

MONOTONE INCLUSION METHODS FOR A CLASS OF SECOND-ORDER NON-POTENTIAL MEAN-FIELD GAMES*

LEVON NURBEKYAN[†], SITING LIU[‡], AND YAT TIN CHOW[§]

Abstract. We propose a monotone splitting algorithm for solving a class of second-order non-potential mean-field games. Following [1], we introduce a finite-difference scheme and observe that the scheme represents first-order optimality conditions for a primal-dual pair of monotone inclusions. Based on this observation, we prove that the finite-difference system obtains a solution that can be provably recovered by an extension of the celebrated primal-dual hybrid gradient (PDHG) algorithm.

Key words. mean-field games, non-potential, monotone inclusions, primal-dual methods, finite-differences

MSC codes. Primary, 35Q89, 65M06, 35A15, 49N80; Secondary, 35Q91, 35Q93, 91A16, 93A15, 93A16

1. Introduction. The main goal of the paper is to introduce a new algorithm for computing the solutions of the following system of PDE

$$(1.1) \quad \begin{cases} -\partial_t \phi - \nu \Delta \phi + H(t, x, \nabla \phi, \rho) = f(t, x, \rho), \\ \partial_t \rho - \nu \Delta \rho - \nabla \cdot (\rho \nabla_q H(t, x, \nabla \phi, \rho)) = 0, \\ \rho(0, x) = \rho_0(x), \quad \phi(T, x) = g(x, \rho(T, \cdot)). \end{cases}$$

Here, we assume periodic boundary conditions; that is, $(t, x) \in [0, T] \times \mathbb{T}^d$, where $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ is the d -dimensional flat torus. Furthermore, $\nu \geq 0$ is the noise (viscosity) parameter, H is the Hamiltonian, f is the mean-field coupling function, and g is the terminal cost function.

System (1.1) characterizes an equilibrium configuration for a continuum of agents that play a non-cooperative differential game. Such games are called mean-field games (MFG) and were independently introduced in [17, 18, 19] and [15, 14]. In this context, $\rho(t, x)$ and $\phi(t, x)$ are, respectively, the distribution and optimal cost of the agents at time t and location x .

We introduce a numerical method for (1.1) in the spirit of Benamou-Brenier technique for solving optimal transportation and MFG systems [5, 6, 7, 9, 8]. In these works, the authors observe that when H is separable; that is, $H(x, \rho, q) = H_0(x, q) - f(x, \rho)$, and suitable convexity, monotonicity, and structural assumptions are met, (1.1) can be seen as a first-order optimality condition for a convex-concave saddle-point problem. Such MFG are called *variational* or *potential*. Hence, one can use various convex optimization algorithms for computing the solutions of (1.1) such as the alternating direction method of multipliers (ADMM) [5, 6, 7] and primal-dual hybrid gradient (PDHG) [9, 8].

*Submitted to the editors DATE.

Funding: S. Liu was supported by Air Force Office of Scientific Research (AFOSR) MURI Grant FA9550-18-502 and Office of Naval Research (ONR) Grant N00014-20-1-2787; Y. T. Chow was supported by the Regents Faculty Fellowship and Omnibus Research and Travel Grant from the University of California Riverside.

[†]Department of Mathematics, Emory University, Atlanta, GA (lnurbek@emory.edu, <https://sites.google.com/view/lnurbek/home>).

[‡]Department of Mathematics, University of California Los Angeles, Los Angeles, CA (siting6@math.ucla.edu, <https://sites.google.com/view/siting6ucla/home>).

[§]Department of Mathematics, University of California Riverside, Riverside, CA (yattinc@ucr.edu, <https://ytchow.github.io>).

Here, we go beyond the potential and separable settings and provide a version of PDHG algorithm to solve (1.1). Our essential observation is that under the so-called Lasry-Lions monotonicity condition (1.1) can be seen as a primal-dual pair of monotone inclusions where the monotone maps are not subdifferential maps in general. We then solve (1.1) by a PDHG variant for monotone inclusions [22]. The possibility of solving non-potential MFG using monotone inclusion variants of PDHG was hinted on in [21] for nonlocal MFG systems.

For related work on non-separable MFG systems we refer to [20, 10] for policy iteration and Newton's methods, and [3, 13, 12] for monotone flows. In [20, 10], the analyses strongly rely on the ellipticity and do not handle first-order systems ($\nu = 0$) and non-smooth mean-field interactions ($\epsilon = 0$ in (2.7)). In contrast, we expect our methods to extend to these singular cases *mutatis mutandis* due to the variational nature of our techniques. Numerical experiments in subsection 4.2 and subsection 4.3 support our claim.

Works [3, 13, 12] address finite-state and stationary problems. Additionally, the algorithms in these works are forward only and do not come with convergence guarantees for the time-discretizations of the monotone flows.

2. Numerical analysis.

2.1. A finite-difference scheme. We follow [1, 8] for introducing a semi-implicit scheme for (1.1). For simplicity, we assume that $d = 2$. Let $h, \Delta t > 0$ be such that $N_h = \frac{1}{h} \in \mathbb{N}$ and $N_T = \frac{T}{\Delta t} \in \mathbb{N}$. We then introduce uniform space-time grids

$$x_{ij} = (ih, jh), \quad t_k = k\Delta t, \quad 0 \leq i, j \leq N_h - 1, \quad 0 \leq k \leq N_T.$$

To enforce periodicity, we assume that $x_{ij} = x_{i'j'}$ whenever $i \equiv i' \pmod{N_h}$ and $j \equiv j' \pmod{N_h}$. Next, for a grid-function $f_{ij}^k = f(x_{ij}, t_k)$ we denote by

$$(D_1 f^k)_{ij} = \frac{f_{(i+1)j}^k - f_{ij}^k}{h}, \quad (D_2 f^k)_{ij} = \frac{f_{i(j+1)}^k - f_{ij}^k}{h},$$

$$[D_h f^k]_{ij} = ((D_1 f^k)_{ij}, (D_1 f^k)_{(i-1)j}, (D_2 f^k)_{ij}, (D_2 f^k)_{i(j-1)}),$$

$$(\Delta_h f^k)_{ij} = \frac{f_{(i-1)j}^k + f_{(i+1)j}^k + f_{i(j-1)}^k + f_{i(j+1)}^k - 4f_{ij}^k}{4h^2},$$

$$(D_t f_{ij})^k = \frac{f_{ij}^{k+1} - f_{ij}^k}{\Delta t}.$$

We then introduce a discretization of the Hamiltonian $H_h(t, x, q_1, q_2, q_3, q_4, \rho)$ that satisfies the following conditions

- *Monotonicity:* H_h is nonincreasing in q_1, q_3 and nondecreasing in q_2, q_4 .
- *Consistency:* $H_h(t, x, q_1, q_1, q_2, q_2, \rho) = H(t, x, q, \rho)$ for all $t \in [0, T]$, $x \in \mathbb{T}^2$, $\rho \geq 0$, and $q = (q_1, q_2) \in \mathbb{R}^2$.
- *Differentiability:* $\nabla_q H_h$ is continuous.
- *Convexity:* $(q_1, q_2, q_3, q_4) \mapsto H_h(t, x, q_1, q_2, q_3, q_4, \rho)$ is convex.
- *Lasry-Lions monotonicity:* A structural condition that yields existence and uniqueness of solutions of (1.1) is the so called Lasry-Lions monotonicity;

that is,

$$(2.1) \quad \begin{pmatrix} -\partial_\rho H(t, x, q, \rho) & \frac{\rho \nabla_q^\top H(t, x, q, \rho)}{2} \\ \frac{\rho \nabla_q H(t, x, q, \rho)}{2} & \rho \nabla_q^2 H(t, x, q, \rho) \end{pmatrix} \geq 0,$$

for all $t \in [0, T]$, $(x, q) \in \mathbb{T}^2 \times \mathbb{R}^2$, $\rho \geq 0$. We require the same condition on the discretized Hamiltonian; that is,

$$(2.2) \quad \begin{pmatrix} -\partial_\rho H_h(t, x, q_1, q_2, q_3, q_4, \rho) & \frac{\rho \nabla_q^\top H(t, x, q_1, q_2, q_3, q_4, \rho)}{2} \\ \frac{\rho \nabla_q H(t, x, q_1, q_2, q_3, q_4, \rho)}{2} & \rho \nabla_q^2 H(t, x, q_1, q_2, q_3, q_4, \rho) \end{pmatrix} \geq 0,$$

for all $t \in [0, T]$, $(x, q_1, q_2, q_3, q_4) \in \mathbb{T}^2 \times \mathbb{R}^4$, $\rho \geq 0$.

With these ingredients at hand, the discretization of (1.1) introduced in [1] reads as follows:

$$(2.3) \quad \begin{cases} -(D_t \phi_{ij})^k - \nu(\Delta_h \phi^k)_{ij} + H_h(t_k, x_{ij}, [D_h \phi^k]_{ij}, \rho_{ij}^{k+1}) = f(t_k, x_{ij}, \rho_{ij}^{k+1}) \\ (D_t \rho_{ij})^k - \nu(\Delta_h \rho^{k+1})_{ij} - \mathcal{B}_{ij}(t_k, \phi^k, \rho^{k+1}, \rho^{k+1}) = 0, \\ \rho_{ij}^0 = (\rho_0)_{ij}, \quad \phi_{ij}^{N_T} = g(x_{ij}, \rho_{ij}^{N_T}), \quad 0 \leq i, j \leq N_h - 1, \quad 0 \leq k \leq N_T - 1, \end{cases}$$

where

$$(\rho_0)_{ij} = \frac{1}{h^2} \int_{\|x - x_{ij}\|_\infty \leq h/2} \rho_0(x) dx.$$

and

$$\begin{aligned} & \mathcal{B}_{ij}(t, \phi, \rho, \eta) \\ &= \frac{1}{h} (\rho_{ij} \partial_{q_1} H_h(t, x_{ij}, [D_h \phi]_{ij}, \eta_{ij}) - \rho_{(i-1)j} \partial_{q_1} H_h(t, x_{(i-1)j}, [D_h \phi]_{(i-1)j}, \eta_{(i-1)j})) \\ &+ \frac{1}{h} (\rho_{(i+1)j} \partial_{q_2} H_h(t, x_{(i+1)j}, [D_h \phi]_{(i+1)j}, \eta_{(i+1)j}) - \rho_{ij} \partial_{q_2} H_h(t, x_{ij}, [D_h \phi]_{ij}, \eta_{ij})) \\ &+ \frac{1}{h} (\rho_{ij} \partial_{q_3} H_h(t, x_{ij}, [D_h \phi]_{ij}, \eta_{ij}) - \rho_{i(j-1)} \partial_{q_3} H_h(t, x_{i(j-1)}, [D_h \phi]_{i(j-1)}, \eta_{i(j-1)})) \\ &+ \frac{1}{h} (\rho_{i(j+1)} \partial_{q_4} H_h(t, x_{i(j+1)}, [D_h \phi]_{i(j+1)}, \eta_{i(j+1)}) - \rho_{ij} \partial_{q_4} H_h(t, x_{ij}, [D_h \phi]_{ij}, \eta_{ij})). \end{aligned}$$

2.2. A discrete energy. Our goal is to formulate the discrete system (2.3) as a primal-dual pair of monotone inclusions. Following [8], we denote by $\mathcal{M} = \mathbb{R}^{(N_T+1) \times N_h \times N_h}$, $\mathcal{W} = (\mathbb{R}^4)^{N_T \times N_h \times N_h}$, $\mathcal{U} = \mathbb{R}^{N_T \times N_h \times N_h}$ and introduce operators $A : \mathcal{M} \rightarrow \mathcal{U}$, $B : \mathcal{W} \rightarrow \mathcal{U}$ as follows:

$$\begin{aligned} (A\rho)_{ij}^k &= (D_t \rho_{ij})^k - \nu(\Delta_h \rho^{k+1})_{ij}, \\ (Bw)_{ij}^k &= (D_1 w^{k,1})_{(i-1)j} + (D_1 w^{k,2})_{ij} + (D_2 w^{k,3})_{i(j-1)} + (D_2 w^{k,4})_{ij}, \end{aligned}$$

for $0 \leq i, j \leq N_h - 1$ and $0 \leq k \leq N_T - 1$. Direct calculations [1, 8] show that

$$\begin{aligned} (B^* \phi)_{ij}^k &= -[D_h \phi^k]_{ij}, \quad 0 \leq k \leq N_T - 1, \\ (A^* \phi)_{ij}^0 &= -\frac{1}{\Delta t} \phi_{ij}^0, \\ (A^* \phi)_{ij}^k &= -(D_t \phi_{ij})^{k-1} - \nu(\Delta_h \phi^{k-1})_{ij}, \quad 1 \leq k \leq N_T - 1, \\ (A^* \phi)_{ij}^{N_T} &= \frac{\phi_{ij}^{N_T-1}}{\Delta t} - \nu(\Delta_h \phi^{N_T-1})_{ij}. \end{aligned}$$

Furthermore, we denote by L_h the Legendre dual of H_h defined as

$$(2.4) \quad L_h(t, x, v_1, v_2, v_3, v_4, \eta) = \sup_{q_1, q_2, q_3, q_4} \left(- \sum_{i=1}^4 q_i v_i - H_h(t, x, q_1, q_2, q_3, q_4, \eta) \right),$$

and denote by E_h the *perspective function* [11] of L_h ; that is,

$$(2.5) \quad E_h(t, x, \rho, w, \eta) = \begin{cases} \rho L_h \left(t, x, \frac{w}{\rho}, \eta \right), & \rho > 0, \\ (\text{rec } L_h(t, x, \cdot, \eta))(w), & \rho = 0, \\ \infty, & \text{otherwise,} \end{cases}$$

where the *recession function* $\text{rec } L_h(t, x, \cdot, \eta)$ is defined as

$$(\text{rec } L_h(t, x, \cdot, \eta))(w) = \lim_{s \rightarrow \infty} \frac{L_h(t, x, v_0 + sw, \eta)}{s}.$$

Above, $v_0 \in \mathbb{R}^4$ is an arbitrary point such that $L_h(t, x, v_0, \eta) < \infty$.

Next, we define $J : \mathcal{M} \times \mathcal{W} \times \mathcal{M} \rightarrow \mathbb{R} \cup \{\infty\}$ as follows:

$$(2.6) \quad \begin{aligned} J(\rho, w, \eta) = & \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} E_h(t_k, x_{ij}, \rho_{ij}^{k+1}, w_{ij}^k, \eta_{ij}^{k+1}) + \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} F(t_k, x_{ij}, \rho_{ij}^{k+1}) \\ & + \frac{1}{\Delta t} \sum_{0 \leq i, j \leq N_h - 1} G(x_{ij}, \rho_{ij}^{N_T}) + \mathbf{1}_{\rho^0 = \rho_0}, \end{aligned}$$

where $\partial_\rho F(t, x, \rho) = f(t, x, \rho)$, and $\partial_\rho G(x, \rho) = g(x, \rho)$. Additionally, $\mathbf{1}_E$ is the convex characteristic function defined as

$$\mathbf{1}_E(y) = \begin{cases} 0, & y \in E, \\ \infty, & y \notin E. \end{cases}$$

Remark 2.1. A remarkable property of perspective functions is that $E_h(t, x, \cdot, \cdot, \eta)$ is a convex lower semicontinuous function [11].

2.3. A congestion model. For concreteness, we consider the MFG model with congestion discussed in [2]. More specifically, assume that

$$(2.7) \quad H(t, x, q, \eta) = \frac{|q|^\beta}{\beta(\eta + \epsilon)^\alpha},$$

for some $\epsilon > 0$, and

$$(2.8) \quad 1 < \beta \leq 2, \quad 0 < \alpha \leq \frac{4(\beta - 1)}{\beta}.$$

These conditions together with the monotonicity of $\rho \mapsto f(t, x, \rho)$ and $\rho \mapsto g(x, \rho)$, and suitable technical assumptions yield the existence and uniqueness of weak solutions for (1.1) as analyzed in [2].

Following [1, 8], we discretize H in (2.7) as follows

$$(2.9) \quad H_h(t, x, q_1, q_2, q_3, q_4, \eta) = \frac{((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2)^{\beta/2}}{\beta(\eta + \epsilon)^\alpha}$$

and denote by $K = \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_-$.

LEMMA 2.2. *Assume that H_h is given by (2.9). Then for every $(t, x) \in \mathbb{R} \times \mathbb{T}^d$ and $\eta \geq 0$ we have that*

$$(2.10) \quad L_h(t, x, v_1, v_2, v_3, v_4, \eta) = \begin{cases} (\eta + \epsilon)^{\alpha(\beta' - 1)} \frac{(v_1^2 + v_2^2 + v_3^2 + v_4^2)^{\beta'/2}}{\beta'}, & (v_1, v_2, v_3, v_4) \in K, \\ \infty, & \text{otherwise,} \end{cases}$$

where $\beta' = \frac{\beta}{\beta - 1}$. Furthermore, we have that

$$(2.11) \quad E_h(t, x, \rho, w, \eta) = \begin{cases} (\eta + \epsilon)^{\alpha(\beta' - 1)} \frac{|w|^{\beta'}}{\beta' \rho^{\beta' - 1}}, & \rho > 0, w \in K, \\ 0, & (\rho, w) = (0, 0), \\ \infty, & \text{otherwise,} \end{cases}$$

where $w = (w_1, w_2, w_3, w_4)$.

Proof. We start by computing L_h . Suppose that $(v_1, v_2, v_3, v_4) \notin K$; for instance, let $v_1 < 0$. Then we have that

$$L_h(t, x, v_1, v_2, v_3, v_4, \eta) \geq \sup_{q_1 > 0} (-q_1 v_1 - H_h(t, x, q_1, 0, 0, 0, \eta)) = \sup_{q_1 > 0} (-q_1 v_1) = \infty.$$

Now assume that $(v_1, v_2, v_3, v_4) \in K$. The first order optimality conditions in the concave program (2.4) yield

$$v_i = -\partial_{q_i} H_h(t, x, q_1, q_2, q_3, q_4, \eta), \quad 1 \leq i \leq 4.$$

Furthermore, we have that

$$\begin{aligned} \partial_{q_i} H_h(t, x, q_1, q_2, q_3, q_4, \eta) &= -\frac{((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2)^{\beta/2 - 1}}{(\eta + \epsilon)^\alpha} q_i^-, \quad i = 1, 3, \\ \partial_{q_i} H_h(t, x, q_1, q_2, q_3, q_4, \eta) &= \frac{((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2)^{\beta/2 - 1}}{(\eta + \epsilon)^\alpha} q_i^+, \quad i = 2, 4. \end{aligned}$$

Hence, the first order optimality conditions yield

$$\begin{aligned} v_i &= \frac{((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2)^{\beta/2 - 1}}{(\eta + \epsilon)^\alpha} q_i^- \geq 0, \quad i = 1, 3, \\ v_i &= -\frac{((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2)^{\beta/2 - 1}}{(\eta + \epsilon)^\alpha} q_i^+ \leq 0, \quad i = 2, 4, \end{aligned}$$

and

$$\begin{aligned} v_1^2 + v_2^2 + v_3^2 + v_4^2 &= \frac{((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2)^{\beta - 1}}{(\eta + \epsilon)^{2\alpha}}, \\ -q_1 v_1 - q_2 v_2 - q_3 v_3 - q_4 v_4 &= \frac{((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2)^{\beta/2}}{(\eta + \epsilon)^\alpha}. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} L_h(t, x, v_1, v_2, v_3, v_4) &= \left(1 - \frac{1}{\beta}\right) \frac{((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2)^{\beta/2}}{(\eta + \epsilon)^\alpha} \\ &= \left(1 - \frac{1}{\beta}\right) \frac{(\eta + \epsilon)^{\frac{\alpha\beta}{\beta - 1}} (v_1^2 + v_2^2 + v_3^2 + v_4^2)^{\frac{\beta}{2(\beta - 1)}}}{(\eta + \epsilon)^\alpha} \\ &= (\eta + \epsilon)^{\alpha(\beta' - 1)} \frac{(v_1^2 + v_2^2 + v_3^2 + v_4^2)^{\beta'/2}}{\beta'}. \end{aligned}$$

Next, we need to compute the recession function of $L_h(t, x, \cdot, \eta)$. Since

$$L_h(t, x, 0, 0, 0, 0, \eta) = 0,$$

we have that $v_0 = (0, 0, 0, 0) \in \text{dom } L_h(t, x, \cdot, \eta)$, and so we can compute the recession function via

$$(\text{rec } L_h(t, x, \cdot, \eta))(w) = \lim_{s \rightarrow \infty} \frac{L_h(t, x, sw, \eta)}{s}.$$

Since $L_h(t, x, w, \eta) = \infty$ for $w \notin K$ we have that

$$(\text{rec } L_h(t, x, \cdot, \eta))(w) = \infty, \quad \forall w \notin K.$$

Furthermore, for $w \in K$ we have that

$$\begin{aligned} (\text{rec } L_h(t, x, \cdot, \eta))(w) &= \lim_{s \rightarrow \infty} \frac{L_h(t, x, sw, \eta)}{s} = \lim_{s \rightarrow \infty} (\eta + \epsilon)^{\alpha(\beta' - 1)} \frac{s^{\beta'} |w|^{\beta'}}{s} \\ &= \begin{cases} 0, & w = 0, \\ \infty, & w \neq 0. \end{cases} \end{aligned}$$

Summarizing, we find that

$$(\text{rec } L_h(t, x, \cdot, \eta))(w) = \begin{cases} 0, & w = 0, \\ \infty, & w \neq 0, \end{cases}$$

and (2.11) follows readily. \square

In [2], the authors point out that conditions (2.8) yield that H satisfies the Lasry-Lions monotonicity condition (2.1). Here, we show that H_h preserves that property.

LEMMA 2.3. *Let H_h be as in (2.9), and (2.8) hold. Then H_h satisfies the Lasry-Lions monotonicity condition (2.2).*

Proof. Denoting by

$$(2.12) \quad \psi(q_1, q_2, q_3, q_4) = \frac{((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2)^{\beta/2}}{\beta},$$

we have that

$$H_h(t, x, q, \rho) = (\rho + \epsilon)^{-\alpha} \psi(q),$$

where we denote by $q = (q_1, q_2, q_3, q_4)$. Then we have that

$$\begin{aligned} \partial_\rho H_h &= -\alpha(\rho + \epsilon)^{-\alpha-1} \psi(q), & \nabla_q H_h &= (\rho + \epsilon)^{-\alpha} \nabla \psi(q), \\ \rho \nabla_q \partial_\rho H_h &= -\alpha \rho (\rho + \epsilon)^{-\alpha-1} \nabla \psi(q), & \rho \nabla_q^2 H_h &= \rho (\rho + \epsilon)^\alpha \nabla^2 \psi(q). \end{aligned}$$

Hence, we have to prove that

$$\begin{pmatrix} \alpha(\rho + \epsilon)^{-\alpha-1} \psi(q) & -\frac{1}{2} \alpha \rho (\rho + \epsilon)^{-\alpha-1} \nabla^\top \psi(q) \\ -\frac{1}{2} \alpha \rho (\rho + \epsilon)^{-\alpha-1} \nabla \psi(q) & \rho (\rho + \epsilon)^\alpha \nabla^2 \psi(q) \end{pmatrix} \geq 0$$

for all $\rho \geq 0$ and $q \in \mathbb{R}^4$. Note that the inequality is trivial when $\rho = 0$ because $\psi(q) \geq 0$ for all $q \in \mathbb{R}^4$. Hence, we can assume that $\rho > 0$. Factoring out a positive number $\alpha\rho(\rho + \epsilon)^{-\alpha-1}$ we arrive at an equivalent inequality

$$\begin{pmatrix} \frac{\psi(q)}{2} & -\frac{\nabla^\top \psi(q)}{2} \\ -\frac{\nabla \psi(q)}{2} & \frac{\rho + \epsilon}{\alpha} \nabla^2 \psi(q) \end{pmatrix} \geq 0.$$

Since ψ is convex, we have that $\nabla^2 \psi(q) \geq 0$. Additionally, from (2.8) we have that

$$\frac{\rho + \epsilon}{\alpha} \geq \frac{\beta \rho}{4(\beta - 1)},$$

and so it is sufficient to prove that

$$\begin{pmatrix} \frac{\psi(q)}{2} & -\frac{\nabla^\top \psi(q)}{2} \\ -\frac{\nabla \psi(q)}{2} & \frac{\beta \rho}{4(\beta - 1)} \nabla^2 \psi(q) \end{pmatrix} \geq 0.$$

Next, since the previous inequality must hold for all $q \in \mathbb{R}^4$, we obtain an equivalent statement if we replace q by ρq ; that is,

$$\begin{pmatrix} \frac{\psi(\rho q)}{2} & -\frac{\nabla^\top \psi(\rho q)}{2} \\ -\frac{\nabla \psi(\rho q)}{2} & \frac{\beta \rho}{4(\beta - 1)} \nabla^2 \psi(\rho q) \end{pmatrix} \geq 0, \quad \forall \rho > 0, q \in \mathbb{R}^4.$$

Denoting by $\psi_\rho(q) = \frac{\psi(\rho q)}{\rho}$, we find that the previous expression is nothing but

$$\begin{pmatrix} \frac{\psi_\rho(q)}{2} & -\frac{\nabla^\top \psi_\rho(q)}{2} \\ -\frac{\nabla \psi_\rho(q)}{2} & \frac{\beta}{4(\beta - 1)} \nabla^2 \psi_\rho(q) \end{pmatrix} \geq 0, \quad \forall \rho > 0, q \in \mathbb{R}^4.$$

Furthermore, due to the homogeneity of ψ , we have that $\psi_\rho(q) = \rho^{\beta-1} \psi(q)$, and so

$$\begin{pmatrix} \frac{\psi_\rho(q)}{2} & -\frac{\nabla^\top \psi_\rho(q)}{2} \\ -\frac{\nabla \psi_\rho(q)}{2} & \frac{\beta}{4(\beta - 1)} \nabla^2 \psi_\rho(q) \end{pmatrix} = \rho^{\beta-1} \begin{pmatrix} \frac{\psi(q)}{2} & -\frac{\nabla^\top \psi(q)}{2} \\ -\frac{\nabla \psi(q)}{2} & \frac{\beta}{4(\beta - 1)} \nabla^2 \psi(q) \end{pmatrix}.$$

The latter means that we need to prove that

$$\begin{pmatrix} \frac{\psi(q)}{2} & -\frac{\nabla^\top \psi(q)}{2} \\ -\frac{\nabla \psi(q)}{2} & \frac{\beta}{4(\beta - 1)} \nabla^2 \psi(q) \end{pmatrix} \geq 0, \quad \forall q \in \mathbb{R}^4.$$

For every $(\xi, \chi) \in \mathbb{R} \times \mathbb{R}^4$ we have to prove that

$$\begin{aligned} & (\xi, \chi)^\top \begin{pmatrix} \frac{\psi(q)}{2} & -\frac{\nabla^\top \psi(q)}{2} \\ -\frac{\nabla \psi(q)}{2} & \frac{\beta}{4(\beta - 1)} \nabla^2 \psi(q) \end{pmatrix} (\xi, \chi) \\ &= \xi^2 \psi(q) - \xi \nabla \psi(q) \cdot \chi + \frac{\beta}{4(\beta - 1)} \chi^\top \nabla^2 \psi(q) \chi \geq 0 \end{aligned}$$

When $\psi(q) = 0$ we have that $\nabla \psi(q) = 0$, and the inequality follows from $\nabla^2 \psi(q) \geq 0$. Hence, assuming that $\psi(q) > 0$ and optimizing with respect to ξ , we obtain an equivalent inequality

$$-\frac{|\nabla \psi(q) \cdot \chi|^2}{4\psi(q)} + \frac{\beta}{4(\beta - 1)} \chi^\top \nabla^2 \psi(q) \chi \geq 0,$$

or

$$\chi^\top \left(\frac{\beta}{\beta-1} \psi(q) \nabla^2 \psi(q) - \nabla \psi(q) \otimes \nabla \psi(q) \right) \chi \geq 0.$$

Denoting by

$$\tilde{q} = (-q_1^-, q_2^+, -q_3^-, q_4^+), \quad \text{for } q = (q_1, q_2, q_3, q_4),$$

we have that

$$\nabla \psi(q) = |\tilde{q}|^{\beta-2} \tilde{q}, \quad \nabla^2 \psi(q) = (\beta-2) |\tilde{q}|^{\beta-4} \tilde{q} \otimes \tilde{q} + |\tilde{q}|^{\beta-2} M(q),$$

where

$$M(q) = \text{diag}(\text{Hvs}(-q_1), \text{Hvs}(q_2), \text{Hvs}(-q_3), \text{Hvs}(q_4)),$$

and Hvs is the Heaviside step function. Hence we obtain that

$$\begin{aligned} & \frac{\beta}{\beta-1} \psi(q) \nabla^2 \psi(q) - \nabla \psi(q) \otimes \nabla \psi(q) \\ &= \frac{\beta-2}{\beta-1} |\tilde{q}|^{2\beta-4} \tilde{q} \otimes \tilde{q} + \frac{1}{\beta-1} |\tilde{q}|^{2\beta-2} M(q) - |\tilde{q}|^{2\beta-4} \tilde{q} \otimes \tilde{q} \\ &= \frac{|\tilde{q}|^{2\beta-4}}{\beta-1} (|\tilde{q}|^2 M(q) - \tilde{q} \otimes \tilde{q}), \end{aligned}$$

and therefore we need to prove that

$$\begin{aligned} & ((q_1^-)^2 + (q_2^+)^2 + (q_3^-)^2 + (q_4^+)^2) \\ & \times (\chi_1^2 \text{Hvs}(-q_1) + \chi_2^2 \text{Hvs}(q_2) + \chi_3^2 \text{Hvs}(-q_3) + \chi_4^2 \text{Hvs}(q_4)) \\ & \geq (-q_1^- \chi_1 + q_2^+ \chi_2 - q_3^- \chi_3 + q_4^+ \chi_4)^2. \end{aligned}$$

The latter simply follows from the Cauchy-Schwarz inequality

$$\sum_{i=1}^4 a_i^2 \cdot \sum_{i=1}^4 b_i^2 \geq \left(\sum_{i=1}^4 a_i b_i \right)^2$$

applied to

$$\begin{aligned} a_1 &= -q_1^-, \quad a_2 = q_2^+, \quad a_3 = -q_3^-, \quad a_4 = q_4^+, \\ b_1 &= \chi_1 \text{Hvs}(-q_1), \quad b_2 = \chi_2 \text{Hvs}(q_2), \quad b_3 = \chi_3 \text{Hvs}(-q_3), \quad b_4 = \chi_4 \text{Hvs}(q_4), \end{aligned}$$

and taking into account that

$$q^- \text{Hvs}(-q) = q^-, \quad q^+ \text{Hvs}(q) = q^+, \quad \text{Hvs}(q)^2 = \text{Hvs}(q), \quad \forall q \in \mathbb{R}.$$

This completes the proof. \square

2.4. Properties of the discrete energy. Here, we discuss key properties of the function J defined in (2.6).

LEMMA 2.4. Assume that $\eta \geq 0$, and $F(t, x, \cdot)$, $G(x, \cdot)$ are convex and continuous for $\rho \geq 0$. Furthermore, for $(\lambda, y, z) \in \{0, 1\} \times \mathcal{M} \times \mathcal{W}$ denote by

$$(2.13) \quad J_{\lambda, y, z}(\rho, w, \eta) = \lambda \left(\frac{|\rho|^2}{2} - \rho \cdot y + \frac{|w|^2}{2} - w \cdot z \right) + J(\rho, w, \eta).$$

Then the following statements are true.

1. The convex program

$$(2.14) \quad \inf_{(1-\lambda)(A\rho+Bw)=0} J_{\lambda, y, z}(\rho, w, \eta)$$

admits a minimizer for all $(\lambda, y, z) \in \{0, 1\} \times \mathcal{M} \times \mathcal{W}$.

2. When $\nu > 0$ and $\lambda = 0$ all minimizers of (2.14) satisfy

$$(2.15) \quad \rho_{ij}^{k+1} > 0, \quad \forall 0 \leq i, j \leq N_h - 1, \quad 0 \leq k \leq N_T - 1.$$

Remark 2.5. The convex program (2.14) is analogous to the problem $(P_{h, \Delta t})$ in [8] with the difference of having extra parameters λ, y, z, η . Hence, the proof techniques used in [8, Theorem 3.1] and auxiliary lemmas extend to setting of Lemma 2.4.

Remark 2.6. The convex program (2.14) is a shorthand for considering two programs

$$\inf_{A\rho+Bw=0} J(\rho, w, \eta)$$

and

$$\inf_{\rho, w} \left\{ \frac{|\rho|^2}{2} - \rho \cdot y + \frac{|w|^2}{2} - w \cdot z + J(\rho, w, \eta) \right\}$$

simultaneously. Both are important for our further analysis.

Proof of Lemma 2.4. 1. We argue by the direct method of calculus of variations. Fix a triple $(\lambda, y, z) \in \{0, 1\} \times \mathcal{M} \times \mathcal{W}$. According to [8, Lemma 3.1] there exists $(\tilde{\rho}, \tilde{w}) \in \mathcal{M} \times \mathcal{W}$ such that

$$(2.16) \quad \tilde{\rho}_{ij}^{k+1} > 0, \quad \tilde{w}_{ij}^k \in \text{int}(K), \quad \forall 0 \leq i, j \leq N_h - 1, \quad 0 \leq k \leq N_T - 1,$$

and

$$(2.17) \quad A\tilde{\rho} + B\tilde{w} = 0, \quad \tilde{\rho}_{ij}^0 = (\rho_0)_{ij}, \quad \forall 1 \leq i, j \leq N_h - 1.$$

Hence, $J_{\lambda, y, z}(\tilde{\rho}, \tilde{w}, \eta) < \infty$, and so

$$\inf_{(1-\lambda)(A\rho+Bw)=0} J_{\lambda, y, z}(\rho, w, \eta) \leq J_{\lambda, y, z}(\tilde{\rho}, \tilde{w}, \eta) < \infty.$$

- (a) When $\lambda = 0$, for every (ρ, w) such that $\rho \geq 0$, $\rho^0 = \rho_0$, and $A\rho + Bw = 0$ one has that [8, Section 3]

$$\sum_{0 \leq i, j \leq N_h - 1} \rho_{ij}^k = \frac{1}{h^2}, \quad \forall 0 \leq k \leq N_T,$$

and so

$$(2.18) \quad 0 \leq \rho_{ij}^k \leq \frac{1}{h^2}, \quad \forall 0 \leq i, j \leq N_h - 1, \quad 0 \leq k \leq N_T.$$

Thus, taking into account the continuity of F, G and the fact that $E_h \geq 0$, we find that

$$\inf_{(1-\lambda)(A\rho+Bw)=0} J_{\lambda,y,z}(\rho, w, \eta) = \inf_{A\rho+Bw=0} J(\rho, w, \eta) > -\infty.$$

(b) When $\lambda = 1$, the convexity of $F(t, x, \cdot)$ and $G(t, x, \cdot)$ yields

$$(2.19) \quad \begin{aligned} F(t_k, x_{ij}, \rho_{ij}^{k+1}) &\geq F(t_k, x_{ij}, 1) + (\rho_{ij}^{k+1} - 1)d_{ij}^{k+1} \\ G(x_{ij}, \rho_{ij}^{N_T}) &\geq G(x_{ij}, 1) + (\rho_{ij}^{N_T} - 1)r_{ij} \end{aligned}$$

for some $d_{ij}^{k+1} \in \partial_\rho F(t_k, x_{ij}, 1)$ and $r_{ij} \in \partial_\rho G(x_{ij}, 1)$. Thus, applying $E_h \geq 0$ in combination with the Cauchy-Schwarz inequality, we obtain

$$J_{1,y,z}(\rho, w, \eta) \geq \frac{|\rho|^2}{2} - \rho \cdot \tilde{y} + \frac{|w|^2}{2} - w \cdot z - c \geq -\frac{|\tilde{y}|^2}{2} - \frac{|z|^2}{2} - c,$$

where

$$(2.20) \quad \begin{aligned} \tilde{y}_{ij}^{k+1} &= y_{ij}^{k+1} - d_{ij}^{k+1}, \quad 0 \leq k \leq N_T - 2 \\ \tilde{y}_{ij}^{N_T} &= y_{ij}^{N_T} - d_{ij}^{N_T} - r_{ij}, \\ c &= \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} (d_{ij}^{k+1} - F(t_k, x_{ij}, 1)) \\ &\quad + \frac{1}{\Delta t} \sum_{0 \leq i, j \leq N_h - 1} (r_{ij}^{N_T} - G(x_{ij}, 1)). \end{aligned}$$

Hence

$$\inf_{(1-\lambda)(A\rho+Bw)=0} J_{\lambda,y,z}(\rho, w, \eta) = \inf_{\rho, w} J_1(\rho, w, \eta) > -\infty.$$

From (a), (b) above, we find that the infimum in (2.14) is always finite. Let (ρ_n, w_n) be a minimizing sequence for (2.14). Then we have that

$$J_{\lambda,y,z}(\rho_n, w_n, \eta) \leq C, \quad \forall n \geq 1,$$

for some $C > 0$. Our goal is to show that the sequence $\{(\rho_n, w_n)\}$ is bounded and extract a convergent subsequence. Again, let us discuss two cases.

(a) When $\lambda = 0$, taking into account (2.18) for ρ_n and the continuity of F, G , we have that

$$(2.21) \quad E_h(t_k, x_{ij}, (\rho_n)_{ij}^{k+1}, (w_n)_{ij}^k, \eta_{ij}^{k+1}) \leq \tilde{C},$$

for all $n \geq 1$, $0 \leq i, j \leq N_h - 1$, $0 \leq k \leq N_T - 1$, and some $\tilde{C} > 0$. Hence if $(\rho_n)_{ij}^{k+1} > 0$ then

$$(2.22) \quad |(w_n)_{ij}^k| \leq \left(\frac{\beta'}{\epsilon^{\alpha(\beta'-1)}} \left(\frac{1}{h^2} \right)^{\beta'-1} \tilde{C} \right)^{\frac{1}{\beta}}.$$

If $(\rho_n)_{ij}^{k+1} = 0$ then we necessarily have that $(w_n)_{ij}^k = 0$. Thus, in any case (2.22) holds for all i, j, k .

(b) Let $\lambda = 1$, and \tilde{y}, c, d, r be as in (2.19) and (2.20). Then we have that

$$\begin{aligned} C &\geq J_{1,y,z}(\rho_n, w_n, \eta) \geq \frac{|\rho_n|^2}{2} - \rho_n \cdot \tilde{y} + \frac{|w_n|^2}{2} - w_n \cdot z - c \\ &\geq \frac{|\rho_n|^2}{4} + \frac{|w_n|^2}{4} - |\tilde{y}|^2 - |z|^2 - c, \end{aligned}$$

and so

$$\frac{|\rho_n|^2}{4} + \frac{|w_n|^2}{4} \leq |\tilde{y}|^2 + |z|^2 + c + C, \quad \forall n \geq 1.$$

Thus, for both cases $\lambda = 0$ and $\lambda = 1$ the minimizing sequence $\{(\rho^n, w^n)\}$ is precompact, and there exists a (subsequential) limit $(\rho^*, w^*) \in \mathcal{M} \times \mathcal{W}$. Since $\rho_n \geq 0$, $(\rho_n)^0 = \rho_0$ and $(1 - \lambda)(A\rho_n + Bw_n) = 0$, we have that

$$\rho^* \geq 0, \quad (\rho^*)^0 = \rho_0, \quad (1 - \lambda)(A\rho^* + Bw^*) = 0.$$

Furthermore, taking into account the lower semicontinuity of $E_h(t, x, \cdot, \cdot, \eta)$ (see Remark 2.1) and the continuity of F, G we obtain that

$$J_{\lambda,y,z}(\rho^*, w^*, \eta) \leq \liminf_{n \rightarrow \infty} J_{\lambda,y,z}(\rho_n, w_n, \eta),$$

and so (ρ^*, w^*) is a minimizer for (2.14).

2. See the proof of the analogous statement in [8, Lemma 3.2]. \square

Denote by

$$(2.23) \quad \mathcal{P}_0 = \left\{ \eta \in \mathcal{M} : \eta \geq 0, \eta^0 = \rho_0, \sum_{0 \leq i, j \leq N_h - 1} \eta_{ij}^k = \frac{1}{h^2}, \forall 1 \leq k \leq N_T \right\},$$

and

$$(2.24) \quad \mathcal{P}_1 = \{ \eta \in \mathcal{M} : \eta \geq 0, \eta^0 = \rho_0, |\eta|^2 \leq c_1 \},$$

where

$$\begin{aligned} c_1 &= 4 \left(|\tilde{y}|^2 + |z|^2 + c + \frac{|\bar{\rho}|^2}{2} - \bar{\rho} \cdot y + \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} F(t_k, x_{ij}, 1) \right. \\ &\quad \left. + \frac{1}{\Delta t} \sum_{0 \leq i, j \leq N_h - 1} G(x_{ij}, 1) \right), \\ \bar{\rho}^0 &= \rho_0, \quad \bar{\rho}_{ij}^{k+1} = 1, \quad \forall 0 \leq k \leq N_T - 1, \end{aligned}$$

and \tilde{y}, c, d, r are as in (2.19) and (2.20).

Furthermore, for $\eta \geq 0$ and $(\lambda, y, z) \in \{0, 1\} \times \mathcal{M} \times \mathcal{W}$ denote by

$$(2.25) \quad \mathcal{S}_{\lambda,y,z}(\eta) = \left\{ \rho^* : \exists w^* \in \mathcal{W} \text{ s.t. } (\rho^*, w^*) \in \underset{(1-\lambda)(A\rho+Bw)=0}{\operatorname{argmin}} J_{\lambda,y,z}(\rho, w, \eta) \right\}.$$

LEMMA 2.7. *Assume that $F(t, x, \cdot)$ and $G(x, \cdot)$ are convex and continuous for $\rho \geq 0$, and $(\lambda, y, z) \in \{0, 1\} \times \mathcal{M} \times \mathcal{W}$. Then the set valued map $\mathcal{S}_{\lambda, y, z} : \mathcal{P}_\lambda \rightrightarrows \mathcal{P}_\lambda$ admits a fixed point; that is, there exists $\rho^* \in \mathcal{P}_\lambda$ such that $\rho^* \in \mathcal{S}_{\lambda, y, z}(\rho^*)$.*

Proof. Our strategy is to apply Kakutani's fixed point theorem [16]. Hence, fix a triple $(\lambda, y, z) \in \{0, 1\} \times \mathcal{M} \times \mathcal{W}$.

1. Lemma 2.4 yields that $\mathcal{S}_{\lambda, y, z}(\eta) \neq \emptyset$ for all $\eta \geq 0$.

(a) When $\lambda = 0$, we have that for all $\rho^* \in \mathcal{S}_{\lambda, y, z}(\eta)$ there exists $w^* \in \mathcal{W}$ such that

$$(\rho^*, w^*) \in \underset{A\rho+Bw=0}{\operatorname{argmin}} J(\rho, w, \eta).$$

Hence, $\rho^* \in \mathcal{P}_0$, and so $\mathcal{S}_{0, y, z}(\eta) \subset \mathcal{P}_0$ for all $\eta \in \mathcal{P}_0$.

(b) When $\lambda = 1$, we have that for all $\rho^* \in \mathcal{S}_{\lambda, y, z}(\eta)$ there exists $w^* \in \mathcal{W}$ such that

$$(\rho^*, w^*) \in \underset{\rho, w}{\operatorname{argmin}} J_{1, y, z}(\rho, w, \eta),$$

and so

$$J_{1, y, z}(\rho^*, w^*, \eta) \leq J_{1, y, z}(\bar{\rho}, 0, \eta).$$

Hence, applying lower bounds on $J_{1, y, z}$ as in Lemma 2.4 we obtain

$$\frac{|\rho^*|^2}{4} + \frac{|w^*|^2}{4} \leq |\tilde{y}|^2 + |z|^2 + c + J_{1, y, z}(\bar{\rho}, 0, \eta).$$

Noting that

$$\begin{aligned} & J_1(\bar{\rho}, 0, \eta) \\ &= \frac{|\bar{\rho}|^2}{2} - \bar{\rho} \cdot y + \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} F(t_k, x_{ij}, 1) + \frac{1}{\Delta t} \sum_{0 \leq i, j \leq N_h - 1} G(x_{ij}, 1), \end{aligned}$$

we conclude that $\rho^* \in \mathcal{P}_1$, and so $\mathcal{S}_{1, y, z}(\eta) \subset \mathcal{P}_1$ for all $\eta \in \mathcal{P}_1$.

Summarizing (a), (b) above, we conclude that $\mathcal{S}_{\lambda, y, z}(\eta) \subset \mathcal{P}_\lambda$ for all $\eta \in \mathcal{P}_\lambda$. Additionally, the convexity and continuity of $F(t, x, \cdot)$ and $G(x, \cdot)$ yield that $J_{\lambda, y, z}(\cdot, \cdot, \eta)$ is convex and lower semicontinuous, and so

$$\underset{(1-\lambda)(A\rho+Bw)=0}{\operatorname{argmin}} J_{\lambda, y, z}(\rho, w, \eta)$$

is a closed convex set. Thus, $\mathcal{S}_{\lambda, y, z}(\eta)$ is also closed and convex.

2. Assume that $\{(\eta_n, \rho_n)\} \subset \mathcal{P}_\lambda \times \mathcal{P}_\lambda$ is such that $\rho_n \in \mathcal{S}_{\lambda, y, z}(\eta_n)$ for all $n \geq 1$, and

$$\lim_{n \rightarrow \infty} (\eta_n, \rho_n) = (\hat{\eta}, \hat{\rho}).$$

Since \mathcal{P}_λ is compact, we have that $\hat{\eta} \in \mathcal{P}_\lambda$. Furthermore, let $w_n \in \mathcal{W}$ be such that

$$(\rho_n, w_n) \in \underset{(1-\lambda)(A\rho+Bw)=0}{\operatorname{argmin}} J_{\lambda, y, z}(\rho, w, \eta_n).$$

Fix a $(\tilde{\rho}, \tilde{w}) \in \mathcal{M} \times \mathcal{W}$ such that (2.16) and (2.17) hold. Then we have that

$$J_{\lambda,y,z}(\rho_n, w_n, \eta_n) \leq J_{\lambda,y,z}(\tilde{\rho}, \tilde{w}, \eta_n), \quad \forall n \geq 1.$$

Furthermore, we have that

$$\lim_{n \rightarrow \infty} J_{\lambda,y,z}(\tilde{\rho}, \tilde{w}, \eta_n) = J_{\lambda,y,z}(\tilde{\rho}, \tilde{w}, \hat{\eta}) < \infty.$$

Hence, there exists $C > 0$ such that

$$-\lambda \left(\frac{|y|^2}{2} + \frac{|z|^2}{2} \right) + J(\rho_n, w_n, \eta_n) \leq J_{\lambda,y,z}(\rho_n, w_n, \eta_n) \leq C, \quad \forall n \geq 1,$$

and so

$$J(\rho_n, w_n, \eta_n) \leq C + \lambda \left(\frac{|y|^2}{2} + \frac{|z|^2}{2} \right), \quad \forall n \geq 1.$$

Since $\{\rho_n\}$ is a bounded sequence and $F(t, x, \cdot)$, $G(x, \cdot)$ are continuous, we obtain that

$$(2.26) \quad 0 \leq E_h(t_k, x_{ij}, (\rho_n)_{ij}^{k+1}, (w_n)_{ij}^k, (\eta_n)_{ij}^{k+1}) \leq \tilde{C},$$

for $0 \leq i, j \leq N_h - 1$, $0 \leq k \leq N_T - 1$, and some $\tilde{C} > 0$. Next, denote by

$$\ell(v_1, v_2, v_3, v_4) = \begin{cases} \frac{(v_1^2 + v_2^2 + v_3^2 + v_4^2)^{\beta'/2}}{\beta'}, & (v_1, v_2, v_3, v_4) \in K, \\ \infty, & \text{otherwise.} \end{cases}$$

The corresponding recession function is then

$$\tilde{\ell}(\rho, w_1, w_2, w_3, w_4) = \begin{cases} \frac{(w_1^2 + w_2^2 + w_3^2 + w_4^2)^{\beta'/2}}{\beta' \rho^{\beta'-1}}, & (w_1, w_2, w_3, w_4) \in K, \\ 0, & (\rho, w_1, w_2, w_3, w_4) = (0, 0, 0, 0, 0), \\ \infty, & \text{otherwise.} \end{cases}$$

Note that for $\eta \geq 0$ we have that

$$(2.27) \quad \begin{aligned} L_h(t, x, v_1, v_2, v_3, v_4, \eta) &= (\eta + \epsilon)^{\alpha(\beta'-1)} \ell(v_1, v_2, v_3, v_4), \\ E_h(t, x, w_1, w_2, w_3, w_4, \eta) &= (\eta + \epsilon)^{\alpha(\beta'-1)} \tilde{\ell}(\rho, w_1, w_2, w_3, w_4). \end{aligned}$$

Taking into account $\eta_n \geq 0$ and (2.26) we obtain

$$(2.28) \quad 0 \leq \tilde{\ell}((\rho_n)_{ij}^{k+1}, (w_n)_{ij}^k) \leq \frac{\tilde{C}}{\epsilon^{\alpha(\beta'-1)}}.$$

In particular, we find that (2.22) holds. Thus $\{(\rho_n, w_n)\}$ is precompact, and, possibly through a subsequence, we have that

$$\lim_{n \rightarrow \infty} (\rho_n, w_n) = (\hat{\rho}, \hat{w}).$$

Hence, we have that

$$\begin{aligned} \hat{\rho}^0 &= \lim_{n \rightarrow \infty} (\rho_n)^0 = \rho_0, \quad \hat{\rho} = \lim_{n \rightarrow \infty} \rho_n \geq 0, \\ (1 - \lambda)(A\hat{\rho} + B\hat{w}) &= \lim_{n \rightarrow \infty} (1 - \lambda)(A\rho_n + Bw_n) = 0. \end{aligned}$$

Additionally, using the lower semicontinuity of $\tilde{\ell}$, we find that

$$\begin{aligned} & \liminf_{n \rightarrow \infty} E_h(t_k, x_{ij}, (\rho_n)_{ij}^{k+1}, (w_n)_{ij}^k, (\eta_n)_{ij}^{k+1}) \\ & \geq \liminf_{n \rightarrow \infty} ((\eta_n)_{ij}^{k+1} + \epsilon)^{\alpha(\beta' - 1)} \cdot \liminf_{n \rightarrow \infty} \tilde{\ell}((\rho_n)_{ij}^{k+1}, (w_n)_{ij}^k) \\ & \geq (\hat{\eta}_{ij}^{k+1} + \epsilon)^{\alpha(\beta' - 1)} \tilde{\ell}(\hat{\rho}_{ij}^{k+1}, \hat{w}_{ij}^k) = E_h(t_k, x_{ij}, \hat{\rho}_{ij}^{k+1}, \hat{w}_{ij}^k, \hat{\eta}_{ij}^{k+1}), \end{aligned}$$

for all $0 \leq i, j \leq N_h - 1$, and $0 \leq l \leq N_T - 1$. But then we have that

$$\begin{aligned} & \liminf_{n \rightarrow \infty} J_{\lambda, y, z}(\rho_n, w_n, \eta_n) \\ & \geq \liminf_{n \rightarrow \infty} \lambda \left(\frac{|\rho_n|^2}{2} - \rho_n \cdot y + \frac{|w_n|^2}{2} - w_n \cdot z \right) \\ & \quad + \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} \liminf_{n \rightarrow \infty} E_h(t_k, x_{ij}, (\rho_n)_{ij}^{k+1}, (w_n)_{ij}^k, (\eta_n)_{ij}^{k+1}) \\ & \quad + \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} \liminf_{n \rightarrow \infty} F(t_k, x_{ij}, (\rho_n)_{ij}^{k+1}) \\ & \quad + \frac{1}{\Delta t} \sum_{0 \leq i, j \leq N_h - 1} \liminf_{n \rightarrow \infty} G(x_{ij}, (\rho_n)_{ij}^{N_T}) \\ & \geq \lambda \left(\frac{|\hat{\rho}|^2}{2} - \hat{\rho} \cdot y + \frac{|\hat{w}|^2}{2} - \hat{w} \cdot z \right) \\ & \quad + \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} E_h(t_k, x_{ij}, \hat{\rho}_{ij}^{k+1}, \hat{w}_{ij}^k, \hat{\eta}_{ij}^{k+1}) \\ & \quad + \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} F(t_k, x_{ij}, \hat{\rho}_{ij}^{k+1}) + \frac{1}{\Delta t} \sum_{0 \leq i, j \leq N_h - 1} G(x_{ij}, \hat{\rho}_{ij}^{N_T}) \\ & = J_{\lambda, y, z}(\hat{\rho}, \hat{w}, \hat{\eta}). \end{aligned}$$

Now fix $(\rho, w) \in \mathcal{M} \times \mathcal{W}$ be such that $\rho^0 = \rho_0$, and $(1 - \lambda)(A\rho + Bw) = 0$. Our goal is to prove that

$$J_{\lambda, y, z}(\hat{\rho}, \hat{w}, \hat{\eta}) \leq J_{\lambda, y, z}(\rho, w, \hat{\eta}),$$

so that $\hat{\rho} \in \mathcal{S}_{\lambda, y, z}(\hat{\eta})$.

If $J_{\lambda, y, z}(\rho, w, \hat{\eta}) = \infty$, the inequality is trivially true. Hence, assume that $J_{\lambda, y, z}(\rho, w, \hat{\eta}) < \infty$. Note that this implies that $J_{\lambda, y, z}(\rho, w, \eta) < \infty$ for all $\eta \geq 0$. We have that

$$J_{\lambda, y, z}(\rho_n, w_n, \eta_n) \leq J_{\lambda, y, z}(\rho, w, \eta_n), \quad \forall n \geq 1,$$

but then

$$\begin{aligned} J_{\lambda, y, z}(\hat{\rho}, \hat{w}, \hat{\eta}) & \leq \liminf_{n \rightarrow \infty} J_{\lambda, y, z}(\rho_n, w_n, \eta_n) \leq \liminf_{n \rightarrow \infty} J_{\lambda, y, z}(\rho, w, \eta_n) \\ & = \lim_{n \rightarrow \infty} J_{\lambda, y, z}(\rho, w, \eta_n) = J_{\lambda, y, z}(\rho, w, \hat{\eta}). \end{aligned}$$

Hence, $\mathcal{S}_{\lambda, y, z}$ maps all $\eta \in \mathcal{P}_\lambda$ to nonempty closed convex subsets of \mathcal{P}_λ and has a closed graph. Therefore, Kakutani's theorem [16] implies that $\mathcal{S}_{\lambda, y, z}$ has a fixed point. \square

COROLLARY 2.8. Assume that $F(t, x, \cdot)$ and $G(x, \cdot)$ are convex and continuous for $\rho \geq 0$, and for $(\rho, w) \in \mathcal{M} \times \mathcal{W}$ denote by

$$(2.29) \quad \mathcal{K}(\rho, w) = \begin{cases} \partial_{(\rho, w)} J(\rho, w, \eta) \Big|_{\eta=\rho}, & \rho \geq 0, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Then \mathcal{K} is maximally monotone.

Remark 2.9. Note that in (2.29) we first differentiate J with respect to (ρ, w) and then plug in $\eta = \rho$. This should not be confused with $\partial_{(\rho, w)} J(\rho, w, \rho)$.

Proof of Corollary 2.8. By Minty's theorem [4, Theorem 3.5.8] we have to prove that \mathcal{K} is monotone, and $I + \mathcal{K}$ is surjective. We start with the latter.

1. *Surjectivity.* Assume that $(y, z) \in \mathcal{M} \times \mathcal{W}$ are arbitrary. By Lemma 2.7 we have that there exists $\rho^* \in \mathcal{P}_1$ such that $\rho^* \in \mathcal{S}_{1, y, z}(\rho^*)$. Hence, there exists $w^* \in \mathcal{W}$ such that

$$(\rho^*, w^*) \in \underset{\rho, w}{\operatorname{argmin}} J_{1, y, z}(\rho, w, \rho^*).$$

Consequently, we have that

$$(0, 0) \in \partial_{(\rho, w)} J_{1, y, z}(\rho^*, w^*, \eta) \Big|_{\eta=\rho^*}.$$

Furthermore, we have that

$$J_{1, y, z}(\rho, w, \eta) = \frac{|\rho|^2}{2} - \rho \cdot y + \frac{|w|^2}{2} - w \cdot z + J(\rho, w, \eta),$$

and

$$(\rho, w) \mapsto \frac{|\rho|^2}{2} - \rho \cdot y + \frac{|w|^2}{2} - w \cdot z$$

is continuously differentiable. Hence, we obtain

$$\partial_{\rho, w} J_{1, y, z}(\rho, w, \eta) = (\rho, w) - (y, z) + \partial_{(\rho, w)} J(\rho, w, \eta),$$

and so

$$(0, 0) \in (\rho^*, w^*) + \mathcal{K}(\rho^*, w^*) - (y, z) \iff (y, z) \in (\rho^*, w^*) + \mathcal{K}(\rho^*, w^*).$$

2. *Monotonicity.* The convexity of $F(t, x, \cdot)$ and $G(x, \cdot)$ and the separable structure of J at the grid points yield that it is sufficient to prove that

$$e_h(t, x, \rho, w) = \begin{cases} \partial_{\rho, w} E_h(t, x, \rho, w, \eta) \Big|_{\eta=\rho}, & \rho \geq 0 \\ \emptyset, & \text{otherwise} \end{cases}$$

is monotone. Applying [11, Proposition 2.3], we obtain

$$(2.30) \quad e_h(t, x, \rho, w) = \begin{cases} \left\{ (-H_h(t, x, q, \rho), -q) : q \in -\partial_v L_h(t, x, \frac{w}{\rho}, \rho) \right\}, & \rho > 0, \\ \left\{ (\mu, -q) : \mu + H_h(t, x, q, \rho) \leq 0 \right\}, & (\rho, w) = (0, 0), \\ \emptyset, & \text{otherwise.} \end{cases}$$

Now assume that

$$(y_i, z_i) \in e_h(t, x, \rho_i, w_i), \quad i = 1, 2.$$

We have to show that

$$(y_2 - y_1)(\rho_2 - \rho_1) + (z_2 - z_1) \cdot (w_2 - w_1) \geq 0.$$

Since the case $(\rho_1, w_1) = (\rho_2, w_2)$ is trivial, we assume that $(\rho_1, w_1) \neq (\rho_2, w_2)$. Hence, up to swapping (ρ_1, w_1) and (ρ_2, w_2) , there are two possibilities:

(a) $\rho_1, \rho_2 > 0$. In this case, (2.30) yields that there exist q_1, q_2 such that

$$y_i = -H_h(t, x, q_i, \rho_i), \quad z_i = -q_i, \quad q_i \in -\partial_v L_h(t, x, \frac{w_i}{\rho_i}, \rho_i), \quad i = 1, 2.$$

But then the Legendre duality yields that

$$w_i = -\rho_i \nabla_q H_h(t, x, q_i, \rho_i), \quad i = 1, 2,$$

and so we have to prove that

$$\begin{aligned} & (-H_h(t, x, q_2, \rho_2) + H_h(t, x, q_1, \rho_1))(\rho_2 - \rho_1) \\ & + (\rho_2 \nabla_q H_h(t, x, q_2, \rho_2) - \rho_1 \nabla_q H_h(t, x, q_1, \rho_1)) \cdot (q_2 - q_1) \geq 0. \end{aligned}$$

This inequality is equivalent to the monotonicity of the map

$$(2.31) \quad (\rho, q) \mapsto (-H_h(t, x, q, \rho), \rho \nabla_q H_h(t, x, \rho, q)), \quad \rho \geq 0, \quad q \in \mathbb{R}^4.$$

Since the map is continuously differentiable, its monotonicity is equivalent to the positive definiteness of the symmetric part of its Jacobian. Note that the Jacobian is given by

$$\begin{pmatrix} -\partial_\rho H_h & -\nabla_q^\top H_h \\ \nabla_q H_h + \rho \nabla_q \partial_\rho H_h & \rho \nabla_q^2 H_h \end{pmatrix},$$

and so its symmetric part is

$$\begin{pmatrix} -\partial_\rho H_h & \frac{1}{2} \rho \nabla_q^\top \partial_\rho H_h \\ \frac{1}{2} \rho \nabla_q \partial_\rho H_h & \rho \nabla_q^2 H_h \end{pmatrix}.$$

Thus, the monotonicity of (2.31) is equivalent to the Lasry-Lions monotonicity condition of H_h , which is proven in Lemma (2.3).

(b) $\rho_1 = 0$, $\rho_2 > 0$. In this case, (2.30) yields

$$w_1 = 0, \quad y_1 \leq -H_h(t, x, q_1, 0), \quad z_1 = -q_1,$$

and

$$y_2 = -H_h(t, x, q_2, \rho_2), \quad z_2 = -q_2, \quad q_2 \in -\partial_v L_h(t, x, \frac{w_2}{\rho_2}, \rho_2).$$

Hence, we have that

$$w_i = -\rho_i \nabla_q H_h(t, x, q_i, \rho_i), \quad i = 1, 2,$$

and we need to show that

$$\begin{aligned} & (-H_h(t, x, q_2, \rho_2) - y_1)(\rho_2 - \rho_1) \\ & + (\rho_2 \nabla_q H_h(t, x, q_2, \rho_2) - \rho_1 \nabla_q H_h(t, x, q_1, \rho_1)) \cdot (q_2 - q_1) \geq 0. \end{aligned}$$

But inequalities $y_1 \leq -H_h(t, x, q_1, 0)$ and $\rho_2 - \rho_1 = \rho_2 > 0$ yield that

$$\begin{aligned} & (-H_h(t, x, q_2, \rho_2) - y_1)(\rho_2 - \rho_1) \\ & + (\rho_2 \nabla_q H_h(t, x, q_2, \rho_2) - \rho_1 \nabla_q H_h(t, x, q_1, \rho_1)) \cdot (q_2 - q_1) \\ \geq & (-H_h(t, x, q_2, \rho_2) + H_h(t, x, q_1, \rho_1))(\rho_2 - \rho_1) \\ & + (\rho_2 \nabla_q H_h(t, x, q_2, \rho_2) - \rho_1 \nabla_q H_h(t, x, q_1, \rho_1)) \cdot (q_2 - q_1) \geq 0, \end{aligned}$$

where the last inequality follows from the monotonicity of (2.31) proven in (a). \square

2.5. First-order optimality conditions.

LEMMA 2.10. *Let $\eta \geq 0$, and assume that $F(t, x, \cdot)$ and $G(x, \cdot)$ are convex and continuous for $\rho \geq 0$. Then the following statements are true.*

1. $(\rho, w) \in \mathcal{M} \times \mathcal{W}$ is a minimizer for (2.14) with $\lambda = 0$ if and only if

$$(2.32) \quad (A^* \phi, B^* \phi) \in \partial_{(\rho, w)} J(\rho, w, \eta), \quad A\rho + Bw = 0,$$

for some $\phi \in \mathcal{U}$.

2. If (2.15) holds and $F(t, x, \cdot)$ and $G(x, \cdot)$ are differentiable for $\rho > 0$, (2.32) is equivalent to

$$(2.33) \quad \begin{cases} -(D_t \phi_{ij})^k - \nu(\Delta_h \phi^k)_{ij} + H_h(x_{ij}, \eta_{ij}^{k+1}, [D_h \phi^k]_{ij}) = f(t_k, x_{ij}, \rho_{ij}^{k+1}), \\ w_{ij}^k = -\rho_{ij}^{k+1} \nabla_q H_h(x_{ij}, \eta_{ij}^{k+1}, [D_h \phi^k]), \\ (D_t \rho_{ij})^k - \nu(\Delta_h \rho^{k+1})_{ij} + (Bw)_{ij}^k = 0, \\ \rho_{ij}^0 = (\rho_0)_{ij}, \quad \phi_{ij}^{N_T} = g(x_{ij}, \rho_{ij}^{N_T}), \quad 0 \leq i, j \leq N_h - 1, \quad 0 \leq k \leq N_T - 1. \end{cases}$$

3. If (2.15) holds and $F(t, x, \cdot)$ and $G(x, \cdot)$ are differentiable for $\rho > 0$, (2.3) is equivalent to

$$(2.34) \quad (A^* \phi, B^* \phi) \in \mathcal{K}(\rho, w), \quad A\rho + Bw = 0,$$

where \mathcal{K} is defined in (2.29).

Proof. 1. Since (2.14) is a convex program with affine constraints, KKT conditions are necessary and sufficient for optimality. To obtain the KKT conditions, we introduce a Lagrange multiplier, ϕ , for the constraint $A\rho + Bw = 0$ and find

$$\begin{aligned} \inf_{A\rho+Bw=0} J(\rho, w, \eta) &= \inf_{(\rho, w) \in \mathcal{M} \times \mathcal{W}} \sup_{\phi \in \mathcal{U}} \{J(\rho, w, \eta) - \langle A\rho + Bw, \phi \rangle\} \\ &= \inf_{(\rho, w) \in \mathcal{M} \times \mathcal{W}} \sup_{\phi \in \mathcal{U}} \{J(\rho, w, \eta) - \langle \rho, A^* \phi \rangle - \langle w, B^* \phi \rangle\}, \end{aligned}$$

which yields (2.32).

2. Next, assume that (2.15) holds. In what follows, we use that for every $\rho > 0$, $\eta \geq 0$, and $w \in K$ one has that [11, Proposition 2.3]

$$\partial_{(\rho, w)} E_h(t, x, \rho, w, \eta) = \left\{ (-H_h(t, x, q, \eta), -q) : q \in -\partial_v L_h(t, x, \frac{w}{\rho}, \eta) \right\}.$$

Hence (2.32) is equivalent to

(2.35)

$$\begin{aligned} -[D_h \phi^k]_{ij} &\in \partial_v L_h(t_k, x_{ij}, \frac{w_{ij}^k}{\rho_{ij}^{k+1}}, \eta_{ij}^{k+1}), \\ -(D_t \phi_{ij})^k - \nu(\Delta_h \phi^k)_{ij} &= -H_h(t_k, x_{ij}, [D_h \phi^k]_{ij}, \eta_{ij}^{k+1}) + f(t_k, x_{ij}, \rho_{ij}^{k+1}), \\ \frac{\phi_{ij}^{N_T-1}}{\Delta t} - \nu(\Delta_h \phi^{N_T-1})_{ij} &= -H_h(t_{N_T-1}, x_{ij}, [D_h \phi^{N_T-1}]_{ij}, \eta_{ij}^{N_T}) \\ &\quad + f(t_{N_T-1}, x_{ij}, \rho_{ij}^{N_T}) + \frac{g(x_{ij}, \rho_{ij}^{N_T})}{\Delta t}. \end{aligned}$$

for all $0 \leq i, j \leq N_h - 1$ and $0 \leq k \leq N_T - 1$. The second and third equalities can be combined in a system

(2.36)

$$\begin{cases} -(D_t \phi_{ij})^k - \nu(\Delta_h \phi^k)_{ij} + H_h(t_k, x_{ij}, [D_h \phi^k]_{ij}, \eta_{ij}^{k+1}) = f(t_k, x_{ij}, \rho_{ij}^{k+1}), \\ \phi_{ij}^{N_T} = g(x_{ij}, \rho_{ij}^{N_T}), \quad \forall 0 \leq i, j \leq N_h - 1, 0 \leq k \leq N_T - 1. \end{cases}$$

Next, using the convex duality again, we find that the first inclusion in (2.35) is equivalent to the equality

$$\frac{w_{ij}^k}{\rho_{ij}^{k+1}} = -\nabla_q H_h(t_k, x_{ij}, [D_h \phi^k]_{ij}, \eta_{ij}^{k+1}),$$

for all $0 \leq i, j \leq N_h - 1$ and $0 \leq k \leq N_T - 1$, since $H_h(t, x, q, \eta)$ is continuously differentiable in q for $\eta \geq 0$. Combining this previous identity with $A\rho + Bw = 0$ we obtain

$$(2.37) \quad \begin{cases} (D_t \rho_{ij})^k - \nu(\Delta_h \rho^{k+1})_{ij} + (Bw)_{ij}^k = 0, \\ w_{ij}^k = -\rho_{ij}^{k+1} \nabla_q H_h(x_{ij}, \eta_{ij}^{k+1}, [D_h \phi^k]), \\ \rho_{ij}^0 = \rho_{0,ij}, \quad 0 \leq i, j \leq N_h - 1, 0 \leq k \leq N_T - 1. \end{cases}$$

Finally, combining (2.36) and (2.37) one obtains (2.33).

3. This follows from part 2 and the definition of \mathcal{K} . \square

COROLLARY 2.11. *Assume that $\nu > 0$, and $F(t, x, \cdot)$ and $G(x, \cdot)$ are convex and continuous for $\rho \geq 0$ and differentiable for $\rho > 0$. Then (2.3) admits a solution.*

Proof. By Lemma 2.7 we have that there exist (ρ^*, w^*) such that

$$(\rho^*, w^*) \in \underset{A\rho + Bw = 0}{\operatorname{argmin}} J(\rho, w, \rho^*).$$

Furthermore, by Lemma 2.4 we have that

$$(\rho^*)_{ij}^{k+1} > 0, \quad 0 \leq k \leq N_T - 1, \quad 0 \leq i, j \leq N_h - 1.$$

Hence, Lemma 2.10 yields ϕ^* such that (ρ^*, w^*, ϕ^*) solves (2.3). \square

3. A pair of primal-dual monotone inclusions. The formulation (2.34) is the basis of our computational method. We first recall monotone inclusion version of PDHG. Following [22], assume that $\mathcal{H}_1, \mathcal{H}_2$ are Hilbert spaces, and $M : \mathcal{H}_1 \rightarrow 2^{\mathcal{H}_1}$, $N : \mathcal{H}_2 \rightarrow 2^{\mathcal{H}_2}$ are maximally monotone operators, and $C : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ is a nonzero bounded linear operator. Now consider the following pair of primal-dual monotone inclusions

$$(3.1) \quad \begin{aligned} & \text{find } x \in \mathcal{H}_1 \text{ s.t. } 0 \in Mx + C^*(N(Cx)) \quad (\text{P}) \\ & \text{find } y \in \mathcal{H}_2 \text{ s.t. } y \in N(Cx), \quad -C^*y \in Mx, \text{ for some } x \in \mathcal{H}_1 \quad (\text{D}) \end{aligned}$$

When M, N are subdifferentials of proper convex lower semicontinuous functions; that is, $M = \partial f_1$, $N = \partial f_2$, (3.1) reduces to a convex-concave saddle point problem

$$(3.2) \quad \begin{aligned} \inf_x \{f_1(x) + f_2(Cx)\} &= \inf_x \sup_y \{f_1(x) + \langle Cx, y \rangle_{\mathcal{H}_2} - f_2(y)\} \\ &= \sup_y \{-f_1^*(-C^*y) - f_2(y)\}. \end{aligned}$$

System (3.1) can be solved by an extension of the celebrated PDHG algorithm:

$$(3.3) \quad \begin{cases} x^{n+1} = (I + \tau M)^{-1} (x^n - \tau C^* y^n) \\ \tilde{x}^{n+1} = 2x^{n+1} - x^n \\ y^{n+1} = (I + \sigma N^{-1})^{-1} (y^n + \sigma C \tilde{x}^{n+1}) \end{cases}$$

where $\tau, \sigma > 0$ are such that $\tau\sigma < \frac{1}{\|C\|^2}$.

Our goal is to formulate (2.34) in the form of (3.1) and solve it via (3.3).

THEOREM 3.1. *Assume that $F(t, x, \cdot)$, $G(x, \cdot)$ are convex and continuous for $\rho \geq 0$, and denote by*

$$(3.4) \quad \begin{aligned} M(\phi) &= 0, \quad C(\phi) = (A^*\phi, B^*\phi), \quad \phi \in \mathcal{U}, \\ N(\rho, w) &= \mathcal{K}^{-1}(\rho, w), \quad (\rho, w) \in \mathcal{M} \times \mathcal{W}. \end{aligned}$$

Then the following statements are true.

1. M, N are maximally monotone, and (2.34) is equivalent to the pair of primal-dual inclusions

$$(3.5) \quad \begin{aligned} & \text{find } \phi \text{ s.t. } 0 \in M(\phi) + C^*(N(C\phi)) \quad (\text{P}) \\ & \text{find } (\rho, w) \text{ s.t. } (\rho, w) \in N(C\phi), \quad -C^*(\rho, w) \in M(\phi), \text{ for some } \phi \quad (\text{D}) \end{aligned}$$

2. The problem (3.5) admits at least one solution.
3. When $\nu > 0$ and $F(t, x, \cdot)$ and $G(x, \cdot)$ are differentiable for $\rho > 0$, (3.5) is equivalent to (2.3).
4. When $\nu > 0$ and $F(t, x, \cdot)$ and $G(x, \cdot)$ are differentiable for $\rho > 0$, the algorithm

$$(3.6) \quad \begin{cases} \phi^{n+1} = \phi^n - \tau(A\rho^n + Bw^n) \\ \tilde{\phi}^{n+1} = 2\phi^{n+1} - \phi^n \\ (\rho^{n+1}, w^{n+1}) = (I + \sigma\mathcal{K})^{-1} \left(\rho^n + \sigma A^* \tilde{\phi}^{n+1}, w^n + \sigma B^* \tilde{\phi}^{n+1} \right) \end{cases}$$

yields a solution for (2.3).

Proof. 1. Trivially, we have that M is monotone, and $I+M = I$ is surjective. Hence, by Minty's theorem [4, Theorem 3.5.8] M is maximally monotone. Furthermore, by Corollary 2.8 we have that \mathcal{K} is maximally monotone, and so $N = \mathcal{K}^{-1}$ is also maximally monotone [4, Section 3.5.2]. Furthermore, the equivalence of (2.34) and (3.5) follows directly from the definitions of N and C since

$$(\rho, w) \in N(C(\phi)) \iff (A^*\phi, B^*\phi) \in \mathcal{K}(\rho, w).$$

2. By Lemma 2.7 we have that there exists (ρ^*, w^*) such that

$$(\rho^*, w^*) \in \underset{A\rho + Bw = 0}{\operatorname{argmin}} J(\rho, w, \rho^*).$$

Hence, by part 1 in Lemma 2.10 we have that there exists $\phi^* \in \mathcal{U}$ such that

$$(A^*\phi^*, B^*\phi^*) \in \partial_{(\rho, w)} J(\rho, w, \eta) \Big|_{\eta = \rho^*} = \mathcal{K}(\rho^*, w^*), \quad A\rho^* + Bw^* = 0,$$

which is precisely (2.34). Thus, we conclude by part 1 above.

3. This follows from part 3 in Lemma 2.10 and part 1 above.
4. This follows from [22], and part 3 above. \square

3.1. Grid-size independent time-steps. The convergence of (3.6) might be very slow for fine grids. The reason is that $\tau, \sigma > 0$ must satisfy $\tau\sigma < \frac{1}{\|C\|^2}$. But $C(\phi) = (A^*\phi, B^*\phi)$ is a discretization of differential operator $(-\partial_t\phi - \nu\Delta\phi, -\nabla\phi)$ that is unbounded. Thus, $\|C\| \rightarrow \infty$ in the continuous limit; that is, when the grid gets finer. Consequently, $\tau, \sigma > 0$ should be very small slowing down the convergence.

To remedy this previous issue, we will choose the norm in \mathcal{U} in such a way that $\|C\| = 1$, and so $\tau, \sigma > 0$ should just satisfy $\tau\sigma < 1$. This procedure will be equivalent to a suitable preconditioning of the ϕ update in (3.6).

Consider the following inner product in \mathcal{U} :

$$\langle \phi_1, \phi_2 \rangle_\star = \langle C\phi_1, C\phi_2 \rangle = \langle C^*C\phi_1, \phi_2 \rangle = \langle (AA^* + BB^*)\phi_1, \phi_2 \rangle, \quad \forall \phi_1, \phi_2 \in \mathcal{U},$$

where the inner products without \star are the standard ones. Note that $\langle \cdot, \cdot \rangle_\star$ is bilinear, and

$$0 = \|\phi\|_\star^2 = \|C\phi\|^2 = \|A^*\phi\|^2 + \|B^*\phi\|^2$$

yields $\phi = 0$. Hence, $\langle \cdot, \cdot \rangle_\star$ is indeed an inner product on \mathcal{U} .

Now define M, C, N as before in (3.4) but let us calculate the adjoint of $C : \mathcal{U} \rightarrow \mathcal{M} \times \mathcal{W}$ with respect to $\langle \cdot, \cdot \rangle_*$. We have that

$$\langle C^*(\rho, w), \phi \rangle = \langle (\rho, w), C\phi \rangle = \langle C_*^*(\rho, w), \phi \rangle_* = \langle C^* C C_*^*(\rho, w), \phi \rangle$$

for all ρ, w, ϕ , and so

$$C_*^* = (C^* C)^{-1} C^* = C^\dagger$$

where C^\dagger is the Moore-Penrose pseudoinverse. Hence, (2.34) has another equivalent formulation as a primal dual pair of monotone inclusions

$$(3.7) \quad \begin{aligned} & \text{find } \phi \text{ s.t. } M(\phi) \in C^\dagger(N(C\phi)) \quad (\text{P}) \\ & \text{find } (\rho, w) \text{ s.t. } (\rho, w) \in N(C\phi), \quad -C^\dagger(\rho, w) \in M(\phi), \text{ for some } \phi \quad (\text{D}) \end{aligned}$$

Accordingly, we have the following PDHG algorithm

$$(3.8) \quad \begin{cases} \phi^{n+1} = \phi^n - \tau(AA^* + BB^*)^{-1}(A\rho^n + Bw^n) \\ \tilde{\phi}^{n+1} = 2\phi^{n+1} - \phi^n \\ (\rho^{n+1}, w^{n+1}) = (I + \sigma\mathcal{K})^{-1}(\rho^n + \sigma A^* \tilde{\phi}^{n+1}, w^n + \sigma B^* \tilde{\phi}^{n+1}) \end{cases}$$

Note that the only difference between (3.8) and (3.6) is the ϕ update step, which is preconditioned with $C^*C = AA^* + BB^*$. The advantage is that

$$\|C\|_* = \sup_{\|\phi\|_* \leq 1} \|C\phi\| = \sup_{\|C\phi\| \leq 1} \|C\phi\| = 1,$$

and so $\tau, \sigma > 0$ in (3.7) must satisfy $\tau\sigma < 1$. Therefore, the magnitude of τ, σ and the convergence rate of the algorithm are independent of the grid.

3.2. Further remarks on the algorithm. Note that $C^*C = AA^* + BB^*$ is the discretization of the continuous operator

$$(-\partial_t - \nu\Delta)(\partial_t - \nu\Delta) + \nabla \cdot (-\nabla) = -\partial_t^2 - \Delta + \nu^2\Delta^2 \geq 0$$

Hence, the ϕ update in (3.7) involves a solution of fourth-order space-time PDE that can be done via FFT.

Furthermore, the (ρ, w) update in (3.8) decouples to 3-dimensional, or $(d+1)$ -dimensional for $d \neq 2$, monotone inclusion problems at grid points that can be solved efficiently by standard algorithms. Often, w can be eliminated leading to one-dimensional problems with respect to ρ at the grid points. This happens, for example, when solving optimal transport or potential MFG problems [5, 6, 7, 9, 8] or the congestion model [2] considered here.

When the convex duals or $F(t, x, \cdot)$ and $G(x, \cdot)$ are readily available, we can simply the (ρ, w) updates by a further splitting. To this end, assume that $F(t, x, \cdot)$ and $G(x, \cdot)$ are convex and continuous in $\rho \geq 0$, and

$$F(t, x, \rho) = \infty, \quad G(x, \rho) = \infty, \quad \forall \rho < 0,$$

and consider

$$(3.9) \quad F^*(t, x, a) = \sup_{\rho \geq 0} \{a \cdot \rho - F(t, x, \rho)\}, \quad G^*(x, b) = \sup_{\rho \geq 0} \{b \cdot \rho - G(x, \rho)\},$$

which are proper convex lower semicontinuous functions with respect to a, b , respectively. Next, consider

$$(3.10) \quad \mathcal{F}^*(a) = \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} F^*(t_k, x_{ij}, a_{ij}^k), \quad \mathcal{G}^*(b) = \sum_{0 \leq i, j \leq N_h - 1} \frac{G^*(x_{ij}, b_{ij} \Delta t)}{\Delta t}.$$

Note that $\mathcal{F}^*, \mathcal{G}^*$ are also proper convex lower semicontinuous functions. Furthermore, denote by

$$(3.11) \quad \tilde{\mathcal{J}}(\rho, w, \eta) = \sum_{\substack{0 \leq i, j \leq N_h - 1 \\ 0 \leq k \leq N_T - 1}} E_h(t_k, x_{ij}, w_{ij}^k, \rho_{ij}^{k+1}, \eta_{ij}^{k+1}) + \mathbf{1}_{\rho^0 = \rho_0},$$

and consider

$$\begin{aligned} \tilde{\mathcal{K}}(\rho, w) &= \partial_{(\rho, w)} \tilde{\mathcal{J}}(\rho, w, \eta) \Big|_{\eta = \rho}, \quad (\rho, w) \in \mathcal{M} \times \mathcal{W}, \\ \tilde{\mathcal{N}}(\rho, w) &= \tilde{\mathcal{K}}^{-1}(\rho, w), \quad (\rho, w) \in \mathcal{M} \times \mathcal{W}, \\ \tilde{\mathcal{M}}(a, b, \phi) &= (\partial_a \mathcal{F}^*(a), \partial_b \mathcal{G}^*(b), 0), \quad (a, b, \phi) \in \mathcal{U} \times \mathbb{R}^{N_h \times N_h} \times \mathcal{U}. \end{aligned}$$

Finally, we define $\tilde{\mathcal{C}} : \mathcal{U} \times \mathbb{R}^{N_h \times N_h} \times \mathcal{U} \rightarrow \mathcal{M} \times \mathcal{W}$ as

$$\tilde{\mathcal{C}}(a, b, \phi) = (A^* \phi - (0, a^{-(N_T-1)}, a^{N_T-1} + b), B^* \phi),$$

where

$$a^k = (a_{ij}^k)_{i,j=0}^{N_h-1}, \quad a^{-k} = (a_{ij}^l)_{i,j=0, l \neq k}^{N_h-1}.$$

Hence, we have that $\tilde{\mathcal{C}}^* : \mathcal{M} \times \mathcal{W} \rightarrow \mathcal{U} \times \mathbb{R}^{N_h \times N_h} \times \mathcal{U}$ is given by

$$\tilde{\mathcal{C}}^*(\rho, w) = (-\rho^{-0}, -\rho^{N_T}, A\rho + Bw).$$

Then (2.34) can be written as

(3.12)

$$\text{find } (a, b, \phi) \text{ s.t. } 0 \in \tilde{\mathcal{M}}(a, b, \phi) + \tilde{\mathcal{C}}^*(\tilde{\mathcal{N}}(\tilde{\mathcal{C}}(a, b, \phi))) \quad (\text{P})$$

$$\text{find } (\rho, w) \text{ s.t. } (\rho, w) \in \tilde{\mathcal{N}}(\tilde{\mathcal{C}}(a, b, \phi)), -\tilde{\mathcal{C}}^*(\rho, w) \in \tilde{\mathcal{M}}(a, b, \phi), \text{ for some } (a, b, \phi) \quad (\text{D})$$

Consequently, the PDHG algorithm is

$$(3.13) \quad \begin{cases} a^{n+1} = (I + \tau \partial_a \mathcal{F}^*)^{-1} (a^n + \tau(\rho^n)^{-0}) \\ b^{n+1} = (I + \tau \partial_b \mathcal{G}^*)^{-1} (b^n + \tau(\rho^n)^{N_T}) \\ \phi^{n+1} = \phi^n - \tau(A\rho^n + Bw^n) \\ (\tilde{a}^{n+1}, \tilde{b}^{n+1}, \tilde{\phi}^{n+1}) = 2(a^{n+1}, b^{n+1}, \phi^{n+1}) - (a^n, b^n, \phi^n) \\ (\rho^{n+1}, w^{n+1}) = (I + \sigma \tilde{\mathcal{K}})^{-1} \left((\rho^n, w^n) + \sigma \tilde{\mathcal{C}}(\tilde{a}^{n+1}, \tilde{b}^{n+1}, \tilde{\phi}^{n+1}) \right) \end{cases}$$

The advantage of (3.13) over (3.6) is that the (ρ, w) update simplifies due to a simpler function $\tilde{\mathcal{K}}$ at the expense of often simple proximal updates of dual variables a, b . As before, except the ϕ update all other updates are decoupled at the grid points.

Additionally, we can obtain a grid independent convergence rate by introducing an inner product

$$\langle (a_1, b_1, \phi_1), (a_2, b_2, \phi_2) \rangle_* = \langle a_1, a_2 \rangle + \langle b_1, b_2 \rangle + \langle (AA^* + BB^*)\phi_1, \phi_2 \rangle,$$

which leads

$$\tilde{C}_*^* = (-\rho^{-0}, -\rho^{N_T}, (AA^* + BB^*)^{-1}(A\rho + Bw))$$

and a preconditioned version of (3.13):

$$(3.14) \quad \begin{cases} a^{n+1} = (I + \tau \partial_a \mathcal{F}^*)^{-1} (a^n + \tau (\rho^n)^{-0}) \\ b^{n+1} = (I + \tau \partial_b \mathcal{G}^*)^{-1} (b^n + \tau (\rho^n)^{N_T}) \\ \phi^{n+1} = \phi^n - \tau (AA^* + BB^*)^{-1} (A\rho^n + Bw^n) \\ (\tilde{a}^{n+1}, \tilde{b}^{n+1}, \tilde{\phi}^{n+1}) = 2(a^{n+1}, b^{n+1}, \phi^{n+1}) - (a^n, b^n, \phi^n) \\ (\rho^{n+1}, w^{n+1}) = (I + \sigma \tilde{\mathcal{K}})^{-1} \left((\rho^n, w^n) + \sigma \tilde{C} (\tilde{a}^{n+1}, \tilde{b}^{n+1}, \tilde{\phi}^{n+1}) \right) \end{cases}$$

4. Numerical Experiments. In this section, we present two numerical experiments to demonstrate the effectiveness of the proposed numerical scheme and its robustness with respect to singular limits $\nu = 0, \epsilon = 0$. We conduct the following experiments on a 2-dimensional torus $\mathbb{T}^2 = [0, 1] \times [0, 1]$ and a time interval $[0, T] = [0, 1]$.

4.1. Convergence of the algorithm. In this example, we numerically verify the grid-independent time steps as discussed in subsection 3.1. We apply the proposed monotone PDHG approach to the following mean-field game system with congestion with $\alpha = 1, \beta = 2$ in (2.7), $\epsilon = 0.1, \nu = 0.1$. The initial distribution is given as follows

$$\begin{aligned} \rho_0(x) &= \hat{\rho}_{c(0.25, 0.25)}(x) + \hat{\rho}_{c(0.75, 0.25)}(x) + \hat{\rho}_{c(0.25, 0.75)}(x) + \hat{\rho}_{c(0.75, 0.75)}(x), \\ \hat{\rho}_{c(a_1, a_2)}(x) &= \hat{c} \exp(-100((x_1 - a_1)^2 + (x_2 - a_2)^2)), \end{aligned}$$

where $\hat{c} > 0$ is a constant scalar such that $\int \rho_0(x) dx = 1$. For the mean-field game system, the terminal condition g is given by

$$g(x) = 0.1 (\sin(\pi(x_1 - 0.5)))^2 + 0.1 (\sin(2\pi(x_2 - 0.25)))^2.$$

In Figure 1, we show the numerical results with discretized grid $N_h = 40, N_t = 32$. We observe that the density moves towards points $(0.5, 0.25)$ and $(0.5, 0.75)$, which are the minima of the terminal cost function g . We also see the diffusion during the time evolution that is induced by the viscosity parameter $\nu = 0.1$.

To show that the convergence rate is independent of the grid-size, we set error tolerance $\varepsilon_1 = 10^{-3}, \varepsilon_2 = 10^{-4}, \varepsilon_3 = 10^{-5}$, and compute the number of iterations the algorithm needs to guarantee that the discretized continuity equation (2.37) satisfies

$$\text{Err}^n := \|\text{discretized continuity equation at } n\text{-th iteration}\|_{L^2} < \varepsilon.$$

Here, the norm $\|\cdot\|_{L^2}$ has been properly scaled with the spatial and temporal discretization $\Delta h, \Delta t$. We perform computations for various grid-sizes and summarize the results in Figure 2.

We fixed the optimization step-sizes $\sigma = \tau = 0.5$. Although preconditioning increases the computational cost per iteration, it is important to emphasize that it ultimately enhances the computational efficiency. To this end, we consider a scenario

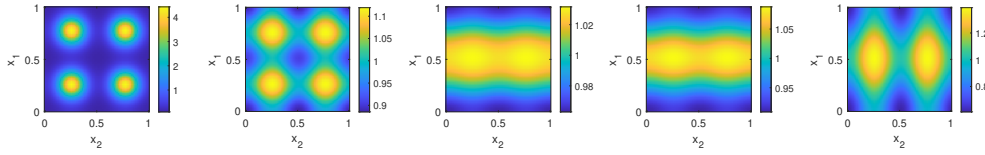


FIG. 1. This figure shows the density evolution of $\rho(x,t)$ in the MFG model with congestion parameters $\alpha = 1, \beta = 2, \epsilon = 0.1$ and viscosity $\nu = 0.1$, as detailed in subsection 4.1. Moving from left to right, at time points $t = 0, 0.25, 0.5, 0.75, 1$, the density starts in the form of 4 Gaussians and eventually concentrates at points $(0.5, 0.25)$ and $(0.5, 0.75)$.

N_h	N_t	$\varepsilon_1 = 10^{-3}$	$\varepsilon_2 = 10^{-4}$	$\varepsilon_3 = 10^{-5}$
20	16	107 (1.34 s)	137 (1.69 s)	196 (2.37 s)
32	20	117 (4.67 s)	147 (5.83 s)	197 (7.77 s)
40	32	117 (11.19 s)	157 (14.62 s)	197 (18.57 s)

FIG. 2. We record the number of iterations our algorithm requires to reach specific error thresholds, ε , across various mesh densities, along with the computational time needed (in seconds). Since the algorithm is not optimized for parallel execution, the computational time increases with the refinement of the mesh.

without preconditioning with $N_h = 20$ and $N_t = 16$, and the algorithm needs 59723 iterations (taking 508.73 seconds) to achieve a residual error smaller than $\varepsilon_1 = 10^{-3}$, and 110093 iterations (taking 932.74 seconds) to reach a residual error smaller than $\varepsilon_2 = 10^{-4}$. This example highlights how suitable choices of norms and inner products in the PDHG improves the overall efficiency of the proposed algorithm.

4.2. Viscosity Effect. In this section, we conduct experiments to demonstrate the impact of the viscosity parameter ν in the MFG model with congestion. The parameters α, β, ϵ are the same as in the previous section, and we compute the solutions for $\nu = 0, 0.02, 0.1$. The computations demonstrate that ν significantly affects the solution, as depicted in Figure 3. A stronger diffusion results in a more widespread solution. More importantly, the algorithm is robust with respect to small values of ν . This phenomenon is explained by the variational nature of the algorithm that can handle singular problems.

4.3. Congestion Effect. In this section, we explore scenarios where $\nu = 0$ and the congestion parameter is small ϵ . The goal of these experiments is to study the robustness of the algorithm with respect to further singularity introduced by small values of ϵ . Figure 4 illustrates that with an increasing congestion coefficient ϵ , the density tends to exhibit reduced movement. Consequently, when $\epsilon = 5$, the density undergoes only slight deformation. Additionally, as $\epsilon \rightarrow 0$, the solution consistently demonstrates behavior akin to the leftmost case with $\epsilon = 0$. As before, the algorithm demonstrates robust performance in all cases.

REFERENCES

- [1] Y. ACHDOU AND I. CAPUZZO-DOLCETTA, *Mean field games: numerical methods*, SIAM J. Numer. Anal., 48 (2010), pp. 1136–1162, <https://doi.org/10.1137/090758477>, <https://doi.org/10.1137/090758477>.
- [2] Y. ACHDOU AND A. PORRETTA, *Mean field games with congestion*, Ann. Inst. H. Poincaré C Anal. Non Linéaire, 35 (2018), pp. 443–480, <https://doi.org/10.1016/j.anihpc.2017.06.001>,

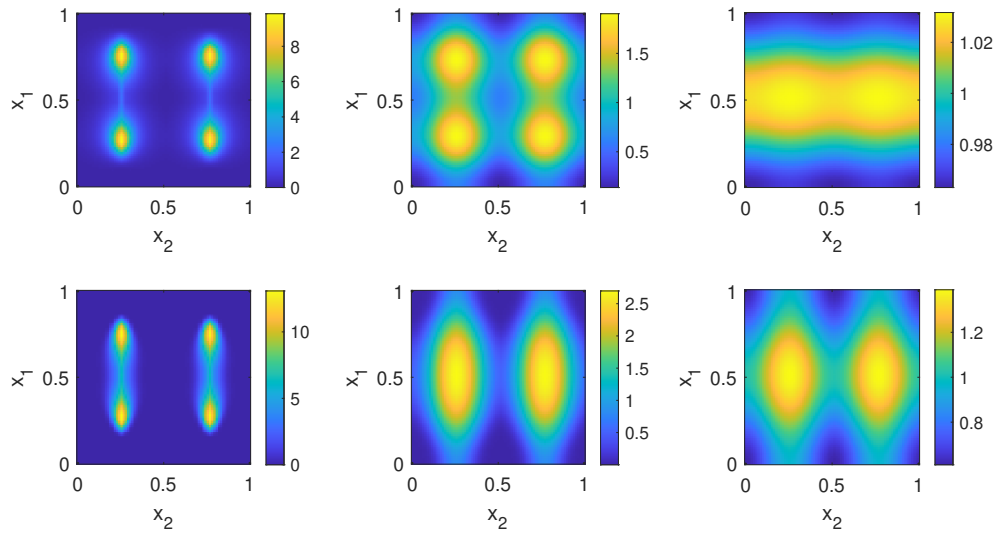


FIG. 3. This figure shows the numerical results of density evolution of congestion model with $\nu = 0, 0.02, 0.1$ (from left to right): $\rho(0.5, x)$ (top) and $\rho(1, x)$ (bottom).

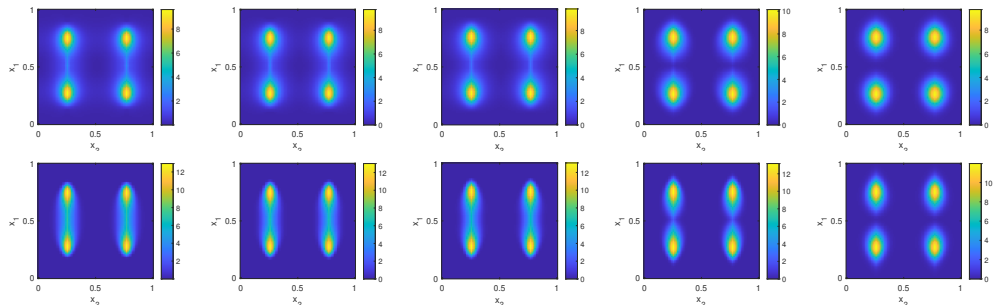


FIG. 4. This figure shows the numerical results of density evolution of congestion model with $\epsilon = 0, 0.02, 0.1, 1, 5$ (from left to right): $\rho(0.5, x)$ (top) and $\rho(1, x)$ (bottom).

<https://doi.org/10.1016/j.anihpc.2017.06.001>.

[3] N. ALMULLA, R. FERREIRA, AND D. GOMES, *Two numerical approaches to stationary mean-field games*, *Dynamic Games and Applications*, 7 (2017), pp. 657–682, <https://doi.org/10.1007/s13235-016-0203-5>, <https://doi.org/10.1007/s13235-016-0203-5>.

[4] J.-P. AUBIN AND H. FRANKOWSKA, *Set-valued analysis*, vol. 2 of *Systems & Control: Foundations & Applications*, Birkhäuser Boston, Inc., Boston, MA, 1990.

[5] J.-D. BENAMOU AND Y. BRENIER, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, *Numer. Math.*, 84 (2000), pp. 375–393, <https://doi.org/10.1007/s002110050002>, <https://doi.org/10.1007/s002110050002>.

[6] J.-D. BENAMOU AND G. CARLIER, *Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations*, *J. Optim. Theory Appl.*, 167 (2015), pp. 1–26, <https://doi.org/10.1007/s10957-015-0725-9>.

[7] J.-D. BENAMOU, G. CARLIER, AND F. SANTAMBROGIO, *Variational mean field games*, in *Active particles. Vol. 1. Advances in theory, models, and applications*, *Model. Simul. Sci. Eng. Technol.*, Birkhäuser/Springer, Cham, 2017, pp. 141–171, https://doi.org/10.1007/978-3-319-49996-3_4.

- [8] L. BRICEÑO ARIAS, D. KALISE, Z. KOBEISSI, M. LAURIÈRE, A. MATEOS GONZÁLEZ, AND F. J. SILVA, *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, vol. 65 of ESAIM Proc. Surveys, EDP Sci., Les Ulis, 2019, pp. 330–348.
- [9] L. M. BRICEÑO ARIAS, D. KALISE, AND F. J. SILVA, *Proximal methods for stationary mean field games with local couplings*, SIAM J. Control Optim., 56 (2018), pp. 801–836, <https://doi.org/10.1137/16M1095615>, <https://doi.org/10.1137/16M1095615>.
- [10] F. CAMILLI AND Q. TANG, *A convergence rate for the newton’s method for mean field games with non-separable hamiltonians*, 2023, <https://arxiv.org/abs/2311.05416>.
- [11] P. L. COMBETTES, *Perspective functions: properties, constructions, and examples*, Set-Valued Var. Anal., 26 (2018), pp. 247–264, <https://doi.org/10.1007/s11228-017-0407-x>, <https://doi.org/10.1007/s11228-017-0407-x>.
- [12] D. A. GOMES AND J. SAÚDE, *Numerical methods for finite-state mean-field games satisfying a monotonicity condition*, Applied Mathematics & Optimization, 83 (2021), pp. 51–82, <https://doi.org/10.1007/s00245-018-9510-0>, <https://doi.org/10.1007/s00245-018-9510-0>.
- [13] D. A. GOMES AND X. YANG, *The hessian riemannian flow and newton’s method for effective hamiltonians and mather measures*, ESAIM: M2AN, 54 (2020), pp. 1883–1915, <https://doi.org/10.1051/m2an/2020036>, <https://doi.org/10.1051/m2an/2020036>.
- [14] M. HUANG, P. E. CAINES, AND R. P. MALHAMÉ, *Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ϵ -Nash equilibria*, IEEE Trans. Automat. Control, 52 (2007), pp. 1560–1571, <https://doi.org/10.1109/TAC.2007.904450>, <https://doi.org/10.1109/TAC.2007.904450>.
- [15] M. HUANG, R. P. MALHAMÉ, AND P. E. CAINES, *Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle*, Commun. Inf. Syst., 6 (2006), pp. 221–251, <http://projecteuclid.org/euclid.cis/1183728987>.
- [16] S. KAKUTANI, *A generalization of Brouwer’s fixed point theorem*, Duke Math. J., 8 (1941), pp. 457–459, <http://projecteuclid.org/euclid.dmj/1077492791>.
- [17] J.-M. LASRY AND P.-L. LIONS, *Jeux à champ moyen. I. Le cas stationnaire*, C. R. Math. Acad. Sci. Paris, 343 (2006), pp. 619–625, <https://doi.org/10.1016/j.crma.2006.09.019>, <https://doi.org/10.1016/j.crma.2006.09.019>.
- [18] J.-M. LASRY AND P.-L. LIONS, *Jeux à champ moyen. II. Horizon fini et contrôle optimal*, C. R. Math. Acad. Sci. Paris, 343 (2006), pp. 679–684, <https://doi.org/10.1016/j.crma.2006.09.018>, <https://doi.org/10.1016/j.crma.2006.09.018>.
- [19] J.-M. LASRY AND P.-L. LIONS, *Mean field games*, Jpn. J. Math., 2 (2007), pp. 229–260, <https://doi.org/10.1007/s11537-007-0657-8>, <https://doi.org/10.1007/s11537-007-0657-8>.
- [20] M. LAURIÈRE, J. SONG, AND Q. TANG, *Policy iteration method for time-dependent mean field games systems with non-separable hamiltonians*, Applied Mathematics & Optimization, 87 (2023), p. 17, <https://doi.org/10.1007/s00245-022-09925-5>, <https://doi.org/10.1007/s00245-022-09925-5>.
- [21] S. LIU AND L. NURBEKYAN, *Splitting methods for a class of non-potential mean field games*, 2020, <https://arxiv.org/abs/2007.00099>.
- [22] B. C. VŨ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681, <https://doi.org/10.1007/s10444-011-9254-8>, <https://doi.org/10.1007/s10444-011-9254-8>.