MODELING THE CARDIAC ELECTROMECHANICAL FUNCTION: A MATHEMATICAL JOURNEY

ALFIO QUARTERONI, LUCA DEDÈ, AND FRANCESCO REGAZZONI

Abstract. In this paper we introduce the electromechanical mathematical model of the human heart. After deriving it from physical first principles, we discuss its mathematical properties and the way numerical methods can be set up to obtain numerical approximations of the (otherwise unachievable) mathematical solutions. The major challenges that we need to face—e.g., possible lack of initial and boundary data, the trade off between increasing the accuracy of the numerical model and its computational complexity—are addressed. Numerical tests here presented have a twofold aim: to show that numerical solutions match the expected theoretical rate of convergence, and that our model can provide a preliminary valuable tool to face problems of clinical relevance.

Contents

1. Introduction 372

Part 1. From the physical problem to the mathematical model 374

2. Cardiac anatomy and physiology at a glance 374
3. Modeling cardiac electrophysiology 376
4. Modeling active force generation 379
5. Modeling cardiac mechanics 381
6. Modeling blood circulation 383
7. A fully coupled cardiac electromechanics model 384

Part 2. Numerical approximation 386

8. Why should we turn to numerical approximation 386
9. Space discretization 387
10. Time discretization 388

Part 3. Numerical solution 390

11. Numerical results for cardiac electromechanics 390
12. Towards clinical applications 393
13. Conclusions and a look ahead 396
About the authors
References 398

Received by the editors December 7, 2020.
2020 Mathematics Subject Classification. Primary 65-02.
This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 740132, iHEART—An Integrated Heart Model for the simulation of the cardiac function, P.I., A. Quarteroni).

©2022 American Mathematical Society
1. Introduction

Historical tips. Since the dawn of history, the heart and its functioning have fascinated the best minds on our planet, through a long journey marked by intuitions and dazzles. Three centuries BC, Aristotle claimed that the blood vessels transmitted animal heat from the heart to the periphery. Shortly thereafter, Praxagoras of Kos sensed the different role of arteries and veins and assumed that the arteries transported air, and veins blood. It was Galen in the second century AD to recognize that arteries also carry blood.

Thirteen centuries later, the great Leonardo Da Vinci distinguished for the first time ventricles from atria (*auriculae*), describing the *moti del core* as a result of the contraction of the atria that coincides with the ventricular diastole, while, conversely, the ventricular systole occurs along with the expansion of the atria. In the seventeenth century Harvey correctly described the blood circulation for the first time. Thanks to the beating heart vivisection of various animals, he recognized that the heart valves in the veins were designed to activate only if the blood flowed to the heart and therefore concluded that “the blood has a movement, and it is circular”. In the following century, the great mathematicians Leonhard Euler and Daniel Bernoulli made decisive steps forward. In 1730 Bernoulli, professor of mathematics and anatomy at the University of Basel, formulated his famous *vis viva* equation between pressure, density, and velocity of blood. In 1775 Euler, in a work entitled *Principia pro motu sanguinis per arteria determinando*, wrote a series of differential equations—still fundamental and used in areas as diverse as the gas motion in pipes or aircraft aerodynamics design—to describe the evolution of blood flow and pressure in an idealized cylindrical blood vessel. These equations have then been generalized and today still set the ground for a one-dimensional model of blood flow in the complex network of arteries and veins of our circulatory system.

An integrated heart model. The ambition of today’s mathematicians is to travel the furrow traced by Euler for creating a virtual, immaterial heart, made only of equations, able to reproduce all the vital cardiac processes: the propagation of the electric potential, the contraction and relaxation of the myocardium (the heart muscle), the blood dynamics in atria and ventricles, the coronaries perfusion of the myocardium, the valve dynamics, and all their mutual interactions. This extraordinary complexity, which mysteriously determines the harmonious synchronization producing the heartbeat, can be turned into a mathematical model, thanks to the first principles of physics, that is Newton’s laws, thermodynamics, chemical kinetics, Navier–Stokes for blood dynamics and Maxwell’s for the propagation of the electric field, etc. However, this is not enough, as the corresponding equations need to be supplemented with constitutive equations that characterize the specific nature of the cardiac tissue at hand. One such relationship relates stresses and strains for the cardiac muscles and accounts for the material properties of the myocardium. Other equations are necessary to characterize the way cardiomyocytes (the elementary components of the cardiac muscle) contract and relax once stimulated by electrical signal propagating through the myocardium. The result is a system of coupled nonlinear differential equations, ordinary differential equations (ODEs), and partial differential equations (PDEs): once properly solved—and this will inevitably call into play computers, as a matter of fact supercomputers!—this
system ought be suitable to simulate the human heart function. The solution of this integrated heart model (IHM) represents a scientifically arduous challenge: the dream is however to provide researchers in applied mathematics, bioengineering, and life sciences, as well as physiologists, cardiologists, and cardiac surgeons, with an effective and powerful tool aimed to the qualitative and quantitative study of the heart, both in physiological or pathological conditions, and at making advances in the diagnosis and treatment of heart diseases.

In this paper, we provide an in-depth and rigorous mathematical presentation of the most crucial aspect of the entire IHM: the so-called electromechanical (EM) mathematical model. We will first derive it from basic principles. In particular, we will highlight the way the activation force triggered by the transmembrane electric potential is generated at the cardiomyocytes level and then affects the whole muscle deformation. We will then recall the main mathematical results inherent to the well-posedness of the electromechanical model components.

The role of data. A mathematical model, if properly designed, enjoys the property of universality, that is its validity is independent of the specific context to which it is applied. Of course, what makes the mathematical model most valuable is its ability to adapt to any specific context: in our case, virtually to every possible individual, thanks to the data that characterize it. Data are needed to feed models. For an IHM, data concern the geometrical shape of a specific heart, initial and boundary conditions that must supplement the differential equations, parameters that characterize material properties, just to mention a few. Data is a crucial issue. Shapes can be retrieved from medical images (such as computer tomography or magnetic resonance imaging) after a good deal of geometric pre-processing. Initial and boundary conditions are seldom available, as they would require invasive (and unnecessary) clinical examinations on patients. Also material parameters are not easy to obtain for specific patients due to the difficulty of producing ad hoc measurements. To compensate for the insufficiency (or lack) of data, mathematicians have historically came up with brilliant ideas such as parameters identification, variational or Kalman filter-based techniques for data assimilation. Moreover, whereas the mathematical model is deterministic in itself, data could be uncertain. It is therefore of paramount importance to quantify how this uncertainty propagates from the input data to the outputs. Dealing with uncertainty in data has required ad hoc, often sophisticated, mathematical tools as well.

Road to solution. Once the differential system is completed by the missing data, closed-form solutions seldom (almost never) exist. The first step is an a priori mathematical analysis to understand in which regimes a solution exists, whether it is unique and if it continuously depends on data (this is the well-posedness in the sense of Hadamard).

Once our problem is set upon rigorous mathematical ground, the next step consists of approximating it by reducing it to a finite dimension, say N (typically very large, of an order of millions or more). A mathematically rigorous strategy consists in projecting the variational formulation (that is the weak formulation) of the differential system onto finite-dimensional subspaces of the Sobolev spaces (typically $H^1$, see, e.g., [3]) to which the weak solution of the original system belongs. To make sure that we have suitably operated, our new finite-dimensional (numerical) problem must enjoy the properties of stability and convergence. Stability is the
continuous dependence (uniformly, with respect to $N$), in suitable norms, of the solution to the problem’s data. Convergence is the property of the numerical solution to tend to the exact (unknown) one once $N$ tends to infinity. Error estimates represent a further peculiar feature of numerical problems. Besides being convergent, the numerical solution should converge to the exact one in suitable norms with a specific rate of convergence with respect to $1/N$. In practice, for models of this kind of complexity, a tradeoff between accuracy and computational effort has to be pursued. Needless to say, as $N$ is unavoidably extremely large (often reaching hundreds of millions), supercomputers are necessary at this stage. Indeed, the simulation of a single heartbeat (about one second of physical time) may require several hours of computation on a big, modern supercomputer.

But what is all that for? From a mathematician’s perspective, mathematical and numerical models of the heart function represent an inexhaustible source of mathematical challenges. From an applied mathematician’s perspective, the excitement is further enhanced by the ambition to provide doctors with a “tool” that may help them with diagnosis and treatment.

In this paper we will illustrate the paradigm (data, mathematical model; numerical approximation, computer simulation), and we will show in a few examples the way we can nowadays provide doctors with solutions of clinical relevance. In this regard cardiac electromechanics provides an excellent testing ground, more and more as the state of the art in mathematical modeling, numerical methods, and computational platforms advances. As a matter of fact, starting from the visionary works of C. S. Peskin [62, 63], this topic has inspired an impressive body of research from many research groups around the world (see, e.g., [10, 11, 17, 21, 25, 32, 37, 39, 57, 63, 83, 85, 86, 90, 91]).

Paper outline. The outline of the paper is as follows. In Part 1 we will give a short introduction to the elementary concepts of cardiac anatomy and physiology. Then we will introduce the mathematical model of cardiac electrophysiology and active force generation. We will then focus on the modeling of cardiac mechanics and will conclude with the multiscale model of cardiac electromechanics. Part 2 will be devoted to the finite-dimensional approximation in space (by the finite element method) and in time. We will then comment on the use of efficient solvers for the associated nonlinear algebraic systems of very large size. Part 3 will then illustrate the numerical solutions that we can achieve, as well as some examples that are relevant to clinical applications. Conclusions and possible generalizations follow.

Part 1. From the physical problem to the mathematical model

2. Cardiac anatomy and physiology at a glance

The human heart is a sophisticated machine, whose functional role is pumping blood throughout the body to its cells, supplying the organs with oxygen, and removing the metabolic waste. The contraction of the heart muscle is due to that of the cardiomyocytes, the cardiac cells, once activated by electrical processes. As a matter of fact, at each heartbeat, an electrical potential propagates through the heart tissue until it reaches each cardiomyocyte of the atria and ventricles (the heart chambers). This synchronized propagation activates an orderly contraction of the heart chambers and generates the heart rhythm. More specifically, each heartbeat is triggered by an electrical signal, originating from the sino-atrial node, the heart’s
Figure 1. (a) The heart is made of four chambers (two atria and two ventricles). The heart contraction is triggered by the sino-atrial (SA) node, from which an electric signal propagates through the atria. Then it reaches the atrioventricular (AV) node and it propagates through the Purkinje fibers, thus reaching all the ventricular cells. (b) The cell membrane of cardiomyocytes is scattered by ion channels, which selectively allow the transit of specific ion species. (c) Among these ionic species, calcium serves as the intracellular messenger by triggering the contraction of sarcomeres. The latter are made of an almost crystalline arrangement of thin filaments and thick filaments. The sliding between the two families of filaments yields to the contraction of the muscle.

natural pacemaker consisting of a group of self-excitable cells that is located in the upper part of the right atrium (see Figure 1). The signal propagates from one cell to another through the two atria and reaches the atrioventricular node, located between the atria and the ventricles. The atrioventricular node acts as a filter for signal propagation to ensure that the contraction of the ventricles begins once the blood has passed from the atria into the same ventricles. The electrical signal then propagates from the atrioventricular node through the Purkinje fibers, a high-conductibility network of fibers, reaching the myocardial cells.

In order to derive a mathematical description of these macroscopic processes, it is necessary to start from a description of the microscopic mechanisms that generate them, following a micro-to-macro approach. This process starts from the elementary components of the cardiac muscle tissue, cardiomyocytes (i.e., cardiac muscle cells). These are in fact excitable: when electrically stimulated, the electrochemical balance of the cell membrane changes, giving rise to a sequence of biochemical processes that determine a significant variation in cell potential: rapid depolarization followed by a repolarization. This phenomenon, known as action potential,
is due to the opening and closing of ion channels, located in the cell membrane. The latter becomes permeable to different ions (calcium, potassium, magnesium) thanks to the transmembrane potential, that is the difference in voltage between the internal and external part of the cell. The ionic fluxes determine a variation of the transmembrane potential and have a feedback effect on the voltage difference itself. Among the various ionic species involved in the dynamics of the action potential, calcium ions play an important role. Calcium represents the trigger for muscle contraction: calcium ions induce a complex chain of reactions generating an active force inside the cardiomyocytes. Finally, thanks to a process of transmission through different space-time scales, the active force at the microscopic scales (that of cardiomyocytes) generates a resultant force at the organ level, thus giving rise to the contraction of ventricles and atria. The combined effect of active and passive force (i.e., the reaction of the myocardium to mechanical stress), as well as the coordinated action of the atria and ventricles, governs the blood fluid dynamics in the four chambers as well as the valve dynamics. The building blocks of this complex network of interactions are summarized in Figure 2. In Sections 3, 4, 5, and 6 starting from the microscale and progressively climbing the hierarchy of scales up to the macroscale, we present mathematical models describing the above mentioned processes.

3. Modeling cardiac electrophysiology

As mentioned, the driver of the cardiac function is electrophysiology, i.e., the result of chemical and electrical processes taking place at different spatial scales, from subcellular to the whole organ scale. A mathematical description of these processes is based on the translation into mathematical terms of the principles ruling the electro-chemical activity of ions species at the finest scale; then, by progressively climbing up the hierarchy of spatial scales, a set of equations describing the tissue-level electrophysiological activity is derived.

Let $\Omega_0 \subset \mathbb{R}^3$ be an open connected set, denoting the region of space occupied by the cardiac tissue. This region can be split into two subregions, namely the inner space of cells (intracellular space) and the region located outside the cell membranes (extracellular space). The intracellular and extracellular spaces are separated by the cell membranes, which are scattered by ion channels. An ion channel can be understood as an opening in the cell membrane that selectively allows the transit of specific ion species. The movement of electrically charged ions generates an electric current. Starting from Maxwell’s laws of electromagnetism and Einstein’s theory on Brownian motion [29], one can derive a system of PDEs describing the ion dynamics. An analytical solution of this system can be obtained under the (realistic) assumption that the ion dynamics in the direction tangential to the cell membrane is negligible compared to the transmembrane one. By further neglecting the space variations across the membrane and focusing exclusively on space-averaged equations, this allows us to link the electric current crossing the membrane (denoted by $I_{\text{ion}}$) with the concentration of ionic species and with the transmembrane potential, defined as $v = u_i - u_e$, where $u_i$ and $u_e$ denote the electric potential in the intracellular and extracellular spaces, respectively.

The membrane of cardiomyocytes is selectively permeable to multiple ionic species, such as sodium, potassium, and calcium, whose fluxes are regulated by the opening and closing of the ion channels. To track the opening and closing of these channels,
as well as the resulting electric current, several models have been proposed in the literature. These models are written in the general form

$$\frac{d\mathbf{z}_{\text{ion}}}{dt} = \Phi_{\text{ion}}(v, \mathbf{z}_{\text{ion}}), \quad I_{\text{ion}} = \mathcal{I}_{\text{ion}}(v, \mathbf{z}_{\text{ion}}),$$

where $\mathcal{I}_{\text{ion}}$ and $\Phi_{\text{ion}}$ are suitably defined functions and $\mathbf{z}_{\text{ion}}(t) \in \mathbb{R}^{N_{\text{ion}}}$ is a vector collecting the so-called ionic variables, describing ion channels, concentrations, or simply phenomenological variables. For the latter, a typical example concerns recovery variables, namely variables that account for the tissue refractoriness, without a one-to-one relationship with measurable physical quantities. Ideally, system (3.1) applies at any cell location. Notable examples are provided by the Aliev–Panfilov model \[4\], which envisages a single ionic variable (i.e., $N_{\text{ion}} = 1$) that accounts

Figure 2. The cardiac function is the result of several physical processes acting in concert, including electrophysiology, active and passive mechanics, perfusion, fluid dynamics, and valve dynamics. These physical processes are tightly coupled together; they are visually conveyed being grouped in the same box. We also highlight the way the integrated heart model of the left ventricle interacts with external circulation. Specifically, red arrows denote oxygenated blood fluxes, blue arrows denote de-oxygenated blood fluxes. Solid arrows represent the main blood pathways (namely, systemic and pulmonary circulations), while dashed arrows represent coronary circulation, consisting in the blood vessels that perfuse the myocardium itself.
for tissue refractoriness; the Ten Tusscher–Panfilov model [89], which includes 18 variables describing ionic concentrations and the probability of opening of specific ion channels; the O’Hara–Rudy model [59], featuring $N_{\text{ion}} = 41$ variables, which was built and calibrated upon data coming from more than 100 undiseased human hearts.

To complete a model describing the electrical activity across the cell membrane, we need to relate the total ionic current $I_{\text{ion}}$ to the dynamics of the transmembrane potential $v$. With this aim, the cell membrane is modeled as a capacitor with capacity $C_m$: when a current $I$ crosses the cell membrane for an infinitesimal time interval $dt$, the transmembrane potential difference decreases by $I/C_m dt$ (by convention, current and potential difference are taken with opposite signs).

Furthermore, equation (3.1) describes the electrical activity of a single cell. However, in the cardiac tissue, the nearly three billion cells are not electrically isolated. Indeed, they are connected by the so-called gap junctions, which link the intracellular and extracellular spaces of adjacent cells and which allow the electrical signal to propagate from one cell to another. The following equation, known as the monodomain model [24], accounts for these effects by means of a diffusive term:

\[
\chi_m C_m \partial_t v - \nabla \cdot (D \nabla v) + \chi_m I_{\text{ion}} = I_{\text{app}} \quad \text{in } \Omega_0 \times (0, T].
\]

In the above equation $I_{\text{app}}$ represents an externally applied current, while $\chi_m$ denotes the area-to-volume ratio (i.e., the amount of membrane area per volume of tissue), and it allows us to relate the microscopic cell dynamics to quantities related to a macroscopical characterization of the tissue. The tensor $D$, known as the conductivity tensor, characterizes the electrical properties of the tissue. A feature of cardiac tissue is that the electrical signal propagates with a larger speed in the fibers’ direction (i.e., $f_0$) than in the orthogonal directions. To account for such an anisotropic property, $D$ is typically defined as

\[
D = \sigma_f f_0 \otimes f_0 + \sigma_s s_0 \otimes s_0 + \sigma_n n_0 \otimes n_0,
\]

where $\sigma_f$, $\sigma_s$, and $\sigma_n$ represent the conductivity coefficients in the fibers, sheets, and cross-fibers directions. The triplet $(f_0, s_0, n_0)$ not only characterizes the electrical behavior of the tissue, but also its mechanical response, as we will describe in Section 5.

The applied current $I_{\text{app}}$ is typically prescribed as a function with support restricted to a short time interval (usually a few milliseconds) at the beginning of each heartbeat and to a few spots located close to the surface of the myocardium, representing the end points of the Purkinje fibers. The typical solution of (3.2) is in the form of a traveling wave for the variable $v$, originating from the locations where $I_{\text{app}}$ is applied. When a point of the domain is reached by the wavefront, $v$ quickly raises (depolarization), then features a plateau, and finally returns to its resting position (repolarization); this yields the so called action potential. Due to the quick depolarization, the potential wave is characterized by a very steep front.

The monodomain model can be derived from the bidomain model, in which two separate variables are associated with the electric potential of the intracellular and extracellular space. In its turn, the bidomain model can be rigorously derived by assuming the existence of the extracellular and intracellular spaces as two simply connected subdomains and by employing a homogenization technique [24].

A general comment is in order. Even when it is derived from physical principles, the mathematical meaningfulness of a model is not guaranteed a priori yet.
Specifically, one should check that the problem is well-posed from a mathematical standpoint, that is, a solution exists, it is unique, and it depends continuously on data (the importance of the latter assumption is related to the uncertainty that unavoidably affects the measurements of these data). Moreover, the existence of solutions is an essential requirement to address the numerical approximation of a model. As a matter of fact, even if approximate solutions exist, they would be meaningless in the absence of an exact solution to which they can converge. In this regard, we report that a well-posedness result for the bidomain model coupled with the FitzHugh–Nagumo model [33] is presented in [34]. Another well-posedness result, that makes use of a fixed-point argument, is provided in [33].

4. Modeling active force generation

Due to the action potential dynamics, in the first stages of each heartbeat the calcium concentration inside cardiomyocytes (denoted by $[\text{Ca}^{2+}]_i$, where $i$ stands for intracellular) quickly raises by nearly one order of magnitude before returning to its resting value in nearly a half second. In fact calcium acts as a cellular messenger by triggering the contraction of cardiomyocytes. At the microscopic scale, cardiomyocytes are organized in sarcomeres, cylinder shaped contractile elements of the size of nearly 2 $\mu$m. Sarcomeres are composed of myofilaments, thin filaments made of the proteins troponin, tropomyosin, and actin, and thick filaments made of the protein myosin. The active force is generated by the interaction between actin and myosin, molecular motors capable of transforming the chemical energy into mechanical work [47]. The long-standing hypothesis that muscle contraction was led by folding of elongated protein filaments was challenged by the discovery that the filaments’ length remains constant during contraction and that it is instead the mutual sliding between the two families of filaments (thin and thick) that makes the muscle contract [45,46]. This discovery, known as sliding filaments theory, was made independently by two research teams: on one side, the British biologist H. Huxley and biophysicist J. Hanson, working at MIT; on the other, the British physiologist A. F. Huxley (Nobel prize winner in 1963 for his work on the action potential) and the German physician R. Niedergerke, working at the University of Cambridge. The two teams decided to publish their work in two consecutive articles in the same issue of Nature [45,46].

Many models describing these microscopic mechanisms have been proposed in the literature. In general terms, these force generation models are written in the form

$$\frac{dz_{\text{act}}}{dt} = \Phi_{\text{act}} \left( z_{\text{act}}, [\text{Ca}^{2+}]_i, SL, \frac{dSL}{dt} \right), \quad T_a = \eta(z_{\text{act}}),$$

where $z_{\text{act}}(t) \in \mathbb{R}^{N_{\text{act}}}$ denotes a vector collecting the state variables associated with the dynamics of the contractile proteins and $T_a$ denotes the generated active force. The state variables $z_{\text{act}}$ may be associated either with the probability of the regulatory proteins of being in a given state [51,70,78,95], or with the elongation of the contractile proteins [72,97], or finally they may be phenomenological variables [51,55]. The dynamics of $z_{\text{act}}(t)$ is driven by that of the calcium concentration ($[\text{Ca}^{2+}]_i$) and of the sarcomere length, denoted by $SL$. As a matter of fact, when the sarcomere is elongated (i.e., for large $SL$), the generated force tends to increase; moreover, when the sarcomere is shortening (i.e., when $dSL/dt < 0$), the generated force is lower than in steady-state conditions (i.e., when $dSL/dt = 0$).
In recent years, many efforts have been devoted to the development of force generation models [50, 56, 70, 77, 80]. The main challenge facing modelers is to describe the complex subcellular processes underlying force generation in a compact way [74, 77]. In fact, despite several models that describe in a biophysically detailed manner the force generation mechanisms that have been proposed, these are not suitable for numerical simulations because of the overwhelming computational cost associated with their numerical approximation. These difficulties are mainly related to two reasons.

The first reason is the stochastic nature of the processes in play, due to the scale at which they occur (the spatial scales are so small that the effects of thermal agitation cannot be overlooked) [81]. For this reason, the output of a force generation model is not a deterministic solution, but rather a probability distribution on the space of the trajectories of the system. Clearly, this dramatically increases the dimensionality of the problem.

The second reason lies in the nearest-neighbor interactions taking place between the proteins that make up the sarcomere [48, 77]. Because of these interactions, in fact, mean-field approximations, which are widely used in the context of molecular dynamics to lower the computational burden of numerical simulations, cannot be adopted without compromising the validity of the results. In fact, these nearest-neighbor interactions are at the basis of an important phenomenon, namely the cooperativity of contractile units, for which the calcium-activated units favor the activation of neighboring units, thus generating a highly nonlinear relationship between calcium concentration and generated force. These cooperative processes, of fundamental importance for an efficient functioning of the organism, cannot be grasped by mean-field models, in which a single representative unit is considered and the other units are taken into account through the average effect they have on the representative unit [77].

To give a concrete example, let us consider the model proposed in the seminal work [78], in which the authors demonstrated the importance of a spatially explicit description of the sarcomere proteins. In this model, each contractile unit is described by a 4-state Markov chain. Considering a single myofibril with \( N = 32 \) units, we have \( 4^N \approx 2 \cdot 10^{19} \) possible states. Hence, to characterize a probability distribution over this set, we need an order of \( 10^{19} \) variables. In conclusion, to simulate on a computer the dynamics of this stochastic model, more than \( 10^5 \) petabytes (i.e., \( 10^{20} \) bytes) would be required just to store the system state in the computer memory. This corresponds to more than 30000 times the storage capacity of the largest supercomputer in the world (up to June 2020). This situation is not uncommon: a mathematical model is available; however, its practical interest is very limited because of the overwhelming computational cost of its numerical approximation. In these cases, suitable modeling assumptions and/or mathematical tools are employed to derive an approximate, yet computationally feasible, reduced mathematical model.

In the case at hand, in order to capture the cooperative effects without explicitly tracking the joint probability of units of the whole filament, several strategies have been proposed in the literature. Among these, we mention Monte Carlo sampling techniques [96, 97] and the reduction of the number of unknowns obtained by grouping together suitable subsets of the states [18, 49, 95]. Alternatively, in

---

the transition rates of the Markov chain were expressed as nonlinear functions of the calcium concentration, to phenomenologically reproduce the cooperative behavior. In [71,72], we introduced a physically motivated assumption of conditional independence of the stochastic processes associated with units that are far from each other along the filament, given the state of the intermediate units. With this assumption, we can describe the protein dynamics through a nonlinear system of nearly 2000 ODEs, thus allowing us to numerically approximate the solution of one heartbeat in just a few seconds of computational time.

A different approach is that of phenomenological models [44,50,55,79]: these do not derive equations from first principles, rather they are built by fitting the measured data with simple laws, chosen a priori by the modeler. The numerical solution of these models, typically expressed as systems of few ODEs, feature a lower computational cost than physics-based models. However, they provide a poorer insight than physically detailed models and they are characterized by a hampered predictive power under experimental conditions different from those under which they are built.

5. Modeling cardiac mechanics

During each heartbeat, the heart muscle undergoes large deformations (up to few centimeters). To model the myocardium displacement, the strain of the tissue must be related to the internal stress induced by cardiomyocytes’ contraction and to the pressure exerted by the blood onto the endocardium, the internal surface of the cardiac chambers. The conceptual framework to describe this phenomenon is continuum mechanics, of which we recall basic notions in the following. For more detail, we refer the interested readers to, e.g., [8,58].

Let \( \Omega_0 \subset \mathbb{R}^3 \) be an open connected set, which we denote as the reference (or undeformed) configuration, designating the region of space occupied by an elastic body at rest. In our setting this will be the region occupied by the heart muscle, say at the end of the diastole (the time \( t = 0 \) of our simulations). We consider a time-dependent deformation map \( \varphi : \Omega_0 \times [0,T] \to \mathbb{R}^d \), such that \( \mathbf{x} = \varphi(\mathbf{X},t) \) (spatial coordinate) represents the position occupied by the point \( \mathbf{X} \in \Omega_0 \) (material coordinate) at time \( t \). We thus define the displacement field \( \mathbf{d}(\mathbf{X}) := \varphi(\mathbf{X}) - \mathbf{X} \) and the deformation gradient tensor \( \mathbf{F}(\mathbf{X},t) = \nabla \varphi(\mathbf{X},t) = \mathbf{I} + \nabla \mathbf{d} \), where \( \mathbf{I} \) denotes the identity tensor and \( \nabla \) is the gradient operator in the reference (material) coordinate. The deformation map is assumed to be smooth enough (typically twice continuously differentiable, but weaker regularity is admitted), injective, and orientation preserving (i.e., its Jacobian \( J = \det \mathbf{F} > 0 \) for any \( \mathbf{X} \in \Omega_0 \)). By denoting by \( \text{Lin} \) the vector space of the linear transformations from \( \mathbb{R}^d \) into itself, let us introduce the subset \( \text{Lin}^+ := \{ \mathbf{A} \in \text{Lin} \text{ s.t. } \det \mathbf{A} > 0 \} \).

By Newton’s second law, the displacement field \( \mathbf{d}(t) \) satisfies the momentum balance equation,

\[
\rho \partial_{tt} \mathbf{d} - \nabla \cdot \mathbf{P} = \mathbf{h} \quad \text{in } \Omega_0 \times (0,T],
\]

endowed with suitable boundary conditions, where \( \rho \) is the mass density of the body and where \( \mathbf{h} \) represents an externally applied load (force per unit volume). The tensor \( \mathbf{P} \), known as the first Piola–Kirchhoff stress tensor (or simply the Piola stress tensor), encodes the internal stresses of the body. For the sake of model closure, the stress tensor \( \mathbf{P} \) is defined via a constitutive law, which is a relationship
linking the state of strain of the body with its state of stress. This law can possibly
depend on the rate of strain (e.g., in the case of visco-elastic materials). For elastic
materials the stress tensor can be written in terms of the strain, i.e., $P = P_F(F)$. Here we focus on hyperelastic materials, characterized by a strain energy density $W$, such that $\int_{\Omega_0} W(X) dV_0$ provides the total elastic energy stored by the body as a consequence of the deformation, where $W(X) = W_F(F(X))$ for some $W_F : \text{Lin}^+ \to \mathbb{R} \cup \{+\infty\}$. By definition, hyperelastic materials are such that

\begin{equation}
(5.2) \\
P = \frac{\partial W}{\partial F}.
\end{equation}

The passive mechanical response of the heart is significantly anisotropic, due to the presence of fibers. Many anisotropic constitutive laws have been proposed in the literature to describe the cardiac tissue (see, e.g., [39,40,43,92]), accounting for the different elastic response along the three directions $f_0, s_0, a n d n_0$, the three preferential directions that also characterize the electrical diffusivity of the cardiac tissue (see Section 3). These energies typically feature an exponential dependence on the strain to model the large stiffening of the tissue when it is over-stretched. As an example, the material model of [92] is defined by the strain energy density function

\begin{equation}
(5.3) \\
W = \frac{C}{2} \left( e^Q - 1 \right) + \frac{B}{2} (J - 1) \log J,
\end{equation}

where $C > 0$ is a constant and $E_{ab} = E a_0 \cdot b_0$, for $a, b \in \{f, s, n\}$, are the entries of the Green–Saint Venant tensor $E = \frac{1}{2}(F^T F - I)$ in the $(f_0, s_0, n_0)$ frame of reference. The constant $B > 0$ is the bulk modulus which weights the volumetric term $\frac{1}{2}(J - 1) \log J$ by penalizing those deformations that would lead to a change of volume occupied by the tissue ($J = \text{det} F \neq 1$). The latter term leads to a quasi-incompressible formulation, as little volume variations are allowed. The strongly incompressible formulation is an alternative, in which the balance of the momentum equation (5.1) is coupled with the constraint $J = 1$, and the Piola stress tensor is redefined as $P = \frac{\partial W}{\partial F} - pJ F^{-T}$, thus yielding a saddle-point problem, wherein the pressure $p$ acts as a Lagrange multiplier.

The cardiac tissue is in fact an active material. This means that its internal stress is not uniquely identified by strain, but rather it can be produced by microscopic mechanisms that turn the chemical energy of ATP into mechanical work. The models presented in Section 4 describe these processes at the microscale and allow to obtain $T_a$, a scalar that represents the magnitude of active force per unit area. This quantity must be related to the macroscopic balance of the momentum equation (5.1), by writing the Piola stress tensor as

\begin{equation}
(5.4) \\
P = P^{\text{pass}} + P^{\text{act}},
\end{equation}

namely as the sum of a passive term (given by $P^{\text{pass}} = \frac{\partial W}{\partial F}$) and an active term $P^{\text{act}}$ which must be constitutively defined by suitably upscaling the microscopically generated stress. In particular, if we suppose that cardiac muscle fibers, aligned along $f_0$, generate a force per unit area of magnitude $T_a$, directed along the fibers’ direction in the current configuration $f := \frac{Ff_0}{|Ff_0|}$, we get

\begin{equation}
(5.5) \\
P^{\text{act}} = T_a \frac{Ff_0 \otimes f_0}{|Ff_0|}.
\end{equation}
Alternative to the *active stress* approach \[5,4\] is the *active strain* one, wherein activation is modeled as a prescribed strain \[5,6\]. We only consider here the active stress approach, which is by far the most widely used in the literature.

For the mathematical well-posedness of nonlinear elasticity models, we refer interested readers to \[12,53\]. Existence results require suitable hypotheses on the energy density function \(W\). The celebrated existence result by J. Ball \[12\], based on the notion of polyconvexity (a weaker notion than convexity) and on suitable growth and regularity conditions, is based on the direct method of calculus of variations \[27\]. These results, derived for passive materials, cannot be directly applied to heart mechanics, as the cardiac tissue is in fact an active material. The well-posedness of active strain models has been studied in \[6\], while hypotheses for the well-posedness of active stress models have been derived in \[70\].

6. Modeling blood circulation

The balance of momentum equation \(5.1\) must be supplemented with suitable boundary conditions to account for the forces acting between the myocardium and the surrounding environment. In particular, the internal surfaces of cardiac chambers (the *endocardium*) are in contact with the blood, which exerts pressure. Since the blood dynamics in the chambers is tightly related to that of the whole vascular network, it is necessary to develop models of the entire circulatory system. For this purpose, several blood circulation models, with different degrees of accuracy, have been proposed. These range from three-dimensional fluid-structure interaction (FSI) models, where the blood flow in blood vessels is described by the Navier–Stokes equations \[62,65,87,88,94\], to zero-dimensional models (whose variables only depend on time, but not on spatial coordinates) \[65,75\].

In the latter family of models, also known as lumped parameters models, the circulatory system is split into a finite number of compartments and an average pressure is associated with each of them. Then, equations describing the dynamics of these pressure variables and of the blood volume contained in each compartment are derived by the principles of conservation of mass and momentum. Lumped parameters models may be limited to a subset of the circulatory systems \[19,36\], or may describe the whole closed-loop cardio-circulatory system \[16,42,75\]. In general terms, they are written as systems of ODEs, in the form

\[
\frac{dz_{\text{circ}}}{dt} = \Phi_{\text{circ}} (z_{\text{circ}}, t),
\]

where the vector \(z_{\text{circ}}(t) \in \mathbb{R}^{N_{\text{circ}}}\) collects the state variables (pressure and volumes). The right-hand side may depend on the time variable, to account for the different phases of the cardiac cycle. A concrete example of \(6.1\) is available in \[75\]. The zero-dimensional model of blood circulation \(6.1\) must be suitably coupled with the three-dimensional model of cardiac mechanics (see Section 5). With this aim, we require that the volume enclosed by the ventricular cavity when the domain \(\Omega_0\) is moved by the displacement \(d\) (which we denote by \(V_{3D}^{LV}(d)\)) equals the one predicted by the zero-dimensional circulation model (which we denote by \(V_{0D}^{LV}(z_{\text{circ}})\)). This coupling is enforced via a Lagrange multiplier, which, in the case at hand, represents the blood pressure inside the ventricular cavity \(p_{LV}\) (more details will be provided in Section 7).
An intermediate class of models for vessel circulation consists of 1D models, in which only the main direction of blood flow is explicitly represented and the equations are averaged in the orthogonal directions. The development of 1D models for blood circulation owes its origin to Leonhard Euler, back in 1775 [30]. The mathematical models of the circulatory system also need to account for the dynamics of the four cardiac valves. The latter regulate blood flow between the atria and the corresponding ventricles, between the right ventricle and the pulmonary artery, and between the left ventricle and the ascending aorta. Their sudden opening and closing is driven by the pressure gradient. The motion of cardiac valves can be described with different levels of detail, ranging from three-dimensional FSI models [61, 63] to zero-dimensional models [16]. For the sake of space, three-dimensional and one-dimensional circulation models will not be further discussed in this paper.

7. A FULLY COUPLED CARDIAC ELECTROMECHANICS MODEL

The mathematical models discussed in Sections 3, 4, 5, and 6 describe different physical processes occurring at different spatio-temporal scales along each heartbeat, which cannot however be understood independently of one another. These are indeed connected though a complex and fascinating web of interactions, feedback loops, and self-regulation mechanisms aimed at preserving the physiological working regime of the heart and responding to external stimuli (see Figure 3). In mathematical terms, in order to describe the whole cardiac function, these models must be coupled to yield a unique system of PDEs and ODEs.

Let us consider for simplicity a single chamber, namely the left ventricle. Hence, the open connected domain $\Omega_0 \subset \mathbb{R}^3$ represents the region of space occupied by the left ventricle at rest (e.g., at the end of the diastole). Its boundary is split into $\partial \Omega_0 = \Gamma_{\text{endo}}^0 \cup \Gamma_{\text{epi}}^0 \cup \Gamma_{\text{base}}^0$, where $\Gamma_{\text{endo}}^0$, $\Gamma_{\text{epi}}^0$, and $\Gamma_{\text{base}}^0$ are disjoint sets respectively denoting the endocardial and epicardial surfaces and the ventricular base, namely the artificial boundary located where the left ventricle geometry is

![Figure 3. Diagram illustrating our multiphysics cardiac electromechanics model. The 3D left ventricle picture has the purpose of displaying the finite-element grid as well as the different boundary components.](image-url)
cut (see Figure 3). Then, the fully coupled electromechanical problem consists in finding

\begin{equation}
\begin{aligned}
v &: \Omega_0 \times [0, T] \to \mathbb{R}, \\
d &: \Omega_0 \times [0, T] \to \mathbb{R}^3, \\
z_{\text{ion}} &: \Omega_0 \times [0, T] \to \mathbb{R}^{N_{\text{ion}}}, \\
z_{\text{act}} &: \Omega_0 \times [0, T] \to \mathbb{R}^{N_{\text{act}}}, \\
z_{\text{circ}} &: [0, T] \to \mathbb{R}^{N_{\text{circ}}}, \\
p_{LV} &: [0, T] \to \mathbb{R}
\end{aligned}
\end{equation}

such that we have, in \( \Omega_0 \times (0, T] \),

\begin{equation}
\begin{aligned}
\chi_m C_m \partial_t v - \nabla \cdot \left( J F^{-1} D F^{-T} \nabla v \right) + \chi_m f_{\text{ion}}(v, z_{\text{ion}}) &= I_{\text{app}}, \\
\rho \partial_t d - \nabla \cdot P &= 0, \\
\partial_t z_{\text{ion}} &= F_{\text{ion}}(v, z_{\text{ion}}), \\
\partial_t z_{\text{act}} &= F_{\text{act}}(z_{\text{act}}, [Ca^{2+}]_i(z_{\text{ion}}), SL(d), \partial_t SL(d))
\end{aligned}
\end{equation}

with boundary conditions

\begin{equation}
\begin{aligned}
(J F^{-1} D F^{-T} \nabla v) \cdot N &= 0 &\text{on } \partial \Omega_0 \times (0, T], \\
PN + K_{\text{epi}} d + C_{\text{epi}} \partial_t d &= 0 &\text{on } \Gamma_0^{\text{epi}} \times (0, T], \\
PN &= -p_{LV} J F^{-T} N &\text{on } \Gamma_0^{\text{endo}} \times (0, T], \\
PN &= p_{LV} |J F^{-T} N| &\frac{\int_{\Gamma_0^{\text{endo}}} J F^{-T} N d\Gamma_0}{\int_{\Gamma_0^{\text{base}}} |J F^{-T} N| d\Gamma_0} &\text{on } \Gamma_0^{\text{base}} \times (0, T],
\end{aligned}
\end{equation}

and coupled with the following equations in \( (0, T] \),

\begin{equation}
\frac{d z_{\text{circ}}}{dt} = \Phi_{\text{circ}}(z_{\text{circ}}, p_{LV}, t),
\end{equation}

with initial conditions (in \( \Omega_0 \times \{0\} \))

\begin{equation}
v = v_0, \\
d = d_0, \\
\partial_t d = 0, \\
z_{\text{ion}} = z_{\text{ion},0}, \\
z_{\text{act}} = z_{\text{act},0}, \\
z_{\text{circ}} = z_{\text{circ},0}.
\end{equation}

We remark that the displacement of the myocardium (encoded in the variable \( d \)) affects the conductivity properties of the tissue (thanks to the so-called mechano-electrical feedback), for which (3.3) is rephrased in the deformed configuration, as the tensor

\begin{equation}
D = \sigma_f \frac{F_{f_0} \otimes F_{f_0}}{|F_{f_0}|^2} + \sigma_s \frac{F_{s_0} \otimes F_{s_0}}{|F_{s_0}|^2} + \sigma_n \frac{F_{n_0} \otimes F_{n_0}}{|F_{n_0}|^2},
\end{equation}

namely the pullback of the conductivity tensor in the reference configuration \( \Omega_0 \) is replaced into (3.2).

The state variable of the ionic model (i.e., \( z_{\text{ion}} \)) provides the value of \([Ca^{2+}]_i\) at each point of the computational domain and at each time \( t \), which is an input for the force generation model to be solved at each point in \( \Omega_0 \). To provide the missing input to the latter model, namely the sarcomere length and its time derivative, we observe that the elongation of sarcomeres can be obtained as \( SL(d) = SL_0 |F_{f_0}| \), where \( SL_0 \) is the reference sarcomere length.

On its turn, the state variable of the force generation model (i.e., \( z_{\text{act}} \)) provides the magnitude of \( T_a \), the active force generated at a given point of \( f_0 \) and at a given time \( t \in [0, T] \), which is employed within the equation describing the mechanics of the myocardium. The associated boundary conditions model the interaction of the cardiac tissue with the surrounding tissue. Specifically, the boundary condition imposed on \( \Gamma_0^{\text{epi}} \) encodes an elastic (\( K_{\text{epi}} \)) and viscous (\( C_{\text{epi}} \)) interaction with the
pericardium, a thin membrane enclosing the heart. Conversely, the boundary condition enforced on $\Gamma_{\text{endo}}^0$ models the stress exerted by the blood contained in the cavity, whose pressure—tracked by the variable $p_{\text{LV}}$—is assumed to be constant in space. The variable $p_{\text{LV}}$ acts in fact as a Lagrange multiplier, enforcing the geometrical compatibility condition $V_{\text{LV}}^3(d) = V_{\text{LV}}^0(z_{\text{circ}})$ at each time $t \in (0, T]$. Finally, the boundary condition imposed on $\Gamma_{\text{base}}^0$—known as energy-consistent boundary condition—accounts for the stress exerted by the upper portion of the heart (that we have deliberately disregarded for the sake of simplicity) on $\Omega_0$, consistently with the principles of energy conservation (see [73] for a thorough derivation). We remark that different sets of boundary conditions, representing different means of interaction with the surrounding tissue, could be considered [36,73].

Despite the large amount of work on cardiac electromechanical models, there are still many open questions about the mathematical well-posedness of these systems of equations. In [60] the authors studied the existence of solutions for the fully coupled electromechanical model, considering different kinds of mechano-electrical feedback in an active stress setting. Concerning the active strain setting, instead, in [7], the authors proved existence of weak solutions and uniqueness of regular solutions for a simplified electromechanical model, where the nonlinear elastic model is linearized and where the FitzHugh–Nagumo ionic model is employed. In [15] these results have been generalized to the more realistic Beeler–Reuter and Luo–Rudy ionic models. The latter results are based on the Faedo–Galerkin method and compactness arguments. (A similar approach, named the Galerkin method, stands also at the basis of the numerical approximation and will be outlined in Section 9.)

Part 2. Numerical approximation

8. Why should we turn to numerical approximation

The electromechanical model (7.1)–(7.5) is in principle able to encode the EM heart function. As a matter of fact, reality is more complex than that. Indeed, (7.1)–(7.5) has been proposed for the left ventricle. A model with a similar mathematical structure can be derived for the right ventricle, as well as for each one of the two atria. The complete EM model would require using different constitutive laws for the atria (their tissue is much thinner than in ventricles), accounting for the electrical signal activation (from the sino-atrial node to the sino-ventricular one, then spreading through the whole myocardium), and coupling the four chambers together; see [65]. In this work, we focus only on the left ventricle model (7.1)–(7.5), as the associated mathematical considerations will highlight what is relevant for the complete model too.

There are a few other issues that need to be addressed, though. This is precisely where applied mathematicians must put their hands into the “real matter”: How do we generate nonhomogeneous boundary data? How do we provide an initial state to our system (i.e., initial data for our initial boundary value problem)? How do we prescribe a patient-specific problem’s coefficients (like fibers orientation) without subjecting a patient to unnecessary expensive and invasive clinical exams and measurements? Here is where we must depart from the beauty and lightness of PDE theory where problems are invariably formulated under the following incipit: “Let us consider the following PDE with regular enough boundary conditions given by (some Dirichlet or Neumann data) and initial conditions (some suitable initial
functions). As a matter of fact, boundary and initial functions are often missing for a living system, and even when available they are often incomplete and discrete (that is available only at few selected points) their “regularity” is hard to attribute, actually it might not even be appropriate to talk about functional spaces for boundary or initial data. In practice, this shortcoming (a troublesome issue) is overcome by resorting to suitable mathematical procedures aimed at providing part of the missing data “coherently” with the original physical problem. See for instance [65] and [69]. See also Sections 11 and 13 where some of these issues will be addressed in the context of our numerical simulations.

From a theoretical standpoint, the first and foremost question we should address is whether (7.1)–(7.5) is well-posed. To our knowledge, a rigorous existence and uniqueness theorem, under realistic assumptions on boundary data, initial conditions, and the problem’s coefficients, is not provable as yet. Partial results do exist concerning the electrophysiology subproblem (7.2) with suitable boundary conditions [34,93], as well as for the passive mechanics problem (5.1) under suitable boundary conditions [12,27]. See [7,15,60] for existence results on the coupled electromechanical model under simplifying assumptions. This may raise questions about the actual solvability of problem (7.1)–(7.5) and suggests a cautious attitude to mathematicians. On the other side, if we are confident—based on the derivation of (7.1)–(7.5) from physics first principles and from biophysically motivated assumptions—that this model represents a faithful description of the reality, then reality suggests to us that “a solution” should exist, and is unique, under physiological (living) conditions. The above heuristic considerations prompt us to try to find numerical solutions (exact, analytic solutions in closed form are impossible to obtain, unless under severe, unrealistic simplistic assumptions).

Generally speaking, the approximation of an initial boundary value problem (IBVP) (either scalar, or vectorial, as the case of (7.1)–(7.5)) can be operated following different strategies. To avoid unnecessary (and lengthy) generalization, here we will stick with sequential approximations, first in space and next in time (simultaneous space-time approximations can be pursued as well).

9. SPACE DISCRETIZATION

Space approximation has the aim of reducing the given IBVP to a system of ODEs (i.e., an IVP). This again can be operated according to many different paradigms. Here we will use the Galerkin method that consists in projecting the weak (variational) formulation of our IBVP at every time \( t \) into a finite dimensional subspace of the (Sobolev) space, say \( V \), to which the exact (weak) solution belongs. More specifically we will use the Galerkin finite element (FE) method for which the finite-dimensional space, say \( V_h \), is made of piecewise polynomial functions (either continuous or discontinuous) on a partition of the computational domain into simplexes (triangles in 2D, tetrahedra in 3D); see, e.g., [22,68]. The subindex \( h \) denotes the characteristic diameter of the simplexes: the smaller the \( h \), the finer the FE partition, the higher the dimension, say \( N_h \), of the vector space \( V_h \), the more accurate (in principle) the numerical solution. Accordingly, we will denote by \( y_h(\mathbf{X},t) \) the FE approximation of the solution \( y(\mathbf{X},t) \) of the IBVP. By suitably defining a basis of the vector space \( V_h \), say \( \{ \varphi_i \}_{i=1}^{N_h} \), the numerical solution is expressed as a linear combination of this basis through \( N_h \) (unknown) coefficients that depend on the time variable \( t \), that is to say \( y_h(\mathbf{X},t) = \sum_{i=1}^{N_h} \varphi_i(\mathbf{X}) y_{i,h}(t) \), where...
\[
y(t) = (y_{1, h}(t), \ldots, y_{N_h, h}(t))^T \in \mathbb{R}^{N_h} \text{ for all } t \in [0, T].
\]
By computing the result of this projection against the \(N_h\) basis functions, after simple algebraic manipulation we end up with a system of \(N_h\) differential equations for the unknown coefficients, known as the semidiscretized problem,

\[
\begin{cases}
\mathcal{M} \frac{d\mathbf{y}}{dt}(t) + A \mathbf{y}(t) + \mathbf{F}(\mathbf{y}(t)) = \mathbf{f}(t) & \text{ in } (0, T], \\
\mathbf{y}(0) = \mathbf{y}_0,
\end{cases}
\tag{9.1}
\]
where \(A \in \mathbb{R}^{N_h \times N_h}\), \(\mathcal{M} \in \mathbb{R}^{N_h \times N_h}\), and \(\mathbf{F} : \mathbb{R}^{N_h} \to \mathbb{R}^{N_h}\) are three operators corresponding to the linear, mass, and nonlinear terms of the IBVP, respectively.

We recall that the electromechanical model considered in this paper is in fact a multiphysics problem, being made of several subproblems coupled together to form problem (7.1)–(7.5). In this setting, the semidiscretized problem (9.1) is still valid, provided that the unknowns vector \(\mathbf{y}\) collects the coefficients associated with the FE discretization of all the unknowns of problem (7.1)–(7.5).

10. Time discretization

At this stage we need to perform a further approximation: the time-derivatives are discretized using suitable differential quotients (i.e., Taylor developments), after having introduced a suitable discretization of the time interval into subintervals of (say, uniform) length \(\Delta t\). The smaller \(\Delta t\), the finer the time interval partition, the more accurate (with respect to \(\Delta t\)) the expected solution. The result is a system of \(N_h\) nonlinear algebraic equations that need to be solved at every time step.

Let us partition for simplicity the time interval \([0, T]\) in \(N_t\) time instances \(t^n = n \Delta t\), for \(n = 0, \ldots, N_t\), where the time step size is \(\Delta t = \frac{T}{N_t}\). We indicate with \(\mathbf{y}^n = (y_{1,h}^n, \ldots, y_{N_h,h}^n)^T \in \mathbb{R}^{N_h}\) the approximation of \(\mathbf{y}(t^n)\) for all \(n = 0, \ldots, N_t\); in this manner, the combined FE and time discretizations applied to \(y\) will yield the approximation \(y(t^n) \approx \sum_{i=1}^{N_h} \varphi_i y_{i,h}^n\). By way of example, let us consider first-order accurate backward differentiation formulas (BDF)—a family of multistep methods for the approximation of ODEs (see [67])—by starting from the semidiscrete problem (9.1). This yields to a sequence of algebraic problems in the unknowns \(\mathbf{y}^n\) for \(n = 0, \ldots, N_t\):

\[
\begin{cases}
\frac{1}{\Delta t} \mathcal{M} \mathbf{y}^{n+1} + A \mathbf{y}^{n+1} + \mathbf{F}^{n+1} = \frac{1}{\Delta t} \mathcal{M} \mathbf{y}^n + \mathbf{f}^{n+1} & \text{ for } n = 0, \ldots, N_t - 1, \\
\mathbf{y}^0 = \mathbf{y}_0.
\end{cases}
\tag{10.1}
\]

At this stage we recall that in the case of multiphysics coupled problems, the unknown vector \(\mathbf{y}\) collects the coefficients associated with several functions, related to different physical entities and to different differential operators. In this setting, the time discretization considered above (see (10.1)) represents an example of the so-called monolithic approach, as the semidiscrete problem (9.1) is discretized in time as a unique monolithic system, disregarding the underlying separation in different core models; see, e.g., [33]. The monolithic approach, however, is not the only possible one in the context of coupled problems. Alternative approaches are based on the so-called operator splitting strategy: if an IBVP is governed by an operator, say \(L\), that is the sum of two (or several) operators, say \(L_1\) and \(L_2\), the operator splitting strategy consists in advancing sequentially the single operators \(L_i\) from one time-step to the next one. Different algorithms are of course available to accomplish
this strategy; see, e.g., [20]. The net result is that at any given step of the algorithm
only one operator is dealt with at a time. Framing this in a different terminology,
operator splitting techniques allow us to split a multiphysics (coupled) problem into
a sequence of subproblems, each one dealing with a single type of physics (that is
with a single core mathematical model).

The numerical solution of every subproblem allows us to lower the computa-
tional complexity inherent to the monolithic approach; see, e.g., [28]. Moreover,
different time and spatial scales can be more easily accommodated in an operator
splitting context, by operating ad hoc choices of FE subspaces, spatial partitions of
computational domains, temporal discretizations ([82]). However, operator split-
ting methods can sometimes require a more severe restriction on the time step
to ensure stability, and they invariably introduce a further error with respect to
\( \Delta t \), called a splitting error. The choice between monolithic methods and operator
splitting methods heavily depends on the IBVP at hand, its multiscale nature, the
way solutions depend on data, and the availability of suitable numerical arsenal to
face the solution of the whole problem using a monolithic approach. As a matter
of fact, when using monolithic methods, one faces a potentially gigantic nonlinear
algebraic system that is severely ill conditioned—a concept crucial in numerical
analysis that, roughly speaking, expresses the way roundoff errors due to inexact
machine arithmetic propagate all along the solution process. Ill-conditioning—that
can be quantified through the ratio between the largest and smallest moduli of
the eigenvalues of the Jacobian matrices of the nonlinear algebraic systems—can
be cured by preconditioning strategies, i.e., by a suitable scaling of the differen-
tial operators to be implemented at the algebraic level on the Jacobian matrices.
Preconditioning, in its turn, calls for scalable algorithms, i.e., the realization of
preconditioning (scaling) matrices whose associated linear systems can be solved
on parallel computer architectures by a full exploitation of the parallelism [38,91].

From a mathematical standpoint, several questions are in order. How should
we choose our FE spaces and our temporal discretizations in order for our fully
discrete nonlinear algebraic system to be nonsingular. And what about solution
uniqueness? Moreover, in case we have ensured a unique solution, is it stable with
respect to coefficients and data perturbations, uniformly with respect to both \( \Delta t \)
and \( h \), in suitable Sobolev norms? Furthermore, with respect to the same norms, is
it convergent to the exact (weak) solution when both \( \Delta t \) and \( h \) tend to zero? And if
so, which are the convergence rates with respect to \( \Delta t \) and \( h \)? Proving these results
requires sophisticated arguments of functional analysis, a priori estimates for the
original IBVP, theoretical estimates for piecewise polynomial (FE) approximations
in Sobolev norms [64,68].

Error estimates with precise convergence rates are important (whenever possible)
to drive the choice of the FE partition in space and the discretization of the time
interval to ensure the desired level of accuracy with the least computational effort.
Indeed, the tradeoff is always the one between improving the accuracy and lowering
the computational complexity. Roughly speaking, the latter is measured in terms
of the size of the nonlinear algebraic systems and the computational time needed
for the determination of the numerical solution. In this respect, a crucial decision
is whether the given IBVP ([7.1]–[7.6]) is approximated as a whole (monolithically)
or by segregating the mechanical model from the electrophysiology one.
Part 3. Numerical solution

11. Numerical results for cardiac electromechanics

We present numerical solutions of the electromechanical problem (7.1)–(7.5) obtained by a FE discretization in space and a segregated scheme in time. Further details on the numerical schemes can be found in [76]. This numerical algorithm has been implemented in life$^x$ (https://lifex.gitlab.io/lifex), a high-performance C++ library developed within the iHEART project. To meet the significant demand of computational resources associated to the numerical approximation of this multiscale and multiphysics problem, we rely on the high-performance computing resources available at MOX, Politecnico di Milano (48 Intel Xeon ES-2640 CPUs), thanks to the parallel implementation of the life$^x$ code.

11.1. Convergence analysis in space. As mentioned above, an essential property of a numerical method is convergence, that is the ability of the method to produce a numerical solution arbitrarily close to the exact one, provided sufficiently fine discretization parameters ($h$ and $\Delta t$) are used. For many of the numerical methods known in the literature, it is also possible to find a convergence estimate, i.e., an upper bound on the error as a function of the discretization parameters. After a numerical method is implemented in a software code, the following test is often performed. One considers a problem for which an exact solution is known (typically on a simple geometry and for suitable data) and solves the numerical problem many times, by progressively refining the discretization parameters. Afterwards, the errors with respect to the exact solution are computed and the error trend is plotted on a graph, in comparison with the theoretical trend. This analysis is an example of so-called numerical verification. Its purpose is twofold: on the one hand it verifies that the implementation is bug free, on the other hand it validates the theoretical result.

The complexity of the models underlying the coupled problem of cardiac electromechanics is such that an exact solution is never known, even in the simplest cases. Nevertheless, a numerical verification of convergence can be done in the following way. We consider a sequence (indexed by $k = 0, 1, \ldots$) of refinements of the space-discretization parameter $h$, defined by $h_k = \alpha^k h_0$, for a suitable reduction factor $\alpha \in (0, 1)$. Let $u_h$ denote the numerical solution associated with $h$, and let us suppose that a numerical method has convergence order $p$. This means that, for $h \to 0^+$, we have $u_h = u_{\text{ex}} + ch^p + o(h^p)$, where $u_{\text{ex}}$ is the exact solution and where $c = \frac{d}{d(h^p)} u_h$. Hence, the difference between the numerical solutions obtained for two consecutive refinements of $h$ asymptotically decreases according to the estimate

$$u_{h_{k+1}} - u_{h_k} \simeq c (h_{k+1}^p - h_k^p) = c h_0^p (\alpha^p - 1) \alpha^k p.$$

This entails that the logarithm of the norm of the solution increment from step $k$ to step $k+1$ is an affine function of $k$:

$$\log \| u_{h_{k+1}} - u_{h_k} \| \simeq \log (\| c \| h_0^p (1 - \alpha^p)) + kp \log \alpha.$$

In conclusion, plotting the increment norms $\| u_{h_{k+1}} - u_{h_k} \|$ in a semilogarithmic plane, we expect a line with slope equal to the product between the convergence order $p$ and the reduction factor $\alpha$.

---

2iHEART—An Integrated Heart Model for the simulation of the cardiac function, European Research Council (ERC) grant agreement No 740132, P.I., A. Quarteroni.
Returning to the specific case of cardiac electromechanics, let us consider a tissue slab, that is a parallelepiped with dimensions $10 \times 10 \times 3$ mm, with fibers and sheets oriented as the $x$ and $y$ axis, respectively. We stimulate the tissue by applying a Gaussian impulse in a corner of the domain, so that an electrical impulse travels through the slab and makes it contract. Because of the different geometry than the one considered for (7.1)–(7.5), we have to adapt the boundary conditions of the mechanical subproblem. Specifically, we anchor the displacement of the tissue on one of the two vertical faces orthogonal to $f_0$ by setting a homogeneous Dirichlet boundary condition on the displacement (i.e., $d = 0$), we set the epicardial boundary conditions of (7.3) on the remaining vertical faces, while we leave the horizontal faces unloaded (homogeneous Neumann boundary conditions, i.e., $PN = 0$).

We fix $\Delta t = 1 \times 10^{-4}$ s, and, starting from $h_0 = 1$ mm, we progressively reduce $h$ by twelve refinements with a factor $\alpha = 2^{-1/3}$, so that $h$ is halved every three refinements. For each value of $k$, we numerically approximate the solution of (7.1)–(7.5) with piecewise linear finite elements and with a fully segregated scheme. In practice, when performing the numerical convergence study, one typically focuses on some quantity of interest (QoI). In the current case, we focus on the maximum displacement of the surface opposite the slab anchoring. The time evolution of this QoI is an index determined by all the processes involved in the model, being the result of the propagation of the electrical signal, the consequent generation of force, and finally of the elastic response of the tissue. In Figure 4 we show the time evolution of the QoI numerically obtained for $k$ ranging between 0 and 12. As expected, the curves visually converge towards a limit curve.

In Figure 5 we plot both the $L^\infty(0,T)$ (i.e., the maximum increment in the time window) and the $L^2(0,T)$ norm of the increments between consecutive curves as function of $k$. We verify in this manner that the QoI converges with order one:
indeed, the curves shown in Figure 5 for both the norms follow the same behavior of the reference dashed line with slope equal to $\log(\alpha)$.

11.2. Left ventricle electromechanics. We now turn to an electromechanical simulation in a realistic geometry. We consider a computational domain derived from the Zygote heart model [2], representing that of an average 21-year-old healthy caucasian man. The computational mesh is represented in Figure 3. It is made of $346 \cdot 10^3$ vertices. Since each vertex corresponds to 24 unknowns, the discretized problem features more than 8 million unknowns for each time step. We discretize the time interval $(0, 0.8\, \text{s})$ with a time step size $\Delta t = 2 \times 10^{-4} \, \text{s}$. The fibers, sheets and cross-fibers directions ($f_0$, $s_0$, and $n_0$), represented in Figure 6, are generated by exploiting the algorithm proposed in [13].

To trigger the propagation of the electrical potential, we apply an electric stimulus $I_{\text{app}}$ in three spherical regions located in the endocardium to mimic the action of the Purkinje fibers. This originates a traveling wave that quickly covers the whole ventricle. Figure 7 shows the variable describing the intracellular calcium concentration $[\text{Ca}^{2+}]_i$ at different times. An output of the simulation which is commonly used in clinical applications is shown in Figure 8, that is the so-called activation time. It corresponds to the time delay from trigger of the stimulus and the time when the action potential wave reaches a given point of the domain $\Omega_0$. More precisely, it is defined as the time corresponding the steepest growth of transmural potential (i.e., $t_{\text{act}}(t, x) = \arg\max_{t \in [0, T]} \partial_t u(t, x)$).

The increase of calcium concentration represented in Figure 7 triggers the microscopic mechanisms of active force generation. As a consequence, the active tension $T_a$ increases significantly and the ventricle contracts. Figure 9 depicts the values of $T_a$ at different times and the corresponding mechanical deformation of the ventricle. The contracting ventricle features a characteristic twisting movement, which is to
be ascribed to the presence of fibers (see Figure 6), accompanied by a thickening of the cardiac wall.

12. Towards clinical applications

As of 2016, cardiovascular diseases stand as the leading cause of death in the world [1], with nearly 18 million individuals deceased, 85% of which are due to heart
attack and stroke. Still, recent advancements in clinical practice are steadily improving patients’ conditions and the treatment of cardiac diseases. In this context, mathematical and numerical modeling are increasingly playing a role by providing medical doctors and clinicians with valuable diagnostic and predictive tools for virtually exploring therapeutic scenarios or even planning surgical interventions. We are actually experiencing the emergence of a new discipline known as computational medicine, the meeting point of mathematics, scientific computing, data science, engineering, and medicine.

Using integrated cardiac or electromechanics modeling to address clinical questions is still at the outset of computational medicine, due to the relative complexity and novelty of these models. Exploiting instead a few core models is becoming more customary in recent years, as it is occurring for cardiac electrophysiology to deal with electric dysfunctions; see, e.g., [9]. We report on the following two examples of clinical applications being addressed by the iHEART team at Politecnico di Milano.

In collaboration with the Arrhythmology and Cardiac Electrophysiology Unit of IRCCS San Raffaele Hospital in Milan (Italy), we investigated factors that favor triggering and sustenance of arrhythmias, as for example ventricular tachycardia. Indeed, structural abnormalities of the tissue that can be modeled as localized alterations of the electrical conductivities $\sigma_f$, $\sigma_s$, and $\sigma_n$ in (3.3), generated by an episode of acute myocardial infarction, can create conditions for triggering a re-entry circuit near the scar area. Along this circuit, the electric impulse can continually reactivate the ventricles in an accelerated and uncontrolled manner, in some cases leading to sudden cardiac arrest, a fatal condition if not immediately treated. Conventional therapy consists of an electrophysiological study with catheter ablation, an invasive clinical procedure that identifies and eliminates those channels—called isthmuses—that sustain the circuits. A similar procedure allows the treatment of atrial fibrillation, a disorganization of the electric signal in the left atrium, which can be linked to progressive tissue damage. The numerical simulation of cardiac arrhythmias is very challenging: the electric potential exhibits a complex pattern with spiral waves or along re-entry circuits, unlike the ordered propagation typical of normal pacing. We performed numerical simulations of instability mechanisms, which are capable of generating ventricular tachycardia and atrial fibrillation, in the presence of heterogeneity in the tissue [35]. These, in combination with electrical measurements on the patient, could help in identifying the channels at the base of the re-entry circuits, and, possibly in the near future, to improve ablative procedures.
In collaboration with the Cardiology and Radiology Departments of S. Maria del Carmine Hospital in Rovereto (Trento, Italy), our team has taken the first steps towards the optimization of the cardiac resynchronization therapy (CRT). CRT consists in implanting a device capable of restoring the correct pacing after the synchronous heartbeat has been compromised by conduction disturbances or tissue scars. A mapping of the coronary veins of the left ventricle is carried out by inserting a catheter–electrode through the blood vessels to detect electrical activity. This mapping serves to identify the point of later activation wherein the left electrode of the CRT is positioned. Our research team has developed a computational pipeline...
to reduce the duration of this invasive clinical procedure and to calculate the point of later activation by enabling a deeper exploration aside the points mapped by the catheter [34].

In the papers previously mentioned, a few validation results are reported by comparing some specific outputs provided by the numerical models with corresponding indicators or biomarkers collected in a clinical environment. For the sake of brevity, in this paper we do not deal with the topic of validation, which is, nonetheless, a step of utmost importance for a clinical exploitation of such computational tools.

13. Conclusions and a look ahead

So far we have limited ourselves to studying the direct (forward) problem, namely problem (7.1)–(7.5), and to showing how its solution can be approximated in an input-output framework. Actually, many topics have remained in the background, even if they are fundamental in facing a problem of such complexity. Major challenges facing cardiac modeling include parameter inference from uncertain experimental measurements, model personalization to patient data, model selection, model discrepancy from reality, and the way these factors affect the confidence in model prediction. Two kinds of uncertainties are usually considered: Epistemic uncertainty is due to either lack of knowledge (e.g., a quantity or a coefficient that was not possible to measure) or experimental errors (e.g., a quantity that was not correctly measured). Aleatory uncertainty is instead due to stochasticity (e.g., every individual reacts differently each time, which represents an intrinsic kind of uncertainty) or to variability between individuals (every individual behaves differently than everyone else, which is an extrinsic uncertainty).

Epistemic uncertainty can be addressed by uncertainty quantification (UQ) mathematical techniques. We distinguish between forward UQ, which deals with the propagation of the uncertainty on the parameters on the outputs of the model, and backward UQ, which studies how the measurement errors on the outputs affect the estimation of the parameters.

Aleatory and epistemic uncertainty can be faced in the framework of reduced order models (ROM). ROM can in fact be used with multiple purposes. A canonical usage is to allow the simultaneous approximation of the whole solution manifold of a boundary value problem (but it could be as well an IVP or an IBVP) that is made of the set of solutions corresponding to the whole range of parameters (or coefficients) that characterize the given boundary value problem. In abstract terms, using for the sake of readability an algebraic standpoint, let us denote by \( y(t; \mu) \) the solution of a time-dependent nonlinear problem \( \mu \in P \subset \mathbb{R}^p \) denotes a set of input parameters) of the form

\[
\begin{align*}
\mathcal{M}(\mu) \frac{dy}{dt}(t, \mu) + \mathcal{A}(\mu) y(t; \mu) + \mathcal{F}(y(t; \mu); \mu) &= f(t; \mu) \quad \text{in} \ (0, T], \\
y(0; \mu) &= y_0(\mu),
\end{align*}
\]

(13.1)

deriving, e.g., from the finite-element discretization of a parametrized IBVP (similar to (9.1)).

The dimensional reduction is made possible by exploiting the parametric dependence of the solution manifold, that is, the set \( M_h = \{ y(t; \mu) : t \in (0, T], \mu \in P \} \), thanks to the evaluation of a database of solutions, or snapshots, for selected parameter values, and to a Galerkin projection onto the finite-dimensional space spanned
by a set of selected basis functions. These basis functions can be selected either through a greedy approach or by means of proper orthogonal decomposition (POD): in the case of a stationary problem, the basis functions are the snapshots themselves if a greedy algorithm is used, or else the first singular vectors of the snapshot matrix if POD is used; see, e.g., [23,41,66]. One possible ROM approximation of (13.1) might be

\[
\begin{cases}
\mathcal{M}_N(\mu) \frac{dy_N}{dt}(t, \mu) + \mathcal{A}_N(\mu) y_N(t; \mu) + W_N^T \mathbf{f}(W_N y_N(t; \mu); \mu) = f_N(t; \mu) \\
y_N(0; \mu) = y_{N,0}(\mu),
\end{cases}
\]

in \((0, T]\),

where \(\mathcal{M}_N(\mu) = W_N^T \mathcal{M}(\mu) W_N\), \(\mathcal{A}_N(\mu) = W_N^T \mathbf{A}(\mu) W_N\), and \(f_N(\mu) = W_N^T \mathbf{f}(t; \mu)\), and \(W_N\) is a rectangular matrix whose columns encode the snapshot solutions. For time-dependent problems, the parameter space can still be sampled by one of the two techniques mentioned, whereas POD is usually exploited to reduce trajectories of the system over the time interval. In mathematical terms, the solution manifold is approximated by a linear subspace \(V_N\) generated by a handful of \(N \ll N_h\) (snapshot) basis functions corresponding to a wise choice of \(N\) representative parameters. These basis functions are computed offline by solving \(N\) finite-element problems, each one consisting of a large number \((N_h)\) of equations. This heavy computational task can however be computed once and for all (offline). Then, for every new value of the parameter \(\mu\) (say, \(\mu_{\text{new}}\)), the corresponding solution \(u_{\text{new}}\) is obtained by solving a Galerkin problem of reduced size \(N\) in the reduced subspace \(V_N\). This online computation is generally very cheap (thanks to the small size of the reduced subspace) and is carried out online. Thanks to this very distinguishing feature, ROM enable the solution in “real time” of large FE problems, allowing us therefore to cultivate the dream of a real-time interaction between mathematicians and clinicians.

Needless to say, many mathematical and numerical issues need to be addressed in order to make the previous strategy theoretically rigorous and computationally efficient. From a theoretical standpoint, the original IBVP should be reducible: this means that it should be possible to accurately approximate its solution manifold by a single linear subspace of (low) dimension \(N\). This property, which depends on the regularity of the solution manifold \(M_h \in V\) with respect to the parameter variations, can be assessed by investigating the so-called Kolmogorov \(N\)-width, that is the minimum distance (in a suitable Sobolev norm) of an arbitrary element of the manifold from linear subspaces of dimension \(N\):

\[
d_N(M_h, V) = \inf_{V_N \subseteq V} \sup_{\dim V_N = N} \inf_{y \in M_h} \| y - y_N \|_V.
\]

For ample classes of boundary value problems, the Kolmogorov \(N\)-width (KNW) decreases exponentially fast with respect to \(N\). In those less benign cases where the KNW is much slower to decrease with respect to \(N\), nonlinear approximation ROM techniques can be set up. In this case the solution manifold is approximated by a set of local linear subspaces that better exploit the local (nonuniform) regularity properties of the solution manifold [23,41,66].
From the computational viewpoint, efficiency relies on the possibility of using an offline-online computational splitting, where the online phase requires computational time depending only on the reduced dimension $N$ and not on the full-order dimension $N_h$ (this property will enable real-time simulations). Precisely, the arrays appearing in (13.2) can be efficiently assembled in a rapid online phase by combining parameter-independent quantities stored during a more expensive offline phase, provided that the dependence of the high-fidelity arrays, such as $A(\mu)$ and $f(t; \mu)$, is affine.

When this is not the case, or at any rate when the original problem is nonlinear, further approximations should be operated, and extra reduced spaces should be used to independently approximate the nonlinear terms. This procedure is sometimes called hyper-reduction [31] and should obviously be accompanied by extra investigation on the corresponding approximation properties of the hyper-reduced solution.

ABOUT THE AUTHORS

Alfio Quarteroni is professor of numerical analysis at Politecnico di Milano and professor emeritus of École polytechnique fédérale de Lausanne, Switzerland. He has authored 25 books and over 400 papers in international scientific journals and conference proceedings; he is editor in chief of two book series published by Springer. He received many awards and honors in recognition of his work, and is a member of the Italian Academy of Science, the European Academy of Science, and the Academia Europaea. His research interests include mathematical modelling and its applications to fluid mechanics, geophysics, medicine, and the improvement of sports performance.

Luca Dedè is professor in numerical analysis at Politecnico di Milano. His research involves the development of efficient and accurate numerical methods for the approximation of partial differential equations, multiphysics, and multiscale problems, as well as problems in computational fluid dynamics and computational medicine.

Francesco Regazzoni is junior researcher at Politecnico di Milano. His research interests cover the mathematical modeling and numerical approximation of multiscale problems, as well as the combination of machine learning with numerical analysis and reduced order modeling.

REFERENCES


A. Tagliabue, L. Dedè, and A. Quarteroni, *Fluid dynamics of an idealized left ventricle: the extended Nitsche’s method for the treatment of heart valves as mixed time varying


MOX–Dipartimento di Matematica, Politecnico di Milano, Milano, Italy; and Mathematics Institute, École Polytechnique Fédérale de Lausanne, Switzerland

Email address: alfio.quarteroni@polimi.it

MOX–Dipartimento di Matematica, Politecnico di Milano, Milano, Italy

Email address: luca.dede@polimi.it

MOX–Dipartimento di Matematica, Politecnico di Milano, Milano, Italy

Email address: francesco.regazzoni@polimi.it