Preconditioning the augmented Lagrangian method for instationary mean field games with diffusion

Roman Andreev

PRECONDITIONING THE AUGMENTED LAGRANGIAN METHOD FOR INSTATIONARY MEAN FIELD GAMES WITH DIFFUSION

ROMAN ANDREEV

Abstract. We apply the augmented Lagrangian method to the convex optimization problem of the instationary variational mean field games with diffusion. The system is first discretized with space-time tensor product piecewise polynomial bases. This leads to a sequence of linear problems posed on the space-time cylinder that are second order in the temporal variable and fourth order in the spatial variable. To solve these large linear problems with the preconditioned conjugate gradients method we propose a parameter-robust preconditioner that is based on a temporal transformation coupled with a spatial multigrid. Numerical examples illustrate the method.

1. Introduction

Mean field games and related models describe a wide range of social phenomena such as crowd motion, opinion dynamics, vaccination rates, stability of marriage, etc., and, moreover, appear as the equations to be solved in each time-step of the so-called JKO time-stepping scheme for gradient flows. In the instationary stochastic case introduced in [17], mean field games is a coupled system of a transport-diffusion equation for a density (of crowd, opinion, etc.) with a nonlinear equation for the value function running in the opposite temporal direction. Existing numerical methods for mean field games are based for instance on finite volumes [1, 2, 3], the dynamic programming principle [10, 12], or on convex duality [8]; further references can be found in [9]. In this work we reconsider the convex duality formulation of [11] and the ALG2 splitting method of [15, 8] including nonzero diffusion. This leads to a sequence of linear problems posed on the space-time cylinder that are second order in time and fourth order in space. To solve these space-time problems with an iterative method such as conjugate gradients, we develop a preconditioner based on a temporal decoupling transformation and a spatial multigrid. The resulting preconditioner is robust in all parameters, including the diffusion coefficient.

The paper is structured as follows. In Section 2 we introduce the mean field games model, its convex formulation and the ALG2 method. In Section 3 we describe the discrete version of ALG2. In Section 4 we comment on preconditioning of the space-time linear problems. The numerical experiments in Section 5 conclude the paper.

2. ALG2 with diffusion

2.1. **Mean field games.** Let $D \subset \mathbb{R}^d$ be a cube. Let $T > 0$. An in [11, 8] we consider the coupled system of partial differential equations

$$\mathrm{KFP}[\rho,\phi] := \partial_t \rho - \nu^2 \Delta \rho + \operatorname{div}(\rho \nabla H(t,x,\nabla\phi)) = 0, \tag{1a}$$

$$\mathrm{HJB}[\rho,\phi] := \partial_t \phi + \nu^2 \Delta \phi + H(t,x,\nabla\phi) = A'(t,x,\rho), \tag{1b}$$

$$\text{s.t.} \qquad \rho(0) = \rho_0 \quad \text{and} \quad \phi(T) = -\Gamma'(x,\rho(T)). \tag{1c}$$

We will refer to this system as the mean field games (MFGs) equations. Here, A and Γ are convex real valued functions of the third variable $\rho \geq 0$ (equal to $+\infty$ for negative ρ) and the indicated derivatives are with respect to this variable. The essential assumption on the Hamiltonian H is convexity with respect to the third variable. The unknowns are the density ρ and the cost ϕ, both space-time dependent real valued functions. We omit the dependence on (t,x) in the notation

Date: April 11, 2016.

I thank Y. Achdou and J.-D. Benamou for their technical advice. Support: ANR-12-MONU-0013 "ISOTACE".

where convenient. Periodic boundary conditions are often assumed but here we will be interested in no-flow boundary conditions on the density ρ. These are implemented by requiring

$$\nabla H(t,x,\nabla\phi)\cdot\mathbf{n}=0 \quad \text{and} \quad \nabla\rho\cdot\mathbf{n}=0 \quad \text{on} \quad \partial D \tag{2}$$

where $\mathbf{n}$ is the outward normal to the spatial boundary. In particular, the total mass $\int_D \rho$ is conserved in time, and we assume that $\rho_0 \geq 0$ with nonzero total mass. We restrict ourselves to radially symmetric Hamiltonians (in the third variable); hence the first condition of (2) amounts to

$$\nabla\phi\cdot\mathbf{n}=0 \quad \text{on} \quad \partial D. \tag{3}$$

The main innovation of this work with respect to the numerical method proposed in [8] is the presence of the diffusion coefficient

$$\nu>0, \tag{4}$$

which is a positive constant, uniform in space-time.

The principal feature of the MFGs equations is that the Kolmogorov–Fokker–Planck (KFP) equation evolves forward in time with an explicit initial condition at $t=0$ and the Hamilton–Jacobi–Bellman (HJB) equation evolves backward in time with a possibly implicit initial condition at $t=T$. The general mathematical interpretation of the KFP is in the weak sense and that of the HJB is in the viscosity sense, but we will mostly proceed in a formal way, in particular assuming the regularity (10)–(11) below.

The MFGs equations arise as follows. Consider the transport equation (slightly overloading the notation) $\mathrm{KFP}[\rho,\mathbf{v}] := \partial_t\rho - \nu^2\Delta\rho + \operatorname{div}(\rho\mathbf{v}) = 0$, where the space-time dependent vector field $\mathbf{v}$ is a control variable. We fix $\mathbf{v}$ by minimizing the functional

$$J_2(\rho,\mathbf{v}) := \int_0^T \int_D \{L(t,x,\mathbf{v})\rho + A(t,x,\rho)\} + \int_D \Gamma(x,\rho(T)), \tag{5}$$

where L is the Lagrangian related to the Hamiltonian by convex duality,

$$H(\mathbf{p}) := L^\star(\mathbf{p}) := \sup_{\mathbf{v}\in\mathbb{R}^d}\{\mathbf{p}\cdot\mathbf{v} - L(\mathbf{v})\}. \tag{6}$$

To that end we look for saddle points of the space-time Lagrangian $J_2(\rho,\mathbf{v}) + \langle\phi, \mathrm{KFP}[\rho,\mathbf{v}]\rangle$. Proceeding in a formal way, the derivative with respect to $\mathbf{v}$ gives the relation $\nabla L(\mathbf{v}) = \nabla\phi$, at least where $\rho\neq 0$. By duality $H = L^\star$, this implies the feedback strategy

$$\mathbf{v} = \nabla H(\nabla\phi) \tag{7}$$

and the representation $H(\nabla\phi) = \mathbf{v}\cdot\nabla\phi - L(\mathbf{v})$. Employing the latter in the derivative of the space-time Lagrangian with respect to ρ gives the HJB equation with its terminal condition.

2.2. **The predual problem.** We will work with the function spaces

$$\mathsf{H} := L_2(D), \quad \mathsf{V} := H^1_{\mathrm{Neu}}(D), \quad \mathsf{W} := H^2(D)\cap\mathsf{V}. \tag{8}$$

We write $\|\cdot\|_{J\times D}$ / $\|\cdot\|_D$ to indicate the $L_2(J\times D)$ / $L_2(D)$ norm. On V, which incorporates homogeneous Neumann boundary conditions, we use the norm given by

$$\|\chi\|_{\mathsf{V}}^2 := \|\chi\|_D^2 + \nu^2\|\nabla\chi\|_D^2, \quad \chi\in\mathsf{V}. \tag{9}$$

Introducing the Bochner–Sobolev space

$$\mathsf{X} := L_2((0,T);\mathsf{W})\cap H^1((0,T);\mathsf{H}), \tag{10}$$

we suppose that

$$\phi\in\mathsf{X}. \tag{11}$$

We abbreviate $L_2 := L_2((0,T)\times D)$. Recall the canonical isometry $L_2 \cong L_2((0,T);L_2(D))$. We identify $\mathsf{H}\cong\mathsf{H}'$ via the Riesz isomorphism, obtaining the Gelfand triple $\mathsf{V}\hookrightarrow\mathsf{H}\cong\mathsf{H}'\hookrightarrow\mathsf{V}'$, and in particular the embedding $\mathsf{V}\hookrightarrow\mathsf{V}'$. Set

$$\mathsf{Y} := L_2\times L_2^d\times\mathsf{V}. \tag{12}$$

Elements of Y will be denoted by

$$\sigma = (a, b, c) \quad \text{or} \quad \lambda = (\rho, m, e) \tag{13}$$

and those of Y' by $\tilde{\lambda} = (\tilde{\rho}, \tilde{m}, \tilde{e})$. Following [8] we define the linear operator

$$\Lambda : \mathsf{X} \to \mathsf{Y}, \quad \phi \mapsto (\partial_t \phi + \nu^2 \Delta \phi, \nabla \phi, -\phi(T)). \tag{14}$$

This operator is injective because testing $\partial_t \phi + \nu^2 \Delta \phi = 0$ by ϕ and integrating by parts gives $\|\phi(t)\|_{\mathsf{H}} \leq \|\phi(T)\|_{\mathsf{H}}$, and therefore $\Lambda \phi = 0$ implies $\phi = 0$. This argument motivates the third term of Λ. The key observation made and exploited in [11, 8] is that the constrained optimization problem

$$\boxed{\operatorname*{argmin}_{(\phi,\sigma)\in \mathsf{X}\times\mathsf{Y}} \{\mathcal{F}(\phi) + \mathcal{G}(\sigma)\} \quad \text{s.t.} \quad \sigma = \Lambda\phi} \tag{15}$$

with

$$\mathcal{F}(\phi) := \int_D \phi(0)\rho_0, \quad \mathcal{G}(\sigma) := \int_0^T \int_D A^\star(a + H(b)) + \int_D \Gamma^\star(c) \tag{16}$$

is a reformulation of the MFGs equations (1). Here, $A^\star/\Gamma^\star$ is the convex conjugate of A/Γ. More precisely, the dual of this problem is the minimization problem discussed in the previous subsection. Indeed, by the general theory of duality [14], the (usual) dual problem of (15) reads

$$\inf_{\tilde{\lambda}\in Y'} \{\mathcal{F}^\star(-\Lambda'\tilde{\lambda}) + \mathcal{G}^\star(\tilde{\lambda})\}, \tag{17}$$

where $\Lambda' : Y' \to X'$ is the adjoint. Since $H(b) = \sup_{\tilde{m}}\{b \cdot \tilde{m}/\tilde{\rho} - L(\tilde{m}/\tilde{\rho})\}$ for any $\tilde{\rho} > 0$, we have

$$A^\star(a + H(b)) = \sup_{\tilde{\rho}}\{(a + H(b))\tilde{\rho} - A(\tilde{\rho})\} = \sup_{\tilde{\rho},\tilde{m}}\{(a, b) \cdot (\tilde{\rho}, \tilde{m}) - [L(\tilde{m}/\tilde{\rho})\tilde{\rho} + A(\tilde{\rho})]\}, \tag{18}$$

where the supremum over $\tilde{\rho}$ is restricted to $\tilde{\rho} > 0$. Thus

$$\mathcal{G}^\star(\tilde{\lambda}) = \mathcal{G}^\star(\tilde{\rho}, \tilde{m}, \tilde{e}) = \int_{(0,T)\times D} \{L(\tilde{m}/\tilde{\rho})\tilde{\rho} + A(\tilde{\rho})\} + \int_D \Gamma(\tilde{e}). \tag{19}$$

Concerning $\mathcal{F}^\star$, we observe

$$\langle -\Lambda\phi, \tilde{\lambda}\rangle = \langle \phi, \partial_t\tilde{\rho} - \nu^2\Delta\tilde{\rho}\rangle - \langle \phi, \tilde{\rho}\rangle|_{t=0}^{t=T} + \langle \phi, \operatorname{div}\tilde{m}\rangle + \langle \phi(T), \tilde{e}\rangle - \mathrm{BT}, \tag{20}$$

where the spatial boundary terms are collected in

$$\mathrm{BT} := \int_J \int_{\partial D} \left(\nu^2\tilde{\rho}\nabla\phi - (\nu^2\nabla\tilde{\rho} - \tilde{m})\phi\right) \cdot \mathbf{n}. \tag{21}$$

Invoking the spatial homogeneous Neumann boundary conditions (3) we find that

$$\mathcal{F}^\star(-\Lambda'\tilde{\lambda}) \text{ is finite (and equals zero)} \tag{22}$$

if and only if the equations

$$\partial_t\tilde{\rho} - \nu^2\Delta\tilde{\rho} + \operatorname{div}\tilde{m} = 0, \quad \tilde{\rho}(0) = \rho_0, \quad \nu^2\nabla\tilde{\rho} \cdot \mathbf{n} = \tilde{m} \cdot \mathbf{n} \quad \text{and} \quad \tilde{\rho}(T) = \tilde{e} \tag{23}$$

are satisfied in the weak sense. In other words, the optimization in (17) is really performed over the variables $\tilde{\rho}$ and $\tilde{m}$ subject to these conditions. Writing $\mathbf{v} := \tilde{m}/\tilde{\rho}$ we find that (17) is just the problem of minimizing (5) subject to $\mathrm{KFP}[\rho, \mathbf{v}] = 0$ from the previous subsection.

2.3. **On existence of minimizers.** The following two classical results [14, Remark 4.2] concern the optimization problems (15)–(17).

Theorem 2.1. *Suppose $\mathcal{F}$ and $\mathcal{G}$ are convex. Suppose there exists ϕ_0 such that $\mathcal{F}(\phi_0)$ and $\mathcal{G}(\Lambda\phi_0)$ are finite and $\mathcal{G}$ is continuous at $\Lambda\phi_0$. Then* $\inf(15) = \inf(17)$ *and* (17) *has a minimizer.*

Theorem 2.2. $(\phi, \tilde{\lambda})$ *solves* (15)–(17) *and* $\inf(15) = \inf(17)$ *iff* $(-\Lambda'\tilde{\lambda}, \tilde{\lambda}) \in \partial\mathcal{F}(\phi) \times \partial\mathcal{G}(\Lambda\phi)$.

Typically, $A^\star$ and $\Gamma^\star$ will be convex nondecreasing functions, and together with convexity of H those conditions suffice to show convexity of F and G directly. The two inclusions from Theorem 2.2 correspond to the KFP (1a) and the HJB (1b) equations, respectively.

The built-in regularity assumption $\phi \in \mathsf{X}$ in (15) implies that $\phi(0) \in \mathsf{V}$, so that $F(\phi)$ is well-defined even for $\rho_0 \in \mathsf{V}'$. Let $\sigma = (a, b, c) = \Lambda\phi$. Then $a \in L_2$, $b \in \mathsf{V}^d$ and $c \in \mathsf{V}$, and certain growth conditions on $A^\star$, $\Gamma^\star$ and H ensure that $G(\sigma)$ is well-defined. Let us focus on $d = 2$ spatial dimensions. Then we may assume quadratic growth for $A^\star$ and $\Gamma^\star$ and arbitrary polynomial growth for H in the last variable. For an arbitrary $\sigma \in \mathsf{Y}$ not in the range of Λ we may extend the definition of G by $+\infty$ but this would preclude the usage of Theorem 2.1. Alternatively, we could restrict Y, say $\mathsf{Y} = L_2 \times L_2((0,T); H^1(D)) \times \mathsf{V}$, but we chose not follow this road in this paper in order to simplify the numerics in Step B of ALG2 below. Another possibility is to assume convexity and at most quadratic growth for $A^\star$ and $A^\star \circ H$ ensuring continuity of G on Y so that Theorem 2.1 could be used but this would rule out the desirable situation of the standard quadratic Hamiltonian H together with a quadratic cost A in (5). In any case Theorem 2.1 does not address existence for the primal problem (15), and seeing (17)–(15) as the primal–dual pair is problematic because $F^\star$ is discontinuous wherever it is finite (unless again, we modify Y suitably).

Existence and stability in space-time Lebesgue spaces have been obtained using Theorem 2.1 in [11], and by operating directly on the MFGs equations for example in [18, Theorem 2.7] and [3, Theorem 3.1].

Another approach to existence of minimizers in (15) is to verify coercivity of the functional under suitable assumptions on the data (which would also play a role in establishing Γ-convergence of numerical solutions [8, Section 3.1]). This, however, does not seem possible because the term $\int_D \Gamma^\star(-\phi(T))$ does not provide control on the spatial derivates of $\phi(T)$, which would be necessary to control the norm of ϕ in X (see e.g. [6, Theorem 4.1] for that kind of statement). We believe the mesh-dependent convergence rate of ALG2, see Section 5.4, is a manifestation of this fact.

2.4. **ALG2 formulation.** In order to solve the optimization problem (15), in the augmented Lagrangian method one looks for saddle points $(\phi, \sigma, \lambda) \in \mathsf{X} \times \mathsf{Y} \times \mathsf{Y}$ of the augmented Lagrangian

$$L_r(\phi, \sigma, \lambda) = \mathcal{F}(\phi) + \mathcal{G}(\sigma) + (\Lambda\phi - \sigma, \lambda)_{\mathsf{Y}} + \tfrac{r}{2}\|\Lambda\phi - \sigma\|^2_{\mathsf{Y}}, \tag{24}$$

where $r \geq 0$; these are characterized by

$$L_r(\phi, \sigma, \cdot) \leq L_r(\phi, \sigma, \lambda) \leq L_r(\cdot, \cdot, \lambda). \tag{25}$$

It is elementary to check that the Lagrangians L_r and L_0 have the same saddle points. Moreover, any saddle point furnishes a solution to (15) and (17). With the notation of the previous subsection it would have been natural to have the duality pairing $\langle \Lambda\phi - \sigma, \tilde{\lambda}\rangle$ with $\tilde{\lambda} \in \mathsf{Y}'$ instead of the Y scalar product but below it is more convenient to work with the Riesz representative $\lambda \in \mathsf{Y}$ given by $(\lambda, \cdot)_{\mathsf{Y}} = \tilde{\lambda}$. The first two components of λ and $\tilde{\lambda}$ are the same.

Consider the function $h_r(\lambda) := \inf_{\phi,\sigma} L_r(\phi, \sigma, \lambda)$. Then, λ in the saddle point characterization (25) maximizes this function. At any point $\bar{\lambda}$, its gradient ascent direction is $(\Lambda\bar{\phi} - \bar{\sigma})$ where $(\bar{\phi}, \bar{\sigma}) := \operatorname{argmin} L_r(\cdot, \cdot, \bar{\lambda})$. This suggests a steepest descent algorithm for finding the optimal λ. This is the algorithm "ALG1" in [15, Section 3.1]. The subsequent algorithm "ALG2" [15, Section 3.2] is a modification where the minimization of ϕ and σ is decoupled. It proceeds by iterating the following three steps.

A. Minimize L_r with respect to the first component by solving the elliptic problem

$$\text{Find} \quad \phi^{(k+1)} \in \mathsf{X} \quad \text{s.t.} \quad \langle \tfrac{\partial L_r}{\partial \phi}(\phi^{(k+1)}, \sigma^{(k)}, \lambda^{(k)}), \tilde{\phi}\rangle = 0 \quad \forall \tilde{\phi} \in \mathsf{X}. \tag{26}$$

B. Proximal step:

$$\sigma^{(k+1)} := \operatorname*{argmin}_{\sigma \in \mathsf{Y}} \{G(\sigma) + \tfrac{r}{2}\|\bar{\sigma}^{(k+1)} - \sigma\|^2_{\mathsf{Y}}\} \quad \text{for} \quad \bar{\sigma}^{(k+1)} := \Lambda\phi^{(k+1)} + \tfrac{1}{r}\lambda^{(k)}. \tag{27}$$

C. Multiplier update:

$$\lambda^{(k+1)} := \lambda^{(k)} + r(\Lambda\phi^{(k+1)} - \sigma^{(k+1)}). \tag{28}$$

It is worth noting that convergence of the algorithm is robust under perturbations in Step A and Step B if those perturbations decay sufficiently fast with the iteration [13, Theorem 8]. We now comment on each of those steps separately.

2.4.1. *Step A.* Since $\mathcal{F}$ is linear, (26) amounts to

$$\text{Find} \quad \phi^{(k+1)} \in \mathsf{X} : (\Lambda\phi^{(k+1)}, \Lambda\tilde{\phi})_{\mathsf{Y}} = (\sigma^{(k)} - \tfrac{1}{r}\lambda^{(k)}, \Lambda\tilde{\phi})_{\mathsf{Y}} - \tfrac{1}{r}\mathcal{F}(\tilde{\phi}) \quad \forall \tilde{\phi} \in \mathsf{X}. \tag{29}$$

The only term involving the unknown $\phi^{(k+1)}$ is (we suppress the iteration index)

$$(\Lambda\phi, \Lambda\tilde{\phi})_{\mathsf{Y}} = \langle (\partial_t + \nu^2\Delta)\phi, (\partial_t + \nu^2\Delta)\tilde{\phi}\rangle + \langle \nabla\phi, \nabla\tilde{\phi}\rangle + (\phi(T), \tilde{\phi}(T))_{\mathsf{V}}. \tag{30}$$

Expanding the first term on the right hand side we obtain

$$\langle (\partial_t + \nu^2\Delta)\phi, \ldots\rangle = \langle \partial_t\phi, \partial_t\tilde{\phi}\rangle + \nu^4\langle \Delta\phi, \Delta\tilde{\phi}\rangle - \nu^2\langle \nabla\phi(t), \nabla\tilde{\phi}(t)\rangle|_{t=0}^{t=T}, \tag{31}$$

having used integration by parts in time and space on the term $\langle \partial_t\phi, \Delta\tilde{\phi}\rangle$. The boundary term $\nu^2\int_{J\times\partial D}(\phi\nabla\tilde{\phi} - \tilde{\phi}\nabla\phi)\cdot\mathbf{n}$ disappears for any combination of homogeneous/periodic Dirichlet/Neumann spatial boundary conditions. The negative ν^2 term cancels by the definition of the norm on V, so we are left with

$$\|\Lambda\phi\|_{\mathsf{Y}}^2 = \|\partial_t\phi\|_{J\times D}^2 + \nu^4\|\Delta\phi\|_{J\times D}^2 + \|\nabla\phi\|_{J\times D}^2 + \|\phi(T)\|_D^2 + \nu^2\|\nabla\phi(0)\|_D^2. \tag{32}$$

Consider the operator $A := \Lambda'\Lambda : \mathsf{X} \to \mathsf{X}'$, where the adjoint is with respect to the Y scalar product; hence $\langle A\phi, \tilde{\phi}\rangle =$ (32). Below we will iteratively invert a discretized version of A, and will therefore require a good preconditioner. To that end observe that the symmetric operator $C : \mathsf{X} \to \mathsf{X}'$ defined by omitting the last term, viz.

$$\langle C\phi, \phi\rangle := \|\partial_t\phi\|_{J\times D}^2 + \|\phi(T)\|_D^2 + \|\nabla\phi\|_{J\times D}^2 + \nu^4\|\Delta\phi\|_{J\times D}^2, \tag{33}$$

is equivalent to A by the argument given in [4, Section 2.5] for $-\Delta$ instead of $\nu^4\Delta^2$.

2.4.2. *Step B.* Recall the notation $\sigma = (a, b, c)$ and $\lambda = (\rho, m, e)$. Step B consists of two decoupled proximal subproblems, the first one being

$$(a^{(k+1)}, b^{(k+1)}) := \operatorname*{argmin}_{(a,b)\in L_2\times L_2^d} \left\{ \textstyle\int_D A^\star(a + H(b)) + \frac{r}{2}\|(\bar{a}^{(k+1)}, \bar{b}^{(k+1)}) - (a, b)\|_{L_2\times L_2^d}^2 \right\} \tag{34}$$

with the "priors" $\bar{a}^{(k+1)} := (\partial_t + \nu^2\Delta)\phi^{(k+1)} + \frac{1}{r}\rho^{(k)}$ and $\bar{b}^{(k+1)} := \nabla\phi^{(k+1)} + \frac{1}{r}m^{(k)}$, and the second one being

$$c^{(k+1)} := \operatorname*{argmin}_{c\in V} \left\{ \textstyle\int_D \Gamma^\star(c) + \frac{r}{2}\|\bar{c}^{(k+1)} - c\|_{\mathsf{V}}^2 \right\} \tag{35}$$

with the "prior" $\bar{c}^{(k+1)} := -\phi^{(k+1)}(T) + \frac{1}{r}e^{(k)}$. The first subproblem is as in [8, Section 4.2] but the second is not due to the non-L_2 norm, which is a consequence of the choice (12) of the space Y.

2.4.3. *Step C.* Step C is a straightforward update.

3. Discretization

3.1. **Discrete spaces.** We discretize the quantities (ϕ, σ, λ) in (24) using tensor products of piecewise polynomial functions on an interval. We use the following notation, leaving the underlying mesh to be specified separately:

- P1. Continuous piecewise affine functions (hat functions);
- P2. Continuous piecewise quadratic functions;
- D0. Piecewise constant functions (top functions);
- D1. Piecewise polynomials of degree one with no interelement continuity;
- B2. Piecewise polynomials of degree two with C^1 continuity.

For simplicity, we assume here that the spatial dimension is $d = 2$. We write $\mathrm{P2} \otimes \mathrm{B2}^{(2)}$ for the function space spanned by the products of P2 functions in time with B2 functions in each spatial dimension, etc. We will look for a discrete saddle point $(\phi_h, \sigma_h, \lambda_h)$ as follows

$$\phi_h \in \mathsf{X}_h := \mathrm{P2} \otimes \mathrm{B2}^{(2)} \tag{36a}$$

$$a_h, \rho_h \in \mathsf{A}_h := \mathrm{D1} \otimes \mathrm{D0}^{(2)} \tag{36b}$$

$$b_h, m_h \in \mathsf{B}_h := \mathrm{D0} \otimes [(\mathrm{P1} \otimes \mathrm{D0}) \times (\mathrm{D0} \otimes \mathrm{P1})] \tag{36c}$$

$$c_h, e_h \in \mathsf{C}_h := \mathrm{B2}^{(2)}. \tag{36d}$$

Furthermore, we set $\mathsf{Y}_h := \mathsf{A}_h \times \mathsf{B}_h \times \mathsf{C}_h$ with the norm Y. The motiation for taking B2 in (36a) is mainly $H^2(D)$-conformity required for the regularity (11). It also interacts well with the choice D0 in (36b) because D0 simulaneously approximates functions in B2 and their second derivatives well, which plays a role in (42) and leads to the dimension formula (38) below. An alternative to the P2–D1 combination in (36a)–(36b) is to take P$(p+1)$–D(p) for any degree $p \geq 0$, so that (42) below corresponds to a so-called continuous Galerkin time-stepping scheme. The choice of the spatial component in (36c) allows integration by parts in space in an expression like $\langle m_h, -\nabla\phi_h\rangle$ and approximates $\nabla\phi_h$ well. Finally, the space in (36d) is simply the trace of (36a) at $t = T$, and is sufficiently regular for the proximal step (48) to be well-defined.

We impose the no-flux boundary conditions for the density ρ through (cf. Section 2.1):

(37a) homogeneous Neumann spatial boundary conditions on X_h;

(37b) homogeneous Dirichlet boundary conditions on the P1 components of B_h.

With these boundary conditions, the number of spatial degrees of freedom of X_h (i.e., $\dim \mathrm{B2}^{(2)}$) and those of A_h (i.e., $\dim \mathrm{D0}^{(2)}$) coincide, and therefore

$$\dim \mathsf{X}_h = \dim \mathsf{A}_h + \dim \mathsf{C}_h. \tag{38}$$

The operator Λ is approximated by the (injective) operator

$$\Lambda_h : \mathsf{X}_h \to \mathsf{Y}_h, \quad \Lambda_h\phi := (Q_1(\partial_t + \nu^2\Delta)\phi), Q_2\nabla\phi, -\phi(T)), \tag{39}$$

where Q_1 is the L_2-orthogonal projection onto A_h and Q_2 is the componentwise L_2-orthogonal projection onto B_h. These projections have no effect in the term $(\Lambda\phi_h - \sigma_h, \lambda_h)_{\mathsf{Y}}$ but they do affect the r-term of the augmented Lagrangian (24).

Our aim is therefore to solve the discrete optimization problem (cf. (15))

$$\boxed{\operatorname*{argmin}_{(\phi_h,\sigma_h)\in\mathsf{X}_h\times\mathsf{Y}_h} \{\mathcal{F}_h(\phi_h) + \mathcal{G}_h(\sigma_h)\} \quad \text{s.t.} \quad \sigma_h = \Lambda_h\phi_h} \tag{40}$$

with $\mathcal{F}_h := \mathcal{F}$ and $\mathcal{G}_h := \mathcal{G}$ from (16) (or convex approximations thereof). The analog of the continuous dual problem (17) now reads

$$\inf_{\tilde\lambda_h\in\mathsf{Y}_h'} \{\mathcal{F}_h^\star(-\Lambda_h'\tilde\lambda_h) + \mathcal{G}_h^\star(\tilde\lambda_h)\}, \tag{41}$$

with $\mathcal{F}_h^\star(-\Lambda_h'\tilde\lambda_h) =$

$$\sup_{\phi_h\in\mathsf{X}_h} \{-\langle\rho_h, (\partial_t + \nu^2\Delta)\phi_h\rangle - \langle m_h, \nabla\phi\rangle - \langle\rho_0, \phi_h(0)\rangle + \langle\tilde e_h, \phi_h(T)\rangle\}, \tag{42}$$

where, notably, the supremum is taken over discrete functions only. Here, $\tilde\lambda_h = (\rho_h, m_h, \tilde e_h)$. It is possible to integrate by parts in space on the term $\langle m_h, -\nabla\phi_h\rangle$ owing to the choice of the discrete spaces (36a) and (36c). Therefore, the supremum (42) is finite (and equals zero) if and only if the triple $(\rho_h, m_h, \tilde e_h) \in \mathsf{Y}_h'$ satisfies the following discrete analog of (1a),

$$\langle\rho_h, (-\partial_t - \nu^2\Delta)\phi\rangle = \langle -\operatorname{div} m_h, \phi\rangle + \langle\rho_0, \phi(0)\rangle - \langle\tilde e_h, \phi(T)\rangle \quad \forall\phi \in \mathsf{X}_h. \tag{43}$$

In particular, fixing m_h and restricting the test functions to $\phi(T) = 0$, the dimension count (38) suggests that ρ_h is well-defined by (43) and approximates the solution ρ of $\partial_t\rho - \nu^2\Delta\rho = -\operatorname{div} m_h$ with $\rho(0) = \rho_0$ and $\nu^2\nabla\rho\cdot\mathbf{n} = 0$ on ∂D in the space-time ultraweak sense (spatial and temporal derivatives are on the discrete test function; the analysis can be done along the lines of [7, 5]). The

initial condition and the no-flux boundary condition are thereby injected naturally. Integration by parts in time shows that admitting nonzero $\phi(T)$ additionally ensures $\langle \tilde{e}_h, \chi\rangle = \langle \rho_h(T), \chi\rangle$ for all $\chi \in \mathsf{C}_h$, that is $\tilde{e}_h$ is determined as the $L_2(D)$-orthogonal projection of $\rho_h(T)$ onto C_h.

Consider $\phi := b_n \otimes (1 \otimes 1)$, where $b_n \in \mathrm{P2}$ is a quadratic (nonzero) bubble on the n-th temporal element that vanishes outside that element. Using this ϕ in (43) yields $\int_J b_n \partial_t \int_D \rho_h = 0$. Since $\partial_t \int_D \rho_h$ is constant, it must equal zero. This implies mass conservation for the discrete density.

3.2. **Discretized ALG2.** We run ALG2 from Section 2.4 on the discrete augmented Lagrangian

$$L_r(\phi_h, \sigma_h, \lambda_h) = \mathcal{F}_h(\phi_h) + \mathcal{G}_h(\sigma_h) + (\Lambda_h\phi_h - \sigma_h, \lambda_h)_{\mathsf{Y}} + \tfrac{r}{2}\|\Lambda_h\phi_h - \sigma_h\|^2_{\mathsf{Y}} \tag{44}$$

obtained from (40).

3.2.1. *Discrete Step A.* The iterate $\phi_h^{(k+1)} \in \mathsf{X}_h$ is defined by the linear variational problem

$$(\Lambda_h\phi_h^{(k+1)}, \Lambda_h\tilde{\phi})_{\mathsf{Y}} = (\sigma_h^{(k)} - \tfrac{1}{r}\lambda_h^{(k)}, \Lambda_h\tilde{\phi})_{\mathsf{Y}} - \tfrac{1}{r}\mathcal{F}_h(\tilde{\phi}) \quad \forall \tilde{\phi} \in \mathsf{X}_h. \tag{45}$$

With $A_h := \Lambda_h'\Lambda_h$, where the adjoint is with respect to the Y scalar product, this can be written as

$$A_h\phi_h^{(k+1)} = b_h^{(k+1)} := \Lambda_h'(\sigma_h^{(k)} - \tfrac{1}{r}\lambda_h^{(k)}) - \tfrac{1}{r}\mathcal{F}_h. \tag{46}$$

3.2.2. *Discrete Step B.* The prior is $\bar{\sigma}_h^{(k+1)} = (\bar{a}_h^{(k+1)}, \bar{b}_h^{(k+1)}, \bar{c}_h^{(k+1)}) := \Lambda_h\phi_h^{(k+1)} + \frac{1}{r}\lambda_h^{(k)}$. The minimization problem (34) is a pointwise minimization in space-time. Even if the data are discrete functions, the minimizer need not lie in the discrete spaces. We first perform the minimization on collocation nodes on each space-time element that are together unisolvent for a piecewise polynomial space. Specifically, we use the 2-node Gauss–Legendre quadrature points on each one-dimensional element (hence 8 nodes per space-time element) to characterize a function in the discrete space $\mathsf{Z}_h := \mathrm{D1} \times \mathrm{D1}^2$. We write $\mathcal{N}(\mathsf{Z}_h) \subset J \times D$ for those collocation nodes. Then we project the result onto the original discrete spaces. The procedure is thus as follows.

(1) For each collocation node $n \in \mathcal{N}(\mathsf{Z}_h)$ let $(a_h(n), b_h(n))$ denote the solution to the pointwise minimization problem

$$\operatorname*{argmin}_{(a,b)\in\mathbb{R}\times\mathbb{R}^d} \left\{A^\star(a + H(b)) + \tfrac{r}{2}|(\bar{a}_h^{(k+1)}(n), \bar{b}_h^{(k+1)}(n)) - (a,b)|^2\right\}. \tag{47}$$

(2) Construct the intermediate $(a_h, b_h) \in \mathsf{Z}_h \times \mathsf{Z}_h^2$ from the values on the collocation nodes.
(3) Project (orthogonally in $L_2 \times [L_2]^d$) the intermediate (a_h, b_h) onto $\mathsf{A}_h \times \mathsf{B}_h$ to obtain the new iterates $(a_h^{(k+1)}, b_h^{(k+1)})$.

Let I_1 denote the operator that constructs a $\mathrm{D1}^2$ function from its values on the spatial collocation nodes $\mathcal{N}(\mathrm{D1}^2) \subset D$. The minimization problem (35) is replaced by

$$c_h^{(k+1)} := \operatorname*{argmin}_{c\in\mathsf{C}_h} \left\{\int_D \mathrm{I}_1\Gamma^\star(c) + \tfrac{r}{2}\|\bar{c}_h^{(k+1)} - c\|^2_{\mathsf{V}}\right\}. \tag{48}$$

3.2.3. *Discrete Step C.* This is the update $\lambda^{(k+1)} := \lambda^{(k)} + r(\Lambda_h\phi_h^{(k+1)} - \sigma_h^{(k+1)})$. Since the range of Λ_h is contained in Y_h by the definition (39), we have $\lambda^{(k+1)} \in \mathsf{Y}_h$ whenever $\lambda^{(k)} \in \mathsf{Y}_h$.

4. Preconditioning the Step A of ALG2

4.1. **Basic preconditioner.** In Section 2.4.1 we introduced the symmetric operator $C : \mathsf{X} \to \mathsf{X}'$,

$$\langle C\phi, \phi\rangle := \|\partial_t\phi\|^2_{J\times D} + \|\phi(T)\|^2_D + \|\nabla\phi\|^2_{J\times D} + \nu^4\|\Delta\phi\|^2_{J\times D}, \tag{33}$$

and argued that it is equivalent to the operator $A := \Lambda'\Lambda$. We discretize this operator to obtain C_h, defined by $C_h\phi_h := (C\phi_h)|_{\mathsf{X}_h}$, and use it as a preconditioner for the discrete operator $A_h = \Lambda_h'\Lambda_h$ in (46)–(51). The inversion of this preconditioner becomes more tractable by passing to a temporal basis of P2 in (36a) that is orthogonal with respect to both scalar products

$$(\theta, \tilde{\theta})_1 := (\theta', \tilde{\theta}')_{L_2(0,T)} + \theta(T)\tilde{\theta}(T) \quad \text{and} \quad (\theta, \tilde{\theta})_0 := (\theta, \tilde{\theta})_{L_2(0,T)}. \tag{49}$$

Such a basis $\{\theta_\omega\}_\omega$ is obtained by solving the generalized eigenvalue problem

$$\text{Find} \quad (\theta_\omega, \omega) \in \mathrm{P2} \times (0,\infty) \quad \text{s.t.} \quad (\theta_\omega, \tilde{\theta})_1 = \omega^2(\theta_\omega, \tilde{\theta})_0 \quad \forall \tilde{\theta} \in \mathrm{P2}. \tag{50}$$

The first pairing is indeed positive definite on $H^1(0,T)$ due to the T-term. In this temporal basis, the preconditioner C_h is block-diagonal and each block is the $\mathrm{B}2^2$ discretization C_h^ω, given by $C_h^\omega v := (B^\omega v)|_{\mathrm{B}2^2}$, of the spatial symmetric positive definite biharmonic operator $C^\omega := \omega^2 - \Delta + \nu^4\Delta^2$ with the homogeneous Neumann boundary conditions (and parameterized by the temporal frequency $\omega > 0$). Thus, applying the inverse of the space-time preconditioner C_h amounts to solving a series of *independent* problems of the form $C_h^\omega u = f$, which can be done in parallel.

4.2. **Multigrid-in-space preconditioner.** Instead of solving $C_h^\omega u = f$ one can replace the inverse of C_h^ω by a multigrid cycle (or another approximation such as the incomplete Cholesky factorization). We follow the geometric multigrid procedure of [16, Section 4.1]. Starting with the finest spatial mesh with $2^{L_x} \times 2^{L_y}$ uniform rectangular elements, the mesh hierarchy is defined by isotropic coarsening the mesh until there is a dimension with at most two elements. As the prolongation we use the natural embedding, the restriction is its adjoint. For the pre- and post-smoother we use the "scaled mass matrix smoother", defined as the preconditioned Richardson iteration $\mathbf{v} \mapsto \mathbf{v} + (\lambda^\omega_{\max}\mathbf{M})^{-1}(\mathbf{f} - \mathbf{C}^\omega\mathbf{v})$, where $\mathbf{M}$ is the mass matrix, $\mathbf{C}^\omega = \omega^2\mathbf{M} + \mathbf{A} + \nu^4\mathbf{B}$ is the discretization of C_h^ω on the current level and for the given temporal frequency ω, and $\lambda^\omega_{\max}$ is the maximal eigenvalue (precomputed numerically) of the generalized eigenvalue problem $\mathbf{C}^\omega\mathbf{v} = \lambda^\omega_{\max}\mathbf{M}\mathbf{v}$. Note that the choice of the basis is irrelevant here. For the computation of $\lambda^\omega_{\max}$, the observation $\lambda^\omega_{\max} = \lambda^0_{\max} + \omega$ is useful.

The contraction factor of this multigrid is robust in the parameters ν and ω, as well as in the mesh width: see Figure 1.

The multigrid method (with semi-coarsening in space) as proposed in [2] for preconditioning linearized mean field games systems did not seem to be robust in ν.

5. Numerical examples

5.1. **Implementation.** We give here some details on the implementation.

As the starting values for the ALG2 iteration we use the zero vector.

The solution to (46) is approximated by the preconditioned conjugate gradient method,

$$\phi_h^{(k+1)} := \phi_h^{(k)} + \mathrm{PCG}[A_h, (b_h^{(k+1)} - A_h\phi_h^{(k)}), \epsilon_{\mathrm{pcg}}, \mathrm{maxit}_{\mathrm{pcg}}, C_h^{-1}], \tag{51}$$

where C_h^{-1} is the approximation of the inverse of the preconditioner as described in Section 4.1 (basic preconditioner) or Section 4.2 (multigrid preconditioner). The PCG iteration is initialized with the zero vector. We use the relative residual tolerance $\epsilon_{\mathrm{pcg}} = 10^{-1}$, meaning that the iteration terminates once the residual in the C_h^{-1} norm (defined by $\|r\|^2 := \langle C_h^{-1}r, r\rangle$) is reduced by the factor ϵ_{pcg}. The maximal iteration number is set to $\mathrm{maxit}_{\mathrm{pcg}} = 100$, but it is never attained in our computations.

The discrete spatial biharmonic problems in the basic preconditioner (§4.1) are solved in Matlab with the backslash operator.

The eigenvalue $\lambda^\omega_{\max}$ in the scaled mass matrix smoother in Section 4.2 is obtained from $\lambda_{\max} = \lambda^0_{\max} + \omega$, where $\lambda^0_{\max}$ is approximated using the Matlab `eigs` routine on $\mathbf{M}^{-1}\mathbf{C}^0$ with flag `'LM'`, the options `opt.isreal = 1`, `opt.issym = 0`, and with the default tolerance (the matrices are assembled in the B-spline basis with a modification at the boundary ensuring the homogeneous Neumann boundary condition). Unless specified otherwise, we use the W cycle with $n_{\mathrm{smooth}} = 5$ pre- and post-smoothing steps on each level.

The minimizer of (47) is approximated by a Newton iteration on the derivative of the argument with 20 iterations initializing with the values of the prior. The use of the Newton iteration (as opposed to the implementation of the proximal operator as in [8, Appendix]) is justified because the second derivative of the functional is still continuous in our examples. This allows to vectorize the Newton iteration in our Matlab implementation, so that Step B takes significantly less time than Step A.

The minimization problem (48) is solved approximately by a trust-region algorithm with Hessian implemented in the Matlab optimization toolbox routine `fminunc`, and the result is used as the

starting value for the same procedure until the relative improvement in the $\|\cdot\|_{\mathsf{V}}$ norm is less than 10^{-10}.

5.2. **Example 1.** In this example we consider the interval $D = (-2, 2)$. The initial density $\rho_0 = \mathbb{1}_{(-2,-1)}$ is the indicator function of $(-2, -1) \subset D$. The terminal cost is $\Gamma(\rho) = 10^3 \times \frac{1}{2}(\rho - \rho_T)^2$, where $\rho_T = \mathbb{1}_{(1,2)}$ is the target terminal state. The running cost is $A(\rho) = \frac{1}{2}\rho^2$. The Hamiltonian is $H(t, x, p) = \frac{1}{2}|p|^2 - 10^3 \times \mathbb{1}_{t\le 3/4}\mathbb{1}_{|x|\le 1/2}$, making the movement in the "no-go area" delimited by $|x| \le 1/2$ relatively costly as long as $t \le 3/4$. This leads to the formation of a strong peak of the density $\rho(t, x)$ for $t \nearrow 3/4$ and $x \nearrow -1/2$. In order to resolve this behavior, we geometrically refine the initial $2^8 \times 2^8$ equidistant mesh around $t = 3/4$ by halving the two temporal intervals adjacent to $t = 3/4$ ten times; the same is done for $t = 0$ and for the spatial mesh around $x = -1/2$. The resulting mesh has $(2^8 + 30) \times (2^8 + 20) = 78'936$ space-time elements. We use the basic preconditioner from Section 4.1. We perform $1'000$ ALG2 iterations. The diffusion coefficient is first set to $\nu = 10^{-1}$. The resulting density ρ behaves as expected: the initial density ρ_0 and the targeted terminal density ρ_T are well-captured, and the "no-go area" $\{t \le 3/4\} \times \{|x| \le 1/2\}$ remains almost mass-free. Now we change the diffusion to $\nu = 1$. The density is now smoothed out in space, in particular eliminating the peak at $t = 3/4$.

5.3. **Example 2.** Here $D = (-2, 2) \times (-1/2, 1/2)$. The initial density is $\rho_0 = \mathbb{1}_{x\le -1}$ in the left part of the domain. The terminal cost is $\Gamma(\rho) := 10^3 \times \frac{1}{2}(\rho - \rho_T)^2$, where $\rho_T := \mathbb{1}_{x\ge 1}$ is the target terminal state in the right part of the domain. The running cost is $A(\rho) = \frac{1}{2}\rho^2$. The Hamiltonian is $H(t, x, p) = \frac{1}{2}|p|^2 - 10^3 \times \mathbb{1}_{\square}$, restricting the movement in the rectangular area $\square := \{|x| \le 1/2\} \times \{|y| \le 1/4\}$. The computational mesh has $2^5 \times (2^6 \times 2^5)$ space-time elements. We use the basic preconditioner from Section 4.1. We perform $1'000$ ALG2 iterations. Temporal snapshots of the density for $\nu = 10^{-1}$ and $\nu = 1$ are shown in Figure 3.

5.4. **Empirical convergence study.** We empirically investigate the convergence of the ALG2 on the example from Section 5.3. As the reference solution we take the discrete solution computed on the finer mesh with $2^6 \times (2^7 \times 2^6)$ space-time elements with $5'000$ ALG2 iterations and the multigrid preconditioner (§4.2). In Figure 4 we show the error of the discrete density $\rho_h^{(k)}$ and the discrete cost $\phi_h^{(k)}$ as the ALG2 iteration progresses, varying the spatial resolution (keeping the temporal resolution at 2^5 elements), the PCG tolerance ϵ_{pcg} in (51), and the diffusion coefficient ν. The preconditioner used is the basic preconditioner (§4.1) but the multigrid preconditioner produces (§4.2) essentially the same results. The convergence rate is clearly mesh dependent, indicating nonuniform convexity (with respect to the discretization) of the functional to be minimized, see the discussion in Section 2.3. Moreover, for the larger diffusion coefficient $\nu = 1$ we observe a somewhat nonmonotonic convergence and $1'000$ ALG2 iterations seem insufficient; this indicates that $5'000$ ALG2 iterations for the reference solution were also not quite enough. The average number of PCG iterations is reported in Table 1.

The application of the preconditioner typically consumes the bulk of the time. The basic preconditioner (with backslash for the solution of the bi-Laplace equations) takes around 7.5s on the $2^5 \times (2^6 \times 2^5)$ mesh and around 375s on the $2^5 \times (2^7 \times 2^6)$ mesh; the multigrid preconditioner 0.9s and 8s, respectively.

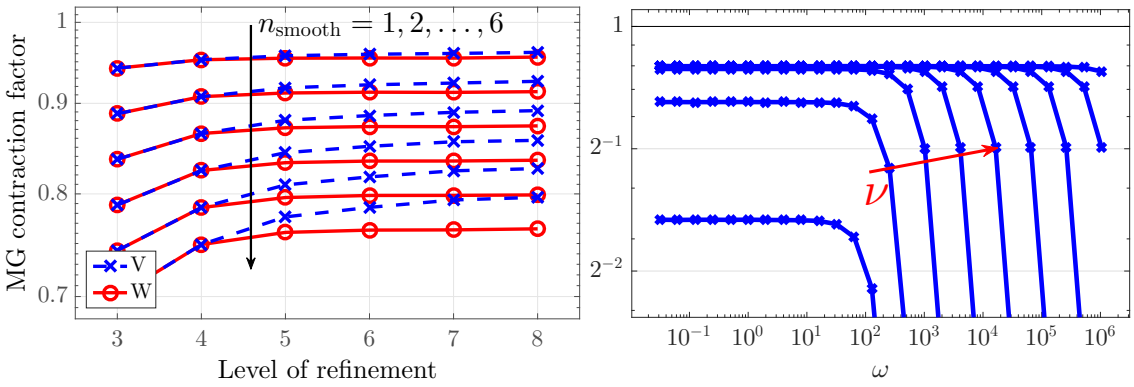

FIGURE 1. Contraction factor $\rho(\mathrm{Id} - \mathrm{MG} \circ \mathrm{OP})$ of the multigrid from Section 4.2 for $\mathrm{OP} := \omega^2 - \Delta + \nu^4\Delta^2$ on a uniform $2^L \times 2^L$ mesh. Left: as a function of the refinement level L and for varying number n_{smooth} of pre- and post-smoothing iterations (V and W cycle), computed as the maximum over $\nu = 2^{-5}, \ldots, 2^3$ and $\omega = 2^{-5}, \ldots, 2^{20}$. Right: as a function of ω for the W cycle with $n_{\text{smooth}} = 5$ starting on $L = 8$ and for varying $\nu = 2^{-5}, \ldots, 2^3$.

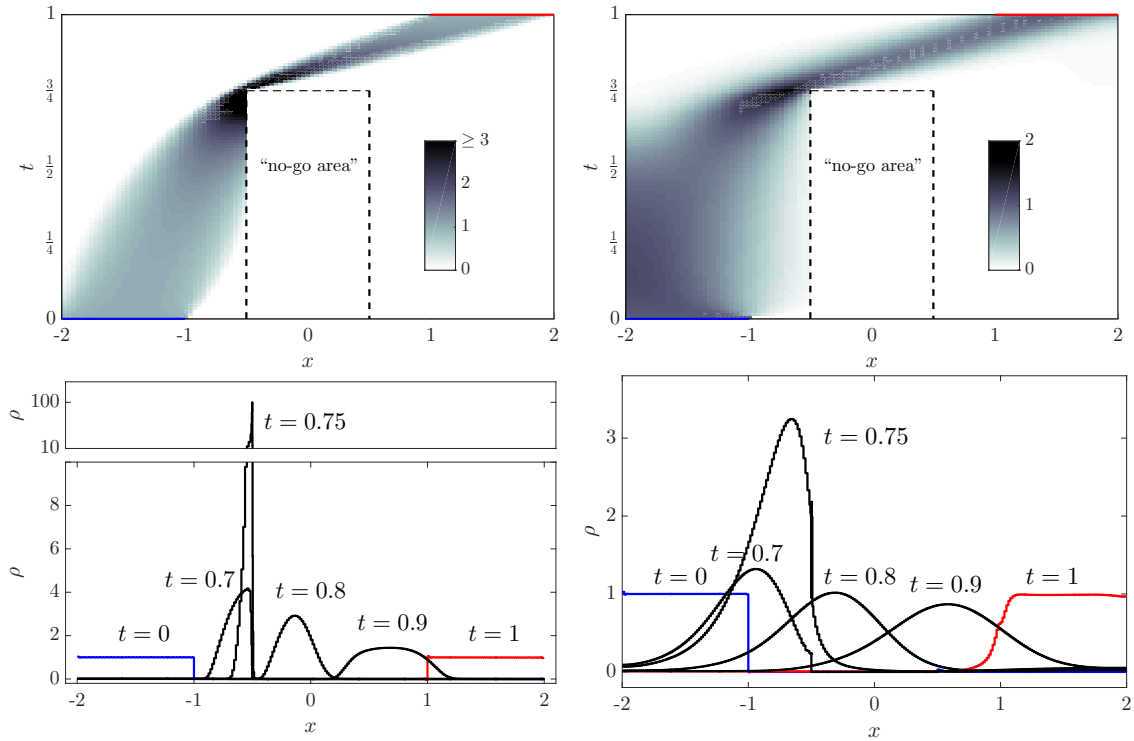

FIGURE 2. The example from Section 5.2. Density ρ in space-time (top) and its temporal snapshots (bottom) with diffusion coefficient $\nu = 10^{-1}$ (left) and $\nu = 1$ (right).

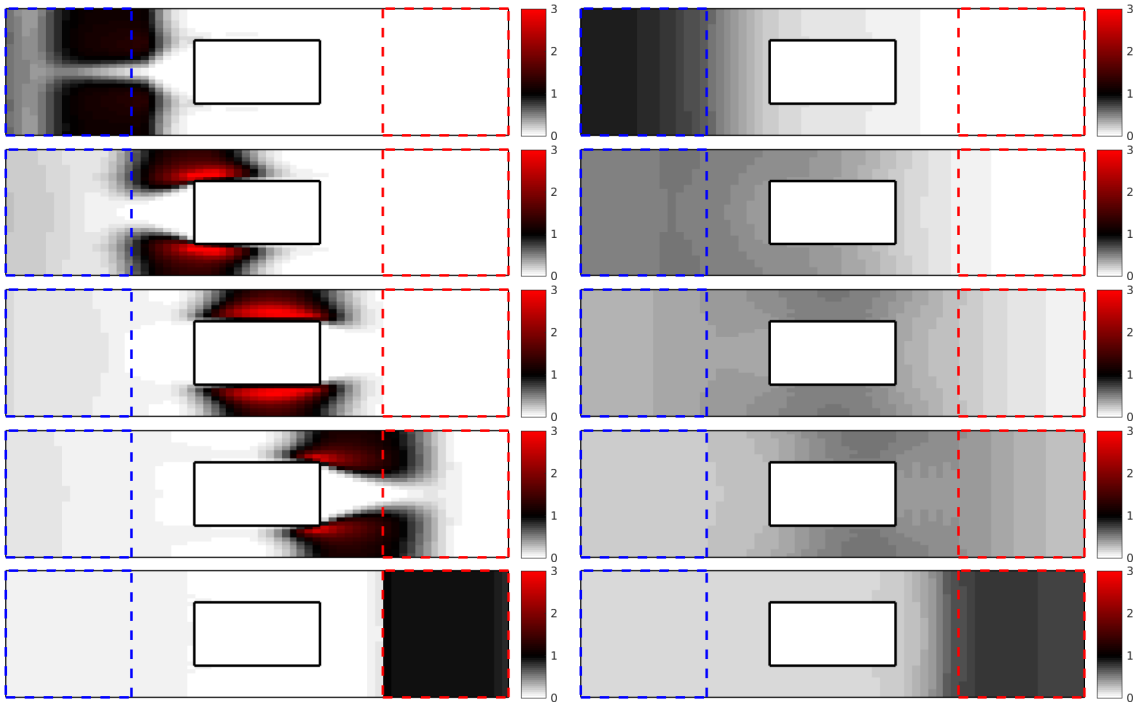

FIGURE 3. The example from Section 5.3. Temporal snapshots $\rho_h(t,\cdot)$ of the computed density at $t = n/13$ for $n = 1, 4, 7, 10, 13$ (top to bottom). Diffusion coefficient $\nu = 10^{-1}$ (left) and $\nu = 1$ (right).

$\nu = 10^{-1}$	Basic, $\epsilon_{\text{pcg}} =$		MG, $\epsilon_{\text{pcg}} =$	
(L_x, L_y)	10^{-1}	10^{-2}	10^{-1}	10^{-2}
$(4,3)$	3.0	5.0	3.0	5.4
$(5,4)$	3.0	5.3	3.0	6.9
$(6,5)$	3.0	6.7	3.0	7.0

$\nu = 1$	Basic, $\epsilon_{\text{pcg}} =$		MG, $\epsilon_{\text{pcg}} =$	
(L_x, L_y)	10^{-1}	10^{-2}	10^{-1}	10^{-2}
$(4,3)$	3.0	4.0	4.6	9.8
$(5,4)$	3.0	4.7	3.7	9.7
$(6,5)$	3.0	4.7	3.0	7.2

TABLE 1. Average number of PCG iterations over the first 500 ALG2 steps in the convergence study in Section 5.4. Left/Right: $\nu = 0.1$ / $\nu = 1$. For $\nu = 1$ with the MG preconditioner, the number of PCG iterations tends to increase in the course of the ALG2 iteration.

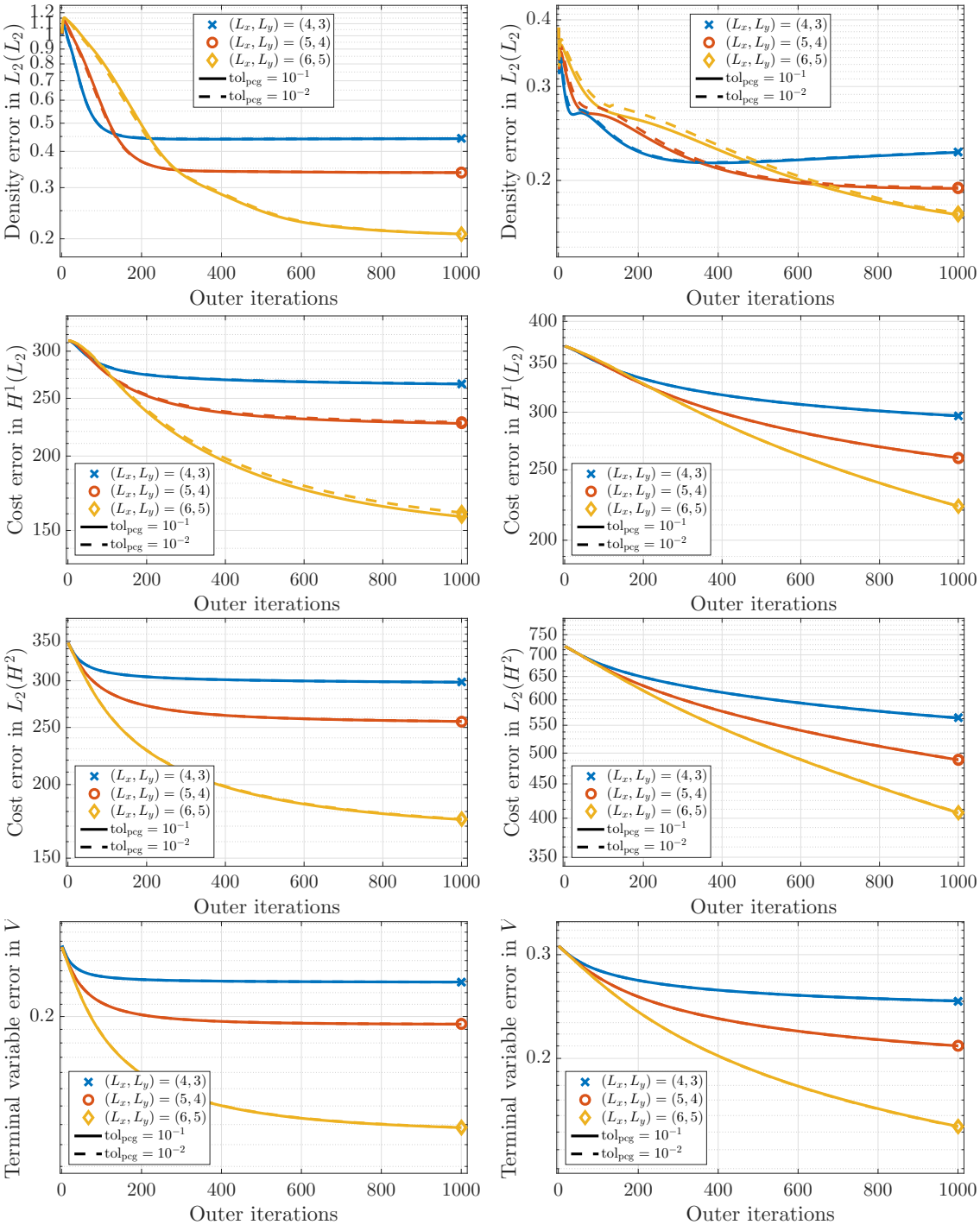

FIGURE 4. Convergence of the discrete ALG2 iteration on the example from Section 5.3 with the basic preconditioner (§4.1) for different spatial resolutions and PCG tolerances in (51). Top to bottom: the discrete density $\rho_h^{(k)}$ in $L_2(J; L_2(D))$, the discrete cost function $\rho_h^{(k)}$ in $H^1(J; L_2(D))$ and in $L_2(J; H^2(D))$, and the terminal variable $e_h^{(k)}$ in V. Left/Right: diffusion coefficient $\nu = 10^{-1}$ and $\nu = 1$. The multigrid precondtioner (§4.2) leads to very similar results.

References

[1] Yves Achdou and Italo Capuzzo-Dolcetta. Mean field games: numerical methods. *SIAM J. Numer. Anal.*, 48(3):1136–1162, 2010. 1

[2] Yves Achdou and Victor Perez. Iterative strategies for solving linearized discrete mean field games systems. *Netw. Heterog. Media*, 7(2):197–217, 2012. 1, 8

[3] Yves Achdou and Alessio Porretta. Convergence of a Finite Difference Scheme to Weak Solutions of the System of Partial Differential Equations Arising in Mean Field Games. *SIAM J. Numer. Anal.*, 54(1):161–186, 2016. 1, 4

[4] Roman Andreev. Wavelet-in-time multigrid-in-space preconditioning of parabolic evolution equations. *SIAM J. Sci. Comput.*, 38(1):A216–A242, 2016. 5

[5] Roman Andreev and Julia Schweitzer. Conditional space-time stability of collocation Runge–Kutta for parabolic evolution equations. *Electron. Trans. Numer. Anal.*, 41:62–80, 2014. 6

[6] Wolfgang Arendt and Ralph Chill. Global existence for quasilinear diffusion equations in isotropic nondivergence form. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 9(3):523–539, 2010. 4

[7] Ivo Babuška and Tadeusz Janik. The h-p version of the finite element method for parabolic equations. II. The h-p version in time. *Numer. Meth. Part. D. E.*, 6:343–369, 1990. 6

[8] Jean-David Benamou and Guillaume Carlier. Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *J Optimiz Theory App*, pages 1–26, 2015. 1, 2, 3, 4, 5, 8

[9] Jean-David Benamou, Guillaume Carlier, and Filippo Santambrogio. Variational mean field games, 2016. To be published in a special volume on "active particles". 1

[10] Fabio Camilli and Francisco Silva. A semi-discrete approximation for a first order mean field game problem. *Netw. Heterog. Media*, 7(2):263–277, 2012. 1

[11] Pierre Cardaliaguet, P. Jameson Graber, Alessio Porretta, and Daniela Tonon. Second order mean field games with degenerate diffusion and local coupling. *NoDEA Nonlinear Differential Equations Appl.*, 22(5):1287–1317, 2015. 1, 3, 4

[12] Elisabetta Carlini and Francisco J. Silva. A fully discrete semi-Lagrangian scheme for a first order mean field game problem. *SIAM J. Numer. Anal.*, 52(1):45–67, 2014. 1

[13] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55(3, Ser. A):293–318, 1992. 5

[14] Ivar Ekeland and Roger Témam. *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999. 3

[15] Michel Fortin and Roland Glowinski. *Augmented Lagrangian methods*, volume 15 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1983. 1, 4

[16] Wolfgang Hackbusch. *Multigrid methods and applications*, volume 4. Springer-Verlag, Berlin, 1985. Second printing 2003. 8

[17] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris*, 343(10):679–684, 2006. 1

[18] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007. 4

(R. Andreev) Université Paris Diderot, Sorbonne Paris Cité, LJLL (UMR 7598 CNRS), F-75205 Paris, France

E-mail address: roman.andreev@upmc.fr