A Refined Mean Field Approximation of Synchronous Discrete-Time Population Models

Nicolas Gast[a,*], Diego Latella[b], Mieke Massink[b]

[a] *Inria, Univ. Grenoble Alpes, F-38000 Grenoble, France*
[b] *CNR-ISTI, I-56124 Pisa, Italy*

Abstract

Mean field approximation is a popular method to study the behaviour of stochastic models composed of a large number of interacting objects. When the objects are asynchronous, the mean field approximation of a population model can be expressed as an ordinary differential equation. When the objects are (clock-) synchronous the mean field approximation is a discrete time dynamical system. We focus on the latter.

We study the accuracy of mean field approximation when this approximation is a discrete-time dynamical system. We extend a result that was shown for the continuous time case and we prove that expected performance indicators estimated by mean field approximation are $O(1/N)$-accurate. We provide simple expressions to effectively compute the asymptotic error of mean field approximation, for finite time-horizon and steady-state, and we use this computed error to propose what we call a *refined* mean field approximation. We show, by using a few numerical examples, that this technique improves the quality of approximation compared to the classical mean field approximation, especially for relatively small population sizes.

Keywords: Mean field approximation; Discrete time population models; Accuracy of approximation.

1. Introduction

Stochastic models are often used to model and analyse the performance of computer (and many other) systems. A particularly rich and popular class of models is given by stochastic population models. These have been used, for instance, to model biological systems [1], epidemic spreading [2] or queuing networks [3]. These systems are composed of a set of homogeneous objects interacting with one another. These models have a high expressive power, but an exact analysis of any such a model is often computationally prohibitive when

*Corresponding author
Email address: nicolas.gast@inria.fr (Nicolas Gast)

the number of objects of the system grows. This results in the need for approximation techniques.

A popular technique is to use mean field approximation. The idea behind mean field approximation is to replace the study of the original stochastic system by the one of a, much simpler, deterministic dynamical system. The success of mean field approximation can be explained by multiple factors : (a) it is fast – many models can be solved in closed form [3, 4, 5, 6] or easily solved numerically [7, 8, 9] – (b) it is proven to be asymptotically optimal as the number of objects in the system goes to infinity [10, 11, 12, 13, 14]; and (c) it is often very accurate also for systems of moderate size, composed of $N \approx 100$ objects.

The mean field approximation of a given model is constructed by considering the limit of the original stochastic model as the number of objects N goes to infinity. There can be two types of limits. The first type arises when the dynamics of the objects are asynchronous. In this case the mean field approximation is given by a continuous time dynamical system (often a system of ordinary differential equations) – this is the most studied case *e.g.* [10, 12, 14]. The second type arises when the objects are synchronous. In this case the mean field approximation is a discrete time dynamical system [11, 15, 16]. We focus on the latter.

Contributions. Our main contribution is an extension to (synchronous) DTMC population models of the results proposed in [17] for (asynchronous) CTMC population models, thus providing a new approximation technique that is significantly more accurate than classical mean field approximation, especially for relatively small systems. Our results apply to the classical model of [11, 15, 18]. We prove our result for the transient and the steady state dynamics. Moreover, it retains an interesting feature of mean field approximation by being computationally non-intensive.

More precisely, if $M_i^{(N)}(t)$ denotes the proportion of objects in a state i at time t, then the classical result of [11] states that, as N grows large, if the vector $M^{(N)}(0)$ converges almost surely to m, for some vector m, then the vector $M^{(N)}(t)$ converges almost surely to a deterministic quantity $\mu(t)$ that satisfies a recurrence equation of the form $\mu(t+1) = \mu(t)\mathbf{K}(\mu(t))$ with $\mu(0) = m$. We show that, for any twice differentiable function h, there exists a constant $V_{t,h}$ such that

$$\lim_{N\to\infty} N(\mathbb{E}\left[h(M_t^{(N)})\right] - h(\mu(t))) = V_{t,h}. \tag{1}$$

We provide an algorithm to compute the constant $V_{t,h}$ by a linear dynamical system that involves the first and second derivative of the functions $m \mapsto m\mathbf{K}(m)$ and h. We also show that if the function $m \mapsto m\mathbf{K}(m)$ has a unique fixed point $\mu(\infty)$ that is globally exponentially stable, then the same result holds for the steady-state : in this case, $V_{\infty,h} = \lim_{t\to\infty} V_{t,h}$ exists and can be expressed as the solution of a discrete-time Lyapunov equation that involves the first and second derivative of $m \mapsto m\mathbf{K}(m)$ and h evaluated at the point $\mu(\infty)$.

By using these results, we define a quantity $h(\mu(t)) + V_{t,h}/N$ that we call the *refined mean field* approximation. As opposed to the classical mean field

approximation, this approximation depends on the system size N. We illustrate our theoretical results with four different examples. While these examples all show that our refined model is clearly more accurate than the classical approximations, they illustrate different characteristics. The first two examples are cases where the dynamical system has a unique exponentially stable attractor. In these examples, refined mean field provides performance estimates that are extremely accurate (the typical error between $M^{(N)}$ and μ is less than 1% for $N = 10$). The third example is different as it is a case when the stochastic system has two absorbing states. In this case, the refined mean field is still more accurate than the classical mean field approximation but remains far from the exact values for $N = 10$. It is only for larger values of N that the refined mean field provides a very accurate estimate. Finally, the fourth example is a case where the mean field approximation has a unique attractor that is not exponentially stable. We observe that in this case the refined approximation provides an accurate approximation of $\mathbb{E}[M^{(N)}(t)]$ for small values of t but fails to predict correctly what happens when t is large compared to N. In fact in this case, one cannot refine the steady-state expectation by a term in $O(1/N)$ because the convergence is only in $O(1/\sqrt{N})$ in this case.

This suggests that, when using a mean field or refined mean field approximation, one has to be careful : the approximations of a system with more than one stable equilibrium or a unique but non-exponentially stable equilibrium is likely to be inaccurate for small values of N, even when one focuses on the transient behaviour.

Related work. Our results extend the recent results of [17]. The authors of [17] study the steady-state of stochastic models that have a continuous-time mean field approximation. They show that Equation (1) is true in this case and provide a numerical algorithm to compute the constant. Our paper has two theoretical contributions with respect to [17] : First we show that the results also hold for models that have a discrete-time mean field approximation, and second we show how to derive these equations for the transient and steady state regimes of such systems. This means that our results remain in the realm of discrete-time models whereas in [17] it is shown how some discrete-time models can be transformed into density-dependent (continuous time) population models by replacing time steps with steps that last for a random time that is exponentially distributed with mean $1/N$, where N is the population size. The resulting continuous time model can then be analysed using the approximation techniques for CTMC population models discussed in [17].

The results of [17] and the one of the current paper follow from a series of recent results concerning the rate of convergence of stochastic models to their mean field approximation [19, 20, 21, 22]. The key idea behind these works is to study the convergence of the generator of the stochastic processes to the one of its mean field approximation and to use this convergence rate to obtain a bound on Equation (1). For the steady-state regime, this is made possible by using Stein's methods [23, 24, 25]. Note that the approach taken in the current paper is fundamentally different from the one that is usually used to

obtain convergence rates, like [13, 26, 15] in which the authors focus on sample path convergence and obtain bounds on the convergence of the expected distance $\mathbb{E}[\|M^{(N)} - \mu\|]$ between the stochastic system and its mean field approximation. When focusing on sample path convergence, the refinement of the mean field approximation would be to consider an additive term of $1/\sqrt{N}$ times a Gaussian noise as for example in [13] and not a $1/N$ term as in this paper.

This paper and the simulations it contained are fully reproducible : `https://github.com/ngast/RefinedMeanField_SynchronousPopulation`.

Outline. The rest of the paper is organised as follows. In Section 2, we introduce the model that we study. In Section 3 we provide the main results and in particular Theorem 1. In Sections 4, 5, 6 and 7, we provide a few numerical examples that demonstrate the accuracy of the refined mean field approximation and its limits. Finally, we conclude in Section 8.

2. Preliminaries

In this section we introduce some terminology and notation as well as some preliminary definitions, setting the context for the rest of the paper.

2.1. Notations

We let $\mathbb{N}$ denote the set of natural numbers and $\mathbb{R}^n_{\geq 0}$ the set of n-tuples of non-negative real numbers; we conventionally see any such an n-tuple $m = (m_1, \ldots, m_n)$ as a row-vector, i.e. a $1 \times n$ matrix. We let $\mathcal{U}^n \subset \mathbb{R}^n_{\geq 0}$ be the unit simplex of $\mathbb{R}^n_{\geq 0}$, that is $\mathcal{U}^n = \{m \in [0,1]^n \mid m_1 + \ldots + m_n = 1\}$.

For function $f : \mathbb{R}^n \to \mathbb{R}^p$ continuous and twice differentiable, with $f(m) = (f_1(m), \ldots, f_p(m))$, we denote by Df and $D^2 f$ its first and second derivatives, respectively. $Df(m)$ is the $p \times n$ (function) matrix such that $(Df(m))_{ij} = \frac{\partial f_i(m)}{\partial m_j}$. $D^2 f(m)$ is the $p \times n \times n$ tensor such that $(D^2 f(m))_{ijk} = \frac{\partial^2 f_i(m)}{\partial m_j \partial m_k}$.

Moreover, for a $p \times n \times n$ tensor P and a $n \times n$ matrix Q, we let $P \cdot Q$ be the (row) vector in $\mathbb{R}^p$ such that $(P \cdot Q)_i = \sum_{j,k=1}^n (P)_{ijk}(Q)_{jk}$, for $i = 1, \ldots, p$. In addition, for $p \times n$ matrix A and $n \times m$ matrix B we use the notation AB for standard matrix product: $(AB)_{ij} = \sum_{k=1}^n (A)_{ik}(B)_{kj}$; obviously this includes the case of vector inner product $u^T v$ for u and v being $n \times 1$ (column) vectors, where A^T is the transpose of A, i.e. $(A^T)_{ij} = (A)_{ji}$; the vector outer product $u \otimes v$, sometimes also denoted by $u\, v^T$, is the matrix such that $(u \otimes v)_{i,j} = u_i v_j$.

Finally, for any vector v, we let $\|v\|$ denote the norm of v and $\|v\|^2$ denote the square of $\|v\|$. Note that the choice of the specific norm is left unspecified because the results presented in the present paper hold for any norm[1]. Note that in the proofs, what we denote by $E, o(\|E\|^2), \mathbb{E}[o(\|E\|^2)]$ are p-dimensional vectors while their norm $\|E\|, \|E\|^2, \|\mathbb{E}[o(\|E\|^2)]\| \in \mathbb{R}$ are real non-negative numbers.

[1] Of course, technical details in the proofs may depend on the specific norm (see e.g. footnote 4 in the proof of Lemma 4)

2.2. Synchronous Mean Field Model

We consider a system of $0 < N \in \mathbb{N}$ identical[2] interacting objects; N is called the *size* of the system. We assume that the number of local states of each object is finite[3], say n; for the sake of simplicity, in the sequel we let the set of states of local objects be $\{1, \ldots, n\}$, when not specified otherwise. Time is *discrete* and the behaviour of the system is characterised by a (time homogeneous) discrete time Markov chain (DTMC) $X^{(N)}(t) = (X_1^{(N)}(t), \ldots, X_N^{(N)}(t))$, where $X_i^{(N)}(t)$ is the state of object i at time t, for $i = 1, \ldots, N$.

We define the occupancy measure at time t as the row-vector $M^{(N)}(t) = (M_1^{(N)}(t), \ldots, M_n^{(N)}(t))$ where, for $j = 1, \ldots, n$, $M_j^{(N)}(t)$ is the *fraction* of objects in state j at time t, over the total population of N objects:

$$M_j^{(N)}(t) = \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_i^{(N)}(t)=j\}}$$

where $1_{\{x=j\}}$ is equal to 1 if $x = j$ and 0 otherwise.

At each time step $t \in \mathbb{N}$ each object performs a local transition, that may also be a self-loop to its current state. The transition probabilities of an object state depend on the current local state of the object and may depend also on $M(t)$.

We let $\mathbf{K}(m)$ denote the one-step transition probability $n \times n$ matrix of an object in the system: $\mathbf{K}_{ij}(m)$ is the probability for the object to jump from state i to state j in the system when the occupancy measure vector is m. We assume that, given the occupancy measure, the transitions made by two objects are independent. Our model is identical to the one of [11] up to the fact that the authors of [11] add a continuous resource to the model and allow object transition matrix $\mathbf{K}$ to depend also on the size N of the system (in which case they assume that the sequence of transition matrices $\mathbf{K}^N$ converges to a function $\mathbf{K}$ as N goes to infinity). To simplify the exposition, in this paper we consider a case without resource and we assume $\mathbf{K}(m)$ is a continuous function of m that does not depend on N. The results presented in this paper could be extended to the more general case where $\mathbf{K}(m)$ is also a function of N and could be modified to incorporate a resource, essentially by replacing our equation for the variance Γ by the one presented in [15, Equation (7)].

Below we recall Theorem 4.1 of [11] on classical mean field approximation, under the simplifying assumptions mentioned above:

[2]It is worth pointing out here that the requirement that the N objects be identical can be relaxed, since a system with *different* classes of *identical* objects can easily be modelled by considering an equivalent system with instances of an object whose set of states is the union of those of the original objects and similarly for the set of its transitions, as shown in the example in Section 5.

[3]In fact, the same theoretical results could be derived for infinite dimensional models with two additional assumptions to cope with the fact that the simplex $\mathcal{U}^\infty$ is not compact : (i) imposing all functions to be *uniformly* continuous and (ii) Imposing tightness assumptions.

Theorem 4.1 of [11] (Convergence to Mean Field) *Assume that the initial occupancy measure $M^{(N)}(0)$ converges almost surely to the deterministic limit $\mu(0)$. Define $\mu(t)$ iteratively by (for $t \geq 0$):*

$$\mu(t+1) = \mu(t)\,\mathbf{K}(\mu(t)). \tag{2}$$

Then for any fixed time t, almost surely:

$$\lim_{N\to\infty} M^{(N)}(t) = \mu(t).$$

In the sequel, we will write $M(t)$ or simply M instead of $M^{(N)}(t)$, leaving N and t implicit, when this does not cause confusion.

3. Refined Deterministic Approximation Theorem

In this section, we present our main results (Theorem 1 and Theorem 2) and provide their proofs.

3.1. First Main Result : Transient Behaviour

The iterative procedure of Theorem 4.1 of [11] can be formalised as a t-indexed family of functions Φ_t from $\mathcal{U}^n$ to $\mathcal{U}^n$, where the functions $\Phi_t : \mathcal{U}^n \to \mathcal{U}^n$ are defined as follows:

$$\Phi_0(m) = m; \qquad (\Phi_1(m))_j = \sum_{i=1}^{n} m_i \mathbf{K}_{ij}(m); \qquad \Phi_{t+1}(m) = \Phi_1(\Phi_t(m))$$

This implies that for all $m \in \mathcal{U}^n$ and $t \in \mathbb{N}$, we have $\Phi_1(\Phi_t(m)) = \Phi_t(\Phi_1(m))$.

In the following we assume that Φ_1 is continuous and twice differentiable with respect to m and that its second derivative is continuous (note that as $\mathcal{U}^n$ is compact, this implies that Φ_1 and its first two derivative are uniformly continuous). Moreover, in what follows, we assume that $M^{(N)}(0)$ converges to $\mu(0)$ (a deterministic value) as N goes to infinity and we let $\mu(t)$ be defined as in Equation (2), or, equivalently, $\mu(t+1) = \Phi_1(\mu(t)) = \Phi_{t+1}(\mu(0))$.

Our main theorem can be stated as follows.

Theorem 1. *Assume that the function Φ_1 is twice differentiable with continuous second derivative and that $M^{(N)}(0)$ converges weakly to $\mu(0)$. Let A_t and B_t be respectively the $n \times n$ matrix $A_t = (D\Phi_1)(\mu(t))$ and the $n \times n \times n$ tensor $B_t = (D^2\Phi_1)(\mu(t))$. Then for any continuous and twice differentiable function with continuous second derivative $h : \mathcal{U}^n \to \mathbb{R}^p_{\geq 0}$ we have:*

$$\lim_{N\to\infty} N\mathbb{E}\left[h(M^{(N)}(t)) - h(\Phi_t(M^{(N)}(0)))\right] = Dh(\mu(t))V_t + \frac{1}{2}D^2h(\mu(t)) \cdot W_t,$$

where V_t is an $n \times 1$ vector and W_t is an $n \times n$ matrix, defined as follows:

$$\begin{aligned} V_{t+1} &= A_t V_t + \tfrac{1}{2} B_t \cdot W_t \\ W_{t+1} &= \Gamma(\mu(t)) + A_t W_t A_t^T, \end{aligned}$$

with $V_0 = 0$, $W_0 = 0$ and $\Gamma(m)$ is the following $n \times n$ matrix:

$$\begin{aligned}\Gamma_{jj}(m) &= \sum_{i=1}^{n} m_i \mathbf{K}_{ij}(m)(1 - \mathbf{K}_{ij}(m)) \\ \Gamma_{jk}(m) &= -\sum_{i=1}^{n} m_i \mathbf{K}_{ij}(m)\mathbf{K}_{ik}(m)\end{aligned}$$

The key idea of the proof is to use a Taylor expansion of $\mathbb{E}[h(\Phi_1(m))]$ around $\Phi_1(m)$. We postpone the proof to Section 3.3.1.

One of the main consequences of Theorem 1 is that it allows us to compute precisely a development in $O(1/N)$ of the mean and the covariance of the vector $M^{(N)}(t)$. This first order development is what we call the *refined mean field approximation*. In our numerical simulations, we will show that this refined approximation can greatly improve the accuracy of the original mean field approximation when the number of entities N is relatively small.

Corollary 1. *Let $t \in \mathbb{N}$. Under the assumptions of Theorem 1, and denoting $\mu(t) = \Phi_t(m)$, it holds that*

(i) For any coordinate i and any time-step

$$\mathbb{E}\left[M_i^{(N)}(t)\right] = \mu_i(t) + \frac{(V_t)_i}{N} + o\left(\frac{1}{N}\right).$$

(ii) For any pair of coordinates i, j, the co-variance satisfies

$$\operatorname{cov}\left(M_i^{(N)}(t), M_j^{(N)}(t)\right) = \frac{1}{N}(W_t)_{i,j} + o\left(\frac{1}{N}\right).$$

3.2. Second Main Result : Steady-State

Mean field approximation can also be used to characterise the steady-state behaviour of a population model. It has been shown that, in the case of continuous time or discrete-time mean field approximation, if this approximation has a unique attractor, then the stationary distribution of the system of size N concentrates on this attractor. In this section, we refine this result by computing the rate of convergence and by defining the refined approximation in this case.

We say that a point $\mu(\infty)$ is an exponentially stable attractor if

- For any $m \in \mathcal{U}^n$: $\lim_{t\to\infty} \Phi_t(m) = \mu(\infty)$ (*i.e.* it is a global attractor).
- There exists an open neighbourhood V of $\mu(\infty)$ and two constants a, b such that all $m \in V$: $\|\Phi_t(m) - \mu(\infty)\| \leq ae^{-bt}\|m - \mu(\infty)\|$ (*i.e.* it is exponentially stable).

Theorem 2. *Assume that $M^{(N)}$ has a unique stationary distribution (for each N), that the function Φ_1 is twice differentiable and that the flow has a unique exponentially stable attractor $\mu(\infty)$. Then there exists a $n \times 1$ vector V_∞ and a $n \times n$ matrix W_∞ such that the constants V_t and W_t defined in Theorem 1 satisfy:*

$$\lim_{t\to\infty} V_t = V_\infty \qquad \text{and} \qquad \lim_{t\to\infty} W_t = W_\infty$$

Moreover

(i) W_∞ *is the unique solution of the discrete-time Lyapunov equation:*

$$A_\infty W A_\infty^T - W + \Gamma(\mu(\infty)) = 0$$

and V_∞ *is uniquely determined by*

$$V_\infty = \frac{1}{2}(I - A_\infty)^{-1} B_\infty W_\infty,$$

where $A_\infty = D\Phi_1(\mu(\infty))$, $B_\infty = D^2\Phi_1(\mu(\infty))$ *and* I *is the identity matrix.*

(ii) for any twice differentiable function h*, we can exchange the limits :*

$$\begin{aligned}\lim_{N\to\infty}\lim_{t\to\infty} N\big(\mathbb{E}[h(M(t))] - h(\Phi_t(M^{(N)}(0)))\big) \\ = \lim_{t\to\infty}\lim_{N\to\infty} N\big(\mathbb{E}[h(M(t))] - h(\Phi_t(M^{(N)}(0)))\big) \\ = Dh(\mu(\infty))V_\infty + \frac{1}{2}D^2h(\mu(\infty))\cdot W_\infty,\end{aligned}$$

This result is interesting for at least two reasons. First, it generalises Theorem 1 to the case of stationary distribution. Second, it is also the first result that provides a rate of convergence for the steady-state distribution of a model that has a discrete-time mean field approximation (to the best of our knowledge, the rate of convergence had only been obtained for the finite-time horizon in [15]). We postpone the proof to Section 3.3.2.

3.3. Proofs

To ease notation, in all the proofs we denote $M^N(0)$ by m, unless specified otherwise. In particular, when we write $\mathbb{E}[h(M(t))]$ we formally mean $\mathbb{E}[h(M^{(N)}(t))|M^{(N)}(0) = m]$.

3.3.1. Proof of Theorem 1

One of the key ingredients to prove our result is to study what happens for $t = 1$. This is what we do in Lemma 1. Then Theorem 1 will follow by using an induction on t. Note that one of the main technicalities in all these lemmas is to prove that the convergence is uniform in the initial condition. This is why we make use of various functions $\varepsilon(N)$ or $\varepsilon_{t,g}(N)$ that control the *uniform* convergence to 0.

Lemma 1. *Let* $h : \mathcal{U}^n \to \mathbb{R}^p_{\geq 0}$ *(for* $p \geq 1$*), be a twice differentiable function such that* D^2h *is continuous. Then, there exists a function* $\varepsilon(N)$ *such that* $\lim_{N\to\infty}\varepsilon(N) = 0$ *and for all* $M^{(N)}(0) = m \in \mathcal{U}^n$*, the following holds:*

$$\left|\left|N\big(\mathbb{E}[h(M(1))] - h(\Phi_1(m))\big) - \frac{1}{2}(D^2h)(\Phi_1(m))\cdot\Gamma(m)\right|\right| \leq \varepsilon(N).$$

where $\Gamma(m)$ *is defined in Theorem 1.*

PROOF. The key idea of this proof is to consider the Taylor expansion of h for $M(1)$ in the neighbourhood of $\Phi_1(m)$. Let $E = M(1) - \Phi_1(m)$. We get the following :

$$h(M(1)) - h(\Phi_1(m)) = (Dh)(\Phi_1(m))E + \frac{1}{2}(D^2h)(\Phi_1(m)) \cdot (E \otimes E) + o(||E||^2)$$

Taking the expectation on both sides, we get :

$$\begin{aligned}&\mathbb{E}[h(M(1))] - h(\Phi_1(m)) \\ &= (Dh)(\Phi_1(m))\mathbb{E}[E] + \frac{1}{2}(D^2h)(\Phi_1(m)) \cdot \mathbb{E}[(E \otimes E)] + \mathbb{E}[o(||E||^2)]\end{aligned}$$

The result follows by using Lemma 3, that establishes that $\mathbb{E}[E] = 0$ and $N\mathbb{E}[(E \otimes E)] = \Gamma(m)$ and Lemma 4 that shows $||N\mathbb{E}[o(||E||^2)]|| \leq \varepsilon(N)$.

The following lemma is a direct generalisation of Lemma 1 and it will be used in the proof of Theorem 1. The proof is essentially the same as that of Lemma 1 and exploits the time-homogeneity of the Markov chain.

Lemma 2. *Let $h : \mathcal{U}^n \to \mathbb{R}^p_{\geq 0}$ be a twice differentiable function whose second derivative is continuous. Then there exists a function $\varepsilon(N)$ such that $\lim_{N\to\infty} \varepsilon(N) = 0$ and for all $t \in \mathbb{N}$, $N \in \mathbb{N}_{>0}$ and $M^{(N)}(t) = m' \in \mathcal{U}^n$, the following holds:*

$$\left|\left|N\big(\mathbb{E}[h(M(t+1))|M(t) = m'] - h(\Phi_1(m'))\big) - \frac{1}{2}(D^2h)(\Phi_1(m')) \cdot \Gamma(m')\right|\right| \leq \varepsilon(N)$$

Lemma 3. *Under the assumptions of Lemma 1, we obtain that $E = \mathbb{E}[M^{(N)}(1) - \Phi_1(m) \mid M^{(N)}(0) = m]$ satisfies $\mathbb{E}[E] = 0$ and $\mathbb{E}[E \otimes E] = \Gamma(m)/N$.*

PROOF. We observe that by definition of our model, $M_j(1)$ is the following random variable:

$$M_j(1) = \frac{1}{N}\sum_{i=1}^{n} \widehat{B}_{ij} \tag{3}$$

where $(\widehat{B}_{i,.})$ is a random vector with multinomial distribution, with parameters Nm_i and $(\mathbf{K}_{i,.}(m))$. The variables are independent for different values of i (in particular, if $i \neq i'$ we have $\operatorname{cov}\left(\widehat{B}_{ij}, \widehat{B}_{i'k}\right) = 0$). Moreover, for all i and all $j \neq k$:

$$\begin{aligned}\mathbb{E}\left[\widehat{B}_{ij}\right] &= Nm_i\mathbf{K}_{ij}(m) \\ \operatorname{var}\left(\widehat{B}_{ij}\right) &= Nm_i\mathbf{K}_{ij}(m)(1 - \mathbf{K}_{ij}(m)) \\ \operatorname{cov}\left(\widehat{B}_{ij}, \widehat{B}_{ik}\right) &= Nm_i\mathbf{K}_{ij}(m)\mathbf{K}_{ik}(m).\end{aligned}$$

This implies that

$$\mathbb{E}[M_j(1)] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{n}\widehat{B}_{ij}\right] = \frac{1}{N}\sum_{i=1}^{n}\mathbb{E}\left[\widehat{B}_{ij}\right] = \Phi_1(m)_j. \tag{4}$$

The case of $E \otimes E$ makes again use of Equation (3). Note that by (4), $E_j = M_j(1) - \Phi_1(m)_j = M_j(1) - \mathbb{E}[M_j(1)]$. This shows that $\mathbb{E}[(E \otimes E)_{jk}] = \mathrm{cov}\,(M_j(1), M_k(1))$, *i.e.* the covariance of $M_j(1)$ and $M_k(1)$. We consider the case $k = j$ and the case $k \neq j$ separately.

Case $k = j$.

$$\begin{aligned} N\mathbb{E}\left[(E \otimes E)_{jj}\right] &= N\mathrm{var}\left(M_j(1)\right) \\ &= N\mathrm{var}\left(\frac{1}{N}\sum_{i=1}^{n}\widehat{B}_{ij}\right) \\ &= \frac{1}{N}\sum_{i=1}^{n}\mathrm{var}\left(\widehat{B}_{ij}\right) \\ &= \sum_{i=1}^{n} m_i \mathbf{K}_{ij}(m)(1 - \mathbf{K}_{ij}(m)), \end{aligned}$$

where the one but last equality comes from the independence of the variables $\widehat{B}_{ij}$ for $i \in \{1 \dots n\}$.

Case $k \neq j$. This case is similar.

$$\begin{aligned} N\mathbb{E}\left[(E \otimes E)_{jk}\right] &= N\mathrm{cov}\left(M_j(1), M_k(1)\right) \\ &= N\sum_{i=1}^{n}\sum_{i'=1}^{n}\frac{1}{N^2}\mathrm{cov}\left(\widehat{B}_{ij}, \widehat{B}_{i'k}\right) \\ &= -\sum_{i=1}^{n} m_i \mathbf{K}_{ij}(m)\mathbf{K}_{ik}(m), \end{aligned}$$

where in the double sum, only the terms $i = i'$ are non-zero because $\widehat{B}_{ij}$ and $\widehat{B}_{i'k}$ are independent when $i \neq i'$.

Lemma 4. *Under the assumptions of Lemma 1 and using the notations of Lemma 1, there exists a function* $\varepsilon(N)$ *such that* $\lim_{N\to\infty}\varepsilon(N) = 0$ *and that* $||N\mathbb{E}[o(||E||^2)]|| \leq \varepsilon(N)$.

Proof. First of all we note that, as D^2h is continuous, it is uniformly continuous (because $\mathcal{U}$ is compact). Hence, the term $o(||E||^2) \in \mathbb{R}^p$ is uniform in m and $||E||^2$, *i.e.*, there exist a function $\delta : \mathbb{R}^n \to \mathbb{R}$ and a constant $\gamma > 0$ such that $\lim_{||e||\to 0}\delta(e) = 0$, $\delta(e) \leq \gamma$ for all $e \in \mathbb{R}^n$, and $||o(||E||^2)|| \leq ||E||^2\delta(E)$. This implies $\mathbb{E}[||o(||E||^2)||] \leq \mathbb{E}[||E||^2\delta(E)]$. The proof proceeds with the following derivation:

First, note that $\lim_{||e||\to 0}\delta(e) = 0$ implies that for all $\epsilon > 0$, there exists $a_\epsilon > 0$ such that $\delta(e) \le \epsilon$ for all e such that $||e|| \le a_\epsilon$. Therefore

$$\begin{aligned}\mathbb{E}[||E||^2\delta(E)] &\le \mathbb{E}[||E||^2\delta(E)\mathbf{1}_{||E||\ge a_\epsilon}] + \mathbb{E}[||E||^2\epsilon\mathbf{1}_{||E||<a_\epsilon}]\\ &\le \eta\,\gamma\,\mathbb{E}[\mathbf{1}_{||E||\ge a_\epsilon}] + \epsilon\mathbb{E}[||E||^2],\end{aligned}$$

for some constant[4] $\eta < \infty$.

As indicated by Equation (3), $E_j = \sum_i(\widehat{B}_{ij}/N - m_i K_{ij}(m))$ is the sum of the n independent random variables $(\widehat{B}_{ij}/N - m_i K_{ij}(m))$ and $\widehat{B}_{ij}$ has a binomial distribution of parameters $(Nm_i, K_{ij}(m))$. Hence, E_j can be expressed as the sum of N independent Bernoulli random variables. By Hoeffding's inequality,

$$\mathbf{P}\left[\|E_j\| \ge t\right] \le e^{-2Nt^2}.$$

This implies that $\mathbf{P}[\|E\| \ge a_\epsilon] \le \sum_{j=1}^n \mathbf{P}[\|E_j\| \ge a_\epsilon/n] \le ne^{-2Na_\epsilon/n^2}$.

Moreover, by Lemma 3, $\mathbb{E}[||E||^2] \le 1/N$. This shows that

$$\begin{aligned}\mathbb{E}[||E||^2\delta(E)] &\le \eta\,\gamma\,\mathbb{P}(||E|| \ge a_\epsilon) + \epsilon\mathbb{E}[||E||^2]\\ &\le \eta\,\gamma\,ne^{-2Na_\epsilon/n^2} + \frac{\epsilon}{N}\end{aligned}$$

The assert follows by using $\varepsilon(N) = \inf_{\epsilon>0}(N\,\eta\,\gamma\,ne^{-2Na_\epsilon/n^2} + \epsilon)$.

Lemma 5. *Under the assumptions of Lemma 1 and that $M^{(N)}(0)$ converges weakly to $\mu(0)$ as N goes to infinity, for any continuous function $g : \mathcal{U}^n \to \mathbb{R}^p_{\ge 0}$ and all t, there exists a function $\varepsilon_{t,g}(N)$ such that $\lim_{N\to\infty}\varepsilon_{t,g}(N) = 0$ and*

$$\|\mathbb{E}[g(M(t))] - g(\mu(t))\| \le \varepsilon_{t,g}(N).$$

Proof. We proceed by induction on t. The lemma holds for $t = 0$ because $M^{(N)}(0)$ converges weakly to $\mu(0)$. As g is continuous, there exists a function $\delta : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\|g(m) - g(m')\| \le \delta(\|m - m'\|)$ and $\lim_{r\to 0}\delta(r) = 0$. Moreover, as g and Φ_t are continuous, $g \circ \Phi_t$ is also uniformly continuous (and the continuity is uniform since $\mathcal{U}^n$ is compact). Hence

$$\begin{aligned}&\mathbb{E}\left[\|g(M(t+1)) - g(\mu(t+1))\|\right]\\ &= \mathbb{E}\left[\|g(M(t+1)) - g(\Phi_1(M(t))) + g(\Phi_1(M(t))) - g(\mu(t+1))\|\right]\\ &\le \mathbb{E}\left[\delta(\|M(t+1) - \Phi_1(M(t))\|)\right] + \mathbb{E}\left[g\circ\Phi_1(M(t)) - g\circ\Phi_1(\mu(t))\right]\\ &\le \mathbb{E}\left[\mathbb{E}\left[\delta(\|E\|) \mid M(t)\right]\right] + \varepsilon_{t,g\circ\Phi_1},\end{aligned}$$

where $E = M(t+1) - \Phi_1(M(t))$ converges to 0 (uniformly in M(t)) by Lemma 3.

[4] The specific value of η depends on the norm used; for instance $\eta = 1$ for the infinity norm.

Proof (Of the main theorem). We proceed by induction on t. The theorem clearly holds for $t = 0$ because $\Phi_0(M^{(N)}(0)) = M^{(N)}(0)$ by definition of Φ_0. Assume that the theorem now holds for some $t \geq 0$. We have :

$$\begin{aligned} N\left(\mathbb{E}[h(M(t+1))] - h(\mu(t+1))\right) =& N\mathbb{E}[h(M(t+1)) - h(\Phi_1(M(t)))] \\ &+ N(\mathbb{E}[h(\Phi_1(M(t)))] - h(\mu(t+1))). \end{aligned}$$

We will analyse the two lines separately. For the first line, the idea is to use Lemma 1. Indeed this line is equal to

$$\begin{aligned} &\mathbb{E}\left[N\mathbb{E}\left[h(M(t+1)) - h(\Phi_1(M(t))) \mid M(t)\right]\right] \\ &= \mathbb{E}\left[\frac{1}{2}(D^2h)(\Phi_1(M(t))) \cdot \Gamma(M(t))\right] + \Theta(N), \end{aligned}$$

where by Lemma 1, $\Theta(N)$ is such that $\|\Theta(N)\| \leq \varepsilon(N)$. By Lemma 5 with $g = (D^2h)(\Phi_1) \cdot \Gamma$, as N goes to infinity, this quantity converges to

$$\frac{1}{2}(D^2h)(\Phi_1(\mu(t))) \cdot \Gamma(\mu(t)) = \frac{1}{2}(D^2h)(\mu(t+1)) \cdot \Gamma(\mu(t)). \tag{5}$$

For the second line, the idea is to apply the induction hypothesis to $h \circ \Phi_1$ which can be done because the $h \circ \Phi_1$ is twice differentiable (because both h and Φ_1 are). This shows that

$$\begin{aligned} &N(\mathbb{E}[h(\Phi_1(M(t)))] - h(\mu(t+1))) \\ &= N(\mathbb{E}[h(\Phi_1(M(t)))] - h(\Phi_1(\mu(t)))) \\ &= D(h \circ \Phi_1)(\mu(t))V_t + \frac{1}{2}D^2(h \circ \Phi_1)(\mu(t)) \cdot W_t + \varepsilon_{t,h\circ\Phi_1}(N) \end{aligned}$$

The first term can be dealt with by applying the chain rule $D(h \circ \Phi_1) = (Dh)(\Phi_1)(D\Phi_1)$ which shows that:

$$\begin{aligned} D(h \circ \Phi_1)(\mu(t))V_t &= (Dh)(\Phi_1(\mu(t)))(D\Phi_1)(\mu(t))V_t \\ &= (Dh)(\mu(t+1))(D\Phi_1)(\mu(t))V_t \\ &= (Dh)(\mu(t+1))A_tV_t. \end{aligned} \tag{6}$$

For the second term, we apply the product rule and again the chain rule:

$$\begin{aligned} \frac{1}{2}D^2(h \circ \Phi_1) \cdot W_t &= \frac{1}{2}D\big((Dh)(\Phi_1)(D\Phi_1)\big) \cdot W_t \\ &= \Big(D\big((Dh)(\Phi_1)\big) \cdot (D\Phi_1) + (Dh)(\Phi_1) \cdot D(D\Phi_1)\Big) \cdot \frac{1}{2}W_t \\ &= \Big((D^2h)(\Phi_1) \cdot (D\Phi_1)(D\Phi_1)^T + (Dh)(\Phi_1)(D^2\Phi_1)\Big) \cdot \frac{1}{2}W_t \end{aligned}$$

By applying the last function at the point $\mu(t)$ we get that:

$$\begin{aligned} \frac{1}{2}D^2(h \circ \Phi_1)(\mu(t)) \cdot W_t =& (D^2h)(\Phi_1(\mu(t))) \cdot (D\Phi_1)(\mu(t))\frac{1}{2}W_t(D\Phi_1)^T(\mu(t)) \\ &+ (Dh)(\Phi_1(\mu(t)))(D^2\Phi_1(\mu(t))) \cdot \frac{1}{2}W_t, \end{aligned}$$

which, using the definition of $\mu(t+1) = \Phi_1(\mu(t))$ and the assumptions $A_t = (D\Phi_1)(\mu(t))$ and $B_t = (D^2\Phi_1(\mu(t)))$, is the same as:

$$\frac{1}{2}(D^2h)(\mu(t+1)) \cdot A_t W_t A_t^T + \frac{1}{2}(Dh)(\mu(t+1))(B_t \cdot W_t). \tag{7}$$

The theorem holds by combining Equations (5), (6) and (7).

3.3.2. Proof of Theorem 2

PROOF. The proof is inspired by the proof of [17, Theorem 3.1] and uses ideas of Stein's method. Because many details are similar to the proof of Theorem 1, we skip some details of computation in this proof.

Let h be a twice-differentiable function and let G_h be the function defined for all m by:

$$G_h(m) = \sum_{t=0}^{\infty} [h(\Phi_t(m)) - h(\mu(\infty))].$$

$G_h(m)$ is well defined because $\mu(\infty)$ is exponentially stable attractor.

By construction, for any m we have

$$\begin{aligned} G_h(m) &= h(m) - h(\mu(\infty)) + \sum_{t=1}^{\infty} [h(\Phi_t(m)) - h(\mu(\infty))] \\ &= h(m) - h(\mu(\infty)) + G_h(\Phi_1(m)) \end{aligned}$$

The above equation is a discrete time Poisson equation and implies that for any m:

$$h(m) - h(\mu(\infty)) = G_h(m) - G_h(\Phi_1(m)) \tag{8}$$

Assume that at time 0, the initial state $M(0)$ is distributed according to the stationary distribution of the system of size N. By the definition of stationarity, at time 1, $M(1)$ is also distributed according to the same stationary distribution and we have $\mathbb{E}[G_h(M(0))] = \mathbb{E}[G_h(M(1))]$.

By using (8) and then the above equation, we get:

$$\begin{aligned} \mathbb{E}\left[M(0)\right] - h(\mu(\infty)) &= \mathbb{E}\left[M(0) - h(\mu(\infty))\right] \\ &= \mathbb{E}\left[G_h(M(0)) - G_h(\Phi_1(M(0)))\right] \\ &= \mathbb{E}\left[G_h(M(1)) - G_h(\Phi_1(M(0)))\right] \end{aligned}$$

By Lemma 1, this shows that :

$$\begin{aligned} N(\mathbb{E}\left[M(\infty)\right] - h(\mu(\infty))) &= \mathbb{E}\left[G_h(M(1)) - G_h(\Phi_1(M(0)))\right] \\ &= \mathbb{E}\left[\frac{1}{2}D^2(G_h)(\Phi_1(M(0))) \cdot \Gamma(M(0))\right] + o(1) \\ &= \frac{1}{2}D^2(G_h)(\Phi_1(\mu(\infty))) \cdot \Gamma(\mu(\infty)) + o(1), \end{aligned}$$

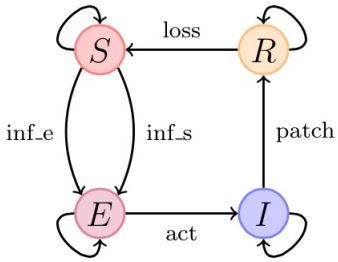


Figure 1: SEIR model of individual object

where the last equality comes from the fact that the stationary distribution of the system of size N converges weakly to a Dirac measure in $\mu(\infty)$ as N goes to infinity (see [15, Corollary 14]).

To conclude the proof, the only remaining step is to compute the second differential of G_h which can be expressed as the infinite sum:

$$D^2(G_h)(\mu(\infty)) = \sum_{t=0}^{\infty} D^2(h \circ \Phi_t)(\mu(\infty)).$$

The expressions for V_∞ and W_∞ come from plugging the above equations into Equation (5), (6) and (7). The uniqueness of the solution of the Lyapunov equation is due to the fact that the fixed point $\mu(\infty)$ is exponentially stable and therefore also linearly stable.

4. Refined Mean Field Model for SEIR

In this section we provide a simple example that illustrates the results for the refined mean field model of the simple computer epidemic SEIR example presented in [14]. Each object in the model consists of four local states: Susceptible (S), Exposed (E), Infected (I) (and active) and Recovered (R). The four-state SEIR model of an individual object is shown in Figure 1.

Its discrete time evolution is given by the following probability transition matrix $\mathbf{K}$ in which m_S, m_E, m_I and m_R denote the fraction of objects in the system that are in local state S, E, I and R, respectively:

$$\mathbf{K}(m_S, m_E, m_I, m_R) = \begin{pmatrix} 1-(\alpha_e+\alpha_i m_I) & \alpha_e+\alpha_i m_I & 0 & 0 \\ 0 & 1-\alpha_a & \alpha_a & 0 \\ 0 & 0 & 1-\alpha_r & \alpha_r \\ \alpha_l & 0 & 0 & 1-\alpha_l \end{pmatrix}$$

In other words, a susceptible becomes exposed with probability $(\alpha_e + \alpha_i m_I)$ – *i.e.*, α_e denotes the external and α_i the internal infection probability –; An exposed node activates his infection with probability α_a; An infected recovers

with probability α_r; and α_l is the probability to loose the protection against infection.

4.1. Computation of A, B and Γ

We illustrate how to apply Theorem 1 in its simplified form, when h is the identity function, as in Corollary 1–*(i)*. The first step is to compute the Jacobian and the Hessian of the function Φ_1 for a generic occupancy measure vector m at time step t. Written as a column vector, the function $\Phi_1(m) = mK(m)$ is given by

$$\Phi_1(m) = \begin{pmatrix} m_S(1-\alpha_e-\alpha_i m_I)+\alpha_l m_R \\ m_S(\alpha_e+\alpha_i m_I)+(1-\alpha_a)m_E \\ m_E\alpha_a+(1-\alpha_r)m_I \\ \alpha_r m_I+(1-\alpha_l)m_R \end{pmatrix}$$

Hence, the Jacobian is the following 4×4 matrix:

$$D(\Phi_1)(m_S,m_E,m_I,m_R) = \begin{pmatrix} 1-(\alpha_e+\alpha_i m_I) & 0 & -\alpha_i m_S & \alpha_l \\ \alpha_e+\alpha_i m_I & 1-\alpha_a & \alpha_i m_S & 0 \\ 0 & \alpha_a & 1-\alpha_r & 0 \\ 0 & 0 & \alpha_r & 1-\alpha_l \end{pmatrix}$$

The Hessian is a $4 \times 4 \times 4$ tensor. We provide them as 4 matrices of 4×4, one for each function $(\Phi_1)_j$, where $j \in \{S, E, I, R\}$:

$$D^2((\Phi_1)_S)(m_S,m_E,m_I,m_R) = \begin{pmatrix} 0 & 0 & -\alpha_i & 0 \\ 0 & 0 & 0 & 0 \\ -\alpha_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$D^2((\Phi_1)_E)(m_S,m_E,m_I,m_R) = \begin{pmatrix} 0 & 0 & \alpha_i & 0 \\ 0 & 0 & 0 & 0 \\ \alpha_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrices for I and R are two 4×4 zero-matrices. The 4×4 matrix Γ depends on K and the occupancy measure m as defined in Lemma 1. The refined mean field approximation of the occupancy measure vector is thus given by $\mathbb{E}[M^{(N)}(t)] \approx \mu(t) + V_t/N$, where V_t is computed recursively, according to Theorem 1.

4.2. Dynamics of the SEIR Model and its Approximations

We consider a model with the following parameter values for the local transition probabilities: $\alpha_e = 0.01, \alpha_i = 0.08, \alpha_r = 0.02, \alpha_l = 0.01$ and $\alpha_a = 0.04$. Initially, $M(0) = (0.2, 0.2, 0.2, 0.4)$. Figure 2 shows the results for the classical mean field approximation, the refined mean field approximation and the average of 100,000 runs of a stochastic simulation of the model obtained. The results are

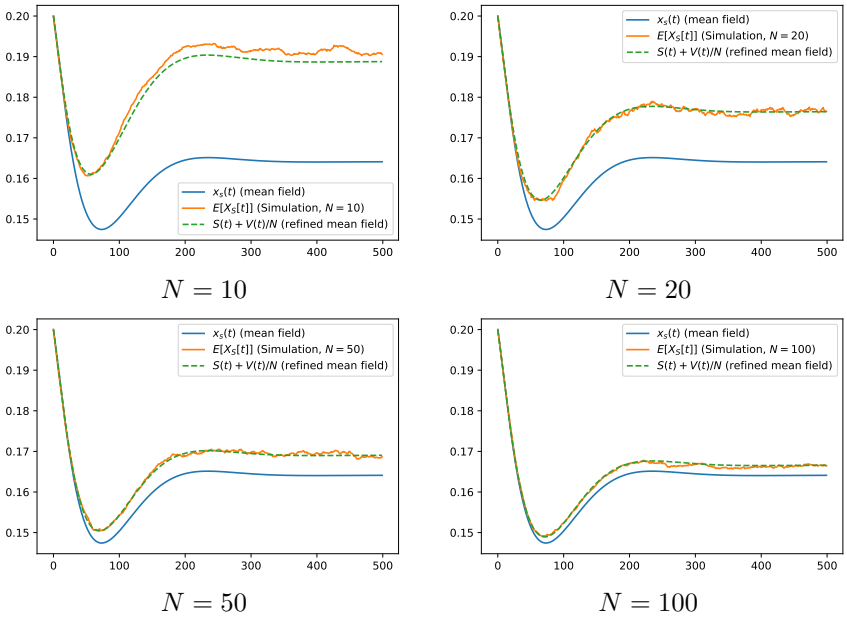


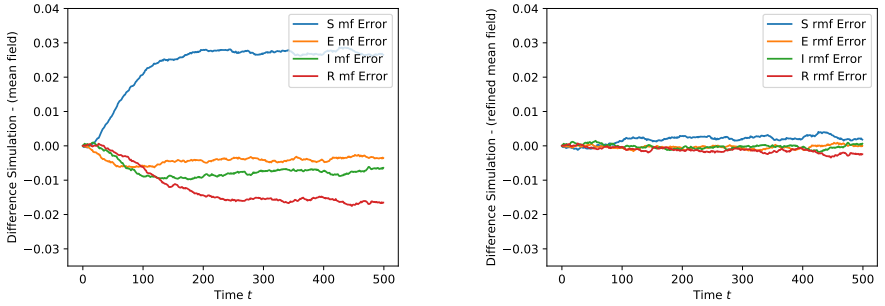
Figure 2: Evolution of the fraction of objects in state S for population sizes $N = 10$, $N = 20$, $N = 50$ and $N = 100$. The figures compare the classical mean field approximation (obtained with `python`–`numpy`) with the refined one and with the average of 10,000 simulation runs of the system.

given for population size $N = 10$, $N = 20$, $N = 50$ and $N = 100$, respectively; time t ranges from 0 to 500 time units.

We observe that, as exepcted, the gap between the classical mean field approximation and the simulation is relatively small and decreases with N. Still, for $N = 10$, we observe a clear difference between the classical mean field approximation and the simulation, whereas the refined mean field provides a much closer approximation (in this case, the graphs overlap almost everywhere). With the increase of the population size N both approximations converge to the same value, as well as the value obtained by simulation: for $N \geq 50$, the curves are almost indistinguishable.

To highlight the differences, we plot in Figure 3 the difference between the two approximations with respect to the simulation : On the left panel, we plot a function of time for the quantities $\mathbb{E}[M(t)] - \mu(t)$; On the right we plot $\mathbb{E}[M(t)] + V_t/N - \mu(t)$ (in both cases for $N = 10$). We observe that the refined mean field approximation (right panel) is an order of magnitude closer to the value obtained by simulation: while the error of the classical mean field approximation can be larger than 0.05, the error of the refined mean field approximation remains always smaller than 0.01.

These two figures illustrate that Theorem 1 is not just valid asymptotically, but actually it refines the classical mean field approximation for relatively small



Error of the mean field approximation Error of the refined mean field approx.

Figure 3: SEIR model: Quantification of the difference (error) between the simulation results and the classical mean field approximation (left) and the refined mean field result (right), respectively, for $N = 10$. The results show simulation value minus mean field value.

values of N. To go further, we study the steady-state distribution in Table 1 in which we display the average proportion of objects in states S, E, I or R estimated by simulation, refined mean field approximation and classical mean field approximation. This illustrates the approximation accuracy for the steady state of the SEIR example for each local state of an object. As for the two previous figures, this table illustrates that the refined mean field approximation provides very accurate estimates of the true stationary distribution even for very small values of N, which shows that the asymptotic results presented in Theorem 2 are also useful for small values of N. These results are in line with the results presented in [17] for continuous time mean field models.

State	S	E	I	R
Simulation ($N = 10$)	0.191	0.115	0.231	0.462
Refined mean field ($N = 10$)	0.189	0.116	0.232	0.464
Mean field ($N = 10$)	0.164	0.119	0.239	0.478

Table 1: SEIR model: Comparison of the accuracy of the mean field and refined mean field approximation. The columns show the average proportion of objects in the states susceptible (S), exposed (E), infected (I) and recovered (R), respectively. Each item in the table was computed by measuring the occupancy measure for times $t = 1000$, i.e. when the systems' occupancy measure has reached a sufficiently stable value. Simulation values are averages over $100,000$ simulations.

5. Refined Mean Field Model for WSN

The next example concerns a simple model of a wireless sensor network [26]. Such networks are composed of wireless sensor nodes and gateways. This example serves two purposes. First it shows that the assumption of homogeneous

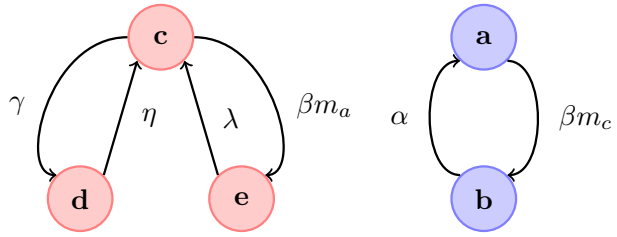


Figure 4: WSN model of individual objects: Sensor Node (left) and Gateway (right)

objects is not restrictive: In this example, there are two classes of objects, which is represented by having a block-diagonal matrix K. Second, we use it to consider a function h that is not just the projection on one coordinate.

Wireless sensor nodes have three local states. In the initial state, e, a sensor node waits for detecting an event of interest and collects data for that event. After that, the node moves to state c to communicate its data to an available gateway. The communication attempt may timeout if no gateway is available. In that case the sensor node moves to state d, introducing some delay before moving back to state c for a further communication attempt.

Gateway nodes have two states. Initially they are in state a and available to receive data from a sensor node. Upon connection to a sensor node they move to state b during which they are busy processing the data. When in state b they are temporarily unavailable for communication with other sensor nodes. After processing the batch of data they move back to state a.

We consider a model where objects have five local states $\{a, b, c, d, e\}$, where a and b are states of a gateway node and c, d and e are states of a sensor node, i.e. each object in the model can behave either as a gateway or as a sensor, but it cannot change its behaviour from that of a gateway to that of a sensor, or vice-versa. A system is then composed of N (syntactical) homogeneous objects with a fixed fraction $m_G = m_a + m_b$ of gateway nodes and a fraction $m_W = m_c + m_d + m_e$ of wireless sensor nodes, such that $m_W = 1 - m_G$. To keep the model simple for the purpose of illustrating the refined mean field approach, we do not consider interference due to collision in the communication between nodes and gateways.

The probability transition matrix is given below:

$$\mathbf{K}(m) = \begin{pmatrix} 1-\beta m_c & \beta m_c & 0 & 0 & 0 \\ \alpha & 1-\alpha & 0 & 0 & 0 \\ 0 & 0 & 1-\gamma-\beta m_a & \gamma & \beta m_a \\ 0 & 0 & \eta & 1-\eta & 0 \\ 0 & 0 & \lambda & 0 & 1-\lambda \end{pmatrix},$$

where α denotes the probability of the gateway to get again available, β the probability of data communication between the gateway and a sensor node, λ the probability that a sensor node is ready to send data, γ the probability that

a sensor node performs a time-out and η the probability that a delayed sensor node tries to communicate again.

In the example we will use the following values for the above parameters: $\alpha = 0.09, \beta = 0.9, \lambda = 0.09, \gamma = 0.01$, and $\eta = 0.01$, and let $M^{(N)}(t)$ denote, as usual, the occupancy measure process of the WSN model (leaving N and t implicit for the sake of notation simplicity). We are interested in the average response time of a sensor node, i.e. the time a sensor node needs to wait to be able to communicate its data to the gateway. This expected response time can be defined as the fraction between the sensor nodes that are already waiting to communicate their data, i.e. the sensor nodes in state c and state d, and the new sensor nodes that became ready in the current time step, i.e. λ times the nodes in local state e:

$$\mathbb{E}[R] = \mathbb{E}\left[\frac{(M_c + M_d)}{\lambda M_e}\right],$$

With reference to Theorem 1, we define $h(x_1, x_2, x_3, x_4, x_5) = \frac{(x_3+x_4)}{\lambda x_5}$.

5.1. Computation of A, B and Γ *for the WSN model*

In the sequel, we make reference to $m = (m_a, m_b, m_c, m_d, m_e) \in \mathcal{U}^5$. The Jacobian of function Φ_1:

$$D(\Phi_1)(m) = \begin{pmatrix} 1-\beta m_c & \alpha & -\beta m_a & 0 & 0 \\ \beta m_c & 1-\alpha & \beta m_a & 0 & 0 \\ -\beta m_c & 0 & 1-\gamma-\beta m_a & \eta & \lambda \\ 0 & 0 & \gamma & 1-\eta & 0 \\ \beta m_c & 0 & \beta m_a & 0 & 1-\lambda \end{pmatrix}$$

The Hessian of the function Φ_1 satisfies $D^2((\Phi_1)_d)(m) = 0$:

$$D^2((\Phi_1)_a)(m) = D^2((\Phi_1)_c)(m) = \begin{pmatrix} 0 & 0 & -\beta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$D^2((\Phi_1)_b)(m) = D^2((\Phi_1)_e)(m) = \begin{pmatrix} 0 & 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The Jacobian of function h is $D(h)(m) = (0, 0, \frac{1}{\lambda m_e}, \frac{1}{\lambda m_e}, -\frac{m_c+m_d}{\lambda m_e^2})$ and its Hessian is

$$D^2(h)(m) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{\lambda m_e^2} \\ 0 & 0 & 0 & 0 & -\frac{1}{\lambda m_e^2} \\ 0 & 0 & -\frac{1}{\lambda m_e^2} & -\frac{1}{\lambda m_e^2} & \frac{2*(m_c+m_d)}{\lambda * m_e^3} \end{pmatrix}.$$

The 5×1 vector V_t and the 5×5 matrix W_t are computed recursively according to Theorem 1, using the new Jacobian and Hessian for function Φ; thus the refined mean field approximation of the measure of interest is given by $\mathbb{E}[h(M^{(N)}(t))] \approx h(\Phi_t(m)) + (D(h) \cdot V_t + \frac{1}{2} * (D^2(h) \cdot W_t))/N$.

5.2. Results

In Figure 5 various approximations of the expected response time for the WSN model are shown, for time values t ranging from 0 to 400 time units. We consider a relatively small system with 15 nodes (10 sensor and 5 gateway nodes). Recall that the function h is defined by : $h(x_1, x_2, x_3, x_4, x_5) = \frac{(x_3+x_4)}{\lambda x_5}$. We compare five curves :

1. The blue curve labelled Classic Mean Field (1) (obtained with Octave) shows the expected response time when this is approximated by defining $\mathbb{E}[R]$ as in [26]:
$$\mathbb{E}[R] = \frac{(m_c + m_d)}{\lambda m_e} = h(m),$$
where m_c, m_d and m_e denote the classical mean field approximation values for the fractions of sensor nodes being in state c, d and e, respectively.
2. The red curve (2) is the expectation of $h(M)$, computed by stochastic simulation:
$$\mathbb{E}\left[h(M)\right] = \min(\mathbb{E}\left[\frac{(M_c + M_d)}{\lambda M_e}\right], 100)$$
In the case of individual simulation runs it may of course happen that $M_e = 0$ occasionally. This is why we defined $\mathbb{E}[h(M)]$ as the minimum between the actual value and 100. The latter is the value one obtains when one but all nodes are waiting and the last node is getting ready for communication too: $M_c + M_d = 9$ and $M_e = 1$, i.e. $9/(0.09*1) = 100$.
3. The orange curve (3) shows the expected response time approximated using the refined mean field approximation of Theorem 1 with the function h.

For comparison, we also compute two other quantities :

4. The purple curve (4) shows the response time approximated as follows, using the refined mean field approximation for the fraction of sensor nodes in each state (i.e. we use Theorem 1 with the identity function and then apply h):
$$h(rmf) = \frac{(rmf_c + rmf_d)}{\lambda rmf_e}$$
where $rmf_c = \mu_c + V_c/N$ denotes the refined mean field approximation of the fraction of sensor nodes in state c, and similarly for rmf_d and rmf_e .

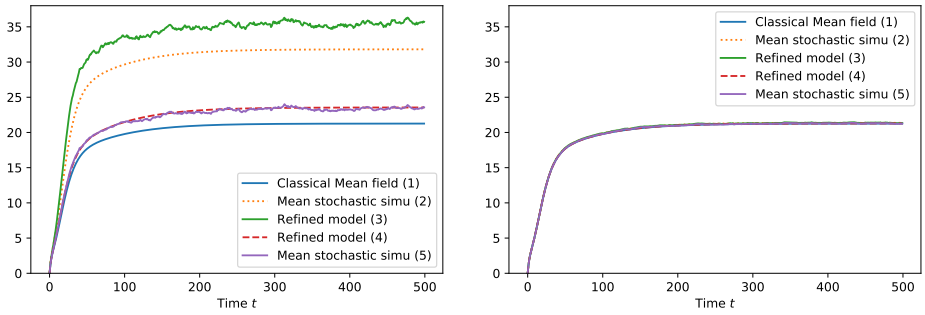
5. Finally, the green curve (5) shows the expected response time defined as:

$$h(\mathbb{E}[M]) = \frac{(\mathbb{E}[M_c] + \mathbb{E}[M_d])}{\lambda \mathbb{E}[M_e]}$$

where $\mathbb{E}[M_c]$ is the average fraction of sensor nodes in state c obtained via the average of 100,000 individual simulation runs of the model. $\mathbb{E}[M_d]$ and $\mathbb{E}[M_e]$ are obtained in a similar way.

In Figure 5(a), we plot these various curves for 15 nodes in total. We make two observations. First, in this case the value obtained by simulation (red curve (2)) is almost 50% larger than the classic mean field approximation (1) whereas the refined approximation (3) is much closer. Second, the purple curve (4) is close to the green curve (5) but quite far away from the red (2) and orange (3) curves. This shows that when applying Theorem 1, computing a refined model for $\mathbb{E}[h(M)]$ and for $h(\mathbb{E}[M])$ might lead to very different results.

Of course, the larger N gets (in an otherwise equal model), the closer the orange (3) and red (2) curves will get to the blue curve (1), i.e. the classic mean field approximation, as illustrated in Figure 5(b). In both cases, all curves collapse into to a single curve.



(a) $N = 15$: 10 sensors; 5 gateways (b) $N = 1500$: 1000 sensors; 500 gateways

Figure 5: Expected response time $E[R]$ for a sensor node to communicate its data to a GW for a WSN model with N nodes of which $2N/3$ are sensor nodes and $N/3$ are gateway nodes. The red line (2) is an average over 20,000 simulations for $N = 15$ and 1000 for $N = 1500$.

6. Refined Mean Field Model for Majority Rule Decision-making

The example in this section concerns a model for collective decision-making. The model is inspired by the work of Montes de Oca et al. (see [27, 28] and references therein). Collective decision-making is a process whereby the members of a group decide on a course of action by consensus. Such collective decision-making processes have also been applied in swarm robotics. In particular, in

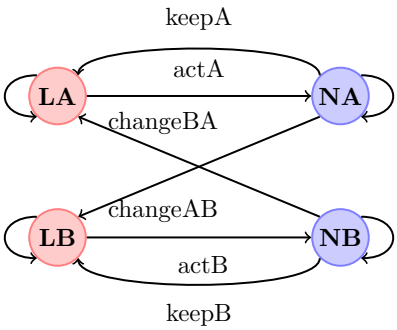


Figure 6: Majority rule differential latency model of an individual object. Latent states are red, non-latent ones blue.

that context the robots where asked to choose between two actions that have the same effect but differ in their execution times [27].

One strategy of collective decision making is the use of the majority rule. In this strategy the agents in a population are initially divided into two groups. One in which all members have opinion A and one where all members have opinion B. In every step three agents are selected randomly from the total population to form a temporary team. The team applies the majority rule such that all its members adopt the opinion held by the majority (i.e. at least two) of the members, after which they return to the total population until the population has reached a consensus on one of the two opinions.

In the majority rule strategy extended with differential latency the agents in the population are not all the time available for team formation. Both types of agents are assumed to perform an action with a certain duration during which they cannot participate in team formation. For example, agents with opinion B perform such actions taking (on average) relatively more time than those with opinion A. In [27] such latency periods for agents with opinion A and B are modelled by random variables with exponential distributions with rate λ_A and λ_B respectively. For simplicity, it can also be assumed that the A-type actions take 1 time unit on average (i.e. $\lambda_a = 1$) and that B-type actions take $1/\lambda$ time units on average, where λ takes a value in $(0, 1]$. This variant of self-organised collective decision-making is known as majority rule with differential latency (MRDL).

In the following we develop a probabilistic, discrete time variant of the MRDL strategy which we call MRDL-DT. In this variant agents can have either opinion A or opinion B, and in both cases they can be either be latent or not, leading to a partition of the population into exactly four classes: LA (latent A), NA (non-latent A), LB (latent B), NB (non-latent B). It is assumed that if an agent is latent it cannot be selected for team formation. The four state MRDL-DT model of an individual object of the population is shown in Figure 6. The name of the states indicate in which class the object is.

The behaviour of an individual object is as follows. Initially the object is latent, and, assuming it has opinion A (state LA), it finishes its job and becomes available for team-formation (transition $actA$) moving to state NA with probability $1/q$, for appropriate q. When in NA it gets selected in a team with two other members. If the two other members have opinion B it changes its opinion into B and moves to state LB. This can happen with a probability $3m_{NB}^2$ where factor 3 models the fact that we abstract from the exact order in which the members of the team are selected, which can happen in 3 different ways. In alternative the two other members can have both opinion A, or one opinion A and the other opinion B. In that case the opinion of the object does not change and the object moves back to LA with probability $\frac{3}{q}(m_{NA}^2 + m_{NA}m_{NB})$.

If the object is in state LB it becomes available for team formation, moving to state NB with a probability λ/q, where λ is a value in $(0, 1]$. The latter models the relative longer duration of activity B with respect to A. The behaviour in state NB is similar to that in state NA, except that now the opinion may change from B to A.

The discrete time evolution of the model is given by probability transition matrix $\mathbf{K}$ in which m_{LA}, m_{NA}, m_{LB} and m_{NB} denote the fraction of objects in the system that are in local state LA, NA, LB and NB, respectively and $m = (m_{LA}, m_{NA}, m_{LB}, m_{NB})$:

$$\mathbf{K}(m) = \begin{pmatrix} 1 - \frac{1}{q} & \frac{1}{q} & 0 & 0 \\ \frac{3}{q}(m_{NA}^2 + m_{NA}m_{NB}) & NA(m) & \frac{3}{q}m_{NB}^2 & 0 \\ 0 & 0 & 1 - \frac{\lambda}{q} & \frac{\lambda}{q} \\ \frac{3}{q}m_{NA}^2 & 0 & \frac{3}{q}(m_{NB}^2 + m_{NA}m_{NB}) & NB(m) \end{pmatrix}$$

where

$$NA(m) = 1 - \frac{3}{q}(m_{NB}^2 + m_{NA}^2 + m_{NA}m_{NB})$$

and

$$NB(m) = 1 - \frac{3}{q}(m_{NA}^2 + m_{NB}^2 + m_{NA}m_{NB}).$$

Since we are dealing with clock-synchronous discrete systems, we also introduced the discretisation factor $q = 10$ so that only a fraction of the population is moving from the latent to the non-latent state at any time.

In the example we use the values $q = 10$ and λ taking values 1.0, 0.5 and 0.25 in the various analyses, modelling that task B takes the same time as task A, or twice as much time or four times as much time as task A, *on average*, respectively. We are interested in the evolution of the consensus, C_A, on opinion A as a function of the initial values and the differential latency λ. Let

$$\mathbb{E}[C_A] = \mathbb{E}[M_{LA} + M_{NA}] = \mathbb{E}[h(M_{LA}, M_{NA}, M_{LB}, M_{NB})]$$

where, with reference to Theorem 1, $h(x_1, x_2, x_3, x_4) = x_1 + x_2$.

6.1. Computation of A, B and Γ for the MRDL-DT model

As before, we first need to compute the Jacobian and the Hessian of the function Φ_1 for a generic occupancy measure vector m at time step t. The Jacobian of function Φ_1 is:

$$D(\Phi_1)(m) = \begin{pmatrix} 1-1/q & \frac{9}{q}m_{NA}^2 + \frac{12}{q}m_{NA}m_{NB} & 0 & \frac{6}{q}m_{NA}^2 \\ 1/q & JNA(m) & 0 & -\frac{3}{q}m_{NA}^2 - \frac{6}{q}m_{NA}m_{NB} \\ 0 & \frac{6}{q}m_{NB}^2 & 1-\frac{\lambda}{q} & \frac{12}{q}m_{NB}m_{NA} + \frac{9}{q}m_{NB}^2 \\ 0 & -\frac{3}{q}m_{NB}^2 - \frac{6}{q}m_{NA}m_{NB} & \frac{\lambda}{q} & JNB(m) \end{pmatrix}$$

where

$$JNA(m) = 1 - (\frac{9}{q}m_{NA}^2 + \frac{6}{q}m_{NA}m_{NB} + \frac{3}{q}m_{NB}^2)$$

and

$$JNB(m) = 1 - (\frac{3}{q}m_{NA}^2 + \frac{9}{q}m_{NB}^2 + \frac{6}{q}m_{NA}m_{NB}).$$

The Hessian of function Φ_1 is:

$$D^2((\Phi_1)_{LA})(m) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{18}{q}m_{NA} + \frac{12}{q}m_{NB} & 0 & \frac{12}{q}m_{NA} \\ 0 & 0 & 0 & 0 \\ 0 & \frac{12}{q}m_{NA} & 0 & 0 \end{pmatrix}$$

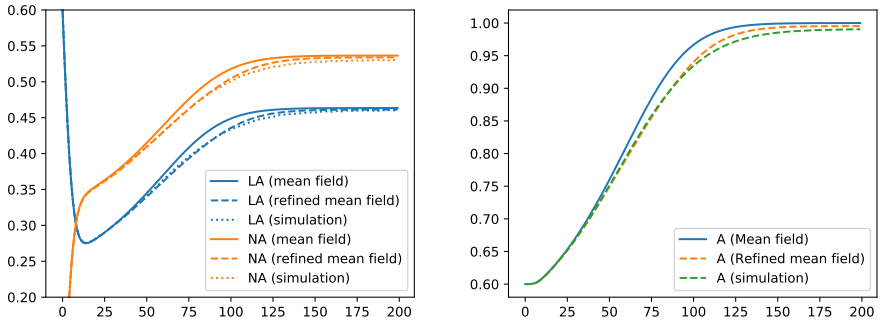
$$D^2((\Phi_1)_{NA})(m) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\frac{18}{q}m_{NA} - \frac{6}{q}m_{NB} & 0 & -\frac{6}{q}m_{NA} - \frac{6}{q}m_{NB} \\ 0 & 0 & 0 & 0 \\ 0 & -\frac{6}{q}m_{NA} - \frac{6}{q}m_{NB} & 0 & -\frac{6}{q}m_{NA} \end{pmatrix}$$

$$D^2((\Phi_1)_{LB})(m) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{12}{q}m_{NB} \\ 0 & 0 & 0 & 0 \\ 0 & \frac{12}{q}m_{NB} & 0 & \frac{18}{q}m_{NB} + \frac{12}{q}m_{NA} \end{pmatrix}$$

$$D^2((\Phi_1)_{NB})(m) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\frac{6}{q}m_{NB} & 0 & -\frac{6}{q}m_{NA} - \frac{6}{q}m_{NB} \\ 0 & 0 & 0 & 0 \\ 0 & -\frac{6}{q}m_{NA} - \frac{6}{q}m_{NB} & 0 & -\frac{18}{q}m_{NB} - \frac{6}{q}m_{NA} \end{pmatrix}$$

6.2. Results for the MRDL-DT example

We first show some results for a medium size population of $N = 160$. Figure 7 shows the dynamics of the fractions of the population having opinion A and B, for both the latent and non-latent objects, for the first 200 time units. Similarly to the previous examples, a good correspondence can be observed between the results of the mean of 1000 simulation runs and the mean field approximations.



(a) Latent and non-latent

(b) Dynamics of opinion A

Figure 7: Dynamics of opinion A, latent and non-latent. Classical mean field (plain lines), refined mean field (dashed lines) and simulation results (dotted lines) of MRDL model with 160 objects, $\lambda = 1.0$, q=10 and initially $0.6 * 160 = 96$ have opinion A and the population is initially latent.

Also in this case the refined mean field provides a better approximation than the classical mean field approximation for the period ranging approximately from 50 to 150 time units.

A similar improved correspondence can be observed when considering the aggregated populations with opinion A or opinion B, respectively, as shown in Figure 7(b).

However, if a much smaller population is considered, e.g. $N = 32$, the mean field approximation differs considerably from the simulation results and the refined approximation does not really improve the accuracy of the approximation. This is what can be observed in Figure 8, where the results for $N = 32$ are shown for a model without differential latency (i.e. $\lambda = 1.0$), for $t \in [0, 500]$. This can be explained as follows : For a large population, a system that is initially biased towards one opinion will reach consensus on this opinion. For a small population, however, a system that is initially biased towards one opinion still can reach consensus on the other opinion due to intrinsic stochastic fluctuations. This cannot be taken into account in a mean field population model. However, both analytical models, derived from a master equation approach, and simulation show that the probability to reach consensus on a given opinion rapidly converges to a step function for a growing population size N [28], where the critical density is given by $m_A = \frac{\lambda}{(1+\lambda)}$, where m_A is the initial fraction of the population with opinion A. For large N, if $m_A > \frac{\lambda}{(1+\lambda)}$ the system almost surely reaches consensus on A, whereas for $m_A < \frac{\lambda}{(1+\lambda)}$ it almost surely reaches consensus on B. This explains why for larger populations the mean field approximations become increasingly accurate as shown in Figure 7.

This example shows a limit of the refined mean field approximation: when a system has multiple equilibrium point (and in particular when there are multiple absorbing states as in this example), the dynamics of the mean field approxi-

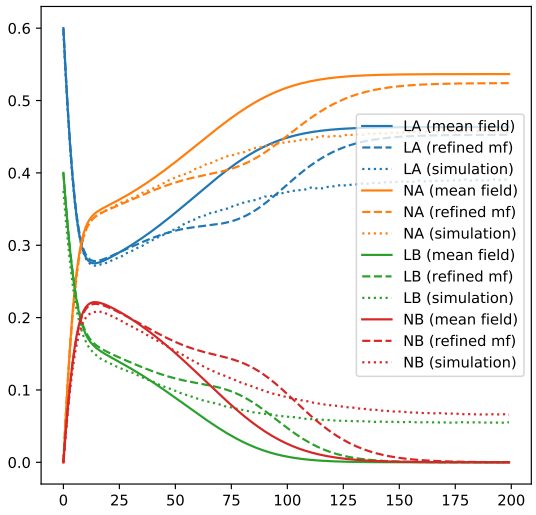


Figure 8: Classical mean field (plain lines), refined mean field (dashed lines) and simulation results (dotted lines) of MRDL model with 32 objects, $\lambda = 1.0$, q=10 and initially $\lfloor 0.6 * 32 \rfloor = 9$ have opinion A and the population is initially latent.

mation depends on the initial state of the system: For a given initial state the mean field will always follow the same trajectory (it is a deterministic system). When the system is large, the random fluctuations will remain small and the corresponding stochastic system will stay in the same basin of attraction. In the case of a small population, however, the dynamics will be greatly affected by the random fluctuations. These fluctuations can lead the system to another basin of attraction than the original one.

7. Non-Exponentially Stable Equilibrium : Accuracy versus Time

In the previous sections, the dynamical system $m = \Phi_1(m)$ has either one exponentially stable attractor (Section 4) or multiple exponentially stable attractors (Section 5). When the attractor is unique and exponentially stable, the accuracy of the mean field or refined mean field approximation is uniform in time (Theorem 2(ii)). In this section, we study the case of a system that has a unique attractor but that is not exponentially stable. We show that, in this case, the accuracy of the (refined) mean field approximation is no longer uniform in time.

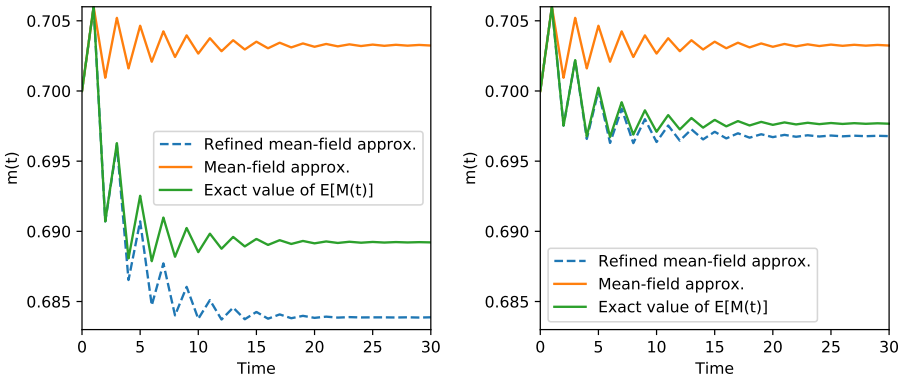
We consider a system with N objects in which each object is in state 0 or 1. An object in state 1 goes to state 0 with probability 1 and an object in state 0 goes to 1 with probability αm_0, where $\alpha \in (0, 1)$ is a parameter. The transition

matrix K is

$$K(m) = \begin{bmatrix} 1 - \alpha m_0 & \alpha m_0 \\ 1 & 0 \end{bmatrix}$$

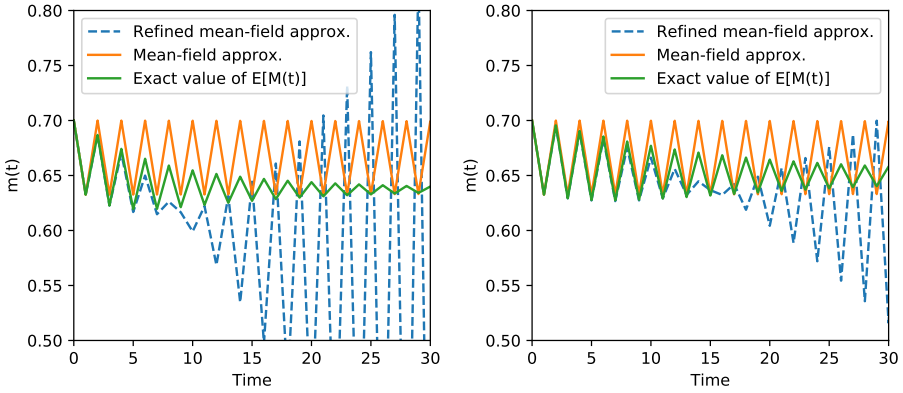
The function $\Phi_1 : m \mapsto mK(m)$ has a unique fixed point whose first component is $\mu_0(\infty) = (\sqrt{1+4\alpha} - 1)/(2\alpha)$. This fixed point is exponentially stable if and only if $\alpha < 0.75$.

7.1. Transient regime and accuracy for large t



(a) $N = 10$ (b) $N = 30$

Figure 9: Exponentially stable case ($\alpha = 0.6$).



(a) $N = 10$ (b) $N = 30$

Figure 10: Non-exponentially stable case ($\alpha = 0.75$).

In Figure 9 and Figure 10, we plot the first component of the mean field $\mu(t)$ and refined mean field approximation $\mu(t) + V(t)/N$ as well an exact value

of $\mathbb{E}[M(t)]$ for $N = 10$ and $N = 30$. The initial value is $m = (0.7, 0.3)$. The exact value of $\mathbb{E}[M(t)]$ was computed by a numerical method that uses the fact that the system with N objects can be described by a Markov chain with $N+1$ states.

These figures show that the refined approximation always improves the accuracy compared to the classical mean field approximation for small values of t, both for $\alpha = 0.6$ and $\alpha = 0.75$. The situation for large values of t is quite different. On the one hand, when the fixed point is exponentially stable ($\alpha = 0.6$, Figure 9), the refined approximation is very accurate for all values of t. On the other hand, when the fixed point is not exponentially stable ($\alpha = 0.75$, Figure 10), the refined approximation seems to be unstable and is not a good approximation of $\mathbb{E}[M(t)]$ for values of t that are too large compared to N ($t > 7$ for $N = 10$ or $t > 12$ for $N = 30$).

7.2. Steady-state convergence

To explore how the non-exponentially stable case affects the accuracy of mean field approximation, we now study in more details the steady-state convergence when $\alpha = 0.75$. It is known (see for example [15, Corollary 14]) that when the dynamical system $m = \Phi_1(m)$ has a unique attractor $\mu(\infty)$, then the steady-state expectation $\mathbb{E}[M^{(N)}]$ converges to $\mu(\infty)$ as N goes to infinity. Theorem 2 shows that if in addition the attractor is exponentially stable then $\mathbb{E}[M^{(N)}] \approx \mu(\infty) + V/N$.

In Figure 11, we show that the latter no longer holds when the mean field system has a unique attractor that is *not* exponentially stable. We consider the same model with $\alpha = 0.75$ for which $\mu(\infty) = 2/3$. We plot in Figure 11 $\sqrt{N}(\mathbb{E}[M^{(N)}] - \mu(\infty))$, where we computed $\mathbb{E}[M^{(N)}]$ by inverting the transition matrix of the system of size N. This figure shows that $\sqrt{N}(\mathbb{E}[M^{(N)}] - \mu(\infty))$ does not converge to 0 as N goes to infinity but seems to converge to approximately -0.0975 (as indicated by the fitted line in orange). This suggests that for this model, one has in steady-state :

$$\mathbb{E}\left[M^{(N)}\right] \approx \mu(\infty) - \frac{0.0975}{\sqrt{N}} + \frac{0.14}{N}. \tag{9}$$

Note that the constants -0.0975 and 0.14 were obtained by a purely numerical method that consist in finding the best curve of the form $a + b/\sqrt{N}$ that fits $\sqrt{N}(\mathbb{E}[M^{(N)} - \mu(\infty)])$. For now, we do not known if there exists a systematic way to obtain these values for another model that would also have a non-exponentially equilibrium point. We left this question for future work.

We remark that the convergence of $\mathbb{E}[M^{(N)}]$ to $\mu(\infty)$ observed Equation (9) is in $O(1/\sqrt{N})$ and not in $O(1/N)$. This model satisfies all the assumptions of Theorem 2 but one : The attractor $\mu(\infty)$ is not exponentially stable. This suggests that having an exponentially stable attractor is needed to obtain a convergence in $1/N$.

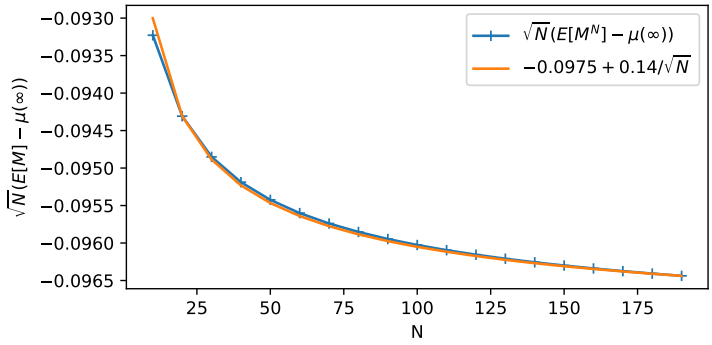


Figure 11: Non-exponentially stable case ($\alpha = 0.75$) : convergence of $\mathbb{E}[M(t)]$ to $\mu(\infty)$.

8. Conclusion

In this paper we studied population models composed of (clock-)synchronous objects. A classical method to study such systems is to consider the mean field approximation. By studying the accuracy of this deterministic approximation, we developed a new approximation, that we call the *refined* mean field approximation. We illustrated on a few examples that this approximation can greatly improve the accuracy of the classical mean field limit, also for systems with a relatively small size ($10 - 20$ objects). Yet, this refined approximation has some limitations when the deterministic approximation has multiple basins of attraction or has a unique attractor that is not exponentially stable.

The proposed refined approximation is given by a set of linear equations that scales as the square of the dimension of the model (but does not depend on the system size). For now, we limited our study to relatively small models, for which the Jacobian and Hessian can be computed in closed form. We are currently investing means to make this computation automatic which will allow us to study large-scale examples.

References

[1] D. Wilkinson, Stochastic Modelling for Systems Biology, Chapman & Hall, 2006.

[2] H. Andersson, T. Britton, Stochastic Epidemic Models and Their Statistical Analysis, Springer-Verlag, 2000.

[3] N. D. Vvedenskaya, R. L. Dobrushin, F. I. Karpelevich, Queueing system with selection of the shortest of two queues: An asymptotic approach, Problemy Peredachi Informatsii 32 (1) (1996) 20–34.

[4] M. Mitzenmacher, The power of two choices in randomized load balancing, IEEE Trans. Parallel Distrib. Syst. 12 (10) (2001) 1094–1104. doi:10.

1109/71.963420.
URL https://doi.org/10.1109/71.963420

[5] J. N. Tsitsiklis, K. Xu, On the power of (even a little) centralization in distributed processing, in: A. Merchant, K. Keeton, D. Rubenstein (Eds.), SIGMETRICS 2011, Proceedings of the 2011 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, San Jose, CA, USA, 07-11 June 2011 (Co-located with FCRC 2011), ACM, 2011, pp. 161–172. doi:10.1145/1993744.1993759.
URL http://doi.acm.org/10.1145/1993744.1993759

[6] W. Minnebo, B. Van Houdt, A fair comparison of pull and push strategies in large distributed networks, IEEE/ACM Trans. Netw. 22 (3) (2014) 996–1006. doi:10.1109/TNET.2013.2270445.
URL https://doi.org/10.1109/TNET.2013.2270445

[7] L. Massoulié, M. Vojnović, Coupon replication systems, SIGMETRICS Perform. Eval. Rev. 33 (1) (2005) 2–13. doi:10.1145/1071690.1064215.
URL http://doi.acm.org/10.1145/1071690.1064215

[8] N. Gast, G. Bruno, A mean field model of work stealing in large-scale systems, SIGMETRICS Perform. Eval. Rev. 38 (1) (2010) 13–24. doi:10.1145/1811099.1811042.
URL http://doi.acm.org/10.1145/1811099.1811042

[9] B. Van Houdt, A mean field model for a class of garbage collection algorithms in flash-based solid state drives, in: M. Harchol-Balter, J. R. Douceur, J. Xu (Eds.), ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS '13, Pittsburgh, PA, USA, June 17-21, 2013, ACM, 2013, pp. 191–202. doi:10.1145/2465529.2465543.
URL http://doi.acm.org/10.1145/2465529.2465543

[10] T. G. Kurtz, Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes, Journal of Applied Probability 7 (1970) 49–58.

[11] J. L. Boudec, D. D. McDonald, J. Mundinger, A generic mean field convergence result for systems of interacting objects, in: Fourth International Conference on the Quantitative Evaluation of Systems (QEST 2007), 17-19 September 2007, Edinburgh, Scotland, UK, IEEE Computer Society, 2007, pp. 3–18. doi:10.1109/QEST.2007.8.
URL https://doi.org/10.1109/QEST.2007.8

[12] M. Benaïm, J. L. Boudec, A class of mean field interaction models for computer and communication systems, Perform. Eval. 65 (11-12) (2008) 823–838. doi:10.1016/j.peva.2008.03.005.
URL https://doi.org/10.1016/j.peva.2008.03.005

[13] N. Gast, B. Gaujal, Markov chains with discontinuous drifts have differential inclusion limits, Perform. Eval. 69 (12) (2012) 623–642. doi:10.1016/j.peva.2012.07.003.
URL https://doi.org/10.1016/j.peva.2012.07.003

[14] L. Bortolussi, J. Hillston, D. Latella, M. Massink, Continuous approximation of collective system behaviour: A tutorial, Perform. Eval. 70 (5) (2013) 317–349. doi:10.1016/j.peva.2013.01.001.
URL https://doi.org/10.1016/j.peva.2013.01.001

[15] N. Gast, B. Gaujal, A mean field approach for optimization in discrete time, Discrete Event Dynamic Systems 21 (1) (2011) 63–101. doi:10.1007/s10626-010-0094-3.
URL https://doi.org/10.1007/s10626-010-0094-3

[16] P. Tinnakornsrisuphap, A. M. Makowski, Limit behavior of ECN/RED gateways under a large number of TCP flows, in: Proceedings IEEE INFOCOM 2003, The 22nd Annual Joint Conference of the IEEE Computer and Communications Societies, San Franciso, CA, USA, March 30 - April 3, 2003, IEEE, 2003, pp. 873–883. doi:10.1109/INFCOM.2003.1208925.
URL https://doi.org/10.1109/INFCOM.2003.1208925

[17] N. Gast, B. Van Houdt, A refined mean field approximation, Proc. ACM Meas. Anal. Comput. Syst. 1 (2) (2017) 33:1–33:28. doi:10.1145/3154491.
URL http://doi.acm.org/10.1145/3154491

[18] D. Latella, M. Loreti, M. Massink, On-the-fly PCTL fast mean-field approximated model-checking for self-organising coordination, Sci. Comput. Program. 110 (2015) 23–50. doi:10.1016/j.scico.2015.06.009.
URL https://doi.org/10.1016/j.scico.2015.06.009

[19] N. Gast, Expected values estimated via mean-field approximation are 1/n-accurate: Extended abstract, in: B. E. Hajek, S. Oh, A. Chaintreau, L. Golubchik, Z. Zhang (Eds.), Proceedings of the 2017 ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems, Urbana-Champaign, IL, USA, June 05 - 09, 2017, ACM, 2017, p. 50. doi:10.1145/3078505.3078523.
URL http://doi.acm.org/10.1145/3078505.3078523

[20] L. Ying, On the approximation error of mean-field models, in: Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science, SIGMETRICS '16, ACM, New York, NY, USA, 2016, pp. 285–297. doi:10.1145/2896377.2901463.
URL http://doi.acm.org/10.1145/2896377.2901463

[21] L. Ying, Stein's method for mean field approximations in light and heavy traffic regimes, POMACS 1 (1) (2017) 12:1–12:27. doi:10.1145/3084449.
URL http://doi.acm.org/10.1145/3084449

[22] V. N. Kolokoltsov, J. Li, W. Yang, Mean Field Games and Nonlinear Markov Processes, ArXiv e-printsarXiv:1112.3744.

[23] C. Stein, Approximate computation of expectations, Lecture Notes-Monograph Series 7 (1986) i–164.
URL http://www.jstor.org/stable/4355512

[24] A. Braverman, J. Dai, Stein's method for steady-state diffusion approximations of $M/Ph/n + M$ systems, The Annals of Applied Probability 27 (1) (2017) 550–581.

[25] A. Braverman, J. G. Dai, J. Feng, Stein's method for steady-state diffusion approximations: An introduction through the erlang-a and erlang-c models, Stochastic Systems 6 (2) (2016) 301–366. arXiv:https://doi.org/10.1287/15-SSY212, doi:10.1287/15-SSY212.
URL https://doi.org/10.1287/15-SSY212

[26] L. Bortolussi, R. A. Hayden, Bounds on the deviation of discrete-time markov chains from their mean-field model, Perform. Eval. 70 (10) (2013) 736–749. doi:10.1016/j.peva.2013.08.012.
URL https://doi.org/10.1016/j.peva.2013.08.012

[27] M. A. M. de Oca, E. Ferrante, A. Scheidler, C. Pinciroli, M. Birattari, M. Dorigo, Majority-rule opinion dynamics with differential latency: a mechanism for self-organized collective decision-making, Swarm Intelligence 5 (3-4) (2011) 305–327. doi:10.1007/s11721-011-0062-z.
URL https://doi.org/10.1007/s11721-011-0062-z

[28] A. Scheidler, Dynamics of majority rule with differential latencies, Phys. Rev. E 83 (2011) 031116–1 – 031116–4. doi:10.1103/PhysRevE.83.031116.
URL https://link.aps.org/doi/10.1103/PhysRevE.83.031116