

Newton methods for mean field games: A numerical study

Elisabetta Carlini^a, Ahmad Zorkot^{b,*}

^a Dipartimento di Matematica Guido Castelnuovo, Sapienza Università di Roma, Rome, 00185, Italy

^b Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, Vienna, 1090, Austria

ARTICLE INFO

Keywords:

Second order mean field games
Newton iterations
Semi-Lagrangian schemes
Finite difference schemes
Numerical experiments

ABSTRACT

We address the numerical solution of second-order Mean Field Game problems through Newton iterations in infinite dimensions, introduced in [1], where quadratic convergence of the method was rigorously established. Building upon this theoretical framework, we develop new numerical discretization techniques, including both a finite difference and a semi-Lagrangian scheme, that enable an effective computational implementation of the infinite-dimensional iterations. The proposed methods are tested on several benchmark problems, and the resulting numerical experiments demonstrate their robustness, accuracy, and efficiency. A comparative analysis between the two schemes and existing approaches from the literature is also presented, highlighting the potential of Newton-based solvers for MFG systems.

1. Introduction

Mean field games (MFG) offer a powerful framework for modeling the strategic behavior of a large number of indistinguishable agents who interact through the statistical distribution of their states. Introduced independently by Lasry and Lions [2,3,48] and by Huang, Malhamé, and Caines [4,5], MFG has become a central tool for the analysis of Nash equilibria in stochastic differential games with infinitely many players. MFG models have found applications in a broad range of fields, including economics, finance, traffic flow, crowd dynamics, and the social sciences. We refer to [6–10] for general overviews of the theory and its applications.

From an analytical viewpoint, the classical formulation of MFGs introduced in [2,48] consists of a coupled system of partial differential equations (PDEs). The backward component is a Hamilton–Jacobi–Bellman (HJB) equation associated with the optimal control problem of a representative agent, complemented by a terminal condition, while the forward component is a Fokker–Planck (FP) equation governing the evolution of the population density and subject to an initial condition. Comprehensive presentations of this framework can be found in the monographs [7,9,11], the survey [10], and the lecture notes [12].

In this work, we focus on the following time-dependent, second-order Mean Field Game system with local coupling, posed on the d -dimensional torus $\mathbb{T}^d$:

$$\begin{cases} -\partial_t u - \nu \Delta u + H(x, Du) = F(m(t, x)) & \text{in } [0, T] \times \mathbb{T}^d, \\ \partial_t m - \nu \Delta m - \operatorname{div}(m D_p H(x, Du)) = 0 & \text{in } [0, T] \times \mathbb{T}^d, \\ m(0, x) = m_0(x), u(T, x) = G(x) & \text{in } \mathbb{T}^d, \end{cases} \quad (1.1)$$

where $T > 0$ denotes the time horizon, $\nu > 0$ is the diffusion coefficient, H is a convex Hamiltonian, and F is a coupling term depending locally on the density m . The choice of the torus $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ allows us to avoid the treatment of boundary conditions.

Assume that the following hypotheses hold:

* Corresponding author.

E-mail addresses: carlini@mat.uniroma1.it (E. Carlini), ahmad.zorkot@univie.ac.at (A. Zorkot).

- (i) the initial density satisfies $m_0 \in L^\infty(\mathbb{T}^d)$, $m_0 \geq 0$, and $\int_{\mathbb{T}^d} m_0(x) dx = 1$;
- (ii) the Hamiltonian $H(x, p)$ is such that $H(x, p)$ and its derivative $H_p(x, p)$ are of class C^1 , moreover, $H(x, p)$ is convex with respect to the momentum variable p ;
- (iii) the Hamiltonian satisfies a global Lipschitz condition in p , i.e. there exists a constant $\beta > 0$ such that

$$|H_p(x, p)| \leq \beta \quad \forall (x, p) \in \mathbb{T}^d \times \mathbb{R}^d;$$

- (iv) the coupling functions $F(m)$ and $G(m)$ are continuous and bounded from below.

Under these assumptions, there exists a classical solution (u, m) to the MFG system (1.1) such that (u, m) belongs to the parabolic Hölder space $C^{2+\alpha, 1+\alpha/2}(Q)$ for some $\alpha \in (0, 1)$. Furthermore, if the coupling functions $F(\cdot)$ and $G(\cdot)$ are non-decreasing with respect to the density m , the solution is unique; see [13, Theorem 1.11]. In addition, the density $m(t, \cdot)$ remains a probability measure on the unit torus $\mathbb{T}^d$ for all $t \in [0, T]$.

The numerical approximation of MFG systems has been the subject of intensive research in recent years. For second-order (stochastic) problems such as (1.1), finite difference schemes have been investigated in [14–16], while semi-Lagrangian approaches have been developed in [17, 18]. More recently, data-driven techniques, including deep learning, reinforcement learning, and tensor-based methods, have been applied to MFG models in [19–22]. In the first-order (deterministic) setting, we refer to [23–28]. Broader surveys of numerical methods for MFGs can be found in [29, 30] and the references therein.

The main focus of the present work is the application of Newton's method in infinite-dimensional function spaces to the continuous MFG system (1.1). This approach, originally proposed in [1] (see also [31] for the stationary case), reformulates the coupled PDE system as a nonlinear operator equation and constructs successive approximations by linearization around the current iterate. Each Newton step therefore requires the solution of a coupled forward–backward linear parabolic system. A key theoretical result established in [1] is the quadratic local convergence of the continuous Newton iterations under suitable regularity and proximity assumptions.

For clarity of presentation, we introduce the discretization in the case of a separable Hamiltonian. Nevertheless, the proposed methodology naturally extends to non-separable Hamiltonians, and this extension is illustrated through a dedicated numerical experiment.

To implement Newton's method for system (1.1), we introduce the operator

$$F : (u, m) \rightarrow \begin{pmatrix} -\partial_t u - v \Delta u + H(\cdot, Du) - F(m) \\ \partial_t m - v \Delta m - \operatorname{div}(m D_p H(\cdot, Du)) \\ u(T, \cdot) - G(\cdot) \\ m(0, \cdot) - m_0(\cdot) \end{pmatrix}, \quad (1.2)$$

so that system (1.1) can be written compactly as

$$F(u, m) = 0.$$

Assuming sufficient smoothness of the data, the Newton iterates (u^n, m^n) for $n \geq 1$ are defined by

$$JF(u^{n-1}, m^{n-1})(u^n - u^{n-1}, m^n - m^{n-1}) = -F(u^{n-1}, m^{n-1}), \quad (1.3)$$

where JF denotes the Jacobian of F , given explicitly by

$$JF(u, m)(v, \rho) = \begin{pmatrix} -\partial_t v - v \Delta v + D_p H(\cdot, Du) Dv & -F'(m) \rho \\ -\operatorname{div}(m D_{pp} H(\cdot, Du) Dv) & \partial_t \rho - v \Delta \rho - \operatorname{div}(D_p H(\cdot, Du) \rho) \\ v(T, \cdot) & 0 \\ 0 & \rho(0, \cdot) \end{pmatrix}, \quad (1.4)$$

for Hölder continuous functions v and ρ defined on $[0, T] \times \mathbb{R}^d$.

Setting $q^n = D_p H(\cdot, Du^{n-1})$ and substituting (1.4) into (1.3), we obtain the linearized system

$$\begin{cases} -\partial_t u^n - v \Delta u^n + q^n Du^n = q^n Du^{n-1} - H(x, Du^{n-1}) + F(m^{n-1}) \\ \quad + F'(m^{n-1})(m^n - m^{n-1}) & \text{in } [0, T] \times \mathbb{T}^d, \\ \partial_t m^n - v \Delta m^n - \operatorname{div}(m^{n-1} q^n) = \operatorname{div}(m^{n-1} D_{pp} H(x, Du^{n-1})(Du^n - Du^{n-1})) & \text{in } [0, T] \times \mathbb{T}^d, \\ m^n(x, 0) = m_0(x), \quad u^n(x, T) = G(x) & \text{in } \mathbb{T}^d. \end{cases} \quad (1.5)$$

It was shown in [1] that, under appropriate assumptions, the sequence generated by these Newton iterations converges locally to the unique solution of (1.1), with a quadratic rate provided the initial guess is sufficiently close to the exact solution. Despite these strong theoretical guarantees, the numerical behavior of the infinite-dimensional Newton method has not been extensively investigated. The aim of the present work is to address this gap by designing suitable discretization schemes and assessing their performance through numerical experiments.

To discretize and solve the linearized systems (1.5), we consider two complementary numerical strategies: a semi-Lagrangian (SL) method and an implicit finite difference scheme. For both approaches, we establish well-posedness results and evaluate their efficiency and accuracy numerically. We also compare our methodology with the Newton-based finite difference scheme introduced in [14], highlighting the distinction between the paradigms of “iterate then discretize”, where Newton's method is applied at the

continuous level, and “discretize then iterate”, where Newton’s method is applied directly to the discrete nonlinear system. Our experiments indicate that both strategies yield comparable accuracy and computational cost, while the former offers a simpler and more flexible discretization framework.

In contrast to approaches such as [14,17], which discretize the fully nonlinear MFG system (1.1) and subsequently apply Newton’s method at the discrete level, our strategy applies standard finite difference and semi-Lagrangian discretizations only to the linear systems arising at each Newton iteration. This results in a significantly simpler numerical structure. Moreover, our results show that the semi-Lagrangian scheme exhibits greater robustness in the hyperbolic regime, corresponding to small values of ν . Finite-difference-based schemes, on the other hand, may experience numerical instabilities in this setting, in agreement with known limitations discussed in [32,33], and often require continuation strategies to gradually decrease ν .

Since Newton’s method is only locally convergent, its effectiveness strongly depends on the choice of the initial guess. To enhance robustness when standard Newton iterations fail, we incorporate a globalized Newton strategy based on a line-search procedure, following [34]. This globalization technique, driven by a suitable merit function, proves particularly effective for the finite-difference-based Newton solver in the hyperbolic regime.

The remainder of the paper is organized as follows. Section 2 is devoted to the construction of a semi-Lagrangian discretization of the linearized system. In Section 3, we introduce an implicit finite difference scheme. Finally, Section 4 presents the numerical experiments and a comparative analysis of the proposed methods.

2. A semi-Lagrangian scheme

In this section, we discretize the iterative system (1.5) by means of a semi-Lagrangian scheme in the two dimensional state-space and we prove the well-posedness of the discrete system.

We refer to [17,18] for the early work on approximating the second-order MFG systems using a semi-Lagrangian scheme, and to [35] for a semi-Lagrangian scheme applied to parabolic equations.

2.1. Notations and definitions

In what follows, to simplify the discussion, we take $d = 2$ and consider the Hamiltonian

$$H(x, p) = \frac{|p|^2}{2} - V(x), \quad \text{for } (x, p) \in \mathbb{T}^2 \times \mathbb{R}^2. \quad (2.1)$$

where V is a given bounded potential.

Given two positive integers N_t and N_h , we define the time step $\Delta t = \frac{T}{N_t}$ and the space step $h = \frac{1}{N_h}$. We also introduce the index sets

$$\mathcal{I}_{\Delta t} := \{0, \dots, N_t\}, \quad \mathcal{I}_{\Delta t}^* := \mathcal{I}_{\Delta t} \setminus \{N_t\}, \quad \mathcal{I}_h := \{0, \dots, N_h - 1\}.$$

We define the discrete time grid $\mathcal{G}_{\Delta t} := \{t_k = k\Delta t \mid k \in \mathcal{I}_{\Delta t}\}$ and discrete space grid $\mathcal{G}_h := \{x_{i,j} = (ih, jh) \mid i, j \in \mathcal{I}_h\}$. We denote by $B(\mathcal{G}_h)$ and $B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)$ the two sets of functions defined in $\mathcal{G}_h$ and $\mathcal{G}_{\Delta t} \times \mathcal{G}_h$ respectively.

The objective is to construct two discrete functions $u_{[i,j]}^{n,k}$ and $m_{[i,j]}^{n,k}$ approximating the solution (u^n, m^n) of (1.5) at the grid points $(t_k, x_{i,j})$ for all $k \in \mathcal{I}_{\Delta t}$ and $i, j \in \mathcal{I}_h$. The index operator $[\cdot, \cdot]$, defined by $[i, j] = ((i + N_h) \bmod N_h, (j + N_h) \bmod N_h)$, accounts for periodic boundary conditions. For notational simplicity we will also write $(u_{i,j}^{n,k}, m_{i,j}^{n,k}) = (u_{[i,j]}^{n,k}, m_{[i,j]}^{n,k})$. Given a grid function $v \in B(\mathcal{G}_h)$, we introduce the first order central differences operators

$$\begin{aligned} (D_1 v)_{i,j} &= \frac{v_{i+1,j} - v_{i-1,j}}{2h} \quad \text{for } i, j \in \mathcal{I}_h, \\ (D_2 v)_{i,j} &= \frac{v_{i,j+1} - v_{i,j-1}}{2h} \quad \text{for } i, j \in \mathcal{I}_h, \end{aligned} \quad (2.2)$$

and define the discrete gradient operator D_h as

$$(D_h v)_{i,j} = ((D_1 v)_{i,j}, (D_2 v)_{i,j})^\top \quad \text{for } i, j \in \mathcal{I}_h. \quad (2.3)$$

Then, for a discrete vector field $p = (p_1, p_2) \in B(\mathcal{G}_h)^2$, we define the discrete divergence operator div_h as

$$(\text{div}_h(vp))_{i,j} = \frac{1}{2h} (v_{i+1,j}(p_1)_{i+1,j} - v_{i-1,j}(p_1)_{i-1,j} + v_{i,j+1}(p_2)_{i,j+1} - v_{i,j-1}(p_2)_{i,j-1}), \quad i, j \in \mathcal{I}_h. \quad (2.4)$$

Given $\phi \in B(\mathcal{G}_h)$, we define its piecewise linear interpolant as

$$I[\phi](x) = \sum_{i,j \in \mathcal{I}_h} \beta_{i,j}(x) \phi_{i,j} \quad \text{for all } x \in \mathbb{T}^2,$$

where $\phi_{i,j} = \phi(x_{i,j})$, and $\beta_{i,j}$ denote the piecewise linear finite element basis associated to the grid $\mathcal{G}_h$. For any function $\varphi : \mathbb{T}^2 \rightarrow \mathbb{R}$, let $\varphi|_{\mathcal{G}_h}$ denote its restriction to the grid $\mathcal{G}_h$. If $\varphi \in C^2(\mathbb{T}^2)$ and its second-order derivatives are bounded, then, by [36, Remark 3.4.2], the interpolation error satisfies

$$\|\varphi - I[\varphi|_{\mathcal{G}_h}]\|_\infty \leq Ch^2, \quad (2.5)$$

where $C > 0$ depends only on φ .

2.2. Semi-Lagrangian scheme for the backward equation

Building on the notations and framework introduced in the previous section, we now apply the semi-Lagrangian discretization to approximate the iterative system (1.5). We start with the backward equation, which can be written as follows:

$$\begin{cases} -\partial_t u^n - \frac{\sigma^2}{2} \Delta u^n + q^n Du^n - G^n(t, x) = 0 & \text{in } [0, T] \times \mathbb{T}^2, \\ u^n(x, T) = G(x) & \text{in } \mathbb{T}^2, \end{cases} \quad (2.6)$$

where

$$\sigma = \sqrt{2\nu}, \quad q^n(t, x) = Du^{n-1}(t, x)$$

and

$$G^n(t, x) = \frac{|q^n(t, x)|^2}{2} + V(x) + F(m^{n-1}(t, x)) + F'(m^{n-1}(t, x))(m^n(t, x) - m^{n-1}(t, x)), \quad \text{for } (t, x) \in [0, T] \times \mathbb{T}^2.$$

By the Feynman-Kac formula (see e.g. [37]), the solution u^n to (2.6), admits the following representation

$$u^n(t, x) = \mathbb{E} \left[\int_t^T G^n(s, X^{t,x}(s)) ds + G(X^{t,x}(T)) \right], \quad \text{for } (t, x) \in [0, T] \times \mathbb{T}^2, \quad (2.7)$$

where $X^{t,x}$ denotes characteristics solving

$$\begin{cases} dX(s) = q^n(s, X(s)) + \sigma dW(s) & \text{for } s \in (t, T), \\ X(t) = x. \end{cases} \quad (2.8)$$

We explain now how to construct a SL approximation using the technique shown in [38]. First, notice that (2.7) imply that for every $k \in I_{\Delta t}^*$, we have

$$u^n(t_k, x) = \mathbb{E} \left[\int_{t_k}^{t_{k+1}} G^n(s, X^{t_k,x}(s)) ds + u^n(t_{k+1}, X^{t_k,x}(\Delta t)) \right]. \quad (2.9)$$

We denote by $q_{i,j}^{n,k} \in \mathbb{R}^2$ an appropriate approximation of the drift term $q^n(t_k, x_{i,j})$ for each $k \in I_{\Delta t}$ and $i, j \in I_h$. Then, we approximate the expectation in (2.9) (see e.g. [39]) as

$$\mathbb{E} [u^n(t_k, X^{t_k,x_{i,j}}(\Delta t))] = \frac{1}{4} \sum_{\ell=1}^4 u^n(t_k, y^\ell(x_{i,j}, q_{i,j}^{n,k})) + O((\Delta t)^2), \quad (2.10)$$

where

$$y^\ell(x_{i,j}, q_{i,j}^{n,k}) = (x_{i,j} + \Delta t q_{i,j}^{n,k} + \sqrt{2\Delta t} \sigma e^\ell)_p \quad \text{for } \ell = 1, \dots, 4, \quad (2.11)$$

with e^ℓ representing four vectors of $\mathbb{R}^2$ with one component equal to ± 1 and the other null, and

$$(z)_p = (z_1 - \lfloor z_1 \rfloor, z_2 - \lfloor z_2 \rfloor) \quad \text{for all } z = (z_1, z_2) \in \mathbb{R}^2,$$

denotes the periodic projection on $\mathbb{T}^2$ of $z \in \mathbb{R}^2$.

Finally, by combining (2.7), (2.9), and (2.10), and using the rectangular rule to approximate the integral term in (2.9), we define the semi-Lagrangian scheme for Eq. (2.6) as follows. Given $q^n \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)^2$ and $m^n, m^{n-1} \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)$, find $u^n \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)$ such that

$$\begin{cases} u_{i,j}^{n,k} = S_{k,(i,j)}^n(u^{n,k+1}) & \text{for all } k \in I_{\Delta t}^*, i, j \in I_h, \\ u^{n,N_t} = G(x_{i,j}), \end{cases} \quad (2.12)$$

where, for every $f \in B(\mathcal{G}_h)$, $k \in I_{\Delta t}^*$, and $i, j \in I_h$,

$$S_{k,(i,j)}^n(f) := \frac{1}{4} \sum_{\ell=1}^4 I[f](y^\ell(x_{i,j}, q_{i,j}^{n,k})) + \Delta t G^n(t_k, x_{i,j}).$$

Let $\phi \in C_0^2(\mathbb{T}^2)$, and let us denote by C a positive real number which can depend only on ϕ . From (2.11) and Taylor expansion (see e.g. [35]) we have

$$\left| \frac{1}{4} \sum_{\ell=1}^4 \phi(y^\ell(x_{i,j}, q_{i,j}^{n,k})) - \left(\phi(x_{i,j}) + \Delta t \frac{\sigma^2}{2} \Delta \phi(x_{i,j}) + \Delta t q_{i,j}^{n,k} D \phi(x_{i,j}) \right) \right| \leq C(\Delta t)^2 \quad \text{for all } i, j \in I_h. \quad (2.13)$$

Let u^n be a smooth solution to (2.6), with bounded derivatives. Then, using (2.5) and (2.13), for every $k \in I_{\Delta t}^*$ and $i, j \in I_h$, the consistency error of scheme (2.12) satisfies

$$T_{\Delta t,h}(t_k, x_{i,j}) := \frac{1}{\Delta t} (u^n(t_{k+1}, x_{i,j}) - S_{k,(i,j)}^n(u^n(t_{k+1}))) = O\left(\Delta t + \frac{h^2}{\Delta t}\right) \quad (t, x) \in (0, T) \times \mathbb{T}^2.$$

2.3. A semi-Lagrangian scheme for the forward equation

Let us now consider the second equation in system (1.5)

$$\begin{cases} \partial_t m^n - \frac{\sigma^2}{2} \Delta m^n - \operatorname{div}(m^n q^n) = \operatorname{div}(m^{n-1}(t, x)(Du^n(t, x) - Du^{n-1}(t, x))) & \text{in } [0, T] \times \mathbb{T}^2, \\ m^n(0, x) = m_0(x) & \text{in } \mathbb{T}^2. \end{cases} \quad (2.14)$$

Following the same analogue in [17], we propose the following scheme to approximate (2.14). Given $q^n, D_h u^{n-1}, D_h u^n \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)^2$, and $m^{n-1} \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)$, find $m^n \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)$ such that

$$\begin{cases} m_{i,j}^{n,k+1} = (S_{k,(i,j)}^n)^*(m^{n,k}) & \text{for all } k \in \mathcal{I}_{\Delta t}^*, i, j \in \mathcal{I}_h, \\ m_{i,j}^{n,0} = m_0(x_{i,j}), \end{cases} \quad (2.15)$$

where, for a given function $f \in B(\mathcal{G}_h)$, $k \in \mathcal{I}_{\Delta t}^*$ and $i, j \in \mathcal{I}_h$

$$(S_{k,(i,j)}^n)^*(f) = \frac{1}{4} \sum_{\ell=1}^4 I^*[f](y^\ell(x_{i,j}, q_{i,j}^{n,k})) + \Delta t (\operatorname{div}_h(m^{n-1,k+1}(D_h u^{n,k+1} - D_h u^{n-1,k+1})))_{i,j},$$

where, for every $i, j \in \mathcal{I}_h$, $I^*[f](y^\ell(x_{i,j}, q_{i,j}^{n,k}))$ is the adjoint operator of $f \rightarrow I[f](y^\ell(x_{i,j}, q_{i,j}^{n,k}))$.

Let $\phi, \psi \in C^4(\mathbb{T}^2)$, $\varphi \in C^3(\mathbb{T}^2)$, and set $p = D\phi - D\psi$. Using the definitions of the discrete gradient and divergence (2.3)-(2.4), we have

$$\operatorname{div}_h(\varphi(D_h\phi - D_h\psi))_{i,j} = \frac{1}{2h} (\varphi_{i+1,j} p_{1,i+1,j} - \varphi_{i-1,j} p_{1,i-1,j} + \varphi_{i,j+1} p_{2,i,j+1} - \varphi_{i,j-1} p_{2,i,j-1}).$$

Defining the smooth function $w = \varphi p$, a Taylor expansion at $x_{i,j}$ yields

$$\frac{w_1(x_{i+1,j}) - w_1(x_{i-1,j})}{2h} = \partial_x w_1(x_{i,j}) + O(h^2), \quad \frac{w_2(x_{i,j+1}) - w_2(x_{i,j-1})}{2h} = \partial_y w_2(x_{i,j}) + O(h^2),$$

and hence

$$\operatorname{div}_h(\varphi p)_{i,j} = \operatorname{div}(\varphi p)(x_{i,j}) + O(h^2).$$

Therefore, under these regularity assumptions, the discrete term

$$\operatorname{div}_h(\varphi(D_h\phi - D_h\psi))$$

is a second-order consistent approximation

$$\operatorname{div}(\varphi(D\phi - D\psi)),$$

showing that the source term (2.3) is $O(h^2)$ consistent in space.

2.4. The discrete Newton system

Given $v \in B(\mathcal{G}_h)$, we denote by $V \in \mathbb{R}^{N_h^2}$ the vectors such that

$$V_{i+jN_h} = v_{i,j} \quad \forall i, j \in \mathcal{I}_h. \quad (2.16)$$

Then the discrete gradient operator $D_h : \mathbb{R}^{N_h} \rightarrow \mathbb{R}^{2N_h}$, defined in (2.3), is such that

$$D_h V = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} V = \begin{pmatrix} D_1 V \\ D_2 V \end{pmatrix}, \quad (2.17)$$

where the matrices $D_1, D_2 \in \mathbb{R}^{N_h \times N_h}$ corresponds to the operator defined in (2.2).

For a discrete vector field $p = (p_1, p_2) \in \mathbb{R}^{2N_h}$, the discrete divergence operator $\operatorname{div}_h : \mathbb{R}^{2N_h} \rightarrow \mathbb{R}^{N_h}$ corresponding to (2.4) is such that

$$\operatorname{div}_h(p) = D_1 p_1 + D_2 p_2 = \begin{bmatrix} D_1 & D_2 \end{bmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}. \quad (2.18)$$

We also denote by $\operatorname{diag}(V) \in \mathbb{R}^{N_h \times N_h}$ the diagonal matrix whose diagonal entries correspond to the components of V .

For $k \in \mathcal{I}_{\Delta t}^*$, we define the vectors $U^{n,k}, U^{N_t} \in \mathbb{R}^{N_h^2}$ by

$$U_{i+jN_h}^{n,k} = u_{i,j}^{n,k}, \quad U_{i+jN_h}^{N_t} = G(x_{i,j}) \quad \forall i, j \in \mathcal{I}_h. \quad (2.19)$$

Similarly, for $k \in \mathcal{I}_{\Delta t} \setminus \{0\}$, we set $M^{n,k}, M^0 \in \mathbb{R}^{N_h^2}$ as

$$M_{i+jN_h}^{n,k} = m_{i,j}^{n,k}, \quad M_{i+jN_h}^0 = m_0(x_{i,j}) \quad \forall i, j \in \mathcal{I}_h. \quad (2.20)$$

Finally, for $k \in I_{\Delta t}^*$, we define $Q^{n,k} = (Q^{1,n,k}, Q^{2,n,k}) \in \mathbb{R}^{2N_h^2}$ as

$$Q_{i+jN_h}^{1,n,k} = q_{i,j}^{1,n,k}, \quad Q_{i+jN_h}^{2,n,k} = q_{i,j}^{2,n,k} \quad \text{for all } i, j \in I_h. \quad (2.21)$$

With this notation, scheme (2.12) can be written in matrix form as

$$\begin{cases} U^{n,k} = \mathcal{A}(Q^{n,k}) U^{n,k+1} + \Delta t \mathcal{W}^k M^{n,k} + \Delta t B^k, & \text{for all } k \in I_{\Delta t}^*, \\ U^{n,N_t} = U^{N_t}, \end{cases} \quad (2.22)$$

where

$$(\mathcal{A}(Q^{n,k}))_{iN_h+j,p+N_hq} = \frac{1}{2d} \sum_{\ell=1}^4 \beta_{p,q}(\mathcal{Y}^\ell(x_{i,j}, q_{i,j}^{n,k})), \quad \text{for all } i, j, p, q \in I_h \quad (2.23)$$

$$(\mathcal{W}^k)_{iN_h+j,p+N_hq} = (\text{diag}(F'(M^{n-1,k})))_{iN_h+j,p+N_hq}, \quad (2.24)$$

$$\begin{aligned} (B^k)_{i+N_hj} &= \frac{|q_{i,j}^{n,k}|^2}{2} - V(x_{i,j}) \\ &+ F((M^{n-1,k})_{i+N_hj}) - F'((M^{n-1,k})_{i+N_hj})(M^{n-1,k})_{i+N_hj} \quad \text{for all } i, j \in I_h. \end{aligned} \quad (2.25)$$

Scheme (2.15) can be written in matrix form as

$$\begin{cases} M^{n,k+1} = \mathcal{A}^*(Q^{n,k}) M^{n,k} + \Delta t \mathcal{Z}^{k+1} U^{n,k+1} + \Delta t C^{k+1} & k \in I_{\Delta t}^*, \\ M^{n,0} = M^0, \end{cases} \quad (2.26)$$

where $\mathcal{A}(Q)^*$ denotes the transpose of $\mathcal{A}(Q)$ given by (2.23), for every $k \in I_{\Delta t}^*$, $\mathcal{Z}^{k+1}$ is $N_h^2 \times N_h^2$ matrix such that, given $V \in \mathbb{R}^{N_h^2}$ and let $\mathcal{M} = M^{n-1,k+1}$

$$\mathcal{Z}^{k+1} V = \text{div}_h(\text{diag}(\mathcal{M}) D_h U) = \text{diag}(\mathcal{M}) \text{div}_h(D_h V) + \text{diag}(D_1 \mathcal{M}) D_1 V + \text{diag}(D_2 \mathcal{M}) D_2 V. \quad (2.27)$$

This shows that $\mathcal{Z}^{k+1}$ can be written in matrix form as

$$\mathcal{Z}^{k+1} = \text{diag}(\mathcal{M}) L_h + \text{diag}(D_1 \mathcal{M}) D_1 + \text{diag}(D_2 \mathcal{M}) D_2. \quad (2.28)$$

where $L_h := D_1 D_1 + D_2 D_2$ is the discrete 2D Laplacian operator. Moreover C^{k+1} is the vector in $\mathbb{R}^{N_h^2}$

$$(C^{k+1})_{i+N_hj} = -(\mathcal{Z}^{k+1} U^{n-1,k+1})_{i+N_hj}. \quad (2.29)$$

Remark 2.1. [Mass preservation] The semi-Lagrangian adjoint operator $\mathcal{A}^*(Q^{n,k})$ is the transpose of a row-stochastic matrix $\mathcal{A}(Q^{n,k})$ representing linear interpolation along characteristics. Let $(i, j), (p, q) \in I_h$ and define the linear indices $\zeta = i + N_h j, l = p + N_h q$, then let A be the matrix with entries $A_{\zeta,l} := (\mathcal{A}(Q^{n,k}))_{\zeta,l}$. By definition,

$$A_{\zeta,l} \geq 0, \quad \sum_l A_{\zeta,l} = 1 \quad \forall \zeta,$$

so A is row-stochastic. Its adjoint satisfies $(A^*)_{l,\zeta} = A_{\zeta,l}$, and for any discrete density M we have

$$\sum_l (A^* M)_l = \sum_l \sum_\zeta A_{\zeta,l} M_\zeta = \sum_\zeta M_\zeta \sum_l A_{\zeta,l} = \sum_\zeta M_\zeta.$$

Furthermore, the linearization term $\mathcal{Z}^{k+1}(U^{n,k+1} - U^{n-1,k+1})$, defined via the discrete divergence div_h (2.4), sums to zero under periodic boundary conditions. Hence, if $\sum_\zeta M_\zeta^{n,0} = 1$, then for all k ,

$$\sum_\zeta M_\zeta^{n,k} = 1,$$

so the scheme is mass-preserving.

Remark 2.2. [Positivity] The linearization term $\text{div}(m^{n-1}(Du^n - Du^{n-1}))$ acts as a source and is generally not sign-definite. Although the transport part preserves positivity,

$$M_l^{n,k} \geq 0, \quad \forall l \implies (\mathcal{A}(Q^{n,k})^* M^{n,k})_l \geq 0, \quad \forall l$$

the full scheme may produce small negative values of order $O(\Delta t)$ due to the source term, so global non-negativity of the density cannot be guaranteed.

Remark 2.3. [Stability of the scheme (2.26)] If the source term in (2.26) satisfies

$$\|\mathcal{Z}^{k+1}(U^{n,k+1} - U^{n-1,k+1})\|_{\ell^1} \leq C,$$

for a positive $C > 0$, we can show that the scheme (2.26) is stable. Indeed, since adjoint operator $\mathcal{A}^*(Q^{n,k})$ preserves mass and has positive entries, we obtain

$$\|M^{n,k+1}\|_{\ell^1} \leq \|M^{n,k}\|_{\ell^1} + C \Delta t,$$

and by iterating backward in time, we get

$$\|M^{n,k+1}\|_{\ell^1} \leq \|M^{n,0}\|_{\ell^1} + CT.$$

The uniform bound of the source term is a crucial step, as it would allow one to prove convergence of the full discrete scheme. We leave this analysis for future work.

The final discrete Newton iteration system reads as follows. Given $(U^{n-1,k}, M^{n-1,k}, Q^{n,k})_{k \in I_{\Delta t}}$, find $(U^{n,k}, M^{n,k})_{k \in I_{\Delta t}}$ such that

$$\begin{cases} U^{n,k} = \mathcal{A}(Q^{n,k})U^{n,k+1} + \Delta t \mathcal{W}^k M^{n,k} + \Delta t B^k & k \in I_{\Delta t}^*, \\ M^{n,k+1} = \mathcal{A}^*(Q^{n,k})M^{n,k} + \Delta t \mathcal{Z}^{k+1} U^{n,k+1} + \Delta t C^{k+1} & k \in I_{\Delta t}^*, \\ U^{n,N_t} = U^{N_t}, \\ M^{n,0} = M^0, \end{cases} \quad (2.30)$$

where $\mathcal{A}(Q^{n,k})$, $\mathcal{W}^k$, B^k , $\mathcal{Z}^{k+1}$, and C^{k+1} are given by (2.23)–(2.25), (2.28), and (2.29) respectively.

2.5. Well-posedness

To establish the well-posedness of the system (2.30), we represent it in matrix form. For this purpose, let $\bar{U}$ and $\bar{M}$ be the vectors in $\mathbb{R}^{(N_t+1)N_h^2}$ defined by

$$(\bar{U}^n)_{kN_h^2+iN_h+j} = (U^{n,k})_{i+N_hj}, \quad (\bar{M}^n)_{kN_h^2+iN_h+j} = (M^{n,k})_{i+N_hj}, \quad k \in I_{\Delta t}, i, j \in I_h. \quad (2.31)$$

Next, define the block matrices $\mathbf{A}$ and $\mathbf{W}$ as

$$\mathbf{A} = \begin{bmatrix} \mathcal{I} & -\mathcal{A}^0 & \mathcal{O} & \cdots & \mathcal{O} \\ \mathcal{O} & \mathcal{I} & -\mathcal{A}^1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathcal{I} & -\mathcal{A}^{N_t-1} \\ \mathcal{O} & \cdots & \cdots & \mathcal{O} & \mathcal{I} \end{bmatrix}, \quad \mathbf{W} = \Delta t \begin{bmatrix} \mathcal{W}^0 & \mathcal{O} & \cdots & \cdots & \mathcal{O} \\ \mathcal{O} & \mathcal{W}^1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \mathcal{W}^{N_t-1} & \mathcal{O} \\ \mathcal{O} & \cdots & \cdots & \mathcal{O} & \mathcal{O} \end{bmatrix}.$$

Here, $\mathcal{I}$ and $\mathcal{O}$ denote the $N_h^2 \times N_h^2$ identity and zero matrices, respectively. If $F' > 0$ and $m > 0$, then the first N_t diagonal blocks of $\mathbf{W}$ are positive, provided that M^{n-1} is positive. We also define the block diagonal matrix $\mathbf{Z}$ as

$$\mathbf{Z} = \Delta t \begin{bmatrix} \mathcal{O} & \cdots & \cdots & \cdots & \mathcal{O} \\ \mathcal{O} & \mathcal{Z}^1 & \mathcal{O} & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathcal{O} \\ \mathcal{O} & \cdots & \cdots & \mathcal{O} & \mathcal{Z}^{N_t} \end{bmatrix},$$

where, for every $k \in I_{\Delta t}^*$, $\mathcal{Z}^k$ are defined in (2.28).

Finally, let

$$\mathbf{B} = \Delta t \left[B^0, \dots, B^{N_t-1}, \frac{1}{\Delta t} U^{N_t} \right]^\top, \quad \mathbf{C} = \Delta t \left[\frac{1}{\Delta t} M^0, C^0, \dots, C^{N_t-1} \right]^\top,$$

with B^k and C^k given in (2.25) and (2.29), respectively.

Then, the discrete Newton system (2.30) can be written as the following system:

$$\begin{bmatrix} \mathbf{A} & -\mathbf{W} \\ -\mathbf{Z} & \mathbf{A}^\top \end{bmatrix} \begin{bmatrix} \bar{U}^n \\ \bar{M}^n \end{bmatrix} = \begin{bmatrix} \mathbf{B} \\ \mathbf{C} \end{bmatrix}. \quad (2.32)$$

Proposition 2.1. Assume that $F' > 0$ and $\bar{M}^{n-1} > 0$ for any $n \in \mathbb{N}$. Then, there exists a unique solution $(\bar{U}^n, \bar{M}^n)$ to the system (2.32).

Proof. Suppose that $\mathbf{B} = 0$ and $\mathbf{C} = 0$, then (2.32) reads as

$$\begin{cases} \mathbf{A} \bar{U}^n - \mathbf{W} \bar{M}^n = 0, \\ -\mathbf{Z} \bar{U}^n + \mathbf{A}^\top \bar{M}^n = 0 \end{cases} \quad (2.33)$$

Multiplying the first equation by $(\bar{M}^n)^\top$ and the second one by $(\bar{U}^n)^\top$ one gets

$$\begin{cases} (\bar{M}^n)^\top \mathbf{A} \bar{U}^n - (\bar{M}^n)^\top \mathbf{W} \bar{M}^n = 0 \\ -(\bar{U}^n)^\top \mathbf{Z} \bar{U}^n + (\bar{U}^n)^\top \mathbf{A}^\top \bar{M}^n = 0. \end{cases}$$

Subtracting both equations, we obtain

$$(\bar{M}^n)^\top \mathbf{W} \bar{M}^n - (\bar{U}^n)^\top \mathbf{Z} \bar{U}^n = 0. \quad (2.34)$$

Since $F' > 0$, the block $\mathcal{W}^k$ is positive definite for all $k \in \mathcal{I}_{\Delta t}$. Moreover, the block $\mathcal{Z}^k$ is negative definite for all $k \in \mathcal{I}_{\Delta t}$, since $\bar{M}^{n-1} > 0$ and it is the sum of a negative definite matrix and a skew symmetric matrix and then, by [40, Remark 1], it is negative definite.

Hence, it follows from (2.34) that

$$(\bar{U}^n)_{kN_h^2+iN_h+j} = 0 \quad \text{for all } k \in \mathcal{I}_{\Delta t}^*, i, j \in \mathcal{I}_h, \quad (2.35)$$

and

$$(\bar{M}^n)_{kN_h^2+iN_h+j} = 0 \quad \text{for all } k \in \mathcal{I}_{\Delta t}^*, i, j \in \mathcal{I}_h. \quad (2.36)$$

Finally, replacing (2.35) and (2.36) in (2.33), we get $\bar{U}^n = 0$ and $\bar{M}^n = 0$, and then the existence of a unique solution to (2.32). $\square$

Remark 2.4. The assumption $F' > 0$ is a standard monotonicity condition in MFG with separable Hamiltonian, which guarantees uniqueness of the solution to (1.1); see [1, Remark 2.1]. In the second-order case ($\nu > 0$), the diffusion ensures that the density remains positive. At the discrete level, the condition $M^{n-1} > 0$ should be interpreted as a positivity requirement on the density at the previous Newton iterate. However, the numerical scheme does not, in general, verify this property (see Remark 2.2) and in practice this condition may fail particularly when ν is small. In Section 4, we propose a framework to address this issue.

Remark 2.5. Note that the blocks of $\mathbf{W}$ would be dense matrices if $F(m(t, x))$ is replaced by a nonlocal operator $f[m(t, \cdot)]$. In this case, we use the notation $\frac{\delta f}{\delta m}$ for the flat derivative of f (see e.g. [8]). The condition $F' > 0$ can then be replaced by assuming that $\frac{\delta f}{\delta m}$ is Lipschitz continuous and that f satisfies

$$\int_{\mathbb{T}^d} (f[m](x) - f[m'](x)) d(m - m')(x) \geq 0 \quad \text{for all } m, m'.$$

A typical example is a nonlocal coupling with smoothing effect

$$f(x, m) = \int_{\mathbb{T}^d} \Phi(z, (\rho * m)(z)) \rho(x - z) dz,$$

where $*$ denotes the usual convolution on $\mathbb{T}^d$, Φ is a smooth and nondecreasing with respect to m , and ρ is a smooth, even function with compact support. In this case, writing $\Phi = \Phi(x, \theta)$, the flat derivative of f reads

$$\frac{\delta f}{\delta m}(x, m, y) = \int_{\mathbb{T}^d} \frac{\partial \Phi}{\partial \theta}(z, (\rho * m)(z)) \rho(x - z) \rho(z - y) dz.$$

3. A finite differences scheme

In this section, we discretize the iterative system (1.5) using an implicit finite difference scheme in two dimensions and prove its well-posedness. The equation for v is discretized using upwind differences for the drift term and central differences for the second-order spatial derivatives. The equation for m is then approximated by taking the adjoint of the resulting scheme.

In order to discretize the system, we define the discrete time derivative of $v \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)$ as

$$D_t v_{i,j}^k = \frac{v_{i,j}^{k+1} - v_{i,j}^k}{\Delta t} \quad \text{for all } k \in \mathcal{I}_{\Delta t}^*, i, j \in \mathcal{I}_h, \quad (3.1)$$

Given $q^n \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)^2$ and $u^{n-1}, m^{n-1} \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)$, we discretize the first equation in (1.5) using an implicit scheme finite difference scheme, which reads for all $k \in \mathcal{I}_{\Delta t}^*$ and $i, j \in \mathcal{I}_h$ as

$$\begin{cases} -D_t u_{i,j}^{n,k} - \nu(L_h u^{n,k})_{i,j} + q_{i,j}^{n,k} (D_h u^{n,k})_{i,j} = \frac{1}{2} |(q^{n,k})_{i,j}|^2 + V(x_{i,j}) + F_{i,j}(m^{n,k+1}, m^{n-1,k+1}), \\ u_{i,j}^{n,T} = G(x_{i,j}), \end{cases} \quad (3.2)$$

where L_h is the discrete Laplace operator, already defined in Section 2.4, and for every $\zeta, \tilde{\zeta} \in B(\mathcal{G}_h)$

$$F_{i,j}(\zeta, \tilde{\zeta}) = F'(\tilde{\zeta}_{i,j})(\zeta_{i,j} - \tilde{\zeta}_{i,j}) + F(\tilde{\zeta}_{i,j}).$$

Then, we consider the following approximation of the forward equation in (1.5)

$$\begin{cases} D_t m_{i,j}^{n,k} - \nu(L_h m^{n,k+1})_{i,j} - (\text{div}_h(q^{n,k} m^{n,k+1}))_{i,j} = \text{div}_h(m^{n-1,k+1} (D_h u^{n,k} - D_h u^{n-1,k}))_{i,j}, \\ m_{i,j}^{n,0} = m_0(x_{i,j}). \end{cases} \quad (3.3)$$

Remark 3.1. Notice that, given $q \in B(\mathcal{G}_{\Delta t} \times \mathcal{G}_h)^2$, the operator $B(\mathcal{G}_h) \ni m \rightarrow (-\nu(L_h m) - (\text{div}_h(qm))) \in B(\mathcal{G}_h)$ is the adjoint of the operator $B(\mathcal{G}_h) \ni u \rightarrow (-\nu(L_h u) + q D_h u) \in B(\mathcal{G}_h)$.

Combining (3.2) and (3.3), and using the vector notations (2.19)–(2.21), we get the following fully discrete upwind scheme for the Newton iterations system (1.5)

$$\begin{cases} \mathcal{D}^k U^{n,k} = U^{n,k+1} + \Delta t \mathcal{W}^{k+1} M^{n,k+1} + \Delta t \tilde{\mathcal{B}}^{k+1} & \text{for } k \in \mathcal{I}_{\Delta t}^*, \\ (\mathcal{D}^k)^T M^{n,k+1} = M^{n,k} + \Delta t \mathcal{Z}^k U^{n,k} + \Delta t \tilde{\mathcal{C}}^{k+1} & \text{for } k \in \mathcal{I}_{\Delta t}^*, \\ U^{n,N_t} = U^T, \\ M^{n,0} = M^0, \end{cases} \quad (3.4)$$

where D^k is the $N_h^2 \times N_h^2$ matrix

$$D^k = I - \Delta t(\nu L_h + Q_1 D_1 + Q_2 D_2),$$

and $\tilde{C}^{k+1}, \tilde{B}^{k+1}$ are vectors in $\mathbb{R}^{N_h^2}$ defined as

$$\begin{aligned} (\tilde{C}^{k+1})_{i+N_h j} &= -(\mathcal{Z}^{k+1} U^{n-1,k})_{i+N_h j}, \\ (\tilde{B}^{k+1})_{i+N_h j} &= \frac{|(q^{n,k})_{i,j}|^2}{2} - V(x_{i,j}) \\ &\quad + F((M^{n-1,k+1})_{i+N_h j}) - F'((M^{n-1,k+1})_{i+N_h j})(M^{n-1,k+1})_{i+N_h j} \quad \text{for } i, j \in \mathcal{I}_h. \end{aligned}$$

Finally, define $\tilde{U}$ and $\tilde{M} \in \mathbb{R}^{(N_t)N_h^2}$ by

$$\begin{aligned} (\tilde{U}^n)_{kN_h^2+iN_h+j} &= (U^{n,k})_{i+N_h j} \quad \text{for } k \in \mathcal{I}_{\Delta t}, i, j \in \mathcal{I}_h, \\ (\tilde{M}^n)_{kN_h^2+iN_h+j} &= (M^{n,k})_{i+N_h j} \quad \text{for } k \in \mathcal{I}_{\Delta t} \setminus \{0\}, i, j \in \mathcal{I}_h. \end{aligned}$$

In order to write system (3.4) in a matrix way, as in (2.32), we define first the matrices

$$\tilde{\mathbf{A}} = \begin{bmatrix} D^0 & -I & \mathcal{O} & \dots & \mathcal{O} \\ \mathcal{O} & D^1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -I \\ \mathcal{O} & \dots & \dots & \mathcal{O} & D^{N_t-1} \end{bmatrix}, \quad \tilde{\mathbf{W}} = \Delta t \begin{bmatrix} \mathcal{W}^1 & \mathcal{O} & \dots & \dots & \mathcal{O} \\ \vdots & \mathcal{W}^2 & \ddots & \ddots & \mathcal{O} \\ \mathcal{O} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathcal{O} & \dots & \dots & \dots & \mathcal{W}^{N_t} \end{bmatrix},$$

and

$$\tilde{\mathbf{Z}} = \Delta t \begin{bmatrix} \mathcal{Z}^0 & \mathcal{O} & \dots & \dots & \mathcal{O} \\ \mathcal{O} & \mathcal{Z}^1 & \ddots & \ddots & \mathcal{O} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathcal{O} & \mathcal{O} & \dots & \dots & \mathcal{Z}^{N_t-1} \end{bmatrix}.$$

Hence, (3.4) is equivalent to the system

$$\begin{bmatrix} \tilde{\mathbf{A}} & -\tilde{\mathbf{W}} \\ -\tilde{\mathbf{Z}} & \tilde{\mathbf{A}}^* \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{U}}^n \\ \tilde{\mathbf{M}}^n \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{B}} \\ \tilde{\mathbf{C}} \end{bmatrix}, \quad (3.5)$$

where $\tilde{\mathbf{B}} = \Delta t [B^1, \dots, B^{N_t} + \frac{1}{\Delta t} U^{N_t}]^\top$ and $\tilde{\mathbf{C}} = \Delta t [\frac{1}{\Delta t} M^0 + C^1, \dots, C^{N_t}]^\top$. Then, arguing as in the proof of Proposition 2.1, one can show the following well-posedness result.

Proposition 3.1. *Suppose that $\tilde{M}^{n-1} > 0$ for any $n \in \mathbb{N}$. Then, there exists a unique solution $(\tilde{U}^n, \tilde{M}^n)$ to the system (3.5).*

4. Numerical tests

In this section, we assess the performance of the proposed numerical schemes through a series of tests in both one and two spatial dimensions, considering cases with separable and non-separable Hamiltonians. In all tests, the Newton iterations are terminated when the following error metrics fall below a prescribed tolerance τ , set to 10^{-4} in all the experiments. Specifically, we monitor the relative variation between two consecutive Newton iterates for both the value function and the distribution, according to:

$$E(m^n) := \|m^{n+1} - m^n\|_\infty \quad \text{and} \quad E(u^n) := \|u^{n+1} - u^n\|_\infty. \quad (4.1)$$

The systems defined by Eqs. (2.32) and (3.5) are solved using Block Gauss–Seidel iterations. The iterations are terminated once the uniform norms of the differences between consecutive solutions fall below a prescribed threshold δ , which is set to 10^{-4} . This ensures comparable accuracy for both discretization procedures. We compare three approaches for solving system (1.1): the semi-Lagrangian scheme (2.30) (Newton–SL), the finite difference scheme (3.4) (Newton–FD), and the finite difference schemes from [14] solved via Newton iterations (FD–Newton). Algorithm 1 implements the Newton method for both Newton–SL and Newton–FD. We note that the semi-Lagrangian scheme used in this work is explicit and does not require a parabolic CFL condition, whereas the finite difference schemes are implicit and thus impose no restriction on the time step. For accuracy purposes, the semi-Lagrangian scheme is run with a time step of the form $\Delta t = O(h^{3/2})$ (see [41,42] for a detailed analysis). We now present a series of four numerical tests designed to illustrate the behavior and performance of the proposed Newton-based schemes. The first test considers a one-dimensional MFG with a known reference solution, allowing us to validate the accuracy of the schemes and compare the performance of Newton–SL, Newton–FD, and FD–Newton. The second test focuses on one-dimensional MFG with smooth initial and terminal conditions, considered for two different diffusion coefficients ($\nu = 0.4$ and $\nu = 0.02$), highlighting the stability and robustness of the schemes. In the small-diffusion case ($\nu = 0.02$), we observe a break-down of the standard Newton iterations for the finite-difference schemes (Newton–FD and FD–Newton) due to a poor initialization, whereas Newton–SL remains stable. To overcome this difficulty, we employ a globalized version of Newton’s method, which is activated whenever the local iteration fails; this global strategy is described in detail in the

following subsection. Given the robust and accurate performance of the Newton–SL scheme observed in the tests 1–2, the subsequent experiments involving non-separable Hamiltonian are conducted using only this discretization approach. The third test examine one-dimensional MFG system with non-separable Hamiltonian, and finally the fourth test examine a two-dimensional MFG systems with a separable Hamiltonian. Both tests demonstrate the flexibility and robustness of the Newton–SL scheme in handling more general and nonlinear Hamiltonian structures. Together, these tests provide a comprehensive evaluation of the schemes, highlighting their accuracy, efficiency, and stability across different problem types, dimensions, and diffusion regimes.

Algorithm 1 Local Newton iterations for mean field games.

```

1: Input: Initial guesses  $u^0$ ,  $m^0$ , and tolerance  $\tau$ 
2:  $q^0 \leftarrow D_h u^0$ 
3:  $n \leftarrow 0$ 
4: repeat
5:   Compute  $(m^{n+1}, u^{n+1})$  as the solution of (2.30) (Newton–SL) or (3.4) (Newton–FD)
6:    $E(m^n) \leftarrow \|m^{n+1} - m^n\|_\infty$ 
7:    $E(u^n) \leftarrow \|u^{n+1} - u^n\|_\infty$ 
8:   Update  $q^n \leftarrow D_h u^n$ 
9:    $n \leftarrow n + 1$ 
10: until  $E(m^n) \geq \tau$  or  $E(u^n) \geq \tau$ 
11: Output: Approximated solution  $(u^{n+1}, m^{n+1})$  to the MFG system (1.1)
  
```

Remark 4.1. [Block Gauss–Seidel for the discrete Newton system(2.32)]

To construct a Block Gauss–Seidel iteration, we perform a splitting of the Newton matrix (2.32) as

$$M = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{Z} & \mathbf{A}^\top \end{bmatrix}, \quad N = \begin{bmatrix} \mathbf{0} & \mathbf{W} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (4.2)$$

The corresponding Block Gauss–Seidel iteration is, for $\ell \geq 0$,

$$X^{(\ell+1)} = M^{-1} N X^{(\ell)} + M^{-1} b, \quad X^{(\ell)} = \begin{bmatrix} \bar{U}^{(\ell)} \\ \bar{M}^{(\ell)} \end{bmatrix}, \quad b = \begin{bmatrix} \mathbf{B} \\ \mathbf{C} \end{bmatrix}. \quad (4.3)$$

given an initial condition $X^{(0)}$.

Each diagonal block of M corresponds to the natural temporal propagation of U backward in time and M forward in time, while the coupling terms $\mathbf{W}$ and $\mathbf{Z}$ are contained in N and in the bottom-left block of M respectively. The iteration can be implemented as a *backward-forward sweep*:

1. *Backward sweep for U :*

$$U_{(\ell+1)}^{n, N_i} = U^{N_i}, \quad U_{(\ell+1)}^{n, k} = \mathcal{A}(Q^{n, k}) U_{(\ell+1)}^{n, k+1} + \Delta t \mathcal{W}^k M_{(\ell)}^{n, k} + \Delta t B^k, \quad k = N_i - 1, \dots, 0.$$

2. *Forward sweep for M :*

$$M_{(\ell+1)}^{n, 0} = M^0, \quad M_{(\ell+1)}^{n, k+1} = \mathcal{A}^*(Q^{n, k}) M_{(\ell+1)}^{n, k} + \Delta t \mathcal{Z}^{k+1} U_{(\ell+1)}^{n, k+1} + \Delta t C^{k+1}, \quad k = 0, \dots, N_i - 1.$$

The block-diagonal matrix M is invertible because $\mathbf{A}$ and $\mathbf{A}^\top$ are invertible. The convergence of the Block Gauss–Seidel is guaranteed if $\rho((\mathbf{A}^\top)^{-1} \mathbf{Z} \mathbf{A}^{-1} \mathbf{W}) < 1$. It can be shown that $\|\mathbf{A}^{-1}\|_\infty \leq C N_i$, $\|\mathbf{W}\|_\infty \leq C \Delta t$ and that $\|\mathbf{Z}\|_\infty = O\left(\frac{\Delta t}{h^2}\right)$, however these bounds are not sufficient condition to ensure convergence, and it seems difficult to improve these estimates in general. Nonetheless, in all our numerical experiments, the Block Gauss–Seidel method converges systematically when the discrete density preserves positivity.

A comparison with GMRES and BiCGSTAB is reported in Table 2 of Section 4.3.2. The results show that all the methods maintain convergence, however, Block Gauss–Seidel needs the lowest CPU time and the least number of iterations. This behavior indicates that the effective spectral radius of the iteration matrix remains strictly smaller than one in the tested regime.

Remark 4.2. For the semi-Lagrangian scheme, we approximate the drift term q^n using the following discrete formulation:

$$q_{i,j}^{n,k} = (D\rho_\varepsilon * u^{n-1,k})_{i,j} \quad \text{for all } k \in \mathcal{I}_{\Delta t}, i, j \in \mathcal{I}_h,$$

where $*$ denotes the discrete circular convolution between two grid functions $A, B \in \mathbb{R}^{N_h \times N_h}$ (see e.g [43, Chapter 8]), and, for $\varepsilon > 0$, the kernel ρ_ε is a Gaussian mollifier defined by

$$\rho_\varepsilon(x) = \frac{1}{\sqrt{2\pi\varepsilon}} e^{-\frac{|x|^2}{2\varepsilon^2}}, \quad \text{for all } x \in \mathbb{R}^d.$$

This approach, which is common in the framework of SL schemes (e.g see [17,24]), enhances the accuracy of the scheme and allows us to avoid the constraints of a parabolic CFL condition, which would otherwise limit the time step. For the finite difference scheme, we use instead the direct formulation:

$$q_{i,j}^{n,k} = (D_h u^{n-1,k})_{i,j} \quad \text{for all } k \in \mathcal{I}_{\Delta t}, i, j \in \mathcal{I}_h,$$

where D_h is given by (2.3).

4.1. Global Newton algorithm

In this section, we describe the global Newton method, which extends the local convergence properties of the standard Newton method to a global setting. The key idea behind this approach is to modify the Newton step when the current iterate is far from the solution, ensuring that each iteration makes sufficient progress. The global convergence of the Newton method in infinite dimensions is proved in [34]. We adapt this framework to our MFG problem to enhance robustness and guarantee convergence even when starting from rough or distant initial guesses. First, we define the *search direction* $d^n := (v^n, \rho^n)$ as the solution to the linear system

$$JF(u^n, m^n)d^n = -F(u^n, m^n). \quad (4.4)$$

Then, the globalized Newton method generates sequences $\{(u^n, m^n)\}$, $\{d^n\}$, and $\{\alpha^n\}$, related by the iteration:

$$(u^{n+1}, m^{n+1}) = (u^n, m^n) + \alpha^n d^n, \quad n \in \mathbb{N},$$

where the step length $\alpha^n > 0$ is determined by a *line-search* procedure using a suitably defined merit function. Following [34], we use the squared L^2 -norm of F as the merit function:

$$\begin{aligned} \Theta(u, m) := \frac{1}{2} \|F(u, m)\|_2^2 &= \int_{[0, T] \times \mathbb{T}^d} |-\partial_t u - v \Delta u + H(x, Du) - F(m(t, x))|^2 dx dt \\ &+ \int_{[0, T] \times \mathbb{T}^d} |\partial_t m - v \Delta m - \operatorname{div}(m D_p H(x, Du))|^2 dx dt \\ &+ \int_{\mathbb{T}^d} |u(T, x) - G(x)|^2 dx + \int_{\mathbb{T}^d} |m(0, x) - m_0(x)|^2 dx. \end{aligned} \quad (4.5)$$

A reasonable solution for the system should correspond to a root of the merit function, i.e., $\Theta(u, m) = 0$, which implies that $F(u, m) = 0$ is satisfied, meaning that the system of equations is fulfilled. The step length α^n is determined by enforcing the *Armijo condition*, which guarantees a sufficient decrease of the merit function at each iteration. Specifically, this condition requires

$$\Theta(u^{n+1}, m^{n+1}) \leq \Theta(u^n, m^n) + c \alpha^n \Theta'(u^n, m^n)(d^n),$$

where $0 < c < 1/2$ is a constant controlling the sufficient decrease, and Θ' denotes the Fréchet derivative of Θ . Moreover, when the direction d^n satisfies the Newton Eq. (4.4), it holds that

$$\Theta'(u^n, m^n)d^n = -2\Theta(u^n, m^n) = -\|F(u^n, m^n)\|^2. \quad (4.6)$$

The global Newton procedure for solving (1.1) is summarized in Algorithm 2. Notice that the Newton step (4.4) leads to the linear parabolic system for $d^n = (v^n, \rho^n)$, as given in (1.5). At the discrete level, the search direction d^n can be computed using either the Newton-SL or the Newton-FD scheme.

Algorithm 2 Global Newton iterations for mean field games.

- 1: **Input:** Initial guesses u^0, m^0 , parameters $\beta \in (0, 1)$, $c \in (0, \frac{1}{2})$, and tolerance $\tau > 0$
 - 2: $q^0 \leftarrow D_h u^0$
 - 3: $n \leftarrow 0$
 - 4: **repeat**
 - 5: Compute $d^n = (m^{n+1/2}, u^{n+1/2})$ s the solution of (2.30) (Newton-SL) or (3.4) (Newton-FD)
 - 6: Find the smallest integer $i_n \geq 0$ such that the step length $\alpha_n = \beta^{i_n}$ satisfies the Armijo condition:

$$\Theta((m^n, u^n) + \alpha_n d^n) \leq \Theta(m^n, u^n) + c \alpha^n \Theta'(m^n, u^n)d^n$$
 - 7: Update the iteration $(m^{n+1}, u^{n+1}) \leftarrow (m^n, u^n) + \alpha_n d^n$
 - 8: $E(m^n) \leftarrow \|m^{n+1} - m^n\|_\infty$
 - 9: $E(u^n) \leftarrow \|u^{n+1} - u^n\|_\infty$
 - 10: Update $q^n \leftarrow D_h u^n$
 - 11: $n \leftarrow n + 1$
 - 12: **until** $E(m^n) \geq \tau$ or $E(u^n) \geq \tau$
 - 13: **Output:** Approximated solution (u^{n+1}, m^{n+1}) to the MFG system (1.1)
-

Remark 4.3. The globalized Newton method typically requires more iterations than the local Newton method. Therefore, in our numerical experiments, we first apply the local Newton iterations, and only if they fail to converge, we switch to the globalized version.

4.2. Test 1: accuracy comparison for an MFG problem with reference solution

In this test (see [44]), we consider the MFG system (1.1) on the time-space domain $[0, 0.05] \times]0, 1[$ with periodic boundary conditions, $v = 0.05$, and $H(x, p) = \frac{|p|^2}{2}$. The coupling term is giving by

$$F(m(x)) = 4 \min(4, m) - 3m_0(x), \quad \text{for all } x \in]0, 1[.$$

Table 1

Test 1. Comparison of Newton–SL, FD–Newton, and Newton–FD schemes: errors, CPU times, and number of iterations for the solution (u, m) .

Newton–SL with $\Delta t = h^{3/2}/2$				
h	$E_\infty(u)$	$E_\infty(m)$	Time	Iterations
$2.50 \cdot 10^{-2}$	$5.51 \cdot 10^{-2}$	$1.64 \cdot 10^{-1}$	0.61s	6
$1.25 \cdot 10^{-2}$	$2.40 \cdot 10^{-2}$	$1.16 \cdot 10^{-1}$	2.77s	7
$6.25 \cdot 10^{-3}$	$1.83 \cdot 10^{-2}$	$6.61 \cdot 10^{-2}$	13.92s	7
$3.125 \cdot 10^{-3}$	$4.50 \cdot 10^{-3}$	$1.41 \cdot 10^{-2}$	80.60s	7
FD–Newton with $\Delta t = h/4$				
h	$E_\infty(u)$	$E_\infty(m)$	Time	Iterations
$2.50 \cdot 10^{-2}$	$1.23 \cdot 10^{-1}$	$3.11 \cdot 10^{-2}$	2.23s	7
$1.25 \cdot 10^{-2}$	$6.21 \cdot 10^{-2}$	$1.63 \cdot 10^{-2}$	18.32s	8
$6.25 \cdot 10^{-3}$	$3.14 \cdot 10^{-2}$	$8.75 \cdot 10^{-3}$	92.91s	8
$3.125 \cdot 10^{-3}$	$1.77 \cdot 10^{-2}$	$9.54 \cdot 10^{-3}$	597.21s	8
Newton–FD with $\Delta t = h/4$				
h	$E_\infty(u)$	$E_\infty(m)$	Time	Iterations
$2.50 \cdot 10^{-2}$	$1.532 \cdot 10^{-1}$	$3.42 \cdot 10^{-2}$	1.48s	7
$1.25 \cdot 10^{-2}$	$6.71 \cdot 10^{-2}$	$1.83 \cdot 10^{-2}$	12.27s	7
$6.25 \cdot 10^{-3}$	$3.37 \cdot 10^{-2}$	$9.51 \cdot 10^{-3}$	68.10s	7
$3.125 \cdot 10^{-3}$	$1.91 \cdot 10^{-2}$	$7.38 \cdot 10^{-3}$	436.01s	7

while the initial condition is given by

$$m_0(x) = \begin{cases} 4 \sin^2(2\pi(x - 1/4)) & \text{if } x \in [1/4, 3/4], \\ 0 & \text{otherwise,} \end{cases}$$

and the terminal condition G is set to zero. To evaluate the accuracy errors, we compare the approximate solutions with a reference solution computed using a high-order value iteration scheme [18] on a fine mesh with spatial step $h = 6.67 \times 10^{-4}$ and time step $\Delta t = h^{3/2}$. The accuracy of each scheme is assessed by computing the error in the discrete uniform norm. In our simulations, we set $\Delta t = h^{3/2}/2$ for the Newton–SL method, and $\Delta t = h/4$ for both Newton–FD and FD–Newton. It is worth noting that the semi-Lagrangian scheme is unconditionally stable, and the chosen time step is primarily dictated by accuracy considerations; in particular, we adopt the same time step used in [18], from which the reference solution is taken. On the other hand, the finite-difference scheme is implicit, allowing for relatively large time steps without any CFL restriction, so that the time step selection in this case is also driven by accuracy rather than stability constraints.

Table 1 reports the numerical results obtained for different grid resolutions h , including the errors, CPU times, and number of Newton iterations for the three methods under consideration. The discrete uniform errors are defined as

$$E_\infty(u) = \|u^n - u_{\text{ref}}\|_\infty, \quad E_\infty(m) = \|m^n - m_{\text{ref}}\|_\infty,$$

where $(u_{\text{ref}}, m_{\text{ref}})$ denotes the reference solution, and (u^n, m^n) denotes the numerical solution obtained by the Newton method at the iteration when the stopping criterion is satisfied.

We first observe that the Newton–SL approach requires fewer iterations and consistently less computational time compared to the two finite-difference-based schemes, while achieving an accuracy comparable to theirs. Between FD–Newton and Newton–FD, the results are very close in terms of both accuracy and CPU time, with Newton–FD showing a slight advantage in efficiency.

4.3. Test 2: a MFG problem varying the diffusion

In this test, we consider two MFG systems, also numerically studied in [45], defined on the time–space domain $[0, 0.01] \times (0, 1)$ with periodic boundary conditions and diffusion coefficients $\nu = 0.4$ and $\nu = 0.02$, respectively. The Hamiltonian is given by $H(x, p) = \frac{|p|^2}{2} - V(x)$, and the corresponding data, illustrated in Fig. 1, are

$$\begin{aligned} m_0(x) &= 1 + \frac{1}{2} \cos(2\pi x), \\ G(x) &= \sin(4\pi x) + 0.1 \cos(10\pi x), \\ V(x) &= 200 \cos(2\pi x) - 10 \cos(4\pi x), \\ F(m) &= m^2. \end{aligned}$$

The spatial and temporal discretization parameters are set to $h = 6.25 \cdot 10^{-3}$ and $\Delta t = h/4$ for the two finite-difference schemes, while $\Delta t = h^{3/2}/2$ is used for the Newton–SL method.

The purpose of this test is to assess the performance of the proposed methods in both a highly diffusive regime and a nearly deterministic setting, and to investigate how the diffusion parameter influences the convergence behavior and stability of the Newton

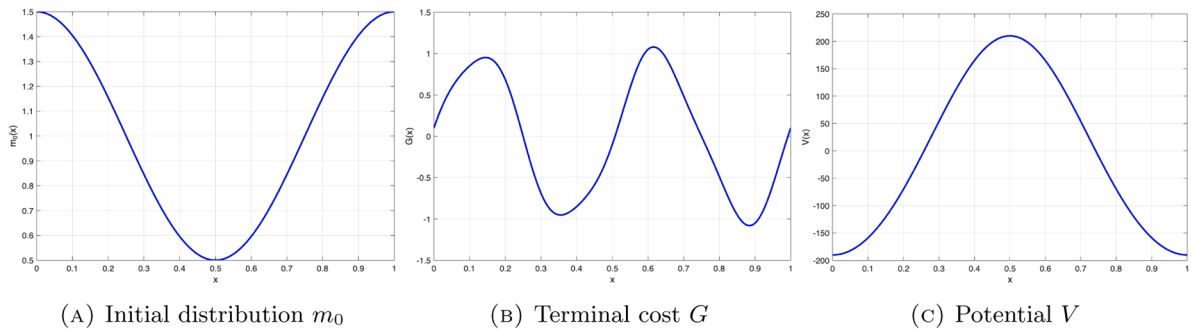


Fig. 1. Test 2. The initial data.

Table 2

Iteration counts and CPU time for the different solvers across both diffusion regimes ($\nu = 0.4$ and $\nu = 0.02$).

Solver	$\nu = 0.4$		$\nu = 0.02$	
	Iterations	CPU (s)	Iterations	CPU (s)
Block Gauss–Seidel	6	0.054	8	0.946
GMRES	6	4.251	13	5.914
BiCGSTAB	6	1.832	13	4.382

iterations. Furthermore, we provide a comparative analysis against standard iterative solvers to evaluate the relative computational efficiency and robustness of our approach across these different regimes.

4.3.1. Performance analysis of linear solvers

At each Newton step, the linear system (2.32) is solved using three different iterative solvers: Block Gauss–Seidel, GMRES, and BiCGSTAB. This comparison is performed for both the diffusive regime ($\nu = 0.4$) and the nearly deterministic setting ($\nu = 0.02$) to ensure the robustness of the solver selection.

Table 2 reports the corresponding iteration counts and CPU times for both cases. In the case $\nu = 0.4$, all three methods satisfy the stopping criterion in 6 iterations. However, as the diffusion parameter decreases to $\nu = 0.02$, the Krylov subspace methods (GMRES and BiCGSTAB) require significantly more iterations and computational time, whereas the Block Gauss–Seidel method remains highly efficient. Across both regimes, Block Gauss–Seidel yields the lowest CPU time. Consequently, the results presented in the remainder of this study correspond to the Block Gauss–Seidel solver.

4.3.2. Test 2.1 with $\nu = 0.4$

We start with the case $\nu = 0.4$. Fig. 2 shows, in logarithmic scale, the evolution of the Newton iteration errors $E(m^n)$ and $E(u^n)$ for the three schemes: Newton–SL, Newton–FD, and FD–Newton. All the three schemes exhibit comparable convergence behavior, whereas FD–Newton shows a slightly faster decay of the ℓ^∞ -norm of successive iterates.

In Fig. 3 we compare the convergence rate of the three Newton algorithms by plotting $\log(m^{n+1})$ against $\log(m^n)$. The estimated convergence rate is computed as the slope of the least-squares linear fit of the data points $(\log(m^n), \log(m^{n+1}))$, which provides a numerical estimate of the exponent p in the relation $m^{n+1} \approx C(m^n)^p$. This results in $p = 1.408$ for Newton–SL, $p = 1.393$ for Newton–FD and $p = 1.663$ for FD–Newton. For a quadratically exact convergent method we expect $p = 2$, hence an ideal slope of 2 in this log–log plot.

Fig. 4 displays the space–time evolution of the approximated distributions computed by the three schemes. It can be observed that, at the final time, the density concentrates around the points where the terminal cost G , shown in Fig. 1a, attains its minima. Finally, Fig. 5 presents the corresponding space–time plots of the approximated value functions obtained with the same schemes. The solutions appear smooth, as expected for the relatively large diffusion coefficient $\nu = 0.4$. Overall, the three methods produce visually consistent results, in good agreement with the reference test reported in [45], confirming the reliability and coherence of the proposed discretizations. The near identity of the results for Newton–FD and FD–Newton can be explained by the fact that both strategies employ the same finite difference spatial operators, leading to similar truncation errors and numerical diffusion profiles. The Newton–SL scheme behaves differently because it is based on a distinct discretization principle derived from the Feynman–Kac representation, which evolves the solution along characteristics. In FD schemes, numerical diffusion is mainly introduced by the upwind stencils used to stabilize the advection term, resulting in grid-based artificial dissipation. In contrast, in SL schemes the dominant source of diffusion arises from the interpolation step at the foot of the characteristics, yielding a different error distribution. Moreover, Newton–SL is run with a smaller time step, $\Delta t = \mathcal{O}(h^{3/2})$, compared to the FD scheme, which uses $\Delta t = \mathcal{O}(h)$. This higher temporal resolution, combined with the interpolation-based dissipation mechanism, results in a slightly less diffusive approximation in the Newton–SL solution and explains the small variations observed in Fig. 4.

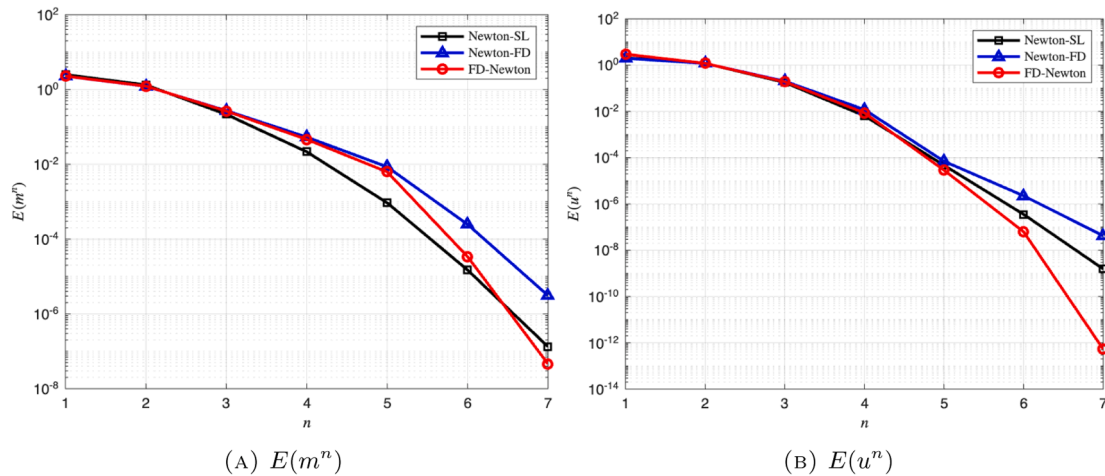


Fig. 2. Test 2.1. Convergence history of the Newton iterations for Newton-SL, Newton-FD, and FD-Newton.

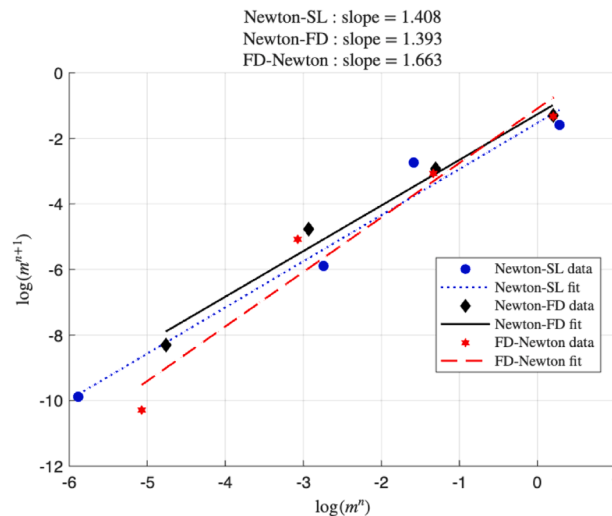


Fig. 3. Log-log plot of $\log(m^n)$ versus $\log(m^{n+1})$ for the three Newton-type schemes (Newton-SL, Newton-FD and FD-Newton).

4.3.3. Test 2.2

We now consider a smaller diffusion coefficient, $\nu = 0.02$. As shown in Fig. 6a, Newton-SL iterations converge after seven steps, reaching the prescribed tolerance. In contrast, both Newton-FD and FD-Newton iterations experience breakdown after only a few iterations, as the computed distribution becomes negative. This behavior highlights the increased robustness of the Newton-SL scheme in regimes characterized by low diffusion. Fig. 6b-c display the space-time evolution of the approximated distribution and the corresponding value function obtained with Newton-SL.

Since both FD-Newton and Newton-FD fail when applying local iterations, we employ the global Newton method described in Section 4.1, using the parameters $\beta = 1/2$ and $c = 1/3$ for the line-search procedure. These values ensure sufficient decrease of the merit function and robust globalization of the iterations. As expected, both schemes converge under this approach. Fig. 7 shows the Newton residual error, the merit function Θ (4.5), and the Armijo step size α_n along the globalized Newton iterations for the Newton-FD scheme. Fig. 8 displays the time evolution of the approximated distribution and value function computed by the global Newton-FD scheme (at convergence). For brevity, we omit the analogous results for FD-Newton, whose performance is comparable to that of Newton-FD. These results confirm the effectiveness of the globalized Newton strategy in stabilizing the discretizations and achieving convergence even in challenging small-diffusion MFG regimes.

Remark 4.4. In [46], the authors solve a finite difference discretization of the MFG system by employing Newton's method combined with a continuation strategy with respect to the diffusion parameter ν (see also [32,33]). This approach is particularly effective for handling problems with small diffusion values. Specifically, the problem is first solved for a large value of ν , and the corresponding solution is then used as the initial guess to solve, still by Newton's method, the discrete MFG system with a smaller diffusion coefficient. The procedure is repeated until the target (small) viscosity is reached. We point out that such a continuation approach has a similar

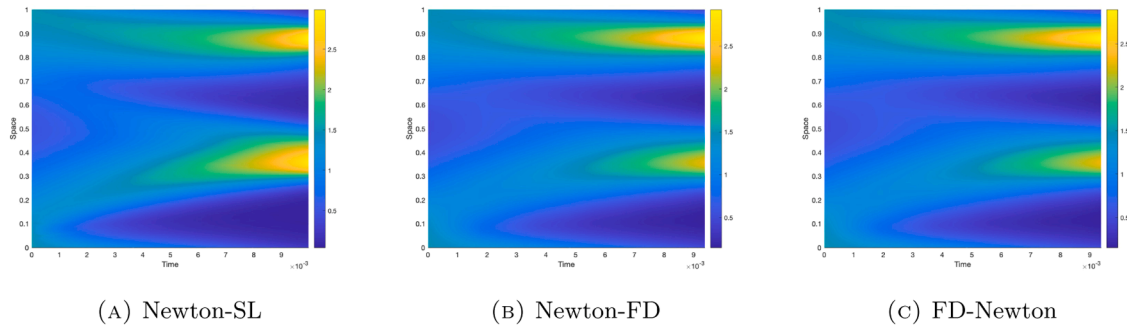


Fig. 4. Test 2.1. Space-time evolution of the approximated distributions m computed by Newton-SL, Newton-FD, and FD-Newton.

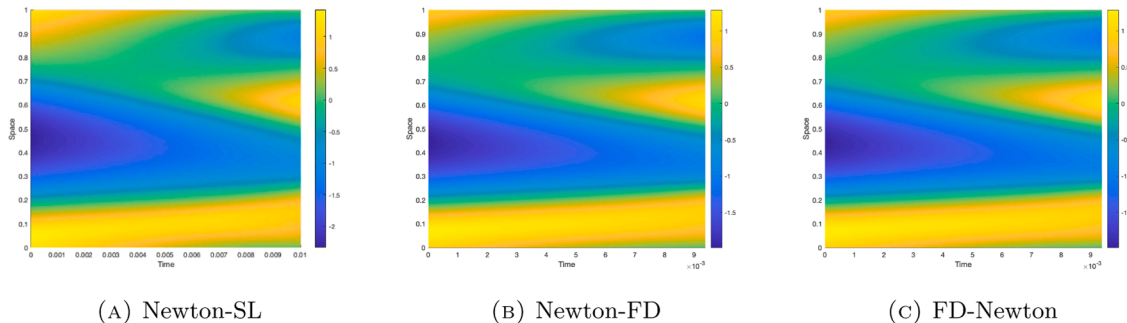


Fig. 5. Test 2.1. Space-time evolution of approximated value function u computed by Newton-SL, Newton-FD, and FD-Newton methods.

effect to the global Newton iterations, as both techniques aim to improve the robustness of convergence when the local Newton method fails due to poor initialization.

Remark 4.5. A simple alternative to the globalized Newton method is the relaxed Newton method, which uses a fixed damping parameter $\alpha \in (0, 1)$ at every iteration. The globalized Newton method with Armijo line search can be viewed as an adaptive generalization: it automatically selects the damping parameter α_n at each iteration. As shown in Fig. 7, the step size is small in the early iterations where the initial guess is far from the solution, and increases to $\alpha_n = 1$ as the method approaches convergence. This shows that damping is necessary in the early phase, but becomes unnecessary near the solution, a behavior that a fixed $\alpha < 1$ cannot capture.

4.4. Test 3: a MFG problem with non separable Hamiltonian

We consider a one-dimensional MFG system featuring a non-separable Hamiltonian,

$$\begin{cases} -\partial_t u - 0.05 \Delta u + \frac{1}{2(1+4m)^\gamma} |Du|^2 = \zeta m & \text{in } [0, 1] \times [0, 1], \\ \partial_t m - 0.05 \Delta m - \operatorname{div} \left(\frac{m Du}{(1+4m)^\gamma} \right) = 0 & \text{in } [0, 1] \times [0, 1], \\ G(x) = 10 \min\{(x - 0.3)^2, (x - 0.7)^2\} & \text{in } [0, 1], \\ m_0(x) = 4 \text{ for } x \in [0.375, 0.625], m_0(x) = 0 & \text{otherwise} \end{cases}$$

In this test, we compute the solution of the system using the Newton-SL scheme, with parameters $\gamma = 1.5$, $\zeta = 1$, spatial step $h = 5 \cdot 10^{-3}$, and time step $\Delta t = h^{3/2}/2$.

In Fig. 9, we display the time evolution of the approximated density and value function at the time instants $t_k \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. We observe that the population distribution splits into two groups: one moving toward the left target and the other toward the right target. However, due to the presence of congestion and crowd-aversion effects, neither group fully concentrates on its respective target at the final time. For this non-separable Hamiltonian test, the simulation obtained with the Newton-SL scheme is stable and accurate, and remains fully consistent with the reference outcomes reported in [45].

We then vary the value of the diffusion coefficient ν to investigate its influence on the solution. Fig. 10 reports the terminal density $m(T, x)$ and the initial value function $u(0, x)$ for $\nu \in \{0.2, 0.05, 0.005\}$. In particular, we can see that decreasing ν reduces diffusive smoothing and yields a more pronounced separation/concentration of the distribution at final time. Finally, Table 3 reports the corresponding iteration counts and CPU times for each value of ν .

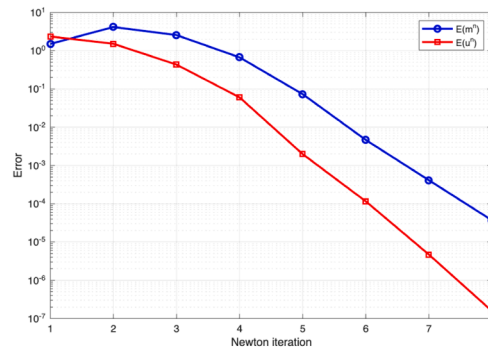
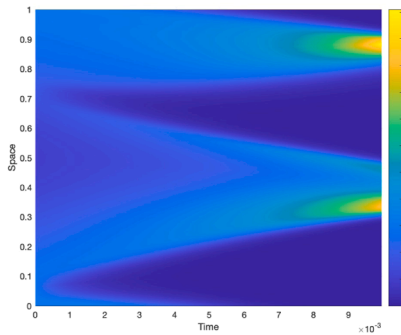
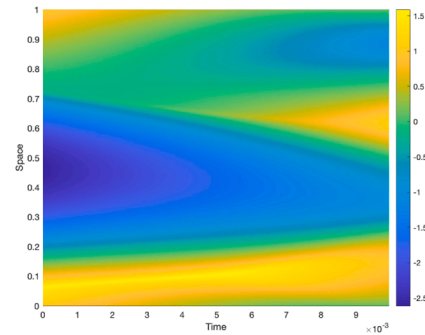
(A) Errors $E(m^n)$ and $E(u^n)$ (B) Evolution of m (C) Evolution of u

Fig. 6. Convergence history of the Newton-SL iterations and space-time evolution of the approximated distribution m and value function u for $\nu = 0.02$.

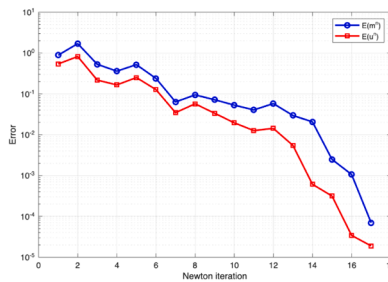
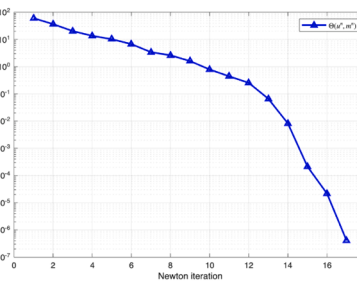
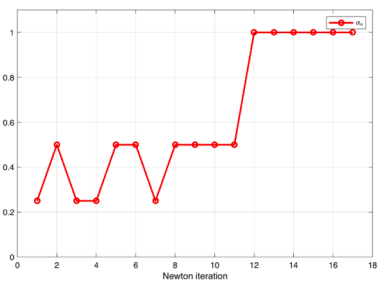
(A) Errors $E(m^n)$ and $E(u^n)$ (B) Merit function $\Theta(u^n, m^n)$ (C) Step size α^n

Fig. 7. Test 2.2. From left to right: Convergence history for the global Newton iterations for Newton-FD with $\nu = 0.02$, the merit function $\Theta(u^n, m^n)$, and the Armijo step size α^n .

4.5. Test 4: a MFG with separable Hamiltonian in dimension two

We now consider a two-dimensional MFG system, previously investigated numerically in [45]. The computational domain is $[0, 1] \times (0, 1)^2$, with diffusion coefficient $\nu = 0.4$, and the following data:

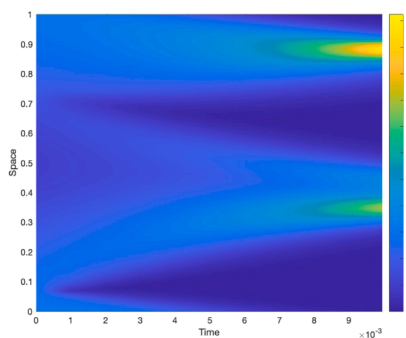
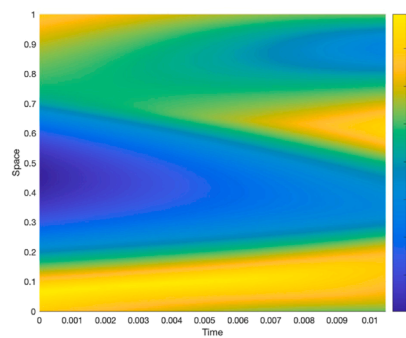
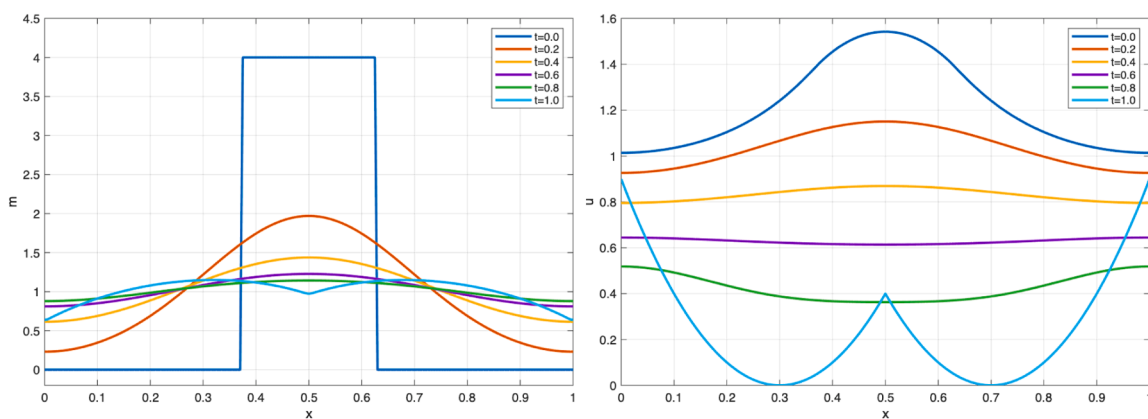
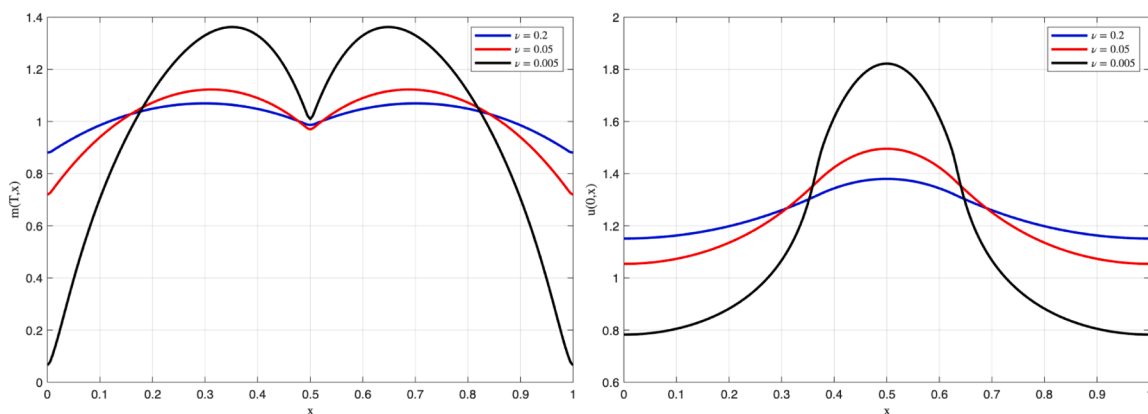
$$m_0(x_1, x_2) = 1 + 12 \cos(2\pi x_1) + 12 \cos(2\pi x_2),$$

$$G(x_1, x_2) = \cos(2\pi x_1) + \cos(2\pi x_2),$$

$$H(x_1, x_2, p) = |p|^2 + \sin(2\pi x_1) + \sin(2\pi x_2) + \cos(4\pi x_1),$$

$$F(m) = m^2.$$

We set $h = 10^{-2}$ and $\Delta t = h^{3/2}/2$. The evolution of the approximated distribution m , computed with the Newton-SL scheme, is displayed in Fig. 11 at times $t_k \in \{0, 0.5, 0.75, 1\}$. Fig. 11b-c illustrate the stationary profile attained by the density at intermediate times, which can be interpreted as a manifestation of the turnpike effect. Turnpike phenomena for mean field games with local couplings

(A) Evaluation of m (B) Evaluation of u **Fig. 8.** Test 2.2. Space-time evolution of the approximated distribution m and value function u by global Newton-FD for $\nu = 0.02$.**Fig. 9.** Test 3. Time evolution of the approximated density (left) and approximated value function (right).**Fig. 10.** Test 3. The approximated density at final time (left) and the approximated value function at initial time (right) for different values of ν .**Table 3**Iteration counts and CPU time for different diffusion coefficients ν .

ν	Number of iterations	CPU (s)
0.2	5	2.631
0.05	6	4.064
0.005	14	12.83

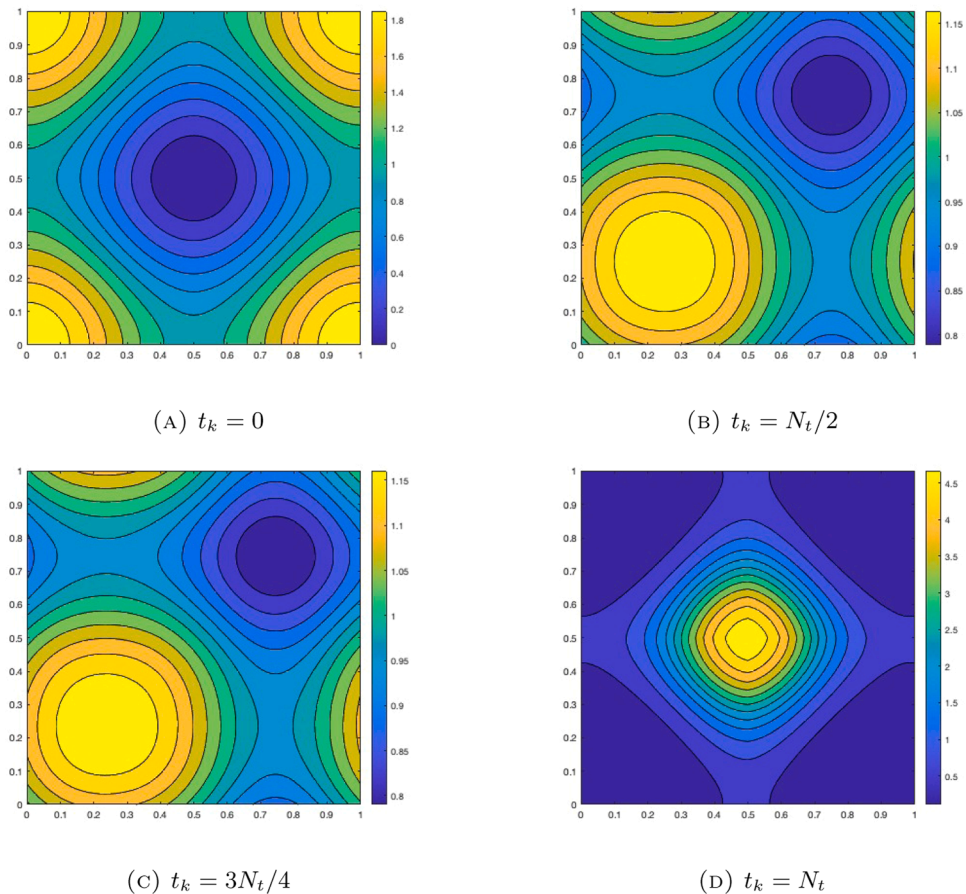


Fig. 11. Test 4. Time evolution of the approximated distribution m at with Newton-SL at different times.

have been discussed in [47]. The results are consistent with those reported in [45], confirming the accuracy of the proposed scheme in reproducing two-dimensional MFG dynamics.

5. Conclusion

In this work, we have proposed a new technique for numerically solving a second-order mean field games system with local coupling by applying Newton iterations in infinite dimensions. Our study is based on a series of numerical experiments comparing two possible strategies: performing Newton iterations at the continuous level and then discretizing, or discretizing first and applying Newton iterations to the resulting finite-dimensional non linear system. A first remark is that the former strategy leads to a relatively simple scheme, since each Newton step reduces to approximate a system of two linear parabolic PDEs. Both Newton-SL and Newton-FD follow a “iteration then discretization” strategy: the Newton direction is computed after discretization, and therefore the iterations do not follow the exact gradient of the continuous problem. As a consequence, the ideal quadratic convergence predicted at the continuous level is not fully observed in practice. In contrast, a “discretization then iteration” strategy, such as FD-Newton, builds the Newton iterations directly from the discrete problem and, in our tests, tends to follow more closely the true discrete gradient, sometimes displaying a behaviour closer to quadratic convergence. Nevertheless, one should not expect the “discretization then iteration” strategy to systematically yield a globally smaller numerical error. Discretizing before or after the Newton step introduces different truncation errors, which may counterbalance any local improvement in the convergence rate. To improve the performance of Newton-based schemes for MFG systems, a promising direction is to employ higher-order discretization schemes within an iterate-then-discretize strategy (e.g., high-order semi-Lagrangian or high-order finite difference methods). This allows the numerical solution to remain closer to the continuous formulation, thereby significantly enhancing both the convergence speed and the overall accuracy of the solver. Finally, as expected, the Newton-SL scheme exhibits strong robustness in advection-dominated regimes, whereas both Newton-FD and FD-Newton require a stabilization strategy to prevent breakdown. In this regard, the use of a globalized Newton method proved particularly effective, especially for problems with small viscosity coefficients.

Acknowledgements

The first author were partially supported by Italian Ministry of Instruction, University and Research (MIUR) (PRIN Project2022238YY5, “Optimal control problems: analysis, approximation”) and by INdAM–GNCS Project, codice CUP_E53C24001950001.

Data availability

Data will be made available on request.

References

- [1] F. Camilli, Q. Tang, On the quadratic convergence of Newton’s method for mean field games with non-separable Hamiltonian, *Dyn. Games Appl.* 15 (2025) 534–557.
- [2] J.M. Lasry, P.L. Lions, Jeux à champ moyen. II – horizon fini et contrôle optimal, *Comptes Rendus. Mathématique* 343 (10) (2006) 679–684.
- [3] J.M. Lasry, P.L. Lions, Mean field games, *Jpn. J. Math.* 2 (03) (2007).
- [4] M. Huang, R.P. Malham’e, P.E. Caines, Large population stochastic dynamic games: closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle, *Commun. Inf. Syst.* 6 (3) (2006) 221–252.
- [5] M. Huang, P.E. Caines, R.P. Malham’e, Individual and mass behaviour in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions, in: 42nd IEEE International Conference on Decision and Control (IEEE Cat. No.03CH37475), 1, 2003.
- [6] R.A. Carmona, Applications of mean field games in financial engineering and economic theory, *Proc. Symp. Appl. Math.* 78 (2021) 165–218.
- [7] R. Carmona, F. Delarue, Probabilistic Theory of Mean Field Games with Applications. I, volume 83 of Probability Theory and Stochastic Modelling, Springer, Cham, 2018.
- [8] P. Cardaliaguet, F. Delarue, J.M. Lasry, P.L. Lions, The Master Equation and the Convergence Problem in Mean Field Games, volume 201 of Ann. Math. Stud., Princeton, NJ: Princeton University Press, 2019.
- [9] R. Carmona, F. Delarue, Probabilistic Theory of Mean Field Games with Applications. II, volume 84 of Probability Theory and Stochastic Modelling, Springer, Cham, 2018.
- [10] D.A. Gomes, J. Saúde, Mean field games models—a brief survey, *Dyn. Games Appl.* 4 (2) (2014) 110–154.
- [11] D.A. Gomes, E.A. Pimentel, V. Voskanyan, Regularity Theory for Mean-Field Game Systems, SpringerBriefs in Mathematics, Springer, Cham, 2016.
- [12] Y. Achdou, P. Cardaliaguet, F. Delarue, A. Porretta, F. Santambrogio, Mean Field Games, volume 2281 of Lecture Notes in Mathematics, Springer, Cham; Centro Internazionale Matematico Estivo (C.I.M.E.), Florence, 2020.
- [13] P. Cardaliaguet, A. Porretta, An introduction to mean field game theory, in: Mean Field Games, volume 2281 of Lecture Notes in Math., Springer, Cham, 2020, pp. 1–158.
- [14] Y. Achdou, I. Capuzzo-Dolcetta, Mean field games: numerical methods, *SIAM J. Numer. Anal.* 48 (3) (2010) 1136–1162.
- [15] Y. Achdou, F. Camilli, I. Capuzzo-Dolcetta, Mean field games: numerical methods for the planning problem, *SIAM J. Control Optim.* 50 (2012).
- [16] Y. Achdou, A. Porretta, Convergence of a finite difference scheme to weak solutions of the system of partial differential equations arising in mean field games, *SIAM J. Numer. Anal.* 54 (1) (2016) 161–186.
- [17] E. Carlini, F.J. Silva, On the discretization of some nonlinear Fokker–Planck–Kolmogorov equations and applications, *SIAM J. Numer. Anal.* 56 (4) (2018) 2148–2177.
- [18] E. Calzola, E. Carlini, F.J. Silva, A high-order scheme for mean field games, *J. Comput. Appl. Math.* 445 (2024) 115769.
- [19] R. Carmona, M. Laurière, Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games I: the ergodic case, *SIAM J. Numer. Anal.* 59 (3) (2021) 1455–1485.
- [20] R. Carmona, M. Laurière, Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games: II—the finite horizon case, *Ann. Appl. Probab.* 32 (6) (2022) 4065–4105.
- [21] A. Angiuli, J.-P. Fouque, M. Laurière, Unified reinforcement Q-learning for mean field game and control problems, *Math. Control Signals Syst.* 34 (2) (2022) 217–271.
- [22] E. Carlini, L. Saluzzi, High order tensor-train-based schemes for high-dimensional mean field games, 2025.
- [23] E. Carlini, F.J. Silva, A fully discrete semi-Lagrangian scheme for a first order mean field game problem, *SIAM J. Numer. Anal.* 52 (1) (2014) 45–67.
- [24] E. Carlini, F.J. Silva, A. Zorkot, A Lagrange–Galerkin scheme for first order mean field game systems, *SIAM J. Numer. Anal.* 62 (2024) 167–198.
- [25] S. Hadikhannloo, F.J. Silva, Finite mean field games: fictitious play and convergence to a first order continuous mean field game, *J. Math. Pures Appl.* 132 (2019) 369–397.
- [26] J. Gianatti, F.J. Silva, A. Zorkot, Approximation of deterministic mean field games under polynomial growth conditions on the data, *J. Dyn. Games* 11 (2) (2024) 131–149.
- [27] J. Gianatti, F.J. Silva, Approximation of deterministic mean field games with control-affine dynamics, *Found. Comput. Math.* 24 (6) (2024) 2017–2061.
- [28] E. Carlini, V. Coscetti, A semi-Lagrangian scheme for first-order mean field games based on monotone operators, *IMA J. Numer. Anal.*, (2026) drag031.
- [29] Y. Achdou, M. Laurière, Mean field games and applications: numerical aspects, in: Mean Field Games, volume 2281 of Lecture Notes in Math., Springer, Cham, 2020.
- [30] M. Laurière, Numerical methods for mean field games and mean field type control, in: Mean Field Games, volume 78 of Proc. Sympos. Appl. Math., Amer. Math. Soc., Providence, RI, 2021.
- [31] J. Berry, O. Ley, F.J. Silva, Approximation and perturbations of stable solutions to a stationary mean field game system, *J. Math. Pures Appl.* 194 (2025) 103666.
- [32] Y. Achdou, M. Bardi, M. Cirant, Mean field games models of segregation, *Math. Models Methods Appl. Sci.* 27 (01) (2017) 75–113.
- [33] Y. Achdou, Z. Kobeissi, Mean field games of controls: finite difference approximations, *Math. Eng.* 3 (3) (2021) 1–35.
- [34] M. Gerds, Global convergence of a nonsmooth Newton method for control-state constrained optimal control problems, *SIAM J. Optim.* 19 (1) (2008) 326–350.
- [35] L. Bonaventura, E. Calzola, E. Carlini, R. Ferretti, Second order fully semi-Lagrangian discretizations of advection-diffusion-reaction systems, *J. Sci. Comput.* 88 (1) (2021) 23.
- [36] A. Quarteroni, A. Valli, Numerical Approximation of Partial Differential Equations, Springer Verlag, 1994.
- [37] J. Yong, X.Y. Zhou, Stochastic Controls. Hamiltonian Systems and HJB Equations, volume 43 of Appl. Math. (N. Y.), New York, NY: Springer, 1999.
- [38] M. Falcone, R. Ferretti, Semi-Lagrangian Approximation Schemes for Linear and Hamilton–Jacobi Equations, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014.
- [39] P.E. Kloeden, E. Platen, The numerical solution of stochastic differential equations, The Journal of the Australian Mathematical Society. Series B. Applied Mathematics 20 (1977) 8–12.
- [40] C.R. Johnson, Positive definite matrices, *Am. Math. Mon.* 77 (3) (1970) 259–264.
- [41] R. Ferretti, On the relationship between semi-Lagrangian and Lagrange–Galerkin schemes, *Numer. Math.* 124 (2013) 31–56.
- [42] R. Ferretti, Equivalence of semi-Lagrangian and Lagrange–Galerkin schemes under constant advection speed, *J. Comput. Math.* 28 (2010) 461–473.
- [43] A.V. Oppenheim, R.W. Schaffer, Discrete-Time Signal Processing, Pearson Education, 3rd edition, 2010.
- [44] P. Bojan, T. Vladimir, Central schemes for mean field games, *Commun. Math. Sci.* (2015) 13.

- [45] H. Li, Y. Fan, L. Ying, A simple multiscale method for mean field games, *J. Comput. Phys.* 439 (2021) 110385.
- [46] Y. Achdou, M. Laurière, *Mean Field Games and Applications: Numerical Aspects*, Springer International Publishing, Cham, 2020.
- [47] M. Cirant, A. Porretta, Long time behavior and turnpike solutions in mildly non-monotone mean field games, *ESAIM, Control Optim. Calc. Var.* 27 (2021) 86.
- [48] J.M. Lasry, P.L. Lions, Jeux à champ moyen. I – Le cas stationnaire, *Comptes Rendus. Mathématique* 343 (9) (2006) 619–625.