

# Crowd-averse robust mean-field games: approximation via state space extension

D. Bauso and T. Mylvaganam and A. Astolfi

**Abstract**—We consider a population of dynamic agents, also referred to as players. The state of each player evolves according to a linear stochastic differential equation driven by a Brownian motion and under the influence of a control and an adversarial disturbance. Every player minimizes a cost functional which involves quadratic terms on state and control plus a cross-coupling mean-field term measuring the congestion resulting from the collective behavior, which motivates the term “crowd-averse”. Motivations for this model are analyzed and discussed in three main contexts: a stock market application, a production engineering example, and a dynamic demand management problem in power systems. For the problem in its abstract formulation, we illustrate the paradigm of robust mean-field games. Main contributions involve first the formulation of the problem as a robust mean-field game; second, the development of a new approximate solution approach based on the extension of the state space; third, a relaxation method to minimize the approximation error. Further results are provided for the scalar case, for which we establish performance bounds, and analyze stochastic stability of both the microscopic and the macroscopic dynamics.

## I. INTRODUCTION

We illustrate the robust mean-field game approach on a population of dynamic agents that wish to regulate their state to zero. The robust approach is particularly useful to account for model mis-specifications, uncertainties, or irrational behaviors on the part of the players. Each agent’s state evolves according to a linear stochastic differential equation (SDE) driven by a Brownian motion and under the influence of a control and an adversarial disturbance. The control minimizes a cost functional which involves quadratic terms on state and control plus a cross-coupling mean-field term involving the control of the single player and the mean control. The mean state is referred to as *common state*. The structure

The work of D. Bauso has been supported by the PRIN 20103S5RN3 “Robust decision making in markets and organization”. The work of T. Mylvaganam is conducted as part of the “STABLE-NET” research project funded by the RCUK’s Energy Programme (contract no: EP/L014343/1).

Dario Bauso is with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Mappin Street Sheffield, S1 3JD, United Kingdom, and the Dipartimento di Ingegneria Chimica, Gestionale, Informatica, Meccanica, Università di Palermo, V.le delle Scienze, 90128 Palermo, Italy. D. Bauso has done part of this work while he was academic visitor at the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK. email: (Email: d.bauso@sheffield.ac.uk)

T. Mylvaganam is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK (Email: thulasi.mylvaganam06@imperial.ac.uk).

A. Astolfi is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK and the Dipartimento di Ingegneria Civile e Ingegneria Informatica, Università di Roma “Tor Vergata”, Via del Politecnico, 1 00133 Roma, Italy (Email: a.astolfi@ic.ac.uk).

of the cost functional, involving the common state in the cross-term, has a straightforward interpretation in terms of pricing and mechanism or incentive design. Justifications for the model are provided in the context of three different applications: stock market, production, and power systems. In the latter example, for instance, the mean-field cross term is useful as it allows the redistribution of the control load away from peak “hours” thus reducing congestion, from which the term “crowd-averse”. Indeed every player pays a cost from controlling its own system when the population as a whole has a high mean control.

We highlight three main contributions. First, we establish a robust mean-field system for the considered game under adversarial disturbances. The resulting solution is referred to as *worst-disturbance feedback mean-field equilibrium*. Second, we provide a new approximate solution to compute such an equilibrium. The approach is based on the extension of the state space in the same spirit as [?], [?]. The method assumes that any player has an internal reference model for the common state. The proposed method relies only on the solution of a differential Riccati equation. Preliminarily to the illustration of the method, we also investigate on the solution of the Hamilton-Jacobi-Isaacs (HJI) equation under the assumption that the time evolution of the common state is given. We show that the problem reduces to solving three matrix equations. The state space extension is a heuristic method justified by *bounded rationality* and *limited computation capabilities* on the part of the players. The players agree on the internal models and act based on the estimate of the common state as obtained from the internal model. The quality of the approximation method depends on the accuracy of the internal reference model for the common state. Thus, as third contribution, we study a relaxation method aiming at minimizing the approximate error.

A few other contributions of the paper are obtained for the scalar case. The first result describes performance bounds. A second result establishes that the microscopic dynamics is exponential asymptotic stable almost surely. A third result shows that the macroscopic dynamics, involving the common state, is exponentially asymptotic stable.

Mean-field games were formulated by Lasry and Lions in [?] and independently by M.Y. Huang, P. E. Caines and R. Malhamé in [?], [?]. The mean-field theory of dynamical games is a modeling framework at the interface of differential game theory, mathematical physics, and  $H_\infty$ -optimal control that tries to capture the mutual influence between a crowd and its individuals. Mean-field games arise in several application domains such as economics, physics, biology, and network engineering (see [?], [?], [?], [?], [?], [?], [?]).

From a mathematical point of view the mean-field ap-

proach leads to a system of two PDEs. The first PDE is the Hamilton-Jacobi-Bellman equation. The second PDE is the Fokker-Planck equation which describes the density of the players [?], [?]. Explicit solutions in terms of mean-field equilibria are not common unless the problem has a linear-quadratic structure, see [?], and are extended to more general cases in [?]. In this sense, a variety of solution schemes has been recently proposed based on discretization and/or numerical approximations [?].

There has been substantial progress in the theory of weak solutions [?] and classical solutions [?], [?] for mean-field games. *Robust mean-field games* have been formulated in [?], [?]. The connection between robustness and risk-sensitivity is discussed in [?], [?]. Therein, an explicit solution in the case of the affine-exponentiated-Gaussian mean-field game is given. In this paper we provide an explicit approximate solution under milder assumptions on the distribution, which need not necessarily be Gaussian.

The paper is organized as follows. In Section ?? we formulate the problem. In Section ?? we provide some motivations. In Section ?? we derive the mean-field game. In Section ?? we introduce the approximate solution approach. In Section ?? we deal with the minimization of the approximation error. In Section ?? we provide additional results for the scalar case. In Section ?? we carry out some numerical studies. Finally in Section ?? we provide some conclusions. Preliminary results of the work presented herein have been published in [?].

**Notation.** We denote by  $(\Omega, \mathcal{F}, \mathbb{P})$  a complete probability space. We let  $\mathcal{B}$  be a finite-dimensional Brownian motion defined on this probability space. Let  $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$  be its natural filtration augmented by all the  $\mathbb{P}$ -null sets (sets of measure-zero with respect to  $\mathbb{P}$ ). We use  $\partial_x$  and  $\partial_{xx}^2$  to denote the first and second partial derivatives with respect to  $x$ , respectively. Given a vector  $x \in \mathbb{R}^n$  and a matrix  $Q \in \mathbb{R}^{n \times n}$  we denote by  $\|x\|_Q^2$  the weighted two-norm  $x^T Q x$ . The symbol  $Q_{i \bullet}$  denotes the  $i$ th row of a given matrix  $Q$ . We denote by  $Diag(x)$  the diagonal matrix in  $\mathbb{R}^{n \times n}$  whose entries in the main diagonal are the components of  $x$ .

## II. PROBLEM SET-UP

In this section we first introduce the model considered and then formulate the problem studied. Consider a game with an infinite number of homogeneous players. For each player let  $x_0 \in \mathbb{R}^n$  be its initial state, which is realized according to the probability distribution  $m_0$ . The state of the player at time  $t$ , denoted by  $x_t \in \mathbb{R}^n$ , evolves according to a controlled stochastic process over a finite horizon  $T > 0$ , i.e.

$$dx_t = [Ax_t + Bu_t + M\zeta_t]dt + \Sigma(x_t)d\mathcal{B}_t, \quad (1)$$

where  $u_t \in \mathbb{R}^r$  is the control input,  $\zeta_t \in \mathbb{R}^w$  is an adversarial disturbance,  $\mathcal{B}_t \in \mathbb{R}^n$  is an  $n$ -dimensional Brownian motion, which is independent across its components, independent of the initial state  $x_0$ , and independent across players and time. The matrices  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times r}$ , and  $M \in \mathbb{R}^{n \times w}$ , and  $\Sigma(\cdot) = Diag((\sigma_i x_i)_{i=1, \dots, n}) \in \mathbb{R}^{n \times n}$  for given scalars  $\sigma_i$ , all full column rank. For each player, the *admissible controls* and *disturbances* are square integrable and are adapted to

the filtration generated by the initial states and the Brownian motion, and possibly also adapted to some aggregate filtration associated with other players' dynamics.

To introduce a macroscopic description of the game, consider probability density functions on the state, control and disturbance spaces:

$$\begin{cases} m : \mathbb{R}^n \times [0, +\infty] \rightarrow [0, +\infty], (x, t) \mapsto m(x, t), \\ \int_{\mathbb{R}^n} m(x, t) dx = 1 \text{ for every } t, \\ z : \mathbb{R}^r \times [0, +\infty] \rightarrow [0, +\infty], (u, t) \mapsto z(u, t), \\ \int_{\mathbb{R}^r} z(u, t) du = 1 \text{ for every } t. \\ \tilde{z} : \mathbb{R}^w \times [0, +\infty] \rightarrow [0, +\infty], (\zeta, t) \mapsto \tilde{z}(\zeta, t), \\ \int_{\mathbb{R}^w} \tilde{z}(\zeta, t) d\zeta = 1 \text{ for every } t. \end{cases}$$

In the following we use the compact notation  $m_t = m(x, t)$ .

At a given time  $t$ , function  $m(\cdot)$  describes the density of players in a given state  $x \in \mathbb{R}^n$ , and  $z(\cdot)$  describes the density of players using a given control  $u \in \mathbb{R}^r$ . Define the mean state, control, and disturbance at time  $t$  as

$$\begin{aligned} \bar{m}_t &:= \int_{\mathbb{R}^n} x m(x, t) dx, & \bar{z}_t &:= \int_{\mathbb{R}^r} u z(u, t) du, \\ \bar{\zeta}_t &:= \int_{\mathbb{R}^w} \zeta \tilde{z}(\zeta, t) d\zeta. \end{aligned}$$

The mean  $\bar{m}_t$  represents an aggregate or macroscopic description of the state of the system and therefore, henceforth, we will refer to it as the *common state*.

Finally we introduce a cost functional with penalty on the final state  $g_T(\cdot)$ , stage cost function  $c(\cdot)$ , and quadratic penalty on the unknown disturbance:

$$J(x_0, u_t, \bar{m}_t, \zeta_t) = \mathbb{E} \left( g_T(x_T) + \int_0^T c(x_t, u_t, \bar{m}_t) dt - \frac{1}{2} \gamma^2 \int_0^T \|\zeta_t\|^2 dt \right). \quad (2)$$

Players wish to stabilize their states to zero, and therefore we set the stage cost as follows:

$$c(x_t, u_t, \bar{m}_t) = \frac{1}{2} [\|u_t\|_{R+S(\bar{m}_t)}^2 + \|x_t\|_Q^2],$$

where, for given nonnegative scalar weights  $\hat{h}_i, i = 1, \dots, r$ ,

$$\begin{aligned} S(\bar{m}_t) &= Diag((\hat{h}_i |\hat{k}_i \bullet \bar{m}_t|)_{i=1, \dots, r}) \\ &= Diag((\hat{h}_i |\bar{z}_t, i|)_{i=1, \dots, r}), \end{aligned} \quad (3)$$

for some  $\hat{k} \in \mathbb{R}^{r \times n}$ .

The term  $\frac{1}{2} \|x_t\|_Q^2$ , with  $Q > 0$ , is the cost of a non-zero state, and  $\frac{1}{2} \|u_t\|_R^2$ , with  $R > 0$ , accounts for a penalty on the control energy. The penalty on the final state  $g_T(x)$  is, in general, convex with minimum in zero, thus penalizing non-zero states at the end of the horizon. Let us take for it the expression  $g_T(x) = \frac{1}{2} \|x\|_\Phi^2$ , where  $\Phi$  is a positive matrix.

The term  $\frac{1}{2} \|u_t\|_{S(\bar{m}_t)}^2$  represents a cross-term coupling the control of each player and the common state of the population. The common state is in turn related to the mean control as explained next. By taking the mathematical expectation of both sides of (??) we can derive the following dynamics for the common state:

$$\frac{d}{dt} \bar{m}_t = A \bar{m}_t + B \bar{z}_t + M \bar{\zeta}_t.$$

Considering a deterministic disturbance  $\zeta_t$ , and using indistinguishability, the mean of the average control solves the equation:

$$\bar{z}_t = (B^T B)^{-1} \left[ B^T \left( \frac{d}{dt} \bar{m}_t \right) - B^T A \bar{m}_t - B^T M \bar{\zeta}_t \right].$$

A relation between  $\frac{d}{dt}\bar{m}_t$  and  $\bar{m}_t$  is yet to be introduced. We will see that both  $\frac{d}{dt}\bar{m}_t$  and  $\zeta_t$  can be approximated by linear functions in  $\bar{m}_t$  and therefore we can rewrite

$$\bar{z}_t = \hat{k}\bar{m}_t, \quad (4)$$

for some  $\hat{k} \in \mathbb{R}^{r \times n}$ . This is useful in the definition of the cost functional to reduce the number of independent variables, in that we turn all functions of  $\bar{z}_t$  as functions explicitly dependent on  $\bar{m}_t$ . The above preamble leads to the following robust mean-field game problem.

**Problem 1: (Robust mean-field problem)** Let  $x_0$  be independent of  $\mathcal{B}$  and with density  $m_0(x)$ . Let  $m_t$  be the mean-field trajectory. The robust mean-field problem in  $\mathbb{R}^n$  and  $(0, T]$  is given by

$$\left\{ \begin{array}{l} \inf_{\{u_t\}_t} \sup_{\{\zeta_t\}_t} J(x_0, u_t, \bar{m}_t, \zeta_t) \\ dx_t = [Ax_t + Bu_t + M\zeta_t] dt + \Sigma(x_t)d\mathcal{B}_t. \end{array} \right.$$

### III. MOTIVATIONS

In this section we first motivate the role of the common state in the cross-term and then we discuss three different applications or extensions of the model introduced Section ???. It turns out that the structure of the cost functional has a straightforward interpretation in terms of pricing and mechanism, or incentive, design.

We first reframe the model within the context of stock market literature. Here the continuous-time stochastic model having the structure of a geometric Brownian motion finds its natural collocation as a classical model for the random stock price with specific volatility. This model is used in the derivation of the well-know Black and Scholes equation, this being a renowned model for the value of a European call option [?]. The stock market model offers the opportunity to indulge in the derivation and interpretation of the cross-term in the objective function (??) as shadow price of a global optimization problem solved by a regulatory authority. Under this perspective, the problem takes a two-layer or leader-follower hierarchical structure with a leader, the regulatory agent setting the price, and multiple followers, the players adapting their strategies to the new price. Two other examples that accommodate well this hierarchical problem are a production model and a power model. In the first one, the regulatory authority is concerned with guaranteeing that the total supply is equal to the total demand. In the second example, the regulatory authority aims at stabilizing the mains frequency in a population of thermostatically controlled loads by using a dynamic demand management policy.

In all of the above examples, the mean-field term can be viewed as an incentive to encourage socially desirable behavior on the part of the players.

#### A. Pricing

Consider a central planner that aims at solving the following infinite horizon linear quadratic optimization problem

$$\min_{\pi_t} \frac{1}{2} \int (\|\bar{z}_t\|_{\Xi}^2 + \|\pi_t\|_{\Pi}^2) dt \quad (5)$$

subject to  $\dot{\bar{z}}_t = \pi_t,$

where  $\Xi$  and  $\Pi$  are opportune positive definite matrices. The term  $\pi_t$  represents the pricing action. Thus the above constraint models the influence of the pricing action on the common state evolution. If the central planner adopts a closed-loop state feedback control policy, we get

$$\min_{\hat{\phi}} \frac{1}{2} \int (\|\bar{z}_t\|_{\Xi}^2 + \|\hat{\phi}\bar{z}_t\|_{\Pi}^2) dt \quad (6)$$

subject to  $\dot{\bar{z}}_t = \hat{\phi}\bar{z}_t.$

In order to compute the optimal feedback control, let us consider the Hamiltonian function

$$H(\bar{z}_t, \lambda) = \frac{1}{2} \left( \|\bar{z}_t\|_{\Xi}^2 + \|\hat{\phi}\bar{z}_t\|_{\Pi}^2 \right) + \lambda \hat{\phi}\bar{z}_t.$$

From Pontryagin maximum principle, we know that the co-state  $\lambda$  satisfies the adjoint equation

$$\dot{\lambda} = - \frac{\partial H(\bar{z}_t, \lambda)}{\partial \bar{z}} = (\Xi + \hat{\phi}^T \Pi \hat{\phi}) \bar{z}_t + \lambda \hat{\phi}.$$

At the equilibrium, obtained by taking the LHS equal to zero, we obtain for the adjoint equation

$$\lambda_{\infty} = \frac{\partial v_{\infty}(\bar{z})}{\partial \bar{z}} = \hat{\phi}^{-1} (\Xi + \hat{\phi}^T \Pi \hat{\phi}) \bar{z}_t.$$

Let us take  $h = 2\hat{\phi}^{-1} (\Xi + \hat{\phi}^T \Pi \hat{\phi})$  and we have

$$\lambda_{\infty} = \frac{1}{2} h \bar{z}_t = \frac{1}{2} \begin{bmatrix} \hat{h}_1 & \dots & 0 \\ \dots & \ddots & \vdots \\ \dots & \dots & \hat{h}_r \end{bmatrix} \begin{bmatrix} \bar{z}_1 \\ \vdots \\ \bar{z}_r \end{bmatrix} = \begin{bmatrix} \hat{h}_1 \bar{z}_1 \\ \vdots \\ \hat{h}_r \bar{z}_r \end{bmatrix}.$$

The significance of the above derivation is that the term  $Diag((\hat{h}_i | \hat{k}_i \bar{m}_t)_{i=1, \dots, r}) = Diag((\hat{h}_i | \bar{z}_t, i)_{i=1, \dots, r})$  in (??) represents the shadow price of the global constraint  $\bar{z}_t = 0$ , namely, the price paid by the community due to the violation of the constraint  $\bar{z}_t = 0$ , and which the regulatory authority charges to the individuals in order to enforce the satisfaction of the constraint.

#### B. Stock market model [?]

The most plausible interpretation of dynamics (??) lies in a stock market context. This is a financial market involving a risky asset, referred to as stock, and a risk-averse asset, denoted as bond. A largely adopted dynamical model of the price of the stock  $\hat{S}_t$  is given by

$$d\hat{S}_t = \hat{S}_t [\mu dt + \sigma d\mathcal{B}_t],$$

where  $\mathcal{B}_t$  is the standard one-dimensional Brownian motion, and  $\mu$  and  $\sigma$  are given constants. Let us also assume that the bond price has a constant interest rate, denote it  $\hat{r}$ , i.e.,

$$d\mathcal{B}_t = \mathcal{B}_t [\hat{r} dt].$$

For each time  $t$ , we denote by  $X_t$  the money invested in the bond,  $Y_t$  the investments in the stock,  $l_t$  the rate of transfer from the bond holdings to the stock,  $\omega_t$  the rate of opposite transfers and  $\hat{c}_t$  the rate of consumption. The time evolution of  $X_t$  and  $Y_t$  is then given by

$$\begin{aligned} dX_t &= \hat{r} X_t dt - l_t dt + \omega_t dt - \hat{c}_t dt \\ dY_t &= Y_t [\mu dt + \sigma d\mathcal{B}_t] + l_t dt - \omega_t dt. \end{aligned} \quad (7)$$

Let the portfolio be obtained as sum of the invested money in the bond and at the stock, namely

$$Z_t = X_t + Y_t,$$

and denote the relative amount invested in the stock by

$$\pi_t = \frac{Y_t}{Z_t}.$$

It can be shown that the time evolution of the portfolio follows a stochastic differential equation in the form of a geometric Brownian motion of type:

$$dZ_t = Z_t[(\hat{r} + \pi_t(\mu - \hat{r}))dt + \pi_t\sigma d\mathcal{B}_t] - \hat{c}_t dt.$$

Note that the above has the same structure as (??) once we take  $Z_t$  represented by state  $x_t$ , the consumption  $\hat{c}_t$  by the control variable  $u_t$ , and for the parameters we take  $A = \hat{r} + \pi_t(\mu - \hat{r})$ . Here the central planner is a regulatory authority. The cost functional (??) aims at incentivizing the players to consume or spend the value of their portfolio within the end of the game while at the same time regulatory decisions apply taxes to consumption if the average consumption rate exceeds a predefined target value.

### C. Production [?], [?]

Suppose we have a continuum of producers with initial reserve  $x_0$  distributed according to  $m_0$ . Let the state be the reserve of raw material or resource available at a given time. Let the control be the production rate by a single producer and the adversarial disturbance be a disturbance parameter reflecting the taxation or inflation on the produced quantity.

A scalar version of equation (??) can be adapted to describe the variation of the reserve at time  $t$  given the current reserve and the consumed resource quantity:

$$dx_t = [-u_t + \sigma\zeta_t]dt + \sigma x_t d\mathcal{B}_t, \quad (8)$$

where  $\sigma$  is a given scalar. The term  $\sigma\zeta_t$  captures the negative and uncertain influence of taxation, or inflation, on the reserve. The stochastic term  $\sigma x_t d\mathcal{B}_t$  captures model misspecification due to the fact that the estimation of the reserves is not perfect or that reserves are random.

The cost functional,  $\hat{c}_t \bar{m}_t$  is the sale price of the final product and the cross-term is related to the income collected from producing and selling the quantity  $u_t$ ;  $\frac{Q}{2}(x_t)^2$  accounts for a production energy consumed,  $Q > 0$  and  $Ru_t^2$  is a known linear taxation on production. The penalty on the final state  $g(x_T)$  can be assumed quadratic in the reserve, so that unexploited reserve at the end of the horizon is penalized.

Finally, the term  $\sigma_t \zeta_t$  is intended to capture the negative and uncertain influence of taxation or inflation on the production. The shadow price  $\lambda_\infty = \frac{1}{2}h\bar{z}_t$  is obtained from a global constraint on demand/supply equilibrium,

$$\int_{\mathbb{R}} uz(u, t) du = D_t = 0.$$

In other words, around the equilibrium, when demand is equal to supply, we assume inelastic demand and take the mean demand equal to zero. Note that such a global balance of demand and supply is particularly significant in the power market [?].

As regards the cost functional (??), this involves a penalty term on production rate, namely  $\frac{1}{2}Ru_t^2$  and storage,  $\frac{1}{2}Qx_t^2$ . The additional cross-term  $\frac{1}{2}h\bar{z}_t u_t^2$  aims at penalizing production when the total supply exceeds the total demand, and

vice versa, to encourage production in the opposite case. The penalty on final state is a convex nonnegative penalty term accounting for unexploited reserve at the end of the horizon.

The cost functional (??) can be modified in different ways without compromising the results of this paper. A common expression in production models with a large number of producers [?], which finds its roots in the Cournot duopoly, appears as

$$c(x_t, u_t, \bar{z}_t, \zeta_t) = -h(\bar{z}_t, \zeta_t)u_t + \frac{1}{2}[(b + h\bar{z}_t)u_t^2 + ax_t^2].$$

Here the cross-term  $h(\bar{z}_t, \zeta_t)$  is the sale price of the manufactured product and thus  $h(\bar{z}_t, \zeta_t)u_t$  is the income collected from producing and selling the quantity  $u_t$ .

A slight change in the sign of the coefficients in (??), and the dynamics mirrors a classical multi-retailer inventory control equation describing the evolution of the inventory over time [?]:

$$dx_t = [u_t - \sigma\zeta_t]dt + \sigma x_t d\mathcal{B}_t. \quad (9)$$

In the above, the control is the reordered quantity and the disturbance is the unknown market demand. A classical scenario is where the transportation cost is shared among all retailers who reorder at a given time instant. A certain level of coordination of the retailers' replenishment strategies may lead to individual costs reduction. The cross mean-field term in the objective function (??) accounts for the reduced cost when orders are placed jointly. The other two terms are the cost of reordering and the cost of inventory shortage or inventory holding. We can generalize the framework to any application where multiple players share a service facility as airport facilities or telephone systems, drilling for oil, cooperative farming, and fishing (see also the references on cost-sharing games in [?])

### D. Dynamic demand management [?], [?]

Players are electrical appliances, say for instance heating or cooling appliances, and their state  $X(s)$  is the temperature at time  $t \leq s \leq T$ , where  $[t, T]$  is the time horizon window. Each appliance can be in one of two modes, *ON* or *OFF*, thus the control variable is a measurable function of time  $\pi_{ON}(\cdot)$  defined as  $s \mapsto \{0, 1\}$  and such that  $\pi_{ON}(s) = 1$  means that, at time  $s$ , the appliance is set to *ON* and  $\pi_{ON}(s) = 0$  means that the appliance is set to *OFF*. Dynamics (??) describes the time evolution of the temperature of each appliance. To see this, consider that when the appliance is *ON* the temperature decreases exponentially up to a fixed lower temperature whereas in *OFF* position the temperature increases exponentially up to a higher temperature. Then, the temperature of each appliance evolves according to the following differential equations, for  $t < s < T$ :

$$X'(s) = \begin{cases} -\alpha(X(s) - X_{ON}) & \text{if } \pi_{ON}(s) = 1 \\ -\alpha(X(s) - X_{OFF}) & \text{if } \pi_{ON}(s) = 0 \end{cases}, \quad (10)$$

with initial state  $X(t) = x$  and where  $\alpha > 0$  is a given scalar (the rate) and  $X_{ON}$ ,  $X_{OFF}$  are the steady-state temperatures of the appliances when in state *ON* or *OFF*, respectively. Here, considering a same rate for the two states has the only meaning of simplifying future computations.

Let us convexify the control set and consider the control of a single agent as the probability of setting the appliance  $ON$ , thus we have  $u(t) \in U := [0, 1]$  where  $U$  is the control set. It turns out that (??) can be rewritten in the form  $X' = f(X, u)$  where  $f : \mathbb{R} \times U \rightarrow \mathbb{R}$  is the following affine dynamics:

$$\begin{cases} X'(s) = -\alpha X(s) + \sigma u(s) + c, & s > t, \\ X(t) = x, \end{cases} \quad (11)$$

where  $x \in [X_{ON}, X_{OFF}]$ ,  $t \in [0, T]$  are the initial state and the initial time, respectively,  $\sigma := -\alpha(X_{OFF} - X_{ON})$ ,  $c := \alpha X_{OFF}$ . For sake of simplicity and without loss of generality we will take  $X_{OFF} = -X_{ON}$ . Indeed, we can always select lower and upper bounds of the temperature symmetric with respect to  $x_{ref}$ . In addition, note that the closed set  $[X_{ON}, X_{OFF}]$  is invariant and that the two extremes are not reachable from any other interior point. Hence, it is not restrictive to assume that no appliances have the temperatures  $X_{ON}$  and  $X_{OFF}$ .

Each controller is given a cost function that accounts for i) the energy consumption, which is captured by the penalty on the control, ii) the deviation of the mains frequency from the nominal value, represented by the cross-term, and iii) the deviation of the agent's temperature from the reference value, described by the penalty on the state. With respect to goal ii), the cross mean-field term incentivizes the appliances to switch to  $OFF$  if the mains frequency is below the nominal value and to switch to  $ON$  if the mains frequency is above the nominal value. This model is a simple one which can be adapted to the case in which the deviation is on the power rather than on the frequency.

#### IV. THE RESULTING MEAN-FIELD GAME

In this section we formulate the problem considered as a robust mean-field game. To this purpose, let  $v_t(x_t)$  be the (upper) value of the robust optimization problem under worst-case disturbance starting at time  $t$  from state  $x_t$ . The next theorem provides the mean-field system associated to the robust mean-field game introduced in Problem ???. The proof, which is given in the appendix, makes use of the definition of the Hamiltonian function, given by

$$H(x_t, p, \bar{m}_t) = \inf_{u_t} \left\{ c(x_t, u_t, \bar{m}_t) + p^T (Ax_t + Bu_t) \right\},$$

where  $p$  is the co-state, and of the robust Hamiltonian (see [?, Chaps 4,8] and [?]), which is obtained as

$$\tilde{H}(x_t, p, \bar{m}_t) = H(x_t, p, \bar{m}_t) + \sup_{\zeta_t} \left\{ p^T M \zeta_t - \frac{1}{2} \gamma^2 \zeta_t^T \zeta_t \right\}.$$

*Theorem 1:* The mean-field system associated to the robust mean-field game for the crowd-averse system is de-

scribed by the equations:

$$\begin{cases} \partial_t v_t + \frac{1}{2} (\partial_x v_t)^T \left[ -B(R + S(\bar{m}_t))^{-1} B^T \right. \\ \left. + \frac{1}{\gamma^2} M M^T \right] \partial_x v_t + (\partial_x v_t)^T A x_t \\ \left. + \frac{1}{2} x_t^T Q x_t + \frac{1}{2} \sum_{i=1}^n \sigma_i^2 x_i^2 \partial_{x_i x_i}^2 v_t = 0, \right. \\ \text{in } \mathbb{R}^n \times [0, T], \\ \\ v_T(x) = \frac{1}{2} x^T \Phi x, \text{ in } \mathbb{R}^n, \\ \\ \partial_t m_t + \partial_x \left[ m_t \left( A x_t - B(R + S(\bar{m}_t))^{-1} B^T \partial_x v_t \right. \right. \\ \left. \left. + \frac{1}{\gamma^2} M M^T \partial_x v_t \right) \right] - \frac{1}{2} \sum_{i=1}^n \partial_{x_i x_i}^2 [\sigma_i^2 x_i^2 m_t] = 0, \\ \text{in } \mathbb{R} \times [0, T], \\ \\ m_0(x) = d(x) \text{ in } \mathbb{R}^n, \\ \\ \bar{m}_t := \int_{\mathbb{R}} x m_t dx, \text{ in } [0, T], \end{cases} \quad (12)$$

Furthermore, the optimal control and worst disturbance are

$$u_t^* = -(R + S(\bar{m}_t))^{-1} B^T \partial_x v_t, \quad \zeta_t^* = \frac{1}{\gamma^2} M^T \partial_x v_t. \quad (13)$$

*Proof:* Given in the appendix.  $\square$

Any solution of the above system of equations is referred to as *worst-disturbance feedback mean-field equilibrium*. The significance of the above result is that to find the optimal control input we need to solve the two coupled PDEs in (??) in  $v_t$  and  $m_t$  with given boundary conditions. This is usually done by iteratively solving the HJI equation for fixed  $m_t$  and by entering the optimal  $u_t^*$  and  $\zeta_t^*$  obtained from (??) in the FPK equation in (??), until a fixed point in  $v_t$  and  $m_t$  is reached.

Since the Bellman equation depends explicitly on the mean of the mean-field and not on the other moments, one can reduce the mean-field system to a lower dimensional system. The reduced mean-field system associated to the robust mean-field game for the problem under study is (??) complemented with the following additional equations

$$\begin{cases} \frac{d}{dt} \bar{m}_t = A \bar{m}_t + B \bar{u}_t^* + M \bar{\zeta}_t^*, \text{ in } [0, T], \\ \bar{m}_0 = \int_{\mathbb{R}^n} x m_0 dx, \end{cases} \quad (14)$$

where

$$\begin{cases} \bar{u}_t^* := \int_{\mathbb{R}} (-B(R + S(\bar{m}_t))^{-1} B^T \partial_x v_t) m_t dx, \\ \text{in } [0, T], \\ \bar{\zeta}_t^* := \int_{\mathbb{R}} (\frac{1}{\gamma^2} M^T \partial_x v_t) m_t dx, \text{ in } [0, T], \end{cases} \quad (15)$$

and where  $\bar{u}_t^* = \bar{z}_t$  is the mean of the optimal individual state feedback control.

The resulting mean-field game maintains the same structure as in the deterministic case, but now both PDEs involve the second order derivatives of the value function  $v(\cdot)$  and the density  $m(\cdot)$  as shown in Table ??.

Because of the presence of the second derivatives, the above game is called second-order mean-field game. The advection equation is now renamed *Kolmogorov-Fokker-Planck (KFP)* equation.

$$\begin{array}{ccc}
\partial_t v_t + \tilde{H}(x_t, \partial_x v_t, \bar{m}_t) + \frac{1}{2} \sum_{i=1}^n \sigma_i^2 x_i^2 \partial_{x_i x_i}^2 v_t = 0, & & \text{(HJB)-backward} \\
\downarrow v_t & & \uparrow \bar{m}_t \\
\partial_t m_t + \partial_x \left( m_t \partial_p \tilde{H}(x, \partial_x v_t, m) \right) - \frac{1}{2} \sum_{i=1}^n \partial_{x_i x_i}^2 [\sigma_i^2 x_i^2 m_t] = 0 & & \text{(Kolmogorov)-forward}
\end{array}$$

TABLE I  
STRUCTURE OF THE MEAN-FIELD GAME.

## V. ON MEAN-FIELD EQUILIBRIA AND THEIR APPROXIMATION

In this section, we introduce an approximate solution approach based on a state-space extension and internal reference model. The approach does not require the existence of a fixed-point for the system in Table ???. The approach provides a description of the microscopic and macroscopic evolution of the system when all players agree on modeling the “environment” using the aforementioned reference models. We then study equilibria and stability properties based on such reference models.

### A. Considerations on the parametrized HJI

We now investigate the solution of the HJI equation under the assumption that the time evolution of the common state is given. We show that the problem reduces to solving three matrix equations. To see this, by isolating the HJI part of (??) for fixed  $m_t$ , for  $t \in [0, T]$ , we have

$$\begin{cases}
\partial_t v_t + \frac{1}{2} (\partial_x v_t)^T \left[ -B(R + S(\bar{m}_t))^{-1} B^T \right. \\
\left. + \frac{1}{\gamma^2} M M^T \right] \partial_x v_t + (\partial_x v_t)^T A x_t + \frac{1}{2} x_t^T Q x_t \\
+ \frac{1}{2} \sum_{i=1}^n \sigma_i^2 x_i^2 \partial_{x_i x_i}^2 v_t = 0, \\
\text{in } \mathbb{R} \times [0, T[, \\
v_T(x) = \frac{1}{2} x^T \Phi x, \text{ in } \mathbb{R}^n.
\end{cases} \quad (16)$$

Consider the following value function

$$v_t(x) = \frac{1}{2} x^T P_t x + \Psi_t x + \chi_t, \quad (17)$$

so that (??) can be rewritten as

$$\begin{cases}
\frac{1}{2} x^T \dot{P}_t x + \dot{\Psi}_t x + \dot{\chi}_t \\
+ (P_t x + \Psi_t)^T \left[ -B(R + S(\bar{m}_t))^{-1} B^T \right. \\
\left. + \frac{1}{\gamma^2} M M^T \right] (P_t x + \Psi_t) + (P_t x + \Psi_t)^T A x \\
+ \frac{1}{2} x_t^T Q x_t + \frac{1}{2} \sum_{i=1}^n \sigma_i^2 x_i^2 P_{ii}(t) = 0 \text{ in } \mathbb{R} \times [0, T[, \\
P_T = \Phi, \quad \Psi_T = 0, \quad \chi_T = 0.
\end{cases}$$

The boundary conditions are obtained by imposing that

$$v_T(x) = \frac{1}{2} x^T P_T x + \Psi_T x + \chi_T = \frac{1}{2} x^T \Phi x.$$

Again, since this is an identity in  $x$ , it reduces to three

equations:

$$\begin{cases}
\dot{P}_t + 2P_t^T \left[ -B(R + S(\bar{m}_t))^{-1} B^T \right. \\
\left. + \frac{1}{\gamma^2} M M^T \right] P_t + 2P_t^T A \\
+ Q + \dot{P}_t = 0 \text{ in } [0, T[, \quad P_T = \Phi, \\
\dot{\Psi}_t + P_t^T \left[ -B(R + S(\bar{m}_t))^{-1} B^T + \frac{1}{\gamma^2} M M^T \right] \Psi_t \\
+ \Psi_t^T \left[ -B(R + S(\bar{m}_t))^{-1} B^T + \frac{1}{\gamma^2} M M^T \right] P_t \\
+ \Psi_t^T A = 0 \text{ in } [0, T[, \quad \Psi_T = 0, \\
\dot{\chi}_t + \Psi_t^T \left[ -B(R + S(\bar{m}_t))^{-1} B^T \right. \\
\left. + \frac{1}{\gamma^2} M M^T \right] \Psi_t = 0 \text{ in } [0, T[, \quad \chi_T = 0,
\end{cases} \quad (18)$$

where

$$\begin{aligned}
\tilde{P} &= \text{Diag}((\sigma_i^2 P_{ii})_{i=1, \dots, n}) \\
&= \begin{bmatrix} \sigma_1^2 P_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ \dots & \dots & \sigma_n^2 P_{nn} \end{bmatrix}.
\end{aligned} \quad (19)$$

For the optimal control and worst-case disturbance we have

$$\begin{cases}
u_t^*(x) = (R + S(\bar{m}_t))^{-1} B^T (P_t x + \Psi_t), \\
\zeta_t^*(x) = \frac{1}{\gamma^2} M^T (P_t x + \Psi_t).
\end{cases} \quad (20)$$

Existence of a solution for the equations (??) is guaranteed under standard assumptions on convexity-concavity of the value function with respect to the control and the disturbance [?, Chap. 8]. This also justifies the choice of the quadratic structure for the value function (??).

### B. Internal reference model and state space extension

In this section, we study the problem in the extended state space involving both the state of the player and the average state distribution. The main idea is illustrated in Fig. ?. In the mean-field system (??) the gradient  $\partial_x v_t$  is parametrized in the average distribution  $\bar{m}_t$ , which evolves according to a nonlinear differential equation. Then, we replace the dynamics of  $\bar{m}_t$  with two linear dynamics on the new variables  $\hat{m}_t$  and  $\tilde{m}_t$  (dashed and dotted trajectories) that upper and lower bound the nonlinear dynamics of  $\bar{m}_t$  (solid). In the extended state space, the state variable evolves according to the equations

$$\begin{cases}
dx_t = [A x_t + B u_t + M \zeta_t] dt + \Sigma(x_t) d\mathcal{B}_t, \\
\dot{\hat{m}}_t = A \hat{m}_t + B \bar{u}_t + M \bar{\zeta}_t,
\end{cases} \quad (21)$$

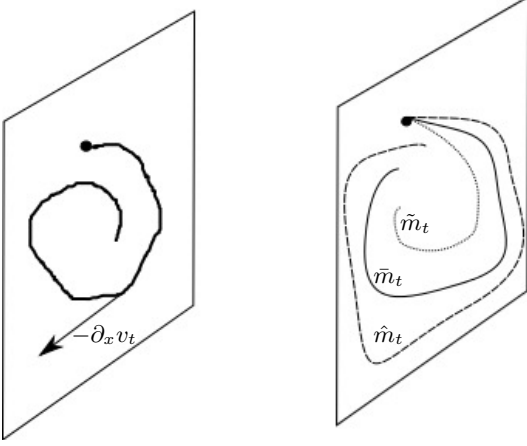


Fig. 1. Extended state space: the gradient  $\partial_x v_t$  depends on  $\tilde{m}_t$ , which is upper and lower bounded by  $\hat{m}_t$  (dashed) and  $\tilde{m}_t$  (dotted) respectively.

which can be rewritten in matrix form as

$$\begin{bmatrix} dx_t \\ d\tilde{m}_t \end{bmatrix} = \left( A \begin{bmatrix} x_t \\ \tilde{m}_t \end{bmatrix} + B \begin{bmatrix} u_t \\ \tilde{u}_t \end{bmatrix} + M \begin{bmatrix} \zeta_t \\ \tilde{\zeta}_t \end{bmatrix} \right) dt + \begin{bmatrix} \Sigma(x_t) dB_t \\ 0 \end{bmatrix}.$$

The main idea is that each player has an internal model for the common state. In particular, each player approximates the evolution of the common state through a dynamics of type

$$\begin{cases} \frac{d}{dt} \tilde{m}_t = \Theta_t \tilde{m}_t, & \text{for all } t \in [0, T], \\ \tilde{m}_0 = \tilde{m}_0. \end{cases} \quad (22)$$

Though this introduces an approximation, it must be said that second-order systems are commonly used to approximate higher-order dynamics.

By substituting the current average distribution  $\tilde{m}_t$  by its estimate  $\hat{m}_t$ , the problem at hand can be rewritten as

$$\inf_{\{u_t\}_t} \sup_{\{\zeta_t\}_t} \int_0^T \frac{1}{2} \left[ \|u_t\|_{R+S(\tilde{m}_t)}^2 + \|\tilde{m}_t\|_{\tilde{Q}}^2 + (\|x_t\|_Q - \gamma^2 \|\zeta_t\|^2) \right] dt$$

$$\begin{bmatrix} dx_t \\ d\tilde{m}_t \end{bmatrix} = \left( \begin{bmatrix} A & 0 \\ 0 & \Theta_t \end{bmatrix} \begin{bmatrix} x_t \\ \tilde{m}_t \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u_t^* + \begin{bmatrix} M \\ 0 \end{bmatrix} \zeta_t^* \right) dt + \begin{bmatrix} \Sigma(x_t) dB_t \\ 0 \end{bmatrix},$$

for some positive definite matrix  $\tilde{Q}$ .

Reformulating the problem in terms of the extended state

$$X_t = \begin{bmatrix} x_t^T & \tilde{m}_t^T \end{bmatrix}^T,$$

yields the linear quadratic problem:

$$\inf_{\{\tilde{u}_t\}_t} \sup_{\{\zeta_t\}_t} \int_0^T \left[ \frac{1}{2} \left( X_t^T \tilde{Q} X_t + \tilde{R} \tilde{u}_t^2 - \tilde{\Gamma} \zeta_t^2 \right) \right] dt$$

$$dX_t = \left( \tilde{A} X_t + \tilde{B} \tilde{u}_t + C \zeta_t \right) dt + D dB_t,$$

where

$$\begin{aligned} \tilde{Q} &= \begin{bmatrix} Q & 0 \\ 0 & \tilde{Q} \end{bmatrix}, & \tilde{R} &= R + S(\tilde{m}_t), & \tilde{\Gamma} &= \gamma^2, \\ \tilde{A} &= \begin{bmatrix} A & 0 \\ 0 & \Theta_t \end{bmatrix}, & \tilde{B} &= \begin{bmatrix} B \\ 0 \end{bmatrix}, \\ C &= \begin{bmatrix} M \\ 0 \end{bmatrix}, & D &= \begin{bmatrix} \Sigma(x_t) \\ 0 \end{bmatrix}. \end{aligned}$$

The idea is therefore to consider a new value function  $\mathcal{V}_t(x, \tilde{m})$  (in compact form  $\mathcal{V}_t(X)$ ) in the extended state space which satisfies

$$\begin{cases} \partial_t \mathcal{V}_t(X) + H(X, \partial_X \mathcal{V}_t(X)) + \frac{1}{2} \frac{1}{\gamma^2} M M^T |\partial_X \mathcal{V}_t(X)|^2 \\ + \frac{1}{2} \sum_{i=1}^n \sigma_i^2 x_i^2 \partial_{x_i}^2 \mathcal{V}_t(X) = 0, & \text{in } \mathbb{R}^2 \times [0, T], \\ \mathcal{V}_T(X) = g(x) & \text{in } \mathbb{R}^2. \end{cases}$$

Assume that  $\mathcal{V}_T(X)$  is given by the quadratic form

$$\mathcal{V}_t(X) = \frac{1}{2} \begin{bmatrix} x_t^T & \tilde{m}_t^T \end{bmatrix} \underbrace{\begin{bmatrix} \mathcal{P}_{11}(t) & \mathcal{P}_{12}(t) \\ \mathcal{P}_{21}(t) & \mathcal{P}_{22}(t) \end{bmatrix}}_{\mathcal{P}_t} \begin{bmatrix} x_t \\ \tilde{m}_t \end{bmatrix},$$

where the matrix  $\mathcal{P}_t$  is the solution of the differential Riccati equation

$$\begin{aligned} \dot{\mathcal{P}}_t + \mathcal{P}_t \tilde{A} + \tilde{A}^T \mathcal{P}_t + \tilde{Q} \\ - 2\mathcal{P}_t (\tilde{B} \tilde{R}^{-1} \tilde{B}^T - C \tilde{\Gamma}^{-1} C^T) \mathcal{P}_t + W_t = 0, \end{aligned} \quad (18)$$

and where

$$\begin{aligned} \tilde{B} \tilde{R}^{-1} \tilde{B}^T - C \tilde{\Gamma}^{-1} C^T \\ = \begin{bmatrix} B(R + S(\tilde{m}_t))^{-1} B^T - \frac{1}{\gamma^2} M M^T & 0 \\ 0 & 0 \end{bmatrix}, \\ W_t = \begin{bmatrix} \tilde{\mathcal{P}}_{11}(t) & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Here  $\tilde{\mathcal{P}}_{11}(t)$  is as in (??).

Note that in the stationary case the above differential equation simplifies to

$$\begin{aligned} \mathcal{P} \tilde{A} + \tilde{A}^T \mathcal{P} + \tilde{Q} - 2\mathcal{P} (\tilde{B} \tilde{R}^{-1} \tilde{B}^T \\ - C \tilde{\Gamma}^{-1} C^T) \mathcal{P} + W = 0. \end{aligned} \quad (17)$$

The above algebraic Riccati equation is then associated to the infinite horizon formulation of the game under study.

Back to the finite horizon game, let  $\mathcal{P}_t$  be the solution of the differential Riccati equation (??), then the optimal control is given by

$$\begin{aligned} \tilde{u}_t &= -\tilde{R}^{-1} \tilde{B}^T \mathcal{P}_t X_t \\ &= -(R + S(\tilde{m}_t))^{-1} [B^T \ 0] \\ &\quad \cdot \begin{bmatrix} \mathcal{P}_{11}(t) & \mathcal{P}_{12}(t) \\ \mathcal{P}_{21}(t) & \mathcal{P}_{22}(t) \end{bmatrix} \begin{bmatrix} x_t \\ \tilde{m}_t \end{bmatrix} \\ &= -(R + S(\tilde{m}_t))^{-1} B^T \\ &\quad \cdot (\mathcal{P}_{11}(t) x_t + \mathcal{P}_{12}(t) \tilde{m}_t), \end{aligned} \quad (17)$$

and the worst-case disturbance is

$$\begin{aligned} \tilde{\zeta}_t &= \tilde{\Gamma}^{-1} C^T \mathcal{P}_t X_t \\ &= \frac{1}{\gamma^2} [\Sigma^T \ 0] \begin{bmatrix} \mathcal{P}_{11}(t) & \mathcal{P}_{12}(t) \\ \mathcal{P}_{21}(t) & \mathcal{P}_{22}(t) \end{bmatrix} \begin{bmatrix} x_t \\ \tilde{m}_t \end{bmatrix} \\ &= \frac{1}{\gamma^2} \Sigma^T (\mathcal{P}_{11}(t) x_t + \mathcal{P}_{12}(t) \tilde{m}_t). \end{aligned} \quad (17)$$

From (??) and (??), we can approximate the mean-field equilibrium, which is captured by the evolution of  $\tilde{m}_t$  over the horizon  $(0, T)$ , as follows:

$$\frac{d}{dt}\bar{m}_t = \left[ A + \left( -B(R + S(\bar{m}_t))^{-1}B^T + \frac{1}{\gamma^2}MM^T \right) \cdot (P_{11}(t) + P_{12}(t)) \right] \bar{m}_t \quad t \in (0, T], \quad x_0 \in \mathbb{R}.$$

The equation above corresponds to saying that the mean distribution converges to zero in absence of the stochastic disturbances (the Brownian motion), under the assumption that all the eigenvalues of the matrix  $\left[ A + \left( -B(R + S(\bar{m}_t))^{-1}B^T + \frac{1}{\gamma^2}MM^T \right) (P_{11}(t) + P_{12}(t)) \right]$  have strictly negative real parts.

## VI. MINIMIZING THE APPROXIMATION ERROR

In this section we introduce a relaxation method aiming at minimizing the approximate error. Indeed, the quality of the approximate problem solved in the previous section depends on the accuracy of the internal reference model for the common state (??).

Let us start by noting that the common state  $\bar{m}_t$  and its approximation  $\tilde{m}_t$  evolve according to

$$\dot{\tilde{m}}_t = \left[ A + (-B(R + S(\tilde{m}_t))^{-1}B^T + \frac{1}{\gamma^2}MM^T) \cdot (P_{11}(t) + P_{12}(t)) \right] \tilde{m}_t, \quad (16)$$

$$\dot{\tilde{m}}_t = \Theta_t \tilde{m}_t. \quad (17)$$

Henceforth, for sake of conciseness, let us denote

$$\Theta' = \left[ A + (-B(R + S(\tilde{m}_t))^{-1}B^T + \frac{1}{\gamma^2}MM^T) \cdot (P_{11}(t) + P_{12}(t)) \right].$$

Consequently, dynamics (??) can be rewritten in compact form as

$$\dot{\tilde{m}}_t = \Theta'_t \tilde{m}_t.$$

Dynamics (??) is obtained from averaging the optimal control (??) and worst-case disturbance (??) in order to obtain  $\bar{u}_t^*$  and  $\bar{\zeta}_t^*$ , respectively. These values are then substituted in the dynamics for the average distribution  $\frac{d}{dt}\bar{m}_t = A\bar{m}_t + B\bar{u}_t^* + \Sigma\bar{\zeta}_t^*$ . The best approximation, namely, the one with the best bound, is given by the following least-square error minimization

$$\min_{\Theta_0, \{\underline{u}_t\}_{t \in [0, T]}} \int_0^T [\|\Theta'_t - \Theta_t\|^2 + \underline{u}_t^T \underline{\rho} \underline{u}_t] dt \quad (16)$$

$$\text{subject to} \quad \begin{aligned} \dot{\Theta}_t &= \underline{u}_t, \\ P_{11}(t) + P_{12}(t) &\text{ from (??).} \end{aligned}$$

In other words, the functional (??) accounts for the approximation error using the internal model. Actually, the internal model returns a predicted common state which differs from the exact value.

Obviously, for  $\underline{\rho} \rightarrow \infty$  we force  $\Theta_t$  to be constant and we limit to consider the best constant value for  $\Theta$  that minimizes the least mean square error.

The least mean square problem (??) yields a gradient algorithm based on the following update law for  $\Theta_t(\tau)$ :

$$\dot{\Theta}_t(\tau) = \underline{k} \left( \left[ A + (-B(R + S(\tilde{m}_t))^{-1}B^T + \frac{1}{\gamma^2}MM^T) \cdot (P_{11}(t) + P_{12}(t)) \right] - \Theta_t(\tau) \right),$$

where  $\underline{k}$  is the update coefficient.

At the equilibrium  $\Theta_t^*$ , setting the LHS  $\dot{\Theta}_t(\tau)$  equal to zero, the resulting solution is the least mean square solution, namely the solution at minimum distance from

$$\Theta_t^* = \left[ A + (-B(R + S(\tilde{m}_t))^{-1}B^T + \frac{1}{\gamma^2}MM^T) \cdot (P_{11}(t) + P_{12}(t)) \right].$$

If the least mean square error, i.e., the optimal cost of (??) is null, then the solution is  $\Theta_t^*$ .

We are now interested in investigating conditions under which the equilibrium point is asymptotically stable. This implies that starting from any solution in a bounded neighborhood of  $\Theta_t^*$ , the resulting solution converges asymptotically to  $\Theta_t^*$ . For this to be true, for any  $\Theta \neq \Theta_t^*$  in a neighborhood of the equilibrium point, it must hold

$$\frac{\partial}{\partial \Theta} \left( \left[ A + (-B(R + S(\tilde{m}_t))^{-1}B^T + \frac{1}{\gamma^2}MM^T) \cdot (P_{11}(t) + P_{12}(t)) \right] - \Theta_t(\tau) \right) < 0.$$

Two main considerations arise. First, when the coefficient tends to infinity, the RHS is negative since  $\Theta'$  is bounded. Second, when the coefficient tends to zero, the RHS is positive since  $\Theta'$  is bounded as well. Both considerations can be schematically summarized as follows:

- C1**  $\Theta_t \rightarrow \infty$  the quantity  $\Theta' = -\left[ A + (-Bb^{-1}B^T + \frac{1}{\gamma^2}MM^T)(P_{11}(t) + P_{12}(t)) \right]$  is bounded and therefore  $\dot{\Theta} < 0$
- C2**  $\Theta_t \rightarrow 0$  the quantity  $\Theta' = -\left[ A + (-B(R + S(\tilde{m}_0))^{-1}B^T + \frac{1}{\gamma^2}MM^T)(P_{11}(t) + P_{12}(t)) \right]$  is bounded and positive and therefore  $\dot{\Theta} > 0$

Considerations **C1** and **C2** guarantee the existence and allow the computation of a lower and an upper bound for the cost.

## VII. SCALAR CASE

In this section we provide some results for the scalar case. The first result describes performance bounds. The second result establishes that the microscopic dynamics is exponentially asymptotically stable almost surely. The third result shows that the macroscopic dynamics is exponentially asymptotically stable.

### A. Performance bounds

Consider the scalar version of the dynamics (??):

$$dx_t = [\alpha x_t + \beta u_t + \mu \zeta_t] dt + \sigma(x_t) dB_t. \quad (13)$$

In the scalar case  $Q \in \mathbb{R}$ ,  $\bar{Q} \in \mathbb{R}$ ,  $R = \mathbb{R}$  and  $S(\tilde{m}_t) = \hat{h}\tilde{m}_t = s\tilde{m}_t \in \mathbb{R}$ . From **C1** and **C2** we can deduce that there exist two variables that approximate from above and from below the evolution of the common state.

First, consider the following assumption.

*Assumption 1:* There exists  $\underline{\theta}$  and  $\kappa$  such that

$$\kappa \bar{m}_t \geq \frac{d}{dt} \bar{m}_t = \alpha \bar{m}_t + \beta \bar{u}_t^* + \sigma \bar{\zeta}_t^* \geq \underline{\theta} \bar{m}_t, \quad \text{for all } t \in [0, T].$$



Possible values for  $\kappa$  and  $\underline{\theta}$  are the one obtained with maximal and minimal congestion, namely,

$$\underline{\theta} = \left[ \alpha + \left( -\frac{\beta^2}{R} + \frac{\sigma^2}{2\gamma^2} \right) (P_{11}(t) + P_{12}(t)) \right], \quad (14)$$

$$\kappa = \left[ \alpha + \left( -\frac{\beta^2}{R + s\bar{m}_0} + \frac{\sigma^2}{2\gamma^2} \right) (P_{11}(t) + P_{12}(t)) \right]. \quad (15)$$

In other words, the main idea is to approximate the mean distribution  $\bar{m}_t$  from below by  $\tilde{m}_t$  and from above by  $\hat{m}_t$ . To do this, we wish the following condition to hold:

$$\tilde{m}_t \leq \bar{m}_t \leq \hat{m}_t, \quad \text{for all } t \in [0, T]. \quad (15)$$

The above reasoning is particularly meaningful when  $\sigma = 0$ , in which case we consider the following dynamics:

$$\begin{cases} \frac{d}{dt} \tilde{m}_t = \left( \alpha - \frac{2\beta(P_{11} + P_{12})}{R + s\tilde{m}_t} \right) \tilde{m}_t, \\ \frac{d}{dt} \hat{m}_t = \left( \alpha - \frac{2\beta(P_{11} + P_{12})}{R} \right) \hat{m}_t := -\theta_t \hat{m}_t, \\ \frac{d}{dt} \tilde{m}_t = \left( \alpha - \frac{2\beta(P_{11} + P_{12})}{R + s\tilde{m}_0} \right) \tilde{m}_t := -\kappa \tilde{m}_t, \\ \tilde{m}_0 = \hat{m}_0 = \bar{m}_0. \end{cases} \quad (15)$$

In the above system of equations, we have set

$$\begin{cases} \theta_t = -\alpha + \frac{2\beta(P_{11} + P_{12})}{R}, \\ \kappa = -\alpha + \frac{2\beta(P_{11} + P_{12})}{R + s\tilde{m}_0}. \end{cases} \quad (15)$$

We are then in the position to establish the following result, which provides a lower bound for the value function in (??) when  $\sigma = 0$ .

*Theorem 2:* Let  $\sigma = 0$ . Then  $\mathcal{V}_t(X)$  approximates  $v(x)$  from below, i.e.,

$$\mathcal{V}_t(X) \leq v_t(x), \quad \forall X, x, t. \quad (15)$$

Furthermore, the approximation error is upper bounded as established by the inequality below

$$\begin{aligned} & \frac{d}{dt} (v_t(x) - \mathcal{V}_t(x)) \\ & \leq s \left( \frac{2\beta(P_{11} + P_{12})}{R} \right)^2 \bar{m}_0^3 \left[ e^{-3\kappa t} - e^{-(\theta_t - 2\kappa)t} \right]. \end{aligned} \quad (15)$$

*Proof:* Given in the appendix.  $\square$

### B. Exponential asymptotic stability of microscopic dynamics

The stochastic differential equation describing the closed-loop system has an exponentially and asymptotically stable equilibrium. To see this from (??)-(??) rewrite the dynamics for  $x_t$  in (??) as

$$\begin{aligned} dx_t &= [\alpha x_t + \beta u_t^* + \sigma \zeta_t^*] dt + \sigma x_t d\mathcal{B}_t \\ &= \left[ \alpha x_t + \left( -\frac{2\beta^2}{R + s\tilde{m}_t} + \frac{\sigma^2}{\gamma^2} \right) (P_{11}(t)x_t + P_{12}(t)\tilde{m}_t) \right] dt \\ & \quad + \sigma x_t d\mathcal{B}_t \quad t \in (0, T], x_0 \in \mathbb{R}, \end{aligned}$$

and consider the following assumption, ensuring that the evolution of the state is bounded from above by an exponential decay.

*Assumption 2:* There exists  $\kappa > 0$  such that

$$-\kappa x_t \geq \left[ \alpha + \left( -\frac{2\beta^2}{R + s\tilde{m}_t} + \frac{\sigma^2}{\gamma^2} \right) P_{11}(t) \right] x_t + \left[ \left( -\frac{2\beta^2}{R + s\tilde{m}_t} + \frac{\sigma^2}{\gamma^2} \right) P_{12}(t) \right] \tilde{m}_t. \quad (14)$$

The analysis is then performed within the framework of stochastic stability theory [?]. To this end, consider the infinitesimal generator

$$\mathcal{L} = \frac{1}{2} \sigma^2 x_t^2 \frac{d^2}{dx_t^2} - \kappa x_t \frac{d}{dx_t}, \quad (15)$$

and the Lyapunov function  $V(x) = x^2$ . The stochastic derivative of  $V(x)$  is obtained by applying (??) to  $V(x)$ , which yields

$$\mathcal{L}V(x_t) = \lim_{dt \rightarrow 0} \frac{\mathbb{E}V(x_{t+dt}) - V(x_t)}{dt} = [\sigma^2 - 2\kappa]x_t^2.$$

*Proposition 7.1:* [[?]] Let Assumption ?? hold. If  $V(x) \geq 0$ ,  $V(0) = 0$  and  $\mathcal{L}V(x) \leq -\eta V(x)$  on  $\hat{Q}_\epsilon := \{x : V(x) \leq \epsilon\}$  for some  $\eta > 0$  and for arbitrarily large  $\epsilon$ , then the origin is asymptotically stable “with probability one”, and

$$P_{x_0} \left\{ \sup_{T \leq t < +\infty} x_t^2 \geq \lambda \right\} \leq \frac{V(x_0)e^{-\psi T}}{\lambda}$$

for some  $\psi > 0$ .

From Proposition ?? we have the following result, establishing exponential stochastic stability of the mean-field equilibrium.

*Corollary 7.1:* Let Assumption ?? hold. If  $[\sigma^2 - 2\kappa] < 0$  then  $\lim_{t \rightarrow \infty} x_t = 0$  almost surely and

$$P_{x_0} \left\{ \sup_{T \leq t < +\infty} x_t^2 \geq \lambda \right\} \leq \frac{V(x_0)e^{-\psi T}}{\lambda}$$

for some  $\psi > 0$ .

### C. Mean-field equilibrium for macroscopic dynamics

Let Assumption ?? hold. We can approximate the mean-field equilibrium, which is captured by the evolution of  $\bar{m}_t$  over the horizon  $(0, T]$ , as follows:

$$\frac{d}{dt} \bar{m}_t \leq -\kappa \bar{m}_t, \quad t \in (0, T], m_0 \in \mathbb{R} \times [0, T],$$

which yields the upper bound for  $\bar{m}_t$ :

$$\bar{m}_t \leq \bar{m}_0 e^{-\kappa t}, \quad t \in (0, T], x_0 \in \mathbb{R}.$$

Essentially, the inequality above describes converging linear dynamics which upper bound the time evolution of  $\bar{m}_t$ , for all  $t \in (0, T]$ . As a result

$$\begin{aligned} \frac{d}{dt} \bar{m}_t &= \left[ \alpha + \left( -\frac{2\beta^2}{R + s\bar{m}_t} + \frac{\sigma^2}{\gamma^2} \right) (P_{11}(t) + P_{12}(t)) \right] \bar{m}_t \\ & \quad t \in (0, T], x_0 \in \mathbb{R}. \end{aligned}$$

Actually, we can derive a differential equation describing the evolution of the mean distribution which represents a bound, namely

$$\begin{cases} \bar{m}_t = \bar{m}_0 e^{\rho t} \\ \rho = \alpha + \left( -\frac{2\beta^2}{R + s\bar{m}_t} + \frac{\sigma^2}{\gamma^2} \right) (P_{11}(t) + P_{12}(t)). \end{cases}$$

The equation above corresponds to saying that the mean distribution converges exponentially to zero in absence of stochastic disturbances (the Brownian motion), under the assumption that  $\rho$  is strictly negative.

$\alpha$	$\beta$	$Q$	$R$	$\theta_t$	$\bar{Q}$	$\gamma$	$\bar{m}_0$
0.1	2	2	1	10	0.1	1	80

TABLE II  
SIMULATION PARAMETERS

### VIII. NUMERICAL STUDIES

In this section we present numerical studies of the robust mean-field game with dynamics (??). In particular, we study in more detail the numerical example introduced in [?]. After computing optimal controls and worst-case disturbances using the heuristic method illustrated earlier, we simulate the macroscopic evolution of the population of players. Consider a system consisting of  $n = 10^3$  indistinguishable players, where each player seeks to minimize a cost functional of the form (??) subject to an adversary disturbance. *I.e.* consider Problem ???. Note that for this scalar problem we use the same notation as in Section ??, *i.e.*  $S(\bar{m}_t) = s\bar{m}_t$ . An approximate solution to the mean-field game is found using the method introduced in Section ???. The matrix

$$P = \begin{bmatrix} P_{11}(\bar{m}) & 0 \\ 0 & \frac{\bar{Q}}{4\theta_t} \end{bmatrix},$$

with  $P_{11}(\bar{m}) = \sqrt{(\sigma^2 + 2\alpha)^2 + 8(\frac{\beta^2}{R+s\bar{m}} - \frac{\sigma^2}{2\gamma^2})} + \sigma^2 + 2\alpha$ , is the positive definite solution to the algebraic Riccati equation (??). The resulting control (??) and the disturbance (??) are adopted using this solution. The set of states is discretized and numerical results are obtained using the algorithm in Table ???. The quantities  $\sigma(x_t)$  and  $\mu$  determine the influence of the Brownian motion  $\mathcal{B}_t$  and the disturbance  $\zeta_t$ , respectively. In this example consider the case in which  $\sigma(x_t)$  is constant, *i.e.*  $\sigma(x_t) = \sigma$  and we set  $\mu = \sigma$ . Simulations have been run for two different values of  $\sigma$ , namely  $\sigma_0 = 0$ ,  $\sigma_1 = 0.1$ . The selection  $\sigma = \sigma_0$  corresponds to the case in which there is no disturbance and dynamics (??) is deterministic. The simulations have also been run for two different values of  $s$ , namely  $s_1 = 0.5$  and  $s_2 = 1.5$ . Note that large values of  $s$  correspond to large penalties when congestion occurs. The remainder of the parameters are as shown in Table ??.

Figure ?? shows the time histories of the states of the players with the weights  $s = s_1$  (top row) and  $s = s_2$  (bottom row) and the parameters  $\sigma = \sigma_0$  (left column) and  $\sigma = \sigma_1$  (right column). Figure ?? shows the distribution,  $m_t$ , of the players' states at different times for the four different selections of parameters. The initial and final distributions are indicated by the dashed and solid curves, respectively, whereas the distribution at intermediate times are denoted by the dotted curves. Figure ?? shows the time histories of the mean,  $\bar{m}_t$ , (left) and the standard deviation (right) for  $s = s_1$  (top) and  $s = s_2$  (bottom). The solid curves correspond to  $\sigma = \sigma_1$  whereas the dashed lines correspond to  $\sigma = \sigma_0$ .

Note that in all four cases the players successfully drive their states to zero. However, for a given value of the parameter  $s$ , the convergence is fastest in the absence of noise and disturbances, *i.e.* when  $\sigma = \sigma_0$ . Figure ?? shows the time histories of the control actions (??) of the players with

**Input:** Set of parameters as in Table ??.

**Output:** Distribution function  $m_t$ , mean  $\bar{m}_t$  and standard deviation  $\text{std}(m_t)$ .

- 1 : **Initialize.** Generate  $x_0$  given  $\bar{m}_0$  and  $\text{std}(m_0)$
- 2 : **for** time  $t = 0, 1, \dots, T - 1$  **do**
- 3 :   **if**  $t > 0$ , **then** compute  $m_t$ ,  $\bar{m}_t$ , and  $\text{std}(m_t)$
- 4 :   **end if**
- 5 :   **for** player  $i = 1, \dots, n$  **do**
- 6 :     Compute control  $\tilde{u}$  using current  $\bar{m}_t$
- 7 :     compute new state  $x_{t+1}$  by executing (??)
- 8 :   **end for**
- 9 : **end for**
- 10 : **STOP**

TABLE III  
SIMULATION ALGORITHM

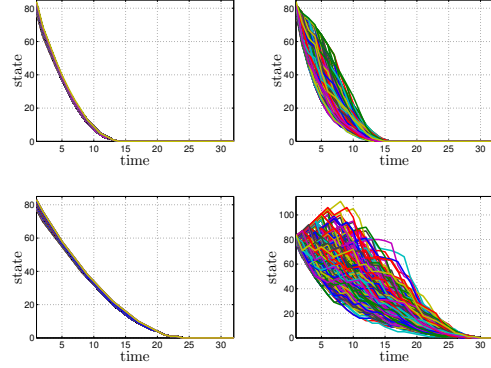


Fig. 2. Time histories of the state of each player. Top row:  $s = s_1$ , bottom row:  $s = s_2$ , left column:  $\sigma = \sigma_0$ , right column:  $\sigma = \sigma_1$ .

$s = s_1$  (top row) and  $s = s_2$  (bottom row), and  $\sigma = \sigma_0$  (left column) and  $\sigma = \sigma_1$  (right column). For the case in which  $\sigma = \sigma_1$ , it is clear that when  $s = s_1$  is selected the players put a larger effort at the beginning of the simulation than when  $s = s_2$  is selected, and the same is true for  $\sigma = \sigma_0$ . Since  $s_2 > s_1$ , this implies that in the former case a larger penalty is incurred when congestion occurs and therefore one would expect the players to stall to avoid this, resulting in the convergence to the zero equilibrium being somewhat slower. The simulations are consistent with this, as for a given value of  $\sigma$  it takes more time for the players to drive their states to zero when the parameter  $s = s_2$  is selected in place of  $s = s_1$ .

### IX. CONCLUDING REMARKS

We have illustrated robust mean-field games as a paradigm for crowd-averse systems. We have discussed these systems in the context of stock market, production engineering, and dynamic demand management in power systems. We have presented a new approximation method based on the extension of the state space.

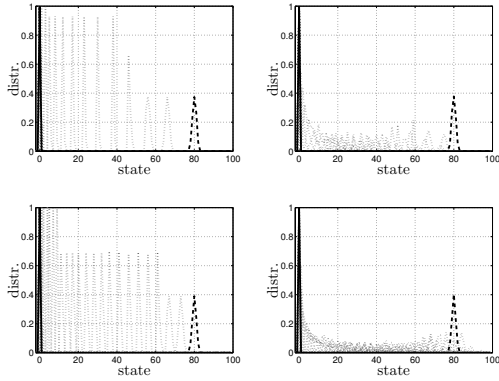


Fig. 3. The initial (dashed line), final (solid line) and intermediate (dotted lines) distribution,  $m_t$ , of the states of the players. Top row:  $s = s_1$ , bottom row:  $s = s_2$ , left column:  $\sigma = \sigma_0$ , right column:  $\sigma = \sigma_1$ .

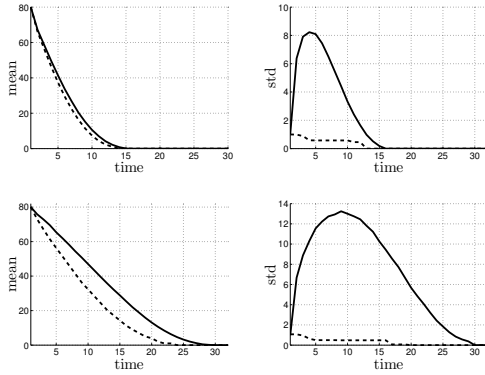


Fig. 4. Time histories of the mean  $\bar{m}_t$  (left) and the standard deviation (right) of the states of the players for  $s = s_1$  (left),  $s = s_2$  (right),  $\sigma = \sigma_0$  (dashed line) and  $\sigma = \sigma_1$  (solid line).

We can extend our study in at least three directions. These include i) the extension of the approximation method to more general cost functionals, ii) the study of the case with “local” mean-field interactions rather than “global” as in the current scenario, and iii) the analysis of crowd-seeking scenarios in contrast to the crowd-averse cases analyzed in this paper.

## REFERENCES

- [1] Y. Achdou, F. Camilli, and I. Capuzzo Dolcetta, “Mean field games: numerical methods for the planning problem”, *SIAM Journal on Control and Optimization*, vol. 50, pp. 77–109, 2012.
- [2] F. Bagagiolo and D. Bauso, “Mean-field games and dynamic demand management in power grids”, *Dynamic Games and Applications*, vol. 4, no. 2, pp. 155–176, 2014.
- [3] M. Bardi, “Explicit solutions of some Linear-Quadratic Mean Field Games”, *Network and Heterogeneous Media*, vol. 7, pp. 243–261, 2012.
- [4] T. Başar and P. Bernhard, *H<sup>∞</sup>-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, Birkhäuser, Boston, MA, 1995.

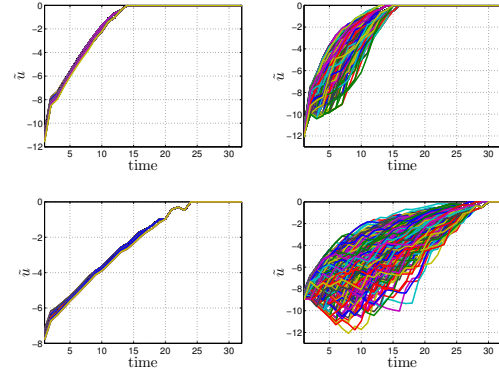


Fig. 5. Time histories of the control actions  $\tilde{u}$  of the players. Top row:  $s = s_1$ , bottom row:  $s = s_2$ , left column:  $\sigma = \sigma_0$ , right column:  $\sigma = \sigma_1$ .

- [5] D. Bauso, H. Tembine, and T. Başar, “Robust Mean Field Games with Application to Production of an Exhaustible Resource”, *Proc. of 7th IFAC Symposium on Robust Control Design*, Aalborg, Denmark, 2012.
- [6] D. Bauso, H. Tembine, and T. Başar, “Robust Mean Field Games”, *Dynamic Games and Applications*, online, doi:10.1007/s13235-015-0160-4.
- [7] D. Bauso, Q. Zhu and T. Başar, “Mixed Integer Optimal Compensation: Decompositions and Mean-Field Approximations”, *Proc. of 2012 American Control Conference*, Montreal, Canada, 2012.
- [8] F. Black and M. Scholes, “The Pricing of Options and Corporate Liabilities”, *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.
- [9] P. Cardaliaguet and P.J. Graber and A. Porretta, and D. Tonon, “Second order mean field games with degenerate diffusion and local coupling”, arXiv preprint arXiv:1407.7024v1, 2014.
- [10] D.A. Gomes and J. Saúde, “Mean Field Games Models - A Brief Survey”, *Dynamic Games and Applications*, vol. 4, no. 2, pp. 110-154, 2014.
- [11] D.A. Gomes and E. Pimentel and H. Sánchez-Morgado, “Time-dependent mean-field games in the subquadratic case”, *Communications in Partial Differential Equations*, vol. 40, no. 1, pp. 40–76, 2015.
- [12] D.A. Gomes and E. Pimentel and H. Sánchez-Morgado, “Time dependent mean-field games in the superquadratic case”, arXiv preprint arXiv:1311.6684v3, 2014.
- [13] D. Bauso and T. Mylvaganam and A. Astolfi, “Approximate solutions for crowd-averse robust mean-field games”, *2014 European Control Conference*, pp.1217-1222, 2014.
- [14] O. Gueant, J. M. Lasry, and P. L. Lions, “Mean-field games and applications”, *Paris-Princeton Lectures*, Springer, pp. 1–66, 2010.
- [15] M.Y. Huang, P.E. Caines, and R.P. Malhamé, “Large Population Stochastic Dynamic Games: Closed Loop Kean-Vlasov Systems and the Nash Certainty Equivalence Principle”, *Communications in Information and Systems*, vol. 6, no. 3, pp. 221–252, 2006.
- [16] M.Y. Huang, P.E. Caines, and R.P. Malhamé, “Large population cost-coupled LQG problems with non-uniform agents: individual-mass behaviour and decentralized  $\epsilon$ -Nash equilibria”, *IEEE Trans. on Automatic Control*, vol. 52, no. 9, pp. 1560–1571, 2007.
- [17] A. Lachapelle, J. Salomon, and G. Turinici, “Computation of Mean Field Equilibria in Economics”, *Math. Models Meth. Appl. Sci.*, vol. 20, pp. 1–22, 2010.
- [18] J.-M. Lasry and P.-L. Lions, “Mean field games”, *Japanese Journal of Mathematics*, vol. 2, pp. 229–260, 2007.

- [19] K. A. Loparo and X. Feng, "Stability of stochastic systems". *The Control Handbook*, CRC Press, pp. 1105–1126, 1996.
- [20] Z. Ma, D. S. Callaway, and I. A. Hiskens, "Decentralized Charging Control of Large Populations of Plug-in Electric Vehicles", *IEEE Transactions on Control System Technology*, vol. 21, no.1, pp. 67–78, 2013.
- [21] R. C. Merton, "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science* (The RAND Corporation), vol. 4, no. 1, pp. 141–183, 1973.
- [22] R. Pesenti, and D. Bauso, "Mean Field Linear Quadratic Games with Set Up Costs", *Dynamic Games and Applications*, vol. 3, no. 1, pp. 89–104, March 2013.
- [23] M. Roozbehani, M. A. Dahleh, and S. K. Mitter, "Volatility of Power Grids Under Real-Time Pricing", *IEEE Transactions on Power Systems*, vol. 27, no. 4, 2012.
- [24] M. Sassano, and A. Astolfi, "Dynamic Approximate Solutions of the HJ Inequality and of the HJB Equation for Input-Affine Nonlinear Systems", *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2490–2503, October 2012.
- [25] M. Sassano, and A. Astolfi, "Approximate finite-horizon optimal control without PDEs", *Systems & Control Letters*, vol. 62 pp. 97–103, 2013.
- [26] H. Tembine, Q. Zhu, and T. Başar, "Risk-sensitive mean-field stochastic differential games", *Proc. of 2011 IFAC World Congress*, Milan, Italy, August 29 - September 2, 2011.
- [27] H. Tembine, Q. Zhu, and T. Başar, "Risk-sensitive mean-field games", *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 835–850, 2014.
- [28] H. Tembine, J. Y. Le Boudec, R. ElAzouzi, E. Altman, "Mean field asymptotic of Markov decision evolutionary games", *IEEE Games 2009*.

#### APPENDIX

##### *Proof of Theorem ??.*

We first prove condition (??). To this end write the Hamiltonian as:

$$H(x_t, \partial_x v_t, m_t) = \inf_u \left\{ \frac{1}{2} (\|u_t\|_{R+S(\bar{m}_t)}^2 + \|x_t\|_Q^2) + \partial_x v_t (Ax_t + Bu_t) \right\}. \quad (11)$$

Differentiating with respect to  $u_t$  gives

$$(R + S(\bar{m}_t))u_t + B^T \partial_x v_t = 0, \quad (11)$$

which yields (??).

For the robust Hamiltonian we then have

$$\tilde{H}(x_t, p, \bar{m}_t) = H(x_t, p, \bar{m}_t) + \sup_{\zeta_t} \left\{ p^T M \zeta_t - \frac{1}{2} \gamma^2 \zeta_t^T \zeta_t \right\}$$

whose solution in terms of  $\zeta_t$  is given by  $\zeta_t^* = \frac{1}{\gamma^2} M^T \partial_x v_t$ . We now prove (??). Substituting the above expression for  $\zeta_t^*$  back in the expression for the robust Hamiltonian we then have

$$\tilde{H}(x_t, p, \bar{m}_t) = H(x_t, p, \bar{m}_t) + \frac{1}{2} \frac{1}{\gamma^2} \partial_x v_t^T M M^T \partial_x v_t.$$

Then the mean-field system associated to the robust mean-field game introduced in Problem ?? is given by

$$\begin{cases} \partial_t v_t + H(x_t, \partial_x v_t, \bar{m}_t) + \frac{1}{2\gamma^2} \partial_x v_t^T M M^T \partial_x v_t \\ + \frac{1}{2} \sum_{i=1}^n \sigma_i^2 x_i^2 \partial_{x_i x_i}^2 v_t = 0, \text{ in } \mathbb{R}^n \times [0, T], \\ v_T(x) = \frac{1}{2} x^T \Phi x, \text{ in } \mathbb{R}^n, \\ \partial_t m_t + \partial_x (m_t \partial_p H(x, \partial_x v_t, m)) \\ + \frac{1}{\gamma^2} \partial_x (m_t M M^T \partial_x v_t) \\ - \frac{1}{2} \sum_{i=1}^n \partial_{x_i x_i}^2 [\sigma_i^2 x_i^2 m_t] = 0, \text{ in } \mathbb{R}^n \times [0, T], \\ m_0(x) = d(x) \text{ in } \mathbb{R}^n, \\ \bar{m}_t := \int_{\mathbb{R}} x m_t dx, \text{ in } [0, T], \end{cases} \quad (10)$$

where  $d(x)$  is the initial population state distribution and  $g$  the terminal payoff.

First note that the second and last equations are the boundary conditions and derive straightforwardly from Bellman equations and the evolution of the state.

To prove the first equation, which is a PDE corresponding to the HJI equation, replace  $u$  in the Hamiltonian (??) by its expression (??), i.e.

$$\begin{aligned} H(x_t, \partial_x v_t, m_t) &= \frac{1}{2} (\|u_t^*\|_{R+S(\bar{m}_t)}^2 + \|x_t\|_Q^2) \\ &\quad + (\partial_x v_t)^T (Ax_t + Bu_t^*) \\ &= \frac{1}{2} \left[ x_t^T Q x_t + u_t^{*T} \left( R + S(\bar{m}_t) \right) u_t^* \right] \\ &\quad + (\partial_x v_t)^T A x_t + (\partial_x v_t)^T B u_t^* \\ &= -\frac{1}{2} (\partial_x v_t)^T [B(R + h\bar{z}_t)^{-1} B^T] \partial_x v_t \\ &\quad + (\partial_x v_t)^T A x_t + \frac{1}{2} x_t^T Q x_t. \end{aligned}$$

Using the above expression of the Hamiltonian in the HJI equation in (??), we obtain the HJI in (??).

To prove the third equation, which is a PDE representing the FPK equation, we simply bring (??) into the FPK equation in (??), and this concludes the proof.

##### *Proof of Theorem ??.*

Let us start by noting that, from convexity on  $\bar{m}$ , we can write the following HJI inequality as

$$\begin{cases} \partial_t v_t + c(x_t, \tilde{u}_t, \tilde{m}_t) - \gamma^2 \tilde{\zeta}_t^2 + \partial_x v_t (\alpha x_t \\ + \beta \tilde{u}_t + \sigma \tilde{\zeta}_t) + \frac{1}{2} \sigma^2 x^2 \partial_{xx}^2 v_t \leq 0, \\ \text{in } \mathbb{R} \times [0, T], \\ v_T(x) = \frac{1}{2} \Phi x^2, \text{ in } \mathbb{R}. \end{cases} \quad (10)$$

In other words

$$\tilde{V}_t(x) := \int_0^T (c(x_t, \tilde{u}_t, \tilde{m}_t) - \gamma^2 \tilde{\zeta}_t^2) dt \leq v_t(x).$$

Then, for the approximation error we have

$$\begin{aligned} e(t) &:= v_t(x) - \mathcal{V}_t(X) \\ &\leq \int_0^T s \tilde{u}_\tau^2 (\bar{m}_\tau - \hat{m}_\tau) d\tau \\ &\leq \int_0^T s \tilde{u}_\tau^2 \bar{m}_0 \left( \frac{\bar{m}_\tau}{\bar{m}_0} - \frac{\hat{m}_\tau}{\bar{m}_0} \right) d\tau, \end{aligned} \quad (10)$$

for any  $\bar{m}_t$ ,  $\hat{m}_t$ , and  $\tilde{m}_t$  satisfying (??). Now, from (??)-(??), the above inequalities can be rewritten as

$$e(t) \leq \int_0^t s \tilde{u}_\tau^2 \bar{m}_0 \left[ e^{-\kappa\tau} - e^{-\theta_t\tau} \right] d\tau, \quad (10)$$

from which, after differentiating with respect to  $t$  and substituting  $\tilde{u}_t$  with the expression in (??), we obtain

$$\begin{aligned} \dot{e}(t) &\leq s \left( \frac{2\beta(P_{11}+P_{12})}{b} \tilde{m}_t \right)^2 \bar{m}_0 \left[ e^{-\kappa\tau} - e^{-\theta_t\tau} \right] \\ &\leq s \left( \frac{2\beta(P_{11}+P_{12})}{b} \right)^2 \left( \frac{\hat{m}_t}{\bar{m}_0} \right)^2 \bar{m}_0^3 \left[ e^{-\kappa\tau} - e^{-\theta_t\tau} \right] \\ &\leq s \left( \frac{2\beta(P_{11}+P_{12})}{b} \right)^2 \bar{m}_0^3 \left[ e^{-3\kappa\tau} - e^{(-\theta_t-2\kappa)\tau} \right], \end{aligned}$$

which proves the thesis.



**Dario Bauso** received his Laurea degree in Aeronautical Engineering in 2000 and his Ph.D. degree in Automatic Control and Systems Theory in 2004 from the University of Palermo, Italy. Since 2015 he has been with the Department of Automatic Control and Systems Engineering, The University of Sheffield (UK), where he is currently Reader in Control and Systems Engineering. Since 2005 he has also been with the Dipartimento di Ingegneria

Chimica, Gestionale, Informatica, Meccanica, University of Palermo (Italy), where he is currently Associate Professor of Operations Research. From 2012 to 2014 he was also research fellow at the Department of Mathematics, University of Trento (Italy). He has been academic visitor in several universities. From October 2001 to June 2002, he was a Visiting Scholar at the Mechanical and Aerospace Engineering Department, University of California, Los Angeles. In 2010 he was short-term visiting scholar at the Department of Automatic Control of Lund University and at the Laboratory of Information and Decision Systems of the Massachusetts Institute of Technology. In 2013 he was visiting lecturer at the Department of Engineering Science, University of Oxford (UK), and at the Department of Electrical and Electronic Engineering Department of Imperial College London, London (UK).

His research interests are in the field of Optimization, Optimal and Distributed Control, and Game Theory. Since 2010 he is member of the Conference Editorial Board of the IEEE Control Systems Society. Since 2011 he is Associate Editor of Dynamic Games and Applications. In 2012 he has been appointed Associate Editor for the IEEE Transactions on Automatic Control. He has also been general chair of the 6th Spain, Italy, and Netherlands Meeting on Game Theory (SING 6), and general chair of the 4th Workshop on Stochastic Methods in Game Theory, International School of Mathematics G. Stampacchia, to be held in Erice, Italy, in September 2013.



**Alessandro Astolfi** was born in Rome, Italy, in 1967. He graduated in electrical engineering from the University of Rome in 1991. In 1992 he joined ETH-Zurich where he obtained a M.Sc. in Information Theory in 1995 and the Ph.D. degree with Medal of Honor in 1995 with a thesis on discontinuous stabilization of nonholonomic systems. In 1996 he was awarded a Ph.D. from the University of Rome "La Sapienza" for his work on nonlinear robust

control. Since 1996 he has been with the Electrical and Electronic Engineering Department of Imperial College London, London (UK), where he is currently Professor in Nonlinear Control Theory and Head of the Control and Power Group. From 1998 to 2003 he was also an Associate Professor at the Dept. of Electronics and Information of the Politecnico of Milano. Since 2005 he has also been a Professor at Dipartimento di Ingegneria Civile e Ingegneria Informatica, University of Rome Tor Vergata. He has been a visiting lecturer in "Nonlinear Control" in several universities, including ETH-Zurich (1995-1996); Terza University of Rome (1996); Rice University, Houston (1999); Kepler University, Linz (2000); SUP-ELEC, Paris (2001).

His research interests are focused on mathematical control theory and control applications, with special emphasis for the problems of discontinuous stabilization, robust and adaptive control, observer design and model reduction. He is the author of more than 120 journal papers, of 30 book chapters and of over 240 papers in refereed conference proceedings. He is the recipient of the IEEE CSS A. Ruberti Young Researcher Prize (2007) and of the IEEE CSS George S. Axelby Outstanding Paper Award (2012). He is a "Distinguished Member" of the IEEE CSS. He is the author (with D. Karagiannis and R. Ortega) of the monograph "Nonlinear and Adaptive Control with Applications" (Springer-Verlag).

He is Associate Editor of Automatica, the International Journal of Control and the International Journal of Adaptive Control and Signal Processing. He is Senior Editor of the IEEE Trans. on Automatic Control and Editor-in-Chief of the European Journal of Control. He has also served in the IPC of various international conferences. He is currently the Chair of the IEEE CSS Conference Editorial Board.



**Thulasi Mylvaganam** was born in Bergen, Norway, in 1988. She received the M.Eng degree in electrical and electronic engineering from Imperial College London, UK, in 2010. In 2014 she completed her Ph.D. degree with the Control and Power research group at the department of Electrical and Electronic Engineering, Imperial College London, where she is currently a postdoctoral research assistant. Her research interests include nonlinear differential games,

multi-agent systems and their applications.