Communications in Contemporary Mathematics
(2026) 2640007 (42 pages)
© World Scientific Publishing Company
DOI: 10.1142/S0219199726400079

Mean field games without rational expectations

Benjamin Moll
Department of Economics, London School of Economics
London WC2A 2AE, UK
b.moll@lse.ac.uk

Lenya Ryzhik*
Department of Mathematics, Stanford University
Stanford, CA 94305, USA
ryzhik@stanford.edu

Received 15 June 2025
Revised 27 January 2026
Accepted 22 February 2026
Published 5 March 2026

Mean Field Game (MFG) models implicitly assume "rational expectations", meaning that the heterogeneous agents being modeled correctly know all relevant transition probabilities for the complex system they inhabit. When there is common noise, it becomes necessary to solve the "Master equation", in which the infinite-dimensional density of agents is a state variable. The rational expectations assumption and the implication that agents solve Master equations are unrealistic in many applications. We show how to instead formulate MFGs with non-rational expectations. Departing from rational expectations is particularly relevant in "MFGs with a low-dimensional coupling", i.e. MFGs in which agents' running reward function depends on the density only through low-dimensional functionals of this density. This happens, for example, in most macroeconomics MFGs in which these low-dimensional functionals have the interpretation of "equilibrium prices". In MFGs with a low-dimensional coupling, departing from rational expectations allows for completely sidestepping the Master equation and for instead solving much simpler finite-dimensional HJB equations. We introduce an adaptive learning model as a particular example of non-rational expectations and discuss its properties.

Keywords: Mean field games; rational expectations; master equation; adaptive learning.

Mathematics Subject Classification 2020: 35Q89, 91A26

1. Introduction

Economists and mathematicians cast models with a large number of interacting agents as Mean Field Games (MFGs), a coupled system of a backward-in-time

*Corresponding author.

Hamilton–Jacobi–Bellman (HJB) equation for agents' value function and a forward-in-time Fokker–Planck type equation for the agents' density. These equations describe the Nash equilibrium of a game played by a large number of agents experiencing fluctuations that are independent from each other. When there is common noise, the backward–forward stochastic coupling becomes more complicated and, to find their optimal strategy, the model agents need to solve a Master equation, that is, an equation in which the infinite-dimensional density is a state variable. This Master equation therefore suffers from an extreme version of the curse of dimensionality and has been nicknamed "Monster equation" due to its complexity.[a] While this complexity makes the Master equation a fascinating mathematical object, it also limits its practical applicability.

An underappreciated fact is that MFGs not only impose Nash equilibrium but also assume "rational expectations", meaning that the heterogeneous agents being modeled correctly know all relevant transition probabilities governing the complex system they inhabit. We argue that this assumption is unrealistic in many applications and show how to instead formulate a more general class of MFGs with non-rational expectations. Furthermore, we show that departing from rational expectations can drastically simplify the complexity of agents' optimization problems in certain applications and may allow sidestepping the unrealistically complex infinite-dimensional Master equation altogether.

After spelling out general MFGs without rational expectations, we focus on a class of MFGs we term "MFGs with a low-dimensional coupling". In these MFGs, the agents' running reward function depends on the density only through low-dimensional functionals of this density, for example a low-dimensional vector of moments of this density. That is, model agents do not directly "care about" the density (in the sense that their rewards do not depend on it) and instead care only about the low-dimensional functionals. Most MFGs in macroeconomics are examples of this class of MFGs, with these low-dimensional functionals corresponding to "equilibrium prices".[b] We show that, in MFGs with a low-dimensional coupling, departing from rational expectations generally results in a much simpler finite-dimensional HJB equation in place of the infinite-dimensional Master equation.

One of this paper's arguments is that MFGs with rational expectations and the Master equation are unrealistically complex as models of human decision-making. MFGs with a low-dimensional coupling illustrate this clearly: under the rational expectations assumption, the low-dimensional coupling neither simplifies the formulation of the MFG nor that of the Master equation in any straightforward way. Indeed, to compute the low-dimensional functionals ("prices" in macroeconomics

[a] Of course, this complexity does not arise in MFGs in which individual states take only a small number of possible values (say two or three) so that the Master equation is finite- and low-dimensional.

[b] In economics, these MFGs with coupling via low-dimensional equilibrium prices are known as "heterogeneous agent models".

applications) in the rational expectations regime, one needs to compute the full density of the agents and there is no closed system that includes the "prices" alone and not the full agents' density. The agents being modeled are assumed to perform the same computations. It is for this reason that the Master equation is an infinite-dimensional PDE despite model agents only "caring about" much lower-dimensional "prices". The present paper criticizes the use of the Master equation in MFGs with a low-dimensional coupling and calls for developing alternative low-dimensional approximations that take advantage of these models' special structure. This part of the paper is a "mathematics translation" using the language of partial differential equations (PDEs) of an economics paper [39] which criticizes the use of the rational expectations assumption in macroeconomics MFGs.[c] We should also be clear that the main focus of this paper is on modeling and not on the mathematical analysis of the proposed models.

Our main results are contained in Sec. 5 which formulates MFGs without rational expectations. Sections 2–4 contain background material and building blocks that are useful for understanding such MFGs. In particular, before considering the full MFG case, Sec. 4 introduces the idea of departing from rational expectations in the simpler case of a single agent solving a stochastic control problem.

As we explain in Sec. 5, the key feature of MFGs without rational expectations is that each agent uses a perceived trajectory of the future empirical density of the other agents that does not necessarily coincide with the density's actual equilibrium trajectory. This results in a model that, formally, still has the backward–forward feature of the MFG but, for a specified perceived trajectory of the density, at any given time, the agents' strategy can be computed solely by using the backward-in-time HJB equation, without resorting to the forward-in-time density equation. This system is the analogue of what economists call a "temporary equilibrium" [23, 24, 26, 35]. After spelling out this more general system we show that, as expected, we recover the familiar backward–forward MFG system in the special case of rational expectations, i.e. when agents' perceived trajectory of the density of the other agents coincides with the density's actual equilibrium trajectory. Analogously, we show how to formulate MFGs without rational expectations in the case with common noise and that we recover the Master equation in the special case with rational expectations.

In Sec. 6, we consider non-rational expectations in MFGs with a low-dimensional coupling. Each agent now uses a perceived trajectory of the future vector of "prices" that does not necessarily coincide with the actual trajectory of equilibrium prices. The result is again a system in which, for a specified perceived trajectory of these prices, at any given time, the agents' strategy can be computed from the HJB

[c] In macroeconomics MFGs, forward-looking decision-makers are assumed to forecast equilibrium prices by forecasting functionals of infinite-dimensional densities. But it seems self-evident that real-world households and firms do not forecast prices in this way and instead solve simpler approximate problems.

equation without resorting to the Fokker–Planck equation. Importantly, in the case with common noise, for a specified perceived law of motion of future prices that imposes the Markov property, agents solve a simple finite-dimensional HJB equation. That is, departing from rational expectations can completely sidestep the infinite-dimensional Master equation.

Sections 5 and 6 consider MFGs in which agents hold beliefs about future evolution or future prices that are specified outside the model (the "temporary equilibrium" idea). In Sec. 7, we instead explain how such beliefs may be determined "inside the model" via some form of learning. Specifically, we introduce an adaptive learning model that has the same key property as the models with exogenously-specified beliefs we just discussed: at any given time, and given the current prices, the agents' strategy can be computed solely from the backward-in-time HJB equation, without resorting to the forward-in-time density equation. However, in contrast to the models with exogenous beliefs, with adaptive learning, agents update their beliefs in the face of new information so that, over time, their perceived trajectory of equilibrium prices may approximate the corresponding actual trajectory. Finally, we also discuss some other promising directions, in particular reinforcement learning (RL) and other stochastic approximation algorithms.

In Sec. 8, for the sake of exposition and completeness, we explain how the arguments of this paper can be adapted to the discrete-time case. Section 9 concludes.

2. Background: Mean Field Games and Rational Expectations

In this section, we briefly review the basics of MFGs. The authors of [12–16, 32, 43] provide more complete treatments. We use the standard formulation in the MFG literature with small modifications explained below. We then briefly explain the rational expectations assumption.

2.1. *Backward–forward MFG system*

Let us recall the setup of MFGs. Consider a system of $N \gg 1$ individual agents (players) at positions (states) $X_{i,t} \in \mathbb{R}^n$, $i = 1, \ldots, N$, with $0 \le t \le T$, where T is a fixed terminal time that is sometimes taken as $T = +\infty$.

Given $t \ge 0$ and initial state $x \in \mathbb{R}^n$ (which differs across agents), each agent's state evolves according to the stochastic differential equation (SDE)

$$dX_{i,s} = \alpha_{i,s} ds + \sqrt{2\nu} dB_{i,s}, \quad X_{i,t} = x, \;\; t \le s \le T. \tag{2.1}$$

Here $\alpha_{i,s} \in A \subset \mathbb{R}^n$ is a control which is optimally chosen by each agent (in a way prescribed below), $B_{i,s}$ is a standard n-dimensional Brownian motion, and $\nu \ge 0$ is a parameter measuring the strength of the fluctuations affecting individual agents. The individual Brownian motions $B_{i,s}$, $i = 1, \ldots, N$, are independent and capture *idiosyncratic* risk. The intuition is that this risk averages to a deterministic effective mean-field effect when $N \gg 1$, as far as the evolution of the agents' density is concerned.

To choose the control $\alpha_{i,s}$ in (2.1), each agent solves an optimization problem for the value function $u_N : \mathbb{R}^n \times [0,T] \to \mathbb{R}$ defined by

$$u_N(x,t) = \max_{\alpha_i \in A} \mathbb{E}\left[\int_t^T e^{-\rho t} R(X_{i,s}, \alpha_{i,s}, m_N(s,\cdot))ds + e^{-\rho(T-t)} V(X_{i,T}, m_N(T,\cdot))\right] \tag{2.2}$$

subject to $X_{i,s}$ solving (2.1). Here,

$$m_N(t,x) = \sum_{i=1}^N \delta(x - X_{i,t}), \quad x \in \mathbb{R}^n, \tag{2.3}$$

is the empirical measure of the collection of agents, $\rho \geq 0$ is a discount rate, $R(x,\alpha,m)$ is a running reward function that depends on the state $X_{i,s}$, the control $\alpha_{i,s}$ and the empirical density $m_N(s,\cdot)$, and $V(x,m)$ is a prescribed terminal value at time T. The expectation in (2.2) is taken with respect to the idiosyncratic noises $B_{i,s}$, $1 \leq i \leq N$.

In this setting, the dynamics of all agents are identical except that they experience different realizations of $B_{i,t}$. That is, they are *ex-ante* identical (the functions R, V, and so on are identical for all agents) but *ex-post* heterogeneous in $X_{i,t}$ because of the different realizations of idiosyncratic risk $B_{i,t}$. One small modification to the standard formulation in the MFG literature is that here the agents maximize their objectives rather than minimize them.

Note that each agent's running reward R depends on the overall system's state, the empirical density m_N. Agents optimally choose the control α taking the future evolution of m_N as given. We denote the optimal policies, that is, the optimally chosen α in (2.2), by π (see (2.6)). Because R depends on m_N, so do the optimal policies π and the value function u_N. In turn, the evolution of the empirical density m_N depends on each agent's optimal policy that appears in (2.1). We are thus considering the Nash equilibrium of a game between a large number of statistically identical players.

The backward–forward MFG system arises in the limit $N \to +\infty$ of a large number of agents and is a coupled system of a HJB equation for the limit $u(x,t)$ of the value functions $u_N(x,t)$ and a Fokker–Planck equation for the limiting empirical density of the agents $m(x,t)$:

$$\begin{aligned} \rho u - \partial_t u &= H(x, \nabla u, m) + \nu \Delta u && \text{in } \mathbb{R}^n \times (0,T), \\ \partial_t m &= -\mathrm{div}(\nabla_\lambda H(x, \nabla u, m)\ m) + \nu \Delta m && \text{in } \mathbb{R}^n \times (0,T), \\ m(0) &= m_0, \quad u(x,T) = V(x, m(T)) && \text{in } \mathbb{R}^n. \end{aligned} \tag{2.4}$$

Here, H is the Hamiltonian[d]

$$H(x,\lambda,m) := \max_{\alpha \in A} \{R(x,\alpha,m) + \lambda \cdot \alpha\}, \quad \lambda \in \mathbb{R}^n, \tag{2.5}$$

[d] We use λ instead of the more standard p to denote the dual "momentum" variable because p will denote the price vector below.

and the policy function (optimal control) of each agent, defined as

$$\pi(x,t) \equiv \arg\max_{\alpha\in A}\{R(x,\alpha,m) + \alpha\cdot\nabla u(x,t)\}, \tag{2.6}$$

is given by

$$\pi(x,t) = \nabla_\lambda H(x,\nabla u(x,t),m(t)). \tag{2.7}$$

2.2. *Master equation with common noise*

We next introduce *common noise* (in MFG terminology) or *aggregate uncertainty* (in macroeconomics terminology). Typically, as, for example, in [3, 4, 13], this is done by directly adding an additional noise to the dynamics of individual agents in (2.1)

$$dX_{i,t} = \alpha_{i,t}dt + \sqrt{2\nu}dB_{i,t} + \sqrt{2\beta}dW_t, \quad X_{i,0} = x. \tag{2.8}$$

Here, the extra noise W_t is identical for all agents. In this formulation, the agents density m_t satisfies a *stochastic* Fokker–Planck equation

$$dm_t = [-\mathrm{div}(\nabla_\lambda H(x,\nabla_x u_t,m_t)\ m_t) + (\nu+\beta)\Delta_x m_t]dt - \mathrm{div}(m_t\sqrt{2\beta}dW_t), \tag{2.9}$$

rather than the standard Fokker–Planck equation that appears in (2.4).

For the sake of simplicity of the presentation, we will discuss now a slightly different model, where the common noise does not directly affect the evolution of the individual agents themselves. Instead, in this model, the common noise directly affects the running reward function. This will affect the optimal policy π and thus the evolution of the individual agents as well. Bertucci and Meynard [7, 8] use a similar approach.

Specifically, we introduce an additional state variable $Z_t \in \mathbb{R}^k$ ("the aggregate state"), with some $k \ll N$, that evolves according to an SDE:

$$dZ_t = \mu_z(Z_t)dt + \sqrt{2\beta}dW_t, \quad Z_0 = z, \tag{2.10}$$

with some drift $\mu_z(z)$ and $\beta \geq 0$. Here, W_t is a standard k-dimensional *common* Brownian motion that — in contrast to the idiosyncratic Brownian motions $B_{i,t}$ — affects *all agents* simultaneously via Z_t.

The main assumption is that the running reward function R in (2.2) now depends on the aggregate state Z_t:

$$u_N(x,t) = \max_{\alpha_i\in A}\mathbb{E}\left[\int_t^T e^{-\rho t}R(X_{i,s},Z_s,\alpha_{i,s},m_N(s,\cdot))ds \right.$$
$$\left. + e^{-\rho(T-t)}V(X_{i,T},Z_T,m_N(T,\cdot))\right]. \tag{2.11}$$

As usual in the common noise setting, the admissible controls $\alpha_{i,s}$ in (2.11) need to be $\mathcal{F}_s$-measurable: they cannot depend on the future. This, of course, is also true for the optimal control $\pi(X_{i,t},Z_t,m_t,t)$.

If the reward function R is non-separable between Z_t and $\alpha_{i,t}$ (which is the relevant assumption in macroeconomics applications), the optimal policy π will depend on the aggregate state Z_t and therefore so will the dynamics of $X_{i,t}$:

$$\begin{aligned} dX_{i,t} &= \pi(X_{i,t}, Z_t, m_t, t)dt + \sqrt{2\nu}dB_{i,t}, \\ dZ_t &= \mu_z(Z_t)dt + \sqrt{2\beta}dW_t. \end{aligned} \tag{2.12}$$

Let us comment that this setting is a special case of the standard common noise formulation in (2.8) but with a degenerate diffusion. To see this, introduce $\widetilde{X}_{i,t} \in \mathbb{R}^d$, with $d = n + k$, with the two components

$$\widetilde{X}_{i,t} = \begin{bmatrix} X_{i,t} \\ Z_t \end{bmatrix}.$$

Note that the second component is identical for all agents. Then (2.12) is the special case of (2.8) in which the first component $X_{i,t}$ is only affected by the idiosyncratic noise but not by the common noise, and the second component Z_t is only affected by the common noise but not by the idiosyncratic noise. Furthermore, the second component is not controlled: $\alpha_{Z,t} \equiv 0$.

The state of the system ("the economy") is now a pair (m_t, Z_t) which evolves as

$$dm_t = [-\mathrm{div}_x(\nabla_\lambda H(x, Z_t, \nabla_x u_t, m_t)\ m_t) + \nu\Delta_x m_t]dt, \tag{2.13}$$

$$dZ_t = \mu_z(Z_t)dt + \sqrt{2\beta}dW_t. \tag{2.14}$$

In contrast to the standard MFG formulation with a common noise (2.9), the Fokker–Planck equation (2.13) for m_t is not a stochastic partial differential equation (SPDE) but a partial differential equation (PDE). Nevertheless, the solution to the system (2.13)–(2.14) is, formally, an infinite-dimensional (degenerate) diffusion, and m_t itself is a stochastic object as it depends on the diffusion Z_t. Because m_t is stochastic, it is now necessary to include it as a state variable in the agents' value function

$$U(x, z, m, t).$$

Note that $m \in \mathcal{P}(\mathbb{R}^n)$, the space of probability measures with support in $\mathbb{R}^n$, which is an infinite-dimensional space. Hence, the Master equation is a HJB equation for the value function U set in infinite-dimensional space:

$$\begin{aligned} \rho U - \partial_t U &= H(x, z, \nabla_x U, m) + \nu\Delta_x U + \beta\Delta_z U \\ &\quad + \int_{\mathbb{R}^n} [\nabla_m U](y) \underbrace{[-\mathrm{div}_y(\nabla_\lambda H(y, z, \nabla_y U, m)\ m) + \nu\Delta_y m]\,(y)}_{\text{drift of probability measure } m \text{ at point } y \text{ from (2.13)}} dm(y) \\ &\qquad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathcal{P}(\mathbb{R}^n) \times (0, T), \\ U(x, z, m, T) &= V(x, z, m) \quad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathcal{P}(\mathbb{R}^n). \end{aligned} \tag{2.15}$$

Here $\nabla_m U$ denotes the derivative of U with respect to the measure m — see [13] for a precise definition — and $[\nabla_m U](y)$ denotes the derivative of U with respect to m *at point* y. Also note that we set the drift of the aggregate state $\mu_z(z) \equiv 0$ for notational simplicity and we will continue to do so going forward.

The stochastic Fokker–Planck equation (2.9) that comes from the standard common noise MFG formulation (2.8) further complicates the Master equation (2.15) with additional second-order derivatives in m [13]. The formulation in this section results in a simpler first-order Master equation and nests all typical macroeconomics applications.

2.3. *Rational expectations*

Rational expectations are a modeling assumption introduced by Muth [40] and popularized in the 1970s by Bob Lucas, Ed Prescott, Tom Sargent and others. It has since been the standard assumption for modeling expectations in macroeconomics. See [39] for key references and a brief historical discussion. The macroeconomics definition of rational expectations is as follows: *Agents have rational expectations if they form expectations over outcomes using the correct objective probability distributions of those outcomes. Hence, subjective probability distributions equal objective probability distributions.*

The rational expectations assumption is best thought of as a consistency requirement between expectations and model reality. Arguably a better name for the assumption is "model-consistent expectations" [44]. Related, it is important to emphasize that "rational expectations" are distinct from the concept of "rationality" which, in economics terminology, simply means that agents maximize some objective function. While the MFGs we consider below relax rational expectations, all of them retain rationality.

In the context of MFGs, the rational expectations assumption is about the expectation operator $\mathbb{E}$ that appears in the optimization procedure for the objective functions (2.2) and (2.11). If this expectation operator uses objective, model-consistent probability distributions for the behavior of all other agents as well as the idiosyncratic and common noise, then the MFG assumes rational expectations.

One key takeaway is that, as we explain in more detail in the next sections, *all* existing MFGs models in the mathematics literature (that we are aware of) implicitly assume rational expectations.

3. Mean Field Games with a Low-Dimensional Coupling

In Sec. 2, the running reward function $R(x, z, \alpha, m)$ that appears in the optimization problem (2.11) depends on the empirical measure $m(t)$ in a general, unrestricted fashion. However, in many applications, in particular in macroeconomics, this dependence is simpler: the running reward function depends on the empirical measure $m(t)$ only through a *low-dimensional* vector $p_t \in \mathbb{R}^\ell$, with some fixed $\ell \ll N$, that is a functional of m_t.

That is, the running reward and the terminal condition in (2.11) are given by $\widetilde{R}(x, z, \alpha, p)$ and $\widetilde{V}(x, z, p)$, so that agents optimize

$$u_N(x,t) = \max_{\alpha_i \in A} \mathbb{E}\left[\int_t^T e^{-\rho t}\widetilde{R}(X_{i,s}, Z_s, \alpha_{i,s}, p_s)ds + e^{-\rho(T-t)}\widetilde{V}(X_{i,T}, Z_T, p_T)\right] \tag{3.1}$$

subject to $X_{i,s}$ solving (2.1), where

$$p_t = P^*(m_t, Z_t), \tag{3.2}$$

for a fixed functional

$$P^* : \mathcal{P}(\mathbb{R}^n) \times \mathbb{R}^k \to \mathbb{R}^\ell. \tag{3.3}$$

We refer to such MFGs as *MFGs with a low-dimensional coupling.* Note that in MFGs with a low-dimensional coupling, model agents do not directly "care about" the infinite-dimensional density m_t in the sense that it does not enter their running reward functions or terminal conditions. Instead, they only "care about" the much lower-dimensional vector p_t.

The next two subsections present applications from macroeconomics that take this form. Other examples are dynamic games with a large number of players in which the running reward depends only on particular moments of the distribution, say, the first moments

$$\overline{X}_j(m) = \int x_j dm(x,t), \quad j = 1, \dots, n.$$

Of course, an MFG with a low-dimensional coupling is just a special case of the general MFGs discussed in Sec. 2 with the running reward and terminal value of the form

$$R(x,z,\alpha,m) = \widetilde{R}(x,z,\alpha,P^*(m,z)), \quad V(x,z,m) = \widetilde{V}(x,z,P^*(m,z)). \tag{3.4}$$

Therefore, under the rational expectations assumption, this low-dimensional coupling does not really simplify the analysis and the backward–forward MFG and Master equation still take the same form. However, as we will show below, low-dimensional coupling can drastically simplify the model's complexity when agents have non-rational expectations.

3.1. *Macroeconomics MFGs*: *Low-dimensional coupling through prices*

Typical macroeconomics MFGs, known as "heterogeneous agent models," are MFGs with a low-dimensional coupling. Usually, the running reward function $R(x, z, \alpha, m)$ depends on the empirical measure $m(t)$ only through a *low-dimensional* price vector $p_t \in \mathbb{R}^\ell$, with some fixed $\ell \ll N$. The prices may represent the actual prices of goods, or correspond to wages or interest rates, that is, prices of other variables like labor and capital. The underlying macroeconomic assumption is that the system stays in what is known as a *competitive equilibrium*. In that case, the prices are set by the intersection of demand and supply ("market clearing") and are determined by the empirical measure of the agents via a set of ℓ "market clearing" conditions (demand equals supply for each of ℓ goods):

$$M(p_t, m_t, Z_t) = 0, \tag{3.5}$$

with a given relation $M : \mathbb{R}^\ell \times \mathcal{P}(\mathbb{R}^n) \times \mathbb{R}^k \to \mathbb{R}^\ell$. Under the assumption that (3.5) can be inverted, this gives rise to a unique mapping which takes the form in (3.2) and which we will refer to as the *equilibrium price function*.

To summarize, the reward function depends on the measure m_t only through the (low-dimensional) price vector p_t that determines the optimal strategy in (2.2). In a competitive equilibrium, the price vector is directly related to m_t either explicitly by (3.2) or implicitly by (3.5).

3.2. *Example of a macroeconomics MFG*

A typical example of a macroeconomics MFG that has this structure is the model described in [1, Sec. 5] which is a continuous-time version of the Krusell and Smith [31] model. The corresponding model without common noise is described in [1, Sec. 2; 2] and which is a continuous-time version of the Aiyagari–Bewley–Huggett model [5, 9, 27].

In this model, the state of the agents is characterized by their income and wealth, so that the agents' positions are parametrized by $x = (x_1, x_2) \in \mathbb{R}^2$. Here, x_1 is wealth and x_2 is income, and $n = 2$ in the setting of Sec. 2. Furthermore, there are $\ell = 2$ prices $p_t = (p_{1,t}, p_{2,t}) \in \mathbb{R}^2$ where $p_{1,t}$ is the interest rate and $p_{2,t}$ the wage. There is $k = 1$ aggregate state $Z_t \in \mathbb{R}$ which has the interpretation of the (logarithm of) productivity of a so-called "representative firm" (see below).

Wealth and income evolve according to a system of SDEs

$$\begin{aligned} dX_{1,i,t} &= (P_1^*(m_t, Z_t)X_{1,i,t} + P_2^*(m_t, Z_t)X_{2,i,t} - C_{i,t})dt, \\ dX_{2,i,t} &= \mu(X_{2,i,t})dt + (2\nu)^{1/2}dB_{i,t}. \end{aligned} \tag{3.6}$$

Here, $C_{i,t}$ is the ith agent's consumption that serves as a control in this setting and μ is a drift coefficient. The price functionals P_1^* and P_2^* are the (scalar) interest rate and wage which depend on the measure m via an equilibrium condition explained below. Agents choose their consumption to maximize a utility function

$$\mathbb{E}\int_0^T e^{-\rho t}U(C_{i,t})dt \quad \text{where } U(c) = \frac{c^{1-\gamma}}{1-\gamma}, \quad \gamma > 0 \tag{3.7}$$

subject to (3.6) and a state constraint $X_{1,i,t} \geq 0$.

The Hamiltonian (2.5) is then

$$H(x, z, \lambda, m) = \max_c\{u(c) + \lambda_1(P_1^*(m, z)x_1 + P_2^*(m, z)x_2 - c) + \lambda_2\mu(x_2)\}.$$

The Hamiltonian H is nonlinear and non-separable between x, z, λ and m. At the same time, it depends on m only through the two-dimensional prices $P^*(m, z) : \mathcal{P}(\mathbb{R}^2) \times \mathbb{R} \to \mathbb{R}^2$.

The price functionals (equilibrium wage and interest rate) are given in this model by

$$P_1^*(m, z) = \partial_{\overline{X}_1} F(\overline{X}_1(m), \overline{X}_2(m), z), \quad P_2^*(m, z) = \partial_{\overline{X}_2} F(\overline{X}_1(m), \overline{X}_2(m), z),$$

where

$$F(\overline{X}_1, \overline{X}_2, z) = e^z \sqrt{\overline{X}_1 \overline{X}_2}, \quad \overline{X}_1(m) = \int x_1 dm(x, t), \quad \overline{X}_2(m) = \int x_2 dm(x, t). \tag{3.8}$$

The function F has the interpretation of the production function of a so-called "representative" firm, z that of (the logarithm of) the firm's productivity, and the derivatives $\partial_{\overline{X}_1} F$ and $\partial_{\overline{X}_2} F$ those of "marginal products". The dependence of the prices merely on the first moments of m, that is, on $\overline{X}_1(m)$ and $\overline{X}_2(m)$, is special to this particular application. Other macroeconomics applications feature (considerably) more complicated price functionals.

3.3. *The case without common noise*

For future reference it will also be useful to spell out MFGs with a low-dimensional coupling but without common noise. This is simply the case in which neither the running reward R nor the low-dimensional functional p depends on the aggregate state z, i.e. this functional is simply given by $p_t = P^*(m_t)$ with $P^* : \mathcal{P}(\mathbb{R}^n) \to \mathbb{R}^\ell$. Equivalently, the case without common noise sets $Z_t = 0$ for all t. In the macroeconomics MFG of Secs. 3.1 and 3.2, the underlying assumption is that neither the running reward R nor the market clearing condition M depend on the aggregate state Z. The particular example in Sec. 3.2 is the Aiyagari–Bewley–Huggett model analyzed in [1, 2].

3.4. *Backward–forward system and Master equation for MFGs with low-dimensional coupling*

The corresponding backward–forward MFG system in the case of a low-dimensional coupling takes exactly the same form (2.4) but with $R(x, \alpha, m) = \widetilde{R}(x, \alpha, P^*(m))$ in the Hamiltonian (2.5) and terminal value $V(x, m) = \widetilde{V}(x, P^*(m))$. The same is true for the Master equation for the value function $U(x, z, m, t)$ which takes the form (2.15) but with $R(x, z, \alpha, m) = \widetilde{R}(x, z, \alpha, P^*(m, z))$ in the Hamiltonian and terminal value $V(x, z, m) = \widetilde{V}(x, z, P^*(m, z))$.

Note that the special structure of MFGs with a low-dimensional coupling neither simplifies the backward–forward MFG system nor the Master equation in any straightforward way. In particular the infinite-dimensional measure $m \in \mathcal{P}(\mathbb{R}^n)$ is still a state variable in the Master equation. What is noteworthy about this is that the Master equation is an infinite-dimensional PDE despite model agents only "caring about" much lower-dimensional "prices". As we explain in the following, the root cause of this feature is the rational expectations assumption.

4. Non-Rational Expectations in Simple Control Problems

Before specifying what rational expectations — and departures from such expectations — mean in the context of large systems of heterogeneous agents (i.e. MFGs) let us first consider the simpler case of a single agent solving a stochastic control problem. We turn to MFGs in Sec. 5.

4.1. *A simple stochastic control problem in an evolving environment*

Consider a single agent solving a stochastic control problem in a prescribed time-dependent environment:

$$u(x,t) = \max_{\alpha \in A} \mathbb{E}\left[\int_t^T e^{-\rho(\tau-t)} R(X_\tau, \alpha_\tau, \beta_\tau) d\tau + e^{-\rho(T-t)} V(X_T)\right] \tag{4.1}$$

subject to X_τ solving an SDE

$$dX_\tau = \alpha_\tau d\tau + \sqrt{2\nu} dB_\tau, \quad X_t = x, \;\; t \leq \tau \leq T. \tag{4.2}$$

Here, α_τ is the control used on the time interval $t \leq \tau \leq T$, and β_τ represents a known time-dependent environment that the agent cannot control.

For future purposes, it will be useful to write the HJB equation for the value function u in terms of the infinitesimal generator which summarizes the transition probabilities of the process for X_t:

$$\mathcal{A}_\pi := \pi \cdot \nabla + \nu \Delta, \tag{4.3}$$

where $\pi(x,t)$ is the agent's policy. The HJB equation is then

$$\rho u - \partial_t u = \max_{\alpha \in A} \{R(x, \alpha, \beta_t) + \mathcal{A}_\alpha u\} \quad \text{in } \mathbb{R}^n \times (0,T), \tag{4.4}$$

with the terminal condition

$$u(x,T) = V(x) \quad \text{in } \mathbb{R}^n, \tag{4.5}$$

The associated optimal policy function is

$$\pi(x,t) \equiv \arg\max_{\alpha \in A} \{R(x, \alpha, \beta_t) + \alpha \cdot \nabla u(x,t)\}. \tag{4.6}$$

To explain what rational expectations mean in this setting, it is useful to spell out what this optimal control problem and the associated HJB equation look like without imposing the rational expectations assumption. In this model, expectations may be non-rational in two respects: first, the agent may have incorrect beliefs about its own dynamics for a given control α_t. Second, she may have incorrect beliefs about the environment β_t in the future. We consider these two cases separately below.

4.2. *Expectations about the evolution of the agent's state*

Let us fix a terminal time T and for the moment fix some time $t \in (0, T)$. For simplicity of notation, we will consider in this subsection a time-independent environment, so that

$$\beta_\tau \equiv \text{const} \quad \text{all } \tau. \tag{4.7}$$

As just discussed, the agent has rational expectations about the process for X_τ for $\tau > t$ if she forms expectations about X_τ using the correct objective transition probabilities. These transition probabilities are summarized via the infinitesimal generator. Rational expectations mean that the agent's beliefs are summarized by the correct generator $\mathcal{A}_\pi$ defined in (4.3) that determines the evolution of the actual state X_t in accordance with (4.2).

Non-rational expectations mean that the agent has some other subjective beliefs about the future evolution of X_τ for $\tau > t$ summarized by a different generator $\widehat{\mathcal{A}}_\pi$. For example, the agent may believe that the state follows an alternative diffusion process

$$d\widehat{X}_\tau \;=\; \widehat{\mu}(\widehat{X}_\tau, \alpha_\tau, \tau)d\tau + \sqrt{2\widehat{\nu}(\widehat{X}_\tau, \alpha_\tau, \tau)}dB_\tau, \quad \tau \geq t, \tag{4.8}$$

$$\widehat{X}_t = x \tag{4.9}$$

instead of the process (4.2) in which case the generator is

$$\widehat{\mathcal{A}}_\pi := \widehat{\mu}(x, \pi, t) \cdot \nabla + \widehat{\nu}(x, \pi, t)\Delta. \tag{4.10}$$

In particular, it may be the case that either $\widehat{\mu}(x, \alpha, \tau) \neq \alpha$ or $\widehat{\nu}(x, \alpha, \tau) \neq \nu$. This happens, for example, if the agent does not have the full information about the idiosyncratic noise, or presumes the existence of an additional drift in the problem.

Agents' policies $\pi(x, t)$ are now determined from the optimization problem

$$\widehat{u}(x, t) = \max_{\alpha \in A} \widehat{\mathbb{E}}\left[\int_t^T e^{-\rho(\tau - t)} R(\widehat{X}_\tau, \alpha_\tau) d\tau + e^{-\rho(T-t)} V(\widehat{X}_T)\right], \tag{4.11}$$

supplemented by (4.8)–(4.9) for the evolution of $\widehat{X}_\tau$ on the same time interval $t \leq \tau \leq T$. We denote the expectations operator by $\widehat{\mathbb{E}}$ to highlight that these subjective expectations are, in general, different from the objective (rational) expectations operator $\mathbb{E}$. That is, $\mathbb{E}$ refers to the expectation with respect to the trajectories generated by the evolution (4.2), and $\widehat{\mathbb{E}}$ to those generated by the perceived evolution (4.8).

The HJB equation with non-rational expectations is therefore

$$\rho\widehat{u} - \partial_s \widehat{u} = \max_{\alpha \in A}\{R(x, \alpha) + \widehat{\mathcal{A}}_\alpha \widehat{u}\} \quad \text{in } \mathbb{R}^n \times (t, T), \tag{4.12}$$

$$\widehat{u}(x, T) = V(x) \quad \text{in } \mathbb{R}^n. \tag{4.13}$$

The associated policy with non-rational expectations is

$$\pi(x,t) = \arg\max_{\alpha\in A}\{R(x,\alpha) + \alpha\cdot\nabla_x \widehat{u}(x,t)\}, \tag{4.14}$$

rather than by (4.6). Note the difference between the two value functions that appear in (4.6) and (4.14).

Let us comment that the actual probability density $\rho(x,t)$ of the agent who follows the strategy $\pi(x,t)$ defined by (4.14) is

$$\partial_t \rho = \mathcal{A}_\pi^* \rho. \tag{4.15}$$

Note the difference between the operator $\widehat{\mathcal{A}}_\pi$ that appears in the HJB equation (4.12) and the operator $\mathcal{A}_\pi$ whose adjoint appears in (4.15).

The special case of rational expectations of the evolution of X_t. Rational expectations mean that the generator $\widehat{\mathcal{A}}_\pi$ that appears in the HJB equation (4.12) coincides with the actual generator $\mathcal{A}_\pi$. In other words, the perceived evolution (4.8) coincides with the actual evolution (4.2) for any given control $\alpha_{\tau,t}$. With this assumption, the HJB equation (4.12) becomes

$$\begin{aligned} &\rho u - \partial_t u = \max_{\alpha\in A}\{R(x,\alpha) + \mathcal{A}_\alpha u\} && \text{in } \mathbb{R}^n\times(0,T), \\ &u(x,T) = V(x) && \text{in } \mathbb{R}^n, \end{aligned} \tag{4.16}$$

which is the same as (4.4) for the case of constant β_t. Hence, for the special case of rational expectations, we have recovered the standard HJB equation.

4.3. *Expectations about the evolution of the environment*

Let us now consider the same optimization problem but reintroduce the time-dependent environment β_t. At a time $t\in(0,T)$ the agent has access to the past trajectory

$$\beta_{\le t} = \{\beta(t') : \ 0\le t'\le t\}. \tag{4.17}$$

She uses this information to make a prediction

$$\widehat{\beta}_{s,t} = \Theta(s,t;\beta_{\le t}), \quad s>t, \tag{4.18}$$

with some given function Θ that depends on the running time s, the starting time t and the path $\beta_{\le t}$ that was observed prior to the time t. Non-rational expectations in this context mean that

$$\widehat{\beta}_{s,t} \ne \beta_s, \quad \text{for some } s>t, \tag{4.19}$$

i.e. that the agent's perceived trajectory of her external environment does not coincide with the environment's actual trajectory.

Let us assume for simplicity of notation that, while the environment may be predicted incorrectly, as in (4.19), the perceived law of motion of the state X_t is correct and the agent assumes that her trajectory is given by (4.2):

$$dX_{\tau,t} = \alpha_{\tau,t}d\tau + \sqrt{2\nu}dB_\tau, \quad X_{t,t}=x, \ \ t\le\tau\le T. \tag{4.20}$$

In other words, we have both $\widehat{\mu} = \alpha$ and $\widehat{\nu} = \nu$ in (4.8). Then, the agent is solving the optimization problem

$$u(x,t) = \max_{\alpha \in A} \mathbb{E}\left[\int_t^T e^{-\rho(\tau-t)} R(X_{\tau,t}, \alpha_{\tau,t}, \widehat{\beta}_{\tau,t}) d\tau + e^{-\rho(T-t)} V(X_{T,t})\right] \tag{4.21}$$

subject to (4.20) and (4.18).[e] As the environment $\widehat{\beta}_{\tau,t}$ now depends both on the running time τ and on the starting time t, in order to formulate the corresponding HJB equation, it is convenient to fix $t \in (0,T)$ and introduce an auxiliary optimization problem

$$\widehat{u}(x,s;t) = \max_{\alpha \in A} \mathbb{E}\left[\int_s^T e^{-\rho(\tau-s)} R(X_{\tau,s}, \alpha_{\tau,s}, \widehat{\beta}_{\tau,t}) d\tau + e^{-\rho(T-s)} V(X_{T,t})\right], \quad t < s < T, \tag{4.22}$$

subject to (4.20) and (4.18). Note that the prediction of the future environment $\widehat{\beta}_{\tau,t}$ is made at the time t and does not depend on the intermediate times s. Then, the perceived future optimal policy of the agent is, similarly to (4.6), given by

$$\widehat{\pi}(x,s;t) = \arg\max_{\alpha \in A}\{R(x,\alpha,\widehat{\beta}_{s,t}) + \alpha \cdot \nabla_x \widehat{u}(x,s;t)\}, \quad t \le s \le T. \tag{4.23}$$

In particular, at the time $s = t$ we have

$$\begin{aligned}\pi(x,t) &= \arg\max_{\alpha \in A}\{R(x,\alpha,\widehat{\beta}_{t,t}) + \alpha \cdot \nabla_x \widehat{u}(x,t;t)\}\\ &= \arg\max_{\alpha \in A}\{R(x,\alpha,\beta_t) + \alpha \cdot \nabla_x u(x,t)\},\end{aligned} \tag{4.24}$$

with

$$u(x,t) = \widehat{u}(x,t;t), \tag{4.25}$$

being the optimal value defined by (4.21). This is the actual policy that the agents are following. As a result of this optimization problem, an agent's state evolves as a diffusion

$$dX_t = \pi(X_t,t)dt + \sqrt{2\nu} dB_t, \tag{4.26}$$

with the policy $\pi(x,t)$ given by (4.24).

To find the function $u(x,t)$ one needs to solve a backward-in-time HJB equation for the value function $\widehat{u}(x,s;t)$. It takes the form

$$\begin{aligned}\rho\widehat{u}(x,s;t) - \partial_s \widehat{u}(x,s;t) &= \max_{\alpha \in A}\{R(x,\alpha,\widehat{\beta}_{s,t}) + \mathcal{A}_\alpha \widehat{u}(x,s;t)\} \quad \text{in } \mathbb{R}^n \times (t,T),\\ \widehat{u}(x,T;t) &= V(x) \quad \text{in } \mathbb{R}^n.\end{aligned} \tag{4.27}$$

[e] The value $u(x,t)$ also depends on the entire prediction $\widehat{\beta}_{\tau,t}, \tau \ge t$ but we suppress this dependence for simplicity.

Once the function $\widehat{u}(x, s; t)$ is found, one defines $u(x, t)$ by (4.25), which, in turn, gives the policy (4.24) that appears in the actual dynamics (4.26).

The special case of rational expectations for the evolution of the environment. Rational expectations on the evolution of the environment mean that

$$\widehat{\beta}_{s,t} \equiv \beta_s, \quad \text{for all } s > t. \tag{4.28}$$

Note that, for deterministic variables like the one considered here, rational expectations simply mean that agents have *perfect foresight* about the evolution of these variables. If that is the case, then the function $\widehat{u}(x, s; t)$ that solves (4.27) does not depend on the starting time t and neither does the policy $\pi(x, s; t)$ that also appears in (4.27). Under the assumption (4.28), the HJB equation (4.27) therefore becomes the standard HJB equation (4.4) with

$$\widehat{u}(x, s; t) = u(x, s). \tag{4.29}$$

5. Mean Field Games Without Rational Expectations

As already noted, *all* existing MFGs models in the mathematics literature we are aware of implicitly assume rational expectations. We now explain this in more detail and show how to formulate MFGs without rational expectations. We first consider the case without common noise, that is, the backward–forward MFG system (2.4), and then cover the Master equation (2.15).

5.1. *Non-rational expectations in the backward–forward MFG system*

We now show how to formulate a generalization of the backward–forward MFG system of Sec. 2.1 without making the rational expectations assumption. We focus on the more interesting case of non-rational expectations about the evolution of the agent's external environment, in this case the evolution of the density m. The case of non-rational expectations about agents' own states $X_{i,t}$ is analogous to the treatment in Sec. 4.2. As in the preceding section, after spelling out the model with general (not necessarily rational) expectations, we show that we recover the standard backward–forward MFG system in the special case with rational expectations.

Let us fix a terminal time T and for the moment fix some time $t \in (0, T)$. We assume that to predict the future empirical density of the other agents, an individual agent uses a perceived law of motion

$$\partial_s \widehat{m}(x, s; t) = \mathcal{B}^* \widehat{m}(x, s; t), \quad s \geq t, \tag{5.1}$$

$$\widehat{m}(x, t; t) = m(x, t). \tag{5.2}$$

Here, $\widehat{m}(x, s; t)$ is a prediction for the empirical measure of other agents for $s \geq t$, and $\mathcal{B}$ is the generator of a Markov process that an agent believes the other agents are following. The initial condition in (5.2) at $s = t$ comes from the actual observed density $m(x,t)$ at the time t. Note that the agents are constantly updating their prediction $\widehat{m}(x,s;t)$ for a given future time s, by changing the initial condition in (5.2) as t grows. In other words, $\widehat{m}(x,s;t)$ and $\widehat{m}(x,s;t')$ are, in general, different if $s > t > t'$. Going forward, we sometimes write $\widehat{m}_{s,t}$ for conciseness.

Note that this setup is exactly analogous to the case of non-rational expectations about the evolution of the environment in Sec. 4.3. Therefore, so is the remainder of the discussion. Similarly to (4.22), agents' policies $\pi(x,t)$ are determined from the perceived optimization problem

$$\widehat{u}(x,s;t) = \max_{\alpha_i \in A} \mathbb{E}\left[\int_s^T e^{-\rho(\tau-s)} R(X_{i,\tau}, \alpha_{\tau,s}, \widehat{m}_{\tau,t}) d\tau + e^{-\rho(T-s)} V(X_{i,T,s}, \widehat{m}_{T,t})\right], \quad t \le s \le T, \tag{5.3}$$

subject to (2.1) for the evolution of $X_{i,t}$ and (5.1)–(5.2) for the evolution of $\widehat{m}_{s,t}$ on the time interval $t \le s \le T$. As a result, the perceived future optimal policy of the agent is, similarly to (4.23), given by

$$\widehat{\pi}(x,s;t) = \arg\max_{\alpha \in A}\{R(x,\alpha,\widehat{m}_{s,t}) + \alpha \cdot \nabla_x \widehat{u}(x,s;t)\}, \quad t \le s \le T. \tag{5.4}$$

At the time $s = t$ we have

$$\pi(x,t) = \widehat{\pi}(x,t;t) = \arg\max_{\alpha \in A}\{R(x,\alpha,m_t) + \alpha \cdot \nabla_x \widetilde{u}(x,t)\}. \tag{5.5}$$

with $\widetilde{u}(x,t) = \widehat{u}(x,t;t)$. This is the actual policy that the agents are following.

As a result of this optimization problem, an idiosyncratic state evolves as a diffusion

$$dX_{i,t} = \pi(X_{i,t},t)dt + \sqrt{2\nu}dB_{i,t}. \tag{5.6}$$

To summarize, agents believe that for all $s > t$ the distribution of the other agents will evolve according to (5.1)–(5.2), and this is what shows up in the continuation values for $s \ge t$ in the definition (5.3) of the perceived value function $\widehat{u}(x,s;t)$. However, when they choose their actual policy $\pi(x,t)$ at time t, given by (5.5), they use the actual realized density m_t (which will generally differ from any previous estimates of $\widehat{m}_{t,t'}$ with $t' < t$) and the perceived value function $\widetilde{u}(x,t)$.

Agents' actual policies $\pi(x,t)$ defined in (5.5) give rise to the generator

$$\mathcal{A}_\pi := \pi \cdot \nabla + \nu\Delta \tag{5.7}$$

that determines the evolution of the actual density $m(x,t)$ in accordance with (5.6):

$$\partial_t m = \mathcal{A}_\pi^* m, \tag{5.8}$$

which is different from the evolution (5.1)–(5.2) for the perceived future density unless $\mathcal{B} = \mathcal{A}_\pi$.

Now, the backward–forward system of equations for the value function $\widehat{u}(s,x;t)$, the perceived density $\widehat{m}(x,s;t)$ and the actual density $m(x,t)$ becomes

$$\rho\widehat{u}(x,s;t) - \partial_s \widehat{u}(x,s;t) = \max_{\alpha \in A}\{R(x,\alpha,\widehat{m}_{s,t}) + \mathcal{A}_\alpha \widehat{u}(x,s;t)\} \quad \text{in } \mathbb{R}^n \times (t,T),$$

$$\begin{aligned}
\partial_s \widehat{m}(x,s;t) &= \mathcal{B}^* \widehat{m}(x,s;t) \quad \text{in } \mathbb{R}^n \times (t,T),\\
\pi(x,t) &= \arg\max_{\alpha\in A}\{R(x,\alpha,m(t)) + \alpha\cdot\nabla_x \widehat{u}(x,t;t)\},\\
\partial_t m &= \mathcal{A}_\pi^* m \quad \text{in } \mathbb{R}^n \times (0,T),\\
m(x,0) &= m_0(x), \quad \widehat{m}(x,t;t) = m(x,t),\\
&\qquad\qquad \widehat{u}(x,T;t) = V(x,\widehat{m}_{T,t}) \quad \text{in } \mathbb{R}^n.
\end{aligned} \tag{5.9}$$

The system (5.9) is the analogue of what economists call a "temporary equilibrium".

Definition. *Temporary equilibrium* at a particular time t is defined as allocations and policies such that (i) agents optimize *given expectations of future variables* (including the density) that are specified in the model but that are *not necessarily rational*, (ii) the economy is in Nash equilibrium at time t.

This idea was originally developed contemporaneously by Hicks [26] and Lindahl [35], and has been further developed by Grandmont [23, 24].

Remark (Connection to the Master equation). Note that the system (5.9) could also be written in terms of a Master equation, with the measure m as a state variable. However, this would defeat the purpose of this approach: the advantage of (5.9) is exactly that we do not have to compute the dynamics in the infinite-dimensional space of probability measures but rather only find the solutions to (5.9) for the measures m_t that one encounters in the course of actual evolution.

Rational expectations in the Backward–Forward MFG System. Rational expectations in the context of the model we have discussed above mean that the generator $\mathcal{B}$ that appears in Eqs. (5.1) and (5.9) for the perceived density $\widehat{m}(x,s;t)$ coincides with the actual generator $\mathcal{A}_\pi$ that appears in the evolution equation (5.7) for the actual density $m(x,t)$:

$$\mathcal{B} = \mathcal{A}_\pi. \tag{5.10}$$

Let us now show that with this assumption, the system (5.9) reduces to the familiar MFG system (2.4) that we write in the form

$$\begin{aligned}
\rho u - \partial_t u &= \max_{\alpha\in A}\{R(x,\alpha,m(t)) + \mathcal{A}_\alpha u\} \quad \text{in } \mathbb{R}^n \times (0,T),\\
\pi(x,t) &= \arg\max_{\alpha\in A}\{R(x,\alpha,m(t)) + \alpha\cdot\nabla u(x,t)\},\\
\partial_t m &= \mathcal{A}_\pi^* m \quad \text{in } \mathbb{R}^n \times (0,T),\\
m(0) &= m_0, \quad u(x,T) = V(x,m(T)) \quad \text{in } \mathbb{R}^n.
\end{aligned} \tag{5.11}$$

Indeed, if (5.10) holds, we deduce from the second and fourth equations in (5.9), together with the initial condition $\widehat{m}(x,t;t) = m(x,t)$ that

$$\widehat{m}(x,s;t) = m(x,s) \quad \text{for all } t \in [0,T] \text{ and } s \in [t,T], \tag{5.12}$$

so that the perceived future density $\widehat{m}(x,s;t)$ does not depend on the starting time t and coincides with the actual density $m(x,s)$. The perceived value function $\widehat{u}(x,s;t)$ is then independent of t as well, so that we have

$$\widehat{u}(x,s;t) = \widehat{u}(x,s;s) = \widetilde{u}(x,s). \tag{5.13}$$

The perceived policies $\widehat{\pi}(x,s;t)$ are also independent of t because we can use (5.12) and (5.13) to write

$$\begin{aligned} \widehat{\pi}(x,s;t) &= \arg\max_{\alpha \in A}\{R(x,\alpha,\widehat{m}_{s,t}) + \alpha \cdot \nabla_x \widehat{u}(x,s;t)\} \\ &= \arg\max_{\alpha \in A}\{R(x,\alpha,m(s)) + \alpha \cdot \nabla_x \widehat{u}(x,s;s)\} = \pi(x,s). \end{aligned} \tag{5.14}$$

It follows from the above and the first equation in (5.9) that the value function $\widetilde{u}(s,x)$ satisfies the backward HJB equation

$$\rho\widetilde{u}(x,s) - \partial_s \widetilde{u}(x,s) = R(x,\pi(x,s),m(s)) + \mathcal{A}_\pi \widetilde{u}(x,s) \quad \text{in } \mathbb{R}^n \times (s,T). \tag{5.15}$$

Together with the forward Fokker–Planck equation

$$\partial_t m = \mathcal{A}_\pi^* m \quad \text{in } \mathbb{R}^n \times (0,T), \tag{5.16}$$

we have recovered the MFG system (5.11) or, equivalently, (2.4).

This shows clearly that the backward–forward MFG system (2.4) implicitly assumes that agents have rational expectations about the process $X_{i,t}$ for all agents $i = 1,\ldots,N$, and hence for the evolution of the density $m(x,t)$.

Remark (Non-rational expectations about agents' own states). As already noted, one could also allow for non-rational expectations about agents' own states $X_{i,t}$. Analogous to the treatment in Sec. 4.2, this simply involves replacing the generator $\mathcal{A}_\alpha$ in the first equation in (5.9) by some other generator $\widehat{\mathcal{A}}_\alpha$ capturing each agent's beliefs about the evolution of her own state, i.e. a perceived law of motion like (4.8) and (4.9).

Remark (Heterogeneous beliefs). It is reasonable to expect that different agents may have different beliefs about the evolution of the density of the other agents. In fact, substantial belief heterogeneity is one of the most prevalent empirical findings in the macroeconomics literature on household and firm expectations, see the discussion and references in [39] (the finding is often summarized under the terminology "disagreement"). In our context, different individual agents may use different forms of the operator $\mathcal{B}$, that appears in (5.1) to predict the future evolution of the density of the other agents. This will, in turn, affect the actual generator $\mathcal{A}_\pi$ governing the evolution of the actual density $m(t)$ of the agents. We will revisit this issue below in the discussion of adaptive learning in Sec. 7.1.

Remark (The unrealism of rational expectations). Rational expectations impose that agents know the correct objective transition probabilities not only for their own individual states but also for the evolution of every other agent's states, i.e. for the entire complex system they inhabit as a whole. Specifically, the implicit assumption is that each agent *knows all other agents' optimal policies* $\pi(y,\cdot)$ for all $y \in \mathbb{R}^n$ and *uses these to (correctly) forecast the evolution of everyone else's state and hence the measure* m*!* For these reasons, the rational expectations assumption is arguably a stretch in complex environments like MFG models.

Remark (The evolution of the density with heterogeneous agents). It is also worth pointing out that the rational expectations assumption (5.10) is the reason why the operator $\mathcal{A}_\pi^*$ in the Fokker–Planck equation in (5.11) is the adjoint of the operator $\mathcal{A}_\pi$ in the HJB equation in (5.11). Without rational expectations this would not be the case: while the Fokker–Planck equation for the actual density $m(x,t)$ in (5.9) necessarily features (the adjoint of) the correct generator $\mathcal{A}_\pi$ because it reflects the actual realized dynamics of $X_{i,t}$ and the associated evolution of the measure m, the HJB equation for the perceived value function $\widehat{u}(s,x;t)$ in (5.9) features $\mathcal{A}_\pi$ only under the rational expectations assumption (5.10). Otherwise it is driven by the perceived strategy $\widehat{\pi}$.

The rational expectations assumption (5.10) may seem to superficially simplify a "complicated" system (5.9) to a "simpler" (and, definitely, shorter) system (5.11). However, while the latter may produce a higher value function for the agents, it comes at a high computational cost. If the operator $\mathcal{B}$ that appears in the evolution equation for the perceived density $\widehat{m}(x,s;t)$ in (5.9) is independent from the perceived value function $\widehat{u}(x,s;t)$ then (5.9) may be solved in a single pass for a fixed $t < T$. First, one solves the forward-in-time equation

$$\begin{aligned} \partial_s \widehat{m}(x,s;t) &= \mathcal{B}^* \widehat{m}(x,s;t) \quad \text{in } \mathbb{R}^n \times (t,T), \\ \widehat{m}(x,t;t) &= m(x,t), \end{aligned} \tag{5.17}$$

to find the perceived future density of the agents. This is followed by solving the backward-in-time HJB equation

$$\begin{aligned} \rho\widehat{u}(x,s;t) - \partial_s \widehat{u}(x,s;t) &= \max_{\alpha\in A}\{R(x,\alpha,\widehat{m}_{s,t}) + \mathcal{A}_\alpha \widehat{u}(x,s;t)\} \quad \text{in } \mathbb{R}^n \times (t,T), \\ \widehat{u}(x,T;t) &= V(x,\widehat{m}(T,t)), \end{aligned} \tag{5.18}$$

to find the perceived value function $\widehat{u}(x,s;t)$. This requires no iterations that are normally used to solve the forward–backward MFG problems. With $\widehat{u}(x,t;t)$ in hand, one can compute the actual policy $\pi(x,t)$ given by (5.5), and continue the evolution of the actual agents' density $m(x,t)$ forward in time. Thus, from the computational-cost point of view, dropping the rational expectations assumption may end up being beneficial.

5.2. *Non-rational expectations in Mean Field Games with common noise*

We now repeat the exercise for the case with common noise and show how to formulate MFGs without rational expectations in this case. To this end, recall that the evolution of the states $(X_{i,t}, Z_t)$ is given by (2.12) where $\pi(X_{i,t}, Z_t, m_t, t)$ is the optimal policy with common noise.

Analogous to (5.1) and (5.2), with non-rational expectations, the agent incorrectly believes that the density and aggregate state evolve according to

$$\begin{aligned} d\widehat{m}_{s,t} &= \mathcal{B}^*_{\widehat{Z}_s}\widehat{m}_{s,t}ds, \quad s \geq t, \\ d\widehat{Z}_{s,t} &= \sqrt{2\widehat{\beta}(\widehat{Z}_{s,t}, \tau)}dW_s, \quad s \geq t, \end{aligned} \tag{5.19}$$

with $\widehat{m}_{t,t} = m_t$ and $\widehat{Z}_{t,t} = Z_t$, instead of the correct evolution (2.13) and (2.14). As already noted, we set the drift $\mu_z(z) \equiv 0$ in (2.12) for simplicity. Note that, in general, the perceived generator $\mathcal{B}_{\widehat{Z}_s}$ may depend on the perceived aggregate state $\widehat{Z}_s$.

This leads to the following infinite-dimensional Master equation:

$$\begin{aligned} \rho\widehat{U} - \partial_t\widehat{U} &= \max_{\alpha\in A}\{R(x, z, \alpha, m) + \mathcal{A}_{\alpha,z}\widehat{U}\} + \widehat{\beta}(z, t)\Delta_z\widehat{U} \\ &\quad + \int_{\mathbb{R}^n}[\nabla_m\widehat{U}](y)[\mathcal{B}^*_z m](y)dm(y) \quad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathcal{P}(\mathbb{R}^n) \times (0, T), \\ \widehat{U}(x, z, m, T) &= V(x, z, m) \quad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathcal{P}(\mathbb{R}^n). \end{aligned} \tag{5.20}$$

As before, $\mathcal{A}_{\pi,z}$ in the term $\mathcal{A}_{\pi,z}\widehat{U}$ summarizes agents' beliefs about the evolution of the individual state $X_{i,t}$. In contrast, $[\mathcal{B}^*_z m](y)$ summarizes their beliefs about evolution of measure m at point y which may, in general, differ from the actual evolution.

After solving for the optimal policies $\pi(x, z, m, t)$, the evolution of the actual density of the agents can be found from the coupled system of SDE

$$dm_t = \mathcal{A}^*_{\pi,Z_t}m_t dt, \quad dZ_t = \sqrt{2\beta}dW_t, \tag{5.21}$$

with

$$\mathcal{A}_{\pi,z} = \pi(x, z, m, t)\nabla_x + \nu\Delta_x.$$

Rational expectations. Exactly as in Sec. 5.1, rational expectations mean that the generator $\mathcal{B}$ that appears in Eqs. (5.19) and (5.20) for the perceived density $\widehat{m}_{s,t}$ and the perceived value function $\widehat{U}$ coincides with the actual generator $\mathcal{A}_{\pi,z}$ that appears in the evolution equation (5.21) for the actual density $m(x, t)$:

$$\mathcal{B}_z = \mathcal{A}_{\pi,z}, \tag{5.22}$$

and, in addition, that the perceived diffusivity for the aggregate state Z_t equals the actual diffusivity, $\widehat{\beta}(z,t) = \beta$. In that case, (5.20) becomes

$$\begin{aligned}
\rho U - \partial_t U &= \max_{\alpha \in A}\{R(x,z,\alpha,m) + \mathcal{A}_\alpha U\} + \beta \Delta_z U \\
&\quad + \int_{\mathbb{R}^n} [\nabla_m U](y)[\mathcal{A}^*_{\pi,z} m](y)dm(y) \quad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathcal{P}(\mathbb{R}^n) \times (0,T), \\
U(x,z,m,T) &= V(x,z,m) \quad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathcal{P}(\mathbb{R}^n), \\
\mathcal{A}_{\pi,z} &= \pi \nabla_x + \nu \Delta_x, \\
\pi(x,z,m,t) &= \arg\max_{\alpha \in A}\{R(x,z,\alpha,m) + \mathcal{A}_{\alpha,z} U\}.
\end{aligned} \tag{5.23}$$

Importantly, analogous to the discussion in the preceding subsection, the implicit assumption underlying rational expectations is that each agent knows not only the correct stochastic process for their own individual state $X_{i,t}$ but also that of the state of all other agents, i.e. their optimal policies in the future.

Remark (Nash equilibrium, "common knowledge", and the computational complexity of the Master equation). MFGs impose Nash equilibrium, i.e. that each agent plays a best response to their prediction of every other agent's strategy. With rational expectations, each agent *knows* the other agents' strategies and then solves the infinite-dimensional Master equation to compute this best response. In fact, the Master equation not only imposes that all agents know each other's strategies, but they also know that they all know, and so on, *ad infinitum*. This assumption is called "common knowledge" in the economics literature. In that sense, the policies coming from the solution to the Master equation are optimal for a given individual agent only if the other agents also have rational expectations and if everyone knows that they do. Without that social compact in which everyone predicts an identical future, solving the infinite-dimensional Master equation is suboptimal and may even be harmful. While requiring such "common knowledge" is a common issue for Nash equilibria, the setting of the infinite-dimensional Master equation is somewhat special because computing the Nash equilibrium requires a huge computational cost that may not be accessible to all agents in the system.

6. Non-Rational Expectations in MFGs with Low-Dimensional Coupling

In this section, we consider MFGs that have the special structure in Sec. 3: agents' rewards depend on the measure m only through a low-dimensional functional.

6.1. *The case without common noise*

The case without common noise is analogous to Sec. 5.1 and we therefore cover it only briefly. As in Sec. 3, the running reward and terminal value are given by

$\widetilde{R}(X_{i,t}, \alpha_{i,s}, p_t)$ and $\widetilde{V}(X_{i,t}, p_t)$ for a low-dimensional vector $p_t \in \mathbb{R}^\ell$ with $p_t = P^*(m_t)$. Going forward we drop the tildes from $\widetilde{R}$ and $\widetilde{V}$ for notational simplicity.

With rational expectations, agents understand the dependence of p_t on m_t and therefore use the functional P^* together with the correct evolution for m_t to predict future values of p_t. That is, agents have perfect foresight over the future trajectory of p_t. With non-rational expectations, agents instead perceive some other trajectory

$$\widehat{p}_{s,t} = \Theta(s, t; p_{\leq t}), \quad s > t, \quad \text{with } p_{t,t} = p(t).$$

In particular note that agents' perceived trajectory of future "prices" $\widehat{p}_{s,t}$, $s > t$ generally depends on past realizations $p_{\leq t}$ which model agents have already observed by the time t.

Benchmark with rational expectations. We first spell out the backward–forward MFG system with rational expectations using the generator notation introduced in the preceding section. We will refer back to this benchmark system at various future points of the paper. The system is

$$\begin{aligned} \rho u - \partial_t u &= \max_{\alpha \in A}\{R(x, \alpha, p(t)) + \mathcal{A}_\alpha u\} && \text{in } \mathbb{R}^n \times (0, T), \\ \pi(x,t) &= \arg\max_{\alpha \in A}\{R(x, \alpha, p(t)) + \alpha \nabla u(x,t)\}, \\ \partial_t m &= \mathcal{A}^*_\pi m && \text{in } \mathbb{R}^n \times (0, T), \\ p(t) &= P^*(m(t)) && \text{in } (0, T), \\ m(0) &= m_0, \quad u(x,T) = V(x, P^*(m_T)) && \text{in } \mathbb{R}^n. \end{aligned} \tag{6.1}$$

Non-rational expectations. Analogously to Sec. 5.1, the backward–forward MFG system without rational expectations is

$$\begin{aligned} \rho \widehat{u}(x,s;t) - \partial_s \widehat{u}(x,s;t) &= \max_{\alpha \in A}\{R(x, \alpha, \widehat{p}_{s,t}) + \mathcal{A}_\alpha \widehat{u}(x,s;t)\} && \text{in } \mathbb{R}^n \times (t, T), \\ \widehat{p}_{s,t} &= \Theta(s, t; p_{\leq t}) && \text{in } (t, T), \\ \pi(x,t) &= \arg\max_{\alpha \in A}\{R(x, \alpha, p(t)) + \alpha \cdot \nabla_x \widehat{u}(x,t;t)\}, \\ \partial_t m &= \mathcal{A}^*_\pi m && \text{in } \mathbb{R}^n \times (0, T), \\ p(t) &= P^*(m(t)) && \text{in } (0, T), \\ m(0,x) &= m_0(x), \quad \widehat{u}(x,T;t) = V(x, \widehat{p}_{T,t}) && \text{in } \mathbb{R}^n. \end{aligned} \tag{6.2}$$

6.2. *Common noise: Sidestepping the Master equation in MFGs with a low-dimensional coupling*

In the presence of a common noise Z_t, as in Sec. 3, the running reward is $R(X_{i,t}, Z_t, \alpha_{i,s}, p_t)$ for a low-dimensional vector $p_t = P^*(m_t, Z_t) \in \mathbb{R}^\ell$. With rational expectations, agents understand the dependence of p_t on m_t and Z_t and therefore use the functional P^* together with the correct stochastic processes for m_t and Z_t to predict future values of p_t. This leads to the Master equation (6.5), with essentially no simplifications despite the low-dimensional coupling.

With non-rational expectations agents instead perceive some other stochastic process for the pair (p_t, Z_t). In the simplest case, they simply perceive p_t to evolve according to a completely exogenous stochastic process

$$d\widehat{p}_{s,t} = \widehat{\mu}_p(\widehat{p}_{s,t})ds + \widehat{\sigma}_p(\widehat{p}_{s,t})dW_t, \quad s \geq t, \quad \widehat{p}_{t,t} = p_t. \tag{6.3}$$

In more complicated cases, agents perceive a joint stochastic process for p_t, Z_t and other variables (which could, in principle, include m_t).

In the case of agents perceiving the simple process (6.3), instead of writing a Master equation, we can write a much simpler, standard finite-dimensional HJB equation for a value function $\widehat{U}(x, z, p, t)$. Denoting the generator corresponding to (6.3) by $\widehat{\mathcal{A}}_p$ we have

$$\begin{aligned} \rho\widehat{U} - \partial_t\widehat{U} &= \max_{\alpha \in A}\{R(x, z, \alpha, p) + \mathcal{A}_\alpha\widehat{U}\} \\ &\quad + \beta\Delta_z\widehat{U} + \widehat{\mathcal{A}}_p\widehat{U} \quad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^\ell \times (0, T), \\ U(x, z, p, T) &= V(x, z, p) \quad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^\ell, \\ \mathcal{A}_\alpha &= \alpha\nabla_x + \nu\Delta_x. \end{aligned} \tag{6.4}$$

Therefore, in MFGs with a low-dimensional coupling, departing from rational expectations can completely sidestep the Master equation. Of course, the case considered here is just an illustrative example. In particular, note that the perceived law of motion (6.3) is specified completely "outside the model" which leaves open the question of where this perceived law of motion "comes from" in the first place.

6.3. *The trouble with the Master equation in MFGs with a low-dimensional coupling*

The corresponding Master equation for the value function $U(x, z, m, t)$ is instead

$$\begin{aligned} \rho U - \partial_t U &= \max_{\alpha \in A}\{R(x, z, \alpha, P^*(m, z)) + \mathcal{A}_\alpha U\} + \beta\Delta_z U \\ &\quad + \int_{\mathbb{R}^n} [\nabla_m U](y)[\mathcal{A}^*_{\pi,z} m](y)dm(y) \quad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathcal{P}(\mathbb{R}^n) \times (0, T), \\ U(x, z, m, T) &= V(x, z, m) \quad \text{in } \mathbb{R}^n \times \mathbb{R}^k \times \mathcal{P}(\mathbb{R}^n), \\ \mathcal{A}_{\pi,z} &= \pi\nabla_x + \nu\Delta_x, \\ \pi(x, z, m, t) &= \arg\max_{\alpha \in A}\{R(x, z, \alpha, P^*(m, z)) + \alpha \cdot \nabla_x U(x, z, m, t)\}. \end{aligned} \tag{6.5}$$

This is the same Master equation as described in Sec. 3.4 but using the generator notation used in the present section. As already noted there, the special structure of MFGs with a low-dimensional coupling does not simplify the Master equation in any straightforward way. In particular the infinite-dimensional measure $m \in \mathcal{P}(\mathbb{R}^n)$ is still a state variable in agents' value function.

The reason this happens is the rational expectations assumption. Intuitively, because agents are forward-looking, they need to forecast future prices p_t, a low-dimensional object. But they understand that p_t depends on the infinite-dimensional measure m_t via the functional (3.2). Therefore, agents forecast the measure m_t in order to forecast prices p_t. Furthermore, as before, they forecast m_t using their knowledge of all other agents' policies $\pi(y)$. Note that all of this happens despite agents not even directly "caring about" the distribution m_t.

Related, note that actual equilibrium prices p_t do not follow a Markov process (if they did, one could write a finite-dimensional HJB equation with prices p as the state variables also in the rational expectations case); instead only (m_t, Z_t) has the Markov property and prices are instead a complicated nonlinear functional of this Markov state. Agents with rational expectations therefore (unrealistically) forecast the Markov state (m_t, Z_t) in order to forecast the non-Markovian prices p_t.

Considering the case of macroeconomics MFGs, Moll [39] argues that we should not make our lives so hard. It seems self-evident that real-world households and firms do not forecast prices by forecasting cross-sectional distributions and instead solve simpler problems. Instead of solving "Monster equations" we should replace the rational expectations assumption and solve the simpler equations corresponding to households' and firms' actual price-forecasting behavior.

7. A Way Forward: Learning in MFGs

As we have argued, MFGs with rational expectations and the Master equation are unrealistically complex as models of human decision-making. We have also seen that departing from rational expectations may hold the promise of sidestepping the infinite-dimensional Master equation altogether. However, in the MFGs with non-rational expectations we have considered thus far, agents held beliefs about future evolution or future prices that were specified outside the model (the "temporary equilibrium" idea). Therefore, these specified beliefs may end up being completely disconnected from the actual evolution of these equilibrium objects, i.e. there may be a disconnect between beliefs and "model reality" and agents' expectations may be systematically disappointed. A related issue is that a model with exogenously specified beliefs is subject to a version of the so-called "Lucas critique" [37]: when there is a change in economic policy (which would typically correspond to a change in a model parameter), one should expect agents' beliefs to change as well and this belief updating needs to be modeled.

How can we model non-rational expectations that are endogenous to the actual equilibrium prices but that, nevertheless, sidestep the Master equation and allow for computing standard finite-dimensional HJB equations for agents' value functions? Put differently: how can we formulate, in a systematic way, models of agents' behavior in situations with a low-dimensional coupling that lead to equations that (i) approximate agents' real-world behavior, and (ii) sidestep computing the solutions to a Master equation with the infinite-dimensional state $m \in \mathcal{P}(\mathbb{R}^n)$ and the associated curse of dimensionality? This is the challenge posed by Moll [39].

A natural answer is to add some form of *learning* to the model. That is, instead of imposing — as the rational expectations assumption does — that agents *know* the correct (and extremely complex) transition probabilities of equilibrium prices, we instead impose that agents *learn* about these transition probabilities over time. This approach has a long tradition in the economics literature, typically in the form of "least-squares learning" [10, 20, 38], and has recently been applied in the MFG literature to the case without common noise [6, 33, 34, 49], mostly in the form of RL.

Before proceeding to describing this way forward, let us also note that one promise of modeling learning in this way is to "kill two birds with one stone": to develop variants of MFGs with a low-dimensional coupling that are easier to compute and analyze, while, at the same time, making these models more realistic and more likely to generate interesting macroeconomic phenomena.

7.1. *Adaptive learning without common noise*

Adaptive learning is a special case of the non-rational beliefs about the future we have discussed in Secs. 5 and 6. More specifically, this is a special form of the future prices predictor $\Theta(s,t;p_{\le t})$ that appears in the system (6.2). One simple version of adaptive learning is least-squares learning [10, 20, 38]. Jacobson [29] implements such an approach in a heterogeneous-agent model.

In the simplest version of least-squares learning, at any time t, agents simply expect prices to remain constant at some value $\bar{p}$, i.e. $p_s = \bar{p}$ for all $s > t$. However, they update their estimate of this constant value over time as they collect new data on actual realized prices and we denote agents' time-t estimate of $\bar{p}$ by $\widehat{p}_t$. Specifically, agents compute $\widehat{p}_t$ as a simple backward-looking average:

$$\widehat{p}_t = \frac{1}{t}\int_0^t p_s ds.$$

Differentiating, we have

$$\dot{\widehat{p}}_t = \frac{1}{t}(p_t - \widehat{p}_t). \tag{7.1}$$

Other forms of learning besides least-squares learning are possible as well, such as, for example, an ordinary differential equation of the form

$$\dot{\widehat{p}}_t = L(p_t, \widehat{p}_t). \tag{7.2}$$

For example, the ODE

$$\dot{\widehat{p}}_t = \alpha(p_t - \widehat{p}_t),$$

with a constant $\alpha > 0$ rather than the factor of $1/t$, as in (7.1), is what is called "adaptive expectations" [11]. Note that expected prices will generally differ from actual prices, i.e. there is no longer perfect foresight.

In more complicated versions of least-squares learning, agents have a parametric "perceived law of motion" of prices

$$\dot{p}_t = \mu_p(p_t, \theta), \tag{7.3}$$

where $\theta \in \mathbb{R}^d$ is a parameter vector which parameterizes their beliefs about the evolution of p_t. For example the perceived law of motion function μ_p could be linear:

$$\dot{p}_t = \theta_0 + \theta_1 p_t.$$

Agents then update their estimate $\widehat{\theta}_t \in \mathbb{R}^d$ of the parameter vector θ recursively over time

$$\dot{\widehat{\theta}}_t = L(p_t, \widehat{\theta}_t), \quad \widehat{\theta}_0 \text{ given}, \tag{7.4}$$

which plays a similar role to the ODE (7.2). For example, if L is linear, $\widehat{\theta}_t$ could be a backward-looking least-squares estimator. Note that p_t that appear in (7.4) are the actual observed prices at time t.

Agents' policies $\pi(x, t)$ are determined from the following optimization problem that is analogous to (5.3). We fix $t \in [0, T]$ and for each $s \in [t, T]$ consider the perceived value function

$$\widehat{u}(x, s; t) = \max_{\alpha_i \in A} \mathbb{E}\left[\int_s^T e^{-\rho(\tau - s)} R(X_{i,\tau}, \alpha_{i,\tau}, \widehat{p}_{\tau;t}) d\tau + e^{-\rho(T-s)} V(X_{i,T}, \widehat{p}_{T,t})\right] \tag{7.5}$$

subject to (2.1) for the evolution of $X_{i,t}$ and where $\widehat{p}_{\tau;t}$ are the perceived future prices that evolve according to

$$\begin{aligned} \frac{d\widehat{p}_{\tau;t}}{d\tau} &= \mu_p(\widehat{p}_{\tau;t}, \widehat{\theta}_t), \quad t \le \tau \le T, \\ \widehat{p}_{t;t} &= p_t. \end{aligned} \tag{7.6}$$

Here, p_t are the actual observed prices at time t, and $\widehat{\theta}_t$ is the solution to (7.4). We emphasize that the parameter $\widehat{\theta}_t$ in (7.6) is fixed for $t \le \tau \le T$ and does not depend on τ. That is, from the point of view of an agent at time t, $\widehat{\theta}_t$ is just a fixed parameter vector that parameterizes the perceived law of motion μ_p of future prices $\widehat{p}_{\tau;t}$, with $\tau \ge t$. Note that the dependence of the perceived value function $\widehat{u}(x, s; t)$ on the time t is solely through the parameter θ_t that appears in the perceived evolution (7.6) of the future prices.

The optimization problem (7.5) gives rise to the perceived future policy

$$\widehat{\pi}(x, s; t) = \arg\max_{\alpha \in A} \{R(x, \alpha, \widehat{p}_{s;t}) + \alpha \cdot \nabla_x \widehat{u}(x, s; t)\}. \tag{7.7}$$

The actual policy that the agents are following is, on the other hand,

$$\pi(x,t) = \widehat{\pi}(x,t;t) = \arg\max_{\alpha \in A}\{R(x,\alpha,p_t) + \alpha \cdot \nabla_x \widehat{u}(x,t;t)\}. \tag{7.8}$$

As above, the interpretation is that agents believe that for all $s > t$ prices will evolve according to (7.6) and this is what shows up in their continuation values. However, when they choose their policy at time t, they see the actual realized prices p_t (which will generally differ from any previous estimates of p_t).

Going forward we drop the hat from $\widehat{\theta}_t$ for notational simplicity but keep in mind that this is really a time-varying estimate of the parameter θ in (7.3). With this notation in hand, the backward–forward MFG system with adaptive learning becomes the following version of (6.2):

$$\begin{aligned}
\rho\widehat{u}(x,s;t) - \partial_s \widehat{u}(x,s;t) &= \max_{\alpha\in A}\{R(x,\alpha,\widehat{p}_{s;t}) + \mathcal{A}_\alpha \widehat{u}(x,s;t)\} && \text{in } \mathbb{R}^n \times (t,T),\\
\frac{d\widehat{p}_{s;t}}{ds} &= \mu_p(\widehat{p}_{s;t},\theta_t) && \text{in } (t,T),\\
\widehat{p}_{t;t} &= p_t, &&\\
\pi(x,t) &= \arg\max_{\alpha\in A}\{R(x,\alpha,p_t) + \alpha\cdot\nabla_x\widehat{u}(x,t;t)\}, &&\\
\partial_t m &= \mathcal{A}_\pi^* m && \text{in } \mathbb{R}^n \times (0,T),\\
\dot{\theta}_t &= L(p_t,\theta_t), \quad \theta_0 \text{ given}, && \text{in } (0,T),\\
p_t &= P^*(m(t)) && \text{in } (0,T),\\
m(0,x) &= m_0(x), \quad \widehat{u}(x,T;t) = V(x,\widehat{p}_{T,t}) && \text{in } \mathbb{R}^n.
\end{aligned} \tag{7.9}$$

Here, in the context of adaptive learning, the low dimensionality of the coupling is crucial: instead of a highly complex infinite-dimensional analog of the Master equation, we get a finite-dimensional system (7.9). The forward-in-time equation for $m(t,x)$ in the system (7.9) is coupled to the backward-in-time equation for the perceived value function $\widehat{u}(x,s;t)$ solely via the actual price p_t that appears in the strategy $\pi(x,t)$ and in the parameter θ_t that appears in the HJB equation for $\widehat{u}(x,s;t)$. This coupling is not as severe as in MFGs with rational expectations since, as we have mentioned previously, the solution to (7.9) does not require any iterations. It does require, however, to solve, for each $t \in [0,T]$, a separate HJB equation on time interval $t \le s \le T$.

The Hamilton–Jacobi–Bellman equation in the price space. An alternative way of writing (7.9) is to introduce the perceived value function $\widehat{U}(x,p,s;\theta)$ defined for all prices $p \in \mathbb{R}^\ell$ and parameters $\theta \in \mathbb{R}^d$:

$$\widehat{U}(x,p,s;\theta) = \max_{\alpha_i\in A}\mathbb{E}\left[\int_s^T e^{-\rho(\tau-s)}R(X_{i,\tau,s},\alpha_{i,\tau,s},\widehat{p}_{\tau,s})d\tau + e^{-\rho(T-s)}V(X_{i,T,s},\widehat{p}_{T,s})\right], \tag{7.10}$$

defined for $s \leq T$. Note that the perceived value function no longer depends on the time t at which the parameter θ_t is fixed for $s > t$. As we have mentioned, the only dependence in (7.5) on t comes from the parameter θ_t that is fixed for $s > t$. Here, the agents are allowed to take various values of θ, which, in turn, allows us to get rid of the dependence on t. The perceived future prices $\widehat{p}_{\tau,s}$ evolve according to a generalization of (7.6)

$$\begin{aligned} \frac{d\widehat{p}_{\tau,s}}{d\tau} &= \mu_p(\widehat{p}_{\tau,s}, \theta), \quad s \leq \tau \leq T, \\ \widehat{p}_{s,s} &= p, \end{aligned} \tag{7.11}$$

but now with the parameter θ set to the value that appears in the argument of $\widehat{u}(x, p, \theta, s)$ in (7.10).

The perceived value function $\widehat{U}(x, p, s; \theta)$ defined in (7.10) solves a single HJB equation

$$\begin{aligned} \rho\widehat{U}(x, p, s; \theta) - \partial_s\widehat{U}(x, p, s; \theta) &= H(x, p, \nabla_x\widehat{U}) + \nu\Delta_x\widehat{U}(x, p, s; \theta) \\ &\quad + \mu_p(p, \theta) \cdot \nabla_p\widehat{U}(x, p, s; \theta), \\ H(x, p, \lambda) &= \max_{\alpha \in A}\{R(x, \alpha, p) + \alpha \cdot \lambda\}, \end{aligned} \tag{7.12}$$

with $(x, p, s) \in \mathbb{R}^n \times \mathbb{R}^\ell \times (0, T)$ and $\theta \in \mathbb{R}^d$.

Note that the arguments of the value function include not only the low-dimensional prices $p \in \mathbb{R}^\ell$ but also the "fairly high-dimensional" parameter vector $\theta \in \mathbb{R}^d$. While including p comes at a low computational cost, including θ comes at a significantly higher additional computational cost. However, note that Eqs. (7.12) for different θ are decoupled from each other. As we will remark below, this is something one can exploit to reduce computational costs.

The HJB equation (7.12) for the perceived value function gives rise to the perceived policy

$$\widehat{\pi}(x, p, s; \theta) = \arg\max_{\alpha \in A}\{R(x, \alpha, p) + \alpha \cdot \nabla_x\widehat{U}(x, p, s; \theta)\}. \tag{7.13}$$

Analogously to above, the actual time-t policy is then given by the perceived policy evaluated at the time-t price p_t and parameter θ_t:

$$\pi(x, t) = \widehat{\pi}(x, p_t, t; \theta_t). \tag{7.14}$$

The prices p_t on the right side of (7.14) are given by

$$p_t = P^*(m(t)). \tag{7.15}$$

The measure $m(x, t)$ and the parameter θ_t, in turn, solve the forward-in-time problems

$$\begin{aligned} \partial_t m(x, t) &= \mathcal{A}_\pi^* m(x, t), \\ \dot{\theta}_t &= L(p_t, \theta_t), \end{aligned} \tag{7.16}$$

with the policy π given by (7.14) and with initial conditions $m_0(x)$ and θ_0. The system (7.14)–(7.16) is driven by the solution to (7.12) via the policy $\widehat{\pi}$ that appears in (7.14).

Remark on computational cost of (7.12). The price for solving a single standard HJB equation for $\widehat{U}$ is the need to solve (7.12) for all θ in the region of interest. Thus, the computational complexity of this formulation is controlled by the dimension d of the parameter space $\theta \in \mathbb{R}^d$ which may be considerably higher than in (7.9). However, one can take advantage of the fact that the HJB equations (7.12) are decoupled for different θ here. In particular, this means that one can solve (7.12) only for the θ's one actually encounters. Specifically, starting from an initial condition θ_0, solve the HJB equation (7.12) for $\theta = \theta_0$; then update θ according to (7.16), solve the HJB equation again for this new θ, and so on.

Internalized learning. In the HJB equation (7.12), agents do not take into account that the parameter vector $\widehat{\theta}_t$ evolves according to the learning rule (7.4). Hence, in (7.12) learning is "external" to agents. An alternative approach pursued in part of the economics literature (e.g., [17]) is to instead assume that agents "internalize" learning, taking into account the evolution of $\widehat{\theta}_t$. The corresponding variant of (7.12) with internalized learning is

$$\begin{aligned} &\rho\widehat{U} - \partial_s\widehat{U} = H(x,p,\nabla_x\widehat{U}) + \nu\Delta_x\widehat{U} + \mu_p(p,\theta)\cdot\nabla_p\widehat{U} + L(p,\theta)\nabla_\theta\widehat{U}, \\ &H(x,p,\lambda) = \max_{\alpha\in A}\{R(x,\alpha,p) + \alpha\cdot\lambda\}, \end{aligned} \tag{7.17}$$

with $(x,p,s) \in \mathbb{R}^n \times \mathbb{R}^\ell \times (0,T)$ and $\theta \in \mathbb{R}^d$. Both formulations are sensible from an economic perspective and simply correspond to different assumptions about agents' sophistication.

Belief heterogeneity. In the system (7.9), the parameter θ_t, that couples its forward- and backward-in-time components, summarizes agents' beliefs about the evolution of the future prices. Although these beliefs vary over time, and θ_t evolves in (7.9) according to a differential equation, the assumption leading to (7.9) is that all agents share the *same* beliefs θ_t. As remarked in Sec. 5.1, it is natural to allow for heterogeneity in beliefs which is an important feature of real-world data on empirical measures of such beliefs.

A convenient feature of the formulation (7.12) in which θ is a state variable is that it is easy to extend to the case of belief heterogeneity which we model as follows. Beliefs θ differ across the population and there is an initial joint density of states and beliefs $m_0(x,\theta)$. Starting from time $t = 0$, beliefs evolve according to the following generalization of (7.4) which allows for learning to depend on the individual state $X_{i,t}$ as well:

$$\dot{\theta}_{i,t} = L(p_t, X_{i,t}, \theta_{i,t}), \quad \theta_{i,0} \sim m_0(x,\theta). \tag{7.18}$$

The value function of an agent with beliefs θ still satisfies the same HJB equation (7.12). The actual policy of that agent is then given by

$$\pi(x,\theta,t)=\widehat{\pi}(x,p_t,t;\theta)=\arg\max_{\alpha\in A}\{R(x,\alpha,p_t)+\alpha\cdot\nabla_x\widehat{u}(x,p_t,t;\theta)\},\quad (7.19)$$

meaning that their trajectories are

$$dX_{i,t}=\pi(x,\theta,t)d\tau+\sqrt{2\nu}dB_{i,\tau}.$$

The prices p_t on the right side of (7.19) are still given by

$$p_t=P^*(m(t)).\quad (7.20)$$

The measure $m(x,\theta,t)$, in turn, solves the forward-in-time problem

$$\partial_t m(x,\theta,t)=\mathcal{A}_\pi^* m(x,\theta,t)-\mathrm{div}_\theta(L(p_t,x,\theta)m(x,\theta,t)),\quad (7.21)$$

with the policy π given by (7.19).

It is worth contrasting the system (7.19)–(7.21) with belief heterogeneity with its counterpart (7.14)–(7.16) for the case of homogeneous beliefs. The structure of the two systems is exactly the same, except that the system (7.19)–(7.21) tracks a joint distribution for (x,θ) rather than a marginal distribution for x.

7.2. *Adaptive learning with common noise*: *Sidestepping the Master equation*

Starting from the second formulation in the preceding section with (p,θ) as state variables, the generalization to the case of common noise is straightforward. The key point of this section will be that modeling adaptive learning in MFGs with a low-dimensional coupling and common noise allows for sidestepping the infinite-dimensional Master equation.

Agents' perceived law of motion for prices is

$$d\widehat{p}_{s,t}=\mu_p(\widehat{p}_{s,t},Z_s,\theta)ds+\sigma_p(\widehat{p}_{s,t},Z_s,\theta)dB_s,\quad (7.22)$$

where $\theta\in\mathbb{R}^d$ is a parameter vector. The agents' estimate of θ is still updated according to the learning rule (7.4).

Denoting the generator corresponding to (7.22) by $\mathcal{A}_p(z,\theta)$, we have the following HJB equation for the perceived value function $\widehat{U}(x,z,p,t;\theta)$ which is similar to (7.12):

$$\begin{aligned}&\rho\widehat{U}-\partial_s\widehat{U}=H(x,z,p,\nabla_x\widehat{U})+\nu\Delta_x\widehat{U}+\mathcal{A}_p(z,\theta)\widehat{U}+\beta\Delta_z\widehat{U},\\&H(x,z,p,\lambda)=\max_{\alpha\in A}\{R(x,z,\alpha,p)+\alpha\cdot\lambda\},\end{aligned}\quad (7.23)$$

with $(x,z,p,t)\in\mathbb{R}^n\times\mathbb{R}^k\times\mathbb{R}^\ell\times(0,T),\theta\in\mathbb{R}^d$, and with corresponding policy

$$\widehat{\pi}(x,z,p,t;\theta)=\arg\max_{\alpha\in A}\{R(x,z,\alpha,p)+\alpha\cdot\nabla_x\widehat{U}(x,z,p,t;\theta)\}.\quad (7.24)$$

The important observation is that (7.23) is a standard finite-dimensional HJB equation rather than an infinite-dimensional Master equation.

To solve for the evolution of the density m_t, equilibrium prices p_t, and the learned parameter estimates θ_t, one can then proceed in the same fashion as in the case without common noise. First, define

$$\pi(x, Z_t, t) = \widehat{\pi}(x, Z_t, p_t, t; \theta_t), \tag{7.25}$$

where the price p_t is given by

$$p_t = P^*(m_t, Z_t). \tag{7.26}$$

Then solve the following forward-in-time system:

$$\begin{aligned} dm_t &= \mathcal{A}^*_{\pi_t, Z_t} m_t dt, \\ dZ_t &= \sqrt{2\beta} dB_t, \\ \dot{\theta}_t &= L(p_t, \theta_t), \end{aligned} \tag{7.27}$$

with the policy π given by (7.25) and with initial condition (m_0, Z_0, θ_0).

Remark on computational cost of (7.23)**.** The same remark as in the case without common noise applies: one does not actually have to solve (7.23) for all values of $\theta \in \mathbb{R}^d$. This is because the equations for different θ are decoupled from each other. Starting from θ_0, and simulating (7.27) forward in time, it is therefore sufficient to solve (7.23) only for the θ's one actually encounters.

Either way, this computational cost should be compared to the cost of computing the infinite-dimensional Master equation with rational expectations. The former is clearly lower regardless of the exact computational strategy for (7.23).

7.3. *Other directions*: *Reinforcement learning and other stochastic approximation methods*

One other approach for sidestepping the Master equation is to approximate the value function in Sec. 6.3 using ideas from the literature on RL. RL means learning value or policy functions of incompletely-known Markov decision processes (MDPs) via Monte Carlo simulation [45]. RL is typically formulated in discrete time but there exist continuous-time formulations [19, 30, 47].

Yang *et al.* [50] and Wibault *et al.* [48] apply RL ideas to discrete-time MFGs with a low-dimensional coupling and common noise of the type studied here. Specifically, Yang *et al.* [50] propose a "Structural Policy Gradient" (SPG) method which leverages the known structure of individual dynamics (the generator $\mathcal{A}_\alpha$) while treating prices p_t via simulation. While they impose that policy functions depend only on current prices, Wibault *et al.* [48] keep track of full price histories using recurrent neural networks as in recurrent RL [25, 41].

More generally, a promising approach could be to approximate these value functions using a "stochastic approximation algorithm" (e.g., [36, 42]) of which RL is a special case [28, 46].

8. Discrete Time

There is also a literature on discrete-time MFGs (e.g., [21, 22]) and this formulation may be useful for the application of some promising approaches to the challenge posed in this paper such as RL. We therefore briefly repeat our paper's arguments in discrete time. For brevity we skip the case without common noise and focus directly on the more challenging case with common noise. For simplicity, we also focus on the case with finite state and action spaces, though this simplification is not essential.

8.1. *Setup*: *Discrete-time MFGs with common noise*

Everything is analogous to the continuous-time setup in Sec. 2. Consider a system of $N \gg 1$ individual agents (players) at positions (states) $X_{i,t} \in \mathcal{X} \subset \mathbb{R}^n$, $i = 1, \dots, N$, where $\mathcal{X}$ is a finite state space, i.e. $X_{i,t}$ can take only finitely many possible values. Time is discrete, $t = 0, 1, \dots, T$, where T is a fixed terminal time that is sometimes taken as $T = +\infty$. As above, we consider the limit $N \to +\infty$ of a large number of agents. Anticipating this limit we here write equations directly in terms of the limiting density which we denote by $m_t(x) \in \mathcal{P}(\mathcal{X})$, the space of probability measures with support in $\mathcal{X}$. Given x can take only finitely many values, this density is simply a high- but finite-dimensional vector (essentially a "histogram"). The setup and notation are close to [33, 48].

Agents receive a period reward $R(X_{i,t}, Z_t, \alpha_{i,t}, m_t)$ and their state evolves according to a Markov process

$$X_{i,t+1} \sim \mathcal{T}_x(\cdot|X_{i,t}, Z_t, \alpha_{i,t}, m_t). \tag{8.1}$$

As above, $\alpha_{i,t} \in A \subset \mathbb{R}^n$ is a control (but with A a finite action space) and $Z_t \in \mathcal{Z} \subset \mathbb{R}^k$ is the aggregate state (common noise) that affects all agents and which evolves according to an exogenous Markov process:

$$Z_{t+1} \sim \mathcal{T}_z(\cdot|Z_t). \tag{8.2}$$

Agents maximize the cumulative discounted reward:

$$u_{i,0} = \max_{\alpha_i \in A} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R(X_{i,t}, Z_t, \alpha_{i,t}, m_t) + \gamma^T V(X_{i,T}, Z_T, m_T)\right] \tag{8.3}$$

subject to (8.1) and (8.2) and where V is a terminal value and $0 < \gamma \leq 1$ a discount factor.

As above, the optimal policy induces the density m_t to evolve over time. This evolution is easiest to spell out for a slight generalization of the problem (8.3) in which we allow for stochastic policies of the form:

$$\alpha_{i,t} \sim \pi_t(\cdot|X_{i,t}, Z_t, m_t), \tag{8.4}$$

where π_t is the probability distribution over actions $\alpha_{i,t}$ conditional on the states $(X_{i,t}, Z_t, m_t)$. Given the optimal policy π_t, the density m_t then evolves according to a Chapman–Kolmogorov equation (the discrete-time analogue of a Fokker–Planck equation)

$$m_{t+1}(x) = \sum_{\tilde{x},\tilde{\alpha}} m_t(\tilde{x})\pi_t(\tilde{\alpha}|\tilde{x}, Z_t, m_t)\mathcal{T}_x(x|\tilde{x}, \tilde{\alpha}, Z_t, m_t) \tag{8.5}$$

which we can also write

$$m_{t+1} = \mathbf{A}^{\mathrm{T}}_{\pi_t,Z_t} m_t, \tag{8.6}$$

where $\mathbf{A}_{\pi_t,Z_t}$ is the transition matrix of x induced by the optimal policy π_t. This equation is the discrete-time counterpart to the Fokker–Planck equation (5.21), with the transition matrix $\mathbf{A}_{\pi_t,Z_t}$ being the counterpart to the generator $\mathcal{A}_{\pi,Z_t}$ (and the matrix transpose that to the operator adjoint). Note that, because Z_t evolves according to the Markov process (8.2), the density m_{t+1} is itself a stochastic process.

Mean Field Games with a low-dimensional coupling. As in Sec. 3, we pay special attention to the class of "MFGs with a low-dimensional coupling". Using analogous notation, agents choose policies π to maximize

$$u_{i,0} = \max_{\alpha_i \in A} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t \widetilde{R}(X_{i,t}, Z_t, \alpha_{i,t}, p_t) + \gamma^T \widetilde{V}(X_{i,T}, Z_T, p_T)\right] \tag{8.7}$$

subject to

$$X_{i,t+1} \sim \mathcal{T}_x(\cdot|X_{i,t}, Z_t, \alpha_{i,t}, p_t), \tag{8.8}$$

and (8.2) where

$$p_t = P^*(m_t, Z_t), \tag{8.9}$$

for a fixed functional $P^* : \mathcal{P}(\mathcal{X}) \times \mathcal{Z} \to \mathbb{R}^\ell$. Again note that model agents do not directly "care about" the density m_t in the sense that it does not enter their running reward functions; instead they only "care about" the much lower-dimensional vector p_t. As discussed in Secs. 3.1 and 3.2 such MFGs arise naturally in macroeconomics where they are known as "heterogeneous agent models" (e.g., [18, 31]). In this case, the vector p_t has the interpretation of "equilibrium prices" that are determined from some "market clearing" conditions analogous to (3.5).

Rational expectations. As above, standard formulations of this MFG impose rational expectations: agents know not only the correct transition probabilities $\mathcal{T}_x$ and $\mathcal{T}_z$ but also the high-dimensional transition matrix $\mathbf{A}_{\pi_t,Z_t}$ which governs the evolution of the complex system they inhabit.

Master equation in discrete-time MFGs. In both general MFGs and MFGs with a low-dimensional coupling, under rational expectations, the agents' optimization problem gives rise to the Master equation (i.e. Bellman equation on the space of probability measures):

$$
\begin{aligned}
U_t(x,z,m) &= \max_{\alpha} R(x,z,\alpha,m) + \gamma \mathbb{E}_{x',z'}[U_{t+1}(x',z',m')|x,z,m] \\
&\text{subject to} \\
x' &\sim \mathcal{T}_x(\cdot|x,z,\alpha,m), \\
z' &\sim \mathcal{T}_z(\cdot|z), \\
m' &= \mathbf{A}^{\mathrm{T}}_{\pi_t,z} m, \\
U_T(x,z,m) &= V(x,z,m).
\end{aligned}
\tag{8.10}
$$

This equation is the exact discrete-time counterpart to (5.23) and it similarly features the state variable $m \in \mathcal{P}(\mathcal{X})$, the space of probability measures with support in $\mathcal{X}$. It therefore suffers from an extreme version of the curse of dimensionality.

Section 6.3 criticized the continuous-time Master equation in MFGs with a low-dimensional coupling. All the same criticisms also apply to its discrete-time counterpart.

8.2. *Discrete-time MFGs without rational expectations*

General MFGs without rational expectations. As above, we assume that agents have rational expectations about the evolution of their own individual state $X_{i,t}$, i.e. that they know the correct transition probabilities $\mathcal{T}_x$. However, like in Sec. 5.2, we allow for the possibility that they may have non-rational expectations about the evolution of the density m_t and the aggregate state Z_t. In particular, agents believe that the density and aggregate state evolve according to the perceived laws of motion

$$
\begin{aligned}
\widehat{m}_{s+1,t} &= \mathbf{B}^{\mathrm{T}}_{\widehat{Z}_t} \widehat{m}_{s,t}, \quad s \geq t, \\
\widehat{Z}_{s+1,t} &\sim \widehat{\mathcal{T}}_z(\cdot|\widehat{Z}_{s,t}), \quad s \geq t,
\end{aligned}
\tag{8.11}
$$

with $\widehat{m}_{t,t} = m_t$ and $\widehat{Z}_{t,t} = Z_t$ rather than (8.2) and (8.6). Under this assumption, the agents' optimization problem gives rise to a Master equation that is just like (8.10) but with $\mathcal{T}_z$ replaced by $\widehat{\mathcal{T}}_z$ and $\mathbf{A}_{\pi,z}$ replaced by $\mathbf{B}_{\widehat{z}}$. Rational expectations impose that

$$
\mathbf{B}_z = \mathbf{A}_{\pi,z} \quad \text{and} \quad \widehat{\mathcal{T}}_z = \mathcal{T}_z
$$

so that, in the special case of rational expectations, we recover the usual Master equation (8.10).

Sidestepping the Master equation in MFGs with a low-dimensional coupling. It is again most interesting to consider non-rational expectations in MFGs with a low-dimensional coupling because this holds the promise of sidestepping the Master equation altogether.

With rational expectations, agents understand the dependence of p_t on m_t and Z_t and therefore use the function P^* together with the correct stochastic processes for m_t and Z_t to predict future values of p_t. This leads to the Master equation (8.10), with essentially no simplifications despite the low-dimensional coupling.

With non-rational expectations, agents instead perceive some other stochastic process for the pair (p_t, Z_t). Like in Sec. 6.2, they could simply perceive p_t to evolve according to an exogenous Markov process

$$\widehat{p}_{s+1,t} \sim \widehat{\mathcal{T}}_p(\cdot|\widehat{p}_{s,t}), \quad s \geq t, \quad \widehat{p}_{t,t} = p_t. \tag{8.12}$$

In more complicated cases, agents may perceive a joint stochastic process for p_t, Z_t and other variables.

In the case of agents perceiving the simple process (8.12), instead of writing a Master equation, we can write a much simpler, standard finite-dimensional Bellman equation (as above, going forward, we drop the tildes from $\widetilde{R}$ and $\widetilde{V}$ for notational simplicity):

$$\begin{aligned}
\widehat{U}_t(x,z,p) &= \max_{\alpha} R(x,z,\alpha,p) + \gamma \mathbb{E}_{x',z',p'}[U_{t+1}(x',z',p')|x,z,p] \\
&\quad \text{subject to} \\
&\qquad x' \sim \mathcal{T}_x(\cdot|x,z,\alpha,p), \\
&\qquad z' \sim \mathcal{T}_z(\cdot|z), \\
&\qquad p' \sim \widehat{\mathcal{T}}_p(\cdot|p), \\
\widehat{U}_T(x,z,p) &= V(x,z,p).
\end{aligned} \tag{8.13}$$

Therefore, in MFGs with a low-dimensional coupling, departing from rational expectations can completely sidestep the Master equation. Of course, the case considered here is just an illustrative example. In particular, note that the perceived law of motion (6.3) is specified completely "outside the model" which leaves open the question where this perceived law of motion "comes from" in the first place.

8.3. *The challenge and a Markov reward process*

As discussed in Sec. 7 and [39], the challenge is: how can we formulate, in a systematic way, models of agents' behavior in situations with a low-dimensional coupling that lead to equations that (i) approximate agents' real-world behavior, and (ii) sidestep computing the solutions to a Master equation with the infinite-dimensional state $m \in \mathcal{P}(\mathcal{X})$ and the associated curse of dimensionality?

A Markov Reward Process with all the difficulty. To understand the key difficulty, it is useful to consider a simplified version of the model with no actions α, i.e. a Markov Reward Process (MRP) rather than an MDP. This material is adapted from ongoing work by Yang *et al.* [50].

To this end, eliminate actions α and assume that the state $X_{i,t}$ evolves exogenously according to a transition matrix $\mathbf{A}_{Z_t}$ that depends on the aggregate state Z_t, implying that (m_t, Z_t) evolve as

$$m_{t+1} = \mathbf{A}_{Z_t}^{\mathrm{T}} m_t, \quad Z_{t+1} \sim \mathcal{T}_z(\cdot|Z_t). \tag{8.14}$$

Similarly, replace the running reward and terminal value in (8.7) by a reward $R(p)$ and terminal value $V(p)$ that depend only on the low-dimensional price vector $p_t \in \mathbb{R}^{\ell}$. As before, the vector p_t is still determined by (8.9).

The simplified problem is to compute the value of a MRP: given p_t evolving according to (8.9) and (m_t, Z_t) according to (8.14), compute the expected present discounted value (PDV) of rewards:

$$u_0 = \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R(p_t) + \gamma^T V(p_T)\right]. \tag{8.15}$$

This problem contains all the difficulty of the more complicated problem in MFGs with a low-dimensional coupling.

The "correct" way — in the rational expectations sense — of computing the value of this MRP is to solve a Master equation for the value function $U_t(z, m)$:

$$\begin{aligned} U_t(z,m) &= R(P^*(m,z)) + \gamma \mathbb{E}_{z'}[U_{t+1}(z',m')|z,m] \\ &\text{subject to} \\ z' &\sim \mathcal{T}_z(\cdot|z), \\ m' &= \mathbf{A}_z^{\mathrm{T}} m, \\ U_T(z,m) &= V(P^*(m,z)). \end{aligned} \tag{8.16}$$

This Master equation illustrates the trouble with the rational expectations assumption: even though the reward function is only a function of a low-dimensional vector $p_t \in \mathbb{R}^{\ell}$, computing the PDV in (8.15) requires solving a Bellman equation on the space of probability measures $\mathcal{P}(\mathcal{X})$.

The difficulty: prices p_t are not Markov. As discussed in Sec. 6, the key difficulty is that the vector p_t does not follow a Markov process; instead only (m_t, Z_t) has the Markov property and p_t is instead a complicated nonlinear functional of this Markov state. Agents with rational expectations therefore (unrealistically) forecast the Markov state (m_t, Z_t) in order to forecast the non-Markovian p_t.

8.4. *Adaptive learning in discrete time*

Finally, we show how to write the adaptive learning model of Sec. 7 in discrete time. The economics literature typically formulates such models in discrete time [10, 20, 29, 38]. The key assumption is that, to forecast prices, agents use a *perceived* *law of motion* in the form of a Markov process:

$$\widehat{p}_{s+1,t} \sim \widehat{\mathcal{T}}_p(\cdot|\widehat{p}_{s,t}, Z_s, \theta), \quad s \geq t, \quad \widehat{p}_{t,t} = p_t, \tag{8.17}$$

where $\theta \in \mathbb{R}^d$ is a parameter vector. The key difference to the Markov process (8.12) in Sec. 8.2 is that the process is endogenous to the model because agents learn the parameter vector θ over time from past observations of p_t. Specifically, they form an estimate $\widehat{\theta}_t$ of θ which they update using the learning rule

$$\widehat{\theta}_{t+1} = L(p_t, \widehat{\theta}_t). \tag{8.18}$$

For example, (8.17) could be a vector autoregressive (VAR) process for the vector of prices and (8.18) a recursive least-squares estimator for the parameters of this VAR.

Dropping the hat and subscript from $\widehat{\theta}_t$ for notational simplicity (but keeping in mind that this is really a time-varying estimate of the parameter θ), the agents' optimization problem given a current estimate θ is

$$\widehat{U}_t(x, z, p; \theta) = \max_{\alpha_i \in A} \mathbb{E}\left[\sum_{s=t}^{T} \gamma^{s-t} R(X_{i,s}, Z_t, \alpha_{i,s}, \widehat{p}_{s,t}) + \gamma^{T-t} V(X_{i,T}, Z_T, \widehat{p}_{T,t})\right] \tag{8.19}$$

subject to

$$\widehat{p}_{s+1,t} \sim \widehat{\mathcal{T}}_p(\cdot|\widehat{p}_{s,t}, Z_s, \theta), \quad s \geq t, \tag{8.20}$$

$$\widehat{p}_{t,t} = p \tag{8.21}$$

and subject to (8.2) and (8.8).

The corresponding Bellman equation is

$$\begin{aligned} \widehat{U}_t(x, z, p; \theta) &= \max_{\alpha} \; R(x, z, \alpha, p) + \gamma \mathbb{E}_{x', z', p'}[U_{t+1}(x', z', p'; \theta)|x, z, p] \\ &\quad \text{subject to} \\ &\qquad x' \sim \mathcal{T}_x(\cdot|x, z, \alpha, p), \\ &\qquad z' \sim \mathcal{T}_z(\cdot|z), \\ &\qquad p' \sim \widehat{\mathcal{T}}_p(\cdot|p, z, \theta), \\ \widehat{U}_T(x, z, p; \theta) &= V(x, z, p), \end{aligned} \tag{8.22}$$

with corresponding optimal policy $\widehat{\pi}_t(x, z, p, \theta)$. The key observation is that this is a standard finite-dimensional Bellman equation rather than an infinite-dimensional Master equation. Analogous to the discussion in Sec. 7.1, the Bellman equation (8.22) assumes that learning is external to agents and one can instead formulate a variant with internalized learning (e.g., [17]). To obtain this variant, we simply replace θ in $U_{t+1}(x', z', p'; \theta)$ by θ' and add the law of motion $\theta' = L(p, \theta)$ to the constraint set.

With the solution in hand, the evolution of the density m_t and equilibrium prices p_t are found as follows. First, define

$$\pi_t(x, Z_t) = \widehat{\pi}_t(x, Z_t, p_t; \theta_t), \tag{8.23}$$

where the price p_t is given by

$$p_t = P^*(m_t, Z_t). \tag{8.24}$$

Then solve the following forward-in-time system:

$$\begin{aligned} m_{t+1} &= \mathbf{A}^{\mathrm{T}}_{\pi_t, Z_t} m_t, \\ Z_{t+1} &\sim \mathcal{T}_z(\cdot|Z_t), \\ \theta_{t+1} &= L(p_t, \theta_t), \end{aligned} \tag{8.25}$$

with the policy π_t given by (8.23) and with initial condition (m_0, Z_0, θ_0). A similar problem is solved by Jacobson [29].

Remark on computational cost of (8.22)**.** As noted in Sec. 7, one does not actually have to solve (8.22) for all values of $\theta \in \mathbb{R}^d$. This is because the equations for different θ are decoupled from each other. Starting from θ_0, and simulating (8.25) forward in time, it is therefore sufficient to solve (8.22) only for the θ's one actually encounters.

Remark on relation to Krusell and Smith [31] algorithm. There is a link between this adaptive learning approach (specifically, the variant with least-squares learning) and the algorithm of Krusell and Smith [31]. In both approaches, decision-makers use a perceived law of motion like (8.17) and estimate its coefficients via least squares. A difference is that adaptive learning updates the coefficient estimate θ_{t+1} from θ_t recursively over time so that solving for the MFG equilibrium and belief updating are done "in one sweep" via solving (8.25) forward in time.

9. Conclusion

This paper has shown how to formulate MFGs without rational expectations, i.e. without the assumption that agents know all relevant transition probabilities for the complex system they inhabit. Instead of using the correct transition probabilities, agents instead use some other "non-rational" transition probabilities when solving their optimization problems. We show how to write the corresponding equations describing the Nash equilibrium of the MFG, both for the case with and without common noise. In the special case of rational expectations we recover the standard backward–forward MFG system and MFG Master equation.

Departing from rational expectations is particularly relevant when there is common noise in "MFGs with a low-dimensional coupling", i.e. MFGs in which agents' running reward function depends on the density only through low-dimensional functionals, which are typical in macroeconomics. In MFGs with a low-dimensional coupling, departing from rational expectations allows for completely sidestepping the

Master equation and for instead solving finite-dimensional HJB equations. We introduced an adaptive learning model as a particular example of non-rational expectations and discussed its properties.

Acknowledgments

We thank Charles Bertucci, Larry Christiano, Jean-Michel Lasry and Peter Tankov for useful comments. BM was partially supported by the Leverhulme Trust. LR was partially supported by NSF grant DMS-2205497 and by ONR grant N00014-22-1-2174.

ORCID

Benjamin Moll https://orcid.org/0009-0003-6067-359X
Lenya Ryzhik https://orcid.org/0000-0003-1892-4861

References

[1] Y. Achdou, F. J. Buera, J.-M. Lasry, P.-L. Lions and B. Moll, Partial differential equation models in macroeconomics, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **372**(2028) (2014) 20130397.

[2] Y. Achdou, J. Han, J.-M. Lasry, P.-L. Lions and B. Moll, Income and wealth distribution in macroeconomics: A continuous-time approach, *Rev. Econ. Stud.* **89**(1) (2021) 45–86.

[3] S. Ahuja, Wellposedness of mean field games with common noise under a weak monotonicity condition, *SIAM J. Control Optim.* **54**(1) (2016) 30–48.

[4] S. Ahuja, W. Ren and T.-W. Yang, Forward–backward stochastic differential equations with monotone functionals and mean field games with common noise, *Stochastic Process. Appl.* **129**(10) (2019) 3859–3892.

[5] S. R. Aiyagari, Uninsured idiosyncratic risk and aggregate saving, *Q. J. Econ.* **109**(3) (1994) 659–684.

[6] C. Bertucci, Mean field games with incomplete information, Working paper (2023).

[7] C. Bertucci and C. Meynard, Noise through an additional variable for mean field games master equation on finite state space, Working paper (2024).

[8] C. Bertucci and C. Meynard, A study of common noise in mean field games, Working paper (2024).

[9] T. Bewley, Stationary monetary equilibrium with a continuum of independently fluctuating consumers, in *Contributions to Mathematical Economics in Honor of Gerard Debreu*, eds. W. Hildenbrand and A. Mas-Collel (North-Holland, Amsterdam, 1986).

[10] M. Bray, Learning, estimation, and the stability of rational expectations, *J. Econ. Theory* **26**(2) (1982) 318–339.

[11] P. D. Cagan, The monetary dynamics of hyperinflation, in *Studies in the Quantity Theory of Money*, ed. M. Friedman (The University of Chicago Press, Chicago, 1956), pp. 25–117.

[12] P. Cardaliaguet, Notes on mean-field games (from P.-L. Lions' lectures at Collège de France) (2013), https://www.ceremade.dauphine.fr/˜cardaliaguet/MFG20130420.pdf.

[13] P. Cardaliaguet, F. Delarue, J.-M. Lasry and P.-L. Lions, *The Master Equation and the Convergence Problem in Mean Field Games*, Annals of Mathematics Studies, Vol. 201 (Princeton University Press, 2019).

[14] P. Cardaliaguet and A. Porretta, An introduction to mean field game theory, in *Mean Field Games*, Lecture Notes in Mathematics, Vol. 2281 (Springer, Cham, 2020), pp. 1–158.

[15] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications. I: Mean Field FBSDEs, Control, and Games*, Probability Theory and Stochastic Modelling, Vol. 83 (Springer, Cham, 2018).

[16] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications. II: Mean Field Games with Common Noise and Master Equations*, Probability Theory and Stochastic Modelling, Vol. 84 (Springer, Cham, 2018).

[17] L. J. Christiano, M. S. Eichenbaum and B. K. Johannsen, Slow learning, NBER Working Paper No. 32358, National Bureau of Economic Research (2024).

[18] W. J. Den Haan, Heterogeneity, aggregate uncertainty, and the short-term interest rate, *J. Bus. Econ. Stat.* **14**(4) (1996) 399–411.

[19] K. Doya, Reinforcement learning in continuous time and space, *Neural Comput.* **12**(1) (2000) 219–245.

[20] G. W. Evans and S. Honkapohja, *Learning and Expectations in Macroeconomics* (Princeton University Press, 2001).

[21] D. A. Gomes, J. Mohr and R. R. Souza, Discrete time, finite state space mean field games, *J. Math. Pures Appl.* **93**(3) (2010) 308–328.

[22] D. A. Gomes, J. Mohr and R. R. Souza, Continuous time finite state mean field games, *Appl. Math. Optim.* **68**(1) (2013) 99–143.

[23] J.-M. Grandmont, Temporary general equilibrium theory, *Econometrica* **45**(3) (1977) 535–572.

[24] J.-M. Grandmont, Temporary equilibrium, in *General Equilibrium* (Springer, 1989), pp. 297–304.

[25] M. Hausknecht and P. Stone, Deep recurrent Q-learning for partially observable MDPs, preprint (2015), arXiv:1507.06527.

[26] J. R. Hicks, *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory* (Clarendon Press, 1939), https://benjaminmoll.com/Hicks_Value_and_Capital/.

[27] M. Huggett, The risk-free rate in heterogeneous-agent incomplete-insurance economies, *J. Econ. Dyn. Control* **17**(5–6) (1993) 953–969.

[28] T. Jaakkola, M. Jordan and S. Singh, Convergence of stochastic iterative dynamic programming algorithms, *Adv. Neural Inf. Process. Syst.* **6** (1993) 703–710.

[29] M. M. Jacobson, Beliefs, aggregate risk, and the U.S. housing boom, Finance and Economics Discussion Series 2022-061, Board of Governors of the Federal Reserve System (U.S.) (2025).

[30] Y. Jia and X. Y. Zhou, Q-learning in continuous time, *J. Mach. Learn. Res.* **24**(1) (2023) 7675–7735.

[31] P. Krusell and A. A. Smith, Income and wealth heterogeneity in the macroeconomy, *J. Polit. Econ.* **106**(5) (1998) 867–896.

[32] J.-M. Lasry and P.-L. Lions, Mean field games, *Jpn. J. Math.* **2** (2007) 229–260.

[33] M. Laurière, S. Perrin, S. Girgin, P. Muller, A. Jain, T. Cabannes, G. Piliouras, J. Perolat, R. Elie, O. Pietquin and M. Geist, Scalable deep reinforcement learning algorithms for mean field games, in *Proc. 39th Int. Conf. Machine Learning*, eds. K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu and S. Sabato, Proceedings of Machine Learning Research, Vol. 162 (PMLR, 2022), pp. 12078–12095.

[34] M. Laurière, S. Perrin, J. Pérolat, S. Girgin, P. Muller, R. Élie, M. Geist and O. Pietquin, Learning in mean field games: A survey, preprint (2022), arXiv:2205.12944.

[35] E. Lindahl, *Studies in the Theory of Money and Capital* (Allen and Unwin, 1939).

[36] L. Ljung, Analysis of recursive stochastic algorithms, *IEEE Trans. Automat. Control* **22**(4) (1977) 551–575.

[37] R. E. Lucas, Econometric policy evaluation: A critique, *Carnegie-Rochester Conf. Ser. Public Policy* **1** (1976) 19–46.

[38] A. Marcet and T. J. Sargent, Convergence of least squares learning mechanisms in self-referential linear stochastic models, *J. Econ. Theory* **48**(2) (1989) 337–368.

[39] B. Moll, The trouble with rational expectations in heterogeneous agent models: A challenge for macroeconomics, Economic Journal Lecture, Royal Economic Society (2025), https://benjaminmoll.com/challenge/.

[40] J. F. Muth, Rational expectations and the theory of price movements, *Econometrica* **29**(3) (1961) 315–335.

[41] T. Ni, B. Eysenbach and R. Salakhutdinov, Recurrent model-free RL can be a strong baseline for many POMDPs, preprint (2021), arXiv:2110.05038.

[42] H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Stat.* **22**(3) (1951) 400–407.

[43] L. Ryzhik, Lecture notes for a reading course on mean-field games (2018), https:// math.stanford.edu/˜ryzhik/STANFORD/MEAN-FIELD-GAMES/ notes-mean-field.pdf.

[44] H. A. Simon, Rationality as process and as product of thought, *Am. Econ. Rev.* **68**(2) (1978) 1–16.

[45] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, 2018).

[46] J. N. Tsitsiklis, Asynchronous stochastic approximation and Q-learning, *Mach. Learn.* **16** (1994) 185–202.

[47] H. Wang, T. Zariphopoulou and X. Y. Zhou, Reinforcement learning in continuous time and space: A stochastic control approach, *J. Mach. Learn. Res.* **21**(198) (2020) 1–34.

[48] C. Wibault, S. Towers, T. Wibault, J. Duque, J. Forkel, G. Whittle, A. Schaab, Y. Yang, C. Wang, M. Osborne, B. Moll and J. Foerster, Recurrent structural policy gradient for partially observable mean field games, preprint (2026), arXiv:2602.20141.

[49] R. Xu, Y. Min, T. Wang, M. I. Jordan, Z. Wang and Z. Yang, Finding regularized competitive equilibria of heterogeneous agent macroeconomic models via reinforcement learning, in *Proc. 26th Int. Conf. Artificial Intelligence and Statistics*, eds. F. Ruiz, J. Dy and J.-W. van de Meent, Proceedings of Machine Learning Research, Vol. 206 (PMLR, 2023), pp. 375–407.

[50] Y. Yang, C. Wang, A. Schaab and B. Moll, Structural reinforcement learning for heterogeneous agent macroeconomics, preprint (2025), arXiv:2512.18892.