Advances in Applied Mathematics and Mechanics
Adv. Appl. Math. Mech., Vol. **11**, No. 6, pp. 1358-1375

DOI: 10.4208/aamm.OA-2018-0186
December 2019

Analysis of Finite Difference Approximations of an Optimal Control Problem in Economics

Alexander Lapin[1,2], Shuhua Zhang[2], Sergey Lapin[3] and Na Yan[4,*]

[1] *Institute of Computational Mathematics and Information Technologies, Kazan Federal University, Kazan 420008, Russia*
[2] *Coordinated Innovation Center for Computable Modeling in Management Science, Tianjin University of Finance and Economics, Tianjin 300222, China*
[3] *Department of Mathematics and Statistics, Washington State University, Pullman, WA 99163, USA*
[4] *Business School, Nankai University, Tianjin 300071, China*

Received 26 August 2018; Accepted (in revised version) 21 April 2019

Abstract. We consider an optimal control problem which serves as a mathematical model for several problems in economics and management. The problem is the minimization of a continuous constrained functional governed by a linear parabolic diffusion-advection equation controlled in a coefficient in advection part. The additional constraint is non-negativity of a solution of state equation. We construct and analyze several mesh schemes approximating the formulated problem using finite difference methods in space and in time. All these approximations keep the positivity of the solutions to mesh state problem, either unconditionally or under some additional constraints to mesh steps. This allows us to remove corresponding constraint from the formulation of the discrete problem to simplify its implementation. Based on theoretical estimates and numerical results, we draw conclusions about the quality of the proposed mesh schemes.

AMS subject classifications: 65M06, 65M12, 65M60

Key words: Mean field game, optimal control problem, parabolic diffusion-advection equation, finite difference methods.

1 Introduction

The theory of optimal control is involved in many models in economics. An actual and common method of solving such problems is mean field game (MFG) formulation. This approach was proposed in [1,2] and describes situations of equilibrium by considering a

*Corresponding author.
Emails: shuhua55@126.com (S. H. Zhang), doforget@sina.com (N. Yan)

continuum of players (also called agents) through two forward/backward coupled PDE system. In this system the first equation is a transport equation (or, Fokker-Planck equation) describing the evolution of the distribution of agents, while the second equation corresponds to a Hamilton-Jacobi-Bellman equation derived from the optimization of a criterion by means of a control. MFG formulation of the problems are thoroughly used both for their theoretical study and numerical solution (see [3–8] and the bibliography therein).

The link between MGF and optimal control takes place in the so-called potential case (see [7] for the details). In this case equilibrium system solution of MFG is a critical point of an optimal control problem governed by transport equation.

In this paper we consider a problem of minimization of a cost functional

$$J(m,\alpha)=\int_0^T\int_0^1 \mathrm{e}^{-rt}m\big(f(m)+\frac{\alpha^2}{2}\big)dxdt, \qquad r=\mathrm{const}\geq 0,$$

with respect to pair state-control (m,α), which satisfy a linear diffusion-advection state equation

$$\frac{\partial m}{\partial t}-\frac{\sigma^2}{2}\frac{\partial^2 m}{\partial x^2}-\frac{\partial}{\partial x}(\alpha m)=0$$

controlled in a coefficient of advection term. Some "economical meanings" of the problem and concrete examples of the functions in this formulation can be found in [8,9].

In the case of differentiable function $f(m)$ we can use Lagrange function to write first order optimality conditions for the considered minimization problem. It is just MFG formulation of the problem, which resulting system consists of two coupled forward-backward parabolic equations and a term that represents the derivative J'_α. In [8] this system was approximated by using finite elements in space and implicit scheme in time, and solved by an iterative solution method. In [9] the initial minimization problem was approximated and solved by a new monotone iterative algorithm. State equation–transport equation–was approximated by a finite difference scheme in space and by explicit approximation in time.

One of the most important property of the solutions m of any mesh approximation of state equation is its positivity. This corresponds to its meaning as the density of agents and, moreover, it is necessary from mathematical point of view, because the cost functional is not bounded from below for functions m which can have the negative values. In [9] this property is saved for the constructed approximation by introducing an additional constraint which connects mesh parameters and sought solution. In the article [8] there were no theoretical studies of the constructed approximations.

The aim of this article is construction and investigation theoretically and numerically several finite difference approximations of the mentioned above optimal control problem. To approximate state equation we use so-called summation equality in space variable and one of the following approximations in time: fully implicit (backward Euler), semi-implicit or fully explicit (forward Euler). The solutions of all constructed approximations

of state equation have positive solutions and keep an analogue of the mass balance condition. We prove the positivity of a solution m of the implicit scheme for any control α. On the other hand, the solutions of semi-implicit and explicit schemes are positive if some additional constraints connecting max-norm of control α and mesh steps are satisfied. Since control function is one of the unknowns in the problem, we cannot satisfy the positivity condition *a priori*. Despite that it is only sufficient condition, violating it can lead to a wrong solution. An example of such situation is given in the article.

We prove the existence of a solution of mesh optimal control problem for all considered approximations of the state equation. The solution is not generally unique, non-uniqueness is demonstrated by a calculation example.

The theoretical results are complemented by numerical tests. To find the extremal points of the mesh cost functions quasi-Newton method with line search procedure for finding an optimal iterative parameter is used.

In this paper we use, for simplicity, uniform meshes, however all the results are also valid for irregular meshes or different (staggered) grids for the approximation of the state and control functions.

2 Differential problem

Let $Q_T=(0,1)\times(0,T]$, $\sigma=\text{const}\neq 0$, and the functions $m(x,t)$ and $\alpha(x,t)$ be defined in closed domain $\bar{Q}_T=[0,1]\times[0,T]$ and satisfy the following initial-boundary value problem:

$$\frac{\partial m}{\partial t}-\frac{\sigma^2}{2}\frac{\partial^2 m}{\partial x^2}-\frac{\partial}{\partial x}(\alpha m)=0 \qquad \text{for} \quad (x,t)\in Q_T, \tag{2.1a}$$

$$\frac{\partial m}{\partial x}=0 \quad \text{for} \quad x=0, \quad x=1 \quad \text{and} \quad t\in(0,T], \tag{2.1b}$$

$$m(x,0)=m_0(x)\geq 0. \tag{2.1c}$$

The functions $\alpha(x,t)$ and $m(x,t)$ are control and state functions, respectively. We impose additional state constraint:

$$m(x,t)\geq 0, \quad \forall (x,t)\in Q_T. \tag{2.2}$$

Control function $\alpha(x,t)$ satisfies the boundary condition

$$\alpha(0,t)=\alpha(1,t)=0 \quad \text{for all} \quad t\in(0,T]. \tag{2.3}$$

Due to (2.3) any solution m of problem (2.1) satisfies the following mass balance property:

$$\int_0^1 m(x,t)dx=\int_0^1 m_0(x)dx=\text{const}, \quad \forall t\in[0,T]. \tag{2.4}$$

To define a weak solution of the initial-boundary value problem (2.1) we introduce several functional spaces (see [10,11]): Lebesgue space $L^2(0,1)$, Sobolev space $H^1(0,1)$ and its

conjugate $(H^1(0,1))^*$, subspace $H_0^1(0,1)\subset H^1(0,1)$ of functions vanishing on the boundary. Below the brackets $\langle\cdot,\cdot\rangle$ mean the dual pairing between $(H^1(0,1))^*$ and $H^1(0,1)$. For the functions $u(t):[0,T]\to V$ which are measurable on the segment $[0,T]$ with the values in a Banach space V we use the spaces $L^2(0,T;V)$, $L^\infty(Q_T)$ and

$$W(0,T)=\Big\{u\in L^2(0,T;H^1(0,1)):\frac{\partial u}{\partial t}\in L^2(0,T;(H^1(0,1))^*)\Big\}$$

with the norm

$$\|u\|_{W(0,T)}=\|u\|_{L^2(0,T;H^1(0,1))}+\Big\|\frac{\partial u}{\partial t}\Big\|_{L^2(0,T;(H^1(0,1))^*)}.$$

For given functions $\alpha\in L^\infty(Q_T)$ and $m_0\in L^2(0,1)$ we define as a weak solution of (2.1) a function $m\in W(0,T)$, which satisfies the following variational equality:

$$\int_0^T\langle\frac{\partial m}{\partial t},v\rangle dt+\int_{Q_T}\Big(\frac{\sigma^2}{2}\frac{\partial m}{\partial x}\frac{\partial v}{\partial x}+\alpha m\frac{\partial v}{\partial x}\Big)dxdt=0,\quad\forall v\in H^1(0,1),\tag{2.5a}$$

$$m(x,0)=m_0(x).\tag{2.5b}$$

Remark 2.1. Obviously, for $\alpha\in L^\infty(Q_T)$ we loose the connection of problem (2.5) with initial-boundary value problem (2.1) because the boundary conditions (2.3) cannot be defined for such functions α. Let α belongs to the space $V_\alpha=L^\infty(Q_T)\cap L^2(0,T;H_0^1(0,1))$ with the norm $\|\alpha\|_{V_\alpha}=\|\alpha\|_{L^\infty(Q_T)}+\|\alpha\|_{L^2(0,T;H_0^1(0,1))}$. Then it satisfies (2.3) in a weak sense and (2.5) is just a weak formulation of (2.1).

Problem (2.5) has a unique solution which satisfies the following stability estimate (cf., e.g., [12,13]):

$$\|m\|_{W(0,T)}\le M\|m_0\|_{L^2(0,1)}\tag{2.6}$$

with constant M depending on $\|\alpha\|_{L^\infty(Q_T)}$ and T.

We define the cost functional

$$J(m,\alpha)=\int_0^T\int_0^1 \mathrm{e}^{-rt}m\Big(f(m)+\frac{\alpha^2}{2}\Big)dxdt,\qquad r=\mathrm{const}\ge 0,\tag{2.7}$$

where the function $f(m)=f(x,t,m)$ satisfies the following assumptions:

$$\begin{aligned}&f(m)=f(x,t,m)\ \text{ is continuous, }\ |f(m)|\le d_1+d_2m,\quad (d_1,d_2\ge 0),\\&\quad\text{for all }\ (x,t)\in\bar{Q}_T\quad\text{and}\quad m\ge 0.\end{aligned}\tag{2.8}$$

Under assumptions (2.8) the functional $J(m,\alpha)$ is well-defined for functions $\alpha\in L^\infty(Q_T)$ and $m\in L^2(Q_T)$ such that $m(x,t)\ge 0$ a.e. in Q_T.

Theorem 2.1. *Let C_α be a positive constant and*

$$\begin{aligned}K=\{&(m,\alpha)\in W(0,T)\times L^\infty(Q_T):(m,\alpha)\ \textit{satisfies}\ (2.5);\\&m(x,t)\ge 0\ \textit{a.e. in}\ Q_T;\ |\alpha(x,t)|\le C_\alpha\ \textit{a.e. in}\ Q_T\}.\end{aligned}$$

For any $m_0 \in L^2(\Omega), m_0 \geq 0$, there exists a solution to the minimization problem

$$\min_{(m,\alpha)\in K} J(m,\alpha). \tag{2.9}$$

Proof. For $\alpha \equiv 0$ and $m_0 \geq 0$ state equation (2.5) becomes heat equation for which the solution $m(0)$ is known to be non-negative. It means that $(m(0),0) \in K$, i.e., $K \neq \emptyset$. Due to estimate (2.6) the set K is bounded in $W(0,T)\times L^\infty(Q_T)$. Assumptions (2.8) ensure boundedness of functional $J(m,\alpha)$ on K.

Let $\{(m_n,\alpha_n)\} \in K$ be a minimizing sequence. Let us denote by $B_\alpha = \{\alpha \in L^\infty(Q_T) : \|\alpha\|_{L^\infty(Q_T)} \leq C_\alpha\}$ the closed ball. Since $\alpha_n \in B_\alpha$, estimate (2.6) guarantees that the sequence $\{m_n\}$ is bounded in $W(0,T)$. The space $W(0,T)$ is compactly embedded in $L^2(Q_T)$ (see [14]). Thus, we can extract a subsequence (we keep for it the previous notation $\{(m_n,\alpha_n)\}$), such that

$$\begin{aligned} &\alpha_n \to \alpha^* \in L^\infty(Q_T) \ \text{ weakly in } \ L^2(Q_T) \ \text{ and a.e. in } \ Q_T; \\ &m_n \to m^* \in W(0,T) \ \text{ weakly in } \ W(0,T), \ \text{ strongly in } \ L^2(Q_T) \\ &\text{and a.e. in } \ Q_T. \end{aligned} \tag{2.10}$$

The limit functions satisfy the inequalities $|\alpha^*(x,t)| \leq C_\alpha$ a.e. in Q_T and $m^*(x,t) \geq 0$ a.e. in Q_T. Moreover, the pair (m^*,α^*) satisfies the state equation (2.5). These results yield $(m^*,\alpha^*) \in K$.

Due to (2.8) the function $mf(m)$ defines a continuous Nemytskii operator from $L^2(Q_T)$ to $L^1(Q_T)$ (cf., e.g. [15,16]), so,

$$\int_0^T \int_0^1 \mathrm{e}^{-rt} m_n f(m_n) dx dt \to \int_0^T \int_0^1 \mathrm{e}^{-rt} m^* f(m^*) dx dt. \tag{2.11}$$

Further, the sequence of non-negative functions $\mathrm{e}^{-rt} m_n \frac{\alpha_n^2}{2} \in L^1(Q_T)$ is bounded and converges a.e. to $\mathrm{e}^{-rt} m^* \frac{(\alpha^*)^2}{2} \in L^1(Q_T)$. By Fatou's lemma (cf., e.g., [17])

$$\liminf_{n\to\infty} \int_{Q_T} \mathrm{e}^{-rt} m_n \frac{\alpha_n^2}{2} dx dt \geq \int_{Q_T} \mathrm{e}^{-rt} m^* \frac{(\alpha^*)^2}{2} dx dt. \tag{2.12}$$

Due to (2.11) and (2.12), the inequality

$$\liminf_{n\to\infty} J(m_n,\alpha_n) \geq J(m^*,\alpha^*)$$

holds, that is (m^*,α^*) is a minimal point of the cost functional. □

Remark 2.2. As mentioned in Remark 2.1, it is reasonable to take $\alpha \in V_\alpha$. The existence result of Theorem 2.1 remains valid in the case of the set

$$\begin{aligned} K = \{(m,\alpha) \in W(0,T)\times V_\alpha : &(m,\alpha) \ \text{ satisfies (2.5);} \\ &m(x,t) \geq 0 \ \text{ a.e. in } \ Q_T; \ \|\alpha\|_{V_\alpha} \leq C_\alpha\}. \end{aligned}$$

3 Approximation

Let $\omega_x=\{x_i=ih,\ i=0,\cdots,N,\ Nh=1\}$ and $\omega_t=\{t_j=j\Delta t,\ j=0,\cdots,M,\ M\Delta t=T\}$ be uniform meshes in space and in time, $\omega_t^0=\omega_t\setminus\{t=0\}$. By V_x we denote the space of mesh functions (vectors) defined on ω_x, and by u_i the value of a mesh function $u\in V_x$ at the point $x_i\in\omega_x$. Let V_x^0 be the subspace of V_x of functions that satisfy the homogeneous Dirichlet condition

$$\alpha\in V_x^0\Leftrightarrow\alpha\in V_x\quad\text{and}\quad\alpha_0=\alpha_N=0. \tag{3.1}$$

If $u(t):\omega_t\to V_x$, then $u^j=u(t_j)\in V_x$ and u_i^j is the value of mesh function $u(x,t)$ at the point $(x_i,t_j)\in\omega_x\times\omega_t$. For the mesh functions $u,v\in V_x$ we use the following notations for combined quadrature formulas approximating integral over $[0,1]$:

$$(u,v]=\sum_{i=1}^{N}hu_iv_i,\quad[u,v)=\sum_{i=0}^{N-1}hu_iv_i,\quad(u,v)=\sum_{i=1}^{N-1}hu_iv_i,$$
$$[u,v]=\frac{1}{2}(u,v]+\frac{1}{2}[u,v).$$

The difference quotients are denoted by $\partial v_i=h^{-1}(v_{i+1}-v_i)$ and $\bar\partial v_i=h^{-1}(v_i-v_{i-1})$. The notations $\alpha^+=0.5(|\alpha|+\alpha)$ and $\alpha^-=0.5(|\alpha|-\alpha)$ are used for positive and negative parts of α, and $m\gg0$ means that all components of a vector m are non-negative.

3.1 Mesh optimal control problem with implicit approximation of state equation

For a given function $\alpha(t):\omega_t^0\to V_x^0$ we find function $m(t):\omega_t\to V_x$, which satisfies the following summation equality approximating state equation (2.5):

$$\begin{cases}\left[\dfrac{m^j-m^{j-1}}{\Delta t},v\right]+\dfrac{\sigma^2}{2}[\partial m^j,\partial v)+((\alpha^j)^+m^j,\bar\partial v)\\ \qquad-((\alpha^j)^-m^j,\partial v)=0,\quad\forall v\in V_x,\quad\forall j=1,\cdots,M,\\ m^0=m_0.\end{cases} \tag{3.2}$$

Summation equality (3.2) is an implicit form of the following system of linear algebraic equations (the backward Euler finite difference scheme):

$$\begin{cases}\rho\dfrac{m^j-m^{j-1}}{\Delta t}+A_0m^j+A_1(\alpha^j)m^j=0,\quad j=1,\cdots,M,\\ m^0=m_0.\end{cases} \tag{3.3}$$

Above $\rho\in V_x$ has components $\rho_i=\{1\text{ for }1\le i\le N-1;1/2\text{ for }i=0\text{ and }i=N\}$ and the matrices A_0, $A_1(\alpha)$ are defined by the following equalities for the mesh functions $m\in V_x$,

$\alpha \in V_x^0$:

$$(A_0 m)_i = \frac{\sigma^2}{2h^2} \begin{cases} -m_{i+1} + 2m_i - m_{i-1}, & 1 \le i \le N-1, \\ m_0 - m_1, & i = 0, \\ m_N - m_{N-1}, & i = N, \end{cases} \tag{3.4a}$$

$$(A_1(\alpha) m)_i = \frac{1}{h} \begin{cases} -\alpha_{i+1}^+ m_{i+1} + |\alpha_i| m_i - \alpha_{i-1}^- m_{i-1}, & 1 \le i \le N-1, \\ -\alpha_1^+ m_1, & i = 0, \\ -\alpha_{N-1}^- m_{N-1}, & i = N. \end{cases} \tag{3.4b}$$

Note that $A_0 m$ approximates the diffusive part $-\frac{\sigma^2}{2}\frac{\partial^2 m}{\partial x^2}$ of the elliptic operator in state equation while $A_1(\alpha)m$ approximates the advective part presented in the form $-\frac{\partial(\alpha m)}{\partial x} = -\frac{\partial(\alpha^+ m)}{\partial x} + \frac{\partial(\alpha^- m)}{\partial x}$.

Lemma 3.1. *Problem* (3.2) *has a unique solution* $m = m(\alpha, t) : \omega_t \to V_x$ *for any* m_0 *and any* $\alpha(t) : \omega_t^0 \to V_x^0$. *This solution has the following properties:*

1. *Mass balance property:*
$$[m^j, 1] = [m_0, 1], \quad \forall j = 1, \cdots, M. \tag{3.5}$$

2. *Strict positivity of a solution:*
$$\text{if } m_0 \gg 0 \quad \text{and} \quad [m_0, 1] > 0, \quad \text{then } m_i^j > 0, \quad \forall i, \quad \forall j \ge 1. \tag{3.6}$$

3. *Stability: If* $m_0 \gg 0$ *then mesh scheme* (3.2) *is stable uniformly with respect to* $\alpha(t)$ *in the mesh* $C([0,T]; L^1)$*-norm:*
$$\max_j [|m^j|, 1] (= \max_j [m^j, 1]) = [m_0, 1]. \tag{3.7}$$

Proof. Unknown vector m^j is a solution of the system of linear equations

$$A m^j = \frac{\rho}{\Delta t} m^{j-1} \text{ with matrix } A = \frac{\rho}{\Delta t} I + A_0 + A_1(\alpha^j), \text{ and identity matrix } I.$$

We investigate the properties of the transpose matrix

$$A^T = \frac{\rho}{\Delta t} I + A_0 + A_1^T(\alpha)$$

for any vector α. Since

$$(A_1^T(\alpha) v)_i = \frac{1}{h} \begin{cases} -\alpha_i^+ (v_{i+1} - v_i) + \alpha_i^- (v_i - v_{i-1}), & 1 \le i \le N-1, \\ 0, & i = 0, \ i = N, \end{cases}$$

it is easy to verify that the transpose matrix A^T is strictly diagonally dominant, has positive diagonal and non-positive off-diagonal entries. Because of this, the matrix A^T is

regular M-matrix, and hence the matrix A is also M-matrix. So, there exists a unique solution m^j.

Choosing $v\equiv 1$ in (3.2), we get the equality (3.5):

$$[m^j,1]=[m^{j-1},1],\quad \forall j\Rightarrow [m^j,1]=[m_0,1],\quad \forall j.$$

Since A is a M-matrix, then nonnegativity of m^{j-1} guarantees nonnegativity of m^j. Let us prove property (3.6), i.e., strict positivity of a solution in the case of $m_0\gg 0$ and $[m_0,1]>0$. By definition matrix A^T is tridiagonal, has strictly positive main diagonal, strictly negative over-diagonal and under-diagonal elements. Due to these properties A^T is irreducible M-matrix and all entries of A^{-T} are strictly positive (cf., e.g., [18, 19]). As a consequence, all entries of A^{-1} are also strictly positive. If vector $m_0\gg 0$ and $[m_0,1]>0$ then there exists at least one positive component of m_0, whence all components of $m^1=A^{-1}m_0$ are positive. By recurrence, m^j has positive components for other time levels $j=2,\cdots,M$.

Finally, stability property (3.7) is a simple consequence of (3.5) and (3.6). □

Now we define mesh cost function approximating (2.7):

$$J_h(m,\alpha)=\sum_{j=1}^{M}\Delta t\mathrm{e}^{-rt^j}\left[m^j,f^j(m^j)+\frac{(\alpha^j)^2}{2}\right]$$

and a set of admissible pairs (m,α):

$$K_{impl}=\{(m,\alpha):m(t):\omega_t\to V_x,\alpha(t):\ \omega_t^0\to V_x^0,\ (m,\alpha)\text{ satisfy (3.2)}\}.$$

Theorem 3.1. *Let $m_0\gg 0$. Then mesh optimal control problem*

$$\textit{find}\quad \min_{(m,\alpha)\in K_{impl}} J_h(m,\alpha) \tag{3.8}$$

has at least one solution (m,α).

Proof. Obviously, $m_0\equiv 0$ implies $m\equiv 0$ for any α and problem (3.8) becomes empty. Because of this we assume that $[m_0,1]>0$.

Let us decompose the cost function into the sum $J_h(m,\alpha)=J_{1h}(m)+J_{2h}(m,\alpha)$, where

$$J_{1h}(m)=\sum_{j=1}^{M}\Delta t\ \mathrm{e}^{-rt^j}[m^j,f^j(m^j)],\qquad J_{2h}(m,\alpha)=\frac{1}{2}\sum_{j=1}^{M}\Delta t\ \mathrm{e}^{-rt^j}[m^j,(\alpha^j)^2].$$

Let a pair (m,α) satisfy (3.2), then due to (3.6) and (3.7) vector m belongs to the bounded set $K_0=\{0\le m_i^j\le m_{\max}\}$, where $m_{\max}=h^{-1}[m_0,1]$ does not depend on α. Since the function $f(m)$ is continuous then there exists a constant C_0 such that

$$|J_{1h}(m)|\le C_0\quad \text{for all}\ \ m\in K_0.$$

Together with the inequality $J_{2h}(m,\alpha)\geq 0$ this ensures that function $J_h(m,\alpha)$ is bounded below. Let $m(0)$ be the solution of (3.2) with $\alpha=0$, then $J_h(m(0),0)\leq C_0$. Thus, function J_h has a finite infimum.

Define vector q^j with coordinates $q_i^j=\alpha_i^j m_i^j$. Due to (3.6) $m_i^j>0$ for all i and all $j\geq 1$, so, we can write

$$J_{2h}(m,\alpha)=J_{2h}(m,q)=\frac{1}{2}\sum_{j=1}^{M}\Delta t\mathrm{e}^{-rt^j}[(m^j)^{-1},(q^j)^2].$$

Let $\{(m^{(k)},\alpha^{(k)})\}$ be a minimizing sequence and $q^{(k)}=\alpha^{(k)}m^{(k)}$. Since $\{m^{(k)}\}$ is bounded and $0\leq J_{2h}(m^{(k)},\alpha^{(k)})\leq 2C_0$, then the sequence $\{q^{(k)}\}$ is also bounded, and there exists a subsequence of $\{(m^{(k)},q^{(k)}\}$ (we keep the same notation for it), which converges to (m,q). The set K_0 is closed, so, $m\in K_0$. Passing to the limit in equation (3.2) written for $\{(m^{(k)},q^{(k)}\}$ we prove that (m,q) satisfies the equation

$$\rho\frac{m^j-m^{j-1}}{\Delta t}+A_0m^j+Bq^j=0,\qquad j=1,\cdots M,$$

where

$$(Bq)_i=\frac{1}{h}\begin{cases}-q_{i+1}^{+}+|q_i|-q_{i-1}^{-}, & 1\leq i\leq N-1,\\ -q_1^{+}, & i=0,\\ -q_{N-1}^{-}, & i=N.\end{cases}$$

Now we define $\alpha_i^j=\frac{q_i^j}{m_i^j}$ if $m_i^j>0$ and α_i^j is arbitrary if $m_i^j=0$. Then the pair (m,α) satisfies the equation

$$\rho\frac{m^j-m^{j-1}}{\Delta t}+A_0m^j+A_1(\alpha^j)m^j=0,\qquad j\geq 1.$$

However, due to Lemma 3.1 the unique solution m of this equation has strictly positive coordinates for any α. It means that vector α is uniquely defined by the equalities $\alpha_i^j=\frac{q_i^j}{m_i^j}$ for all i and j, and the pair (m,α) is a minimizer of the function J_h. □

Remark 3.1. It is worth to note that the result of Theorem 3.1 is valid without any additional constraint for the mesh functions m and α although the function $J_h(m,\alpha)$ is not coercive and the set K_{impl} is not bounded. The properties (3.5) and (3.6) of the state function m play the crucial role in proving the existence result.

3.2 Mesh optimal control problem with semi-implicit and explicit approximations of state equation

Let the matrices A_0, A_1 be defined by (3.4a) and (3.4b). The semi-implicit and explicit (forward Euler) finite difference schemes are defined by the following systems of the

equations:

$$\begin{cases} \rho\dfrac{m^j-m^{j-1}}{\Delta t}+A_0m^j+A_1(\alpha^{j-1})m^{j-1}=0, \quad j=1,\cdots M, \\ m^0=m_0, \end{cases} \tag{3.9a}$$

$$\begin{cases} \rho\dfrac{m^j-m^{j-1}}{\Delta t}+A_0m^{j-1}+A_1(\alpha^{j-1})m^{j-1}=0, \quad j=1,\cdots,M, \\ m^0=m_0. \end{cases} \tag{3.9b}$$

Remark 3.2. We use α^{j-1} in the equation on the j-th time level only to emphasize that in approximation we take $\alpha(x,t)$ on the $(j-1)$-th time level.

Lemma 3.2. *Finite difference schemes* (3.9a) *and* (3.9b) *have unique solutions* $m=m(\alpha)$ *which satisfy the properties* (3.5)-(3.7). *under the following additional constraints:*

$$\max_{i,j}|\alpha_i^j|<\frac{h}{2\Delta t} \quad \textit{for problem} \ \ (3.9a), \tag{3.10a}$$

$$\max_{i,j}|\alpha_i^j|<\frac{h}{2\Delta t}-\frac{\sigma^2}{2h} \quad \textit{for problem} \ \ (3.9b). \tag{3.10b}$$

Proof. First, we consider semi-implicit finite difference scheme (3.9a). For $j\geq 1$ vector m^j is a solution of the system of linear algebraic equations

$$\Big(\frac{\rho}{\Delta t}I+A_0\Big)m^j=\Big(\frac{\rho}{\Delta t}I-A_1(\alpha^{j-1})\Big)m^{j-1}.$$

Matrix $\frac{\rho}{\Delta t}I+A_0$ is irreducible, strictly diagonally dominant M-matrix, because of this there exists inverse matrix $(\frac{\rho}{\Delta t}I+A_0)^{-1}$ with strictly positive entries. In turn, matrix

$$B=\frac{\rho}{\Delta t}I-A_1(\alpha^{j-1})$$

in the right-hand side of the system has non-negative off-diagonal entries and its diagonal entries $\frac{\rho_i}{\Delta t}-\frac{|\alpha_i^{j-1}|}{h}$ are strictly positive if condition (3.10a) is satisfied. Let $m^{j-1}\geq 0$ and $[m^{j-1},1]>0$, then vector Bm^{j-1} has non-negative coordinates and at least one positive component. Thus,

$$m^j=\Big(\frac{\rho}{\Delta t}I+A_0\Big)^{-1}Bm^{j-1}$$

has strictly positive coordinates and property (3.6) is proved.

Choosing $v\equiv 1$ in the summation equality, we get the equality $[m^j,1]=[m^{j-1},1]$ for all j. Stability property (3.7) is simple consequence of (3.5) and (3.6).

Let now m is a solution of explicit finite difference scheme (3.9b). For $j\geq 1$ vector m^j satisfies the equation

$$\frac{\rho}{\Delta t}m^j=\Big(\frac{\rho}{\Delta t}I-A_0-A_1(\alpha^{j-1})\Big)m^{j-1}.$$

Matrix $\frac{\rho}{\Delta t}I-A_0-A_1(\alpha^{j-1})$ has non-negative off-diagonal entries. Its diagonal entries $\frac{\rho_i}{\Delta t}-\frac{|\alpha_i^{j-1}|}{h}-\rho_i\frac{\sigma^2}{h^2}$ are strictly positive if (3.10b) holds. Using these properties, we continue, as described above. □

In the case of semi-implicit and explicit approximations of the state equation the cost function is defined by the equality:

$$\tilde{J}_h(m,\alpha)=\sum_{j=1}^{M}\Delta t\mathrm{e}^{-rt^j}\Big[m^j,f^j(m^j)+\frac{(\alpha^{j-1})^2}{2}\Big].$$

The corresponding mesh optimal control problems are:

$$\begin{cases}\text{find}\quad \min\limits_{(m,\alpha)\in K_{semi}}\tilde{J}_h(m,\alpha),\\ K_{semi}=\{(m,\alpha)\ \text{ satisfy state equation (3.9a) and } \alpha \text{ satisfies (3.10a)}\},\end{cases}\tag{3.11a}$$

$$\begin{cases}\text{find}\quad \min\limits_{(m,\alpha)\in K_{expl}}\tilde{J}_h(m,\alpha),\\ K_{expl}=\{(m,\alpha)\,\text{satisfy state equation (3.9b) and } \alpha \text{ satisfies (3.10b)}\}.\end{cases}\tag{3.11b}$$

Theorem 3.2. *There exist solutions to problems* (3.11a) *and* (3.11b).

Proof. Due to the properties (3.5) and (3.6) of the solutions of state problems (3.9a) and (3.9b), we can proceed as in the proof of Theorem 3.1. □

Remark 3.3. The existence of solutions of minimization problems governed by semi-implicit and explicit schemes can be proved by using Weierstrass theorem on the minimization of continuous functions on compact sets. In fact, due to constraints (3.10a) and (3.10b) we can consider α from a compact set

$$A=\Big\{\alpha:\max_{i,j}|\alpha_i^j|\le C=C(h,\Delta t)\Big\}.$$

The solutions $m(\alpha)$ of (3.9a) and (3.9b) depend continuously on α and the function $\tilde{J}_h(m,\alpha)$ is continuous with respect to m and α. As a consequence, function $I(\alpha)\equiv\tilde{J}_h(m(\alpha),\alpha)$ is continuous on A, whence the result.

4 Numerical results

We have carried out numerical experiments for the state equation (2.1) with $\sigma^2/2=0.07$, $T=1$ and the cost function (2.7) with $f(x,t,m)=\frac{x}{0.1+m}+S(t)(1-0.8x)$, a continuous function $S(t)$ defined below and $r=0$. It is easy to see that $f(x,t,m)$ satisfies assumptions (2.8) with $d_2=0$. Note that f is correctly defined for negative $m>-0.1$.

4.1 Problems with unique solutions

Calculations were performed with function $S(t)\equiv 10$ and different initial values m_0. We present the calculated results for m_0 which corresponds to Gaussian distribution centered in $x=0.5$, "triangle" function and Dirac-type mesh function m_0 (Fig. 1).

For our calculation, finite difference schemes were used on the series of meshes using implicit, semi-implicit and explicit approximations of the state equation. The implementation of the constructed non-linear optimization problems was executed by quasi-Newton method with line search procedure for finding an optimal iterative parameter. We used different initial guesses α^0. The stopping criteria was a small value of max-norm of the gradient of the cost function. We also controlled the speed of the cost function decay and number of iterations.

Below we present the results of calculations on the mesh with $\Delta x=1/128$, $\Delta t=1/128$ when using the implicit scheme for the state equation. We observed the same optimal solutions (α,m) when using semi-implicit or explicit approximations for the state equation with corresponding constraints (3.10a) or (3.10b), respectively, which ensure the positivity of m. The iterative process for implicit scheme converges monotonically and by the ninth iteration we obtain reasonable accuracy (Table 1).

The conclusions of the calculations:

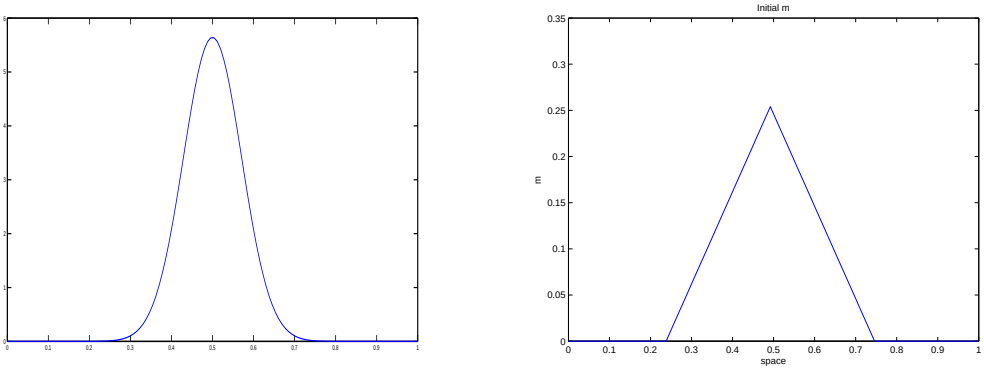

Figure 1: Gaussian and "triangle" initial value m_0.

Table 1: Behavior of the iterative method for implicit scheme with $(\Delta x=1/128,\ \Delta t=1/128)$.

Iteration number	Value of cost function	Max-norm of gradient
0	6.42581	0.00215
1	5.2663	0.0012
2	4.15197	0.00058
...	...	...
8	3.92331	0.000109
9	3.92076	8.41×10^{-5}

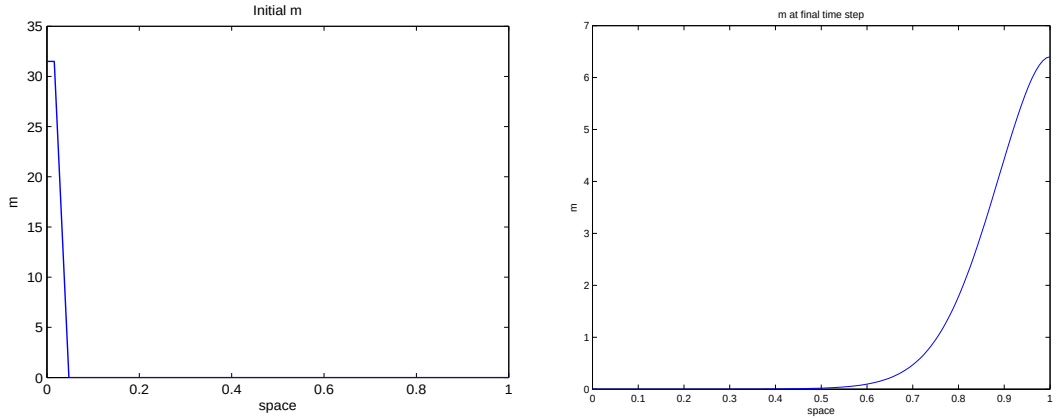

Figure 2: "Dirac" initial value m_0 (left) and the shape of final value $m(x,T)$ for Gaussian initial value m_0.

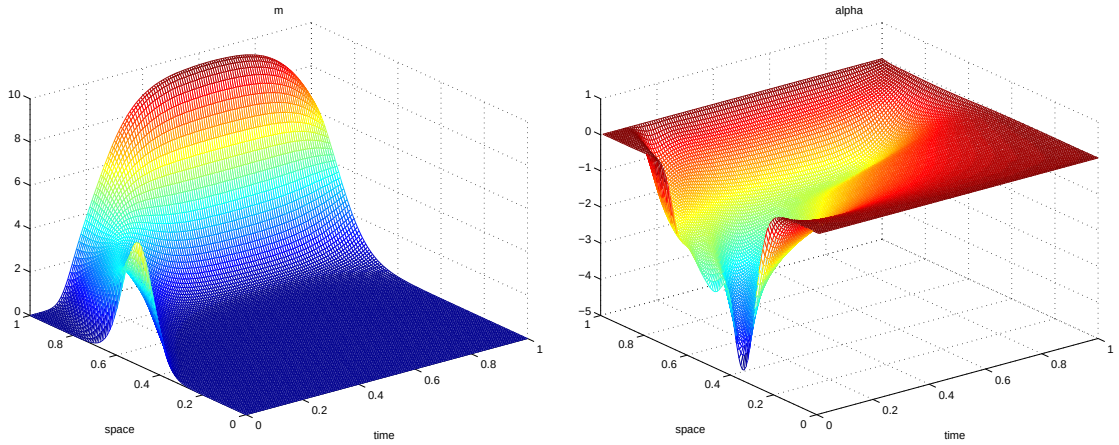

Figure 3: Behavior of m and α for Gaussian initial value m_0.

1. The number of iterations to achieve the required accuracy does not significantly depend on the initial guess α^0, and the choice $\alpha^0 \equiv 0$ was very reasonable.

2. For all fixed m_0, h and Δt, the calculated optimal solutions (m,α) were unique for the input data considered: the difference between solutions calculated from different initial guesses in the iteration method was negligibly small compared with the values of these solutions.

4.2 Semi-implicit scheme with violated constraint to mesh steps

First, we use semi-implicit scheme on the mesh with $(\Delta x = 1/32, \Delta t = 1/64)$ for the same problem as in Section 4.1. We take initial guess for the iterations to be $\alpha^0 = 0$.

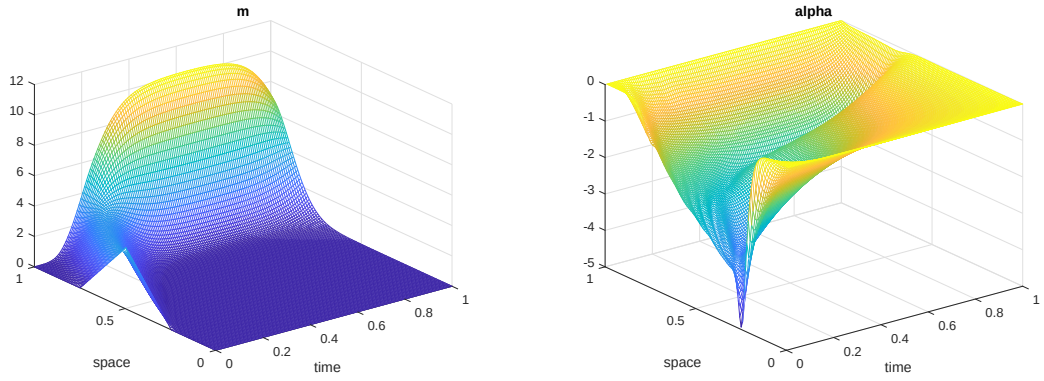

Figure 4: Behavior of m and α for "triangle" initial value m_0.

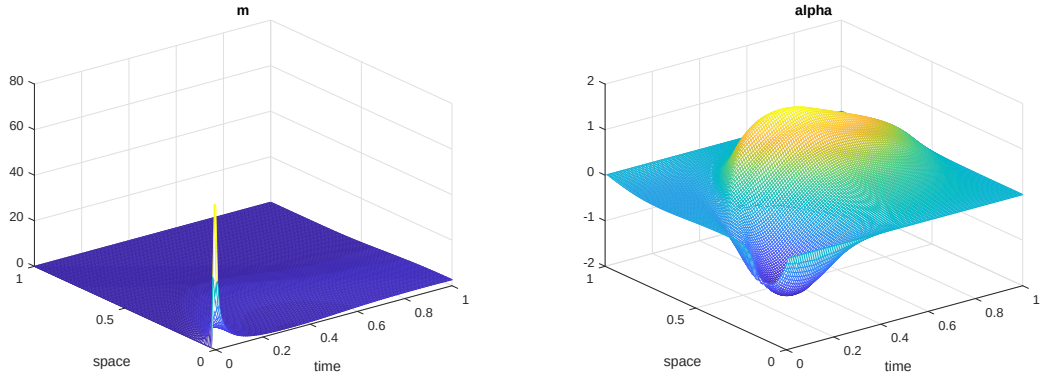

Figure 5: Behavior of m and α for "Dirac" initial value m_0.

The optimal solution is calculated above and we know that the maximal value of optimal α is about 5. So, the sufficient condition of the non-negativity of m for semi-implicit approximation

$$\max_{i,j}|\alpha_i^j| \leq \frac{h}{2\Delta t} = 1$$

is violated. Table 2 shows that the method "falls apart" on fourth iteration. The following result was obtained in the calculations: already at 2-nd iteration m becomes negative in several mesh points; at the 3-rd iteration the set of the mesh points, where $m<0$ enlarges and modules of their values increase; by the 4-th iteration the cost function approaches $-\infty$ (recall that the function $f(m)$ is defined for negative $m>-0.1$ and tends to $-\infty$ when $m\downarrow -0.1$).

Table 2: Divergence of the iterative method for semi-implicit scheme with $(\Delta x=1/32,\ \Delta t=1/64)$.

Iteration number	Value of cost function	Max-norm of gradient
0	6.42318	0.0187
1	5.33449	0.0118
2	4.15197	0.0292
3	4.13176	0.371
4	-1.5017×10^{67}	7.2×10^{67}

4.3 Non-uniqueness of solution

For a given pair of $S(t)$ and m_0 in the Fig. 6 we found two local minima. They have been reached by starting in the iterative method from different initial guesses α^0.

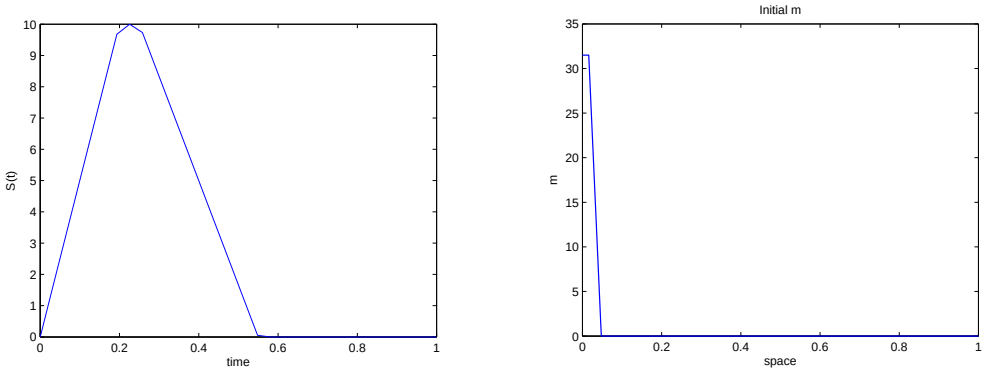

Figure 6: $S(t)$ and m_0 in the problem with non-unique solution.

5 Concluding remarks on the approximations of the problem

The fully implicit scheme (3.2) approximating state equation has the obvious advantages in comparison with semi-implicit and explicit approximations, namely, it is unconditionally stable and keep the positivity of m from the initial condition m_0 without any constraint for mesh steps or/and control function α.

As opposed to implicit scheme, semi-implicit (3.9a) and explicit (3.9b) schemes have to satisfy some constraints connecting control function α and mesh steps h, Δt to be stable and to keep the positivity of m. These constraints can be treated as limits for the mesh step Δt: $\Delta t\le\frac{h}{2\max\limits_{i,j}|\alpha_i^j|}$ for semi-implicit scheme and $\Delta t\le\frac{h^2}{\sigma^2+2h\max\limits_{i,j}|\alpha_i^j|}$ for explicit scheme. Since we do not know a priori estimate for α, we must control the positivity of m and use adaptive mesh refinement. As a consequence, we may need to recalculate the problem from

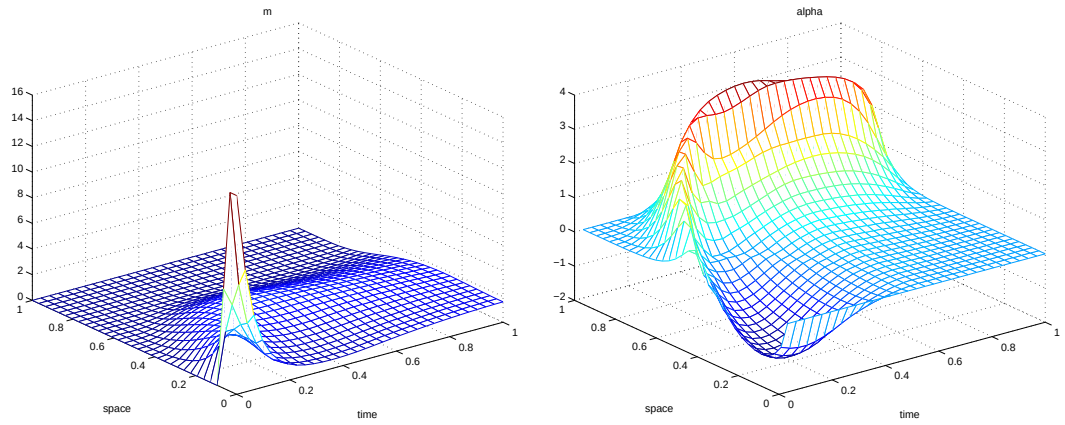

Figure 7: Behavior of m and α when initial guess in the iterative method is $\alpha=0$.

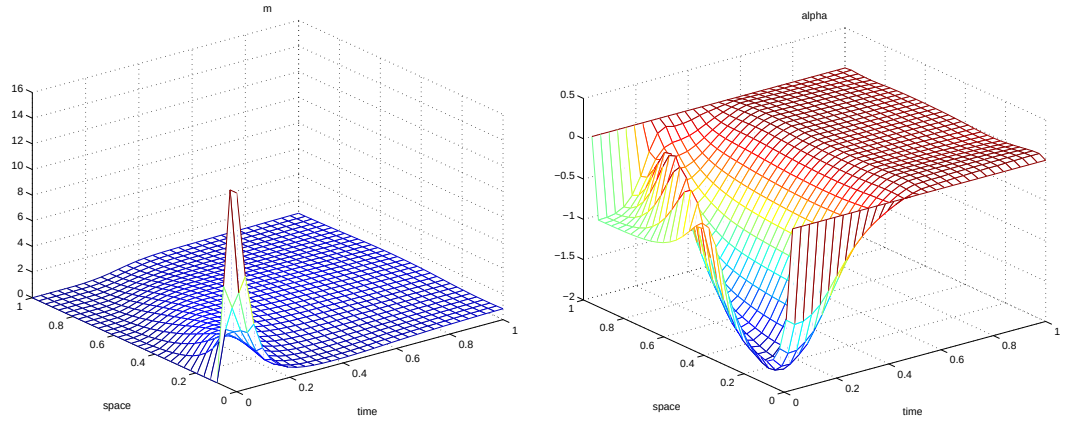

Figure 8: Behavior of m and α when initial guess in the iterative method is $\alpha=-1$ in the internal points of ω_x.

the very beginning after each refinement of the mesh, which leads to a time-consuming algorithm.

To construct well-posed discrete optimal control problem we can add the state constraint $m\geq 0$ to its formulation. But it is well known that implementing a state constrained optimal control problem is much more difficult task than implementing an unconstrained problem.

Acknowledgements

Shuhua Zhang was supported by the National Basic Research Program (No. 2012CB955804), the Major Research Plan of the National Natural Science Foundation of China (No. 91430108), the Natural Science Foundation of China (No. 11771322),

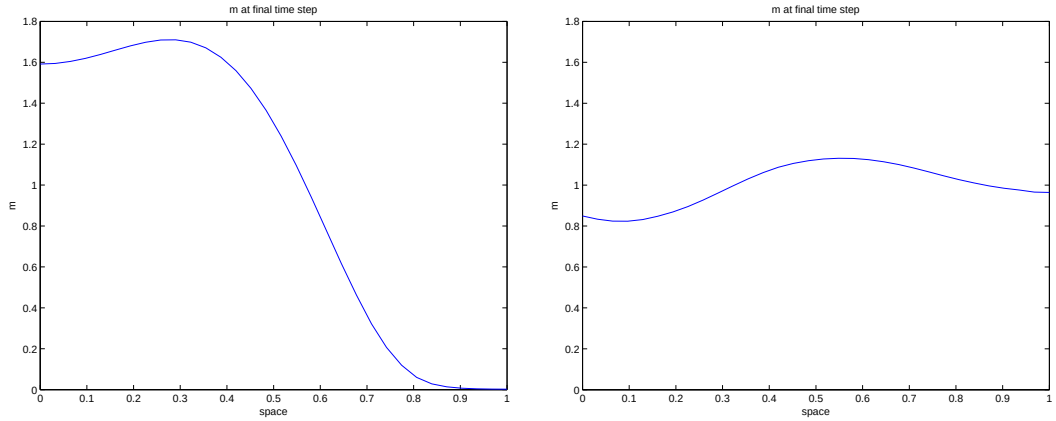

Figure 9: Final state of m when initial guess in the iterative method is $\alpha=0$ (left) and $\alpha=-1$ in the internal points of ω_x.

and the Major Program of Tianjin University of Finance and Economics (No. ZD1302). Alexander Lapin was supported by Russian Foundation of Basic Researches (No. 16-01-00408) and by program "1000 Talents" of China.

References

[1] J.-M. LASRY AND P.-L. LIONS, *Mean field games. I. The stationary case,* C. R. Math. Acad. Sci. Paris, 343 (2006), pp. 619–625.

[2] J.-M. LASRY AND P.-L. LIONS, *Mean feld games. II. Finite horizon and optimal control,* C. R. Math. Acad. Sci. Paris, 343 (2006), pp. 679–684.

[3] Y. ACHDOU AND I. CAPUZZO-DOLCETTA, *Mean field games: numerical methods,* SIAM J. Numer. Anal., 48(3) (2010), pp. 1136–1162.

[4] Y. ACHDOU, F. CAMILLI AND I. CAPUZZO-DOLCETTA, *Mean field games: convergence of a finite difference method,* SIAM J. Numer. Anal., 51(5) (2013), pp. 2585–2612.

[5] Y. ACHDOU, *Hamilton-Jacobi equations: approximations, numerical analysis and applications,* Lecture Notes in Math. In: P. Loreti, N. A. Tchou, (eds.) Finite Difference Methods for Mean Field Games, Vol. 2074, pp. 1–47, Springer, Heidelberg, 2013.

[6] O. GUÈANT, *New numerical methods for mean field games with quadratic costs,* Networks and Heterogeneous Media, 7 (2012), pp. 315–336.

[7] A. LACHAPELLE, Quelques Problèmes de Transport et de Contrôle en Èconomie: Aspects Théoriques et Numériques, Mathématiques, Université Paris Dauphine–Paris IX, 2010. Francais.

[8] SH. CHANG, X. WANG AND A. SHANANIN, *Modeling and computation of mean field equilibria in producers' game with emission permits trading communications,* Nonlinear Science and Numerical Simulation V, 37 (2016), pp. 238–248.

[9] A. LACHAPELLE, J. SALOMON AND G. TURINICI, *Computation of mean field equilibria in economics,* Math. Models Methods Appl. Sci., 20 (2010), pp. 567–588.

[10] J.-L. LIONS AND E. MAGENES, Non-Homogeneous Boundary Value Problems and Applications, Springer, 1972.
[11] J.-L. LIONS, Quelques Méthodes de Résolution des Problèmes aux Limites non Linéaires– Paris, Dunod, 1969.
[12] O. LADYZHENSKAYA, V. SOLONNIKOV AND N. URAL'CEVA, Linear and Quasilinear Equations of Parabolic Type, Transl. Math. Monographs, vol. 23, AMS, Providence, RI, 1968.
[13] A. QUARTERONI AND A. VALLI, Numerical Approximation of Partial Differential Equations, Springer, 1997.
[14] R. TEMAM, Navier-Stokes Equations, North-Holland, 1984.
[15] M. A. KRASNOSEL'SKII, Topological Methods in the Theory of Nonlinear Integral Equations (Pure and Applied Mathematics Monograph) Pergamon, 1st edition, (January 1, 1964).
[16] M. RENARDY AND R. C. ROGERS, An Introduction to Partial Differential Equations, Texts in Applied Mathematics 13 (Second ed.), New York, Springer–Verlag, 2004.
[17] N. L. CAROTHERS, Real Analysis, New York: Cambridge University Press, 2000.
[18] V. V. VOEVODIN AND Y. A. KUZNETSOV, Matrices and Calculations, Nauka, Moscow, 1984 (in Russian).
[19] A. BERMAN AND R. J. PLEMONS, *Nonnegative matrices in the mathematical sciences,* Classics in Applied Mathematics, Vol. 9, SIAM, 1994.