

# Numerical methods for mean field games based on Gaussian processes and Fourier features

Chenchen Mou<sup>a</sup>, Xianjin Yang<sup>b,c,\*</sup>, Chao Zhou<sup>d</sup>

<sup>a</sup> Department of Mathematics, City University of Hong Kong, Hong Kong, China

<sup>b</sup> Yau Mathematical Sciences Center, Tsinghua University, Haidian District, Beijing, 100084, China

<sup>c</sup> Beijing Institute of Mathematical Sciences and Applications, Huairou District, Beijing, 101408, China

<sup>d</sup> Department of Mathematics and Risk Management Institute, National University of Singapore, Singapore

## ARTICLE INFO

### Article history:

Received 12 December 2021

Received in revised form 10 March 2022

Accepted 28 March 2022

Available online 1 April 2022

### Keywords:

Mean field games

Gaussian process

Fourier features

## ABSTRACT

In this article, we propose two numerical methods, the Gaussian Process (GP) method and the Fourier Features (FF) algorithm, to solve mean field games (MFGs). The GP algorithm approximates the solution of a MFG with maximum *a posteriori* probability estimators of GPs conditioned on the partial differential equation (PDE) system of the MFG at a finite number of sample points. The main bottleneck of the GP method is to compute the inverse of a square gram matrix, whose size is proportional to the number of sample points. To improve the performance, we introduce the FF method, whose insight comes from the recent trend of approximating positive definite kernels with random Fourier features. The FF algorithm seeks approximated solutions in the space generated by sampled Fourier features. In the FF method, the size of the matrix to be inverted depends only on the number of Fourier features selected, which is much less than the size of sample points. Hence, the FF method reduces the precomputation time, saves the memory, and achieves comparable accuracy to the GP method. We give the existence and the convergence proofs for both algorithms. The convergence argument of the GP method does not depend on any monotonicity condition, which suggests the potential applications of the GP method to solve MFGs with non-monotone couplings in future work. We show the efficacy of our algorithms through experiments on a stationary MFG with a non-local coupling and on a time-dependent planning problem. We believe that the FF method can also serve as an alternative algorithm to solve general PDEs.

© 2022 Elsevier Inc. All rights reserved.

## 1. Introduction

Mean field games (MFGs) [25–31] study the behavior of a large population of rational and indistinguishable agents as the number of agents goes to infinity. The Nash equilibrium of a typical MFG is formulated by a coupled system of two partial differential equations (PDEs), a Hamilton–Jacobi–Bellman (HJB) equation and a Fokker–Plank (FP) equation. The HJB equation gives the value function of agents and the FP equation determines agents' distribution. Recently, MFG models have found widespread applications, see [18–21,24,33,34]. Up to now, the well-posedness of MFGs is well understood in various settings and the first results date back to the original works of Lasry and Lions and have been given in the course of Lions

\* Corresponding author at: Yau Mathematical Sciences Center, Tsinghua University, Haidian District, Beijing, 100084, China.

E-mail addresses: [chencmou@cityu.edu.hk](mailto:chencmou@cityu.edu.hk) (C. Mou), [yxjmath@gmail.com](mailto:yxjmath@gmail.com) (X. Yang), [matzc@nus.edu.sg](mailto:matzc@nus.edu.sg) (C. Zhou).

at Collège de France (see [36]). However, very few MFG models admit explicit solutions. Hence, numerical computations of MFGs play an essential role in obtaining quantitative descriptions of underlying models.

Here, we propose two new algorithms to solve MFGs. The first one, called the Gaussian Process (GP) method, applies the algorithm in [16] to solve MFGs. The authors in [16] give a GP regression framework to solve nonlinear PDEs. The solution to a PDE is found by solving an optimal recovery problem, whose minimizer is viewed as a maximum *a posteriori* probability (MAP) estimator of a GP conditioned on the PDE evaluated at sample points. The main bottleneck of the GP method lies in the computation of the inverse of a square gram matrix, whose size is proportional to the product of the size of sample points times the number of linear operators in the PDE. To improve the performance, we propose the Fourier Features (FF) method, where the optimal recovery problem seeks minimizers in the space generated by sampled trigonometric functions. The FF method is based on the recent trend of approximating positive definite kernels with randomized trigonometric functions in Gaussian regressions, see [43,46,48]. Using the technique we propose in Remark 2.10, the dimension of the matrix needed to be inverted in the FF algorithm depends only on the number of sampled Fourier features, which is less than the size of sample points. The numerical experiments show that the FF algorithm reduces the precomputation time and the amount of storage with comparable accuracy compared to the GP method.

Meanwhile, we also prove the convergence of our algorithms. The proofs for the GP method are based on the compactness arguments in [16] and do not depend on any monotonicity condition that guarantees stability and uniqueness of the solution to a typical MFG. This feature implies the potential applications of the GP method to MFGs with non-monotone couplings. On the other hand, since the Fourier features space the FF method uses lacks compactness, the same arguments of the GP method cannot be adapted to the setting of the FF method. Instead, the Lasry–Lions monotonicity arguments provide the uniform bounds for the errors of numerical solutions and lead to the convergence of the FF algorithm. In future work, we plan to investigate the convergence of the FF method under the setting of MFGs without monotone couplings.

### 1.1. Related works

By now, there have been various numerical methods for MFGs. Here, we briefly summarize the numerical algorithms solving MFGs that are closely related to the methods proposed in this paper. For clarification, we group them into different categories. The first group consists of mesh-based algorithms:

- Finite difference methods. As far as we know, the first finite difference method for MFGs is introduced in [2] for both stationary and time-dependent MFGs. The Lax-Friedrichs or Godunov type schemes are introduced to approximate Hamiltonians. The Fokker–Plank equation is discretized in such a way that it preserves the adjoint structure in the MFG. The authors proved the existence and uniqueness of the discretized algebraic systems. One can solve the discretized equations using iterative methods [5]. To know more about details and applications of the finite difference method, we refer readers to [1,3,4,32].
- Optimization algorithms. In the seminal paper [31], Lasry and Lions explain the concept of a variational MFG, which can be interpreted as the optimality condition of a PDE-driven optimization problem. From this point of view, several techniques from Optimization have been introduced to solve the PDE constrained minimization problem. In [7], the authors propose augmented Lagrangian methods. Later, the performances of different algorithms are compared in [9,10], and the Chambolle–Pock algorithm outperforms other methods.
- Monotone flows. In [6], the authors have proposed the monotone flow method to solve stationary MFGs. Later, in [22], the authors apply the monotone flow to solve stationary and time-dependent MFGs with finite states. The insight is that solving a MFG with monotonicity is equivalent to computing a zero of a monotone operator, which is the stationary point of the monotone flow. However, the monotone flow may not guarantee the non-negativity of the probability density in the flow. In [23], the authors design the Hessian Riemannian flow method to preserve positivity.

Despite the rigorous theories behind the algorithms mentioned above, they are mesh-based methods and prone to the curse of dimensionality. The following methods are mesh-free methods developed recently to keep pace with growing problem sizes:

- Lagrangian Methods. As far as we know, the first Lagrangian method for MFGs appears in [40]. The authors use Fourier expansions of the non-local coupling term in the HJB equation and parameterize the MFG. The parameterized MFG is reformulated as the optimality condition of a convex optimization problem over a finite-dimensional subspace of continuous curves, which is independent of the dimension. Based on this work, the authors in [37,38] further develop this idea to solve non-potential MFGs with mixed non-local couplings. Even though the methods proposed in our paper use kernels and Fourier features, we work in a different way and approximate the solution of a MFG in reproducing kernel Hilbert spaces (RKHSs) or Fourier feature spaces.
- Neural Networks. Since neural networks are compositions of nonlinear maps and have the power to represent sufficient complicated functions, the authors in [12,13] use neural networks to approximate unknowns in PDEs and solve ergodic MFGs and time-dependent MFGs. At the same time, in [45], the authors use the Lagrangian Method in [40] to reformulate variational MFGs and solve the resulting system using methods from neural networks. For more applications of neural networks in MFGs, see [14,35].

We refer readers to the recent comprehensive surveys [4,32] for more details and applications of numerical methods for MFGs.

Compared to the above-mentioned algorithms, our methods for MFGs admit the following features:

1. The GP and the FF algorithms are meshfree and flexible to the shape of domains, compared to algorithms in the first group mentioned above. Especially, when we solve MFGs in the whole Euclidean space, to approximate derivatives, mesh-based algorithms have to impose artificial conditions on the boundary of the domain we choose to work on. Our methods parameterize the values of derivatives evaluated at sample points and solve reformulated finite-dimensional minimization problems as in (2.9) and (2.32). Hence, we do not impose extra conditions to deal with derivatives near the boundary.
2. Compared to the neural network methods, our algorithms base on theories of RKHSs and Fourier series, for which the math backgrounds are well understood. On the other hand, our algorithms are equivalent to the neural network methods in the following perspective: a neural network with a single inner layer, with the activation functions being feature maps parameterized at sample points or being random sampled trigonometric functions, and with a linear output layer can be viewed as a function in RKHSs or the Fourier features space, and vice versa.
3. The choice of the kernel for the GP method and the selection of Fourier features have a profound impact on the convergence and the accuracy of our approximations. We leave the study of hyperparameter learning (see [17,42]) to future work.
4. The convergence of the GP method does not depend on the Lasry–Lions monotonicity condition of the coupling terms in MFGs, which is required by the monotone flow [6,22,23] and the Lagrangian methods [37,38,40]. Hence, we plan to study the application of the GP method to solve MFGs with displacement monotonicity [39] or non-monotone couplings in future work.
5. In general, it is less costly to compute a linear map of a Fourier feature than to calculate the same linear transformation of a kernel function. For instance, in Subsection 4.1, we need to parametrize the linear operator  $L = (1 - \Delta)^{-1}(1 - \Delta)^{-1}$  at sample points, where  $\Delta$  is a Laplacian operator. In the GP method, the representer theorem gives expressions involving the computation of  $LK_2(x, \cdot)$  for  $x \in \mathbb{T}^2$ , where  $K_2$  is the kernel of the RKHS we choose to find an approximation for the probability density. Since  $LK_2(x, \cdot)$  does not admit an explicit formula, one can use the fast Fourier transform to compute it numerically. On the other hand, for any  $\omega \in \mathbb{N}^2$ , we observe that  $L \sin(\omega^T x) = \sin(\omega^T x)/(1 + |\omega|^2)^2$  and  $L \cos(\omega^T x) = \cos(\omega^T x)/(1 + |\omega|^2)^2$  for  $x \in \mathbb{T}^2$ . Hence, if we choose trigonometric functions as features in the FF method, the action of  $L$  on a Fourier feature admits an explicit formula and is easier to compute.
6. The main bottleneck of the GP method is to compute the Cholesky decomposition of the gram matrix, whose size increases as the number of sample points grows. The FF method relieves the pressure by approximating functions in the FF space and by using the technique in Remark 2.10. However, as the dimension of the problem at hand increases, we should also enlarge the FF space in the absence of information about the solution. Hence, a clever selection of features is required to apply the FF method in large dimensions. In this paper, we choose the Fourier series in the periodic settings and use random Fourier features for the non-periodic cases. We leave the study of other selections to future work.

## 1.2. Outline

This article is organized as follows. In Section 2, we present our algorithms by solving a one-dimensional MFG which admits an explicit smooth solution. We also give the existence and convergence analysis of our methods. A simple numerical experiment follows and verifies the efficacy of our methods. Section 3 presents the general frameworks of the GP method and the FF algorithm. We also show the existence and convergence for the GP and the FF methods in general settings. The proofs related to the GP method are shown in Section 3 and those of the FF method are presented in Appendix A. Numerical experiments follow in Section 4. Conclusions and future work appear in Section 5.

**Notations.** For a vector  $v$  with real values, we denote by  $|v|$  the Euclidean norm of  $v$  and by  $v^T$  the transpose of  $v$ . We represent the inner product of two real valued vectors  $u$  and  $v$  by  $\langle u, v \rangle$  or  $u^T v$ . Let  $\mathbb{Z}^d$  be the space of  $d$ -dimensional vectors with integer elements. Given  $N \in \mathbb{N}$ , we define  $\mathbb{Z}_N^d$  as the set  $\{i|i \in \mathbb{Z}^d, 0 < |i| \leq N\}$ . Let  $\Omega$  be a subset of  $\mathbb{R}^d$ , let  $\text{int } \Omega$  be the interior of  $\Omega$ , and let  $\partial\Omega$  be the boundary of  $\Omega$ . We denote by  $L^2(\Omega)$  the space of square-integrable functions on  $\Omega \subset \mathbb{R}^d$ . Meanwhile, let  $H^1(\Omega)$  be the space such that  $\forall f \in H^1(\Omega)$ ,  $f \in L^2(\Omega)$  and  $\nabla f \in (L^2(\Omega))^d$ . Furthermore, we represent the space of functions with finite sup norm by  $L^\infty(\Omega)$ . For a normed vector space  $V$ , we denote by  $\|\cdot\|_V$  the norm of  $V$ . Let  $\mathcal{U}$  be a Banach Space endowed with a quadratic norm  $\|\cdot\|_{\mathcal{U}}$ . We denote by  $\mathcal{U}^*$  the dual of  $\mathcal{U}$  and by  $[\cdot, \cdot]$  the duality pairing. We assume that there exists a covariance operator  $\mathcal{K}_{\mathcal{U}} : \mathcal{U}^* \mapsto \mathcal{U}$ , which is linear, bijective, symmetric ( $[\mathcal{K}_{\mathcal{U}}\phi, \psi] = [\mathcal{K}_{\mathcal{U}}\psi, \phi]$ ), and positive ( $[\mathcal{K}_{\mathcal{U}}\phi, \phi] > 0$  for  $\phi \neq 0$ ), such that

$$\|u\|_{\mathcal{U}}^2 = [\mathcal{K}^{-1}u, u], \forall u \in \mathcal{U}.$$

Let  $\phi_1, \dots, \phi_P$  be  $P \in \mathbb{N}$  elements of  $\mathcal{U}^*$  and let  $\phi := (\phi_1, \dots, \phi_P)$  be an element of the product space  $(\mathcal{U}^*)^{\otimes P}$ . Then, for  $u \in \mathcal{U}$ , the pairing  $[\phi, u]$  is denoted by

$$[\phi, u] := ([\phi_1, u], \dots, [\phi_P, u]).$$

Furthermore, for  $\mathbf{u} := (u_1, \dots, u_S) \in \mathcal{U}^{\otimes S}$ ,  $S \in \mathbb{N}$ , we represent by  $[\phi, \mathbf{u}] \in \mathbb{R}^{P \times S}$  the matrix with entries  $[\phi_i, u_j]$ . Finally, we denote by  $C$  a positive real number whose value may change line by line.

## 2. An appetizer: solving a stationary MFG with an explicit solution

To show our algorithms, we apply our methods to solve the one-dimensional stationary MFG in [6], which admits an explicit solution. Let  $\mathbb{T}$  be the one-dimensional torus and be characterized by  $[0, 1)$ . Given smooth functions  $V : \mathbb{T} \mapsto \mathbb{R}$  and  $b : \mathbb{T} \mapsto \mathbb{R}$ , we find  $(u, m, \bar{H}) \in C^\infty(\mathbb{T}) \times C^\infty(\mathbb{T}) \times \mathbb{R}$  solving

$$\begin{cases} \frac{u_x^2}{2} + V(x) + b(x)u_x = \ln m + \bar{H}, & \text{on } \mathbb{T}, \\ -(m(u_x + b(x)))_x = 0, & \text{on } \mathbb{T}, \\ \int_{\mathbb{T}} m \, dx = 1, \int_{\mathbb{T}} u \, dx = 0, \end{cases} \quad (2.1)$$

where  $u$  is the value function,  $m$  presents the probability density of the population, and  $\bar{H}$  is a real number. According to [6], when  $\int_{\mathbb{T}} b(x) \, dx = 0$ , (2.1) admits a unique smooth solution, which has the following explicit formulas

$$\begin{cases} u(x) = -\int_0^y b(y) \, dy + \int_{\mathbb{T}} \int_0^x b(y) \, dy \, dx, \\ m(x) = \frac{e^{V(x) - \frac{b^2(x)}{2}}}{\int_{\mathbb{T}} e^{V(y) - \frac{b^2(y)}{2}} \, dy}, \\ \bar{H} = \ln \left( \int_{\mathbb{T}} e^{V(y) - \frac{b^2(y)}{2}} \, dy \right). \end{cases} \quad (2.2)$$

For ease of presentation, we denote

$$H(x, p) := V(x) + \frac{p^2}{2} + b(x)p, \quad \forall x \in \mathbb{T}, p \in \mathbb{R}. \quad (2.3)$$

### 2.1. The Gaussian process method

Here, we use the GP method in [16] to solve (2.1). First, we sample a collection of  $M$  points  $\{x_i\}_{i=1}^M$  in  $[0, 1)$ . Let  $\mathcal{U}$  and  $\mathcal{V}$  be the RKHSs associated with positive definite kernels  $K_1$  and  $K_2$ , respectively. We denote by  $\|\cdot\|_{\mathcal{U}}$  the norm of  $\mathcal{U}$  and represent the norm of  $\mathcal{V}$  by  $\|\cdot\|_{\mathcal{V}}$ . Following [16], we approximate the solution  $(u^*, m^*, \bar{H}^*)$  of (2.1) by a minimizer of the following optimal recovery problem

$$\begin{cases} \min_{(u, m, \bar{H}) \in \mathcal{U} \times \mathcal{V} \times \mathbb{R}} \|u\|_{\mathcal{U}}^2 + \|m\|_{\mathcal{V}}^2 + |\bar{H}|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M m(x_i) - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M u(x_i) \right|^2 \\ \text{s.t. } H(x_i, u_x(x_i)) = \ln m(x_i) + \bar{H}, \quad \forall i = 1, \dots, M, \\ m_x(x_i)(u_x(x_i) + b(x_i)) + m(x_i)(u_{xx}(x_i) + b_x(x_i)) = 0, \quad \forall i = 1, \dots, M, \end{cases} \quad (2.4)$$

where  $\beta > 0$  is a penalization parameter. We assume that the kernels  $K_1$  and  $K_2$  are properly chosen such that  $\mathcal{U} \subset C^2(\mathbb{T})$  and  $\mathcal{V} \subset C^1(\mathbb{T})$ . Hence, the constraints of (2.4) are well-defined. Let  $(u^\dagger, m^\dagger, \bar{H}^\dagger)$  be a minimizer to (2.4), whose existence is given by Theorem 2.3 later. Then,  $u^\dagger$  and  $m^\dagger$  can be viewed as MAP estimators of two GPs conditioned on the MFG at the sample points  $\{x_i\}_{i=1}^M$  [16]. We note that the conditioned GPs are not Gaussian since the constraints in (2.4) are nonlinear.

The key idea of the GP method is to characterize the minimizer of (2.4) via a finite-dimensional representation formula using the representer theorem [41, Sec. 17.8]. Following [16], we rewrite (2.4) as a two-level optimization problem

$$\begin{cases} \min_{z \in \mathbb{R}^{3M}, \rho \in \mathbb{R}^{2M}, \lambda \in \mathbb{R}} \begin{cases} \min_{(u, m, \bar{H}) \in \mathcal{U} \times \mathcal{V} \times \mathbb{R}} \|u\|_{\mathcal{U}}^2 + \|m\|_{\mathcal{V}}^2 + |\bar{H}|^2 \\ \text{s.t. } u(x_i) = z_i^{(1)}, u_x(x_i) = z_i^{(2)}, u_{xx}(x_i) = z_i^{(3)}, \quad \forall i = 1, \dots, M, \\ m(x_i) = \rho_i^{(1)}, m_x(x_i) = \rho_i^{(2)}, \bar{H} = \lambda, \quad \forall i = 1, \dots, M, \\ + \beta \left| \frac{1}{M} \sum_{i=1}^M \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M z_i^{(1)} \right|^2 \end{cases} \\ \text{s.t. } H(x, z_i^{(2)}) = \ln \rho_i^{(1)} + \lambda, \quad \forall i = 1, \dots, M, \\ \rho_i^{(2)}(z_i^{(2)} + b(x_i)) + \rho_i^{(1)}(z_i^{(3)} + b_x(x_i)) = 0, \quad \forall i = 1, \dots, M, \end{cases} \quad (2.5)$$

where

$$\mathbf{z} = (z_1^{(1)}, \dots, z_M^{(1)}, z_1^{(2)}, \dots, z_M^{(2)}, z_1^{(3)}, \dots, z_M^{(3)}) \quad (2.6)$$

and

$$\boldsymbol{\rho} = (\rho_1^{(1)}, \dots, \rho_M^{(1)}, \rho_1^{(2)}, \dots, \rho_M^{(2)}). \quad (2.7)$$

Let  $\delta_x$  be the Dirac delta function concentrated at  $x$ . We define  $\phi_i^{(1)} = \delta_{x_i}$ ,  $\phi_i^{(2)} = \delta_{x_i} \circ \partial_x$ , and  $\phi_i^{(3)} = \delta_{x_i} \circ \partial_{xx}$  for  $i = 1, \dots, M$ . Let  $\boldsymbol{\phi}^{(1)}$ ,  $\boldsymbol{\phi}^{(2)}$ , and  $\boldsymbol{\phi}^{(3)}$  be the  $M$ -dimensional vectors with entries  $\phi_i^{(1)}$ ,  $\phi_i^{(2)}$ , and  $\phi_i^{(3)}$ , separately, and let  $\boldsymbol{\phi}$  be the  $(3M)$ -dimensional vector obtained by concatenating  $\boldsymbol{\phi}^{(1)}$ ,  $\boldsymbol{\phi}^{(2)}$ , and  $\boldsymbol{\phi}^{(3)}$ . Denote by  $\boldsymbol{\psi}$  the  $(2M)$ -dimensional vector obtained by concatenating  $\boldsymbol{\phi}^{(1)}$  and  $\boldsymbol{\phi}^{(2)}$ . For ease of presentation, we also write  $\phi_i$  and  $\psi_i$  as the  $i$ th components of  $\boldsymbol{\phi}$  and  $\boldsymbol{\psi}$ , separately.

For  $k = 1, 2$ , let  $K_k(x, \boldsymbol{\phi})$  be the vector with entries  $\int_{\mathbb{T}} K_k(x, x') \phi_i(x') dx'$  and let  $K_k(\boldsymbol{\phi}, \boldsymbol{\phi})$ , called the gram matrix, be with entries  $\int_{\mathbb{T}} \int_{\mathbb{T}} K_k(x, x') \phi_i(x') \phi_j(x') dx dx'$ . Similarly, we define  $K_k(x, \boldsymbol{\psi})$  and  $K_k(\boldsymbol{\psi}, \boldsymbol{\psi})$ . By the representer theorem (see [41, Sec. 17.8]), the first level optimization problem in (2.5) yields

$$\begin{cases} u(x) = \langle K_1(x, \boldsymbol{\phi}), K_1(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z} \rangle, \\ m(x) = \langle K_2(x, \boldsymbol{\psi}), K_2(\boldsymbol{\psi}, \boldsymbol{\psi})^{-1} \boldsymbol{\rho} \rangle, \\ \overline{H} = \lambda. \end{cases} \quad (2.8)$$

Thus, we have

$$\begin{cases} \|u\|_{\mathcal{U}}^2 = \mathbf{z}^T K_1(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z}, \\ \|m\|_{\mathcal{V}}^2 = \boldsymbol{\rho}^T K_2(\boldsymbol{\psi}, \boldsymbol{\psi})^{-1} \boldsymbol{\rho}. \end{cases}$$

Hence, we can formulate (2.5) as a finite-dimensional optimization problem

$$\begin{cases} \min_{\mathbf{z} \in \mathbb{R}^{3M}, \boldsymbol{\rho} \in \mathbb{R}^{2M}, \lambda \in \mathbb{R}} \mathbf{z}^T K_1(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z} + \boldsymbol{\rho}^T K_2(\boldsymbol{\psi}, \boldsymbol{\psi})^{-1} \boldsymbol{\rho} + |\lambda|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M z_i^{(1)} \right|^2 \\ \text{s.t. } H(x, z_i^{(2)}) = \ln \rho_i^{(1)} + \lambda, \forall i = 1, \dots, M \\ \rho_i^{(2)} (z_i^{(2)} + b(x_i)) + \rho_i^{(1)} (z_i^{(3)} + b_x(x_i)) = 0, \forall i = 1, \dots, M. \end{cases} \quad (2.9)$$

To deal with the nonlinear constraints in (2.9), we introduce a prescribed penalization parameter  $\gamma > 0$  and consider the following relaxation

$$\begin{aligned} \min_{\mathbf{z} \in \mathbb{R}^{3M}, \boldsymbol{\rho} \in \mathbb{R}^{2M}, \lambda \in \mathbb{R}} \mathbf{z}^T K_1(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z} + \boldsymbol{\rho}^T K_2(\boldsymbol{\psi}, \boldsymbol{\psi})^{-1} \boldsymbol{\rho} + |\lambda|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M z_i^{(1)} \right|^2 \\ + \gamma \sum_{i=1}^M |e^{H(x, z_i^{(2)}) - \lambda} - \rho_i^{(1)}|^2 + \gamma \sum_{i=1}^M |\rho_i^{(2)} (z_i^{(2)} + b(x_i)) + \rho_i^{(1)} (z_i^{(3)} + b_x(x_i))|^2. \end{aligned} \quad (2.10)$$

The problem (2.10) is the foundation of the GP method for solving (2.1). We use the Gauss–Newton method to solve (2.10), which is detailed in Section 3 of [16].

**Remark 2.1.** In (2.10), we use the exponential form for the constraints from the HJB equation to avoid possible numerical issues when evaluating the logarithm function.

**Remark 2.2.** The matrices  $K_1(\boldsymbol{\phi}, \boldsymbol{\phi})$  and  $K_2(\boldsymbol{\psi}, \boldsymbol{\psi})$  are ill-conditioned in general. To compute  $K_1(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1}$  and  $K_2(\boldsymbol{\psi}, \boldsymbol{\psi})^{-1}$ , we perform the Cholesky decomposition on  $K_1(\boldsymbol{\phi}, \boldsymbol{\phi}) + \eta_1 R_1$  and  $K_2(\boldsymbol{\psi}, \boldsymbol{\psi}) + \eta_2 R_2$ , where  $\eta_1 > 0$ ,  $\eta_2 > 0$  are chosen regularization constants, and  $R_1$ ,  $R_2$  are block diagonal nuggets constructed using the approach introduced in [16]. In the numerical experiments, we precompute the Cholesky decomposition of  $K_1(\boldsymbol{\phi}, \boldsymbol{\phi}) + \eta_1 R_1$  and  $K_2(\boldsymbol{\psi}, \boldsymbol{\psi}) + \eta_2 R_2$ , and store Cholesky factors for further uses.

Following the arguments in [16], the next theorem shows that there exists a solution to (2.4).

**Theorem 2.3.** The minimization problem in (2.4) admits a minimizer  $(u^\dagger, m^\dagger, \overline{H}^\dagger)$  such that

$$\begin{cases} u^\dagger(x) = \langle K_1(x, \boldsymbol{\phi}), K_1(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z}^\dagger \rangle, \\ m^\dagger(x) = \langle K_2(x, \boldsymbol{\psi}), K_2(\boldsymbol{\psi}, \boldsymbol{\psi})^{-1} \boldsymbol{\rho}^\dagger \rangle, \\ \overline{H}^\dagger = \lambda^\dagger, \end{cases} \quad (2.11)$$

where  $(\mathbf{z}^\dagger, \boldsymbol{\rho}^\dagger, \lambda^\dagger)$  is a minimizer of (2.9).

**Proof.** By the above arguments, (2.4) is equivalent to (2.9). Hence, the key is to prove the existence of a minimizer to (2.9). The argument here is similar to the proof of Theorem 1.1 in [16], which proves the existence of a minimizer for the problem of the following form

$$\begin{cases} \min_{\mathbf{z}} \mathbf{z}^T \Theta^{-1} \mathbf{z} \\ \text{s.t. } G(\mathbf{z}) = 0. \end{cases} \quad (2.12)$$

Here,  $\Theta$  is assumed to be an invertible gram matrix and  $G$  is continuous. The arguments of [16] proceed by constructing a vector  $\mathbf{z}_*$  from the solution to the corresponding PDE such that  $G(\mathbf{z}_*) = 0$ . Then, (2.12) is equivalent to

$$\begin{cases} \min_{\mathbf{z}} \mathbf{z}^T \Theta^{-1} \mathbf{z} \\ \text{s.t. } \mathcal{C} = \{\mathbf{z} | G(\mathbf{z}) = 0\} \cap \{\mathbf{z} | \mathbf{z}^T \Theta^{-1} \mathbf{z} \leq \mathbf{z}_*^T \Theta^{-1} \mathbf{z}_*\}. \end{cases}$$

Since  $\mathcal{C}$  is compact (by the continuity of  $G$ ) and nonempty ( $\mathcal{C}$  contains  $\mathbf{z}_*$ ), the objective function  $\mathbf{z}^T \Theta^{-1} \mathbf{z}$  admits a minimum in  $\mathcal{C}$ .

The above arguments can be extended to our MFG setting. Let  $(u^*, m^*, \bar{H}^*)$  be the solution to (2.1). We define the tuple  $(\mathbf{z}_*, \boldsymbol{\rho}_*, \lambda_*)$  such that  $\mathbf{z}_*$  is the vector with entries  $z_{*,i}^{(1)} = u^*(x_i)$ ,  $z_{*,i}^{(2)} = u_x^*(x_i)$ , and  $z_{*,i}^{(3)} = u_{xx}^*(x_i)$ ,  $\boldsymbol{\rho}_*$  consists of elements  $\rho_{*,i}^{(1)} = m^*(x_i)$  and  $\rho_{*,i}^{(2)} = m_x^*(x_i)$ , and  $\lambda_* = \bar{H}^*$  for  $i = 1, \dots, M$ . Then,  $(\mathbf{z}_*, \boldsymbol{\rho}_*, \lambda_*)$  satisfies the constraints in (2.9). For a little abuse of notations, we denote by  $G(\mathbf{z}, \boldsymbol{\rho}, \lambda) = 0$  the constraints in (2.9) and define

$$\begin{aligned} \mathcal{C}_1 = & \left\{ (\mathbf{z}, \boldsymbol{\rho}, \lambda) | \mathbf{z}^T K_1(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z} + \boldsymbol{\rho}^T K_2(\boldsymbol{\psi}, \boldsymbol{\psi})^{-1} \boldsymbol{\rho} + |\lambda|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M z_i^{(1)} \right|^2 \right. \\ & \left. \leq \mathbf{z}_*^T K_1(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z}_* + \boldsymbol{\rho}_*^T K_2(\boldsymbol{\psi}, \boldsymbol{\psi})^{-1} \boldsymbol{\rho}_* + |\lambda_*|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M \rho_{*,i}^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M z_{*,i}^{(1)} \right|^2 \right\}. \end{aligned}$$

Then, (2.9) is equivalent to

$$\begin{cases} \min_{\mathbf{z}, \boldsymbol{\rho}, \lambda} \mathbf{z}^T K_1(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z} + \boldsymbol{\rho}^T K_2(\boldsymbol{\psi}, \boldsymbol{\psi})^{-1} \boldsymbol{\rho} + |\lambda|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M z_i^{(1)} \right|^2 \\ \text{s.t. } (\mathbf{z}, \boldsymbol{\rho}, \lambda) \in \mathcal{C} := \{(\mathbf{z}, \boldsymbol{\rho}, \lambda) | G(\mathbf{z}, \boldsymbol{\rho}, \lambda) = 0\} \cap \mathcal{C}_1. \end{cases} \quad (2.13)$$

By the continuity of  $G$  in  $(\mathbf{z}, \boldsymbol{\rho}, \lambda)$  and the fact that  $(\mathbf{z}_*, \boldsymbol{\rho}_*, \lambda_*) \in \mathcal{C}$ ,  $\mathcal{C}$  is compact and non-empty. Hence, the objective function (2.13) achieves a minimum on  $\mathcal{C}$ . Thus, (2.9) admits a minimizer. We conclude (2.11) by (2.8).  $\square$

Using a similar argument as in [16], we obtain the following convergence theory.

**Theorem 2.4.** Assume that the kernels  $K_1$  and  $K_2$  are chosen such that  $\mathcal{U} \subset \subset H^{s_1}(\mathbb{T})$  and  $\mathcal{V} \subset \subset H^{s_2}(\mathbb{T})$  for some  $s_1 > 3$  and  $s_2 > 2$ . Let  $(u^*, m^*, \bar{H}^*)$  be the solution to (2.1). Denote by  $(u_{M,\beta}^\dagger, m_{M,\beta}^\dagger, \bar{H}_{M,\beta}^\dagger)$  a minimizer of (2.4) with  $M$  different sample points  $\{x_i\}_{i=1}^M$  and the penalization parameter  $\beta$ . Suppose further that as  $M \rightarrow \infty$ ,

$$\sup_{x \in \mathbb{T}} \min_{1 \leq i \leq M} |x - x_i| \rightarrow 0. \quad (2.14)$$

Then, as  $\beta$  and  $M$  go to infinity, up to a subsequence,  $(u_{M,\beta}^\dagger, m_{M,\beta}^\dagger, \bar{H}_{M,\beta}^\dagger)$  converges to  $(u^*, m^*, \bar{H}^*)$  pointwisely in  $\mathbb{T}$  and in  $H^{t_1}(\mathbb{T}) \times H^{t_2}(\mathbb{T}) \times \mathbb{R}$  for any  $t_1 \in (3, s_1)$  and any  $t_2 \in (2, s_2)$ .

**Proof.** The argument is a direct adaptation of the proof of Theorem 1.2 in [16]. Let  $(u^*, m^*, \bar{H}^*)$  be the solution of (2.1). Clearly,  $(u^*, m^*, \bar{H}^*)$  satisfies the constraints in (2.4). According to (2.2),  $u^*$  and  $m^*$  are smooth. Thus, a minimizer  $(u_{M,\beta}^\dagger, m_{M,\beta}^\dagger, \bar{H}_{M,\beta}^\dagger)$  to (2.4) with  $M$  different sample points and the penalization constant  $\beta$  satisfies

$$\begin{aligned} & \|u_{M,\beta}^\dagger\|_{\mathcal{U}}^2 + \|m_{M,\beta}^\dagger\|_{\mathcal{V}}^2 + |\bar{H}_{M,\beta}^\dagger|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M m_{M,\beta}^\dagger(x_i) - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M u_{M,\beta}^\dagger(x_i) \right|^2 \\ & \leq \|u^*\|_{\mathcal{U}}^2 + \|m^*\|_{\mathcal{V}}^2 + |\bar{H}^*|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M m^*(x_i) - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M u^*(x_i) \right|^2. \end{aligned} \quad (2.15)$$

By  $\int_{\mathbb{T}} u^* dx = 0$ ,  $\int_{\mathbb{T}} m^* dx = 1$  and (2.14), there exist a sequence  $\{(\beta_p, M_p)\}_{p=1}^\infty$  and a constant  $C$  such that

$$\beta_p \left| \frac{1}{M_p} \sum_{i=1}^{M_p} m^*(x_i) - 1 \right|^2 + \beta_p \left| \frac{1}{M_p} \sum_{i=1}^{M_p} u^*(x_i) \right|^2 \leq C, \text{ for } p \geq 1. \quad (2.16)$$

Thus, using (2.15) and (2.16), we get

$$\|u_{M,\beta}^\dagger\|_{\mathcal{U}}^2 + \|m_{M,\beta}^\dagger\|_{\mathcal{V}}^2 + |\bar{H}_{M,\beta}^\dagger|^2 \leq C. \quad (2.17)$$

Meanwhile, by  $\mathcal{U} \subset \subset H^{s_1}$  and  $\mathcal{V} \subset \subset H^{s_2}$ ,  $s_1 > 3$  and  $s_2 > 2$ , there exists a constant  $C$  such that

$$\|u_{M,\beta}^\dagger\|_{H^{s_1}(\mathbb{T})} \leq C \|u_{M,\beta}^\dagger\|_{\mathcal{U}} \text{ and } \|m_{M,\beta}^\dagger\|_{H^{s_2}(\mathbb{T})} \leq C \|m_{M,\beta}^\dagger\|_{\mathcal{V}}. \quad (2.18)$$

For any  $t_1 \in (3, s_1)$  and any  $t_2 \in (2, s_2)$ , we have  $H^{s_1}(\mathbb{T}) \subset \subset H^{t_1}(\mathbb{T})$  and  $H^{s_2}(\mathbb{T}) \subset \subset H^{t_2}(\mathbb{T})$ . Thus, according to (2.17) and (2.18), there exists a limit  $(u_\infty^\dagger, m_\infty^\dagger, \bar{H}_\infty^\dagger) \in H^{t_1}(\mathbb{T}) \times H^{t_2}(\mathbb{T}) \times \mathbb{R}$  such that, up to a subsequence,

$$u_{M_p,\beta_p}^\dagger \rightarrow u_\infty^\dagger \text{ in } H^{t_1}(\mathbb{T}),$$

$$m_{M_p,\beta_p}^\dagger \rightarrow m_\infty^\dagger \text{ in } H^{t_2}(\mathbb{T}),$$

and

$$\bar{H}_{M_p,\beta_p}^\dagger \rightarrow \bar{H}_\infty^\dagger \text{ in } \mathbb{R},$$

as  $p \rightarrow \infty$ . Since  $H^{t_1}(\mathbb{T}) \subset \subset C^2(\mathbb{T})$  and  $H^{t_2}(\mathbb{T}) \subset \subset C^1(\mathbb{T})$ , the limit  $(u_\infty^\dagger, m_\infty^\dagger, \bar{H}_\infty^\dagger)$  satisfies the constraints in (2.4). By (2.14), the collection of points  $\{x_i\}_{i=1}^{M_p}$  forms a dense subset of  $\mathbb{T}$  as  $M_p \rightarrow \infty$ . Thus, dividing both sides of (2.15) by  $\beta$  and passing  $M$  and  $\beta$  to infinity, we get  $\int_{\mathbb{T}} m_\infty^\dagger dx = 1$  and  $\int_{\mathbb{T}} u_\infty^\dagger dx = 0$ . Hence, by the regularity of  $u_\infty^\dagger$  and  $m_\infty^\dagger$ ,  $(u_\infty^\dagger, m_\infty^\dagger, \bar{H}_\infty^\dagger)$  is a classical solution to (2.9) and by the uniqueness of the solution to (2.1),  $(u_\infty^\dagger, m_\infty^\dagger, \bar{H}_\infty^\dagger) = (u^*, m^*, \bar{H}^*)$ . The limit  $(u_\infty^\dagger, m_\infty^\dagger, \bar{H}_\infty^\dagger)$  is independent of the choice of the samples and the convergent subsequence. Therefore, we conclude that  $(u_{M,\beta}^\dagger, m_{M,\beta}^\dagger, \bar{H}_{M,\beta}^\dagger)$  converges to  $(u^*, m^*, \bar{H}^*)$  pointwisely in  $\mathbb{T}$  and in  $H^{t_1}(\mathbb{T}) \times H^{t_2}(\mathbb{T}) \times \mathbb{R}$ .  $\square$

## 2.2. The Fourier features method

The main bottleneck of solving (2.9) is to compute the inverses of  $K_1(\phi, \phi)$  and  $K_2(\psi, \psi)$ , whose dimensions increase with the product of the size of samples and the number of linear operators in the MFG. Hence, the computational cost of the GP method grows dramatically as the number of samples increases. We propose the FF method based on the idea of approximating kernels with randomized trigonometric functions (see [43,46,48]), which reduces the precomputation time and the amount of storage significantly.

The idea of our Fourier Features method comes from the following observations. According to (2.11), the function  $u^\dagger$  in the minimizer of (2.4) has the following form

$$u^\dagger(x) = \sum_{i=1}^M a_i K_1(x, x_i) + \sum_{i=1}^M b_i \partial_y K_1(x, x_i) + \sum_{i=1}^M c_i \partial_y^2 K_1(x, x_i), \quad (2.19)$$

where  $a_i, b_i, c_i \in \mathbb{R}$  are coefficients determined by  $K_1(\phi, \phi)^{-1} \mathbf{z}^\dagger$ . We assume that  $K_1$  is shift-invariant and properly scaled. According to the Bochner theorem [43,44], the Fourier transform  $p$  of  $K_1$  is a probability distribution and

$$K_1(x - y) = E_w[\cos(\omega(x - y))], \forall x, y \in \mathbb{R}, \quad (2.20)$$

where  $w$  is a random variable following  $p$ . Combining (2.19) and (2.20), we get

$$u^\dagger(x) = \sum_{i=1}^M a_i E_w[\cos(w(x - x_i))] + \sum_{i=1}^M b_i E_w[w \sin(w(x - x_i))] - \sum_{i=1}^M c_i E_w[w^2 \cos(w(x - x_i))].$$

Thus, drawing  $N/2$  samples  $\{w_j\}_{j=1}^{N/2}$  from  $p$ , using trigonometric identities, and by the law of large numbers, we obtain

$$\begin{aligned} u^\dagger(x) &\approx \sum_{j=1}^{N/2} \sum_{i=1}^M (a_i \cos(w_j x_i) - b_i w_j \sin(w_j x_i) - c_i w_j^2 \cos(w_j x_i)) \cos(w_j x) \\ &\quad + \sum_{j=1}^{N/2} \sum_{i=1}^M (-a_i \sin(w_j x_i) + b_i w_j \cos(w_j x_i) + c_i w_j^2 \sin(w_j x_i)) \sin(w_j x). \end{aligned} \quad (2.21)$$

From (2.21), we observe that  $u^\dagger$  can be approximated by a linear combination of randomized trigonometric functions. The same conclusion holds for  $m^\dagger$  in (2.11). Hence, (2.21) motivates us to approximate solutions of MFGs in the space generated by sampled trigonometric functions.

Next, we present the FF method by solving (2.1). To deal with the periodic boundary condition in (2.1), inspired by [46], we approximate  $u$  and  $m$  by functions in the space generated by the Fourier series. Non-periodic settings are discussed in Section 3, where we present a general framework of the FF method.

Given  $N \in \mathbb{N}$ , we define the set

$$\mathcal{G}^N = \left\{ \phi \left| \phi(x) = c + \sum_{i=1}^N \alpha_i \sin(2\pi i x) + \sum_{i=1}^N \beta_i \cos(2\pi i x), c, \alpha_i, \beta_i \in \mathbb{R} \right. \right\}. \quad (2.22)$$

We call  $\mathcal{G}^N$  the Fourier features space. Meanwhile, we consider functions  $\sin(2\pi i x)$  and  $\cos(2\pi i x)$  for  $i = 1, \dots, N$  as features in  $\mathcal{G}^N$ . Given  $\gamma > 0$ , we define the functional

$$J_\gamma(u, m, \bar{H}) = \int_{\mathbb{T}} |u|^2 dx + \int_{\mathbb{T}} |m|^2 dx + \bar{H}^2 + \gamma \mathcal{Q}(u, m, \bar{H}), \quad (2.23)$$

where

$$\mathcal{Q}(u, m, \bar{H}) = \int_{\mathbb{T}} |e^{H(x, u_x) - \bar{H}} - m|^2 dx + \left| \int_{\mathbb{T}} u dx \right|^2 + \left| \int_{\mathbb{T}} m dx - 1 \right|^2 + \int_{\mathbb{T}} |m(u_{xx} + b_x(x)) + m_x(u_x + b(x))|^2 dx. \quad (2.24)$$

Then, we approximate the solution of (2.1) by a minimizer of the following problem

$$\min_{u^N \in \mathcal{G}^N, m^N \in \mathcal{G}^N, \bar{H}^N \in \mathbb{R}} J_\gamma(u^N, m^N, \bar{H}^N). \quad (2.25)$$

The following theorem gives the existence of a solution to (2.25).

**Theorem 2.5.** *The minimization problem (2.25) admits a minimizer.*

**Proof.** Let  $N \in \mathbb{N}$ . We define

$$\xi(x) = [1, \sin(2\pi x), \dots, \sin(2\pi Nx), \cos(2\pi x), \dots, \cos(2\pi Nx)]^T. \quad (2.26)$$

Then, for any  $(u^N, m^N) \in \mathcal{G}^N \times \mathcal{G}^N$ , there exists  $\alpha \in \mathbb{R}^{2N+1}$  and  $\beta \in \mathbb{R}^{2N+1}$  such that

$$u^N = \alpha^T \xi \text{ and } m^N = \beta^T \xi.$$

Therefore, (2.25) is equivalent to

$$\min_{\alpha \in \mathbb{R}^{2N+1}, \beta \in \mathbb{R}^{2N+1}, \bar{H}^N \in \mathbb{R}} J_\gamma(\alpha^T \xi, \beta^T \xi, \bar{H}^N). \quad (2.27)$$

We observe that the objective function in (2.27) is co-coercive and continuous. Thus, (2.27) admits a solution. Therefore, there is a minimizer to (2.25).  $\square$

Then, we prove the convergence of the minimizer of (2.25) to the solution of (2.1) as  $\gamma$  and  $N$  go to infinity. First, we give a bound for the minimum of  $J_\gamma$  in the following theorem.

**Theorem 2.6.** *Let  $(u^*, m^*, \bar{H}^*)$  be the solution to (2.1), let  $\mathcal{G}^N$  be as in (2.22), and let  $J_\gamma$  be as in (2.23) for  $\gamma > 0$ . Then, for any sufficiently small  $\varepsilon > 0$ , there exist a constant  $C > 0$ , and functions  $(u^N, m^N) \in \mathcal{G}^N \times \mathcal{G}^N$  such that*

$$J_\gamma(u^N, m^N, \bar{H}^*) \leq 2\|u^*\|_{L^2(\mathbb{T})}^2 + 2\|m^*\|_{L^2(\mathbb{T})}^2 + |\bar{H}^*|^2 + C(1 + \gamma)\varepsilon^2. \quad (2.28)$$

We refer readers to Appendix A for the proof of Theorem 2.6, which directly yields the following corollary.

**Corollary 2.7.** *Let  $\mathcal{G}^N$  be as in (2.22) and let  $\mathcal{Q}$  be given in (2.24). For any sufficient small  $\varepsilon > 0$ , there exist a constant  $C > 0$ , a sufficiently large  $\gamma > 0$ , and  $N > 0$  such that any minimizer  $(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma}) \in \mathcal{G}^N \times \mathcal{G}^N \times \mathbb{R}$  of (2.25) corresponding to  $\gamma$  satisfies*

$$\mathcal{Q}(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma}) \leq C\varepsilon. \quad (2.29)$$

See Appendix A for the proof of the above corollary. The following theorem shows that any minimizer of (2.25) converges to the solution of (2.1) as  $N$  and  $\gamma$  increase.

**Theorem 2.8.** *There exists a sequence  $\{(N_i, \gamma_i)\}_{i=1}^{\infty}$  such that any minimizer  $(u^i, m^i, \bar{H}^i)$  of (2.25) corresponding to  $(N_i, \gamma_i)$  satisfies  $u^i \rightarrow u^*$  in  $H^1(\mathbb{T})$ ,  $m^i \rightarrow m^*$  in  $L^1(\mathbb{T})$ , and  $\bar{H}^i \rightarrow \bar{H}^*$  in  $\mathbb{R}$  as  $i$  goes to infinity.*

The proof is given in Appendix A. Next, we propose a numerical method to solve (2.27), since (2.25) and (2.27) are equivalent. Let  $N \in \mathbb{N}$  and take  $M$  samples  $\{x_i\}_{i=1}^M$ . Let  $\zeta$  be given in (2.26). Introducing penalization parameters  $\gamma > 0$  and  $\beta > 0$ , we reformulate (2.27) as an equivalent two-level optimization problem

$$\begin{aligned} \min_{\mathbf{z} \in \mathbb{R}^{3M}, \boldsymbol{\rho} \in \mathbb{R}^{2M}, \lambda \in \mathbb{R}} \quad & \begin{cases} \min_{\boldsymbol{\alpha} \in \mathbb{R}^{2N+1}, \boldsymbol{\beta} \in \mathbb{R}^{2N+1}, \bar{H} \in \mathbb{R}} & \|\boldsymbol{\alpha}\|^2 + \|\boldsymbol{\beta}\|^2 + |\bar{H}|^2 \\ \text{s.t. } & \boldsymbol{\alpha}^T \boldsymbol{\zeta}(x_i) = z_i^{(1)}, \boldsymbol{\alpha}^T \boldsymbol{\zeta}_x(x_i) = z_i^{(2)}, \boldsymbol{\alpha}^T \boldsymbol{\zeta}_{xx}(x_i) = z_i^{(3)}, \\ & \boldsymbol{\beta}^T \boldsymbol{\zeta}(x_i) = \rho_i^{(1)}, \boldsymbol{\beta}^T \boldsymbol{\zeta}_x(x_i) = \rho_i^{(2)}, \bar{H} = \lambda, \end{cases} \\ & + \beta \left| \frac{1}{M} \sum_{i=1}^M \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M z_i^{(1)} \right|^2 + \gamma \sum_{i=1}^M |e^{H(x_i, z_i^{(2)}) - \lambda} - \rho_i^{(1)}|^2 \\ & + \gamma \sum_{i=1}^M |\rho_i^{(2)}(z_i^{(2)} + b(x_i)) + \rho_i^{(1)}(z_i^{(3)} + b_x(x_i))|^2, \end{aligned} \quad (2.30)$$

where  $\mathbf{z}$  and  $\boldsymbol{\rho}$  have forms in (2.6) and (2.7). In (2.30), we use two penalization parameters to take into account different variability of the constraints. Let  $\boldsymbol{\phi}, \boldsymbol{\psi}$  be defined as in Subsection 2.1. Denote by  $\mathcal{U}^*$  the dual of the Banach space  $\mathcal{U}$  and by  $[\cdot, \cdot]$  the duality pairing. Let  $[\boldsymbol{\phi}, \boldsymbol{\zeta}]$  and  $[\boldsymbol{\psi}, \boldsymbol{\zeta}]$  be the matrices with entries  $[\phi_i, \zeta_j]$  and  $[\psi_i, \zeta_j]$ , separately. Then, the first level optimization problem of (2.30) yields

$$\boldsymbol{\alpha} = [\boldsymbol{\phi}, \boldsymbol{\zeta}]^T ([\boldsymbol{\phi}, \boldsymbol{\zeta}][\boldsymbol{\phi}, \boldsymbol{\zeta}]^T)^{-1} \mathbf{z}, \boldsymbol{\beta} = [\boldsymbol{\psi}, \boldsymbol{\zeta}]^T ([\boldsymbol{\psi}, \boldsymbol{\zeta}][\boldsymbol{\psi}, \boldsymbol{\zeta}]^T)^{-1} \boldsymbol{\rho}, \bar{H} = \lambda. \quad (2.31)$$

From (2.31), we get

$$\|\boldsymbol{\alpha}\|^2 = \mathbf{z}^T ([\boldsymbol{\phi}, \boldsymbol{\zeta}][\boldsymbol{\phi}, \boldsymbol{\zeta}]^T)^{-1} \mathbf{z}, \|\boldsymbol{\beta}\|^2 = \boldsymbol{\rho}^T ([\boldsymbol{\psi}, \boldsymbol{\zeta}][\boldsymbol{\psi}, \boldsymbol{\zeta}]^T)^{-1} \boldsymbol{\rho}.$$

Hence, (2.30) is equivalent to

$$\begin{aligned} \min_{\mathbf{z} \in \mathbb{R}^{3M}, \boldsymbol{\rho} \in \mathbb{R}^{2M}, \lambda \in \mathbb{R}} \quad & \mathbf{z}^T ([\boldsymbol{\phi}, \boldsymbol{\zeta}][\boldsymbol{\phi}, \boldsymbol{\zeta}]^T)^{-1} \mathbf{z} + \boldsymbol{\rho}^T ([\boldsymbol{\psi}, \boldsymbol{\zeta}][\boldsymbol{\psi}, \boldsymbol{\zeta}]^T)^{-1} \boldsymbol{\rho} + |\lambda|^2 \\ & + \gamma \sum_{i=1}^M |e^{H(x_i, z_i^{(2)}) - \lambda} - \rho_i^{(1)}|^2 + \gamma \sum_{i=1}^M |\rho_i^{(2)}(z_i^{(2)} + b(x_i)) + \rho_i^{(1)}(z_i^{(3)} + b_x(x_i))|^2, \\ & + \beta \left| \frac{1}{M} \sum_{i=1}^M \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M} \sum_{i=1}^M z_i^{(1)} \right|^2. \end{aligned} \quad (2.32)$$

Then, we apply the Gauss–Newton method to solve (2.32).

**Remark 2.9.** In general,  $[\boldsymbol{\phi}, \boldsymbol{\zeta}][\boldsymbol{\phi}, \boldsymbol{\zeta}]^T$  and  $[\boldsymbol{\psi}, \boldsymbol{\zeta}][\boldsymbol{\psi}, \boldsymbol{\zeta}]^T$  are ill-conditioned. Hence, we introduce two regularization parameters  $\mu_1 > 0$  and  $\mu_2 > 0$ , and consider the inverses of  $[\boldsymbol{\phi}, \boldsymbol{\zeta}][\boldsymbol{\phi}, \boldsymbol{\zeta}]^T + \mu_1 I$  and  $[\boldsymbol{\psi}, \boldsymbol{\zeta}][\boldsymbol{\psi}, \boldsymbol{\zeta}]^T + \mu_2 I$ .

**Remark 2.10.** We propose an alternative method other than the Cholesky factorization to compute  $[\boldsymbol{\phi}, \boldsymbol{\zeta}^N][\boldsymbol{\phi}, \boldsymbol{\zeta}^N]^T + \mu_1 I$  for  $\mu_1 > 0$ , which is the cornerstone for the FF method to save the precomputation time and the memory. For simplicity, denote  $A := [\boldsymbol{\phi}, \boldsymbol{\zeta}^N][\boldsymbol{\phi}, \boldsymbol{\zeta}^N]^T$ . To solve  $(AA^T + \mu_1 I)^{-1}$ , we first perform QR decomposition on  $A$ . We get

$$A = QR = [Q_1 \ Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1,$$

where  $Q \in \mathbb{R}^{3M \times 3M}$  is orthonormal,  $R \in \mathbb{R}^{3M \times (2N+1)}$ ,  $Q_1 \in \mathbb{R}^{3M \times (2N+1)}$ ,  $Q_2 \in \mathbb{R}^{3M \times (3M-2N-1)}$ , and  $R_1 \in \mathbb{R}^{(2N+1) \times (2N+1)}$ . Then, we have

$$AA^T + \mu_1 I = Q R R^T Q^T + \mu_1 I = Q R R^T Q^T + \mu_1 Q Q^T.$$

Thus, we get

**Table 1**

$L^\infty$  errors of numerical approximations for the solution of (2.1), given the number of sample points  $M = 256$ .

|                     | GP                    | FF                    |
|---------------------|-----------------------|-----------------------|
| Errors of $u$       | $3.08 \times 10^{-6}$ | $9.09 \times 10^{-5}$ |
| Errors of $m$       | $3.70 \times 10^{-2}$ | $1.68 \times 10^{-2}$ |
| Errors of $\bar{H}$ | $7.83 \times 10^{-4}$ | $9.87 \times 10^{-4}$ |

$$\begin{aligned} (AA^T + \mu_1 I)^{-1} &= Q(RR^T + \mu_1 I)^{-1}Q^T = Q \begin{bmatrix} R_1 R_1^T + \mu_1 I & 0 \\ 0 & \mu_1 I \end{bmatrix}^{-1} Q^T \\ &= Q_1(R_1 R_1^T + \mu_1 I)^{-1}Q_1^T + \frac{1}{\mu_1}Q_2Q_2^T. \end{aligned}$$

Using

$$I = Q Q^T = Q_1 Q_1^T + Q_2 Q_2^T,$$

we have

$$(AA^T + \mu_1 I)^{-1} = Q_1(R_1 R_1^T + \mu_1 I)^{-1}Q_1^T + \frac{1}{\mu_1}(I - Q_1 Q_1^T). \quad (2.33)$$

Since  $R_1 R_1^T + \mu_1 I \in \mathbb{R}^{(2N+1) \times (2N+1)}$ , computing the right-hand side of (2.33) consumes less CPU time than calculating the left-hand side of (2.33) if  $N < M$ . In the FF method, we precompute the QR decomposition of  $A$  and the Cholesky decomposition of  $R_1 R_1^T + \mu_1 I$ , and store the results for further uses. We apply the same technique to compute the inverse of  $[\psi, \zeta][\psi, \zeta]^T + \mu_2 I$ .

**Remark 2.11.** The Woodbury matrix identity admits the same advantage as what we stated for (2.33) in Remark 2.10. However, the Woodbury formula is numerically unstable in our experiments, especially for small values of  $\mu_1$  and  $\mu_2$ .

### 2.3. Numerical results

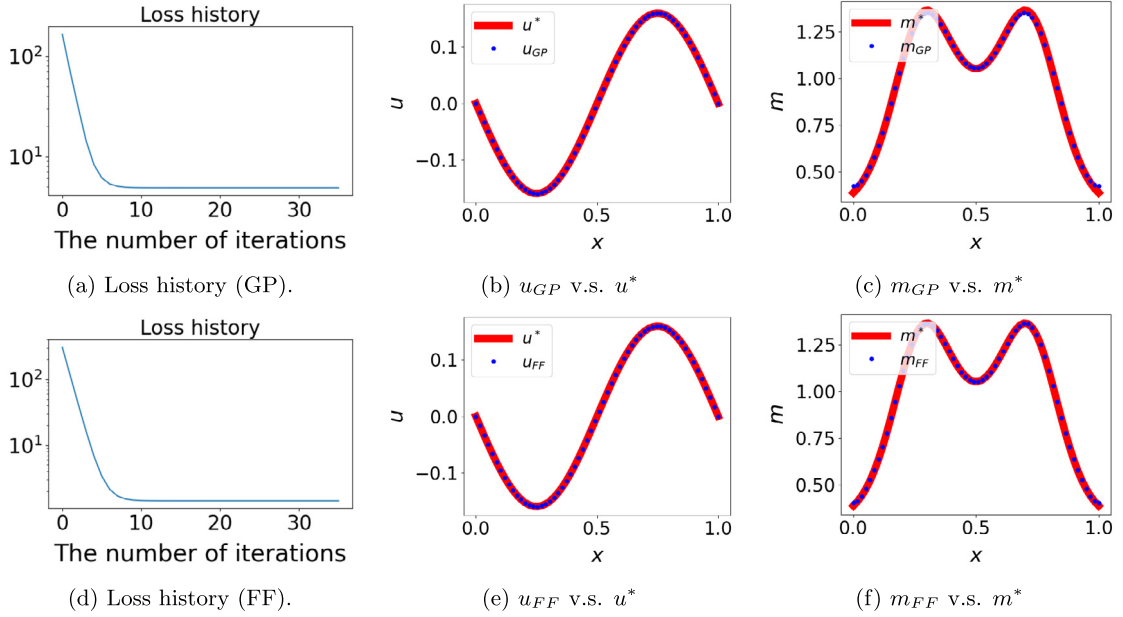
To demonstrate the efficacy of our algorithms, we show here a simple numerical experiment by solving (2.1). We perform the calculation using MacBook Air 2015 (4 GB RAM, Intel Core i5 CPU). Let  $V(x) = \sin(\pi x)$  and  $b(x) = \cos(2\pi x)$  for  $x \in \mathbb{T}$ . For the GP method, we choose the periodic kernel used in [46] for both  $K_1$  and  $K_2$ , i.e.,

$$K_1(x, y) = K_2(x, y) = e^{\frac{\cos(2\pi(x-y))-1}{\sigma^2}}, \quad \forall x, y \in \mathbb{T},$$

with lengthscale  $\sigma > 0$ . We denote by  $(u_{GP}, m_{GP}, \bar{H}_{GP})$  the numerical result of the GP method. For the FF algorithm, given  $N \in \mathbb{N}$ , we approximate the solution of (2.1) in the space  $\mathcal{G}^N$  given by (2.22). We represent the numerical solution of the FF method by  $(u_{FF}, m_{FF}, \bar{H}_{FF})$ . Let  $(u^*, m^*, \bar{H}^*)$  be the solution of (2.1) given by (2.2). We choose  $\sigma = 0.6$  and  $N = 10$ . For both methods, we use the Gauss–Newton method and take the step size 0.4. We set the regularization parameters  $\eta_1 = \eta_2 = \mu_1 = \mu_2 = 10^{-6}$  in Remarks 2.2 and 2.9. Meanwhile, we choose the penalization constants  $\beta = 10^{-6}$  and  $\gamma = 1$  in (2.10) and (2.30). Both algorithms start from the same initial point and stop after 35 iterations. In Fig. 1, we show the numerical solutions of the GP method and the FF algorithm when we take uniformly distributed  $M = 256$  sample points. We present  $L^\infty$  errors in Table 1. We see that the FF algorithm is comparable to the GP method in terms of accuracy. Table 2 records the CPU time consumed by the Cholesky and the QR decomposition by both methods, which implies that the FF algorithm costs less precomputation time than the GP method. We see that the CPU time consumed by the Cholesky decomposition in the FF method does not increase with the number of samples since we choose a fixed number of base functions. Surprisingly, the CPU time of the QR decomposition for the FF method even decreases when  $M = 2048$ . We attribute it to the tall-and-skinny property of  $[\phi, \zeta][\phi, \zeta]^T$  and  $[\psi, \zeta][\psi, \zeta]^T$  when  $M \gg N$ , i.e.,  $[\phi, \zeta][\phi, \zeta]^T$  and  $[\psi, \zeta][\psi, \zeta]^T$  have more rows than columns.

### 3. The general frameworks

This section presents the general frameworks of the GP method and the FF algorithm for solving MFGs. We mainly state the settings for stationary MFGs. The arguments can be naturally adapted to the time-dependent cases. Numerical experiments on both stationary and time-dependent MFGs follow in Section 4.



**Fig. 1.** Numerical results for (2.1): (a), (d) the histories of loss functions; (b), (e) the numerical results  $u_{GP}$  and  $u_{FF}$  v.s. the explicit solution  $u^*$ ; (c), (f) the numerical approximations  $m_{GP}$  and  $m_{FF}$  v.s. the explicit solution  $m^*$ . (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

**Table 2**

The precomputation time (in seconds) for solving (2.1) as  $M$  increases.

| $M$  | GP       | FF   |          |
|------|----------|------|----------|
|      | Cholesky | QR   | Cholesky |
| 256  | 0.70     | 0.30 | 0.18     |
| 512  | 0.78     | 0.32 | 0.18     |
| 1024 | 1.59     | 0.34 | 0.18     |
| 2048 | 8.08     | 0.24 | 0.18     |

### 3.1. The general forms of MFGs

Let  $\Omega$  be a subset of  $\mathbb{R}^d$ . Suppose that the stationary MFGs of our interest have the form

$$\begin{cases} \mathcal{P}(u^*, m^*, \bar{H}^*)(x) = 0, \forall x \in \text{int } \Omega, \\ \mathcal{B}(u^*, m^*)(x) = 0, \forall x \in \partial \Omega, \\ \int_{\Omega} u^* dx = 0, \int_{\Omega} m^* dx = 1. \end{cases} \quad (3.1)$$

Here,  $\mathcal{P}$  is a nonlinear differential operator and  $\mathcal{B}$  represents a boundary operator. We assume that (3.1) admits a unique classical solution  $(u^*, m^*, \bar{H}^*)$ . If the solution of (3.1) is not smooth enough, we suggest using the vanishing viscosity method [11] or regularizing the MFG with smooth mollifiers [15] to get a system with a solution of stronger regularity and applying our numerical methods to compute approximated solutions.

**Remark 3.1.** In time-dependent settings, let  $\Omega$  be a space-time domain. We consider MFGs taking the form

$$\begin{cases} \mathcal{P}(u^*, m^*)(x) = 0, \forall x \in \text{int } \Omega, \\ \mathcal{B}(u^*, m^*)(x) = 0, \forall x \in \partial \Omega, \end{cases} \quad (3.2)$$

and assume that  $(u^*, m^*)$  is a unique classical solution to (3.2).

### 3.2. The Gaussian process method

Using the method in [16], we approximate  $(u^*, m^*)$  in the solution of (3.1) by two GPs conditioned on PDEs at sampled collocation points in  $\Omega$ . Then, we compute the solution by calculating the MAP points of such conditioned GPs. More

precisely, we take a set of samples  $\{x_i\}_{i=1}^M$  in such a way that  $x_1, \dots, x_{M_\Omega} \in \text{int } \Omega$  and  $x_{M_\Omega+1}, \dots, x_M \in \partial \Omega$  for  $1 \leq M_\Omega \leq M$ . Let  $\mathcal{U}$  and  $\mathcal{V}$  be Banach Spaces with associated covariance operators  $\mathcal{K}_\mathcal{U} : \mathcal{U}^* \mapsto \mathcal{U}$  and  $\mathcal{K}_\mathcal{V} : \mathcal{V}^* \mapsto \mathcal{V}$ , respectively. Following [16], we introduce a penalization parameter  $\beta > 0$  and consider the following problem

$$\begin{cases} \min_{(u, m, \bar{H}) \in \mathcal{U} \times \mathcal{V} \times \mathbb{R}} \|u\|_{\mathcal{U}}^2 + \|m\|_{\mathcal{V}}^2 + |\bar{H}|^2 + \beta \left| \frac{1}{M_\Omega} \sum_{i=1}^{M_\Omega} u(x_i) \right|^2 + \beta \left| \frac{1}{M_\Omega} \sum_{i=1}^{M_\Omega} m(x_i) - 1 \right|^2 \\ \text{s.t. } \mathcal{P}(u, m, \bar{H})(x_i) = 0, \quad \text{for } i = 1, \dots, M_\Omega, \\ \mathcal{B}(u, m)(x_j) = 0, \quad \text{for } j = M_\Omega+1, \dots, M. \end{cases} \quad (3.3)$$

We further make a similar assumption to Assumption 3.1 in [16] on  $\mathcal{P}$  and  $\mathcal{B}$ .

**Assumption 1.** For  $1 \leq Q_b \leq Q$  and  $1 \leq D_b \leq D$ , there exist bounded and linear operators  $L_1, \dots, L_{Q_b} \in \mathcal{L}(\mathcal{U}; C(\partial \Omega))$ ,  $L_{Q_b+1}, \dots, L_Q \in \mathcal{L}(\mathcal{U}; C(\text{int } \Omega))$ ,  $J_1, \dots, J_{D_b} \in \mathcal{L}(\mathcal{V}; C(\partial \Omega))$ ,  $J_{D_b+1}, \dots, J_D \in \mathcal{L}(\mathcal{V}; C(\text{int } \Omega))$ , and continuous nonlinear maps  $P$  and  $B$  such that

$$\begin{cases} \mathcal{P}(u^*, m^*, \bar{H}^*)(x) = P(L_{Q_b+1}(u^*)(x), \dots, L_Q(u^*)(x), J_{D_b+1}(m^*)(x), \dots, J_D(m^*)(x), \bar{H}^*), \forall x \in \text{int } \Omega, \\ \mathcal{B}(u^*, m^*)(x) = B(L_1(u^*)(x), \dots, L_{Q_b}(u^*)(x), J_1(m^*)(x), \dots, J_{D_b}(m^*)(x)), \forall x \in \partial \Omega. \end{cases} \quad (3.4)$$

Following [16], under Assumption 1, we define functionals  $\phi_i^{(q)} \in \mathcal{U}^*$  and  $\psi_i^{(p)} \in \mathcal{V}^*$  as

$$\phi_i^{(q)} = \delta_{x_i} \circ L_q, \text{ where } \begin{cases} M_\Omega + 1 \leq i \leq M, & \text{if } 1 \leq q \leq Q_b, \\ 1 \leq i \leq M_\Omega, & \text{if } Q_b+1 \leq q \leq Q, \end{cases}$$

and

$$\psi_i^{(p)} = \delta_{x_i} \circ J_p, \text{ where } \begin{cases} M_\Omega + 1 \leq i \leq M, & \text{if } 1 \leq p \leq D_b, \\ 1 \leq i \leq M_\Omega, & \text{if } D_b+1 \leq p \leq D. \end{cases}$$

For ease of presentation, we denote by  $\phi^{(q)}$  the vector consisting of  $\phi_i^{(q)}$  and define

$$\phi = (\phi^{(1)}, \dots, \phi^{(Q)}) \in (\mathcal{U}^*)^{\otimes N_\mathcal{U}}, \text{ where } N_\mathcal{U} = (M - M_\Omega)Q_b + M_\Omega(Q - Q_b). \quad (3.5)$$

Similarly, we concatenate  $\psi_i^{(p)}$  to get the vector  $\psi^{(p)}$  and denote

$$\psi = (\psi^{(1)}, \dots, \psi^{(D)}) \in (\mathcal{V}^*)^{\otimes N_\mathcal{V}}, \text{ where } N_\mathcal{V} = (M - M_\Omega)D_b + M_\Omega(D - D_b). \quad (3.6)$$

According to Assumption 1, we define the nonlinear map  $G$  such that for any  $u \in \mathcal{U}$ ,  $m \in \mathcal{V}$ ,  $\bar{H} \in \mathbb{R}$ ,

$$(G([\phi, u], [\psi, m], \bar{H}))_i = \begin{cases} P([\phi_i^{(Q_b+1)}, u], \dots, [\phi_i^{(Q)}, u], [\psi_i^{(D_b+1)}, m], \dots, [\psi_i^{(D)}, m], \bar{H}) & \text{if } i \in \{1, \dots, M_\Omega\}, \\ B([\phi_i^{(1)}, u], \dots, [\phi_i^{(Q_b)}, u], [\psi_i^{(1)}, m], \dots, [\psi_i^{(D_b)}, m], \bar{H}) & \text{if } i \in \{M_\Omega+1, \dots, M\}. \end{cases} \quad (3.7)$$

Hence, we can rewrite (3.3) as

$$\begin{cases} \min_{(u, m, \bar{H}) \in \mathcal{U} \times \mathcal{V} \times \mathbb{R}} \|u\|_{\mathcal{U}}^2 + \|m\|_{\mathcal{V}}^2 + |\bar{H}|^2 + \beta \left| \frac{1}{M_\Omega} \sum_{i=1}^{M_\Omega} u(x_i) \right|^2 + \beta \left| \frac{1}{M_\Omega} \sum_{i=1}^{M_\Omega} m(x_i) - 1 \right|^2 \\ \text{s.t. } G([\phi, u], [\psi, m], \bar{H}) = 0. \end{cases} \quad (3.8)$$

**Remark 3.2.** Similarly, in time-dependent settings, we consider

$$\begin{cases} \min_{(u, m) \in \mathcal{U} \times \mathcal{V}} \|u\|_{\mathcal{U}}^2 + \|m\|_{\mathcal{V}}^2, \\ \text{s.t. } G([\phi, u], [\psi, m]) = 0, \end{cases} \quad (3.9)$$

where  $G$  is an analog to the one in (3.7)

The following theorem gives the foundation for the GP method to solve (3.8).

**Theorem 3.3.** Suppose that Assumption 1 holds. Let  $N_{\mathcal{U}}$ ,  $\phi$ ,  $N_{\mathcal{V}}$ ,  $\psi$ , and  $G$  be as in (3.5), (3.6), and (3.7). Define matrices  $\Theta \in \mathbb{R}^{N_{\mathcal{U}} \times N_{\mathcal{U}}}$ ,  $\Psi \in \mathbb{R}^{N_{\mathcal{V}} \times N_{\mathcal{V}}}$  such that

$$\Theta_{ij} = [\phi_i, \mathcal{K}_{\mathcal{U}} \phi_j], 1 \leq i, j \leq N_{\mathcal{U}}, \text{ and } \Psi_{ks} = [\psi_k, \mathcal{K}_{\mathcal{V}} \psi_s], 1 \leq k, s \leq N_{\mathcal{V}}.$$

Assume further that  $\Theta$  and  $\Psi$  are invertible. Let  $\chi = (\chi_1, \dots, \chi_{N_{\mathcal{U}}})$  and  $\eta = (\eta_1, \dots, \eta_{N_{\mathcal{V}}})$  be the vectors with elements

$$\chi_i = \sum_{n=1}^{N_{\mathcal{U}}} \Theta_{in}^{-1} \mathcal{K}_{\mathcal{U}} \phi_n, i = 1, \dots, N_{\mathcal{U}}, \text{ and } \eta_i = \sum_{n=1}^{N_{\mathcal{V}}} \Psi_{in}^{-1} \mathcal{K}_{\mathcal{V}} \psi_n, i = 1, \dots, N_{\mathcal{V}},$$

where  $\Theta_{in}^{-1}$  and  $\Psi_{in}^{-1}$  are the elements of  $\Theta^{-1}$  and  $\Psi^{-1}$  at the  $i$ th row and the  $n$ th column. Then,  $(u^\dagger, m^\dagger, \bar{H}^\dagger)$  is a solution to (3.8) if and only if

$$u^\dagger = \chi^T z^\dagger, m^\dagger = \eta^T \rho, \text{ and } \bar{H}^\dagger = \lambda, \quad (3.10)$$

where  $(z^\dagger, \rho^\dagger, \lambda)$  is a minimizer to

$$\begin{cases} \min_{z \in \mathbb{R}^{N_{\mathcal{U}}}, \rho \in \mathbb{R}^{N_{\mathcal{V}}}, \lambda \in \mathbb{R}} z^T \Theta^{-1} z + \rho^T \Psi^{-1} \rho + |\lambda|^2 + \beta \left| \frac{1}{M_{\Omega}} \sum_{i=1}^{M_{\Omega}} z_i^{(1)} \right|^2 + \beta \left| \frac{1}{M_{\Omega}} \sum_{i=1}^{M_{\Omega}} \rho_i^{(1)} - 1 \right|^2 \\ \text{s.t. } G(z, \rho, \lambda) = 0. \end{cases} \quad (3.11)$$

**Proof.** We conclude using similar arguments as in the proof of Theorem 2.3. Let  $(u^*, m^*, \bar{H}^*)$  be the solution to (3.1) and define  $z_* = [\phi, u^*]$ ,  $\rho_* = [\psi, m^*]$  and  $\lambda_* = \bar{H}^*$ . Then,  $G(z_*, \rho_*, \lambda_*) = 0$ . Thus, the minimization problem (3.11) can be restricted to the form of (2.13). Hence, (3.11) admits a minimizer. Following nearly identical steps to the derivation of (2.9), we conclude (3.10).  $\square$

Next, following [16], we have the convergence theorem.

**Theorem 3.4.** Assume that Assumption 1 holds and that the MFG in (3.1) has a unique classical solution  $(u^*, m^*, \bar{H}^*)$  in the space  $\mathcal{U} \times \mathcal{V} \times \mathbb{R}$ . Assume further that  $\mathcal{U} \subset \subset \mathcal{H}_1 \subset \subset C^1(\Omega) \cap C^{t_1}(\partial\Omega)$  and that  $\mathcal{V} \subset \subset \mathcal{H}_2 \subset \subset C^2(\Omega) \cap C^{t_2}(\partial\Omega)$ , where  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are Banach spaces and  $t_1, t_1', t_2$ , and  $t_2'$  are sufficiently large. Denote by  $\{x_i\}_{i=1}^M$  the collection of samples with  $M$  points. Suppose further that as  $M \rightarrow \infty$ ,

$$\sup_{x \in \text{int } \Omega} \min_{1 \leq i \leq M_{\Omega}} |x - x_i| \rightarrow 0 \text{ and } \sup_{x \in \partial\Omega} \min_{M_{\Omega}+1 \leq i \leq M} |x - x_i| \rightarrow 0. \quad (3.12)$$

Given  $M$  and  $\beta$ , let  $(u_{M,\beta}^\dagger, m_{M,\beta}^\dagger, \bar{H}_{M,\beta}^\dagger)$  be a minimizer of (3.8). Then, as  $M$  and  $\beta$  go to infinity, up to a sub-sequence,  $(u_{M,\beta}^\dagger, m_{M,\beta}^\dagger, \bar{H}_{M,\beta}^\dagger)$  converges to  $(u^*, m^*, \bar{H}^*)$  pointwisely in  $\Omega$  and in  $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathbb{R}$ .

**Proof.** The argument is similar to the proof of Theorem 2.4. Given  $M$  sample points and the penalization parameter  $\beta$ , denote by  $(u_{M,\beta}^\dagger, m_{M,\beta}^\dagger, \bar{H}_{M,\beta}^\dagger)$  a minimizer to (3.8). Using the fact that the classical solution  $(u^*, m^*, \bar{H}^*)$  to (3.1) satisfies the constraints in (3.8), we prove that there exists a sequence  $\{(M_p, \beta_p)\}_{p=1}^\infty$  such that  $\|u_{M_p,\beta_p}^\dagger\|_{\mathcal{U}}$ ,  $\|m_{M_p,\beta_p}^\dagger\|_{\mathcal{V}}$ , and  $|\bar{H}_{M_p,\beta_p}|$  are uniformly bounded for all  $p \geq 1$ , and that  $\int_{\Omega} m_{M_p,\beta_p}^\dagger dx \rightarrow 1$  and  $\int_{\Omega} u_{M_p,\beta_p}^\dagger dx \rightarrow 0$ . Since  $\mathcal{U} \subset \subset \mathcal{H}_1$ ,  $\mathcal{V} \subset \subset \mathcal{H}_2$ , and a closed bounded set in  $\mathbb{R}$  is compact, the sequence  $(u_{M_p,\beta_p}^\dagger, m_{M_p,\beta_p}^\dagger, \bar{H}_{M_p,\beta_p}^\dagger)$  converge, up to a sub-sequence, to a limit  $(u_\infty^\dagger, m_\infty^\dagger, \bar{H}_\infty^\dagger)$  in  $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathbb{R}$  as  $p \rightarrow \infty$ . Using  $\mathcal{H}_1 \subset \subset C^1(\Omega) \cap C^{t_1'}(\partial\Omega)$  and that  $\mathcal{H}_2 \subset \subset C^2(\Omega) \cap C^{t_2'}(\partial\Omega)$ , we conclude that  $(u_\infty^\dagger, m_\infty^\dagger, \bar{H}_\infty^\dagger)$  satisfies the constraints in (3.3) at all points in  $\{x_i\}_{i=1}^M$ . Due to (3.12),  $\{x_i\}_{i=1}^M$  is dense in  $\Omega$  as  $M \rightarrow \infty$ . Hence,  $(u_\infty^\dagger, m_\infty^\dagger, \bar{H}_\infty^\dagger)$  solves (3.1). Since the solution to (3.1) is unique, we conclude that  $(u_\infty^\dagger, m_\infty^\dagger, \bar{H}_\infty^\dagger) = (u^*, m^*, \bar{H}^*)$ .  $\square$

**Remark 3.5.** To deal with the nonlinear constraints of (3.11), we introduce two penalization parameters,  $\gamma > 0$  and  $\beta > 0$ , and consider the relaxation formulation similar to (2.10).

**Remark 3.6.** Under similar Assumptions of Theorems 3.3 and 3.4, (3.9) admits a solution  $(u_M^\dagger, m_M^\dagger)$  for given sample size  $M$ , and  $(u_M^\dagger, m_M^\dagger)$  converges to the classical solution of (3.2) as  $M$  goes to infinity.

### 3.3. The Fourier features method

Let  $\mathcal{O} = \{\zeta(\cdot, \omega) | \omega \in \mathcal{W}\}$  be a family of base functions parametrized over the set  $\mathcal{W}$ , where

$$\sup_{x \in \Omega} |\zeta(x, \omega)| \leq 1, \forall \omega \in \mathcal{W} \text{ and } \int_{\Omega} \zeta^2(x, \omega) dx < \infty. \quad (3.13)$$

We propose to approximate  $u^*$  and  $m^*$  in the solution of (3.1) by linear combinations of functions sampled from  $\mathcal{O}$ . More precisely, given  $N \in \mathbb{N}$ , we take samples  $\{\omega_i\}_{i=1}^N$  and  $\{\tau_i\}_{i=1}^N$  from  $\mathcal{W}$  and define vector valued functions  $\zeta^N$  and  $\vartheta^N$  by

$$\zeta^N(\cdot) = [\zeta(\cdot; \omega_1), \dots, \zeta(\cdot; \omega_N)]^T \text{ and } \vartheta^N(\cdot) = [\zeta(\cdot; \tau_1), \dots, \zeta(\cdot; \tau_N)]^T. \quad (3.14)$$

Then, we define spaces

$$\mathcal{G}_{\zeta^N} = \{\alpha^T \zeta^N | \alpha \in \mathbb{R}^N\} \text{ and } \mathcal{G}_{\vartheta^N} = \{\beta^T \vartheta^N | \beta \in \mathbb{R}^N\}. \quad (3.15)$$

Meanwhile, for  $\zeta$  satisfying (3.13), we equip the spaces  $\mathcal{G}_{\zeta^N}$  and  $\mathcal{G}_{\vartheta^N}$  with the norms

$$\|u_{\alpha}\|_{\mathcal{G}_{\zeta^N}} = |\alpha|, \forall u_{\alpha} = \alpha^T \zeta^N \in \mathcal{G}_{\zeta^N} \text{ and } \|m_{\beta}\|_{\mathcal{G}_{\vartheta^N}} = |\beta|, \forall m_{\beta} = \beta^T \vartheta^N \in \mathcal{G}_{\vartheta^N}.$$

By (3.13), the norms  $\|\cdot\|_{\mathcal{G}_{\zeta^N}}$  and  $\|\cdot\|_{L^2(\Omega)}$  are equivalent for the space  $\mathcal{G}_{\zeta^N}$ . Then, we approximate the solution of (3.1) by the minimizer of the following problem

$$\begin{aligned} \min_{u_{\alpha} \in \mathcal{G}_{\zeta^N}, m_{\beta} \in \mathcal{G}_{\vartheta^N}, \bar{H} \in \mathbb{R}} & \|u_{\alpha}\|_{L^2(\Omega)}^2 + \|m_{\beta}\|_{L^2(\Omega)}^2 + \gamma \|\mathcal{P}(u_{\alpha}, m_{\beta}, \bar{H})\|_{L^2(\Omega)}^2 \\ & + \gamma \|\mathcal{B}(u_{\alpha}, m_{\beta})\|_{L^2(\Omega)}^2 + \gamma \left| \int_{\Omega} m_{\beta} dx - 1 \right|^2 + \gamma \left| \int_{\Omega} u_{\alpha} dx \right|^2, \end{aligned} \quad (3.16)$$

where  $\mathcal{P}$  and  $\mathcal{B}$  are given in Assumption 1, and  $\gamma > 0$ .

**Remark 3.7.** When the domain  $\Omega = \mathbb{T}^d$ , we choose  $\mathcal{O}$  such that

$$\mathcal{O} = \{\zeta(\cdot, \omega) = \cos(2\pi \langle a, \cdot \rangle + b) | \omega = (a, b) \in \mathbb{Z}^d \times \{0, \pi/2\}\}.$$

Then, given  $N \in \mathbb{N}$ , we select a subset  $\{\omega_i = (a_i, b_i)\}_{i=1}^N$  in  $\mathbb{Z}^d \times \{0, \pi/2\}$ . Thus, the vector valued function  $\zeta^N$  in (3.14) becomes

$$\zeta^N(\cdot) = [\cos(2\pi \langle a_1, \cdot \rangle + b_1), \dots, \cos(2\pi \langle a_N, \cdot \rangle + b_N)]^T.$$

Hence, for  $\alpha = \{\alpha_1, \dots, \alpha_N\} \in \mathbb{R}^N$ , the function  $u_{\alpha} \in \mathcal{G}_{\zeta^N}$  in (3.15) has the following form

$$u_{\alpha}(x) = \sum_{i=1}^N \alpha_i \cos(2\pi a_i^T x + b_i).$$

The same construction holds for the space  $\mathcal{G}_{\vartheta^N}$ .

**Remark 3.8.** When the domain  $\Omega$  is non-periodic, we choose

$$\mathcal{O} = \{\zeta(\cdot, \omega) = \cos(2\pi \langle \omega, \cdot \rangle) | \omega \in \mathbb{R}^d\} \cup \{\zeta(\cdot, \omega) = \sin(2\pi \langle \omega, \cdot \rangle) | \omega \in \mathbb{R}^d\}.$$

Then, we take  $N/2$ ,  $N \in \mathbb{N}$  and  $N$  is even, samples  $\{\omega_i\}_{i=1}^{N/2}$  from  $\mathbb{R}^d$  in such a way that the nonlinear map  $\zeta^N$  in (3.14) is defined as

$$\zeta^N(x) = \sqrt{\frac{2}{N}} [\sin(\omega_1^T x), \dots, \sin(\omega_{N/2}^T x), \cos(\omega_1^T x), \dots, \cos(\omega_{N/2}^T x)]^T, \forall x \in \mathbb{R}^d.$$

To sample  $\{\omega_i\}_{i=1}^{N/2}$ , we use the method of orthogonal random features [48]. More precisely, the matrix  $W = [\omega_1, \dots, \omega_{N/2}]^T$  satisfies

$$W = \frac{1}{\varsigma} S Q, \quad (3.17)$$

where  $\varsigma > 0$ ,  $Q$  is a uniformly distributed random orthogonal matrix, and  $S$  is a diagonal matrix with entries sampled i.i.d. from  $\chi$ -distribution with  $d$  degrees of freedom. We refer readers to [48] for more details about the construction of  $W$ .

Then, given  $\alpha = \{\alpha_1, \dots, \alpha_N\} \in \mathbb{R}^N$  and  $u_\alpha \in \mathcal{G}_{\varsigma^N}$  in (3.15), we obtain

$$u_\alpha(x) = \sum_{i=1}^{N/2} \alpha_i \sin(\omega_i^T x) + \sum_{i=N/2+1}^N \alpha_i \cos(\omega_i^T x).$$

We use the same method to build the space  $\mathcal{G}_{\vartheta^N}$ .

**Remark 3.9.** In a time-dependent setting, the space-time domain is non-periodic. Hence, we use features given in Remark 3.8 and approximate the solution to (3.2) by the minimizer of the following problem

$$\min_{u_\alpha \in \mathcal{G}_{\varsigma^N}, m_\beta \in \mathcal{G}_{\vartheta^N}, \bar{H} \in \mathbb{R}} \|u_\alpha\|_{L^2(\Omega)}^2 + \|m_\beta\|_{L^2(\Omega)}^2 + \gamma \|\mathcal{P}(u_\alpha, m_\beta, \bar{H})\|_{L^2(\Omega)}^2 + \gamma \|\mathcal{B}(u_\alpha, m_\beta)\|_{L^2(\Omega)}^2.$$

A numerical example for a time-dependent planning MFG is shown in Section 4.

The following theorem gives the existence of a solution to (3.16).

**Theorem 3.10.** Under Assumption 1, the minimization problem (3.16) admits a minimizer for any given  $\gamma \geq 0$ .

**Proof.** The arguments are the same as in the proof of Theorem 2.5. First, we convert (3.16) into an equivalent minimization problem similar to (2.27). Then, from the lower semi-continuity and the coercivity of the objective function, we conclude that a minimizer exists.  $\square$

Next, we study the convergence of (3.16) as  $N$  and  $\gamma$  go to infinity. We do not try to provide the most general convergence result for all MFGs since the spaces in (3.15) lack compactness. Hence, we cannot apply the arguments of Theorem 3.4 here. Instead, we prove the validity of our method in concrete setups of interests. In the rest of this subsection, we build the convergence results of the FF method applied to a stationary MFG with a Lipschitz coupling and a unique smooth solution, see Subsection 4.1 for a numerical experiment. We postpone the study of the convergence of the FF method in settings of time-dependent MFGs to future work.

More precisely, given a smooth function  $V : \mathbb{R}^d \mapsto \mathbb{R}$  and a functional  $F : C(\mathbb{T}^d) \mapsto C(\mathbb{T}^d)$ , we consider the following MFG

$$\begin{cases} -\Delta u + \frac{|\nabla u|^2}{2} + V(x) = F[m](x) + \bar{H}, & \text{in } \mathbb{T}^d, \\ -\Delta m - \operatorname{div}(m \nabla u) = 0, & \text{in } \mathbb{T}^d, \\ \int_{\mathbb{T}^d} m \, dx = 1, \quad \int_{\mathbb{T}^d} u \, dx = 0. \end{cases} \quad (3.18)$$

We assume that (3.18) admits a smooth solution. In addition, we suppose further that (3.18) satisfies the following assumption, which guarantees the uniqueness of the solution to (3.18).

**Assumption 2.** There exists a constant  $L_F > 0$  such that for any  $m, \mu \in C(\mathbb{T}^d)$ ,

$$\sup_{x \in \mathbb{T}^d} |F[m](x) - F[\mu](x)| \leq L_F \|m - \mu\|_{L^\infty(\mathbb{T}^d)}. \quad (3.19)$$

Meanwhile,  $F$  is monotone, i.e., for any  $m, \mu \in C(\mathbb{T}^d)$ ,  $m = \mu$  if and only if

$$\int_{\mathbb{T}^d} (F[m](x) - F[\mu](x))(m(x) - \mu(x)) \, dx \leq 0.$$

Next, we study the convergence of our method when we use the Fourier series to approximate  $u$  and  $m$ . Given  $N \in \mathbb{N}$ , we define the Fourier features space

$$\mathcal{G}^N = \left\{ \phi \left| \phi(x) = c + \sum_{i \in \mathbb{Z}_N^d} \alpha_i \sin(2\pi i^T x) + \sum_{i \in \mathbb{Z}_N^d} \beta_i \cos(2\pi i^T x), c, \alpha_i, \beta_i \in \mathbb{R} \right. \right\} \quad (3.20)$$

and equip  $\mathcal{G}^N$  with the norm  $\|\cdot\|_{\mathcal{G}^N}$  defined as

$$\|\phi\|_{\mathcal{G}^N} = |c|^2 + \sum_{i \in \mathbb{Z}_N^d} |\alpha_i|^2 + |\beta_i|^2, \forall \phi \in \mathcal{G}^N.$$

For ease of presentation, given  $\gamma > 0$ , we define the following functional

$$J_\gamma(u, m, \bar{H}) = \int_{\mathbb{T}^d} |u|^2 dx + \int_{\mathbb{T}^d} |m|^2 dx + \bar{H}^2 + \gamma \mathcal{Q}(u, m, \bar{H}), \quad (3.21)$$

where

$$\begin{aligned} \mathcal{Q}(u, m, \bar{H}) = & \int_{\mathbb{T}^d} \left| -\Delta u + \frac{|\nabla u|^2}{2} + V(x) - F[m](x) - \bar{H} \right|^2 dx \\ & + \int_{\mathbb{T}^d} \left| -\Delta m - \operatorname{div}(m \nabla u) \right|^2 dx + \left| \int_{\mathbb{T}^d} u dx \right|^2 + \left| \int_{\mathbb{T}^d} m dx - 1 \right|^2. \end{aligned} \quad (3.22)$$

Then, (3.16) is equivalent to

$$\min_{u^N \in \mathcal{G}^N, m^N \in \mathcal{G}^N, \bar{H}^N \in \mathbb{R}} J_\gamma(u^N, m^N, \bar{H}^N). \quad (3.23)$$

Using the same arguments as in the proof of Theorem 3.10, we conclude that (3.23) admits a minimizer. Next, we show the convergence of a minimizer of (3.23). First, we give an upper bound for the minimum of (3.23) in the following theorem.

**Theorem 3.11.** *Let  $(u^*, m^*, \bar{H}^*)$  be the solution to (3.18) and the functional  $J_\gamma$  be as in (3.21) for  $\gamma > 0$ . Suppose that Assumption 2 holds. Then, for any sufficiently small  $\varepsilon > 0$ , there exist a constant  $C > 0$ ,  $N \in \mathbb{N}$ , and functions  $(u^N, m^N) \in \mathcal{G}^N \times \mathcal{G}^N$  such that*

$$J_\gamma(u^N, m^N, \bar{H}^*) \leq 2\|u^*\|_{L^2(\mathbb{T}^d)}^2 + 2\|m^*\|_{L^2(\mathbb{T}^d)}^2 + |\bar{H}^*|^2 + C(1 + \gamma)\varepsilon^2. \quad (3.24)$$

See Appendix A for the proof of the above theorem. The following corollary follows directly from Theorem 3.11 and proves that there exists a minimizer  $(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma})$  of (3.23) such that  $\mathcal{Q}(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma}) \rightarrow 0$  as  $N$  and  $\gamma$  go to infinity.

**Corollary 3.12.** *Let  $\mathcal{Q}$  be as in (3.22). Under the same assumptions of Theorem 3.11, for any  $\varepsilon > 0$ , there exist a constant  $C > 0$ , sufficiently large  $\gamma > 0$  and  $N > 0$ , and a minimizer of (3.23), which is denoted by  $(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma}) \in \mathcal{G}^N \times \mathcal{G}^N \times \mathbb{R}$ , satisfies*

$$\mathcal{Q}(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma}) \leq C\varepsilon.$$

We give the proof of Corollary 3.12 in Appendix A. The following theorem shows the convergence of minimizers of (3.23) to the solution of (3.18) as  $N$  and  $\gamma$  go to infinity. The proof is presented in Appendix A.

**Theorem 3.13.** *Let  $(u^*, m^*, \bar{H}^*)$  be the solution to (3.18). Under the assumptions of Theorem 3.11, there exists a sequence  $\{(N_i, \gamma_i)\}_{i=1}^\infty$  such that the sequence  $\{(u^i, m^i, \bar{H}^i)\}_{i=1}^\infty$ , where  $(u^i, m^i, \bar{H}^i)$  is a minimizer of (3.23) corresponding to  $N_i$  and  $\gamma_i$ , satisfies  $u^i \rightarrow u^*$  in  $H^1(\mathbb{T}^d)$ ,  $m^i \rightarrow m^*$  in  $H^1(\mathbb{T}^d)$ , and  $\bar{H}^i \rightarrow \bar{H}^*$  in  $\mathbb{R}$  as  $i \rightarrow \infty$ .*

Next, we propose a method to solve (3.16). We take  $M$  samples  $\{x_i\}_{i=1}^M$  in the domain  $\Omega$  such that  $\{x_i\}_{i=1}^{M_\Omega} \subset \operatorname{int} \Omega$  and  $\{x_i\}_{i=M_\Omega+1}^M \subset \partial \Omega$  for  $1 \leq M_\Omega \leq M$ . To capture different variability of the constraints, we introduce two penalization parameters  $\gamma, \beta$ , and consider

$$\begin{aligned} \min_{u_\alpha \in \mathcal{G}_{\zeta N}, m_\beta \in \mathcal{G}_{\theta N}, \bar{H} \in \mathbb{R}} & \|u_\alpha\|_{\mathcal{G}_{\zeta N}}^2 + \|m_\beta\|_{\mathcal{G}_{\theta N}}^2 + |\bar{H}|^2 + \beta \left| \frac{1}{M_\Omega} \sum_{i=1}^{M_\Omega} m_\beta(x_i) - 1 \right|^2 \\ & + \beta \left| \frac{1}{M_\Omega} \sum_{i=1}^{M_\Omega} u_\alpha(x_i) \right|^2 + \gamma \sum_{i=1}^{M_\Omega} |\mathcal{P}(u_\alpha, m_\beta, \bar{H})(x_i)|^2 + \gamma \sum_{i=M_\Omega+1}^M |\mathcal{B}(u_\alpha, m_\beta)(x_i)|^2. \end{aligned} \quad (3.25)$$

Let  $N_{\mathcal{U}}$  and  $N_{\mathcal{V}}$  be as in (3.5) and (3.6). Under Assumption 1, we reformulate (3.25) into an equivalent two-level minimization problem

$$\begin{aligned} \min_{\mathbf{z} \in \mathbb{R}^{N_{\mathcal{U}}}, \boldsymbol{\rho} \in \mathbb{R}^{N_{\mathcal{V}}}, \lambda \in \mathbb{R}} \quad & \gamma \|G(\mathbf{z}, \boldsymbol{\rho}, \lambda)\|^2 + \beta \left| \frac{1}{M_{\Omega}} \sum_{i=1}^{M_{\Omega}} \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M_{\Omega}} \sum_{i=1}^{M_{\Omega}} z_i^{(1)} \right|^2 \\ & + \begin{cases} \min_{\boldsymbol{\alpha} \in \mathbb{R}^N, \boldsymbol{\beta} \in \mathbb{R}^N, \bar{H} \in \mathbb{R}} \|\boldsymbol{\alpha}\|^2 + \|\boldsymbol{\beta}\|^2 + \bar{H}^2, \\ \text{s.t. } [\boldsymbol{\phi}, \boldsymbol{\zeta}^N] \boldsymbol{\alpha} = \mathbf{z}, [\boldsymbol{\psi}, \boldsymbol{\vartheta}^N] \boldsymbol{\beta} = \boldsymbol{\rho}, \bar{H} = \lambda. \end{cases} \end{aligned} \quad (3.26)$$

The first level optimization problem gives

$$\begin{cases} \boldsymbol{\alpha} = [\boldsymbol{\phi}, \boldsymbol{\zeta}^N]^T ([\boldsymbol{\phi}, \boldsymbol{\zeta}^N][\boldsymbol{\phi}, \boldsymbol{\zeta}^N]^T)^{-1} \mathbf{z}, \\ \boldsymbol{\beta} = [\boldsymbol{\psi}, \boldsymbol{\vartheta}^N]^T ([\boldsymbol{\psi}, \boldsymbol{\vartheta}^N][\boldsymbol{\psi}, \boldsymbol{\vartheta}^N]^T)^{-1} \boldsymbol{\rho}, \\ \bar{H} = \lambda. \end{cases}$$

Hence, (3.26) is equivalent to

$$\begin{aligned} \min_{\mathbf{z} \in \mathbb{R}^{N_{\mathcal{U}}}, \boldsymbol{\rho} \in \mathbb{R}^{N_{\mathcal{V}}}, \lambda \in \mathbb{R}} \quad & \gamma \|G(\mathbf{z}, \boldsymbol{\rho}, \lambda)\|^2 + \beta \left| \frac{1}{M_{\Omega}} \sum_{i=1}^{M_{\Omega}} \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M_{\Omega}} \sum_{i=1}^{M_{\Omega}} z_i^{(1)} \right|^2 \\ & + \mathbf{z}^T ([\boldsymbol{\phi}, \boldsymbol{\zeta}^N][\boldsymbol{\phi}, \boldsymbol{\zeta}^N]^T)^{-1} \mathbf{z} + \boldsymbol{\rho}^T ([\boldsymbol{\psi}, \boldsymbol{\vartheta}^N][\boldsymbol{\psi}, \boldsymbol{\vartheta}^N]^T)^{-1} \boldsymbol{\rho} + \lambda^2. \end{aligned}$$

**Remark 3.14.** In general, the matrices  $[\boldsymbol{\phi}, \boldsymbol{\zeta}^N][\boldsymbol{\phi}, \boldsymbol{\zeta}^N]^T$  and  $[\boldsymbol{\psi}, \boldsymbol{\vartheta}^N][\boldsymbol{\psi}, \boldsymbol{\vartheta}^N]^T$  are ill-conditioned. Hence, we choose two regularization parameters,  $\mu_1 > 0$  and  $\mu_2 > 0$ , and consider

$$\begin{aligned} \min_{\mathbf{z} \in \mathbb{R}^{N_{\mathcal{U}}}, \boldsymbol{\rho} \in \mathbb{R}^{N_{\mathcal{V}}}, \lambda \in \mathbb{R}} \quad & \gamma \|G(\mathbf{z}, \boldsymbol{\rho}, \lambda)\|^2 + \beta \left| \frac{1}{M_{\Omega}} \sum_{i=1}^{M_{\Omega}} \rho_i^{(1)} - 1 \right|^2 + \beta \left| \frac{1}{M_{\Omega}} \sum_{i=1}^{M_{\Omega}} z_i^{(1)} \right|^2 \\ & + \mathbf{z}^T ([\boldsymbol{\phi}, \boldsymbol{\zeta}^N][\boldsymbol{\phi}, \boldsymbol{\zeta}^N]^T + \mu_1 I)^{-1} \mathbf{z} \\ & + \boldsymbol{\rho}^T ([\boldsymbol{\psi}, \boldsymbol{\vartheta}^N][\boldsymbol{\psi}, \boldsymbol{\vartheta}^N]^T + \mu_2 I)^{-1} \boldsymbol{\rho} + \lambda^2. \end{aligned} \quad (3.27)$$

**Remark 3.15.** When necessary, we also eliminate equality constraints in (3.27) as discussed in Section 3.3 of [16].

**Remark 3.16.** According to Theorems 3.4 and 3.13, using the GP and the FF methods, the non-negativity of the probability measure  $m$  is guaranteed at the limit. Thus, unless the coupling term is not well defined when  $m$  is non-positive (see Remark 2.1 for discussions about a MFG with a log coupling), we do not impose extra non-negativity constraints on the Gauss–Newton iterations. We will also see that the numerical results of the probability measures are non-negative in the next section.

#### 4. Numerical results

In this section, we implement our methods to solve a non-local stationary MFG in Subsection 4.1 and a time-dependent planning problem in Subsection 4.2. The runtimes are measured using MacBook Air 2015 (4 GB Ram, Intel Core i5 CPU). Our implementation is based on the code of [16],<sup>1</sup> which uses Python with the JAX package for automatic differentiation. Our experiments count only on CPUs. Additional speedups can be achieved by using accelerated hardware such as Graphics Processing Units (GPU).

##### 4.1. A second-order non-local stationary MFG

We consider a variant of the non-local stationary MFG in Subsection 6.2.5 of [2]. More precisely, given  $\nu > 0$ , we want to find  $(u, m, \bar{H})$  solving

$$\begin{cases} -\nu \Delta u + H(x, \nabla u) = ((1 - \Delta)^{-1} (1 - \Delta)^{-1} m)(x) + \bar{H}, & \text{in } \mathbb{T}^2, \\ -\nu \Delta m - \operatorname{div}(m D_p H(x, \nabla u)) = 0, & \text{in } \mathbb{T}^2, \\ \int_{\mathbb{T}^2} u \, dx = 0, \int_{\mathbb{T}^2} m \, dx = 1, \end{cases} \quad (4.1)$$

<sup>1</sup> <https://github.com/yifanc96/NonlinearPDEs-GPsolver.git>.

**Table 3**

The precomputation time (in seconds) consumed by the GP method and the FF algorithm for solving (4.1) when  $\nu = 0.1$  and  $N = 10$  as  $M$  increases.

| $M$  | GP       | FF   |          |
|------|----------|------|----------|
|      | Cholesky | QR   | Cholesky |
| 400  | 1.29     | 0.35 | 0.31     |
| 800  | 6.31     | 0.50 | 0.21     |
| 1200 | 38.66    | 1.09 | 0.23     |
| 1600 | 88.79    | 1.92 | 0.30     |

where

$$H(x, p) = \sin(2\pi x_1) + \sin(2\pi x_2) + \cos(4\pi x_1) + |p|^2, \forall x \in \mathbb{T}^2, p \in \mathbb{R}^2.$$

We use the GP method and the FF algorithm proposed in Section 3 to solve (4.1). In the experiments, we write (4.1) in the form of (3.4) with  $Q_b = D_b = 0$ ,  $Q = 4$ ,  $D = 5$ , and with linear operators  $L_1(u) = u$ ,  $L_2(u) = \partial_x u$ ,  $L_3(u) = \partial_y u$ ,  $L_4(u) = \Delta u$ ,  $J_1(m) = m$ ,  $J_2(m) = \partial_x m$ ,  $J_3(m) = \partial_y m$ ,  $J_4(m) = \Delta m$ , and  $J_5(m) = (1 - \Delta)^{-1}(1 - \Delta)^{-1}m$ . We compute the action of  $J_5$  on a kernel by the Fast Fourier transform. The GP method uses the periodic kernels

$$K_1((x_1, x_2), (y_1, y_2)) = K_2(x_1, x_2, y_1, y_2) = e^{\frac{1}{\sigma^2}(\cos(2\pi(x_1 - y_1)) + \cos(2\pi(x_2 - y_2)) - 2)}$$

with lengthscale  $\sigma = 0.2$ . For the FF method, we fix  $N \in \mathbb{N}$ , use the basis

$$\begin{aligned} \xi^N(x_1, x_2) &= \vartheta^N(x_1, x_2) = [1, \sin(2\pi x_1 + 2\pi x_2), \dots, \sin(2\pi i x_1 + 2\pi j x_2), \\ &\dots, \sin(2\pi N x_1 + 2\pi N x_2), \cos(2\pi x_1 + 2\pi x_2), \\ &\dots, \cos(2\pi i x_1 + 2\pi j x_2), \dots, \cos(2\pi N x_1 + 2\pi N x_2)]^T, \forall i, j = 1, \dots, N, \end{aligned}$$

and approximate  $u$  and  $m$  by functions in the space  $\mathcal{G}_N$  as in (3.15). We choose the regularization parameters  $\eta_1 = \eta_2 = \mu_1 = \mu_2 = 10^{-5}$  in Remarks 2.2 and 2.9. To measure the accuracy of our algorithms, we identify  $\mathbb{T}^2$  with  $[0, 1) \times [0, 1)$ , discretize the domain with  $100 \times 100$  uniformly distributed grid points, and use the FD method in [2] to solve (4.1) with high accuracy. The GP method and the FF algorithm use the same sample points and start from the same initial values. We denote by  $(u_{GP}, m_{GP}, \bar{H}_{GP})$  and  $(u_{FF}, m_{FF}, \bar{H}_{FF})$  the numerical solutions of the GP method and the FF algorithm, separately.

In Fig. 2, we show the numerical results of both algorithms for  $\nu = 0.1$  after 36 iterations. We take the same  $M = 400$  samples for both the GP method and the FF algorithm, which is shown in Fig. 2d. We use the Gauss–Newton iteration with step size 1 to solve the optimization problems, and choose  $N = 10$  for the FF method. We set the penalization parameters  $\beta = 1$  and  $\gamma = 10^{-15}$  for both algorithms. Figs. 2b, 2e, 2c, and 2f plot the graphs of  $u_{GP}$ ,  $u_{FF}$ ,  $m_{GP}$ , and  $m_{FF}$ , separately. The convergence histories of the Gauss–Newton iterations are presented in Figs. 2g and 2j, which verify the convergence of our algorithms. We attribute the non-monotone decreasing of the loss curves to the non-linearity of the objective functions. The contours of pointwise errors are shown in Figs. 2h, 2i, 2k, and 2l. We see that the errors are smaller in smoother areas. The errors of  $\bar{H}_{GP}$  and  $\bar{H}_{FF}$  are given in Fig. 2a.

For larger values of  $\nu$ , the solutions are smoother. Then, fewer bases are enough for the FF method to achieve higher accuracy, which is shown in Fig. 3. We take  $M = 400$ ,  $\nu = 1$ , and  $N = 2$ . Meanwhile, we set the penalization parameters  $\beta = 1$ ,  $\gamma = 10^{-15}$  for the FF method and  $\beta = 10^{-2}$ ,  $\gamma = 10^{-15}$  for the GP algorithm. The Gauss–Newton method uses step size 1 for both methods and stops after 5 iterations. Figs. 2 and 3 imply that the selection of parameters depends on the data of the model and suggest the need to study hyperparameter learning in future work.

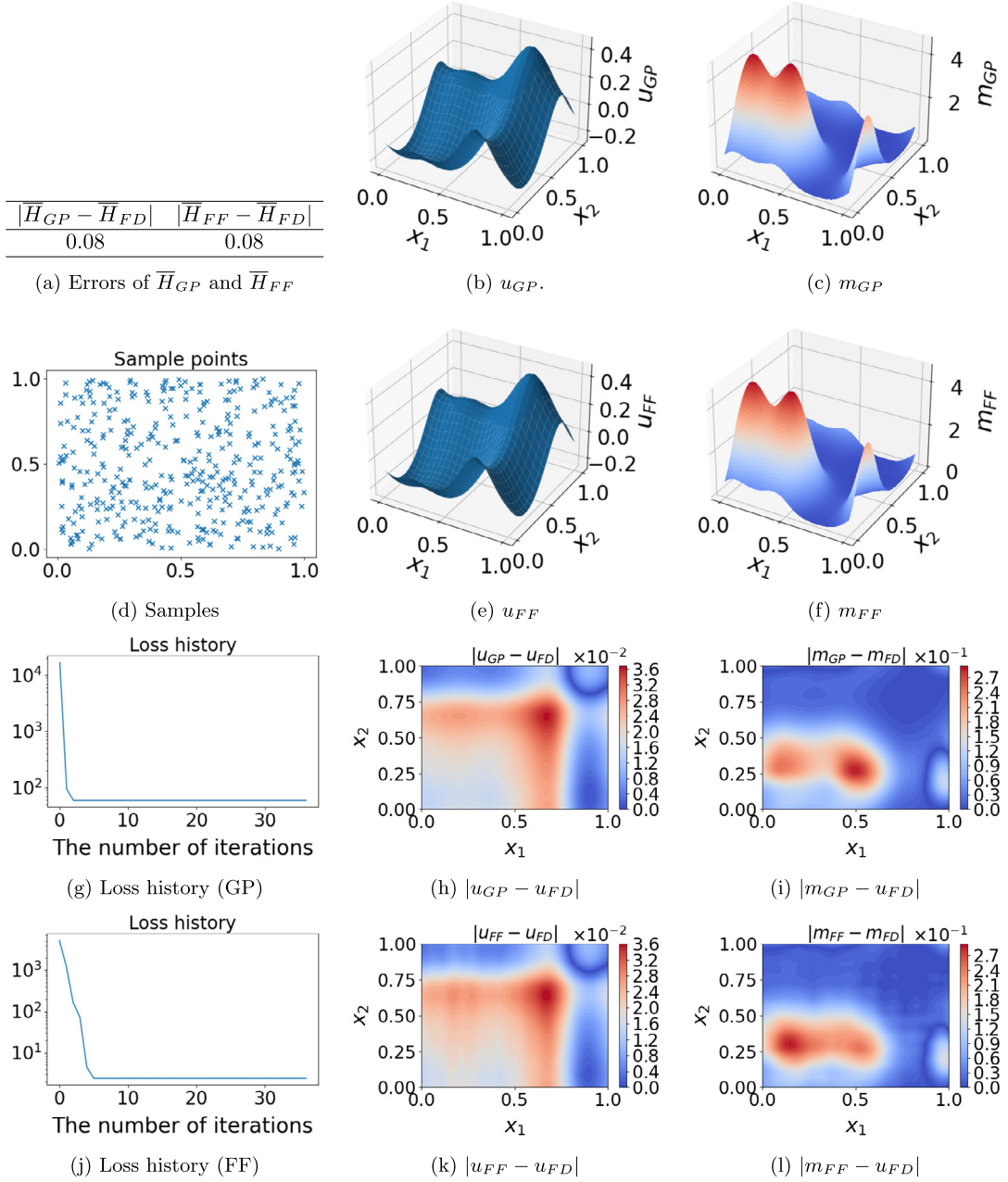
Table 3 records the CPU time of performing the Cholesky and the QR decomposition for both algorithms as  $M$  increases. We set  $\nu = 0.1$  and  $N = 10$ . We see that the FF algorithm outperforms the GP method in the precomputation stage.

#### 4.2. A planning problem

We consider a planning MFG, a variant of the crowd motion model given in [45]. Let  $\rho_G(\cdot, \mu_0, \sigma_0)$  be the probability density function of a one-dimensional Gaussian with mean  $\mu_0 \in \mathbb{R}$  and standard variance  $\sigma_0$ . We seek  $(u, m)$  solving

$$\begin{cases} -\frac{\partial u}{\partial t} + \frac{1}{2}|u_x|^2 = 0.01m, & \text{in } (0, 1) \times \mathbb{R}, \\ \frac{\partial m}{\partial t} - (mu_x)_x = 0, & \text{in } (0, 1) \times \mathbb{R}, \\ m(0, x) = \rho_0(x), m(1, x) = \rho_1(x), & \text{in } \mathbb{R}, \end{cases} \quad (4.2)$$

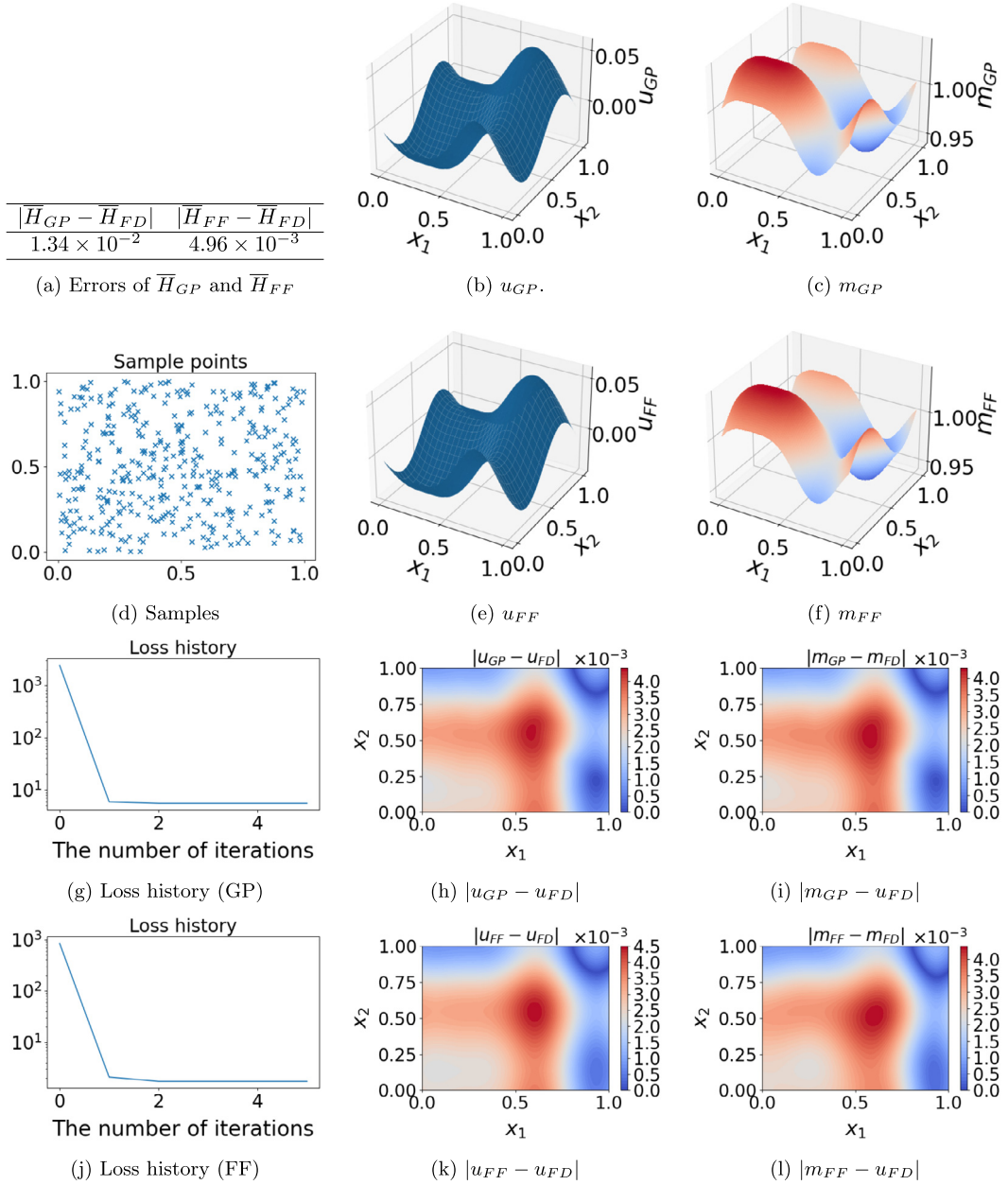
where  $\rho_0(x) = \rho_G(x, 0.5, 0.1)$  and  $\rho_1(x) = \rho_G(x, -0.5, 0.1)$ . Since the time and the space domains have different variability, following [16], we use the anisotropic kernel



**Fig. 2.** The numerical solutions of the non-local stationary MFG (4.1) when  $\nu = 0.1$ . (a) The errors of  $\bar{H}_{GP}$  and  $\bar{H}_{FF}$  with respect to the reference  $\bar{H}_{FD}$ ; (b), (c), (e), (f) Numerical values of  $u_{GP}$ ,  $m_{GP}$ ,  $u_{FF}$ , and  $m_{FF}$ ; (d) the samples used by the GP and the FF methods; (g), (j) the evolution of the losses, which are the objective functions in (3.11) and (3.27); (h), (i), (k), (l) The contours of the pointwise errors of  $u_{GP}$ ,  $m_{GP}$ ,  $u_{FF}$ , and  $m_{FF}$ .

$$\kappa((t, s), (t', s')) = \exp(-(s - s')^2 / \sigma_1^2 - (t - t')^2 / \sigma_2^2)$$

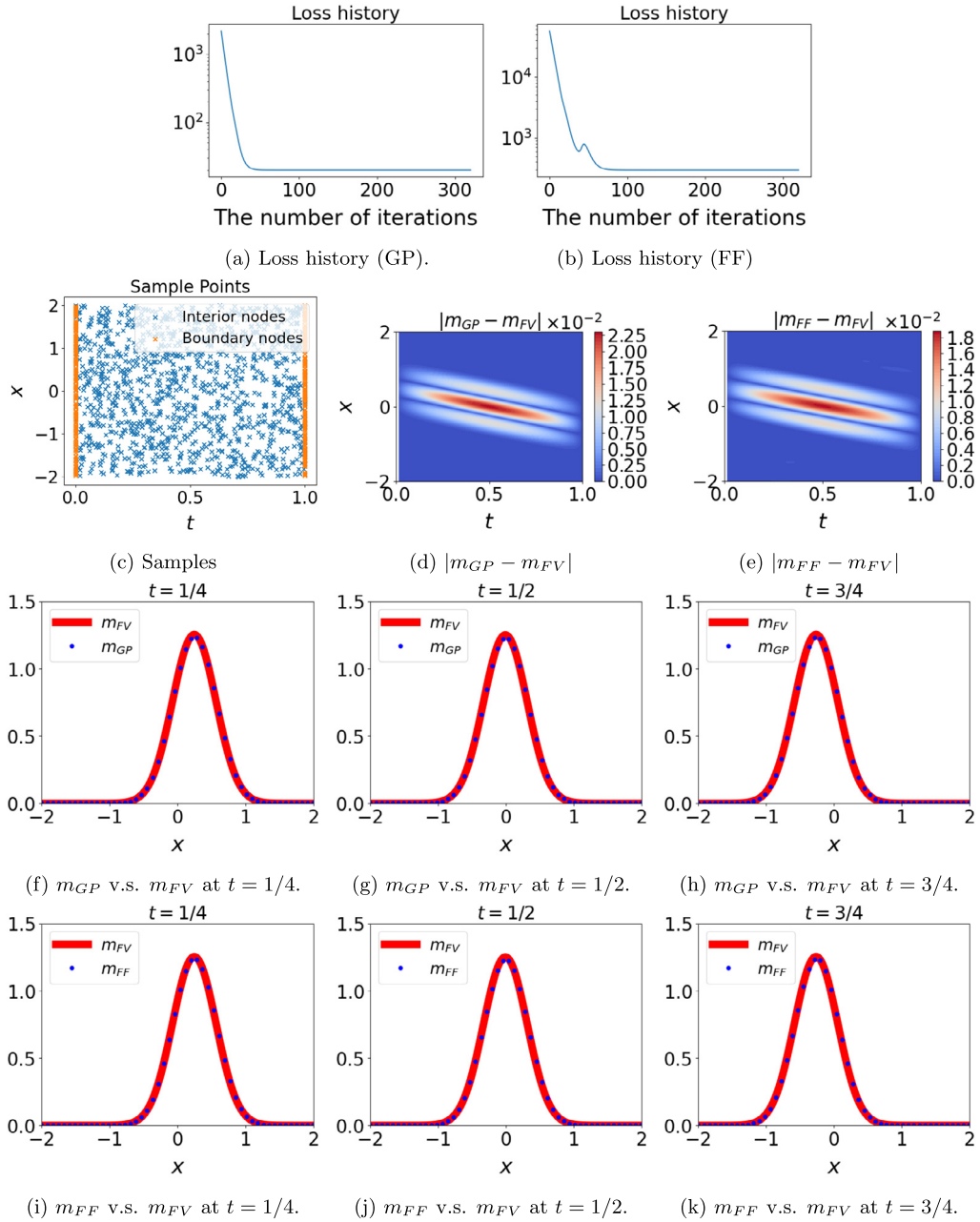
for the GP method, where  $(\sigma_1, \sigma_2) = (1/\sqrt{5}, 1/\sqrt{2})$ . We choose the regularization parameters  $\eta_1 = \eta_2 = \mu_1 = \mu_2 = 10^{-5}$  in Remarks 2.2 and 2.9. For the FF method, we use orthogonal random Fourier features stated in Remark 3.8. In (3.17), we choose  $\zeta = 0.2$  and select 200 random Fourier features both for approximating  $u$  and  $m$ . We write (4.2) in an analog form of (3.4) with  $Q_b = D_b = 0$ ,  $Q = 4$ ,  $D = 3$ , and with linear operators  $L_1(u) = u$ ,  $L_2(u) = \partial_t u$ ,  $L_3(u) = \partial_x u$ ,  $L_4(u) = \partial_x^2 u$ ,  $J_1(m) = m$ ,  $J_2(m) = \partial_t m$ , and  $J_3(m) = \partial_x m$ . To solve the optimization problems, we apply the Gauss-Newton method with step size 0.1. Both algorithms stop after 320 iterations. In the experiments, we uniformly sample  $M_\Omega = 1200$  points in  $(0, 1) \times (-2, 2)$ , 200 samples in  $\{0\} \times (-2, 2)$ , and 200 points in  $\{1\} \times (-2, 2)$ . To show the accuracy of our methods, we



**Fig. 3.** The numerical solutions of the non-local stationary MFG (4.1) when  $\nu = 1$ . (a) The errors of  $\overline{H}_{GP}$  and  $\overline{H}_{FF}$  with respect to the reference  $\overline{H}_{FD}$ ; (b), (c), (e), (f) Numerical values of  $u_{GP}$ ,  $m_{GP}$ ,  $u_{FF}$ , and  $m_{FF}$ ; (d) the samples used by the GP and the FF methods; (g), (j) the evolution of the losses, which are the objective functions in (3.11) and (3.27); (h), (i), (k), (l) The contours of the pointwise errors of  $u_{GP}$ ,  $m_{GP}$ ,  $u_{FF}$ , and  $m_{FF}$ .

discretize the domain  $[0, 1] \times [-2, -2]$  with  $64 \times 512$  uniformly distributed grid points, solve (4.2) via the Eulerian solver<sup>2</sup> used in [45], and save the results as a reference. Denote by  $(u_{GP}, m_{GP})$  and  $(u_{FF}, m_{FF})$  the numerical solutions of the GP method and the FF algorithm, separately. We represent the result of the Eulerian solver by  $(u_{FV}, m_{FV})$ . Fig. 4 plots numerical results for the planning MFG. We plot the histories of Gauss–Newton iterations in Figs. 4a–4b. Both the GP and the FF methods use the same set of samples, which is shown in Fig. 4c. Since we care more about the evolution of the probability density in the planning problem, we show the pointwise error between  $m_{GP}$  and  $m_{FV}$  in Fig. 4d, and plot the error between  $m_{FF}$  and  $m_{FV}$  in Fig. 4e. We see that the errors are larger near the peaks of the probability density. This

<sup>2</sup> <https://github.com/EmoryMLIP/MFGnet.jl.git>.



**Fig. 4.** Numerical results for the Planing MFG (4.2): (a), (b) The histories of Gauss–Newton iterations; (c) uniformly sampled points in the space-time domain; (d), (e) contours of the pointwise errors of the numerical solutions; (f)–(k) time slices of the numerical solutions and the reference at  $t = 1/4, 1/2, 3/4$ .

phenomenon suggests the need to study non-uniformly distributed sample points in the future. Figs. 4f–4k compare various time slices of  $m_{GP}$ ,  $m_{FF}$ , and  $m_{FV}$  at time  $t = 1/4, 1/2, 3/4$  to highlight the accuracy of our methods.

## 5. Conclusions and future work

This paper presents two meshless algorithms, the GP method and the FF algorithm, to solve MFGs. The GP method adapts the algorithm in [16] to solve MFGs and finds numerical solutions in RKHSs. The convergence analysis of the GP method does not rely on the Lasry–Lions monotonicity condition. Hence, we plan to apply the GP method to solve MFGs with other monotonicity conditions or with non-monotone couplings in future work. To get better performance, we introduce the FF algorithm seeking approximations in Fourier features spaces. Compared to the GP method, the FF algorithm consumes less precomputation time without losing accuracy by using a prescribed number of base functions. Since the Fourier features

space lacks compactness, we cannot use the same convergence arguments of the GP method. Instead, the Lasry–Lions monotonicity ensures the boundedness of numerical errors and gives the convergence of the FF method. We plan to study the convergence of the FF algorithm in displacement monotone or non-monotone settings in future work. We believe that one can also use the FF method to solve general PDEs. As we have observed in the numerical experiments, the choices of base functions and parameters in our methods profoundly influence the accuracy of numerical solutions. Hence, in future work, we also plan to investigate methods of hyperparameter learning and different sampling techniques.

### CRedit authorship contribution statement

**Chenchen Mou:** Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – review & editing. **Xi-anjin Yang:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Chao Zhou:** Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Visualization, Writing – review & editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgement

We want to thank Yifan Chen for sharing the code of [16] online and discussing the implementation. We also thank Professor Lars Ruthotto for sharing the Matlab code about the crowd motion model in [45] and thank Mathieu Laurière for discussions about the finite difference methods of MFGs. C. Mou gratefully acknowledges the support by CityU Start-up Grant 7200684 and Hong Kong RGC Grant ECS 9048215. C. Zhou is supported by Singapore MOE (Ministry of Education), AcRF Grants R-146-000-271-112 and R-146-000-284-114, and NSFC Grant No. 11871364.

### Appendix A. Proofs of the results

**Proof of Theorem 2.6.** Let  $(u^*, m^*, \bar{H}^*)$  be the solution of (2.1). By the explicit formulas in (2.2),  $u^*$  and  $m^*$  are smooth. Thus, according to the approximation theorem of Fourier series [47, Chapter 14], for any  $\epsilon > 0$ , there exist  $N \in \mathbb{N}$  and functions  $u^N, m^N \in \mathcal{G}^N$  such that

$$\max_{i \leq 2} \sup_{x \in \mathbb{T}} |d^{(i)} u^N(x) - d^{(i)} u^*(x)| \leq \epsilon \quad (\text{A.1})$$

and

$$\max_{i \leq 1} \sup_{x \in \mathbb{T}} |d^{(i)} m^N(x) - d^{(i)} m^*(x)| \leq \epsilon, \quad (\text{A.2})$$

where  $d^{(i)}$  represents the  $i$ th order derivative. Hence, by the definition of  $\mathcal{Q}$  in (2.24), we have

$$\begin{aligned} \mathcal{Q}(u^N, m^N, \bar{H}^*) &= \int_{\mathbb{T}} |e^{H(x, u_x^N) - \bar{H}^*} - m^N|^2 dx + \left| \int_{\mathbb{T}} u^N dx \right|^2 + \left| \int_{\mathbb{T}} m^N dx - 1 \right|^2 \\ &\quad + \int_{\mathbb{T}} |m^N(u_{xx}^N + b_x(x)) + m_x^N(u_x^N + b(x))|^2 dx \\ &\leq \int_{\mathbb{T}} |e^{H(x, u_x^N) - \bar{H}^*} - e^{H(x, u_x^*) - \bar{H}^*}|^2 dx + \int_{\mathbb{T}} |m^N - m^*|^2 dx \\ &\quad + \left| \int_{\mathbb{T}} u^N dx - \int_{\mathbb{T}} u^* dx \right|^2 + \left| \int_{\mathbb{T}} m^N dx - \int_{\mathbb{T}} m^* dx \right|^2 \\ &\quad + \int_{\mathbb{T}} |m^N(u_{xx}^N + b_x(x)) - m^*(u_{xx}^* + b_x(x))|^2 dx \\ &\quad + \int_{\mathbb{T}} |m_x^N(u_x^N + b(x)) - m_x^*(u_x^* + b(x))|^2 dx. \end{aligned} \quad (\text{A.3})$$

Next, we estimate the terms at the right-hand side (RHS) of (A.3). From (A.1) and (A.2), we get

$$\int_{\mathbb{T}} |m^N - m^*|^2 dx + \left| \int_{\mathbb{T}} u^N dx - \int_{\mathbb{T}} u^* dx \right|^2 + \left| \int_{\mathbb{T}} m^N dx - \int_{\mathbb{T}} m^* dx \right|^2 \leq 3\epsilon^2. \quad (\text{A.4})$$

For the first term at the RHS of (A.3), we have

$$\begin{aligned} & \int_{\mathbb{T}} |e^{H(x, u_x^N) - \bar{H}^*} - e^{H(x, u_x^*) - \bar{H}^*}|^2 dx \\ &= e^{-2\bar{H}^*} \int_{\mathbb{T}} \left| \int_0^1 de^{H(x, \hat{u}_x^N)} \right|^2 dx \quad (\text{where } \hat{u}_x := su_x^N + (1-s)u_x^*, s \in [0, 1]) \\ &= e^{-2\bar{H}^*} \int_{\mathbb{T}} \left| \int_0^1 e^{H(x, \hat{u}_x^N)} D_p H(x, \hat{u}_x^N) (u_x^N - u_x^*) ds \right|^2 dx. \end{aligned} \quad (\text{A.5})$$

Using (A.1), we get

$$\sup_{x \in \mathbb{T}} |\hat{u}_x^N(x)| \leq \sup_{x \in \mathbb{T}} s|u_x^N - u_x^*| + \sup_{x \in \mathbb{T}} |u_x^*| \leq s\epsilon + \sup_{x \in \mathbb{T}} |u_x^*|. \quad (\text{A.6})$$

Thus, by the definition of  $H$  in (2.3), (A.1), (A.5), and (A.6), there exists a constant  $C > 0$  such that

$$\int_{\mathbb{T}} |e^{H(x, u_x^N) - \bar{H}^*} - e^{H(x, u_x^*) - \bar{H}^*}|^2 dx \leq C\epsilon^2. \quad (\text{A.7})$$

Next, we use (A.1) and (A.2), and get

$$\int_{\mathbb{T}} |m^N(u_{xx}^N + b_x(x)) - m^*(u_{xx}^* + b_x(x))|^2 dx \leq \int_{\mathbb{T}} |(m^N - m^*)(u_{xx}^N + b_x(x))|^2 dx + \int_{\mathbb{T}} |m^*(u_{xx}^N - u_{xx}^*)|^2 dx \leq C\epsilon^2. \quad (\text{A.8})$$

Similarly, there exists a constant  $C > 0$  such that

$$\int_{\mathbb{T}} |m_x^N(u_x + b(x)) - m_x^*(u_x^* + b(x))|^2 dx \leq C\epsilon^2. \quad (\text{A.9})$$

Combining (A.4), (A.7), (A.8), (A.9), we get from (A.3) that

$$\mathcal{Q}(u^N, m^N, \bar{H}^*) \leq C\epsilon^2. \quad (\text{A.10})$$

Using (A.1) and (A.2) again, we get

$$\int_{\mathbb{T}} \|u^N\|^2 dx \leq 2\|u^*\|_{L^2(\mathbb{T})}^2 + 2\epsilon^2 \quad \text{and} \quad \int_{\mathbb{T}} \|m^N\|^2 dx \leq 2\|m^*\|_{L^2(\mathbb{T})}^2 + 2\epsilon^2. \quad (\text{A.11})$$

Therefore, we conclude (2.28) by combining (A.10), (A.11), and the definition of  $J_\gamma$  in (2.23).  $\square$

**Proof of Corollary 2.7.** Let  $(u^*, m^*, \bar{H}^*)$  be the solution of (2.1). By Theorem 2.6, for given small  $\epsilon > 0$  and  $\gamma > 0$ , there exists a constant  $C$  and functions  $(u^N, m^N) \in \mathcal{G}^N \times \mathcal{G}^N$  such that

$$J_\gamma(u^N, m^N, \bar{H}^*) \leq 2\|u^*\|_{L^2}^2 + 2\|m^*\|_{L^2}^2 + |\bar{H}^*|^2 + C(1 + \gamma)\epsilon^2. \quad (\text{A.12})$$

Let  $(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma})$  be a minimizer in (2.25) given  $N$  and  $\gamma$ . Then, we have

$$J_\gamma(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma}) \leq J_\gamma(u^N, m^N, \bar{H}^*).$$

Thus, using (A.12) and the definition of  $J_\gamma$  in (2.23), we have

$$\gamma \mathcal{Q}(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma}) \leq 2\|u^*\|_{L^2}^2 + 2\|m^*\|_{L^2}^2 + |\bar{H}^*|^2 + C(1 + \gamma)\epsilon^2,$$

which yields

$$\mathcal{Q}(u^{N,\gamma}, m^{N,\gamma}, \bar{H}^{N,\gamma}) \leq \frac{2}{\gamma} (\|u^*\|_{L^2}^2 + \|m^*\|_{L^2}^2 + |\bar{H}^*|^2) + C\left(\frac{1}{\gamma} + 1\right)\epsilon^2.$$

Therefore, we conclude (2.29) by passing  $\gamma$  to infinity.  $\square$

**Proof of Theorem 2.8.** By Theorem 2.6 and Corollary 2.7, there exists a sequence  $\{(N_i, \gamma_i)\}_{i=1}^\infty$  such that the set  $\{(u^i, m^i, \bar{H}^i)\}_{i=1}^\infty$ , where  $(u^i, m^i, \bar{H}^i)$  is a minimizer of (2.25) given  $N_i$  and  $\gamma_i$ , satisfies

$$\mathcal{Q}(u^i, m^i, \bar{H}^i) \rightarrow 0 \text{ as } i \rightarrow \infty. \quad (\text{A.13})$$

Meanwhile, there exists a constant  $C$  such that for  $i \geq 1$ ,

$$J_{\gamma_i}(u^i, m^i, \bar{H}^i) \leq C. \quad (\text{A.14})$$

Thus, by (A.14) and the definition of  $J_{\gamma_i}$  in (2.23), there exists a constant  $C$  such that

$$\|u^i\|_{L^2(\mathbb{T})} \leq C, \quad \|m^i\|_{L^2(\mathbb{T})} \leq C, \quad \text{and } |\bar{H}^i| \leq C. \quad (\text{A.15})$$

Then, there exists  $\bar{H}^\infty \in \mathbb{R}$  such that, up to a subsequence,  $\bar{H}^i$  converges to  $\bar{H}^\infty$  in  $\mathbb{R}$ . Let  $h^i$  and  $g^i$  be functions such that

$$\begin{cases} \frac{(u_x^i)^2}{2} + V(x) + b(x)u_x^i = \ln(m^i + h^i(x)) + \bar{H}^i, \\ -(m^i(u_x^i + b(x)))_x = g^i(x). \end{cases} \quad (\text{A.16})$$

By (A.13), we have

$$\lim_{i \rightarrow \infty} \int_{\mathbb{T}} m^i dx = 1, \quad \lim_{i \rightarrow \infty} \|h^i\|_{L^2(\mathbb{T})} \rightarrow 0 \quad \text{and} \quad \lim_{i \rightarrow \infty} \|g^i\|_{L^2(\mathbb{T})} \rightarrow 0. \quad (\text{A.17})$$

Hence, integrating the first equation of (A.16), we get from (A.17) and (A.15) that

$$\|u_x^i\|_{L^2(\mathbb{T})} \leq C.$$

Thus,  $u^i \in H^1(\mathbb{T})$ . Since  $u^i$  is the finite linear combination of trigonometric functions, there exists a constant  $C$  such that

$$\|u^i\|_{C^1(\mathbb{T})} \leq C \|u^i\|_{H^1(\mathbb{T})} \leq C. \quad (\text{A.18})$$

Next, we prove the convergence of  $(u^i, m^i, \bar{H}^i)$  using the Lasry-Lions monotonicity argument. Let  $(u^*, m^*, \bar{H}^*)$  be the solution to (2.1). Then, from (A.16), we get

$$\begin{cases} \frac{(u_x^i)^2}{2} - \frac{(u_x^*)^2}{2} + b(x)u_x^i - b(x)u_x^* = \ln(m^i + h^i(x)) - \ln m^* + \bar{H}^i - \bar{H}^*, \\ -(m^i(u_x^i + b(x)))_x + (m^*(u_x^* + b(x)))_x = g^i(x). \end{cases} \quad (\text{A.19})$$

Multiplying the first equation of (A.19) by  $m^i + h^i - m^*$  and the second equation in (A.19) by  $u^i - u^*$ , subtracting the resulting equations, and integrating by parts, we obtain

$$\begin{aligned} & \frac{1}{2} \int_{\mathbb{T}} (m^i + h^i) |u_x^* - u_x^i|^2 dx + \frac{1}{2} \int_{\mathbb{T}} m^* |u_x^* - u_x^i|^2 dx + \int_{\mathbb{T}} (\ln(m^i + h^i) - \ln m^*) (m^i + h^i - m^*) dx \\ &= (\bar{H}^* - \bar{H}^i) \int_{\mathbb{T}} (m^i + h^i - m^*) dx + \int_{\mathbb{T}} g^i (u^i - u^*) dx + \int_{\mathbb{T}} h^i (u_x^i + b(x)) (u_x^i - u_x^*) dx. \end{aligned} \quad (\text{A.20})$$

By passing  $i \rightarrow 0$ , (A.17), (A.18), and (A.20) yield

$$\lim_{i \rightarrow 0} \frac{1}{2} \int_{\mathbb{T}} (m^i + h^i) |u_x^* - u_x^i|^2 dx + \frac{1}{2} \int_{\mathbb{T}} m^* |u_x^* - u_x^i|^2 dx + \int_{\mathbb{T}} (\ln(m^i + h^i) - \ln m^*) (m^i + h^i - m^*) dx = 0. \quad (\text{A.21})$$

From the first equation of (A.16), we get

$$m^i + h^i(x) = e^{\frac{1}{2}(u_x^i + b(x))^2 + V(x) - \bar{H}^i - b^2(x)} \geq e^{-C} \min_{x \in \mathbb{T}} e^{V(x) - b^2(x)}.$$

Then, the left-hand side of (A.21) is non-negative, and  $m^i + h^i$  is uniformly bounded below. Hence, (A.21) yields

$$u_x^i \rightarrow u_x^* \text{ in } L^2(\mathbb{T}).$$

Thus, up to a sub-sequence,  $u_x^i$  converges to  $u_x^*$  pointwisely. By the first equation of (A.16), there exists a function  $m^\infty$  such that, up to a sub-sequence,

$$m^i \rightarrow m^\infty = e^{\frac{(u_x^*)^2}{2} + V(x) + b(x)u_x^* - \overline{H}^\infty} \text{ pointwisely.}$$

Since  $m^i$  is uniformly bounded above, by the dominated convergence theorem, we conclude that  $m^i$  converges to  $m^\infty$  in  $L^1(\mathbb{T})$ . We note that  $(u^*, m^\infty, \overline{H}^\infty)$  satisfies (2.1). By the uniqueness of the solution, we conclude that  $m^\infty = m^*$  and  $\overline{H}^\infty = \overline{H}^*$ .  $\square$

**Proof of Theorem 3.11.** The arguments here are similar to the proof of Theorem 2.6. Using the approximation theorem of Fourier series and (3.19), we perform similar computations as in (A.3)-(A.11). Then, we conclude (3.24) by the definition of  $J_\gamma$ .  $\square$

**Proof of Corollary 3.12.** The proof is the same as in the arguments of Corollary 2.7.  $\square$

**Corollary A.1.** Let  $\{f^i\}_{i=1}^\infty$  be a sequence such that  $f^i \in \mathcal{G}^i$  defined in (3.20). If  $\{f^i\}_{i=1}^\infty$  is uniformly bounded in  $L^2(\mathbb{T}^d)$ . Then, there exists a constant  $C > 0$  such that

$$\|f^i\|_{L^\infty(\mathbb{T}^d)} \leq C.$$

Moreover, there exists a continuous function  $f$  such that, up to a sub-sequence,  $f^i$  converges to  $f$  in  $L^\infty(\mathbb{T}^d)$  as  $i \rightarrow \infty$ .

**Proof.** Since  $f^i \in \mathcal{G}^i$  for  $i \geq 1$ , there exist real numbers  $\{\alpha_j^i\}_{j \in \mathbb{Z}_i^d}$ ,  $\{\beta_j^i\}_{j \in \mathbb{Z}_i^d}$ , and  $c^i$  such that

$$f^i(x) = c^i + \sum_{j \in \mathbb{Z}_i^d} \alpha_j^i \sin(2\pi j^T x) + \sum_{j \in \mathbb{Z}_i^d} \beta_j^i \cos(2\pi j^T x), x \in \mathbb{T}^d.$$

Thus, we have

$$\begin{aligned} (f^i(x))^2 &= (c^i)^2 + \sum_{k, j \in \mathbb{Z}_i^d} \alpha_j^i \alpha_k^i \sin(2\pi j^T x) \sin(2\pi k^T x) + \sum_{k, j \in \mathbb{Z}_i^d} \beta_j^i \beta_k^i \cos(2\pi j^T x) \cos(2\pi k^T x) \\ &\quad + 2 \sum_{k, j \in \mathbb{Z}_i^d} \beta_j^i \alpha_k^i \cos(2\pi j^T x) \sin(2\pi k^T x) + c^i \sum_{j \in \mathbb{Z}_i^d} \alpha_j^i \sin(2\pi j^T x) + c^i \sum_{j \in \mathbb{Z}_i^d} \beta_j^i \cos(2\pi j^T x). \end{aligned} \quad (\text{A.22})$$

Since  $\forall i, j \in \mathbb{Z}^d$ ,

$$\int_{\mathbb{T}^d} \cos(2\pi i^T x) dx = \int_{\mathbb{T}^d} \sin(2\pi i^T x) dx = 0,$$

$$\int_{\mathbb{T}^d} \sin(2\pi j^T x) \sin(2\pi i^T x) dx = \begin{cases} 0, & \text{if } i \neq j \text{ or } i = j = 0, \\ \frac{1}{2}, & \text{if } i = j \neq 0, \end{cases}$$

$$\int_{\mathbb{T}^d} \cos(2\pi j^T x) \cos(2\pi i^T x) dx = \begin{cases} 0, & \text{if } i \neq j, \\ \frac{1}{2}, & \text{if } i = j \neq 0, \\ 1, & \text{if } i = j = 0, \end{cases}$$

and

$$\int_{\mathbb{T}^d} \cos(2\pi j^T x) \sin(2\pi i^T x) dx = 0,$$

we get from (A.22) that

$$\int_{\mathbb{T}^d} (f^i)^2 dx = (c^i)^2 + \frac{1}{2} \sum_{j \in \mathbb{Z}_i^d} (\alpha_j^i)^2 + \frac{1}{2} \sum_{j \in \mathbb{Z}_i^d} (\beta_j^i)^2. \quad (\text{A.23})$$

By (A.23) and the fact that  $\{f^i\}_{i=1}^\infty$  is uniformly bounded in  $L^2(\mathbb{T}^d)$ , there exists a constant  $C > 0$  such that

$$\int_{\mathbb{T}^d} (f^i)^2 dx = (c^i)^2 + \frac{1}{2} \sum_{j \in \mathbb{Z}_i^d} (\alpha_j^i)^2 + \frac{1}{2} \sum_{j \in \mathbb{Z}_i^d} (\beta_j^i)^2 \leq C. \quad (\text{A.24})$$

Thus, using (A.24) and Young's inequality, we obtain

$$\|f^i\|_{L^\infty(\mathbb{T}^d)}^2 \leq (|c^i| + \sum_{j \in \mathbb{Z}_i^d} |\alpha_j^i| + \sum_{j \in \mathbb{Z}_i^d} |\beta_j^i|)^2 \leq C \int_{\mathbb{T}^d} (f^i)^2 dx \leq C.$$

Hence,  $\{f^i\}_{i=1}^\infty$  is uniformly bounded in  $L^\infty(\mathbb{T}^d)$ .

Meanwhile, by (A.24), there exist real numbers  $c$ ,  $\{\alpha_j\}_{j=1}^\infty$ , and  $\{\beta_j\}_{j=1}^\infty$  such that, up to subsequences,  $c^i \rightarrow c$ ,  $\alpha_j^i \rightarrow \alpha_j$ , and  $\beta_j^i \rightarrow \beta_j$ , for all  $j \in \mathbb{Z}_i^d$ , as  $i \rightarrow \infty$ . Furthermore, for any  $\epsilon > 0$ , there exists  $N \in \mathbb{N}$  such that for any  $|i| \geq N$ ,

$$\sum_{j \in \mathbb{Z}_i^d, |j| \geq N} |\alpha_j^i| \leq \epsilon, \quad \sum_{j \in \mathbb{Z}_i^d, |j| \geq N} |\beta_j^i| \leq \epsilon, \quad (\text{A.25})$$

$$\sum_{j \in \mathbb{Z}^d, |j| \geq N} |\alpha_j| \leq \epsilon, \quad \text{and} \quad \sum_{j \in \mathbb{Z}^d, |j| \geq N} |\beta_j| \leq \epsilon. \quad (\text{A.26})$$

We define

$$f(x) = c + \sum_{j \in \mathbb{Z}^d \setminus \{0\}} \alpha_j \sin(2\pi j^T x) + \sum_{j \in \mathbb{Z}^d \setminus \{0\}} \beta_j \cos(2\pi j^T x)$$

and

$$f_N(x) = c + \sum_{j \in \mathbb{Z}_N^d} \alpha_j \sin(2\pi j^T x) + \sum_{j \in \mathbb{Z}_N^d} \beta_j \cos(2\pi j^T x).$$

Thus, for  $|i| > N$ , by (A.25) and (A.26), there exists a constant  $C > 0$  such that

$$\begin{aligned} \|f^i - f\|_{L^\infty(\mathbb{T}^d)} &\leq \|f^i - f_N\|_{L^\infty(\mathbb{T}^d)} + \|f_N - f\|_{L^\infty} \\ &\leq |c^i - c| + \sum_{j \in \mathbb{Z}_N^d} |\alpha_j^i - \alpha_j| + \sum_{j \in \mathbb{Z}_N^d} |\beta_j^i - \beta_j| \\ &\quad + \sum_{j \in \mathbb{Z}_i^d, |j| > N} |\alpha_j^i| + \sum_{j \in \mathbb{Z}_i^d, |j| > N} |\beta_j^i| + \sum_{j \in \mathbb{Z}^d, |j| > N} |\alpha_j| + \sum_{j \in \mathbb{Z}^d, |j| > N} |\beta_j| \leq C\epsilon. \end{aligned}$$

Therefore, we conclude that  $f^i$  converges to  $f$  in  $L^\infty(\mathbb{T}^d)$ .  $\square$

**Proof of Theorem 3.13.** By Theorem 3.11 and Corollary 3.12, there exist a sequence  $\{(N^i, \gamma^i)\}_{i=1}^\infty$  and a constant  $C$  such that the sequence  $\{(u^i, m^i, \bar{H}^i)\}_{i=1}^\infty$ , where  $(u^i, m^i, \bar{H}^i)$  is a minimizer of (3.23) given  $N^i$  and  $\gamma^i$ , satisfies

$$\|u^i\|_{L^2(\mathbb{T}^d)} \leq C, \quad \|m^i\|_{L^2(\mathbb{T}^d)} \leq C, \quad \text{and} \quad |\bar{H}^i| \leq C. \quad (\text{A.27})$$

Meanwhile, let  $h^i$  and  $g^i$  be functions such that

$$\begin{cases} -\Delta u^i + \frac{|\nabla u^i|^2}{2} + V(x) - F[m^i] - \bar{H}^i = h^i(x), \\ -\Delta m^i - \operatorname{div}(m^i \nabla u^i) = g^i(x). \end{cases} \quad (\text{A.28})$$

Then,

$$\lim_{i \rightarrow \infty} \|h^i\|_{L^2(\mathbb{T}^d)} \rightarrow 0 \quad \text{and} \quad \lim_{i \rightarrow \infty} \|g^i\|_{L^2(\mathbb{T}^d)} \rightarrow 0. \quad (\text{A.29})$$

Next, we study the regularity and the convergence of  $\{u^i\}_{i=1}^\infty$ . We split our arguments into three claims and prove each claim.

**Claim 1.** There exists a constant  $C > 0$  such that  $\|u^i\|_{C^1(\mathbb{T}^d)} \leq C$  for all  $i$  and exists a multi-valued function  $\varsigma \in (L^\infty(\mathbb{T}^d))^d$  such that, up to a subsequence,  $\nabla u^i$  converges to  $\varsigma$  in  $(L^\infty(\mathbb{T}^d))^d$ .

Integrating the first equation in (A.28), we get

$$\int_{\mathbb{T}^d} \frac{|\nabla u^i|^2}{2} dx = - \int_{\mathbb{T}^d} V(x) dx + \int_{\mathbb{T}^d} F[m^i] dx + \bar{H}^i + \int_{\mathbb{T}^d} h^i(x) dx. \quad (\text{A.30})$$

Thus, for  $i$  large enough, using Assumption 2, (A.27), (A.29), and the smoothness of  $V$ , we get that  $\nabla u^i$  is uniformly bounded in  $(L^2(\mathbb{T}^d))^d$ . Hence, by Corollary A.1 and (A.27), there exists a constant  $C$  such that  $\|u^i\|_{C^1(\mathbb{T}^d)} \leq C$  and exists a function  $\zeta \in (L^\infty(\mathbb{T}^d))^d$  such that, up to a sub-sequence,  $\nabla u^i$  converges in  $(L^\infty(\mathbb{T}^d))^d$  to  $\zeta$ .

The following claim gives the convergence of  $\{m^i\}_{i=1}^\infty$  and the properties of the limit.

**Claim 2.** *There exists a continuous function  $m^\infty$  such that, up to a subsequence,  $m^i$  converges to  $m^\infty$  in  $H^1(\mathbb{T}^d)$  and in  $L^\infty(\mathbb{T}^d)$  as  $i \rightarrow \infty$ . Moreover,  $\int_{\mathbb{T}^d} m^\infty dx = 1$  and  $m^\infty \geq 0$ .*

By (A.27) and Corollary A.1, there exists a continuous function  $m^\infty \in L^\infty(\mathbb{T}^d)$  such that  $m^i \rightarrow m^\infty$  in  $L^\infty(\mathbb{T}^d)$ . By Corollary 3.12,  $\lim_{i \rightarrow \infty} \int_{\mathbb{T}^d} m^i dx = 1$ . Thus,  $\int_{\mathbb{T}^d} m^\infty dx = 1$ . Next, we show that  $m^\infty \geq 0$ .

Multiplying the second equation of (A.28) by  $m^i$  and integrating by parts, we get

$$\begin{aligned} \int_{\mathbb{T}^d} |\nabla m^i|^2 dx &= - \int_{\mathbb{T}^d} m^i \langle \nabla u^i, \nabla m^i \rangle dx + \int_{\mathbb{T}^d} g^i m^i dx \\ &\leq \frac{1}{2} \int_{\mathbb{T}^d} (m^i)^2 |\nabla u^i|^2 dx + \frac{1}{2} \int_{\mathbb{T}^d} |\nabla m^i|^2 dx + \int_{\mathbb{T}^d} g^i m^i dx, \end{aligned} \quad (\text{A.31})$$

where we use Young's inequality in the above inequality. Hence, by (A.27), (A.29), Claim 1, and (A.31), there exists a constant  $C$  such

$$\|\nabla m^i\|_{L^2(\mathbb{T}^d)} \leq C.$$

Then, by Corollary A.1, up to a subsequence,  $\nabla m^i$  converges to a vector field  $\mathcal{E} \in (L^\infty(\mathbb{T}^d))^d$ . Since  $m^i \rightarrow m^\infty$  in  $L^\infty(\mathbb{T}^d)$ ,  $m^\infty$  is differentiable and  $\mathcal{E} = \nabla m^\infty$ . Hence,  $m^i$  converges to  $m^\infty$  in  $H^1(\mathbb{T}^d)$ . We multiply the second equation of (A.28) by  $\varphi \in C^\infty(\mathbb{T}^d)$ , integrate it by parts, and get

$$\int_{\mathbb{T}^d} \langle \nabla m^i, \nabla \varphi \rangle dx + \int_{\mathbb{T}^d} m^i \langle \nabla u^i, \nabla \varphi \rangle dx = \int_{\mathbb{T}^d} g^i \varphi dx.$$

Passing  $i$  to infinity, we obtain

$$\int_{\mathbb{T}^d} \langle \nabla m^\infty, \nabla \varphi \rangle dx + \int_{\mathbb{T}^d} m^\infty \langle \zeta, \nabla \varphi \rangle dx = 0.$$

Hence,  $m^\infty$  is a weak solution to

$$-\Delta m^\infty - \operatorname{div}(\zeta m^\infty) = 0 \text{ in } \mathbb{T}^d \text{ and } \int_{\mathbb{T}^d} m^\infty dx = 1.$$

Thus, by the ergodic theory,  $m^\infty \geq 0$  (see Section 1.4 of [8]).

Finally, we have the convergence of a subsequence of  $(u^i, m^i, \bar{H}^i)_{i=1}^\infty$  to the solution of (3.18). The proof is based on the Lasry-Lions monotonicity argument.

**Claim 3.** *Up to a subsequence,  $u^i \rightarrow u^*$  in  $H^1(\mathbb{T}^d)$ ,  $m^i \rightarrow m^*$  in  $H^1(\mathbb{T}^d)$ , and  $\bar{H}^i$  converges to  $\bar{H}^*$  in  $\mathbb{R}$  as  $i$  goes to infinity.*

Denote  $w^i = u^i - u^*$ ,  $\rho^i = m^i - m^*$ ,  $\tau^i = \bar{H}^i - \bar{H}^*$ . Then, we have

$$\begin{cases} -\Delta w^i + \langle \frac{\nabla u^i + \nabla u^*}{2}, \nabla w^i \rangle = F[m^i] - F[m^*] + \tau^i + h^i(x), \\ -\Delta \rho^i - \operatorname{div}(\nabla w^i m^i) - \operatorname{div}(\rho^i \nabla u^*) = g^i(x). \end{cases} \quad (\text{A.32})$$

Multiplying the first equation in (A.32) by  $\rho^i$  and integrating by parts, we obtain

$$\int_{\mathbb{T}^d} \langle \nabla w^i, \nabla \rho^i \rangle dx + \int_{\mathbb{T}^d} \left\langle \frac{\nabla u^i + \nabla u^*}{2}, \nabla w^i \rho^i \right\rangle dx = \int_{\mathbb{T}^d} (F[m^i] - F[m^*]) \rho^i dx + \tau^i \int_{\mathbb{T}^d} \rho^i dx + \int_{\mathbb{T}^d} h^i(x) \rho^i dx. \quad (\text{A.33})$$

Then, we integrate the second equation of (A.32) multiplied with  $w^i$  over  $\mathbb{T}^d$  and get

$$\int_{\mathbb{T}^d} \langle \nabla \rho^i, \nabla w^i \rangle dx + \int_{\mathbb{T}^d} |\nabla w^i|^2 m^i dx + \int_{\mathbb{T}^d} \langle \rho^i \nabla u^*, \nabla w^i \rangle dx = \int_{\mathbb{T}^d} g^i(x) w^i dx. \quad (\text{A.34})$$

Subtracting (A.33) from (A.34), we have

$$\frac{1}{2} \int_{\mathbb{T}^d} |\nabla w^i|^2 (m^i + m^*) dx + \int_{\mathbb{T}^d} (F[m^i] - F[m^*]) \rho^i dx = -\tau^i \int_{\mathbb{T}^d} \rho^i dx - \int_{\mathbb{T}^d} h^i(x) \rho^i dx + \int_{\mathbb{T}^d} g^i(x) w^i dx.$$

Using (A.29),  $\int_{\mathbb{T}^d} m^* dx = \int_{\mathbb{T}^d} m^\infty dx = 1$ , and Claim 2, we obtain

$$\lim_{i \rightarrow \infty} \frac{1}{2} \int_{\mathbb{T}^d} |\nabla w^i|^2 (m^i + m^*) dx + \int_{\mathbb{T}^d} (F[m^i] - F[m^*]) (m^i - m^*) dx = 0.$$

Hence, we have

$$\frac{1}{2} \int_{\mathbb{T}^d} |\nabla w^i|^2 (m^\infty + m^*) dx + \int_{\mathbb{T}^d} (F[m^i] - F[m^*]) (m^i - m^*) dx \rightarrow 0.$$

Since  $m^\infty \geq 0$  and  $m^* \geq 0$ , we get

$$\lim_{i \rightarrow 0} \int_{\mathbb{T}^d} (F[m^i] - F[m^*]) (m^i - m^*) dx = 0.$$

By Assumption 2 and Claim 2,

$$\int_{\mathbb{T}^d} (F[m^\infty] - F[m^*]) (m^\infty - m^*) dx = 0.$$

Thus, we have  $m^\infty = m^*$ . By Claim 2, we conclude that, up to a sub-sequence,  $m^i$  converges to  $m^*$  in  $H^1(\mathbb{T}^d)$ . Next, we show the convergence of  $u^i$ .

Multiplying the first equation in (A.32) by  $w^i$  and integrating by parts, we get

$$\int_{\mathbb{T}^d} |\nabla w^i|^2 = - \int_{\mathbb{T}^d} \left\langle \frac{\nabla u^i + \nabla u^*}{2}, \nabla w^i w^i \right\rangle dx + \int_{\mathbb{T}^d} (F[m^i] - F[m^*]) w^i dx + \tau^i \int_{\mathbb{T}^d} w^i dx + \int_{\mathbb{T}^d} h^i(x) w^i dx. \quad (\text{A.35})$$

Hence, passing  $i$  to infinity, we get from (A.35) that  $\nabla w^i$  converges to 0 in  $L^2(\mathbb{T}^d)$ . Thus,  $u^i$  converges to  $u^*$  in  $H^1(\mathbb{T}^d)$ . Finally, using (A.30),  $\bar{H}^i$  converges to  $\bar{H}^*$  in  $\mathbb{R}$ .  $\square$

## References

- [1] Y. Achdou, F. Camilli, I. Capuzzo-Dolcetta, Mean field games: numerical methods for the planning problem, *SIAM J. Control Optim.* 50 (1) (2012) 77–109.
- [2] Y. Achdou, I. Capuzzo-Dolcetta, Mean field games: numerical methods, *SIAM J. Numer. Anal.* 48 (3) (2010) 1136–1162.
- [3] Y. Achdou, J.-M. Lasry, Mean field games for modeling crowd motion, in: *Contributions to Partial Differential Equations and Applications*, Springer, 2019, pp. 17–42.
- [4] Y. Achdou, M. Laurière, Mean field games and applications: numerical aspects, in: *Mean Field Games, 2020*, pp. 249–307.
- [5] Y. Achdou, V. Perez, Iterative strategies for solving linearized discrete mean field games systems, *Netw. Heterog. Media* 7 (2) (2012) 197.
- [6] N. Almulla, R. Ferreira, D. Gomes, Two numerical approaches to stationary mean-field games, *Dyn. Games Appl.* 7 (4) (2017) 657–682.
- [7] J.-D. Benamou, G. Carlier, Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations, *J. Optim. Theory Appl.* 167 (1) (2015) 1–26.
- [8] A. Bensoussan, G. Blankenship, Singular perturbations in stochastic control, in: *Singular Perturbations and Asymptotic Analysis in Control Systems*, Springer, 1987, pp. 171–260.
- [9] L.M. Briceño-Arias, D. Kalise, Z. Kobeissi, M. Laurière, A.M. González, F. Silva, On the implementation of a primal-dual algorithm for second order time-dependent mean field games with local couplings, *ESAIM Proc. Surv.* 65 (2019) 330–348.
- [10] L.M. Briceño-Arias, D. Kalise, F. Silva, Proximal methods for stationary mean field games with local couplings, *SIAM J. Control Optim.* 56 (2) (2018) 801–836.
- [11] P. Cardaliaguet, Notes on mean field games, Technical report, 2010.

- [12] R. Carmona, M. Laurière, Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games I: the ergodic case, *SIAM J. Numer. Anal.* 59 (3) (2021) 1455–1485.
- [13] R. Carmona, M. Laurière, Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games: II—the finite horizon case, *Ann. Appl. Probab.* (2021), in press.
- [14] R. Carmona, M. Laurière, Deep learning for mean field games and mean field control with applications to finance, *arXiv preprint*, arXiv:2107.04568, 2021.
- [15] A. Cesaroni, M. Cirant, S. Dipierro, M. Novaga, E. Valdinoci, On stationary fractional mean field games, *J. Math. Pures Appl.* 122 (2019) 1–22.
- [16] Y. Chen, B. Hosseini, H. Owghadi, A.M. Stuart, Solving and learning nonlinear PDEs with Gaussian processes, *J. Comput. Phys.* 447 (2021) 110668.
- [17] Y. Chen, H. Owghadi, A. Stuart, Consistency of empirical Bayes and kernel flow for hierarchical parameter estimation, *Math. Comput.* (2021).
- [18] D. Evangelista, R. Ferreira, D. Gomes, L. Nurbekyan, V. Voskanyan, First-order, stationary mean-field games with congestion, *Nonlinear Anal.* 173 (2018) 37–74.
- [19] H. Gao, A. Lin, R.A. Banez, W. Li, Z. Han, S.J. Osher, H.V. Poor, Belief and opinion evolution in social networks: a high-dimensional mean field game approach, in: *ICC 2021–IEEE International Conference on Communications*, IEEE, 2021, pp. 1–6.
- [20] D. Gomes, L. Nurbekyan, E. Pimentel, *Economic Models and Mean-Field Games Theory*, Publicacoes Matematicas, IMPA, Rio, Brazil, 2015.
- [21] D. Gomes, J. Saúde, A mean-field game approach to price formation, *Dyn. Games Appl.* 11 (1) (2021) 29–53.
- [22] D. Gomes, J. Saúde, Numerical methods for finite-state mean-field games satisfying a monotonicity condition, *Appl. Math. Optim.* 83 (1) (2021) 51–82.
- [23] D. Gomes, X. Yang, The Hessian Riemannian flow and Newton's method for effective Hamiltonians and Mather measures, *ESAIM: Math. Model. Numer. Anal.* 54 (6) (2020) 1883–1915.
- [24] O. Guéant, J.-M. Lasry, P.-L. Lions, Mean field games and applications, in: *Paris-Princeton Lectures on Mathematical Finance 2010*, Springer, 2011, pp. 205–266.
- [25] M. Huang, P.E. Caines, R.P. Malhamé, An invariance principle in large population stochastic dynamic games, *J. Syst. Sci. Complex.* 20 (2) (2007) 162–172.
- [26] M. Huang, P.E. Caines, R.P. Malhamé, Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized  $\epsilon$ -Nash equilibria, *IEEE Trans. Autom. Control* 52 (9) (2007) 1560–1571.
- [27] M. Huang, P.E. Caines, R.P. Malhamé, The Nash certainty equivalence principle and McKean-Vlasov systems: an invariance principle and entry adaptation, in: *2007 46th IEEE Conference on Decision and Control*, IEEE, 2007, pp. 121–126.
- [28] M. Huang, R.P. Malhamé, P.E. Caines, Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle, *Commun. Inf. Syst.* 6 (3) (2006) 221–252.
- [29] J.-M. Lasry, P.-L. Lions, Jeux à champ moyen. I—le cas stationnaire, *C. R. Math.* 343 (9) (2006) 619–625.
- [30] J.-M. Lasry, P.-L. Lions, Jeux à champ moyen. II—horizon fini et contrôle optimal, *C. R. Math.* 343 (10) (2006) 679–684.
- [31] J.-M. Lasry, P.-L. Lions, Mean field games, *Jpn. J. Math.* 2 (1) (2007) 229–260.
- [32] M. Laurière, Numerical methods for mean field games and mean field type control, *arXiv preprint*, arXiv:2106.06231, 2021.
- [33] W. Lee, S. Liu, W. Li, S. Osher, Mean field control problems for vaccine distribution, *arXiv preprint*, arXiv:2104.11887, 2021.
- [34] W. Lee, S. Liu, H. Tembine, W. Li, S.J. Osher, Controlling propagation of epidemics via mean-field control, *SIAM J. Appl. Math.* (2020).
- [35] A.T. Lin, S.W. Fung, W. Li, L. Nurbekyan, S.J. Osher, Alternating the population and control neural networks to solve high-dimensional stochastic mean-field games, *Proc. Natl. Acad. Sci.* 118 (31) (2021).
- [36] P.-L. Lions, *Cours au Collège de France*, 2007–2013.
- [37] S. Liu, M. Jacobs, W. Li, L. Nurbekyan, S.J. Osher, Computational methods for first-order nonlocal mean field games with applications, *SIAM J. Numer. Anal.* 59 (5) (2021) 2639–2668.
- [38] S. Liu, L. Nurbekyan, Splitting methods for a class of non-potential mean field games, *J. Dyn. Games* (2020) 2020.
- [39] A.R. Mészáros, C. Mou, Mean field games systems under displacement monotonicity, *arXiv preprint*, arXiv:2109.06687, 2021.
- [40] L. Nurbekyan, J. Saúde, Fourier approximation methods for first-order nonlocal mean-field games, *Port. Math.* 75 (3) (2019) 367–396.
- [41] H. Owghadi, C. Scovel, *Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design*, vol. 35, Cambridge University Press, 2019.
- [42] H. Owghadi, G.R. Yoo, Kernel flows: from learning kernels from data into the abyss, *J. Comput. Phys.* 389 (2019) 22–47.
- [43] A. Rahimi, B. Recht, Random features for large-scale kernel machines, in: *NIPS*, vol. 3, Citeseer, 2007, p. 5.
- [44] W. Rudin, *Fourier Analysis on Groups*, Courier Dover Publications, 2017.
- [45] L. Ruthotto, S.J. Osher, W. Li, L. Nurbekyan, S.W. Fung, A machine learning framework for solving high-dimensional mean field game and mean field control problems, *Proc. Natl. Acad. Sci.* 117 (17) (2020) 9183–9193.
- [46] A. Tompkins, F. Ramos, Periodic kernel approximation by index set Fourier series features, in: *Uncertainty in Artificial Intelligence*, PMLR, 2020, pp. 486–496.
- [47] A. Vasy, *Partial Differential Equations: An Accessible Route Through Theory and Applications*, 2015.
- [48] F.X. Yu, A.T. Suresh, K.M. Choromanski, D.N. Holtmann-Rice, S. Kumar, Orthogonal random features, *Adv. Neural Inf. Process. Syst.* 29 (2016) 1975–1983.