldsymbol{\chi}))_Q = -(\varphi, \boldsymbol{\chi}).$$

Then as before,

$$\left(\mathrm{D_t}\,\rho^n, \varphi\right) = \left(\mathbf{A}\varepsilon\left(\mathbf{R}_h\,\mathrm{D_t}\,U^n - \mathrm{D_t}\,U^n\right), \varepsilon(\Psi)\right) + \left(\mathbf{B}\varepsilon\left(\rho^n\right), \varepsilon\left(\mathbf{R}_h\Psi\right)\right) =: R_3 + R_4,$$

where

$$R_3 \leq C\|\mathbf{R}_h\,\mathrm{D_t}\,U^n - \mathrm{D_t}\,U^n\|_{\boldsymbol{V}}\|\Psi\|_{\boldsymbol{V}} \leq Ch^2\|\mathrm{D_t}\,U^n\|_{\mathbf{H}^2}.$$

For R_4, we note that $\|\Psi\|_{\mathbf{H}^2} \leq C\|\varphi\| \leq C$, so that by using integration by parts and observing that both ρ^n and Ψ are zero on $\partial\Omega$ we get,

$$\begin{aligned} R_4 &\leq \left(\mathbf{B}\varepsilon(\rho^n), \varepsilon(\mathbf{R}_h\Psi - \Psi)\right) + \left(\mathbf{B}\varepsilon(\rho^n), \varepsilon(\Psi)\right) \\ &\leq C\|\rho^n\|_{\boldsymbol{V}}\|\mathbf{R}_h\Psi - \Psi\|_{\boldsymbol{V}} + C\|\rho^n\|\|\Psi\|_{\mathbf{H}^2} + \|\rho^n\|_{\mathbf{L}^2(\partial\Omega)}\|\Psi\|_{\mathbf{H}^1(\partial\Omega)} \\ &\leq Ch^2 + C\|\rho^n\|. \end{aligned}$$

Hence similarly to the calculation for the $\boldsymbol{V}$-norm, Grönwall's lemma implies that

$$\|\mathrm{D_t}\,\rho^n\| \leq Ch^2\left(\|\mathrm{D_t}\,U^n\|_{\mathbf{H}^2} + Ck\sum_{j=1}^{n-1}\|\mathrm{D_t}\,U^j\|_{\mathbf{H}^2}\right),$$

so that

$$\|\rho^n\| + \|\mathrm{D_t}\,\rho^n\| \leq Ch^2.$$

To bound η^n, we also need a bound on the second derivative of ρ^n. For this, we apply $\mathrm{D_t}$ to (3.6) and then follow the same procedure as above. This shows that

$$\|\mathrm{D_t^2}\,\rho^n\|_{\boldsymbol{V}} \le Ch\left(\|\mathrm{D_t^2}\,U^n\|_{\mathbf{H}^2} + Ck\sum_{j=1}^{n-1}\|\mathrm{D_t^2}\,U^j\|_{\mathbf{H}^2}\right),$$

and similarly for the $\mathbf{L}^2$-norm, but with h^2 instead of h. We do not have pointwise $\mathbf{H}^2$-regularity of $\mathrm{D_t^2}\,U^n$ from Theorem 3.2, but we may estimate the sum by

$$k\sum_{j=1}^{n-1}\|\mathrm{D_t^2}\,U^j\|_{\mathbf{H}^2} \le \left(k\sum_{j=1}^{n-1}\|\mathrm{D_t^2}\,U^j\|_{\mathbf{H}^2}^2\right)^{1/2} \le C,$$

and conclude that

$$\|\mathrm{D_t^2}\,\rho^n\| + h\|\mathrm{D_t^2}\,\rho^n\|_{\boldsymbol{V}} \le Ch^2 + Ch^2\|\mathrm{D_t^2}\,U^n\|_{\mathbf{H}^2}. \tag{3.7}$$

Here the $\|\mathrm{D_t^2}\,U^n\|_{\mathbf{H}^2}$-term is not necessarily finite, but since this bound will only be used inside a sum it causes no problems.

Now for η^n, by using (3.6) to exchange W^n for U^n and then (2.9), (2.12), we get

$$\begin{aligned}\left(\mathrm{D_t^2}\,\eta^n, \boldsymbol{\chi}\right) + \left(\mathbf{A}\varepsilon\left(\mathrm{D_t}\,\eta^n\right) + \mathbf{B}\varepsilon(\eta^n), \varepsilon(\boldsymbol{\chi})\right)&\\ = \left(\mathrm{D_t^2}\,U^n - \mathrm{D_t^2}\,W^n, \boldsymbol{\chi}\right) + \left(\mathbf{M}e_{\theta,h}^n, \varepsilon(\boldsymbol{\chi})\right)&\\ = -\left(\mathrm{D_t^2}\,\rho^n, \boldsymbol{\chi}\right) + \left(Me_{\theta,h}^n, \varepsilon(\boldsymbol{\chi})\right).&\end{aligned}$$

Choosing $\boldsymbol{\chi} = \mathrm{D_t}\,\eta^n \in \boldsymbol{S}_h$, by (3.7) we get, after canceling a $C_2\|\mathrm{D_t}\,\eta^n\|_{\boldsymbol{V}}^2$ term,

$$\mathrm{D_t}\|\mathrm{D_t}\,\eta^n\|^2 + C_2\|\mathrm{D_t}\,\eta^n\|_{\boldsymbol{V}}^2 + \mathrm{D_t}\|\eta^n\|_{\mathbf{B}}^2 \le C\left(h^4 + h^4\|\mathrm{D_t^2}\,U^n\|_{\mathbf{H}^2}^2 + \|e_{\theta,h}^n\|^2\right),$$

so summing and noting again that $k\sum_{j=1}^{n-1}\|\mathrm{D_t}\,U^j\|_{\mathbf{H}^2}^2 \le C$, we have

$$\|\mathrm{D_t}\,\eta^n\|^2 + k\sum_{j=1}^{n-1}\|\mathrm{D_t}\,\eta^j\|_{\boldsymbol{V}}^2 + \|\eta^n\|_{\boldsymbol{V}}^2 \le Ch^4 + Ck\sum_{j=1}^{n-1}\|e_{\theta,h}^j\|^2.$$

Finally, combining the bounds for ρ^n, η^n and their first derivatives leads to the statement of the lemma. □

Remark 3.2 We note that the regularity given in Theorem 3.2 is not enough to show $\|\mathrm{D_t}\,e_{u,h}^n\|_{\boldsymbol{V}}^2 \le Ch^2 + Ck\sum_{j=1}^{n}\|e_{\theta,h}^j\|^2$, but such a bound is not required for the proof of the next theorem.

Theorem 3.3 *Let Assumptions 3.1–3.4 be satisfied and* (Θ^n, Φ^n, U^n) *and* $(\Theta_h^n, \Phi_h^n, U_h^n)$ *be solutions to Eqs. (2.7)–(2.9) and (2.10)–(2.12), respectively. Then there are positive constants* k_0 *and* h_0 *such that if* $k < k_0$ *and* $h < h_0$ *then for* $n = 1, \dots, N$,

$$\|e_{\theta,h}^n\| + \|e_{\phi,h}^n\| + \|\mathrm{D_t}\, e_{u,h}^n\| \leq Ch^2 \quad \text{and} \quad \|e_{\theta,h}^n\|_{H^1} + \|e_{\phi,h}^n\|_{H^1} + \|\mathrm{D_t}\, e_{u,h}^n\|_{\boldsymbol{V}} \leq Ch,$$

with the constant C *independent of* k, h *and* n.

Proof The idea is, similarly to the time-discrete case, essentially to write down the equation for $e_{\theta,h}^n$, test it with $e_{\theta,h}^n$, express the errors $e_{u,h}^n$ and $e_{\phi,h}^n$ in terms of $e_{\theta,h}^j$ by Lemma 3.2 and its potential-analogue, and finally use Grönwall's lemma. However, since $e_{\theta,h}^n$ does not belong to the finite element space, we need to introduce instead

$$e_h^n = \Theta_h^n - R_h \Theta^n,$$

where R_h denotes the Ritz projection onto S_h. Due to Theorem 3.2 we then have $\|e_{\theta,h}^n\| \leq \|e_h^n\| + \|R_h\Theta^n - \Theta^n\| \leq \|e_h^n\| + Ch^2$. It follows that for all $\chi \in S_h$,

$$\begin{aligned}\left(\mathrm{D_t}\, e_h^n, \chi\right) + \left(\nabla\theta_h^n, \nabla\chi\right) &= \left(\mathrm{D_t}\left(\Theta^n - R_h\Theta^n\right), \chi\right) + \left(R_\phi, \chi\right) \\ &\quad - \left(M{:}\varepsilon\left(\mathrm{D_t}\, e_{u,h}^{n-1}\right), \chi\right),\end{aligned}$$

where R_ϕ contains terms related to the potential ϕ. Choosing $\chi = e_h^n$, we know from [22] that

$$\left(R_\phi, e_h^n\right) \leq Ch^3 + Ch^4\|\mathrm{D_t}\,\Theta^n\|_{H^2}^2 + Ch^{-1}\|e_h^{n-1}\|^4 + C\|e_h^{n-1}\|^2 + \frac{1}{4}\|e_h^n\|_{H^1}^2,$$

and we also have by (3.4) that

$$\left(M : \varepsilon\left(\mathrm{D_t}\, e_{u,h}^{n-1}\right), e_h^n\right) \leq C\|\mathrm{D_t}\, e_{u,h}^{n-1}\|^2 + \frac{1}{4}\|e_h^n\|_{H^1}^2.$$

We additionally know that $\|e_h^0\| = \|I_h\theta_0 - \theta_0\| \leq Ch^2 < h^{1/2}$ if $h < h_0$. Assuming that $\|e_h^m\| \leq h^{1/2}$ for $m = 1, \dots, n-1$ therefore means that

$$\mathrm{D_t}\|e_h^m\|^2 + \|e_h^m\|_{H^1}^2 \leq Ch^3 + Ch^4\|\mathrm{D_t}\,\Theta^m\|_{H^2}^2 + C\|e_h^{m-1}\|^2 + C\|\mathrm{D_t}\, e_{u,h}^{m-1}\|^2$$

for $m = 1, \dots, n$, which after summation and usage of Lemma 3.2 yields

$$\begin{aligned}\|e_h^m\|^2 + k\sum_{j=1}^{m}\|e_h^j\|_{H^1}^2 &\leq Ch^3 + Ch^4 + Ck\sum_{j=1}^{m-1}\|e_h^j\|^2 + Ck\sum_{j=1}^{m-1}\|\mathrm{D_t}\, e_{u,h}^j\|^2 \\ &\leq Ch^3 + Ck\sum_{j=1}^{m-1}\left(\|e_h^j\|^2 + Ck\sum_{i=1}^{j}\|e_h^i\|^2\right).\end{aligned}$$

If we now set $g_m = \max_{1\le j\le m}\left(\|e_h^j\|^2 + Ck\sum_{i=1}^{j}\|e_h^i\|^2\right)$ we have

$$g_m \le Ch^3 + Ck\sum_{j=1}^{m-1} g_j,$$

to which we may apply Grönwall's lemma to acquire

$$\|e_h^n\|^2 + Ck\sum_{j=1}^{n}\|e_h^j\|^2 \le \tilde{C}h^3.$$

Hence if $\tilde{C}h^{5/2} \le 1$ we have that $\|e_h^n\| \le h^{1/2}$. Thus by induction $\|e_h^n\| \le h^{1/2}$ holds for all n such that $0 \le n \le N$. But then also the other calculations just performed are valid for $1 \le n \le N$, so in fact $\|e_h^n\| \le h^{3/2}$. This preliminary bound may be used as in [22, p. 631] to show $\|e_{\phi,h}^n\| \le Ch$ and to improve the bound of the quadratic potential term to

$$\left(R_\phi, e_h^n\right) \le Ch^4 + Ch^4\|\mathrm{D_t}\,\Theta^n\|_{H^2}^2 + C\|e_h^{n-1}\|^2 + \frac{1}{4}\|e_h^n\|_{H^1}^2.$$

Hence,

$$\|e_h^n\|^2 + k\sum_{j=1}^{n}\|e_h^j\|_{H^1}^2 \le Ch^4 + Ck\sum_{j=1}^{m-1}\left(\|e_h^j\|^2 + Ck\sum_{i=1}^{j}\|e_h^i\|^2\right),$$

and once more applying Grönwall's lemma to g_n shows that

$$\|e_h^n\|^2 + k\sum_{j=1}^{n}\|e_h^j\|_{H^1}^2 \le Ch^4.$$

This proves $\|e_{\theta,h}^n\| \le Ch^2$, and from [22] we find $\|e_{\phi,h}^n\| + h\|e_{\phi,h}^n\|_{H^1} \le Ch^2$. Applying Lemma 3.2 gives $\|\mathrm{D_t}\,e_{u,h}^n\| \le Ch^2$. Finally, by inverse inequalities we find also that $\|e_{\theta,h}^n\|_{H^1} + \|\mathrm{D_t}\,e_{u,h}^n\|_V \le Ch$. □

Proof (of Theorem 3.1) This follows directly from Theorem 3.2 and Theorem 3.3 upon observing that, e.g.

$$\|\mathrm{D_t}\,U_h^n - \dot{u}_n\| \le \|e_{u,h}\| + \|e_u\| + \|\mathrm{D_t}\,u_n - \dot{u}_n\|,$$

where the last term is bounded in the proper way due to the regularity assumptions on the solution to the continuous system. □

4 Numerical experiments

We have implemented both the method based on (2.10)–(2.12) and the corresponding fully implicit method based on implicit Euler, using FEniCS (see e.g. [4,26]). These implementations were then used to verify our theoretical results by applying them to the following test examples.

4.1 Problem 1

First consider the two-dimensional problem with $\Omega = (0,1)^2$, $\mathbf{M} = I$, $f = [0,0]^T$ and the viscosity and elasticity tensors given in Voigt notation by

$$\mathbf{A} = \mathbf{B} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We take the electrical conductivity to be given by

$$\sigma(\theta) = 2.5 - \arctan(5\theta - 10),$$

which has a rather steep slope close to $\theta = 2$. The initial conditions are given by $\theta_0(x,y) = 0$ and $u_0(x,y) = v_0(x,y) = [0,0]^T$. These functions also define the Dirichlet boundary conditions for θ and u, while for ϕ they are given by $\phi_b(x,y) = 5(1-x)$.

We discretize Ω by first subdividing it into squares and then dividing each square into four triangles. With N_x squares in each dimension, each triangle has diameter $h = 1/N_x$ and the full grid has $4N_x^2$ triangles. We take $N_x \in \{4, 8, 16, 32, 64\}$. Since the error should be $\mathrm{O}(h^2 + k)$, we choose the number of time steps to be $N_t = N_x^2/2$. With the final time $T = 1$, this gives $k = 2h^2$. We emphasize here that the time steps could be taken much larger than this, but illustrating the error is then less straightforward. Finally, because the exact solution of the problem is not available we cannot compute the exact errors. Instead, we compare the different approximations to a reference approximation $(\Theta_{\mathrm{ref}}, \Phi_{\mathrm{ref}}, U_{\mathrm{ref}})$ computed by the implicit Euler scheme with $N_x = 128$ and $N_t = 8192$.

Figure 1 shows the errors

$$\max_{1\le n\le N_t} \|\Theta_h^n - \Theta_{\mathrm{ref}}(t_n)\|_{L^2}, \quad \max_{1\le n\le N_t} \|\Phi_h^n - \Phi_{\mathrm{ref}}(t_n)\|_{L^2} \quad \text{and} \quad \max_{1\le n\le N_t} \|U_h^n - U_{\mathrm{ref}}(t_n)\|_{\mathbf{L}^2} \tag{4.1}$$

for the different discretizations on a logarithmic scale, for both the semi-implicit method (left) and the method based on implicit Euler (right). These clearly exhibit the expected error behaviour predicted by Theorem 3.3, except for the first points where the grid is very coarse. We also note that the errors are very similar in size, which means that the semi-implicit method is much more efficient. A peculiar effect in this

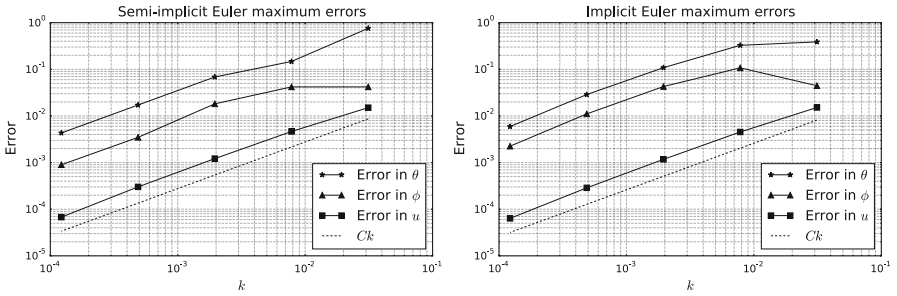

Fig. 1 The errors (4.1) for the problem defined in Sect. 4.1, computed by the semi-implicit method (*left*) and the implicit Euler method (*right*)

case is that the semi-implicit errors in θ and ϕ are actually less than the implicit Euler errors, though this does not hold for the error in u.

4.2 Problem 2

In the second experiment, we investigated the influence of the viscosity on the errors. To this end, we employ the same data as presented in Sect. 4.1 except for the viscosity operator which we set to

$$\mathbf{A} = \gamma \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(in Voigt notation). In this case, we used $N_x \in \{4, 8, 16, 32\}$ with $N_t = N_x^2/4$ and took $N_x = 64$, $N_t = 1024$ for the reference approximation. We only used the semi-implicit scheme here. The first observation is that varying γ has essentially no effect on the errors in θ and ϕ. This is to be expected, as the influence of u on θ is not so large. We therefore omit the plots of these errors, and instead present the error in u for different values of γ in Fig. 2.

We observe that the error clearly increases as γ is decreased, which is to be expected. Indeed, an inspection of the convergence proof indicates that the L^2-error should be inversely proportional to the coercivity constant of $\mathbf{A}$, and thus also of γ. This is, however, in the worst case. In the current situation, Fig. 2 indicates that even $\gamma = 0$ would be perfectly feasible, though smaller step sizes might be necessary to enter the asymptotic regime.

4.3 Problem 3

For our last numerical experiment, we consider a 3D problem arising from an engineering application, inspired by [16,17]. We let Ω be as in Fig. 3, which also shows a typical spatial tetrahedral discretization. This represents a micro-electro-mechanical system (MEMS) used for precise positioning on small scales. When an electric current is passed through the device from the upper-left connector to the lower-left connector,

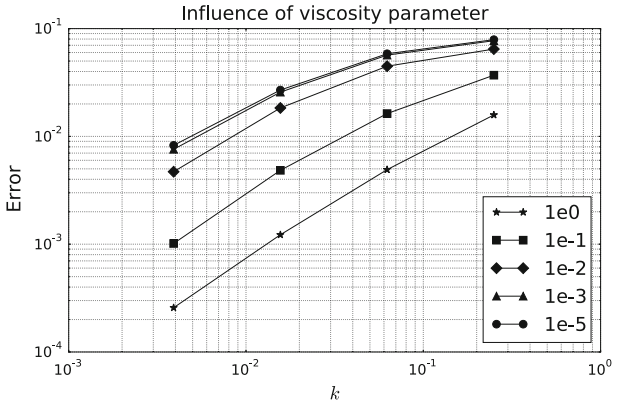

Fig. 2 The errors $\max_{1\leq n\leq N_t} \|U_h^n - U_{\text{ref}}^n\|_{\mathbf{L}^2}$ for the problem defined in Sect. 4.2, computed by the semi-implicit method. The different curves correspond to the different values of $\gamma \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}\}$

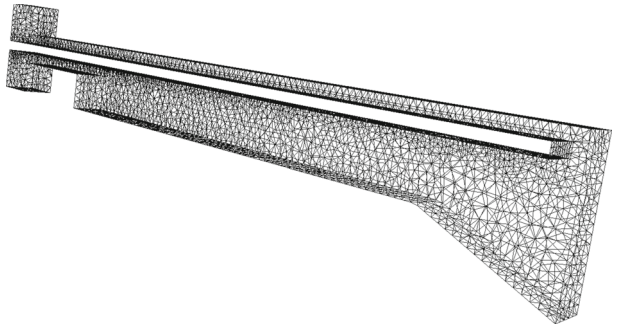

Fig. 3 A mesh for the problem described in Sect. 4.3. The outer dimensions are $192 \times 27 \times 9\,\mu\text{m}$

it heats up. This causes a deformation, which due to the asymmetrical design of the component makes the tip move downwards.

We employ homogeneous Neumann boundary conditions everywhere except for at the left-most edge of the two connectors. These correspond to the component being insulated and stress-free. On the left-most edge we choose the Dirichlet boundary conditions

$$\theta = 0, \quad \phi = \begin{cases} 50, & z > 0 \\ 0, & z < 0 \end{cases}, \quad \text{and} \quad u = v = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

corresponding to the component being clamped and having a potential difference applied between the two connectors. The equations, including physical constants, are

$$\rho c \dot{\theta} = \nabla \cdot \left(\mathbf{K} \nabla \theta \right) + \sigma(\theta) |\nabla \phi|^2 - \Theta_0 \mathbf{M} : \varepsilon(\dot{u}), \tag{4.2}$$

$$0 = \nabla \cdot \left(\sigma(\theta) \nabla \phi \right), \tag{4.3}$$

$$\rho \ddot{u} = \nabla \cdot \left(\mathbf{A} \varepsilon(\dot{u}) + \mathbf{B} \varepsilon(u) - \mathbf{M} \theta \right) + f. \tag{4.4}$$

Here, ρ denotes the density, c the specific heat capacity, $\mathbf{K} = k\mathbf{I}$ the thermal conductivity matrix, $\mathbf{M} = m\mathbf{I}$ the thermal expansion matrix and σ the electrical conductivity. Additionally, θ indicates the deviation from the ambient temperature $\Theta_0 = 293.15\,\mathrm{K}$.

We choose the elasticity and viscosity operators to be given on Lamé parameter form:

$$\mathbf{A}\varepsilon(\dot{u}) = 2\eta_1\varepsilon(\dot{u}) + \eta_2 \operatorname{tr}\varepsilon(\dot{u})\mathbf{I} \quad \text{and} \quad \mathbf{B}\varepsilon(u) = 2\mu\varepsilon(u) + \lambda \operatorname{tr}\varepsilon(u)\mathbf{I},$$

where

$$\mu = \frac{E}{2(1+\nu)} \quad \text{and} \quad \lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}$$

are given in terms of Poisson's ratio ν and Young's modulus E, and η_1, η_2 are corresponding viscosity parameters. Here, tr denotes the trace of a matrix; $\operatorname{tr}\tau = \tau_{11} + \tau_{22}$.

The parameter values we have used, similar to the material properties of silicon, are listed in Table 1. In addition to this, we take $f = [0, 0, 0]^T$ and choose the electrical conductivity as

$$\sigma(\theta) = \frac{38 \cdot 10^6}{27}\left(3000 + 550\left(\frac{\pi}{2} + \arctan\frac{\theta_1 - 250}{250}\right)\right)^{-1}\mathrm{S\,m^{-1}},$$

where $\theta_1 = \Theta_0 + \theta$.

We solve the problem until the time $T = 0.1$ using the semi-implicit method for different spatial and temporal discretizations. The maximum sizes h of the tetrahedrons that were used and the corresponding number of vertices are listed in Table 2. The time steps were again taken proportional to h^2 but modified slightly to yield an integer number of steps. Since the temporal grids thus generated are not refinements of each other, we measured the error as the sum of the errors at only the points $t_j = \boldsymbol{j} \cdot 10^{-2}$ for $j = 1, \ldots, 10$. These errors are listed in Table 2, and also plotted in Fig. 4. While we cannot apply Theorem 3.3 directly, due to the mixed boundary conditions and the non-convexity of the domain, we observe that we still acquire almost $\mathrm{O}(h^2 + k)$ convergence. The curves wiggle because $k = Ch^2$ is only approximately satisfied, and the different magnitudes of the errors reflect the relative sizes of the solution components. The larger error in θ for the coarsest mesh indicates that it violates either the $k < k_0$ or $h < h_0$ mesh size limitations.

Table 1 Parameter values utilized in Problem 3

Parameter	Value	Unit	Parameter	Value	Unit
ρ	$2.33 \cdot 10^3$	$\mathrm{kg\,m^{-3}}$	c	$0.70 \cdot 10^3$	$\mathrm{J\,kg^{-1}\,K^{-1}}$
k	158	$\mathrm{W\,m^{-1}\,K^{-1}}$	m	$1.33 \cdot 10^5$	$\mathrm{N\,m^{-2}\,K^{-1}}$
ν	0.01	1	E	$150 \cdot 10^7$	$\mathrm{N\,m^{-2}}$
η_1	$1 \cdot 10^6$	$\mathrm{N\,s\,m^{-2}}$	η_2	$5 \cdot 10^6$	$\mathrm{N\,s\,m^{-2}}$

Table 2 Spatial and temporal discretizations parameters as well as maximal errors for the MEMS problem (Sect. 4.3) at the time points $t_j = j \cdot 10^{-2}$ for $j = 1, \ldots, 10$. The last row corresponds to the reference approximation

h	k	Vertices	Error in θ	Error in ϕ	Error in u
$4.82 \cdot 10^{-6}$	$5.00 \cdot 10^{-3}$	5219	$1.44 \cdot 10^{-1}$	$1.42 \cdot 10^{-1}$	$9.65 \cdot 10^{-1}$
$3.56 \cdot 10^{-6}$	$3.33 \cdot 10^{-3}$	7510	$2.74 \cdot 10^{-3}$	$2.18 \cdot 10^{-3}$	$2.00 \cdot 10^{-1}$
$2.80 \cdot 10^{-6}$	$2.00 \cdot 10^{-3}$	11,783	$1.60 \cdot 10^{-3}$	$1.27 \cdot 10^{-3}$	$1.24 \cdot 10^{-1}$
$2.39 \cdot 10^{-6}$	$1.67 \cdot 10^{-3}$	18,719	$1.22 \cdot 10^{-3}$	$9.52 \cdot 10^{-4}$	$8.89 \cdot 10^{-2}$
$2.01 \cdot 10^{-6}$	$1.11 \cdot 10^{-3}$	28,473	$7.72 \cdot 10^{-4}$	$6.00 \cdot 10^{-4}$	$5.04 \cdot 10^{-2}$
$1.33 \cdot 10^{-6}$	$5.26 \cdot 10^{-4}$	85,310	–	–	–

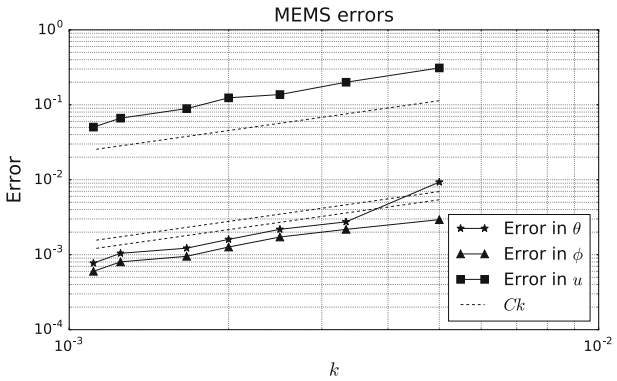

Fig. 4 Maximal errors at the time points $t_j = j \cdot 10^{-2}$ for $j = 1, \ldots, 10$ for the MEMS problem defined in Sect. 4.3. The lines wiggle because $k = Ch^2$ is only approximately satisfied

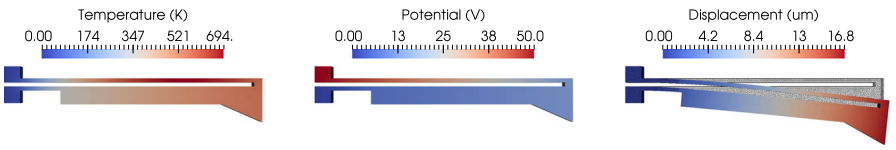

Fig. 5 The approximation to the solution of the problem defined in Sect. 4.3 at $t = T$ and with the finest spatial and temporal discretization. In the right-most plot, the grid has been deformed according to the computed displacement and then super-imposed over the original mesh to illustrate the deformation. We note that the grid is never deformed in the actual computations (this figure is in color in the electronic version of the article)

Finally, Fig. 5 shows the approximations Θ_h^N, Φ_h^N and U_h^N at T, viewed from the side. At this point in time the solutions have just reached their steady state, and we see that the body deforms in the expected fashion.

5 Conclusions and outlook

We have presented a fully discrete numerical method for the fully coupled thermo-viscoelastic thermistor problem (1.1)–(1.3) and proved optimal convergence orders in both space and time. These theoretical results are validated by experimental results.

We reiterate that mixed boundary conditions and re-entrant corners might lead to order reductions. In that case an adaptive mesh refinement strategy may be used, which requires a good a posteriori error estimate. It is possible that the ideas in [3] regarding this can be extended to the present, deformable case.

As illustrated by Sect. 4.3, a typical thermistor is not convex, so a further item that could be improved in the analysis is therefore the shape of the computational domain itself. In this direction we note that the stationary version of the non-deformable problem has been studied in [17,19] for very general domains. It is our ambition to extend these ideas to the time-dependent deformable case in the future.

Finally, a similar analysis would apply also for higher-order methods both in time and space. See e.g. [24] for a Crank–Nicolson-approach to the non-deformable Joule heating problem. However, such an analysis would require extra regularity assumptions that are unfeasible in real-world engineering applications.

Acknowledgements Funding was provided by Vetenskapsrådet (Grant No. 2015-04964).

References

1. Akrivis, G., Larsson, S.: Linearly implicit finite element methods for the time-dependent Joule heating problem. BIT **45**(3), 429–442 (2005). doi:10.1007/s10543-005-0008-1
2. Allegretto, W., Xie, H.: Existence of solutions for the time-dependent thermistor equations. IMA J. Appl. Math. **48**(3), 271–281 (1992). doi:10.1093/imamat/48.3.271
3. Allegretto, W., Yan, N.: A posteriori error analysis for FEM of thermistor problems. Int. J. Numer. Anal. Model. **3**(4), 413–436 (2006)
4. Alnæs, M.S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M.E., Wells, G.N.: The FEniCS project version 1.5. Arch. Numer. Softw. **3**(100), 9–23 (2015). doi:10.11588/ans.2015.100.20553
5. Antontsev, S.N., Chipot, M.: The thermistor problem: existence, smoothness uniqueness, blowup. SIAM J. Math. Anal. **25**(4), 1128–1156 (1994). doi:10.1137/S0036141092233482
6. Chen, X.: Existence and regularity of solutions of a nonlinear nonuniformly elliptic system arising from a thermistor problem. J. Partial Differ. Equ. **7**(1), 19–34 (1994)
7. Ciarlet, P.G.: The finite element method for elliptic problems. In: *Classics in Applied Mathematics*, vol. **40**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2002). doi:10.1137/1.9780898719208. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)]
8. Cimatti, G.: Remark on existence and uniqueness for the thermistor problem under mixed boundary conditions. Q. Appl. Math. **47**(1), 117–121 (1989)
9. Cimatti, G.: Existence of weak solutions for the nonstationary problem of the Joule heating of a conductor. Ann. Mat. Pura Appl. **4**(162), 33–42 (1992). doi:10.1007/BF01759998
10. Duvaut, G., Lions, J.L.: Inequalities in Mechanics and Physics. Springer, Berlin (1976)
11. Elliott, C.M., Larsson, S.: A finite element model for the time-dependent Joule heating problem. Math. Comp. **64**(212), 1433–1453 (1995). doi:10.2307/2153363

12. Fernández, J.R.: Numerical analysis of the quasistatic thermoviscoelastic thermistor problem. M2AN Math. Model. Numer. Anal. **40**(2), 353–366 (2006). doi:10.1051/m2an:2006016
13. Fernández, J.R., Kuttler, K.L.: A dynamic thermoviscoelastic problem: an existence and uniqueness result. Nonlinear Anal. **72**(11), 4124–4135 (2010). doi:10.1016/j.na.2010.01.044
14. Fernández, J.R., Kuttler, K.L.: A dynamic thermoviscoelastic problem: numerical analysis and computational experiments. Q. J. Mech. Appl. Math. **63**(3), 295–314 (2010). doi:10.1093/qjmam/hbq012
15. Grisvard, P.: Elliptic Problems in Nonsmooth Domains, Monographs and Studies in Mathematics, vol. 24. Pitman (Advanced Publishing Program), Boston (1985)
16. Henneken, V.A., Tichem, M., Sarro, P.M.: In-package MEMS-based thermal actuators for microassembly. J. Micromech. Microeng. **16**, 107–115 (2006). doi:10.1088/0960-1317/16/6/S17
17. Holst, M.J., Larson, M.G., Målqvist, A., Söderlund, R.: Convergence analysis of finite element approximations of the Joule heating problem in three spatial dimensions. BIT **50**(4), 781–795 (2010). doi:10.1007/s10543-010-0287-z
18. Howison, S.D., Rodrigues, J.F., Shillor, M.: Stationary solutions to the thermistor problem. J. Math. Anal. Appl. **174**(2), 573–588 (1993). doi:10.1006/jmaa.1993.1142
19. Jensen, M., Målqvist, A.: Finite element convergence for the Joule heating problem with mixed boundary conditions. BIT **53**(2), 475–496 (2013)
20. Kuttler, K.L., Shillor, M., Fernández, J.R.: Existence for the thermoviscoelastic thermistor problem. Differ. Equ. Dyn. Syst. **16**(4), 309–332 (2008). doi:10.1007/s12591-008-0017-z
21. Larsson, S., Thomée, V., Wahlbin, L.B.: Finite-element methods for a strongly damped wave equation. IMA J. Numer. Anal. **11**(1), 115–142 (1991). doi:10.1093/imanum/11.1.115
22. Li, B., Sun, W.: Error analysis of linearized semi-implicit Galerkin finite element methods for nonlinear parabolic equations. Int. J. Numer. Anal. Model. **10**(3), 622–633 (2013)
23. Li, B., Yang, C.: Uniform BMO estimate of parabolic equations and global well-posedness of the thermistor problem. Forum Math. Sigma **3**, e26 (2015). doi:10.1017/fms.2015.29
24. Li, B., Gao, H., Sun, W.: Unconditionally optimal error estimates of a Crank–Nicolson Galerkin method for the nonlinear thermistor equations. SIAM J. Numer. Anal. **52**(2), 933–954 (2014). doi:10.1137/120892465
25. Lin, Y.P., Thomée, V., Wahlbin, L.B.: Ritz–Volterra projections to finite-element spaces and applications to integrodifferential and related equations. SIAM J. Numer. Anal. **28**(4), 1047–1070 (1991). doi:10.1137/0728056
26. Logg, A., Mardal, K.A., Wells, G.N., et al.: Automated Solution of Differential Equations by the Finite Element Method. Springer, Berlin (2012). doi:10.1007/978-3-642-23099-8
27. Nitsche, J.A.: On Korn's second inequality. RAIRO Anal. Numér. **15**(3), 237–248 (1981)
28. Thomée, V.: Galerkin Finite Element Methods for Parabolic Problems, Springer Series in Computational Mathematics, vol. 25, 2nd edn. Springer, Berlin (2006)
29. Thomée, V., Zhang, N.Y.: Error estimates for semidiscrete finite element methods for parabolic integro-differential equations. Math. Comp. **53**(187), 121–139 (1989). doi:10.2307/2008352
30. Wu, X., Xu, X.: Existence for the thermoelastic thermistor problem. J. Math. Anal. Appl. **319**(1), 124–138 (2006). doi:10.1016/j.jmaa.2006.01.076
31. Yuan, G.W., Liu, Z.H.: Existence and uniqueness of the C^{α} solution for the thermistor problem with mixed boundary value. SIAM J. Math. Anal. **25**(4), 1157–1166 (1994). doi:10.1137/S0036141092237893