Continuous analogue to iterative optimization for PDE-constrained inverse problems

R. Boiger, A. Fiedler, J. Hasenauer & B. Kaltenbacher

Published online: 10 Jul 2018.

Submit your article to this journal ↗

View Crossmark data ↗

Continuous analogue to iterative optimization for PDE-constrained inverse problems

R. Boiger[a,b]*, A. Fiedler[c,d]*, J. Hasenauer[c,d] and B. Kaltenbacher[a]

[a]Institute of Mathematics, Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria; [b]Materials Center Leoben Forschung Gmbh, Leoben, Austria; [c]Chair of Mathematical Modeling of Biological Systems, Center for Mathematics, Technische Universität München, Garching, Germany; [d]Institute of Computational Biology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany

ABSTRACT

The parameters of many physical processes are unknown and have to be inferred from experimental data. The corresponding parameter estimation problem is often solved using iterative methods such as steepest descent methods combined with trust regions. For a few problem classes also continuous analogues of iterative methods are available. In this work, we expand the application of continuous analogues to function spaces and consider PDE (partial differential equation)-constrained optimization problems. We derive a class of continuous analogues, here coupled ODE (ordinary differential equation)–PDE models, and prove their convergence to the optimum under mild assumptions. We establish sufficient bounds for local stability and convergence for the tuning parameter of this class of continuous analogues, the retraction parameter. To evaluate the continuous analogues, we study the parameter estimation for a model of gradient formation in biological tissues. We observe good convergence properties, indicating that the continuous analogues are an interesting alternative to state-of-the-art iterative optimization methods.

1. Introduction

Partial differential equations (PDEs) are used in various application areas to describe physical, chemical, biological, economic, or social processes. The parameters of these processes are often unknown and need to be estimated from experimental data [1,2]. The maximum likelihood and maximum a posteriori parameter estimates are given by the solutions of PDE-constrained optimization problems. Since PDE-constrained optimization problems are in general nonlinear, non-convex as well as computationally challenging, efficient and reliable optimization methods are required.

Over the last decades, a large number of numerical methods for PDE-constrained optimization problems have been proposed (see [3–8] and references therein). Beyond methodological contributions, there is numerous literature on parameter estimation in

special applications available [9]. Most of the theoretical and applied work focuses on iterative methods, which generate a discrete sequence of points along which the objective function decreases. In this manuscript, we pursue an alternative route and consider continuous analogues (see [10–14] and references therein).

Continuous analogues are formulations of iterative optimization methods in terms of differential equations and have been derived for a series of optimizers for real-valued optimization problems, including the Levenberg–Marquardt and the Newton–Raphson method [11,15]. For many constrained and unconstrained optimization problems, continuous analogues exhibit larger regions of attraction and more robust convergence than discrete iterative methods [15]. In a recent study, the concept of continuous analogues has been employed to develop a tailored parameter optimization method for ordinary differential equation (ODE) models with steady-state constraints [16]. This method outperformed iterative constrained optimization methods with respect to convergence and computational efficiency [16], and complemented methods based on computer algebra [17]. Although continuous analogues are generally promising, we are not aware of continuous analogues for solving PDE-constrained optimization problems.

In this manuscript, we formulate a continuous analogue for PDE-constrained optimization problems based on the structure of the PDE and the local geometry of its solution space. The continuous analogue is a coupled system of ODEs and PDEs, which has the optima of the PDE-constrained optimization problem as equilibrium points. We establish local stability and convergence under mild assumptions. The continuous analogue enables us to use adaptive numerical methods for solving optimization problems with PDE constraints. Beyond the generalization of previous results for ODEs (i.e. [16]) to function spaces, we provide rationales and constraints for the choice of tuning parameters, in particular of the retraction factor. The methods are applied to a model of gradient formation in biological tissues.

The manuscript is structured as follows . In Sections 2 and 3, we introduce the considered class of models and PDE-constrained optimization problems. For these classes, we propose a continuous analogue to descent methods in Section 4. In Section 5, we prove convergence of the continuous analogue to local optima. The assumptions under which stability and convergence to local optima are achieved are discussed in Section 6. The properties of the continuous analogues are studied in Section 7 for a model of gradient formation in biological tissues.

2. Mathematical model

We consider parameter estimation for models with elliptic and parabolic PDE constraints. In the parabolic case, the initial condition u_0 of the parabolic PDE is defined as the solution of an elliptic PDE:

$$u_t(t) = C(\theta, t, u(t)) \quad t \in]0, T[$$
$$u(0) = u_0, \tag{1}$$

with

$$0 = C_0(\theta, u_0), \tag{2}$$

in which $u(t)$ is the time-dependent solution of the elliptic PDE, and C and C_0 are operators. The considered combination of parabolic and elliptic PDEs is flexible and allows for the analysis of many practically relevant scenarios. Here, u_0 denotes a stable steady state of the unperturbed system, while u denotes the transient solution of the perturbed system starting in the steady state of the unperturbed system u_0. The observables of the models are defined via observation operators $B(\theta, t, u(t))$ and $B_0(\theta, u_0)$

$$y(t) = B(\theta, t, u(t)), \quad y_0 = B_0(\theta, u_0). \tag{3}$$

We work in the following functional analytic setting. Let V be a separable Banach space with dual space V^* and let V be continuously and densely embedded into a Hilbert space H, such that $V \subseteq H \cong H^* \subseteq V^*$ forms a Gelfand triple. Furthermore, we consider the initial value $u_0 \in V$ and the transient state $u \in W(0, T) = L^2(0, T; V) \cap H^1(0, T; V^*)$. The parameters θ are assumed to be finite dimensional and real-valued, $\theta \in \mathbb{R}^{n_\theta}$. For each t and θ, the observation operators B, B_0 are mappings, $B : \mathbb{R}^{n_\theta} \times]0, T[\times V \to Z$, $B_0 : \mathbb{R}^{n_\theta} \times V \to Z_0$, where the observation spaces Z, Z_0 are Hilbert spaces. The operators C and C_0 are mappings, $C : \mathbb{R}^{n_\theta} \times]0, T[\times V \to V^*$ and $C_0 : \mathbb{R}^{n_\theta} \times V \to V^*$. Existence of a weak solution holds, e.g. under the following assumptions on the differential operator C ([18] p.770 ff.):

Assumption 2.1: *There exists $\varepsilon_\theta > 0$ such that for all $\theta \in B_{\varepsilon_\theta}(\theta^*) = \{\theta \in \mathbb{R}^{n_\theta} \mid \|\theta - \theta^*\| < \varepsilon_\theta\} \subseteq \mathbb{R}^{n_\theta}$, the following holds.*

- *The operator $-C(\theta, t, \cdot)$ is monotone and hemicontinuous for each $t \in]0, T[$.*
- *The operator $-C(\theta, t, \cdot)$ is coercive for each $t \in]0, T[$, i.e. there exist constants $c_1 > 0$ and $c_2 \geq 0$ such that*

$$- \langle C(\theta, t, v), v \rangle_{V^*, V} \geq c_1 \|v\|_V^2 - c_2 \quad \text{for all } v \in V, \ t \in]0, T[.$$

- *The operator $C(\theta, \cdot, \cdot)$ satisfies a growth condition, i.e. there exist a non-negative function $c_3 \in L^2(0, T)$ and a constant $c_4 > 0$ such that*

$$\|C(\theta, t, v)\|_{V^*} \leq c_3(t) + c_4 \|v\|_V \quad \text{for all } v \in V, \ t \in]0, T[.$$

- *The function $t \mapsto \langle C(\theta, \cdot, w), v \rangle_{V^*, V}$ is measurable on $]0, T[$ for all $v, w \in V$.*

For the differential operator C_0, the first two items in Assumption 2.1 are assumed to hold, cf. Assumption 5.2 below.

In mathematical biology, the differential operator C is often semilinear and describes a reaction–diffusion–advection equation,

$$C(\theta, t, u) = f(k, u) - \nabla_x \cdot (vu - D\nabla_x u),$$

in which $u(t) \in V$ is a concentration vector, $x \in \Omega \subseteq \mathbb{R}^n$ is the spatial location and $f : \mathbb{R}^{n_k} \times V \to V^*$ is the reaction term. The parameters $\theta = (k, v, D)$ are the velocity vector $v \in \mathbb{R}^n$, the diffusion matrix $D \in \mathbb{R}^{n \times n}$ and the kinetic parameters $k \in \mathbb{R}^{n_k}$. If D is positive definite and f fulfils a certain growth condition, then this operator satisfies Assumption 2.1.

3. Parameter estimation problem

We consider the estimation of the unknown model parameters θ from noise-corrupted measurements of the observables y. To obtain estimates for the unknown parameters, an objective function is minimized, e.g. the negative log-likelihood function or the sum-of-squared-residuals. In the following, we distinguish two cases.

3.1. Elliptic and parabolic PDE constraints

In the general case, observations are available for the initial state and the transient phase. The objective function $\tilde{J}$ depends on the parameters and the parameter-dependent solutions of the parabolic and the elliptic PDE, $\tilde{J} : \mathbb{R}^{n_\theta} \times V \times W(0,T) \to \mathbb{R}$, e.g. $\tilde{J}(\theta, u_0, u) = \frac{1}{2} \|y_0 - B_0(\theta, u_0)\|_{Z_0}^2 + \frac{1}{2} \int_0^T \|y(t) - B(\theta, t, u)\|_Z^2 \, dt$. The optimization problem is given by

$$
\begin{aligned}
\min_{\theta, u_0, u} \quad & \tilde{J}(\theta, u_0, u) \\
\text{s.t.} \quad & u_t = C(\theta, t, u) \quad u(0) = u_0 \\
& 0 = C_0(\theta, u_0).
\end{aligned} \tag{4}
$$

3.2. Elliptic PDE constraint

In many applications, only experimental data for the steady state of a process are available. Possible reasons include fast equilibration of the process and limitations of experimental devices. In this case, the problem is simplified and the objective function J depends only on the parameters and parameter-dependent solutions of the elliptic PDE, $J : \mathbb{R}^{n_\theta} \times V \to \mathbb{R}$, e.g. $J(\theta, u_0) = \frac{1}{2} \|y_0 - B_0(\theta, u_0)\|_{Z_0}^2$. The optimization problem is given by

$$
\begin{aligned}
\min_{\theta, u_0} \quad & J(\theta, u_0) \\
\text{s.t.} \quad & 0 = C_0(\theta, u_0).
\end{aligned} \tag{5}
$$

The reduced formulation of the optimization problem (5) is given by

$$
\min_\theta \quad j(\theta) := J(\theta, \varphi_0(\theta)), \tag{6}
$$

in which $\varphi_0(\theta)$ denotes the parameter-dependent solution of $C_0(\theta, u_0) = 0$, and $j : \mathbb{R}^{n_\theta} \to \mathbb{R}$ denotes the reduced objective function.

4. Continuous analogue of descent methods for PDE-constrained problems

Optimization problems of types (4) and (5) are currently often solved using iterative descent methods, combined with trust regions. In the following, we develop a continuous analogue of an iterative descent method for PDEs. For simplicity, we first consider the case of an elliptic PDE constraint (5) and afterwards generalize the results to the case of mixed parabolic and elliptic PDE constraints (4).

4.1. Elliptic PDE Constraint

To solve optimization problem (5), we derive a coupled ODE–PDE system. The trajectory of this continuous analogue evolves in parameter and state space on the manifold defined by $C_0(\theta, u_0) = 0$ towards a local minimum. To evolve on the manifold, the continuous analogue exploits the first-order geometry of the manifold, i.e. its tangent space.

Mathematically, the first-order geometry of $C_0(\theta, \varphi_0(\theta)) = 0$ is defined by the sensitivity equations

$$\frac{\partial C_0}{\partial u_0}(\theta, \varphi_0(\theta))\frac{\partial \varphi_0}{\partial \theta_i}(\theta) + \frac{\partial C_0}{\partial \theta_i}(\theta, \varphi_0(\theta)) = 0, \quad i \in \{1, \ldots, n_\theta\}. \tag{7}$$

The sensitivity equations can be reformulated to

$$\frac{\partial \varphi_0}{\partial \theta_i}(\theta) = -\left(\frac{\partial C_0}{\partial u_0}(\theta, \varphi_0(\theta))\right)^{-1}\frac{\partial C_0}{\partial \theta_i}(\theta, \varphi_0(\theta)), \quad i \in \{1, \ldots, n_\theta\},$$

provided the inverse of $(\partial C_0/\partial u_0)(\theta, \varphi_0(\theta))$ exists. We extend $\nabla_\theta \varphi_0$ to points (θ, u_0) not necessarily lying on the solution manifold of $C_0 = 0$ by defining the operator $S_0 : \mathbb{R}^{n_\theta} \times V \to V^{n_\theta}$ that provides the solution of the sensitivity equations

$$\frac{\partial C_0}{\partial u_0}(\theta, u_0)S_{0i}(\theta, u_0) + \frac{\partial C_0}{\partial \theta_i}(\theta, u_0) = 0, \quad i \in \{1, \ldots, n_\theta\}, \tag{8}$$

for given θ and u_0. With (7), it holds that

$$S_0(\theta, \varphi_0(\theta)) = \nabla_\theta \varphi_0(\theta), \tag{9}$$

provided the solutions to (7) and (8) are unique.

In order to couple changes in θ with appropriate changes in u_0, we now use the fact that the function $\nabla_\theta \varphi_0(\theta)$ provides the first-order term of the Taylor series expansion of the steady state with respect to the parameter vector θ,

$$\varphi_0(\theta + r\Delta\theta) = \varphi_0(\theta) + \nabla_\theta \varphi_0(\theta)r\Delta\theta + o(r) \quad \text{as } r \to 0, \ r \in \mathbb{R}. \tag{10}$$

Defining $\hat{\varphi}_0(r) := \varphi_0(\theta + r\Delta\theta)$ for some $\Delta\theta \in \mathbb{R}^{n_\theta}$ and differentiating (10) with respect to r yields

$$\frac{d\hat{\varphi}_0}{dr}(r) = \nabla_\theta \varphi_0(\theta)\Delta\theta + o(1) = S_0(\theta + r\Delta\theta, \hat{\varphi}_0(r))\Delta\theta + o(1) \quad \text{as } r \to 0. \tag{11}$$

This relation motivates the formulation of the coupled ODE–PDE model

$$\begin{aligned}
\frac{d\theta}{dr}(r) &= g(\theta, u_0), \quad \theta(0) = \theta_0, \\
\frac{du_0}{dr}(r) &= S_0(\theta, u_0)\frac{d\theta}{dr}(r) = S_0(\theta, u_0)g(\theta, u_0), \quad u_0(0) = u_{0,0},
\end{aligned} \tag{12}$$

using the artificial time parameter r. For a change in the parameters $d\theta/dr$, the update in u_0 is chosen according to (11). Solutions of this dynamical system evolve on the manifold $C_0(\theta(r), u_0(r)) = 0$ for arbitrary parameter update directions $g : \mathbb{R}^{n_\theta} \times V \to \mathbb{R}^{n_\theta}$,

provided that the initial state is on the manifold $C_0(\theta_0, u_{0,0}) = 0$. The state variables of this coupled ODE–PDE system are θ and u_0, and the path variable is r. To solve optimization problem (6), g is chosen as an arbitrary descent direction satisfying

$$\nabla j(\theta)^{\mathrm{T}} g(\theta, u_0) < 0,$$

more precisely satisfying Assumption 5.5 below. For example, g can be chosen as a steepest descent direction

$$g = \operatorname*{argmin}_{\|v\|_* \le 1} \nabla j(\theta)^{\mathrm{T}} v \tag{13}$$

for some norm $\|\cdot\|_*$. For the Euclidian norm, we obtain the gradient descent direction,

$$g_i(\theta, u_0) := -\frac{\partial J}{\partial \theta_i}(\theta, u_0) - \left\langle \frac{\partial J}{\partial u_0}(\theta, u_0), S_{0i}(\theta, u_0) \right\rangle_{V^*, V} =: d_i(\theta, u_0), \quad i \in \{1, \dots, n_\theta\}, \tag{14}$$

in which we substituted $\varphi_0(\theta)$ by u_0 extending the definition also to states u_0 that are not on the steady-state manifold. Likewise, defining $\|v\|_*^2 := v^{\mathrm{T}} H(\theta, u_0) v$ with some positive definite matrix $H(\theta, u_0)$, so that

$$g(\theta, u_0) = H(\theta, u_0)^{-1} d(\theta, u_0) \tag{15}$$

leads to a descent direction. Using, e.g. the Hessian of j,

$$H_{ij}(\theta, u_0) = \frac{\partial^2 J}{\partial \theta_i \partial \theta_j}(\theta, u_0) + \left\langle \frac{\partial^2 J}{\partial u_0 \partial \theta_j}(\theta, u_0), S_{0i}(\theta, u_0) \right\rangle_{V^*, V}$$

$$+ \left\langle \frac{\partial^2 J}{\partial u_0 \partial \theta_i}(\theta, u_0), S_{0j}(\theta, u_0) \right\rangle_{V^*, V} + \frac{\partial^2 J}{\partial u_0^2}(\theta, u_0) \left(S_{0i}(\theta, u_0), S_{0j}(\theta, u_0) \right)$$

$$+ \left\langle \frac{\partial J}{\partial u_0}(\theta, u_0), T_{0i,j}(\theta, u_0) \right\rangle_{V^*, V}, \quad i, j \in \{1, \dots, n_\theta\}, \tag{16}$$

with the second-order sensitivities defined by

$$\frac{\partial C_0}{\partial u_0}(\theta, u_0) T_{0i,j}(\theta, u_0) + \frac{\partial^2 C_0}{\partial u_0^2}(\theta, u_0) \left(S_{0i}(\theta, u_0), S_{0j}(\theta, u_0) \right) + \frac{\partial^2 C_0}{\partial u_0 \partial \theta_j}(\theta, u_0) S_{0i}(\theta, u_0)$$

$$+ \frac{\partial^2 C_0}{\partial u_0 \partial \theta_i}(\theta, u_0) S_{0j}(\theta, u_0) + \frac{\partial^2 C_0}{\partial \theta_i \partial \theta_j}(\theta, u_0) = 0, \quad i, j \in \{1, \dots, n_\theta\},$$

leads to Newton, Gauss-Newton (upon skipping the $T_{0i,j}$ terms) or quasi-Newton methods (where approximations to the Hessian (16) are computed via low rank updates). As the Hessian is not guaranteed to be positive definite, regularization with a scaled identity matrix, $H_\mu(\theta, u_0) = H(\theta, u_0) + \mu I$ with $\mu > 0$, might be useful. However, how to choose μ and possible continuous update rules are out of the scope of this work. The coupled ODE–PDE systems (12) can be solved using numerical time-stepping methods. These numerical methods might, however, accumulate errors resulting in the divergence of the state $(\theta(r), u_0(r))$ from the steady-state manifold. Additionally, the initial state, $u_{0,0}$, might not be on the steady-state manifold. To account for this, we include the retraction

term $\lambda C_0(\theta, u_0)$ in the evolution equation of u_0, with retraction factor $\lambda > 0$. This yields the following continuous analogue of a descent method for optimization problems with elliptic PDE constraints,

$$\frac{\mathrm{d}\theta}{\mathrm{d}r}(r) = g(\theta, u_0), \quad \theta(0) = \theta_0,$$

$$\frac{\mathrm{d}u_0}{\mathrm{d}r}(r) = S_0(\theta, u_0)g(\theta, u_0) + \lambda C_0(\theta, u_0), \quad u_0(0) = u_{0,0}. \tag{17}$$

As, for fixed θ, the equation $C_0(\theta, u_0) = 0$ defines a stable steady state of the PDE (2), the retraction term stabilizes the manifold. For $\lambda \gg 1$, the system should first converge to the steady state $\varphi_0(\theta_0)$ for the initial parameter θ_0 and then move along the manifold to a local optimum θ^* as illustrated in Figure 1.

4.2. Elliptic and parabolic PDE constraints

The continuous analogue for descent with elliptic PDE constraints can be generalized to problems with parabolic and elliptic PDE constraints. One possibility for doing so is to
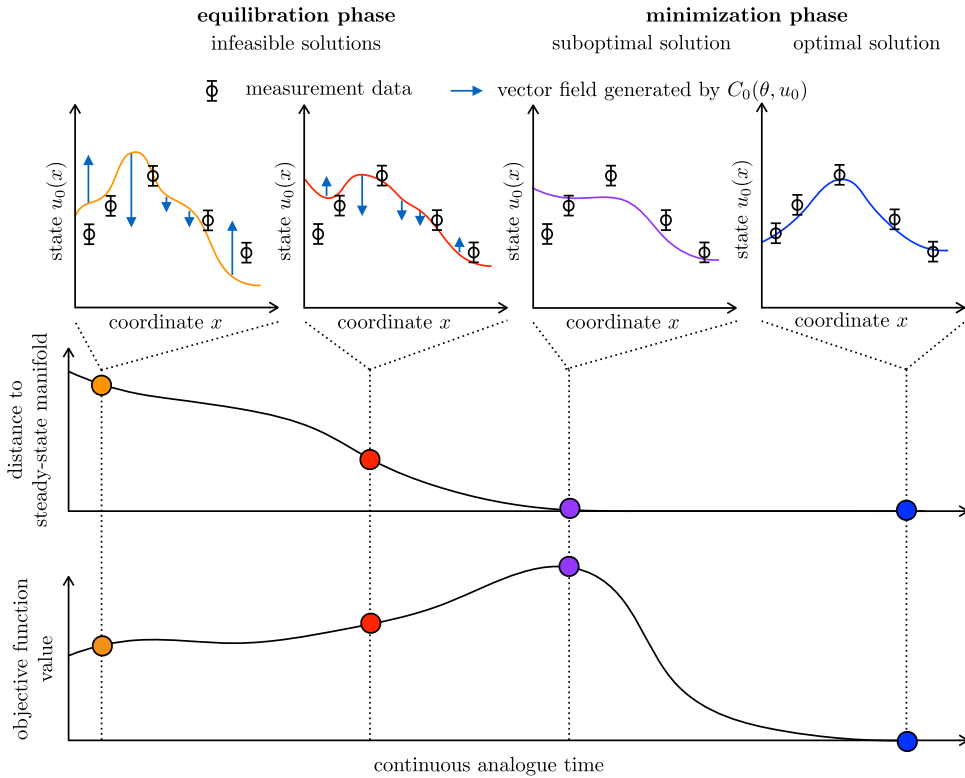
Figure 1. The state of the system is illustrated along the trajectory of (17). In the first phase, the equilibration phase, the system converges to the manifold. The solution is not feasible during this phase as the equality constraint, $C_0(\theta, u_0) = 0$, is violated. In the course of the equilibration, the objective function value might increase. In the second phase, the minimization phase, the objective function is minimized along the steady-state manifold.

consider the partially reduced problem

$$\min_{\theta, u_0} \quad \tilde{\mathcal{J}}(\theta, u_0) := \tilde{J}(\theta, u_0, \varphi(\theta, u_0))$$

$$\text{s.t.} \quad 0 = C_0(\theta, u_0), \tag{18}$$

in which $u = \varphi(\theta, u_0)$ denotes the solution to $u_t = C(\theta, t, u)$ with $u(0) = u_0$. Given this formulation, we can use continuous analogue (17) with

$$\tilde{g}_i(\theta, u_0) = -\frac{\partial \tilde{\mathcal{J}}}{\partial \theta_i}(\theta, u_0) - \frac{\partial \tilde{\mathcal{J}}}{\partial u_0}(\theta, u_0) S_{0i}(\theta, u_0), \quad i \in \{1, \ldots, n_\theta\}. \tag{19}$$

To avoid the need for the solution operator $\varphi : \mathbb{R}^{n_\theta} \times V \to W(0, T)$, alternatively, a continuous analogue of the full problem can also be formulated. This is beyond the scope of this study, though, and will be subject of future research.

5. Local stability and convergence to a local optimum

The behaviour of the coupled ODE–PDE system (17) introduced in the previous section depends on the properties of the objective function and the PDE model, as well as the retraction factor λ. To prove that a solution of (17) with an appropriate retraction factor λ is well defined and converges to the local minimizer $(\theta^*, u_0^*) = (\theta^*, \varphi_0(\theta^*))$ of (5), we impose the following assumptions.

Assumption 5.1: *The descent direction vanishes at the minimizer θ^* of the optimization problem $\min_\theta j(\theta)$,*

$$g(\theta^*, \varphi_0(\theta^*)) = 0.$$

Assumption 5.2: *There exists $\varepsilon_\theta > 0$ such that for all $\theta \in B_{\varepsilon_\theta}(\theta^*) = \{\theta \in \mathbb{R}^{n_\theta} \mid \|\theta - \theta^*\| < \varepsilon_\theta\} \subseteq \mathbb{R}^{n_\theta}$, the following holds.*

- *The operator $-C_0(\theta, \cdot)$ is monotone and hemicontinuous.*
- *The operator $-C_0(\theta, \cdot)$ is coercive.*

Assumption 5.3: *The function $C_0(\theta, u_0)$ is locally uniformly monotonically decreasing, i.e. there exist $\gamma_c > 0$ and $\varepsilon_{u_0} > 0$ such that*

$$\left\langle C_0(\theta, u_0^1) - C_0(\theta, u_0^2), u_0^1 - u_0^2 \right\rangle_{V^*, V} \le -\gamma_c \|u_0^1 - u_0^2\|_V^2$$

for all $\theta \in B_{\varepsilon_\theta}(\theta^*)$ and $u_0^1, u_0^2 \in B_{\varepsilon_{u_0}}(\varphi_0(\theta^*)) := \{u_0 \in V \mid \|u_0 - \varphi_0(\theta^*)\|_V < \varepsilon_{u_0}\}$.

Assumption 5.4: *The sensitivity $S_0(\theta, u_0)$ is locally Lipschitz continuous with respect to u_0, i.e. there exists $L_{S_0} \ge 0$ such that*

$$\|S_0(\theta, u_0^1) - S_0(\theta, u_0^2)\|_{V^{n_\theta}} \le L_{S_0} \|u_0^1 - u_0^2\|_V$$

for all $\theta \in B_{\varepsilon_\theta}(\theta^*)$ and $u_0^1, u_0^2 \in B_{\varepsilon_{u_0}}(\varphi_0(\theta^*))$.

Assumption 5.5: *The mapping $\theta \mapsto g(\theta, \varphi_0(\theta))$ is uniformly monotonically decreasing on $B_{\varepsilon_\theta}(\theta^*)$, i.e. there exists $\gamma_g > 0$ such that*

$$(g(\theta, \varphi_0(\theta)) - g(\theta^*, \varphi_0(\theta^*)))^{\mathrm{T}}(\theta - \theta^*) \leq -\gamma_g \|\theta - \theta^*\|^2$$

for all $\theta \in B_{\varepsilon_\theta}(\theta^*)$.

Assumption 5.6: *The descent direction g is locally Lipschitz continuous with respect to u_0, i.e. there exists $L_g \geq 0$ such that*

$$\|g(\theta, u_0^1) - g(\theta, u_0^2)\| \leq L_g \|u_0^1 - u_0^2\|_V$$

for all $\theta \in B_{\varepsilon_\theta}(\theta^*)$ and $u_0^1, u_0^2 \in B_{\varepsilon_{u_0}}(\varphi_0(\theta^*)) \subseteq V$.
Moreover, g is uniformly bounded on $B_{\varepsilon_\theta}(\theta^*) \times B_{\varepsilon_{u_0}}(\varphi_0(\theta^*))$, i.e there exists $K_g \geq 0$ such that

$$\|g(\theta, u_0)\| \leq K_g$$

for all $(\theta, u_0) \in B_{\varepsilon_\theta}(\theta^*) \times B_{\varepsilon_{u_0}}(\varphi_0(\theta^*))$.

5.1. Elliptic PDE constraints

Using Assumptions 5.1–5.6 and the existence of a weak solution (Assumption 2.1), we can prove the following theorem on stability and convergence for the continuous analogue of the descent method for elliptic PDE constraints.

Theorem 5.1: *Let Assumptions 5.1–5.6 be satisfied. Then there exists a $\lambda^* \geq 0$ such that for all $\lambda > \lambda^*$ solutions to (17) are well defined for all $r > 0$ and the local minimizer (θ^*, u_0^*) of the optimization problem (5) is a locally exponentially stable steady state of the system (17).*

Proof: Define $\tilde{\theta} := \theta - \theta^*$ and $\tilde{u}_0 = u_0 - \varphi_0(\theta)$, with $\theta \in B_{\varepsilon_\theta}(\theta^*)$ and $u_0 \in B_{\varepsilon_{u_0}}(\varphi_0(\theta^*))$, where $\varphi_0(\theta^*) = u_0^*$ exists, because of Assumption 5.2. We further define a Lyapunov function $\mathcal{V}(r) = \frac{1}{2}\|\tilde{\theta}(r)\|^2 + \frac{1}{2}\|\tilde{u}_0(r)\|_H^2$. To prove Theorem 5.1, we will show that the Lyapunov function decreases exponentially. The derivative along the trajectories is given by

$$\frac{\mathrm{d}}{\mathrm{d}r}\mathcal{V}(r) = \frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{1}{2}\|\tilde{\theta}(r)\|^2\right) + \frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{1}{2}\|\tilde{u}_0(r)\|_H^2\right).$$

First, we bound the first summand from above, using (17), and Assumptions 5.1, 5.5 and 5.6,

$$\frac{\mathrm{d}}{\mathrm{d}r}\frac{1}{2}\|\tilde{\theta}\|^2 = \left(\frac{\mathrm{d}}{\mathrm{d}r}\tilde{\theta}\right)^{\mathrm{T}}\tilde{\theta}$$

$$= (g(\theta, u_0) - g(\theta, \varphi_0(\theta)))^{\mathrm{T}}\tilde{\theta} + (g(\theta, \varphi_0(\theta)) - g(\theta^*, \varphi_0(\theta^*)))^{\mathrm{T}}\tilde{\theta}$$

$$\leq \|g(\theta, u_0) - g(\theta, \varphi_0(\theta))\| \|\tilde{\theta}\| - \gamma_g\|\theta - \theta^*\|^2$$

$$\leq L_g\|\tilde{u}_0\| \|\tilde{\theta}\| - \gamma_g\|\tilde{\theta}\|^2.$$

Second, we bound the second summand from above, using (17) and $C_0(\theta, \varphi_0(\theta)) = 0$, as well as the fact that by Assumption 5.3 we have (9),

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}r} \frac{1}{2} \|\tilde{u}_0\|_H^2 &= \left(\frac{\mathrm{d}\tilde{u}_0}{\mathrm{d}r}, \tilde{u}_0 \right)_{V^*,V} = \left(\frac{\mathrm{d}u_0}{\mathrm{d}r} - \nabla_\theta \varphi_0(\theta) \frac{\mathrm{d}\theta}{\mathrm{d}r}, \tilde{u}_0 \right)_{V^*,V} \\
&= \left((S_0(\theta, u_0) - S_0(\theta, \varphi_0(\theta))) g(\theta, u_0), \tilde{u}_0 \right)_{V^*,V} \\
&\quad + \lambda \left(C_0(\theta, u_0) - C_0(\theta, \varphi_0(\theta)), \tilde{u}_0 \right)_{V^*,V} \\
&\leq \| (S_0(\theta, u_0) - S_0(\theta, \varphi_0(\theta))) \|_{V^{n_\theta}} \|g(\theta, u_0)\| \|\tilde{u}_0\|_V \\
&\quad + \lambda \left(C_0(\theta, u_0) - C_0(\theta, \varphi_0(\theta)), \tilde{u}_0 \right)_{V^*,V}.
\end{aligned}
$$

With Assumptions 5.3, 5.4 and 5.6, we get

$$
\frac{\mathrm{d}}{\mathrm{d}r} \frac{1}{2} \|\tilde{u}_0\|_H^2 \leq (L_{S_0} K_g - \lambda \gamma_c) \|\tilde{u}_0\|_V^2.
$$

Hence, we can estimate the derivative of the Lyapunov function,

$$
\frac{\mathrm{d}}{\mathrm{d}r} \mathcal{V}(r) \leq -(-L_{S_0} K_g + \lambda \gamma_c) \|\tilde{u}_0\|_V^2 + L_g \|\tilde{u}_0\|_V \|\tilde{\theta}\| - \gamma_g \|\tilde{\theta}\|^2.
$$

To show that $\mathcal{V}$ decays exponentially, we have to show that

$$
\frac{\mathrm{d}}{\mathrm{d}r} \mathcal{V}(r) \leq -a \mathcal{V}(r)
$$

for some $a > 0$. Based on our estimates, proving Theorem 5.1 reduces to finding $a > 0$ with

$$
0 \leq \left(-L_{S_0} K_g + \lambda \gamma_c - \frac{a}{2} \right) \|\tilde{u}_0\|_V^2 - L_g \|\tilde{u}_0\|_V \|\tilde{\theta}\| + \left(\gamma_g - \frac{a}{2} \right) \|\tilde{\theta}\|^2. \tag{20}
$$

We want this inequality to be valid without restrictions on $\|\tilde{\theta}\|$ or $\|\tilde{u}_0\|_V$. Due to the last term, we can therefore only consider values of a that are smaller than $2\gamma_g$. Hence, (20) is equivalent to

$$
0 \leq \left(\sqrt{\gamma_g - \frac{a}{2}} \|\tilde{\theta}\| - \frac{L_g}{2\sqrt{\gamma_g - \frac{a}{2}}} \|\tilde{u}_0\|_V \right)^2 + \left(-L_{S_0} K_g + \lambda \gamma_c - \frac{a}{2} - \frac{L_g^2}{4(\gamma_g - \frac{a}{2})} \right) \|\tilde{u}_0\|_V^2.
$$

Since the first term in the inequality is greater or equal to 0, we have to find $a > 0$ such that

$$
\lambda \gamma_c - \frac{a}{2} - L_{S_0} K_g - \frac{L_g^2}{4(\gamma_g - \frac{a}{2})} \geq 0.
$$

Multiplying with $4\gamma_g - 2a$, we obtain a quadratic inequality for a

$$
a^2 + 2(-\gamma_g - \lambda \gamma_c + L_{S_0} K_g)a + (4\lambda \gamma_c \gamma_g - 4L_{S_0} K_g \gamma_g - L_g^2) \geq 0.
$$

The roots of the quadratic polynomial are given by

$$
a_{1,2} = \gamma_g + \lambda \gamma_c - L_{S_0} K_g \pm \sqrt{d},
$$

with discriminant

$$d = (\gamma_g - \lambda\gamma_c + L_{S_0}K_g)^2 + L_g^2.$$

The discriminant is always positive, therefore, $a_1 = \gamma_g + \lambda\gamma_c - L_{S_0}K_g - \sqrt{d} < a_2 = \gamma_g + \lambda\gamma_c - L_{S_0}K_g + \sqrt{d}$ are real roots. In the following, we will assume that $a_1 > 0$, which can be achieved by choosing λ such that $\lambda > \lambda^* = L_{S_0}K_g/\gamma_c + L_g^2/(4\gamma_g\gamma_c) \geq 0$. This choice is justified as follows. As the square root, $\sqrt{d}$, is always positive, $\gamma_g + \lambda\gamma_c - L_{S_0}K_g > 0$, i.e. $\lambda > (L_{S_0}K_g - \gamma_g)/\gamma_c$ needs to hold to ensure $a_1 > 0$. Squaring both sides of the inequality

$$\gamma_g + \lambda\gamma_c - L_{S_0}K_g > \sqrt{d} \tag{21}$$

yields

$$(\gamma_g + \lambda\gamma_c - L_{S_0}K_g)^2 > (\gamma_g - \lambda\gamma_c + L_{S_0}K_g)^2 + L_g^2$$

$$\Leftrightarrow \lambda > \frac{L_g^2}{4\gamma_g\gamma_c} + \frac{L_{S_0}K_g}{\gamma_c}.$$

Taking

$$\lambda > \lambda^* := \max\left\{\frac{L_{S_0}K_g - \gamma_g}{\gamma_c}, \frac{L_g^2}{4\gamma_g\gamma_c} + \frac{L_{S_0}K_g}{\gamma_c}\right\} = \frac{L_g^2}{4\gamma_g\gamma_c} + \frac{L_{S_0}K_g}{\gamma_c}$$

ensures $a_1 > 0$.

Therefore, a either fulfils $0 < a < a_1$ with $a < 2\gamma_g$ or $a_2 < a < 2\gamma_g$, provided $\lambda > \lambda^*$. Hence, we distinguish the following three cases:

(1) $2\gamma_g < a_1 < a_2$,
(2) $a_1 < a_2 \leq 2\gamma_g$,
(3) $a_1 \leq 2\gamma_g \leq a_2$,

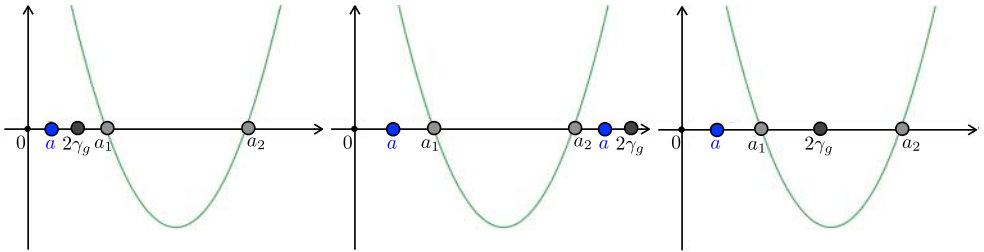for the relation of $2\gamma_g$, a_1 and a_2 as illustrated in Figure 2.

Figure 2. The function $f(a) = a^2 + (-2\gamma_g - 2\lambda\gamma_c + 2L_{S_0}K_g)a + (4\lambda\gamma_c\gamma_g - 4L_{S_0}K_g\gamma_g - L_g^2)$ is illustrated with the two roots a_1 and a_2 and the three different positions of $2\gamma_g$, as well as possible positions of a.

Case (1): $2\gamma_g < a_1$ is equivalent to

$$\lambda\gamma_c - L_{S_0}K_g - \gamma_g > \sqrt{d}.$$

If the term $\lambda\gamma_c - L_{S_0}K_g - \gamma_g$ is negative, the inequality cannot be valid. The term $\lambda\gamma_c - L_{S_0}K_g - \gamma_g$ is non-negative if $\lambda \geq (\gamma_g + L_{S_0}K_g)/\gamma_c$. In this case, we can square the inequality and get a contradiction ($0 > L_g^2$).

Case (2): $a_2 \leq 2\gamma_g$ is equivalent to

$$\sqrt{d} \leq \gamma_g + L_{S_0}K_g - \lambda\gamma_c.$$

This leads to a contradiction with the same arguments as in case (1).

Case (3): $a_1 \leq 2\gamma_g$ is equivalent to

$$-\gamma_g + \lambda\gamma_c - L_{S_0}K_g \leq \sqrt{d}.$$

The left-hand side $-\gamma_g + \lambda\gamma_c - L_{S_0}K_g$ is non-negative for all $\lambda \geq (\gamma_g + L_{S_0}K_g)/\gamma_c$. With squaring, we get $0 \leq L_g^2$. On the other hand, if the term $-\gamma_g + \lambda\gamma_c - L_{S_0}K_g$ is negative, that is $\lambda < (\gamma_g + L_{S_0}K_g)/\gamma_c$, we have $\sqrt{d} > 0$. This is true for all λ, because $d > 0$. In total, we find $a_1 \leq 2\gamma_g$ for all $\lambda > 0$. Analogously we get for a_2 that $a_2 \geq 2\gamma_g$ is fulfilled for all $\lambda > 0$. Hence, we know that $a_1 \leq 2\gamma_g \leq a_2$ holds for all $\lambda > 0$ and only case (3) is valid.

Altogether, we find that a lies in the interval $[0, a_1]$ provided $\lambda > \lambda^*$. In this case, it also holds that

$$\frac{d}{dr}\mathcal{V}(r) \leq -\frac{a}{2}\left(\|\tilde{u}_0(r)\|_V^2 + \|\tilde{\theta}(r)\|^2\right) \leq -\frac{a}{2K_{V\to H}^2}\|\tilde{u}_0(r)\|_H^2 - \frac{a}{2}\|\tilde{\theta}(r)\|^2 \leq -\tilde{a}\mathcal{V}(r),$$

with $\tilde{a} = -(a/2)\min\{1/K_{V\to H}^2, 1\}$, where $K_{V\to H}$ is the embedding constant. ∎

Remark 5.1: To tune the choice of the retraction factor λ, we now consider the fact that the value of a determines the speed at which $\mathcal{V}(r)$ decreases, thus a convenient choice of the retraction factor $\lambda > \lambda^*$ maximizes a to yield the fastest exponential decay. In our case, this means maximizing $a(\lambda) = a_1 = \gamma_g + \lambda\gamma_c - L_{S_0}K_g - \sqrt{d(\lambda)}$ with respect to λ. An elementary computation yields

$$\frac{da}{d\lambda}(\lambda) = \gamma_c + \frac{\gamma_g\gamma_c - \lambda\gamma_c^2 + L_{S_0}K_g\gamma_c}{\sqrt{d(\lambda)}} \geq 0$$

with equality iff $L_g = 0$, thus $\lambda \mapsto a(\lambda)$ is monotonically increasing (strictly, if $L_g > 0$) and therefore

$$\sup_{\lambda\in(\lambda^*,\infty)} a(\lambda) = \lim_{\lambda\to\infty} a(\lambda) = \lim_{\lambda\to\infty} \gamma_g + \lambda\gamma_c - L_{S_0}K_g - \sqrt{(\gamma_g - \lambda\gamma_c + L_{S_0}K_g)^2 + L_g^2}$$

$$= 2\gamma_g.$$

In case $L_g = 0$, we have $a(\lambda) = \gamma_g + \lambda\gamma_c - L_{S_0}K_g - |\gamma_g - \lambda\gamma_c + L_{S_0}K_g|$. Distinguishing the two cases for the absolute value yields the maximal value $a(\lambda) = 2\gamma_g$, attained at all $\lambda \geq (\gamma_g - L_{S_0}K_g)/\gamma_c$.

This shows that (unless $L_g = 0$) the exponential decay is maximized by choosing $\lambda > \lambda^*$ as large as possible. Nevertheless, in practice, λ should not be chosen too large in order to avoid stiffness of system (17).

Remark 5.2: The proof provides a lower bound for the retraction factor λ, namely $\lambda > \lambda^* = L_g^2/(4\gamma_g\gamma_c) + L_{S_0}K_g/\gamma_c$. In specific applications, it might not always be possible to explicitly compute all involved constants. If this is the case, an alternative Lyapunov function can be used to derive a lower bound for λ. A possible candidate for this Lyapunov function is

$$\mathcal{V}(r) = j(\theta(r)) - j(\theta^*) + \tfrac{1}{2}\|u_0 - \varphi_0(\theta(r))\|_H^2. \tag{22}$$

With this choice and analogous computations as above, different lower bounds involving different constants can be derived. The lower bound for the retraction factor can be estimated as $\lambda > \hat{\lambda}^* = \hat{L}/\gamma_c + \hat{L}_g^2/(4\gamma_c)$ with

$$\hat{L} = \begin{cases} \dfrac{((S_0(\theta, u_0) - S_0(\theta, \varphi_0(\theta)))g(\theta, u_0), u_0 - \varphi_0(\theta))_{V^*, V}}{\|u_0 - \varphi_0(\theta)\|_V^2}, & \text{if } u_0 \neq \varphi_0(\theta), \\ 0, & \text{else,} \end{cases} \tag{23}$$

and

$$\hat{L}_g = \begin{cases} \dfrac{g(\theta, \varphi_0(\theta))^{\mathrm{T}}(g(\theta, \varphi_0(\theta)) - g(\theta, u_0))}{\|u_0 - \varphi_0(\theta)\|_V\|g(\theta, \varphi_0(\theta))\|}, & \text{if } u_0 \neq \varphi_0(\theta) \text{ and } g(\theta, \varphi_0(\theta)) \neq 0, \\ 0, & \text{else.} \end{cases} \tag{24}$$

This bound depends on θ, the current parameter estimates during computation, and therefore requires a posteriori adaptation of the retraction factor. A practical implementation of such a retraction factor choice involves the evaluation of functionals of the residuals $g(\theta, \varphi_0(\theta)) - g(\theta, u_0), u_0 - \varphi_0(\theta)$ as well as sensitivities $(S_0(\theta, u_0) - S_0(\theta, \varphi_0(\theta)))g(\theta, u_0)$ (which can be done approximately, on a coarser computational mesh, and using adjoint techniques) is subject of future work.

Remark 5.3: If the cost function is locally convex but not locally uniformly convex, the parameters are identifiable but not strictly identifiable [19]. In this case, $g(\theta, \varphi_0)$ defined by (14) is not uniformly monotone, but just monotone, i.e. formally $\gamma_g = 0$. In this case, using a projection P on the orthogonal complement of the null space of $\nabla^2 j(\theta^*)$ might facilitate the proof of convergence on this subspace. A possible Lyapunov function in this case is given by $\mathcal{V}(r) = \tfrac{1}{2}\|Pg(\theta, \varphi_0(\theta))\|^2 + \tfrac{1}{2}\|u_0 - \varphi_0(\theta)\|_X^2$. Denoting the smallest positive eigenvalue of $(\mathrm{d}g/\mathrm{d}\theta)(\theta\varphi_0(\theta))$ with $\underline{\mu}$, we require that $\xi^{\mathrm{T}}P(\mathrm{d}g/\mathrm{d}\theta)(\theta, \varphi_0(\theta))\xi \leq -\underline{\mu}\|P\xi\|^2$ for all $\xi \in \mathbb{R}^{n_\theta}$. Then a retraction factor λ should be chosen $\lambda > \lambda^* = \hat{L}/\gamma_c + \hat{L}_g^2/(4\gamma_c\underline{\mu})$, with

$$\hat{L} = \begin{cases} \dfrac{((S_0(\theta, u_0) - S_0(\theta, \varphi_0(\theta)))g(\theta, u_0), u_0 - \varphi_0(\theta))_{V^*, V}}{\|u_0 - \varphi_0(\theta)\|_V^2}, & \text{if } u_0 \neq \varphi_0(\theta), \\ 0, & \text{else,} \end{cases} \tag{25}$$

and

$$\hat{L}_g = \begin{cases} \dfrac{g(\theta, \varphi_0(\theta))^{\mathrm{T}}(g(\theta, \varphi_0(\theta)) - g(\theta, u_0))}{\|u_0 - \varphi_0(\theta)\|_V \|g(\theta, \varphi_0(\theta))\|}, & \text{if } u_0 \neq \varphi_0(\theta) \text{ and } g(\theta, \varphi_0(\theta)) \neq 0, \\ 0, & \text{else.} \end{cases} \quad (26)$$

However, the null space of $\nabla^2 j(\theta^*)$ depends on the unknown optimal parameter θ^* and can in general not be assessed a priori, thus leaving this approach for further investigation.

Remark 5.4: Locality in Theorem 5.1 is only imposed in terms of the size of $\varepsilon_\theta, \varepsilon_{u_0}$ in Assumptions 5.2–5.6. As a matter of fact, Assumptions 5.2, 5.3, 5.4, 5.6 will typically hold for (θ, u_0) values in a larger neighbourhood of the solution, possibly after imposing certain constraints on the parameter values such as nonnegativity, in order to ensure that the model defined by C_0 is valid. Smallness of $\varepsilon_\theta, \varepsilon_{u_0}$ will mainly be required in Assumption 5.5 to guarantee, together with Assumption 5.1, that the descent direction defined by g really points towards $(\theta_*, \varphi_0(\theta_*))$.

To enlarge the convergence radius, globalization strategies for the solution of nonlinear equations or optimization problems can be employed. In particular, as the analysis in [20,21] indicates, the strategy of using a particular time-stepping algorithm together with an appropriate step size control in place of the continuous flow defined by (17), might enable to get rid of the locality assumption in Theorem 5.1. Indeed, with

$$F(\theta, u_0) = \begin{pmatrix} -d(\theta, u_0) \\ -\lambda C_0(\theta, u_0) \end{pmatrix}, \quad M(\theta, u_0) = \begin{pmatrix} H(\theta, u_0) & 0 \\ -S_0(\theta, u_0)H(\theta, u_0) & I \end{pmatrix}^{-1},$$

for the choice (15), method (17) can at least formally be cast into the framework of [20, Equation (5.1)], see also the introduction of [21] for the infinite dimensional setting relevant here. However, the detailed conditions from [21] would still have to be verified for our special setting, which will be subject of future research.

5.2. Elliptic and parabolic PDE constraints

As we consider the partially reduced form of the optimization problem with elliptic and parabolic PDE constraints (18), the results established for the elliptic problem can be easily transferred given the existence of a solution operator for the parabolic problem.

Theorem 5.2: *Let Assumptions* 2.1, 5.1–5.6 *be satisfied with g replaced by $\tilde{g}$ according to* (19). *Then there exists a $\lambda^* > 0$ such that for all $\lambda > \lambda^*$ solutions to*

$$\begin{aligned} \frac{\mathrm{d}\theta}{\mathrm{d}r}(r) &= \tilde{g}(\theta, u_0), \quad \theta(0) = \theta_0, \\ \frac{\mathrm{d}u_0}{\mathrm{d}r}(r) &= S_0(\theta, u_0)\tilde{g}(\theta, u_0) + \lambda C_0(\theta, u_0), \quad u_0(r) = u_{0,0}, \end{aligned} \quad (27)$$

are well defined for all $r > 0$ and the local minimizer (θ^*, u_0^*) of the optimization problem (4) *is a locally exponentially stable steady state of the system* (27).

With the setting introduced in the last paragraph of Section 4, the result directly follows from Theorem 5.1.

6. Discussion of the assumptions

In this section, we discuss Assumptions 5.1–5.6 in more detail. In particular, we provide sufficient conditions for these assumptions which show that the assumptions are rather weak and fulfilled by many application problems.

Remark 6.1 (Assumption 5.1): For the choice (14) of g, this condition is satisfied due to the identity

$$0 = \nabla_\theta j(\theta^*) = -\nabla_\theta J(\theta, \varphi(\theta^*)) - \frac{\partial J}{\partial u_0}(\theta, \varphi(\theta^*)) S_0(\theta, \varphi(\theta^*)).$$

This remains valid more generally for descent directions defined by (15), such as Newton-type methods.

Remark 6.2 (Assumptions 5.2 and 5.3): Dynamical systems in engineering, physics and life sciences will typically exhibit locally exponentially stable steady-state solutions. This implies that these systems are locally uniformly monotone. This is especially true for the optimal parameters θ^* and thus (for smooth $C_0(\theta, u_0)$) also in a neighbourhood of $(\theta^*, \varphi_0(\theta^*))$. Hence, around the true parameters, Assumption 5.3 and thus also the monotonicity part of Assumption 5.2 is locally fulfilled for most real-world systems.

An additional consequence of uniform monotonicity, hemicontinuity and coercivity of C_0 according to Assumption 5.2 is uniform monotonicity, hemicontinuity and coercivity of its linearization and thus, by the Theorem of Browder and Minty, existence and boundedness of the inverse $((\partial C_0/\partial u_0)(\theta^*, u_0^*))^{-1} : V^* \to V$, is assured, as will be used below.

Remark 6.3 (Assumption 5.4): The Lipschitz continuity of S is satisfied for example if

- $\partial C_0/\partial u_0 : \mathbb{R}^{n_\theta} \times V \to \mathcal{L}(V, V^*)$ is continuous,
- $\nabla_\theta C_0$ and $\partial C_0/\partial u_0$ are locally Lipschitz continuous with respect to u_0 in a neighbourhood of (θ^*, u_0^*), with Lipschitz constants $L_{C_0,\theta}$ and L_{C_0,u_0}, respectively,
- $\nabla_\theta C_0$ is uniformly bounded on this neighbourhood by $K_{C_0,\theta}$,
- $(\partial C_0/\partial u_0)(\theta^*, u_0^*)^{-1}$ is bounded by K_{C_0,u_0}, as discussed in Remark 6.2,

provided that C_0 is continuously (Frèchet) differentiable. This can be seen as follows.
We want to show that $\|S_{0_i}(\theta, u_0^1) - S_{0_i}(\theta, u_0^2)\|_V \leq L_{S_i} \|u_0^1 - u_0^2\|_V$ for all $i \in \{1, \ldots, n_\theta\}$, $\theta \in B_{\varepsilon_\theta}(\theta^*)$ and $u_0^1, u_0^2 \in B_{\varepsilon_u}(\varphi_0(\theta^*))$, where S_{0_i} is the ith component of S_0.
Using $S_{0_i}(\theta, u_0) = ((\partial C_0/\partial u_0)(\theta, u_0))^{-1}((\partial C_0/\partial \theta_i)(\theta, u_0))$ for $i \in \{1, \ldots, n_\theta\}$ yields

$$\begin{aligned}
&\left\| S_{0_i}(\theta, u_0^1) - S_{0_i}(\theta, u_0^2) \right\|_V \\
&\leq \left\| \left(\frac{\partial C_0}{\partial u_0}(\theta, u_0^1) \right)^{-1} \right\|_{V^* \to V} \left\| \frac{\partial C_0}{\partial \theta_i}(\theta, u_0^1) - \frac{\partial C_0}{\partial \theta_i}(\theta, u_0^2) \right\|_{V^*} \\
&\quad + \left\| \left(\frac{\partial C_0}{\partial u_0}(\theta, u_0^1) \right)^{-1} - \left(\frac{\partial C_0}{\partial u_0}(\theta, u_0^2) \right)^{-1} \right\|_{V^* \to V} \left\| \frac{\partial C_0}{\partial \theta_i}(\theta, u_0^2) \right\|_{V^*}.
\end{aligned}$$

With $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$, we can further estimate

$$\left\| S_{0_i}(\theta, u_0^1) - S_{0_i}(\theta, u_0^2) \right\|_V$$

$$\leq \left\| \left(\frac{\partial C_0}{\partial u_0}(\theta, u_0^1) \right)^{-1} \right\|_{V^* \to V} \left(\left\| \frac{\partial C_0}{\partial \theta_i}(\theta, u_0^1) - \frac{\partial C_0}{\partial \theta_i}(q, u_0^2) \right\|_{V^*} \right.$$

$$\left. + \left\| \frac{\partial C_0}{\partial u_0}(\theta, u_0^2) - \frac{\partial C_0}{\partial u_0}(\theta, u_0^1) \right\|_{V \to V^*} \left\| \left(\frac{\partial C_0}{\partial u_0}(\theta, u_0^1) \right)^{-1} \right\|_{V^* \to V} \left\| \frac{\partial C_0}{\partial \theta_i}(\theta, u_0^2) \right\|_{V^*} \right).$$

The fact that the inverses $((\partial C_0/\partial u_0)(\theta, u_0^1))^{-1}$, $((\partial C_0/\partial u_0)(\theta, u_0^2))^{-1}$ exist and are bounded follows from the regularity of $(\partial C_0/\partial u_0)(\theta^*, \varphi_0(\theta^*))$ and a perturbation argument. Taking $L = (\partial C_0/\partial u_0)(\theta^*, \varphi_0(\theta^*))$ and $M_i = (\partial C_0/\partial u_0)(\theta, u_0^i)$, $i = 1,2$ and using that by continuity of $(\partial C_0/\partial u_0)(\theta, u_0)$ in a neighbourhood of (θ^*, u_0^*) and by possibly decreasing $\varepsilon_\theta, \varepsilon_{u_0}$, we get

$$\|M_i - L\| = \left\| \frac{\partial C_0}{\partial u_0}(\theta, u_0^i) - \frac{\partial C_0}{\partial u_0}(\theta^*, \varphi_0(\theta^*)) \right\|_{V \to V^*} < \frac{1}{\|L^{-1}\|_V}$$

for any $(\theta, u_0^i) \in B_{\varepsilon_\theta}(\theta^*) \times B_{\varepsilon_u}(\varphi_0(\theta^*))$. Therewith the operator $M_i^{-1} = ((\partial C_0/\partial u_0)(\theta, u_0))^{-1}$ exists and is bounded by

$$\|M_i^{-1}\|_V = \left\| \left(\frac{\partial C_0}{\partial u_0}(\theta, u_i) \right)^{-1} \right\|_V \leq \frac{\|L^{-1}\|}{1 - \|L^{-1}\|\|L - M_i\|} \leq K_{C_0, u_0}.$$

Using this bound and local Lipschitz continuity with respect to u_0 of the derivatives $\partial C_0/\partial \theta_i$ and $\partial C_0/\partial u_0$ results in

$$\left\| S_{0_i}(\theta, u_0^1) - S_{0_i}(\theta, u_0^2) \right\|_V \leq K_{C_0, u_0} \left(L_{C_0, \theta_i} + K_{C_0, u_0} K_{C_0, \theta_i} L_{C_0, u_0} \right) \left\| u_0^1 - u_0^2 \right\|_V$$

and hence

$$\left\| S_0(\theta, u_0^1) - S_0(\theta, u_0^2) \right\|_V \leq L_{S_0} \left\| u_0^1 - u_0^2 \right\|_V$$

with $L_{S_0} = \sum_{i=1}^{n_\theta} \{ K_{C_0, u_0}(L_{C_0, \theta_i} + K_{C_0, u_0} K_{C_0, \theta_i} L_{C_0, u_0}) \}$, where K_{C_0, θ_i} denotes the bound of the derivative w.r.t. θ_i of the steady-state residual.

Remark 6.4 (Assumption 5.5): If the descent direction g is defined by (14), then, according to [19], Assumption 5.5 is equivalent to practical identifiability, which implies local structural identifiability of the parameter vector θ. The same holds true for more general choices (15), such as Newton-type methods, under a uniform positivity condition on $H(\theta, u_0)$.

If local structural identifiability cannot be guaranteed, one can still use regularization [22], e.g. by adding a term $\alpha(\theta - \theta_p)^T \Gamma^{-1}(\theta - \theta_p)$ with positive definite Γ and positive α to the cost function $\mathcal{J}(\theta, u_0)$ in (5) for which (17) yields a minimizer $(\theta^*(\alpha), u_0^*(\alpha))$. Regularization theory, e.g. [22], provides convergence of $\theta^*(\alpha)$ to a parameter $\bar{\theta}$ that is consistent with the observations as $\alpha \to 0$.

Remark 6.5 (Assumption 5.6): As an example, we check Assumption 5.6 for the case of a gradient descent-based update of θ, i.e. g according to (14). We find

$$\left\| g(\theta, u_0^1) - g(\theta, u_0^2) \right\|$$

$$= \left\| -\nabla_\theta J(\theta, u_0^1) - \frac{\partial J}{\partial u_0}(\theta, u_0^1) S_0(\theta, u_0^1) + \nabla_\theta J(\theta, u_0^2) + \frac{\partial J}{\partial u_0}(\theta, u_0^2) S_0(\theta, u_0^2) \right\|$$

$$\leq \left\| \nabla_\theta J(\theta, u_0^1) - \nabla_\theta J(\theta, u_0^2) \right\| + \left\| \frac{\partial J}{\partial u_0}(\theta, u_0^1) - \frac{\partial J}{\partial u_0}(\theta, u_0^2) \right\|_{V \to \mathbb{R}} \left\| S_0(\theta, u_0^1) \right\|_{V^{n_\theta}}$$

$$+ \left\| \frac{\partial J}{\partial u_0}(\theta, u_0^2) \right\|_{V \to \mathbb{R}} \left\| S_0(\theta, u_0^1) - S_0(\theta, u_0^2) \right\|_{V^{n_\theta}}.$$

Here, S_0 is Lipschitz continuous according to Remark 6.3. Also boundedness of S_0 is fulfilled under the assumptions made in Remark 6.3, namely that $((\partial C_0/\partial u_0)(\theta, u_0))^{-1}$ and $\nabla_\theta C_0$ are locally uniformly bounded.

$$\left\| S_0(\theta, u_0^1) \right\|_{V^{n_\theta}} \leq \left\| \left(\frac{\partial C_0}{\partial u_0}(\theta, u_0^1) \right)^{-1} \right\|_{V^* \to V} \left\| \nabla_\theta C_0(\theta, u_0^1) \right\|_{V^{*n_\theta}}$$

$$\leq K_{C_0,u_0} K_{C_0,\theta} =: K_{S_0}. \tag{28}$$

If J is differentiable and the derivatives $\nabla_\theta J$ and $\partial J/\partial u_0$ are locally Lipschitz continuous with respect to u_0 with Lipschitz constants $L_{J,\theta;u_0}$ and $L_{J,u_0;u_0}$, respectively, and $\partial j/\partial u_0$ is uniformly bounded on $B_{\varepsilon_\theta}(\theta^*) \times B_{\varepsilon_u}(\varphi_0(\theta^*))$, we can conclude

$$\left\| g(\theta, u_0^1) - g(\theta, u_0^2) \right\| \leq L_g \| u_0^1 - u_0^2 \|_V,$$

with $L_g = L_{J,\theta;u_0} + K_{J,u_0} L_{S_0} + L_{J,u_0;u_0} K_{S_0}$.

The more general choice (15) requires additionally Lipschitz continuity of $H(\theta, u_0)$, which in case of Newton-type methods with (16) amounts to higher smoothness of J and C_0.

Thus we have shown the following proposition.

Proposition 6.1: *Let Assumption 6.1 be satisfied. Then there exist $\varepsilon_\theta, \varepsilon_{u_0} > 0$ such that on $B_{\varepsilon_\theta}(\theta^*) \times B_{\varepsilon_u}(\varphi_0(\theta^*))$, Assumptions 5.1–5.6 are fulfilled.*

Assumption 6.1:

(1) (a) $C_0(\theta, \cdot)$ is locally uniformly monotonically decreasing.
   (b) C_0 is continuous and continuously (Fréchet) differentiable in a neighbourhood of (θ^*, u_0^*).
   (c) $\nabla_\theta C_0$ and $\partial C_0/\partial u_0$ are locally Lipschitz continuous w.r.t. u_0 and $\nabla_\theta C_0$ is bounded in a neighbourhood of (θ^*, u_0^*).
(2) $g(\theta, u_0)$ is defined by (14) and θ is locally structurally identifiable.
(3) (a) J is continuously (Fréchet) differentiable in a neighbourhood of (θ^*, u_0^*).
   (b) $\partial J/\partial u_0$ and $\nabla_\theta J$ are locally Lipschitz continuous with respect to u_0.
   (c) $\partial J/\partial u_0$ is bounded in a neighbourhood of (θ^*, u_0^*).

7. Application

To illustrate the continuous analogue of the descent method, we use it to study CCL21 gradient formation in biological tissues. This process is highly relevant in immune response [23,24] and described by a reaction–diffusion equation [25]. In the following, we outline the model, estimate its parameters using the approach proposed in this paper and analyse the results.

7.1. Model formulation

CCL21 gradients are necessary for the guidance of dendritic cells towards lymphatic vessels [26]. They are formed by the combination of several biological processes. The chemokine CCL21 is produced in the lymphatic vessels, which cover a subset domain Ω_L of the domain Ω of interest, $\Omega_L \subset \Omega$. The source term is defined via the function

$$Q(x) = \begin{cases} 1, & \text{for } x \in \Omega_L, \\ 0, & \text{otherwise.} \end{cases}$$

The concentration of free CCL21 is denoted by u. Free CCL21 binds to a sugar whose concentration is denoted by s. The binding yields immobilized CCL21 whose concentration is denoted by c. The parameters $k_1, k_{-1}, D, \gamma, \alpha$ denote the binding and unbinding rates, the diffusion coefficient, the degradation rate and the production rate of CCL21 from the lymphatic vessels, respectively. A PDE model for the process has been developed in [25] and is given by

$$
\begin{aligned}
u_t - D\Delta u &= \alpha Q - k_1 us + k_{-1}c - \gamma u, \\
\dot{s} &= -k_1 us + k_{-1}c, \\
\dot{c} &= k_1 us - k_{-1}c,
\end{aligned}
\tag{29}
$$

for $t \in]0, T[$ and $x \in \Omega$, with initial conditions $u(0, x) = c(0, x) = 0$, $s(0, x) = s_0$ and no-flux boundary conditions $(\partial/\partial \nu)u = 0$ where ν is the outer normal on Ω. The parameter s_0 denotes the initial sugar concentration.

As the formation of the gradient is apparently fast, we consider the steady state of (29). By considering the PDE for the time evolution of s and c, we find that $c = s_0 u_s/(1 + u_s)$ and $s = s_0/(u_s + 1)$ with the scaled CCL21 concentration $u_s := k_1 u/k_{-1}$. Using the additional reformulation $\tilde{D} = D/\gamma$, $\tilde{\alpha} = \alpha k_1/(\gamma k_{-1})$, the scaled steady-state concentration of CCL21, u_s, has to fulfil $0 = \tilde{D}\Delta \tilde{\Delta} u_s + \tilde{\alpha} Q - u_s$ and the boundary conditions $(\partial/\partial \nu)u_s = 0$.

For the considered process, imaging data have been collected [27]. These images provide information about the localization of the lymphatic vessels (encoded in Q) and the concentration of immobilized CCL21. As the measured intensity values are corrupted by background fluorescence and as the data are not normalized, we model the readout following [25] as

$$y_i = s_l \left(b + \int_{A_i} c(t, x)\, dx \right),$$

where b denotes the intensity of the background fluorescence, s_l is a scaling constant and $A_i \subset \Omega$ is the domain of the pixel k. As the parameters are structurally non-identifiable,

we reformulate the models in terms of $\tilde{b} = s_l b$ and $\tilde{s}_0 = s_l s_0$ in the parameter estimation to one parameter and just consider $\tilde{b}$ and $\tilde{s}_0$.

The optimization problem is then given by

$$\min_{\theta, u_s} \quad J(\theta, u_s) = \frac{1}{2} \left\{ \sum_{i=1}^{M} \log(2\pi \sigma_i^2 \bar{y}_i) + \left(\frac{\log(\bar{y}_i) - \log(y_i)}{\sigma_i} \right)^2 \right\} \qquad (30)$$

$$\text{s.t.} \quad -\tilde{D}\Delta u_s + u_s = \tilde{\alpha} Q, \quad x \in \Omega$$

$$\frac{\partial}{\partial \nu} u_s = 0, \quad x \in \partial\Omega$$

$$y_i = \tilde{b} + \int_{A_i} \frac{\tilde{s}_0 u_s(x)}{u_s(x) + 1} \, dx = \tilde{b} + \tilde{s}_0 h_i(u_s),$$

where $\Omega \subseteq \mathbb{R}^2$, σ_i is the scale parameter of the log-normally distributed measurement error and $h_i(u_s) = \int_{A_i} (u_s(x)/(u_s(x) + 1)) \, dx$, $i = 1, \ldots, M$. The parameter vector θ is given by $\theta = (\tilde{D}, \tilde{\alpha}, \tilde{s}_0, \tilde{b}, \sigma) \in \mathbb{R}^{n_\theta}$, with $n_\theta = 5$.

All parameters are assumed to be non-negative due to their biological meaning. The spaces V and V^* for which we examine the problem are $V = H^1(\Omega)$ and $V^* = H^1(\Omega)^*$. The operator C_0 is given by $C_0(\theta, u_s) = \tilde{D}\Delta u_s - u_s + \tilde{\alpha} Q$. For these spaces and operators, it can easily be checked that all assumptions for applying the method (17) are satisfied.

7.2. Numerical implementation

For the numerical simulation of the biological process, we employed a finite element discretization of the PDE model. The discretization was obtained using the MATLAB PDE toolbox and accounts for the stationary topology of the model (Figure 3A). The mesh consists of 2170 elements and the concentrations in these elements are the state variables of the discretization. For parameter optimization using the coupled ODE–PDE model (17), the same mesh was employed and the states of the discretized PDE were coupled with
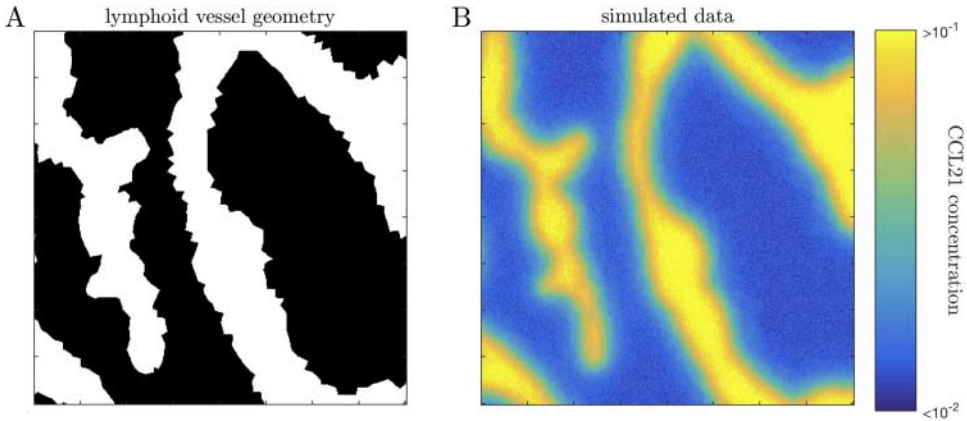

Figure 3. (A) Geometry of a lymphoid vessel obtained from biological imaging data [27]. (B) Simulated data of the CCL21 gradient generated by simulating model (29).

the ODE for the parameters. This yields a model with 2170+5 equations. The simulation-based method for parameter estimation was implemented in MATLAB extending the routine published in [16]. The numerical simulation was performed using the MATLAB ODE solver ode15s, an implicit scheme applicable to stiff problems. To accelerate the calculations, we implemented the Jacobian of the coupled ODE–PDE model. The simulation of the continuous analogue was terminated, if the gradient of the right-hand side became small, i.e. $\|C_0(\theta, u_0)\|_{V^*}/\|u_0\|_V \leq 10^{-6}$. Furthermore, simulations were interrupted whenever the objective function value became complex, which can happen due to the log-transformation of the output.

7.3. Simulated data

To evaluate the convergence properties of the proposed algorithm for the models, we considered published simulated data for the ground truth (similar to [28]). The geometry of lymphatic tissue was extracted from the available imaging data [27] using the MATLAB PDE toolbox. On this geometry, the discretized PDE was simulated using biologically plausible parameter values (Table 1). The simulated data for CCL21 gradient formation process were corrupted by noise to obtain a plausible scenario (Figure 3B).

7.4. Optimization

The objective function for most parameter estimation problems is non-convex and can be multimodal. For this reason, we employed multi-start local optimization using the continuous analogue for which we have established local convergence in this paper. The starting points for the local optimizations were sampled using a latin hypercube approach with lower and upper bounds provided in Table 1. We used a linear parametrization for the states and a log-parametrization $\xi = \log(\theta)$ for the parameters following previous evaluations for biochemical systems [29]. We did not implement any bounds for values of parameter or states. The implementation of the multi-start local optimization is based upon the MATLAB toolbox PESTO [30]. The implementation of the objective function and finite element schemes was adapted from [25]. For the local optimization with the continuous analogue, we chose the negative gradient as descent direction.

As a reference, we performed also multi-start local optimization using a discrete iterative optimization method. We used the state-of-the-art optimizer `fmincon.m` with the starting points sampled for the continuous analogue and the interior point algorithm implemented in the MATLAB Optimization Toolbox. This interior point algorithm employs either a Newton step, where the Hessian is approximated by the

Table 1. True parameters, estimated parameters and parameter ranges for the latin hypercube sampling for the CCL21 model.

Name	True value	Estimates	Lower bound sampling	Upper bound sampling
D	$8.50\ 10^1$	$8.51\ 10^1$	$1.50\ 10^{-2}$	$2.50\ 10^4$
α	$2.40\ 10^{-1}$	$2.36\ 10^{-1}$	$4.50\ 10^{-5}$	$2.00\ 10^1$
s_0	$2.10\ 10^{-1}$	$2.13\ 10^{-1}$	$2.50\ 10^{-3}$	$3.00\ 10^0$
b	$1.30\ 10^{-2}$	$1.30\ 10^{-2}$	$2.50\ 10^{-3}$	$1.00\ 10^0$
σ	$5.00\ 10^{-2}$	$4.99\ 10^{-2}$	$2.50\ 10^{-3}$	$3.00\ 10^0$

Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, or a conjugate gradient step using a trust region [31–33]. The optimizer was provided with the objective function, the non-linear constraint, as well as the corresponding derivatives. We used the same parametrization as for the continuous analogue and additionally constrained parameter values for $\xi = \log(\theta)$ in the optimization by the same upper and lower bounds used for the sampling (Table 1). The value of u_0 at the nodes of the mesh for the finite element discretization was constrained using upper and lower bounds for the optimization to lie in $[-1, 3]$. A total of 2000 iterations and 4000 function evaluations was allowed.

7.5. Comparison of continuous analogue and discrete iterative procedure

We performed 100 local optimization runs with the continuous analogue for a retraction factor $\lambda = 10^7$ and a discrete iterative method (Figure 4A). Both methods found the same best parameter value (Table 1) and achieved a good fit to the data. The assessment of the results revealed a good convergence of the continuous analogue. Almost 90% of the runs achieved an objective function value which was comparable with the best objective
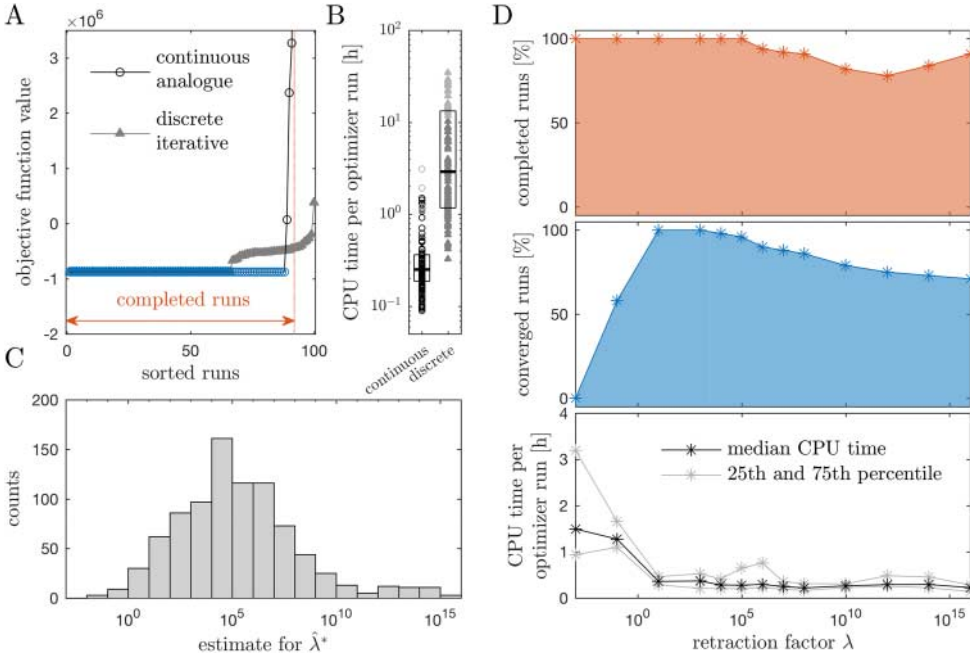

Figure 4. Results of parameter estimation for CCL21 model. (A) Sorted objective function values for the multi-start optimization with continuous analogue ($\lambda = 10^7$) and discrete iterative procedure. Converged runs are indicated in blue. (B) CPU time needed per optimizer run for the optimization using the continuous analogue and the discrete iterative procedure (lighter grey colour indicates runs which stopped because the maximal number of iterations was reached). The box covers the range between the 25th and the 75th percentile of the distribution. The median CPU time is indicated by a line. (C) Histogram of values for $\hat{\lambda}^*$ (Remark 5.2) obtained for 1000 points sampled in parameter-state space. (D) Percentage of completed runs (top), converged runs (middle) and median as well as 25th and 75th percentile of the runtime of completed runs (bottom) for different values of λ. For each value of λ, 100 local optimization runs were performed.

function value found across all runs (relative difference $< 0.001\%$). Overall, 96% of the runs finished successfully, meaning that either the optimization was stopped because the stopping criterion was fulfilled or the maximum number of iterations was reached, while 4% of the runs stopped prematurely.

The discrete iterative optimization converged for 66% of the runs to the optimal value (Figure 4A). Accordingly, the success rate was substantially lower than for the proposed continuous analogue. Of the runs which did not converge to the global optimum 25 runs were stopped because the maximal number of iterations was reached.

For the considered problem, the continuous analogue outperformed the discrete iterative method regarding the CPU time (Figure 4B). We found a median CPU time of 15 minutes for the continuous solver and 174 minutes for the discrete iterative procedure. In light of the fact that the discrete iterative method uses second-order information, it is interesting to observe that a continuous analogue using the negative gradient is more efficient. One possible explanation is that the efficiency of the continuous analogue is a result of the application of sophisticated numerical solvers. The adaptive, implicit solver `ode15s`, which is provided with the analytical Jacobian of the ODE–PDE model, might facilitate large step-sizes and fast convergence. Indeed, the Jacobian also provides second-order information.

7.6. Evaluation of retraction factor influence

As an analytical calculation of the bound for the retraction factor was not possible, we sampled 1000 points in parameter-state space and evaluated the estimate for the lower bound $\hat{\lambda}^*$ (Remark 5.2). The histogram of the resulting values for $\hat{\lambda}^*$ is presented in Figure 4C. The values for $\hat{\lambda}^*$ span many orders of magnitude, and the distribution peaks at 10^4. This result indicated that for different starting points very different retraction factors might be ideal.

To investigate the convergence properties for the different values of the retraction factor λ, we performed 100 local optimization runs for a range of different retraction factors. For each retraction factor, we assessed the number of completed runs and the number of converged runs (Figure 4D). Interestingly, as λ was increased the percentage of completed runs decreased. Yet, for large retraction factors many of the completed runs also converged, while for small retraction factors no runs converged as the maximum number of iterations becomes too large. The median CPU time for the optimization of one run decreased for increasing values of λ (Figure 4D). Notably, for the small values, the median CPU time was nearly six to seven times higher than the smallest one. The quantiles indicate that also the variability was higher for small values of λ. These results indicated that the retraction factor should be chosen large enough but not too large.

In summary, the analysis of the model of CCL21 gradient formation revealed that the retraction factor λ has a substantial influence on the convergence properties as well as the run time. For low values of λ starts did not converge while for large values of λ increasing stiffness of the problem could be observed. In an intermediate regime, which could here also be found by random sampling, we found the best convergence properties.

8. Conclusion and outlook

Parameter estimation is an important problem in a wide range of applications. Robustness and performance of the available iterative methods is, however, often limited. In

this study, we introduced continuous analogues of descent methods for optimization with PDE constraints. For these continuous analogues, we proved local convergence of their solutions to the optima. The necessary assumptions are fulfilled for a wide range of application problems, rendering the results interesting for several research fields.

We demonstrated the applicability of continuous analogues for a model of gradient formation in biological tissues and compared them with an iterative discrete procedure. The results highlight the potential of the continuous analogues, e.g. a high convergence rate and lower computation times than the discrete iterative procedure. For the comparison, we used the MATLAB optimization routine `fmincon.m`, a state-of-the-art discrete iterative procedure. Alternatives would be IPOPT or KNITRO. As `fmincon` is a generic interior point method, there might apparently be approaches which are efficient for the considered PDE-constrained problems (see also the no free lunch theorem [34]). The evaluation of the influence of the retraction factor revealed the importance of an appropriate choice of the retraction factor as well as the issue of premature stopping. In this study, we provide a lower bound for λ which ensures local convergence. As this bound might, however, be conservative and can only be assessed pointwise, the use of adaptive methods might be interesting. To address the issue of premature stopping, bounds for parameters and state variables have to be implemented, e.g. by including log-barrier functions [35] in the objective function or through projection into the feasible space.

In the application problem, we only considered elliptic PDE constraints as for the proposed continuous analogues parabolic constraints can be encapsulated in the objective function. This changes the objective function landscape and indirectly influences the convergence. Conceptually, it should also be possible to formulate continuous analogues which do not require a solution operator for the parabolic PDE but also have the solution of the parabolic PDE as a state variable. This mathematically more elegant approach is left for future research.

In conclusion, this study presented continuous analogues for a new problem class. Similar to other problem classes for which continuous analogues have been established [15,16], we expect an improvement of convergence and computation time. The continuous analogues for optimization also complement recent work on simulation-based methods for uncertainty analysis [36]. The efficient implementation of these methods in easily accessible software packages should be a focus of future research as it would render the methods available to a broad community.

The method and its analysis apply as they are to the case of infinite dimensional parameters θ. However, in that situation, the inverse problem of identifying θ is often ill-posed, so the assumption of practical identifiability (cf. Assumption 5.5 and Remark 6.4) might not be satisfied. To restore stability, regularization can be employed, as pointed out in Remark 6.4.

Future research in this context will be concerned with globalization strategies, such as those proposed in [20,21], cf. Remark 5.4.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

References

[1] Isakov V. Inverse problems for partial differential equations. 2nd ed. New York, NY: Springer; 2006. (Applied Mathematical Sciences; 127).

[2] Tarantola A. Inverse problem theory and methods for model parameter estimation. Philadelphia: SIAM Society for Industrial and Applied Mathematics; 2005.

[3] Banks H, Kunisch K. Estimation techniques for distributed parameter systems. Boston, Basel, Berlin: Birkhäuser; 1989. (Systems & Control: Foundations & Applications; 1).

[4] Bock H, Carraro T, Jäeger W, et al. Model based parameter estimation: theory and applications. Berlin, Heidelberg: Springer; 2013. (Contributions in Mathematical and Computational Sciences; 4).

[5] Carvalho EP, Martínez J, Martínez JM, et al. On optimization strategies for parameter estimation in models governed by partial differential equations. Math Comput Simul. 2015;114:14–24.

[6] Hinze M, Pinnau R, Ulbrich M, et al. Optimization with PDE constraints. Netherlands: Springer; 2009. (Mathematical Modelling: Theory and Applications; 23).

[7] Ito K, Kunisch K. Lagrange multiplier approach to variational problems and applications. Philadelphia: SIAM Society for Industrial and Applied Mathematics; 2008. (Advances in Design and Control; 15).

[8] Xun X, Cao J, Mallick B, et al. Parameter estimation of partial differential equation models. J Am Stat Assoc. 2013;108:1009–1020.

[9] Nielsen BF, Lysaker M, Grøttum P. Computing ischemic regions in the heart with the bidomain model – first steps towards validation. IEEE Trans Med Imaging. 2013 Jun;32(6):1085–1096.

[10] Airapetyan RG, Ramm AG, Smirnova AB. Continuous methods for solving nonlinear ill-posed problems. Operator Theory Appl. 2000;25:111.

[11] Kaltenbacher B, Neubauer A, Ramm AG. Convergence rates of the continuous regularized Gauss-Newton method. J Inv Ill-Posed Problems. 2002;10:261–280.

[12] Tanabe K. Continuous Newton–Raphson method for solving and underdetermined system of nonlinear equations. Nonlinear Anal Theory Methods Appl. 1979;3(4):495–503.

[13] Tanabe K. A geometric method in nonlinear programming. J Optim Theory Appl. 1980;30(2):181–210.

[14] Watson L, Bartholomew-Biggs M, Ford J. Optimization and nonlinear equations. Journal of computational and applied mathematics. Vol. 124. Amsterdam: The Netherlands; 2001.

[15] Tanabe K. Global analysis of continuous analogues of the Levenberg–Marquardt and Newton–Raphson methods for solving nonlinear equations. Ann Inst Statist Math. 1985;37(Part B):189–203.

[16] Fiedler A, Raeth S, Theis FJ, et al. Tailored parameter optimization methods for ordinary differential equation models with steady-state constraints. BMC Syst Biol. 2016 Aug;10:80.

[17] Rosenblatt M, Timmer J, Kaschek D. Customized steady-state constraints for parameter estimation in non-linear ordinary differential equation models. Front Cell Dev Biol. 2016;4:41.

[18] Zeidler E. Nonlinear functional analysis and its applications II/B: nonlinear monotone operators. New York, NY: Springer; 1990.

[19] Faller D, Klingmüller U, Timmer J. Simulation methods for optimal experimental design in systems biology. Simulation. 2003;79(12):717–725.

[20] Potschka A. Backward step control for global Newton-type methods. SIAM J Numer Anal. 2016;54(1):361–387.

[21] Potschka A. Backward step control for Hilbert space problems. 2017; ArXiv:1608.01863 [math.NA].

[22] Engl H, Hanke M, Neubauer A. Regularization of inverse problems. Dordrecht: Kluwer Academic Publishers; 2000. (Mathematics and Its Applications; 375).

[23] Alvarez D, Vollmann EH, von Andrian UH. Mechanisms and consequences of dendritic cell migration. Immunity. 2008;29(3):325–342. 2017/05/19.

[24] Mellman I, Steinman RM. Dendritic cells. Cell. 2001;106(3):255–258. 2017/05/19.

[25] Hock S, Hasenauer J, Theis FJ. Modeling of 2D diffusion processes based on microscopy data: parameter estimation and practical identifiability analysis. BMC Bioinformatics. 2013;14(10):1–6.

[26] Schumann K, Lammermann T, Bruckner M, et al. Immobilized chemokine fields and soluble chemokine gradients cooperatively shape migration patterns of dendritic cells. Immunity. 2010 May;32(5):703–713.

[27] Weber M, Hauschild R, Schwarz J, et al. Interstitial dendritic cell guidance by haptotactic chemokine gradients. Science. 2013;339:328–332.

[28] Hross SParameter estimation and uncertainty quantification for reaction-diffusion models in image based systems biology [dissertation]. München: Technische Universität München; 2016.

[29] Raue A, Schilling M, Bachmann J, et al. Lessons learned from quantitative dynamical modeling in systems biology. PLoS One. 2013 Sep;8(9):e74335.

[30] Stapor P, Weindl D, Ballnus B, et al. Pesto: parameter estimation toolbox. Bioinformatics. 2017 Oct;btx676–btx676. Available from: http://dx.doi.org/10.1093/bioinformatics/btx676.

[31] Byrd RH, Hribar ME, Nocedal J. An interior point algorithm for large-scale nonlinear programming. SIAM J Optim. 1999;9:877–900.

[32] Byrd RH, Gilbert JC, Nocedal J. A trust region method based on interior point techniques for nonlinear programming. Math Program. 2000 Nov;89(1):149–185. Available from: http://dx.doi.org/10.1007/PL00011391.

[33] Waltz RA, Morales JL, Nocedal J, et al. An interior algorithm for nonlinear optimization that combines line search and trust region steps. Math Program. 2006 Jul;107(3):391–408. Available from: http://dx.doi.org/10.1007/s10107-004-0560-5.

[34] Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE Trans Evol Comput. 1997 Apr;1(1):67–82.

[35] Boyd S, Vandenberghe L. Convex optimisation. UK: Cambridge University Press; 2004.

[36] Boiger R, Hasenauer J, Hross S, et al. Integration based profile likelihood calculation for PDE constrained parameter estimation problems. Inverse Prob. 2016 Dec;32(12):125009.