



**HAL**  
open science

# Forward-backward algorithm for functions with locally Lipschitz gradient: applications to mean field games

Luis M. Briceno-Arias, Francisco José Silva, Xianjin Yang

► **To cite this version:**

Luis M. Briceno-Arias, Francisco José Silva, Xianjin Yang. Forward-backward algorithm for functions with locally Lipschitz gradient: applications to mean field games. 2023. hal-04137639

**HAL Id: hal-04137639**

**<https://hal.science/hal-04137639>**

Preprint submitted on 22 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Forward-backward algorithm for functions with locally Lipschitz gradient: applications to mean field games.

Luis M. Briceño-Arias<sup>1†</sup>, Francisco J. Silva<sup>2\*</sup>, Xianjin Yang<sup>3†</sup>

<sup>1</sup>Departamento de Matemática, Universidad Técnica Federico Santa María, Santiago, Chile.

<sup>2</sup>Laboratoire XLIM, Université de Limoges, Limoges, 87060, France.

<sup>3</sup>Computing and Mathematical Sciences, California Institute of Technology, California, USA.

\*Corresponding author(s). E-mail(s): [francisco.silva@unilim.fr](mailto:francisco.silva@unilim.fr);  
Contributing authors: [luis.briceno@usm.cl](mailto:luis.briceno@usm.cl); [yxjmath@caltech.edu](mailto:yxjmath@caltech.edu);

†These authors contributed equally to this work.

## Abstract

In this paper, we provide a generalization of the forward-backward splitting algorithm for minimizing the sum of a proper convex lower semicontinuous function and a differentiable convex function whose gradient satisfies a locally Lipschitz-type condition. We prove the convergence of our method and derive a linear convergence rate when the differentiable function is locally strongly convex. We recover classical results in the case when the gradient of the differentiable function is globally Lipschitz continuous and an already known linear convergence rate when the function is globally strongly convex. We apply the algorithm to approximate equilibria of variational mean field game systems with local couplings. Compared with some benchmark algorithms to solve these problems, our numerical tests show similar performances in terms of the number of iterations but an important gain in the required computational time.

**Keywords:** Constrained convex optimization, Forward-Backward splitting, Locally Lipschitz gradient, Mean field games.

**MSC Classification:** 65K05 , 90C25 , 90C90 , 91-08 , 91A16 , 49N80 , 35Q89.

# 1 Introduction

In this paper, we aim to solve the following problem.

**Problem 1.** Let  $\mathcal{H}$  be a real Hilbert space, let  $\psi: \mathcal{H} \rightarrow ]-\infty, +\infty]$  be a proper lower semicontinuous convex function, and let  $\varphi: \mathcal{H} \rightarrow \mathbb{R}$  be a convex Gâteaux differentiable function such that, for every  $x \in \mathcal{H}$  and every  $M \in ]0, +\infty[$ , there exist  $\mu_M(x) \in [0, +\infty[$  and  $L_M(x) \in [\mu_M(x), +\infty[$  such that, for all  $y, z \in B_M(x)$ ,

$$\mu_M(x)\|y - z\|^2 \leq \langle y - z \mid \nabla\varphi(y) - \nabla\varphi(z) \rangle \leq L_M(x)\|y - z\|^2, \quad (1)$$

where  $B_M(x)$  is the open ball centered at  $x$  with radius  $M$ . The problem is to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \psi(x) + \varphi(x), \quad (2)$$

under the assumption that the set of solutions to (2), denoted by  $S \subset \mathcal{H}$ , is nonempty.

This problem appears in several domains, including mean field games [18], optimal transport problems [12, 48], image and signal processing [19, 29], control theory [8], among others (see also [29] and the references therein for more applications). In the particular case when  $\nabla\varphi$  is globally Lipschitz continuous, a standard algorithm for solving (2) is the *forward-backward splitting* (FBS), which finds its roots in the projected gradient method [41] (case  $\psi = \iota_C$  for some nonempty closed convex set  $C$ ). In the context of variational inequalities appearing in some PDEs, a generalization of the projected gradient method is proposed in [14, 44, 53]. FBS combines a gradient step (forward) on  $\varphi$  and a proximal (backward) step on  $\psi$ . More precisely, given  $x_0 \in \mathcal{H}$ , FBS iterates

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \text{prox}_{\gamma\psi}(x_n - \gamma\nabla\varphi(x_n)), \quad (3)$$

where  $\gamma > 0$  is known as the step-size and, for every  $f \in \Gamma_0(\mathcal{H})$ ,  $\text{prox}_f: \mathcal{H} \rightarrow \mathcal{H}$  assigns to every  $x \in \mathcal{H}$  the unique solution to the lower semicontinuous strongly convex function  $f + \|\cdot - x\|^2/2$ . The weak convergence of the sequence generated by FBS to a solution to (2) is guaranteed provided that  $\gamma \in ]0, 2/L[$ , where  $L > 0$  is the globally Lipschitz constant of  $\nabla\varphi$  (see, e.g., [9, Theorem 23.14]). One of the central arguments to prove the convergence is Baillon-Haddad's theorem [7]. It asserts that globally Lipschitz continuous gradients of convex functions are cocoercive, from which it is proved that  $\{x_n\}_{n \in \mathbb{N}}$  is a Féjerian sequence, i.e., its distance to any solution is decreasing with  $n \in \mathbb{N}$ . If, in addition,  $\varphi$  is strongly convex, FBS converges linearly to the unique solution to (2) [19, 54]. However, the globally Lipschitz continuity on  $\nabla\varphi$  is quite restrictive in applications.

An approach to weaken the globally Lipschitz continuity of  $\nabla\varphi$  is to use linesearch procedures to compute the step-size at each iteration of FBS (see, e.g., [10, 11, 52] and the references therein). In this setting, the convergence is then ensured under weaker conditions on  $\nabla\varphi$  as, e.g., the uniform continuity on weakly compact sets [52, Theorem 3.18]. However, each evaluation made in the linesearch procedure can be costly, e.g., when the proximity operator of  $\psi$  is not simple to compute. Moreover, depending on the linesearch parameters, the resulting step-size can be very small, affecting the efficiency of the algorithm.

In this paper, we provide a new approach to guarantee the convergence of FBS when  $\nabla\varphi$  satisfies (1). Our approach relies on a recent refinement of Baillon-Haddad's theorem for convex sets [49], which allows us to use the cocoercivity property of  $\nabla\varphi$  in balls. This permits us to use similar arguments than in the globally Lipschitz continuous case to prove the convergence of FBS under the assumption (1). In addition, we can estimate an upper bound for the step-size and a linear convergence rate in the presence of strong convexity, as in the globally Lipschitz case. We also recover the classic convergence result for FBS when  $\nabla\varphi$  is globally Lipschitz continuous, as well as the linear convergence rate in [19, 54] when, in addition,  $\varphi$  is strongly convex.

Another contribution of this paper is the application of the proposed algorithm to approximate equilibria of variational Mean Field Games (MFGs) with local couplings. The main purpose of MFGs theory, introduced independently by Lasry and Lions in [37–39], and by Caines, Huang, and Malhamé in [35, 36], is to describe the asymptotic behaviour of Nash equilibria of non-cooperative symmetric dynamic differential games with a large number of indistinguishable players. We refer to [4, 24, 25, 32, 34], and the references therein, for a general overview on MFGs theory including analytic and probabilistic aspects, as well as their numerical approximation and applications in crowd motion, economics, and finance. A particular class of MFGs, called *variational* MFGs, characterizes the aforementioned Nash equilibria in terms of the first order optimality condition of an associated optimization problem. This viewpoint opens the door to the application of variational techniques to establish the existence of MFG equilibria [21, 23, 45, 46] and to approximate them numerically by using convex optimization methods [1, 13, 15, 16, 18, 40, 43].

In the framework of ergodic and variational MFGs with monotone local couplings, we consider the finite difference discretization introduced in [6]. Under assumptions ensuring the existence of a unique classical solution to the MFGs system, it is shown in [2] that the solution to the finite difference scheme converges as the discretization step tends to zero. Let us also mention the contribution [5] dealing with the convergence of solutions to the scheme in the framework of weak solutions. It turns out that this finite difference discretization preserves the variational structure, i.e. it corresponds to the optimality condition of a convex optimization problem. We apply the forward-backward algorithm to two dual formulations of this problem and we compare their performance in the case of a first order system involving a logarithmic coupling and which admits an explicit solution (see [6]). In the case of second order ergodic MFGs with power and logarithmic couplings, the forward-backward method is compared with state-of-the-art algorithms and we show, in this particular instance, a similar behavior in terms of the numbers of iterations but an important improvement in terms of computational time.

The remainder of the article is organized as follows. Section 2 reviews the necessary notation and some background in convex analysis for later use. We present in Section 3 our main result on the global convergence of the forward-backward algorithm. Section 4 recalls the ergodic MFG system with local couplings, its finite difference approximation, and its variational formulation. We also compute two dual formulations for which the forward-backward algorithm will be applied in Section 5, devoted to

numerical tests and comparisons with other benchmark algorithms in terms of number of iterations and computational time.

## 2 Preliminaries

Throughout this paper,  $\mathcal{H}$  is a real Hilbert space endowed with the inner product  $\langle \cdot | \cdot \rangle$  and associated norm  $\|\cdot\|$ . The weak and the strong convergences are denoted by  $\rightharpoonup$  and  $\rightarrow$ , respectively. Given  $x \in \mathcal{H}$  and  $M > 0$ , the open and closed ball centered at  $x$  with radius  $M$  are denoted by  $B_M(x)$  and  $\overline{B}_M(x)$ , respectively. Let  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$ . The domain of  $f$  is  $\text{dom } f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$  and  $f$  is proper if  $\text{dom } f \neq \emptyset$ . Denote by  $\Gamma_0(\mathcal{H})$  the class of proper lower semicontinuous convex functions from  $\mathcal{H}$  to  $] - \infty, +\infty]$ . Suppose that  $f \in \Gamma_0(\mathcal{H})$ . The Fenchel conjugate of  $f$  is

$$f^* : \mathcal{H} \rightarrow ]-\infty, +\infty] : u \mapsto \sup_{x \in \mathcal{H}} (\langle x | u \rangle - f(x)). \quad (4)$$

We have  $f^* \in \Gamma_0(\mathcal{H})$  and  $f^{**} = f$ . The subdifferential of  $f$  is the set-valued operator

$$\partial f : \mathcal{H} \rightarrow 2^{\mathcal{H}} : x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H}) \quad \langle y - x | u \rangle + f(x) \leq f(y)\} \quad (5)$$

and  $\text{dom } \partial f = \{x \in \mathcal{H} \mid \partial f(x) \neq \emptyset\}$ . The *proximity operator* of  $f$  is

$$\text{prox}_f : \mathcal{H} \rightarrow \mathcal{H} : x \mapsto \arg \min_{y \in \mathcal{H}} \left( f(y) + \frac{1}{2} \|x - y\|^2 \right), \quad (6)$$

which, by the Fermat's rule [9, Theorem 16.3], is characterized by

$$(\forall x \in \mathcal{H})(\forall p \in \mathcal{H}) \quad p = \text{prox}_f x \quad \Leftrightarrow \quad x - p \in \partial f(p). \quad (7)$$

For every  $\gamma \in ]0, +\infty[$ , [9, Theorem 14.3(ii)] implies that

$$\text{prox}_{\gamma f} = \text{Id} - \gamma \text{prox}_{f^*/\gamma} \circ (\text{Id} / \gamma), \quad (8)$$

where  $\text{Id} : \mathcal{H} \rightarrow \mathcal{H}$  denotes the identity operator. Moreover, by [9, Proposition 12.28],  $\text{prox}_{\gamma f}$  is firmly nonexpansive, i.e., for all  $x, y \in \mathcal{H}$ ,

$$\|\text{prox}_{\gamma f} x - \text{prox}_{\gamma f} y\|^2 + \|(\text{Id} - \text{prox}_{\gamma f})x - (\text{Id} - \text{prox}_{\gamma f})y\|^2 \leq \|x - y\|^2. \quad (9)$$

In addition,  $f$  is supercoercive if

$$\lim_{|x| \rightarrow +\infty} f(x)/|x| = +\infty. \quad (10)$$

It follows from [9, Proposition 14.15] and [9, Proposition 18.9] that if  $f$  is supercoercive and strictly convex, then

$$\text{dom } f^* = \mathcal{H} \text{ and } f^* \text{ is G\^ateaux differentiable on } \mathcal{H}. \quad (11)$$

Let  $\mathcal{O} \subset \mathcal{H}$  be an open convex set and let  $\beta \in ]0, +\infty[$ . Then,  $f$  is  $\beta$ -strongly convex on  $\mathcal{O}$  if  $f - \beta\|\cdot\|^2/2$  is convex on  $\mathcal{O}$ . Now, suppose that  $f$  is Gâteaux differentiable on  $\mathcal{O}$ . We say that  $\nabla f$  is  $\beta$ -Lipschitz continuous on  $\mathcal{O}$  if

$$(\forall x \in \mathcal{O})(\forall y \in \mathcal{O}) \quad \|\nabla f(x) - \nabla f(y)\| \leq \beta\|x - y\| \quad (12)$$

and it is  $1/\beta$ -cocoercive on  $\mathcal{O}$  if

$$(\forall x \in \mathcal{O})(\forall y \in \mathcal{O}) \quad \langle \nabla f(x) - \nabla f(y) \mid x - y \rangle \geq \frac{1}{\beta} \|\nabla f(x) - \nabla f(y)\|^2. \quad (13)$$

A crucial result in our theorem is the following enhanced version of the Baillon-Haddad theorem [7].

**Theorem 2.** [49, Theorem 3.1] *Let  $\mathcal{O} \subset \mathcal{H}$  be a nonempty convex open set, let  $f: \mathcal{O} \rightarrow \mathbb{R}$  be a lower semicontinuous convex function, and let  $\beta \in ]0, +\infty[$ . Then the following are equivalent:*

- (i)  *$f$  is Gâteaux differentiable in  $\mathcal{O}$  and  $\nabla f$  is  $\beta$ -Lipschitz continuous on  $\mathcal{O}$ .*
- (ii)  *$\frac{\beta}{2}\|\cdot\|^2 - f$  is convex on  $\mathcal{O}$ .*
- (iii)  *$f$  is Gâteaux differentiable in  $\mathcal{O}$  and  $\nabla f$  is  $1/\beta$ -cocoercive on  $\mathcal{O}$ .*

Let  $C \subset \mathcal{H}$  be a nonempty closed convex set. Then

$$C^- = \{u \in \mathcal{H} \mid (\forall x \in C) \langle x \mid u \rangle \leq 0\} \quad (14)$$

is the polar cone to  $C$ ,

$$\iota_C : \mathcal{H} \rightarrow ]-\infty, +\infty] : x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C \end{cases} \quad (15)$$

is the indicator function of  $C$ , and

$$\sigma_C : \mathcal{H} \rightarrow ]-\infty, +\infty] : u \mapsto \sup_{x \in C} \langle x \mid u \rangle \quad (16)$$

is the support function of  $C$ , which satisfies  $\sigma_C = (\iota_C)^*$ . We also denote by  $P_C = \text{prox}_{\iota_C}$  the projection operator onto  $C$  and by  $d_C(x) = \|x - P_C x\|$  the distance of  $x \in \mathcal{H}$  to  $C$ .

For further background on convex analysis in Hilbert spaces, the reader is referred to [9].

### 3 Forward-backward algorithm for locally Lipschitz functions

We start with some remarks concerning Problem 1.

**Remark 1.** *Consider the setting of Problem 1.*

(i) Without loss of generality, we assume that, for every  $x \in \mathcal{H}$ ,

$$0 < M' \leq M \quad \Rightarrow \quad [L_{M'}(x) \leq L_M(x) \text{ and } \mu_{M'}(x) \geq \mu_M(x)]. \quad (17)$$

(ii) Note that (1) is equivalent to the  $\mu_M(x)$ -strong convexity of  $\varphi$  on  $B_M(x)$  and the Lipschitz continuity of  $\nabla\varphi$  on  $B_M(x)$ . Indeed, observe that [9, Proposition 17.7(iii)] implies that the first inequality in (1) is equivalent to the convexity of  $\varphi - \mu_M(x)\|\cdot\|^2/2$  on  $B_M(x)$  and, thus, the  $\mu_M(x)$ -strong convexity of  $\varphi$  on  $B_M(x)$  (in the case  $\mu_M(x) = 0$  it reduces to convexity). Similarly, the second inequality in (1) is equivalent to the convexity of  $L_M(x)\|\cdot\|^2/2 - \varphi$  on  $B_M(x)$ . Hence, it follows from Theorem 2 that this is equivalent to the  $L_M^{-1}(x)$ -cocoercivity of  $\nabla\varphi$  on  $B_M(x)$  or the Lipschitz continuity of  $\nabla\varphi$  on  $B_M(x)$ .

Now we characterize the solution set as fixed points of a suitable operator.

**Proposition 3.** *In the context of Problem 1,*

$$(\forall x^* \in \mathcal{H}) \quad x^* \in S \Leftrightarrow (\forall \gamma > 0) \quad x^* = \text{prox}_{\gamma\psi}(x^* - \gamma\nabla\varphi(x^*)). \quad (18)$$

Moreover, if we have  $\mu_M(x^*) > 0$  for some  $x^* \in S$  and  $M > 0$ , then the solution is unique.

*Proof.* The first assertion follows from the convexity of  $\psi + \varphi$ , Fermat's rule,  $\text{dom } \varphi = \mathcal{H}$ , and (7). Now suppose that there exist  $x^* \in S$  and  $M > 0$  such that  $\mu_M(x^*) > 0$  and that there exists  $y^* \in S$  such that  $y^* \neq x^*$ . Then, it follows from the convexity of  $\psi + \varphi$  that  $z^* = x^* + \frac{M(y^* - x^*)}{2\|y^* - x^*\|} \in B_M(x^*) \cap S$ . Moreover, it follows from [9, Theorem 17.10(iii)] that first inequality in (1) implies that  $\varphi|_{B_M(x^*)}$  is strictly convex and, thus,  $(\psi + \varphi)|_{B_M(x^*)}$  is strictly convex. Hence, since  $x^* \neq z^*$  are in  $S \cap B_M(x^*)$ , we obtain a contradiction and the uniqueness follows.  $\square$

**Proposition 4.** *In the context of Problem 1, set*

$$(\forall \gamma > 0) \quad T_\gamma: x \mapsto \text{prox}_{\gamma\psi}(x - \gamma\nabla\varphi(x)), \quad (19)$$

let  $x \in \mathcal{H}$ , let  $M > 0$ , and let  $\gamma \in ]0, 2/L_M(x)[$ . Then, the following hold:

(i) For all  $y, z \in B_M(x)$  we have

$$\begin{aligned} \|T_\gamma y - T_\gamma z\|^2 &\leq \|y - z\|^2 - \gamma \left( \frac{2}{L_M(x)} - \gamma \right) \|\nabla\varphi(y) - \nabla\varphi(z)\|^2 \\ &\quad - \|y - T_\gamma y - (z - T_\gamma z) - \gamma(\nabla\varphi(y) - \nabla\varphi(z))\|^2. \end{aligned} \quad (20)$$

(ii) Suppose that  $\mu_M(x) > 0$ . Then, for all  $y, z \in B_M(x)$ , we have

$$\|T_\gamma y - T_\gamma z\| \leq \rho_M(x, \gamma)\|y - z\|, \quad (21)$$

where

$$\rho_M: (x, \gamma) \mapsto \max\{|1 - \gamma\mu_M(x)|, |1 - \gamma L_M(x)|\} \in ]0, 1[. \quad (22)$$

*Proof.* Define, for every  $\mu \in [0, \mu_M(x)]$ ,

$$\phi_\mu : y \mapsto \varphi(y) - \frac{\mu}{2} \|y\|^2. \quad (23)$$

Since  $\varphi$  is Gâteaux differentiable, then  $\phi_\mu$  is Gâteaux differentiable and  $\nabla\phi_\mu = \nabla\varphi - \mu \text{Id}$ . Hence, we obtain from (1) that, for all  $y, z \in B_M(x)$ ,

$$0 \leq \langle y - z \mid \nabla\phi_\mu(y) - \nabla\phi_\mu(z) \rangle \leq (L_M(x) - \mu) \|y - z\|^2. \quad (24)$$

Therefore, it follows from [9, Proposition 17.7(iii)] that  $\phi_\mu|_{B_M(x)}$  and  $((L_M(x) - \mu) \|\cdot\|^2/2 - \phi_\mu)|_{B_M(x)}$  are convex. Hence, Theorem 2 implies that, for all  $y, z \in B_M(x)$ ,

$$\langle y - z \mid \nabla\phi_\mu(y) - \nabla\phi_\mu(z) \rangle \geq \frac{1}{L_M(x) - \mu} \|\nabla\phi_\mu(y) - \nabla\phi_\mu(z)\|^2. \quad (25)$$

Now, fix  $y$  and  $z$  in  $B_M(x)$  and let  $\gamma > 0$ . It follows from (9), the identity  $\text{Id} - \gamma\nabla\varphi = (1 - \gamma\mu)\text{Id} - \gamma\nabla\phi_\mu$ , and (25) that

$$\begin{aligned} & \|T_\gamma y - T_\gamma z\|^2 \\ & \leq \|y - z - \gamma(\nabla\varphi(y) - \nabla\varphi(z))\|^2 - \|y - T_\gamma y - (z - T_\gamma z) - \gamma(\nabla\varphi(y) - \nabla\varphi(z))\|^2 \\ & = \|(1 - \gamma\mu)(y - z) - \gamma(\nabla\phi_\mu(y) - \nabla\phi_\mu(z))\|^2 \\ & \quad - \|y - T_\gamma y - (z - T_\gamma z) - \gamma(\nabla\varphi(y) - \nabla\varphi(z))\|^2 \\ & = (1 - \gamma\mu)^2 \|y - z\|^2 - 2\gamma(1 - \gamma\mu) \langle y - z \mid \nabla\phi_\mu(y) - \nabla\phi_\mu(z) \rangle \\ & \quad + \gamma^2 \|\nabla\phi_\mu(y) - \nabla\phi_\mu(z)\|^2 - \|y - T_\gamma y - (z - T_\gamma z) - \gamma(\nabla\varphi(y) - \nabla\varphi(z))\|^2 \\ & \leq (1 - \gamma\mu)^2 \|y - z\|^2 \\ & \quad - (2\gamma(1 - \gamma\mu) - \gamma^2(L_M(x) - \mu)) \langle y - z \mid \nabla\phi_\mu(y) - \nabla\phi_\mu(z) \rangle \\ & \quad - \|y - T_\gamma y - (z - T_\gamma z) - \gamma(\nabla\varphi(y) - \nabla\varphi(z))\|^2 \\ & = (1 - \gamma\mu)^2 \|y - z\|^2 - \gamma(2 - \gamma(L_M(x) + \mu)) \langle y - z \mid \nabla\phi_\mu(y) - \nabla\phi_\mu(z) \rangle \\ & \quad - \|y - T_\gamma y - (z - T_\gamma z) - \gamma(\nabla\varphi(y) - \nabla\varphi(z))\|^2. \quad (26) \end{aligned}$$

(i): Since (26) holds for every  $\mu \in [0, \mu_M(x)]$ , by setting  $\mu = 0$  we have  $\nabla\phi_\mu = \nabla\varphi$  and the result follows from (25) and (26).

(ii): We have two cases.

- Suppose that  $\gamma \in ]0, 2/(L_M(x) + \mu)[$ : Since the convexity of  $\phi_\mu$  implies  $\langle y - z \mid \nabla\phi_\mu(y) - \nabla\phi_\mu(z) \rangle \geq 0$ , we deduce from (26) that

$$\|T_\gamma y - T_\gamma z\| \leq |1 - \gamma\mu| \|y - z\|. \quad (27)$$

- Suppose that  $\gamma \in ]2/(L_M(x) + \mu), 2/L_M(x)[$ : It follows from (24) and (26) that

$$\|T_\gamma y - T_\gamma z\|^2 \leq ((1 - \gamma\mu)^2 + \gamma(L_M(x) - \mu)(\gamma(L_M(x) + \mu) - 2)) \|y - z\|^2$$



$$= (1 - \gamma L_M(x))^2 \|y - z\|^2. \quad (28)$$

Since previous results hold for every  $\mu \in [0, \mu_M(x)]$ , by setting  $\mu = \mu_M(x)$ , the result follows from (27) and (28).  $\square$

Now we state our main result.

**Theorem 5.** *In the context of Problem 1, let  $x_0 \in \mathcal{H}$ , let  $\varepsilon > 0$ , set  $M_0 := \sup_{x^* \in S} \|x^* - x_0\| + \varepsilon$ , and suppose that  $\mathcal{L}_{M_0} := \sup_{x^* \in S} L_{M_0}(x^*) \in ]0, +\infty[$ . Let  $\gamma \in ]0, 2/\mathcal{L}_{M_0}[$ , and consider the routine*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \text{prox}_{\gamma\psi}(x_n - \gamma \nabla \varphi(x_n)). \quad (29)$$

Then, the following holds:

- (i) There exists  $x^* \in S$  such that  $(x_n)_{n \in \mathbb{N}} \subset B_{M_0}(x^*)$  and  $x_n \rightarrow x^*$ .
- (ii) Suppose that there exists  $x^* \in S$  such that  $\mu_{M_0}(x^*) > 0$ . Then,  $S = \{x^*\}$ ,  $(x_n)_{n \in \mathbb{N}} \subset B_{M_0}(x^*)$ , and

$$(\forall n \in \mathbb{N}) \quad \|x_n - x^*\| \leq (\rho_{M_0}(x^*, \gamma))^n \|x_0 - x^*\|, \quad (30)$$

where  $\rho_{M_0}(x^*, \gamma) \in ]0, 1[$  is defined in (22).

*Proof.* Fix  $x^* \in S$ . We have  $\gamma \in ]0, 2/L_{M_0}(x^*)[$  and Proposition 3 yields  $x^* = T_\gamma x^*$ . Let us first prove by recurrence that  $(x_n)_{n \in \mathbb{N}} \subset \overline{B}_{M_0 - \varepsilon}(x^*)$ . Indeed, since  $\|x_0 - x^*\| \leq \sup_{y^* \in S} \|x_0 - y^*\| = M_0 - \varepsilon$ ,  $x_0 \in \overline{B}_{M_0 - \varepsilon}(x^*)$ . Suppose that  $x_n \in \overline{B}_{M_0 - \varepsilon}(x^*)$  for some  $n \in \mathbb{N}$ . Then, it follows from (19) and Proposition 4(i) that

$$\|x_{n+1} - x^*\| = \|T_\gamma x_n - T_\gamma x^*\| \leq \|x_n - x^*\| \leq M_0 - \varepsilon \quad (31)$$

and, hence,  $x_{n+1} \in \overline{B}_{M_0 - \varepsilon}(x^*)$ . Therefore, we conclude  $(x_n)_{n \in \mathbb{N}} \subset \overline{B}_{M_0 - \varepsilon}(x^*)$  and Proposition 4(i) yields, for every  $n \in \mathbb{N}$ ,

$$\begin{aligned} \|x_{n+1} - x^*\|^2 &= \|T_\gamma x_n - T_\gamma x^*\|^2 \\ &\leq \|x_n - x^*\|^2 - \gamma(2/L_{M_0}(x^*) - \gamma) \|\nabla \varphi(x_n) - \nabla \varphi(x^*)\|^2 \\ &\quad - \|x_n - x_{n+1} - \gamma(\nabla \varphi(x_n) - \nabla \varphi(x^*))\|^2. \end{aligned} \quad (32)$$

We deduce from [9, Lemma 5.31] that  $(\|x_n - x^*\|)_{n \in \mathbb{N}}$  is convergent and that

$$\sum_{n \in \mathbb{N}} \|\nabla \varphi(x_n) - \nabla \varphi(x^*)\|^2 < +\infty \quad (33)$$

and

$$\sum_{n \in \mathbb{N}} \|x_n - x_{n+1} - \gamma(\nabla \varphi(x_n) - \nabla \varphi(x^*))\|^2 < +\infty, \quad (34)$$

which yields  $\nabla \varphi(x_n) \rightarrow \nabla \varphi(x^*)$  and

$$x_n - x_{n+1} = (\text{Id} - T_\gamma)x_n \rightarrow 0. \quad (35)$$

Now, let  $x$  be a weak accumulation point of  $(x_n)_{n \in \mathbb{N}}$ , say  $x_{k_n} \rightharpoonup x$ . Since  $\overline{B}_{M_0-\varepsilon}(x^*)$  is weakly closed,  $x$  and  $(x_{k_n})_{n \in \mathbb{N}}$  are in  $\overline{B}_{M_0-\varepsilon}(x^*)$ . Hence, it follows from the nonexpansiveness of  $T_\gamma$  in  $B_{M_0}(x^*)$ , guaranteed by Proposition 4(i), that

$$\begin{aligned}
& \|x - T_\gamma x\|^2 \\
&= \|x_{k_n} - T_\gamma x\|^2 - \|x_{k_n} - x\|^2 - 2\langle x_{k_n} - x \mid x - T_\gamma x \rangle \\
&= \|x_{k_n} - T_\gamma x_{k_n}\|^2 + \|T_\gamma x_{k_n} - T_\gamma x\|^2 + 2\langle x_{k_n} - T_\gamma x_{k_n} \mid T_\gamma x_{k_n} - T_\gamma x \rangle \\
&\quad - \|x_{k_n} - x\|^2 - 2\langle x_{k_n} - x \mid x - T_\gamma x \rangle \\
&\leq \|(\text{Id} - T_\gamma)x_{k_n}\|^2 + 2\langle (\text{Id} - T_\gamma)x_{k_n} \mid x_{k_{n+1}} - T_\gamma x \rangle - 2\langle x_{k_n} - x \mid x - T_\gamma x \rangle. \quad (36)
\end{aligned}$$

Therefore, (35),  $x_{k_n} \rightharpoonup x$ , and boundedness of  $(x_{k_n})_{n \in \mathbb{N}}$  imply that the right hand side of (36) tends to 0 as  $n \rightarrow +\infty$ , which yields  $x = T_\gamma x$  and, thus,  $x \in S$ , in view of Proposition 3. Then, (i) follows from [9, Lemma 2.47].

(ii): The uniqueness of the solution  $x^*$  follows from Proposition 3. Since  $\mu_{M_0}(x^*) > 0$ , Proposition 4(ii) implies that  $(x_n)_{n \in \mathbb{N}} \subset B_{M_0}(x^*)$ , following the same argument used in i. Therefore, for every  $n \in \mathbb{N}$ ,

$$\|x_{n+1} - x^*\| = \|T_\gamma x_n - T_\gamma x^*\| \leq \rho_{M_0}(x^*, \gamma) \|x_n - x^*\| \leq (\rho_{M_0}(x^*, \gamma))^{n+1} \|x_0 - x^*\| \quad (37)$$

and the proof is complete.  $\square$

**Remark 2.** (i) Note that, for every  $x^*$  and  $M > 0$ ,  $\rho_M(x^*, \cdot)$  is decreasing in  $]0, 2/(L_M(x^*) + \mu_M(x^*))]$  and increasing in  $]2/(L_M(x^*) + \mu_M(x^*)), 2/L_M(x^*)]$ . Therefore, the optimal convergence rate is obtained by choosing  $\gamma^* = 2/(L_M(x^*) + \mu_M(x^*))$ , which yields

$$\rho_M(x^*, \gamma^*) = \frac{L_M(x^*) - \mu_M(x^*)}{L_M(x^*) + \mu_M(x^*)}. \quad (38)$$

- (ii) In the particular case when  $\nabla \varphi$  is globally Lipschitz continuous with constant  $L > 0$ , we have, for every  $M > 0$ ,  $L_M(\cdot) \equiv \mathcal{L}_M = L$  and we recover the convergence of the classical FBS (see, e.g., [29, Theorem 3.4]). In addition, if  $\varphi$  is  $\rho$ -strongly convex in  $\mathcal{H}$  for some  $\rho > 0$ , we recover the linear convergence rate in [19, 54].
- (iii) In general,  $M_0$  and  $\mathcal{L}_{M_0}$  in Theorem 5 are difficult to compute exactly since the solution set  $S$  is not known. However, it is possible to over-estimate them by knowing a priori bounds on the solutions. Indeed, if  $S \subset U$ , where  $U \subset \mathcal{H}$  is an a priori region in which the solutions are known to be, we have

$$M_0 \leq M := \sup_{x \in U} \|x - x_0\| \quad \text{and} \quad \mathcal{L}_{M_0} \leq \sup_{x \in U} L_M(x). \quad (39)$$

In the locally strongly convex case, if the unique solution  $x^*$  is known to be in an a priori set  $U$ , we can also under-estimate  $\mu_M(x^*)$  by  $\inf_{x \in U} \mu_M(x)$  provided that the latter is strictly positive.

## 4 Application to ergodic variational mean field games

Consider the following stationary MFGs system with monotone local couplings

$$\begin{aligned} -\nu \Delta u + \mathcal{H}(\nabla u) + \lambda &= f(x, m(x)) \text{ in } \mathbb{T}^d, \\ -\nu \Delta m - \operatorname{div}(\nabla \mathcal{H}(\nabla u)m) &= 0 \quad \text{in } \mathbb{T}^d, \\ \int_{\mathbb{T}^d} u(x) dx = 0, \int_{\mathbb{T}^d} m(x) dx &= 1, m > 0. \end{aligned} \quad (40)$$

Here,  $\nu > 0$ ,  $\mathbb{T}^d$  denotes the  $d$ -dimensional torus,  $u, m: \mathbb{T}^d \rightarrow \mathbb{R}$  and  $\lambda \in \mathbb{R}$  are the unknowns, the Hamiltonian  $\mathcal{H}: \mathbb{R}^d \rightarrow \mathbb{R}$  is convex and differentiable, and  $\mathbb{T}^d \times ]0, +\infty[ \ni (x, \rho) \mapsto f(x, \rho) \in \mathbb{R}$  is continuous and strictly increasing with respect to  $\rho$ . Moreover, we suppose that

$$(\forall x \in \mathbb{T}^d)(\forall \rho \in ]0, +\infty[) \int_0^\rho f(x, \rho') d\rho' < \infty, \quad (41)$$

$$(\forall x \in \mathbb{T}^d) \lim_{\rho \rightarrow +\infty} f(x, \rho) = +\infty. \quad (42)$$

Under suitable assumptions on the growth of  $\mathcal{H}$  and on the regularity of  $f$ , system (40) corresponds to the optimality system of a convex variational problem (see e.g. [39, 46]) and admits a unique smooth solution  $(u^*, m^*, \lambda^*)$  (see e.g. [22, 27, 30, 31, 33, 50]).

In this section, we consider the finite difference scheme proposed in [3] to approximate the solution to (40) in the two-dimensional case  $d = 2$  and when

$$(\forall p \in \mathbb{R}^d) \quad \mathcal{H}(p) = \ell^*(\|p\|), \quad (43)$$

where  $\ell \in \Gamma_0(\mathbb{R})$  is non-negative, supercoercive, increasing, and strictly convex on its domain  $\operatorname{dom} \ell = [0, +\infty[$ . The non-negativity of  $\ell$  over its domain yields that  $\ell^*$  is increasing. Moreover, by (11) and [9, Corollary 17.44], the supercoercivity and the strict convexity of  $\ell$  imply that  $\operatorname{dom} \ell^* = \mathbb{R}$  and that  $\ell^*$  is differentiable on  $\mathbb{R}$ .

**Remark 3.** Let  $\gamma \in [1, +\infty[$  and set  $\gamma' = \gamma/(\gamma - 1)$ . A typical example of a Hamiltonian satisfying the previous assumptions is given by  $\mathcal{H}(p) = \|p\|^{\gamma'}/\gamma'$  for all  $p \in \mathbb{R}^d$ , which is obtained from (43) with  $\ell(r) = |r|^\gamma/\gamma + \iota_{[0, +\infty[}(r)$  for all  $r \in \mathbb{R}$ .

Let  $N \in \mathbb{N}$ , set  $h = 1/N$  and set  $\mathbb{T}_h = \{x_{i,j} = (ih, jh) \mid i, j = 0, \dots, N-1\}$ . Let  $\mathcal{M}_h$  be the set of real-valued functions defined on  $\mathbb{T}_h$ , set  $\mathcal{W}_h = \mathcal{M}_h^4$ , and set  $\mathcal{Y}_h = \{z \in \mathcal{M}_h \mid \sum_{i,j=0}^{N-1} z(x_{i,j}) = 0\}$ . For notational simplicity, given  $z \in \mathcal{M}_h$ , we will write  $z_{i,j}$  for  $z(x_{i,j})$ . The discrete differential operators  $D_1: \mathcal{M}_h \rightarrow \mathcal{Y}_h$ ,  $D_2: \mathcal{M}_h \rightarrow \mathcal{Y}_h$ ,

$D_h: \mathcal{M}_h \rightarrow \mathcal{Y}_h^4$ ,  $\Delta_h: \mathcal{M}_h \rightarrow \mathcal{Y}_h$ , and  $\text{div}_h: \mathcal{W}_h \rightarrow \mathcal{Y}_h$  are defined as

$$\begin{aligned} (D_1 z)_{i,j} &= \frac{z_{i+1,j} - z_{i,j}}{h}, & (D_2 z)_{i,j} &= \frac{z_{i,j+1} - z_{i,j}}{h}, \\ [D_h z]_{i,j} &= ((D_1 z)_{i,j}, (D_1 z)_{i-1,j}, (D_2 z)_{i,j}, (D_2 z)_{i,j-1}), \\ (\Delta_h z)_{i,j} &= \frac{z_{i-1,j} + z_{i+1,j} + z_{i,j-1} + z_{i,j+1} - 4z_{i,j}}{h^2}, \\ (\text{div}_h(w))_{i,j} &= (D_1 w^1)_{i-1,j} + (D_1 w^2)_{i,j} + (D_2 w^3)_{i,j-1} + (D_2 w^4)_{i,j}, \end{aligned}$$

for all  $z \in \mathcal{M}_h$ ,  $w \in \mathcal{W}_h$ ,  $i, j = 0, \dots, N-1$ , and the sums between the indexes are taken modulo  $N$ .

Let us set  $\mathbb{R}_+ = [0, +\infty[$  and  $\mathbb{R}_- = ]-\infty, 0]$ . Let  $K := \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_- \subset \mathbb{R}^4$ , denote by  $P_K(p)$  the euclidean projection of  $p \in \mathbb{R}^4$  onto the closed and convex cone  $K$ , and set

$$(\forall C \in \{K, \mathbb{R}^4\}) \quad \zeta_C: \mathbb{R}^4 \rightarrow \mathbb{R}^4: \xi \mapsto \begin{cases} \frac{\ell^{*'}(\|P_C \xi\|)}{\|P_C \xi\|} P_C \xi, & \text{if } P_C \xi \neq 0; \\ 0, & \text{if } P_C \xi = 0. \end{cases} \quad (44)$$

The finite difference discretization of system (40) reads as follows:

$$\begin{aligned} -\nu(\Delta_h u)_{i,j} + \ell^*(\|P_K(-[D_h u]_{i,j})\|) + \lambda &= f(x_{i,j}, m_{i,j}), \quad \text{for all } i, j = 0, \dots, N-1, \\ -\nu(\Delta_h m)_{i,j} + (\text{div}_h(m \zeta_K(-[D_h u])))_{i,j} &= 0, \quad \text{for all } i, j = 0, \dots, N-1, \\ \sum_{i,j=0}^{N-1} u_{i,j} = 0, \quad h^2 \sum_{i,j=0}^{N-1} m_{i,j} = 1, \quad m_{i,j} > 0, & \text{for all } i, j = 0, \dots, N-1. \end{aligned} \quad (45)$$

As for (40), system (45) corresponds to the optimality condition for the solution to a convex variational problem. In order to define this problem, set

$$b: \mathbb{R} \times \mathbb{R}^4 \rightarrow ]-\infty, +\infty]: (\rho, \omega) \mapsto \begin{cases} \rho \ell\left(\frac{\|\omega\|}{\rho}\right), & \text{if } \rho > 0; \\ 0, & \text{if } (\rho, \omega) = (0, 0); \\ +\infty & \text{otherwise} \end{cases} \quad (46)$$

and, for every  $i, j = 0, \dots, N-1$ , define

$$F_{i,j}(\rho) = \begin{cases} \int_0^\rho f(x_{i,j}, \rho') d\rho', & \text{if } \rho \in [0, +\infty[; \\ +\infty, & \text{otherwise.} \end{cases} \quad (47)$$

Notice that (41) yields  $\text{dom } F_{i,j} = [0, +\infty[$ , and, since  $f(x_{i,j}, \cdot)$  is strictly increasing, the function  $F_{i,j}$  is strictly convex. Moreover, it follows from (41), (42), and (47) that

$F_{i,j}$  is supercoercive. Thus, as for  $\ell$ , we have that  $\text{dom } F_{i,j}^* = \mathbb{R}$  and  $F_{i,j}^*$  is increasing and differentiable on  $\mathbb{R}$ .

**Problem 6.** *The problem is to*

$$\begin{aligned} \min_{(m,w) \in \mathcal{M}_h \times \mathcal{W}_h} & \sum_{i,j=0}^{N-1} \left( b(m_{i,j}, w_{i,j}) + F_{i,j}(m_{i,j}) \right) \\ \text{s.t.} & \begin{cases} -\nu \Delta_h m + \text{div}_h w = 0, \\ h^2 \sum_{i,j=0}^{N-1} m_{i,j} = 1, \\ w_{i,j} \in K, \quad \text{for all } i, j = 0, \dots, N-1. \end{cases} \end{aligned} \quad (48)$$

Note that, in view of [51],  $b \in \Gamma_0(\mathbb{R} \times \mathbb{R}^4)$  and hence Problem 4.2 is convex.

**Proposition 7.** *System (45) has a unique solution  $(m^h, u^h, \lambda^h)$ . Moreover, setting*

$$w^h = m^h \zeta_K(-[D_h u^h]), \quad (49)$$

*we have that  $(m^h, w^h)$  is the unique solution to Problem 6.*

*Proof.* When  $\mathcal{H} = \|\cdot\|^2/2$ , the proof of the previous result is given in [15, Proposition 4.1]. The case when  $\mathcal{H}$  is given by (43) follows from exactly the same arguments.  $\square$

In what follows, our aim will be to apply the results in Section 3 to a dual formulation of Problem 6. For this purpose, the following result will play an important role.

**Proposition 8.** *In the context of Problem 6, let  $C \in \{K, \mathbb{R}^4\}$  and, for every  $i, j \in \{0, \dots, N-1\}$ , define*

$$\phi_{i,j}: \mathbb{R} \times \mathbb{R}^4 \rightarrow ]-\infty, +\infty]: (\rho, \omega) \mapsto b(\rho, \omega) + F_{i,j}(\rho) + \iota_C(\omega). \quad (50)$$

*Then, for every  $(\rho^*, \omega^*) \in \mathbb{R} \times \mathbb{R}^4$ , the following hold:*

- (i)  $\phi_{i,j}^*(\rho^*, \omega^*) = F_{i,j}^*(\rho^* + \ell^*(\|P_C \omega^*\|))$ .
- (ii)  $\phi_{i,j}^*$  is differentiable in  $\mathbb{R} \times \mathbb{R}^4$  and

$$\nabla \phi_{i,j}^*(\rho^*, \omega^*) = F_{i,j}^{*'}(\rho^* + \ell^*(\|P_C \omega^*\|)) \begin{pmatrix} 1 \\ \zeta_C(\omega^*) \end{pmatrix}, \quad (51)$$

*where  $\zeta_C$  is defined in (44).*

*Proof.* Fix  $i$  and  $j$  in  $\{0, \dots, N-1\}$  and  $(\rho^*, \omega^*) \in \mathbb{R} \times \mathbb{R}^4$ .

(i): Set  $\hat{\rho} = 1$  and  $\hat{\omega} = (1, -1, 1, -1)$ . Since  $b$  is continuous at  $(\hat{\rho}, \hat{\omega})$  and  $F_{i,j}(\hat{\rho}) + \iota_C(\hat{\omega}) < +\infty$ , it follows from [9, Proposition 15.5(iv) & Theorem 15.3], [9,

Proposition 13.30 & Example 13.3(i)-(ii)], and [28, Proposition 2.3(iv)] that

$$\begin{aligned}
\phi_{i,j}^*(\rho^*, \omega^*) &= \inf_{(\rho, \omega) \in \mathbb{R} \times \mathbb{R}^4} b^*(\rho^* - \rho, \omega^* - \omega) + F_{i,j}^*(\rho) + \iota_{C^-}(\omega) \\
&= \inf_{\substack{(\rho, \omega) \in \mathbb{R} \times C^-, \\ \rho^* + \ell^*(\|\omega^* - \omega\|) \leq \rho}} F_{i,j}^*(\rho) \\
&= \inf_{\omega \in C^-} F_{i,j}^*(\rho^* + \ell^*(\|\omega^* - \omega\|)) \\
&= F_{i,j}^*(\rho^* + \ell^*(\|\omega^* - P_{C^-} \omega^*\|)),
\end{aligned}$$

where, in the last equality, we have used that  $\mathbb{R} \ni t \mapsto F_{i,j}^*(\rho^* + \ell^*(t)) \in \mathbb{R}$  is increasing. The result follows from the identity  $P_C = \text{Id} - P_{C^-}$ .

(ii): Using that, for all  $\xi \in \mathbb{R}^4$ ,  $\|P_C \xi\| = d_{C^-}(\xi)$ ,  $\ell^{*'}(0) = 0$ , and  $P_C \xi = 0$  if and only if  $\xi \in C^-$ , it follows from [9, Example 17.33] that  $\ell^* \circ \|\cdot\| \circ P_C$  is differentiable and  $\nabla(\ell^* \circ \|\cdot\| \circ P_C) = \zeta_C$ . Thus, by i,  $\phi_{i,j}^*$  is differentiable in  $\mathbb{R} \times \mathbb{R}^4$  and (51) holds.  $\square$

For every  $C \in \{K, \mathbb{R}^4\}$ , we define

$$\begin{aligned}
D_C = \left\{ (m, w) \in \mathcal{M}_h \times \mathcal{W}_h \mid -\nu \Delta_h m + \text{div}_h w = 0, \right. \\
\left. h^2 \sum_{i,j=0}^{N-1} m_{i,j} = 1, (\forall i, j \in \{0, \dots, N-1\}) \quad w_{i,j} \in C \right\}. \quad (52)
\end{aligned}$$

Note that Problem 6 can be written equivalently as

$$\min_{(m, w) \in D_{C_1}} \sum_{i,j=0}^{N-1} (b(m_{i,j}, w_{i,j}) + F_{i,j}(m_{i,j}) + \iota_{C_2}(w_{i,j})) \quad (53)$$

either if  $(C_1, C_2) = (K, \mathbb{R}^4)$  or  $(C_1, C_2) = (\mathbb{R}^4, K)$ . In view of [9, Proposition 13.30 & Definition 15.19] and Proposition 8, formulation (53) leads to the following Fenchel-Rockafellar dual problem

$$\min_{(\theta, v) \in \mathcal{M}_h \times \mathcal{W}_h} \sigma_{D_{C_1}}(-\theta, -v) + \sum_{i,j=0}^{N-1} F_{i,j}^*(\theta_{i,j} + \ell^*(\|P_{C_2} v_{i,j}\|)). \quad (54)$$

Note that (54) can be written as (2), where  $\psi = \sigma_{D_{C_1}} \circ (-\text{Id})$ ,  $\varphi: (\theta, v) \mapsto \sum_{i,j=0}^{N-1} \phi_{i,j}^*(\theta_{i,j}, v_{i,j})$ , and, for every  $i, j \in \{0, \dots, N-1\}$ ,  $\phi_{i,j}^*: (\theta_{i,j}, v_{i,j}) \mapsto F_{i,j}^*(\theta_{i,j} + \ell^*(\|P_{C_2} v_{i,j}\|))$ . By Proposition 8(ii) we deduce that  $\varphi$  is differentiable. Under additional assumptions on  $\ell^{*'}$  and  $(F_{i,j}^*)'_{0 \leq i,j \leq N-1}$ , one can prove that (54) is a particular instance of Problem 1. In the next section, we provide some examples and explicit computations of Lipschitz and strong convexity constants for the function  $\varphi$ .

Notice that, by [9, Proposition 19.4] and Proposition 8(ii), we recover the unique solution  $(m^h, w^h)$  to Problem 6 from a solution  $(\theta^h, v^h)$  to Problem 54 through the

expressions

$$(\forall i, j \in \{0, \dots, N-1\}) \quad m_{i,j}^h = F_{i,j}^{*'}(\theta_{i,j}^h + \ell^*(\|P_C v_{i,j}^h\|)) \quad \text{and} \quad w_{i,j}^h = m_{i,j}^h \zeta_C(v_{i,j}^h). \quad (55)$$

Since (49) implies that

$$(\forall i, j \in \{0, \dots, N-1\}) \quad \ell^*(\|P_K(-[D_h u]_{i,j})\|) = \ell^*(\ell'(\|w_{i,j}^h\|/m_{i,j}^h)), \quad (56)$$

one can compute  $(u^h, \lambda^h)$  by solving the linear system consisting in the first equation in (45) together with the condition  $\sum_{i,j=0}^{N-1} u_{i,j}^h = 0$ .

Therefore, in what follows, we focus on approximating the solution  $(\theta^h, v^h)$  to (54) for which we will consider the following routine.

**Algorithm 9.** *Let  $(\theta_0, v_0) \in \mathcal{M}_h \times \mathcal{W}_h$ , let  $\gamma \in ]0, +\infty[$ , and consider the recurrence:*

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \text{for } i, j = 0, \dots, N-1 \\ t_{i,j,n} = \theta_{i,j,n} - \gamma F_{i,j}^{*'}(\theta_{i,j,n} + \ell^*(\|P_{C_2} v_{i,j,n}\|)) \\ \nu_{i,j,n} = v_{i,j,n} - \gamma F_{i,j}^{*'}(\theta_{i,j,n} + \ell^*(\|P_{C_2} v_{i,j,n}\|)) \zeta_{C_2}(v_{i,j,n}) \\ (\theta_{n+1}, v_{n+1}) = \text{prox}_{\gamma \sigma_{D_{C_1}}}((t_{i,j,n}, \nu_{i,j,n})_{0 \leq i,j \leq N-1}). \end{cases} \quad (57)$$

## 5 Numerical results

In this section, we consider some numerical experiments in two instances. In the first experiment, we consider the first order mean field game studied in [6], in which the exact solution is known. In this context we compare the relative exact error of Algorithm 9 when  $(C_1, C_2) = (K, \mathbb{R}^4)$  and when  $(C_1, C_2) = (\mathbb{R}^4, K)$ . The second experiment is devoted to a second-order mean field game in which the coupling term is the sum of a power function with an entropic penalization. We compare state-of-the-art algorithms to both algorithms tested in the first experiment. We test a bunch of different values of time steps and record the best performance corresponding to the minimum number of iterations of each algorithm considered in each numerical comparison.

### 5.1 A first order mean field game with a logarithmic coupling

We consider here the MFGs system

$$\frac{1}{2}|\nabla u|^2 + \lambda = \log(m) - \sin(2\pi x) - \sin(2\pi y) \quad \text{in } \mathbb{T}^2, \quad (58)$$

$$\text{div}(m \nabla u) = 0 \quad \text{in } \mathbb{T}^2, \quad \int_{\mathbb{T}^2} u(x) dx = 0, \quad \int_{\mathbb{T}^2} m(x) dx = 1, \quad m > 0. \quad (59)$$

Its finite difference discretization (45) is given by

$$\begin{aligned} \frac{1}{2} \|P_K(-[D_h u]_{i,j})\|^2 + \lambda &= \log(m_{i,j}) - \sin(2\pi i h) - \sin(2\pi j h), \\ &\text{for all } i, j = 0, \dots, N-1, \\ (\operatorname{div}_h(m P_K(-[D_h u])))_{i,j} &= 0, \quad \text{for all } i, j = 0, \dots, N-1, \\ \sum_{i,j=0}^{N-1} u_{i,j} &= 0, \quad h^2 \sum_{i,j=0}^{N-1} m_{i,j} = 1, \quad m_{i,j} > 0, \quad \text{for all } i, j = 0, \dots, N-1. \end{aligned} \quad (60)$$

Note that, since  $\nu = 0$  in (60), one cannot apply Proposition 7. However, one can easily check (see [18, Remark 2.1]) that (7) corresponds to the optimality condition of Problem 6, where  $\nu = 0$ ,  $\ell = |\cdot|^2/2 + \iota_{[0,+\infty[}$ , for every  $i, j \in \{0, \dots, N-1\}$ ,  $F_{i,j}: \rho \mapsto \rho(\ln(\rho) - 1) - c_{i,j}\rho$  and  $c_{i,j} = \sin(2\pi i h) + \sin(2\pi j h)$ . One of the main interests of system (7) is that it admits an explicit solution (see [6]) which allows us to establish the precise error for its approximation. In particular, the explicit solution to the dual problem (54) in this case is

$$\theta_{i,j}^* = t^*, \quad \text{for all } i, j = 0, \dots, N-1, \quad \text{where } t^* = -\ln\left(h^2 \sum_{i,j=0}^{N-1} e^{c_{i,j}}\right), \quad \text{and } v^* = 0. \quad (61)$$

Moreover, since  $\ell^* = \max\{\operatorname{sgn}(\cdot) \cdot |\cdot|^2/2, 0\}$  and  $F_{i,j}^*: \rho^* \mapsto e^{\rho^* + c_{i,j}}$ , the dual formulation (54) reduces to

$$\min_{(\theta, v) \in \mathcal{M}_h \times \mathcal{W}_h} \sigma_{D_{C_1}}(-\theta, -v) + \sum_{i,j=0}^{N-1} e^{\theta_{i,j} + \frac{1}{2} \|P_{C_2} v_{i,j}\|^2 + c_{i,j}}. \quad (62)$$

Now, in the case when  $C_1 = K$  and  $C_2 = \mathbb{R}^4$ , let us prove that (62) is a particular instance of Problem 1 which is also strongly convex. Indeed, let  $M_0 > 0$ , for every  $i, j \in \{0, \dots, N-1\}$ , set

$$L_{M_0, i, j}(\theta_{i,j}^*, v_{i,j}^*) = e^{\theta_{i,j}^* + c_{i,j}} g(M_0^2) \leq e^{\theta_{i,j}^* + c_{i,j} + M_0} (3 + M_0^2) \quad (63)$$

and

$$\mu_{M_0, i, j}(\theta_{i,j}^*, v_{i,j}^*) = e^{\theta_{i,j}^* + c_{i,j} - M_0}, \quad (64)$$

where  $g: M \mapsto \max_{x \in [0, M]} e^{\sqrt{M-x}} (1 + x/2 + (2 + x/2)\sqrt{x/(4+x)})$ . Then,  $\phi_{i,j}^*: (\rho^*, \omega^*) \mapsto e^{\rho^* + \|\omega^*\|^2/2 + c_{i,j}}$  is  $\mu_{M_0, i, j}(\theta_{i,j}^*, v_{i,j}^*)$ -strongly convex, differentiable in  $B((\theta_{i,j}^*, v_{i,j}^*), M_0)$ . Moreover, in view of Proposition 8(ii),  $\ell^{*l} = \max\{\operatorname{Id}, 0\}$ ,  $\zeta_{C_2} = \operatorname{Id}$ , and  $\nabla \phi_{i,j}^*: (\rho^*, \omega^*) \mapsto e^{\rho^* + \|\omega^*\|^2/2 + c_{i,j}} (1, \omega^*)^\top$  is  $L_{M_0, i, j}(\theta_{i,j}^*, v_{i,j}^*)$ -Lipschitz continuous in  $B((\theta_{i,j}^*, v_{i,j}^*), M_0)$ . Therefore,  $\varphi: (\theta, v) \mapsto \sum_{i,j=0}^{N-1} \phi_{i,j}^*(\theta_{i,j}, v_{i,j})$  is  $\mu_{M_0}(\theta^*, v^*)$ -strongly convex and  $\nabla \varphi$  is  $L_{M_0}(\theta^*, v^*)$ -Lipschitz continuous in  $B((\theta^*, v^*), M_0)$ , where



$$\begin{aligned}
L_{M_0}(\theta^*, v^*) &= \max_{i,j=0,\dots,N-1} L_{M_0,i,j}(\theta_{i,j}^*, v_{i,j}^*) = e^{t^* + \bar{c}} g(M_0^2) \leq e^{t^* + \bar{c} + M_0} (3 + M_0^2), \\
\mu_{M_0}(\theta^*, v^*) &= \min_{i,j=0,\dots,N-1} \mu_{M_0,i,j}(\theta_{i,j}^*, v_{i,j}^*) = e^{t^* + \underline{c} - M_0} > 0, \quad (65)
\end{aligned}$$

$\underline{c} = \min_{i,j=0,\dots,N-1} c_{i,j}$ , and  $\bar{c} = \max_{i,j=0,\dots,N-1} c_{i,j}$ . In this case, Algorithm 9 reduces to

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \text{for } i, j = 0, \dots, N-1 \\ \left[ \begin{array}{l} t_{i,j,n} = \tilde{\theta}_{i,j,n} - \gamma e^{\tilde{\theta}_{i,j,n} + \|\tilde{v}_{i,j,n}\|^2/2 + c_{i,j}} \\ \nu_{i,j,n} = (1 - \gamma e^{\tilde{\theta}_{i,j,n} + \|\tilde{v}_{i,j,n}\|^2/2 + c_{i,j}}) \tilde{v}_{i,j,n} \\ (\tilde{\theta}_{n+1}, \tilde{v}_{n+1}) = \text{prox}_{\gamma \sigma_{D_K}}((t_{i,j,n}, \nu_{i,j,n})_{0 \leq i,j \leq N-1}), \end{array} \right. \end{cases} \quad (\text{DFB0})$$

where  $(\tilde{\theta}_0, \tilde{v}_0) \in \mathcal{M}_h \times \mathcal{W}_h$  and  $\gamma > 0$ . In this context, if  $\gamma \in ]0, 2/L_{M_0}(\theta^*, v^*)[$ , with  $M_0 = \|(\tilde{\theta}_0, \tilde{v}_0) - (\theta^*, v^*)\|$ , Theorem 5(ii) guarantees the linear convergence of the sequence  $(\tilde{\theta}_n, \tilde{v}_n)_{n \in \mathbb{N}}$  generated by (DFB0) to  $(\theta^*, v^*)$ . In addition, the optimal rate is achieved by choosing  $\gamma^*$  from Remark 2(i). Observe that, in view of (8), the computation of  $\text{prox}_{\gamma \sigma_{D_K}}$  is obtained from the projection onto  $D_K$ , which needs a subroutine.

On the other hand, if  $C_1 = \mathbb{R}^4$  and  $C_2 = K$ ,  $\varphi: (\theta, v) \mapsto \sum_{i,j=0}^{N-1} e^{\theta_{i,j} + \|P_K v_{i,j}\|^2/2 + c_{i,j}}$  is convex, differentiable and

$$\begin{aligned}
\nabla \varphi: (\theta, v) &\mapsto (e^{\theta_{i,j} + \|P_K v_{i,j}\|^2/2 + c_{i,j}} (1, P_K v_{i,j})^\top)_{0 \leq i,j \leq N-1} \\
&= \nabla \varphi(\theta, (P_K v_{i,j})_{0 \leq i,j \leq N-1}) \quad (66)
\end{aligned}$$

is  $L_{M_0}(\theta^*, v^*)$ -Lipschitz continuous in  $B((\theta^*, v^*), M_0)$ , for every  $M_0 > 0$ . However, for every  $M_0 > 0$ ,  $\varphi$  is not strongly convex in  $B((\theta^*, v^*), M_0)$  (i.e.,  $\mu_{M_0}(\theta^*, v^*) = 0$ ). Indeed, let  $\theta = \theta^*$  and let  $v \in (\mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+)^N \setminus \{0\}$ . Then, for every  $i, j \in \{0, \dots, N-1\}$ ,  $P_K v_{i,j} = 0$  and (61) yields

$$\begin{aligned}
&(\nabla \varphi(\theta, v) - \nabla \varphi(\theta^*, v^*))_{i,j} \\
&= e^{\theta_{i,j} + \|P_K v_{i,j}\|^2/2 + c_{i,j}} (1, P_K v_{i,j})^\top - e^{\theta_{i,j}^* + \|P_K v_{i,j}^*\|^2/2 + c_{i,j}} (1, P_K v_{i,j}^*)^\top \\
&= e^{t^* + c_{i,j}} (1, 0)^\top - e^{t^* + c_{i,j}} (1, 0)^\top = 0 \quad (67)
\end{aligned}$$

and hence

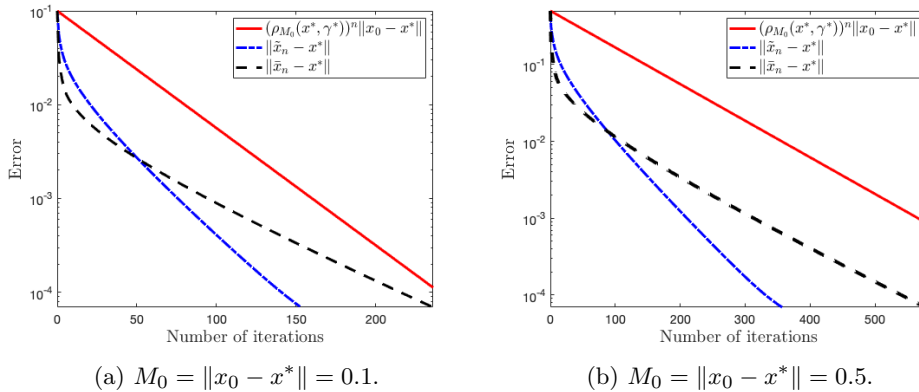
$$\langle \nabla \varphi(\theta, v) - \nabla \varphi(\theta^*, v^*) \mid (\theta - \theta^*, v - v^*) \rangle = 0. \quad (68)$$

Then, (62) reduces to Problem 1 but without any strong convexity. In this case, Algorithm 9 reduces to

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \text{for } i, j = 0, \dots, N-1 \\ \left[ \begin{array}{l} t_{i,j,n} = \bar{\theta}_{i,j,n} - \gamma e^{\bar{\theta}_{i,j,n} + \|P_K \bar{v}_{i,j,n}\|^2/2 + c_{i,j}} \\ \nu_{i,j,n} = \bar{v}_{i,j,n} - \gamma e^{\bar{\theta}_{i,j,n} + \|P_K \bar{v}_{i,j,n}\|^2/2 + c_{i,j}} P_K \bar{v}_{i,j,n} \\ (\bar{\theta}_{n+1}, \bar{v}_{n+1}) = \text{prox}_{\gamma \sigma_{D_{\mathbb{R}^4}}}((t_{i,j,n}, \nu_{i,j,n})_{0 \leq i,j \leq N-1}), \end{array} \right. \end{cases} \quad (\text{DFB1})$$

where  $(\bar{\theta}_0, \bar{v}_0) \in \mathcal{M}_h \times \mathcal{W}_h$  and  $\gamma > 0$ . In this context, if  $\gamma \in ]0, 2/L_{\bar{M}_0}(\theta^*, v^*)[$ , with  $\bar{M}_0 = \|(\bar{\theta}_0, \bar{v}_0) - (\theta^*, v^*)\|$ , Theorem 5(i) guarantees the convergence of the sequence  $(\bar{\theta}_n, \bar{v}_n)_{n \in \mathbb{N}}$  generated by (DFB1) to  $(\theta^*, v^*)$ . Our result does not guarantee linear convergence, but the computation of  $\text{prox}_{\gamma\sigma_{D_{\mathbb{R}^4}}}$  is obtained from the projection onto  $D_{\mathbb{R}^4}$ , which only needs a numerically efficient matrix inversion.

In Figure 1 we compare the numerical behavior of the algorithms in (DFB0), (DFB1), and the theoretical upper bound obtained in Theorem 5(ii). We set  $\tilde{x}_n = (\tilde{\theta}_n, \tilde{v}_n)$  and  $\bar{x}_n = (\bar{\theta}_n, \bar{v}_n)$  for the iterates computed with (DFB0) and (DFB1), respectively. We consider the step-size  $\gamma^*$  achieving the optimal convergence rate in Remark 2(i) and, for (DFB1), we set  $\gamma = 1.99/L_{\bar{M}_0}(\theta^*, v^*)$ . We also set the error tolerance to  $7 \cdot 10^{-5}$  and  $\tilde{x}_0 = (\tilde{\theta}_0, \tilde{v}_0) = (\bar{\theta}_0, \bar{v}_0) = \bar{x}_0$ . In each case, the initial point  $x_0 = \tilde{x}_0, \bar{x}_0$  is chosen by perturbing  $x^* = (\theta^*, v^*)$  randomly and such that  $\|x_0 - x^*\| \in \{0.1, 0.5\}$ . We observe that the numerical and theoretical linear convergence rates are closer for closer starting points, while in both cases the error of (DFB1) decreases slowly with respect to the number of iterations. However, since (DFB0) involves sub-iterations in order to compute the projection onto  $D_K$ , it is actually much slower in terms of computational time than (DFB1) as it can be perceived in Table 1.



**Fig. 1:** Numerical comparison of the theoretical upper bound and the errors of the algorithms in (DFB0) and (DFB1) for a tolerance of  $7 \cdot 10^{-5}$  and two initial conditions.

## 5.2 A second order mean field game with logarithmic and power couplings

Let  $\nu \in ]0, +\infty[$ ,  $\alpha \in ]1, 2]$ , and  $\epsilon \in [0, +\infty[$ . We consider the MFGs system

$$\begin{aligned} -\nu\Delta u + \frac{1}{2}\|\nabla u\|^2 + \lambda &= m^{\alpha-1} + \epsilon \log m + \tilde{H}(x, y) \quad \text{in } \mathbb{T}^2, \\ -\nu\Delta m - \text{div}(m\nabla u) &= 0 \quad \text{in } \mathbb{T}^2, \end{aligned}$$

$\ x_0 - x^*\ $	DFB0			DFB1		
	$\gamma$	Time (s)	Iterations	$\gamma$	Time (s)	Iterations
0.1	0.3748	170.343	153	0.3748	3.609	236
0.5	0.2132	566.695	357	0.2132	8.486	577

**Table 1:** Performance of (DFB0) and (DFB1), for different initialization when the number of grid nodes (DoF) is  $60^2$  and the tolerance is  $7 \cdot 10^{-5}$ .

$$\int_{\mathbb{T}^2} u(x) dx = 0, \quad \int_{\mathbb{T}^2} m(x) dx = 1, \quad m > 0, \quad (69)$$

where, for all  $(x, y) \in \mathbb{T}^2$ ,

$$\tilde{H}(x, y) = -\frac{\sin(2\pi x) + \sin(2\pi y) + \cos(4\pi x)}{2}.$$

We consider the finite difference approximation (45) of the system above, which corresponds to the optimality condition of Problem 6 with  $\ell = |\cdot|^2/2 + \iota_{[0, +\infty[}$  and

$$F_{i,j}: \rho \mapsto \begin{cases} \frac{1}{\alpha} \rho^\alpha + c_{i,j} \rho + \epsilon \rho (\log \rho - 1), & \text{if } \rho > 0; \\ 0, & \text{if } \rho = 0; \\ +\infty, & \text{if } \rho < 0, \end{cases} \quad (70)$$

where, for every  $i, j \in \{0, \dots, N-1\}$ ,  $c_{i,j} = \tilde{H}(ih, jh)$ .

$\nu$	CP			DR			DFB0			DFB1		
	Time (s)	Iterations	$\gamma$	Time (s)	Iterations	$\gamma$	Time (s)	Iterations	$\gamma$	Time (s)	Iterations	$\gamma$
0.1	5.285	26	1.05	5.2094	26	1.05	34.4958	23	0.6	1.6456	30	0.65
0.5	3.1873	16	1.05	3.1924	16	1.05	17.4105	12	0.55	0.50957	10	0.55

**Table 2:** Performance comparison among CP, DR, (DFB0), and (DFB1), for varying values of  $\nu$  when  $\epsilon = 0$  and the number of grid nodes (DoF) is  $60^2$ .

Since, for every  $i, j \in \{0, \dots, N-1\}$ ,  $F_{i,j}$  and  $\ell$  are supercoercive and strictly convex on their domains, Proposition 8(ii) implies that  $\phi_{i,j}^*: (\theta_{i,j}, v_{i,j}) \mapsto F_{i,j}^*(\theta_{i,j} + \|v_{i,j}\|^2/2)$  is everywhere defined, differentiable, and

$$\nabla \phi_{i,j}^*: (\theta_{i,j}, v_{i,j}) \mapsto (F_{i,j}^*)'(\theta_{i,j} + \|v_{i,j}\|^2/2) \begin{pmatrix} 1 \\ v_{i,j} \end{pmatrix}. \quad (71)$$

$\nu$	CP			DR			DFB0			DFB1		
	Time (s)	Iterations	$\gamma$	Time (s)	Iterations	$\gamma$	Time (s)	Iterations	$\gamma$	Time (s)	Iterations	$\gamma$
0.1	5.6857	26	0.95	5.5486	26	0.95	32.6036	22	0.7	1.4765	27	0.75
0.5	3.5337	17	0.95	3.5753	17	0.95	13.831	10	0.65	0.48746	10	0.65

**Table 3:** Performance comparison among CP, DR, (DFB0), and (DFB1), for varying values of  $\nu$  when  $\epsilon = 0.1$  and the number of grid nodes (DoF) is  $60^2$ .

$\nu$	CP			DR			DFB0			DFB1		
	Time (s)	Iterations	$\gamma$	Time (s)	Iterations	$\gamma$	Time (s)	Iterations	$\gamma$	Time (s)	Iterations	$\gamma$
0.1	7.9191	25	0.8	7.561	25	0.8	25.3507	17	1	1.3145	20	1.1
0.5	5.6428	20	0.75	5.4536	20	0.75	15.2802	8	1.05	0.49585	8	1.1

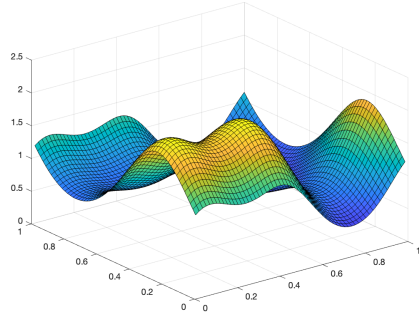
**Table 4:** Performance comparison among CP, DR, (DFB0), and (DFB1), for varying values of  $\nu$  when  $\epsilon = 0.5$  and the number of grid nodes (DoF) is  $60^2$ .

In the case when  $\epsilon > 0$ , we have, for every  $i, j \in \{0, \dots, N-1\}$ ,  $\text{dom } \partial F_{i,j} = \text{ran } (F_{i,j}^*)' = ]0, +\infty[$  and, therefore, for every  $(\rho, y) \in \mathbb{R} \times ]0, +\infty[$ ,

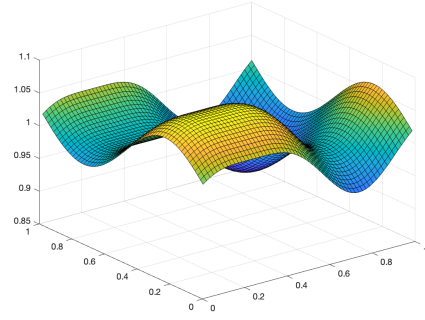
$$\begin{aligned}
& y = (F_{i,j}^*)'(\rho) \\
\Leftrightarrow & \rho = F_{i,j}'(y) = y^{\alpha-1} + c_{i,j} + \epsilon \ln(y) \\
\Leftrightarrow & \left(\frac{\alpha-1}{\epsilon}\right)(\rho - c_{i,j}) + \ln\left(\frac{\alpha-1}{\epsilon}\right) = \left(\frac{\alpha-1}{\epsilon}\right)y^{\alpha-1} + \ln\left(\left(\frac{\alpha-1}{\epsilon}\right)y^{\alpha-1}\right) \\
\Leftrightarrow & \left(\frac{\alpha-1}{\epsilon}\right)e^{\left(\frac{\alpha-1}{\epsilon}\right)(\rho - c_{i,j})} = \left(\left(\frac{\alpha-1}{\epsilon}\right)y^{\alpha-1}\right)e^{\left(\frac{\alpha-1}{\epsilon}\right)y^{\alpha-1}} \\
\Leftrightarrow & y = \left(\frac{\epsilon}{\alpha-1}\right)^{\frac{1}{\alpha-1}} W_0\left(\left(\frac{\alpha-1}{\epsilon}\right)e^{\left(\frac{\alpha-1}{\epsilon}\right)(\rho - c_{i,j})}\right)^{\frac{1}{\alpha-1}}, \tag{72}
\end{aligned}$$

where  $W_0$  stands for the principal branch of the Lambert W-function. On the other hand, in the case  $\epsilon = 0$ , we have  $\text{dom } \partial F_{i,j} = \text{ran } (F_{i,j}^*)' = [0, +\infty[$ . Altogether, it follows from simple computations that

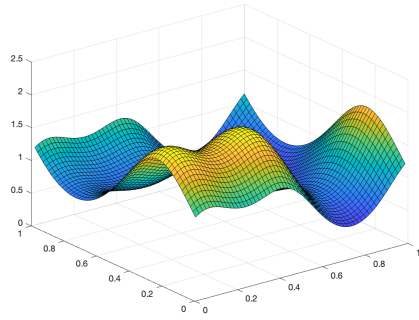
$$(F_{i,j}^*)': \rho \mapsto \begin{cases} 0, & \text{if } \epsilon = 0 \text{ and } \rho < c_{i,j}; \\ (\rho - c_{i,j})^{\frac{1}{\alpha-1}}, & \text{if } \epsilon = 0 \text{ and } \rho \geq c_{i,j}; \\ \left(\frac{\epsilon}{\alpha-1}\right)^{\frac{1}{\alpha-1}} W_0\left(\left(\frac{\alpha-1}{\epsilon}\right)e^{\left(\frac{\alpha-1}{\epsilon}\right)(\rho - c_{i,j})}\right)^{\frac{1}{\alpha-1}}, & \text{if } \epsilon > 0. \end{cases} \tag{73}$$



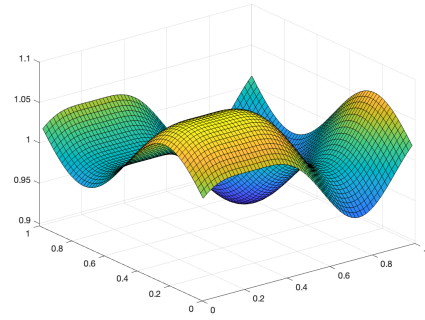
(a)  $\epsilon = 0, \nu = 0.1$ .



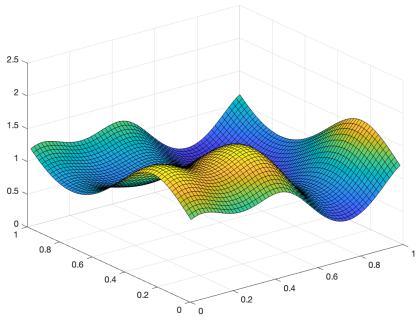
(b)  $\epsilon = 0, \nu = 0.5$ .



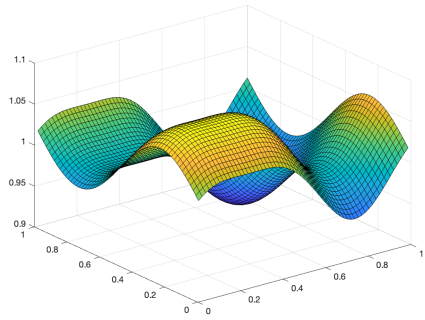
(c)  $\epsilon = 0.1, \nu = 0.1$ .



(d)  $\epsilon = 0.1, \nu = 0.5$ .



(e)  $\epsilon = 0.5, \nu = 0.1$ .



(f)  $\epsilon = 0.5, \nu = 0.5$ .

**Fig. 2:** Numerical results of (DFB1) for different values of  $\epsilon$  and  $\nu$ .

Since  $\alpha \in ]1, 2]$ , for every  $\epsilon \geq 0$  and  $i, j \in \{0, \dots, N - 1\}$ , the function  $(F_{i,j}^*)'$  is Lipschitz continuous on bounded sets. Thus, we derive from (71) that Problem 6 is a particular instance of Problem 1.

As in Section 5.1, we denote by (DFB0) and (DFB1) the Algorithm 9 with  $(C_1, C_2) = (K, \mathbb{R}^4)$  and  $(C_1, C_2) = (\mathbb{R}^4, K)$ , respectively. In Tables 2-4, we compare the computational time and number of iterations that algorithms (DFB0), (DFB1), Chambolle-Pock (CP) [26], and Douglas-Rachford (DR) [42] take to approximate a solution for the case when  $\alpha = 1.5$  with a tolerance of  $\varepsilon = h^3$  for the relative error  $\|(\theta_{n+1} - \theta_n, v_{n+1} - v_n)\| / \|(\theta_n, v_n)\|$ . For CP, we use critical step-sizes, i.e.,  $\sigma\gamma = 1$ , where  $\sigma > 0$  and  $\gamma > 0$  are the step-sizes associated with dual and primal variable updates, respectively. This choice is justified by theoretical results ensuring convergence of CP with critical step-sizes [17, 20] and numerical results asserting that the numerical behavior of CP improves as  $\sigma\gamma$  approaches 1 [20]. Moreover, CP and DR show identical numerical behavior, which is justified by the fact that CP coincides with DR for critical step-sizes [26, 47]. We also observe that (DFB0) needs less number of iterations to achieve a relative error of  $\varepsilon = h^3$ , however its computational time is much larger because of the subiterations needed to compute the projection onto  $D_K$ . By considering computational time, (DFB1) is the more efficient algorithm, obtaining a time reduction up to 90% with respect to CP and DR. Moreover, note that the computational time of CP and DR increases as long as the parameter  $\epsilon$  multiplying the entropy penalization increases. On the other hand, (DFB1) has similar computational time and even a smaller number of iterations to achieve the tolerance as  $\epsilon$  increases. We display in Figure 2 the approximation of the equilibrium configurations  $m^h$  that we obtain for different values of  $\epsilon$  and the viscosity parameter  $\nu$ . Comparing the distributions on the left of Figure 2 with those on the right, we observe that, as expected, for larger viscosity parameters the diffusive behavior of equilibria prevails with regard to the individual preferences of the agents.

**Acknowledgments.** The work of Luis M. Briceño-Arias was supported by Agencia Nacional de Investigación y Desarrollo (ANID-Chile) under grants FONDECYT 1190871, FONDECYT 1230257, and by Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for centers of excellence. The work of F. J. Silva was partially supported by l'Agence Nationale de la Recherche (ANR), project ANR-22-CE40-0010, and by KAUST through the subaward agreement ORA-2021-CRG10-4674.6. X. Yang was partly funded by the Air Force Office Scientific Research under MURI award number FA9550-20-1-0358 (Machine Learning and Physics-Based Modeling and Simulation).

## References

- [1] Achdou, Y., Camilli, F., Capuzzo-Dolcetta, I.: Mean field games: numerical methods for the planning problem. *SIAM J. Control Optim.* **50**(1), 77–109 (2012). DOI 10.1137/100790069. URL <http://dx.doi.org/10.1137/100790069>
- [2] Achdou, Y., Camilli, F., Capuzzo-Dolcetta, I.: Mean field games: convergence of a finite difference method. *SIAM J. Numer. Anal.* **51**(5), 2585–2612 (2013). DOI 10.1137/120882421. URL <https://doi.org/10.1137/120882421>

- [3] Achdou, Y., Capuzzo-Dolcetta, I.: Mean field games: numerical methods. *SIAM J. Numer. Anal.* **48**(3), 1136–1162 (2010). DOI 10.1137/090758477
- [4] Achdou, Y., Cardaliaguet, P., Delarue, F., Porretta, A., Santambrogio, F.: Mean field games, *Lecture Notes in Math.*, vol. 2281. Springer, Cham; Centro Internazionale Matematico Estivo (C.I.M.E.), Florence (2020). DOI 10.1007/978-3-030-59837-2. URL <https://doi.org/10.1007/978-3-030-59837-2>
- [5] Achdou, Y., Porretta, A.: 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), 161–186 (2016). DOI 10.1137/15M1015455. URL <https://doi.org/10.1137/15M1015455>
- [6] Almulla, N., Ferreira, R., Gomes, D.: Two numerical approaches to stationary mean-field games. *Dyn. Games Appl.* **7**(4), 657–682 (2017). DOI 10.1007/s13235-016-0203-5. URL <https://doi-org.usm.idm.oclc.org/10.1007/s13235-016-0203-5>
- [7] Baillon, J.B., Haddad, G.: Quelques propriétés des opérateurs angle-bornés et  $n$ -cycliquement monotones. *Israel J. Math.* **26**(2), 137–150 (1977). DOI 10.1007/BF03007664. URL <https://doi.org/10.1007/BF03007664>
- [8] Balakrishnan, A.V.: An operator theoretic formulation of a class of control problems and a steepest descent method of solution. *J. SIAM Control Ser. A* **1**, 109–127 (1963) (1963)
- [9] Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, second edn. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017). DOI 10.1007/978-3-319-48311-5
- [10] Bello Cruz, J.Y., Nghia, T.T.A.: On the convergence of the forward-backward splitting method with linesearches. *Optim. Methods Softw.* **31**(6), 1209–1238 (2016). DOI 10.1080/10556788.2016.1214959. URL <https://doi-org.usm.idm.oclc.org/10.1080/10556788.2016.1214959>
- [11] Bello-Cruz, Y., Li, G., Nghia, T.T.A.: On the linear convergence of forward-backward splitting method: Part I—Convergence analysis. *J. Optim. Theory Appl.* **188**(2), 378–401 (2021). DOI 10.1007/s10957-020-01787-7. URL <https://doi-org.usm.idm.oclc.org/10.1007/s10957-020-01787-7>
- [12] Benamou, J.D., Carlier, G.: Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *J. Optim. Theory Appl.* **167**(1), 1–26 (2015). DOI 10.1007/s10957-015-0725-9
- [13] Benamou, J.D., Carlier, G., Santambrogio, F.: Variational mean field games. In: *Active particles. Vol. 1. Advances in theory, models, and applications, Model.*

Simul. Sci. Eng. Technol., pp. 141–171. Birkhäuser/Springer, Cham (2017)

- [14] Brezis, H., Sibony, M.: Méthodes d’approximation et d’itération pour les opérateurs monotones. *Arch. Rational Mech. Anal.* **28**, 59–82 (1967/1968). DOI 10.1007/BF00281564. URL <https://doi-org.usm.idm.oclc.org/10.1007/BF00281564>
- [15] Briceño-Arias, L., Deride, J., López-Rivera, S., Silva, F.J.: A primal-dual partial inverse algorithm for constrained monotone inclusions: applications to stochastic programming and mean field games. *Appl. Math. Optim.* **87**(2), Paper No. 21, 36 (2023). DOI 10.1007/s00245-022-09921-9. URL <https://doi.org/10.1007/s00245-022-09921-9>
- [16] Briceño-Arias, L., Kalise, D., Kobeissi, Z., Laurière, M., Mateos González, A., Silva, F.J.: On the implementation of a primal-dual algorithm for second order time-dependent mean field games with local couplings. In: *CEMRACS 2017—numerical methods for stochastic models: control, uncertainty quantification, mean-field, ESAIM Proc. Surveys*, vol. 65, pp. 330–348. EDP Sci., Les Ulis (2019). DOI 10.1051/proc/201965330. URL <https://doi.org/10.1051/proc/201965330>
- [17] Briceño-Arias, L., Roldán, F.: Primal-dual splittings as fixed point iterations in the range of linear operators. *J. Global Optim.* **85**(4), 847–866 (2023). DOI 10.1007/s10898-022-01237-w. URL <https://doi-org.usm.idm.oclc.org/10.1007/s10898-022-01237-w>
- [18] Briceño-Arias, L.M., Kalise, D., Silva, F.J.: Proximal methods for stationary mean field games with local couplings. *SIAM J. Control Optim.* **56**(2), 801–836 (2018). DOI 10.1137/16M1095615
- [19] Briceño-Arias, L.M., Pustelnik, N.: Convergence rate comparison of proximal algorithms for non-smooth convex optimization with an application to texture segmentation. *IEEE Signal Processing Letters* **29**, 1337–1341 (2022). DOI 10.1109/LSP.2022.3179169
- [20] Briceño-Arias, L.M., Roldán, F.: Split-Douglas-Rachford algorithm for composite monotone inclusions and split-ADMM. *SIAM J. Optim.* **31**(4), 2987–3013 (2021). DOI 10.1137/21M1395144. URL <https://doi-org.usm.idm.oclc.org/10.1137/21M1395144>
- [21] Cardaliaguet, P.: Weak solutions for first order mean field games with local coupling. In: *Analysis and geometry in control theory and its applications, Springer INdAM Ser.*, vol. 11, pp. 111–158. Springer, Cham (2015). DOI 10.1007/978-3-319-06917-3\_5. URL [https://doi.org/10.1007/978-3-319-06917-3\\_5](https://doi.org/10.1007/978-3-319-06917-3_5)
- [22] Cardaliaguet, P., Lasry, J.M., Lions, P.L., Porretta, A.: Long time average of mean field games. *Netw. Heterog. Media* **7**(2), 279–301 (2012). DOI 10.3934/nhm.2012.7.279. URL <https://doi.org/10.3934/nhm.2012.7.279>



- [23] Cardaliaguet, P., Mészáros, A., Santambrogio, F.: First order mean field games with density constraints: Pressure equals price. *SIAM J. Control Optim.* **54**(5), 2672–2709 (2016). DOI 10.1137/15M1029849
- [24] Carmona, R., Delarue, F.: Probabilistic theory of mean field games with applications. I, *Probability Theory and Stochastic Modelling*, vol. 83. Springer, Cham (2018). Mean field FBSDEs, control, and games
- [25] Carmona, R., Delarue, F.: Probabilistic theory of mean field games with applications. II, *Probability Theory and Stochastic Modelling*, vol. 84. Springer, Cham (2018). Mean field games with common noise and master equations
- [26] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision* **40**(1), 120–145 (2011). DOI 10.1007/s10851-010-0251-1
- [27] Cirant, M.: Multi-population mean field games systems with Neumann boundary conditions. *J. Math. Pures Appl.* (9) **103**(5), 1294–1315 (2015). DOI 10.1016/j.matpur.2014.10.013. URL <https://doi.org/10.1016/j.matpur.2014.10.013>
- [28] Combettes, P.L.: Perspective functions: properties, constructions, and examples. *Set-Valued Var. Anal.* **26**(2), 247–264 (2018). DOI 10.1007/s11228-017-0407-x. URL <https://doi-org.usm.idm.oclc.org/10.1007/s11228-017-0407-x>
- [29] Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **4**(4), 1168–1200 (2005). DOI 10.1137/050626090. URL <https://doi-org.usm.idm.oclc.org/10.1137/050626090>
- [30] Gomes, D.A., Mitake, H.: Existence for stationary mean-field games with congestion and quadratic Hamiltonians. *NoDEA Nonlinear Differential Equations Appl.* **22**(6), 1897–1910 (2015). DOI 10.1007/s00030-015-0349-7. URL <https://doi.org/10.1007/s00030-015-0349-7>
- [31] Gomes, D.A., Patrizi, S., Voskanyan, V.: On the existence of classical solutions for stationary extended mean field games. *Nonlinear Anal.* **99**, 49–79 (2014). DOI 10.1016/j.na.2013.12.016. URL <https://doi.org/10.1016/j.na.2013.12.016>
- [32] Gomes, D.A., Pimentel, E.A., Voskanyan, V.: Regularity theory for mean-field game systems. *SpringerBriefs Math.* Springer (2016). DOI 10.1007/978-3-319-38934-9. URL <https://doi.org/10.1007/978-3-319-38934-9>
- [33] Gomes, D.A., Pires, G.E., Sánchez-Morgado, H.: A-priori estimates for stationary mean-field games. *Netw. Heterog. Media* **7**(2), 303–314 (2012). DOI 10.3934/nhm.2012.7.303. URL <https://doi.org/10.3934/nhm.2012.7.303>
- [34] Gomes, D.A., Saúde, J.: Mean field games models—a brief survey. *Dyn. Games Appl.* **4**(2), 110–154 (2014). DOI 10.1007/s13235-013-0099-2. URL <https://doi.org/10.1007/s13235-013-0099-2>

- [35] Huang, M., Caines, P.E., Malhamé, R.P.: Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized  $\epsilon$ -Nash equilibria. *IEEE Trans. Automat. Control* **52**(9), 1560–1571 (2007). DOI 10.1109/TAC.2007.904450
- [36] Huang, M., Malhamé, R.P., Caines, P.E.: Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.* **6**(3), 221–251 (2006)
- [37] Lasry, J.M., Lions, P.L.: Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris* **343**(9), 619–625 (2006). DOI 10.1016/j.crma.2006.09.019
- [38] Lasry, J.M., Lions, P.L.: Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris* **343**(10), 679–684 (2006). DOI 10.1016/j.crma.2006.09.018. URL <https://doi.org/10.1016/j.crma.2006.09.018>
- [39] Lasry, J.M., Lions, P.L.: Mean field games. *Jpn. J. Math.* **2**(1), 229–260 (2007). DOI 10.1007/s11537-007-0657-8
- [40] Lavigne, P., Pfeiffer, L.: Generalized conditional gradient and learning in potential mean field games (2022). Preprint, arXiv:2209.12772
- [41] Levitin, E.S., Polyak, B.T.: Constrained minimization methods. *USSR Computational mathematics and mathematical physics* **6**(5), 1–50 (1966)
- [42] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16**(6), 964–979 (1979). DOI 10.1137/0716071
- [43] Liu, S., Jacobs, M., Li, W., Nurbekyan, L., Osher, S.J.: Computational Methods for First-Order Nonlocal Mean Field Games with Applications. *SIAM J. Numer. Anal.* **59**(5), 2639–2668 (2021). DOI 10.1137/20M1334668
- [44] Mercier, B.: Inéquations variationnelles de la mécanique, *Publications Mathématiques d’Orsay 80 [Mathematical Publications of Orsay 80]*, vol. 1. Université de Paris-Sud, Département de Mathématique, Orsay (1980)
- [45] Mészáros, A.R., Silva, F.J.: A variational approach to second order mean field games with density constraints: the stationary case. *J. Math. Pures Appl.* (9) **104**(6), 1135–1159 (2015). DOI 10.1016/j.matpur.2015.07.008. URL <https://doi.org/10.1016/j.matpur.2015.07.008>
- [46] Mészáros, A.R., Silva, F.J.: On the variational formulation of some stationary second-order mean field games systems. *SIAM J. Math. Anal.* **50**(1), 1255–1277 (2018). DOI 10.1137/17M1125960

- [47] O'Connor, D., Vandenberghe, L.: On the equivalence of the primal-dual hybrid gradient method and Douglas-Rachford splitting. *Math. Program.* **179**(1-2, Ser. A), 85–108 (2020). DOI 10.1007/s10107-018-1321-1. URL <https://doi-org.usm.idm.oclc.org/10.1007/s10107-018-1321-1>
- [48] Papadakis, N., Peyré, G., Oudet, E.: Optimal transport with proximal splitting. *SIAM J. Imaging Sci.* **7**(1), 212–238 (2014). DOI 10.1137/130920058. URL <https://doi.org/10.1137/130920058>
- [49] Pérez-Aros, P., Vilches, E.: An enhanced Baillon-Haddad theorem for convex functions defined on convex sets. *Appl. Math. Optim.* **83**(3), 2241–2252 (2021). DOI 10.1007/s00245-019-09626-6. URL <https://doi.org/10.1007/s00245-019-09626-6>
- [50] Pimentel, E.A., Voskanyan, V.: Regularity for second-order stationary mean-field games. *Indiana Univ. Math. J.* **66**(1), 1–22 (2017). DOI 10.1512/iumj.2017.66.5944. URL <https://doi.org/10.1512/iumj.2017.66.5944>
- [51] Rockafellar, R.T.: Level sets and continuity of conjugate convex functions. *Trans. Amer. Math. Soc.* **123**, 46–63 (1966). DOI 10.2307/1994612. URL <https://doi.org/10.2307/1994612>
- [52] Salzo, S.: The variable metric forward-backward splitting algorithm under mild differentiability assumptions. *SIAM J. Optim.* **27**(4), 2153–2181 (2017). DOI 10.1137/16M1073741. URL <https://doi-org.usm.idm.oclc.org/10.1137/16M1073741>
- [53] Sibony, M.: Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone. *Calcolo* **7**, 65–183 (1970). DOI 10.1007/BF02575559. URL <https://doi-org.usm.idm.oclc.org/10.1007/BF02575559>
- [54] Taylor, A.B., Hendrickx, J.M., Glineur, F.: Exact worst-case convergence rates of the proximal gradient method for composite convex minimization. *J. Optim. Theory Appl.* **178**(2), 455–476 (2018). DOI 10.1007/s10957-018-1298-1. URL <https://doi-org.usm.idm.oclc.org/10.1007/s10957-018-1298-1>