

Continuous Approximation of a Discrete Situated and Reactive Multi-Agent System: Contribution to Agent Parameterization

Simon Stuker, Françoise Adreit, Jean-Marc Couveignes, Marie-Pierre Gleizes

▶ **To cite this version:**

Simon Stuker, Françoise Adreit, Jean-Marc Couveignes, Marie-Pierre Gleizes. Continuous Approximation of a Discrete Situated and Reactive Multi-Agent System: Contribution to Agent Parameterization. PRIMA 2014 - 17th Pacific Rim international Conference on Principles and Practice of Multi-Agent Systems, Dec 2014, Queenland, Australia. pp.365–380, 10.1007/978-3-319-13191-7_30 . hal-03190229

HAL Id: hal-03190229
https://hal.science/hal-03190229

Submitted on 6 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Continuous Approximation of a Discrete Situated and Reactive Multi-agent System: Contribution to Agent Parameterization

Simon Stuker[1], Françoise Adreit[1],
Jean-Marc Couveignes[2], and Marie-Pierre Gleizes[1]

[1]Institut de Recherche en Informatique de Toulouse,
Université Paul Sabatier - Toulouse III, France
[2]Institut de Mathématiques de Bordeaux,
Université de Bordeaux, France

Abstract. We propose a formal model for situated and reactive multi-agent systems based on correlated discrete random walks. In order to study this model, we construct a continuous approximation ending up on the Fokker-Planck equation. This result allows us to determine an optimal parameterization for the agents, with respect to the system's objective. Numerical simulations confirm the approach from two points of view, the validity of the continuous model and the optimality of the agents' parameterization.

Keywords: Modelling System Dynamics, Validation and Verification of Multi-Agent Systems, Multi-agent Simulation.

1 Introduction

Multi-agent systems offer an interesting approach for solving distributed problems. These problems can be industrial, like the optimal control of a heat engine [4] or resource management [2]. They can also be more academic, like constraint satisfaction problems [21] or constrained optimization of mathematical functions [18,15].

Although multi-agent solutions have proven to be efficient on difficult problems and may be methodically implemented[1], their formal validation is an important obstacle that has yet to be overcome. The main reason is that a multi-agent system can have a complex behaviour, meaning that it cannot be easily deduced from the agents' individual behaviour.

As a result, upon creation of a multi-agent system it is difficult to know if it turns out "functional", meaning that it fulfills its role. Furthermore the influence of the agents' parameters on the system's behaviour may be difficult to understand, and thus difficult to adjust if the system's performance is insufficient.

For these reasons, validating desired properties of the system and adjusting the agents' parameters are often done by experimentation. This step can be time-costly and offers only partial answers, of statistical order.

[1] By methods like TROPOS [6] or ADELFE [25].

In this article we focus on the formal establishment of basic dynamic properties of a restricted class of systems. We concentrate on systems of identical and situated agents. The objective we set for the agents is to have a desired distribution at equilibrium, given by a time-invariant function, that will be referred to as *resource*. Each individual moves randomly, according to some probabilities that depend on its local perceptions. The aim of our approach is to define and parameterize the agent's behaviour so that they position themselves, after some time, according to the target distribution.

This type of problem can be encountered in entomology, when an insect colony tries to harvest scattered food (if we ignore food repatriation), or in swarm robotics, when a group of medical robots needs to deploy in a critical situation where human victims are scattered in space.

We study long-term behaviours of such systems, and address two main questions:

1. Can we locally parameterize the agents so that the system achieves the objective?
2. Given some agent behaviour rule, can we predict their dynamics and measure their accomplishment?

Since discrete approaches have shown to be of limited use (see mentioned work in section 2) in the study of large complex systems, we study these questions by means of continuous approximations inspired by statistical physics.

Therefore, we approximate the system by two successive limits: the first (mean field) is valid when the number of agents goes to infinity, the second (continuity) is valid when the elementary time and space step tend to zero. The resulting approximated system is then studied using PDE results. The combination of these three steps provides satisfactory quantitative answers to questions 1 and 2.

After a brief presentation of related work (section 2), we propose a generic model for situated reactive multi-agent systems (section 3). We show that, under some specific hypotheses, the mathematical model can be approximated by an equation of which we can study some formal properties (section 4). Using this approximation, we manage to parameterize the agents' behaviour so that they collectively achieve the system's goal (section 5). Numerical simulations validate our approach (section 6). The article ends with a discussion of the stated results, and opens on some possible working directions (section 7).

2 Related Work

A vast literature focuses on the study of complex systems. Since multi-agent systems are endowed with some underlying objective, we divide related work in two classes: purely descriptive models, focused on the study of a given system, and control-type models, equipped with some utility function and the implicit task to maximize its value. These two classes of models are appropriate for question 2 and 1 respectively.

2.1 Descriptive Approaches

Interacting particle systems [20] model complex systems as a set of coupled random processes. The special case of Cellular automata and their application to multi-agent systems is studied in [29]. However, it is generally difficult to prove properties of these models.

A traditional population-based approach is to describe a system by the fraction of agents in each state, and its evolution by exchanges between those fractions. Well-known examples are the chemical rate equations and the Lotka-Volterra equation. Application to multi-agent systems without any formal derivations are found in [14,19]. In contrast, [17,1] provide general methods to derive these models from systems of individuals with "mean interaction".

Continuous spatial models close to the one used of this article are used in ecology to study social animal behaviours [26,16,7].

Our approach can be related to all the above works. Starting with a discrete system of random walkers with a given task, we derive some continuous model using a mean field result. However, our model also involves a global reward function that we want to maximize.

2.2 Control-Type Approaches

Markov Decision Processes [23] model stochastic systems equipped with some reward function that is to be maximized over time. This formalism is closely related to game theory [28] and its application to multi-agent systems is studied in [13,27]. However it seems difficult to use these methods for large systems, as the computational complexity increases rapidly with the number of agents [3].

Mean field methods have been used to solve optimal control problems on systems of identical individuals with global control [10,11,5] on small state spaces. Mean field games [12] can be seen as an extension of these methods to spatial systems.

In contrast to [10,11,5], we favour local control, and unlike mean field games we mainly focus on cooperative systems (where agents share identical interests) at equilibrium.

3 Formal Model

In this section we define the formal model that is used to study the class of problems we focus on. This model includes discrete time and state spaces. Agents are randomly distributed at the start, and follow discrete memoryless random walks from then on. Randomness of these movements has multiple motivations. Since agents are completely identical, a deterministic decision rule would lead to gregarious collective behaviour: all agents sharing identical information (e.g. global knowledge) move likewise, which is clearly a suboptimal behaviour. Furthermore, space exploration is often related to diffusion processes, which are mainly modeled by random walks.

3.1 System Description

The agents' **state space** is the unit circle $\mathcal{C}$, evenly divided in n_p positions (Fig. 1) and denoted

$$\mathcal{C}_\delta = \left\{ \frac{2k\pi}{n_p} \right\}_{k\in\{0,\dots,n_p-1\}} = \{k\delta\}_{k\in\{0,\dots,n_p-1\}}$$

where $\delta = \frac{2\pi}{n_p}$ is the spatial step, which will tend to 0 later, so that $\mathcal{C}_\delta \to \mathcal{C}$ (in the sense of Hausdorff's set convergence [22]).

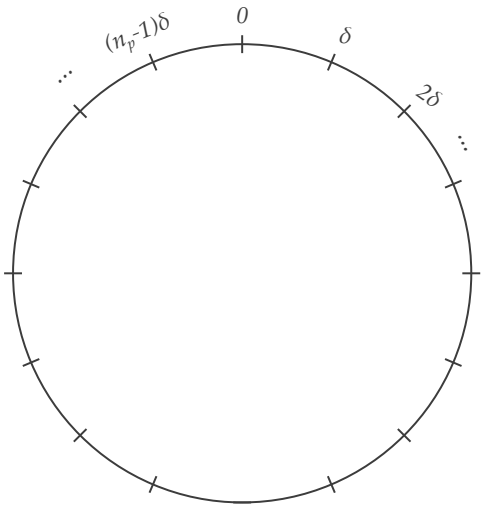


Fig. 1. Agents' positions

The periodic topology and compactness of this domain offer some theoretic advantages for the theorems of this article. Possible extensions are discussed in section 7.

Since agents are assumed to be identical, the system's state is completely described by the number of agents at each position. Let $(X_n(t))_{n=1}^N$ stand for the positions of the N agents at time t. The system's state is described by the *occupation density vector* $u^N(t) = \left(u^N(x,t)\right)_{x\in\mathcal{C}_\delta}$, where

$$u^N(x,t) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{X_n(t)=x} \text{with } \mathbb{1}_{X_n(t)=x} = \begin{cases} 1 \text{ if } X_n(t) = x \\ 0 \text{ else} \end{cases}.$$

Coordinate $u(x,t)$ is the proportion of agents located at position x. During a time interval τ, each agent may

- move by $-\delta$ with probability $p^N_{\delta,\tau}(x,u)$,
- move by $+\delta$ with probability $q^N_{\delta,\tau}(x,u)$,
- stay in place with probability $1 - p^N_{\delta,\tau}(x,u) - q^N_{\delta,\tau}(x,u)$.

These probabilities depend on the agent's current position and the density vector (Fig. 2).

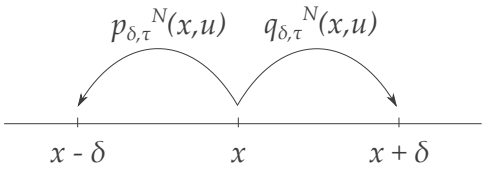


Fig. 2. Agents' movements

At this stage we have defined a generic formal model describing the random movements of a set of agents[2]. The probabilities of the agents' movements have yet to be chosen. They will act as control parameters, and will be adjusted according to some objective function defined in the next paragraph.

3.2 Reward Function

In order to fit the system's goal in our model, we add a reward function r evaluating the system's success in its task (moving towards the resource).

Let f be the periodic function standing for the resource quantity at each point. This function is assumed to be time invariant, strictly positive and normalized so that $\int_{\mathcal{C}} f(x)dx = 1$. Reward function r defined by

$$r : (x, u) \mapsto u(x)\, e^{1-\frac{u(x)}{f(x)}}$$

gives a local representation of the system's accomplishment in each length unit $[x, x + \delta[$ (Fig. 3).

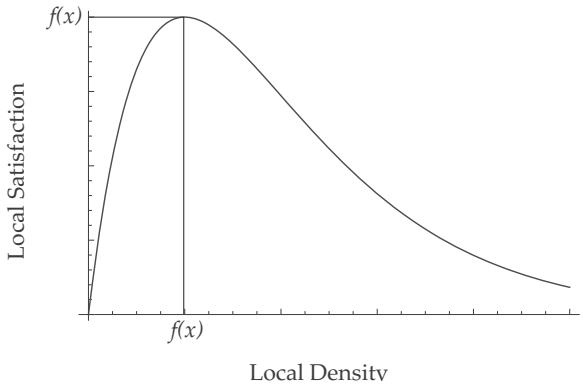


Fig. 3. Agents' reward function

The agents' presence at resources is increasingly rewarded, but overcrowding is punished exponentially. The system's global reward at time t is

$$R_\delta(t) = \sum_{x \in \mathcal{C}_\delta} r(x, u(x, t))\, \delta.$$

[2] Which is not limited to global dependence, since the probability $p^N_{\delta,\tau}(x, u)$ may depend only on the density of agents located at x.

Global reward R is bounded by the total amount of resources $\sum_{x\in\mathcal{C}_\delta} f(x)\ \delta$. Hence, the agents' optimal distribution is exactly the resource distribution.

3.3 Expression of the Addressed Problem

The objective is to find an agent behaviour (functions $p_{\delta,\tau}(x,u)$ and $q_{\delta,\tau}(x,u)$) yielding maximum global reward. In this paper we restrict the objective to some long-term reward $R(T)$ (with a large value for T)

$$\max_{p_{\delta,\tau},q_{\delta,\tau}} \mathbb{E}\left(R_\delta(T)\right),$$

where $\mathbb{E}$ stands for mathematical expectation over the possible system states.

It is a simple task to simulate the system for a given pair of functions $(p_{\delta,\tau}, q_{\delta,\tau})$, and measure its reward. Formal study is however a bit more tricky, since the Markov chain $\left((X_n(t))_{n=1}^N\right)_t$ (or even $(u^n(t))_t$) describing the system's evolution has a state space that increases quickly with the number of agents. To mitigate this difficulty, we suggest a continuous approximation of this discrete model in the next section.

4 Approximation of the Discrete Model

The derivation of the continuous approximation is done according to the two steps described in section 1, and ends up in a partial differential equation known as the *Fokker-Planck equation*. We provide some interesting properties of this equation in a special case.

4.1 Derivation of the Continuous Model

The discrete system defined at section 3 has three characteristic dimensions:

- the number of agents N,
- the spatial step δ,
- the time step τ.

In this section we will make these values tend to some extreme values in order to obtain a limit model.

Statistical Limit. At first, we let the number of agents N tend to infinity. The idea behind this limit is the following: as the number of agents grows larger, the influence of each individual on the density becomes insignificant. As a consequence statistical fluctuations caused by the randomness of the movements have no impact on the system, which tends to behave in a deterministic way. In other words, agents observing the densities are interacting with a macroscopic (and deterministic) quantity on which they have no influence. This approach is known as *mean field theory* [17], and formalized by the following result:

Theorem 1 (Statistical Limit). *Suppose that, when $N \to \infty$ the initial distributions $u^N(0)$ converge almost surely to some limit $u(0)$, and that the probabilities $p^N_{\delta,\tau}(x,u)$, $q^N_{\delta,\tau}(x,u)$ converge uniformly in u to continuous functions $u \mapsto p_{\delta,\tau}(x,u)$, $u \mapsto q_{\delta,\tau}(x,u)$. Then, for each t, the function $u^N(t)$ converges almost surely to the vector $u(t)$ recursively defined by $u(x,0) = u_0(x)$ and*

$$u(x,t+\tau) = u(x,t)(1 - p_{\delta,\tau}(x,u) - q_{\delta,\tau}(x,u)) + u(x-\delta,t)q_{\delta,\tau}(x-\delta,u) + u(x+\delta,t)p_{\delta,\tau}(x+\delta,u) \quad (1)$$

Relation (1) can be seen as a *balance equation*: agents located at x may

- come from the $x-\delta$, with probability $q_{\delta,\tau}(x-\delta,u)$,
- come from the $x+\delta$, with probability $p_{\delta,\tau}(x+\delta,u)$,
- have been immobile, with probability $1 - p_{\delta,\tau}(x,u) - q_{\delta,\tau}(x,u)$.

The macroscopic quantity $u(t)$, called *mean field*, is ruled by as many equations as positions (i.e. n_p) and may be difficult to study if n_p is large.This is why we carry out another approximation in the next section.

Spatiotemporal Limit. Consider a sufficiently smooth solution to (1), at least twice differentiable in t and once in x. A Taylor expansion of order 2 in δ, and of order 1 in t in (1) leaves, after simplification:

$$\tau\frac{\partial}{\partial t}u(x,t) + o(\tau) = -\delta\frac{\partial}{\partial x}\left((p_{\delta,\tau}(x) - q_{\delta,\tau}(x))u(x,t)\right) + \frac{\delta^2}{2}\frac{\partial^2}{\partial x^2}\left((p_{\delta,\tau}(x) + q_{\delta,\tau}(x))u(x,t)\right) + o(\delta^2)$$

If both members are divided by τ and if $\delta, \tau \to 0$, so that $\frac{\delta^2}{\tau}$ is bounded $(0 < a < \frac{\delta^2}{\tau} < b < \infty)$, and if the limits below exist

$$c(x,u) = \lim_{\delta,\tau}\left(\frac{p_{\delta,\tau}(x,u) - q_{\delta,\tau}(x,u)}{\tau}\delta\right) \quad (2)$$

$$d(x,u) = \lim_{\delta,\tau}\left(\frac{p_{\delta,\tau}(x,u) + q_{\delta,\tau}(x,u)}{2\tau}\delta^2\right), \quad (3)$$

we obtain the limit equation:

$$\frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left(c(x,u)u(x,t)\right) + \frac{\partial^2}{\partial x^2}\left(d(x,u)u(x,t)\right) \quad (4)$$

known as the *Fokker-Planck equation* [24]. Its *convection term* $-\frac{\partial}{\partial x}\left(c(x,u)u(x,t)\right)$ accounts for macroscopic displacement, whereas *the diffusion term* $\frac{\partial^2}{\partial x^2}\left(d(x,u)u(x,t)\right)$ accounts for dispersion.

In a similar way, the *convection coefficient* c may be understood as the mean speed of individuals located at position x whereas the *diffusion coefficient* d quantifies the local tendency to leave position x.

Equations (2), (3) are commonly called *scaling conditions*. They show how the system's endogenous parameters need to scale with its characteristic dimensions if we want (4) to be a valid approximation.

Formulation of the Continuous Problem. The discrete model defined at section 3.2 has been approximated by the partial differential equation (4), which is parameterized by the coefficients c and d. By means of the scaling equations (2), (3) we translate[3] the optimal choice of the agents probabilities $p_{\delta,\tau}$ and $q_{\delta,\tau}$ into a continuous problem on c and d:

$$\max_{c,d} \int_{\mathcal{C}} r\left(x, u(x,T)\right) dx$$

under the constraint given by equation (4). However it is not completely obvious that this formulation has really simplified our problem. Equation (4) may be difficult to solve according to the coefficients c and d, and may have solutions that have no physical sense[4]. The next section is devoted to a simple case where these difficulties do not appear.

4.2 Properties of the Linear Fokker-Planck Equation

A well-mastered special case of (4) is when the functions c and d depend only on x[5]. The corresponding Fokker-Planck equation is

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\left(c(x)u\right) + \frac{\partial^2}{\partial x^2}\left(d(x)u\right), \tag{5}$$

and is equipped with a boundary condition constraining solutions to be periodic. In order to show that equation (5) is a "good" representation of our system, we show that it has sufficiently regular solutions[6] that are density functions.

Theorem 2 (Well-Posedness of the Equation)

(i) Suppose the coefficients c, d and the initial distribution u_0 infinitely differentiable, and d bounded from below: $d(x) \geq \epsilon > 0$.
Then equation (5) has a unique solution that is infinitely differentiable.

(ii) If we suppose, in addition to the previous assumptions, that u_0 is positive and $\int_{\mathcal{C}} u_0(x)dx = 1$, then the solution of (5) is positive and

$$\int_{\mathcal{C}} u(x,t)dx = 1, \forall t \geq 0.$$

Dynamics of the solutions of (5) are resumed by the following two theorems:

[3] In a non-bijective way: multiple probabilities $p_{\delta,\tau}$, $q_{\delta,\tau}$ may result in the same c, d.
[4] These *weak solutions* are purely mathematical objects with no concrete interpretation.
[5] In a situation where the agents only observe their position x.
[6] In particular, bounded and without discontinuities.

Theorem 3 (Existence of a Unique Stationary Distribution). *The stationary Fokker-Planck equation*

$$-\frac{\partial}{\partial x}\left(c(x)u_\infty(x)\right) + \frac{\partial^2}{\partial x^2}\left(d(x)u_\infty(x)\right) = 0$$

has a unique periodic solution u_∞ so that $\int_{\mathcal{C}} u_\infty(x)dx = 1$, which is positive.

Thus, for each pair of coefficients c, d there is a unique stable distribution[7].

Theorem 4 (Convergence of Solutions). *Let u be a periodic solution of (5). There are two positive constants A and B such that*

$$\int_{\mathcal{C}} |u(x,t) - u_\infty(x)|dx \leq A\, e^{-Bt}$$

This provides a complete characterization of the solutions' dynamics: either the systems starts at the invariant distribution u_∞ and remains so indefinitely, or it starts at another distribution in which case it converges to u_∞ at exponential speed.

For now, we have made very few assumptions on the coefficients c and d. In the following section, we adjust them to maximize the system's performance.

5 Optimal Strategies

In this section, we establish two strategies with optimal long-term behaviour, with respect to the continuous formulation of the problem defined at section 3. Each of these strategies consists in a judicious choice of the coefficients c and d, and is interpreted in terms of agent behaviour.

5.1 A Resource-Dependent Strategy

We start off with the linear case (sect. 4.2). A simple way to maximize the system's global satisfaction is to identify the stationary solution u_∞ to the optimal distribution f. The results of the previous section then guarantee exponential convergence to the optimal distribution.

Replacing u_∞ by f in the stationary equation, we obtain

$$-\frac{d}{dx}\left(c(x)f(x)\right) + \frac{d^2}{dx^2}\left(d(x)f(x)\right) = 0.$$

This equation has an obvious solution:

$$c(x) = \frac{f'(x)}{f(x)} = \frac{d}{dx}\ln(f(x)),\ d(x) = 1,$$

where c is well defined, since f is assumed to be strictly positive.

[7] This does not imply that the agents are motionless, since movements may keep the global distribution unchanged.

This solution has an interesting meaning. Coefficients c and d correspond to a movement oriented towards the logarithmic gradient of resources. This fact can be related to the Weber-Fechner law [9] in psychophysics: *The intensity of a sensation varies as the logarithm of the stimulus*, translated in our case by: the agents move with an average speed proportional to the logarithmic derivative of their perception (the quantity of surrounding resources).

Even though this strategy is interesting, it has two flaws: agents only consider the amount of surrounding resources, and completely ignore each other, and local knowledge of the target distribution (the resource in this case) and its variations is required. This is why we suggest another strategy in the next section, based only on local and partial information.

5.2 A Strategy Based on Local Rewards

An interesting strategy, inspired by the previous one, is given by:

$$c(x,u) = \frac{\partial}{\partial x} \ln\big(r(x,u(x,t))\big),\ d(x,u) = 1$$

with the reward function $r(x,u) = u(x)e^{1-\frac{u(x)}{f(x)}}$ defined at section 3.2. This means the agents move towards locations where local satisfaction is high. In agreement with Weber-Fechner's law, the intensity of these movements is given by the logarithmic derivative of the perception (i.e. local reward).

The corresponding Fokker-Planck equation is nonlinear, and the results of section 4.2 are of no use. We can however simplify its expression:

Proposition 1 (Simplification). *The nonlinear Fokker-Planck equation*

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\left(u\,\frac{\partial}{\partial x}\ln\big(r(x,u(x,t))\big)\right) + \frac{\partial^2 u}{\partial x^2}$$

reduces to

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left[u\,\frac{\partial}{\partial x}\left(\frac{u}{f}\right)\right] \tag{6}$$

and show that this reduced equation has similar properties to the linear case. These properties are summed up by the following theorem:

Theorem 5. *Suppose that f is a strictly positive, periodic and regular function such that $\int_{\mathcal{C}} f(x)dx = 1$.*

(i) (Existence of regular solutions) If the initial condition u_0 is infinitely differentiable and strictly positive, equation (6) *has a solution that is infinitely differentiable and strictly positive.*

(ii) (Population size conservation) If the initial condition u_0 satisfies $\int_{\mathcal{C}} u_0(x)dx = 1$, then for any regular solution u of (6)*:*

$$\int_{\mathcal{C}} u(x,t)dx = 1, \forall t \geq 0$$

(iii) (Existence of a unique stationary solution) The stationary equation

$$\frac{\partial}{\partial x}\left[u_\infty \frac{\partial}{\partial x}\left(\frac{u_\infty}{f}\right)\right] = 0$$

has a unique positive periodic solution u_∞ so that $\int_{\mathcal{C}} u_\infty(x)dx = 1$ and this solution is f.

(iv) (Convergence of solutions) Let u be a regular, positive and periodic solution of equation (6) so that $\int_{\mathcal{C}} u(x,t)dx = 1$, $\forall t \geq 0$. There are two positive constants A and B such that

$$\int_{\mathcal{C}} |u(x,t) - f(x)|dx \leq A\, e^{-Bt}.$$

Assertions (i) and (ii) show that equation (6) provides a satisfactory representation of our system. Assertions (iii) and (iv) show that solutions of equation (6) have desired long-term behaviour: they converge to the target distribution f at exponential speed.

We stress the fact that this section's strategy is only based on local satisfaction (and its spatial variations) and, in contrast to section 5.1, it involves agent interaction (through the local reward function).

6 Numerical Validation

In this section we validate the above theoretical results by numeric simulations, highlighting two main aspects: validity of the continuous approximation (section 4.1) and long-term optimality of the strategies (section 5).

6.1 Validation of the Continuous Model

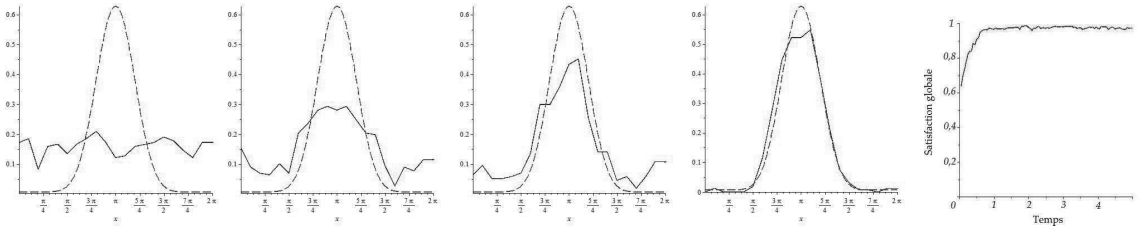
We simulate all three models (discrete, mean field, continuous) for a given set of parameters and characteristic dimensions, and focus on the resource-dependent case (section 5.1).

For all the following simulations, the **resource** is distributed according to the function f proportional to $\cos\left(\frac{x-\pi}{2}\right)^{10} + 0.01$ and normalized so that $\int_{\mathcal{C}} f(x)dx = 1$. It is displayed in hashed marks on every figure to show the objective.

The **discrete system** is shown on Fig. 4. The circle is evenly divided in $n_p = 20$ positions with spatial step $\delta = \frac{2\pi}{20} \simeq 0,3$. Time step τ is set to $\tau = 0,1$. The system contains $N = 500$ agents that are initialized according to independent uniform distributions on $\mathcal{C}_\delta$. Their movement probabilities are

$$p_{\delta,\tau}(x) = \frac{\tau}{2\delta^2}\left(2 + \frac{f(x+\delta) - f(x-\delta)}{2f(x)}\right)$$

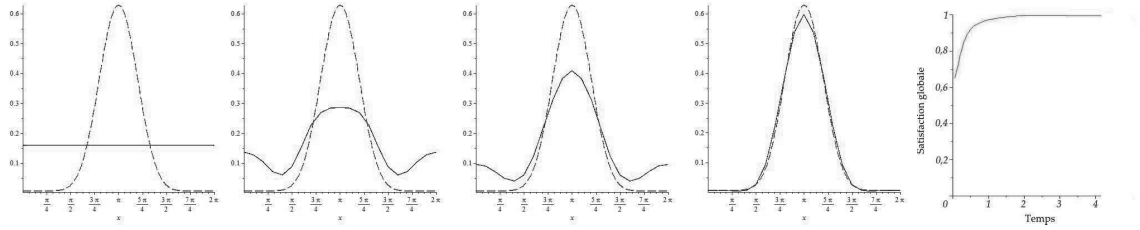
$$q_{\delta,\tau}(x) = \frac{\tau}{2\delta^2}\left(2 - \frac{f(x+\delta) - f(x-\delta)}{2f(x)}\right) \quad (7)$$



Discrete system at times $t = 0$, $t = 0.05$, $t = 0.5$ and $t = 2.5$ Global reward

Fig. 4. Dynamics of the discrete system

The **mean field**, shown on Fig. 5, is initialized according to the discrete uniform distribution $u_0 = (\frac{1}{20}, \ldots, \frac{1}{20})$ and evolves according to induction (1) in Theorem 1.



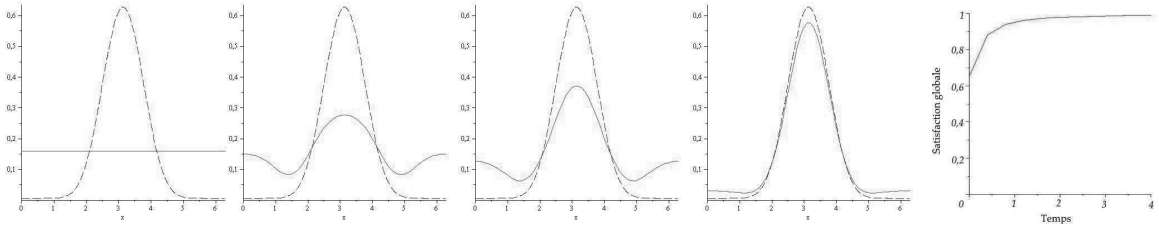
Mean field at times $t = 0$, $t = 0.05$, $t = 0.5$ and $t = 2.5$ Global reward

Fig. 5. Mean field dynamics

The **continuous system**, shown on figure 6, evolves according to the Fokker-Planck equation (4) with

$$c(x, u) = \frac{d}{dx} \ln (f(x)) \quad \text{and} \quad d(x, u) = 1$$

and is initialized as the uniform distribution on $\mathcal{C}$. Note that these coefficients c, d verify scaling conditions (2), (3) with the probabilities of the discrete system (7). We show numerical approximations obtained by the Crank-Nicholson method [8].



System state at times $t = 0$, $t = 0.05$, $t = 0.5$ and $t = 2.5$ Global reward

Fig. 6. Dynamics of the continuous system

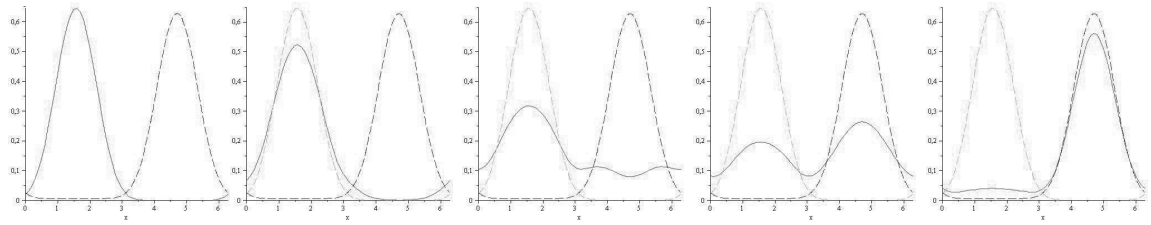
Figures 4, 5, 6 show all three systems have similar dynamics and rewards. This confirms the validity of our approximation.

6.2 Validation of the Optimal Strategies

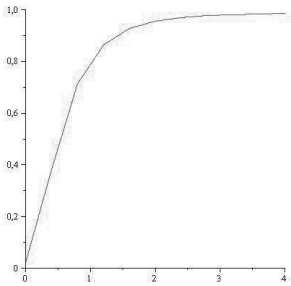
In order to show the optimality of the coefficients c and d found at section 5, we focus on the continuous system. The **resource** is distributed according to the function f defined at 6.1 and the **agents' initial distribution** u_0 is proportional to $\cos\left(\frac{x-3\pi}{2}\right)^{20}$ and normalized so that $\int_{\mathcal{C}} u_0(x)dx = 1$. The functions f and u_0 are shown on each simulation in hashed lines.

The **resource-based strategy**, studied at paragraph 5.1 and displayed on Fig. 7, parameterizes the agents' movements as

$$c(x) = \frac{d}{dx}\ln\left(f(x)\right), \quad d(x) = 1.$$



System state at times $t = 0$, $t = 0.1$, $t = 0.5$, $t = 1$ and $t = 3$

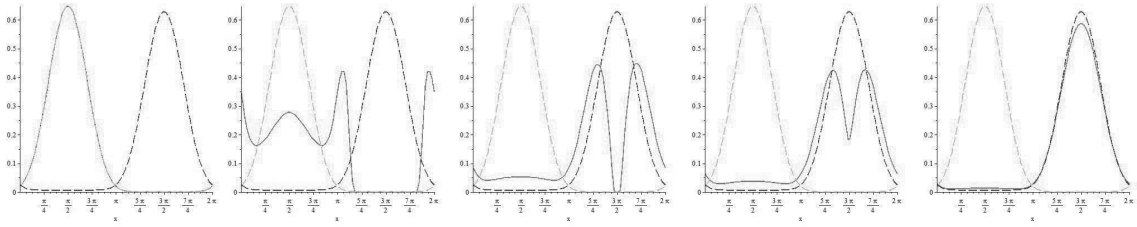


Global reward

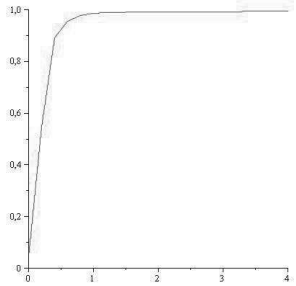
Fig. 7. Dynamics and reward of the system for the resource-based strategy

The **reward-based strategy**, established at part 5.2 and displayed on Fig. 8, corresponds to the coefficients

$$c(u,x) = \frac{\partial}{\partial x}\ln\big(r(x,u(x,t))\big),\ d(u,x) = 1.$$



System state at times $t = 0$, $t = 0.02$, $t = 0.2$, $t = 0.3$ and $t = 1$



Global reward

Fig. 8. Dynamics and reward of the system for the reward-based strategy

Figures 7, 8 show that both suggested strategies have the anticipated dynamics, and converge to the optimal distribution. This fact is highlighted by the optimal reward that increases to its maximal value 1. Furthermore the reward-based strategy appears to converge faster, even though this might not be true in general.

The density curves of figures 7, 8 show that the this strategy favours convection, in particular at early stages, whereas the first strategy is more diffusive.

7 Conclusion and Perspectives

In this article we suggest a method to model and parameterize a situated reactive system of agents that we wish to move towards a given goal distribution. A continuous approximation leads to two strategies with optimal long-term behaviour:

- the first is based on local knowledge of the target distribution, and requires global control,
- the second is autonomous, and relies on local and individual perception of the reward function.

Numerical simulations validate our approach from two points of view: correctness of the approximation (section 6.1) and long-term optimality of the strategies (section 6.2).

Though the setting is very simple, the reasoning we used is easily extended to other types of domains. In this case it is necessary to ensure well-posedness of the limit equations with appropriate boundary conditions [24]. It is also possible to consider dynamic resources that move in space, exhaust over time or resources

that are consumed by the agents. The explicit convergence bounds in theorems 4 and 7 allow a partial extension of our results to such situations if the resources evolve "slowly"[8].

Finally, another possible extension is to consider multiple agent populations, with mean field (density-dependent) interactions. In that case we obtain a system of coupled Fokker-Planck equations, with possible exchange terms between the populations. Examples found in literature [7,26,16], never involve objectives. Studying such systems with shared or diverging interests for each population is an interesting scope we plan to study in future work.

In regard to the optimal parameterizations, we limited this article to long-term global objectives. When facing cumulative rewards, it is necessary to study the limit of the corresponding MDP. The statistical limit towards some *mean field MDP* has been established by [10,5]. We expect the spatio-temporal limit to converge to a *mean field game* [12], whose connections to multi-agent systems have not yet been established.

References

1. Benaim, M., Le Boudec, J.Y.: A class of mean field interaction models for computer and communication systems. Performance Evaluation 65 (2008)
2. Benaouda, A., Zerhouni, N., Varnier, C.: Une approche multi-agents coopératifs pour la gestion des ressources matérielles dans un contexte multi-sites de e-manufacturing
3. Bernstein, D.S., Zilberstein, S., Immerman, N.: The complexity of decentralized control of markov decision processes. In: Proc. of the Sixteenth Conference on Uncertainty in Artificial Intelligence (2000)
4. Boes, J., Migeon, F., Gatto, F.: Self-Organizing Agents for an Adaptive Control of Heat Engines (short paper). In: International Conference on Informatics in Control, Automation and Robotics (ICINCO) (2013)
5. Bordenave, C., Anantharam, V.: Optimal control of interacting particle systems (June 2007), `http://hal.archives-ouvertes.fr/hal-00397327`
6. Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., Mylopoulos, J.: Tropos: An agent-oriented software development methodology. JAAMAS 8(3) (2004)
7. Chen, X., Hambrock, R., Lou, Y.: Evolution of conditional dispersal: a reaction–diffusion–advection model. Journal of Mathematical Biology 57 (2008)
8. Fadugba, S.E., Edogbanya, O.H., Zelibe, S.C.: Crank Nicolson method for solving parabolic PDEs. Int. Journal of Applied Math. and Modeling IJA2M (2013)
9. Fechner, G.: Elemente der Psychophysik. No. Bd. 1 in Elemente der Psychophysik, Breitkopf und Härtel (1860)
10. Gast, N., Gaujal, B.: A mean field approach for optimization in particle systems and applications. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (2009)
11. Gast, N., Gaujal, B., Le Boudec, J.Y.: Mean field for markov decision processes: from discrete to continuous optimization. IEEE Transactions on Automatic Control (2012)

[8] With slower rate than the agents' convergence speed.

12. Guéant, O., Lasry, J.-M., Lions, P.-L.: Mean field games and applications. In: Paris-Princeton Lectures on Mathematical Finance (2011)
13. Guestrin, C., Koller, D., Parr, R.: Multiagent planning with factored MDPs. In: Neural Information Processing Systems (NIPS) (2001)
14. Ijspeert, A.J., Martinoli, A., Billard, A., Gambardella, L.M.: Collaboration through the exploitation of local interactions in autonomous collective robotics: The stick pulling experiment. Autonomous Robots 11 (2001)
15. Jorquera, T., Georgé, J.P., Gleizes, M.P., Régis, C.: A Natural Formalism and a MultiAgent Algorithm for Integrative Multidisciplinary Design Optimization. In: International Conference on Intelligent Agent Technology (IAT) (2013)
16. Jüngel, A.: Diffusive and nondiffusive population models. In: Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences (2010)
17. Le Boudec, J.Y., McDonald, D., Mundinger, J.: A generic mean field convergence result for systems of interacting objects. In: Quantitative Evaluation of Systems (QEST). IEEE (2007)
18. Lepagnot, J., Nakib, A., Oulhadj, H., Siarry, P.: A new multiagent algorithm for dynamic continuous optimization. International Journal of Applied Metaheuristic Computing (IJAMC) (2010)
19. Lerman, K., Galstyan, A.: A general methodology for mathematical analysis of multi-agent systems. ISI-TR-529, USC Information Sciences Institute, Marina del Rey, CA (2001)
20. Liggett, T.M.: Particle Systems. Springer (1985)
21. Liu, J., Jing, H., Tang, Y.: Multi-agent oriented constraint satisfaction. Artificial Intelligence (2002)
22. Munkres, J.R.: Topology: a first course, vol. 23. Prentice-Hall, Englewood Cliffs (1975)
23. Puterman, M.L.: Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons (2009)
24. Risken, H.: Fokker-Planck Equation. Springer (1984)
25. Rougemaille, S., Arcangeli, J.-P., Gleizes, M.-P., Migeon, F.: ADELFE design, AMAS-ML in action. In: Artikis, A., Picard, G., Vercouter, L. (eds.) ESAW 2008. LNCS, vol. 5485, pp. 105–120. Springer, Heidelberg (2009)
26. Shigesada, N., Kawasaki, K., Teramoto, E.: Spatial segregation of interacting species. Journal of Theoretical Biology 79 (1979)
27. Shoham, Y., Leyton-Brown, K.: Multiagent systems: Algorithmic, game-theoretic, and logical foundations. Cambridge University Press (2009)
28. Spaan, M.T., Melo, F.S.: Interaction-driven markov games for decentralized multiagent planning under uncertainty. In: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems, vol. 1 (2008)
29. Spicher, A., Fatès, N.A., Simonin, O., et al.: From reactive multi-agent models to cellular automata-illustration on a diffusion-limited aggregation model. In: Proceedings of ICAART 2009 (2009)

Appendix: References to Mathematical Proofs

Due to spatial constraints, this article cannot contain the proofs of the mentioned results. A document containing these proofs is provided at `ftp://ftp.irit.fr/IRIT/SMAC/DOCUMENTS/PUBLIS/PRIMA2014_Stuker_Proofs.pdf`.