Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-30 (2022) 406–411

A policy iteration method for Mean Field Games

Fabio Camilli *

* *SBAI, Sapienza Università di Roma, Italy (e-mail: fabio.camilli@uniroma1.it).*

Abstract: The policy iteration method is a classical algorithm for solving optimal control problems. We introduce a policy iteration method for Mean Field Games systems and we prove, under a classical monotonicity assumption on the coupling cost, the convergence of this procedure to the solution of the problem.

Keywords: Mean Field Games, policy iteration, convergence rate.

1. INTRODUCTION

The policy iteration method, introduced by Bellman (1957), is an iterative procedure to solve the Hamilton-Jacobi-Bellman (HJB in short) equation. This PDE, which suffers from the so-called "curse of dimensionality", is approximated by a sequence of solutions of linear PDEs, coupled at each step with an optimization problem for the updating of the new policy. It is known that the algorithm converges (see Fleming (1963); Puterman (1981); Santos and Rust (2004)) and, in some cases, it is possible to prove a (local) quadratic rate of convergence of the method (see Bokanowski et al. (2009); Kerimkulov et al. (2020)).

Mean Field Games (MFG in short) theory, introduced in Huang et al. (2006); Lasry and Lions (2007), characterizes Nash equilibria for differential games involving a large (infinite) number of agents. The corresponding mathematical formulation leads to the study of a system of PDEs, composed by a HJB equation, characterizing the value function and the optimal control for the agents; a Fokker-Planck (FP in short) equation, governing the distribution of the population when the agents behave in an optimal way. In the case of a finite horizon problem with periodic boundary conditions, the MFG system reads as

$$\begin{cases} -\partial_t u - \Delta u + H(Du) = \sigma F[m(t)](x) & \text{in } Q \\ \partial_t m - \Delta m - \operatorname{div}(m H_p(Du)) = 0 & \text{in } Q \\ m(x,0) = m_0(x),\ u(x,T) = u_T(x) & \text{in } \mathbb{T}^d \end{cases} \tag{1}$$

where $Q := \mathbb{T}^d \times [0,T]$, $\mathbb{T}^d$ stands for the flat torus $\mathbb{R}^d/\mathbb{Z}^d$, H is a convex Hamiltonian and σ a positive parameter. Let $L(q) = \sup_{p\in\mathbb{R}^d}\{p\cdot q - H(p)\}$ be the Legendre transform of H. We consider the following policy iteration method for (1):

Policy iteration algorithm: Fixed $R > 0$ and given a bounded, measurable vector field $q^{(0)} : \mathbb{T}^d \times [0,T] \to \mathbb{R}^d$ with $\|q^{(0)}\|_{L^\infty(Q)} \le R$, iterate

(i) Solve

$$\begin{cases} \partial_t m^{(n)} - \Delta m^{(n)} - \operatorname{div}(m^{(n)} q^{(n)}) = 0, & \text{in } Q \\ m^{(n)}(x,0) = m_0(x) & \text{in } \mathbb{T}^d. \end{cases}$$

(ii) Solve

$$\begin{cases} -\partial_t u^{(n)} - \Delta u^{(n)} + q^{(n)} Du^{(n)} - L(q^{(n)}) \\ \qquad = \sigma F[m^{(n)}(t)](x) & \text{in } Q \\ u^{(n)}(x,T) = u_T(x) & \text{in } \mathbb{T}^d. \end{cases}$$

(iii) Update the policy

$$q^{(n+1)}(x,t) = \arg\max_{|q|\le R} \left\{ q \cdot Du^{(n)}(x,t) - L(q) \right\}.$$

The main advantage of the method is that, at each iteration, the linear HJB and FP equations are completely decoupled and can be quickly solved with different numerical methods.

In this paper, we will review some of the results proved in Cacace et al. (2021); Camilli and Tang (2022). The first result concerns the convergence of the policy iteration method assuming either that the Hamiltonian is convex and globally Lipschitz, hence in a setting similar to Fleming (1963); Puterman (1981), or if the Hamiltonian has polynomial growth, i.e. $H(p) \simeq |p|^\gamma$, $\gamma > 1$.

Moreover, we study the rate of convergence of method. We obtain, via purely PDE arguments, a linear rate of convergence for the solution of the MFG system. More precisely, the error between two successive iterations of the sequence $\{(u^{(n)}, m^{(n)})\}$ generated by the algorithm improves linearly with respect to the error of the previous iterations, for sufficiently large n and small σ.

Finally, we introduce a suitable finite difference approximation of the MFG system and we show the convergence of the policy iteration method for the discrete problem. Some numerical tests in dimension one and two complete the presentation. For reasons of space, we present the results only in the case of the evolutive system (1). However, similar results can also be proved in the ergodic stationary setting.

The paper is organized as follows. In Section 2, we introduce some notations and recall the convergence result in Cacace et al. (2021). In Section 3, we prove the convergence rate for the MFG policy iteration method. In Section 4 we introduce an approximation of the MFG system while in Section 5 we show some numerical tests.

10.1016/j.ifacol.2022.11.087

2. CONVERGENCE OF THE POLICY ITERATION METHOD

We denote by $L^r(\mathbb{T}^d)$, $1 \le r \le \infty$, the set of r summable periodic functions and by $W^{k,r}(\mathbb{T}^d)$, $k \in \mathbb{N}$ and $1 \le r \le \infty$, the Sobolev space of periodic functions having r-summable weak derivatives up to order k. For any $r \ge 1$, we denote by $W_r^{2,1}(Q)$ the space of functions f such that $\partial_t^\delta D_x^\beta f \in L^r(Q)$ for all multi-indices β and δ such that $|\beta| + 2\delta \le 2$. All these spaces are endowed with the corresponding standard norm.
Defined $W_s^{1,0}(Q)$ as the space of functions such that the norm

$$\|u\|_{W_s^{1,0}(Q)} := \|u\|_{L^s(Q)} + \sum_{|\beta|=1} \|D_x^\beta u\|_{L^s(Q)}$$

is finite, we denote by $\mathcal{H}_s^1(Q)$ the space of functions $u \in W_s^{1,0}(Q)$ with $\partial_t u \in (W_{s'}^{1,0}(Q))'$, where $\frac{1}{s} + \frac{1}{s'} = 1$, equipped with the natural norm

$$\|u\|_{\mathcal{H}_s^1(Q)} := \|u\|_{W_s^{1,0}(Q)} + \|\partial_t u\|_{(W_{s'}^{1,0}(Q))'} .$$

For $\alpha \in (0,1)$, we denote the classical parabolic Hölder space $C^{\alpha,\frac{\alpha}{2}}(Q)$ as the space of functions $u \in C(Q)$ such that

$$[u]_{C^{\alpha,\frac{\alpha}{2}}(Q)} = \sup_{(x_1,t_1),(x_2,t_2)\in Q} \frac{|u(x_1,t_1) - u(x_2,t_2)|}{(\mathrm{d}(x_1,x_2)^2 + |t_1 - t_2|)^{\frac{\alpha}{2}}} < \infty,$$

where $\mathrm{d}(x,y)$ stands for the geodesic distance from x to y in $\mathbb{T}^d$. If $s > d+2$, then $\mathcal{H}_s^1(Q)$ is continuously embedded onto $C^{\delta,\delta/2}(Q)$ for some $\delta \in (0,1)$ (see Cirant and Goffi (2019)).
We describe the assumptions on the data of the problem. Concerning the Hamiltonian, we consider two different settings

(H1) H is differentiable, convex and globally Lipschitz continuous, i.e. there exists a constant $R_0 > 0$ such that

$$|D_p H(p)| \le R_0 \qquad \text{for all } p \in \mathbb{R}^d .$$

(H2) H is of the form

$$H(p) = |p|^\gamma, \qquad \gamma > 1.$$

Recall that

$$H(p) = p \cdot q - L(q) \quad \text{if and only if} \quad q = D_p H(p) .$$

Concerning the coupling cost, we assume that

(F1) F maps continuously $\mathcal{P}(\mathbb{T}^d)$, the set of probability measure on $\mathbb{T}^d$, endowed with with the weak*-convergence, into a bounded subset of $C^{0,1}(\mathbb{T}^d)$. Moreover

$$\int_{\mathbb{T}^d} (F[m_1] - F[m_2]) d(m_1 - m_2) > 0 \quad \text{if } m_1 \ne m_2 \tag{2}$$

for $m_1, m_2 \in \mathcal{P}_1(\mathbb{T}^d)$.

Finally, for the initial and terminal data, we suppose that

(I) $u_T \in W^{2-\frac{2}{r},r}(\mathbb{T}^d)$, $r > d+2$;
$m_0 \in W^{1,s}(\mathbb{T}^d)$, $s > d+2$, $m_0 \ge 0$ and $\int_{\mathbb{T}^d} m_0(x)dx = 1$.

The following result states the convergence of the policy iteration method for (1) in the appropriate functional spaces.

Theorem 1. Let either **(H1)** or **(H2)**, **(F1)** and **(I)** be in force. Then, for R sufficiently large, the sequence $(u^{(n)}, m^{(n)})$, generated by the policy iteration algorithm, converges to the unique solution $(u,m) \in W_r^{2,1}(Q) \times \mathcal{H}_s^1(Q)$ of (1).

A similar convergence result also holds for a local coupling F satisfying

$$\begin{aligned} &F : \mathbb{R}^+ \to \mathbb{R} \text{ is continuous and there exists } C_F \\ &\text{such that } |F(m)| < C_F \text{ for } m \ge 0 \end{aligned} \tag{3}$$

and the monotonicity assumption (2).
Note also that, by the Sobolev embedding of $W_r^{2,1}(Q)$ in $C^{1+\alpha,\frac{1+\alpha}{2}}(Q)$ for $r > d+2$ with

$$\|u\|_{C^{1+\alpha,(1+\alpha)/2}(Q)} \le C\|u\|_{W_r^{2,1}(Q)}$$

and since $q^{(n)} = H_p(Du^{(n-1)})$, it also follows the convergence of policy $q^{(n)}$ to the optimal control $q = H_p(Du)$ in $L^\infty(Q)$ for $n \to \infty$.
The proof of Theorem 1, see Theorems 2.3 and 2.5 in-Cacace et al. (2021) for details, is based on the following a priori estimates for the solution of the linear equations involved in policy iteration method, which allows to prove compactness of the sequence $(u^{(n)}, m^{(n)})$.

Lemma 2. Given $b \in L^\infty(Q;\mathbb{R}^d)$, $f \in L^r(Q)$ and $u_T \in W^{2-\frac{2}{r},r}(\mathbb{T}^d)$ for some $r > 1$, then the problem

$$\begin{cases} -\partial_t u - \Delta u + b(x,t)Du = f(x,t) & \text{in } Q \\ u(x,T) = u_T(x) & \text{in } \mathbb{T}^d \end{cases}$$

admits a unique solution $u \in W_r^{2,1}(Q)$ and it holds

$$\|u\|_{W_r^{2,1}(Q)} \le C(\|f\|_{L^r(Q)} + \|u_T\|_{W^{2-\frac{2}{r},r}(\mathbb{T}^d)}),$$

where C depends on the norm of the coefficients as well as on r, d, T. Furthermore, if $r > d+2$ we have $Du \in C^{\alpha,\alpha/2}$ for some $\alpha \in (0,1)$.

Lemma 3. Given a bounded, measurable vector field $g : Q \to \mathbb{R}^d$ and $m_0 \in L^2(\mathbb{T}^d)$, $m_0 \ge 0$, then the problem

$$\begin{cases} \partial_t m - \Delta m - \mathrm{div}(g(x,t)m) = 0 & \text{in } Q, \\ m(x,0) = m_0(x) & \text{in } \mathbb{T}^d, \end{cases}$$

has a unique non negative solution $m \in \mathcal{H}_2^1(Q)$. Furthermore, if $m_0 \in L^s(\mathbb{T}^d)$, $s \in (1,\infty)$, then $m \in L^\infty(0,T;L^s(\mathbb{T}^d)) \cap \mathcal{H}_2^1(Q)$ and, if $m_0 \in W^{1,s}(\mathbb{T}^d)$, then

$$\|m\|_{\mathcal{H}_s^1(Q)} \le C$$

for some constant $C = C(\|g\|_{L^\infty(Q;\mathbb{R}^d)}, \|m_0\|_{W^{1,s}(\mathbb{T}^d)})$.

3. RATE OF CONVERGENCE FOR THE POLICY ITERATION METHOD

In this section, we study the rate of convergence for the policy iteration method. We replace assumption **(H1)** with

(H3) H is two times differentiable, satisfies **(H1)** and for any $S > 0$, there exists $C_S > 0$ such that

$$H_{pp}(p)q \cdot q \le C_S|q|^2 \quad \text{for any } |p| \le S,\ q \in \mathbb{R}^d.$$

and **(F1)** with

(F2) $F : \mathbb{T}^d \times L^s(\mathbb{T}^d) \to L^r(\mathbb{T}^d)$ and for all $m_1, m_2 \in \mathcal{H}_s^1(Q)$

$$\|F[m_1] - F[m_2]\|_{L^r(\mathbb{T}^d)} \le C_F\|m_1 - m_2\|_{L^s(\mathbb{T}^d)},$$

for $r, s > d+2$. Moreover F satisfies the monotonicity assumption (2).a

The following theorem gives an estimate for the rate of convergence for the policy iteration method, see Camilli and Tang (2022), Theorem 3.1.

Theorem 4. Let either **(H2)** or **(H3)**, **(F2)** and **(I)** be in force and R as in Theorem 1. Then, there exists a constant C, depending only on the data of problem, such that, if $(u^{(n)}, m^{(n)})$ is the sequence generated by the policy iteration method, we have

$$\|m^{(n+1)} - m^{(n)}\|_{C(0,T;L^s(\mathbb{T}^d))} \le C\|q^{(n+1)} - q^{(n)}\|_{L^\infty(Q)},$$
$$\|m^{(n+1)} - m^{(n)}\|_{\mathcal{H}^1_2(Q)} \le C\|q^{(n+1)} - q^{(n)}\|_{L^\infty(Q)},$$

and

$$\|u^{(n+1)} - u^{(n)}\|_{W^{2,1}_r(Q)} \le C\big(\|u^{(n)} - u^{(n-1)}\|^2_{W^{2,1}_r(Q)} + \sigma\|m^{(n+1)} - m^{(n)}\|_{C(0,T;L^s(\mathbb{T}^d))}\big).$$

A key difficulty for obtain a convergence rate using Theorem 4 is that we cannot control the constants C in the estimate of $\|m^{(n+1)} - m^{(n)}\|$ and $\|u^{(n+1)} - u^{(n)}\|$. To address this difficulty, we introduce an additional assumption on the smallness of the constant σ in the coupling cost. Note that this assumption is not needed for the convergence of the policy iteration method but only to get a linear convergence rate to the solution of MFG system in the policy iteration.

Corollary 5. Under the same assumptions of Theorem 4, there exist constants $\sigma_0 > 0$ and $0 < C^* < 1$, such that for sufficiently large n and $\forall \sigma < \sigma_0$,

$$\|u^{(n+1)} - u^{(n)}\|_{W^{2,1}_r(Q)} + \sigma\|m^{(n+1)} - m^{(n)}\|_{C(0,T;L^s(\mathbb{T}^d))} \le C^*\|u^{(n)} - u^{(n-1)}\|_{W^{2,1}_r(Q)}.$$

4. APPROXIMATION OF THE MEAN FIELD GAMES SYSTEM AND DISCRETE POLICY ITERATION METHOD

In this section, we present some details on the numerical approximation of the MFG system and we prove the convergence of the corresponding discrete policy iteration method in a simple setting. We consider the reference case of the Eikonal-diffusion HJB equation, namely we choose the Hamiltonian

$$H(x, Du) = \frac{|Du|^2}{2} - V(x) = \sup_{q\in\mathbb{R}^d}\{q\cdot Du - \frac{1}{2}|q|^2 - V(x)\},$$

where V is a given bounded potential, and we focus on the stationary ergodic problem

$$\begin{cases} -\epsilon\Delta u + H(Du) + \lambda = F(m(x)) & \text{in } \mathbb{T}^d \\ -\epsilon\Delta m - \operatorname{div}(mD_pH(Du)) = 0 & \text{in } \mathbb{T}^d \\ \displaystyle\int_{\mathbb{T}}^d m(x)dx = 1, \quad m \ge 0, \quad \int_{\mathbb{T}}^d u(x)dx = 0\,. \end{cases} \tag{4}$$

where F is a local coupling satisfying (2) and (3).

We define a grid $\mathcal{G}$ on $\mathbb{T}^d$, the vectors U, M approximating respectively u, m at the grid nodes, and the number Λ approximating the ergodic cost λ. Then, we approximate (4) by the following nonlinear problem on $\mathcal{G}$,

$$\begin{cases} -\varepsilon\Delta_\sharp U + \frac{1}{2}|D_\sharp U|^2 + \Lambda = V_\sharp + F_\sharp(M) \\ -\varepsilon\Delta_\sharp M - \operatorname{div}_\sharp(M\, D_\sharp U) = 0 \\ \displaystyle\int_\sharp M = 1\,, \quad M \ge 0\,, \quad \int_\sharp U = 0 \end{cases} \tag{5}$$

where, in order to avoid cumbersome notation, we use the symbol $\sharp$ to denote suitable discretizations of the linear differential operators, evaluations of functions at the grid nodes, and quadrature rules for the integrals. For instance, in dimension $d = 1$, given a uniform discretization of $[0, 1]$ with I nodes x_i, for $i = 0, \dots, I-1$, and space step $h = 1/I$, we have

$$(\Delta_\sharp U)_i = \frac{1}{h^2}\left(U_{[i-1]} - 2U_i + U_{[i+1]}\right),$$

$$(D_\sharp U)_i = (D_L U_i\,,\, D_R U_i) = \frac{1}{h}\left(U_i - U_{[i-1]}\,,\, U_{[i+1]} - U_i\right),$$

where the index operator $[\cdot] = \{(\cdot + I)\, mod\, I\}$ accounts for the periodic boundary conditions. Moreover, using the notation $(\cdot)^+ = \max\{\cdot, 0\}$ and $(\cdot)^- = \min\{\cdot, 0\}$, we have

$$(|D_\sharp U|^2)_i = \left(D_L U_i^+\right)^2 + \left(D_R U_i^-\right)^2,$$

$$\begin{aligned}(\operatorname{div}_\sharp(M\, D_\sharp U))_i =& \frac{1}{h}\left(M_{[i+1]} D_L U^+_{[i+1]} - M_i D_L U_i^+\right) \\ &+ \frac{1}{h}\left(M_i D_R U_i^- - M_{[i-1]} D_R U^-_{[i-1]}\right),\end{aligned}$$

$$(F_\sharp(M))_i = F(M_i)\,, \quad (V_\sharp)_i = V(x_i)\,,$$

$$\int_\sharp M = h\sum_{i=0}^{I-1} M_i\,, \quad \int_\sharp U = h\sum_{i=0}^{I-1} U_i\,.$$

It is worth noting that, at a formal level, $D_\sharp U$ acts in the scheme as a vector field with a number of components $2d$, doubled with respect to dimension d of the problem. This suggests a natural way to approximate the policy q when building the policy iteration algorithm. Indeed, given an initial guess $Q = (Q_L, Q_R) : \mathcal{G} \to \mathbb{R}^{2d}$ and using the notation $Q_\pm = (Q_L^+, Q_R^-)$, we set $Q^{(0)} = Q$ and we iterate on $k \ge 0$ the following steps:

(i) Solve

$$\begin{cases} -\varepsilon\Delta_\sharp M^{(k)} - \operatorname{div}_\sharp(M^{(k)}\, Q^{(k)}) = 0, & \text{on } \mathcal{G} \\ \displaystyle\int_\sharp M^{(k)} = 1\,, \quad M^{(k)} \ge 0. \end{cases}$$

(ii) Solve

$$\begin{cases} -\varepsilon\Delta_\sharp U^{(k)} + Q^{(k)}_\pm \cdot D_\sharp U^{(k)} + \Lambda^{(k)} \\ \qquad = \frac{1}{2}|Q^{(k)}_\pm|^2 + V_\sharp + F_\sharp(M^{(k)}) \quad \text{on } \mathcal{G} \\ \displaystyle\int_\sharp U^{(k)} = 0\,. \end{cases}$$

(iii) Update the policy

$$Q^{(k+1)} = \begin{cases} D_\sharp U^{(k)} & \text{if } |D_\sharp U^{(k)}| \le R \\ \dfrac{D_\sharp U^{(k)}}{|D_\sharp U^{(k)}|} R & \text{if } |D_\sharp U^{(k)}| > R \end{cases} \quad \text{on } \mathcal{G}. \tag{6}$$

The following theorem states the convergence of the above discrete policy iteration, in the case of a quadratic Hamiltonian and in dimension one, but the argument can be extended with similar techniques to any dimension and more general Hamiltonians.

Theorem 6. Let F be a local coupling satisfying -(2)-(3). Then, for R in (6) sufficiently large, the sequence $(U^{(k)}, \Lambda^{(k)}, M^{(k)})$, generated by the policy iteration algorithm, converges to a solution (U, Λ, M) of (5)

For the proof, we refer to (Cacace et al., 2021, Theorem 5.1)

5. NUMERICAL SIMULATIONS

In this section, we provide some numerical tests and we present a comparison with a direct Newton method for a stationary MFG system introduced in Cacace and Camilli (2016). Both algorithms are implemented in C language, employing the free library SuiteSparseQR for solving the linear systems via QR factorization. To check convergence, given a tolerance $\tau > 0$, we rely on the 2-norm of the residual for the discrete nonlinear system

$$\mathcal{F}(U,M,\Lambda) = \begin{pmatrix} -\varepsilon\Delta_\sharp U + \frac{1}{2}|D_\sharp U|^2 - V_\sharp - F_\sharp(M) + \Lambda \\ -\varepsilon\Delta_\sharp M - \mathrm{div}_\sharp(M\, D_\sharp U) \\ \int_\sharp U \\ \int_\sharp M - 1 \end{pmatrix},$$

requiring $\|\mathcal{F}(U^{(k)}, M^{(k)}, \Lambda^{(k)})\|_2 < \tau$. In the following test, we set the problem in dimension $d = 1$, with $\tau = 10^{-8}$, $\varepsilon = 0.3$, $V(x) = \sin(2\pi x) + \cos(4\pi x)$ and $F(m) = m^2$. In particular, the choice of the coupling cost satisfies the monotonicity assumption (2), ensuring uniqueness of solutions for the MFG system. Moreover, we set the initial guess for the Newton method as $U^{(0)} \equiv 0$, $M^{(0)} \equiv 1$ on $\mathcal{G}$ and $\Lambda^{(0)} = 0$, while we take the initial policy $Q^{(0)} \equiv (0,0)$ on $\mathcal{G}$ for the policy iteration algorithm.

Figure 1 shows the solution computed by the policy iteration algorithm on a grid with $|\mathcal{G}| = 200$ nodes, while in Figure 2 we compare the performace of the two methods.

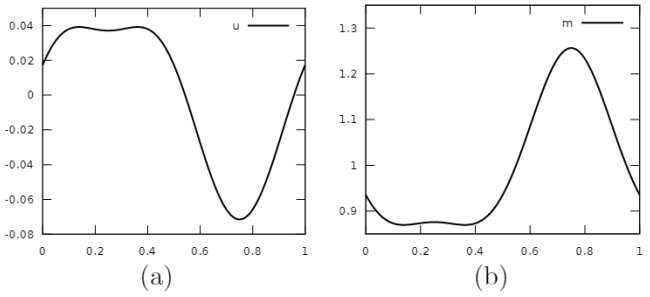

Fig. 1. Policy iteration solution for the stationary MFG system, (a) the corrector u and (b) the density m.

More precisely, in Figure 2a, we show the residuals of the two methods, against the number of iterations needed to reach the given tolerance τ. The Newton method converges in just 5 iterations, while the policy iteration requires 24 iterations. Similarly, in Figure 2b-c-d we show the differences between the solutions of the two methods in the discrete L^2 norm, against the number of iterations. Due to the particular choice of the initial guess, at the first iteration the two methods compute the same solution, but the policy iteration algorithm requires more iterations to reach the same accuracy for the residual. Nevertheless, as reported in Table 1, the policy iteration exhibits a better performance as the number of grid nodes increases, due to the reduced size of the corresponding linear systems (see the averaged CPU times per iteration). We must observe that the comparison is not truly fair, since the update step

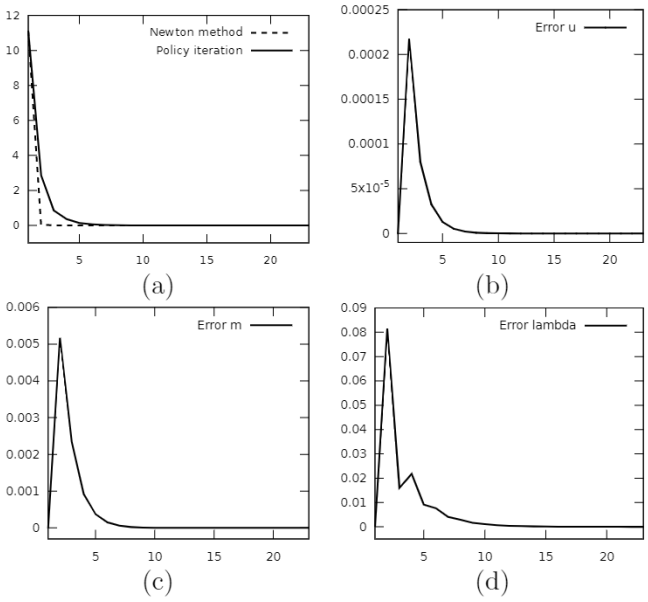

Fig. 2. Policy iteration vs Newton method, (a) MFG system residuals and (b-c-d) differences in the solutions u, m, λ.

for the policy iteration is explicit in this example, with a negligible computational cost. However, in the general case, we expect that the relevant speed-up of the proposed algorithm on large grids can compensate the efforts for the optimization process in step (iii) of the method, since it is a point-wise procedure that can be completely parallelized.

Now, let us consider the evolutive MFG system (1), again in the special case of the Eikonal-diffusion HJB equation, but in dimension $d = 2$. Spatial discretization is performed in both variables as in the one dimensional case, while, for time discretization, we employ an implicit Euler method for both the time-forward FP equation and the time-backward HJB equation. To this end, we introduce a uniform grid on the interval $[0,T]$ with $N+1$ nodes $t_n = n\,dt$, for $n = 0,\dots,N$, and time step $dt = T/N$. Then, we denote by U_n, M_n and Q_n the vectors on $\mathcal{G}$ approximating respectively the solution and the policy at time t_n. In particular, we set on $\mathcal{G}$ the initial condition $M_0 = m_0(\cdot)$ and the final condition $U_N = u_T(\cdot)$. The

	$\lvert\mathcal{G}\rvert$	Its	Av.CPU/It (secs)	Total CPU (secs)
NM	200	5	0.006	0.034
PI	200	24	0.003	0.079
NM	500	5	0.037	0.189
PI	500	25	0.009	0.247
NM	1000	5	0.173	0.865
PI	1000	25	0.036	0.917
NM	2000	5	0.973	4.869
PI	2000	25	0.241	6.039
NM	5000	5	13.662	68.313
PI	5000	25	1.724	43.115
NM	10000	5	123.769	618.845
PI	10000	25	7.917	197.949

Table 1. Policy iteration (PI) vs Newton method (NM) under grid refinement, number of iterations, averaged CPU times per iteration, and total CPU times.

policy iteration algorithm for the fully discretized system is the following: given an initial guess $Q_n^{(0)} : \mathcal{G} \to \mathbb{R}^{2d}$ for $n = 0, \dots, N$, initial and final data M_0, $U_N : \mathcal{G} \to \mathbb{R}$, iterate on $k \geq 0$ up to convergence,

(i) Solve for $n = 0, \dots, N-1$ on $\mathcal{G}$

$$\begin{cases} M_{n+1}^{(k)} - dt(\varepsilon \Delta_\sharp M_{n+1}^{(k)} + \mathrm{div}_\sharp(M_{n+1}^{(k)} Q_{n+1}^{(k)})) = M_n^{(k)} \\ M_0^{(k)} = M_0 \end{cases}$$

(ii) Solve for $n = N-1, \dots, 0$ on $\mathcal{G}$

$$\begin{cases} U_n^{(k)} - dt \left(\varepsilon \Delta_\sharp U_n^{(k)} - Q_{n,\pm}^{(k)} \cdot D_\sharp U_n^{(k)} \right) \\ \quad = U_{n+1} + dt(\frac{1}{2}|Q_{n+1,\pm}^{(k)}|^2 + V_\sharp + F_\sharp(M_{n+1}^{(k)})) \\ U_N^{(k)} = U_N \end{cases}$$

(iii) Update the policy $Q_n^{(k+1)} = D_\sharp U_n^{(k)}$ on $\mathcal{G}$ for $n = 0, \dots, N$, and set $k \leftarrow k+1$.

Note that each iteration of the algorithm now requires the solution of $2N$ linear systems of size $|\mathcal{G}| \times |\mathcal{G}|$.

In the following test, we choose a number of nodes $I = 50$ for each space dimension and $N = 100$ nodes in time, corresponding to 200 linear systems of size 2500×2500 per iteration. We set the final time $T = 1$, the diffusion coefficient $\varepsilon = 0.3$, the coupling cost $F(m) = m^2$ and the potential $V(x_1, x_2) = -|\sin(2\pi x_1)\sin(2\pi x_2)|$. Moreover, to check convergence, we rely on the discrete L^2 squared distance between policies at successive iterations, i.e. we stop the algorithm when $\max_n \int_\sharp |Q_n^{(k+1)} - Q_n^{(k)}|^2 < \tau$, setting the tolerance $\tau = 10^{-8}$. Finally, we take the initial policy $Q_n^{(0)} \equiv (0,0,0,0)$ on $\mathcal{G}$ for $n = 0, \dots, N$, while we define the initial and final data M_0 and U_N approximating on $\mathcal{G}$ the functions $m_0(x_1, x_2) = -u_T(x_1, x_2) = C \exp\left\{-40[(x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2]\right\}$, namely two Gaussian with opposite signs centered at the point $(\frac{1}{2}, \frac{1}{2})$, with $C > 0$ such that $\int_{\mathbb{T}^2} m_0(x)dx = 1$.
The algorithm requires 58 iterations to reach convergence up to τ, with an averaged CPU time per iteration of 7.3 seconds, and a total CPU time of 423 seconds. In Figure 3, we report some relevant frames of the time evolution, by plotting, for n fixed, the solution density M_n in gray scales, and superimposing the optimal dynamics for the FP equation, which is obtained by merging the two-sided components of Q_n, namely $(Q_{n,L}^1 + Q_{n,R}^1, Q_{n,L}^2 + Q_{n,R}^2)$. We remark that, by definition, the absolute minimum of the potential V is achieved at the points $(\frac{1}{4}, \frac{1}{4})$, $(\frac{3}{4}, \frac{1}{4})$, $(\frac{1}{4}, \frac{3}{4})$, $(\frac{3}{4}, \frac{3}{4})$. We observe that the optimal dynamics readily splits the density symmetrically in four parts, pushing them to concentrate around these minimizers, while, in the final part of the time interval $[0, T]$, it forces the density to merge again and concentrate exactly around the point $(1/2, 1/2)$ (i.e. the absolute minimizer of u_T), in order to to satisfy the final condition for the HJB equation. This configuration corresponds to the so called *turnpike phenomenon*. Roughly speaking, it turns out that the solution of the evolutive problem corresponds to approach the solution of the corresponding stationary ergodic problem, standing on this equilibrium as long as possible before moving again towards u_T.

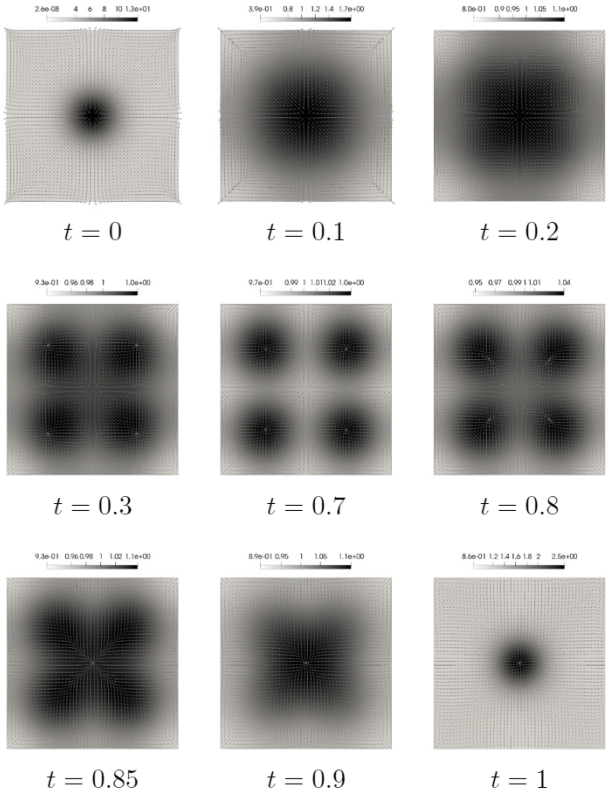

Fig. 3. Solution of the evolutive MFG system at different times, mass density in gray scales and optimal dynamics.

6. CONCLUSION

We presented a policy iteration algorithm for MFG systems, discussing its convergence and showing some error estimates. In this paper we restrict the discussion to MFGs with separable Hamiltonians. Recently in Laurière et al. (2021) the authors considered the convergence rate of policy iteration algorithms for MFGs with non-separable Hamiltonians using contraction fixed point method. In the future, we plan to extend this approach to other types of MFG systems, such as those associated with MFG of Controls.
Another possible direction of research would be to consider a modified version of fictitious play for mean field games. It can be shown to be connected with policy iteration: instead of updating $m^{(n)}$, it is averaged with previous steps of $m^{(n)}$. Fictitious play type method has been especially important for considering multi-agent reinforcement learning models as MFGs.

REFERENCES

Bellman, R., 1957. Dynamic programming. Princeton University Press, Princeton, N. J.

Bokanowski, O., Maroso, S., Zidani, H., 2009. Some convergence results for Howard's algorithm. SIAM J. Numer. Anal. 47 (4), 3001–3026.

Cacace, S., Camilli, F., 2016. A generalized Newton method for homogenization of Hamilton-Jacobi equations. SIAM J. Sci. Comput. 38 (6), A3589–A3617.

Cacace, S., Camilli, F., Goffi, A., 2021. A policy iteration method for mean field games. ESAIM Control Optim. Calc. Var. 27, Paper No. 85, 19.

Camilli, F., Tang, Q., 2022. Rates of convergence for the policy iteration method for mean field games systems. J. Math. Anal. Appl. 512 (1), Paper No. 126138, 18.

Cirant, M., Goffi, A., 2019. On the existence and uniqueness of solutions to time-dependent fractional MFG. SIAM J. Math. Anal. 51 (2), 913–954.

Fleming, W. H., 1963. Some Markovian optimization problems. J. Math. Mech. 12, 131–140.

Huang, M., Malhamé, R. P., Caines, P. E., 2006. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. Commun. Inf. Syst. 6 (3), 221–251.

Kerimkulov, B., Šiška, D., Szpruch, L., 2020. Exponential convergence and stability of Howard's policy improvement algorithm for controlled diffusions. SIAM J. Control Optim. 58 (3), 1314–1340.

Lasry, J.-M., Lions, P.-L., 2007. Mean field games. Jpn. J. Math. 2 (1), 229–260.

Laurière, M., Song, J., Tang, Q., 2021. Policy iteration method for time-dependent mean field games systems with non-separable hamiltonians. arXiv:2110.02552.

Puterman, M. L., 1981. On the convergence of policy iteration for controlled diffusions. J. Optim. Theory Appl. 33 (1), 137–144.

Santos, M. S., Rust, J., 2004. Convergence properties of policy iteration. SIAM J. Control Optim. 42 (6), 2094–2115.