ESAIM: M2AN 50 (2016) 381–414
DOI: 10.1051/m2an/2015048

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

# TIME OPTIMAL CONTROL FOR A REACTION DIFFUSION SYSTEM ARISING IN CARDIAC ELECTROPHYSIOLOGY – A MONOLITHIC APPROACH *, **

Karl Kunisch[1,2], Konstantin Pieper[3] and Armin Rund[1]

**Abstract.** Motivated by the termination of undesirable arrhythmia, a time optimal control formulation for the monodomain equations is proposed. It is shown that, under certain conditions, the optimal solutions of this problem steer the system into an appropriate stable neighborhood of the resting state. Towards this goal, some new regularity results and asymptotic properties for the monodomain equations with the Rogers−McCulloch ionic model are obtained. For the numerical realization, a monolithic approach, which simultaneously optimizes for the optimal times and optimal controls, is presented and analyzed. Its practical realization is based on a semismooth Newton method. Numerical examples and comparisons are included.

**Mathematics Subject Classification.** 35M30, 49K20, 49J52, 90C46.

Received December 1, 2014. Revised May 6, 2015.
Published online February 19, 2016.

## 1. Introduction and problem formulation

This work is focused on time optimal control of the monodomain equations. This reaction diffusion equation is a simplified version of the bidomain equations which were developed in the late 1970's and which, in conjunction with different ionic models, provide the description of the electrophysiological activity of the heart [34]. The monodomain equations are a reaction diffusion system consisting of a partial differential equation for the electrical potential coupled with an ordinary differential equation describing the ionic variables. They allow for challenging wave phenomena, such as reentry waves, which physiologically correspond to undesired arrhythmias. We introduce a control mechanism which models an external stimulus exerted by means of an electrode, with

*Keywords and phrases.* Time optimal control, monodomain equations, semismooth Newton method, reaction diffusion system, asymptotic behavior.

* *The first and second author gratefully acknowledge support from the International Research Training Group IGDK 1754, funded by the German Science Foundation (DFG) and the Austrian Science Fund (FWF).*

** *The third author gratefully acknowledges the Austrian Science Fund (FWF) for financial support under SFB F32, "Mathematical Optimization and Applications in Biomedical Sciences".*

[1] Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, 8010 Graz, Austria.
karl.kunisch@uni-graz.at; armin.rund@uni-graz.at

[2] Radon Institute, Austrian Academy of Sciences, Austria.

[3] Chair of Optimal Control, Technische Universität München, Boltzmannstraße 3, 85748 Garching bei München, Germany.
pieper@ma.tum.de

the goal of dampening the undesired waves. Due to the dynamical properties of the underlying equations, which include, for example, that excited cells need a certain amount of time before they return to rest, the formulation of the optimal control problem as a time optimal problem offers itself as particularly useful one.

The analysis and numerical realization of the mono- and bidomain equations are themselves an active area of research; see, *e.g.* [5, 10, 31]. Their investigation in the context of optimal control has been taken up only recently; see, *e.g.* [24]. The present paper, however, has yet a second focus, namely the practical treatment of time optimal control problems. The analysis of such problems for ordinary differential equations has received an abundance of attention; see, *e.g.* the monograph [17], and the references therein. Time optimal control of partial differential equations was investigated for instance in [14]. Turning to the numerical treatment of time optimal control problems for ordinary differential equations we mention [19] and further literature quoted there. Numerical techniques for solving time optimal control problems for partial differential equations were developed in [20, 21]. In contrast to these latter papers we propose here a joint optimization of the free final time and the control within one combined optimization variable. For this reason we refer to our approach as the "monolithic" optimization algorithm.

While our work focuses on the monodomain equations, many concepts are applicable to a wider class of reactions diffusions systems. Such systems arise frequently in biomathematical modeling [23] and chemical reaction dynamics [29]. What concerns the genuine treatment of numerical methods for open loop optimal control problems very little specialized attention has been paid to such systems (see, however, *e.g.* [4, 8]).

This paper is organized as follows. Section 2 is devoted to existence, regularity and asymptotic behavior of the state equation. Well-posedness of the optimal control problem is discussed in Section 3, and Section 4 is devoted to obtaining and analyzing the optimality system. Section 5 contains the description of the numerical approach to solve the optimality system. The numerical realization is briefly discussed in Section 6. Numerical examples are provided in Section 7.

### 1.1. Monodomain equations

We start by considering the monodomain equations in the form

$$\partial_t v + I_{\text{ion}}(v,w) - \nabla\cdot(\sigma\nabla v) = I_{\text{e}} \qquad \text{in } (0,t_{\text{f}})\times\Omega, \tag{1.1a}$$

$$\partial_t w + G(v,w) = 0 \qquad \text{in } (0,t_{\text{f}})\times\Omega, \tag{1.1b}$$

$$n\cdot\sigma\nabla v = 0 \qquad \text{on } (0,t_{\text{f}})\times\partial\Omega, \tag{1.1c}$$

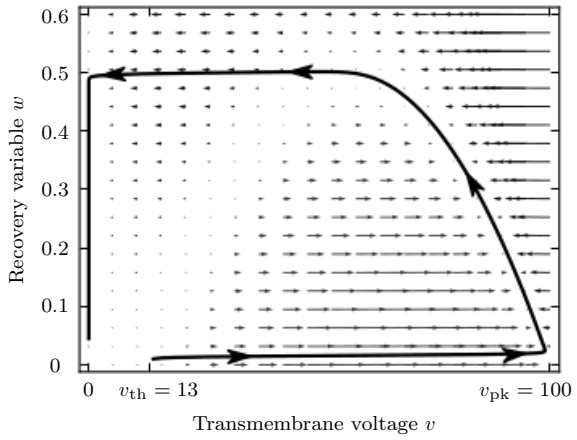
$$v(0) = v_0, \quad w(0) = w_0 \qquad \text{in } \Omega. \tag{1.1d}$$

The independent variables are $x\in\Omega$, with domain $\Omega\subset\mathbb{R}^d$ for $d=2,3$ and outer unit normal vector $n$, and time $t\in(0,t_{\text{f}})$, with final time $t_{\text{f}}>0$. The functions $v(t,x)$ and $w(t,x)$ denote the transmembrane voltage, and the recovery variable (see [31], Sect. 2.4.1), and $\sigma:\Omega\to\mathbb{R}^{d\times d}$ is related to the intracellular conductivity tensor (see [31], Sect. 2.2.3). The functions $I_{\text{ion}}(v,w)$ and $G(v,w)$ are chosen according to the Rogers–McCulloch ionic model (a modified FitzHugh–Nagumo's model) as:

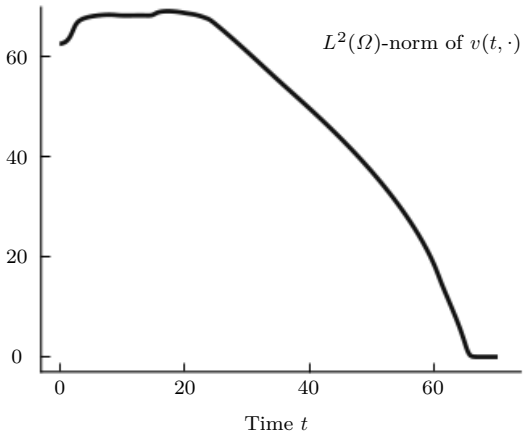
$$I_{\text{ion}}(v,w) = \eta_0\, v\left(1-\frac{v}{v_{\text{th}}}\right)\left(1-\frac{v}{v_{\text{pk}}}\right) + \eta_1\, v\, w, \tag{1.2a}$$

$$G(v,w) = \eta_2\left(\eta_3\, w - \frac{v}{v_{\text{pk}}}\right), \tag{1.2b}$$

with $\eta_i\in\mathbb{R}^+$. A cell at $x\in\Omega$ is referred to as excited if its transmembrane potential exceeds the threshold potential $v_{\text{th}}>0$. Further $v_{\text{pk}}>v_{\text{th}}$ stands for the peak potential. In (1.1a) the forcing function $I_{\text{e}}$ denotes the external (extracellular) stimulus. It describes the current introduced *via* external devices for pacing or defibrillation. We consider devices which consist of several given electrode plates $\Omega_{\text{con},n}$, $n=1,\dots,N_{\text{con}}$. The current signal per electrode plate acts as control and will be designed to control the transmembrane voltage.




(A) Phase diagram and typical solution of the ODE for (1.2).




(B) Spatial norm of $v(t,\cdot)$ over time for a typical example of terminating a reentry wave.

FIGURE 1. Illustrations of the behavior of the nonlinearity (1.2).

Therefore we introduce time dependent control functions $u_n$ and model the external stimulus as

$$I_{\text{e}}(t,x) = \sum_{n=1}^{N_{\text{con}}} \chi_{\Omega_{\text{con},n}}(x)\, u_n(t), \qquad \text{where } t \in (0,t_{\text{f}}). \tag{1.3}$$

It will be convenient to express (1.1) with $I_{\text{e}}$ as in (1.3) and combined state variable $y = (v,w)$ as an abstract control system in the form

$$\begin{aligned} \partial_t y(t) + A(y(t)) &= Bu(t), \quad \text{for } t \in (0,t_{\text{f}}), \\ y(0) &= y_0, \end{aligned} \tag{1.4}$$

with $y_0 = (v_0, w_0)$ as initial condition. We will give a precise definition of this notion in Section 2.

## 1.2. Modeling a successful defibrillation

After a normal activation of the heart the electrical wave dies out and all cells return to the resting state. However, disturbances in the propagation of the impulse may induce a reentry wave, a cycling wave that re-excites the heart muscle cells over and over again. From a physiological point of view, such reentry waves are undesired since they disturb the regular propagation of the electrical impulses, and increase the hearts activation rate. This may lead to arrythmias including atrial or ventricular fibrillation. An example of a reentry wave is investigated in Section 7.

It is our goal to investigate control theory techniques to terminate reentry waves by the external stimulus $I_{\text{e}}$ and to bring the transmembrane potential to the neighborhood of the resting state. From there the natural heart rhythm evolves again following the impulse of the (natural) pacemaker of the human heart, the sinoatrial node.

In Figure 1a we depict the phase diagram of the uncontrolled ordinary differential equation which arises from (1.1) when the second order elliptic operator is eliminated, with parameters chosen as in Section 7. If the initial state is far enough from the origin (*e.g.* $v > v_{\text{th}}$, $w = 0$), then the trajectory rapidly reaches $v_{\text{pk}}$, along the lower part of the curve, from where it slowly moves back to the origin along the upper part of this curve. In the distributed context, *i.e.* for the monodomain equations (1.1), this situation is much more complex since these trajectories are transversed at different times at different spatial locations $x$ leading to interaction which

allows for the evolution of complex wave patterns. Therefore, the asymptotic behavior needs special attention and we shall analyse it in Section 2.3.

For the optimal control formulation this suggests to not constrain or penalize the evolution of the trajectory throughout the complete control time horizon, but rather impose the optimization goal only at the final time $t_\mathrm{f}$. In Figure 1b we depict the $L^2(\Omega)$ norm of an optimized transmembrane potential $v$ as a function of time (terminating the reentry wave in the setup from Sect. 7.2). In view of our discussion of the phaseplane behavior of the corresponding ordinary differential equation, it comes as no surprise that this function is not monotonously decreasing.

Therefore, the aim is to bring the heart muscle tissue approximately to the resting state at some final time $t_\mathrm{f}$ by applying a defibrillation shock. We call the system stabilized, if the controlled transmembrane voltage $v(t,x)$ goes to zero uniformly in $x \in \Omega$ for times $t$ larger than $t_\mathrm{f}$. Since this condition is not directly suitable for an optimal control formulation or for numerical computations, we next establish a control formulation for which we shall be able to verify that it implies the monodomain system to be stabilized. Towards this goal we shall first show that a pointwise condition at the final time of the form

$$v(t_\mathrm{f}, x) \leq v_\mathrm{max} < v_\mathrm{th} \quad \text{for all } x \in \Omega \tag{1.5}$$

ensures that the transmembrane potential goes to zero without further control action, provided that positivity of $v$ and $w$ is guaranteed, see Corollary 2.12. Next we shall replace the pointwise inequality (1.5), which unnecessarily complicates the problem setting, by the weaker integral condition

$$\|v(t_\mathrm{f})\|_{L^2(\Omega)} \leq \delta \tag{1.6}$$

for a sufficiently small $\delta > 0$. A rigorous justification that the corresponding optimal solutions fulfill $\|v(t)\|_{L^\infty(\Omega)} \to 0$ for $t \to \infty$ and thus that the system is stabilized, will be obtained in Theorem 3.4. Finally, in Proposition 3.8 it will be shown that the formulation based on (1.6) is equivalent to a tracking-type formulation, which allows for the design of efficient second order optimization methods.

### 1.3. Time optimal problem formulation

The time frame needed for a successful defibrillation may vary heavily with the problem data. Consequently, the terminal time of the control horizon can not be fixed in advance. The desire of a short defibrillation pulse with low energy is therefore cast as a mixed time optimal control formulation

$$\begin{aligned} \min_{t_\mathrm{f} \geq 0,\, u \in U_\mathrm{ad},\, y=(v,w)} \quad & \int_0^{t_\mathrm{f}} \left(\kappa + \frac{\alpha}{2}\,|u(t)|^2\right) \mathrm{d}t, \\ \text{subject to} \quad & \partial_t y(t) + A(y(t)) = Bu(t) \quad \text{for } t \in I, \\ & y(0) = y_0, \\ & \|v(t_\mathrm{f})\|_{L^2(\Omega)} \leq \delta, \end{aligned} \tag{$\mathcal{P}_\delta$}$$

with pointwise restrictions

$$u \in U_\mathrm{ad} = \{\, u \in U \,|\, |u_n(t)| \leq u_{\mathrm{max},n} \text{ for } n = 1 \ldots N_\mathrm{con}, t \in I \,\},$$

where $u_{\mathrm{max},n} \in \mathbb{R}^+$ are the pointwise bounds for the $n$-th pulse, and $u \in U = L^2(I, \mathbb{R}^{N_\mathrm{con}})$ with $I = (0, t_\mathrm{f})$. The objective prefers small times or pulses with low energy, depending on the choice of the parameters $\kappa > 0$ and $\alpha \geq 0$. In Section 5.3 we shall suppose that $\alpha > 0$ for the analysis of the semismooth Newton method.

To obtain pulses with the desired effect and structure we have to ensure that $\delta > 0$ is chosen small enough. We remark that the choice $\delta = 0$ would require an exact controllability property in the variable $v$. We do not follow up controllability issues for (1.1) here, but point out that, for the combined variable $y = (v, w)$, it was recently shown that the linearized monodomain equations with the FitzHugh–Nagumo's model are not exactly null-controllable, even if the PDE is controlled using $I_\mathrm{e} = u(x,t)$ on all of $\Omega$; see [7].

## 2. Existence and regularity of the state equation

First we recall the existence and regularity results for the solutions of the monodomain equations. We state some basic assumptions.

(1) $\Omega \subset \mathbb{R}^d$ with $d \in \{2, 3\}$ is a bounded domain with Lipschitz boundary.
(2) The conductivity tensor is matrix function $\sigma\colon \Omega \to \mathbb{R}^{d\times d}$ with symmetric, uniformly positive definite and uniformly bounded values.
(3) The initial values are assumed to have the regularity $v_0 \in H^1(\Omega) \cap L^\infty(\Omega)$ and $w_0 \in L^\infty(\Omega)$, except if explicitly mentioned otherwise.

Throughout the paper we abbreviate the spatial spaces without explicitly mentioning the domain $\Omega$, *i.e.* $L^p = L^p(\Omega)$ for any $p \in [1, \infty]$, *etc.* We denote by $V$ the Sobolev space $H^1$ and by $V^*$ its dual. For convenience, we abbreviate as $(\cdot,\cdot) = (\cdot,\cdot)_{L^2}$ the inner product in $L^2$ and as $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_{V^*,V}$ the duality pairing in $V \hookrightarrow L^2 \hookrightarrow V^*$. Furthermore, we use the notations $I = (0, t_\mathrm{f})$ for the time interval and $Q = (0, t_\mathrm{f}) \times \Omega$ for the parabolic cylinder.

### 2.1. Weak solutions

We give the standard weak solution concept for (1.1) that has been considered in the literature.

**Definition 2.1** (Weak formulation of the state equation)**.** For any extracellular stimulus $I_\mathrm{e} \in L^2(I, V^*)$ the tuple $(v, w)$ with

$$v \in L^2(I, V) \cap L^4(Q),\ \partial_t v \in L^2(I, V^*) + L^{4/3}(Q),\ v(0) = v_0, \tag{2.1}$$
$$w \in L^2(Q),\ \partial_t w \in L^2(Q),\ w(0) = w_0 \tag{2.2}$$

is called a weak solution of the monodomain equations as given in (1.1) if

$$\int_I \langle \partial_t v, \varphi_1\rangle + (I_\mathrm{ion}(v, w), \varphi_1) + (\sigma\nabla v, \nabla\varphi_1)\,\mathrm{d}t = \int_I \langle I_\mathrm{e}, \varphi_1\rangle\,\mathrm{d}t, \tag{2.3a}$$
$$\int_I (\partial_t w, \varphi_2) + (G(v, w), \varphi_2)\,\mathrm{d}t = 0 \tag{2.3b}$$

holds for every $(\varphi_1, \varphi_2) \in L^2(I, V) \cap L^4(Q) \times L^2(Q)$.

**Remark 2.2.** The regularity for the time derivative in (2.1) results from from $I_\mathrm{ion}(v, w) \in L^{4/3}(Q)$ for $v \in L^4(Q)$ and $w \in L^2(Q)$. Note, that $L^2(I, V^*) + L^{4/3}(Q) = [L^2(I, V) \cap L^4(Q)]^*$ and is endowed with the corresponding norm. Furthermore, it holds $v \in C(\bar I, L^2)$ for any $v$ with the regularity as in (2.1).

The weak formulation (2.3) was analyzed in [5] for the more general bidomain equations and has been subsequently studied in the context of optimal control (see, *e.g.* [24]).

**Theorem 2.3** (Weak solutions of the state equation)**.** *For $v_0, w_0 \in L^2$ and $I_\mathrm{e} \in L^2(I, V^*)$ the weak formulation (2.3) possesses a unique solution $(v, w)$ satisfying*

$$\|v\|_{L^2(I,V)} + \|v\|^2_{L^4(Q)} + \|\partial_t v\|_{L^2(I,V^*)+L^{4/3}(Q)} + \|v\|_{C(\bar I,L^2)} + \|w\|_{H^1(I,L^2)} \le C\left(1 + \|v_0\|_{L^2} + \|w_0\|_{L^2} + \|I_\mathrm{e}\|_{L^2(I,V^*)}\right),$$

*where the constant $C$ does not depend on $v_0$, $w_0$ or $I_\mathrm{e}$.*

*Proof.* See ([5], Sect. 5) for the existence result and the estimate given above, which is based on a Galerkin argument. The uniqueness result (without additional regularity assumptions) is given in the Appendix; see Proposition A.1. □

Due to the special affine linear form of $G$ we can explicitly give the solutions for the recovery variable $w$ depending on the transmembrane voltage $v$.

**Proposition 2.4** (The solution of the inhibitor equation)**.** *The second component of the monodomain equations for the recovery variable* (1.1b) *with the specific choice of $G$ given in* (1.2b)*, which is*

$$\partial_t w(t,x) + \eta_2\eta_3\, w(t,x) = \frac{\eta_2}{v_{\mathrm{pk}}}\, v(t,x) \quad \text{a.e. in } Q,$$

*has a closed analytic solution, depending on $v$, given by the variation of constants formula*

$$w(t) = W_0(w_0)(t) + W(v)(t) = \mathrm{e}^{-\eta_2\eta_3 t} w_0 + \frac{\eta_2}{v_{\mathrm{pk}}} \int_0^t \mathrm{e}^{-\eta_2\eta_3(t-s)} v(s)\,\mathrm{d}s.$$

*We can verify that for each Banach space $X \subset L^2$ and any $p \in [1,\infty]$ the operators $W_0$ and $W$ are linear and continuous on the spaces*

$$W_0 \colon X \to W^{1,p}(I,X), \quad W \colon L^p(I,X) \to W^{1,p}(I,X).$$

## 2.2. Strong solutions

Due to the cubic nonlinearity appearing in (1.2a), the weak solution concept and corresponding regularity from Theorem 2.3 does not allow for a convenient discussion of the first and second derivatives of the control to state mapping, as we will need in Sections 4 and 5. Since $I_{\mathrm{e}}$ has higher regularity than $L^2(I,V^*)$ in our problem setting (even $I_{\mathrm{e}} \in L^\infty(Q)$), we can derive additional regularity for $v$ using bootstrapping arguments. With this, we can turn to a strong solution concept, which will facilitate the discussion of the nonlinear terms; *cf.* Remark 2.9. To this purpose, we first recall that the second order elliptic operator in the parabolic equation (2.3a) can be understood as an operator

$$(-\nabla\cdot\sigma\nabla) \colon V \to V^*,$$

defined in the usual weak formulation. To formulate the strong solution concept we introduce the space

$$D_2 = \mathrm{dom}_{L^2}(-\nabla\cdot\sigma\nabla) = \left\{ v \in V \middle| -\nabla\cdot\sigma\nabla v \in L^2 \right\}$$

endowed with the graph norm $\|v\|_{D_2} = \| -\nabla\cdot\sigma\nabla v + v\|_{L^2}$. With this definition $D_2$ is a Hilbert space, and $(-\nabla\cdot\sigma\nabla)\colon D_2 \subset L^2 \to L^2$ can be understood as a selfadjoint (unbounded) operator. Due to the low regularity assumptions on $\partial\Omega$ and $\sigma$ we do not have an explicit characterization of $D_2$. However, it is known that the elements of $D_2$ are Hölder continuous, which will be sufficient for our purposes; *cf.* [15].

**Proposition 2.5** (Elliptic regularity for nonsmooth data [16])**.** *Under the general conditions on $\partial\Omega$ and $\sigma$ there exists a constant $\beta > 0$, such that*

$$D_2 \hookrightarrow C^\beta.$$

*Proof.* We apply the elliptic regularity result from [16] to show that the solution operator $(-\nabla\cdot\sigma\nabla + 1)^{-1}$ is bounded from $L^2$ to $C^\beta$ for some $\beta > 0$. □

**Remark 2.6.** In the case that $\partial\Omega$ is $C^2$ and the coefficients of $\sigma$ are $C^1$, this result is the consequence of a classical regularity result for the solution operator $(-\nabla\cdot\sigma\nabla + 1)^{-1}\colon L^2 \to H^2$ and the Sobolev embedding $H^2 \hookrightarrow C^{1/2}$ in spaces up to dimension three.

For the higher regularity and the formulation of the strong solution concept, we introduce the space

$$W_2(0,t_{\mathrm{f}}) = L^2(I,D_2) \cap H^1(I,L^2),$$

which is endowed with the canonical inner product and norm. We have the continuous embedding (see, *e.g.* [22], Sect. 2)

$$W_2(0,t_\mathrm{f}) \hookrightarrow C(\bar{I},[D_2,L^2]_{1/2}) = C(\bar{I},\mathrm{dom}_{L^2}(-\nabla\cdot\sigma\nabla)^{1/2}) = C(\bar{I},V).$$

Note, that the weak suppositions on $\sigma$ and $\partial\Omega$ do not cause additional complications here. More precisely, we apply the trace theorem ([22], Thm. 4.2) for the first embedding and use the characterization/definition of the interpolation space $[D_2,L^2]_{1/2} = \mathrm{dom}_{L^2}(-\nabla\cdot\sigma\nabla)^{1/2}$ (see, *e.g.* [22], Def. 2.1, and [32], Sect. 1.18.10). Furthermore $\mathrm{dom}_{L^2}(-\nabla\cdot\sigma\nabla)^{1/2} = V = H^1$ follows from the definition of $(-\nabla\cdot\sigma\nabla)$ by the weak formulation with continuity and ellipticity of the bilinear from.

**Theorem 2.7** (Additional regularity)**.** *If we have $I_\mathrm{e}\in L^2(Q)$, $v_0\in H^1$ and $w_0\in L^3$, then the solution $(v,w)$ of (2.3) has the regularity $v\in W_2(0,t_\mathrm{f})$ and $w\in H^1(I,L^3)$ with the corresponding estimate*

$$\|v\|_{W_2(0,t_\mathrm{f})} + \|w\|_{H^1(I,L^3)} \le C\left(1+\|v_0\|_{H^1}^2 + \|w_0\|_{L^3}^2 + \|I_\mathrm{e}\|_{L^2(Q)}^2\right). \tag{2.4}$$

*Proof.* We recall that $I_\mathrm{ion}$ is given by $I_\mathrm{ion}(v,w) = R(v)+\eta_1 vw$ with the cubic polynomial $R(\cdot)$ according to (1.2a). It holds that $R'$ is bounded from below by:

$$R'(\cdot) \ge -c_0 = R'\left((v_\mathrm{pk}+v_\mathrm{th})/3\right) = \eta_0\left(1-(v_\mathrm{pk}+v_\mathrm{th})^2/(3v_\mathrm{th}v_\mathrm{pk})\right).$$

Now, we take the weak solution $(v,w)$ from Theorem 2.3 and consider the solution $\tilde{v}$ with $\tilde{v}(0)=v_0$ of the semilinear parabolic equation

$$\partial_t\tilde{v} - \nabla\cdot\sigma\nabla\tilde{v} + R(\tilde{v}) + (1+c_0)\tilde{v} = I_\mathrm{e} + (1+c_0)v - \eta_1 vw, \tag{2.5}$$

which has a strictly monotone nonlinearity $v\mapsto R(v)+(1+c_0)v$. We know that a solution in the sense of (2.3a) is given by $v$. Furthermore, by standard arguments using $R'(\cdot)\ge -c_0$, the solution of (2.5) is unique, which implies $\tilde{v}=v$ (*cf.* Prop. A.1). For the right-hand side in (2.5) it holds that

$$f = I_\mathrm{e} + (1+c_0)v - \eta_1 vw \in L^2(Q).$$

For the last term we use $v\in L^2(I,V)\hookrightarrow L^2(I,L^6)$ from Theorem 2.3, and consequently $w\in L^\infty(I,L^3)$ with Proposition 2.4 (using $w_0\in L^3$). This allows to estimate

$$\|vw\|_{L^2(Q)} \le \|w\|_{L^\infty(I,L^3)}\|v\|_{L^2(I,L^6)} \le \frac{1}{2}\|w\|_{L^\infty(I,L^3)}^2 + \frac{1}{2}\|v\|_{L^2(I,L^6)}^2$$

by Hölder's inequality in time and space and Young's inequality.

Now, we can derive additional regularity of $v$ with a Galerkin argument for the semilinear equation. In the following we give only a sketch of the proof, *i.e.* we do not construct a suitable finite dimensional subspace first. We assume $\partial_t v\in L^2(Q)$. Then we test equation (2.5) with $\partial_t v$ to obtain

$$\|\partial_t v(t)\|_{L^2}^2 + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(\sigma\nabla v(t),\nabla v(t)) + \frac{\mathrm{d}}{\mathrm{d}t}\Psi(v(t)) = (f(t),\partial_t v(t)),$$

where $\Psi(v)=\int_\Omega \psi(v)\,\mathrm{d}x$ and $\psi$ is the quartic polynomial with $\psi'(v)=R(v)+(1+c_0)v$ and $\psi(0)=0$, which is uniformly convex and positive. In fact, by construction we have $\psi(v)\ge (1/2)\,v^2$ for all $v\in\mathbb{R}$ and therefore $\Psi(v)\ge (1/2)\,\|v\|_{L^2}^2$ holds for all $v\in L^4$. Integrating from 0 to $t>0$ we obtain

$$\int_0^t \|\partial_t v\|_{L^2}^2\,\mathrm{d}s + \frac{\gamma}{2}\|\nabla v(t)\|_{L^2}^2 + \Psi(v(t)) \le \frac{1}{2\gamma}\|\nabla v_0\|_{L^2}^2 + \Psi(v_0) + \frac{1}{2}\int_0^t \left(\|f\|_{L^2}^2 + \|\partial_t v\|_{L^2}^2\right)\mathrm{d}t,$$

where $\gamma > 0$ is a constant with $\gamma \|\nabla v\|_{L^2}^2 \leq (\sigma \nabla v, \nabla v) \leq \gamma^{-1} \|\nabla v\|_{L^2}^2$ for all $v \in H^1$. By taking the term $(1/2) \int_0^t \|\partial_t v\|_{L^2}^2 \, \mathrm{d}t$ to the left side and taking the supremum over all $t \in I$ it follows

$$\|\partial_t v\|_{L^2(Q)}^2 + \|v\|_{L^\infty(I,H^1)}^2 \leq C \left( \|v_0\|_{H^1}^2 + \|v_0\|_{L^4}^4 + \|f\|_{L^2(Q)}^2 \right)$$

using $\Psi \geq \|\cdot\|_{L^2}^2$ on the left-hand side, and $\Psi(v_0) \leq C \|v_0\|_{L^4}^4$ on the right-hand side. Now, we go back to (2.5), rewrite it as the elliptic equation

$$-\nabla \cdot \sigma \nabla \tilde{v} + \tilde{v} = f - \partial_t v - R(v) - c_0 v,$$

and obtain $v = \tilde{v} \in L^2(I, D_2)$ by the definition of $D_2$. Together with the Sobolev embedding $\|v_0\|_{L^4} \leq \|v_0\|_{H^1}$, the combined estimates imply (2.4). The estimate for $w$ is simply a consequence of Proposition 2.4. □

Based on the higher regularity from Theorem 2.7 we define a strong solution concept that we will use in the following.

**Definition 2.8** (Strong solutions of the state equation)**.** Suppose that $I_\mathrm{e} \in L^2(Q)$. We call the pair $(v, w)$ with $v \in W_2(0, t_\mathrm{f})$ with $v(0) = v_0 \in V$, and $w \in H^1(I, L^2)$ with $w(0) = w_0 \in L^2$ a (strong) solution of (1.1) if it fulfills equations (2.3) for all test functions in $\varphi_1, \varphi_2 \in L^2(Q)$.

**Remark 2.9.** For Definition 2.8 to make sense, the nonlinear terms have to be well-defined. In fact, for each $v \in W_2(0, t_\mathrm{f})$ and $w \in H^1(I, L^2)$, we can verify with Hölder's inequality that

$$I_\mathrm{ion}(v, w) = R(v) + \eta_1 \, v \, w \in L^2(Q),$$

using $v \in C(\bar{I}, V) \hookrightarrow L^6(Q)$ for the first term, and $v \in L^2(I, D_2) \hookrightarrow L^2(I, L^\infty)$, $w \in L^\infty(I, L^2)$ for the second term.

To make the meaning of the abstract equation (1.4) precise, we introduce the nonlinear operator $A \colon D_2 \times L^2 \to L^2 \times L^2$. It acts on the combined variable $y = (v, w)$ and is given in weak formulation as

$$\langle A(y), \varphi \rangle = (\sigma \nabla v, \nabla \varphi_1) + (I_\mathrm{ion}(v, w), \varphi_1) + (G(v, w), \varphi_2)$$

for any $\varphi = (\varphi_1, \varphi_2) \in V \times L^2$. In the same way, we also define the linear control operator $B \colon \mathbb{R}^{N_\mathrm{con}} \to L^\infty \times \{0\}$, given for $u \in \mathbb{R}^{N_\mathrm{con}}$ as $\langle Bu, \varphi \rangle = (I_\mathrm{e}, \varphi_1)$, where $I_\mathrm{e} = \sum_{n=1}^{N_\mathrm{con}} \chi_{\Omega_{\mathrm{con},n}} \, u_n$. The abstract equation (1.4) is then understood as the weak formulation

$$\int_I \langle \partial_t y, \varphi \rangle + \langle A(y), \varphi \rangle \, \mathrm{d}t = \int_I \langle Bu, \varphi \rangle \, \mathrm{d}t \tag{2.6}$$

for $\varphi \in L^2(Q) \times L^2(Q)$. Together with $y = (v, w) \in W_2(0, t_\mathrm{f}) \times H^1(I, L^2)$ and the initial condition $y(0) = (v_0, w_0) = y_0$, formulation (2.6) is equivalent to Definition 2.8.

**Proposition 2.10** (Bounded solutions)**.** *Suppose that $v_0 \in V \cap L^\infty$, $w_0 \in L^\infty$, and $I_\mathrm{e} \in L^\infty(I, L^2)$. Then the solutions $v$ and $w$ of (2.3) are bounded on the whole cylinder $Q$ with the estimate*

$$\sup_{t \in [0, t_\mathrm{f}]} \left( \|v(t)\|_{L^\infty} + \|w(t)\|_{L^\infty} \right) \leq C \left[ t_\mathrm{f}, \|v_0\|_{L^\infty}, \|v_0\|_{H^1}, \|w_0\|_{L^\infty}, \|I_\mathrm{e}\|_{L^\infty(I,L^2)} \right]. \tag{2.7}$$

*Moreover the solutions are Hölder-continuous (in space and time) for all positive times $t > 0$. Furthermore, there exists $\beta > 0$ such that for all $\varepsilon > 0$ we have the* a priori *estimate*

$$\sup_{t \in [\varepsilon, t_\mathrm{f}]} \left( \|v(t)\|_{C^\beta} + \|w(t)\|_{C^\beta} \right) \leq C \left[ \varepsilon, t_\mathrm{f}, \|v_0\|_{H^1}, \|w_0\|_{L^3}, \|I_\mathrm{e}\|_{L^\infty(I,L^2)} \right]. \tag{2.8}$$

*In both of the last estimates, the constants depend (continuously) on the quantities in angle brackets and the problem setup.*

*Proof.* For uniform boundedness, Hölder regularity, and the estimates (2.7) and (2.8) we now only have to consider the linear part of the parabolic equation. We are going to apply the regularity results from [15]. For the convenience of the reader, we sketch the main steps. First, we split the solution $v = v_1 + v_2$ into the solutions of the linear equations

$$\partial_t v_1 - \nabla \cdot \sigma \nabla v_1 = 0, \qquad v_1(0) = v_0, \tag{2.9a}$$

$$\partial_t v_2 - \nabla \cdot \sigma \nabla v_2 = I_\mathrm{e} - R(v) - \eta_1 v w, \qquad v_2(0) = 0. \tag{2.9b}$$

The solution of (2.9a) is given using the (analytic) semigroup $v_1(t) = \mathrm{e}^{t\,\nabla\cdot\sigma\nabla} v_0$ (*cf.* [15], Thm. 5.2). The semigroup maps the space $L^2$ continuously to the domain of the elliptic operator $D_2$ for all positive times $t$. According to Proposition 2.5 we have $D_2 \hookrightarrow C^\beta$ for a constant $\beta > 0$. Moreover, the semigroup is contractive on $L^\infty$ (see [15], Thm. 4.12), *i.e.* it holds $\|v_1(t)\|_{L^\infty} \le \|v_0\|_{L^\infty}$ for all $t > 0$.

For the solution of (2.9b) we apply a result on maximal parabolic regularity. First we see that by (2.4) we have $g = I_\mathrm{e} - R(v) - \eta_1 v w \in L^\infty(I, L^2)$. In fact, for $I_\mathrm{e}$ this is covered by the assumption. For the other terms we use $v \in W_2(0, t_\mathrm{f}) \hookrightarrow L^\infty(I, L^6)$ and $w \in L^\infty(I, L^3)$ combined with Hölder's inequality in space. Now, by maximal parabolic regularity (*cf.* [15], Thm. 7.4) combined with the trace method of interpolation theory (see, *e.g.* [1], Thm. III 4.10.2) we obtain

$$v_2 \in L^q(I, D_2) \cap W^{1,q}(I, L^2) \hookrightarrow C(\bar{I}, (D_2, L^2)_{1/q,q})$$

for any $q < \infty$, together with a corresponding *a priori* estimate. For sufficiently large choice of $q$, the interpolation space $(D_2, L^2)_{1/q,q}$ embeds into $C^{\beta_\theta}$ for some $\beta_\theta \in (0, \beta)$, (*cf.* [15], Lem. 7.1). Combining all the arguments, we obtain (2.7) and (2.8) for $v = v_1 + v_2$. The estimate for $w$ is again a direct consequence of Proposition 2.4. □

### 2.3. Long time behavior

It already follows from Theorems 2.3 and 2.7 that the solution of (1.1) can be defined for arbitrarily large times. In the following, we study the long time behavior of the solutions. We employ the convention that the solution $(v(t), w(t))$ fullfills the dynamics with $I_\mathrm{e}(t) = 0$ for times $t > t_\mathrm{f}$. In a first step we prove that the uncontrolled system is asymptotically stable in a region containing zero.

**Theorem 2.11.** *Let the state $y(t_\mathrm{f}) = (v(t_\mathrm{f}), w(t_\mathrm{f}))$ satisfy*

$$v_\mathrm{min} \le v(t_\mathrm{f}, \cdot) \le v_\mathrm{max}, \quad w_\mathrm{min} \le w(t_\mathrm{f}, \cdot) \quad \text{in } \Omega, \tag{2.10}$$

*with constants $v_\mathrm{max} \ge 0$ and $w_\mathrm{min}, v_\mathrm{min} \le 0$ fulfilling*

$$v_\mathrm{max} \le v_\mathrm{th}, \quad w_\mathrm{min} > -\frac{\eta_0}{\eta_1}\left(1 - \frac{v_\mathrm{max}}{v_\mathrm{th}}\right)\left(1 - \frac{v_\mathrm{max}}{v_\mathrm{pk}}\right), \quad \text{and } v_\mathrm{min} = v_\mathrm{pk}\eta_3\, w_\mathrm{min}.$$

*Then we have*

$$\|v(t)\|_{L^\infty} \to 0, \quad \|w(t)\|_{L^\infty} \to 0 \quad \text{for } t \to \infty \quad \text{(exponentially fast).}$$

*Proof.* First, we will show that the inequalities (2.10) hold also for all $t \ge t_\mathrm{f}$. Consider that

$$I_\mathrm{ion}(v, w) = \rho(v, w)\, v = \left(\eta_0 \left(1 - \frac{v}{v_\mathrm{th}}\right)\left(1 - \frac{v}{v_\mathrm{pk}}\right) + \eta_1\, w\right) v.$$

The term $\rho(v, w)$ is non-negative for sufficiently small $v$ and sufficiently large $w$, *i.e.* we have

$$\rho(v, w) \ge 0 \quad \text{for all} \quad v \le v_\mathrm{th},\ w \ge -\frac{\eta_0}{\eta_1}\left(1 - \frac{v}{v_\mathrm{th}}\right)\left(1 - \frac{v}{v_\mathrm{pk}}\right).$$

Moreover, for entries $v$ and $w$ with $v \le v_\mathrm{max}$, $w_\mathrm{min} \le w$ we even have $\rho(v, w) \ge \rho_\mathrm{min} = \rho(v_\mathrm{max}, w_\mathrm{min}) > 0$.

Define for the given solutions $v$ and $w$ the function $f(t) = \min_{x\in\bar\Omega} \rho(v(t,x), w(t,x))$. Note, that $f$ is continuous on $[t_\mathrm{f}, \infty)$, since $v, w \in C([t_\mathrm{f}, \infty)\times\bar\Omega)$ according to Proposition 2.10. Due to the assumptions on $v(t_\mathrm{f})$ and $w(t_\mathrm{f})$ we have $f(t_\mathrm{f}) \geq \rho_{\min} > 0$. We define $\hat t = \sup\{\, t \in [t_\mathrm{f}, \infty) | f \geq 0 \text{ on } [t_\mathrm{f}, t]\,\}$. By construction, $\rho(v,w) \geq 0$ holds on $[t_\mathrm{f}, \hat t)\times\bar\Omega$. Due to continuity of $f$, a finite value of $\hat t$ implies that $f(\hat t) = 0$. Next, we show that $\hat t < \infty$ is not possible. Therefore, we consider $z = v - v_{\max}$, which fulfills

$$\partial_t z - \nabla\cdot(\sigma\nabla z) + \rho(v,w)\, z = -\rho(v,w) v_{\max}, \tag{2.11}$$

on $(t_\mathrm{f}, \hat t)\times\Omega$ with homogeneous Neumann boundary conditions, and the initial condition $z(t_\mathrm{f}, \cdot) = v(t_\mathrm{f}, \cdot) - v_{\max} \leq 0$. We define $z^+ = \max(0, z)$, which fulfills $z^+(t_\mathrm{f}) = 0$ and $z^+ \in L^2((t_\mathrm{f}, \hat t), H^1)$ (see, *e.g.* [36], Thm. 2.1.11). Then we test (2.11) with $\chi_{(t_\mathrm{f}, t)} z^+$ for some $t \in (t_\mathrm{f}, \hat t)$ to obtain

$$\begin{aligned}\int_{t_\mathrm{f}}^t \langle \partial_t z, z^+\rangle + (\sigma\nabla z, \nabla z^+) + (\rho(v,w)\, z, z^+)\,\mathrm{d}s \\ = \frac12 \|z^+(t)\|_{L^2}^2 + \int_{t_\mathrm{f}}^t (\sigma\nabla z^+, \nabla z^+) + (\rho(v,w)\, z^+, z^+)\,\mathrm{d}s = -\int_{t_\mathrm{f}}^t (\rho(v,w)\, v_{\max}, z^+)\,\mathrm{d}s \leq 0\end{aligned}$$

for all $t \in (t_\mathrm{f}, \hat t)$. Here we have used the fact that $\langle \partial_t z(t), z^+(t)\rangle = (1/2)\,\frac{\mathrm{d}}{\mathrm{d}t}\|z^+(t)\|_{L^2}^2$ (see [9], Lem. 11.2). All the terms on the left-hand side are non-negative, which implies $z^+ = 0$, wherefore $z \leq 0$ and hence $v \leq v_{\max}$ on the cylinder $[t_\mathrm{f}, \hat t)\times\bar\Omega$. With an analogous argument we see that also $v \geq v_{\min}$ holds on $[t_\mathrm{f}, \hat t)\times\bar\Omega$. Considering the recovery variable, we use the solution formula from Proposition 2.4 to see that for all $t \in [t_\mathrm{f}, \hat t)$ it holds

$$\begin{aligned} w(t,\cdot) &= \mathrm{e}^{-\eta_2\eta_3(t-t_\mathrm{f})} w(t_\mathrm{f}, \cdot) + \frac{\eta_2}{v_\mathrm{pk}} \int_{t_\mathrm{f}}^t \mathrm{e}^{-\eta_2\eta_3(t-s)} v(s,\cdot)\,\mathrm{d}s \\ &\geq w_{\min}\mathrm{e}^{-\eta_2\eta_3(t-t_\mathrm{f})} + \frac{v_{\min}}{v_\mathrm{pk}\eta_3}\left(1 - \mathrm{e}^{-\eta_2\eta_3(t-t_\mathrm{f})}\right) = w_{\min}. \end{aligned} \tag{2.12}$$

The inequality uses the assumption $w(t_\mathrm{f}, \cdot) \geq w_{\min}$, the relation $v(t,\cdot) \geq v_{\min}$ proven above, and the explicit calculation of the integral. Finally, the exponentials cancel out due to the choice $v_{\min} = v_\mathrm{pk}\eta_3\, w_{\min}$. Consequently, inequalities (2.10) are valid for all $t \in [t_\mathrm{f}, \hat t)$. This implies $\rho(v(t,\cdot), w(t,\cdot)) \geq \rho_{\min} > 0$ for all $t \in [t_\mathrm{f}, \hat t)$. Therefore, if $\hat t$ is finite, we obtain $f(\hat t) > 0$ for the function $f$ defined above. This contradicts $f(\hat t) = 0$, which implies that $\hat t = \infty$ must hold. In other words, inequalities (2.10) and $\rho(v(t,\cdot), w(t,\cdot)) \geq \rho_{\min} > 0$ are valid for all $t \geq t_\mathrm{f}$.

Now, we consider the variable $z = v - \mathrm{e}^{-\rho_{\min}(t-t_\mathrm{f})} v_{\max}$. We have

$$\partial_t z - \nabla\cdot(\sigma\nabla z) + \rho(v,w)\, z = (\rho_{\min} - \rho(v,w))\,\mathrm{e}^{-\rho_{\min}(t-t_\mathrm{f})} v_{\max}, \tag{2.13}$$

for $t \geq t_\mathrm{f}$ with $\rho(v,w) - \rho_{\min} \geq 0$. Therefore, the right-hand side in (2.13) is non-positive. With the same argument as before we obtain $z(t) \leq 0$ for all $t \geq t_\mathrm{f}$ which implies $v(t) \leq \mathrm{e}^{-\rho_{\min}(t-t_\mathrm{f})} v_{\max}$. A similar argument shows $v(t) \geq \mathrm{e}^{-\rho_{\min}(t-t_\mathrm{f})} v_{\min}$ and therefore it holds that

$$\|v(t)\|_{L^\infty} \leq \mathrm{e}^{-\rho_{\min}(t-t_\mathrm{f})} \max\{\, v_{\max}, -v_{\min}\,\},$$

for $t \geq t_\mathrm{f}$. Considering again (2.12) and inserting the previous estimate for $v$ yields the estimate $\|w(t)\|_{L^\infty} \leq C(\mathrm{e}^{-\eta_1\eta_2(t-t_\mathrm{f})} + \mathrm{e}^{-\rho_{\min}(t-t_\mathrm{f})})$, which completes the proof. □

**Corollary 2.12.** *Suppose that* $0 \leq v(t_\mathrm{f}, \cdot) \leq v_{\max} < v_{\mathrm{th}}$ *and that* $w(t_\mathrm{f}, \cdot) \geq 0$ *in* $\Omega$. *Then we have*

$$\|v(t)\|_{L^\infty} \to 0, \quad \|w(t)\|_{L^\infty} \to 0 \quad \textit{for } t \to \infty \quad (\textit{exponentially fast}).$$

In a second step we consider estimates of the state for all times $t \in \mathbb{R}^+$. First, we show that the $L^\infty$ bounds on $v$ and $w$ as in Proposition 2.10 can be chosen independently of the final time $t_\mathrm{f}$ for uniformly bounded $I_\mathrm{e} \in L^\infty(Q)$. We remark that for admissible controls $u \in U_\mathrm{ad}$ we have $\|I_\mathrm{e}\|_{L^\infty(Q)} \leq \max_n u_{\max,n}$ independently of $t_\mathrm{f}$.

**Lemma 2.13.** *Suppose $u_0, v_0 \in L^\infty$ and $I_\mathrm{e} \in L^\infty(Q)$. For the solution of (1.1) we have the* a priori *estimate*

$$\sup_{t\in[0,\infty)} \left(\|v(t)\|_{L^\infty} + \|w(t)\|_{L^\infty}\right) \leq C\left[\|v_0\|_{L^\infty}, \|w_0\|_{L^\infty}, \|I_\mathrm{e}\|_{L^\infty(Q)}\right],$$

*where the constant depends (continuously) on the quantities in angle brackets and the problem setup.*

*Proof.* The proof uses comparison principles as in the proof of Theorem 2.11. Here, we have to additionally account for the data term $I_\mathrm{e}$, which is nonzero only for $t \leq t_\mathrm{f}$, and bounded by $C_\mathrm{e} = \|I_\mathrm{e}\|_{L^\infty(Q)}$. Furthermore, we have to account for the non-positivity of $\rho(v,w)$, where $I_\mathrm{ion}(v,w) = \rho(v,w)\, v$. As a first observation, we note that there exists constants $\bar\rho, \bar\gamma > 0$, such that

$$\rho(v,w) = \eta_0 \left(1 - \frac{v}{v_\mathrm{th}}\right)\left(1 - \frac{v}{v_\mathrm{pk}}\right) + \eta_1\, w \geq -\bar\rho + \eta_1\, w + \bar\gamma\, v^2 \tag{2.14}$$

for all entries $v, w$. Now we choose the constants $v_\mathrm{max} \geq \|v_0\|_{L^\infty}$ and $w_\mathrm{max} \geq \|w_0\|_{L^\infty}$ as

$$v_\mathrm{max} = \max\left\{1,\, \|v_0\|_{L^\infty},\, v_\mathrm{pk}\eta_3\|w_0\|_{L^\infty},\, \bar\gamma^{-1}\left(C_\mathrm{e} + \bar\rho + \frac{2\eta_1}{v_\mathrm{pk}\eta_3}\right)\right\}, \tag{2.15}$$

$$w_\mathrm{max} = \frac{v_\mathrm{max}}{v_\mathrm{pk}\eta_3}. \tag{2.16}$$

Our goal is to show that $\|v(t)\|_{L^\infty} \leq v_\mathrm{max}$ and $\|w(t)\|_{L^\infty} \leq w_\mathrm{max}$ holds for all $t \geq 0$. Similar to the proof of Theorem 2.11 we initially define $\hat t$ to be the largest time such that $\|w(t)\|_{L^\infty} \leq 2\, w_\mathrm{max}$ holds for all $t \in [0, \hat t]$. As before, due to continuity of $\|w(\cdot)\|_{L^\infty}$ we have $\hat t > 0$. In fact, with Proposition 2.10 (2.7) we have $v \in L^\infty((0,t_\mathrm{f}), L^\infty)$ and with Proposition 2.4 we obtain $w \in W^{1,\infty}((0,t_\mathrm{f}), L^\infty)$. To show the upper bound $v \leq v_\mathrm{max}$ on $(0,\hat t) \times \Omega$, we consider $z = v - v_\mathrm{max}$, which fulfills

$$\partial_t z - \nabla\cdot\sigma\nabla z = I_\mathrm{e} - \rho(v,w)\, v,$$

on $(0,\hat t) \times \Omega$, together with homogeneous Neumann boundary conditions and the initial condition $z(0,\cdot) = v_0(\cdot) - v_\mathrm{max} \leq 0$. Testing with $\chi_{(0,t)} z^+$ for $t \in (0,\hat t)$ as before, we obtain

$$\frac{1}{2}\|z^+(t)\|_{L^2}^2 + \int_0^t (\sigma\nabla z^+, \nabla z^+)\,\mathrm{d}s = \int_0^t (I_\mathrm{e} - \rho(v,w)\, v, z^+)\,\mathrm{d}s. \tag{2.17}$$

We discuss the right-hand side in a pointwise (almost everywhere) fashion. Therefore, we define the set $Q^+ = \{\,(s,x) \in (0,\hat t) \times \Omega | z(s,x) \geq 0\,\}$. We have

$$\rho(v,w) \geq -\bar\rho + \eta_1\, w + \bar\gamma\, v^2 \geq -\bar\rho - 2\eta_1\, w_\mathrm{max} + \bar\gamma\, v_\mathrm{max}^2 \quad \text{in } Q^+,$$

using (2.14), $w \geq -2w_\mathrm{max}$ on $(0,\hat t) \times \Omega$, and $v \geq v_\mathrm{max}$ on $Q^+$. By the choices of $v_\mathrm{max}$ and $w_\mathrm{max}$, we compute

$$\bar\gamma\, v_\mathrm{max}^2 \geq \left(C_\mathrm{e} + \bar\rho + \frac{2\eta_1}{v_\mathrm{pk}\eta_3}\right) v_\mathrm{max} \geq C_\mathrm{e} + \bar\rho + \frac{2\eta_1 v_\mathrm{max}}{v_\mathrm{pk}\eta_3} = C_\mathrm{e} + \bar\rho + 2\eta_1 w_\mathrm{max}.$$

Consequently, we see that $\rho(v,w) \geq C_\mathrm{e}$ on $Q^+$ and thus

$$I_\mathrm{e} - \rho(v,w)\, v \leq C_\mathrm{e} - \rho(v,w)\, v_\mathrm{max} \leq 0 \quad \text{on } Q^+.$$

Therefore (2.17) implies that $\|z^+(t)\|_{L^2}^2 \le 0$ for all $t \in (0,\hat{t})$, which implies $v \le v_{\max}$ on $(0,\hat{t}) \times \Omega$. With similar arguments for $z = -v_{\max} - v$, we obtain also $v \ge -v_{\max}$ on $(0,\hat{t}) \times \Omega$. It remains to verify that

$$\|w(t)\|_{L^\infty} \le \mathrm{e}^{-\eta_2\eta_3 t}\|w_0\|_{L^\infty} + \frac{\eta_2}{v_{\mathrm{pk}}}\int_0^t \mathrm{e}^{-\eta_2\eta_3(t-s)}\|v(s)\|_{L^\infty}\,\mathrm{d}s \le w_{\max}\,\mathrm{e}^{-\eta_2\eta_3 t} + \frac{v_{\max}}{v_{\mathrm{pk}}\eta_3}\left(1-\mathrm{e}^{-\eta_2\eta_3 t}\right) = w_{\max}$$

for all $t \in [0,\hat{t})$. Since the time $\hat{t}$ was chosen maximal, we obtain $\hat{t} = \infty$ which concludes the proof. □

From Lemma 2.13 we deduce that also the Hölder estimate from Proposition 2.10 does not depend on the final time $t_{\mathrm{f}}$.

**Proposition 2.14.** *Suppose $u_0, v_0 \in L^\infty$ and $I_{\mathrm{e}} \in L^\infty(Q)$. Then there exists a $\beta > 0$, such that for any $\varepsilon > 0$ the solution of (1.1) fulfills the* a priori *estimate*

$$\sup_{t\in[\varepsilon,\infty)} \left(\|v(t)\|_{C^\beta} + \|w(t)\|_{C^\beta}\right) \le C\left[\varepsilon, \|v_0\|_{L^\infty}, \|w_0\|_{L^\infty}, \|I_{\mathrm{e}}\|_{L^\infty(Q)}\right],$$

*where the constant depends (continuously) on the quantities in angle brackets and the problem setup.*

*Proof.* We first show the estimate for the variable $v$. We can rewrite the equation for $v$ as

$$\partial_t v - \nabla\cdot\sigma\nabla v + v = I_{\mathrm{e}} + v - I_{\mathrm{ion}}(v,w) = f.$$

With Lemma 2.13, we have $\sup_{t\in[0,\infty)}\|f(t)\|_{L^\infty} \le C[\|v_0\|_{L^\infty}, \|w_0\|_{L^\infty}, \|I_{\mathrm{e}}\|_{L^\infty(Q)}]$. Recall, that the negative of $E = (-\nabla\cdot\sigma\nabla + 1)\colon D_2 \to L^2$ is the generator of an (analytic) semigroup. Therefore, we can write

$$v(t+\varepsilon) = \mathrm{e}^{-\varepsilon E}v(t) + \int_t^{t+\varepsilon} \mathrm{e}^{-(t+\varepsilon-s)E} f(s)\,\mathrm{d}s \tag{2.18}$$

with the variation of constants formula from semigroup theory (see, *e.g.* [25], Sect. 4.2). To estimate $v(t+\varepsilon)$ in a suitable Hölder norm, we introduce the interpolation spaces

$$D_2^\theta = [D_2, L^2]_{1-\theta} = \mathrm{dom}_{L^2}(-\nabla\cdot\sigma\nabla)^\theta.$$

Since $D_2 \hookrightarrow C^{\tilde{\beta}}$ for some $\tilde{\beta} > 0$ according to Proposition 2.5, there exists a $\theta \in (0,1)$, such that $D_2^\theta \hookrightarrow C^\beta$, for some $\beta \in (0,\tilde{\beta})$ (*cf.* the proof of Prop. 2.10). Furthermore, since the semigroup generated by $-E$ is analytic (see [15, Thm. 5.2]), we have the *a priori* estimate

$$\left\|\mathrm{e}^{-tE}\hat{v}\right\|_{D_2^\theta} \le C\,t^{-\theta}\|\hat{v}\|_{L^2}$$

for any $\hat{v} \in L^2$ with a constant $C$ that is independent of $t > 0$ (see, *e.g.* [25, Thm. 6.13]). Applying this to (2.18) results in

$$\begin{aligned}\|v(t+\varepsilon)\|_{C^\beta} \le C\|v(t+\varepsilon)\|_{D_2^\theta} &\le C\left(\varepsilon^{-\theta}\|v(t)\|_{L^2} + \int_t^{t+\varepsilon}(t+\varepsilon-s)^{-\theta}\|f(s)\|_{L^2}\,\mathrm{d}s\right)\\ &\le C\left(\varepsilon^{-\theta}\|v(t)\|_{L^2} + \frac{\varepsilon^{1-\theta}}{1-\theta}\sup_{s\in(t,t+\varepsilon)}\|f(s)\|_{L^2}\right)\\ &\le C\left[\varepsilon, \|v_0\|_{L^\infty}, \|w_0\|_{L^\infty}, \|I_{\mathrm{e}}\|_{L^\infty(Q)}\right],\end{aligned}$$

where the last estimate is due to Lemma 2.13. The corresponding estimate for $w$ is now a direct consequence of the solution formula from Proposition 2.4. □

## 3. Existence of minimizers

In this section we will discuss well-posedness of the optimal control formulation and derive properties of the optimal solutions. Moreover, we state and analyze a related optimal control problem based on a terminal tracking formulation.

### 3.1. Terminal constraint

First we will argue that the time optimal problem ($\mathcal{P}_\delta$) is well posed under the assumption that there exists an admissible point.

**Assumption 3.1.** Suppose that there exists an admissible triple $(\tilde{t_\mathrm{f}}, \tilde{u}, \tilde{y})$ with $\tilde{t_\mathrm{f}} > 0$ and $\tilde{u} \in U_\mathrm{ad}$ such that $\tilde{y}$ solves equation (1.4) and fulfills the terminal condition $\|\tilde{v}(\tilde{t_\mathrm{f}})\|_{L^2} \leq \delta$.

We note that this assumption requires global stabilizability to the origin for (1.4) in the presence of control constraints. At present we know of no precise conditions which ensure that Assumption 3.1 holds. Magnitude of the initial condition, size and number of control plates, and control bounds will certainly play a role. For the monodomain equations with the classical FitzHugh−Nagumo ionic model, we can find some related results: The question of approximate controllability (with distributed control on a subset) is addressed in [6]. Local stabilizability of all steady states of the monodomain equations with an arbitrary number of control plates is shown in [7]. For the setting considered here, specific setups which allow for stabilization are known experimentally; see the examples in Section 7.

**Theorem 3.2.** *Under Assumption 3.1 the problem ($\mathcal{P}_\delta$) possesses at least one optimal solution $\bar{t_\mathrm{f}} \geq 0$, $\bar{u} \in U_\mathrm{ad}$ with corresponding state solution $\bar{y} = (\bar{v}, \bar{w})$. If we choose $\delta$ such that $\delta < \|v_0\|$, then we have additionally that $\bar{t_\mathrm{f}} > 0$ and $\|\bar{v}(\bar{t_\mathrm{f}})\|_{L^2} = \delta$.*

*Proof.* Since the set of admissible points is not empty, we can select an admissible minimizing sequence $(t_{\mathrm{f}k}, u_k, v_k, w_k)$. We take a subsequence with $t_{\mathrm{f}k} \to \bar{t_\mathrm{f}}$ for $k \to \infty$, extend the $u_k$ (admissibly with respect to $u_k \in U_\mathrm{ad}$) to the interval $I_\mathrm{max} = (0, {t_\mathrm{f}}^\mathrm{max})$ for ${t_\mathrm{f}}^\mathrm{max} = \max_k t_{\mathrm{f}k}$ and select a further subsequence such that $u_k \rightharpoonup \bar{u}$ weakly in $L^2(I_\mathrm{max}, \mathbb{R}^{N_\mathrm{con}})$ for some $\bar{u}$. Since $U_\mathrm{ad}$ is closed and convex and thus weakly closed we obtain $\bar{u} \in U_\mathrm{ad}$. By a standard argument involving weak lower semicontinuity of the squared $L^2(I_\mathrm{max}, \mathbb{R}^{N_\mathrm{con}})$ norm and $t_{\mathrm{f}k} \to \bar{t_\mathrm{f}}$ we have that

$$\inf(\mathcal{P}_\delta) = \liminf_{k\to\infty} \int_0^{t_{\mathrm{f}k}} \left(\kappa + \frac{\alpha}{2}|u_k|^2\right) \mathrm{d}t \geq \int_0^{\bar{t_\mathrm{f}}} \left(\kappa + \frac{\alpha}{2}|\bar{u}|^2\right) \mathrm{d}t.$$

It remains to check that the solution corresponding to $\bar{u}$ fulfills the terminal condition at time $\bar{t_\mathrm{f}}$. For this purpose, we also extend the solutions $(v_k, w_k)$ to the interval $I_\mathrm{max}$ by taking the solutions corresponding to the extended $u_k$. By the *a priori* estimate from Theorem 2.7 the $v_k$ are bounded in $W_2(0, {t_\mathrm{f}}^\mathrm{max})$. Therefore $v_k \rightharpoonup \bar{v}$ in $W_2(0, {t_\mathrm{f}}^\mathrm{max})$ holds for a further subsequence. By taking the limit in all the terms of the weak formulation (2.3), which can be justified also for the nonlinear terms, we verify that $\bar{v}$ is the solution corresponding to the control $\bar{u}$. By the embedding $H^1(I_\mathrm{max}, L^2) \hookrightarrow C^{1/2}(I_\mathrm{max}, L^2)$ we also have strong convergence $v_k(t_{\mathrm{f}k}) - v_k(\bar{t_\mathrm{f}}) \to 0$ in $L^2$ due to Hölder continuity in time, which holds uniformly in $k$. Moreover, it holds $v_k(\bar{t_\mathrm{f}}) - \bar{v}(\bar{t_\mathrm{f}}) \rightharpoonup 0$ weakly in $V = H^1$ since the point evaluation in $V$ at $\bar{t_\mathrm{f}}$, which is a bounded linear operator on $W_2(0, {t_\mathrm{f}}^\mathrm{max})$, preserves weak convergence. By the compactness of the embedding $V \hookrightarrow L^2$, this implies

$$v_k(t_{\mathrm{f}k}) \to \bar{v}(\bar{t_\mathrm{f}}) \quad \text{in } L^2,$$

and therefore $\|\bar{v}(\bar{t_\mathrm{f}})\|_{L^2} \leq \delta$ by continuity, which finishes the first part of the proof.

Now assume that $\|\bar{v}(t_\mathrm{f})\|_{L^2} < \delta$ for some $t_\mathrm{f} > 0$ and an optimal control $\bar{u}$. Due to $\|\bar{v}(0)\|_{L^2} = \|v_0\|_{L^2} > \delta$ and the continuity of $t \mapsto \bar{v}(t) \in L^2$, there is a point $t_0 \in (0, t_\mathrm{f})$ with $\|\bar{v}(t_0)\| = \delta$. This immediately implies that $\bar{t_\mathrm{f}} \leq t_0 < t_\mathrm{f}$ for the optimal $\bar{t_\mathrm{f}}$ with corresponding $\bar{u}$ by comparison of values of the objective for $(t_\mathrm{f}, \bar{u})$ and $(\bar{t_\mathrm{f}}, \bar{u})$. This shows that $\|\bar{v}(\bar{t_\mathrm{f}})\| = \delta$. □

**Remark 3.3.** Certainly, a choice of $\delta \geq \|v_0\|$ is not appropriate, since its global minimizer $\bar{t_{\mathrm{f}}} = 0, \bar{u} = 0$ does not help stabilizing the system.

Based on Theorems 2.11 and Proposition 2.14, we can show that for sufficiently small $\delta$ the optimal solutions of $(\mathcal{P}_\delta)$ will stabilize the system. For this purpose, we show that the pointwise uniform condition from Theorem 2.11 can be realized by the condition on the $L^2$ norm by using the smoothing of the parabolic solution operator.

**Theorem 3.4.** *Assume that Assumption 3.1 holds for arbitrary $\delta > 0$. For a sufficiently small choice of $\delta > 0$ the optimal solutions $(\bar{t_{\mathrm{f}}}, \bar{u})$ with corresponding state $\bar{y} = (\bar{v}, \bar{w})$ of problem $(\mathcal{P}_\delta)$ fulfill*

$$\|\bar{v}(t)\|_{L^\infty} \to 0 \quad \text{for } \bar{t_{\mathrm{f}}} \leq t \to \infty,$$

*provided that the condition $\min_{x\in\bar{\Omega}} \bar{w}(\bar{t_{\mathrm{f}}}, x) \geq w_{\min}$ on the recovery variable holds at the final time for some $w_{\min} > -\eta_0/\eta_1$ independent of $\delta$.*

*Proof.* The proof rests on the uniform Hölder continuity of $v$ as shown in Proposition 2.14. It holds for any positive $\varepsilon > 0$ and some $\beta > 0$ that

$$\max_{t\in[\varepsilon,t_{\mathrm{f}}]} \|v(t)\|_{C^\beta} \leq C[\varepsilon, \|v_0\|_{L^\infty}, \|w_0\|_{L^\infty}, u_{\max}].$$

Now, let $(t_{\mathrm{f}\,\delta}, u_\delta)$ with the corresponding $(v_\delta, w_\delta)$ be optimal solutions of $(\mathcal{P}_\delta)$ for given $\delta > 0$. Furthermore, fix some $\hat{\delta} < \|v_0\|$ with a corresponding optimal solution $(t_{\mathrm{f}\,\hat{\delta}}, u_{\hat{\delta}})$ of problem $(\mathcal{P}_{\hat{\delta}})$. Using the optimality of $(t_{\mathrm{f}\,\hat{\delta}}, u_{\hat{\delta}})$ in problem $(\mathcal{P}_{\hat{\delta}})$, it is clear that

$$0 < \int_0^{t_{\mathrm{f}\,\hat{\delta}}} \left(\kappa + \frac{\alpha}{2}\left|u_{\hat{\delta}}\right|^2\right) \mathrm{d}t \leq \int_0^{t_{\mathrm{f}\,\delta}} \left(\kappa + \frac{\alpha}{2}\left|u_\delta\right|^2\right) \mathrm{d}t \leq t_{\mathrm{f}\,\delta}\left(\kappa + \frac{\alpha}{2}\max_n u_{\max,n}^2\right)$$

for all $\delta < \hat{\delta}$, since $(t_{\mathrm{f}\,\delta}, u_\delta)$ is admissible in problem $(\mathcal{P}_{\hat{\delta}})$. This implies $t_{\mathrm{f}\,\delta} \geq t_{\mathrm{f}\min}$ for all $\delta < \hat{\delta}$ with a $t_{\mathrm{f}\min} > 0$ independent of $\delta$. Consequently we have

$$\|v_\delta(t_{\mathrm{f}\,\delta})\|_{C^\beta} \leq C$$

for a constant $C = C[t_{\mathrm{f}\min}, \|v_0\|_{L^\infty}, \|w_0\|_{L^\infty}, u_{\max}]$ independent of $\delta$. Consider the interpolation space $[L^2, C^\beta]_\theta$ for $\theta \in (0,1)$. For $\theta > d/(2\beta + d)$, such that $\beta_\theta = \theta(\beta + d/2) - d/2 > 0$ we have $[L^2, C^\beta]_\theta \hookrightarrow C^{\beta_\theta} \hookrightarrow L^\infty$ (*cf.* the proof of [15], Lem. 7.1). Using the corresponding interpolation inequality we confirm for any such $\theta$ that

$$\|v_\delta(t_{\mathrm{f}\,\delta})\|_{L^\infty} \leq C\,\|v_\delta(t_{\mathrm{f}\,\delta})\|_{L^2}^{1-\theta}\|v_\delta(t_{\mathrm{f}\,\delta})\|_{C^\beta}^{\theta} \leq C\,\delta^{1-\theta}, \tag{3.1}$$

for a generic constant $C$ independent of $\delta$. The conclusion follows now by a sufficiently small choice of $\delta$, Theorem 2.11, and the assumption on $w(t_{\mathrm{f}})$. □

**Remark 3.5.** As a consequence of the last proof we note that under the assumptions of Theorem 3.4, for any $\eta > 0$, the choice $\delta = (\eta/C)^{1/(1-\theta)}$ guarantees that $\|v_\delta(t_{\mathrm{f}\,\delta})\|_{L^\infty} \leq \eta$, where $C$ is the final constant from (3.1). In other words, by an appropriate choice of $\delta = \delta(\eta)$, an arbitrarily small $L^\infty$-norm will be reached at the final time.

**Remark 3.6.** The assumption on the additional lower bound on $w$ in Theorem 3.4 could be easily dropped by an additional constraint on $\|w(t_{\mathrm{f}})\|_{L^2}$ in the problem formulation. We remark that in computations we have never seen values of $w$ which were smaller than a fraction of $-\eta_0/\eta_1$, and we therefore did not include it. Moreover, $w$ is a phenomenological variable, which cannot be observed in practice.

### 3.2. Terminal tracking

For the numerical realization of problem $(\mathcal{P}_\delta)$ we additionally consider the associated problem

$$\begin{aligned} \min_{t_\mathrm{f}\geq 0,\, u\in U_\mathrm{ad},\, y=(v,w)} \quad & \int_0^{t_\mathrm{f}} \left(\kappa + \frac{\alpha}{2}\,|u|^2\right) \mathrm{d}t + \frac{\mu}{2}\|v(t_\mathrm{f})\|_{L^2}^2, \\ \text{subject to} \quad & \partial_t y + A(y) = Bu, \quad y(0) = y_0. \end{aligned} \tag{$\mathcal{P}^\mu$}$$

For this problem we have to select the penalization parameter $\mu > 0$, which now has to be chosen sufficiently large. Under mild conditions this problem leads to the same optimal solutions as $(\mathcal{P}_\delta)$ provided that $\mu > 0$ and $\delta > 0$ are chosen appropriately, see Proposition 3.8. An existence result analogous to Theorem 3.2 can be obtained here using the same methods.

**Proposition 3.7.** *Problem $(\mathcal{P}^\mu)$ possesses at least one optimal solution $\bar{t}_\mathrm{f} \geq 0$, $\bar{u} \in U_\mathrm{ad}$.*

It is easy to see that $(\mathcal{P}_\delta)$ and $(\mathcal{P}^\mu)$ are closely related to each other.

**Proposition 3.8.** *Let $(\bar{t}_\mathrm{f}, \bar{u})$ be an optimal solution of $(\mathcal{P}^\mu)$ for $\mu > 0$ with the corresponding state $\bar{y} = (\bar{v}, \bar{w})$. Then $(\bar{t}_\mathrm{f}, \bar{u})$ is an optimal solution of $(\mathcal{P}_\delta)$ for $\delta = \|\bar{v}(\bar{t}_\mathrm{f})\|_{L^2}$.*

*Proof.* The optimal solution of each problem is admissible to the other problem. Comparing the objective values of both solutions in both problems immediately shows that the global minimizers of $(\mathcal{P}^\mu)$ are global minimizers of $(\mathcal{P}_\delta)$ due to the choice of $\delta$. □

By Proposition 3.8 we know that the minimizers of $(\mathcal{P}^\mu)$ are also minimizers of $(\mathcal{P}_\delta)$ for a suitable choice of $\delta$. Furthermore, we can see that the solutions of the tracking type problem will yield solutions of the original problem for arbitrarily small $\delta$ (provided that admissible points exist for $(\mathcal{P}_\delta)$ with arbitrarily small $\delta$).

**Proposition 3.9.** *Suppose that Assumption 3.1 holds for $(\mathcal{P}_\delta)$ and let $\delta_1 > \delta$. Then, there exists $M > 0$ such that any optimal solution $(t_{\mathrm{f}\mu}, u_\mu, v_\mu, w_\mu)$ of $(\mathcal{P}^\mu)$ for $\mu \geq M$ fulfills $\|v_\mu(t_{\mathrm{f}\mu})\|_{L^2} \leq \delta_1$.*

*Proof.* Let $(\tilde{t}_\mathrm{f}, \tilde{u})$ be an admissible point for problem $(\mathcal{P}_\delta)$ with associated state solutions $(\tilde{v}, \tilde{w})$. Furthermore denote by $(t_{\mathrm{f}\mu}, u_\mu)$ with corresponding $(v_\mu, w_\mu)$ optimal solutions of $(\mathcal{P}^\mu)$ for given $\mu > 0$. By optimality we have for all $\mu > 0$ that

$$f_\mu + \frac{\mu}{2}\|v_\mu(t_{\mathrm{f}\mu})\|_{L^2}^2 \leq \tilde{f} + \frac{\mu}{2}\|\tilde{v}(\tilde{t}_\mathrm{f})\|_{L^2}^2, \tag{3.2}$$

where $f_\mu = \int_0^{t_{\mathrm{f}\mu}} (\kappa + (\alpha/2)\,|u_\mu|^2)\,\mathrm{d}t$ and $\tilde{f} = \int_0^{\tilde{t}_\mathrm{f}} (\kappa + (\alpha/2)\,|\tilde{u}|^2)\,\mathrm{d}t$. We note that $\tilde{f}$ is independent of $\mu$ by construction. Furthermore we have $\|\tilde{v}(\tilde{t}_\mathrm{f})\|_{L^2} \leq \delta$. Dividing both sides of (3.2) by $2/\mu$, we obtain

$$\frac{2}{\mu}\, f_\mu + \|v_\mu(t_{\mathrm{f}\mu})\|_{L^2}^2 \leq \frac{2}{\mu}\, \tilde{f} + \delta^2.$$

Since $f_\mu$ is positive this implies $\|v_\mu(t_{\mathrm{f}\mu})\|_{L^2}^2 \leq \delta_1^2$ for all $\mu \geq 2\tilde{f}/(\delta_1^2 - \delta^2) = M$. □

This shows that minimizers of $(\mathcal{P}_\delta)$ with the desired properties as in Theorem 3.4 can be found by solving $(\mathcal{P}^\mu)$ with sufficiently large $\mu$. Therefore, we will focus on $(\mathcal{P}^\mu)$ in the following.

## 4. Optimality system

In the following we derive optimality conditions for the time optimal control problem. Therefore, we first introduce a time transformation to a fixed reference interval, and then analyze the linearized state equation.

### 4.1. Time transformation

For the derivation of optimality conditions and the numerical realization, we transform the problem to a fixed reference interval $\tilde{I} = (0, T)$ with a new time variable $\tilde{t} \in \tilde{I}$. Analogously to ([27], Sect. 4) we introduce the local "velocity of time" $\nu \in L^\infty(0, T)$ on the reference interval with

$$\nu \in \mathcal{N}_{\mathrm{ad}} = \left\{ \nu \in L^\infty(0,T) | \operatorname{ess\,inf}_{s\in(0,T)} \nu(s) > 0 \right\},$$

which is an open subset of $\mathcal{N} = L^\infty(0, T)$. The true time $t \in (0, t_{\mathrm{f}})$ and the free end time $t_{\mathrm{f}}$ are now given as

$$t = \theta(\tilde{t}) = \int_0^{\tilde{t}} \nu(s)\,\mathrm{d}s \quad \text{and} \quad t_{\mathrm{f}} = \theta(T) = \int_{\tilde{I}} \nu(s)\,\mathrm{d}s. \tag{4.1}$$

Note that the function $\theta$ is strictly monotonously increasing and Lipschitz continuous. We define $\tilde{y} = y(\theta(\tilde{t}))$, $\tilde{u} = u(\theta(\tilde{t}))$ and consider the transformed equation on the reference interval

$$\partial_t \tilde{y} + \nu A(\tilde{y}) = \nu B\tilde{u}, \quad \tilde{y}(0) = y_0, \tag{4.2}$$

which is equivalent to the weak formulation defined for $\tilde{y} \in W_2(0, T) \times H^1(\tilde{I}, L^2)$ as

$$\int_{\tilde{I}} \langle \partial_t \tilde{y}, \varphi\rangle + \nu\langle A(\tilde{y}), \varphi\rangle\,\mathrm{d}t = \int_{\tilde{I}} \nu\langle B\tilde{u}, \varphi\rangle\,\mathrm{d}t$$

for all $\varphi \in L^2(\tilde{I}, L^2) \times L^2(\tilde{I}, L^2)$. It is easy to verify that a solution of (4.2) exists, is unique and coincides with the solution of (1.4) by an appropriate rescaling.

**Proposition 4.1.** *For each $\nu \in \mathcal{N}_{\mathrm{ad}}$ and $\tilde{u} \in U_{\mathrm{ad}}$ the unique solution $\tilde{y}$ to (4.2), with the same regularity as in Theorem 2.7, corresponds to the solution $y$ of (1.4) for $u \in U_{\mathrm{ad}}$ with the relations*

$$y(t) = \tilde{y}(\theta^{-1}(t)), \quad u(t) = \tilde{u}(\theta^{-1}(t)), \tag{4.3}$$

*with the function $\theta\colon (0, T) \to (0, t_{\mathrm{f}})$ from (4.1).*

*Proof.* We give a short sketch of the proof: First we observe that for the unique solution $y$ of (1.4) the function $\tilde{y}$ with $\tilde{y}(\tilde{t}) = y(\theta(\tilde{t}))$ is an element of $W_2(0, T) \times H^1(\tilde{I}, L^2)$ with

$$\partial_t \tilde{y} = \nu\,(\partial_t y \circ \theta) \quad \text{in } L^2(\tilde{I}, L^2).$$

Here, we used $\partial_t \theta = \nu \in L^\infty(0, T)$. Therefore, $\tilde{y}$ solves the rescaled equation (4.2) by a change of variables, which shows existence for (4.2). Uniqueness follows using the inverse transformation. For any solution of (4.2) we obtain with (4.3) a solution of the original problem. Here, we use the fact that

$$\partial_t \theta^{-1} = \frac{1}{\nu \circ \theta^{-1}} \in L^\infty(0, t_{\mathrm{f}}),$$

which is guaranteed by $\nu \in \mathcal{N}_{\mathrm{ad}}$. □

Now, we see that we can rewrite the optimization problem ($\mathcal{P}^\mu$) in the new coordinates as

$$\begin{aligned} \min_{\nu\in\mathcal{N}_{\mathrm{ad}},\, \tilde{u}\in U_{\mathrm{ad}},\, \tilde{y}=(\tilde{v},\tilde{w})} \quad & \int_{\tilde{I}} \nu \left(\kappa + \frac{\alpha}{2}|\tilde{u}|^2\right)\mathrm{d}t + \frac{\mu}{2}\|\tilde{v}(T)\|_{L^2}^2, \\ \text{subject to} \quad & \partial_t \tilde{y} + \nu A(\tilde{y}) = \nu B\tilde{u}, \quad \tilde{y}(0) = y_0. \end{aligned} \tag{$\mathcal{P}_{\mathrm{fix}}$}$$

We can not expect the solutions of ($\mathcal{P}_{\mathrm{fix}}$) to be unique since we have a large amount of freedom in the choice of $\nu$. For algorithmic realization, an adequate specialization of $\nu$ will be introduced in Section 5.1. However, the additional freedom in the choice of $\nu$ will offer a convenient way of deriving the Hamiltonian condition (see Thm. 4.8). Besides, we can easily see how the rescaled problem corresponds with the original one.

**Proposition 4.2.** *If the optimal final time $\bar{t}_f$ from Proposition 3.7 is not zero, the solution set of the rescaled problem ($\mathcal{P}_{\text{fix}}$) coincides with solution set of ($\mathcal{P}^\mu$) by the relations*

$$t_f = \theta(T), \quad u(t) = \tilde{u}(\theta^{-1}(t)), \quad \text{and} \quad y(t) = \tilde{y}(\theta^{-1}(t)).$$

In the following we will omit the superscript $\tilde{}$ for the transformed variables for ease of presentation. In particular we will again denote the time variable on the fixed interval by $t \in I = (0,T)$ and also reuse the symbols $u$, $y$, and $p$ on the fixed time frame.

## 4.2. Differentiability

Since we are going to work with a control reduced approach we start by defining the control to state mapping according to Proposition 4.1 as the operator

$$\begin{aligned} S\colon \mathcal{N}_{\text{ad}} \times U &\to W_2(0,T) \times H^1(I,L^2), \\ S(\nu,u) &= y. \end{aligned}$$

We are going to prove that $S$ is continuously differentiable in a suitable sense. For the proof of the differentiability of the control to state mapping we have to analyze the linearized state equation.

**Definition 4.3** (Linearized state equation)**.** We call a $\delta y = (\delta v, \delta w) \in W_2(0,T) \times H^1(I,L^2)$ with $\delta y(0) = 0$ a solution of the linearized state equation at some function $y = (v,w) \in W_2(0,T) \times H^1(I,L^2)$ for the given data $f = (f_1, f_2) \in L^2(Q) \times L^2(Q)$ if it fulfills

$$\partial_t \delta v + \nu \left( I'_{\text{ion},v}(v,w)\delta v + I'_{\text{ion},w}(v)\delta w - \nabla \cdot \sigma \nabla \delta v \right) = \nu f_1, \tag{4.4a}$$

$$\partial_t \delta w + \nu\, G(\delta v, \delta w) = \nu f_2, \tag{4.4b}$$

in the sense of Definition 2.8 (which incorporates the natural zero boundary condition).

We will also use the following abstract form for (4.4) given by the evolution equation

$$\partial_t \delta y + \nu A'(y)\delta y = \nu f, \qquad \delta y(0) = 0, \tag{4.5}$$

where the spatial operator $A'(y)$ for $y = (v,w)$ is defined in the obvious way according to (4.4).

**Lemma 4.4.** *For every $\nu \in \mathcal{N}_{\text{ad}}$, $y = (v,w) \in L^\infty(Q) \times L^\infty(Q)$ and given data $f = (f_1, f_2) \in L^2(Q) \times L^2(Q)$ the linearized state equation (4.4) has a unique solution $\delta y = (\delta v, \delta w) \in W_2(0,T) \times H^1(I,L^2)$ with the corresponding* a priori *estimate*

$$\|\delta v\|_{W_2(0,T)} + \|\delta w\|_{H^1(I,L^2)} \le C \left( \|f_1\|_{L^2(Q)} + \|f_2\|_{L^2(Q)} \right).$$

*The generic constant $C$ may depend on $\nu$ and $y$, but does not depend on either $f$ or $\delta y$.*

*Proof.* First, we can argue by a similar transformation argument as in Proposition 4.1 that it suffices to consider (4.5) for $\nu \equiv 1$. Then we eliminate the variable $\delta w$ with Proposition 2.4. We obtain an integro-differential equation in terms of $\delta v$ as

$$\partial_t \delta v - \nabla \cdot \sigma \nabla \delta v + I'_{\text{ion},v}(v,w)\delta v + \eta_1 v\, W(\delta v) = f_1 + (v_{\text{pk}}/\eta_2)\, W(f_2),$$

which is equivalent to (4.5). Now, to obtain $L^2$-contractivity we replace $\delta v$ with the variable $\tilde{v} = \mathrm{e}^{-\gamma t}\delta v$ for a $\gamma > 0$. It is easy to check that the corresponding equation for $\tilde{v}$ is given by

$$\partial_t \tilde{v} - \nabla \cdot \sigma \nabla \tilde{v} + I'_{\text{ion},v}(v,w)\tilde{v} + \eta_1 v\, W_\gamma(\tilde{v}) + \gamma \tilde{v} = \mathrm{e}^{-\gamma t} \left( f_1 + (v_{\text{pk}}/\eta_2)\, W(f_2) \right), \tag{4.6}$$

where $W_\gamma$ has the form

$$W_\gamma(\tilde{v})(t) = \frac{\eta_2}{v_{\text{pk}}} \int_0^t \mathrm{e}^{-(\eta_2\eta_3+\gamma)(t-s)} \tilde{v}(s)\, \mathrm{d}s.$$

We are going to show existence of $\tilde{v}$ by an application of the Banach fixed-point theorem. First, we choose $\gamma > 0$ sufficiently large such that $I'_{\mathrm{ion},v}(v,w) + \gamma > 0$ in $Q$, which is possible due to $v, w \in L^\infty(Q)$. Then we define the mapping $F\colon L^2(Q) \to W_2(0,T)$ by $F(\tilde{v}) = v'$, where $v'$ is the solution of the linear parabolic equation

$$\partial_t v' - \nabla \cdot \sigma \nabla v' + I'_{\mathrm{ion},v}(v,w)v' + \gamma v' = -\eta_1 v\, W_\gamma(\tilde{v}) + \mathrm{e}^{-\gamma t}\left(f_1 + (v_{\mathrm{pk}}/\eta_2)\, W(f_2)\right).$$

With standard linear parabolic solution theory, $F$ is well defined with respect to the given spaces (see, *e.g.* [13], Sect. 7.1). The low regularity assumptions on $\sigma$ and $\partial\Omega$ do not cause difficulty due to the appropriate definition of $W_2(0,T)$ (*cf.* Thm. 2.7). The difference $h = F(\tilde{v}_1) - F(\tilde{v}_2)$ for two given $\tilde{v}_1$ and $\tilde{v}_2$ fulfills

$$\partial_t h - \nabla \cdot \sigma \nabla h + I'_{\mathrm{ion},v}(v,w)h + \gamma h = -\eta_1 v\, W_\gamma(\tilde{v}_1 - \tilde{v}_2).$$

By a standard Galerkin estimate (see, *e.g.* [13], Sect. 7.1.2), which is obtained by testing this equation with the solution $h$, we obtain

$$\|h\|_{L^2(Q)} \le C_\gamma \|\eta_1 v\|_{L^\infty(Q)} \|W_\gamma(\tilde{v}_1 - \tilde{v}_2)\|_{L^2(Q)}$$

for a constant $C_\gamma$ depending only on $\gamma$, the coefficients in the parabolic equation, and the domains. By inspection of the proof of the estimate we obtain that $C_\gamma$ can be bounded independently of $\gamma$. In fact, we even have $C_\gamma \to 0$ for $\gamma \to \infty$. Furthermore one can show that

$$\|W_\gamma(\tilde{v}_1 - \tilde{v}_2)\|_{L^2(Q)} \le \frac{\eta_2}{v_{\mathrm{pk}}} \sqrt{\frac{T}{2(\eta_2\eta_3 + \gamma)}}\, \|\tilde{v}_1 - \tilde{v}_2\|_{L^2(Q)}$$

by an application of Hölder's inequality (*cf.* also [8], Sect. 2.2). In combination, we have

$$\|h\|_{L^2(Q)} \le \frac{1}{2} \|\tilde{v}_1 - \tilde{v}_2\|_{L^2(Q)}$$

for a sufficiently large choice of $\gamma > 0$. By the Banach fixed-point theorem there exists a unique $\tilde{v} \in L^2(Q)$ with $\tilde{v} = F(\tilde{v}) \in W_2(0,T)$ with the corresponding *a priori* estimate

$$\|\tilde{v}\|_{W_2(0,T)} \le C\left(\|f_1\|_{L^2(Q)} + \|f_2\|_{L^2(Q)}\right).$$

By construction $\tilde{v}$ solves (4.6) and the corresponding $\delta v$ together with $\delta w = W(\delta v) + (v_{\mathrm{pk}}/\eta_2)\, W(f_2)$ solves (4.4) with the same estimate, albeit with a bigger constant. The regularity of $\delta w$ follows directly from Proposition 2.4. □

Now, we can discuss differentiability of the control to state mapping. Since we only have a solution theory for the tangent equation for states corresponding to controls in $L^\infty(I, \mathbb{R}^{N_{\mathrm{con}}})$, ($v, w \in L^\infty(Q)$ can only be assured for $I_{\mathrm{e}} \in L^\infty(I, L^2)$, *cf.* Prop. 2.10), we only obtain differentiability in a neighborhood of $U_{\mathrm{ad}}$. Note however, that this neighborhood can be chosen with respect to the norm in $U = L^2(I, \mathbb{R}^{N_{\mathrm{con}}})$.

**Theorem 4.5.** *The control to state mapping $S$ is (arbitrarily often) continuously Fréchet differentiable as an operator from*

$$S\colon \mathcal{N}_{\mathrm{ad}} \times U_{\mathrm{ad}} \subset L^\infty(I) \times U \to W_2(0,T) \times H^1(I, L^2).$$

*The first derivative at $(\nu, u) \in \mathcal{N}_{\mathrm{ad}} \times U_{\mathrm{ad}}$ in a direction $(\delta\nu, \delta u) \in L^\infty(I) \times U$ is given as the solution $\delta y$ of the tangent equation, which is given by*

$$\partial_t \delta y + \nu A'(y)\delta y = \delta\nu(Bu - A(y)) + \nu B\delta u, \qquad \delta y(0) = 0. \tag{4.7}$$

*Proof.* We are going to apply the implicit function theorem. To this purpose we consider the state $y$ as the unique solution of the nonlinear equation

$$e(\nu, u, y) = \partial_t y + \nu A(y) - \nu B u = 0, \tag{4.8}$$

where $e\colon L^\infty(I) \times U \times W_2(0,T) \times H^1(I,L^2) \to L^2(Q) \times L^2(Q)$, and argue that $e$ is Fréchet differentiable. First we consider the mapping

$$y = (v, w) \in W_2(0,T) \times H^1(I,L^2) \mapsto I_{\text{ion}}(v,w) = R(v) + \eta_1 vw \in L^2(Q).$$

For the first term we use the embedding $W_2(0,T) \hookrightarrow L^6(Q)$. The differentiability (arbitrarily often) of the superposition operator $R\colon L^6(Q) \to L^2(Q)$ induced by the cubic polynomial can be verified with a direct computation using Hölder's inequality. Similarly we use the embeddings $H^1(I,L^2) \hookrightarrow L^\infty(I,L^2)$ and $L^2(I,D_2) \hookrightarrow L^2(I,L^\infty)$ combined with Hölder's inequality for the second bilinear term. With this it is evident that $e$ is (arbitrarily often) continuously differentiable in the variables $(u,y)$ for fixed $\nu$, since all the other parts are linear. For the total differentiability, we use the following (standard) argument: Since the mapping $(u,y) \mapsto A(y) - Bu$ is continuously differentiable, the function

$$(\nu, u, y) \in L^\infty(I) \times U \times W_2(0,T) \times H^1(I,L^2) \;\mapsto\; \nu\,(A(y) - Bu) \in L^2(Q) \times L^2(Q)$$

is continuously differentiable as well. This is again essentially a consequence of Hölder's inequality, using $\nu \in L^\infty(I)$. A similar statement holds for the higher derivatives.

We have shown the first prerequisite of the implicit function theorem. The second prerequisite requires the partial derivative, given by

$$e'_y(\nu,u,y)(\cdot) = (\partial_t + \nu A'(y))\colon\ W_2(0,T) \times H^1(I,L^2) \to L^2(Q) \times L^2(Q),$$

to be an isomorphism at the point $(\nu, u, S(\nu,u))$. Obviously, the operator is bounded. Bounded invertibility follows from Lemma 4.4, which is applicable since $y = (v,w) \in L^\infty(Q) \times L^\infty(Q)$ for all admissible controls $(\nu,u) \in \mathcal{N}_{\text{ad}} \times U_{\text{ad}}$ according to Theorem 2.7. Therefore the implicit function theorem (see, *e.g.* [12], Thm. 10.2.1) can be applied, *i.e.* for any point $(\nu,u) \in \mathcal{N}_{\text{ad}} \times U_{\text{ad}}$ there exists a neighborhood in $\mathcal{N}_{\text{ad}} \times U$, such that we can uniquely resolve (4.8) for $y$ with continuously differentiable mapping $y = S(\nu,u)$. Since $e$ is arbitrarily often continuously differentiable, this property transfers to the solution operator $S$ (see, *e.g.* [12], Thm. 10.2.3). □

### 4.3. Necessary condition

Next, we give an optimality condition for problem ($\mathcal{P}_{\text{fix}}$), which will be the basis of the optimization algorithm in Section 5. First we introduce the Hamiltonian as;

$$H(u,y,p) = \kappa + \frac{\alpha}{2}\,|u|^2 + \langle Bu - A(y), p\rangle.$$

The corresponding Lagrange function for ($\mathcal{P}_{\text{fix}}$) (in the sense of constrained optimization) is given in terms of the Hamiltonian by:

$$\begin{aligned}\mathcal{L}(\nu,u,y,p) &= \int_I \left(\nu H(u,y,p) - \langle \partial_t y, p\rangle\right) \mathrm{d}t + \frac{\mu}{2}\|v(T)\|_{L^2}^2 \\ &= \int_I \nu\left(\kappa + \frac{\alpha}{2}\,|u|^2\right) \mathrm{d}t + \frac{\mu}{2}\|v(T)\|_{L^2}^2 - \int_I \left(\langle \partial_t y, p\rangle + \nu\langle A(y) - Bu, p\rangle\right) \mathrm{d}t.\end{aligned}$$

Further we define the adjoint equation.

**Definition 4.6** (Adjoint equation)**.** Let $y = (v, w)$ be the solution of (4.2) corresponding to some $\nu \in \mathcal{N}_{\mathrm{ad}}$ and $u \in U_{\mathrm{ad}}$. Then we define the adjoint state $p = (p_1, p_2) \in W_2(0,T) \times H^1(I, L^2)$ as the solution of the adjoint equation

$$-\partial_t p + \nu A'(y)^* p = 0, \qquad p(T) = \mu(v(T), 0). \tag{4.9}$$

in the sense of the usual weak formulation.

For the sake of brevity, we skip the non-abstract form of the adjoint equation. For the full expressions we refer to Appendix A.2. With slight modifications of the proof of Lemma 4.4 one can show existence and regularity.

**Proposition 4.7.** *For $y = (v, w) \in L^\infty(Q) \times L^\infty(Q)$, $\nu \in \mathcal{N}_{\mathrm{ad}}$, and $v(T) \in V$ the adjoint equation (4.9) has a unique solution $p = (p_1, p_2) \in W_2(0,T) \times H^1(I, L^2)$ with the corresponding a priori estimate*

$$\|p_1\|_{W_2(0,T)} + \|p_2\|_{H^1(I,L^2)} \le C \|v(T)\|_{H^1}.$$

With these prerequisites we can give an optimality condition.

**Theorem 4.8** (Optimality conditions)**.** *Let $(\bar\nu, \bar u, \bar y)$ be an optimal solution of ($\mathcal{P}_{\mathrm{fix}}$). Then there exists a unique adjoint state $\bar p$, which fulfills the corresponding adjoint equation (4.9) and a $\lambda \in \partial \chi_{U_{\mathrm{ad}}}(\bar u)$ (the subdifferential of the convex indicator function of $U_{\mathrm{ad}}$ at the point $\bar u$), such that the optimality conditions*

$$H(\bar u(t), \bar y(t), \bar p(t)) = 0 \quad \text{for a.a. } t \in I, \tag{4.10}$$

$$\alpha \bar u + B^* \bar p + \lambda = 0, \tag{4.11}$$

*are fulfilled. For $\alpha > 0$ the optimality condition (4.11) can also be given by the componentwise projection formula*

$$\bar u = P_{\mathrm{ad}}\left(-\frac{1}{\alpha} B^* \bar p\right) = \min\left(\max\left(-\frac{1}{\alpha} B^* \bar p,\, -u_{\max}\right), u_{\max}\right), \tag{4.12}$$

*depending only on the adjoint state $\bar p$.*

*Proof.* We can use standard methods. We introduce the reduced objective functional,

$$j(\nu, u) = \int_I \nu \left(\kappa + \frac{\alpha}{2} |u|^2\right) \mathrm{d}t + \frac{\mu}{2} \|v(T)\|_{L^2}^2, \quad \text{where } y = (v, w) = S(\nu, u).$$

With Theorem 4.5, the chain rule, and the evident differentiability of the explicit part of $j$ we obtain the optimality conditions

$$j'_\nu(\bar\nu, \bar u) = 0 \text{ in } (L^\infty(I))^*, \quad \text{and} \quad j'_u(\bar\nu, \bar u)(\tilde u - \bar u) \ge 0 \text{ for all } \tilde u \in U_{\mathrm{ad}}, \tag{4.13}$$

by a standard result in nonlinear optimization (see, *e.g.* [33]). For the first equality we recall that $\mathcal{N}_{\mathrm{ad}}$ is open in $L^\infty(I)$. Now we compute the specific form of the partial derivatives. For any admissible pair $(\nu, u) \in \mathcal{N}_{\mathrm{ad}} \times U_{\mathrm{ad}}$ we can write $j(\nu, u) = \mathcal{L}(\nu, u, S(\nu, u), p)$ for arbitrary $p \in L^2(I, V) \times L^2(Q)$. Therefore with Theorem 4.5 we obtain the derivative of $j$ in the direction $(\delta\nu, \delta u)$ as

$$j'(\nu, u)(\delta\nu, \delta u) = \mathcal{L}'_\nu(\nu, u, y, p)(\delta\nu) + \mathcal{L}'_u(\nu, u, y, p)(\delta u) + \mathcal{L}'_y(\nu, u, y, p)(\delta y), \tag{4.14}$$

by an application of the chain rule, where $y = S(\nu, u)$ and $\delta y$ is the solution of the tangent equation (4.7). Now we take $p$ to be the adjoint state corresponding to $\nu$ and $y$ as defined in (4.9). With the integration by parts formula in time it is easy to see that (4.9) implies

$$\mathcal{L}'_y(\nu, u, y, p)(\varphi) = \mu(v(T), \varphi_1(T)) - \int_I \left(\langle \partial_t \varphi, p\rangle + \nu \langle A'(y)\varphi, p\rangle\right) \mathrm{d}t = 0,$$

for all $\varphi = (\varphi_1, \varphi_2) \in W_2(0,T) \times H^1(I, L^2)$ with $\varphi(0) = 0$. Therefore, the last term in (4.14) vanishes and we compute

$$j'(\nu, u)(\delta\nu, \delta u) = \int_I (\delta\nu H(u, y, p) + \nu(\alpha u + B^* p) \cdot \delta u)\, \mathrm{d}t. \tag{4.15}$$

Together with the first part of the abstract condition (4.13) and $H(u(\cdot), y(\cdot), p(\cdot)) \in L^1(I)$ we derive the Hamiltonian condition (4.10). From the second part we obtain the variational inequality

$$\int_I \bar\nu\, (\alpha\bar u + B^*\bar p) \cdot (\tilde u - \bar u)\, \mathrm{d}t \geq 0,$$

for all $\tilde u \in U_{\mathrm{ad}}$. With $\inf_I \bar\nu > 0$ and standard methods from convex optimization the variational inequality is equivalent to (4.11) and (4.12), which completes the proof. □

**Corollary 4.9.** *Suppose that $\alpha > 0$. For any optimal $\bar u$ it holds that $\bar u \in H^1(I, \mathbb{R}^{N_{\mathrm{con}}})$. In particular, the optimal controls are continuous in time. Furthermore we have that the Hamiltonian is continuous in time and therefore $H(\bar u(t), \bar y(t), \bar p(t)) = 0$ for all $t \in I$ (the qualifier "a.a." can be dropped).*

*Proof.* We use that $\bar p = (p_1, p_2) \in W_2(0,T) \times H^1(I, L^2)$ and therefore

$$(B^*\bar p)_n = \int_{\Omega_{\mathrm{con},n}} p_1(\cdot, x)\, \mathrm{d}x \in H^1(I) \quad \text{for } n = 1, \dots, N_{\mathrm{con}}.$$

This regularity is preserved for $\bar u = P_{\mathrm{ad}}(-(1/\alpha)\, B^*\bar p)$. From this, and due to $\bar y, \bar p \in C(\bar I, V)$ follows the continuity of the Hamiltonian. □

## 5. Semismooth Newton method

In this section we describe a semismooth Newton method to solve the penalized time optimal problem ($\mathcal{P}^\mu$). We will generally require $\alpha > 0$. Local superlinear convergence of the method is proven.

Note, that the free end time and the control are considered as a combined optimization variable. For this reason we refer to our approach as "monolithic". This is in contrast to the schemes proposed in [21] or [20], where the optimal control is resolved first for a given final time, and then the resulting value function is optimized by another method.

### 5.1. Variation of the end time

To derive an implementable set of necessary conditions, we take the rescaled problem ($\mathcal{P}_{\mathrm{fix}}$) and specialize the general time transformation to the case of a free parameter $\tau$ by choosing the parameterized velocity-of-time as

$$\nu_\tau = 1 + \tau\hat\nu, \tag{5.1a}$$

with a fixed $\hat\nu \in L^\infty(I)$ such that $\int_I \hat\nu\, \mathrm{d}t \neq 0$. In the following, we focus on the straightforward choice $\hat\nu \equiv 1$, and hence

$$\nu_\tau \equiv 1 + \tau, \qquad t_{\mathrm{f}} = \int_I \nu_\tau\, \mathrm{d}t = T(1+\tau).$$

Furthermore, we have $\nu_\tau \in \mathcal{N}_{\mathrm{ad}}$ for all $\tau \in (-1, \infty)$. With this choice of $\nu_\tau$, the optimization problem has the special form

$$\begin{aligned} \min_{\tau \in (-1,\infty),\, u \in U_{\mathrm{ad}},\, y=(v,w)} \quad & \int_I \nu_\tau \left(\kappa + \frac{\alpha}{2}|u|^2\right) \mathrm{d}t + \frac{\mu}{2}\|v(T)\|_{L^2}^2 \\ \text{subject to} \quad & \partial_t y + \nu_\tau A(y) = \nu_\tau B u, \quad y(0) = y_0. \end{aligned} \tag{5.1b}$$

It is easy to verify that for this parameterization, the problems (5.1) and ($\mathcal{P}_{\mathrm{fix}}$) are equivalent.

**Remark 5.1.** In more general situations the choice (5.1a) can be adapted properly. For example, using $\nu = 1 + \tau_1\hat\nu_1 + \tau_2\hat\nu_2$ with adequate characteristic functions $\hat\nu_1, \hat\nu_2$, optimization problems with two different free time points in the combined optimization variable can be treated (*e.g.* observation or switching times).

### 5.2. Newton system

For clarity of presentation, we first describe the optimization procedure without box constraints. It is a straightforward application of Newton's method for minimization problems to the reduced objective functional with some modifications for the free end time. Similar to the proof of Theorem 4.8 we consider the reduced objective $j(\tau, u)$, which is now depending on $\tau$ due to the specialization in (5.1a). It has the form

$$j(\tau, u) = \int_I \nu_\tau \left(\kappa + \frac{\alpha}{2}|u|^2\right) \mathrm{d}t + \frac{\mu}{2}\|v_{\tau,u}(T)\|_{L^2}^2,$$

where $y_{\tau,u} = (v_{\tau,u}, w_{\tau,u}) = S(\nu_\tau, u)$ is the state solution corresponding to $\tau$ and $u$. As in Theorem 4.8 we can compute the derivatives of $j$.

**Proposition 5.2.** *The functional $j$ is arbitrarily often continuously Fréchet differentiable. The gradient of $j$ w.r.t. the inner product in $\mathbb{R} \times U$ is given by*

$$Dj(\tau, u) = \begin{pmatrix} \int_I H(u, y_{\tau,u}, p_{\tau,u})\,\mathrm{d}t \\ \nu_\tau(\alpha u + B^* p_{\tau,u}) \end{pmatrix} = \begin{pmatrix} \int_I (\kappa + (\alpha/2)\ |u|^2 + \langle Bu - A(y_{\tau,u}), p_{\tau,u}\rangle)\,\mathrm{d}t \\ \nu_\tau(\alpha u + B^* p_{\tau,u}) \end{pmatrix}, \tag{5.2}$$

*where $p_{\tau,u}$ is the corresponding adjoint state fulfilling (4.9).*

*Proof.* The first statement is easy to check for the first explicit part of $j$. The differentiablity of the second part is implied by Theorem 4.5 and the chain rule. We follow the same steps as in the proof of Theorem 4.8 and apply the chain rule $\delta\nu = \delta\tau\,\hat{\nu} \equiv \delta\tau$ induced by the parametrization (5.1a) to the representation formula (4.15). □

**Proposition 5.3.** *An application of the Hessian of $j$, given by the symmetric operator*

$$D^2 j(\tau, u) \colon \mathbb{R} \times U \to \mathbb{R} \times U,$$

*can be computed with the representation*

$$D^2 j(\tau, u)(\delta\tau, \delta u) = \begin{pmatrix} \int_I ((\alpha u + B^* p) \cdot \delta u - \langle \delta y, A'(y)^* p\rangle + \langle Bu - A(y), \delta p\rangle)\,\mathrm{d}t \\ \nu_\tau(\alpha \delta u + B^* \delta p) + \delta\tau\,(\alpha u + B^* p) \end{pmatrix}, \tag{5.3}$$

*where $p$ is again the corresponding adjoint state fulfilling (4.9), $\delta y$ is the solution of the tangent equation (4.7), and $\delta p \in W_2(0,T) \times H^1(I, L^2)$ is the solution of the second adjoint equation, given by*

$$\begin{aligned} -\partial_t \delta p + \nu_\tau A'(y)^* \delta p &= -\nu_\tau (A''(y)\delta y)^* p - \delta\tau A'(y)^* p, \\ \delta p(T) &= \mu(\delta v(T), 0). \end{aligned} \tag{5.4}$$

*Proof.* With the same techniques as in Theorem 4.5 we can show that the mapping $(\tau, u) \mapsto p_{\tau,u}$ is (arbitrarily often) continuously differentiable and that the derivative is given by (5.4). Now, we can apply again Theorem 4.5 for the control to state mapping $(\tau, u) \mapsto y_{\tau,u}$ and apply the chain rule to obtain (5.3). □

A Newton method for an unconstrained problem without the restriction $u \in U_{\mathrm{ad}}$ can be based on the Newton update computed as the solution $h \in \mathbb{R} \times U$ of

$$D^2 j(\tau, u)h = -Dj(\tau, u). \tag{5.5}$$

We will go into further detail after incorporating the box constraints in the next section.

**Remark 5.4.** The formulas (5.2) and (5.3) can also be derived from general expressions for an abstract optimal control problem. For instance, the Newton method based on (5.5) fits into the general framework as described in ([18], Chap. 5.2).

**Remark 5.5.** The terms of the second derivative and the equations for $\delta y$ and $\delta p$ which stem from the variation of the free end time contain the expressions $(Bu - A(y))$ and $A'(y)^* p$. Since $y$ and $p$ are solution of the state and adjoint equations, we can replace them by $\partial_t y/\nu$ and $\partial_t p/\nu$ respectively. This will be very convenient for the practical realization in the discrete setting as described in Section 6; see also Appendix A.2.

### 5.3. Box constraints

To efficiently handle the box constraints, we introduce the auxiliary optimization variable $q \in U$ and parameterize the control as

$$u = u_q = P_{\mathrm{ad}}(q). \tag{5.6}$$

By inspection of Theorem 4.8 an equivalent optimality condition for ($\mathcal{P}_{\mathrm{fix}}$) can be given in terms of the optimization variables $(\tau, q)$ with the "normal map" (due to Robinson [28]) defined for our purposes as

$$F(\tau, q) = Dj(\tau, P_{\mathrm{ad}}(q)) + \begin{pmatrix} 0 \\ c\,\nu_\tau(q - P_{\mathrm{ad}}(q)) \end{pmatrix} \tag{5.7}$$

for an arbitrary constant $c > 0$. We can verify that the zeros of $F$ are precisely the points satisfying the first order necessary conditions.

**Proposition 5.6.** *Suppose that for $(\bar\tau, \bar u) \in \mathbb{R}^+ \times U_{\mathrm{ad}}$ and the corresponding $(\bar\nu, \bar y, \bar p) = (\nu_{\bar\tau}, y_{\bar\tau,\bar u}, p_{\bar\tau,\bar u})$ the first order necessary conditions from Theorem 4.8 are fulfilled. Then there exists a $\bar q \in U$ such that $\bar u = P_{\mathrm{ad}}(\bar q)$ and $F(\bar\tau, \bar q) = 0$.*

*Proof.* Define the Lagrange multiplier $\lambda = -(\alpha\bar u + B^*\bar p)$, which is an element of the subdifferential $\partial\chi_{U_{\mathrm{ad}}}(\bar u)$ (*cf.* Thm. 4.8). We set $\bar q = \bar u + c^{-1}\lambda$. To see directly that $P_{\mathrm{ad}}(\bar q) = P_{\mathrm{ad}}(\bar u + c^{-1}\lambda) = \bar u$, we can use the characterization

$$\partial\chi_{U_{\mathrm{ad}}}(\bar u) = \left\{ \lambda \mid \operatorname{supp}\lambda^+ \subset \{\,\bar u = u_{\max}\,\},\ \operatorname{supp}\lambda^- \subset \{\,\bar u = -u_{\max}\}\right\}.$$

Furthermore, we have $c(\bar q - P_{\mathrm{ad}}(\bar q)) = \lambda$ and therefore the second component of $F(\bar\tau, \bar q)$, which is given by $\nu_\tau(\alpha\bar u + B^*\bar p + \lambda)$, is zero. The first component is given by $j'_\tau(\bar\tau, \bar u) = \int_I H(\bar u, \bar y, \bar p)\,\mathrm{d}t$, which is zero. □

In the following we will suppose $\alpha > 0$ and set $c = \alpha$ in the normal map (5.7). By construction we have then that

$$F(\tau, q) = \begin{pmatrix} \int_I H(u_q, y_{\tau,q}, p_{\tau,q})\,\mathrm{d}t \\ \nu_\tau(\alpha q + B^* p_{\tau,q}) \end{pmatrix}, \tag{5.8}$$

since the term $\nu_\tau\alpha P_{\mathrm{ad}}(q) = \nu_\tau\alpha u_q$ cancels out in the second row of (5.7). We can now see that the condition $F(\bar\tau, \bar q) = 0$ implies the relation

$$\bar q = -\frac{1}{\alpha}B^* p_{\bar\tau,\bar q},$$

which directly gives the optimality condition for the control $\bar u = P_{\mathrm{ad}}(\bar q)$. To apply a Newton type method to $F(\tau, q) = 0$ we require the linearization of $F$. Since (5.6) involves a pointwise projection it is non-smooth. Therefore we work with the semismoothness calculus in Banach spaces as in [35]. We introduce the generalized differential of the projection $P_{\mathrm{ad}}$ as

$$DP_{\mathrm{ad}}(q)(\delta q) = \chi_{\mathcal{I}}\delta q = \begin{cases} \delta q & \text{where } |q| \le u_{\max}, \\ 0 & \text{else,} \end{cases}$$

where $\chi_{\mathcal{I}}$ is the indicator function of the "inactive set" given by

$$\mathcal{I} = \{\,(t,n)|\,|q_n(t)| \le u_{\max,n}\,\}.$$

With the central result on semismoothness of superposition operators on Lebesgue spaces, the pointwise projection $P_{\mathrm{ad}}$ is semismooth as an operator from $L^r(I, \mathbb{R}^{N_{\mathrm{con}}}) \to L^2(I, \mathbb{R}^{N_{\mathrm{con}}})$ for any exponent $r \in (2, \infty]$ (see, *e.g.* [35], Thm. 3.49). With a chain rule for semismooth operators and representation (5.3) we compute a representation for the generalized derivative of $F$ at the point $(\tau, q)$ of the form

$$DF(\tau, q)(\delta\tau, \delta q) = \begin{pmatrix} \int_I ((\alpha q + B^* p)\cdot\chi_{\mathcal{I}}\delta q - \langle \delta y, A'(y)^* p\rangle + \langle Bu - A(y), \delta p\rangle)\,\mathrm{d}t \\ \nu_\tau(\alpha\delta q + B^*\delta p) + \delta\tau(\alpha q + B^* p) \end{pmatrix}, \tag{5.9}$$

where $\delta y$ solves the tangent equation (4.7) with $\delta u = \chi_{\mathcal{I}} \delta q$ and $\delta p$ solves the corresponding second adjoint equation (5.4). To be precise, we obtain the following result.

**Proposition 5.7.** *Define the space $U^r = L^r(I, \mathbb{R}^{N_{\text{con}}})$. For any $r \in (2, \infty]$ the mapping $F\colon \mathbb{R} \times U^r \to \mathbb{R} \times U^r$ as given in (5.8) is semismooth with the generalized derivative $DF$ given in (5.9),* i.e. *we have*

$$\|F(\tau + \delta\tau, q + \delta q) - F(\tau, q) - DF(\tau + \delta\tau, q + \delta q)(\delta\tau, \delta q)\|_{\mathbb{R}\times U^r} \in o\left(\|(\delta\tau, \delta q)\|_{\mathbb{R}\times U^r}\right)$$

*for $(\delta\tau, \delta q) \to 0$ in $\mathbb{R} \times U^r$.*

*Proof.* We decompose $F$ into $F = F_2 \circ F_1$ with the definitions

$$\begin{aligned} F_1 &\colon \mathbb{R} \times U^r \to \mathbb{R} \times U^r \times U & (\tau, q) &\mapsto (\tau, q, P_{\text{ad}}(q)), \\ F_2 &\colon \mathbb{R} \times U^r \times U \to \mathbb{R} \times U^r & (\tau, q, u) &\mapsto \begin{pmatrix} \int_I H(u, y_{\tau,u}, p_{\tau,u})\,\mathrm{d}t \\ \nu_\tau(\alpha q + B^* p_{\tau,u}) \end{pmatrix}. \end{aligned}$$

The mapping $F_1$ is Lipschitz continuous and semismooth with generalized derivative $DF_1 = (1, \mathrm{Id}, DP_{\text{ad}})$ according to the properties of $P_{\text{ad}}$ (see, e.g. [35], Thm. 3.49). Furthermore, the function $F_2$ is (arbitrarily often) Fréchet differentiable in a neighborhood of $P_{\text{ad}}(U) = U_{\text{ad}}$, which can be proved in a straightforward way as in Theorem 4.5. We skip the details but mention that the most important step is to show differentiability of

$$(\tau, u) \in \mathbb{R} \times U \mapsto B^* p_{\tau,u} \in H^1(I, \mathbb{R}^{N_{\text{con}}})$$

and then use the embedding $H^1(I, \mathbb{R}^{N_{\text{con}}}) \hookrightarrow U^r$ for any $r \in [1, \infty]$. As a continuously Fréchet differentiable function $F_2$ is semismooth with respect to the classical derivative (see [35], Prop. 3.4). Therefore, we can apply the semismooth chain rule (see [35], Prop. 3.8) to obtain semismoothness of $F = F_2 \circ F_1$.

To obtain the concrete form of $DF$ as given in (5.9) we have applied some obvious algebraic modifications to the first line of (5.9) and used that $\chi_{\mathcal{I}} u = \chi_{\mathcal{I}} q$. □

For the analysis of the semismooth Newton method we also need the bounded invertibility of the operator $DF(\tau, q)$ in a neighborhood of the optimum. For this we require a second order sufficient condition in the optimum.

**Assumption 5.8.** Assume that for $(\bar\tau, \bar u) \in \mathbb{R}^+ \times U_{\text{ad}}$ there exists a constant $\gamma > 0$, such that

$$\left((\delta\tau, \delta u), D^2 j(\bar\tau, \bar u)(\delta\tau, \delta u)\right)_{\mathbb{R}\times U} \geq \gamma \left(\delta\tau^2 + \|\delta u\|_U^2\right)$$

holds for all $(\delta\tau, \delta u) \in \mathbb{R} \times U$.

**Lemma 5.9.** *Suppose that $(\bar\tau, \bar q) \in \mathbb{R}^+ \times U$ are chosen such that Assumption 5.8 holds for $(\bar\tau, P_{\text{ad}}(\bar q))$. Then there exists a neighborhood $N(\bar\tau, \bar q) \subset \mathbb{R} \times U$ of $(\bar\tau, \bar q)$, such that the generalized derivative*

$$DF(\tau, q)\colon \mathbb{R} \times U^r \to \mathbb{R} \times U^r$$

*is uniformly boundedly invertible for all $(\tau, q) \in N(\bar\tau, \bar q)$.*

*Proof.* Define $\bar u = P_{\text{ad}}(\bar q)$. By continuity of $D^2 j(\tau, u)$ in a neighboorhood of $(\bar\tau, \bar u)$, the coercivity condition from Assumption 5.8 also holds in a neighborhood $\tilde N(\bar\tau, \bar u) \subset \mathbb{R} \times U$, with a possibly smaller constant $\tilde\gamma > 0$. We will show, that the neighborhood for $(\bar\tau, \bar q)$ for the invertibility of $DF$ can now be chosen as $N(\bar\tau, \bar q) = \{\, (\tau, u + (\bar q - \bar u)) | (\tau, u) \in \tilde N(\bar\tau, \bar u)) \,\}$. In the following, we fix some $(\tau, q)$ from this neighborhood. It is clear, that for all such $q$ we have $(\tau, P_{\text{ad}}(q)) \in N(\bar\tau, \bar u)$. We can separate $DF$ into the two parts $DF = DF_1 + DF_2$ according to

$$DF(\tau, q)(\delta\tau, \delta q) = \begin{pmatrix} 0 \\ \alpha\nu_\tau \delta q \end{pmatrix} + \begin{pmatrix} \int_I \left(H'_u(\cdot)\chi_{\mathcal{I}}\delta q + H'_y(\cdot)\delta y + H'_p(\cdot)\delta p\right)\mathrm{d}t \\ \nu_\tau B^* \delta p + \delta\tau(\alpha q + B^* p) \end{pmatrix}$$

with $\delta y$ and $\delta p$ as in (5.9). Recall that $\chi_{\mathcal{I}} = DP_{\mathrm{ad}}(q)$. We note that $DF_2$ depends only on the values of $\delta q$ on the inactive set, *i.e.* we have

$$DF_2(\tau,q)(\delta\tau,\delta q) = DF_2(\tau,q)(\delta\tau,\chi_{\mathcal{I}}\delta q) \quad \text{for all } (\delta\tau,\delta q)\in\mathbb{R}\times U.$$

Furthermore $DF_2$ has a smoothing property, *i.e.* it maps $\mathbb{R}\times U$ continuously to the smaller space $\mathbb{R}\times U^r$. We define the subspace of $U$ induced by the inactive set as $U_{\mathcal{I}} = \{\,\chi_{\mathcal{I}}u | u\in U\,\}$ and introduce the pointwise multiplication operator

$$P_{\mathcal{I}}\colon \mathbb{R}\times U\to\mathbb{R}\times U_{\mathcal{I}}, \qquad (\delta\tau,\delta q)\mapsto(\delta\tau,\chi_{\mathcal{I}}\delta q),$$

which is the canonical orthogonal projection to the linear subspace $\mathbb{R}\times U_{\mathcal{I}}$. By comparing the expressions for $D^2 j$ and for $DF$, it is easy to check that

$$P_{\mathcal{I}}\circ DF(\tau,q) = P_{\mathcal{I}}\circ D^2 j(\tau,P_{\mathrm{ad}}(q))\circ P_{\mathcal{I}} \tag{5.10}$$

holds in the sense of equality of operators on $\mathbb{R}\times U$. Having introduced these notations, we turn to the proof. Consider the equation

$$DF(\tau,q)(\delta\tau,\delta q) = f = (f_\tau, f_q) \tag{5.11}$$

for some given right-hand side $f\in\mathbb{R}\times U^r$. To show that it has a solution $(\delta\tau,\delta q)\in\mathbb{R}\times U^r$, we set $\delta u = \chi_{\mathcal{I}}\delta q\in U_{\mathcal{I}}$ and look first for the solution of

$$P_{\mathcal{I}}\circ DF(\tau,q)(\delta\tau,\delta u) = P_{\mathcal{I}} f.$$

With (5.10) and Assumption 5.8 the operator $P_{\mathcal{I}}\circ DF(\tau,q)$ is symmetric and positive definite on $\mathbb{R}\times U_{\mathcal{I}}$ and we obtain a unique solution $(\delta\tau,\delta u)\in\mathbb{R}\times U_{\mathcal{I}}$ with

$$\|(\delta\tau,\chi_{\mathcal{I}}\delta q)\|_{\mathbb{R}\times U} = \|(\delta\tau,\delta u)\|_{\mathbb{R}\times U}\le C\|P_{\mathcal{I}} f\|_{\mathbb{R}\times U}\le C\|f\|_{\mathbb{R}\times U}.$$

It is clear that the constant $C$ can be chosen independently of $(\tau,q)\in N(\bar\tau,\bar q)$ due to uniform ellipticity and boundedness of $D^2 j(\tau,P_{\mathrm{ad}}(q))$. To obtain a full solution, we use the splitting of $DF$ from above and get

$$DF(\tau,q)(\delta\tau,\delta q) = DF_1(\tau,q)(\delta\tau,\delta q) + DF_2(\tau,q)(\delta\tau,\chi_{\mathcal{I}}\delta q).$$

Rearranging (5.11) and using $\chi_{\mathcal{I}}\delta q = \delta u$ yields

$$(0,\alpha\nu_\tau\delta q) = DF_1(\tau,q)(\delta\tau,\delta q) = f - DF_2(\tau,q)(\delta\tau,\delta u), \tag{5.12}$$

which implies that the second component of the full solution has to fulfill the pointwise equation $\alpha\nu_\tau\delta q = f_q - \nu_\tau B^*\delta p - \delta\tau(\alpha q + B^* p)$. This can be solved for $\delta q$ by using $\nu_\tau\in\mathcal{N}_{\mathrm{ad}}$. By using the smoothing property of $DF_2(\tau,q)$ we obtain

$$\|\delta q\|_{U^r}\le C\|f - DF_2(\tau,q)(\delta\tau,\delta u)\|_{\mathbb{R}\times U^r}\le C\left(\|f\|_{\mathbb{R}\times U^r} + \|(\delta\tau,\delta u)\|_{\mathbb{R}\times U}\right),$$

with a generic constant $C$ independent of $(\delta\tau,\delta u)$ and $(\tau,q)\in N(\bar\tau,\bar q)$. Combining this with the previous estimate for $(\delta\tau,\delta u)$ we conclude the proof. □

**Theorem 5.10.** *Suppose that $F(\bar\tau,\bar q) = 0$ and that Assumption 5.8 holds at $(\bar\tau,P_{\mathrm{ad}}(\bar q))$. The semismooth Newton method based on $(\delta\tau,\delta q)$ computed from*

$$DF(\tau,q)(\delta\tau,\delta q) = -F(\tau,q) \tag{5.13}$$

*with the update rule $(\tau^{\mathrm{new}},q^{\mathrm{new}}) = (\tau,q) + (\delta\tau,\delta q)$ converges locally superlinearly towards $(\bar\tau,\bar q)$ in the space $\mathbb{R}\times U^r$.*

*Proof.* We combine Proposition 5.7 and Lemma 5.9 (*cf.*, *e.g.* [35], Thm. 3.13). □

## 6. Practical realization

In the following the necessary background is established that is needed for an efficient numerical solution of the time optimal control problem. The discretization concept and a proper globalization of the semismooth Newton method are outlined.

### 6.1. Discretization

For numerical realization we choose a consistent discretization of the objective, the constraints and the derivatives in the sense that First-Discretize-Then-Optimize methods (FDTO) and First-Optimize-Then-Discretize (FOTD) commute, and that we get the exact discrete derivatives. We achieve this by using a standard FE-Galerkin method in space and a Petrov−Galerkin method in time; *cf.* [3]. Since the space discretization is straightforward, we focus on the time discretization.

The time grid is denoted by $0 = t_0 < \ldots < t_M = T$ with stepsizes $k_m = t_m - t_{m-1}$ and subintervals $I_m = (t_m, t_{m-1}]$. To apply the cG(1) Crank-Nicolson's scheme for both state equations, the trial space for $v$ and $w$ is chosen to consist of continuous piecewise linear functions, *i.e.*

$$v(t)|_{I_m} = v_{m-1} + \frac{t - t_{m-1}}{k_m}(v_m - v_{m-1}),$$

while the test space is set to piecewise constant functions $\psi(t,x)|_{I_m} = \psi_m(x)$. For the adjoint states $p_1$ and $p_2$ the trial and test space are interchanged, *i.e.* $p_1(t,x)|_{I_m} = p_{1,m}(x)$, and due to (4.12) the natural control discretization is given by piecewise constant functions, *i.e.* $u(t)|_{I_m} = u_m \in \mathbb{R}^{N_{\mathrm{con}}}$. With constant $\nu_\tau \equiv 1 + \tau$ we obtain the semidiscrete Lagrange function as

$$\begin{aligned}\mathcal{L}_k(\tau,u,y,p) = \frac{\mu}{2}(v_M, v_M) + \sum_{m=1}^{M} \Bigg\{ & k_m\nu_\tau \left(\kappa + \frac{\alpha}{2}|u_m|^2\right) - (p_{1,m}, v_m - v_{m-1}) \\ & - k_m\nu_\tau \Bigg[ (\nabla p_{1,m}, \sigma \nabla v_{m-1/2}) + \frac{1}{2}\big(p_{1,m}, I_{\mathrm{ion}}(v_m, w_{m-1}) + I_{\mathrm{ion}}(v_{m-1}, w_{m-1})\big) - \sum_{n=1}^{N_{\mathrm{con}}} (p_{1,m}, \chi_{\Omega_{\mathrm{con},n}}) u_{n,m} \Bigg] \\ & - (p_{2,m}, w_m - w_{m-1}) - k_m\nu_\tau \left(p_{2,m}, G(v_{m-1/2}, w_{m-1/2})\right) \Bigg\},\end{aligned}$$

where $v_{m-1/2} = (v_m + v_{m-1})/2$ and $w_{m-1/2} = (w_m + w_{m-1})/2$. Here we additionally modified the discretization of the nonlinearity in the term $I_{\mathrm{ion}}(v_m, w_{m-1})$ to achieve a decoupling of the ODE variable $w$ in the state equation. The time-stepping schemes for the state and adjoint equations can be derived from $\mathcal{L}_k$. Since both will always be realized for $\nu_0 \equiv 1$ according to Section 6.3, they coincide with those derived in [20] for fixed final time. The second derivatives are obtained analogously from the semidiscrete Lagrange function. Compared to those in [20], additional terms arise on the right-hand side of the tangent and the second adjoint equation, as well as in the Hessian evaluation, due to the differentiation w.r.t. the parameter $\tau$.

To get a more convenient representation for the second derivatives, we note that we can replace the left-hand side of the differential equations analogously to Remark 5.5 also on the discrete level. Here we replace the corresponding expressions occurring in the discrete Hamiltonian by $(1/\nu)\partial_t y|_{I_m} = (y_m - y_{m-1})/(\nu k_m)$.

### 6.2. Implementation and globalization of the semismooth Newton method

In the following we describe the optimization algorithm TR-SN (Trust Region Semismooth Newton), which is displayed in Appendix A.3. Therein, the Newton step is computed in a similar way as in Lemma 5.9. First, we compute a step $(\delta\tau, \delta q_1)$ that solves (5.13) up to equivalence on the inactive set. This is done in a matrix-free fashion using the method of conjugate gradients for the system matrix $DF(\tau, q)$ in combination with the inner product $(\cdot,\cdot)_{\mathcal{I}} = (\cdot, P_{\mathcal{I}}\,\cdot)$ induced by the projection to the inactive set as defined above. The use of the CG

method is justifed due to the relation to the Hessian (5.10). Note, that we do not set $\delta q_1$ to zero on the inactive set. In the case of convergence, which is determined according to the energy error (see [11], p. 171f), we derive the full step $(\delta\tau, \delta q)$ from expression (5.12).

To achieve more robust convergence of the semismooth Newton method, it is embedded into a trust region framework inspired by Steihaug-CG (see [30]). As a consequence the CG method is augmented by a radius constraint (on the inactive sets) and two additional termination criteria along the lines of ([30], Sect. 2). The update of the radius is performed as described in Appendix A.3. A more detailed description of this adaptation of Steihaug-CG to a nonsmooth setting can be found in ([26], Sect. 3.5). Note, that a termination of the CG method in the first step (*i.e.*, $(\delta\tau, \delta q_1) = -\theta F(\tau, q)$ for $\theta > 0$) corresponds to a variant of the projected gradient method.

### 6.3. Efficient use of the time transformation

In the practical realization we employ a modification of the time transformation, which allows us to extend existing optimal control codes to the time optimal case more conveniently. In each step, instead of updating $\tau^{\text{new}} = \tau + \delta\tau$, we apply the time transformation to the discretization of the time interval. After computation of an update $(\delta\tau, \delta q)$ from the Newton equation (5.13), it is applied as

$$\delta q^{\text{new}} = q + \delta q, \quad k_m^{\text{new}} = (1 + \delta\tau)k_m, \quad t_m^{\text{new}} = t_{m-1}^{\text{new}} + k_m^{\text{new}}, \quad \tau^{\text{new}} = 0. \tag{6.1}$$

We can verify that this yields an equivalent algorithm (in terms of the iterates $t_{\text{f}}$ and $q$ and the corresponding functional values). Consequently, the state and adjoint solve always work with $\tau = 0$, $\nu \equiv 1$, *i.e.* the corresponding code does not need to be changed with respect to an optimal control problem with fixed $t_{\text{f}}$. The resulting Newton system, adjoint and tangent equations are displayed in Appendix A.2.

## 7. Numerical results

In the following we apply the monolithic TR-SN method to two examples, the stabilization of an excitation wave and of a reentry wave. We investigate the convergence properties and show the successful stabilization of the system. Furthermore, we compare the performance of the proposed optimization method to a non-monolithic method.

The computations are carried out with Lagrange Q1 elements on a quadrilateral grid in space using the finite element library `deal.II` [2]. Concerning the choice of the parameters in the monodomain equations we followed [10] and chose:

| $\eta_0$ | $\eta_1$ | $\eta_2$ | $\eta_3$ | $v_{\text{th}}$ | $v_{\text{pk}}$ | $\sigma$ |
|---|---|---|---|---|---|---|
| 1.5 | 4.4 | 0.012 | 1.0 | 13 | 100 | $\text{diag}\,(3 \times 10^{-3}, 3.1525 \times 10^{-4})$ |

### 7.1. Example 1: Excitation wave

In the first example, an "excitation wave" has to be stabilized *via* problem $(\mathcal{P}^{\mu})$. One control function is applied on two disjoint electrode plates $\Omega_{\text{con},1}$. The initial data of the excitation wave is generated by solving, with an equidistant step size of 0.01, the uncontrolled monodomain equation with the discontinuous initial data $(v_0, w_0) = (101\chi_{\Omega_{\text{exc}}}(x), 0)$ on the time interval $(0, 0.17)$. Here $\chi_{\Omega_{\text{exc}}}(x)$ stands for the characteristic function of the initially excited zone $\Omega_{\text{exc}} = [0.18, 0.22] \times [0.18, 0.24]$. The terminal value $(v(0.17, \cdot), w(0.17, \cdot))$ serves as initial data for the optimal control experiment. Without control, the wave would spread through the domain forming an ellipse, leading to a very large value of the objective functional. The parameter values are given by:

| $\alpha$ | $\kappa$ | $\mu$ | $u_{\max}$ | $\Omega_{\text{exc}}$ | $\Omega_{\text{con},1}$ |
|---|---|---|---|---|---|
| $10^{-5}$ | 1 | 100 | 1000 | $[0.18, 0.22] \times [0.18, 0.24]$ | $[0.0, 0.1] \times [0.1, 0.3] \cup [0.3, 0.4] \times [0.1, 0.3]$ |

The domain $\Omega = (0, 0.4)^2$ is discretized in $64 \times 64$ quadrilateral cells (six refinements). The temporal grid is spaced equidistantly with $N_t = 151$ points. The optimization is initialized with $t_{\text{f}} = t_{\text{f}0} = 6$ (ms) and $q_0 = -u_{\max}\,\chi_{[0,4]}(t)$.

TABLE 1. Monolithic TR-SN run with $u_{\max} = 1000$.

| $n$ | $j_n$ | $F_n$ | #CG | flag | $t_{\mathrm{f}n}$ | $\lvert\mathcal{I}\rvert$ | $s_n$ |
|---|---|---|---|---|---|---|---|
| 0 | 27.0215 | $1.0 \times 10^{0}$ | 2 | 2 | 6.000 | 49 | |
| 1 | 24.0119 | $2.7 \times 10^{-1}$ | 2 | 3 | 5.436 | 150 | |
| 2 | 19.6478 | $3.1 \times 10^{-1}$ | 2 | 3 | 4.377 | 150 | |
| R 3 | 19.6478 | $3.1 \times 10^{-1}$ | 2 | 3 | 4.377 | 150 | |
| 4 | 18.7142 | $2.0 \times 10^{-1}$ | 4 | 3 | 4.037 | 150 | |
| 5 | 18.6004 | $5.2 \times 10^{-2}$ | 4 | 3 | 4.046 | 136 | |
| 6 | 18.4606 | $3.7 \times 10^{-2}$ | 4 | 3 | 4.084 | 129 | |
| 7 | 18.1565 | $2.8 \times 10^{-2}$ | 4 | 3 | 4.110 | 127 | |
| 8 | 17.8473 | $4.7 \times 10^{-1}$ | 6 | 3 | 4.143 | 127 | |
| 9 | 16.7880 | $2.9 \times 10^{-1}$ | 7 | 3 | 4.337 | 128 | |
| 10 | 15.6300 | $7.2 \times 10^{-1}$ | 14 | 0 | 4.682 | 127 | 1.09 |
| 11 | 14.8599 | $7.3 \times 10^{-1}$ | 9 | 0 | 5.208 | 125 | 0.67 |
| 12 | 14.4844 | $3.5 \times 10^{-1}$ | 16 | 0 | 5.277 | 125 | 0.49 |
| 13 | 14.3833 | $6.2 \times 10^{-2}$ | 9 | 0 | 5.328 | 124 | 0.27 |
| 14 | 14.3723 | $5.9 \times 10^{-2}$ | 7 | 0 | 5.319 | 124 | 0.11 |
| 15 | 14.3694 | $5.6 \times 10^{-3}$ | 9 | 3 | 5.317 | 124 | 0.26 |
| 16 | 14.3688 | $1.6 \times 10^{-2}$ | 7 | 0 | 5.309 | 123 | 0.22 |
| 17 | 14.3686 | $3.4 \times 10^{-4}$ | 10 | 0 | 5.309 | 123 | 0.28 |
| 18 | 14.3686 | $1.4 \times 10^{-4}$ | 12 | 0 | 5.308 | 123 | 0.08 |
| 19 | 14.3686 | $9.2 \times 10^{-7}$ | | 0 | 5.308 | 123 | 0.01 |

Numerically, we employ the weighted norm $\|(\delta\tau, \delta q_1)\|^2 = \zeta\,\delta\tau^2 + \|\delta q_1\|_{\mathcal{I}}^2$ on $\mathbb{R} \times U$ and the corresponding inner product. Thereby, too aggressive or too conservative changes of the terminal time in the initial iterates can be avoided *via* a proper choice of $\zeta > 0$. Note that this modification does only affect the steps where the CG method is terminated early. We set $\zeta = 10^4$ and use $\|F(\tau_n, q_n)\| \leq 10^{-5}\|F(\tau_0, q_0)\|$ as relative stopping criterion. The iteration terminates after 19 steps. Table 1 depicts the history of the objective $j_n$, the Newton residual $F_n$, which are given by

$$j_n = j(\tau_n, P_{\mathrm{ad}}(q_n)), \quad \text{and } F_n = \|F(\tau_n, q_n)\|,$$

the number of CG iterates in the Steihaug-CG method together with its exit flag (0 fully converged, 2 negative curvature, 3 large step), the current guess for the terminal time $t_{\mathrm{f}n}$, the number of inactive control points $|\mathcal{I}|$ (out of 150), and an indicator for superlinear convergence of the objective $s_n = (j_n - j_{n-1})/(j_{n-1} - j_{n-2})$. We observe fast decrease in the objective at the beginning of the iterations, together with a reduction of the terminal time. From iteration 10 on, the CG iteration is fully resolved (flag 0). As soon as the inactive set is converged, the first order optimality sharply decreases and we observe superlinear convergence $s_n \to 0$ at the end of the iterations.

Let us turn to the question whether the excitation wave is stabilized by applying the optimal control. The optimal state fulfills $-0.40 \leq \bar{v}(t_{\mathrm{f}}, x) \leq 0.51$, $-0.023 \leq \bar{w}(t_{\mathrm{f}}, \cdot)$. Hence, the system is stabilized according to Theorem 2.11 using

$$v_{\max} = 0.51 < 13, \quad w_{\min} = -0.023 > -0.326, \quad v_{\min} = -2.3.$$

We also ran tests with the initialization for $q$ changed to $q_0 = 0$, *i.e.* we start with an unhindered evolution of the excitation wave. In this case the TR-SN stops after six steps with a different stationary point given by $\bar{t}_{\mathrm{f}} = 0.00025$ and $\bar{j} = 693$, which does not allow defibrillation. But increasing $\zeta$ to $10^7$ recovers $\bar{t}_{\mathrm{f}} = 5.308$ also from $q_0 = 0$, in 52 iterations.

*7.1.1. Comparison to a bilevel method*

For comparison we also solved the same problems with the bilevel method from [20], stopping at the same accuracy of three digits in the optimal terminal time $\bar{t}_{\mathrm{f}}$. The bisection-based bilevel method is started on the

TABLE 2. Total number of state, gradient, and Hessian evaluations.

| | Number of evaluations | | |
|---|---|---|---|
| Method | State | Gradient | Hessian |
| Monolithic | 20 | 19 | 157 |
| Bilevel | 212 | 197 | 1685 |

interval $[4, 6]$. It determines $\bar{t}_f = 5.31$ after 14 evaluations of the lower level problem. The total number of evaluations of state, gradient and Hessian are given in Table 2.

For this example, the monolithic method saves approximately 90 percent of state, gradient, and Hessian evaluations. The large number of required evaluations in the bilevel method is related to the fact that the lower level problem has to fully converge to exclude bad decisions for the upper level. Therefore it invests more Hessian evaluations far away from the optimal terminal time. In contrast, the monolithic method concentrates most of the Hessian evaluations in a neighborhood of the optimal terminal time. The increase in the average run time of an Hessian evaluation in the monolithic method due to additional assembling operations was observed to be insignificant.

### 7.2. Example 2: Reentry wave

In the second example a reentry wave has to be stabilized successfully in minimal time and with minimal energy input. This wave was generated as in ([20], Ex. 1) for $\Omega = (0, 2) \times (0, 0.8)$. We choose the parameters:

| $\alpha$ | $\kappa$ | $\mu$ | $u_{\max}$ | $\Omega_{\text{con},1}$ |
|---|---|---|---|---|
| $10^{-3}$ | 1 | 1000 | 40 | $[0.0, 0.25] \times [0.3, 0.55] \cup [1.75, 2.0] \times [0.3, 0.55]$ |

The discretization is based on $M = 1600$ time steps and $128 \times 64$ cells in space. We initialize the optimizer with $q_0 = -u_{\max}\chi_{[0,70]}(t)$ and $t_{f0} = 100$, and choose again the weighted norm with $\zeta = 10^3$. The TR-SN converges in 68 steps reducing the first order optimality condition $F_n$ to a relative accuracy of $10^{-6}$, see Table 3. Again we observe superlinear convergence in the last steps.

TABLE 3. Monolithic TR-SN run with $\alpha = 10^{-3}$.

| $n$ | $j_n$ | $F_n$ | #CG | $t_{fn}$ | $\lvert\mathcal{I}\rvert$ | $s_n$ |
|---|---|---|---|---|---|---|
| 0 | 156.05 | $4.9 \times 10^{0}$ | 1 | 100.000 | 479 | |
| 10 | 98.9124 | $3.3 \times 10^{-1}$ | 3 | 67.541 | 1600 | |
| 20 | 74.3351 | $1.8 \times 10^{0}$ | 4 | 67.670 | 1600 | |
| 30 | 69.1416 | $2.9 \times 10^{0}$ | 5 | 67.166 | 1600 | |
| 40 | 68.7508 | $2.2 \times 10^{0}$ | 5 | 66.920 | 1600 | |
| 50 | 68.3986 | $5.6 \times 10^{-1}$ | 5 | 66.720 | 1583 | |
| R 60 | 68.2537 | $8.4 \times 10^{-1}$ | 5 | 66.492 | 1579 | |
| 61 | 68.2383 | $7.5 \times 10^{-2}$ | 6 | 66.500 | 1578 | |
| 62 | 68.2329 | $4.3 \times 10^{-2}$ | 10 | 66.525 | 1577 | |
| R 63 | 68.2329 | $4.3 \times 10^{-2}$ | 5 | 66.525 | 1577 | |
| 64 | 68.2312 | $7.9 \times 10^{-2}$ | 6 | 66.521 | 1577 | |
| 65 | 68.2294 | $7.1 \times 10^{-2}$ | 7 | 66.512 | 1576 | 1.13 |
| 66 | 68.2289 | $6.8 \times 10^{-3}$ | 11 | 66.515 | 1576 | 0.28 |
| 67 | 68.2287 | $4.0 \times 10^{-3}$ | 7 | 66.510 | 1576 | 0.26 |
| 68 | 68.2287 | $5.0 \times 10^{-6}$ | | 66.510 | 1576 | 0.00 |

The method delivers the optimal terminal time $\bar{t}_f = 66.51$ and the time optimal control $\bar{u}$ depicted as solid line in Figure 2. The time optimal control exhibits a multi-phasic structure, compared to the monophasic initial

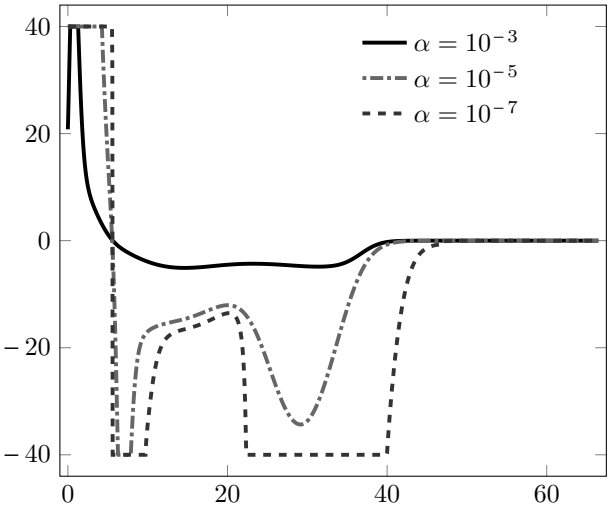



FIGURE 2. Time optimal controls for different $\alpha$.

TABLE 4. Optimization results for different $\alpha$.

| $\alpha$ | $\bar{t_f}$ | $\|\bar{u}\|$ | $\|\bar{v}(x,\bar{t_f})\|_{L^2}$ |
|---|---|---|---|
| $10^{-3}$ | 66.51 | 56 | 0.016 |
| $10^{-5}$ | 65.53 | 151 | 0.016 |
| $10^{-7}$ | 65.48 | 221 | 0.016 |

control. Its energy input to the tissue is very low at $\|\bar{u}\| = 56$, and thus significantly lower compared to the initial control $\|u_0\| = 335$. Again, we confirm the successful stabilization *via* Theorem 2.11 by examination of the optimal state. It fulfills $-0.01 \leq \bar{v}(t_f, \cdot) \leq 0.20$, $0 \leq \bar{w}(t_f, \cdot)$.

The plot in Figure 2 additionally shows the time optimal control for different values of $\alpha$. With decreasing $\alpha$ the controls exhibit a different switching structure, showing additional arcs on the lower bound. Table 4 shows the corresponding parts of the objective. We note a rather large increase in the norms of the optimal controls, when compared to the relatively small reduction of the optimal terminal times $\bar{t_f}$.

Figure 3 shows snapshots of the optimal transmembrane voltage $\bar{v}(t_m, x)$ at different times $t_m$ and comparisons to the uncontrolled transmembrane voltage $v(t_m, x)$ at the same times. The latter depicts the evolution of the reentry wave without applying an extracellular stimulus, *i.e.* $u \equiv 0$. While the uncontrolled reentry wave persists, the time optimal control facilitates a fast propagation of the wave front into the control region by its positive values; see Figure 3a for $t = 1.33$. Afterwards, the negative values of $\bar{u}(t)$ hinder the wave to leave the control region to the north. Thereby, the wave is deflected to the south and falls apart.

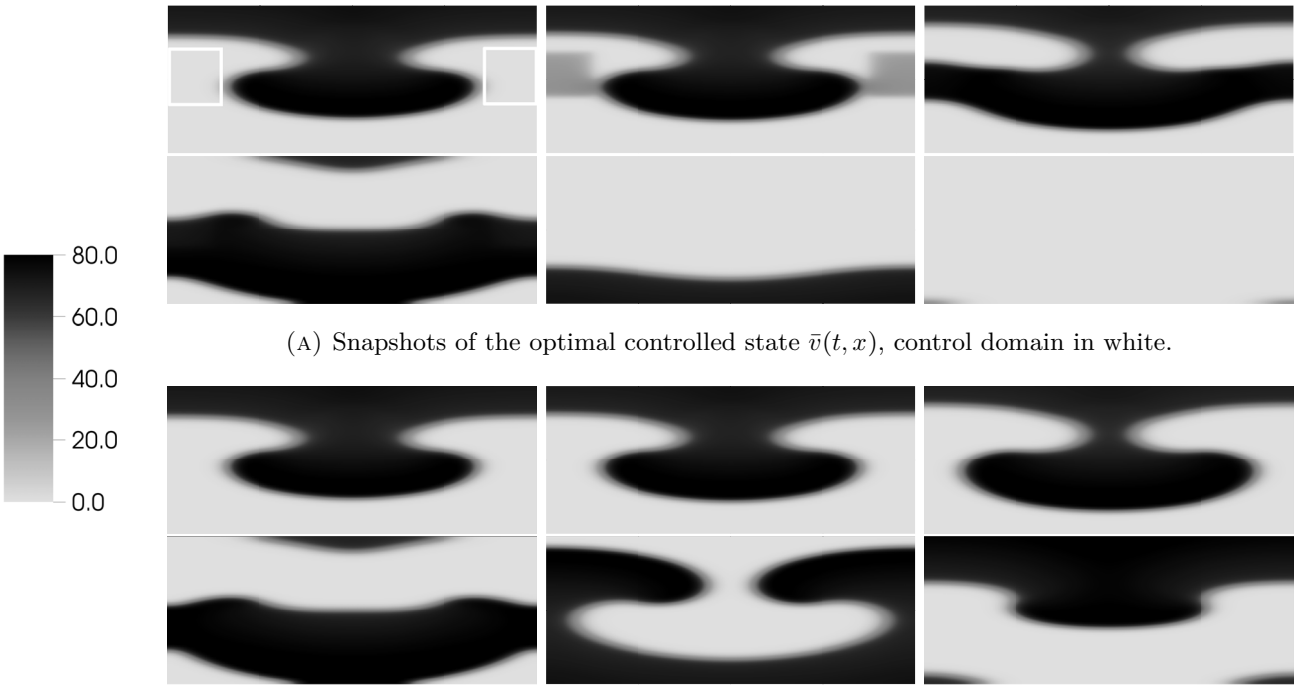
## A. APPENDIX

### A.1. *A priori* analysis of the state equation

We prove the uniqueness result from Theorem 2.3.

**Proposition A.1.** *The solution to* (1.1) *in the sense of Definition* 2.1 *is unique.*

*Proof.* Let $(v_1, w_1)$, $(v_2, w_2)$ be two solutions of (1.1) with the same initial conditions $v_0$, $w_0$ and the same right-hand side $I_e$. We set $\delta v = v_1 - v_2$ and $\delta w = w_1 - w_2$. We subtract the equations for $v_1$ and $v_2$ from each




(A) Snapshots of the optimal controlled state $\bar{v}(t,x)$, control domain in white.

(B) Corresponding snapshots for the uncontrolled reentry wave $v(t,x)$ for $u \equiv 0$.

FIGURE 3. Snapshots of the controlled and uncontrolled state at $t = 0$, 1.33, 6 and $t = 16$, 48, 65 (upper row and lower row, respectively).

other and obtain

$$\partial_t \delta v - \nabla \cdot \sigma \nabla \delta v + R(v_1) - R(v_2) + \eta_1(v_1 w_1 - v_2 w_2) = 0$$

with the cubic nonlinearity $R$ as in the proof of Theorem 2.7. Clearly it holds $v_1 w_1 - v_2 w_2 = \delta v\, w_1 + v_2\, \delta w$. We test this equation with $\chi_{(0,t)}\delta v$ for some $t > 0$ to obtain

$$\int_0^t \left[(\partial_t \delta v, \delta v) + (\sigma \nabla \delta v, \nabla \delta v) + (R(v_1) - R(v_2), \delta v)\right] \mathrm{d}s = -\eta_1 \int_0^t \left[(\delta v\, w_1, \delta v) + (v_2\, \delta w, \delta v)\right] \mathrm{d}s.$$

To estimate the term containing the cubic nonlinearity on the left-hand side we use the estimate from below given by

$$(R(v_1) - R(v_2), \delta v) = \int_0^1 (R'(\theta v_1 + (1-\theta)v_2)\, \delta v, \delta v)\, \mathrm{d}\theta \geq -c_0 \|\delta v\|_{L^2}^2,$$

which is a consequence of $R'(\cdot) \geq -c_0$ (*cf.* the proof of Thm. 2.7). The two terms on the right-hand side are treated with Hölder's inequality in space, which results in

$$\left| \int_0^t (\delta v\, w_1, \delta v)\, \mathrm{d}s \right| \leq \int_0^t \|\delta v^2\|_{L^2} \|w_1\|_{L^2}\, \mathrm{d}s \leq C \int_0^t \|\delta v\|_{L^4}^2\, \mathrm{d}s,$$
$$\left| \int_0^t (v_2\, \delta w, \delta v)\, \mathrm{d}s \right| \leq \int_0^t \|v_2\|_{L^4} \|\delta w\|_{L^2} \|\delta v\|_{L^4}\, \mathrm{d}s \leq \int_0^t \left(\|v_2\|_{L^4}^2 \|\delta w\|_{L^2}^2 + \|\delta v\|_{L^4}^2\right) \mathrm{d}s.$$

for a constant $C$ that is independent of $\delta v$, using that $w_1$ is bounded in $L^\infty(I, L^2)$ by Proposition 2.4. Now, for any $\varphi \in H^1$, we can estimate the $L^4$ norm of $\varphi$ by

$$\|\varphi\|_{L^4} \le \|\varphi\|_{L^2}^{1/4}\|\varphi\|_{L^6}^{3/4} \le c_d^{3/4}\,\|\varphi\|_{L^2}^{1/4}\|\varphi\|_{H^1}^{3/4} \le \frac{c_d^3}{4\,\varepsilon^3}\|\varphi\|_{L^2} + \frac{3\,\varepsilon}{4}\|\varphi\|_{H^1}$$

with $\varepsilon > 0$ arbitrary, which is a consequence of Hölder's inequality, the Sobolev embedding $H^1 \hookrightarrow L^6$ in up to three space dimensions (with constant $c_d$), and Young's inequality. We apply this to $\|\delta v\|_{L^4}^2$ (for each $s \in (0,t)$), which leads to the estimate

$$\left|\int_0^t (\delta v\, w_1, \delta v)\,\mathrm{d}s\right| + \left|\int_0^t (v_2\,\delta w, \delta v)\,\mathrm{d}s\right| \le C\int_0^t \left(\varepsilon^2\|\delta v\|_{H^1}^2 + \varepsilon^{-6}\|\delta v\|_{L^2}^2 + \|v_2\|_{L^4}^2\|\delta w\|_{L^2}^2\right)\mathrm{d}s,$$

for arbitrary $\varepsilon > 0$, where $C$ does not depend on $\delta v, \delta w$, or $\varepsilon$. Combining the estimates, using the integration by parts formula in time and the positive definiteness of the elliptic form, we obtain

$$\|\delta v(t)\|_{L^2}^2 + \int_0^t \gamma\|\delta v\|_{H^1}^2\,\mathrm{d}s \le C\int_0^t \left[(1+\varepsilon^{-6})\|\delta v\|_{L^2}^2 + \varepsilon^2\|\delta v\|_{H^1}^2 + \|v_2\|_{L^4}^2\|\delta w\|_{L^2}^2\right]\mathrm{d}s$$

for $\gamma > 0$ and $C$ independent of $\delta v$ and $\delta w$. By a sufficiently small choice of $\varepsilon$, the $H^1$ norm on the right-hand side can be absorbed into the left-hand side, which leads to

$$\|\delta v(t)\|_{L^2}^2 \le C\int_0^t \left(\|\delta v\|_{L^2}^2 + \|v_2\|_{L^4}^2\|\delta w\|_{L^2}^2\right)\mathrm{d}s.$$

By a quick computation we can also obtain a corresponding inequality for $\delta w$ as

$$\|\delta w(t)\|_{L^2}^2 \le C\int_0^t \left(\|\delta v\|_{L^2}^2 + \|\delta w\|_{L^2}^2\right)\mathrm{d}s.$$

Adding both, we obtain

$$\|\delta v(t)\|_{L^2}^2 + \|\delta w(t)\|_{L^2}^2 \le C\int_0^t \left(1 + \|v_2\|_{L^4}^2\right)\left(\|\delta v(t)\|_{L^2}^2 + \|\delta w(t)\|_{L^2}^2\right)\mathrm{d}s,$$

for every $t \in I$. Note that $s \mapsto (1 + \|v_2(s)\|_{L^4}^2)$ is integrable, since $v_2 \in L^4(Q)$. By Gronwall's inequality, it now follows that $\delta v = \delta w = 0$. □

## A.2. Newton residual, adjoint and tangent equations

Here, we give the full formulas for the Newton residual $F$, its derivative $DF$ and the auxiliary equations. Using the time transformation iteratively as explained in Section 6.3 together with the simplification from Remark 5.5, the Newton system $DF(0,q)(\delta\tau, \delta q) = -F(0,q)$ can be expressed as:

$$F(0,q) = \begin{pmatrix} \int_0^T \left(\kappa + (\alpha/2)\ |P_{\mathrm{ad}}(q)|^2 + \langle \partial_t v, p_1\rangle + \langle \partial_t w, p_2\rangle\right)\mathrm{d}t \\ \left(\alpha q_n + \int_{\Omega_{\mathrm{con},n}} p_1\,\mathrm{d}x\right)_{n=1\dots N_{\mathrm{con}}} \end{pmatrix},$$

$$DF(0,q)\begin{pmatrix}\delta\tau\\ \delta q\end{pmatrix} = \begin{pmatrix} \int_0^T \Big(\sum_{n=1}^{N_{\mathrm{con}}}(\alpha q_n + \int_{\Omega_{\mathrm{con},n}} p_1)\chi_{\mathcal{I}_n}\delta q_n - \langle \delta v, \partial_t p_1\rangle \cdots \\ \qquad \cdots - \langle \delta w, \partial_t p_2\rangle + \langle \partial_t v, \delta p_1\rangle + \langle \partial_t w, \delta p_2\rangle\Big)\mathrm{d}t \\ \left(\alpha\delta q_n + \int_{\Omega_{\mathrm{con},n}} \delta p_1 + \delta\tau(\alpha q_n + \int_{\Omega_{\mathrm{con},n}} p_1)\right)_{n=1\dots N_{\mathrm{con}}} \end{pmatrix}.$$

Furthermore, we give the auxiliary equations in their full, non-abstract formulation (which incorporate the natural boundary conditions). The adjoint equations at current state $(v, w)$ are given by

$$\begin{aligned} -\partial_t p_1 - \nabla \cdot \sigma \nabla p_1 + I'_{\mathrm{ion},v}(v,w)\, p_1 + G'_v\, p_2 &= 0, \\ -\partial_t p_2 + I'_{\mathrm{ion},w}(v)\, p_1 + G'_w\, p_2 &= 0, \\ p_1(T) = \mu\, v(T), \quad p_2(T) &= 0. \end{aligned}$$

The tangent equations at current state $(v, w)$ are given by

$$\begin{aligned} \partial_t \delta v - \nabla \cdot \sigma \nabla \delta v + I'_{\mathrm{ion},v}(v,w)\, \delta v + I'_{\mathrm{ion},w}(v)\, \delta w &= \textstyle\sum_{n=1}^{N_{\mathrm{con}}} \chi_{\mathcal{I}_n} \chi_{\Omega_{\mathrm{con},n}} \delta q_n + \delta\tau\, \partial_t v, \\ \partial_t \delta w + G'_v\, \delta v + G'_w\, \delta w &= \delta\tau\, \partial_t w, \\ \delta v(0) = 0, \quad \delta w(0) &= 0. \end{aligned}$$

The second adjoint equations at current state $(v, w)$ and adjoint state $(p_1, p_2)$ are given by

$$\begin{aligned} -\partial_t \delta p_1 - \nabla \cdot \sigma \nabla \delta p_1 + I'_{\mathrm{ion},v}(v,w)\, \delta p_1 + G'_v\, \delta p_2 &= -I''_{\mathrm{ion},vv}(v)\, p_1\, \delta v - I''_{\mathrm{ion},vw}\, p_1\, \delta w - \delta\tau\, \partial_t p_1, \\ -\partial_t \delta p_2 + I'_{\mathrm{ion},w}(v)\, \delta p_1 + G'_w\, \delta p_2 &= -I''_{\mathrm{ion},wv}\, p_1\, \delta v - \delta\tau\, \partial_t p_2, \\ \delta p_1(T) = \mu\, \delta v(T), \quad \delta p_2(T) &= 0. \end{aligned}$$

### A.3. Optimization algorithm TR-SN

(1) Initialize $q^0$, maximal radius $\Delta_{\max} > 0$, initial radius $0 < \Delta_0 \le \Delta_{\max}$ and set $k = 0$.
(2) Compute $F(0, q^k)$ from (5.8) (state and adjoint solve) and determine inactive sets.
(3) Compute $(\delta\tau, \delta q_1)$ from (5.13) by Steihaug-CG using the inner product on the inactive set $(\cdot,\cdot)_{\mathcal{I}}$.
(4) If CG is converged to a sufficiently small tolerance, compute $\delta q$ from (5.12). Otherwise, set $\delta q = \delta q_1$.
(5) Calculate $\varrho^{\mathrm{act}} = j(0, P_{\mathrm{ad}}(q^k)) - j(\delta\tau, P_{\mathrm{ad}}(q^k + \delta q))$ and decide:
- If $(\varrho^{\mathrm{act}} < -\varepsilon)$ then reject step:
  - Set $q^{k+1} = q^k$ and $\Delta_{k+1} = 0.2\Delta_k$.
- Else accept step:
  - Set $q^{k+1} = q^k + \delta q$.
  - Set $\varphi_k(\delta\tau, \delta q) = \big((\delta\tau, \delta q), F(0, q^k)\big)_{\mathcal{I}} + (1/2)\, \big((\delta\tau, \delta q), H(0, q^k)(\delta\tau, \delta q)\big)_{\mathcal{I}}$.
  - Set $\varrho_k = -\varrho^{\mathrm{act}}/\varphi_k(\delta q)$ and update the radius:

$$\Delta_{k+1} = \begin{cases} \min(2\|\delta q\|_{\mathcal{I}}, \Delta_{\max}), & \text{if } \varrho_k \in [0.7, 1.3] \quad \text{(model good)} \\ 0.5\|\delta q\|_{\mathcal{I}}, & \text{if } \varrho_k \notin [0.25, 1.75] \quad \text{(model bad)} \\ \Delta_k, & \text{else.} \end{cases}$$

  - Apply the time transformation as in (6.1).
(6) If stopping criteria are not fulfilled, set $k = k + 1$ and go to (2).

## References


[1] H. Amann, Linear and Quasilinear Parabolic Problems: Volume I: Abstract Linear Theory. Birkhäuser, Basel (1995).
[2] W. Bangerth, R. Hartmann and G. Kanschat, deal.II – a general purpose object oriented finite element library. *ACM Trans. Math. Softw.* **33** (2007) 24/1–24/27.
[3] R. Becker, D. Meidner and B. Vexler, Efficient numerical solution of parabolic optimization problems by finite element methods. *Optim. Methods Softw.* **22** (2007) 813–833.
[4] A. Borzì and R. Griesse, Distributed optimal control of lambda-omega systems. *J. Numer. Math.* **14** (2006) 17–40.
[5] Y. Bourgault, Y. Coudière and C. Pierre, Existence and uniqueness of the solution for the bidomain model used in cardiac electrophysiology. *Nonlin. Anal. Real World Appl.* **10** (2009) 458–482.

[6] A.J. Brandão, E. Fernández-Cara, P.M. Magalhães and M.A. Rojas-Medar, Theoretical analysis and control results for the Fitzhugh-Nagumo equation. *Electron. J. Differ. Eq.* **2008** (2008) 1–20.
[7] T. Breiten and K. Kunisch, Compensator design for the monodomain equations with the Fitzhugh-Nagumo model. To appear in *ESAIM: COCV* (2016). Doi:10.1051/cocv/2015047
[8] E. Casas, C. Ryll and F. Tröltzsch, Sparse optimal control of the Schlögl and FitzHugh−Nagumo systems. *Comput. Meth. Appl. Math.* **13** (2013) 415–442.
[9] M. Chipot, Elements of Nonlinear Analysis, *Adv. Texts Series*. Springer (2000).
[10] P. Colli Franzone, P. Deuflhard, B. Erdmann, J. Lang and L. Pavarino, Adaptivity in space and time for reaction-diffusion systems in electrocardiology. *SIAM J. Sci. Comput.* **28** (2006) 942–962.
[11] P. Deuflhard and M. Weiser, Numerische Mathematik 3: Adaptive Lösung partieller Differentialgleichungen, De Gruyter Studium. De Gruyter (2011).
[12] J. Dieudonné, Foundations of Modern Analysis. Academic Press (1969).
[13] L.C. Evans, Partial Differential Equations. American Mathematical Society (2010).
[14] H.O. Fattorini, infinite Dimensional Linear Control Systems: The Time Optimal and Norm Optimal Problems. *North-Holland Math. Stud.* Elsevier Science, Amsterdam (2005).
[15] J.A. Griepentrog, H.-C. Kaiser and J. Rehberg, Heat kernel and resolvent properties for second order elliptic differential operators with general boundary conditions on $L^p$. *Adv. Math. Sci. Appl.* **11** (2001) 87–112.
[16] J.A. Griepentrog and L. Recke, Linear elliptic boundary value problems with non-smooth data: Normal solvability on Sobolev-Campanato spaces. *Math. Nachr.* **225** (2001) 39–74.
[17] H. Hermes and J.P. Lasalle, Functional Analysis and Time Optimal Control. *Math. Sci. Eng.* Academic Press, New York (1969).
[18] K. Ito and K. Kunisch, Lagrange Multiplier Approach to Variational Problems and Applications. *SIAM* (2008).
[19] K. Ito and K. Kunisch, Semismooth Newton methods for time-optimal control for a class of ODEs. *SIAM J. Control Optim.* **48** (2010) 3997–4013.
[20] K. Kunisch and A. Rund, Time optimal control of the monodomain model in cardiac electrophysiology. *IMA J. Appl. Math.* (2015).
[21] K. Kunisch and D. Wachsmuth, On time optimal control of the wave equation and its numerical realization as parametric optimization problem. *SIAM J. Control Optim.* **51** (2013) 1232–1262.
[22] J.-L. Lions and E. Magenes, Non-homogeneous Boundary Value Problems and Applications. Vol. I. Springer (1972).
[23] J.D. Murray, Mathematical Biology I. An Introduction. In vol. 17 of *Interdisciplinary Applied Mathematics*. 3rd edition. Springer, New York (2002).
[24] C. Nagaiah, K. Kunisch and G. Plank, Optimal control approach to termination of re-entry waves in cardiac electrophysiology. *J. Math. Biol.* **67** (2013) 359–388.
[25] A. Pazy, Semigroups of Linear Operators and Applications to Partial Differential Equations. In vol. 44 of *Appl. Math. Sci.* Springer-Verlag, New York (1983).
[26] K. Pieper, *Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems*. Ph.D. dissertation, Technische Universität München (2015).
[27] J.-P. Raymond and H. Zidani, Pontryagin's principle for time-optimal problems. *J. Optim. Theory Appl.* **101** (1999) 375–402.
[28] S.M. Robinson, Normal maps induced by linear transformations. *Math. Oper. Res.* **17** (1992) 691–714.
[29] F. Schlögl, Chemical reaction models for non-equilibrium phase transitions. *Z. Phys. A* **253** (1972) 147–161.
[30] T. Steihaug, The conjugate gradient method and trust regions in large scale optimization. *SIAM J. Numer. Anal.* **20** (1983) 626–637.
[31] J. Sundnes, G.T. Lines, X. Cai, B.F. Nielsen, K.-A. Mardal and A. Tveito, Computing the Electrical Activity in the Heart. Springer, Berlin, Heidelberg (2006).
[32] H. Triebel, Interpolation Theory, Function Spaces, Differential Operators. In vol. 18 of *North-Holland Math. Library*. North-Holland Publ., Amsterdam (1978).
[33] F. Tröltzsch, Optimal Control of Partial Differential Equations. In vol. 112 of *Grad. Stud. Math. AMS*, Providence, Rhode Island (2010).
[34] L. Tung, *A Bi-domain Model for Describing Ischemic Myocardial D-c Potentials*. Ph.D. thesis, Massachusetts Institute of Technology (1978).
[35] M. Ulbrich, Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces. *MOS-SIAM Series on Optimization. SIAM* (2011).
[36] W.P. Ziemer, Weakly Differentiable Functions. Sobolev Spaces and Functions of Bounded Variation. Springer (1989).