ON THE IMPLEMENTATION OF A PRIMAL-DUAL ALGORITHM FOR SECOND ORDER TIME-DEPENDENT MEAN FIELD GAMES WITH LOCAL COUPLINGS

L. BRICEÑO-ARIA S, D. KALISE, Z. KOB EISSI, M. LAURIÈRE, Á. MATEOS GONZÁ LEZ, AND F. J. SILVA

Abstract. We study a numerical approximation of a time-dependent Mean Field Game (MFG) system with local couplings. The discretization we consider stems from a variational approach described in [14] for the stationary problem and leads to the finite difference scheme introduced by Achdou and Capuzzo-Dolcetta in [3]. In order to solve the finite dimensional variational problems, in [14] the authors implement the primal-dual algorithm introduced by Chambolle and Pock in [20], whose core consists in iteratively solving linear systems and applying a proximity operator. We apply that method to time-dependent MFG and, for large viscosity parameters, we improve the linear system solution by replacing the direct approach used in [14] by suitable preconditioned iterative algorithms.

1. Introduction

In this work we consider the following MFG system with local couplings

\[
\begin{cases}
-\partial_t u - \nu \Delta u + H(x, \nabla u) = f(x, m(x, t)) & \text{in } \mathbb{T}^d \times [0, T], \\
\partial_t m - \nu \Delta m - \text{div}(\nabla_p H(x, Du)m) = 0 & \text{in } \mathbb{T}^d \times [0, T], \\
m(\cdot, 0) = m_0(\cdot), \quad u(\cdot, T) = g(\cdot, m(\cdot, T)) & \text{in } \mathbb{T}^d.
\end{cases}
\]

(MFG)

In the notation above \( \nu > 0 \), \( d \in \mathbb{N} \), \( \mathbb{T}^d \) is the \( d \)-dimensional torus, \( H : \mathbb{T}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \) is jointly continuous and convex with respect to its second variable, \( f, g : \mathbb{T}^d \times \mathbb{R} \rightarrow \mathbb{R} \) are continuous functions and \( m_0 : \mathbb{T}^d \rightarrow \mathbb{R} \) satisfies \( m_0 \geq 0 \) and \( \int_{\mathbb{T}^d} m_0(x)dx = 1 \).

System (MFG) has been introduced by J.-M. Lasry and P.-L. Lions in [27, 28] in order to describe the asymptotic behaviour of symmetric stochastic differential games as the number of players tends to infinity. Several analytical techniques can be used to prove the existence of solutions to (MFG) under various assumptions on the data. Even if system (MFG) has been introduced recently, there are already too many references about its theoretical study to be exhaustive in this introduction and we content ourselves in referring the interested reader to the monographs [10, 24], the surveys [16, 25] and the references therein for the state of the art of the subject.

A useful argument that can be used to establish the existence of solutions to (MFG) is the variational one, already presented in [28]. The main idea behind is that, at least formally, system (MFG) can be seen as the first order optimality condition associated to minimizers of the following optimization problem

\[
\inf_{(m, w)} \int_0^T \int_{\mathbb{T}^d} \left[ b(x, m(x, t), w(x, t)) + F(x, m(x, t)) \right] dx + \int_{\mathbb{T}^d} G(x, m(x, T))dx
\]

subject to

\[
\begin{cases}
\partial_t m - \nu \Delta m + \text{div}(w) = 0 & \text{in } \mathbb{T}^d \times (0, T), \\
m(\cdot, 0) = m_0(\cdot) & \text{in } \mathbb{T}^d.
\end{cases}
\]

1. Universidad Técnica Federico Santa María, Departamento de Matemática, Av. Vicuña Mackenna 3939, San Joaquín, Santiago, Chile. luis.briceno@usm.cl.

2. Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom. dkaliseb@ic.ac.uk.

3. Laboratoire Jacques-Louis Lions, Univ. Paris Diderot, Sorbonne Paris Cité, UMR 7598, UPMC, CNRS, 75205, Paris, France. ziad.kobeissi@ens-lyon.fr.

4. ORFE, Bendheim Center for Finance, Princeton University, Princeton, NJ 08540, USA. lauriere@princeton.edu.

5. Institut Montpelliérain Alexander Grothendieck (IMAG), Université de Montpellier, UMR 5149, CNRS, 34095, Montpellier, France. alvaro.mateos-gonzalez@umontpellier.fr.

6. Toulouse School of Economics, Université de Toulouse I Capitole, 31015 Toulouse, France and Institut de recherche XLIM-DMI, UMR-CNRS 7252 Faculté des sciences et techniques Université de Limoges, 87060 Limoges, France. francisco.silva@unilim.fr.
provided that they exist). In (P), the functions $b : \mathbb{T}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ and $F, G : \mathbb{T}^d \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ are defined as follows

$$b(x, m, w) = \begin{cases} mH^*(x, -\frac{w}{m}) & \text{if } m > 0, \\ 0 & \text{if } (m, w) = (0, 0), \\ +\infty & \text{otherwise}, \end{cases}$$

$$F(x, m) := \begin{cases} \int_0^m f(x, m') dm' & \text{if } m \geq 0, \\ +\infty & \text{otherwise}, \end{cases} \quad G(x, m) := \begin{cases} \int_0^m g(x, m') dm' & \text{if } m \geq 0, \\ +\infty & \text{otherwise}, \end{cases}$$

where, in the definition of $b$, $H^*(x, \cdot)$ denotes the Legendre-Fenchel conjugate of $H(x, \cdot)$. Under the assumption that $f(x, \cdot)$ and $g(x, \cdot)$ are non-decreasing, problem (P) is shown to be a convex optimization problem and convex duality techniques can be successfully applied in order to provide existence and uniqueness results to (MFG). This argument has been made rigorous in several articles: let us mention [17, 18] in the context of first order MFGs, the paper [19] for degenerate second order MFGs, and finally [29, 30] for ergodic second order MFGs.

The variational approach described above has also been successful in the numerical resolution of system (MFG). In this direction, we mention the article [26] dealing with applications in economics, the paper [1] concerned with the so-called planning problem in MFGs, the works [9, 7] focused on the resolution of a discretization of (P) by the Alternating Direction Method of Multipliers (ADMM) and [14] where several first order methods are implemented and compared for the stationary version of (MFG). Let us mention that the variational approach is closely related to the so-called mean field optimal control problem, for which numerical methods have been studied in [15, 6], among others.

In this paper we consider a finite difference discretization of problem (P). Assuming that $f(x, \cdot)$ and $g(x, \cdot)$ are non-decreasing, the discretization we consider is such that it preserves the convexity properties of problem (P) and the first order optimality conditions for its solutions, which are shown to exist, coincide with the finite difference scheme for MFGs introduced in [3]. A very nice feature of this approach is that the solutions of the resulting discretized MFGs are shown to converge to the solutions of (MFG). We refer the reader to [2], where the convergence result is obtained under the assumption that (MFG) admits a unique classical solution, and to [5] in the framework of weak solutions (see [32] for the definition of this notion). We solve the discrete convex optimization problem by using the primal-dual algorithm introduced in [20]. As was pointed out in [14] (see also [31] in the context of transport problems), the primal-dual algorithm we consider seems to be faster than the ADMM when $\nu$ in (MFG) is small (or null). On the other hand, the efficiency of both methods is arguable when $\nu$ is large. This is due to the fact that, in both algorithms, at each iteration one has to invert a matrix whose condition number importantly increases as the viscosity parameter increases. Naturally, preconditioning strategies (see e.g. [11]) can then be used in order to improve the efficiency of both algorithms. This strategy has been already successfully implemented in [7] for the ADMM.

Our main objective in the present work is to take a closer look at the phenomenon described at the end of previous paragraph when considering the primal-dual algorithm. We have implemented standard indirect methods for solving the linear systems appearing in the computation of the iterates of the primal-dual algorithm. As our numerical results suggest, it is very important to design suitable preconditioning strategies in order to be able to find solutions of the discretization of problem (P) efficiently, and in a robust way with respect to the viscosity. For this, we explore different preconditioning strategies, and in particular, multigrid preconditioning (see also [4, 7], where multigrid strategies have been implemented for other solution methods).

The article is organized as follows. In section 2 we introduce some standard notation and we recall the finite difference scheme for (MFG) introduced in [3]. The variational interpretation of this finite difference scheme is discussed in section 3. Next, in section 4, we recall the primal-dual algorithm introduced in [20] and we consider its application to the discretization of (P). In section 5, we summarize the preconditioning strategies we consider and we discuss a numerical example, which is the time-dependent version of one of the examples treated in [3, 14].
2. Preliminaries and the finite difference scheme in [3]

In this section we introduce some basic notation and present the finite difference scheme introduced in [3], whose efficient resolution will be the subject of this article. For the sake of simplicity, we will assume that \( d = 2 \) and that given \( q > 1 \), with conjugate exponent denoted by \( q’ = q/(q - 1) \), the Hamiltonian \( H : \mathbb{T}^2 \times \mathbb{R}^2 \to \mathbb{R} \) has the form

\[
H(x, p) = \frac{1}{q’} |p|^{q’} \quad \forall \ x \in \mathbb{T}^2, \ p \in \mathbb{R}^2.
\]

In this case the function \( b \) defined in (1.1) takes the form

\[
b(x, m, w) = \begin{cases} \lfloor w \rfloor^{\ast} & \text{if } m > 0, \\ 0 & \text{if } (m, w) = (0, 0), \\ +\infty & \text{otherwise.}
\end{cases}
\]

Let \( N_T, N_h \) be positive integers and set \( \Delta t = T/N_T \), the time step, and \( h = 1/N_h \), the space step.

We associate to these steps a time grid \( \mathcal{T}_\Delta t := \{ t_n = n\Delta t : n = 0, \ldots, N_T \} \) and a space grid \( \mathbb{T}^2_h := \{ x_{i,j} = (ih, jh) : i, j \in \mathbb{Z} \} \). Since \( \mathbb{T}^2_h \) intends to discretize \( \mathbb{T}^2 \), we impose the identification \( z_{i,j} = \bar{z}((i \mod N_h), (j \mod N_h)) \), which allows to assume that \( i, j \in \{ 0, \ldots, N_h - 1 \} \). A function \( y := \mathbb{T}^2 \times [0, T] \to \mathbb{R} \)

is approximated by its values at \( (x_{i,j}, t_n) \in \mathbb{T}^2_h \times \mathcal{T}_\Delta t \), which we denote by \( y^N_{i,j} := y(x_{i,j}, t_n) \). Given \( y : \mathbb{T}^2_h \to \mathbb{R} \) we define the first order finite difference operators

\[
(D_1 y)_{i,j} = \frac{y_{i+1,j} - y_{i,j}}{h}, \quad \text{and} \quad (D_2 y)_{i,j} = \frac{y_{i,j+1} - y_{i,j}}{h}.
\]

For later use, given \( a \in \mathbb{R} \), we set \( a^+ := \max(a, 0) \) and \( a^- := a^+ - a \), and we define

\[
\Delta h y_{i,j} = ((D_1 y)_{i,j}^-, (D_1 y)_{i,j}^+), (D_2 y)_{i,j}^-), (D_2 y)_{i,j}^+).
\]

The discrete Laplace operator \( \Delta_h y : \mathbb{T}^2_h \to \mathbb{R} \) is defined by

\[
(\Delta_h y)_{i,j} = -\frac{1}{h^2} (4y_{i,j} - y_{i+1,j} - y_{i-1,j} - y_{i,j+1} - y_{i,j-1}).
\]

For \( y : \mathbb{T}^2_h \times \mathcal{T}_\Delta t \to \mathbb{R} \) we define the discrete time derivative

\[
D_t y^k_{i,j} = \frac{y^k_{i,j+1} - y^k_{i,j}}{\Delta t}.
\]

The Godunov-type finite difference discretization of (MFG) introduced in [3] is as follows: find \( u, m : \mathbb{T}^2_h \times \mathcal{T}_\Delta \to \mathbb{R} \) such that for all \( 0 \leq i, j \leq N_h - 1 \) and \( 0 \leq k \leq N_T - 1 \) we have

\[
\begin{cases}
-D_t u^k_{i,j} - \nu(\Delta_h u^k)_{i,j} + \frac{1}{q’} |(D_h u^k)_{i,j}|^{q’} = f(x_{i,j}, m^{k+1}_{i,j}), \\
D_t m^{k+1}_{i,j} - \nu(\Delta_h m^{k+1})_{i,j} - \mathcal{T}_{i,j}(u^k, m^{k+1}) = 0,
\end{cases}
\]

(MFG\(_h,\Delta t\))

where

\[
\bar{m}_{i,j} := \int_{|x-x_{i,j}| \leq \frac{h}{2}} m_0(x)dx \geq 0,
\]

and the operator \( \mathcal{T}(u’, m’) : \mathbb{T}^2_h \to \mathbb{R} \), with \( u’, m’ : \mathbb{T}^2_h \to \mathbb{R} \), is defined by
dual operators are given by
for all $0 \leq i,j \leq N$.
Denoting by $R$ the set of non-negative real numbers and by $\mathbb{R}_-$ the set of non-positive real numbers, we define

\[ K := \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_- \quad \text{and we denote by } P_K(v) \text{ the orthogonal projection of } v \in \mathbb{R}^4 \text{ onto } K. \]

Let $\mathcal{M} := \mathbb{R}^{(N_T+1) \times N_h \times N_h}$, $\mathcal{W} := (\mathbb{R}^4)^{N_T \times N_h \times N_h}$, and $\mathcal{U} := \mathbb{R}^{N_T \times N_h \times N_h}$. If $y \in \mathbb{R}^4$, we denote by $y^l$ the $l$-th coordinate of $y$, $1 \leq l \leq 4$. Let $A : \mathcal{M} \to \mathcal{U}$ and $B : \mathcal{W} \to \mathcal{U}$ be defined by

\[
(A_{m,k})_{i,j} = D_t m_{i,j}^k - \nu(\Delta_h m_{i,j}^{k+1}), \\
(B_{w,k})_{i,j} = (D_1 w_{i,j}^{k,(1)})_{i-1,j} + (D_1 w_{i,j}^{k,(2)})_{i,j} + (D_2 w_{i,j}^{k,(3)})_{i,j-1} + (D_2 w_{i,j}^{k,(4)})_{i,j},
\]

for all $0 \leq i,j \leq N_h - 1$ and $0 \leq k \leq N_T - 1$. One can easily check (see e.g. [3]) that the corresponding dual operators are given by

\[
(B^*_{m,k})_{i,j} = -[D_h w_{i,j}^{k}]_{i,j} \quad \text{for all } 0 \leq k \leq N_T - 1, \\
(A^*_{u,k})_{i,j} = -D_t u_{i,j}^{k-1} - \nu(\Delta_h u_{i,j}^{k-1}), \quad \text{if } 1 \leq k \leq N_T - 1,
\]

\[
(A^*_{u,0})_{i,j} = -\frac{1}{\Delta t} u_{i,j}^0, \\
(A^*_{u,N_T})_{i,j} = \frac{1}{\Delta t} u_{i,j}^{N_T-1} - \nu(\Delta_h u_{i,j}^{N_T-1}),
\]

for all $u \in \mathcal{U}$. For latter use, notice that

\[
\text{Ker}(B^*) = \{ u \in \mathcal{U} \mid \forall k = 0, \ldots, N_T - 1 \text{ there exists } c_k \in \mathbb{R} \text{ such that } u_{i,j}^k = c_k \ \forall \ i,j \},
\]

and so

\[
\text{Im}(B) = \text{Ker}(B^*)^\perp = \left\{ u \in \mathcal{U} \mid \sum_{i,j} u_{i,j}^k = 0 \ \forall k = 0, \ldots, N_T - 1 \right\}.
\]

Let us define $\tilde{b} : \mathbb{R} \times \mathbb{R}^4 \to \mathbb{R} \cup \{ +\infty \}$

\[
\tilde{b}(m, w) := \begin{cases} 
\frac{|m|^{\gamma}}{q m_+^{\gamma}}, & \text{if } m > 0, w \in K, \\
0, & \text{if } (m, w) = (0, 0), \\
+\infty, & \text{otherwise},
\end{cases}
\]

with the convention:

\[
\langle D_h u_{i,j}^k \rangle_{i,j}^q = 0 \text{ if } q > 0 \text{ and } [D_h u_{i,j}^k]_{i,j} = 0.
\]
and the functions $B, \mathcal{F}, G : \mathcal{M} \times \mathcal{W} \to \mathbb{R}, \mathcal{G} : \mathcal{M} \times \mathcal{W} \to \mathcal{M} \times \mathbb{R}^{N_h \times N_h}$ as

\[
B(m, w) = \sum_{1 \leq k \leq N_T, 0 \leq i,j \leq N_h-1} \tilde{b}(m_{i,j}^k, w_{i,j}^{k-1}),
\]

(3.5)

\[
\mathcal{F}(m, w) = \sum_{1 \leq k \leq N_T, 0 \leq i,j \leq N_h-1} F(x_{i,j}, m_{i,j}^k) + \frac{1}{\Delta t} \sum_{0 \leq i,j \leq N_h-1} G(x_{i,j}, m_{i,j}^{N_T}),
\]

\[
\mathcal{G}(m, w) := (Am + Bw, m^0).
\]

Note that if $(m, w) \in \mathcal{M} \times \mathcal{W}$ is such that $\mathcal{G}(m, w) = (0, \tilde{m})$ then

\[
h^2 \sum_{i,j} m_{i,j}^k = 0 \quad \forall k = 0, \ldots, N_T.
\]

Indeed, by periodicity, $-\sum_{i,j}(\Delta_h m_{i,j}^{k+1})_{i,j} = 0$ and $\sum_{i,j}(Bw)^k_{i,j} = 0$ for all $k = 0, \ldots, N_T - 1$. This implies that

\[
0 = \sum_{i,j} (Am + Bw)^k_{i,j} = \sum_{i,j} \frac{m_{i,j}^{k+1}}{\Delta t} - \sum_{i,j} \frac{m_{i,j}^k}{\Delta t},
\]

and so $h^2 \sum_{i,j} m_{i,j}^k = h^2 \sum_{i,j} \tilde{m}_{i,j} = 1$ for all $k = 0, \ldots, N_T$.

The discretization of the variational problem $(P)$ we consider is

\[
(P_{h,\Delta t}) \quad \inf_{(m, w) \in \mathcal{M} \times \mathcal{W}} B(m, w) + \mathcal{F}(m), \quad \text{subject to} \quad \mathcal{G}(m, w) = (0, \tilde{m}),
\]

where we recall that $F$ and $G$ in (3.5) are defined in (1.1).

We have the following result

**Theorem 3.1.** For any $\nu > 0$ problem $(P_{h,\Delta t})$ admits at least one solution $(m_{h,\Delta t}^k, w_{h,\Delta t}^k)$ and associated to it there exists $u_{h,\Delta t} : \mathcal{M} \times \mathcal{W} \to \mathbb{R}$ such that $(MFG_{h,\Delta t})$ holds true. Moreover, $(m_{h,\Delta t}^k)_{i,j} > 0$ for all $k = 1, \ldots, N_T, i, j = 0, \ldots, N_h - 1$.

In order to prove the result above, let us first show a lemma that implies the feasibility of the constraints in $(P_{h,\Delta t})$.

**Lemma 3.1.** There exists $(\tilde{m}, \tilde{w}) \in \mathcal{M} \times \mathcal{W}$ such that

\[
G(\tilde{m}, \tilde{w}) = (0, \tilde{m}), \quad \tilde{w}_{i,j}^k \in \text{int}(K) \quad \forall i, j = 1, \ldots, N_h - 1, \quad k = 1, \ldots, N_T - 1,
\]

\[
\tilde{m}_{i,j}^k > 0, \quad \forall i, \quad j = 1, \ldots, N_h - 1, \quad k = 1, \ldots, N_T.
\]

**Proof.** Let us define $\tilde{m}_{i,j}^0 = \bar{m}_{i,j}$ and $\tilde{m}_{i,j}^k := 1$ for all $k = 1, \ldots, N_T$ and $i, j$. Since $h^2 \sum_{i,j} \tilde{m}_{i,j}^k = 1$ for all $k = 0, \ldots, N_T$, by (3.3) and the definition of $A$ we easily get that $Am \in \text{Im}(B)$. Therefore, there exists $\tilde{w} \in \mathcal{W}$ satisfying $G(\tilde{m}, \tilde{w}) = (0, \tilde{m})$. Then, given $\delta > 0$, we set for all $k = 0, \ldots, N_T - 1$ and $i, j$

\[
\tilde{w}_{i,j}^k := \left( w_{i,j}^{k,(1)} + \max_{i,j} w_{i,j}^{k,(1)} + \delta, w_{i,j}^{k,(2)} - \max_{i,j} w_{i,j}^{k,(2)} - \delta, w_{i,j}^{k,(3)} + \max_{i,j} w_{i,j}^{k,(3)} + \delta, w_{i,j}^{k,(4)} - \max_{i,j} w_{i,j}^{k,(4)} - \delta \right),
\]

which satisfies $\tilde{w}_{i,j}^k \in \text{int}(K)$ and $(B\tilde{w})^k = (B\tilde{w})^k$. The result follows.

Now, we prove the existence of solutions to $(P_{h,\Delta t})$.

**Lemma 3.2.** Problem $(P_{h,\Delta t})$ admits at least one solution $(m_{h,\Delta t}^k, w_{h,\Delta t}^k)$ and every such solution satisfies

\[
(m_{h,\Delta t}^k)_{i,j} > 0 \quad \text{for all} \quad k = 1, \ldots, N_T, \quad i, j = 0, \ldots, N_h - 1.
\]

**Proof.** Let $(m^n, w^n)$ be a minimizing sequence for $(P_{h,\Delta t})$. Lemma 3.1 implies that $B(\tilde{m}, \tilde{w}) + \mathcal{F}(m) < +\infty$. Therefore, there exists a constant $C_1 > 0$ such that

\[
B(m^n, w^n) + \mathcal{F}(m^n) \leq C_1 \quad \text{for all} \quad n \in \mathbb{N}.
\]
As a consequence, by definition of \( \hat{b}, (m^n)_{i,j}^k \geq 0 \) for all \( i, j \) and \( k \) and \( (w^n)_{i,j}^k \in K \) for all \( k \). Since \( Am^n + Bw^n = 0 \), relation (3.6) implies that \( h^2 \sum_{i,j} (m^n)_{i,j}^k = 1 \). In particular, there exists \( C_2 > 0 \) (independent of \( n \)) such that \( \sup_{i,j,k} (m^n)_{i,j}^k \leq C_2 \). Using that, if \((m^n)_{i,j}^k > 0\),

\[
\hat{b}(m^n)_{i,j}^k, (w^n)_{i,j}^k \geq \frac{|(w^n)_{i,j}^k|^q}{qC_2^{q+2}},
\]

and that \( F(m^n) \) is uniformly bounded (because \( F \) and \( G \) are continuous and \( m^n \) is bounded), relation (3.8) yields the existence of \( C_3 > 0 \) (independent of \( n \)) such that \( \sup_{i,j,k} |(w^n)_{i,j}^k|^q \leq C_3 \). Thus, there exists \((m^n(t), w^n(t)) \in M \times W \) such that, up to some subsequence, \( m^n \to m^h(t) \) and \( w^n \to w^h(t) \) as \( n \to \infty \). Since \( G(m^n, w^n) = (0, \bar{m}) \) we obtain that \( G(m^n(t), w^n(t)) = (0, \bar{m}) \). The lower semicontinuity of \( B + F \) implies that

\[
B(m^h(t), w^h(t)) + F(m^h(t)) \leq \lim_{n \to \infty} B(m^n, w^n) + F(m^n),
\]

which implies that \((m^h(t), w^h(t)) \) solves (P_h(t)). Finally, if \((m, w) \in M \times W \) solves (P_h(t)) and \( m_{i,j}^{k,1} = 0 \) for some \( i, j \) and \( k = 1, \ldots, N_T \), then, by the definition of \( B \), we must have that \( w_{i,j}^{k,1} = 0 \). Thus, the constraint \((Am + Bw)_{i,j}^{k,1} = 0 \) can be written as

\[
-\frac{m_{i+1,j}^{k,1}}{\Delta t} - \frac{\nu}{k} (m_{i-1,j}^{k,1} + m_{i,j}^{k,1} + m^k_{i,j}) = \frac{w_{i-1,j}^{k,1}}{h^2} + \frac{w_{i,j}^{k,1}}{h^2} - \frac{w_{i+1,j}^{k,1}}{h^2}.
\]

Since the left hand side above is non-positive and the right hand side is non-negative (by definition of \( K \)), we deduce that all the terms above are zero. By repeating the argument at the indexes neighboring \((i, j)\), we deduce that \( m^k \equiv 0 \) and so \( h^2 \sum_{i,j} m_{i,j}^k = 0 \) which, by (3.6), contradicts \( G(m, w) = (0, \bar{m}) \). The result follows.

Proof of Theorem 3.1. By Lemma 3.2 we know that there exists a solution \((m^h(t), w^h(t)) \) to (P_h(t)) and \( m_{i,j}^{h,k} > 0 \) for all \( i, j \). Thus, in order to conclude it suffices to show the existence of \( u^h(t) \) such that (MFG_h(t)) holds true. For notational convenience we will omit the superindexes \( h \) and \( \Delta t \). Define the Lagrangian

\[
L(m, w, u, \lambda) := B(m, w) + F(m) - \langle u, Am + Bw \rangle - \langle \lambda, m^0 - \bar{m} \rangle
\]

\[
= B(m, w) + F(m) - \langle A^* u, m \rangle - \langle B^* u, w \rangle - \langle \lambda, m^0 - \bar{m} \rangle.
\]

Note that the linear mapping \( M \ni m \mapsto (Am, m) \in U \times \mathbb{R}^N \) is invertible as it is shown by its matrix representation (see (4.7) in the next section). As a consequence \( G \) is surjective and, hence, by standard arguments, there exists \((u, \lambda) \in U \times \mathbb{R}^N \) such that

\[
0 = \partial m_{i,j}^k L(m, w, u, \lambda) = -\frac{1}{q} \langle \frac{|w_{i,j}^{k,1}|^q}{(m_{i,j}^k)^q} f(x_{i,j}, m_{i,j}^k) \rangle - [A^* u]_{i,j}^k \quad \forall k = 1, \ldots, N_T - 1, \forall i, j,
\]

\[
0 = \partial m_{i,j}^0 L(m, w, u, \lambda) = -\lambda_{i,j} - [A^* u]_{i,j}^0 \quad \forall i, j,
\]

\[
0 = \partial \nu_{i,j} L(m, w, u, \lambda) = -\frac{1}{q} \langle \frac{|w_{i,j}^{N_T-1}|^q}{(m_{i,j}^{N_T})^q} f(x_{i,j}, m_{i,j}^{N_T}) \rangle + \frac{1}{\Delta t} g(x_{i,j}, m_{i,j}^{N_T}) - [A^* u]_{i,j}^{N_T} \quad \forall i, j,
\]

\[
0 \in \partial u_{i,j}^{k-1} \lambda L(m, w, u, \lambda) = \langle w_{i,j}^{k-1} q - \frac{w_{i,j}^{k-1}}{(m_{i,j}^{k-1})^q} - [B^* u]_{i,j}^{k-1} + N_K (w_{i,j}^{k-1}) \rangle \quad \forall k = 1, \ldots, N_T, \forall i, j,
\]

where we have used definition (3.4) and that \( m_{i,j}^{k} > 0 \) for all \( k = 1, \ldots, N_T \) and all \( i, j \). Defining \( u_{i,j}^{N_T} := g(x_{i,j}, m_{i,j}^{N_T}) \), by the last relation in (3.2), the third relation in (3.10) can be rewritten as

\[
-D_t u_{i,j}^{N_T} - \nu (\Delta_h u_{i,j}^{N_T})_{i,j} + \frac{1}{q} \langle \frac{|w_{i,j}^{N_T}|^q}{(m_{i,j}^{N_T})^q} \rangle = f(x_{i,j}, m_{i,j}^{N_T}) \quad \forall k = 0, \ldots, N_T - 1, \forall i, j,
\]

and hence, by the second relation in (3.2) and the first relation in (3.10), we have that

\[
-D_t u_{i,j}^k - \nu (\Delta_h u_{i,j}^k)_{i,j} + \frac{1}{q} \langle \frac{|w_{i,j}^k|^q}{(m_{i,j}^{k+1})^q} \rangle = f(x_{i,j}, m_{i,j}^{k+1}) \quad \forall k = 0, \ldots, N_T - 1, \forall i, j.
\]
The last relation in (3.10) yields that for all \( k = 1, \ldots, N_T \) and all \( i, j \)

\[
\begin{cases}
\frac{(m^k_i)^{q-1}}{[w^k_{i,j}]^{q-2}}[B^*u]^{k-1}_{i,j} \in w^{k-1}_{i,j} + N_K(w^{k-1}_{i,j}) & \text{if } w^{k-1}_{i,j} \neq 0, \\
[B^*u]^{k-1}_{i,j} \in N_K(0) & \text{if } w^{k-1}_{i,j} = 0,
\end{cases}
\]

which, by (3.2) and under the convention (2.3), is equivalent to

\[
(3.12) \quad w^{k-1}_{i,j} = m^k_{i,j} P_K(-[D_hu]^k_{i,j}^{-1})^\frac{q-1}{q} P_K(-[D_hu]^k_{i,j}) = m^k_{i,j} |[D_hu]^k_{i,j}| \frac{q-1}{q} |[D_hu]^k_{i,j}|.
\]

Shifting the index \( k \), the expression above yields

\[
\frac{1}{q} \frac{|w^k_{i,j}|^q}{(m^k_{i,j})^q} = \frac{1}{q} |[D_hu]^k_{i,j}|^q \quad \forall k = 0, \ldots, N_T - 1, \quad \forall i, j,
\]

which, combined with (3.11), implies the first equation in (MFG_{h,\Delta t}). The second equation in (MFG_{h,\Delta t}) is a consequence of \( A m + B w = 0 \) and the fact that (3.12) provides the identity

\[
(Bw)^k_{i,j} = -T_{i,j}(w^k, m^{k+1}) \quad \forall k = 0, \ldots, N_T - 1, \quad \forall i, j.
\]

The result follows. \( \square \)

**Remark 3.1.** The proof of the existence of solutions to (MFG_{h,\Delta t}) in Theorem 3.1 provides an alternative argument to the one in [3], based on Brouwer fixed-point theorem. The uniqueness of solutions under the assumption that \( f(x, \cdot) \) and \( q(x, \cdot) \) are increasing, with one of them being strictly increasing, is straightforward because, in this case, the cost functional in (P_{h,\Delta t}) is strictly convex.

4. A PRIMAL-DUAL ALGORITHM TO SOLVE (P_{h,\Delta t})

As discussed in [14], for solving the optimization problem

\[
(4.1) \quad \min_{y \in \mathbb{R}^N} \varphi(y) + \psi(y),
\]

and its dual

\[
(4.2) \quad \min_{\sigma \in \mathbb{R}^N} \varphi^*(-\sigma) + \psi^*(\sigma),
\]

where \( \varphi: \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\} \) and \( \psi: \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\} \) are convex l.s.c. proper functions, methods in [13, 20, 21, 22, 23] can be applied with guaranteed convergence under mild assumptions. In [14], devoted to the stationary case, the method proposed in [20] has the best performance when the viscosity parameter is small. This method is inspired by the first-order optimality conditions satisfied by a solution \((\hat{y}, \hat{\sigma})\) to

(4.1)-(4.2) under standard qualification conditions, which reads [34, Theorem 8]

\[
(4.3) \quad \begin{cases}
-\hat{\sigma} \in \partial \varphi(\hat{y}) \\
\hat{y} \in \partial \psi^*(\hat{\sigma})
\end{cases} \iff \begin{cases}
\dot{y} - \tau \hat{\sigma} \in \tau \partial \varphi(\hat{y}) + \hat{y} \\
\hat{\sigma} + \gamma \hat{y} \in \gamma \partial \psi^*(\hat{\sigma}) + \hat{\sigma}
\end{cases} \iff \begin{cases}
\text{prox}_{\tau \varphi}(\hat{y} - \tau \hat{\sigma}) = \hat{y} \\
\text{prox}_{\gamma \psi^*}(\hat{\sigma} + \gamma \hat{y}) = \hat{\sigma},
\end{cases}
\]

where \( \gamma > 0 \) and \( \tau > 0 \) are arbitrary and, given a l.s.c. convex proper function \( \phi: \mathbb{R}^N \to ]-\infty, +\infty[ \),

\[
\text{prox}_{\gamma \phi} x := \arg \min_{y \in \mathbb{R}^N} \left\{ \phi(y) + \frac{|y - x|^2}{2\gamma} \right\} = (I + \partial(\gamma \phi))^{-1}(x) \quad \forall x \in \mathbb{R}^N.
\]

Given \( \theta \in [0, 1], \tau \) and \( \gamma \) satisfying \( \tau \gamma < 1 \), and starting points \((y^0, \hat{y}^0, \sigma^0) \in \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^M\), the iterates \( \{(y^k, \sigma^k)\}_{k \in \mathbb{N}} \) generated by

\[
\begin{align*}
\sigma^{k+1} &= \text{prox}_{\gamma \psi^*}(\sigma^k + \gamma \hat{y}^k), \\
y^{k+1} &= \text{prox}_{\tau \varphi}(y^k - \tau \sigma^{k+1}), \\
\hat{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k)
\end{align*}
\]

converge to a primal-dual solution \((\hat{y}, \hat{\sigma})\) to (4.1)-(4.2) (see, e.g., [20]).

In the case under study, the equations of the time-dependent discretization are very similar to their stationary counterparts (see [14]). Specifically, the discrete linear operators \( A \) and \( B \) defined in (3.1), by an
abuse of notation, are represented by real matrices $A$ and $B$, of dimensions $(N_T \times N_h^2) \times ((N_T + 1) \times N_h^2)$ and $(N_T \times N_h^2) \times (N_T \times 4N_h^2)$, respectively, given by

$$A = \begin{pmatrix}
\frac{-1}{\Delta t} \text{Id}_{N_h^2} & \nu L + \frac{1}{\Delta t} \text{Id}_{N_h^2} & 0 & \cdots & 0 \\
0 & \frac{-1}{\Delta t} \text{Id}_{N_h^2} & \nu L + \frac{1}{\Delta t} \text{Id}_{N_h^2} & \cdots & \vdots \\
\vdots & \cdots & \cdots & \cdots & \ddots \\
0 & \cdots & 0 & \frac{-1}{\Delta t} \text{Id}_{N_h^2} & \nu L + \frac{1}{\Delta t} \text{Id}_{N_h^2}
\end{pmatrix}$$

(4.5)

and

$$B = \begin{pmatrix}
M & 0 & \cdots & 0 \\
0 & M & \cdots & 0 \\
\vdots & \cdots & \cdots & \vdots \\
0 & \cdots & 0 & M
\end{pmatrix},$$

(4.6)

where $L \in \mathcal{M}_{N_h^2 \times N_h^2}(\mathbb{R})$ is the matrix that represents $-\Delta_h$ on the torus $T_h^2$ and $M \in \mathcal{M}_{N_h^2 \times 4N_h^2}(\mathbb{R})$ is the matrix representing the discrete divergence. Denoting by $\tilde{A}$ and $\tilde{B}$ the $((N_T + 1) \times N_h^2) \times ((N_T + 1) \times N_h^2)$ and $((N_T + 1) \times N_h^2) \times (N_T \times 4N_h^2)$ real matrices

$$\tilde{A} = \begin{pmatrix}
\text{Id}_{N_h^2} & 0 & \cdots & 0 \\
A & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & \text{Id}_{N_h^2}
\end{pmatrix} \quad \text{and} \quad \tilde{B} = \begin{pmatrix}
0 & \cdots & 0 \\
B & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & M
\end{pmatrix},$$

(4.7)

the constraint $G(m, w) = (0, \bar{m})$ in $(\mathcal{P}_{h, \Delta t})$ can be rewritten as $C(m, w) = (\bar{m}, 0)$, where $C := [\tilde{A} | \tilde{B}]$.

**Remark 4.1.**

(i) The matrix $\tilde{A}$ is block lower triangular with invertible diagonal blocks and, hence, it is invertible. Indeed, the first diagonal block $\text{Id}_{N_h^2}$ is obviously invertible and the other blocks, given by $\nu L + \frac{1}{\Delta t} \text{Id}_{N_h^2}$, are also invertible because they are strictly diagonally dominant.

(ii) Since $\tilde{A}$ is invertible, the matrix

$$Q := CC^* = \tilde{A}^* + \tilde{B} \tilde{B}^*$$

is positive definite and, hence, invertible.

Therefore, $(\mathcal{P}_{h, \Delta t})$ is a particular instance of (4.1) with

$$\varphi(m, w) := B(m, w) + \mathcal{F}(m), \quad \psi(m, w) := \gamma_{\ker C + \{(m, w)\}}(m, w),$$

where $(m_f, w_f)$ is a feasible vector (provided for instance by Lemma 3.1), and $\gamma_{\ker C + \{(m_f, w_f)\}}$ is the function defined as 0 for all $(m, w) \in \ker C + \{(m_f, w_f)\}$ and $+\infty$, otherwise.

Since $\text{prox}_{\gamma\psi^*} = \text{Id} - \gamma \text{prox}_{\psi^*/\gamma} \circ (\text{Id}/\gamma) = \text{Id} - \gamma \text{prox}_{\psi} \circ (\text{Id}/\gamma)$ (see e.g. [8, Section 24.2]) and $\text{prox}_{\psi} : (m, w) \mapsto (m, w) - C^*Q^{-1}(C(m, w) - (\bar{m}, 0))$, we have

$$\text{prox}_{\gamma\psi^*} : (m, w) \mapsto C^*Q^{-1}(C(m, w) - (\bar{m}, 0)),$$

where $Q$ is defined in (4.8). By setting $\gamma^0 = (m^0, w^0), \bar{y}^0 = (\bar{m}^0, \bar{w}^0), \sigma^0 = (n^0, v^0) \in \mathbb{R}^{N_T \times N_h^2} \times \mathbb{R}^{N_T \times 4N_h^2}$, (4.4) becomes

$$
\begin{cases}
z^{[l+1]} = -Q^{-1}\left(\tilde{A}(n^{[l]} + \gamma \bar{m}^{[l]}) + \tilde{B}(\bar{v}^{[l]} + \gamma \bar{w}^{[l]}) - \gamma(\bar{m}, 0)\right), \\
n^{[l+1]} = \tilde{A}^*z^{[l+1]}, \\
\bar{v}^{[l+1]} = \text{prox}_{\gamma\varphi}(\tilde{B}^{[l]} + \gamma \bar{v}^{[l]}), \\
\bar{w}^{[l+1]} = \text{prox}_{\gamma\varphi}(\tilde{B}^{[l]} + \gamma \bar{w}^{[l]}),
\end{cases}
$$

(4.9)
and, if $\gamma \tau < 1$, the convergence of $(n[l], w[l])$ to a solution $(\bar{n}, \bar{w})$ to $(P_{h,\Delta t})$ is guaranteed together with
the convergence of $(\tilde{n}[l], \tilde{v}[l])$ to some $(\hat{n}, \hat{v})$ as $l \to \infty$. In order to compute the Lagrange multiplier $\hat{u} \in U$
which solves the first equation in (MFG$_{h,\Delta t}$), note that (3.9) can be written equivalently as
\[
\mathcal{L}(m, w, u, \lambda) := \varphi(m, w) - \langle (\lambda, u), \hat{A}m + \hat{B}w \rangle + \langle \lambda, \tilde{m} \rangle
\]
(4.10)
and the optimality condition yields
\[
\left( \begin{array}{c}
\hat{A}^* \\
\hat{B}^*
\end{array} \right) \hat{z} \in \partial \varphi(\tilde{m}, \hat{w}),
\]
where $(\tilde{m}, \hat{w})$ is the primal solution and $\hat{z} = (\hat{\lambda}, \hat{u})$. Therefore, in order to approximate $\hat{z}$, note that from
(4.9) we have
\[
\left( \begin{array}{c}
\left( m[l+1] - m[l] \right) \\
\left( \frac{m[l+1]}{\tau} - \sigma[l+1] \right)
\end{array} \right) \in \partial \varphi(m[l+1], w[l+1])
\]
and, hence, since the algorithm generates converging sequences $m[l] \to \hat{m}$ and $w[l] \to \hat{w}$ and $z[l] \to \hat{z} :=
-\bar{Q}^{-1}(\hat{A} \hat{m} + \gamma \hat{w}) + \bar{B}(\hat{v} + \gamma \hat{w}) - \gamma (\hat{m}, 0)$, the closedness of the graph of $\partial \varphi$ [8, Proposition 20.38] yields
(4.11) and, hence, a good approximation of $\hat{z}$ is $z[l]$ for $l$ large enough. For obtaining $\left[ u[l] \right]_{i,j}^N$, a good
approximation is $\left[ u[l] \right]_{i,j}^N = g(x_{i,j}, [m[l]]_{i,j}^N)$.

Remark 4.2. (i) In order to obtain an efficient algorithm, the computation of $\text{prox}_{\tau \varphi}$ in (4.9) should be
fast. A complete study of $\text{prox}_{\tau \varphi}$ is presented in [14, Section 3.2] showing that its computation depends
on the resolution of a real equation, which can be efficiently solved.
(ii) An important step in (4.9) is the efficient computation of the inverse of $Q$. Different preconditioning
strategies to tackle this issue will be presented in the following section.

5. Preconditioning strategies

At the beginning of each iteration of the primal-dual algorithm (4.9), we require the solution of a linear
system
\[
Qz = b.
\]
The purpose of this section is to discuss preconditioning strategies for the solution of this linear system.
For the stationary setting discussed in [14], the solution of such a system via direct methods as the
backslash (mldivide) command in MATLAB \(^1\) was feasible for relatively fine meshes (up to the order of
100 nodes per space dimension). However, as shown in Table 1, introducing a temporal dimension and
thus increasing the degrees of freedom to $N_T \times N_T$ significantly increases the computation time. Indeed,
the use of backslash on fine space and time grids – e.g. 128\(^2\) space grid points and 40 time steps – requires
an amount of RAM that is prohibitive on the machine used for our performance tests \(^2\), leading to “out of
memory” errors. We mitigate this problem by exploring the solution of (5.1) via preconditioned iterative
methods, which perform efficiently for finer space and time subdivisions and different viscosities.
We begin by illustrating the difficulties associated to the conditioning of the system in (4.9). Table 2 shows
the condition number of the system for different space-time discretizations and viscosity values. Without
any preconditioner, the condition numbers of different discretizations scale up to $10^6$. The same Table
shows that by selecting a suitable preconditioner, such as the modified incomplete Cholesky factorization
\(^{[11]}\) (michol in MATLAB), the conditioning of the system is improved by 4 orders of magnitude.

We have tested different choices of preconditioners and iterative methods for our problem. Since the
matrix $Q$ in our setting is sparse, symmetric, and positive-definite, we have implemented an incomplete
Cholesky factorization with diagonal scaling, a modified incomplete Cholesky factorization, and multigrid
preconditioning. As for the choice of the iterative method, our tests included both conjugate gradient
(\texttt{pcg}), and the biconjugate gradient stabilized method (\texttt{BiCGStab}). The interested reader will find in
\(^{[36], Chapters 6 and 8}\) a thorough description of the aforementioned methods, and in the Appendix of

\(^1\)\url{http://uk.mathworks.com/help/matlab/ref/mldivide.html}
\(^2\)Intel Core i7-4600U @ 2.7GHz, 16GB RAM
approximation of the solution $\bar{x}$ to the linear PDE over the grids $G_0, G_1, \ldots, G_{\ell}$, where for each $k = 0, \ldots, \ell - 1$, $G_k \subseteq G_{k+1}$, we can proceed similarly and obtain a better approximation of the solution to $A_k \bar{x}_k = b_k$. As in the previous case, we begin with the finest grid $G_\ell$ and we perform $\eta_{1}$ smoothing steps to obtain the residual $r_\ell := b_\ell - A_\ell \bar{x}_\ell$. The procedure continues until we get to the coarsest grid.
Algorithm 1 Preconditioned BiCGStab

\[ x_t \leftarrow \text{BiCGStab}(Q_t, b_t, P_L, P_R, x_0, tol) \]

**procedure** BiCGStab \((Q_t, b_t, P_L, P_R, x_0, tol)\)

\[ r_0 := p_0 := Q_t x_0 - b_t; \quad \hat{r}_0 := P_R r_0; \quad \hat{p}_0 := \langle r_0, r_0 \rangle; \quad k := 0 \]

while \(\| r_k \| \geq tol\) do

\[ v_k := Q_P R \hat{p}_k \]

\[ \hat{v}_k := P_L v_k \]

\[ \hat{\alpha}_k := \hat{p}_k / \langle \hat{v}_k, \hat{r}_0 \rangle \]

\[ s_{k+1} := r_k - \hat{\alpha}_k v_k \]

\[ \hat{s}_{k+1} := P_L s_{k+1} \]

\[ t_{k+1} := \alpha P_R \hat{s}_{k+1} \]

\[ \hat{t}_{k+1} := P_L t_{k+1} \]

\[ \hat{\omega}_{k+1} := \langle \hat{s}_{k+1}, \hat{t}_{k+1} \rangle / \langle \hat{t}_{k+1}, \hat{t}_{k+1} \rangle \]

\[ \hat{x}_{k+1} := \hat{x}_k + \hat{\alpha}_k \hat{p}_k + \hat{\omega}_{k+1} \hat{s}_{k+1} \]

\[ r_{k+1} := s_{k+1} - \hat{\omega}_{k+1} t_{k+1} \]

\[ \hat{r}_{k+1} := \hat{s}_{k+1} - \hat{\omega}_{k+1} \hat{t}_{k+1} \]

\[ \hat{\rho}_{k+1} := \langle \hat{r}_{k+1}, \hat{r}_0 \rangle \]

\[ \hat{\beta}_{k+1} := \hat{\alpha}_k / \hat{s}_{k+1} \]

\[ \hat{p}_{k+1} := \hat{r}_{k+1} + \hat{\beta}_{k+1} (\hat{p}_k - \hat{\omega}_{k+1} \hat{v}_k) \]

\[ k := k + 1 \]

return \(x_k := P_R \hat{x}_k\)

---

\(G_0\), where the solution \(e_0\) to the corresponding linear system can be found easily (typically using a direct method). Next, the solution \(e_1\) on the grid \(G_1\) is corrected with the interpolation of \(e_0\). Another post-smoothing is performed to the corrected solution on \(G_1\) and using its interpolation in the grid \(G_2\) we correct the previous solution on this grid. The smoothing, interpolation and correction iterations end once we arrive to the finest grid \(G_T\) to obtain the final approximation of \(\hat{e}_T\). The previous procedure is called a multigrid method with a \(V\)-cycle. An alternative, to obtain a more accurate solution, is to proceed as before going from \(G_\ell\) to \(G_{\ell-1}\) and then for \(k=\ell-1,\ldots,1\) to perform two consecutive coarser-grid corrections, instead of one as in the \(V\)-cycle. The resulting procedure is known as multigrid with a \(W\)-cycle. Finally, in between the \(V\)-cycle and the \(W\)-cycle, we have the \(F\)-cycle, where in the process of going from the coarsest grid to the finest one, if a grid has been reached for the first time, another correction with the coarser grids using a \(V\)-cycle is performed.

In our context, we use one cycle of the multigrid algorithm, which is a linear operator as a function of the residual on the finest grid, as a preconditioner for solving (5.1) with the BiCGStab method. Since \(Q\) is related to the finite difference discretization of the operator \(-\partial_t^2 + \nu^2 \Delta^2 - \Delta\) and \(\nu\) is not necessarily small, as in [4], it is natural to consider the refinements of the grid only in the space variable (we refer the reader to [35] for semi-coarsening multigrid methods in the context of anisotropic operators). We suppose that the spatial mesh is such that \(N_h = H 2^\ell\), with \(H > 1\) and \(\ell\) is a positive integer (in the numerical example in the next section \(H\) will be equal to 2 or 3, \(H^2\) being the number of spatial points in the coarsest grid).

Let us specify the main steps of the multigrid method we use as a preconditioner.

- **Hierarchy of Grids:** Semi-coarsened grids \(G_k\) with size \((N_T + 1) H^2 2^k\) for all \(k = 0 \ldots \ell\).
- **Cycle:** We use the \(F\)-cycle.
  - **Restriction operator:** As in [4], in order to restrict the residual on the grid \(G_k\) to the grid \(G_{k-1}\), we use the second-order operator \(R_k: \mathbb{R}^{(2^k H)^2(N_T+1)} \to \mathbb{R}^{(2^{k-1} H)^2(N_T+1)}\) defined by
    \[ (R_k X)_{i,j}^n = \frac{1}{16} \left( 4X_{2i,2j}^n + 2(X_{2i+1,2j}^n + X_{2i-1,2j}^n + X_{2i,2j+1}^n + X_{2i,2j-1}^n) \right) \]
    for \(n = 0, \ldots, N_T, i, j = 1, \ldots, 2^{k-1} H\).
  - **Interpolation operator:** We denote by \(I_k: \mathbb{R}^{(2^{k-1} H)^2(N_T+1)} \to \mathbb{R}^{(2^k H)^2(N_T+1)}\) the interpolation operator from the grid \(G_{k-1}\) to the grid \(G_k\). We have chosen a standard bilinear interpolation.
operator in the space variable, which is also a second-order operator and dual to the restriction
operator \((I_k = 4R_k^*)\). According to [12], the sum of the orders of \(R_k\) and \(I_k\) has to be at least
equal to the degree of the differential operator. In our context, both are equal to 4.

- **Linear systems on the different grids:** The linear systems are defined by the matrices

\[
Q_k := A_k A_k^* + B_k B_k^*, \quad k = 0, \ldots, \ell,
\]

where we recall that \(A_k\) and \(B_k\) are the finite difference discretizations of \(\partial_t - \nu \Delta\) and \(\operatorname{div}()\),
respectively, on the grid \(G_k\) (see (3.1)).

- **Smother:** Here we have used Gauss-Seidel iterations in the lexicographic order. There is no
reason for choosing the lexicographic order, other than its simplicity.

- **Solving the system on the coarsest grid \(G_0\):** We can use an exact solver such as *backlash* in
MATLAB. Indeed, in \(G_0\) the size of the system is really small with respect to the size of the
system on the grid \(G_\ell\) (in \(G_0\), we can even store the inverse of \(Q_0\) and inversion at this level just
becomes a matrix multiplication).

The multigrid preconditioning procedure is summarized in Algorithm 2.

**Algorithm 2** Multigrid Preconditioner for \(Q_\ell x_\ell = b_\ell\)

- \(P_L : y \mapsto \text{MultigridSolver}(\ell, 0, y, \text{cycle})\)
- \(x_\ell \leftarrow \text{BiCGStab}(Q_\ell, b_\ell, P_L, I_\ell, x_0, \text{tol})\)

**procedure** MultigridSolver\((k, x_k, b_k, \text{cycle})\)

1. if \(k = 0\) then
   - \(x_k \leftarrow Q_0^{-1} b_k\)
2. else
   - \(x_k \leftarrow \text{Perform } \eta_1\) steps of Gauss-Seidel from \(x_k\) with \(b_k\) as second member.
   - \(x_{k-1} \leftarrow 0\)
   - \(x_{k-1} \leftarrow \text{MultigridSolver}(k - 1, x_{k-1}, R_k(b_k - Q_k x_k), \text{cycle})\)
   - if cycle is W then
     - \(x_{k-1} \leftarrow \text{MultigridSolver}(k - 1, x_{k-1}, R_k(b_k - Q_k x_k), \text{cycle})\)
   - if cycle is F then
     - \(x_{k-1} \leftarrow \text{MultigridSolver}(k - 1, x_{k-1}, R_k(b_k - Q_k x_k), V)\)
   - \(x_k \leftarrow x_k + I_k x_{k-1}\)
3. \(x_k \leftarrow \text{Perform } \eta_2\) steps of Gauss-Seidel from \(x_k\) with \(b_k\) as second member.

**return** \(x_k\)

5.2. **Numerical Tests.** In this section we present a test case considered in [3], for which the stationary
solution has been computed numerically in [14] using the primal-dual algorithm presented above. The
setting is as follows: for \((x, y) \in \mathbb{R}^2\) and \(m \in \mathbb{R}_+\),

\[
f(x, y, m) = m^2 - \overline{H}(x, y), \quad \overline{H}(x, y) = \sin(2\pi y) + \sin(2\pi x) + \cos(2\pi x).
\]

We first validate the dynamic behavior of our solution. Figure 1 shows the evolution of the mass at four
different time steps. Starting from a constant initial density, the mass converges to a steady state, and
then, when \(t\) gets close to the final time \(T\), the mass is influenced by the final cost and converges to a
final state. This behavior is referred to as *turnpike phenomenon* in the literature [33]. It is illustrated by
Figure 2, which displays as a function of time \(t\) the distance of the mass at time \(t\) to the stationary state
computed as in [14].

For the multi-grid preconditioner, Table 3 shows the computation times for different discretizations.
It can be observed that finer meshes with 128\(^3\) degrees of freedom are solvable within CPU times which
outperform others methods shown in the Appendix and in [14]. Furthermore, the method is robust with
respect to different viscosities.

From Table 3 we observe that most of the computational time is used for solving the second proximal
operator (the third equality of (4.9)), which does not use a multigrid strategy but which is a pointwise
operator (see Proposition 3.1 of [14]) and thus could be fully parallelizable.
Figure 1. Evolution of the density $m$ obtained with the multi-grid preconditioner for $\nu = 0.5$, $T = 1$, $N_T = 200$ and $N_h = 128$. At $t = 0.12$ the solution is close to the solution of the associated stationary MFG.

Figure 2. Distance to the stationary solution at each time $t \in [0, T]$, for $\nu = 0.5$, $T = 2$, $N_T = 200$ and $N_h = 128$. The distance is computed using the $\ell^2$ norm. The turnpike phenomenon is observed as for a long time frame the time-dependent mass approaches the solution of the stationary MFG.
Unlike the stationary case, low viscosities seem to make the algorithm be slightly slower. However, Table 4 shows that the average number of iterations of BiCGStab stays low regardless of the viscosity. Indeed Table 3 shows that more Chambolle-Pock iterations are needed to converge. The same behavior happens when we use a direct exact solver instead of the multi-grid preconditioned BiCGStab algorithm.

(a) iterations to decrease the residual by a factor $10^{-3}$. (b) iterations to solve the system with an error of $10^{-8}$.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>32 $\times$ 32</th>
<th>64 $\times$ 64</th>
<th>128 $\times$ 128</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>1.65</td>
<td>1.86</td>
<td>2.33</td>
</tr>
<tr>
<td>0.36</td>
<td>1.62</td>
<td>1.90</td>
<td>2.43</td>
</tr>
<tr>
<td>0.2</td>
<td>1.68</td>
<td>1.93</td>
<td>2.59</td>
</tr>
<tr>
<td>0.12</td>
<td>1.84</td>
<td>2.25</td>
<td>2.65</td>
</tr>
<tr>
<td>0.046</td>
<td>1.68</td>
<td>2.05</td>
<td>2.63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>32 $\times$ 32</th>
<th>64 $\times$ 64</th>
<th>128 $\times$ 128</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>3.33</td>
<td>3.40</td>
<td>3.38</td>
</tr>
<tr>
<td>0.36</td>
<td>3.10</td>
<td>3.21</td>
<td>3.83</td>
</tr>
<tr>
<td>0.2</td>
<td>3.07</td>
<td>3.31</td>
<td>4.20</td>
</tr>
<tr>
<td>0.12</td>
<td>3.25</td>
<td>3.73</td>
<td>4.64</td>
</tr>
<tr>
<td>0.046</td>
<td>2.88</td>
<td>3.59</td>
<td>4.67</td>
</tr>
</tbody>
</table>

Table 4. Average number of iterations of the preconditioned BiCGStab with $\eta_1 = \eta_2 = 2, T = 1$ and a tolerance between two iterations of the Chambolle-Pock algorithm equal to $10^{-6}$ in normalized $\ell^2$-norm.

**Concluding Remarks.** In this work we have developed a first-order primal-dual algorithm for the solution of second-order, time-dependent mean field games. The procedure consists of: a variational formulation for the MFG, its discretization via finite differences, the application Chambolle-Pock algorithm to the resulting minimization. While this method has been studied for stationary MFG in [14], its numerical realization for time-dependent MFGs was prohibitive in terms of computing time, as the Chambolle-Pock iteration requires the solution of a large-scale linear system at each iteration. We have overcome this difficulty by studying different preconditioning strategies for the associated linear system. Overall, the multigrid preconditioner with a BiCGStab iteration performs satisfactorily for different discretizations and viscosity values.

**Acknowledgments.** Most of this work was realized while the fourth author was a postdoctoral fellow at NYU Shanghai and supported by a discretionary research fund.

During the first phase of the project, the fifth author was affiliated to the Unité de Mathématiques Pures et Appliquées (UMPA) UMR CNRS 5669, of the École Normale Supérieure de Lyon and to the Project-Team Beagle of the Inria Rhône-Alpes. He wishes to acknowledge funding within the framework of the LabeX MILYON (ANR-10-LABX-0070) of the Université de Lyon, within the program "Investissements d’Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). In addition, his participation to this project has been partially supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement...
No. 639638). His participation to this project has also been partially supported by a CEMRACS 2017 scholarship.

The sixth author thanks the support from the PGMO project VarPDEMFG and from the ANR project MFG ANR-16-CE40-0015-01.

REFERENCES


Appendix

(a) Unpreconditioned

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$N_h$</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5 \times 10^{-4}$</td>
<td>18.5</td>
<td>87.6</td>
<td>448</td>
<td>1720</td>
<td></td>
</tr>
<tr>
<td>$5 \times 10^{-3}$</td>
<td>22.1</td>
<td>98.6</td>
<td>392</td>
<td>1750</td>
<td></td>
</tr>
<tr>
<td>$5 \times 10^{-2}$</td>
<td>20.7</td>
<td>93.4</td>
<td>607</td>
<td>8240</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>24.9</td>
<td>113</td>
<td>[X]</td>
<td>[X]</td>
<td></td>
</tr>
</tbody>
</table>

(b) michol

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$N_h$</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5 \times 10^{-4}$</td>
<td>15.8</td>
<td>77.8</td>
<td>346</td>
<td>1390</td>
<td></td>
</tr>
<tr>
<td>$5 \times 10^{-3}$</td>
<td>12.4</td>
<td>80.9</td>
<td>325</td>
<td>1244</td>
<td></td>
</tr>
<tr>
<td>$5 \times 10^{-2}$</td>
<td>5.47</td>
<td>26.3</td>
<td>138</td>
<td>636</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>3.69</td>
<td>16.5</td>
<td>[X]</td>
<td>[X]</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Conjugate Gradient computation times (s). (a) Unpreconditioned. (b) Preconditioned with modified incomplete Cholesky factorization. Time discretization: $N_T = 40$. [X] indicates no convergence.

(a) Unpreconditioned

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$N_h$</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5 \times 10^{-4}$</td>
<td>2.42</td>
<td>14.2</td>
<td>69.0</td>
<td>294</td>
<td>1210</td>
<td></td>
</tr>
<tr>
<td>$5 \times 10^{-3}$</td>
<td>3.09</td>
<td>16.3</td>
<td>63.9</td>
<td>270</td>
<td>1210</td>
<td></td>
</tr>
<tr>
<td>$5 \times 10^{-2}$</td>
<td>1.41</td>
<td>12.9</td>
<td>61.3</td>
<td>389</td>
<td>5470</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>3.41</td>
<td>16.5</td>
<td>98.9</td>
<td>[X]</td>
<td>[X]</td>
<td></td>
</tr>
</tbody>
</table>

(b) michol

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$N_h$</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5 \times 10^{-4}$</td>
<td>15.7</td>
<td>80.4</td>
<td>412</td>
<td>1890</td>
<td></td>
</tr>
<tr>
<td>$5 \times 10^{-3}$</td>
<td>12.2</td>
<td>82.2</td>
<td>369</td>
<td>1650</td>
<td></td>
</tr>
<tr>
<td>$5 \times 10^{-2}$</td>
<td>5.25</td>
<td>27.2</td>
<td>174</td>
<td>894</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>3.53</td>
<td>18.8</td>
<td>122</td>
<td>2120</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. BiCGStab computation times (s). (a) Unpreconditioned. (b) Preconditioned with modified incomplete Cholesky factorization. Time discretization: $N_T = 40$. [X] indicates no convergence.