Bounds on the deviation of discrete-time Markov chains from their mean-field model

Luca Bortolussi[a,b], Richard A. Hayden[c]

[a]*Dept. of Mathematics and Geosciences, University of Trieste, Italy*
[b]*ISTI - CNR, Via Moruzzi 1, Pisa, Italy*
[c]*Dept. of Computing, Imperial College London, Huxley Building, 180 Queen's Gate, London SW7 2BZ, United Kingdom*

Abstract

We consider a generic mean-field scenario, in which a sequence of population models, described by discrete-time Markov chains (DTMCs), converges to a deterministic limit in discrete time. Under the assumption that the limit has a globally attracting equilibrium, the steady states of the sequence of DTMC models converge to the point-mass distribution concentrated on this equilibrium. In this paper we provide explicit bounds in probability for the convergence of such steady states, combining stochastic bounds on the local error with control-theoretic tools used in the stability analysis of perturbed dynamical systems to bound the global accumulation of error. We also adapt this method to compute bounds on the transient dynamics. The approach is illustrated by a wireless sensor network example.

---

1. Introduction

Stochastic modelling is a well established approach to reasoning quantitatively about a wide variety of computer systems. Owing to their relative ease of tractability, Markov processes have played a prominent role in this context, having been employed for the analysis of many systems, ranging from queueing models [1] to the spread of worm epidemics in a computer network [2]. These models usually fall into the class of population Markov processes, which consist of many similar entities or agents interacting within a common environment [3]. However, the computational tractability of their analysis becomes increasingly problematic as the number of agents to be modelled rises; this affects most discrete-state representations of large-scale systems. One way to tackle this problem is to approximate the limiting behaviour of a population Markov process with a deterministic dynamical system, either in discrete time, by a discrete-time dynamical system (DTDS) [4], or in continuous time, by a set of ordinary differential equations (ODEs) [3, 5].

Mean-field approximation is grounded on theoretical results that guarantee, under mild hypotheses, convergence of the Markov process to its mean-field model when the size of the population of agents goes to infinity [4, 3, 5]. This convergence generally holds only in the transient regime, but can be extended to steady state under some additional regularity assumptions on the limit system. Essentially, convergence to steady state holds when the limit system converges globally to a unique fixed point starting from any initial condition, as time goes to infinity [6, 7].

Mean-field approximation, both in its discrete-time and continuous-time variants, has been widely used recently to analyse different kind of computer systems, including MAC protocols [8], file streaming [9], network epidemics [2], gossip protocols [10], and swarm robotics [11]. Mean field models can also be obtained from high level descriptions of systems, for instance in terms of stochastic process algebras [12, 13]. Furthermore, mean-field techniques have been applied for the computation of passage times [14, 15] and to stochastic model checking [16].

An invaluable companion of mean-field approaches are error bounds, giving guaranteed estimates of the deviation of the stochastic process from its deterministic limit. The best known general bounds [5] are based on a representation of the Markov process as a DTDS or as an ODE (in integral form), perturbed by a noise term, which turns out to be a martingale. Then, one uses exponential martingale inequalities and some functional analysis inequality (usually the Grönwall inequality [17]) to obtain bounds (in probability) on the distance from the mean-field limit. However, the bounds obtained in this way are quite loose, and, furthermore are restricted to the transient behaviour, as they expand exponentially with the time interval considered. The very slack nature of such bounds is due directly to the Grönwall inequality, which ignores the topology of the particular phase space under consideration, instead relying on the general fact that for Lipschitz vector fields, nearby trajectories can separate, at worst, exponentially fast. An alternative approach to obtain very slack steady-state bounds is to set to zero the drift of a suitable Lyapunov function evaluated on the Markov chain and use simple bounds on its magnitude inside and outside of a region of interest [e.g. 18, 19]. The looseness here is due to the fact that this fairly crude technique generally does not improve as $N \to \infty$ and thus does not even agree qualitatively with the mean-field limit in that it cannot be used directly to verify asymptotic concentration of measure on the mean-field fixed point.

In general, practical case studies show a much better behaviour than existing error bounds in the literature would suggest, particularly when mean-field convergence can be extended to steady state. In this case, in fact, the limit system converges to a unique fixed point, hence trajectories tend to converge exponentially fast, rather than diverge, as time increases.

The main observation behind this paper is that error bounds can be greatly improved if this information about the topology of the phase space of the limit system is taken into account. In order to do this, we will exploit the previously mentioned characterisation of Markov processes as perturbed dynamical systems, combined with analytical and numerical techniques coming from control theory. Essentially, the crucial step is to abstract from the precise stochastic nature of the perturbation term in those equations, using the martingale bounds to replace it by a non-deterministic perturbation. This has the effect of combining bounds on noise with a fine-grained treatment of the phase space of the limit model. Hence, we expect not only to obtain tighter bounds on the transient, but also to be able to construct bounds on the steady state by using ideas of stability in control theory [20]. This allows us to obtain bounds on the convergence of the steady state of the Markov process to its limit distribution given by the mean-field model, which, to our knowledge, have never been obtained in general.

In this paper, we begin in this direction, by considering steady-state and transient bounds for discrete-time Markov chains converging to a discrete-time dynamical system [4]. As we will see, a combination of the ergodic theorem, martingale bounds and input-to-state stability (ISS) [21] or numerical reachability algorithms [22, 23, 24] will allow us to obtain general bounds. We will also discuss some examples, and illustrate how to compute these bounds in practice.

Summarising, the contributions of the paper are the following:

- We introduce a general method to obtain bounds for the steady-state convergence, for discrete-time mean-field models, proving its correctness;
- We adapt this method to compute bounds on the transient behaviour;
- We discuss how to compute in practice these bounds, exploiting numerical methods;
- We illustrate the techniques on a wireless sensor network model showing how bounds on useful response-time measures can be obtained.

The paper is organised as follows. In Section 2, we introduce the notation, the class of models we consider, and the mean-field results in discrete time. We also provide a new proof of convergence to steady state. In Section 3, we introduce the core of the method, obtaining the steady-state bound in terms of a property of the stochastic system (a bound on large deviations of the noise term) and of a property of the limit model (the number of steps required to return to a neighbourhood of the fixed point). These two quantities are then discussed separately, in Sections 3.1 and 4, respectively. In particular, in Section 4.1, we focus on the computation of the return time using a method based on Lyapunov functions, and in Section 4.2 we discuss a numerical approach based on the computation of reachable sets. Finally, in Section 6 we discuss bounds for the transient dynamics and in Section 7 we detail future work and final conclusions.

2. Mean field in discrete time

The population Markov model we consider in this paper consists of N interacting objects or agents, each taking values in a finite local state space $\mathcal{S} := \{1, \ldots, S\}$. Time is discrete, namely $t \in \mathbb{Z}_+$, and the state of agent j at time t is given by $X_j^{(N)}(t) \in \mathcal{S}$. The global state of the system is described by the *occupancy measure* $\mathbf{M}^{(N)}(t)$, which is a stochastic process taking values in the unit S-simplex $\Delta^S := \{\mathbf{m} \in \mathbb{R}_+^S : \sum_{i=1}^S m_i = 1\}$ given by $M_i^{(N)}(t) := \frac{1}{N}\sum_{n=1}^N \mathbf{1}_{\{X_n^{(N)}(t)=i\}}$.

The interacting agents in the system evolve synchronously at each step. More specifically, each agent changes state with a transition probability that depends on its local state and on the global state of the system via its occupancy measure $\mathbf{M}^{(N)}(t)$. Marginal state transition probabilities are thus:

$$K_{ij}^{(N)}(\mathbf{m}) := \mathbb{P}\{X_n^{(N)}(t+1) = j \,|\, \mathbf{M}^{(N)}(t) = \mathbf{m}, X_n^{(N)}(t) = i\}$$

and are the same for each object. Let $\mathbf{K}^{(N)}(\mathbf{m})$ be the $S \times S$ state-transition matrix for an individual object.

In order to specify the evolution at the system level, we introduce, for each $t \in \mathbb{Z}_+$, mutually independent random variables $\mathbf{Z}_i^{(N)}(t)$, $i = 1, \ldots, S$ which are distributed according to a multinomial distribution on $\mathcal{S}$ with number of trials $N \cdot M_i^{(N)}(t)$ and probabilities $(K_{i1}^{(N)}(\mathbf{M}^{(N)}(t)), \ldots, K_{iS}^{(N)}(\mathbf{M}^{(N)}(t)))$. Then, we may define:

$$\mathbf{M}^{(N)}(t+1) = \frac{1}{N}\sum_{i=1}^{S} \mathbf{Z}_i^{(N)}(t) \tag{1}$$

If we let the number of objects N go to infinity, we obtain that the occupancy measure $\mathbf{M}^{(N)}(t)$ converges, for each $t \in \mathbb{Z}_+$, to a deterministic limit process $\mu(t)$, namely a discrete-time dynamical system which approximates the behaviour of the Markov chain for large populations, in the sense explained in Theorem 1 below, defined by the mean-field equations:

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

In the previous equation, $\mathbf{K}(\mathbf{m})$ is the N-independent continuous limit of the transition matrices $\mathbf{K}^{(N)}(\mathbf{m})$, which may depend on N. We require that this convergence is uniform on Δ^S. This is the setting considered originally in [4]. In the rest of the paper, we assume that these conditions are in force. Indeed, they are usually satisfied by practical models, and are trivially true for the examples we consider (for which $\mathbf{K}^{(N)}(\mathbf{m})$ does not depend on N). We now give the formal transient convergence theorem in a slightly more general form than given originally in [4], with the proof given in Appendix A.1.1. We require the extra generality for the proof of convergence in steady state which will follow after.

Theorem 1. *[Mean field in discrete time (transient regime)] Under the previous hypotheses, and assuming that $\mathbf{M}^{(N)}(0) \Rightarrow \mu(0)$ as $N \to \infty$ weakly in distribution on $\mathbb{R}^S$, the same weak convergence $\mathbf{M}^{(N)}(t) \Rightarrow \mu(t)$ holds for each $t \geq 0$, where $\mu(t)$ is the* stochastic process *defined by $\mu(t+1) = \mu(t)\mathbf{K}(\mu(t))$.*[1]

Example. We illustrate the mean-field approach by means of a simple example of a wireless sensor network. We consider a network composed of two component types: wireless sensor nodes and gateways. The local states of the model are $\mathcal{S} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}\}$, where $\mathtt{a}, \mathtt{b}$ are the states of the gateways and $\mathtt{c}, \mathtt{d}, \mathtt{e}$ are the states of the wireless sensors. There are N components in total, with a fixed fraction m_G of them being gateway nodes, while the fraction of wireless sensors is $m_W = 1 - m_G$. Wireless sensor nodes begin in state `e` where they may detect an event of interest, after which they move into state `c`. Here they attempt to communicate observed data, but may give up after timing out, moving to state `d`. If they timeout, they try again to transmit the data after a delay. Communication of data requires a synchronization of a sensor with

[1] Additionally, where the initial convergence $\mathbf{M}^{(N)}(0) \to \mu(0)$ is almost sure, the same can be said of the convergence at all time points when both processes are constructed on the same probability space.

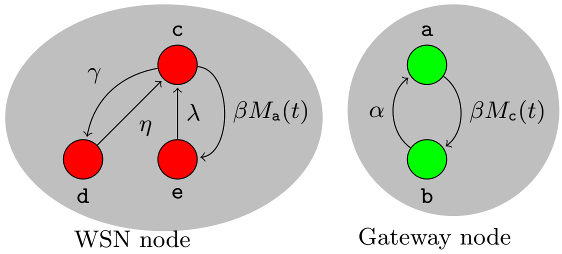


Figure 1: Representation of the behaviour of wireless sensor network (WSN) nodes and gateway nodes in the wireless sensor network model.

a gateway node. A gateway node is ready to receive data when it is in state a, and, once it has received a certain amount of data from wireless sensor nodes, it enters state b where it is temporarily unable to receive further communication from wireless nodes, reflecting a processing delay. The local behaviour of both types of nodes is depicted graphically in Figure 1. We write $\mathbf{M}(t) = (M_{\mathtt{a}}(t), M_{\mathtt{b}}(t), M_{\mathtt{c}}(t), M_{\mathtt{d}}(t), M_{\mathtt{e}}(t))$ for the occupancy measure and the local transition matrix for an individual component is:

$$\mathbf{K}(\mathbf{m}) = \begin{pmatrix} 1-\beta m_{\mathtt{c}} & \beta m_{\mathtt{c}} & 0 & 0 & 0 \\ \alpha & 1-\alpha & 0 & 0 & 0 \\ 0 & 0 & 1-\gamma-\beta m_{\mathtt{a}} & \gamma & \beta m_{\mathtt{a}} \\ 0 & 0 & \eta & 1-\eta & 0 \\ 0 & 0 & \lambda & 0 & 1-\lambda \end{pmatrix} \tag{3}$$

For $\mu(t) = (\mu_{\mathtt{a}}(t), \mu_{\mathtt{b}}(t), \mu_{\mathtt{c}}(t), \mu_{\mathtt{d}}(t), \mu_{\mathtt{e}}(t))$, we obtain the *non-linear* mean-field equations $\mu(t+1) = \mu(t)\mathbf{K}(\mu(t))$.

Convergence to Steady State. We assume in this section deterministic initial conditions $X_n^{(N)}(0)$ inducing a reachable state space for $\mathbf{M}^{(N)}(t)$, say $\mathcal{M}^{(N)} \subset \Delta^S$, where $\mathcal{M}^{(N)}$ is assumed irreducible with each state aperiodic so that each Markov chain is ergodic and thus has a unique stationary distribution, which we denote by the random variable $\mathbf{M}^{(N)}(\infty)$.

Theorem 1 guarantees convergence for any finite step t, but does not necessarily imply convergence of the steady state measures. Mean-field convergence of $\mathbf{M}^{(N)}(\infty)$ to the limit point(s) of the mean-field equation holds only under additional assumptions on the mean-field dynamical system: we consider here the case where the mean-field dynamical system has a unique globally attracting fixed point within some compact set $K \subseteq \Delta^S$ for which $\mathcal{M}^{(N)} \subseteq K$ for each N. This means that there is an $\mathbf{m}^* \in K$ such that, for all fixed $\mu(0) \in K$, $\lim_{t\to\infty} \mu(t) = \mathbf{m}^*$. If this holds, then we have the following theorem, proved in Appendix A.1.2.

Theorem 2. *[Mean field in discrete time (stationary regime)] Under the previous hypotheses, we have* $\mathbf{M}^{(N)}(\infty) \Rightarrow \delta_{\mathbf{m}^*}$*, where* $\delta_{\mathbf{m}^*}$ *is the point mass at* $\mathbf{m}^*$*.*

In the rest of the paper, we will consider only models that satisfy the conditions for the convergence of steady-state measures, providing bounds from the distance of $\mathbf{M}^{(N)}(\infty)$ from $\delta_{\mathbf{m}^*}$, for a *fixed* N. For this reason, we will *drop the superscript* N in order to lighten the notation.

3. Derivation of the steady-state error bound

In this section we present the derivation of the error bound for the distance of the steady state $\mathbf{M}(\infty)$ from $\mathbf{m}^*$. In particular, we will derive a bound for the probability $\mathbb{P}\{\mathbf{M}(\infty) \in \mathcal{E}\}$, where $\mathcal{E}$ is a fixed neighbourhood of $\mathbf{m}^*$, usually a ball centred in $\mathbf{m}^*$ of a given radius ϵ.

The first step in the derivation is to rewrite the formula (1), for $t \in \mathbb{Z}_+$, as:

$$\mathbf{M}(t+1) = \mathbf{M}(t)\mathbf{K}(\mathbf{M}(t)) + \mathbf{D}(t+1) \tag{4}$$

where we define the *local noise* term:

$$\mathbf{D}(t+1) := \mathbf{M}(t+1) - \mathbf{M}(t)\mathbf{K}(\mathbf{M}(t)) \tag{5}$$

which is the stochastic process that captures the error introduced by the mean-field approximation at each time step. We set $\mathbf{D}(0) := \mathbf{0}$ and, as before, assume deterministic initial conditions $X_n(0)$ inducing a reachable state space for $\mathbf{M}(t)$, say $\mathcal{M} \subset \Delta^S$, where $\mathcal{M}$ is also finite and assumed irreducible with each state aperiodic so that the Markov chain is ergodic and thus has a unique stationary distribution, which we denote by the random variable $\mathbf{M}(\infty)$.[2]

It is clear that due to the fixed total population size in the models we are considering here, there will be at least one population conservation law and thus one or more redundant population variables which can be recovered from the others. It is desirable to work with the reduced version of the model equations such that all such redundant population variables have been eliminated. For the occupancy measure of the reduced system we write the vector $\mathbf{M}'(t)$ whose elements are a subset of those of $\mathbf{M}(t)$ tracking the occupancy measures for a defining subset $\mathcal{S}' \subseteq \mathcal{S}$ of the local states. We also write $\mathcal{M}' \subset \mathbb{R}_+^{|\mathcal{S}'|}$ for the reduced reachable global state space.

For instance, in the example of Section 2, we have two populations, gateways and wireless sensors, whose total number remains constant. This gives two conservation laws, $\mu_{\mathtt{b}}(t) = m_G - \mu_{\mathtt{a}}(t)$ for gateways and $\mu_{\mathtt{e}}(t) = m_W - \mu_{\mathtt{c}}(t) - \mu_{\mathtt{d}}(t)$ for wireless sensors, that can be used to eliminate variables $\mu_{\mathtt{b}}$ and $\mu_{\mathtt{e}}$, obtaining the reduced local state space $\mathcal{S}' = \{\mathtt{a}, \mathtt{c}, \mathtt{d}\}$.

In the reduced case, the analogues of Eqs. (2), (4) and (5) are, respectively:

$$\begin{aligned} \mu'(t+1) &= \mathbf{f}(\mu'(t)) \\ \mathbf{M}'(t+1) &= \mathbf{f}(\mathbf{M}'(t)) + \mathbf{D}'(t+1) \\ \mathbf{D}'(t+1) &= \mathbf{M}'(t+1) - \mathbf{f}(\mathbf{M}'(t)) \end{aligned}$$

where $\mathbf{f} : \mathbb{R}_+^{|\mathcal{S}'|} \to \mathbb{R}_+^{|\mathcal{S}'|}$ is such that $\mathbf{f}(\mathbf{m}')$ captures $\mathbf{m}\,\mathbf{K}(\mathbf{m})$ for the components of the reduced system (i.e. we eliminate the equations for the variables not in the reduced system from $\mathbf{m}\,\mathbf{K}(\mathbf{m})$, and use the conservation laws to remove those variables from the remaining equations).

Let $\mathcal{E} \subseteq \mathcal{M}'$ be the subset of the reduced state space in which we wish to bound the steady-state probability mass, with $\mathbf{m}^* \in \mathcal{E}$, so that we are aiming to bound $\mathbb{P}\{\mathbf{M}'(\infty) \in \mathcal{E}\}$ from below. Fix a *local noise threshold* $d > 0$ and define the *deviation times* $\{T_k\}_{k=0}^{\infty}$ by $T_0 := 0$ and $T_k := \inf\{t > T_{k-1} : \|\mathbf{D}'(t)\| > d\}$. We assume that d has been chosen such that $T_k < \infty$ almost surely for all $k \in \mathbb{Z}_+$; equivalently, there must exist some state $\mathbf{m}' \in \mathcal{M}'$ such that given $\mathbf{M}'(t) = \mathbf{m}'$, there is a non-zero probability that $\|\mathbf{D}'(t+1)\| > d$.

The computation of the bound will rely on being able to compute a key estimate $w(d) \in \mathbb{Z}_+$ on the time taken to return to $\mathcal{E}$ after a deviation (the *return time*), such that for all $k \in \mathbb{Z}_+$:

$$\mathbb{P}\{\mathbf{M}'(T_k + l) \in \mathcal{E} \text{ for all } w(d) \leq l < T_{k+1} - T_k\} = 1 \tag{6}$$

We return to considering methods for computing $w(d)$ shortly. Note, however, that if $\mathcal{E}$ is not a neighbourhood of $\mathbf{m}^*$ or if $\mathcal{E}$ is too small or d is too large, it is possible that no such finite $w(d)$ exists, in which case, it would be necessary to either enlarge $\mathcal{E}$ or decrease d. This will become clear when we consider methods for computing $w(d)$ and illustrated in terms of the example in Section 5. For now we assume that such a $w(d)$ can be found for our fixed $\mathcal{E}$ and d.

By the ergodic theorem for discrete-time Markov chains, we have, almost surely:

$$\mathbb{P}\{\mathbf{M}'(\infty) \in \mathcal{E}\} = \lim_{k\to\infty} \frac{1}{k} \sum_{l=0}^{k-1} \mathbf{1}_{\{\mathbf{M}'(l)\in\mathcal{E}\}} = \limsup_{k\to\infty} \frac{1}{k} \sum_{l=0}^{k-1} \mathbf{1}_{\{\mathbf{M}'(l)\in\mathcal{E}\}}$$

[2]Since we are concerned here with bounds on the steady-state error, we need not consider probabilistic initial conditions.

Now taking the limit along the subsequence T_j, we have:

$$\begin{aligned}\mathbb{P}\{\mathbf{M}'(\infty)\in\mathcal{E}\} &= \limsup_{k\to\infty}\frac{1}{T_k+1}\sum_{l=0}^{T_k}\mathbf{1}_{\{\mathbf{M}'(l)\in\mathcal{E}\}} \quad\geq\quad \limsup_{k\to\infty}\frac{1}{T_k+1}\sum_{l=0}^{k-1}(T_{l+1}-T_l-w(d))\vee 0\\ &\geq \limsup_{k\to\infty}\frac{1}{T_k+1}\sum_{l=0}^{k-1}(T_{l+1}-T_l-w(d)) \quad=\quad \limsup_{k\to\infty}\frac{1}{T_k+1}(T_k-kw(d))\\ &\geq 1-w(d)\limsup_{k\to\infty}\frac{k}{T_k}\end{aligned}\tag{7}$$

In the expression above, the first inequality follows because the system is by definition in $\mathcal{E}$ when $t\in[T_l+w(d),T_{l+1})$ for all $l\geq 0$. The second equality is obtained by expanding the summation, while the final inequality follows from standard properties of lim sup and the fact that the sequence T_k is divergent. In order to bound $\limsup_{k\to\infty}\frac{k}{T_k}$, we define the indicator process $A(t):=\mathbf{1}_{\{\|\mathbf{D}'(t)\|>d\}}$ and observe that:

$$\limsup_{k\to\infty}\frac{k}{T_k}=\limsup_{k\to\infty}\frac{1}{T_k}\sum_{t=1}^{T_k}A(t)=\limsup_{l\to\infty}\frac{1}{l}\sum_{t=1}^{l}A(t)$$

Now for $\mathbf{m}'\in\mathcal{M}'$, define $\rho_{\mathbf{m}'}(d):=\mathbb{P}\{\|\mathbf{D}'(t+1)\|>d\,|\,\mathbf{M}'(t)=\mathbf{m}'\}$, which is independent of t. Then, by adding and subtracting $\rho_{\mathbf{M}'(t-1)}(d)$ from each addend, we may write:

$$\limsup_{l\to\infty}\frac{1}{l}\sum_{t=1}^{l}A(t)\leq\limsup_{l\to\infty}\frac{1}{l}\sum_{t=1}^{l}(A(t)-\rho_{\mathbf{M}'(t-1)}(d))+\limsup_{l\to\infty}\frac{1}{l}\sum_{t=1}^{l}\rho_{\mathbf{M}'(t-1)}(d)\tag{8}$$

Now the process $Z(t):=A(t)-\rho_{\mathbf{M}'(t-1)}(d)$ for $t>0$ and $Z(0):=0$ is a martingale difference sequence with respect to $\mathbf{M}(t)$ since for any $t>0$:

$$\mathbb{E}[Z(t)|(\mathbf{M}'(s))_{s<t}]=\mathbb{E}[Z(t)|\mathbf{M}'(t-1)]=\mathbb{E}[A(t)|\mathbf{M}'(t-1)]-\rho_{\mathbf{M}'(t-1)}(d)=0$$

So we may apply the strong law of large numbers for discrete-time martingales [e.g. 25, pg. 238] to show that the first term on the right-hand side of Eq. (8) is zero almost surely and obtain:

$$\limsup_{l\to\infty}\frac{1}{l}\sum_{t=1}^{l}A(t)\leq\limsup_{l\to\infty}\frac{1}{l}\sum_{t=1}^{l}\rho_{\mathbf{M}'(t-1)}(d)\leq\rho(d)$$

where $\rho(d):=\max_{\mathbf{m}'\in\mathcal{M}'}\rho_{\mathbf{m}'}(d)$, which, returning to Eq. (7), gives:

$$\mathbb{P}\{\mathbf{M}'(\infty)\in\mathcal{E}\}\geq 1-w(d)\rho(d)$$

3.1. Bounding $\rho(d)$

Fix $\mathbf{m}\in\mathcal{M}$ and assume that $\mathbf{M}(t)=\mathbf{m}$. For each $i\in S$ and $1\leq n\leq Nm_i$, let P_i^n be mutually-independent random variables whose distribution on $j\in\{1,\ldots,S\}$ is given by $K_{ij}(\mathbf{m})$. Then conditioned on $\mathbf{M}(t)=\mathbf{m}$, $D_j(t+1)$ is equal in distribution to:

$$\frac{1}{N}\sum_{i=1}^{S}\sum_{n=1}^{Nm_i}\mathbf{1}_{\{P_i^n=j\}}-\sum_{i=1}^{S}m_iK_{ij}(\mathbf{m})$$

Let σ^2 be such that $K_{ij}(\mathbf{m})(1-K_{ij}(\mathbf{m}))\leq\sigma^2$ for all $i,j\in\{1,\ldots,S\}$ and $\mathbf{m}\in\mathcal{M}$. Note that we can always take $\sigma^2=0.25$ but, where the model allows, smaller values of σ^2 will result in tighter bounds on $\rho(d)$. An application of a Chernoff inequality [e.g. 26, Theorem 2.6] gives:

$$\mathbb{P}\{|D_j(t+1)| > d \,|\, \mathbf{M}(t) = \mathbf{m}\} \leq 2\exp\left(-\frac{Nd^2}{2(\sigma^2 + d/3)}\right)$$

and thus:

$$\begin{aligned}\mathbb{P}\{\|\mathbf{D}'(t+1)\| > d \,|\, \mathbf{M}'(t) = \mathbf{m}'\} &\leq \mathbb{P}\{\sqrt{|\mathcal{S}'|}\|\mathbf{D}'(t+1)\|_\infty > d \,|\, \mathbf{M}'(t) = \mathbf{m}'\} \\ &\leq 2|\mathcal{S}'|\exp\left(-\frac{Nd^2}{2|\mathcal{S}'|\left(\sigma^2 + d/\sqrt{9|\mathcal{S}'|}\right)}\right)\end{aligned}$$

giving $\rho(d) \leq 2|\mathcal{S}'|\exp\left(-\frac{Nd^2}{2|\mathcal{S}'|\left(\sigma^2+d/\sqrt{9|\mathcal{S}'|}\right)}\right)$.

4. Bounding the return time

In order to compute an appropriate value for the return time $w(d)$, we will consider the perturbed dynamical system (in the sense of [21]):

$$\mu'(t+1) = \mathbf{f}(\mu'(t)) + \mathbf{w}(t) \tag{9}$$

where $\mathbf{w}(t)$ is a non-deterministic input signal. In our situation we will assume that the initial condition $\mu'(0)$ is an arbitrary point in $\mathcal{M}'$ and that $\|\mathbf{w}(t)\| \leq d$ for all $t \in \mathbb{Z}_+$. We may also assume that $\mu(t) \in \mathcal{M}'$ for all $t \in \mathbb{Z}_+$. Our aim is then to find some $w(d) \in \mathbb{Z}_+$ such that $\mu'(k) \in \mathcal{E}$ for all $k \geq w(d)$. It is then straightforward to see that such a $w(d)$ would satisfy Eq. (6). For ease of notation, we assume that the system and state space has been translated relative to $\mathbf{m}^*$ such that $\mathcal{E} = \{\mathbf{m}' \in \mathcal{M}' : \|\mathbf{m}'\| \leq \varepsilon\}$ for some given $\varepsilon > 0$.

In general, bounding the solutions of such systems is a very difficult problem. However, the problem becomes much simpler when the function $\mathbf{f}$ is Lipschitz continuous with Lipschitz constant $L < 1$. In this case, in fact, we have that

$$\|\mu'(t)\| \leq L\|\mu'(t-1)\| + \|\mathbf{w}(t-1)\| \leq L^t\|\mu'(0)\| + d\left(\frac{1-L^t}{1-L}\right) \leq L^t\|\mu'(0)\| + \frac{d}{1-L}, \tag{10}$$

where the second inequality follows by recursively applying the first one to $\|\mu'(t-1)\|$. We then see that $\frac{d}{1-L}$ is an upper bound on the perturbation induced by the noise, while the effect of the initial distance from equilibrium vanishes as $t \to \infty$. From such an expression and the fact that $\|\mu'(0)\| \leq \sqrt{S}$ in $\mathcal{M}'$, $w(d)$ is easily computed as

$$w(d) = \left\lceil \frac{1}{\log(L)} \log\left(\frac{1}{\sqrt{S}}\left(\epsilon - \frac{d}{1-L}\right)\right)\right\rceil, \tag{11}$$

provided $\frac{d}{1-L} < \epsilon$.

Unfortunately, the condition on the Lipschitz constant L of being less than one is not satisfied by many models, and is also dependent on the specific value of model parameters. However, note that when this condition is true, the function $V(\mathbf{x}) = \|\mathbf{x}\|$, which is always non-negative, and null only in zero, is decreasing along $\mathbf{f}$, i.e. $V(\mathbf{f}(\mathbf{x})) < V(\mathbf{x})$. This makes it a *Lyapunov* function for the un-perturbed discrete-time dynamical system. This suggests that Lyapunov functions, behaving similarly to equation (10) with respect to the perturbation term, may give a more general criterion to estimate $w(d)$. This is indeed the case, using a theory similar to that of *input-to-state stability* [21].

4.1. Bounding $w(d)$ using ISS-Lyapunov functions

In this section we assume that $\mathbf{f}$ is a general, potentially non-linear function and we wish to bound solutions of Eq. (9) assuming that $\mu'(0) \in \mathcal{M}'$ is arbitrary; $\mu'(t) \in \mathcal{M}'$ and $\|\mathbf{w}(t)\| \leq d$ for all $t \in \mathbb{Z}_+$.

Let $\mathcal{K}$ be the set of all functions $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ that are continuous, strictly increasing and have $\gamma(0) = 0$; and $\mathcal{K}_\infty$ the set of all functions $\gamma \in \mathcal{K}$ with the additional property $\gamma(s) \to \infty$ as $s \to \infty$. Then assume there exists a continuous function $V : \mathcal{M}' \to \mathbb{R}_+$ corresponding to the system of Eq. (9) where the following holds:

- There exist $\alpha_1, \alpha_2 \in \mathcal{K}_\infty$ such that for all $\mathbf{m}' \in \mathcal{M}'$:

$$\alpha_1(\|\mathbf{m}'\|) \leq V(\mathbf{m}') \leq \alpha_2(\|\mathbf{m}'\|)$$

- There exists a function $\alpha_3 \in \mathcal{K}_\infty$ and a constant $Z \in \mathbb{R}_+$ such that for all $\mathbf{m}' \in \mathcal{M}'$ and $\mathbf{w} \in \mathbb{R}^{|\mathcal{S}'|}$ with $\|\mathbf{w}\| \leq d$:

$$V(\mathbf{f}(\mathbf{m}') + \mathbf{w}) - V(\mathbf{m}') \leq -\alpha_3(\|\mathbf{m}'\|) + Z \tag{12}$$

The requirements on V here are very similar to those of an *ISS-Lyapunov function* in [21], adapted to our specific context. It can then be proved in a similar fashion to Lemma 3.5 of [21] that under the above assumptions, there exists a constant $W \in \mathbb{R}_+$ and a function $\beta : \mathbb{R}_+ \times \mathbb{Z}_+ \to \mathbb{R}_+$ such that:

$$\|\mu'(t)\| \leq \max\{\beta(\|\mu'(0)\|, t), W\} \tag{13}$$

where, for fixed $t \in \mathbb{Z}_+$, $\beta(\cdot, t)$ is non-decreasing and, for fixed $\mathbf{m}' \in \mathcal{M}'$, $\beta(\|\mathbf{m}'\|, \cdot)$ is decreasing with $\beta(\|\mathbf{m}'\|, t) \to 0$ as $t \to \infty$. Specific definitions of W and β can be taken as follows. Fix $Z' > Z$, and write:

$$\beta(r, t) := \alpha_1^{-1}(\bar{\beta}^t(\alpha_2(r)))$$
$$W := \alpha_1^{-1}(\alpha_4^{-1}(Z'))$$

where here $\bar{\beta}^t := \underbrace{\bar{\beta} \circ \bar{\beta} \circ \cdots \circ \bar{\beta}}_{t}$, $\alpha_4 \in \mathcal{K}_\infty$ such that for all $r \in \mathbb{R}_+$, $\alpha_4(r) \leq \alpha_3(\alpha_2^{-1}(r))$ and $\mathbf{1} - \alpha_4$ is non-decreasing;[3] and:

$$\bar{\beta}(r) := \sup_{s \in [0,r]} \{s - (1 - Z/Z')\alpha_4(s)\}$$

Choosing a different parameter Z' may result in improved bounds. A proof of this bound can be found in Appendix A.2.

For some special classes of dynamical system, a Lyapunov function V witnessing global asymptotic stability is known and the condition $V(\mathbf{f}(\mathbf{m}')+\mathbf{w})-V(\mathbf{m}') < 0$ for all $\mathbf{0} \neq \mathbf{m}' \in \mathcal{M}'$ can be proved algebraically without the need to resort to numerical methods. In such cases, it should be fairly straightforward to define appropriate α_3 and Z and to validate Eq. (12) algebraically. However, such cases are relatively rare and it is much more likely that the validation of Eq. (12) will need to be carried out numerically using an appropriate global optimisation technique.

More specifically, given a candidate Lyapunov function V and appropriate α_4 and Z, one can use a global optimization method (like the Global Optimization Toolbox of Matlab [27]) to check that the maximum in the bounded regions $\mathbf{m}' \in \mathcal{M}'$, $\|\mathbf{w}\| \leq d$ of the left hand side minus the right hand side of the ISS bound (13) is non-positive. The disadvantage of this numerical approach is that it can never provide a certificate of correctness that the function V is an ISS-Lyapunov function. However, this may be obtained, in case the ISS bound (13) is polynomial in $\mathbf{m}'$ and $\mathbf{w}$ (as for the worked example in the next section), by using methods

[3] If $\alpha_4 := \alpha_3 \circ \alpha_2^{-1}$ is not already such that $\mathbf{1} - \alpha_4$ is non-decreasing, a result such as [21, Lemma B.1] can always be used to construct a lower-bounding function for which this is true.

of computational algebraic geometry [28]. For instance, one may find the global maximum by looking at the set of zeros of the gradient of the the left hand side minus the right hand side of the ISS bound (13) and using the method of Lagrange multipliers to check for maxima in the boundary of $\mathcal{M}'$. While such a certified approach can be more computationally expensive, the key point to note is that the cost of this operation does not scale with the population size N as would be the case if the underlying Markov chain were to be solved explicitly for its stationary distribution.

4.2. Bounding $w(d)$ computationally

We consider also a third way to identify $w(d)$, which is based on numerical methods providing over-approximations of reachable sets. More precisely, consider again the perturbed system given by equation (9), and let $W_d = [-d, d]^{|\mathcal{S}'|}$, so that $\mathbf{w}(t) \in W_d$. Now, if we apply the function defined by equation (9) to a set $A \subset \mathcal{M}'$, we obtain the set

$$F_d(A) = (\mathbf{f}(A) \oplus W_d) \cap \mathcal{M}',$$

where $\oplus$ is the Minkowski sum[4]. We can then define the sequence of sets E_t as $E_0 = \mathcal{M}'$, the domain of the system, and $E_{t+1} = F_d(E_t)$. It is easy to see that the set E_t is the set of points that can be reached by the perturbed limit system at time t, starting from an arbitrary point of $\mathcal{M}'$. Given a target neighbourhood $\mathcal{E}$, then we can characterise $w(d)$ as the minimum time t_0 such that $\forall t \geq t_0$, $E_t \subset \mathcal{E}$.

Now, suppose we have a computational procedure that for each t computes an over-approximation $\hat{E}_t$ of E_t, i.e. $E_t \subseteq \hat{E}_t$, and that we find a time $\hat{t}$ such that $\forall t \geq \hat{t}$, $\hat{E}_t \subset \mathcal{E}$. It follows that $\hat{t} \geq w(d)$, and thus $\hat{t}$ provides an estimate of $w(d)$.

Computing over-approximations of reachable sets for non-linear functions is, in general, very challenging, yet the problem becomes tractable for multi-affine functions [22], i.e. polynomial functions that are affine in each variable. This class of functions covers not only the example of this paper, but the large class of models in which synchronisation between components is represented in a mass-action style, that is, it is proportional to the number, or density of some other component(s) in the model. A multi-affine function $\mathbf{g} : \mathbb{R}^S \to \mathbb{R}^S$ has the following property [22]: given a hyper-rectangle $R = \prod_{i=1}^{S}[l_i, u_i]$, $[l_i, u_i] \subset \mathbb{R}$, $\mathbf{g}(R)$ is the *convex hull* of the values of $\mathbf{g}$ on the vertices $V(R)$ of R. Therefore, we can construct a rectangle $bbox(\mathbf{g}, R)$ containing $\mathbf{g}(R)$ by taking, for each coordinate i, the maximum and the minimum of g_i, the i-th function of $\mathbf{g}$, over the vertices $V(R)$ of R. More precisely, we can write $bbox(\mathbf{g}, R) = \prod_{i=1}^{S} bbox_i(\mathbf{g}, R)$, where

$$bbox_i(\mathbf{g}, R) = \left[\min_{\mathbf{v}\in V(R)}\{g_i(\mathbf{v})\}, \max_{\mathbf{v}\in V(R)}\{g_i(\mathbf{v})\}\right].$$

Hence, for a multi-affine function $\mathbf{f}$, we can compute (a hyper-rectangle) $\hat{E}_t$ as

$$\hat{E}_t = bbox(\mathbf{f}, \hat{E}_{t-1}) \oplus W_d.$$

Then, to estimate an upper bound for $w(d)$, we just need to compute the sequence $\hat{E}_t$ until it stabilises to a fixed point,[5] then finding the time $\hat{t}$, if any, such that $\forall t \geq \hat{t}$, $\hat{E}_t \subset \mathcal{E}$.

This gives a practical and computationally efficient way to estimate $w(d)$ for multi-affine systems which does not require any knowledge of Lipschitz constants or Lyapunov functions, but that in general may provide worse estimates of $w(d)$ (or require smaller values of d) than Lyapunov-based methods, due to the over-approximation of the reachable set by a hyper-rectangle. Note that a better approximation can be obtained at a higher computational cost by covering the set $\mathbf{g}(R)$ with smaller hyper-rectangles, thus reducing the numerical error. This computational approach may also be applied to more complex functions, using more refined but computationally more expensive methods. For instance, for polynomial functions, one can use over-approximations based on Bernstein polynomials [23], while for general non-linear functions one may attempt to use hybridization-like approaches [24].

[4] The Minkowski sum of two subsets A and B of $\mathbb{R}^S$ is defined as $A \oplus B = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in A, \mathbf{y} \in B\}$.

[5] Practically, one checks that the distance between $\hat{E}_t$ and $\hat{E}_{t+1}$ becomes less that a predefined tolerance level.

Remark 1. If we are able to show numerically for the unperturbed system ($d = 0$) that $\hat{E}_t \subset \mathcal{E}$, $\forall t \geq \hat{t}$, then we also have a proof that the limit dynamical system will eventually enter the set $\mathcal{E}$ and remain there forever. Even if this is does not guarantee existence and uniqueness of a global fixed point, it can be used to ensure that each limit point of the sequence of steady-state measures for finite populations will be supported in $\mathcal{E}$, by a minor adaptation of Theorem 2. However, this is the only required condition to safely apply the steady-state bound of this paper, which can then be used for a larger class of models than those considered here (combined with appropriate numerical routines to compute tight over-approximations of $\hat{E}_t$).

5. Worked example

In this section we illustrate the complete approach for general non-linear mean-field dynamics by means of the wireless sensor network model introduced in Section 2. Recall that the state space of components is $\mathcal{S} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}\}$, hence the occupancy measure is $\mathbf{M}(t) = (M_\mathtt{a}(t), M_\mathtt{b}(t), M_\mathtt{c}(t), M_\mathtt{d}(t), M_\mathtt{e}(t))$, while the mean field equation is $\mu(t + 1) = \mu(t)\mathbf{K}(\mu(t))$, where $\mathbf{K}(\mu(t))$ is given by equation (3).

Since the model consists of two separate component populations (the WSN nodes and the gateway nodes), given the vector of initial occupancy measures $\mathbf{M}(0) = (m_G, 0, 0, 0, m_W)$ where $m_G + m_W = 1$ (to define the reachable state space), the mean-field equations can be reduced to an equivalent 3-dimensional system in terms of just $\mu_\mathtt{a}(t)$, $\mu_\mathtt{c}(t)$ and $\mu_\mathtt{d}(t)$. Furthermore, we assume that the reduced system has a unique fixed point, say μ^*, and define the re-centred quantities:

$$\mu'(t) = (\mu'_\mathtt{a}(t), \mu'_\mathtt{c}(t), \mu'_\mathtt{d}(t)) := (\mu_\mathtt{a}(t) - \mu^*_\mathtt{a}, \mu_\mathtt{c}(t) - \mu^*_\mathtt{c}, \mu_\mathtt{d}(t) - \mu^*_\mathtt{d})$$

It is straightforward then to see that $\mu'(t)$ satisfies $\mu'(t + 1) = \mathbf{f}(\mu'(t))$, where

$$\mathbf{f}(\mu'(t)) = \begin{pmatrix} \mu'_\mathtt{a}(t) - \beta(\mu'_\mathtt{c}(t)\mu'_\mathtt{a}(t) - \mu'_\mathtt{c}(t)\mu^*_\mathtt{a} - \mu'_\mathtt{a}(t)\mu^*_\mathtt{c}) - \alpha\mu'_\mathtt{a}(t) \\ \mu'_\mathtt{c}(t) - \gamma\mu'_\mathtt{c}(t) - \beta(\mu'_\mathtt{c}(t)\mu'_\mathtt{a}(t) - \mu'_\mathtt{c}(t)\mu^*_\mathtt{a} - \mu'_\mathtt{a}(t)\mu^*_\mathtt{c}) + \eta\mu'_\mathtt{d}(t) - \lambda(\mu'_\mathtt{c}(t) - \mu'_\mathtt{d}(t)) \\ \gamma\mu'_\mathtt{c}(t) + \mu'_\mathtt{d}(t) - \eta\mu'_\mathtt{d}(t) \end{pmatrix}^T$$

Fix $\mathcal{E} = \{\mathbf{m}' \in \mathcal{M}' : \|\mathbf{m}'\| \leq \epsilon\}$ for some $\epsilon > 0$ in terms of the re-centred state space and $d \in \mathbb{R}_+$. In order to compute a suitable $w(d)$, we can use one of the three approaches presented in Section 4.

The first attempt is to use the method based on the Lipschitz constant. For parameter values: $\alpha = 0.4$, $\beta = 0.6$, $\eta = 0.3$, $\lambda = 0.2$, $\gamma = 0.2$; and $m_G = 1/3$, $m_W = 2/3$, we find that the Lipschitz constant of the function $\mathbf{f}$ is $L \leq 0.852 < 1$, hence the method is applicable. The constant has been bounded numerically by maximising the norm of the Jacobian matrix of $\mathbf{f}'$. Letting $\epsilon = 0.1$, fixing the local noise level to $d = 0.0147$ and the confidence level δ for the error bound to 0.95, we obtain that $\mathbb{P}\{\mathbf{M}'(\infty) \in \mathcal{E}\} \geq 0.95$ for $N \geq 61,093$. The bound for the population level N is obtained by solving for N the equality $\rho(d)w(d) = \delta$, where $w(d)$ is given by equation (11), and we take $\sigma^2 = 0.25$ in the expression for $\rho(d)$. The local noise level d has been chosen by optimising the bound for N as a function of d, and results in $w(d) = 49$.

Another approach involves finding a Lyapunov-like function. A natural candidate is $\mathbf{V}(\mathbf{m}') := \|\mathbf{m}'\|^2$. In this case, we take, for $K, Z, Z' \in \mathbb{R}_+$ to be determined, and $r \in \mathbb{R}_+$:

$$\begin{aligned} &\alpha_1(r) := \alpha_2(r) := r^2 && \alpha_1^{-1}(r) = \alpha_2^{-1}(r) = \sqrt{r} \\ &\alpha_3(r) := Kr^2 && \\ &\alpha_4(r) = Kr && \alpha_4^{-1}(r) = r/K \\ &\beta(r, t) = (1 - (1 - Z/Z')K)^{t/2}\, r && W = \sqrt{Z'/K} \end{aligned}$$

Now as long as $W \leq \epsilon$ and Eq. (12) hold, as $\|\mathbf{m}'\| \leq \sqrt{3}$ for all $\mathbf{m}' \in \mathcal{M}'$, we can use Eq. (13) to obtain:

$$w(d) = \left\lceil \frac{2\log\left(\epsilon/\sqrt{3}\right)}{\log(1 - (1 - Z/Z')K)} \right\rceil \tag{14}$$

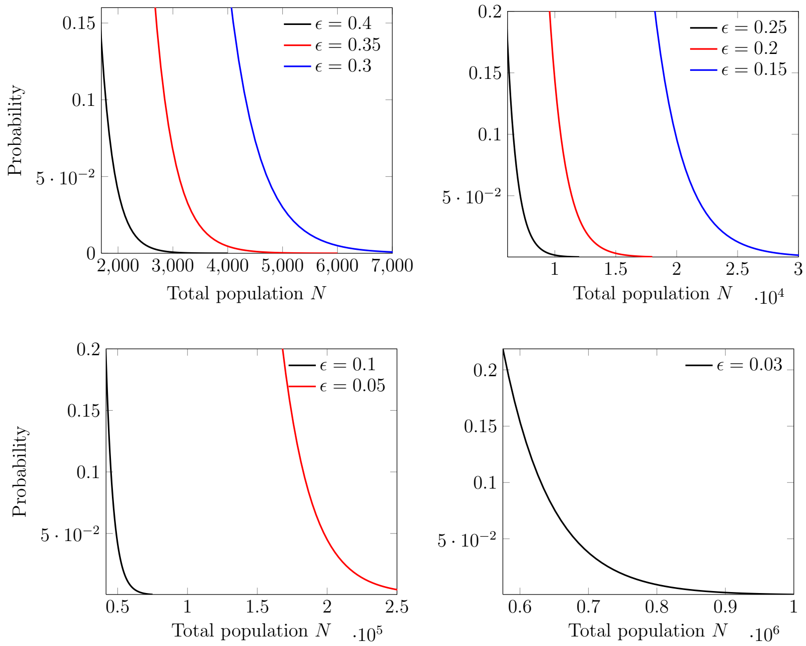


Figure 2: Upper bounds on $\mathbb{P}\{\mathbf{M}'(\infty) \notin \mathcal{E}\}$ for different values of ϵ as the total component population size N varies in the WSN model. In each case here, we have taken $Z' = 1.5Z$.

For given $\epsilon > 0$, in order to enforce validity of the value of $w(d)$ computed using Eq. (14), we must find specific values of K, d, Z and Z' which satisfy jointly both $W \leq \epsilon$ and Eq. (12).

For the same parameter values above ($\alpha = 0.4$, $\beta = 0.6$, $\eta = 0.3$, $\lambda = 0.2$, $\gamma = 0.2$; and $m_G = 1/3$, $m_W = 2/3$), we can compute the bound with the Lyapunov method. In this case, if we take $d = 0.0167$ then Eq. (12) holds for $K = 0.248$ and $Z = 1.65 \times 10^{-3}$. If we choose $Z' = 1.5Z$ and $\epsilon = 0.1$, then we have also $W < \epsilon$ and we may take $w(d) = 67$. This gives us $\mathbb{P}\{\mathbf{M}'(\infty) \in \mathcal{E}\} \geq 0.95$ for $N \geq 48,986$, which is an improvement over the Lipschitz constant-based bound, but at the price of increased complexity.

Finally, we can compute the bounds using the numerical approach based on multi-affine functions. In this case, for the same parameters as above, we get an estimate of $w(d) = 27$ for local noise level $d = 0.013$, giving that $\mathbb{P}\{\mathbf{M}'(\infty) \in \mathcal{E}\} \geq 0.95$ for $N \geq 72,464$. This is the worst bound for N, as the over-approximation error requires to choose a smaller local noise, but it has been obtained purely mechanistically without any requirement to perform global optimisation to validate Lipschitz constants or Lyapunov functions. Furthermore, the computations involved took less than one second on a standard desktop computer (the fixed point is reached after 49 steps for a precision of 10^{-6}).

Figure 2 gives many further examples of the bounds that can be obtained for this model for various values of ϵ and the parameters given above, using the Lyapunov function method to estimate $w(d)$. Here we have used the Global Optimization Toolbox of Matlab [27] to verify the drift equation Eq. (12). We note that once specific K, d, Z and Z' have been found to satisfy $W \leq \epsilon$ and Eq. (12) for a given ϵ, the resulting $w(d)$ is valid for any N since here the function $\mathbf{f}$ does not depend on N. This allows us to use the same values of these parameters for each graph in Figure 2 which we do here for the sake of presentation. It may be that tighter bounds for a specific N can be obtained by varying d with N and thus potentially requiring different values for K, Z and Z' for each N. Indeed, the bounds that can be obtained may be tightened significantly by performing computational local optimisation to minimise the quantity $w(d)\rho(d)$ with respect to the parameters (K, d, Z, Z') subject to the constraints that the drift equation (Eq. (12)) holds and also that $W < \epsilon$.

We consider now a different set of parameters for the model, namely $\alpha = 0.23$, $\beta = 0.89$, $\eta = 0.49$,

$\lambda = 0.18$, $\gamma = 0.02$; and $m_G = 1/3$, $m_W = 2/3$. In this case, by numerically maximising the Jacobian of $\mathbf{f}$, we obtain $L \geq 1.06$, hence the simple Lipschitz constant-based bound cannot be used, further justifying the need for the additional approaches, which can still be applied here. For example, the computational approach gives, for a local noise level $d = 0.009$, a bound for the return time $w(d) = 34$, resulting, for $\epsilon = 0.1$, in the bound $\mathbb{P}\{\mathbf{M}'(\infty) \in \mathcal{E}\} \geq 0.95$ for $N \geq 155,027$. The increased bound for N with respect to the previous set of parameters is due to an increased sensitivity of the function $\mathbf{f}$ to the local perturbations, resulting in smaller values for d.

Finally, we will exploit the results obtained here to give upper and lower bounds on the mean steady-state response time for a wireless node to communicate its data to a gateway node once it has observed it. This would be a useful performance quantity in practice since the network designer may seek to minimise this delay period so as to conserve the energy used by the wireless nodes' radio due to their limited batteries. In order to proceed, we will use the numerical results computed above and displayed in Figure 2 to obtain bounds on the expected occupancy measure $\mathbb{E}[\mathbf{M}(t)]$, which, utilising Little's Law, will yield bounds on the mean response time directly. Specifically, if the response-time random variable of interest is R, then, we have:

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

From an upper bound $\mathbb{P}\{\mathbf{M}'(\infty) \notin \mathcal{E}\} \leq p$, we may obtain the following expectation bounds straightforwardly:

$$\begin{aligned}
\mu_c^* - (p+\epsilon) \leq \mathbb{E}[M_c(t)] &\leq \mu_c^* + (p+\epsilon) \\
\mu_d^* - (p+\epsilon) \leq \mathbb{E}[M_d(t)] &\leq \mu_d^* + (p+\epsilon) \\
m_W - \mu_c^* - \mu_d^* - 2(p+\epsilon) \leq \mathbb{E}[M_e(t)] &\leq m_W - \mu_c^* - \mu_d^* + 2(p+\epsilon)
\end{aligned}$$

and thus:

$$\frac{\mu_c^* + \mu_d^* - 2(p+\epsilon)}{\lambda(m_W - \mu_c^* - \mu_d^* + 2(p+\epsilon))} \leq \mathbb{E}[R] \leq \frac{\mu_c^* + \mu_d^* + 2(p+\epsilon)}{\lambda(m_W - \mu_c^* - \mu_d^* - 2(p+\epsilon))}$$

Figure 3 shows the response-time bounds computed in this manner using the probability bounds $\mathbb{P}\{\mathbf{M}'(\infty) \notin \mathcal{E}\} \leq p$ shown in Figure 2 for various values of ϵ as N is varied. For comparison, we also plot the actual mean response time as computed using stochastic simulation, which in this case is essentially equal to the deterministic approximation given by the mean-field equation's fixed point. However, it is important to note that the bounds computed here provide a guaranteed worst case upper bound on the mean response time, whereas, it is unknown *a priori* how accurate the deterministic mean-field approximation will be for any given N.

6. Error bounds for the transient dynamics

In the previous sections we discussed in detail how to obtain error bounds for the stationary distributions, combining properties of the stochastic process with the analysis of the perturbed limit model. In so doing, we considered several methods to estimate the return time of a perturbed discrete-time dynamical system: one based on the Lipschitz constant, one based on Lyapunov functions, and one based on numerical over-approximation algorithms. We now investigate further some of these approaches, and tune them to estimate the error also on the transient dynamics.

We will proceed in two steps: first, similarly to the steady-state case, we move from the stochastic world to the non-deterministic one, reducing the problem of computing error bounds in the stochastic model to a problem of computing bounds on the dynamics of the perturbed mean-field model. Then we solve the resulting non-deterministic problem.

Fix now a time horizon $T > 0$ and the initial state $\mathbf{m}_0'$. We want to find bounds on the transient dynamics up to time T. Recall that the (reduced) stochastic process can be written as $\mathbf{M}'(t+1) = \mathbf{f}(\mathbf{M}'(t)) + \mathbf{D}'(t+1)$

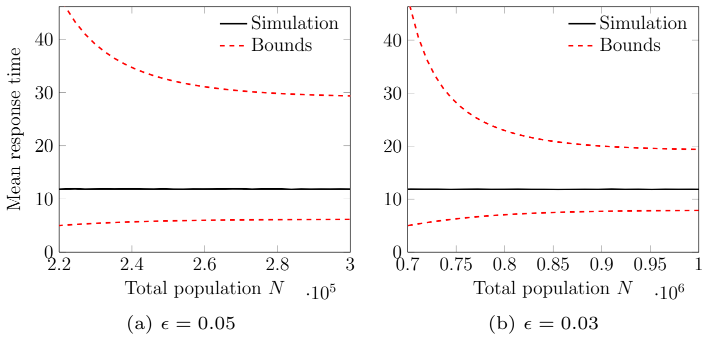


Figure 3: Bounds on the mean steady-state response time for a wireless node to communicate its data to a gateway node, $\mathbb{E}[W]$. Results are computed using the probability bounds in Figure 2 for different values of ϵ and as N varies.

and that for any t and any state $\mathbf{m}' \in \mathcal{M}'$, we have that $\mathbb{P}\{\|\mathbf{D}'(t+1)\| > d \,|\, \mathbf{M}'(t) = \mathbf{m}'\} \leq \rho(d)$, where $\rho(d)$ has been defined in Section 3. Now note that by the tower law of conditional expectation:

$$\begin{aligned}
\mathbb{P}\left\{\max_{t<T} \|\mathbf{D}'(t+1)\| \leq d\right\} &= \mathbb{P}\left(\bigcap_{t<T} \{\|\mathbf{D}'(t+1)\| \leq d\}\right) \\
&= \mathbb{E}[\mathbb{E}[\mathbf{1}_{\{\|\mathbf{D}'(1)\|\leq d\}} \mathbf{1}_{\bigcap_{0<t<T}\{\|\mathbf{D}'(t+1)\|\leq d\}} \,|\, (\mathbf{M}'(t))_{t\leq 1}]] \\
&= \mathbb{E}[\mathbf{1}_{\{\|\mathbf{D}'(1)\|\leq d\}} \mathbb{E}[\mathbf{1}_{\bigcap_{0<t<T}\{\|\mathbf{D}'(t+1)\|\leq d\}} \,|\, (\mathbf{M}'(t))_{t\leq 1}]] \\
&= \mathbb{E}[\mathbf{1}_{\{\|\mathbf{D}'(1)\|\leq d\}} \mathbb{E}[\mathbb{E}[\mathbf{1}_{\bigcap_{0<t<T}\{\|\mathbf{D}'(t+1)\|\leq d\}} \,|\, (\mathbf{M}'(t))_{t\leq 2}] \,|\, (\mathbf{M}'(t))_{t\leq 1}]] \\
&= \mathbb{E}[\mathbf{1}_{\{\|\mathbf{D}'(1)\|\leq d\}} \mathbb{E}[\mathbf{1}_{\{\|\mathbf{D}'(2)\|\leq d\}} \mathbb{E}[\mathbf{1}_{\bigcap_{1<t<T}\{\|\mathbf{D}'(t+1)\|\leq d\}} \,|\, (\mathbf{M}'(t))_{t\leq 2}] \,|\, (\mathbf{M}'(t))_{t\leq 1}]]
\end{aligned}$$

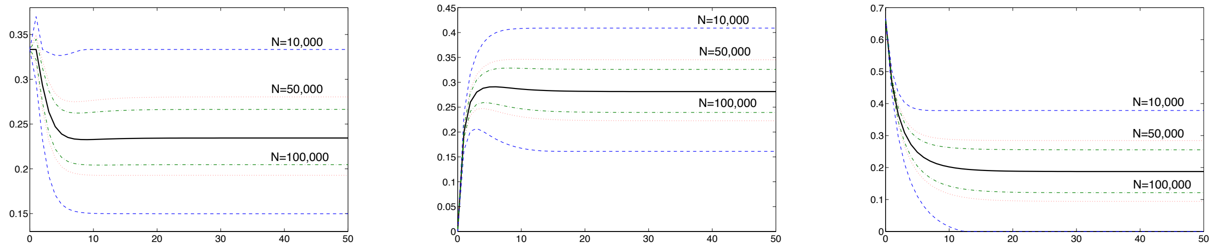
Iterating this line of reasoning, we see that $\mathbb{P}\{\max_{t<T} \|\mathbf{D}'(t+1)\| > d\} \leq 1 - (1-\rho(d))^T$. It follows that, with probability at least $(1-\rho(d))^T$, the local noise level will be uniformly less than d, hence the perturbed system $\mu'(t+1) = \mathbf{f}(\mu'(t)) + \mathbf{w}(t)$ with $\|\mathbf{w}(t)\| \leq d$ and $\mu'(0) = \mathbf{m}'_0$, bounds the evolution of the stochastic model. More precisely, we have that, conditional on $\max_{t<T} \|\mathbf{D}'(t+1)\| > d$, the trajectories of the stochastic model (up to time T) are contained in the trajectories of the perturbed system (up to time T). Hence, if we can bound for each $t \leq T$ the set of states that can be reached by the perturbed dynamical system, we have automatically a bound on the transient dynamics of the stochastic model.

Inspecting again the techniques of Section 4 to bound the return time, we can easily see that the Lipschitz-based method and the numerical method are two tools that can be used to compute over-approximations of the reachable sets. As for the numerical method, we just need to compute the over-approximation $\hat{E}_t$ of the set E_t of reachable states with the method of choice. For multi-affine functions, we can use the approach discussed in Section 4.2. For the Lipschitz constant-based bound, inspecting again equation (10) and using the fact that the state at time zero is fixed, we obtain that

$$E_t \subseteq \mathcal{B}\left(\mu'(t), d\left(\frac{1-L^t}{1-L}\right)\right),$$

where L is the Lipschitz constant of the function $\mathbf{f}$ and $\mathcal{B}\left(\mu'(t), d\left(\frac{1-L^t}{1-L}\right)\right)$ is the ball centred in $\mu'(t)$, the solution of the unperturbed system, with radius $d\left(\frac{1-L^t}{1-L}\right)$. Obviously, if $L > 1$, these bounds will grow exponentially with t.

As an example, consider again the non-linear wireless sensor network model of Section 5, with parameters



(a) $M_{\mathtt{a}}(t)$ (b) $M_{\mathtt{c}}(t)$ (c) $M_{\mathtt{d}}(t)$

Figure 4: Transient error bounds on the dynamics of the wireless sensor network model of Section 5, for parameters $\alpha = 0.4$, $\beta = 0.6$, $\eta = 0.3$, $\lambda = 0.2$, $\gamma = 0.2$, $m_G = 1/3$, $m_W = 2/3$ and different values of N.

$\alpha = 0.4$, $\beta = 0.6$, $\eta = 0.3$, $\lambda = 0.2$, $\gamma = 0.2$, $m_G = 1/3$, $m_W = 2/3$, so that the Lipschitz constant of the function $\mathbf{f}$ is $L \leq 0.852 < 1$, and both the numerical approach and the Lipschitz constant-based bounds can be used. In Figure 4, we show the error bounds up to time $T = 50$, for different values of N. To compute them, we fixed the confidence level to $\delta = 0.95$, and solved the equation $(1 - \rho(d))^T = \delta$ for d, to find the smallest local noise level d resulting in the desired confidence. Then, we plugged the obtained value of d into the numerical approach, starting from an initial state in which all gateways are in state $\mathtt{a}$ and all sensors are in state $\mathtt{d}$. As expected, bounds get tighter as N increases. Note also that the shrinking of the bounds passing from $N = 50,000$ to $N = 100,000$ is smaller than the one from $N = 10,000$ to $N = 50,000$, because of the effects of the over-approximation error intrinsic in the numerical approach.

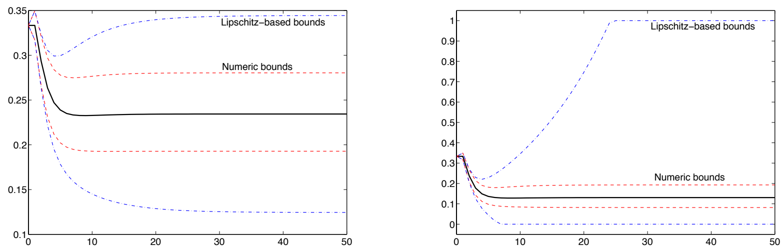
In Figure 5 left, instead, we compare the Lipschitz constant-based bounds with the ones obtained from the numerical approach, for $N = 50,000$. As can be seen, the numerical method provides tighter bounds than the Lipschitz method, differently from the steady-state case.[6] We can also compare the numerically computed bounds with the Lipschitz constant based ones for parameters $\alpha = 0.23$, $\beta = 0.89$, $\eta = 0.49$, $\lambda = 0.18$, $\gamma = 0.02$ (Figure 5 right). In this case, the Lipschitz constant is greater than one, hence the Lipschitz method fails to provide useful bounds, while the numerical approach still works.

We stress that this exponential blow up of error bounds is rather common, making the computational approach the only general method to compute error bounds. Indeed, in continuous time, known error bounds [5] have dependence on the Lipschitz constant L and the number of steps T proportional to $\exp(LT)$.

7. Conclusion

In this paper we have presented a method to compute error bounds (in probability) on both the steady-state and the transient dynamics of a population DTMC model converging to a mean-field limit in discrete time. The method combines stochastic bounds in terms of martingale inequalities and Chernoff inequalities, with control-theoretic methods to study the stability of a system perturbed by non-deterministic noise terms, and with algorithms to over-approximate the set of reachable states. The insight behind this method is that, by abstracting stochastic noise in a non-deterministic fashion, we can use refined control theoretic tools that can take into account the fine-grained nature of the phase space of the mean-field limit, which is instead ignored by standard error bounds [5]. To our knowledge, this has allowed us to obtain for the first time, general bounds for steady-state convergence, and, additionally, transient error bounds which do not explode exponentially.

[6] This inconsistency is due to the fact that bounds for the transient for the Lipschitz method are computed using the 1-norm, while bounds for the steady state for the Lipschitz method use the 2-norm. The numerical approach, instead, is based on over-approximation by hyper-rectangles, which is a closer approach to the 1-norm.



(a) $L < 1$ (b) $L > 1$

Figure 5: Comparison of Lipschitz constant based bounds with those obtained by the numerical approach, for the wireless sensor network model of Section 5 with $N = 50,000$. Left: parameters $\alpha = 0.4$, $\beta = 0.6$, $\eta = 0.3$, $\lambda = 0.2$, $\gamma = 0.2$, $m_G = 1/3$, $m_W = 2/3$, resulting in a Lipschitz constant $L < 1$. Right: parameters $\alpha = 0.23$, $\beta = 0.89$, $\eta = 0.49$, $\lambda = 0.18$, $\gamma = 0.02$, $m_G = 1/3$, $m_W = 2/3$, resulting in a Lipschitz constant $L < 1$.

Our work is related to Freidlin-Wentzell large deviation theory [29], which has also been tuned to discrete time [30]. However, such theories give bounds on large deviations from attractors of the limit mean field models that depend on a cost function defined by talking the infimum of a "cost" over all possible paths linking two states. Such a quantity is extremely difficult to compute in practice, hence large deviation bounds are generally used *implicitly* to prove qualitative properties of the sequence of Markov processes. Our approach, instead, is focussed on providing ways to *explicitly* compute error bounds for steady-state and for transient dynamics, avoiding the exponential blow up with the increase in the time horizon. Ideas taken from the derivation of bounds on the exit times from attractors, however, may be used to tighten the bounds given in this paper. We are currently investigating the feasibility of such an approach.

In the future we will apply this method to more complex examples, but there is also scope for improvements and extensions:

- Extend the steady-state bound to mean-field limits in continuous time;
- Tighten the current bounds, by improving the probability bound $\rho(d)$ on the local noise (which is the quantity which impacts more on the error bound, as N and d appear together in the exponential of $\rho(d)$) and by bounding more effectively the size of large deviations and the return time $w(d)$ in the derivation of the steady-state bound.

Acknowledgements. Work partially funded by the EU-FET project QUANTICOL (nr. 600708), FRA-UniTS and the UK EPSRC project AMPS (ref. EP/G011737/1).

References

[1] L. Kleinrock, Queueing Systems—Theory, Vol. 1, Wiley-Interscience, 1975.

[2] J. Bradley, S. Gilmore, J. Hillston, Analysing distributed Internet worm attacks using continuous state-space approximation of process algebra models, Journal of Computer and System Sciences 74 (6) (2008) 1013–1032.

[3] T. Kurtz, Approximation of population processes, SIAM, 1981.

[4] J.-Y. Le Boudec, 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), IEEE, 2007, pp. 3–18.

[5] R. W. R. Darling, J. R. Norris, Differential equation approximations for Markov chains, Probability Surveys 5 (2008) 37–79.

[6] M. Benaïm, J.-Y. Le Boudec, A class of mean field interaction models for computer and communication systems, Performance Evaluation 65 (11-12) (2008) 823–838.

[7] C. Bordenave, D. McDonald, A. Proutiére, A particle system in interaction with a rapidly varying environment: Mean field limits and applications, NHM 5 (1) (2010) 31–62.

[8] G. Sharma, A. Ganesh, P. Key, Performance analysis of contention based medium access control protocols, IEEE Transactions on Information Theory 55 (4) (2009) 1665–1682.
[9] S. Kumar, L. Massouli, Integrating streaming and file-transfer internet traffic: fluid and diffusion approximations, Queueing Systems 55 (4) (2007) 195–205.
[10] R. Bakhshi, L. Cloth, W. Fokkink, B. Haverkort, Mean-field analysis for the evaluation of gossip protocols, in: Proceedings of Sixth International Conference on the Quantitative Evaluation of Systems (QEST 2009), 2009, pp. 247–256.
[11] A. Martinoli, K. Easton, W. Agassounon, Modeling swarm robotic systems: a case study in collaborative distributed manipulation, The International Journal of Robotics Research 23 (4-5) (2004) 415–436.
[12] R. Hayden, J. Bradley, A fluid analysis framework for a Markovian process algebra, Theor. Comput. Sci. 411 (22-24) (2010) 2260–2297.
[13] L. Bortolussi, A. Policriti, Dynamical systems and stochastic programming: To ordinary differential equations and back, T. Comp. Sys. Biology 11 (2009) 216–267.
[14] R. Hayden, A. Stefanek, J. Bradley, Fluid computation of passage-time distributions in large Markov models, Theor. Comput. Sci. 413 (1) (2012) 106–141.
[15] R. A. Hayden, J. T. Bradley, A. Clark, Performance Specification and Evaluation with Unified Stochastic Probes and Fluid Analysis, IEEE Transactions on Software Engineering 39 (1).
[16] L. Bortolussi, J. Hillston, Fluid model checking, in: Proceedings of CONCUR 2012, 2012.
[17] T. Kurtz, S. Ethier, Markov Processes - Characterisation and Convergence, Wiley, 1986.
[18] R. A. Hayden, Convergence of ODE approximations and bounds on performance models in the steady-state, in: 9th Workshop on Process Algebra and Stochastically Timed Activities (PASTA), 2010, pp. 49–56.
URL http://aesop.doc.ic.ac.uk/pubs/ode-stationary-bounds/
[19] T. Dayar, H. Hermanns, D. Spieler, V. Wolf, Bounding the equilibrium distribution of Markov population models, Numerical Linear Algebra with Applications 18 (6) (2011) 931–946.
[20] H. K. Khalil, Nonlinear systems, MacMillan Pub. Co., 1992.
[21] Z.-P. Jiang, Y. Wang, Input-to-state stability for discrete-time nonlinear systems, Automatica 37 (6) (2001) 857–869.
[22] C. Belta, L. Habets, Controlling a class of nonlinear systems on rectangles, Automatic Control, IEEE Transactions on 51 (11) (2006) 1749–1759.
[23] T. Dang, O. Maler, R. Testylier, Accurate hybridization of nonlinear systems, in: K. H. Johansson, W. Yi (Eds.), HSCC, ACM ACM, 2010, pp. 11–20.
[24] T. Dang, C. L. Guernic, O. Maler, Computing reachable states for nonlinear biological models, Theor. Comput. Sci. 412 (21) (2011) 2095–2107.
[25] W. Feller, An Introduction to Probability Theory and its Applications (Volume II), John Wiley & Sons, 1966.
[26] F. Chung, L. Lu, Complex Graphs and Networks, American Mathematical Society, 2006.
[27] MATLAB, v. 7.14.0 (R2012a), The MathWorks Inc., Natick, Massachusetts, 2012.
[28] D. A. Cox, J. B. Little, D. O'Shea, Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, 2nd edition, Springer, 2006.
[29] A. Shwartz, A. Weiss, Large Deviations for Performance Analysis, Chapman & Hall, 1995.
[30] Y. Kifer, A discrete-time version of the Wentzell-Friedlin theory, The Annals of Probability 18 (4) (1990) pp. 1676–1692.
[31] W. Whitt, Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues, Springer, 2002.
[32] Z.-P. Jiang, Y. Wang, A converse Lyapunov theorem for discrete-time systems with disturbances, Systems & Control Letters 45 (1) (2002) 49–58.

Appendix A. Proofs

Appendix A.1. Mean-field convergence proofs

Appendix A.1.1. Transient regime

Using the Skorohod Representation Theorem [e.g. 31, Theorem 3.2.2], we can construct a common probability space for the $\mathbf{M}^{(N)}(0)$ and $\mu(0)$ for which $\mathbf{M}^{(N)}(0) \to \mu(0)$ almost surely. On the same probability space, we can extend the construction of $\mathbf{M}^{(N)}(t)$ and $\mu(t)$ to all $t \in \mathbb{Z}_+$ by introducing suitable multinomial random variables.

We aim now to show that, on this probability space, $\mathbf{M}^{(N)}(t) \to \mu(t)$ in probability for each $t \geq 0$ which is sufficient to prove the theorem. We proceed by induction; the base case $t = 0$ is immediate. Assume that $\mathbf{M}^{(N)}(t) \to \mu(t)$ in probability for some $t \geq 0$. We have:

$$\|\mathbf{M}^{(N)}(t+1) - \mu(t+1)\| \leq \|\mathbf{M}^{(N)}(t)\mathbf{K}(\mathbf{M}^{(N)}(t)) - \mu(t)\mathbf{K}(\mu(t))\| + \|\mathbf{D}^{(N)}(t+1)\| + \epsilon^{(N)}$$

where $\mathbf{D}^{(N)}(t+1) := \mathbf{M}^{(N)}(t+1) - \mathbf{M}^{(N)}(t)\mathbf{K}(\mathbf{M}^{(N)}(t))$ and the term $\epsilon^{(N)}$ comes from the replacement of $\mathbf{K}^{(N)}$ by $\mathbf{K}$ with the uniformity of their convergence giving $\epsilon^{(N)} \to 0$. The induction hypothesis gives the required convergence in probability of the first term and that $\|\mathbf{D}^{(N)}(t+1)\| \to 0$ in probability follows from any of the bounds in Section 3.1. □

Appendix A.1.2. Steady-state regime

The sequence of stationary distributions $\mathbf{M}^{(N)}(\infty)$ is tight since K is compact. Let $\mathbf{M}^{N_i}(\infty) \Rightarrow \mathbf{M}(\infty)$ be some weakly convergent subsequence. Now let each $\mathbf{M}^{N_i}(0)$ be distributed as $\mathbf{M}^{N_i}(\infty)$, then each $\mathbf{M}^{N_i}(t)$ is stationary. Then by Theorem 1, we have that $\mathbf{M}^{N_i}(t) \Rightarrow \mu(t)$ for each $t \geq 0$, where $\mu(t+1) = \mu(t)\mathbf{K}(\mu(t))$ for $t \geq 0$ and $\mu(0)$ is distributed as $\mathbf{M}(\infty)$. Now since each $\mathbf{M}^{N_i}(t)$ is stationary and by uniqueness of the limit, $\mu(t)$ is stationary and distributed as $\mathbf{M}(\infty)$. Assume for a contradiction that there is some neighbourhood $U \subset K$ of $\mathbf{m}^*$ such that $\mathbb{P}\{\mathbf{M}(\infty) \in K \setminus U\} = \mathbb{P}\{\mu(0) \in K \setminus U\} > 0$. Finally, since the mean-field dynamical system is globally asymptotically stable on the compact set K, it is uniformly globally asymptotically stable there [32, Corollary 3.3], and therefore there is some time $t \in \mathbb{Z}_+$ for which we must have $\mathbb{P}\{\mu(t) \in K \setminus U\} = 0$, which is the required contradiction. □

Appendix A.2. Proof of the dynamical system bound in Section 4

We work with the setup and notation of Section 4 which we do not repeat here for the sake of brevity.

We begin by noting that for any $\mathbf{m}' \in \mathcal{M}'$ and $\mathbf{w} \in \mathbb{R}^{|\mathcal{S}'|}$ with $\|\mathbf{w}\| \leq d$:

$$V(\mathbf{f}(\mathbf{m}') + \mathbf{w}) - V(\mathbf{m}') \leq -\alpha_4(V(\mathbf{m}')) + Z$$

Set now $\mathcal{K} := \{\mathbf{m}' \in \mathcal{M}' \,:\, V(\mathbf{m}') \leq \alpha_4^{-1}(Z')\}$ and assume that for some $t_0 \in \mathbb{Z}_+$, $\mu'(t_0) \in \mathcal{K}$. Then:

$$V(\mu'(t_0+1)) \leq (\mathbf{1} - \alpha_4)(V(\mu'(t_0))) + Z \leq (\mathbf{1} - \alpha_4)(\alpha_4^{-1}(D')) + Z \leq \alpha_4^{-1}(Z')$$

since $\mathbf{1} - \alpha_4$ was chosen to be non-decreasing. Therefore we have that $\mu'(t) \in \mathcal{K}$ for all $t \geq t_0$ by induction.

Now set $t_0 := \min\{t \in \mathbb{Z}_+ \,:\, \mu'(t) \in \mathcal{K}\} \leq \infty$ and for $t \geq t_0$, we have $V(\mu'(t)) \leq \alpha_4^{-1}(Z')$, which gives $\|\mu'(t)\| \leq \alpha_1^{-1}(\alpha_4^{-1}(Z')) = W$. For $t < t_0$, it holds that $\alpha_4(V(\mu'(t))) > Z'$, and hence:

$$\begin{aligned} V(\mu'(t+1)) - V(\mu'(t)) &\leq -\alpha_4(V(\mu'(t))) + Z = -(1 - Z/Z')\alpha_4(V(\mu'(t))) - (Z/Z')\alpha_4(V(\mu'(t))) + Z \\ &\leq -(1 - Z/Z')\alpha_4(V(\mu'(t))) \end{aligned}$$

Recall that for $r \in \mathbb{R}_+$, $\bar{\beta}(r) := \sup_{s \in [0,r]}\{s - (1 - Z/Z')\alpha_4(s)\}$ and note that for $r > 0$, $\bar{\beta}(r) < r$ and $\bar{\beta}$ is non-decreasing. For $t < t_0$, we have $V(\mu'(t+1)) \leq \bar{\beta}(V(\mu'(t)))$ and by a standard comparison principle[7] we have $V(\mu'(t)) \leq z(t)$ for all $t \leq t_0$ where $z(t)$ is the unique solution to the initial value problem $z(t+1) = \bar{\beta}(z(t))$ with $z(0) = V(\mu'(0))$. That is, $z(t) = \bar{\beta}^t(V(\mu'(0)))$, so $V(\mu'(t)) \leq \bar{\beta}^t(V(\mu'(0)))$ for all $t \leq t_0$. Note that $\bar{\beta}^t(V(\mu'(0))) \to 0$ as $t \to \infty$. This is because each $\bar{\beta}^t(V(\mu'(0))) \geq 0$ and the sequence is non-increasing so that we must have $\bar{\beta}^t(V(\mu'(0))) \to c \geq 0$. However, since $\bar{\beta}$ is continuous, we have $\lim_{t\to\infty} \bar{\beta}(\bar{\beta}^t(V(\mu'(0)))) = \bar{\beta}(c) < c$ if $c > 0$ so we must have $c = 0$. Finally, for $t \leq t_0$, we then obtain $\|\mu'(t)\| \leq \alpha_1^{-1}(\bar{\beta}^t(V(\mu'(0)))) \leq \alpha_1^{-1}(\bar{\beta}^t(\alpha_2(\|\mu'(0)\|)))$, from which the desired result follows. □

[7] If the scalar functions $y(t)$ and $x(t)$ satisfy $y(t+1) = f(y(t))$ and $x(t+1) \leq f(x(t))$ for $t < t_0$ with $x(0) \leq y(0)$ where f is non-decreasing, it is straightforward to see that $x(t) \leq y(t)$ for all $t \leq t_0$.