Mean-field approximation of counting processes from a differential equation perspective

Dedicated to Professor Tibor Krisztin on the occasion of his 60th birthday

Dávid Kunszenti-Kovács[*1, 3] and Péter L. Simon[✉2, 3]

[1]MTA Alfréd Rényi Institute of Mathematics, P.O. Box 127, Budapest H–1364, Hungary
[2]Institute of Mathematics, Eötvös Loránd University, 1/C Pázmány P. s., Budapest, H–1117, Hungary
[3]Numerical Analysis and Large Networks Research Group
Hungarian Academy of Sciences, Hungary

Abstract. Deterministic limit of a class of continuous time Markov chains is considered based purely on differential equation techniques. Starting from the linear system of master equations, ordinary differential equations for the moments and a partial differential equation, called Fokker–Planck equation, for the distribution is derived. Introducing closures at the level of the second and third moments, mean-field approximations are introduced. The accuracy of the mean-field approximations and the Fokker–Planck equation is investigated by using two differential equation-based and an operator semigroup-based approach.

Keywords: mean-field model, exact master equation, Fokker–Planck equation.

2010 Mathematics Subject Classification: 35Q84, 47D06, 47N40, 60J28, 65C40.

1 Introduction

Large systems consisting of many identical particles are usually described by stochastic processes, while their deterministic limit can be modeled by differential equations. This widely known and accepted idea is a strongly embedded scientific paradigm and has been justified in several ways from heuristic reasoning to sophisticated and abstract mathematical theories. Despite of these facts, new approaches for deriving deterministic limits, based purely on differential equation techniques, has been developed recently. The aim of this paper is to review and put these approaches in a unified context.

Although the problem can be formulated in very general terms, here we restrict ourselves to a specific situation which is a compromise between tractability and mathematical generality. In the discussion section we will show the ways in which the problem can be generalized. Let

---

[*]Email: daku@renyi.hu
[✉]Corresponding author. Email: simonp@cs.elte.hu

us consider a system of N identical elements or particles each of which can be in one of finitely many states. To be specific we will consider the case with two states denoted by Q and T. The number of particles in state Q and T at time t is denoted by $X_Q(t)$ and $X_T(t)$, respectively. These are considered to be random variables and our main goal is to derive differential equations yielding approximations for the expected value of these variables. The state of the system changes when the state of a particle changes, for which there are two possibilities: transitions $Q \to T$ and $T \to Q$. It will be assumed that the transition rate depends on the proportion of nodes in the different states, hence the state of the whole system can be given by the number of particles of different types. The process is then a birth–death or counting process, in which the the number of particles of each type changes by one in a short time interval. The state of the system can be given by the pair $(k, N - k)$ yielding the number of particles of type Q and T. In fact, the state will be given only by k and the main objects of our study will be the time dependent functions $p_k(t)$, the probabilities of having k particles of type Q at time t. Once the states of the system are given, the transition rates between the states can be defined and the master equations for the probability for each state can be written down. While limiting mean-field ODE models can provide a good approximation of the expected value of the random variables X_Q and X_T, a PDE-based approach is needed if information on the probability distribution of these is desired. The aim of the paper is to review and study the differential equation-based methods that has been developed to estimate the accuracy of mean-field ODE and PDE approximations.

Such and similar questions were studied in the density dependent case by several authors, see [7, 10, 14]. The stochastic convergence of the random variables to the solution of the mean-field equation has been widely studied by using martingale theory [7, 10, 14]. Uniform convergence was proved in [17] by introducing an infinite system of ODEs for the moments. In [5] uniform convergence was also proved by using the approximation theory of operator semigroups and the authors showed that the difference is of order $1/N$. This operator semigroup approach enabled them to approximate not only the expected value but also the distribution p_k itself with a partial differential equation [6]. The approximation is based on introducing a two-variable function v for which $v(t, k/N) \approx p_k(t)$ and deriving the Fokker–Planck equation. Then the master equation can be considered to be the discretization of the Fokker–Planck equation in an appropriate sense. Armbruster developed a simple approach [3], based only on elementary ODE and probability tools, to prove that the accuracy of the mean-field approximation is order $1/N$. It is natural to look for lower and upper bounds for the expected value that can be used for finite N (in contrast to the previous asymptotic results). It is known that in many cases, the mean-field ODE yields an upper bound on the expected value of the process. The first lower bound was derived by Armbruster and Beck [2] where a system of two ODEs yields a lower bound on the expected value in the case of a susceptible–infected–susceptible (SIS) epidemic model on a complete graph. This result was extended to a wider class of Markov chains in [1]. ODE systems yielding order $1/N^2$ approximation can be derived by using a priori assumptions on the distribution p_k, see [13].

The paper is structured as follows. In Section 2 the master equations for the case of two states is formulated and the differential equations for the moments are derived together with the corresponding Fokker–Planck equation. A slightly simplified version of Armbruster's elementary approach [3], is shown in Section 3 in the density dependent case given by quadratic polynomials. The upper and lower bounds in this case are shown in Section 4 based on the approach presented in [1]. The higher order approximation based on a priori distributions is shown in Section 5. The relation of the Fokker–Planck equation and the master equation is

dealt with in Section 6 by using the operator semigroup approach. The possible generalizations are presented in the discussion section.

2 Model formulation

Whereas the master equations can be derived for systems with arbitrarily many particle states, in order to present the methods for estimating the accuracy of mean-field models we will use systems with only two particle states Q and T. Thus the state space of the whole system will be $\{0, 1, \ldots, N\}$, where k represents the state with k particles in state Q and $N - k$ particles in state T, that is $X_Q(t) = k$ and $X_T(t) = N - k$. Transition from state k is possible only to states $k + 1$ and $k - 1$ with rates a_k and c_k, respectively. These processes are called birth–death or one-step processes. Denoting by $p_k(t)$ the probability of state k at time t and assuming that the process is Markovian, the master equations of the process take the form

$$\dot{p}_k = a_{k-1} p_{k-1} - (a_k + c_k) p_k + c_{k+1} p_{k+1}, \qquad k = 0, \ldots, N. \tag{2.1}$$

The equation corresponding to $k = 0$ does not contain the first term in the right-hand side, while that corresponding to $k = N$ does not contain the third term, i.e. $a_{-1} = c_{N+1} = 0$. Moreover, the Markov chain requires a_N and c_0 are set to zero. This will ensure that the sum of each column in the transition matrix is zero.

The infinite size limit, i.e. the case when $N \to \infty$, can be described by differential equations in the so-called density dependent case, when the transition rates a_k and c_k can be given by non-negative, continuous functions $A, C : [0, 1] \to [0, +\infty)$ satisfying $A(1) = 0 = C(0)$ as follows

$$\frac{a_k}{N} = A\left(\frac{k}{N}\right) \quad \text{and} \quad \frac{c_k}{N} = C\left(\frac{k}{N}\right). \tag{2.2}$$

We note that the conditions $A(1) = 0 = C(0)$ ensure $a_N = 0 = c_0$. The case when A and C are polynomials will play a crucial role in our investigation.

Once the system of master equations is solved for p_k, one can determine the (scaled) moments of $X_Q(t)$ as

$$m_n(t) = \sum_{k=0}^{N} \frac{k^n}{N^n} p_k(t). \tag{2.3}$$

The aim of our investigation is to determine or approximate

- the expected value m_1 by deriving a mean-field ODE, and

- the distribution p_k by using a PDE.

This can be carried out on one hand by introducing mean-field approximations by deriving ODEs for the moments and using algebraic relations among them as closures, and on the other hand by deriving the corresponding Fokker–Planck equations [15, 18] that can be considered as the continuous version of the master equation (2.1).

2.1 Differential equations for the moments

In order to derive ODEs for the moments we will make use of the following elementary lemma, the proof of which can be found in [5].

Lemma 2.1. *Let r_k ($k = 0, 1, 2, \ldots$) be a sequence and let $r(t) = \sum_{k=0}^{N} r_k p_k(t)$, where $p_k(t)$ is given by (2.1). Then*

$$\dot{r}(t) = \sum_{k=0}^{N} (a_k(r_{k+1} - r_k) + c_k(r_{k-1} - r_k)) p_k(t).$$

Applying this lemma to the *n*-th moment, i.e. choosing $r_k = \frac{k^n}{N^n}$ leads to the following proposition.

Proposition 2.2. *The scaled n-th moment, m_n, of the distribution p_k determined by the master equation (2.1) satisfies the differential equation*

$$\dot{m}_n = \frac{1}{N^n} \sum_{k=0}^{N} \left[a_k((k+1)^n - k^n) + c_k((k-1)^n - k^n) \right] p_k. \tag{2.4}$$

This proposition will enable us to study the moments. Applying it for $n = 1$ in the density dependent case (2.2) leads to

$$\dot{m}_1 = \sum_{k=0}^{N} \left[A\left(\frac{k}{N}\right) - C\left(\frac{k}{N}\right) \right] p_k. \tag{2.5}$$

In order to derive an approximating differential equation for m_1 one can assume that the order of application of a non-linear function (like A or C) and the expected value can be exchanged. (For a linear function this yields an exact relation while for non-linear ones it is only an approximation.) Thus the closure approximation implies

$$\dot{m}_1 \approx A\left(\sum_{k=0}^{N} \frac{k}{N} p_k\right) - C\left(\sum_{k=0}^{N} \frac{k}{N} p_k\right) = A(m_1) - C(m_1).$$

Introducing y_1 as the approximation of m_1, the approximating closed differential equation takes the form

$$\dot{y}_1 = A(y_1) - C(y_1). \tag{2.6}$$

In the case when A and C are polynomials, the proposition makes it possible to derive an exact system of ODEs for the moments. We will mainly investigate the case of quadratic polynomials when

$$A(z) = A_0 + A_1 z + A_2 z^2 \quad \text{and} \quad C(z) = C_0 + C_1 z + C_2 z^2. \tag{2.7}$$

We note that the conditions $A(1) = 0 = C(0)$ impose restrictions on the coefficients, but we will consider the general case of arbitrary coefficients. Then applying (2.5) and introducing $D_i = A_i - C_i$ we get

$$\dot{m}_1 = \sum_{k=0}^{N} \left[D_0 + D_1 \frac{k}{N} + D_2 \frac{k^2}{N^2} \right] p_k = D_0 + D_1 m_1 + D_2 m_2.$$

In order to derive the differential equation for the second moment we apply (2.4) for $n = 2$ and use density dependence leading to

$$\dot{m}_2 = \sum_{k=0}^{N} \left[\frac{a_k}{N} \frac{2k+1}{N} + \frac{c_k}{N} \frac{1-2k}{N} \right] p_k = \sum_{k=0}^{N} \left[A\left(\frac{k}{N}\right) \frac{2k+1}{N} + C\left(\frac{k}{N}\right) \frac{1-2k}{N} \right] p_k.$$

Now using that A and C are quadratic polynomials and introducing $E_i = A_i + C_i$ we obtain

$$\dot{m}_2 = 2 \sum_{k=0}^{N} \left[D_0 \frac{k}{N} + D_1 \frac{k^2}{N^2} + D_2 \frac{k^3}{N^3} \right] p_k + \frac{1}{N} \sum_{k=0}^{N} \left[E_0 + E_1 \frac{k}{N} + E_2 \frac{k^2}{N^2} \right] p_k$$

$$= 2(D_0 m_1 + D_1 m_2 + D_2 m_3) + \frac{1}{N}(E_0 + E_1 m_1 + E_2 m_2).$$

Thus the exact system for the first two moments, when A and C are quadratic polynomials given in (2.7), takes the form

$$\dot{m}_1 = D_0 + D_1 m_1 + D_2 m_2, \tag{2.8}$$

$$\dot{m}_2 = 2(D_0 m_1 + D_1 m_2 + D_2 m_3) + \frac{1}{N}(E_0 + E_1 m_1 + E_2 m_2), \tag{2.9}$$

where $D_i = A_i - C_i$ and $E_i = A_i + C_i$. This exact system contains the third moment hence needs a closure approximation. The performance of the closure can be investigated analytically in two different ways. On one hand, the difference between the exact and the closed system can be proved to decrease in a given order as N tends to infinity. On the other hand, lower and upper bounds can be derived for the exact value of the moments. These will be dealt with in Sections 3 and 4. We note that the system can be derived similarly in the case when A and C are polynomials of higher degree. In that case, higher order moments will also be included in the right-hand sides.

2.2 Fokker–Planck equation

The Fokker–Planck equation can be considered as the continuous version of the master equation (2.1). We wish to approximate the solution $p_k(t)$ by considering it as a discretization of a continuous function $v(t,z)$ in the interval $[0,1]$, i.e.

$$v \left(t, \frac{k}{N} \right) = p_k(t) \tag{2.10}$$

for $0 \leq k \leq N$. Now we derive an approximating PDE, called the Fokker–Planck equation, for the function $v(\cdot, \cdot)$ based on the ODE given by the master equation. This PDE is traditionally given in the form

$$\partial_t v(t,z) = \frac{1}{2} \partial_{zz}(G(z)v(t,z)) - \partial_z(H(z)v(t,z)) \tag{2.11}$$

see [15, 16, 18]. (We note that the factor $1/2$ is sometimes built in the coefficient function G.) The functions G and H will be determined in such a way that the finite difference discretization of this PDE will yield the master equation (2.1). (In fact, any parabolic type PDE with space dependent coefficients could serve as the continuous version of the master equation.) The Fokker–Planck equation can be derived for more general (not only one-step) processes, see e.g. [15, 16, 18].

The following second order finite difference discretization approximations will be used to relate the PDE and the master equation.

$$f(z - \Delta z) - 2f(z) + f(z + \Delta z) \approx \Delta z^2 f''(z), \qquad f(z + \Delta z) - f(z - \Delta z) \approx 2\Delta z f'(z).$$

Applying these formulas with $z = k/N$ and $\Delta z = 1/N$ to the partial derivatives of the functions $G(z)v(t,z)$ and $H(z)v(t,z)$ with respect to z leads to

$$\partial_t v \left(t, \frac{k}{N} \right) = \frac{N^2}{2}(g_{k+1}x_{k+1} - 2g_k x_k + g_{k-1}x_{k-1}) - \frac{N}{2}(h_{k+1}x_{k+1} - h_{k-1}x_{k-1}), \tag{2.12}$$

where the notations $x_k = v(t, \frac{k}{N})$, $g_k = G(\frac{k}{N})$, $h_k = H(\frac{k}{N})$ are used. The above discretization will be used also for $k = 0$ and $k = N$, hence two artificial mesh points are introduced at $z = -1/N$ and at $z = 1 + 1/N$ (these will be eliminated through the use of the boundary conditions, but there we lose one order of magnitude in the approximation, cf. [6, Section 3]). Differentiating (2.10) with respect to t and using the master equation (2.1) yields

$$\partial_t v\left(t, \frac{k}{N}\right) = \dot{p}_k = a_{k-1}p_{k-1} - (a_k + c_k)p_k + c_{k+1}p_{k+1}. \tag{2.13}$$

This equation can be considered as the discretization of the Fokker–Planck equation. Upon substituting p_k by x_k for all k we arrive at the right-hand side of (2.12). Making the coefficients equal leads to

$$a_k = \frac{N}{2}h_k + \frac{N^2}{2}g_k, \qquad c_k = \frac{N^2}{2}g_k - \frac{N}{2}h_k. \tag{2.14}$$

Under the assumption that $c_0 = a_N = 0$, we thus got that the desired unknown functions G and H have to be defined in such a way that the relations

$$G\left(\frac{k}{N}\right) = g_k = \frac{1}{N^2}(a_k + c_k), \qquad H\left(\frac{k}{N}\right) = h_k = \frac{1}{N}(a_k - c_k)$$

hold. In the density dependent case (2.2), when the coefficients a_k and c_k are given by the functions A and C as $\frac{a_k}{N} = A(\frac{k}{N})$ and $\frac{c_k}{N} = C(\frac{k}{N})$, we obtain that G and H can be given as

$$G(z) = \frac{1}{N}(A(z) + C(z)), \qquad H(z) = A(z) - C(z).$$

Similar derivation leads to the boundary conditions, see [6]. Summarising, the Fokker–Plank equation of the one-step-process given by density dependent coefficients is

$$\partial_t v(t, z) = \frac{1}{2N}\partial_{zz}((A(z) + C(z))v(t, z)) - \partial_z((A(z) - C(z))v(t, z)) \tag{2.15}$$

subject to boundary conditions

$$h\partial_z((A + C)v)(-h, t) - ((A - C)v)(-h, t) = 0, \tag{2.16}$$

$$h\partial_z((A + C)v)(1 + h, t) - ((A - C)v)(1 + h, t) = 0, \tag{2.17}$$

where $h = 1/2N$, and satisfies the initial condition

$$v(0, z) = v_0(z) \tag{2.18}$$

for $z \in [-h, 1 + h]$, where the initial function v_0 corresponds to the initial condition $p_k(0)$ in the sense that $v_0(k/N) = p_k(0)$ for all $0 \leq k \leq N$. We note here that this choice of boundary conditions also ensures that the integral of $v(t, \cdot)$ is constant in time.

2.3 Approximation results

The goal now is to prove that the approximation error, that is the difference between the mean-field approximation y_1 and the exact value m_1, is of order $1/N$. Thus, for a large system the mean-field equation gives a good approximation to the expected value. This question is studied in detail in [17] for the case of SIS epidemics and in [5] for general density dependent processes, where the following theorem is proved.

Theorem 2.3. *Let the coefficients of* (2.1) *be given by* (2.2) *and let p_k be the solution of* (2.1) *satisfying the initial conditions $p_\ell(0) = 1$, $p_j(0) = 0$, $j \neq \ell$ with some $\ell \in \{0, 1, 2, \ldots, N\}$. Let m_1 be the expected value and y_1 be the solution of the mean-field equation* (2.6) *subject to the initial condition $y_1(0) = m_1(0) = \frac{\ell}{N}$. Then for any $t_0 > 0$ there exist a constant K, for which*

$$|y_1(t) - m_1(t)| \leq \frac{K}{N}, \qquad t \in [0, t_0].$$

The solution v of the Fokker–Plank equation was introduced as an approximation of the distribution p_k in the sense that $v(t, k/N) \approx p_k(t)$. In [6] Theorem 4.6 states that this is an order $1/N^2$ approximation on finite time intervals, but we have to settle for less concentrated initial conditions. However, the proof in fact yields a weaker result, as the definition of one of the norms absorbs a factor $1/N$, leading to a loss of one order, see Lemma 6.5 later in this paper. The correct statement is Corollary 2.7 below.

We consider an initial function u_0^N that is essentially the same for each N, and set

$$p_k(0) := \frac{1}{Q_N} u_0^N \left(\frac{k}{N} \right), \quad \text{with} \quad Q_N = \sum_{k=0}^{N} u_0^N \left(\frac{k}{N} \right) \tag{2.19}$$

as the initial condition for the ODE system, where the normalization constant ensures that $(p_0, p_1, \ldots, p_N)$ is a probability distribution. In order to estimate the accuracy of the Fokker–Planck equation precisely, we will need the following assumptions on the coefficient functions A and C and on the initial condition u_0.

Assumptions 2.4. Let $A, C \in C^4[-\eta, 1 + \eta]$ with some $\eta > 0$ satisfying $A + C > 0$, $A(1) = C(0) = 0$. Assume that $A - C$ is positive on $[-\eta, 0]$ and negative on $[1, 1 + \eta]$. Moreover, let $N_0 > 1/2\eta$ be a positive integer such that $2N_0|A(x) - C(x)| > |A'(x) + C'(x)|$ for all $x \in [-\eta, 0] \cup [1, 1 + \eta]$.

Assumptions 2.5. Let $u_0 \in C^2[0, 1]$ be a non-negative function satisfying $u_0(0) = u_0'(0) = u_0(1) = u_0'(1) = 0$, and let $u_0^N \in C^2[-h, 1 + h]$ be obtained as its extension as constant 0 outside of $[0, 1]$.

Theorem 2.6. *Suppose that Assumptions 2.4 and 2.5 hold. Let the coefficients of* (2.1) *be given by* (2.2), *and let q_k be the solution of* (2.1) *satisfying the initial conditions $q_k(0) = u_0(k/N)$ for all $k \in \{0, 1, 2, \ldots, N\}$. Let u^N be the solution of the Fokker–Planck equation* (2.15) *subject to the boundary condition* (2.16)–(2.17) *and initial condition $u^N(0, \cdot) = u_0^N(\cdot)$. Then for any $t_0 > 0$ there exists a constant K independent of N, for which*

$$\left| u^N \left(t, \frac{k}{N} \right) - q_k(t) \right| \leq K, \qquad t \in [0, t_0], \quad k = 0, 1, 2, \ldots, N.$$

The proof, which is based on a Trotter–Kato type result in the context of operator semigroups, is shown in Section 6.

Note that q_k is not a proper distribution since $\sum q_k \neq 1$, however, the above theorem translates easily to a statement concerning the distribution p_k determined by (2.1). Namely, Q_N, given in (2.19), relates q_k to p_k and u^N to v as follows. The functions q_k and p_k are solutions of the same system of linear differential equations (2.1), and they are scalar multiples of each other. According to (2.19) the relation $p_k = q_k/Q_N$ holds for all k. Similarly, u^N and v are

solutions of the Fokker–Planck equation (2.15) belonging to different initial conditions, hence they are related through $v(t, z) = u^N(t, z)/Q_N$. Moreover,

$$Q_N = \sum_{\ell=0}^{N} u_0^N(\ell/N) = \sum_{\ell=0}^{N} u_0(\ell/N) = N \int_0^1 u_0 + o(N),$$

hence we have the approximation of order $1/N$ for $p(t)$ and $v(t, z)$:

$$\max_{0 \le k \le N} |v(t, k/N) - p_k(t)| = \max_{0 \le k \le N} \left| \frac{u^N(t, k/N)}{Q_N} - \frac{q_k(t)}{Q_N} \right| \le \frac{K}{Q_N} \le \frac{K'}{N}.$$

Also, the boundary of the Fokker–Planck equation was chosen in such a way that the integral of the function u^N is constant in time. Therefore, the approximation result given in Theorem 2.6 yields the following estimate between p_k and v.

Corollary 2.7. *Suppose that Assumptions 2.4 and 2.5 hold. Let the coefficients of (2.1) be given by (2.2), and let p_k be the solution of (2.1) satisfying the initial conditions $p_k(0) = u_0(k/N)/Q_N$ for all $k \in \{0, 1, 2, \dots, N\}$, with Q_N given in (2.19). Let v be the solution of the Fokker–Planck equation (2.15) subject to the boundary condition (2.16)–(2.17) and initial condition $v(0, \cdot) = u_0^N(\cdot)/Q_N$. Then for any $t_0 > 0$ there exists a constant K independent of N, for which*

$$\left| v\left(t, \frac{k}{N}\right) - p_k(t) \right| \le \frac{K}{N}, \qquad t \in [0, t_0], \quad k = 0, 1, 2, \dots, N.$$

3   An elementary approximation result for the moments

In this section we show an elementary proof of Theorem 2.3 in the case when A and C are quadratic polynomials, following the ideas of [3]. In that case the first two moments satisfy the differential equations (2.8)–(2.9) and, using (2.6), the mean-field approximation satisfies

$$\dot{y}_1 = D_0 + D_1 y_1 + D_2 y_1^2. \tag{3.1}$$

This approximation is based on the assumption that the distribution p_k is concentrated to a single point, hence the second moment can be approximated as $m_2 \approx m_1^2$. Thus the accuracy of the approximation can be measured in terms of $m_2 - m_1^2$. This motivates to introduce the function

$$h(t) = \frac{m_3(t) - m_1^3(t)}{m_2(t) - m_1^2(t)}$$

and rewrite system (2.8)–(2.9) as follows.

$$\dot{m}_1 = D_0 + D_1 m_1 + D_2 m_1^2 + D_2(m_2 - m_1^2), \tag{3.2}$$

$$\dot{m}_2 = 2m_1(D_0 + D_1 m_1 + D_2 m_1^2) + 2(D_1 + D_2 h)(m_2 - m_1^2) + \frac{1}{N}(E_0 + E_1 m_1 + E_2 m_2). \tag{3.3}$$

Observe that the derivative of m_1^2 is $2m_1(D_0 + D_1 m_1 + D_2 m_1^2 + D_2(m_2 - m_1^2))$ that coincides with the first term in the right-hand side of the second differential equation if $m_2 = m_1^2$. This suggests that it might be useful to introduce the functions w_1 and w_2 satisfying the non-autonomous system (note that h is a time-dependent function)

$$\dot{w}_1 = D_0 + D_1 w_1 + D_2 w_1^2 + D_2(w_2 - w_1^2), \tag{3.4}$$

$$\dot{w}_2 = 2w_1(D_0 + D_1 w_1 + D_2 w_1^2) + 2(D_1 + D_2 h)(w_2 - w_1^2). \tag{3.5}$$

As the observation above shows, the pair $w_1 = y_1$, $w_2 = y_1^2$ solves this system, since the derivative of y_1^2 is $2y_1(D_0 + D_1 y_1 + D_2 y_1^2)$. Now the estimate $|y_1 - m_1| \leq K/N$ can be easily derived from Peano's inequality.

Lemma 3.1 (Peano's inequality). *Suppose that $f, g \colon [0, t_0] \times \Omega \to \mathbb{R}^k$ are Lipschitz continuous functions in their second variable with Lipschitz constant L and $|f(t, x) - g(t, x)| \leq M$ in $[0, t_0] \times \Omega$ with some constant M, where $\Omega \subset \mathbb{R}^k$ is a compact set. If u and v solve the differential equations $\dot{u}(t) = f(t, u(t))$ and $\dot{v}(t) = g(t, v(t))$ and satisfy the same initial condition, $u(0) = v(0)$, then*

$$|u(t) - v(t)| \leq \frac{M}{L} \left(e^{Lt} - 1 \right) \qquad (t \in [0, t_0]).$$

In order to apply this lemma we need the boundedness of h, which follows from the next proposition.

Proposition 3.2. *Let X be a random variable with values in $[0, 1]$. Then*

$$0 \leq E(X^3) - E^3(X) \leq 3(E(X^2) - E^2(X)),$$

where $E(X)$ denotes the expected value of X.

The statement is a direct consequence of Jensen's inequality, see Lemma 4.1, for the function $\varphi(x) = x^3 - 3x^2$, which is concave in the interval $[0, 1]$.

This proposition implies that $0 \leq h(t) \leq 3$ for all t, implying that the right-hand sides of systems (3.2)–(3.3) and (3.4)–(3.5) are Lipschitz continuous. Moreover, the difference of their right-hand sides is not greater than $(|E_0| + |E_1| + |E_2|)/N$. Hence Peano's inequality yields that for any t_0 there is a constant K, such that $|m_1(t) - w_1(t)| \leq K/N$ for all $t \in [0, t_0]$, which implies Theorem 2.3 since $w_1 = y_1$.

4   Upper and lower bounds for the moments

Now we show that not only an approximation but also lower and upper bounds can be given to the expected value $m_1(t)$ in terms of simple differential equations. This will be carried out when A and C are quadratic polynomials with $A_2 \leq C_2$, however, the proof can be generalized to higher degree polynomials with appropriate conditions on the coefficients, see [1]. One of the main ideas is to apply Jensen's inequality to get an estimate of the second moment in terms of the first one. The classical Jensen's inequality [8] and the definition of the expected value yields the probabilistic version of Jensen's inequality.

Lemma 4.1 (Jensen's inequality). *If X is a random variable and $\varphi \colon \mathbb{R} \to \mathbb{R}$ is a convex function, then*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

For concave φ, the reverse inequality holds.

Apply Jensen's inequality for the random variable X, for which $\mathbb{P}(X(t) = k) = p_k(t)$ with a fixed value of t and for $k \in \{0, 1, \ldots, N\}$. If φ is chosen as $\varphi(x) = x^n$, then due to Jensen's inequality,

$$m_1^n = (\mathbb{E}[X/N])^n \leq \mathbb{E}[(X/N)^n] = m_n \quad \text{for } n \geq 1. \tag{4.1}$$

Hence assuming $D_2 \leq 0$, the differential equation of the first moment, (2.8), yields

$$\dot{m}_1(t) = D_0 + D_1 m_1 + D_2 m_2 \leq D_0 + D_1 m_1 + D_2 m_1^2. \tag{4.2}$$

Recall that the mean field equation (2.6) is $\dot{y}_1(t) = D_0 + D_1 y_1 + D_2 y_2$, that is m_1 satisfies the inequality with the same right-hand side as that of the differential equation of y_1. Then the following comparison result yields that $m_1(t) \leq y_1(t)$.

Lemma 4.2 (Comparison). *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a continuous function. Assume that*

- *the initial value problem $\dot{x}_2(t) = f(t, x_2(t))$, $x_2(0) = x_0$ has a unique solution for $t \in [0, T]$;*

- *$\dot{x}_1(t) \leq f(t, x_1(t))$ for $t \in [0, T]$; and $x_1(0) \leq x_0$.*

Then $x_1(t) \leq x_2(t)$ for $t \in [0, T]$.

The result is standard in the theory of ODEs, see e.g. [11].

Let us turn now to the derivation of a lower bound for the expected value m_1. On one hand, we have $m_1^2 \leq m_2$ yielding $m_1 \leq m_2^{1/2}$. On the other hand, applying again Jensen's inequality for the random variable $(X/N)^2$ and for $\varphi(x) = x^{3/2}$, yields $m_2^{3/2} \leq m_3$. Using the differential equation of the second moment (2.9), and assuming the sign conditions $D_0 \geq 0$ and $D_2 \leq 0$ yield

$$\dot{m}_2(t) \leq 2(D_0 m_2^{1/2} + D_1 m_2 + D_2 m_2^{3/2}) + \frac{1}{N} E_m, \tag{4.3}$$

where $E_m = |E_0| + |E_1| + |E_2|$. The right-hand side of this inequality motivates the introduction of the functions z_1 and z_2 as solutions of the system of differential equations given below

$$\dot{z}_1 = D_0 + D_1 z_1 + D_2 z_2 \tag{4.4}$$

$$\dot{z}_2 = 2(D_0 z_2^{1/2} + D_1 z_2 + D_2 z_2^{3/2}) + \frac{1}{N} E_m, \tag{4.5}$$

and satisfying the initial condition $z_j(0) = m_1^j(0)$, for $j = 1, 2$. Hence the comparison lemma yields that $m_2 \leq z_2$. These inequalities enable us to compare z_1 to m_1. Namely, substituting $m_2 \leq z_2$ in (2.8) and exploiting the sign condition $D_2 \geq 0$ leads to

$$\dot{m}_1 \geq D_0 + D_1 m_1 + D_2 z_2. \tag{4.6}$$

Considering the function z_2 as fixed, then (4.6) and (4.5) have the form $\dot{m}_1(t) \geq g(t, m_1(t))$ and $\dot{z}_1(t) = g(t, z_1(t))$ where $g(t, x) = D_0 + D_1 x + D_2 z_2(t)$. Thus, the comparison lemma implies $m_1 \geq z_1$. Summarising, the following theorem holds.

Theorem 4.3. *Consider the master equation (2.1) with density dependent coefficients given in (2.2) and assume that the functions A and C are quadratic polynomials given in (2.7). Let $D_j = A_j - C_j$ be the difference of the coefficients of the polynomials. Let m_1 be the first moment, solving (2.8), y_1 be the mean-field approximation, solving (3.1) and satisfying the same initial condition as m_1, i.e. $y_1(0) = m_1(0)$ and let z_j (for $j = 1, 2$) be the solution of (4.4)–(4.5) subject to the initial condition $z_j(0) = m_1^j(0)$. If $D_0 \geq 0$ and $D_2 \leq 0$, then*

$$z_1(t) \leq m_1(t) \leq y_1(t) \quad and \quad m_2(t) \leq z_2(t)$$

hold for all $t \geq 0$.

We note that it can be also proved that the difference of y_1 and z_1 is of order $1/N$ as N tends to infinity, implying that y_1 is an order $1/N$ approximation of the expected value m_1, see in [1]. In that paper the question of upper and lower bounds is dealt with in more detail, the performance of the bounds is illustrated with numerical examples and the necessity of the sign condition is discussed.

5   Higher order closure based on an a priori distribution

The idea for deriving the simplest mean-field approximation (3.1) starting from the exact equation (2.8) was to use the approximation $m_2 \approx m_1^2$. Now, to get a more accurate approximation, by keeping the differential equation of the first moment exact and by using the closure approximation in the differential equation of the second moment m_2. These kind of algebraic relations between the moments are referred to as higher order closures. The calculation can be carried out relatively easily in the quadratic case, when A and C are quadratic polynomials given in (2.7). In that case the exact differential equations for the first two moments are (2.8)–(2.9). These equations are not closed or self-contained since the second moment depends on the third one and an equation for this is also needed. The novel closure put forward here is based on the empirical observation that $p_k(t)$ is well approximated by a binomial or normal distribution that can also be justified based on stochastic arguments. In the case of the normal distribution $\mathcal{N}(\mu, \sigma^2)$, the parameters μ and σ depend on time and will be specified in terms of the moments of the distribution. The first three moments of the normal distribution can be specified easily in terms of the two parameters and are as follows,

$$M_1 = \mu, \qquad M_2 = \mu^2 + \sigma^2, \qquad M_3 = \mu^3 + 3\mu\sigma^2.$$

The third moment can easily be expressed in terms of the first two moments as

$$M_3 = M_1^3 + 3M_1(M_2 - M_1^2) = 3M_1 M_2 - 2M_1^3.$$

This relation defines the new closure. Using the equations for the first two moments (2.8)–(2.9) and the closure at the level of the third moment yields the new approximating system in the form

$$\dot{x}_1 = D_0 + D_1 x_1 + D_2 x_2,$$
$$\dot{x}_2 = 2(D_0 x_1 + D_1 x_2 + D_2 x_3) + \frac{1}{N}(E_0 + E_1 x_1 + E_2 x_2),$$
$$x_3 = 3x_1 x_2 - 2x_1^3.$$

Summarising, the idea for deriving the higher order closure is to assume an a priori distribution for p_k that leads to an expression for the third moment in terms of the first two moments. As an alternative to normal distribution, one can approximate the distribution p_k with a binomial distribution. Then a similar derivation yields the third moment in terms of the first two moments. Hence an alternative closed system, based on the binomial distribution approximation can be derived, see [13]. These new closed systems were introduced for the case of SIS epidemic propagation in [13], where their performance was also investigated in detail. Extensive numerical study showed that these closures give order $1/N^2$ approximation for the moments, in contrast to the $1/N$ accuracy of the usual mean-field approximation given by (3.1) and the pairwise approximation given by the widely used triple closure. We note that the order $1/N^2$ accuracy for the binomial and normal closures still awaits formal proof.

6   Operator semigroup approach

In this section, we show the main steps of the proof of Theorem 2.6, which relies on operator semigroup theory.

The system of ODEs (2.1) can be written in the form $\dot{p} = A_N p$, where $p = (p_0, p_1, \ldots, p_N)^T$ and A_N is a tridiagonal matrix. The solution of the system can be given as $p(t) = T_N(t)p(0)$, where $T_N(t) = \exp(A_N t)$ is an operator semigroup on $\mathbb{R}^{N+1}$. (We note that it is extendable to $\mathbb{C}^{N+1}$ in a the usual way.) We will show that the solution of the Fokker–Planck equation (2.15) can also be given by using an operator semigroup as $u(t, \cdot) = S_N(t)u(0, \cdot)$. Then we estimate the difference of the solutions by using a Trotter–Kato type result claiming that the semigroups are close to each other if this is known about their generators.

The generator of S_N is given by the right-hand side of the Fokker–Planck equation and will be denoted as

$$\mathcal{A}_N f = \frac{1}{2N}((A+C)f)'' - ((A-C)f)'. \tag{6.1}$$

Carrying out the differentiations and using the boundary conditions (2.16)–(2.17), we get that the domain of this operator is the following subspace of the space of twice continuously differentiable functions

$$D(\mathcal{A}_N) := \left\{ f \in C^2[-h, 1+h] : h((A+C)f)'(z) - ((A-C)f)(z) = 0 \text{ for } z = -h, 1+h \right\},$$

where $h = -1/2N$. Now we introduce the general framework.

6.1 Perturbation result in the abstract setting

Assumptions 6.1. Let $\mathcal{X}_n$, X_n ($n \in \mathbb{N}^+$) be Banach spaces and assume that $P_n : \mathcal{X}_n \to X_n$ are bounded linear operators with $\|P_n\| \leq K$ for some constant $K > 0$. Suppose that the operators A_n, $\mathcal{A}_n$ generate strongly continuous semigroups $(T_n(t))_{t \geq 0}$ and $(S_n(t))_{t \geq 0}$ on X_n and $\mathcal{X}_n$, respectively, and that there are constants $M \geq 0$, $\omega \in \mathbb{R}$ such that the stability condition

$$\|T_n(t)\| \leq M e^{\omega t} \quad \text{holds for all } t \geq 0. \tag{6.2}$$

Under these assumptions, we have the following Trotter–Kato type approximation result (cf., e.g. [4, Proposition 3.8], where the result is stated for $p > 0$, but as seen below, the arguments also hold for $p = 0$).

Proposition 6.2. *Suppose that Assumptions 6.1 hold, that there is a dense subset $Y_n \subset D(\mathcal{A}_n)$ invariant under the semigroup S_n such that $P_n Y_n \subset D(A_n)$, and that Y_n is a Banach space with some norm $\|\cdot\|_{Y_n}$ satisfying*

$$\|S_n(t)\|_{Y_n} \leq M e^{\omega t}.$$

Let further $f \in Y_n$. If there exists a constant $p \geq 0$ with the property that for any $\tau \geq 0$ there exists a $C > 0$ such that for all $\tau \geq t \geq 0$ the estimate

$$\|A_n P_n S_n(t) f - P_n \mathcal{A}_n S_n(t) f\|_{X_n} \leq C \frac{\|S_n(t)f\|_{Y_n}}{n^p}, \tag{6.3}$$

holds, then for each $\tau \geq 0$ there exists some $C' > 0$ such that

$$\|T_n(t)P_n f - P_n S_n(t) f\|_{X_n} \leq C' \frac{\|f\|_{Y_n}}{n^p}$$

for all $0 \leq t \leq \tau$, where C' depends only on C, τ, M and ω.

The statement can be verified as follows. Let $f \in Y_n$, then the function $[0, t] \ni s \mapsto T_n(t-s)P_n S_n(s)f$ is continuously differentiable with derivative

$$T_n(t-s)P_n \mathcal{A}_n S_n(s)f - T_n(t-s)A_n P_n S_n(s)f = T_n(t-s)(P_n \mathcal{A}_n - A_n P_n)S_n(s)f,$$

and the fundamental theorem of calculus yields

$$P_n S_n(t)f - T_n(t)P_n f = \int_0^t T_n(t-s)(P_n \mathcal{A}_n - A_n P_n)S_n(s)f \, \mathrm{d}s.$$

Hence we have

$$
\begin{aligned}
\|T_n(t)P_n f - P_n S_n(t)f\|_{X_n} &\leq \int_0^t \|T_n(t-s)(P_n \mathcal{A}_n - A_n P_n)S_n(s)f\|_{X_n} \, \mathrm{d}s \\
&\leq \int_0^t M e^{\omega(t-s)} \|(P_n \mathcal{A}_n - A_n P_n)S_n(s)f\|_{X_n} \, \mathrm{d}s \\
&\leq \int_0^t M e^{\omega(t-s)} C \frac{\|S_n(s)f\|_{Y_n}}{n^p} \, \mathrm{d}s \leq \int_0^t M e^{\omega(t-s)} C \frac{M e^{\omega s} \|f\|_{Y_n}}{n^p} \, \mathrm{d}s \\
&\leq C' \frac{\|f\|_{Y_n}}{n^p}
\end{aligned}
$$

with $C' = M^2 C \tau e^{\tau|\omega|}$.

6.2 Proof of Theorem 2.6

Now we turn to how this abstract setting applies to our case. Let Assumptions 2.4 and 2.5 hold. For each $N \geq N_0$, choose $X_N := (\mathbb{C}^{N+1}, \|\cdot\|_\infty)$ and $\mathcal{X}_N := C[-h, 1+h]$, with P_N projecting $f \in \mathcal{X}_n$ onto the vector

$$(f(0), f(1/N), \ldots, f(1))^T \in X_N.$$

Clearly $\|P_N\| = 1$. Let further A_N be the transition matrix pertaining to the system of equations (2.1) and $T_N(t) = \exp(A_N t)$. The operator $(\mathcal{A}_N, D(\mathcal{A}_N))$ given by (6.1) generates the analytic operator semigroup $(S_N(t))_{t \geq 0}$ on $\mathcal{X}_N$ that gives the solutions of PDE (2.15) with boundary conditions (2.16)–(2.17), cf. [9, Section VI.4.b]. Since the semigroup is analytic, it leaves $D(\mathcal{A}_N)$ invariant. Thus using the notations of Theorem 2.6 we have

$$q_k(t) = (T_n(t)P_n u^N(0, \cdot))_k, \quad \text{and} \quad u^N(t, k/N) = (P_n S_n(t)u^N(0, \cdot))_k,$$

where the subscript k refers to the k-th coordinate. Now we formulate three lemmas that will allow us to verify that the conditions of Proposition 6.2 hold with p=0, thus the statement of the Theorem follows from it. The proofs of the Lemmas are too technical to be included here, these will be published separately. The first lemma is about the growth bounds of the semigroups in question, together with some of their restrictions.

Lemma 6.3. *Suppose that Assumptions 2.4 hold. Then there exists a constant $d > 0$ such that for any $N \geq N_0$, the following hold:*

1. *the space $Y_N := (D(\mathcal{A}_N), \|\cdot\|_{\mathcal{A}_N - dI})$ is a Banach space with the norm*

$$\|f\|_{\mathcal{A}_N - dI} := \|(\mathcal{A}_N - dI)f\|_{\mathcal{X}_N};$$

2. *for all $t \geq 0$ the following norms are all bounded from above by 1:*

$$\|e^{-dt}T_N(t)\|_{X_N}; \quad \|e^{-dt}S_N(t)\|_{X_N}; \quad \left\|e^{-dt}S_N(t)\Big|_{Y_N}\right\|_{Y_N}; \quad \left\|e^{-dt}S_N(t)\Big|_{Z_N}\right\|_{Z_N},$$

where $Z_N = \{f \in C^1[-h, 1+h] : h((A+C)f)'(z) - ((A-C)f)(z) = 0 \text{ for } z = -h, 1+h\}$.

Next, concerning the error bound between the generators, the second degree Taylor expansion with Lagrange remainder term will be used to estimate the left hand side of (6.3) (writing f instead of $S_N(f)$ for simplicity):

$$\max\{(A_N P_N f - P_N \mathcal{A}_N f)_k : k = 0, 1, \ldots, N\}.$$

Using the tridiagonal form of the matrix A_N, the first term can be written as

$$(A_N P_N f)_k = a_{k-1} f\left(\frac{k-1}{N}\right) - (a_k + c_k) f\left(\frac{k}{N}\right) + c_{k+1} f\left(\frac{k+1}{N}\right).$$

Exploiting the density dependence (2.2) this can be artificially rearranged to

$$(A_N P_N f)_k = \frac{N}{2}\left(((A+C)f)\left(\frac{k-1}{N}\right) - 2((A+C)f)\left(\frac{k}{N}\right) + ((A+C)f)\left(\frac{k+1}{N}\right)\right)$$
$$+ \frac{N}{2}\left((A-C)f)\left(\frac{k-1}{N}\right) - ((A-C)f)\left(\frac{k+1}{N}\right)\right).$$

The second degree Taylor formula with Lagrangian remainder will be used in the form

$$F(z + \Delta z) = F(z) + F'(z)\Delta z + F''(z+\zeta)\frac{\Delta z^2}{2},$$

where ζ is between 0 and Δz. This will be applied with $z = k/N$, $\Delta z = 1/N$, $\Delta z = -1/N$, $F = (A+C)f$ and $F = (A-C)f$ leading to

$$(A_N P_N f)_k = \frac{1}{4N}\left(((A+C)f)''\left(\frac{k}{N} - \zeta_1\right) + ((A+C)f)''\left(\frac{k}{N} + \zeta_2\right)\right)$$
$$- ((A-C)f)'\left(\frac{k}{N}\right) + \frac{1}{4N}\left(((A-C)f)''\left(\frac{k}{N} - \zeta_3\right) - ((A-C)f)''\left(\frac{k}{N} + \zeta_4\right)\right)$$

with ζ_i is between zero and $1/N$.

Now using (6.1) we have

$$(P_N \mathcal{A}_N f)_k = \frac{1}{2N}((A+C)f)''\left(\frac{k}{N}\right) - ((A-C)f)'\left(\frac{k}{N}\right),$$

hence the difference of the two generators can be expressed as

$$(A_N P_N f)_k - (P_N \mathcal{A}_N f)_k = \frac{1}{4N}\left(((A+C)f)''\left(\frac{k}{N} - \zeta_1\right) - ((A+C)f)''\left(\frac{k}{N}\right)\right)$$
$$+ \frac{1}{4N}\left((((A+C)f)''\left(\frac{k}{N} + \zeta_2\right) - ((A+C)f)''\left(\frac{k}{N}\right)\right)$$
$$+ \frac{1}{4N}\left(((A-C)f)''\left(\frac{k}{N} - \zeta_3\right) - ((A-C)f)''\left(\frac{k}{N} + \zeta_4\right)\right).$$

The difference can be estimated as follows:

$$|(A_N P_N f)_k - (P_N \mathcal{A}_N f)_k| \leq \frac{1}{2N}(2\|(A+C)f\|_{C^2} + \|(A-C)f\|_{C^2}).$$

We note, that the estimate could be of order $1/N^2$ if the C^3 norm of the functions was used. However, the analogous calculation carried out at the boundary points yields only order $1/N$ difference, hence the sharper estimate at the inner points is not used. On the boundary, we have to deal with the "virtual" mesh points at $-2h$ and $1 + 2h$, respectively. To do so, we first extend the function $(A + C)f$ (independently from the already existing extension of the functions A and C) beyond our interval $[-h, 1 + h]$ whilst keeping it C^2. Note that this can be done without increasing the C^2 norm by more than a constant factor. This can be carried out for the function $(A - C)f$ in a similar way. Then we discretize the boundary condition at the endpoints using the Taylor expansion, and can thereby replace the function values at the virtual mesh points $-2h$ and $1 + 2h$ in terms of the function values and derivatives at 0 and 1, respectively. Finally, using $C(0) = A(1) = 0$, we obtain the required order of approximation.

Completing these calculations at the boundary points leads to the following statement.

Lemma 6.4. *Suppose that Assumptions 2.4 hold. Then there exists a constant C_0 such that for any $N \geq N_0$ and $f \in D(\mathcal{A}_N)$ we have*

$$\max_{0 \leq k \leq N} |(A_N P_n f)_k - (P_N \mathcal{A}_N f)_k| \leq \frac{C_0}{N}(\|(A+C)f\|_{C^2} + \|(A-C)f\|_{C^2})$$

Using Lemma 6.3, this can be supplemented by the following norm inequality.

Lemma 6.5. *Suppose that Assumptions 2.4 and 2.5 hold. For any $\tau \geq 0$ there exists a constant C_1 such that for any $N \geq N_0$ we have*

$$\|(A+C)u^N(t, \cdot)\|_{C^2} + \|(A-C)u^N(t, \cdot)\|_{C^2} \leq C_1 N \|u_0^N\|_{\mathcal{A}_N - dI}$$

for all $0 \leq t \leq \tau$.

Noticing that

$$\|u_0^N\|_{\mathcal{A}_N - dI} = \|(\mathcal{A}_N - dI)u_0^N\| \leq \|\partial_{zz}((A+C)u_0^N)\| + \|\partial_z((A-C)u_0^N)\| + d\|u_0^N\|$$
$$= \|\partial_{zz}((A+C)u_0)\| + \|\partial_z((A-C)u_0)\| + d\|u_0\| = \text{const.}$$

we have proved that (6.3) holds with $p = 0$. Hence all the conditions of Proposition 6.2 are fulfilled, and that finishes the proof of Theorem 2.6.

7 Discussion

The averaged behaviour of a system consisting of N identical particles, each of which can be in two states, was studied. The starting point of our investigations was the master equation (2.1) formulated in terms of $p_k(t)$, the probability that there are k particles in one of the states at time t. This form of the master equation is based on several restrictive assumptions that can be released giving rise to more general models.

The most obvious extension is to consider particles with more than two states. Now we briefly show how the complexity of the problem changes when each particle can be in one of three states Q, T and R. The state of the system changes when the state of a particle changes,

for which there are six possibilities: transitions $Q \to T$, $Q \to R$, $T \to Q$, $T \to R$, $R \to Q$ and $R \to T$. It is assumed that each transition rate depends on the proportion of nodes in the different states, hence the state of the whole system can be given by the number of particles of different types. The process is then a birth–death or counting process, in which the number of particles of each type changes by one in a short time interval. The state of the system can be given by the triple $(i, j, N - i - j)$ yielding the number of particles of type Q, T and R, or simply by the pair (i, j). Thus the state space consists of the lattice points (with integer coordinates) lying in the two dimensional simplex in $\mathbb{R}^3$. Then the $N \to \infty$ limit will lead to the Fokker–Planck equation with state variable lying in the two dimensional simplex. In general, if each particle is in one of m different states, then the Fokker–Planck equation will be given in the $m - 1$ dimensional simplex.

Another restrictive condition that would be desired to be released is the density dependence, namely that the transition rate depends on the proportion of nodes in the different states. Density dependence enabled us to characterize the state of the whole system by only the number of particles of different types. This condition is based on the assumption that all particles are identical. In a more realistic network context, the mutual position of the particles (nodes) in the network also affects the process, hence the particles (nodes) cannot be considered to be identical. Thus the number of nodes in a given state is not enough to describe the configuration of the system, hence the state space of the system is typically much larger. For example, in the case of two statuses for each particle (node), the total number of configurations is 2^N, compared to $N + 1$ that was used in this paper.

Finally, all transitions were assumed to be Markovian in our case, this assumption leaded us to the master equation in the form of ODEs given in (2.1). For non-Markovian processes the master equations can be delay differential equations or partial integro-differential equations [12].

The limit of the above generalizations for large system size can be the subject of future work.

Acknowledgements

References

[1] B. ARMBRUSTER, Á. BESENYEI, P. L. SIMON, Bounds for the expected value of one-step processes, *Commun. Math. Sci.* **14**(2016), No. 7, 1911–1923.

[2] B. ARMBRUSTER, E. BECK, An elementary proof of convergence to the mean-field equations for an epidemic model, *IMA J. Appl. Math.*, published online on April 13, 2016. url

[3] B. ARMBRUSTER, A simple and general proof for the convergence of Markov processes to their mean-field limits, (2016), available on arXiv:1602.05224.

[4] A. Bátkai, P. Csomós, B. Farkas, A. Ostermann, Operator semigroups for numerical analysis, *Internet-Seminar Manuscript*, 2012, `https://isem-mathematik.uibk.ac.at`.

[5] A. Bátkai, I. Z. Kiss, E. Sikolya, P. L. Simon, Differential equation approximations of stochastic network processes: an operator semigroup approach, *Netw. Heter. Media* **7**(2012), 43–58. MR2908609; url

[6] A. Bátkai, Á. Havasi, R. Horváth, D. Kunszenti-Kovács, P. L. Simon, PDE approximation of large systems of differential equations, *Oper. Matrices* **9**(2015), 147–163. MR3338556; url

[7] A. Bobrowski, *Functional analysis for probability and stochastic processes. An introduction*, Cambridge University Press, Cambridge, 2005. MR2176612

[8] Z. Cvetkovski, *Inequalities. Theorems, techniques and selected problems*, Springer Science & Business Media, Heidelberg, 2012. MR3015124

[9] K.-J. Engel, R. Nagel, *One-parameter semigroups for linear evolution equations*, Graduate Texts in Math., Vol. 194, Springer-Verlag, New York, 2000. MR1721989

[10] S. N. Ethier, T. G. Kurtz, *Markov processes. Characterization and convergence*, John Wiley & Sons, Inc., USA, 2005. MR0838085

[11] J. Hale, *Ordinary differential equations*, Dover Publications, New York, 2009. MR587488

[12] R. A. Hayden, I. Horváth, M. Telek, Mean field for performance models with generally-distributed timed transitions, in: *Quantitative Evaluation of Systems*, Lecture Notes in Computer Science, Vol. 8657, Springer, 2014, pp. 90–105. url

[13] I. Z. Kiss, P. L. Simon, New moment closures based on a priori distributions with applications to epidemic dynamics, *Bull. Math. Biol.* **74**(2012), 1501–1515. MR2930084; url

[14] T. G. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, *J. Appl. Prob.* **7**(1970), 49–58. MR0254917; url

[15] H. Risken, *The Fokker–Planck equation. Methods of solution and applications*, Springer, Berlin Heidelberg, 2012. MR987631; url

[16] M. Scott, *Applied stochastic processes in science and engineering*, University of Waterloo, eBook, 2013.

[17] P. L. Simon, I. Z. Kiss, From exact stochastic to mean-field ODE models: a new approach to prove convergence results, *IMA J. Appl. Math.* **78**(2013), 945–964. MR3116158; url

[18] N. G. van Kampen, *Stochastic processes in physics and chemistry*, Elsevier, Amsterdam, 1992. MR648937