Advances in Computational Mathematics (2024) 50:4
https://doi.org/10.1007/s10444-023-10098-0

# Numerical investigation of agent-controlled pedestrian dynamics using a structure-preserving finite volume scheme

**Jan-Frederik Pietschmann**[1] · **Ailyn Stötzner**[2] · **Max Winkler**[2]

Received: 12 January 2023 / Accepted: 4 December 2023


**Abstract**
We provide a numerical realization of an optimal control problem for pedestrian motion with agents that was analyzed in Herzog et al. (*Appl. Math. Optim.* 88(3):87, 2023). The model consists of a regularized variant of Hughes' model for pedestrian dynamics coupled to ordinary differential equations that describe the motion of agents which are able to influence the crowd via attractive forces. We devise a finite volume scheme that preserves the box constraints that are inherent in the model and discuss some of its properties. We apply our scheme to an objective functional tailored to the case of an evacuation scenario. Finally, numerical simulations for several practically relevant geometries are performed.

**Keywords** Crowd motion · Nonlinear transport · Eikonal equation · ODE-PDE coupling · Optimal control · Finite volume · Projected gradient descent

**Mathematics Subject Classification (2010)** 49K20 · 35Q91 · 35M33 · 65M08

## 1 Introduction

With more and more people living in highly populated areas, the modelling, simulation and control of (large) pedestrian crowds is an important field of research. In this work,

Communicated by: Stefan Volkwein

✉ Max Winkler
max.winkler@mathematik.tu-chemnitz.de

Jan-Frederik Pietschmann
jan-f.pietschmann@uni-a.de

Ailyn Stötzner
ailyn@stoetzner.de

[1] Chair for Inverse Problems, Faculty of Mathematics, Natural Sciences, and Materials Engineering, University of Augsburg, 86159 Augsburg, Germany

[2] Faculty of Mathematics, Technische Universität Chemnitz, 09107 Chemnitz, Germany

Published online: 28 December 2023

we study the optimal control problem for a regularized version of Hughes' model for pedestrian motion [28]. In our approach, the (continuous) crowd can be controlled by a fixed small number of agents that can attract people in their vicinity. In terms of the model, this corresponds to an additional potential term centered at the agents positions. In a previous work [25], we already studied the well-posedness and optimality conditions of this problem while here, we focus on a numerical implementation of the control problem and extensive numerical examples. In particular, we provide and analyze a finite volume scheme that preserves the box constraints inherent in our problem.

To introduce the model, we fix $\Omega \subset \mathbb{R}^2$ to be a bounded domain with $C^4$-boundary $\partial\Omega$. Furthermore $T > 0$ is an arbitrary time horizon and $Q_T := (0,T) \times \Omega$ and $\Sigma_T = (0,T) \times \partial\Omega$ denote the space-time cylinder and its lateral boundary, respectively. The boundary is decomposed into two parts: $\partial\Omega_D$ representing the exits and $\partial\Omega_W$ the part where the domain is constrained by walls. For theoretical purposes (regularity of solutions) we assume $\overline{\partial\Omega_D} \cap \overline{\partial\Omega_W} = \emptyset$ meaning that both boundary parts are separated from each other, see Fig. 1. In a similar way we define $\Sigma_D = (0,T) \times \partial\Omega_D$ and $\Sigma_W = (0,T) \times \partial\Omega_W$.

The unknown variables in our system of equations are the density of the crowd $\rho\colon Q_T \to \mathbb{R}_+$, a potential specifying the current time to escape $\phi\colon Q_T \to \mathbb{R}$. In addition, there are $M$ agents which may influence the motion of the crowd via attractive forces. Their positions are denoted by $x_i\colon (0,T) \to \mathbb{R}^2$, $i = 1,\dots,M$. In addition, each agent is able to regulate the strength by which it acts on the crowd. This is encoded in the intensities $c_i\colon (0,T) \to \mathbb{R}_+$, $i = 1,\dots,M$. We remark that for an agent to control the strength of its action on the crowd is, in most scenarios, not very realistic. Thus one may argue that either a constant strength or the ability to turn attraction on and off for each agent are more plausible choices. However, our experiments show that the first option does not yield very satisfactory results. The second version, on the other hand, would lead to a mixed-integer programming problem and is much more challenging, both from the analysis and the computational point of view. Thus we think that a variable intensity as a first approximation of this more complicated scenario is reasonable and postpone the discrete case to future work. Both the agent trajectories and interaction strength are summarized in a vector $\boldsymbol{x} = (x_1,\dots,x_M)^\top$ and $\boldsymbol{c} = (c_1,\dots,c_M)^\top$, respectively.

The mathematical equations describing the movement of a pedestrian crowd influenced by agents then read as follows. For given agent movement directions $\boldsymbol{u} = (u_1,\dots,u_M)^\top$ with $u_i \in L^\infty(0,T;\mathbb{R}^2)$ the unknowns $\rho, \phi, \boldsymbol{x}$ are related to each other by means of

$$\partial_t \rho - \nabla \cdot (\rho\, \beta(\rho,\phi,\boldsymbol{x},\boldsymbol{c})) = \varepsilon \Delta\rho \qquad \text{in } Q_T, \tag{1.1a}$$

$$-\delta_1 \Delta\phi + |\nabla\phi|^2 = \frac{1}{f(\rho)^2 + \delta_2} \qquad \text{in } Q_T, \tag{1.1b}$$

$$\dot{x}_i(t) = f\left(\rho(t,x_i(t))\right) u_i(t) \quad \text{for } t \in (0,T), \quad i = 1,\dots,M. \tag{1.1c}$$

Moreover, we impose the boundary conditions

$$\begin{aligned} -\big(\varepsilon\,\nabla\rho + \rho\,\beta(\rho,\phi,\boldsymbol{x},\boldsymbol{c})\big)\cdot n &= \gamma\,\rho, & \phi &= 0 \quad \text{on } \Sigma_{\mathrm{D}},\\ \big(\varepsilon\,\nabla\rho + \rho\,\beta(\rho,\phi,\boldsymbol{x},\boldsymbol{c})\big)\cdot n &= 0, & \nabla\phi\cdot n &= 0 \quad \text{on } \Sigma_{\mathrm{W}}, \end{aligned} \tag{1.2}$$

as well as the initial conditions

$$\rho(0,\cdot) = \rho_0 \quad \text{in } \Omega, \quad x_i(0) = x_{i,0} \quad \text{for } i = 1,\dots,M. \tag{1.3}$$

Here, $\varepsilon, \delta_1, \delta_2 > 0$ are regularization parameters and the corresponding terms in the system are needed to guarantee a certain regularity for the solution, see Theorem 2.1.

The domain $\Omega$ is sufficiently large such that $x_i(t) \in \Omega$ on $[0,T]$ for $i = 1,\dots,M$ and $t \in [0,T]$ if $|u_i(t)| \le 1$.

Let us briefly discuss the meaning of the respective terms: Equation (1.1a) states that pedestrians are transported according to the velocity field $\beta$, see (2.1) below, while also performing (little) random motion encoded by the Laplacian of $\rho$. The second equation (1.1b) is a modified and regularized Eikonal equation whose solution is the distance to the closest exit, mitigating areas of high density via the term on the right-hand side. Here, the additional diffusion accounts for the fact that pedestrians do not know their environment exactly. Then, (1.1c) governs the motion of the agents, whose speed is also influenced by the surrounding pedestrian density. The function $f : [0,1] \to [0,1]$ is a density-velocity rule, chosen in such a way that $f(\rho)$ determines the maximum velocity an individual can move if the density in its current position is $\rho$. We choose $f$ to be monotonically decreasing meaning that higher densities lead to slower movements. The velocity field $\beta$ will reflect the fact that pedestrians are, on the one hand, trying the minimize their exit time which amounts to a drift term in the direction of $\nabla\phi$ and on the other hand, they are attracted by the agents which is realized by additional attractive potentials whose center depends on the agents' positions $\boldsymbol{x}$. This results in a velocity which is the sum of two terms. Furthermore, to account for the effect that the velocity will deteriorate in regions of high density, it will be modified by an additional multiplicative factor $f(\rho)$. As in the equations for the motion of the agents, $f$ is monotonically decreasing and becomes zero at a given maximal density.

The boundary conditions (1.2) allow for an outflow with velocity $\gamma$ on parts of the boundary ($\Sigma_D$) while no-flux conditions on the remaining parts are to be interpreted as walls ($\Sigma_W$). A detailed description of the involved non-linearities will be given in the next section, but we also refer to [25] for more details on the model and the regularizing terms.

Analytical properties of the unregularized Hughes' model introduced in [28] (i.e. $\varepsilon = \delta_1 = \delta_2 = 0$ in (1.1)), without control, are difficult because of the low regularity of $\nabla\phi$ that appears on a set depending on the solution $\rho$ of the first equation, but see [3, 4, 21]. Thus, regularized variants have been considered, see [19] for an instance where $\varepsilon = 0$ but $\delta_1,\ \delta_2 \neq 0$. In fact, the result there is obtained as a vanishing viscosity limit $\varepsilon \to 0$. There are also a number of extensions and variants of the model, aiming to understand additional properties, make it more realistic, or consider different settings like graphs, see [9, 13, 14, 16, 18].

Control of systems by means of a small number of agents has received lots of interest recently, both on a discrete level (i.e. one considers a large system of ODEs for the motion of individuals coupled to a small number of equations for the agents), see [11, 26], but also for coupled PDE-ODE systems [1, 2]. Let us emphasize that whenever PDEs are coupled to ODEs in such a fashion that the solution of the PDE needs to evaluated at the solution of the ODE (as in (1.1c)), regularity is needed. While in our case, this is obtained by the additional diffusion in (1.1a), when hyperbolic models for the transport of pedestrian are considered, an additional regularization in the ODE is needed, see, e.g., [6–8]. Finally, let us mention that the ODE-ODE and PDE-ODE perspectives are closely related by means of the so-called mean field limits when the number of agents tends to infinity, see [10, 31] and also the recent overview on control of crowds [5].

For the numerical discretization of (1.1a), we employ, as we think of the parameter $\varepsilon$ being small, a finite volume scheme for the spatial discretization, which may also be interpreted as a discontinuous Galerkin scheme. In combination with the Lax-Friedrichs numerical fluxes the scheme is stable and preserves the bounds $0 \leq \rho \leq 1$ inherent in our model. Such structure-preserving discretizations of PDEs gained much attention, e.g., in the context of chemotaxis problems [22–24, 29, 30, 33]. The previously mentioned articles differ also in the choice of the time-stepping scheme. For the treatment of (1.1) we will use an implicit-explicit finite difference scheme, whereas the diffusion-related terms are established implicitly and the convection-related terms explicitly. The equation (1.1b) is discretized with standard linear finite elements and for (1.1c) we employ a backward Euler scheme.

The paper is organized as follows: In Sect. 2 we give a precise definition of our problem and recall the analytical results from [25]. In Sect. 3 we provide a numerical discretization scheme in space and time and analyze some of its properties, in particular, we show that it preserves physical bounds of the density of pedestrians. Corresponding optimization algorithms are discussed in Sect. 3.5 and Sect. 4 finally provides the results of our numerical experiments.

## 2 The continuous optimal control problem

Let us motivate the remaining quantities arising in the system of equations. The function $f\colon [0, \rho_{\max}] \to \mathbb{R}$ is a density-velocity relation and is assumed to be $W^{3,\infty}(\mathbb{R}) \cap C_c(\mathbb{R})$ with $f(0) = 1$ and $f(\rho_{\max}) = 0$ with $\rho_{\max}$ denoting the maximal density. A usual choice is

$$f(\rho) = \xi\left(1 - \frac{\rho}{\rho_{\max}}\right)$$

with a cutoff function $\xi \in C_c^\infty(-1, 2)$ satisfying $\xi \equiv 1$ on $(0, 1)$. Obviously, a higher density leads to a lower velocity. Throughout the article we set $\rho_{\max} = 1$. The movement direction of the crowd described by the function $-\beta(\rho, \phi, \boldsymbol{x}, \boldsymbol{c})$ is modelled as follows. The primary interest of the crowd is to move either towards the closest

emergency exit, this is the direction $-\nabla\phi$. This is mitigated by the attraction of close by agents which is the direction $-\nabla\phi_K$, where is an agent potential defined by

$$\phi_K(\boldsymbol{x}, \boldsymbol{c}; t, x) := \sum_{i=1}^{M} c_i(t)\, K\big(x - x_i(t)\big).$$

Here, $K(x) = k(|x|)$, $k \in W^{3,\infty}(\mathbb{R})$ is a radially symmetric function and $c_i \in H^1(0,T)$, $i = 1, \dots, M$, are time-dependent intensity functions. Typical choices for attractive agent potentials are either the bump function

$$k(r) = \begin{cases} \exp\big(-\frac{R^2}{R^2 - r^2}\big), & \text{if } r < R, \\ 0, & \text{otherwise} \end{cases}$$

with attraction radius $R > 0$, or the Morse potential

$$k(r) = \mathrm{e}^{-2\,a\,(r - r_a)} - 2\,\mathrm{e}^{-a\,(r - r_a)}$$

with certain parameters $a, r_a > 0$, realizing a repulsion in the near and an attraction in the far range of the agents. This is useful to avoid a high density very close to the respective agent. We refer to [12] for a more detailed discussion on potentials in the context of flocking problems. In the experiments of this article the Morse potential was used as the resulting pedestrian behavior seems more realistic.

These considerations yield a velocity field defined as follows

$$\beta(\rho, \phi, \boldsymbol{x}, \boldsymbol{c}) = v_0 f(\rho)\, h(\nabla(\phi + \phi_K(\boldsymbol{x}, \boldsymbol{c}))) \quad \text{with} \quad h(x) = \widetilde{\min}\{1, |x|\}\frac{x}{|x|}, \quad (2.1)$$

where $\widetilde{\min}$ is a continuously differentiable approximation of the minimum function resulting in $h$ being a smoothed projection into the unit ball in $\mathbb{R}^2$, and the factor $f(\rho)$ again links the allowed movement speed to the current density, this is, $|\beta(\rho, \phi, \boldsymbol{x}, \boldsymbol{c})| \le v_0 f(\rho)$ in $Q_T$.

We briefly introduce the function spaces used in the sequel, see also [17]. For a domain $\Omega \subset \mathbb{R}^2$ we denote by $W^{k,p}(\Omega)$, $k \in \mathbb{N}_0$, $p \in [1, \infty]$, the usual Sobolev spaces and by $W^{k-1/p,p}(\Gamma)$ for $k \ge 1$ the corresponding trace spaces which may be equipped with the Sobolev-Slobodetskij norm. Furthermore, we write $H^k(\Omega) = W^{k,2}(\Omega)$. Special spaces incorporating already boundary conditions are $H^1_D(\Omega) = \{v \in H^1(\Omega) \colon v|_{\partial\Omega_{\mathrm{D}}} \equiv 0\}$ and $W^{2,p}_{\mathrm{DN}}(\Omega) := \{v \in W^{2,p}(\Omega) \colon v|_{\partial\Omega_D} = 0,\ \partial_n v|_{\partial\Omega_W} = 0\}$. For time-dependent functions $v \colon [0,T] \to X$ for some Banach space $X$ we define

$$L^p(0,T;X) := \{v \colon (0,T) \to X \mid \int_0^T \|v(t)\|_X^p \,\mathrm{d}t < \infty\}, \quad p \in [1, \infty),$$

as well as

$$W^{s,p}(0,T;X) := \{v \colon (0,T) \to X \mid \partial_t^\ell v \in L^p(0,T;X), 0 \le \ell \le s\}, \quad s \in \mathbb{N}_0,\ p \in [1,\infty).$$

For the application we have in mind the following spaces

$$W_p^{r,s}(Q_T) := L^p(0,T;W^{r,p}(\Omega)) \cap W^{s,p}(0,T;L^p(\Omega)), \quad p \in [1,\infty), \quad r,s \in \mathbb{N}_0,$$

are of interest which are equipped with the natural norms

$$\|v\|_{W_p^{r,s}(Q_T)} := \left( \|v\|^p_{L^p(0,T;W^{r,p}(\Omega))} + \|v\|^p_{W^{s,p}(0,T;L^p(\Omega))} \right)^{1/p}.$$

Spaces with non-integer $r$ and $s$ are defined as interpolation spaces.

In a previous work [25], a global (in time) well-posedness and regularity result for (1.1)–(1.3) was established. Furthermore, optimality conditions for related optimal control problems where this system occurs as a constraint were derived. For convenience of the reader we briefly summarize the most important results needed in the present article.

First, there holds the following existence and regularity result:

**Theorem 2.1** *Assume that $\rho_0 \in W^{3/2,4}(\Omega)$ and fix $T > 0$. Given arbitrary agent movement directions $\boldsymbol{u} = (u_1,\dots,u_M)^{\mathrm{T}} \in L^\infty(0,T;\mathbb{R}^2)^M$ and intensities $\boldsymbol{c} = (c_1,\dots,c_M) \in H^1(0,T)^M$, there exists a unique strong solution $(\rho,\phi,\boldsymbol{x})$ to (1.1)–(1.3) which satisfies, for any $2 < p < \infty$, $\rho \in W_p^{2,1}(Q_T)$ and $\phi \in L^\infty(0,T;W^{2,p}(\Omega))$. The agent trajectories $x_i$, $i = 1,\dots,M$ are absolutely continuous. Moreover, the a priori estimate*

$$\|\rho\|_{W_p^{2,1}(Q_T)} + \|\phi\|_{L^\infty(0,T;W^{2,p}(\Omega))} \le C\|\rho_0\|_{W^{1,p}(\Omega)},$$

*holds with $C > 0$ depending only on the domain, the bounds for the coefficients and the respective kernel.*

The previous result allows to define an operator, the so-called control–to–state operator,

$$S\colon \mathcal{Q} \to \mathcal{Y}, \qquad \boldsymbol{q} := (\boldsymbol{u},\boldsymbol{c}) \mapsto S(\boldsymbol{q}) = \boldsymbol{y} := (\rho,\phi,\boldsymbol{x})$$

with control and state spaces

$$\mathcal{Q} := \mathcal{U} \times \mathcal{C} := L^\infty(0,T;\mathbb{R}^2)^M \times H^1(0,T)^M,$$
$$\mathcal{Y} := W_p^{2,1}(Q_T) \times \left( L^\infty(0,T;W^{2,p}_{\mathrm{DN}}(\Omega)) \cap W^{1,p}(0,T;W^{1,p}(\Omega)) \right) \times W^{1,s}(0,T;\mathbb{R}^2)^M$$

for $s = \left(\frac{1}{2} - \frac{1}{p}\right)^{-1}$. Furthermore we define the set of admissible controls

$$\mathcal{Q}_{\mathrm{ad}} := \{(\boldsymbol{u},\boldsymbol{c}) \in \mathcal{Q}\colon |u_i(t)| \le 1,\ 0 \le c_i(t) \le 1 \text{ f.a.a. } t \in (0,T) \text{ and all } i = 1,\dots,M\}.$$

The optimal control problem we study in this article reads

$$\text{Minimize} \quad \mathcal{J}(\boldsymbol{y};\boldsymbol{q}) := \int_{\widetilde{Q}_T} e^{\nu t}\,\rho(t,x)\mathrm{d}x\,\mathrm{d}t - \mu \sum_{i=1}^{M}\int_0^T \ln(B(x_i(t)))\,\mathrm{d}t + \frac{\alpha_1}{2\,T}\sum_{i=1}^{M}\|u_i\|^2_{H^1(0,T;\mathbb{R}^2)} + \frac{\alpha_2}{2\,T}\sum_{i=1}^{M}\|c_i\|^2_{H^1(0,T)} \tag{2.2a}$$

$$\text{subject to} \quad \boldsymbol{y} := (\rho,\phi,\boldsymbol{x}) = S(\boldsymbol{q}), \tag{2.2b}$$

$$\boldsymbol{q} := (\boldsymbol{u},\boldsymbol{c}) \in \mathcal{Q}_{\text{ad}}. \tag{2.2c}$$

The objective functional $\mathcal{J}$ aims at a fast evacuation of the crowd. By the factor $e^{\nu t}$ higher densities at a later time are penalized more. The parameter $\nu > 0$ adjusts the urgency of the evacuation. We observe the density in $\widetilde{Q}_T = I \times \widetilde{\Omega}$ where $\widetilde{\Omega} \subset \Omega$ is a subregion which the pedestrians must leave. We use the temporal $H^1$-norm of the agent movement directions and the intensities as a regularization to guarantee the smoothness required in Theorem 2.1. The regularization parameters $\alpha_1, \alpha_2 > 0$ are arbitrary but positive.

The fourth term in the objective is a barrier used to avoid that the agents walk through walls. The barrier function $B \in H^1_D(\Omega)$ is the weak solution of the singularly perturbed problem

$$-\delta_4 \Delta B + B = 1 \quad \text{in } \Omega, \tag{2.3a}$$

$$B = 0 \quad \text{on } \partial\Omega. \tag{2.3b}$$

The barrier function $-\ln(B(x_i(t)))$ tends to infinity if $\operatorname{dist}(x_i(t),\partial\Omega) \to 0$ for some $t \in [0,T]$. For $x_i(t) \in \operatorname{int}\Omega$ there holds $\lim_{\mu\to 0}\mu\,\ln(B(x_i(t))) = 0$. Here we choose $\mu > 0$ to be fixed but small.

The control constraint $(\boldsymbol{u},\boldsymbol{c}) \in \mathcal{Q}_{\text{ad}}$ guarantees that the agents do not move faster than the density in their current position allows and that the intensity is bounded by reasonable values.

We have the following well-posedness result and necessary optimality condition.

**Theorem 2.2** *There exists at least one global solution* $(\boldsymbol{y},\boldsymbol{q}) \in \mathcal{Y}\times\mathcal{Q}_{\text{ad}}$ *of* (2.2).

*Furthermore, each local minimizer* $(\boldsymbol{y},\boldsymbol{q}) \in \mathcal{Y}\times\mathcal{Q}_{\text{ad}}$, $\boldsymbol{y} = (\rho,\phi,\boldsymbol{x})$, $\boldsymbol{q} = (\boldsymbol{u},\boldsymbol{c})$, *of* (2.2) *fulfills for all directions in the tangential cone at* $(\boldsymbol{u},\boldsymbol{c})$, *namely* $\delta\boldsymbol{q} = (\delta\boldsymbol{u},\delta\boldsymbol{c}) \in \mathcal{T}_{\mathcal{Q}_{\text{ad}}}(\boldsymbol{u},\boldsymbol{c})$,

$$\int_{\widetilde{Q}_T} e^{\nu t}\,\delta\rho(t,x)\mathrm{d}x\,\mathrm{d}t + \frac{\alpha_1}{T}\,\boldsymbol{u},\delta\boldsymbol{u}_{H^1(0,T;\mathbb{R}^2)^M} + \frac{\alpha_2}{T}\,(\boldsymbol{c},\delta\boldsymbol{c})_{H^1(0,T)^M} - \mu\sum_{i=1}^{M}\int_0^T \frac{\nabla B(x_i(t))^\top\,\delta x_i(t)}{B(x_i(t))}\,\mathrm{d}t \geq 0,$$

*with* $\boldsymbol{y} = S(\boldsymbol{u}, \boldsymbol{c})$ *and* $\delta \boldsymbol{y} = (\delta\rho, \delta\phi, \delta\boldsymbol{x}) = S'(\boldsymbol{u}, \boldsymbol{c})\,(\delta\boldsymbol{u}, \delta\boldsymbol{c})$ *characterized by the system*

$$\partial_t \delta\rho - \varepsilon \Delta\delta\rho - \nabla \cdot \left( \delta\rho\, \beta(\rho, \phi, \boldsymbol{x}, \boldsymbol{c}) + \rho \left( \frac{\partial \beta(\rho, \phi, \boldsymbol{x}, \boldsymbol{c})}{\partial(\rho, \phi, \boldsymbol{x}, \boldsymbol{c})} (\delta\rho, \delta\phi, \delta\boldsymbol{x}, \delta\boldsymbol{c}) \right) \right) = 0, \tag{2.4a}$$

$$-\delta_1\, \Delta\delta\phi + 2\nabla\phi \cdot \nabla\delta\phi + \frac{2 f(\rho)\, f'(\rho)}{(f^2(\rho) + \delta_2)^2} \delta\rho = 0, \tag{2.4b}$$

$$\dot{\delta x}_i - v_0\, f'(\rho(\cdot, x_i)) \big( \nabla\rho(\cdot, x_i)^{\mathrm{T}} \delta x_i + \delta\rho(\cdot, x_i) \big) u_i = v_0\, f(\rho(\cdot, x_i))\, \delta u_i, \tag{2.4c}$$

*for* $i = 1, \ldots, M$, *together with the boundary conditions* (1.2) *and homogeneous initial conditions*

$$\delta\rho(0, \cdot) = 0 \quad \textit{and} \quad \delta x_i(0) = 0, \quad i = 1, \ldots, M. \tag{2.5}$$

The proof of the theorem above is very close to those of Theorem 3.8 and Theorem 4.4 [25]. The main difference is that the model in [25] only allows to control the velocity $\boldsymbol{u}$ of the agents, yet not their strength $\boldsymbol{c}$. As for the existence proof, this does not impose any additional difficulty due to the uniform $L^\infty$-boundedness of $\boldsymbol{c}$ as a consequence of the embedding $H^1 \hookrightarrow L^\infty$ in one spatial dimension. For the differentiability result, one has to add the derivatives with respect to $\boldsymbol{c}$, yielding an additional term in (2.4a) that, however, can be estimated similarly to the remaining terms.

## 3 Discretization of the state equation

In this section we introduce the numerical scheme used to compute approximate solutions of the forward system (1.1)–(1.3). To this end, we introduce a semi-implicit time-stepping scheme and use a finite volume discretization for the density function $\rho$ and continuous Lagrange finite elements for the potential function.

### 3.1 Space discretization

For the spatial discretization of the system (1.1)–(1.3) we define a family of geometrically conforming triangular meshes $\{\mathcal{T}_h\}_{h>0}$. For each $T \in \mathcal{T}_h$ we denote by $h_T = \mathrm{diam}(T)$ the element diameter and by $\rho_T$ the diameter of the largest inscribed ball in $T$. The mesh parameter is then $h = \max_{T \in \mathcal{T}_h} h_T$. The mesh family is assumed to be shape regular meaning that there is a constant $\kappa > 0$ such that $h_T/\rho_T \leq \kappa$ for all $T \in \mathcal{T}_h$ and all $h > 0$. By $\mathcal{F}_h^{\mathrm{i}}$ we denote the set of interior element edges, by $\mathcal{F}_h^{\mathrm{bd}}$ the boundary edges and write $\mathcal{F}_h := \mathcal{F}_h^{\mathrm{i}} \cup \mathcal{F}_h^{\mathrm{bd}}$. Furthermore, to each edge $F \in \mathcal{F}_h$ we associate the normal vector $n_F$ which is pointing outward in case of a boundary edge and has a fixed orientation in case of an interior edge.

We propose a finite volume scheme for the transport equation. As we use the finite element package FEniCSfor our implementation, we use a notation which is rather

usual for discontinuous Galerkin discretizations, see [20] for an overview. The finite-dimensional function spaces are defined by

$$V_h = \{v \in L^\infty(\Omega)\colon v|_T \in \mathcal{P}^0(T) \text{ for all } T \in \mathcal{T}_h, \},$$
$$W_h = \{v \in C(\overline{\Omega})\colon \ v|_T \in \mathcal{P}^1(T) \text{ for all } T \in \mathcal{T}_h\}, \quad W_{h,\mathrm{D}} := W_h \cap H^1_D(\Omega),$$

where $\mathcal{P}^k(T)$ denotes the space of polynomials on $T$ of degree not larger than $k \in \mathbb{N}_0$. For a function $v\colon \Omega \to \mathbb{R}$, we define interface averages and jumps in the following way

$$\{v\}_F := \frac{1}{2}(v|_{T_1} + v|_{T_2}), \qquad [\![v]\!]_F := v|_{T_1} - v|_{T_2}\,, \qquad \forall F \in \mathcal{F}_h^{\mathrm{i}},$$

where $T_1, T_2 \in \mathcal{T}_h$ are chosen in such a way that $n_F = n_{\partial T_1|F} = -n_{\partial T_2|F}$.

In order to discretize the system (1.1) we use discontinuous approximations for $\rho$ and continuous ones for $\phi$. The unknowns in our semi-discrete scheme are

$$\rho_h(t) \in V_h, \quad \phi_h(t) \in W_{h,\mathrm{D}}, \quad x_1(t), \ldots, x_M(t) \in \mathbb{R}^2, \quad c_1(t), \ldots, c_M(t) \in \mathbb{R}$$

for all $t \in [0, T]$. The approximate transport direction is then given by

$$\beta_h(\rho_h, \phi_h, \boldsymbol{x}, \boldsymbol{c}) := f(\rho_h)\, h(\nabla\phi_h + \phi_K(\boldsymbol{x}, \boldsymbol{c}))$$

with

$$\phi_K(\boldsymbol{x}, \boldsymbol{c}; t, x) := \sum_{j=1}^{M} c_j(t)\, K(x - x_i(t)).$$

The semi-discretization of (1.1a) then reads Find $\rho_h\colon [0, T] \to V_h$ with $\rho_h(0) = \mathrm{proj}_{V_h}(\rho_0)$ and

$$(\partial_t \rho_h(t), v_h)_\Omega + a(\rho_h(t), v_h) + b(\beta_h)(\rho_h(t), v_h) = 0 \qquad \forall v_h \in V_h, t \in (0, T). \quad (3.1)$$

Here, $\mathrm{proj}_{V_h}\colon L^2(\Omega) \to V_h$ is some projection operator, $(\cdot,\cdot)_\Omega$ is the standard $L^2(\Omega)$-inner product and the bilinear forms are defined by

$$a(\rho_h, v_h) = \varepsilon \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} \int_F \tau_F\, [\![\rho_h]\!][\![v_h]\!]\mathrm{d}s + \sum_{F \in \mathcal{F}_h^{\mathrm{bd}}} \chi_{\partial\Omega_D}\gamma \int_F \rho_h\, v_h \mathrm{d}s \quad (3.2a)$$

$$b(\beta_h)(\rho_h, v_h) = -\sum_{F \in \mathcal{F}_h^{\mathrm{i}}} \int_F (\rho_h\, \beta_h)^*_F\, [\![v_h]\!]\, \mathrm{d}s. \quad (3.2b)$$

The parameter $\tau_F$ is defined by

$$\tau_F := \frac{1}{|x_{T_1} - x_{T_2}|},$$

where $x_T$ is the intersection of the orthogonal edge bisectors of $T \in \mathcal{T}_h$. The term $\tau_F\, [\![\rho_h]\!][\![v_h]\!]$ with $v_h = \chi_T$ for some $T \in \mathcal{T}_h$ approximates the diffusive flux $\nabla\rho_h \cdot n_{\partial T}$

over the edge $F \subset T$. The bilinear form $b$ establishes the convective flux $\beta\,\rho\,\cdot n_{\partial T}$. As numerical flux function $(\cdot)^*$, we choose the Lax-Friedrichs flux, see [32], defined by

$$(\rho_h\,\beta)_F^* = \{\rho_h\,\beta\}_F \cdot n_F - \frac{\eta}{2}\,[\![\rho_h]\!]_F. \tag{3.3}$$

The stabilization parameter $\eta \in \mathbb{R}$ is specified later. For the closely related chemotaxis model such an approach has been sucessfully applied in [24, 30]. Of course, also other flux functions are possible, e.g., the central upwind flux [22].

The Eikonal equation (1.1b) is discretized in space using standard linear Lagrange elements which yields

$$\delta_1\left(\nabla\phi_h(t), \nabla w_h\right)_\Omega + \left(|\nabla\phi_h(t)|^2, w_h\right)_\Omega = \left(\frac{1}{f(\rho_h(t))^2 + \delta_2}, w_h\right)_\Omega \qquad \forall w_h \in W_{h,\mathrm{D}}. \tag{3.4}$$

In our numerical experiments we used the Newton solver integrated in FEniCS. The Jacobian is established by automatic differentiation.

The ordinary differential equations for the agent trajectories (1.1c) depend on a point evaluation $\rho(t, x_i(t))$ of a function which is discontinuous in the discrete setting. In particular, this term would not be differentiable with respect to $x_i(t)$. As a remedy, we use instead of a point evaluation, see also [6], the following regularization

$$\rho_h(t, x_i(t)) \approx \eta_{x_i(t)} * \rho_h(t), \quad \text{with} \quad \eta_{x_i(t)} := \frac{\delta_{x_i(t)}}{\delta_{x_i(t)} * 1},$$

where $*$ stands for the convolution integral $\delta_{x_0} * v = \int_\Omega \delta_{x_0}\, v\,\mathrm{d}x$ of the functions $v \in L^1(\Omega)$ and some kernel function $\delta_{x_0} \in C^\infty(\mathbb{R}^2)$. An obvious choice is the regularized Dirac delta function

$$\delta_{x_0}(x) := \frac{1}{2\,\pi\,\zeta}\,\mathrm{e}^{-\frac{\|x-x_0\|^2}{2\,\zeta}}$$

with small locality parameter $\zeta > 0$. Note that for $\zeta \to 0$ there holds $\delta_{x_0} * v \to v(x_0)$ for any $v \in C(\overline{\Omega})$. Furthermore, the regularized Dirac delta fulfills $\int_{\mathbb{R}^2} \delta_{x_0}\mathrm{d}x = 1$ for arbitrary $x_0 \in \mathbb{R}^2$. The discretized ordinary differential equation then reads

$$\dot{x}_i(t) = v_0\, f\left(\eta_{x_i(t)} * \rho_h(t)\right) u_i(t), \quad n = 1, \ldots, N \tag{3.5}$$

and initial conditions $x_i(0) = x_{i,0}$.

### 3.2 Time discretization

For the temproal discretization we cover the time interval $[0, T]$ by an equidistant grid $I_\tau := \{t_n\}_{n=0}^N$ with grid points $t_n := n\,\tau$ and grid size $\tau := T/N$. The spatial

and temporal discretization parameters are summarized in $\sigma = (h, \tau)$. Moreover, we define the space of grid functions

$$H_\tau(V) = \{v\colon I_\tau \to V\},$$

with $V$ an arbitrary linear space. If $V$ is again a function space containing functions $v\colon \Omega \to \mathbb{R}$ we write $v(t_n) = v(t_n, \cdot)$. The functions $\rho_h$, $\phi_h$, $x_i$, $u_i$ and $c_i$ arising in the semi-discrete equations (3.1), (3.4) and (3.5) are approximated by grid functions

$$\rho_\sigma \in H_\tau(V_h), \quad \phi_\sigma \in H_\tau(W_{h,\mathrm{D}}), \quad u_{i,\sigma}, x_{i,\sigma} \in H_\tau(\mathbb{R}^2), \quad c_{i,\sigma} \in H_\tau(\mathbb{R})$$

for $i = 1, \ldots, M$. For brevity we write for $n = 0, \ldots, N$

$$\rho_h^n \coloneqq \rho_\sigma(t_n, \cdot), \quad \phi_h^n \coloneqq \phi_\sigma(t_n, \cdot), \quad x_i^n \coloneqq x_{i,\sigma}(t_n), \quad u_i^n \coloneqq u_{i,\sigma}(t_n)$$

and for the transport vector we use

$$\beta_h^n = \beta_h(\rho_h^n, \phi_h^n, \boldsymbol{x}^n, \boldsymbol{c}^n).$$

We replace the temporal derivative by a difference quotient and use a semi-implicit time-stepping scheme, more precisely, the diffusion-related terms are evaluated implicitly and the convection-related terms explicitly. This yields the fully discrete system

$$(\rho_h^{n+1}, v_h)_\Omega + \tau\, a(\rho_h^{n+1}, v_h) = (\rho_h^n, v_h)_\Omega - \tau\, b(\beta_h^n)(\rho_h^n, v_h), \tag{3.6a}$$

$$\delta_1(\nabla\phi_h^n, \nabla w_h)_\Omega + (|\nabla\phi_h^n|^2, w_h)_\Omega = \left(\frac{1}{f(\rho_h^n)^2 + \delta_2}, w_h\right)_\Omega, \tag{3.6b}$$

$$x_i^{n+1} - x_i^n = \tau\, v_0\, f\left(\eta_{x_i^{n+1}} * \rho_h^{n+1}\right) u_i^{n+1}, \tag{3.6c}$$

for all test functions $v_h \in V_h$, $w_h \in W_{h,\mathrm{D}}$ and indices $i = 1, \ldots, M, n = 0, \ldots, N-1$. Furthermore, the initial conditions are established by means of:

$$\rho_h^0 = \mathrm{proj}_{Vh}(\rho_0), \qquad x_i^0 = x_{i,0},\ i = 1, \ldots, M.$$

Note that the system of equations (3.6) completely decouples and we can compute each variable after the other, in the following order

$$\rho_h^0, \boldsymbol{x}^0 \mapsto \phi_h^0 \mapsto \rho_h^1 \mapsto \boldsymbol{x}^1 \mapsto \phi_h^1 \mapsto \ldots \mapsto \rho_h^{N-1} \mapsto \boldsymbol{x}^{N-1} \mapsto \phi_h^{N-1} \mapsto \rho_h^N \mapsto \boldsymbol{x}^N. \tag{3.7}$$

### 3.3 Quality of discrete solutions

In this section we study some basic properties for the solutions of (3.6). In particular, it is of interest whether the physical bounds observed for the solution of the continuous system (1.1)–(1.3) are transferred to the discrete setting.

The basis functions of the finite element space $V_h$ are denoted by $\{\chi_T\}_{T\in\mathcal{T}_h}$ defined by $\chi_T|_{T'} \equiv \delta_{T,T'}$ for all $T, T' \in \mathcal{T}_h$. Note that by a slight abuse of notation we use the elements of $\mathcal{T}_h$ as indices here. Introducing the matrices $M = (m_{T,T'})_{T,T'\in\mathcal{T}_h}$, $A = (a_{T,T'})_{T,T'\in\mathcal{T}_h}$ and $B^n = (b^n_{T,T'})_{T,T'\in\mathcal{T}_h}$ with entries

$$m_{T,T'} = \begin{cases} |T|, & \text{if } T = T', \\ 0, & \text{otherwise,} \end{cases} \tag{3.8a}$$

$$a_{T,T'} = \begin{cases} \varepsilon \displaystyle\sum_{F\in\mathcal{F}_T\cap\mathcal{F}_h^{\mathrm{i}}} \tau_F\,|F| + \gamma \sum_{F\in\mathcal{F}_T\cap\mathcal{F}_h^{\mathrm{bd}}} |F|, & \text{if } T = T', \\ -\varepsilon\,\tau_F\,|F|, & \text{if } T \neq T' \text{ and } F := \partial T \cap \partial T' \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases} \tag{3.8b}$$

$$b^n_{T,T'} = \begin{cases} -\dfrac{1}{2} \displaystyle\sum_{F\in\mathcal{F}_T\cap\mathcal{F}_h^{\mathrm{i}}} \left( \int_F \beta_h^n|_T \cdot n_{\partial T}\,\mathrm{d}s - \eta\,|F| \right), & \text{if } T = T', \\ -\dfrac{1}{2}\left( \displaystyle\int_F \beta_h^n|_{T'} \cdot n_{\partial T}\,\mathrm{d}s + \eta\,|F| \right), & \text{if } T \neq T' \text{ and } F = \partial T \cap \partial T' \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases} \tag{3.8c}$$

allows to rewrite the system of equations (3.6a) as

$$(M + \tau\,A)\,\vec{\rho}^{\,n+1} = (M - \tau\,B^n)\,\vec{\rho}^{\,n}. \tag{3.9}$$

Here, $\vec{\rho}^{\,n}$, $n = 0, \ldots, N$, are the vector representations of $\rho_h^n$ with respect to the basis $\{\chi_T\}_{T\in\mathcal{T}_h}$. Note that the matrix $B^n$ depends also on $\vec{\rho}^{\,n}$.

**Theorem 3.1** *The numerical scheme* (3.6a) *is mass conserving in the following sense. Assuming that* $\gamma = 0$ *holds, i. e., there are no-flux boundary conditions present at all boundary parts* $\partial\Omega_D$ *and* $\partial\Omega_W$*, the solution* $\rho_\sigma$ *fulfills*

$$\int_\Omega \rho_h^n \mathrm{d}x = \int_\Omega \mathrm{proj}_{V_h}(\rho_0)\mathrm{d}x \quad \forall n = 0, 1, \ldots, N.$$

***Proof*** The assertion is trivially fulfilled for $n = 0$ as the initial condition is established by $\rho_h^0 = \mathrm{proj}_{V_h}(\rho_0)$. In matrix-vector notation the assertion is equivalent to $\vec{1}^\top M\vec{\rho}^{\,n+1} = \vec{1}^\top M\vec{\rho}^{\,n}$. This follows from (3.9) after using

$$\vec{1}^\top A\vec{\rho}^{\,n+1} = \sum_{T\in\mathcal{T}_h}\sum_{T'\in\mathcal{T}_h} a_{T,T'}\rho_{T'}^{n+1} = 0$$

and

$$\vec{1}^\top B^n\vec{\rho}^{\,n} = \sum_{T\in\mathcal{T}_h}\sum_{T'\in\mathcal{T}_h} b_{T,T'}\rho_{T'}^{n}.$$

In this expression, the stabilization terms (the ones multiplied with $\eta$) cancel out. Furthermore, after sorting terms in the sum by the edges $F \in \mathcal{F}_h^{\mathrm{i}}$ and denoting by $T_{F,1}, T_{F,2}$ the two triangles meeting in $F$, we obtain the terms

$$\vec{1}^{\top} B^n \vec{\rho}^n = -\frac{1}{2} \sum_{F \in \mathcal{F}_h^{\mathrm{i}}} \int_F \Big( \beta_h|_{T_{F,1}} n_{\partial T_{F,1}} \rho_{T_{F,1}} + \beta_h|_{T_{F,2}} n_{\partial T_{F,1}} \rho_{T_{F,2}} + \beta_h|_{T_{F,2}} n_{\partial T_{F,2}} \rho_{T_{F,2}} + \beta_h|_{T_{F,1}} n_{\partial T_{F,2}} \rho_{T_{F,1}} \Big) \mathrm{d}s = 0.$$

The last step follows due to $n_{\partial T_{F,2}} = -n_{\partial T_{F,2}}$ which implies $\vec{1}^{\top} B^n \vec{\rho}^n = 0$. □

**Theorem 3.2** *Choose $\eta = 1$ in (3.3) and denote by $\kappa > 0$ the maximal aspect ratio of the mesh family $\mathcal{T}_h$, see* Sect. 3.1. *Let $\tau$ be chosen to satisfy the CFL condition*

$$\tau \leq \frac{\pi}{3\,\kappa^2} \min_{T \in \mathcal{T}_h} h_T. \tag{3.10}$$

*If furthermore, there holds $\mathrm{proj}_{V_h}(\rho_0) \in [0,1]$ a. e. in $\Omega$ and $\beta_h^n = (1 - \rho_h^n)\, \Phi^n$ with $|\Phi^n| \leq 1$, $n = 0, \ldots, N$, the solutions of (3.6) fulfill for all $n = 0, \ldots, N$*

$$\rho_h^n(x) \in [0,1] \quad \textit{f.a.a. } x \in \Omega.$$

***Proof*** The diagonal entries of $(M + \tau\, A)$ are all positive and the off-diagonal entries are non-positive. Moreover, one easily concludes the strict diagonal dominance, this is,

$$\sum_{\substack{T' \in \mathcal{T}_h \\ T' \neq T}} |m_{T,T'} + \tau\, a_{T,T'}| < m_{T,T} + \tau\, a_{T,T}.$$

This implies that $(M + \tau\, A)$ is an M-matrix and consequently, the inverse $(M + \tau\, A)^{-1}$ exists and fulfills $(M + \tau\, A)^{-1} \geq 0$.

Let $n \in \mathbb{N}_0$ be fixed and assume that $\rho_h^n(x) \in [0,1]$ for almost all $x \in \Omega$. We show $\rho_h^{n+1} \geq 0$ by confirming that the right-hand side of (3.9) has non-negative entries only. Assuming that $\rho_T^n \geq 0$, $T \in \mathcal{T}_h$, we show that the entries of $(M - \tau\, B^n)$ are non-negative as well. The entries on the diagonal have the form

$$|T| + \frac{\tau}{2} \sum_{F \in \mathcal{F}_T \cap \mathcal{F}_h^{\mathrm{i}}} \int_F (\beta_h^n|_T \cdot n_{\partial T} - 1)\, \mathrm{d}s \geq |T| - \tau\, |\partial T|,$$

where the first step follows from the assumption (2.1) implying $|\beta_h^n| \leq 1$. To estimate further we take the geometric mesh quantities and the shape regularity $h_T \leq \kappa\, \rho_T$ into account and arrive at

$$\frac{|T|}{|\partial T|} \geq \frac{\pi\, \rho_T^2}{3\, h_T} \geq \frac{\pi\, h_T}{3\, \kappa^2} \geq \tau,$$

where the last step is the CFL condition (3.10). The previous two estimates confirm $(m_{T,T} - \tau\, b^n_{T,T}) \geq 0$ for all $T \in \mathcal{T}_h$. The remaining entries in the matrix, namely $(m_{T,T'} - \tau\, b^n_{T,T'})$ with $F = \partial T \cap \partial T' \neq \emptyset$, have the form

$$\frac{\tau}{2} \int_F (\beta_h^n|_{T'} \cdot n_{\partial T} + 1)\, \mathrm{d}s \geq 0.$$

The non-negativity follows again from $|\beta_h^n| \leq 1$. Combining the previous arguments provides the lower bound $\rho_h^{n+1} \geq 0$.

To show the upper bound we rearrange the equation system (3.9) in the form

$$(M + \tau\, A)\,(\vec{1} - \vec{\rho}^{\,n+1}) = M\,(\vec{1} - \vec{\rho}^{\,n}) + \tau\, B^n\, \vec{\rho}^{\,n} + \tau\, A\, \vec{1}. \tag{3.11}$$

We may rewrite the transport term using $\rho_h^n\, \beta_h^n = (1 - \rho_h^n)\, \widetilde{\beta}_h^n$ with $\widetilde{\beta}_h^n = \rho_h^n\, \Phi$. In (3.11) we reformulate the expression involving $B^n$ by means of

$$\begin{aligned}
[B^n\, \vec{\rho}^{\,n}]_T &= -\sum_{F \in \mathcal{F}_T \cap \mathcal{F}_h^{\mathrm{i}}} \int_F \left( \{(1 - \rho_T^n)\, \widetilde{\beta}_h^n\}_F \cdot n_F + \frac{1}{2}\, [\![1 - \rho_h^n]\!] \right)_F [\![\chi_T]\!]_F\, \mathrm{d}s \\
&= -\frac{1}{2} \sum_{F \in \mathcal{F}_T \cap \mathcal{F}_h^{\mathrm{i}}} \int_F (\widetilde{\beta}_h^n|_T \cdot n_{\partial T} + 1)\, \mathrm{d}s \cdot (1 - \rho_T^n) \\
&\quad - \frac{1}{2} \sum_{F \in \mathcal{F}_T \cap \mathcal{F}_h^{\mathrm{i}}} \int_F (\widetilde{\beta}_h^n|_{T_F} \cdot n_{\partial T} - 1)\, \mathrm{d}s \cdot (1 - \rho_{T_F}^n) \\
&=: [\widetilde{B}^n (1 - \vec{\rho}^{\,n})]_T.
\end{aligned}$$

with $T_F \in \mathcal{T}_h \backslash \{T\}$, $T_F \cap T = F$. With the same arguments like above one can show that the entries of $M + \tau\, \widetilde{B}^n$ are non-negative and together with $1 - \vec{\rho}^{\,n} \geq 0$, $A\, \vec{1} \geq 0$ and the M-matrix property of $M + \tau\, A$ we arrive at the desired bound $1 - \vec{\rho}^{\,n+1} \geq 0$.□

### 3.4 The discrete optimal control problem

Next, we study a discrete version of the optimal control problem (2.2). The control and state variables are grid functions in time and thus, we introduce the discrete $H^1(0, T)$-inner product for functions $u_\tau, v_\tau : \{t_n\}_{n=0}^N \to V$, with $V$ some Hilbert space,

$$(u_\tau, v_\tau)_{H^1(0,T;V),\tau} := \tau \sum_{n=0}^{N} \left(u^n, v^n\right)_{V \times V} + \tau^{-1} \sum_{n=0}^{N-1} \left(u^{n+1} - u^n, v^{n+1} - v^n\right)_{V \times V}.$$

This induces the norm $\|u_\tau\|^2_{H^1(0,T;V),h} := (u_\tau, u_\tau)_{H^1(0,T;V),h}$. For the discrete control space we obtain

$$\begin{aligned}
\mathcal{U}_\sigma &:= \{\boldsymbol{u}_\sigma = (u_{1,\sigma}, \dots, u_{M,\sigma}) \colon\ u_{i,\sigma} \in H_\tau(\mathbb{R}^2) \text{ for } i = 1, \dots, M\} \\
\mathcal{C}_\sigma &:= \{\boldsymbol{c}_\sigma = (c_{1,\sigma}, \dots, c_{M,\sigma}) \colon\ c_{i,\sigma} \in H_\tau(\mathbb{R}) \text{ for } i = 1, \dots, M\}, \\
\mathcal{Q}_\sigma &:= \mathcal{U}_\sigma \times \mathcal{C}_\sigma,
\end{aligned}$$

and the admissible set by

$$\begin{aligned}
\mathcal{U}_{\sigma,\mathrm{ad}} &:= \{\boldsymbol{u}_\sigma \in \mathcal{U}_\sigma \colon |u_i^n| \le 1,\ i = 1, \dots, M,\ n = 0, \dots, N\}, \\
\mathcal{C}_{\sigma,\mathrm{ad}} &:= \{\boldsymbol{c}_\sigma \in \mathcal{C}_\sigma \colon 0 \le c_i^n \le 1, i = 1, \dots, M, n = 0, \dots, N\}, \\
\mathcal{Q}_{\sigma,\mathrm{ad}} &:= \mathcal{U}_{\sigma,\mathrm{ad}} \times \mathcal{C}_{\sigma,\mathrm{ad}}.
\end{aligned}$$

The discrete state space is defined by

$$\mathcal{Y}_\sigma := H_\tau(V_h) \times H_\tau(W_{h,\mathrm{D}}) \times H_\tau(\mathbb{R}^2)^M.$$

With these definitions, the discrete optimal control problem related to (2.2) reads as

$$\begin{aligned}
\text{Minimize}\quad \mathcal{J}_\sigma(\boldsymbol{y}_\sigma, \boldsymbol{q}_\sigma) = \tau \sum_{n=1}^{N} e^{\nu t_n} \int_\Omega \rho_h^n(x)\,\mathrm{d}x - \mu\,\tau \sum_{i=1}^{M}\sum_{n=1}^{N} \ln(\eta_{x_i^n} * B_h) \\
+ \frac{\alpha_1}{2\,T} \sum_{i=1}^{M} \|u_{i,\sigma}\|^2_{H^1(0,T),\tau} + \frac{\alpha_2}{2\,T} \sum_{i=1}^{M} \|c_{i,\sigma}\|^2_{H^1(0,T),\tau} \qquad (3.12a)
\end{aligned}$$

$$\text{subject to}\quad \boldsymbol{y}_\sigma := (\rho_\sigma, \phi_\sigma, \boldsymbol{x}_\sigma) = S_\sigma(\boldsymbol{q}_\sigma), \qquad (3.12b)$$

$$\boldsymbol{q}_\sigma := (\boldsymbol{u}_\sigma, \boldsymbol{c}_\sigma) \in \mathcal{Q}_{\sigma,\mathrm{ad}}, \qquad (3.12c)$$

where is $B_h \in W_h$ the finite element approximation of (2.3) with first-order Lagrange elements. Furthermore, $S_\sigma$ is the solution operator of (3.6). Note that, in order to maintain the differentiability of the barrier term with respect to $x_i^n$, we use a regularization of the point evaluation of the nonsmooth function $B_h$, compare also (3.6c).

We may write the control problem (3.12) in the more compact reduced form

$$j_\sigma(\boldsymbol{q}_\sigma) := \mathcal{J}_\sigma(S_\sigma(\boldsymbol{q}_\sigma), \boldsymbol{q}_\sigma) \to \min! \quad \text{subject to} \quad \boldsymbol{q}_\sigma \in \mathcal{Q}_{\sigma,\mathrm{ad}}. \qquad (3.13)$$

To deduce a necessary optimality condition we apply the Lagrange formalism. The Lagrange function

$$\mathcal{L}_\sigma \colon \mathcal{Y}_\sigma \times \mathcal{Q}_\sigma \times \mathcal{Y}_\sigma \to \mathbb{R}$$

coupling the discrete state equation (3.6) reads

$$\begin{aligned}
&\mathcal{L}_\sigma(\rho_\sigma, \phi_\sigma, \boldsymbol{x}_\sigma; \boldsymbol{u}_\sigma, \boldsymbol{c}_\sigma; \lambda_{\rho,\sigma}, \lambda_{\phi,\sigma}, \lambda_{\boldsymbol{x},\sigma}) = \mathcal{J}_\sigma(\rho_\sigma, \phi_\sigma, \boldsymbol{x}_\sigma; \boldsymbol{u}_\sigma, \boldsymbol{c}_\sigma) \\
&\quad - \int_\Omega (\rho_h^0 - \mathrm{proj}_{V_h}(\rho_0))\, \lambda_{\rho,h}^0 \,\mathrm{d}x \\
&\quad - \sum_{n=0}^{N-1} \left( \int_\Omega (\rho_h^{n+1} - \rho_h^n)\, \lambda_{\rho,h}^{n+1} \,\mathrm{d}x + \tau\, a(\rho_h^{n+1}, \lambda_{\rho,h}^{n+1}) - \tau\, b(\beta_h^n)(\rho_h^n, \lambda_{\rho,h}^{n+1}) \right) \\
&\quad - \sum_{n=0}^{N-1} \tau \left( \delta_1 \int_\Omega \nabla\phi_h^n \cdot \nabla\lambda_{\phi,h}^n \,\mathrm{d}x + \int_\Omega |\nabla\phi_h^n|^2\, \lambda_{\phi,h}^n \,\mathrm{d}x - \int_\Omega \frac{1}{f(\rho_h^n)^2 + \delta_2}\, \lambda_{\phi,h}^n \,\mathrm{d}x \right) \\
&\quad - \sum_{i=1}^{M} \left( (x_i^0 - x_{i,0})^\top \lambda_{\boldsymbol{x},i}^0 + \sum_{n=0}^{N-1} \left( x_i^{n+1} - x_i^n - \tau\, f(\eta_{x_i^{n+1}} * \rho_h^{n+1})\, u_i^{n+1} \right)^\top \lambda_{\boldsymbol{x},i}^{n+1} \right).
\end{aligned}$$

To shorten the notation we write $\boldsymbol{\lambda}_\sigma := (\lambda_{\rho,\sigma}, \lambda_{\phi,\sigma}, \lambda_{\boldsymbol{x},\sigma})$. The adjoint equation system determining these variables for a given control and state is

$$\lambda_{\rho,h}^n \in V_h: \quad \frac{\partial \mathcal{L}_\sigma}{\partial \rho_h^n}(\boldsymbol{y}_\sigma, \boldsymbol{q}_\sigma, \boldsymbol{\lambda}_\sigma)\delta\rho_h = 0 \quad \forall \delta\rho_h \in V_h,\ n = 0, \dots, N, \tag{3.14a}$$

$$\lambda_{\phi,h}^n \in W_{h,\mathrm{D}}: \quad \frac{\partial \mathcal{L}_\sigma}{\partial \phi_h^n}(\boldsymbol{y}_\sigma, \boldsymbol{q}_\sigma, \boldsymbol{\lambda}_\sigma)\delta\phi_h = 0 \quad \forall \delta\phi_h \in W_h,\ n = 0, \dots, N-1, \tag{3.14b}$$

$$\lambda_{\boldsymbol{x}_i}^n \in \mathbb{R}^2: \quad \frac{\partial \mathcal{L}_\sigma}{\partial x_i^n}(\boldsymbol{y}_\sigma, \boldsymbol{q}_\sigma, \boldsymbol{\lambda}_\sigma)\delta x_i = 0 \quad \forall \delta x_i \in \mathbb{R}^2,\ n = 0, \dots, N \tag{3.14c}$$

for $i = 1, \dots, M$. Note that this can be interpreted as a coupled system involving a parabolic PDE and an ODE that run backward in time. We use the automatic differentiation feature in FEniCS in our implementation. As the forward system completely decouples in each time step, so does the adjoint system and we can compute step by step:

$$\lambda_{\boldsymbol{x},h}^N \mapsto \lambda_{\rho,h}^N \mapsto (\lambda_{\phi,h}^N) \mapsto \dots \mapsto \lambda_{\boldsymbol{x},h}^0 \mapsto \lambda_{\rho,h}^0 \mapsto \lambda_{\phi,h}^0.$$

With the adjoint states at hand we can assemble the derivatives of the reduced objective (3.13) and end up with the following optimality condition for (3.12):

**Theorem 3.3** (Necessary optimality condition) *Let $(\boldsymbol{y}_\sigma, \boldsymbol{q}_\sigma) \in \mathcal{Y}_\sigma \times \mathcal{Q}_{\sigma,\mathrm{ad}}$ be a local solution of (3.12). Then, there exists $\boldsymbol{\lambda}_\sigma \in \mathcal{Y}_\sigma$ fulfilling (3.14) and*

$$\begin{aligned}
&\alpha_1 (u_{\sigma,i}, v_{\sigma,i} - u_{\sigma,i})_{H^1(0,T;\mathbb{R}^2),\tau} + \alpha_2 (c_{\sigma,i}, d_{\sigma,i} - c_{\sigma,i})_{H^1(0,T),\tau} \\
&\quad + \tau \sum_{n=1}^{N} (f(\eta_{x_i^n} * \rho_h^n)\, \lambda_{\boldsymbol{x}_i}^n, v_i^n - {u_i}^n)_{\mathbb{R}^2} + \tau \sum_{n=0}^{N-1} \frac{\partial (b(\beta_h^n)(\rho_h^n, \lambda_{\rho,h}^{n+1}))}{\partial c_i^n} (d_i^n - c_i^n) \geq 0
\end{aligned} \tag{3.15}$$

*for all test functions $\boldsymbol{r}_\sigma := (\boldsymbol{v}_\sigma, \boldsymbol{d}_\sigma) \in \mathcal{Q}_{\sigma,\mathrm{ad}}$ and all $i = 1, \dots, M$.*

***Proof*** It is well-known that the variational inequality $j'_\sigma(\boldsymbol{q}_\sigma)(\boldsymbol{r}_\sigma - \boldsymbol{q}_\sigma) \geq 0$ for $\boldsymbol{r}_\sigma \in \mathcal{Q}_{\sigma,\mathrm{ad}}$ is necessary for $\boldsymbol{q}_\sigma$ being a local minimizer of (3.13). Taking into account the equivalence

$$j'_\sigma(\boldsymbol{q}_\sigma)\delta\boldsymbol{q}_\sigma = \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}_\sigma}(\boldsymbol{y}_\sigma, \boldsymbol{q}_\sigma, \boldsymbol{\lambda}_\sigma)\delta\boldsymbol{q}_\sigma \quad \text{if } \boldsymbol{\lambda}_\sigma \text{ solves (3.14)}$$

yields the variational inequality (3.15). □

Our solution algorithm is based on a projected gradient algorithm and it remains to establish a representation of the gradient of $j_\sigma$.

The derivative of the objective (3.12a) towards some direction $\delta\boldsymbol{q}_\sigma = (\delta\boldsymbol{u}_\sigma, \delta\boldsymbol{c}_\sigma) \in \mathcal{Q}$ reads

$$\begin{aligned} j'_\sigma(\boldsymbol{q}_\sigma)\delta\boldsymbol{q}_\sigma = \sum_{i=1}^{M} \Big( & \alpha_1 (u_{i,\sigma}, \delta u_{i,\sigma})_{H^1(0,T),\tau} + \alpha_2 (c_{i,\sigma}, \delta c_{i,\sigma})_{H^1(0,T),\tau} \\ & + \tau \sum_{n=1}^{N} f(\eta_{x_i^n} * \rho_h^n)\, \delta u_i^{n\top} \lambda_{\boldsymbol{x}_i}^n + \tau \sum_{n=0}^{N-1} \frac{\partial \left( b(\beta_h^n)(\rho_h^n, \lambda_{\rho,h}^{n+1}) \right)}{\partial c_i^n} \delta c_i^n \Big). \end{aligned}$$

To obtain a representation of the $H^1(0,T)$, $\tau$-gradient of $j_\sigma$ with respect to $\boldsymbol{u}_\sigma$ and $\boldsymbol{c}_\sigma$, we introduce the grid functions $z_{i,\sigma} : \{t_n\}_{n=0}^N \to \mathbb{R}^2$ and $d_{i,\sigma} : \{t_k\}_{n=0}^N \to \mathbb{R}$, $i = 1, \ldots, M$ solving

$$\left( \frac{1}{\tau^2} \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix} + I_{N+1\times N+1} \right) \begin{pmatrix} z_i^0 \\ z_i^1 \\ z_i^2 \\ \vdots \\ z_i^{N-1} \\ z_i^N \end{pmatrix} = \begin{pmatrix} 0 \\ -f(\rho_h^1(\boldsymbol{x}_i^1))\lambda_{x_i}^0 \\ -f(\rho_h^2(\boldsymbol{x}_i^2))\lambda_{x_i}^1 \\ \vdots \\ -f(\rho_h^{N-1}(\boldsymbol{x}_i^{N-1}))\lambda_{x_i}^{N-2} \\ -f(\rho_h^N(\boldsymbol{x}_i^N))\lambda_{x_i}^{N-1} \end{pmatrix} \tag{3.16a}$$

and

$$\left( \frac{1}{\tau^2} \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix} + I_{N+1\times N+1} \right) \begin{pmatrix} d_i^0 \\ d_i^1 \\ d_i^2 \\ \vdots \\ d_i^{N-1} \\ d_i^N \end{pmatrix} = \begin{pmatrix} -\partial_{c_i^0} b(\beta_h^0)(\rho_h^0, \lambda_{\rho,h}^1) \\ -\partial_{c_i^1} b(\beta_h^1)(\rho_h^1, \lambda_{\rho,h}^2) \\ -\partial_{c_i^2} b(\beta_h^2)(\rho_h^2, \lambda_{\rho,h}^3) \\ \vdots \\ -\partial_{c_i^{N-2}} b(\beta_h^{N-2})(\rho_h^{N-2}, \lambda_{\rho,h}^{N-1}) \\ -\partial_{c_i^{N-1}} b(\beta_h^{N-1})(\rho_h^{N-1}, \lambda_{\rho,h}^N) \\ 0 \end{pmatrix} \tag{3.16b}$$

for $i = 1, \dots, M$. By a simple calculation we then confirm

$$(z_{i,\sigma}, \delta u_{i,\sigma})_{H^1(0,T;\mathbb{R}^2),\tau} = \sum_{n=1}^{N} f(\eta_{x_i^n} * \rho_h^n)\, \delta u_i^{n\top} \lambda_{\boldsymbol{x}_i}^n,$$
$$(d_{i,\sigma}, \delta c_{i,\sigma})_{H^1(0,T),\tau} = \sum_{n=0}^{N-1} \frac{\partial \left( b(\beta_h^n)(\rho_h^n, \lambda_{\rho,h}^{n+1}) \right)}{\partial c_i^n} \delta c_i^n$$

We write $\boldsymbol{z}_\sigma = (z_{1,\sigma}, \dots, z_{M,\sigma}) \in \mathcal{U}$ and $\boldsymbol{d}_\sigma = (d_{1,\sigma}, \dots, d_{M,\sigma}) \in \mathcal{C}$ and get the following representation of the gradient of $j_\sigma$:

$$\nabla_{\boldsymbol{u}_\sigma} j_\sigma(\boldsymbol{q}_\sigma) = \alpha_1 \boldsymbol{u}_\sigma + \boldsymbol{z}_\sigma, \tag{3.17a}$$
$$\nabla_{\boldsymbol{c}_\sigma} j_\sigma(\boldsymbol{q}_\sigma) = \alpha_2 \boldsymbol{c}_\sigma + \boldsymbol{d}_\sigma. \tag{3.17b}$$

This allows an implementation of a projected gradient method which we discuss in the following section.

### 3.5 Optimization algorithms for the discretized problem

For a solution of the discretized optimal control problem 3.12 we propose a *projected gradient algorithm*. In this procedure, for a given initial control $\boldsymbol{q}^{(0)} = (\boldsymbol{u}^{(0)}, \boldsymbol{c}^{(0)}) \in \mathcal{Q}$, the new iterates are successively computed by means of

$$\boldsymbol{u}_\sigma^{(k+1)} = \boldsymbol{\Pi}_{\text{ad}}^u \left( \boldsymbol{u}_\sigma^{(k)} - s_{(k)} \nabla_{\boldsymbol{u}_\sigma} j_\sigma(\boldsymbol{q}_\sigma^{(k)}) \right), \tag{3.18a}$$
$$\boldsymbol{c}_\sigma^{(k+1)} = \boldsymbol{\Pi}_{\text{ad}}^c \left( \boldsymbol{c}_\sigma^{(k)} - s_{(k)} \nabla_{\boldsymbol{c}_\sigma} j_\sigma(\boldsymbol{q}_\sigma^{(k)}) \right), \tag{3.18b}$$

with $\boldsymbol{\Pi}_{\text{ad}}^u \colon \mathcal{U}_\sigma \to \mathcal{U}_{\sigma,\text{ad}}$ and $\boldsymbol{\Pi}_{\text{ad}}^c \colon \mathcal{C}_\sigma \to \mathcal{C}_{\sigma,\text{ad}}$ the $H^1(0,T), \tau$ projections onto the admissible sets $\mathcal{U}_{\sigma,\text{ad}}$ and $\mathcal{C}_{\sigma,\text{ad}}$, respectively, this is,

$$\boldsymbol{\Pi}_{\text{ad}}^u(\boldsymbol{u}_\sigma) := \operatorname*{arg\,min}_{\boldsymbol{v}_\sigma \in \mathcal{U}_{\sigma,\text{ad}}} \frac{1}{2} \| \boldsymbol{u}_\sigma - \boldsymbol{v}_\sigma \|^2_{H^1(0,T;\mathbb{R}^2),\tau}, \tag{3.19a}$$
$$\boldsymbol{\Pi}_{\text{ad}}^c(\boldsymbol{c}_\sigma) := \operatorname*{arg\,min}_{\boldsymbol{d}_\sigma \in \mathcal{C}_{\sigma,\text{ad}}} \frac{1}{2} \| \boldsymbol{c}_\sigma - \boldsymbol{d}_\sigma \|^2_{H^1(0,T),\tau}. \tag{3.19b}$$

A formula for the gradient of $j_\sigma$ has been derived in the previous section already, see (3.17). The step length parameter $s^{(k)} > 0$ is obtained by an Amijo line search [27, Section 2.2.2.1] and must fulfill the sufficient decrease condition

$$\begin{aligned} &j_\sigma \left( \boldsymbol{\Pi}_{\text{ad}} \left( \boldsymbol{q}_\sigma^{(k)} - s_{(k)} \nabla j_\sigma(\boldsymbol{q}_\sigma^{(k)}) \right) \right) \\ &\quad \le j_\sigma(\boldsymbol{q}_\sigma^{(k)}) - \frac{d}{s_{(k)}} \| \boldsymbol{q}_\sigma^{(k)} - \boldsymbol{\Pi}_{\text{ad}} \left( \boldsymbol{q}_\sigma^{(k)} - s^{(k)} \nabla j_\sigma(\boldsymbol{q}_\sigma^{(k)}) \right) \|^2_{\mathcal{Q}_\sigma} \end{aligned} \tag{3.20}$$

with a decrease parameter $d \in (0,1)$ which is usually small (e.g., $10^{-4}$). A reasonable stopping criterion for the projected gradient algorithm is

$$\|\boldsymbol{q}_\sigma^{(k)} - \boldsymbol{\Pi}_{\mathrm{ad}}\left(\boldsymbol{q}_\sigma^{(k)} - \nabla j_\sigma(\boldsymbol{q}_\sigma^{(k)})\right)\|_{\mathcal{Q}_\sigma} \le 10^{-3}.$$

It remains to discuss the realization of the projection operators and we propose a primal dual active set strategy that may also be considered as semismooth Newton method. Note that the operators $\boldsymbol{\Pi}_{\mathrm{ad}}$ are semismooth, see [15]. The evaluation of the projection operator $\boldsymbol{\Pi}_{\mathrm{ad}}^u \colon \mathcal{U}_\sigma \to \mathcal{U}_{\sigma,\mathrm{ad}}$ requires to solve the optimization problem (3.19a). The unknowns (assuming $M = 1$ and omitting the agent's index $i$ for a while) are the coefficients of the functions $\mathcal{U}_\sigma \ni \boldsymbol{\Pi}_{\mathrm{ad}}(u_\sigma) = w_\sigma \simeq \vec{w} \in \mathbb{R}^{(N+1)\times 2}$ for some given $\mathcal{U}_\sigma \ni u_\sigma \simeq \vec{u} \in \mathbb{R}^{(N+1)\times 2}$. We switch to a matrix-vector notation and define

$$w^n := \begin{pmatrix} w_1^n \\ w_2^n \end{pmatrix} := w_\sigma(t_n), \qquad \vec{w}_j := (w_j^0, \dots, w_j^N)^\top,\ j = 1, 2,$$

as well as the matrix $A \in \mathbb{R}^{(N+1)\times(N+1)}$ on the left-hand side of the linear system (3.16) inducing the discrete $H^1(0,T)$, $\tau$-norm.

The Lagrangian for (3.19) reads

$$L(\vec{w}_1, \vec{w}_2, \vec{\lambda}) = \frac{1}{2}\sum_{i=1}^{2}(\vec{w}_i - \vec{u}_i)^\top A(\vec{w}_i - \vec{u}_i) - \frac{1}{2}\lambda^\top\left(|\vec{w}|_*^2 - \vec{1}\right),$$

with $|\vec{w}|_*^2 = (|w^0|^2, \dots, |w^N|^2)^\top$. The Karush-Kuhn-Tucker system for (3.19) then reads

$$\begin{aligned} A\,(\vec{w}_i - \vec{u}_i) - \vec{\lambda}\cdot\vec{w}_i = 0 \qquad i = 1, 2, \\ \frac{1}{2}\left(|\vec{w}|_*^2 - \vec{1}\right) \le 0, \qquad \vec{\lambda} \ge 0, \qquad \frac{1}{2}\vec{\lambda}\cdot(|\vec{w}|_*^2 - 1) = 0, \end{aligned}$$

where $\cdot$ is the component-wise multiplication of two vectors. We reformulate the complementarity condition by means of a nonsmooth equation and arrive at the following equivalent form of the KKT system

$$F(\vec{w}_1, \vec{w}_2, \vec{\lambda}) := \begin{bmatrix} A\,(\vec{w}_1 - \vec{u}_1) - \vec{\lambda}\cdot\vec{w}_1 \\ A\,(\vec{w}_2 - \vec{u}_2) - \vec{\lambda}\cdot\vec{w}_2 \\ \vec{\lambda} - \max\{0, -\frac{1}{2}(|\vec{w}|_*^2 - \vec{1}) + \vec{\lambda}\} \end{bmatrix} = 0. \tag{3.21}$$

This nonlinear system can be solved iteratively by a semismooth Newton method. Given is an initial pair $(\vec{u}^{(0)}, \vec{\lambda}^{(0)})$. Successively, one computes the active and inactive set

$$\begin{aligned} \mathcal{A}^{(k)} &:= \{n \in \{0, \dots, N\} \colon -\frac{1}{2}(|w^n|_2^2 - 1) + \lambda^n > 0\}, \\ \mathcal{I}^{(k)} &:= \{0, \dots, N\} \setminus \mathcal{A}^{(k)}, \end{aligned}$$

solves the Newton system

$$\begin{bmatrix} A - D_{\vec{\lambda}^{(k)}} & 0 & -D_{\vec{w}_1^{(k)}} \\ 0 & A - D_{\vec{\lambda}^{(k)}} & -D_{\vec{w}_2^{(k)}} \\ D_{\mathcal{A}^{(k)}} D_{\vec{w}_1^{(k)}} & D_{\mathcal{A}^{(k)}} D_{\vec{w}_2^{(k)}} & D_{\mathcal{I}^{(k)}} \end{bmatrix} \begin{bmatrix} \delta\vec{w}_1 \\ \delta\vec{w}_2 \\ \delta\vec{\lambda} \end{bmatrix} = - \begin{bmatrix} A\,(\vec{w}_1^{(k)} - \vec{u}_1) - \vec{\lambda}^{(k)} \cdot \vec{w}_1^{(k)} \\ A\,(\vec{w}_2^{(k)} - \vec{u}_2) - \vec{\lambda}^{(k)} \cdot \vec{w}_2^{(k)} \\ \vec{\lambda}^{(k)} - \max\{0, -\frac{1}{2}\,(|\vec{w}^{(k)}|_*^2 - \vec{1}) + \vec{\lambda}^{(k)}\}, \end{bmatrix}$$

with the diagonal matrices $D_{\vec{v}} = \operatorname{diag}(\vec{v})$ for $\vec{v} \in \mathbb{R}^{N+1}$ and $D_{\mathcal{M}} = \operatorname{diag}(\chi_{\mathcal{M}})$ for $\mathcal{M} \subset \{0, \dots, N\}$, and performs the Newton update

$$\vec{w}^{(k+1)} = \vec{w}^{(k)} + \delta\vec{w}, \qquad \vec{\lambda}^{(k+1)} = \vec{\lambda}^{(k)} + \delta\vec{\lambda}.$$

This procedure is repeated for $k = 0, 1, \dots$ until some termination criterion, e. g., $\|F(\vec{w}_1, \vec{w}_2, \vec{\lambda})\| < \mathrm{tol}$, is fulfilled.

## 4 Numerical experiments

This section is devoted to numerical experiments. To establish the discretized system (3.12) the finite element library FEniCS was used, complemented by a Python implementation of the projected gradient method from (3.18) and the Armijo step size rule from (3.20). The computational meshes were created by the mesh generator `mshr` integrated in FEniCS.

### 4.1 Example 1

In a first numerical test we solve the problem Equation (2.2) in the domain $\Omega$ depicted in Fig. 1 with the following parameters.

$$\begin{array}{lllllll} T = 9 & n_T = 300 & \alpha_1 = \alpha_2 = 5 \cdot 10^{-2} & \gamma = 10 & \zeta = 10^{-2} & \mu = 5 \cdot 10^{-2} \\ \varepsilon = 10^{-5} & \delta_1 = 0.2 & \delta_2 = 0.1 & \delta_3 = 10^{-2} & \delta_4 = 0.1 & \mathrm{tol} = 10^{-2}. \end{array}$$

The initial density $\rho_0$ is the sum of 6 Gaussian bells, see also Fig. 1a. The subdomain $\widetilde{\Omega}$ where densities are penalized is chosen to cover the region within the walls. Without any controlled agents, most of the people will squeeze through the 2 smaller emergency exits in the south and north while the large exit in the east is rarely used. To improve the evacuation 3 agents were introduced. The initial control $(\boldsymbol{u}, \boldsymbol{c})$ was chosen in such a way that the agent moves straight to the right outside of the room having a constant the intensity. The projected gradient method reached the desired tolerance after 4372 iterations. The Armijo parameter was halved until (3.20) is fulfilled and the initial Armijo parameter in the subsequent iteration was chosen according to

$$\begin{aligned} s_{(k)}^0 &= s_{(k-1)} \cdot \min\{1.5, 1.1 \cdot \max\{1, \mathrm{res}_{(k-1)}^2 \,/\, \mathrm{res}_{(k)}^2\}\}, \\ \text{with} \quad \mathrm{res}_{(k)} &= \|\boldsymbol{q}_\sigma^{(k)} - \boldsymbol{\Pi}_{\mathrm{ad}}(\boldsymbol{q}_\sigma^{(k)} - \nabla j_\sigma(\boldsymbol{q}_\sigma^{(k)}))\|_{\mathcal{Q}_\sigma}. \end{aligned}$$

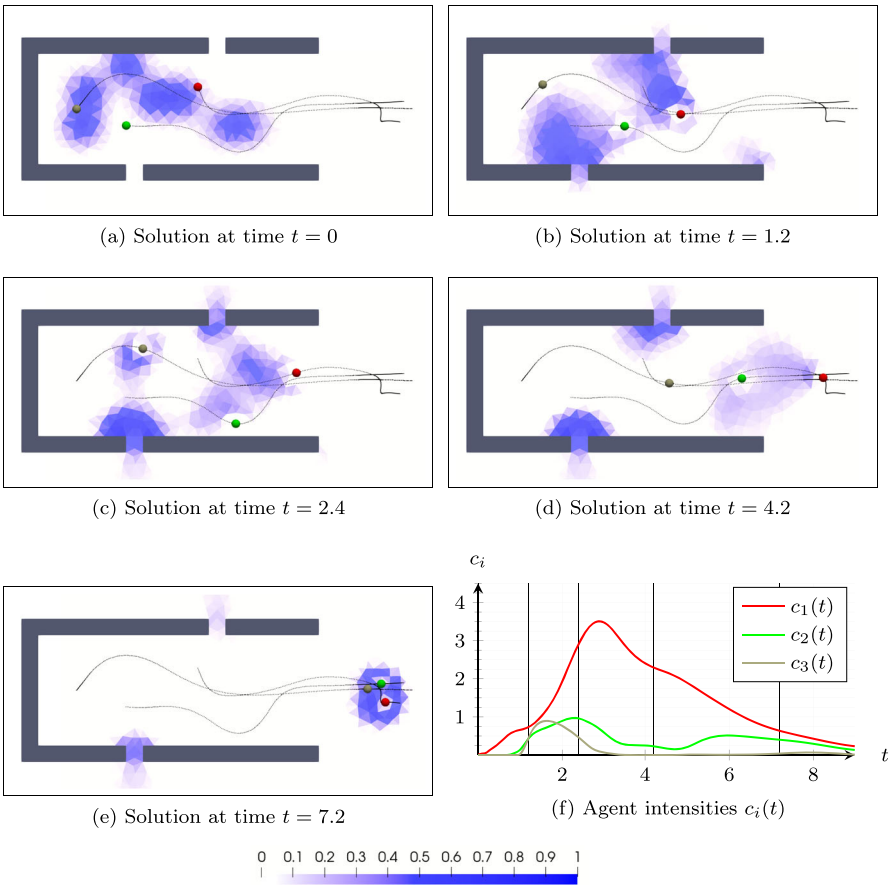
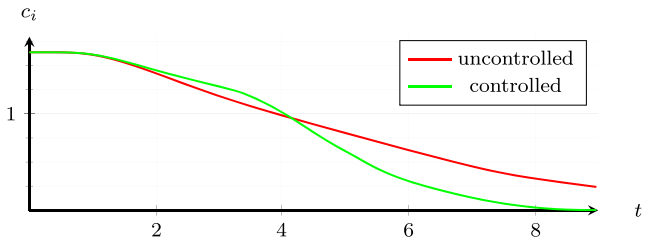


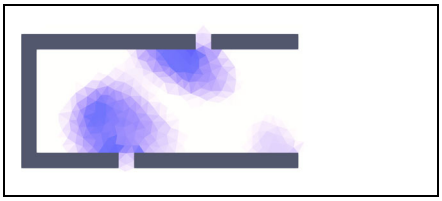

**Fig. 1** Solution of the problem from Sect. 4.1 at different time slices (vertical lines in f). The colored background represents the density $\rho$; the dots are the agent positions; the black curves are the agent trajectories

This provides the possibility to moderately increase the step size while at the same time avoiding too many iterations in the Armijo loop. By this choice at most 2 functional evaluations per iteration are necessary.

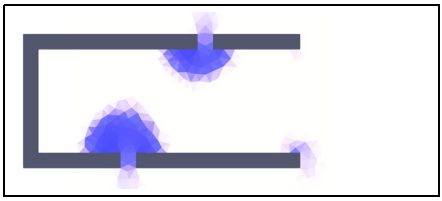
In Fig. 2 we compared the results of the optimized evacuation with the uncontrolled case, meaning that no agents are present resulting in a crowd motion where each individual is aiming at the closest emergency exit. For most individuals in the crowd the small side exits are the closest and the large exit in the east is rarely used. An interesting observation is that the total number of individuals, $m(t) = \int_{\widetilde{\Omega}} \rho(t, x)\, \mathrm{d}x$, is smaller in the uncontrolled case in the early stage but afterwards the controlled evacuation becomes much better. This is preferred by the objective due to the




(a) Total mass $m(t)$



(b) Solution at time $t = 0.9$, uncontrolled



(c) Solution at time $t = 2.1$, uncontrolled

**Fig. 2** Comparison between controlled and uncontrolled case: Plot of the total mass $m(t) := \int_{\widetilde{\Omega}} \rho(t, x)\mathrm{d}x$ against time and solution of the uncontrolled system at different time points

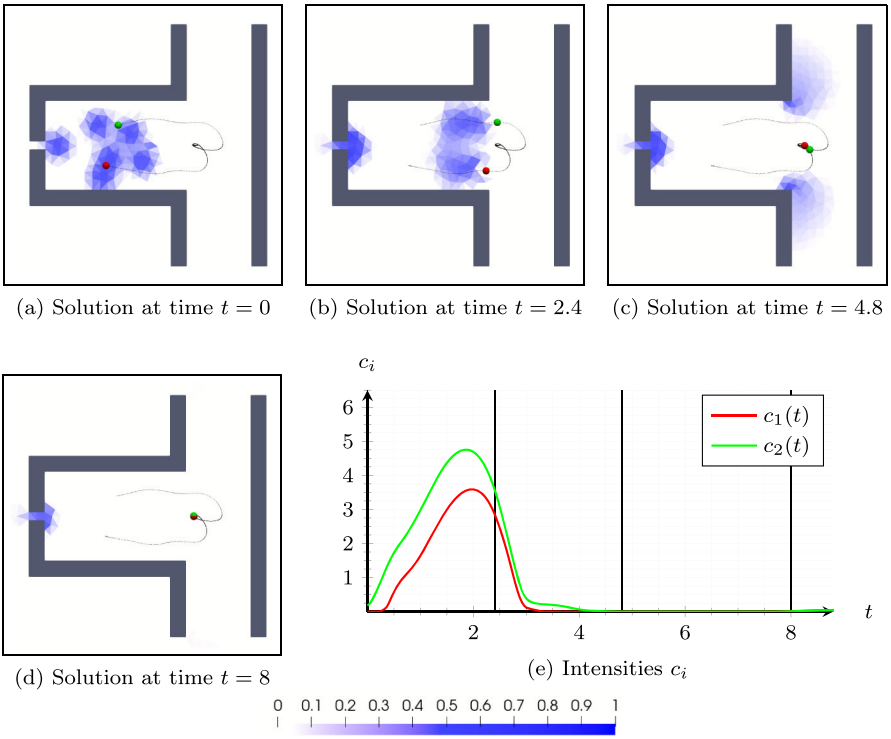
weight function $e^{\nu t}$ penalizing high densities at later time stronger than at an earlier time.

### 4.2 Example 2

In a second example we consider the domain illustrated in Fig. 3. The initial density is concentrated near the slit in the wall on the left-hand side. In an uncontrolled evacuation scenario the majority of the people would leave the domain through this slit causing a massive congestion leading to a very slow evacuation of the crowd. The model and algorithm parameters chosen in the current example are as follows:

$$T = 12 \quad n_T = 300 \quad \alpha_1 = \alpha_2 = 5 \cdot 10^{-2} \quad \gamma = 10 \quad \zeta = 10^{-2} \quad \mu = 5 \cdot 10^{-2}$$
$$\varepsilon = 10^{-5} \quad \delta_1 = 0.2 \quad \delta_2 = 0.1 \quad \delta_3 = 10^{-2} \quad \delta_4 = 0.1 \quad \mathrm{tol} = 10^{-2}.$$

The tolerance was reached after 3889 iterations. This example shows that the evacuation can be significantly improved by using two agents with optimized trajectory and intensity. Interesting is, that the intensity is non-zero only in the time interval $t \in (0, 3)$. The agents attract the people leading them sufficiently far away from the slit in the west and then they stop influencing the crowd. When being sufficiently far away from the slit the people find the way to the larger exits in the north and south on their own by using the movement direction determined by the potential $\phi$.




(a) Solution at time $t = 0$ (b) Solution at time $t = 2.4$ (c) Solution at time $t = 4.8$

(d) Solution at time $t = 8$ (e) Intensities $c_i$

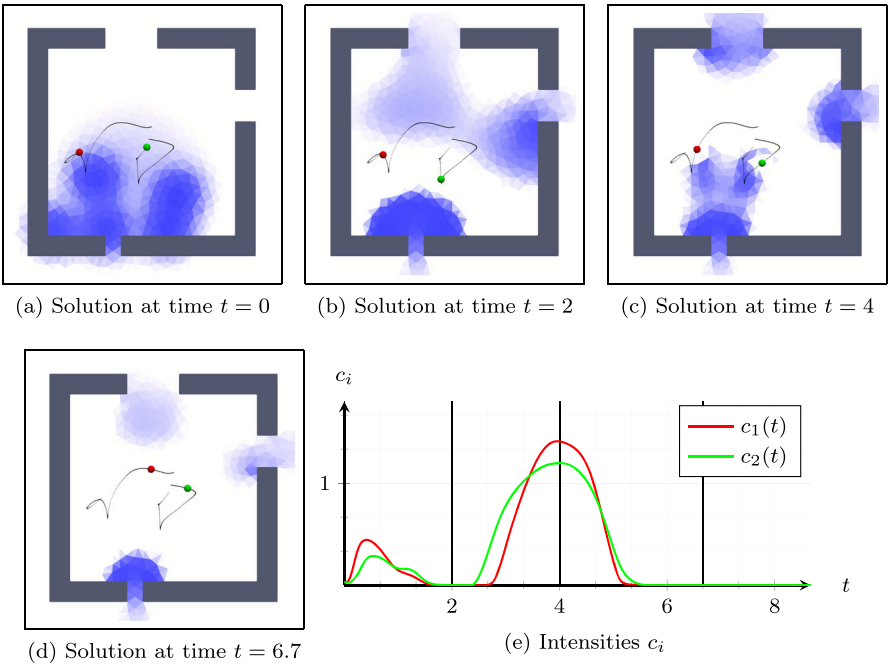
**Fig. 3** Solution of the problem from Sect. 4.2 at different time slices (vertical lines in e) and intensity of the agents

### 4.3 Example 3

In a third example we consider a square-shaped room with exits in the south, east and north. The computational results are illustrated in Fig. 4. The exits have different widths. The model and algorithm parameters are chosen as follows:

$$T = 10 \quad n_T = 300 \quad \alpha_1 = \alpha_2 = 5 \cdot 10^{-2} \quad \gamma = 10 \quad \zeta = 10^{-2} \quad \mu = 5 \cdot 10^{-2}$$
$$\varepsilon = 10^{-5} \quad \delta_1 = 0.2 \quad \delta_2 = 0.1 \quad \delta_3 = 10^{-2} \quad \delta_4 = 0.1 \quad \text{tol} = 10^{-2}.$$

The initial density is concentrated near the small exit and without a control of the crowd motion most of the people are blocking each other while squeezing through this small exit. Two agents were added in this scenario with the aim attracting the people in such a way that more of them find the other two exits in the north and east. The optimization algorithm reached the desired tolerance after 9283 iterations. The computed agent trajectories are quite short. It is interesting to observe that in the time interval $t \in (0, 2)$ the agents just go to an optimal position sufficiently close to the crowd and attract them in the time interval $t \in (2, 5)$, leading some of the people to the center of the room. At this point the agents drive their intensity to zero meaning




(a) Solution at time $t = 0$ (b) Solution at time $t = 2$ (c) Solution at time $t = 4$

(d) Solution at time $t = 6.7$ (e) Intensities $c_i$

**Fig. 4** Solution of the problem from Sect. 4.3 at different time slices (vertical lines in e) and intensity of the agents

that they stop influencing the crowd. However, when being sufficiently far away from the critical exit the people find the route to the less used exits on their own due to the movement rule determined by the potential $\phi$.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data Availability** The source code for the numerical experiments are publicly available on https://github.com/max-winkler/hughes_control.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References


1. Albi, G., Bongini, M., Cristiani, E., Kalise, D.: Invisible control of self-organizing agents leaving unknown environments. SIAM J. Appl. Math. **76**(4), 1683–1710 (2016). https://doi.org/10.1137/15M1017016
2. Albi, G., Fornasier, M., Kalise, D.: A Boltzmann approach to mean-field sparse feedback control. IFAC-PapersOnLine **50**(1), 2898–2903 (2017)
3. Amadori, D., Di Francesco, M.: The one-dimensional Hughes model for pedestrian flow: Riemann-type solutions. Acta Math. Sci. **32**(1), 259–280 (2012). https://doi.org/10.1016/s0252-9602(12)60016-2
4. Amadori, D., Goatin, P., Rosini, M.D.: Existence results for Hughes' model for pedestrian flows. J. Math. Anal. Appl. **420**(1), 387–406 (2014). https://doi.org/10.1016/j.jmaa.2014.05.072
5. Banda, M.K., Herty, M., Trimborn, T.: Recent developments in controlled crowd dynamics. In: Crowd Dynamics, Volume 2, pp. 133–157. Springer International Publishing (2020). https://doi.org/10.1007/978-3-030-50450-2_7
6. Borsche, R., Colombo, R.M., Garavello, M., Meurer, A.: Differential equations modeling crowd interactions. J. Nonlinear Sci. **25**(4), 827–859 (2015). https://doi.org/10.1007/s00332-015-9242-0
7. Borsche, R., Klar, A., Kühn, S., Meurer, A.: Coupling traffic flow networks to pedestrian motion. Math. Models Methods Appl. Sci. **24**(2), 359–380 (2014). https://doi.org/10.1142/S0218202513400113
8. Borsche, R., Meurer, A.: Microscopic and macroscopic models for coupled car traffic and pedestrian flow. J. Comput. Appl. Math. **348**, 356–382 (2019). https://doi.org/10.1016/j.cam.2018.08.037
9. Burger, M., Di Francesco, M., Markowich, P.A., Wolfram, M.T.: Mean field games with nonlinear mobilities in pedestrian dynamics. Discrete Contin. Dyn. Syst. - B **19**(5), 1311–1333 (2014). https://doi.org/10.3934/dcdsb.2014.19.1311
10. Burger, M., Pinnau, R., Totzeck, C., Tse, O., Roth, A.: Instantaneous control of interacting particle systems in the mean-field limit. J. Comput. Phys. **405**, 109181, 20 (2020). https://doi.org/10.1016/j.jcp.2019.109181
11. Caponigro, M., Fornasier, M., Piccoli, B., Trélat, E.: Sparse stabilization and optimal control of the Cucker-Smale model. Math. Control Relat. Fields **3**(4), 447–466 (2013). https://doi.org/10.3934/mcrf.2013.3.447
12. Carillo, J.A., Huang, Y., Martin, S.: Explicit flock solutions for Quasi-Morse potentials. Eur. J. Appl. Math. **25**(5), 553–578 (2014). https://doi.org/10.1017/s0956792514000126
13. Carlini, E., Festa, A., Silva, F.J., Wolfram, M.T.: A semi-Lagrangian scheme for a modified version of the Hughes' model for pedestrian flow. Dyn. Games Appl. **7**(4), 683–705 (2016). https://doi.org/10.1007/s13235-016-0202-6
14. Carrillo, J.A., Martin, S., Wolfram, M.T.: An improved version of the Hughes model for pedestrian flow. Math. Models Methods Appl. Sci. **26**(04), 671–697 (2016). https://doi.org/10.1142/s0218202516500147
15. Christof, C., Wachsmuth, G.: Semismoothness for solution operators of obstacle-type variational inequalities with applications in optimal control. arXiv:2112.12018 (2021)
16. Colombo, R.M., Gokieli, M., Rosini, M.D.: Modeling crowd dynamics through hyperbolic-elliptic equations. In: Non-Linear Partial Differential Equations, Mathematical Physics, and Stochastic Analysis, EMS Series of Congress Reports, pp. 111–128. European Mathematical Society, Zürich (2018). https://doi.org/10.4171/186-1/6
17. Denk, R., Hieber, M., Prüss, J.: Optimal $L^p$-$L^q$-estimates for parabolic boundary value problems with inhomogeneous data. Math. Z. **257**(1), 193–224 (2007). https://doi.org/10.1007/s00209-007-0120-9
18. Di Francesco, M., Fagioli, S., Rosini, M.D., Russo, G.: Deterministic particle approximation of the Hughes model in one space dimension. Kinet. Relat. Models **10**(1), 215–237 (2017). https://doi.org/10.3934/krm.2017009
19. Di Francesco, M., Markowich, P.A., Pietschmann, J.F., Wolfram, M.T.: On the Hughes' model for pedestrian flow: The one-dimensional case. J. Differ. Equ. **250**(3), 1334–1362 (2011). https://doi.org/10.1016/j.jde.2010.10.015
20. Di Pietro, D.A., Ern, A.: Mathematical Aspects of Discontinuous Galerkin Methods. Springer Berlin Heidelberg (2012). https://doi.org/10.1007/978-3-642-22980-0
21. El-Khatib, N., Goatin, P., Rosini, M.D.: On entropy weak solutions of Hughes' model for pedestrian motion. Zeitschrift für Angewandte Mathematik und Physik. ZAMP. J. Appl. Math. Phys. Journal de Mathématiques et de hysique Appliquées **64**(2), 223–251 (2013). https://doi.org/10.1007/s00033-012-0232-x

22. Epshteyn, Y., Kurganov, A.: New interior penalty discontinuous galerkin methods for the Keller-Segel chemotaxis model. SIAM J. Numer. Anal. **47**(1), 386–408 (2009). https://doi.org/10.1137/07070423x
23. Filbet, F.: A finite volume scheme for the Patlak-Keller-Segel chemotaxis model. Numer. Math. **104**(4), 457–488 (2006). https://doi.org/10.1007/s00211-006-0024-3
24. Guo, L., Li, X.H., Yang, Y.: Energy dissipative local discontinuous galerkin methods for Keller-Segel chemotaxis model. J. Sci. Comput. **78**(3), 1387–1404 (2018). https://doi.org/10.1007/s10915-018-0813-8
25. Herzog, R., Pietschmann, J.F., Winkler, M.: Optimal control of hughes' model for pedestrian flow via local attraction. Appl. Math. Optim. **88**(3), 87 (2023). https://doi.org/10.1007/s00245-023-10064-8
26. Himakalasa, A., Wongkaew, S.: Optimal control through leadership of the cucker and smale flocking model with time delays. Complexity **2021**, 1–14 (2021). https://doi.org/10.1155/2021/5545551
27. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE constraints. Mathematical Modelling: Theory and Applications, vol. 23. Springer, New York (2009)
28. Hughes, R.L.: A continuum theory for the flow of pedestrians. Transp. Res. B Methodol. **36**(6), 507–535 (2002). https://doi.org/10.1016/s0191-2615(01)00015-7
29. Ibrahim, M., Saad, M.: On the efficacy of a control volume finite element method for the capture of patterns for a volume-filling chemotaxis model. Comput. Math Appl. **68**(9), 1032–1051 (2014). https://doi.org/10.1016/j.camwa.2014.03.010
30. Li, X.H., Shu, C.W., Yang, Y.: Local discontinuous galerkin method for the Keller-Segel chemotaxis model. J. Sci. Comput. **73**(2–3), 943–967 (2017). https://doi.org/10.1007/s10915-016-0354-y
31. Pinnau, R., Totzeck, C.: Interacting particles and optimization. PAMM **18**(1) (2018). https://doi.org/10.1002/pamm.201800182
32. Rider, W.J., Lowrie, R.B.: The use of classical Lax-Friedrichs Riemann solvers with discontinuous Galerkin methods. pp. 479–486 (2002). https://doi.org/10.1002/fld.334. ICFD Conference on Numerical Methods for Fluid Dynamics, Part II (Oxford, 2001)
33. Strehl, R., Sokolov, A., Kuzmin, D., Turek, S.: A flux-corrected finite element method for chemotaxis problems. Comput. Methods. Appl. Math. **10**(2), 219–232 (2010). https://doi.org/10.2478/cmam-2010-0013