IMA Journal of Applied Mathematics (2015) **80**, 1664–1683
doi:10.1093/imamat/hxv010
Advance Access publication on 23 May 2015

Time optimal control of the monodomain model in cardiac electrophysiology

Karl Kunisch and Armin Rund*
Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstr. 36, 8010 Graz, Austria
*Corresponding author: armin.rund@uni-graz.at

[Received on 3 April 2014; revised on 4 March 2015; accepted on 17 April 2015]

An optimal control approach to a simplified reaction–diffusion system describing cardiac defibrillation is proposed that allows for joint optimization of shape and duration of defibrillation pulses. Within the framework, optimized multi-phasic pulses with low energy, short duration and/or low amplitude can be designed according to specific needs. The approach is based on a novel time optimal control formulation for the monodomain model, which takes into consideration the dynamical system properties of the uncontrolled equation. The highly complex dynamics requires a consistent discretization of first- and second-order information to guarantee effective optimization schemes leading to successful defibrillation. Numerical examples underline the efficiency of the proposed method.

Keywords: time optimal control; reaction–diffusion equation; monodomain model; PDE-constrained optimization; trust region semismooth Newton method.

1. Introduction and problem formulation

Over the last decades significant progress was made in the numerical treatment of open loop optimal control problems governed by distributed parameter systems. The techniques that were developed were adapted for a wide range of important equations, including wave and diffusion equations, the equations of fluid mechanics and fluid–structure interaction models. In contrast, very little attention was paid to reaction–diffusion systems, whose dynamical systems behaviour is significantly different from those of the systems mentioned before. In this paper, we continue our efforts on one particular reaction–diffusion system, which describes the electrical activity of the heart. Compared with our earlier work we propose a new choice of cost functionals, which allows a much wider class of optimized trajectories. This is only possible by using well-conceived numerical optimal control techniques. Due to the rich dynamical systems behaviour for the problems under consideration, ad hoc techniques will simply fail, especially for second-order methods.

Let us briefly describe the physiological background for the problem to be investigated. The heart supplies all organs with blood by rhythmic contractions that are triggered electrically. Disturbances in the formation and/or propagation of electrical impulses may induce reentrant activation patterns which lead to a noticeable increase in the hearts activation rate. Such fast rhythms may lead to fibrillation. To restore a healthy rhythm, the delivery of electrical shocks, referred to as defibrillation, is a reliable therapy. It can be administered by means of implantable cardioverter defibrillators (ICDs), which monitor the heart rate and deliver a discharge, which acts as a control, to restore a normal rhythm.

The bidomain model is a well-accepted continuous and macroscopic description of the electrical activity of cardiac muscle cells. The model consists of two coupled reaction–diffusion equations together with an ODE describing the ionic currents associated with the reaction terms; see e.g. Keener & Sneyd (2009). Assuming the intracellular and extracellular conductivity tensors to be linearly dependent, the model can be simplified to the monodomain model, which results in a substantial reduction of computational effort (Potse *et al.*, 2006; Sundnes *et al.*, 2006). Once a model for the physiological phenomena and their dependence on a control input is fixed, an optimal control approach can be utilized to decide on the optimal shock delivery.

Due to severe physiological constraints, involving time scales, geometry and multi-physics aspects, the current optimal control techniques certainly fall short of addressing all relevant aspects. But the medical technology itself is still changing rapidly, so that certain assumptions, as for instance, the availability of observations or actuator support which is not too small relative to the overall tissue size, may become reality. Current technological advances include, for instance, the development of a new type of ICDs; see e.g. Puri *et al.* (2013). They consist of flexible arrays of leads which act as sensors, gathering information on the electrical state of the heart, and as actuator-electrodes, delivering a defibrillation shock when arrhythmias are detected. In case of a defibrillation therapy, each lead is provided with a defibrillation pulse that has to be designed appropriately, based on the measured data.

Within the optimal control approach to cardiac defibrillation, pulses are designed by solving an optimal control problem constrained by a reaction–diffusion system. The aims of effective defibrillation and minimal detrimental side effects to the patient are modelled within the control objective. By adapting the objective and its parameters, a wide range of goals can be achieved, which makes the optimal control approach a powerful and flexible tool for defibrillation pulse design. The design of the objective is of paramount importance and, together with an efficient numerical realization, are the main innovation of this paper. For the choice of the control objective, several conflicting interests need to be taken into consideration. They include the behaviour of the unforced dynamical system, which for the simplified ODE-FitzHugh Nagumo model states that once the state is sufficiently excited, it must necessarily reach a plateau value before it can return to the stable equilibrium; see e.g. Murray (2002, p. 241f). For the infinite-dimensional system (1) this behaviour can occur at different times at any point in the spatial domain. This suggests to use a control objective (defibrillation) which only involves the terminal time of the control horizon. This leads to a highly ill-conditioned optimal control problem, making exact computation of gradient and Hessian information indispensable.

The topic of numerical simulation of the electrical activity of the heart has inspired much research, so that we can only quote selected references (Franzone *et al.*, 2006; Vigmond *et al.*, 2008). The optimal control approach to cardiac defibrillation was previously investigated in Nagaiah *et al.* (2011) and Götschel *et al.* (2013) for the monodomain model, and in Nagaiah *et al.* (2013) for the bidomain model. Differently from the present paper, these papers consider the case where the shock length is fixed. Moreover, the cost functional for the optimal control formulation involves a reference trajectory. As a consequence, the number of phases of the optimal pulse is determined a priori. The optimal control of reaction–diffusion systems involving wave phenomena was also the focus of the research in Borzì & Griesse (2006) and Casas *et al.* (2013).

The article is organized as follows: the monodomain model is described in Section 2. Section 3 is devoted to the formulation of the optimal control problem. The necessary conditions are obtained in Section 4. In Section 5, the optimization method is presented, which is based on a bilevel formulation together with a trust region semismooth Newton method. Section 6 introduces the numerical framework which is chosen in such a manner that discretizations before or after deriving the necessary optimality conditions commute, and lead to a Galerkin discretizations with the exact discrete derivatives; see Section 6. The proposed techniques are tested by numerical experiments on termination of reentry waves in Section 7. One of the examples also addresses robustness of the computed controls.

2. The controlled state equation

We investigate a sample of heart tissue described by the domain Ω. The electrophysiology is modelled by the monodomain equation using the cell model of Rogers & McCulloch (1994), which is a modified FitzHugh–Nagumo model. For simplicity, we do not consider a conductive bath and therefore model the heart tissue to stay electrically isolated, leading to homogeneous Neumann boundary conditions. Thus, the dynamical system is given by the monodomain model

$$v_t + I(v,w) - \nabla\cdot(\bar{\sigma}_i \nabla v) = I_{\mathrm{e}} \quad \text{a. e. in } Q := (0,t_{\mathrm{f}})\times\Omega, \tag{1a}$$

$$w_t + G(v,w) = 0 \quad \text{a. e. in } Q, \tag{1b}$$

$$\nu\cdot\bar{\sigma}_i\nabla v = 0 \quad \text{on } \Sigma := (0,t_{\mathrm{f}})\times\partial\Omega, \tag{1c}$$

$$v(x,0) = v_0(x), \quad w(x,0) = w_0(x) \quad \text{a. e. in } \Omega. \tag{1d}$$

The independent variables are $x\in\Omega\subset\mathbb{R}^d$, $d=2$, and time $t\in(0,t_{\mathrm{f}})$ with the terminal time $t_{\mathrm{f}}>0$; Ω is a bounded domain with Lipschitz continuous boundary $\partial\Omega$ and unit outer normal ν. The functions $v(t,x)$, $w(t,x)$ denote the transmembrane electric potential and the gating or recovery variable. The intercellular conductivity tensor $\bar{\sigma}_i\in L^\infty(\Omega,\mathbb{R}^{d\times d})$ is assumed to be symmetric and uniformly elliptic. The extracellular stimulation current I_{e} depends on the defibrillation pulse, which has to be controlled. The ionic current $I(v,w)$ and $G(v,w)$ are given as

$$I(v,w) = \eta_0 v\left(1-\frac{v}{v_{\mathrm{th}}}\right)\left(1-\frac{v}{v_{\mathrm{pk}}}\right) + \eta_1 vw, \tag{2a}$$

$$G(v,w) = \eta_2\left(\eta_3 w - \frac{v}{v_{\mathrm{pk}}}\right), \tag{2b}$$

with $\eta_0,\eta_1,\eta_2,\eta_3\in\mathbb{R}^+$. A cell is excited if the transmembrane potential exceeds the threshold potential $v_{\mathrm{th}}>0$. Further, $v_{\mathrm{pk}}>v_{\mathrm{th}}$ is the peak potential. The initial conditions $v_0(x)\in L^2(\Omega)$, $w_0(x)\in L^4(\Omega)$ describe a fibrillatory situation.

The geometric setting represents a layer of heart muscle tissue modelled as a 2D domain Ω. A finite number of electrode plates $\Omega_{\mathrm{con},k}$, $k=1,\ldots,N_{\mathrm{e}}$, are pasted on top. In the common setting, this would just be a pair; alternatively, it can be an array of plates in case of a flexible sensor array. For simplicity, the electrodes are assumed to be homogeneous and perfectly connected to the tissue. For the monodomain model, each electrode is assigned an independent defibrillation pulse $u_k(t)$, whereas a compatibility condition would be needed for bidomain modelling. The extracellular stimulation current I_{e} is therefore modelled as

$$I_{\mathrm{e}}(t,x) = \sum_{k=1}^{N_{\mathrm{e}}} u_k(t)\chi_{\Omega_{\mathrm{con},k}}(x), \tag{3}$$

where $\chi_{\Omega_{\mathrm{con},k}}(x)$ denotes the characteristic function of electrode plate k, $u_k(t)$ the corresponding pulse and $u(t) = (u_1(t),\ldots,u_{N_{\mathrm{e}}}(t))$ the control vector.

3. The optimal control problem

Here, defibrillation will be posed as an optimal control problem. The aim consists in influencing the extracellular stimulation current $I_{\mathrm{e}}(t,x)$ in such a way that the tissue changes to a state where fibrillatory propagation is hindered. Additionally, side effects on the tissue should be kept small. While this description is clear, its particular modelling is involved. We first discuss the choice of the time horizons and then define the optimal control problem.

3.1 *Modelling the time horizon*

After a defibrillation shock has been applied successfully, the heart muscle tissue needs a certain amount of time to reach a non-fibrillatory state, especially in the presence of complicated patterns of reentry waves. Therefore, a successful defibrillation can only be confirmed at a time t_{f} with $t_{\mathrm{f}} \gg T$, where T is the end time of the defibrillation shock. There are several ways how one might incorporate this fact into the optimal control problem.

Nagaiah *et al.* (2013) propose a formulation with a short fixed time horizon $[0,T]$ and enforce the defibrillation on the basis of a tracking functional using a desired trajectory given by an a priori known defibrillation pulse, which brings the tissue to a non-excited state at $t_{\mathrm{f}} \gg T$. Post-optimally, the simulation on (T,t_{f}) is continued to confirm successful defibrillation.

Here, we propose a formulation which is different in several ways. First, we do not rely on a desired trajectory; secondly, the shock duration itself is optimized. Thirdly, defibrillation is quantified in the cost by demanding that at some final time of simulation t_{f} the electric potential is small throughout Ω. Thus, the optimization problem is posed on some fixed horizon $[0,t_{\mathrm{f}}]$ at the end of which defibrillation must be achieved. The defibrillation pulse is applied on the first part $[0,T]$, with T being part of the optimization. Compared with Nagaiah *et al.* (2013), this gives an increased flexibility in how the defibrillation is achieved. In particular the number of phase changes is part of the optimization and is not aligned with some desired trajectory. In addition, for successful defibrillation the system is monitored throughout the time interval $[0,t_{\mathrm{f}}]$, rather than only on $[0,T]$.

From the point of view of numerical optimization, this problem is significantly more challenging, since the elimination of the use of a desired trajectory leads to a drastically reduced coercivity of the optimal control formulation. Our approach will lead to different optimal controls that deliver less energy to the tissue, since the optimal control formulation is more flexible in choosing the pulses.

3.2 *Defibrillation as optimal control problem*

For effective defibrillation at time t_{f}, we aim at bringing as much tissue to the resting state as possible. Then, the next natural activation given by the sinoatrial node or by a pacemaker should be able to reestablish the normal heart rhythm. How to model this terminal condition? The goal is realized by a terminal penalty term.

To model negative side effects of the applied shock I_{e}, three different quantities are considered: the duration, the energy and the amplitude of the pulse. Since the exposure of the patient is related to the duration of the electrical shock, we aim at minimizing the duration T. Moreover, the energy of the pulses $\|u\|_2^2$ has to be minimized. Additionally, we restrict the amplitudes by imposing inequality constraints $u_{\min} \leqslant u_k(t) \leqslant u_{\max}$, since too large amplitudes would result in a local damage to the tissue adjacent to the electrodes.

These considerations suggest the following optimal control problem:

$$\min_{0\leqslant T\leqslant t_{\mathrm{f}},\, u(t)\in U_{\mathrm{ad}}} J(v,u,T) := T + \frac{\mu}{2}\|v(\cdot,t_{\mathrm{f}})\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\sum_{k=1}^{N_{\mathrm{e}}}\|u_k\|^2_{L^2(0,T)}, \tag{4a}$$

$$\text{subject to (1) with } I_{\mathrm{e}} = \sum_{k=1}^{N_{\mathrm{e}}} u_k(t)\chi_{\Omega_{\mathrm{con},k}}(x)\chi_{(0,T)}(t), \tag{4b}$$

with weighting parameters $\mu > 0, \alpha > 0$. The amplitude of the controls are bounded via the set of admissible controls:

$$U_{\mathrm{ad}} := \{u \in U : u_{\min}(t) \leqslant u_k(t) \leqslant u_{\max}(t) \text{ for a. a. } t \in (0,t_{\mathrm{f}}), k = 1,\ldots,N_{\mathrm{e}}\}, \tag{4c}$$

where $u_{\min},\ u_{\max} \in L^\infty(0,t_{\mathrm{f}})$ and $U := L^2(0,t_{\mathrm{f}};\mathbb{R}^N_{\mathrm{e}})$. Equation (4) constitutes a time optimal control problem with a non-linear ODE–PDE system as constraints. The objective (4a) is a scalarized multi-objective formulation favouring successful defibrillation for large μ, small energy inputs for large α and short pulses for small α and μ.

3.3 *Existence*

At first, we recall the existence and regularity results for the solutions of the monodomain model, which are defined next. We introduce $Q = (0,t_{\mathrm{f}}) \times \Omega$ and the Sobolev space $V := H^1(\Omega)$ with its dual V^*. The duality pairing between V and V^* is denoted by $\langle\cdot,\cdot\rangle_{V^*,V}$.

Definition 1 For $I_{\mathrm{e}} \in L^2(0,t_{\mathrm{f}},V^*)$ and $(v_0,w_0) \in L^2(\Omega) \times L^2(\Omega)$, a pair (v,w) is called a weak solution to (1) if $(v,w) \in L^2(0,t_{\mathrm{f}};V) \cap C([0,t_{\mathrm{f}}];L^2(\Omega)) \cap L^4(Q) \times C^1([0,t_{\mathrm{f}}];L^2(\Omega))$, $v_t \in L^2(0,t_{\mathrm{f}};V^*) + L^{4/3}(Q)$, and for a.a. $t \in (0,t_{\mathrm{f}})$ and all $\varphi \in V$,

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\int_\Omega v(t)\varphi\,\mathrm{d}x + \int_\Omega \bar{\sigma}_i\nabla v(t)\nabla\varphi\,\mathrm{d}x + \int_\Omega I(v(t),w(t))\varphi\,\mathrm{d}x = \langle I_{\mathrm{e}}(t),\varphi\rangle_{V^*,V}, \\ w_t(t) + G(v(t),w(t)) = 0 \quad \text{a.e. in } \Omega, \end{cases}$$

where the time derivative is to be understood in the distributional sense.

Existence and uniqueness results for the bidomain equations are considered in e.g. Bourgault *et al.* (2009) and Nagaiah *et al.* (2011). Since we restrict ourselves to the monodomain equation here and since we use a simple form for G, only minor modifications in the proof of these results imply the following proposition, which holds in dimensions 2 and 3; see also Kunisch & Wagner (2011) for the monodomain equation.

Proposition 3.1 Let $I_{\mathrm{e}} \in L^2(0,t_{\mathrm{f}},V^*)$ and $(v_0,w_0) \in L^2(\Omega) \times L^2(\Omega)$. Then System (1) admits a weak solution. Furthermore, there exists a constant C, such that

$$\begin{aligned} &\|v\|^2_{C([0,t_{\mathrm{f}}];L^2)} + \|v\|^2_{L^2(0,t_{\mathrm{f}};V)} + \|v\|^4_{L^4(Q)} + \|v_t\|^{4/3}_{L^{4/3}(Q)+L^2(0,t_{\mathrm{f}};V^*)} + \|w\|^2_{C^1([0,t_{\mathrm{f}}];L^2)} \\ &\quad \leqslant C(1 + \|v_0\|^2_{L^2(\Omega)} + \|w_0\|^2_{L^2(\Omega)} + \|I_{\mathrm{e}}\|^2_{L^2(V^*)}). \end{aligned}$$

If additionally $I_{\mathrm{e}} \in L^\infty(0,t_{\mathrm{f}};V^*)$ and $w_0 \in L^4(\Omega)$ holds, then the weak solution is unique.

This proposition applies in particular to the choice of I_{e} made in (4b). In the following, we prove the existence of a global minimizer of the time optimal control problem (4).

PROPOSITION 3.2 Problem (4) admits a solution (v^*, w^*, u^*, T^*).

Proof. Let $\{(u^n, T^n)\}_{n=1}^{\infty}$ denote a minimizing sequence. This sequence is bounded and hence there exists a subsequence, denoted by the same symbols, and (u^*, T^*) such that $(u^n, T^n) \rightharpoonup (u^*, T^*)$ weakly in $L^2(0, t_{\mathrm{f}}; \mathbb{R}^N_{\mathrm{e}}) \times \mathbb{R}$ with $u^* \in U_{\mathrm{ad}}$.

Let $(v^n, w^n) = (v(u^n), w(u^n))$ denote the associated states of the monodomain model. By Proposition 3.1 there exists a subsequence of (v^n, w^n) denoted by the same indices, and $(\bar{v}, \bar{w}) \in L^2(0, t_{\mathrm{f}}; V) \cap L^4(Q) \times W^{1,2}(0, t_{\mathrm{f}}; L^2(\Omega))$ with $\bar{v}_t \in L^{4/3}(Q) + L^2(0, t_{\mathrm{f}}; V^*)$ such that $v^n \rightharpoonup \bar{v}$ weakly in $L^2(0, t_{\mathrm{f}}; V) \cap L^4(Q)$, $(v^n)_t \rightharpoonup \bar{v}_t$ weakly in $L^{4/3}(Q) + L^2(0, t_{\mathrm{f}}; V^*)$ and $w^n \rightharpoonup \bar{w}$ weakly in $W^{1,2}(Q)$. This puts us in a position to use the arguments in (Bourgault *et al.*, 2009, Section 5.2.3) to argue that we can pass to the limit in the state equations so that $(\bar{v}, \bar{w}) = (v(u^*), w(u^*))$ satisfy (1). The argument in Bourgault *et al.* (2009) is carried out for the bidomain equations and equally applies for the monodomain equation. Since $\{v^n\}$ is bounded in $L^2(0, t_{\mathrm{f}}; V)$ and $\{v^n_t\}$ is bounded in $L^{4/3}(V^*)$, it follows that, possibly on a further subsequence, $v^n(t_{\mathrm{f}}) \to v^*(t_{\mathrm{f}})$ strongly in V^*; see e.g. Constantin & Foias (1988, p. 71). Since $\{v^n(t_{\mathrm{f}})\}$ is bounded in $L^2(\Omega)$, we also have that $v^n(t_{\mathrm{f}}) \rightharpoonup v^*(t_{\mathrm{f}})$ weakly in $L^2(\Omega)$. Now we can pass to the limes inferior in

$$\inf_{0\leqslant T\leqslant t_{\mathrm{f}}, u\in U_{ad}} J(v,u,T) = \underline{\lim}_{n\to\infty} \left(T^n + \frac{\mu}{2}\|v(\cdot\,; u^n, t_{\mathrm{f}})\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\sum_{k=1}^{N_{\mathrm{e}}} \|u^n_k\|^2_{L^2(0,T^n)} \right)$$

$$\geqslant T^* + \frac{\mu}{2}\|v(\cdot\,; u^*, t_{\mathrm{f}})\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\sum_{k=1}^{N_{\mathrm{e}}} \underline{\lim}_{n\to\infty} \|u^n_k\|^2_{L^2(0,T^n)}.$$

To treat the last term, we define

$$\tilde{u}^n_k = \begin{cases} u^n_k & \text{on } (0, T^n) \\ 0 & \text{on } (T^n, t_{\mathrm{f}}) \end{cases}, \quad \tilde{u}^*_k = \begin{cases} u^*_k & \text{on } (0, T^*) \\ 0 & \text{on } (T^*, t_{\mathrm{f}}). \end{cases}$$

It is simple to verify that $\tilde{u}^n_k \rightharpoonup u^*_k$ weakly in $L^2(0, t_{\mathrm{f}})$. Therefore,

$$\underline{\lim}_{n\to\infty} \int_0^{T_n} |u^n_k|^2 = \underline{\lim}_{n\to\infty} \int_0^{t_{\mathrm{f}}} |\tilde{u}^n_k|^2 \geqslant \int_0^{t_{\mathrm{f}}} |\tilde{u}^*_k|^2 = \int_0^{T^*} |u^*_k|^2,$$

consequently

$$\inf_{0\leqslant T\leqslant t_{\mathrm{f}}, u\in U_{ad}} J(v,u,T) \geqslant J(v^*, u^*, T^*),$$

and thus, (u^*, T^*) is a solution to (4). □

4. Necessary conditions

The numerical realization of (4) is based on first-order necessary optimality conditions that an optimal solution $(\bar{u}, \bar{v}, \bar{w}, \bar{T})$ has to fulfil. Applying a formal Lagrangian approach with $p(t,x)$ and $q(t,x)$ as the Lagrange multipliers associated to the parabolic PDE and the ODE, one can proceed in a by now standard manner to obtain the first-order necessary system; see e.g. Tröltzsch (2010), for problems with fixed time horizon and Ito & Kunisch (2010) for time optimal control problems. The first-order necessary system consists of the state equations (1), the adjoint equations (5), the optimality conditions (6) and a transversality condition (7) for the optimal free time $\bar{T}$.

$$
\begin{aligned}
& -p_t - \nabla\cdot(\bar{\sigma}_i \nabla p) + I_v(\bar{v},\bar{w})p + G_v q = 0 \quad \text{in } Q, && \text{(5a)}\\
& -q_t + I_w(\bar{v},\bar{w})\cdot p + G_w \cdot q = 0 \quad \text{in } Q, && \text{(5b)}\\
& \nu\cdot\bar{\sigma}_i\nabla p = 0 \quad \text{on } \Sigma, && \text{(5c)}\\
& p(t_\mathrm{f}) = \mu\bar{v}(t_\mathrm{f}), \quad q(t_\mathrm{f}) = 0 \quad \text{in } \Omega. && \text{(5d)}
\end{aligned}
$$

$$
(\alpha\bar{u}(t) + B^* p(t))\cdot(u(t)-\bar{u}(t)) \geqslant 0 \quad \text{a.a. } t\in(0,\bar{T}) \quad \forall u \in U_{\mathrm{ad}}. \tag{6}
$$

$$
\begin{aligned}
0 = {} & \frac{1}{\bar{T}}\int_0^{\bar{T}} \left(1 + \frac{\alpha}{2}\|\bar{u}\|_2^2 + \langle I_\mathrm{e}(\bar{u}) + \nabla\cdot(\bar{\sigma}_i\nabla v) - I, p\rangle - \langle G, q\rangle\right)\mathrm{d}t \\
& - \frac{1}{t_\mathrm{f}-\bar{T}}\int_{\bar{T}}^{t_\mathrm{f}} (\langle\nabla\cdot(\bar{\sigma}_i\nabla v) - I, p\rangle - \langle G, q\rangle)\,\mathrm{d}t.
\end{aligned} \tag{7}
$$

Here I_v, I_w, G_v, G_w denote the partial derivatives of the model functions (2) and $B^* : L^2(Q) \to U$, $B^* p := (\chi_{(0,\bar{T})}(t)\int_{\Omega_{\mathrm{con},k}} p(t,x)\mathrm{d}x)_{k=1,\ldots,N_\mathrm{e}}$. For the derivation of the transversality condition by a time transformation, we refer the reader to Kunisch *et al.* (2014).

To apply the semismooth Newton method later on, we first reformulate (6) using the projection operator $P_{\mathrm{ad}} : L^2(0,t_\mathrm{f};\mathbb{R}^N_\mathrm{e}) \to L^2(0,t_\mathrm{f};\mathbb{R}^N_\mathrm{e})$, $P_{\mathrm{ad}}(y) = \min(u_{\max}, \max(u_{\min}, y))$ resulting in

$$
\bar{u}(t) = P_{\mathrm{ad}}\left(-\frac{1}{\alpha}B^* p\right) \quad \text{a.a. } t\in(0,\bar{T}). \tag{8}
$$

Secondly, we introduce artificial optimization variables

$$
z \in U, \quad z := (z_k) = -\frac{1}{\alpha}B^* p
$$

and parametrize the controls as $u = P_{\mathrm{ad}}(z)$. Hence, we shift the non-smooth projection operator to the state equation and the objective. Thus, the first-order necessary conditions are equivalent to (1), (5), (7) with eliminated control $u = P_{\mathrm{ad}}(z)$ and

$$
0 = F(z) := \alpha z + B^* p \quad \text{a.a. } t\in(0,\bar{T}). \tag{9}
$$

5. Methods

Time optimal control problems are challenging numerically. To partially appreciate this fact, we note that by means of a time transformation, time optimal problems can be transformed to a fixed time interval, at the expense of an additional non-linearity in the dynamical system. We want to avoid such a new non-linearity since already (1) is known to be rich in structure, allowing wave-like and reentry phenomena, for example.

Therefore, we propose a bilevel approach for solving (4), separating T and the controls u by treating T as parameter in the lower-level problem (LLP):

$$\min_{0 \leqslant T \leqslant t_f} \left(\min_{\substack{u \in U_{\text{ad}} \\ \text{s.t. } (4b)}} J(v, u; T) \right). \tag{10}$$

Obviously, this problem has the same solution as the time optimal control problem (4). For each fixed T the LLP constitutes a terminal tracking problem for a coupled ODE–PDE system with controls acting on a fixed part of the time interval. An alternative all-at-once approach was developed in Kunisch *et al.* (2014).

The bilevel problem will be solved by an iterative method, where the LLP is solved by a semismooth Newton method (TR-SN). It consists of a combination of the reduced Newton method of Hinze & Kunisch (2001) with a globalization based on Steihaug-CG (Steihaug, 1983). The extension to semismooth Newton methods to allow for the control constraints $u(t) \in U_{\text{ad}}$ will be explained in the next section. The method is matrix-free, i.e. the Hessians are not set up explicitly, but we compute only the action of the Hessians and resort to Krylov methods. All forward and backward systems are solved efficiently with time-stepping methods; see Section 6. A globally convergent (derivative-free) direct search method is used for the upper-level minimization problem avoiding the transversality condition (7), which is checked a posteriori.

Before we describe the TR-SN, we note that with the technique of the proof of Proposition 3.2 it is simple to argue the following result.

LEMMA 5.1 The LLP has an optimal solution for every $T > 0$.

The optimality conditions for the LLP consist of (1), (5) and (8) with a fixed current guess for $\bar{T}$, and follow from the results in Kunisch & Wagner (2011).

5.1 *Trust region semismooth Newton method for solving the LLP*

In the following, we describe the solution of the LLP using a trust region semismooth Newton method introduced in Pieper (2015); see also Kunisch *et al.* (2014). Globally convergent semismooth Newton methods can be found, e.g. in Ulbrich (2011), where a step of a first-order method is applied in case the Newton step has to be rejected. In contrast, here a trust region approach inspired by Steihaug (1983) is used, which tries to provide a more gradual transition between a first-order and a second-order method. The method is known to be globally convergent in the unconstrained case; see Steihaug (1983). In the constrained case the update of the trust region radius is currently done heuristically, which performs well in numerical experiments. A proof of global convergence (under possible modifications) is not yet available. Though global convergence for a closely related first-order method (with a fixed step-size) was proved in Pieper (2015, Theorem 3.27).

First, we describe the matrix-free semismooth Newton method. Therefore, we treat all state and adjoint variables as functions of z (as solutions of (1) and (5)), and we define the reduced objective w.r.t. u as $j(u) = J(v(u), u; T)$. Consequently, the reduced optimality condition is $0 = F(z)$ with F from (9). Here, F is non-smooth, but it allows for the application of a semismooth Newton method. Using the semismoothness calculus in Banach spaces from, e.g. Ito & Kunisch (2008) or Ulbrich (2011), we introduce the generalized differential of the projection operator $P_{\text{ad}}(y)$

$$DP_{\text{ad}}(z)(h) = \chi_{\mathscr{I}} h, \tag{11}$$

where $\chi_{\mathscr{I}} h := (\chi_{\mathscr{I}^k} h_k)_k$ and $\chi_{\mathscr{I}^k}$ denotes the indicator function of the inactive set $\mathscr{I}^k = \{t \in (0,T) \mid u_{\min}(t) < z_k(t) < u_{\max}(t)\}$ of component u_k. The generalized derivative of F at point z^n in the direction δz is then given by

$$H(z^n)(\delta z) = \alpha \delta z + B^* \delta p(\delta z). \tag{12}$$

To compute $\delta p(\delta z)$, first the tangent equation depending on δz and incorporating $\chi_{\mathscr{I}^k}$ is solved for δv, δw forward in time, and then the second adjoint equation is solved for δp, δq backward in time, see the end of Appendix A.

Taken together we can formulate the semismooth Newton iteration

$$H(z^n)(\delta z) = -F(z^n), \quad z^{n+1} = z^n + \delta z. \tag{13}$$

While the operator H is in general non-symmetric, it is symmetric with respect to the L^2-inner product of the inactive set $(a,b)_{\mathscr{I}} := \sum_{k=1}^{N_e} \int_0^T \chi_{\mathscr{I}^k} a_k b_k \mathrm{d}t$. Therefore, we compute d by solving (13) with the CG method using $(\cdot,\cdot)_{\mathscr{I}}$ as inner product. By this, we obtain a solution of (13) on the inactive set, i.e. $\chi_{\mathscr{I}}(Hd + F) = 0$. Afterwards, a solution of the full system (13) is obtained by updating the components on the active set according to

$$\delta z = d - \frac{1}{\alpha}(F(z^n) + H(z^n)d). \tag{14}$$

We note that, for $U_{\mathrm{ad}} = U$, the semismooth Newton method coincides with the well-known matrix-free Newton method of Hinze & Kunisch (2001).

For globalization, the method is embedded into a trust region framework analogously to Steihaug (1983). Therefore,we note that the CG method with $(\cdot,\cdot)_{\mathscr{I}}$ computes a particular solution of the quadratic problem

$$\min_{h \in U} \varphi_{z^n}(h) := (h, F(z^n))_{\mathscr{I}} + \tfrac{1}{2}(h, H(z^n)h)_{\mathscr{I}}.$$

We replace this problem by the trust region problem

$$\min_{h \in U} \varphi_{z^n}(h) \quad \text{s.t. } \|h\|_{\mathscr{I}} \leqslant \Delta_n,$$

with trust region radius $\Delta_n > 0$ and $\|h\|_{\mathscr{I}} = \sqrt{(h,h)_{\mathscr{I}}}$. It is solved with Steihaug-CG (Steihaug, 1983, Section 2) using the inner product $(\cdot,\cdot)_{\mathscr{I}}$. The update (14) is done only for a fully converged CG method, hence not for the cases when negative curvature or a large step is encountered. For practical realization the update (14) should be replaced by minimizing the residual in the direction of $r = -F - Hd$ according to

$$\delta z = d + \theta r \quad \text{with } \theta \in \mathbb{R}, \quad \theta = \arg\min(\|H(d + \theta r) + F\|_{L^2(0,T;\mathbb{R}_e^N)}), \tag{15}$$

in order to make the procedure more robust w.r.t. rounding errors.

The update of the trust region radius Δ_n and the decision of accepting or rejecting a step are done analogously to Steihaug (1983); see the full algorithm in Appendix A. Additionally, we modify the trust region method to be monotone, i.e. accepted steps will always yield a decrease in the objective.

5.2 *Direct search method for the upper-level problem*

The upper-level problem is solved with a globally convergent derivative-free optimization method based on bisection. It is assumed that the optimal values $G(T)$ of the LLP are continuous w.r.t T. We start from a triple $L < M < R$ with $G(M) < G(L)$ and $G(M) < G(R)$, i.e. we assume that a minimizer is contained in $[L,R]$. Then both intervals are bisected by $P := (L+M)/2$ and $Q := (M+R)/2$ and $G(P)$, $G(Q)$ are computed. Next, we choose M as minimizer in $\{M,P,Q\}$, tighten both intervals and iterate. Additionally, we skip the computation of $G(Q)$ if $G(P) < G(M)$ holds.

6. Discretization

We give a brief description of the discretization of the LLP. To combine the advantages of First-Discretize-Then-Optimize (FDTO) methods and First-Optimize-Then-Discretize (FOTD) methods, we choose a Galerkin Finite Element (FE) method in space together with a Petrov–Galerkin method in time, which allows for exact discrete derivatives and a natural translation of the optimality conditions from the continuous to the discrete level; see Becker *et al.* (2007). Hence, FDTO and FOTD commute and coincide within our framework, which is very important for trust region Newton methods.

In particular, we choose Lagrange Q1 elements on a quadrilaterally structured grid for spatial and the Crank–Nicolson method in the cG(1)-scheme for temporal discretization; for the latter see e.g. Eriksson *et al.* (1996). Since the spatial discretization is straightforward, we defer it to Appendix B. However, the time discretization is important to gain exact discrete derivatives and decoupling. Therefore, the essential parts are presented in the following, concentrating on the semidiscretization in time.

We aim for an efficient decoupling method to solve the ODE and PDE variables independently per time step. Therefore, we utilize a decoupling of the ODE from the PDE by taking the gating variable explicitly in the PDE. By working thoroughly through the Lagrangian calculus, we reestablish the exact discrete derivatives respecting the decoupling.

A time grid $t_0 < \cdots < t_N$ with step-sizes $\tau_m := t_m - t_{m-1}$ is chosen. The state variables are semidiscretized in time as continuous piecewise linear functions with values $V^m(x) = v(t_m,x)$, $m = 0,\dots,N$ and analogously for w; see Fig. 1. The adjoint and control variables are piecewise constant in time with values $P^m(x)$. Hence we have $p(t,x) = \sum_{m=1}^{N} P^m(x)\chi_{(t_{m-1},t_m]}(t)$, and analogously for $q(t,x)$ and $u_k(t) = \sum_{m=1}^{N} u_k^m \chi_{(t_{m-1},t_m]}(t)$.

Therefore, the semidiscrete Lagrangian $\mathcal{L}$ can be expressed as

$$\begin{aligned}\mathcal{L}(\dots) := T &+ \frac{\mu}{2}\int_\Omega \left(V^N\right)^2 \mathrm{d}x + \frac{\alpha}{2}\sum_{k=1}^{N_\mathrm{e}}\sum_{m=1}^{N}\tau_m (u_k^m)^2 - \sum_{m=1}^{N}\int_\Omega \frac{\tau_m}{2}\nabla P^m\cdot\bar{\sigma}_i\nabla(V^m+V^{m-1})\\ &+ P^m\left[V^m - V^{m-1} - \tau_m\sum_{k=1}^{N_\mathrm{e}}\chi_{\Omega_{\mathrm{con},k}}u_k^m + \frac{\tau_m}{2}I(V^m,W^{m-1}) + \frac{\tau_m}{2}I(V^{m-1},W^{m-1})\right]\mathrm{d}x\\ &- \sum_{m=1}^{N}\int_\Omega Q^m\left[W^m - W^{m-1} + \frac{\tau_m}{2}G(V^m+V^{m-1},W^m+W^{m-1})\right]\mathrm{d}x,\end{aligned}$$

where we leave the inequality constraints as explicit constraints. We again emphasize the decoupling of w at $I(V^m,W^{m-1})$, which later results also in an adapted decoupling in the adjoint and tangent equations. Therefore, the ODE can generally be solved efficiently in a matrix-free manner.

Next, the well-known Lagrange formalism yields a consistent semidiscretization of the tangent, adjoint and second adjoint equation. A subsequent spatial discretization with FE is straightforward and results in the equations in Appendix B.

The FE calculations are done with `deal.II` (Bangerth *et al.*, 2007). The non-linear systems in each time step of the state equation are solved with Newton's method, and the linear systems are solved directly with UMFPACK.

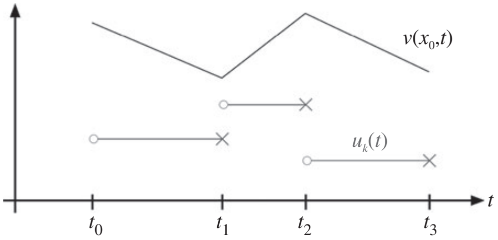

FIG. 1. Ansatz space in time for state (upper curve), control and adjoint variables (lower curve).

7. Numerical experiments

In the following, the proposed formulation and method are tested on several examples. The choice of parameters is inspired by Franzone *et al.* (2006), where one can also find the aforegoing non-dimensionalization. The following parameters are fixed throughout all examples:

η_0	η_1	η_2	η_3	$v_{\rm th}$	$v_{\rm pk}$	$\bar{\sigma}_i$
1.5	4.4	0.012	1.0	13	100	$\mathrm{diag}(3\cdot 10^{-3}, 3.1525\cdot 10^{-4})$

The geometry is set to be a rectangle $\Omega = (0,2)\times(0,0.8)$ of size 2 cm × 0.8 cm, which is discretized into 128 × 64 cells. All computations were done with an equidistant time discretization with step size $\tau = 0.04$ (ms). The stopping criteria are set to $\|F_n\| < \min(10^{-5}, 10^{-5}\|F_0\|)$ for the (trust region) Newton method—where the gradient F_n is the discretization of (9)—and $\|r_k\| < 10^{-5}\|r_0\|^{1.3}$ for the residual of the Steihaug-CG method.

The initial condition (v_0, w_0) describes a reentry wave of the 'figure of eight' type. It is constructed by the usual S1–S2-protocol as follows. Starting by exciting the lower edge $v(x,0) = 101$ if $x_2 \leqslant 1/160$ and 0 otherwise, $w(x,0) = 0$, we integrate the uncontrolled solution until $t = 130$ using a fixed step size $\tau = 0.1$. The solution describes a planar wave front travelling from the bottom up. As soon as the centre gets excitable again, a second stimulus is based on a circle around the midpoint with radius 0.3 for 2 ms, i.e. $I_{\rm e} = 200\chi_{\Omega_{S2}}(x)\chi_{[130,132]}(t)$ with $\Omega_{S2} = B_{0.3}(1,0.4)$. We carry on the simulation without any further stimulus up to $t = 217$ and save both states $v(x,217), w(x,217)$ as future initial conditions for the optimization. The timing and radius of the second stimulus are crucial. For different domains or parameters, one has to adapt it by trial and error; otherwise a reentry wave will not evolve.

In the examples which follow, we address the different demands for optimized pulses, looking for: a short pulse with restricted amplitude in Example 1, a low norm $\|u\|$ in Example 2 and a robust optimized pulse w.r.t. the tensor data in Example 3.

7.1 *Example 1: symmetric defibrillation of a reentry wave*

We start with an axially symmetric problem, where it is possible to defibrillate with just one control pulse, i.e. $N_{\rm e} = 1$. The geometry of the control domain is $\Omega_{\rm con,1} = [0,0.25]\times[0.3,0.55]\cup[1.75,2]\times[0.3,0.55]$; see Fig. 2. The bilevel method was started on the interval $[L,R] = [30,40]$ and convergence was reported for $|R-L| < 4\times 10^{-2}$. The parameters were $t_{\rm f} = 64$, $\alpha = 10^{-3}$, $\mu = 1000$ and $u_{\max} = -u_{\min} = 100$. The initial control is set to $u_0^1 = u_0 = -50$ for the first LLP with $T^1 = 40$. All other LLPs for $k \geqslant 2$ were warm-started with the optimal control of the former LLP $\bar{u}^{k-1}$ restricted to the current interval $[0, T^k]$ or expanded with zero, e.g. $u_0^k(t) = \bar{u}^{k-1}(t)\chi_{[0,\min(T^k,T^{k-1})]}(t)$. An alternative

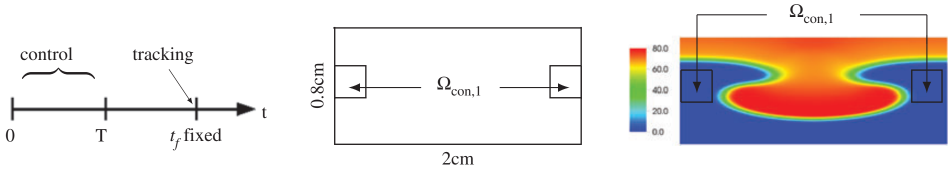

FIG. 2. Time domain, geometry and initial condition of Example 1.

TABLE 1 *TR-SN method for the first LLP with $T = 40$*

n	$j(u_n)$	$\|F(u_n)\|$	#CG	$\|\mathscr{I}\|$
0	38,118	8.3×10^{1}		1000
5	538	6.9×10^{0}	2	555
10	327	1.8×10^{0}	1	738
15	262	6.5×10^{-1}	2	837
20	244	1.6×10^{-2}	13	915
21	244	1.9×10^{-3}	13	915
22	244	1.3×10^{-4}	14	915

procedure to obtain the new initial control $u_0^k(t)$ is to linearly map $\bar{u}^{k-1}$ from $[0, T^{k-1}]$ to $[0, T^k]$ by $u_0^k(t) = \bar{u}^{k-1}(t(T^{k-1}/T^k))$.

The direct search method in the upper level needs 16 function evaluations to converge at $\bar{T} = 34.12$ with $\bar{J} = 238.786$, i.e. 16 LLPs were solved in total. We note that this is not the shortest pulse that effectively defibrillates, since we are facing a multi-objective formulation with three goals. It is an optimal compromise between short duration and low energy. The total number of state, gradient and Hessian evaluations throughout the bilevel run are 78, 71 and 658, respectively. We see that 7 of the 62 TR-Newton steps are rejected. The total number of 559 CG steps yields ≈ 9 CG steps per Newton step; excluding the globalization steps, we observe ≈ 14 CG steps per fully converged CG call.

Typically, the most CPU work is required for the first LLP with $T = R = 40$, since it is not warm-started (see Table 1). The TR–SN method needs 22 steps to converge, reducing the objective from $j(u_0) = 38118$ to $j(u_{22}) = 244$ and reducing the first-order optimality $\|F(u_n)\|$ significantly. The last two columns show the number of CG iterations and the number of inactive time points $|\mathscr{I}|$. All subsequent LLP solves show a fast convergence of the TR–SN method, see e.g. the second LLP solve with $T = L = 30$ in Table 2. Due to the warm-start, only a few globalization steps are needed, where Steihaug-CG is stopped due to too large steps (flag 1) or negative curvatures (flag 2). Afterwards, the CG is fully solved (flag 0) and the number of inactive time points $|\mathscr{I}|$ converges. Superlinear convergence of the objective j is observed from $s_n := (j(u_n) - j(u_{n-1}))/(j(u_{n-1}) - j(u_{n-2}))$ in the last column.

The time optimal control $\bar{u}(t)$ is depicted as the second curve in both graphs of Fig. 3. Additional curves show the time optimal controls for different control bounds $u_{\max}$ (left) and different cost parameters α (right). Apparently, all time optimal controls differ to a large extent from the initial control $u_0(t) = -50\chi_{[0,40]}(t)$, in particular the shape, the duration and the switching structure. Consequently, the corresponding trajectories behave qualitatively different. While the initial control only counteracts the wave propagation due to $u_0 \leqslant 0$, we observe a speed-up of the wave propagation at certain points for the time optimal control, since it features positive values, too.

TABLE 2 *TR-SN method for the second LLP with $T = 30$*

n	$j(u_n)$	$\|F(u_n)\|$	#CG	Flag	$\|\mathscr{I}\|$	s_n
0	255.744	6.2×10^{-1}	0		665	
1	254.084	4.3×10^{-1}	2	1	663	
2	254.084	4.3×10^{-1}	8	2	663	
3	253.574	4.7×10^{-2}	7	1	577	0.31
4	253.533	2.0×10^{-3}	14	0	570	0.08
5	253.533	1.9×10^{-5}	14	0	569	0.00
6	253.533	7.3×10^{-11}	15	0	569	0.00

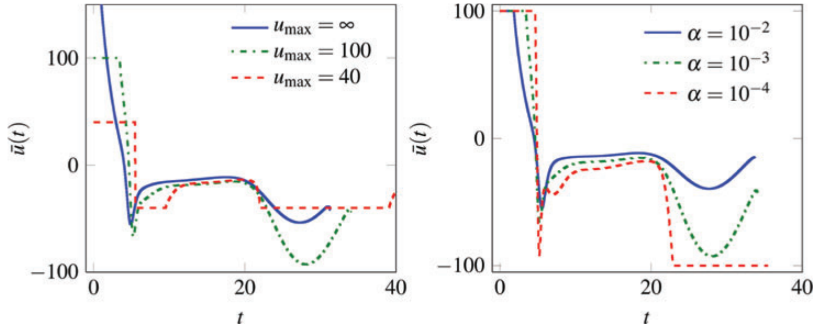

FIG. 3. Time optimal controls for different $u_{\max} = -u_{\min}$ with $\alpha = 10^{-3}$ (left) and different α with $u_{\max} = 100$ (right).

TABLE 3 *Optimal value $\bar{J}$, pulse length $\bar{T}$ and norm of the time optimal pulse for different $u_{\max}$ with $\alpha = 10^{-3}$ (left) and for different α with $u_{\max} = 100$ (right)*

$u_{\max}$	$\bar{J}$	$\bar{T}$	$\|\bar{u}(t)\|_U$	α	$\bar{J}$	$\bar{T}$	$\|\bar{u}(t)\|_U$
∞	130	31	329	10^{-2}	501	33.9	206
100	239	34	334	10^{-3}	239	34.1	334
40	2167	39	217	10^{-4}	173	35.1	436

According to the left plot and Table 3, a lower-bound $u_{\max}$ leads to an increase in the optimal pulse length $\bar{T}$ and the optimal value $\bar{J}$, since the effectivity of the control decreases. On the other hand, reducing the cost parameter α results in a smaller optimal value, a slightly increased pulse length and a larger energy of the optimal pulse.

For a verification we compute the transversality condition (7) both for the initial guess (u_0^1, T^1) and the optimal pair $(\bar{u}, \bar{T})$, which yields -1660 and -0.1, respectively. The comparison shows a relative decrease of 6×10^{-5} in this optimality condition, which underlines the optimality of the computed time optimal control.

7.2 *Example 2: asymmetric defibrillation of a reentry wave*

The second example considers two independent electrode plates with $I_e = \chi_{[0,T]}(t)(u_1(t)\chi_{\Omega_{\text{con},1}}(x) + u_2(t)\chi_{\Omega_{\text{con},2}}(x)$ in an asymmetric setting $\Omega_{\text{con},1} = [0, 0.25] \times [0.4, 0.55]$, $\Omega_{\text{con},2} = [1.75, 2] \times [0.35, 0.4]$;

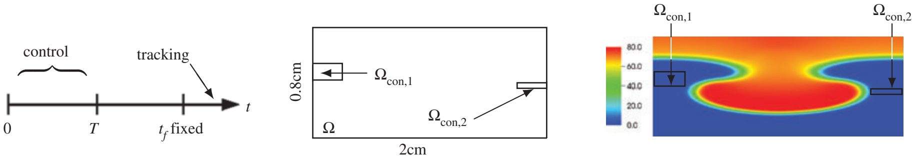

FIG. 4. Time domain, geometry and initial condition $v(0,x)$ of Example 2.

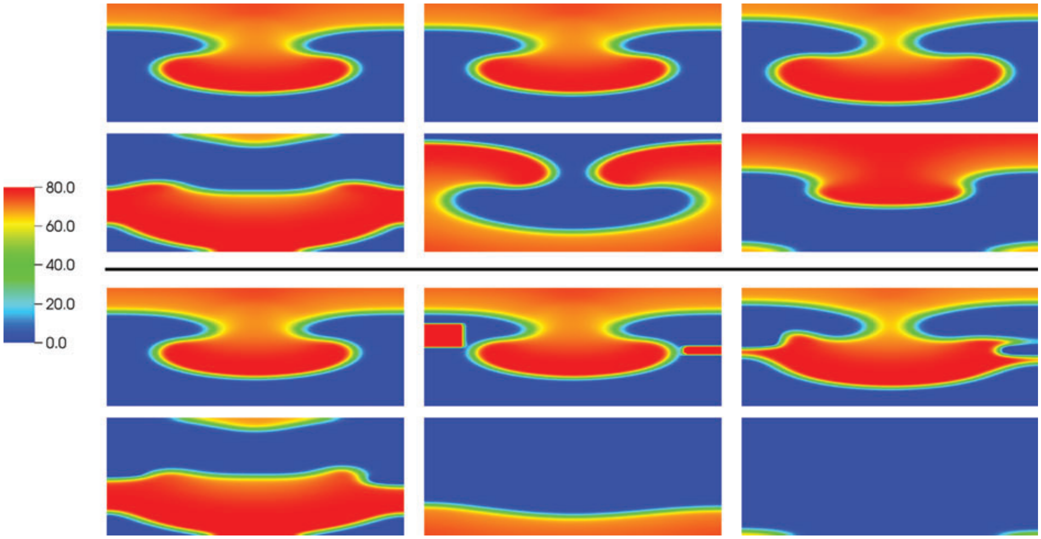

FIG. 5. Snapshots of the uncontrolled wave (above line) compared with the controlled wave (below line). Depicted are snapshots of the $v(t,x)$ at times $t = 0$, 0.12, 6 (upper row) and $t = 16$, 48, 65 (lower row).

see Fig. 4. The parameters are $t_{\mathrm{f}} = 65$, $\alpha = 1 \cdot 10^{-5}$, $\mu = 100$ and $U_{\mathrm{ad}} = U$, i.e. the LLP method coincides with a trust region Newton method.

The bilevel method was started on the interval $[L,R] = [27.5, 37.5]$ with $u_0 = -50$ and convergence was reported for $|R - L| < 4 \times 10^{-2}$. We observe again global convergence of the bilevel method and locally superlinear convergence of each LLP. Figure 5 depicts snapshots of the transmembrane voltage $v(t,x)$ for six different times t, both for the uncontrolled evolution of the reentry wave (above line) and the optimally controlled wave (below line). The uncontrolled wave ($u \equiv 0$) exhibits a periodic behaviour with a period around 75. The controlled wave is influenced heavily at the very beginning of the time horizon. The positive part of the pulses act on the excitable part of the tissue adjacent to the wave front, bringing it to a non-excitable state (see the second plot for $t = 0.12$). Thus, the wave cannot progress upwards, falls apart and leaves the domain. At the terminal time, not a single part of the tissue is excited, which confirms a successful defibrillation.

For checking the gradient and Hessian consistency, we verify the derivatives given by the adjoint calculus via a comparison with finite differences in Table 4 using the initial control $u = -50$ and $d = -F(z)$. For the gradient, the absolute difference $\mathrm{abs} = C - (g,d)$ and the relative difference $\mathrm{rel} = \mathrm{abs}/C$ are computed using the central difference $C = (j(u + \varepsilon d) - j(u - \varepsilon d))/2\varepsilon$. For the Hessian, the differences are $\mathrm{abs} = C - (d, Hd)$, $\mathrm{rel} = \mathrm{abs}/C$ with $C = (j(u + \varepsilon d) - 2j(u) + j(u - \varepsilon d))/\varepsilon^2$.

TABLE 4 *Verification of the gradient and Hessians against finite differences with $U_{\text{ad}} = U$*

	Gradient		Hessian	
ε	abs	rel	abs	rel
1.0e+01	7.9e+03	1.0e+00	1.6e+03	1.0e+00
1.0e+00	5.4e−01	1.7e−02	2.0e−01	2.9e−02
1.0e−01	5.2e−03	1.6e−04	1.9e−03	2.8e−04
1.0e−02	5.2e−05	1.6e−06	1.9e−05	2.8e−06
1.0e−03	5.8e−07	1.8e−08	8.8e−06	1.3e−06
1.0e−04	1.2e−07	3.6e−09	1.4e−03	2.1e−04
1.0e−05	4.6e−07	1.4e−08	1.3e−01	2.0e−02

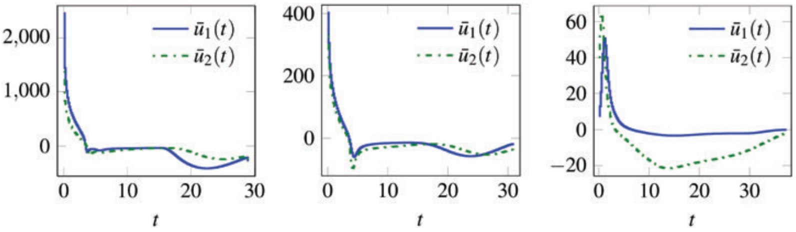

FIG. 6. Time optimal controls $\bar{u}_1(t)$ and $\bar{u}_2(t)$ for different $\alpha = 10^{-5}, 10^{-3}, 10^{0}$ with corresponding norms $\|u\|_U = 2195, 442, 121$.

All columns confirm a quadratic convergence of the finite differences to the adjoint-based values of the first and second derivatives, as well as a very high precision of the gradient and Hessian code. This is crucial for the success of the optimization since optimal control problems with only terminal observation are known to be highly ill-conditioned.

To find time optimal control pulses with consideration for small energy, we successively increase α and depict the corresponding time optimal controls and their energy in Fig. 6. The required energy decreases from 2195 to 121 while maintaining an effective defibrillation pulse. With an increase in α the optimal duration increases as well.

7.3 *Example 2: a robust design*

In the next example, we take into account some uncertainty in the conductivity tensor data, reflecting the fact that they may vary heavily between different settings. As an example, we set $\bar{\sigma}_i = \text{diag}(\sigma \times 10^{-3}, 3.1525 \times 10^{-4})$ and assume that $\sigma \in \mathbb{R}^+$ is a random variable. By extending (Boyd & Vandenberghe, 2009, Secttion 6.4) to optimal control problems, the expectation value of the tracking term enters the objective. Thus we replace J by

$$J_{\mathbb{E}} = T + \frac{\mu}{2}\mathbb{E}(\|v(x,t_{\text{f}};\sigma)\|^2) + \frac{\alpha}{2}\|u\|_U^2. \tag{16}$$

Together with the constraints (1) and $u \in U_{\text{ad}}$, this constitutes a stochastic robust control problem, which in general is computationally demanding. Therefore, we restrict ourselves to the case where σ takes only a finite number of values $\{\sigma^1, \ldots, \sigma^r\}$ with probabilities $P_1, \ldots, P_r \geqslant 0$, $\sum_{k=1}^{r} P_k = 1$. Consequently, the

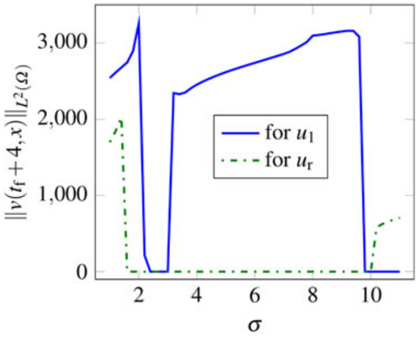

FIG. 7. $\|v(t_{\mathrm{f}}+4,x)\|_{L^2(\Omega)}$ for different values of σ, both for the optimal pulse u_1 and the robust optimal control u_{r}.

objective, the reduced gradient and Hessian change to

$$J_r = T + \frac{\mu}{2}\sum_{k=1}^{r} P_k \|v(x,t_{\mathrm{f}};\sigma^k)\|^2 + \frac{\alpha}{2}\|u\|^2 = \sum_{k=1}^{r} P_k J(v,u,T;\sigma^k), \tag{17a}$$

$$F_r = \sum_{k=1}^{r} P^k F(z;\sigma^k), \qquad H_r(z)\delta z = \sum_{k=1}^{r} P^k H(z;\sigma^k)\delta z. \tag{17b}$$

We see that each call to the objective, the gradient and the Hessian has to be split into r calls to the existing solvers (with different $\bar{\sigma}_i$) followed by a weighted mean. This would allow a parallelization of the code, which, however, is not pursued here.

To investigate the effect of the robustness approach, we compare an optimal control (for fixed $\sigma = 3$) to a robust optimal control in the following. To facilate this comparison, we fix the pulse length $T = t_{\mathrm{f}}$, i.e. we compute only one LLP for both settings. Thus, we compute the solution u_1 of the LLP with fixed $\sigma = 3$ on the one hand, and the robust counterpart u_{r}, which minimizes the LLP incorporating the changes from (17), on the other hand.

We investigate the reentry setting with an electrode placement different from above: $\Omega_{\mathrm{con},1} = [0.05, 0.5] \times [0.45, 0.55]$, $\Omega_{\mathrm{con},2} = [1.8, 1.9] \times [0, 0.45]$. The parameters are set to $\alpha = 10^{-2}$, $\mu = 1000$ and $t_{\mathrm{f}} = 86$. As an example, we test a uniform distribution for $\sigma \in \{2, 4, 6, 8, 10\}$, i.e. $p_j = 1/r \;\forall j$ with $r = 5$.

The optimization yields a robust pulse at the expense of a higher norm: $\|u_{\mathrm{r}}\| = 713$ compared with $\|u_1\| = 189$. To inspect the robustness of the two pulses, we test them for different values of $\sigma = 1 + n/5$, $n = 0, \ldots, 50$. For each value of σ, the monodomain model is solved and successful defibrillation is confirmed at t_{f} and a later time $t = t_{\mathrm{f}} + 4$, to exclude regeneration of a reentry wave. Figure 7 shows the norm $\|v(x, t_{\mathrm{f}} + 4)\|_{L^2(\Omega)}$ over σ. The zero set of the curves corresponds to a successful defibrillation. While u_{r} defibrillates for all $\sigma \in [2, 10]$, the pulse u_1 is found to be successful only for $\sigma \in [2.8, 3]$, and by chance also for $\sigma \in [9.8, 11]$; see Fig. 7.

8. Conclusion and outlook

It was demonstrated that the choice of cost functional reflecting the system dynamics and incorporating time-optimality for the joint optimization of the shape and the duration of defibrillation pulses is effective for the optimal control of the monodomain model. Certainly it would be of interest to extend the proposed methods to the bidomain equations, to realistic geometries and to more complex ionic models.

Acknowledgment

The authors thank Konstantin Pieper, TU Munich for many fruitful discussions, especially concerning Section 5.1.

Funding

The research was supported in part by the Austrian Science Fund (FWF) under grant SFB F32, 'Mathematical Optimization and Applications in Biomedical Sciences'.

References

Bangerth, W., Hartmann, R. & Kanschat, G. (2007) deal.II—a general purpose object oriented finite element library. *ACM Trans. Math. Softw.*, **33**, 24/1–24/27.

Becker, R., Meidner, D. & Vexler, B. (2007) Efficient numerical solution of parabolic optimization problems by finite element methods. *Optim. Methods Softw.*, **22**, 813–833.

Borzì, A. & Griesse, R. (2006) Distributed optimal control of lambda-omega systems. *J. Numer. Math.*, **14**, 17–40.

Bourgault, Y., Coudière, Y. & Pierre, C. (2009) Existence and uniqueness of the solution for the bidomain model used in cardiac electrophysiology. *Nonlinear Anal. Real World Appl.*, **10**, 458–482.

Boyd, S. & Vandenberghe, L. (2009) *Convex Optimization*. Cambridge: Cambridge University Press.

Casas, E., Ryll, C. & Tröltzsch, F. (2013) Sparse optimal control of the Schlögl and FitzHugh–Nagumo systems. *Comput. Meth. Appl. Math.*, **13**, 415–442.

Constantin, P. & Foias, C. (1988) *Navier–Stokes Equations*. Chicago: The University of Chicago Press.

Eriksson, K., Estep, D., Hansbo, P. & Johnson, C. (1996) *Computational Differential Equations*. Cambridge: Cambridge University Press.

Franzone, P., Deuflhard, P., Erdmann, B., Lang, J. & Pavarino, L. (2006) Adaptivity in space and time for reaction–diffusion systems in electrocardiology. *SIAM J. Sci. Comput.*, **28**, 942–962.

Götschel, S., Chamakuri, N., Kunisch, K. & Weiser, M. (2013) Lossy compression in optimal control of cardiac defibrillation. *J. Sci. Comput.*, 1–25.

Hinze, M. & Kunisch, K. (2001) Second order methods for optimal control of time-dependent fluid flow. *SIAM J. Control Optim.*, **40**, 925–946.

Ito, K. & Kunisch, K. (2008) *Lagrange Multiplier Approach to Variational Problems and Applications*. Advances in Design and Control, vol. 15. Philadelphia: SIAM.

Ito, K. & Kunisch, K. (2010) Semismooth Newton methods for time-optimal control for a class of ODEs. *SIAM J. Control Optim.*, **48**, 3997–4013.

Keener, J. & Sneyd, J. (2009) *Mathematical Physiology II: Systems Physiology*. New York: Springer.

Kunisch, K., Pieper, K. & Rund, A. (2014) Time optimal control for a reaction diffusion system arising in cardiac electrophysiology—a monolithic approach. *Technical Report SFB-Report 2014–016*. SFB 'Mathematical Optimization and Biomedical Applications'.

Kunisch, K. & Wagner, M. (2011) Optimal control of the bidomain system (i): the monodomain approximation with the Rogers–McCulloch model. *Nonlinear Anal. Real World Appl.*, **13**, 1525–1550.

Murray, J. D. (2002) Mathematical biology I. An introduction. *Interdisciplinary Applied Mathematics*, vol. 17, 3rd edn. New York: Springer.

Nagaiah, C., Kunisch, K. & Plank, G. (2011) Numerical solution for optimal control of the reaction-diffusion equations in cardiac electrophysiology. *Comput. Optim. Appl.*, **49**, 149–178.

NAGAIAH, C., KUNISCH, K. & PLANK, G. (2013) Optimal control approach to termination of re-entry waves in cardiac electrophysiology. *J. Math. Biol.*, **67**, 359–388.

PIEPER, K. (2015) Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems. *Ph.D. Thesis*, Technische Universität München.

POTSE, M., DUBÉ, B., RICHER, J., VINET, A. & GULRAJANI, R. M. (2006) A comparison of monodomain and bidomain reaction-diffusion models for action potential propagation in the human heart. *IEEE Trans. Biomed. Eng.*, **53**, 2425–2435.

PURI, M., CHAPALAMADUGU, K. C., MIRANDA, A., GELOT, S., ADITHYA, P. C., MORENO, W., LAW, C. & TIPPARAJU, S. M. (2013) Integrated approach for smart implantable cardioverter defibrillator (ICD) device with real time ECG monitoring: use of flexible sensors for localized arrhythmia sensing and stimulation. *Front. Physiol.*, **4**, 1–4.

ROGERS, J. & MCCULLOCH, A. D. (1994) A collocation-Galerkin finite element model of cardiac action potential propagation. *IEEE Trans. Biomed. Eng.*, **41**, 743–757.

STEIHAUG, T. (1983) The conjugate gradient method and trust regions in large scale optimization. *SIAM J. Numer. Anal.*, **20**, 626–637.

SUNDNES, J., NIELSEN, B. F., MARDAL, K. A., CAI, X., LINES, G. T. & TVEITO, A. (2006) On the computational complexity of the bidomain and the monodomain models of electrophysiology. *Ann. Biomed. Eng.*, **34**, 1088–1097.

TRÖLTZSCH, F. (2010) *Optimal Control of Partial Differential Equations*. Graduate Studies in Mathematics, vol. 112. Providence: American Mathematical Society.

ULBRICH, M. (2011) *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. MOS-SIAM Series on Optimization. Philadelphia: SIAM.

VIGMOND, E., DOS, Santos, DEO, M. & PLANK, G. (2008) Solvers for the cardiac bidomain equations. *Prog. Biophys. Mol. Biol.*, **96**, 3–18.

Appendix A. Optimization algorithm TR-SN

1. Initialize z^0, maximal radius $\varDelta_{\max} > 0$, initial radius $0 < \varDelta_0 \leqslant \varDelta_{\max}$ and set $n = 0$.

2. Solve state and adjoint equations for z^n , set up gradient $F(z^n)$ from (9) and determine inactive sets $\mathscr{I}^k = \{t \in (0,T) \mid u_{\min}(t) < z_k^n(t) < u_{\max}(t)\}$.

3. Compute d from (13) by Steihaug-CG using the L^2-inner product on the inactive set $(\cdot,\cdot)_{\mathscr{I}}$.

4. If Steihaug-CG is fully converged (i.e. Steihaug, 1983, (2.3) is fulfilled), then compute δz according to (14). Otherwise set $\delta z = d$.

5. Calculate $j(P_{\text{ad}}(z^n + \delta z))$ and $\varrho_n := \varrho^{\text{act}}/\varrho^{\text{pred}} = (j(P_{\text{ad}}(z^n)) - j(P_{\text{ad}}(z^n + \delta z)))/-\varphi_{z^n}(\delta z)$.

6. Update z:

$$z^{n+1} := \begin{cases} z^n + \delta z, & \text{if } \varrho_n > \alpha_2 \text{ and } \varrho^{\text{act}} > \epsilon \quad \text{(accept)},\\ z^n, & \text{otherwise} \quad \text{(reject)}.\end{cases}$$

7. update radius $\varDelta_n$:

$$\varDelta_{n+1} = \begin{cases} \min(2\|\delta z\|_{\mathscr{I}}, \varDelta_{\max}), & \text{if } \varrho_n \in [0.7, 1.3] \quad \text{(model good)}\\ 0.25\varDelta_n, & \text{elseif } \varrho^{\text{act}} \leqslant \epsilon \quad \text{(no decrease)}\\ 0.5\|\delta z\|_{\mathscr{I}}, & \text{elseif } \varrho_n \notin [0.25, 1.75] \quad \text{(model bad)}\\ \varDelta_n, & \text{else.}\end{cases}$$

8. If stopping criteria are not fulfilled, set $n = n + 1$ and goto 2.

Each Hessian evaluation in 3. is carried out by the following steps.

1. Solve the tangent equation with δz and corresponding states (v^n, w^n) for $u^n = P_{\mathrm{ad}}(z^n)$

$$\begin{aligned}
&\delta v_t - \nabla \cdot (\bar{\sigma}_i \nabla \delta v) + I_v(v_n, w_n)\, \delta v + I_w(v_n)\, \delta w \\
&\qquad = \sum_{k=1}^{N_{\mathrm{e}}} \chi_{\Omega_{\mathrm{con},k}}(x) \chi_{\mathscr{I}^k}(t) \delta z_k(t) \chi_{(0,T)}(t) \quad \text{in } Q, \\
&\delta w_t + G(\delta v, \delta w) = 0 \quad \text{in } Q, \\
&\nu \cdot \bar{\sigma}_i \nabla\, \delta v = 0 \quad \text{on } \Sigma, \\
&\delta v(x, 0) = 0, \quad \delta w(x, 0) = 0 \quad \text{in } \Omega.
\end{aligned}$$

2. Solve the second adjoint equation with p^n the adjoint to (v^n, w^n)

$$\begin{aligned}
&- \delta p_t - \nabla \cdot (\bar{\sigma}_i \nabla \delta p) + I_v(v_n, w_n) \delta p + G_v \delta q = -I_{vv}(v_n) p_n \delta v - I_{vw} p_n \delta w \quad \text{in } Q, \\
&- \delta q_t + I_w(v_n) \delta p + G_w \delta q = -I_{vw} p_n \delta v \quad \text{in } Q, \\
&\nu \cdot \bar{\sigma}_i \nabla \delta p = 0 \quad \text{on } \Sigma, \\
&\delta p(x, t_{\mathrm{f}}) = \mu \delta v, \quad \delta q(x, t_{\mathrm{f}}) = 0 \quad \text{in } \Omega.
\end{aligned}$$

3. Evaluate (12).

Appendix B. Discretization formulas for state, adjoint and second-order solvers

The space is discretized with a FE-Galerkin method using Lagrange-Q1 elements $\{\varphi_i(x), i = 1, \dots, N_x\}$. Hence, we search for FE-coordinates $v_m := (v_m^i)_{i=1,\dots,N_x}$ with $v(t_m, x) = V^m(x) = \sum_{i=1}^{N_x} v_m^i \varphi_i(x)$ and analogously for w, δv, δw, p, q, δp, δq.

As matrices we define the mass matrix $M := (\int_\Omega \varphi_i \varphi_j \mathrm{d}x)_{i,j}$, the negative stiffness matrix $\Delta_\sigma := -(\int_\Omega \nabla \varphi \bar{\sigma}_i \nabla \varphi_j \mathrm{d}x)_{i,j}$ and the Jacobian $J_{m,n} = (\int_\Omega (\partial I / \partial v)(v_m(x), w_n(x)) \varphi_i(x) \varphi_j(x) \mathrm{d}x)_{i,j}$. Further we define the vectors $\vec{\chi}_k := (\int_{\Omega_{\mathrm{con},k}} \varphi_j \, \mathrm{d}x)_j$ and $I_{m,n} := (\int_\Omega I(v_m(x), w_n(x)) \varphi_j(x) \, \mathrm{d}x)_j$. v_0, w_0 are the FE coordinates of the initial states $v_0(x)$, $w_0(x)$. The index m passes through $1, \dots, N$ for state and tangent equations, and through $1, \dots, N-1$ for adjoint (adj.) and second-adjoint equation.

$$\begin{aligned}
\text{state:} \quad &\left[M - \frac{\tau_m}{2} \Delta_\sigma \right] v_m + \frac{\tau_m}{2} I_{m,m-1} = \left[M + \frac{\tau_m}{2} \Delta_\sigma \right] v_{m-1} - \frac{\tau_m}{2} I_{m-1,m-1} \\
&+ \tau_m \sum_{k=1}^{N_{\mathrm{e}}} u_k^m \vec{\chi}_k \chi_{(0,T)}(t_m), \\
&\left[1 + \frac{\tau_m}{2} G_w \right] M w_m = \left[1 - \frac{\tau_m}{2} G_w \right] M w_{m-1} - \frac{\tau_m}{2} G_v M (v_m + v_{m-1}),
\end{aligned}$$

$$
\begin{aligned}
\text{adj.:}\quad & q_N = 0, \quad \left[M - \frac{\tau_N}{2}\Delta_\sigma + \frac{\tau_N}{2}J_{N,N-1}\right] p_N = \mu M v_N, \\
& \left[M - \frac{\tau_m}{2}\Delta_\sigma + \frac{\tau_m}{2}J_{m,m-1}\right] p_m = \left[M + \frac{\tau_{m+1}}{2}\Delta_\sigma - \frac{\tau_{m+1}}{2}J_{m,m}\right] p_{m+1} \\
& \quad - \frac{G_v}{2} M\,(\tau_m q_m + \tau_{m+1} q_{m+1}), \\
& \left[1 + \frac{\tau_m}{2}G_w\right] M q_m = \left[1 - \frac{\tau_{m+1}}{2}G_w\right] M q_{m+1} - \frac{\tau_{m+1}}{2}\int_\Omega I_w(V^{m+1} + V^m)P^{m+1}\varphi_j \mathrm{d}x, \\
\text{tangent:}\quad & \delta v_0 = 0, \quad \delta w_0 = 0, \\
& \left[M - \frac{\tau_m}{2}\Delta_\sigma + \frac{\tau_m}{2}J_{m,m-1}\right] \delta v_m = \left[M + \frac{\tau_m}{2}\Delta_\sigma - \frac{\tau_m}{2}J_{m-1,m-1}\right] \delta v_{m-1} \\
& \quad - \frac{\tau_m}{2}\int_\Omega I_w(V^m + V^{m-1})\delta W^{m-1}\varphi_j \mathrm{d}x + \tau_m \sum_{k=1}^{N_\mathrm{e}} \chi_{\mathscr{I}^k}(t_m)\chi_{(0,T)}(t_m)\delta z_k^m \vec{\chi}_k, \\
& \left[1 + \frac{\tau_m}{2}G_w\right] M\,\delta w_m = \left[1 - \frac{\tau_m}{2}G_w\right] M\,\delta w_{m-1} - \frac{\tau_m}{2}G_v M\,(\delta v_m + \delta v_{m-1}), \\
\text{2nd adj.:}\quad & \delta q_N = 0, \\
& \left[M - \frac{\tau_N}{2}\Delta_\sigma + \frac{\tau_N}{2}J_{N,N-1}\right] \delta p_N = -\frac{\tau_N}{2}\int_\Omega P^N[I_{vv}(V^N)\delta V^N + I_{vw}\delta W^{N-1}]\varphi_j \mathrm{d}x + M\,\delta v_N, \\
& \left[M - \frac{\tau_m}{2}\Delta_\sigma + \frac{\tau_m}{2}J_{m,m-1}\right] \delta p_m = \left[M + \frac{\tau_{m+1}}{2}\Delta_\sigma - \frac{\tau_{m+1}}{2}J_{m,m}\right] \delta p_{m+1} \\
& \quad - \frac{1}{2}G_v M\,(\tau_m \delta q_m + \tau_{m+1}\delta q_{m+1}) - \frac{1}{2}\int_\Omega \{\tau_m P^m[I_{vv}(V^m)\delta V^m + I_{vw}\delta W^{m-1}] \\
& \quad + \tau_{m+1}P^{m+1}[I_{vv}(V^m)\delta V^m + I_{vw}\delta W^m]\}\varphi_j \mathrm{d}x, \\
& \left[1 + \frac{\tau_m}{2}G_w\right] M\,\delta q_m = \left[1 - \frac{\tau_{m+1}}{2}G_w\right] M\,\delta q_{m+1} \\
& \quad - \frac{\tau_{m+1}}{2}\int_\Omega [\delta P^{m+1} I_w(V^{m+1} + V^m) + P^{m+1}(\delta V^{m+1} + \delta V^m)I_{vw}]\varphi_j\, \mathrm{d}x.
\end{aligned}
$$

All solves with a pure mass matrix are avoided by directly updating w_m, respectively, δw_m and by storing Mq_m resp. $M\,\delta q_m$.