Partial mean field limits in heterogeneous networks

Carsten Chong*, Claudia Klüppelberg

Center for Mathematical Sciences, Technical University of Munich, Boltzmannstraße 3, 85748 Garching, Germany

Abstract

We investigate systems of interacting stochastic differential equations with two kinds of heterogeneity: one originating from different weights of the linkages, and one concerning their asymptotic relevance when the system becomes large. To capture these effects, we define a partial mean field system, and prove a law of large numbers with explicit bounds on the mean squared error. Furthermore, a large deviation result is established under reasonable assumptions. The theory will be illustrated by several examples: on the one hand, we recover the classical results of chaos propagation for homogeneous systems, and on the other hand, we demonstrate the validity of our assumptions for quite general heterogeneous networks including those arising from preferential attachment random graph models.
© 2019 Elsevier B.V. All rights reserved.

1. Introduction

The application of mean field theory to large systems of stochastic differential equations (SDEs) was initiated by McKean's seminal work [26–28]. In the classical case, an N-dimensional interacting particle system is governed by SDEs of the form

$$\mathrm{d}X_i^N(t) = \frac{1}{N-1} \sum_{j \neq i} \big(X_j^N(t) - X_i^N(t)\big)\,\mathrm{d}t + \mathrm{d}B_i(t), \quad t \in \mathbb{R}_+,$$

$$X_i^N(0) = X_i(0), \quad i = 1, \ldots, N, \tag{1.1}$$

---

* Corresponding author.
    E-mail addresses: carsten.chong@tum.de (C. Chong), cklu@ma.tum.de (C. Klüppelberg).
    URL: http://www.statistics.ma.tum.de (C. Klüppelberg).

with independent starting random variables $X_i(0)$ and independent Brownian motions B_i. As the number of particles increases, the pair dependencies in this coupled system decrease with order $1/N$ such that a law of large numbers applies (see Theorem 1.4 of [36]). Defining

$$\mathrm{d}\bar{X}_i^N(t) = \frac{1}{N-1} \sum_{j \neq i} \big(\mathbb{E}[\bar{X}_j^N(t)] - \bar{X}_i^N(t) \big) \, \mathrm{d}t + \mathrm{d}B_i(t), \quad t \in \mathbb{R}_+,$$

$$\bar{X}_i^N(0) = X_i(0), \quad i = 1, \ldots, N, \tag{1.2}$$

there exists for every $T \in \mathbb{R}_+$, a constant $C(T) \in \mathbb{R}_+$ independent of N such that

$$\sup_{i=1,\ldots,N} \mathbb{E}\left[\sup_{t \in [0,T]} |X_i^N(t) - \bar{X}_i^N(t)|^2 \right]^{1/2} \leq \frac{C(T)}{\sqrt{N}}. \tag{1.3}$$

In other words, in a large system, the behaviour of a fixed number of particles evolving according to (1.1) is well described by the so-called *mean field system* (1.2), where all stochastic processes are stochastically independent, a phenomenon that is called *propagation of chaos*. Thus, mean field theory provides a model simplification by reducing a many-body problem as in (1.1) to a one-body problem as in (1.2), with explicit L^2-estimates on the occurring error. Moreover, it can be shown that the empirical measure of the particles satisfies a large deviation principle as $N \to \infty$, see [13,25]. There is a huge literature dealing with this or related topics, and we only mention the review papers [20,36].

The systems (1.1) and (1.2) describe statistically equal or exchangeable particles: any permutation of the indices $i \in \{1, \ldots, N\}$ leads to a system with the same distribution (cf. [40]). In particle physics such an assumption is certainly reasonable and underlies many other similar models of mean field type, see for example the two treatises [37,38] for numerous examples.

However, when mean field models are considered in applications other than statistical mechanics, the homogeneity assumption may not be appropriate in all situations. For instance, in [9,22] the processes (1.1) are used to model the wealth of trading agents in an economy, who are typically far from being equal in their trading behaviour (there are "market makers" and others). Similarly, the stochastic Cucker–Smale model that is considered in [2,7] describes the "flocking" phenomenon of individuals. Also here, it seems natural that one or several "leaders" may have a distinguished role, setting them apart from the remaining system. Moreover, in systemic risk modelling, the particles represent financial institutions that interact with each other through mutual exposures, see [6,18,23] for some approaches in this direction. The different players in the banking sector vary considerably in size and importance, which is obvious from the fact that some banks were considered too big to fail during the financial crisis of 2007–08. Further fields of applications where mean field theory is used for interacting particle systems include genetic algorithms [29], neuron modelling (see [19] and references therein) and epidemics modelling [24].

Partly triggered by the examples in the previous paragraph, this paper aims to investigate deviations from homogeneous systems to heterogeneous systems. First, we allow for different interaction rates between pairs (instead of $1/(N-1)$ throughout), and second, we permit the subsistence of a core–periphery structure in the mean field limit, that is, some particles may have a non-vanishing influence even when the system becomes large. Another restriction we will relax in our analysis concerns the driving noises of the interacting SDEs: instead of independence we explicitly allow for different degrees of dependence in the noise terms, even asymptotically. Until now there is only a small amount of literature that generalizes (1.1) in these directions: in [4,11,23,31,32,39] the particles are divided into different groups within which

they are homogeneous (and the number of members in each group must tend to infinity for the law of large numbers), and [10,12], where one major agent exists and propagation of chaos for the minor agents is considered conditioned on the major one. Other papers that consider general heterogeneous systems include [14], where the propagation of chaos result is *assumed*, and [16,17,21], where a law of large numbers for the empirical measure is proved under various conditions. Regarding the last mentioned papers, two aspects are worth commenting on. First, assuming that finitely many core particles do exist in the system, their contribution to the empirical distribution becomes less and less as $N \to \infty$ although their impact may very well stay high. Thus, in this case the empirical distribution may fail to describe the behaviour of the system as a whole. Second, whereas for homogeneous systems the convergence of the empirical measure is equivalent to the existence of a mean field limit in the sense of (1.3) (see e.g. Proposition 2.2(i) of [36]), this is no longer true for heterogeneous systems. For core particles, the left-hand side of (1.3) need not converge to 0 even if the empirical distribution converges, say, to a deterministic limit. For example, in the case of [10,12] with one core particle, an unconditional propagation of chaos result does not hold for this particle without further assumptions (even if it does for the periphery particles).

Due to the two aforementioned reasons, we will *not* work with the empirical distribution in this paper but state and prove mean field limit theorems for the particles on the process level. In Section 2 we start by introducing the precise interacting particle model we want to investigate. Then we define a corresponding *partial mean field model*, for which we prove a law of large numbers type result (Theorem 3.1) with explicit convergence rates in Section 3. It generalizes (1.3) by taking into account the different kinds of heterogeneity due to varying pair interaction rates, a distinction between important/core and less important/periphery pair relationships, and interdependencies between the driving noise terms.

The main difficulty here is to identify the correct rates that govern the distance between the original system and the mean field approximation. As we will see, a total of twelve rates is required, each expressing a connectivity property of the underlying interaction and correlation networks. This is inevitable in contrast to [10,12] where the stochastic dependencies among the particles are annihilated simply by conditioning. In order to elucidate the meaning of each rate, we discuss three exemplary situations in detail. In Section 3.1, in particular in Example 3.4, we show that in the quasi-homogeneous case, all twelve rates typically boil down to a single rate like in (1.3). In Section 3.2, we explain why the prerequisites for Theorem 3.1 in the heterogeneous case are essentially sparsity assumptions on the particle network, which are satisfied for instance if this network is generated from a preferential attachment mechanism, see Section 3.3. In order to show the last statement, we have to derive the asymptotics of the maximal in- and out-degrees of directed preferential attachment graphs, see Lemma 3.8. This result may be of independent interest and generalizes that of [30] for undirected graphs.

The second main result of our paper is a large deviation principle for the difference $X^N - \bar{X}^N$, which is presented in Section 4 as Theorem 4.1. In contrast to homogeneous systems, where such a principle is proved for the empirical measure (see [13,25]), we work on the process level again and therefore need to require the existence of all exponential moments. Section 5 contains the proofs.

2. The model

Before we introduce the model we analyse in this paper, we list a number of notations that will be employed throughout the paper.

$\mathbb{R}_+$	The set $[0, \infty)$ of *positive* real numbers;
$[z]$	The largest integer smaller or equal to $z \in \mathbb{R}$;
$\mathbb{N}$	The natural numbers $\{1, 2, \ldots\}$;
A, x	The typical notation for a matrix $A = (A_{ij} : i, j \in \mathbb{N}) \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ and a vector $x = (x_i : i \in \mathbb{N})' \in \mathbb{R}^{\mathbb{N}}$, with all binary relations such as $\leq$, or operations relying on them such as the absolute value $|\cdot|$ or taking the supremum being understood componentwise when applied to matrices and vectors;
$(\cdot)'$	The transposition operator;
AB, Ax, e^A	Matrix–matrix and matrix–vector multiplication and the matrix exponential, all defined in analogy to the finite-dimensional case, provided that the involved series converge;
$x.y$	The entrywise product $x.y = (x_i y_i : i \in \mathbb{N})'$ for $x, y \in \mathbb{R}^{\mathbb{N}}$;
$|A|_\infty, |x|_\infty$	$|A|_\infty := \sup_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} |A_{ij}|$ and $|x|_\infty := \sup_{i \in \mathbb{N}} |x_i|$ for $A \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ and $x \in \mathbb{R}^{\mathbb{N}}$;
$|A|_\mathrm{d}$	$|A|_\mathrm{d} := \sup_{i \in \mathbb{N}} |A_{ii}|$ for matrices A;
$A^\times$	The matrix A with all diagonal entries set to 0;
I	The identity matrix in $\mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ or $\mathbb{R}^{d \times d}$ for some $d \in \mathbb{N}$;
L^p	The space $L^p(\Omega, \mathcal{F}, \mathbb{P})$, $p \in [1, \infty]$, endowed with the topology induced by $\|X\|_{L^p} := \mathbb{E}[|X|^p]^{1/p}$, and to be understood entrywise when applied to matrix- or vector-valued random variables;
$\mathbb{E}[X], \mathrm{Var}[X]$	Componentwise expectation and variance for random variables in $\mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ or $\mathbb{R}^{\mathbb{N}}$;
$\mathrm{Cov}[X, Y], \mathrm{Cov}[X]$	The matrices whose (ij)th entry is $\mathrm{Cov}[X_i, Y_j]$ and $\mathrm{Cov}[X_i, X_j]$, respectively, when X and Y are random vectors;
x^*	$x^*(t) := \sup_{s \in [0, t]} |x(s)|$ for $t \in \mathbb{R}_+$ and functions $x : \mathbb{R}_+ \to \mathbb{R}$, again considered entrywise when x takes values in $\mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ or $\mathbb{R}^{\mathbb{N}}$;
D_T^d, D_T^∞	The space of $\mathbb{R}^d$-valued (resp. $\mathbb{R}^{\mathbb{N}}$-valued) functions on $[0, T]$ whose coordinates are all càdlàg functions;
C_T^d, C_T^∞	Elements of D_T^d and D_T^∞ where each coordinate is a continuous function;
AC_T^d, AC_T^∞	Elements of D_T^d and D_T^∞ where each coordinate is an absolutely continuous function;
$\mathcal{D}_T^d, \mathcal{D}_T^\infty$	The σ-field on D_T^d (resp. D_T^∞) generated by the evaluation maps $\pi_t(x) = x(t)$, $x \in D_T^d$ (resp. D_T^∞), for $t \in [0, T]$;
U, J_1	The uniform topology and the Skorokhod topology on D_T^d and D_T^∞ (in the latter case they are defined via the product of the d-dimensional topologies);
M_T^d	The space of all $(\theta_1, \ldots, \theta_d)$ where each θ_i is a signed Borel measure on $[0, T]$ of finite total variation $|\theta_i|([0, T])$

Given a stochastic basis $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}(t))_{t \in \mathbb{R}_+}, \mathbb{P})$ satisfying the usual hypotheses of completeness and right-continuity, we investigate a network described by the following *interacting*

particle system (IPS):

$$
\begin{aligned}
\mathrm{d}X_i(t) = {} & \sum_{j=1}^{\infty} a_{ij}(t) X_j(t) \, \mathrm{d}t + \sum_{j=1}^{\infty} \sigma_{ij}(t) X_j(t-) \, \mathrm{d}L_i(t) + \sum_{j=1}^{\infty} f_{ij}(t) \, \mathrm{d}B_j(t) \\
& + \sum_{j=1}^{\infty} \rho_{ij}(t) \, \mathrm{d}M_j(t), \quad t \in \mathbb{R}_+, \quad i \in \mathbb{N},
\end{aligned}
\tag{2.1}
$$

subject to some $\mathcal{F}(0)$-measurable $\mathbb{R}^{\mathbb{N}}$-valued initial condition $X(0)$. We will also use the more compact form

$$
\mathrm{d}X(t) = a(t) X(t) \, \mathrm{d}t + \sigma(t) X(t-).\mathrm{d}L(t) + f(t) \, \mathrm{d}B(t) + \rho(t) \, \mathrm{d}M(t), \quad t \in \mathbb{R}_+,
\tag{2.2}
$$

for (2.1). The ingredients satisfy the following conditions:

- The two measurable functions $t \mapsto a(t)$ and $t \mapsto \sigma(t)$ are decomposed into $a = a^{\mathrm{C}} + a^{\mathrm{P}}$ and $\sigma = \sigma^{\mathrm{C}} + \sigma^{\mathrm{P}}$ such that for all $T \in \mathbb{R}_+$ and $i, j \in \mathbb{N}$,

$$
A_{ij}^{\diamond}(T) := \sup_{t \in [0,T]} |a_{ij}^{\diamond}(t)| < \infty, \quad \Sigma_{ij}^{\diamond}(T) := \sup_{t \in [0,T]} |\sigma_{ij}^{\diamond}(t)| < \infty, \quad \diamond \in \{\mathrm{C}, \mathrm{P}\}. \tag{2.3}
$$

  We define $A_{ij}(T) := A_{ij}^{\mathrm{C}}(T) + A_{ij}^{\mathrm{P}}(T)$ and $\Sigma_{ij}(T) := \Sigma_{ij}^{\mathrm{C}}(T) + \Sigma_{ij}^{\mathrm{P}}(T)$.
- L is an $\mathbb{R}^{\mathbb{N}}$-valued $\mathbb{F}$-Lévy process (i.e. an $\mathbb{F}$-adapted Lévy process whose increments are independent of the past σ-fields in $\mathbb{F}$) with finite second moment and mean 0.
- M is an $\mathbb{R}^{\mathbb{N}}$-valued square-integrable martingale on any finite time interval, and B is an $\mathbb{R}^{\mathbb{N}}$-valued predictable process such that each coordinate process is of locally finite variation. We assume that B and the predictable quadratic variation process $\langle M, M \rangle$ have progressively measurable Lebesgue densities $b \colon \Omega \times \mathbb{R}_+ \to \mathbb{R}^{\mathbb{N}}$ and $c \colon \Omega \times \mathbb{R}_+ \to \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$.
- f is the sum of two deterministic measurable functions $f^{\mathrm{C}}, f^{\mathrm{P}} \colon \mathbb{R}_+ \to \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$, and ρ the sum of two predictable processes $\rho^{\mathrm{C}}, \rho^{\mathrm{P}} \colon \Omega \times \mathbb{R}_+ \to \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$.

Of course, the stochastic integrals behind (2.2) must make sense: each single integral must be well defined *and* the infinite sums must converge in an appropriate sense. A sufficient condition for the existence of the infinite-dimensional integral is the existence of the one-dimensional ones plus the summability of their L^2-norms.

Next, we shall explain the rationale behind the IPS model (2.2) and the specific choices for the involved processes. By the definition given in (2.1), the processes $(X_i : i \in \mathbb{N})'$ are coupled in two ways: first, they interact internally with each other through a drift term (determined by a) and a volatility term (determined by σ in conjunction with L); and second, they are exposed to the same external forces (given by B and M), where f and ρ determine the level of influence these noises have on the particles. In particular, by tuning the parameters a, σ, f and ρ appropriately, one obtains a large range of possible dependence structures for the model (2.2).

The question this paper aims to attack is how and to which degree the complexity of the high-dimensional IPS (2.2) can be reduced. Of course, if each entry of the matrices a, σ, f and ρ is either zero or large, there is no hope in simplifying the model. Therefore, our focus lies on particle networks, where only a small number of pairs have strong interaction, while the majority of links in the system are relatively weak. This is implemented in the decomposition of a, σ, f and ρ into a *core matrix* (superscript C) and a *periphery matrix* part (superscript P). It is important to notice that our distinction between core and periphery is not made on the basis of the particles, but on the linkages between them. This allows for greater modelling flexibility since it includes multi-tier networks in our analysis.

In the presence of non-negligible pair interactions, it is natural to apply the mean field limit only to the links encoded by the periphery matrices. Therefore, we propose the following *partial mean field system* (PMFS) as an approximation to the IPS (2.2):

$$d\bar{X}(t) = \left(a^{\mathrm{C}}(t)\bar{X}(t) + a^{\mathrm{P}}(t)\mathbb{E}[\bar{X}(t)]\right) dt + \left(\sigma^{\mathrm{C}}(t)\bar{X}(t-) + \sigma^{\mathrm{P}}(t)\mathbb{E}[\bar{X}(t)]\right).dL(t)$$
$$+ f^{\mathrm{C}}(t)b(t)\,dt + f^{\mathrm{P}}(t)\mathbb{E}[b(t)]\,dt + \rho^{\mathrm{C}}(t)\,dM(t), \quad t \in \mathbb{R}_+, \tag{2.4}$$
$$\bar{X}(0) = X(0).$$

Written for each row $i \in \mathbb{N}$, this is equivalent to:

$$d\bar{X}_i(t) = \sum_{j=1}^{\infty}\left(a_{ij}^{\mathrm{C}}(t)\bar{X}_j(t) + a_{ij}^{\mathrm{P}}(t)\mathbb{E}[\bar{X}_j(t)]\right) dt$$
$$+ \sum_{j=1}^{\infty}\left(\sigma_{ij}^{\mathrm{C}}(t)\bar{X}_j(t-) + \sigma_{ij}^{\mathrm{P}}(t)\mathbb{E}[\bar{X}_j(t)]\right) dL_i(t)$$
$$+ \sum_{j=1}^{\infty}\left(f_{ij}^{\mathrm{C}}(t)b_j(t) + f_{ij}^{\mathrm{P}}(t)\mathbb{E}[b_j(t)]\right) dt \tag{2.5}$$
$$+ \sum_{j=1}^{\infty}\rho_{ij}^{\mathrm{C}}(t)\,dM_j(t), \quad t \in \mathbb{R}_+,$$
$$\bar{X}_i(0) = X_i(0).$$

It is clear that a priori there is no reason for (2.4) to be a good approximation for (2.2). Therefore, in the next section, we will give precise L^2-estimates in terms of the model coefficients for the difference between the IPS and the PMFS. Moreover, we will determine conditions under which this difference becomes small such that we can indeed speak of a law of large numbers.

3. Law of large numbers

The first main result of this paper assesses the distance between the original IPS (2.2) and the PMFS (2.4). To formulate this, we have to introduce some further notation. For $T \in \mathbb{R}_+$, we define

$$v_a(T) := \sup_{i\in\mathbb{N}}\sum_{j=1}^{\infty} A_{ij}(T), \quad v_{a,\mathrm{d}}(T) := \sup_{i\in\mathbb{N}} A_{ii}^{\mathrm{C}}(T), \qquad v_\sigma(T) := \sup_{i\in\mathbb{N}}\sum_{j=1}^{\infty} \Sigma_{ij}(T),$$

$$v_L := \sup_{i\in\mathbb{N}}\|L_i(1)\|_{L^2}, \qquad v_b(T) := \sup_{i\in\mathbb{N}}\sup_{t\in[0,T]}\|b_i(t)\|_{L^2}, \qquad v_X := \sup_{i\in\mathbb{N}}\|X_i(0)\|_{L^2},$$

$$v_f(T) := \sup_{i\in\mathbb{N}}\sup_{t\in[0,T]}\sum_{j=1}^{\infty}(|f_{ij}^{\mathrm{C}}(t)| + |f_{ij}^{\mathrm{P}}(t)|),$$

$$v_{\rho,M}(T) := \sup_{i\in\mathbb{N}}\sup_{t\in[0,T]}\left(\sum_{j,k=1}^{\infty}\left|\mathbb{E}[\rho_{ij}^{\mathrm{C}}(t)\rho_{ik}^{\mathrm{C}}(t)c_{jk}(t)]\right| + \left|\mathbb{E}[\rho_{ij}^{\mathrm{P}}(t)\rho_{ik}^{\mathrm{P}}(t)c_{jk}(t)]\right|\right)^{1/2},$$

$$\tag{3.1}$$

and introduce the rates

$$r_1(T) := \left| A^{\mathrm{P}}(T) | \mathrm{Cov}[X(0)] | (A^{\mathrm{P}}(T))' \right|_{\mathrm{d}}^{1/2}, \qquad r_2(T) := \left| \Sigma^{\mathrm{P}}(T) | \mathrm{Cov}[X(0)] | (\Sigma^{\mathrm{P}}(T))' \right|_{\mathrm{d}}^{1/2},$$

$$r_3(T) := \left| A^{\mathrm{P}}(T) | \mathrm{Cov}[L(1)] | (A^{\mathrm{P}}(T))' \right|_{\mathrm{d}}^{1/2}, \qquad r_4(T) := \left| \Sigma^{\mathrm{P}}(T) | \mathrm{Cov}[L(1)] | (\Sigma^{\mathrm{P}}(T))' \right|_{\mathrm{d}}^{1/2},$$

$$r_5(T) := \sup_{t \in [0,T]} \left| f^{\mathrm{P}}(t) \mathrm{Cov}[b(t)] (f^{\mathrm{P}}(t))' \right|_{\mathrm{d}}^{1/2}, \qquad r_6(T) := \sup_{t \in [0,T]} \left| \mathbb{E}\left[\rho^{\mathrm{P}}(t) c(t) (\rho^{\mathrm{P}}(t))' \right] \right|_{\mathrm{d}}^{1/2},$$

$$r_7(T) := \left| A^{\mathrm{P}}(T) A^{\mathrm{C}}(T)^{\times} \right|_{\infty}, \qquad r_8(T) := \left| \Sigma^{\mathrm{P}}(T) A^{\mathrm{C}}(T)^{\times} \right|_{\infty},$$

$$r_9(T) := \sup_{s,t \in [0,T]} \left| A^{\mathrm{P}}(T) | f^{\mathrm{C}}(s) \mathrm{Cov}[b(s), b(t)] (f^{\mathrm{C}}(t))' | (A^{\mathrm{P}}(T))' \right|_{\mathrm{d}}^{1/2},$$

$$r_{10}(T) := \sup_{s,t \in [0,T]} \left| \Sigma^{\mathrm{P}}(T) | f^{\mathrm{C}}(s) \mathrm{Cov}[b(s), b(t)] (f^{\mathrm{C}}(t))' | (\Sigma^{\mathrm{P}}(T))' \right|_{\mathrm{d}}^{1/2},$$

$$r_{11}(T) := \sup_{t \in [0,T]} \left| A^{\mathrm{P}}(T) | \mathbb{E}[\rho^{\mathrm{C}}(t) c(t) (\rho^{\mathrm{C}}(t))'] | (A^{\mathrm{P}}(T))' \right|_{\mathrm{d}}^{1/2},$$

$$r_{12}(T) := \sup_{t \in [0,T]} \left| \Sigma^{\mathrm{P}}(T) | \mathbb{E}[\rho^{\mathrm{C}}(t) c(t) (\rho^{\mathrm{C}}(t))'] | (\Sigma^{\mathrm{P}}(T))' \right|_{\mathrm{d}}^{1/2}.$$

$$(3.2)$$

Theorem 3.1. *Fix some $T \in \mathbb{R}_+$, and grant the general model assumptions as given in Section 2. Furthermore, assume that each of the numbers in (3.1) is finite. Then (2.2) and (2.4) have a pathwise unique solution X and $\bar{X}$, respectively, and there exist constants $K(T)$ and $K_\iota(T)$, $\iota = 1, \ldots, 12$, which depend on the model coefficients only through the numbers in (3.1), such that*

$$\sup_{i \in \mathbb{N}} \left\| (X_i - \bar{X}_i)^*(T) \right\|_{L^2} \leq K(T) \sum_{\iota=1}^{12} K_\iota(T) r_\iota(T). \qquad (3.3)$$

The proof of Theorem 3.1 will be given in Section 5. Compared to the homogeneous case of [36], we have to take care of several kinds of heterogeneous dependencies in the system: different weights on the edges, the distinction between core and periphery links, and possibly dependent driving noises. Hence, the main idea behind the proof of Theorem 3.1 is to analyse the different channels along which the error induced by the partial mean field approximation can spread through the system, and then to obtain the total L^2-error by Gronwall's lemma. Taking into account all possible ways core and periphery links can interact with each other, yields twelve rates in contrast to a single one in (1.3). Since our proof makes explicit use of covariances, the linearity of the interaction terms is crucial. We will further comment on this in Remarks 3.5 and 3.7.

Remark 3.2. Our calculations produce the following constants in (3.3):

$$K(T) := \sqrt{2} \exp((T^{1/2} v_a(T) + 2 v_\sigma(T) v_L)^2 T), \qquad (3.4)$$

and

$$K_1(T) := E(T) T, \qquad\qquad K_2(T) := 2 v_L E(T) T^{1/2},$$

$$K_3(T) := \frac{2}{3} E(T) v_\sigma(T) V(T) T^{3/2}, \qquad K_4(T) := \sqrt{2} v_L E(T) v_\sigma(T) V(T) T,$$

$$K_5(T) := T, \qquad\qquad\qquad K_6(T) := 2T^{1/2},$$

$$K_7(T) := \frac{1}{2}E(T)V(T)T^2, \qquad\qquad K_8(T) := \frac{2}{\sqrt{3}}v_L E(T)V(T)$$

$$K_9(T) := \frac{1}{2}E(T)T^2, \qquad\qquad K_{10}(T) := \frac{2}{\sqrt{3}}v_L E(T)T^{3/2},$$

$$K_{11}(T) := \frac{2}{3}E(T)T^{3/2}, \qquad\qquad K_{12}(T) := \sqrt{2}v_L E(T)T,$$

where

$$E(T) := e^{v_{a,\mathrm{d}}(T)},$$
$$V(T) := \sqrt{2}e^{(v_a(T)T^{1/2}+2v_L v_\sigma(T))^2 T}\left(v_X + v_f(T)v_b(T)T + 2v_{\rho,M}(T)T^{1/2}\right). \quad \square$$

Remark 3.3. There are several possibilities to extend Theorem 3.1 without substantially new arguments.

(1) It is straightforward to show that Theorem 3.1 can be extended to the case where the interaction matrices a and σ are replaced by (still deterministic but possibly history-dependent) linear functionals.

(2) Suppose that $L = \Gamma L^0$ with some matrix $\Gamma \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ and some other Lévy process L^0 with finite variance and mean zero. Furthermore, $\Gamma = \Gamma^{\mathrm{C}} + \Gamma^{\mathrm{P}}$ and accordingly $L^{\mathrm{C}} = \Gamma^{\mathrm{C}}L^0$ and $L^{\mathrm{P}} = \Gamma^{\mathrm{P}}L^0$. What one would like to do when passing to the PMFS (2.4) is to replace L there by L^{C}. How does this affect the estimate (3.3) in Theorem 3.1? A similar analysis as for Theorem 3.1 reveals that an extra rate

$$r_{13} := \left|\Gamma^{\mathrm{P}}\mathrm{Cov}[L^0(1)](\Gamma^{\mathrm{P}})'\right|_{\mathrm{d}}^{1/2}$$

appears with constant $K_{13} := 2v_\sigma(T)V(T)T^{1/2}$.

(3) Two further generalizations are discussed in Remarks 3.5 and 3.7. $\quad \square$

It is obvious that the usefulness of Theorem 3.1 depends on the sizes of the rates in (3.2): only if they are small, the PMFS (2.4) is a good approximation to the IPS (2.2). Moreover, there are two different views on Theorem 3.1: first, if we assume that the underlying network of the IPS is static, it gives an upper bound on the L^2-error when the IPS is approximated by the PMFS; and second, if the interaction network (i.e. a, σ, f and ρ) is assumed to evolve according to an index $N \in \mathbb{N}$, Theorem 3.1 gives conditions under which the difference between the IPS and the PMFS converges in the L^2-sense to 0, that is,

$$\sup_{i\in\mathbb{N}} \|(X_i^N - \bar{X}_i^N)^*(T)\|_{L^2} \to 0, \quad N \to \infty$$

(this happens precisely when all rates in (3.2) converge to 0 as $N \to \infty$, and the numbers in (3.1) are majorized independently of N). It is also this second point of view that is the traditional one in mean field analysis and that justifies the title "Law of large numbers" for the current section.

In the following subsections we will study three examples of dynamical networks and the corresponding conditions for the law of large numbers to hold for the PMFS.

3.1. Propagation of chaos

We first discuss the phenomenon of chaos propagation, and our results will particularly extend the results of [18], Section 17.3, [23], Corollary 4.1, and [36], Theorem 1.4, by including inhomogeneous weights in the model. The setting is as follows:

(1) The underlying network changes with $N \in \mathbb{N}$. In particular, we will index X and $\bar{X}$, the coefficients a, σ, f and ρ as well as the rates in (3.2) by N.

(2) All structural assumptions in Section 2 hold and the numbers in (3.1), some of which now depend on N, are uniformly bounded in N.

(3) The core matrices $a^{N,C}(t)$, $\sigma^{N,C}(t)$, $f^{N,C}(t)$ and $\rho^{N,C}(t)$ are diagonal matrices for all times $t \in \mathbb{R}_+$.

(4) For each $N \in \mathbb{N}$, $(L_i, b_i, M_i, \rho_{ii}^{N,C}, X_i^N(0) : i \in \mathbb{N})$ is a sequence of independent random elements (note that the noises indexed by a fixed i may depend on each other).

(5) For each $T \in \mathbb{R}_+$, the following rates converge to 0 as $N \to \infty$:

$$r_a^N(T) := \sup_{i \in \mathbb{N}} \left(\sum_{j=1}^{\infty} (A_{ij}^{N,P}(T))^2 \right)^{1/2}, \quad r_{\rho,M}^N(T) := \sup_{i \in \mathbb{N}} \sup_{t \in [0,T]} \left(\sum_{j=1}^{\infty} \mathbb{E}[(\rho_{ij}^{N,P}(t))^2 c_{jj}(t)] \right)^{1/2},$$

$$r_\sigma^N(T) := \sup_{i \in \mathbb{N}} \left(\sum_{j=1}^{\infty} (\Sigma_{ij}^{N,P}(T))^2 \right)^{1/2}, \quad r_f^N(T) := \sup_{i \in \mathbb{N}} \sup_{t \in [0,T]} \left(\sum_{j=1}^{\infty} (f_{ij}^{N,P}(t))^2 \right)^{1/2}.$$

These hypotheses ensure that all pair dependencies between the processes X_i^N, $i \in \mathbb{N}$, vanish when $N \to \infty$. As a result, in the PMFS, the independence of the particles i at $t = 0$ propagates through all times $t > 0$: the PMFS decouples in contrast to the original IPS.

Example 3.4. In classical mean field theory as in the references mentioned in the introduction, the Nth network consists of exactly N particles. In other words, a_{ij}^N, σ_{ij}^N, f_{ij}^N, ρ_{ij}^N and $X_i^N(0)$ are all 0 for $i > N$ or $j > N$. Moreover, all pair interactions are assumed to be of order $1/N$, that is, we have for each $T \in \mathbb{R}_+$,

$$A_{ij}^{N,P}(T) = \frac{A_{ij}(T)}{N}, \quad \Sigma_{ij}^{N,P}(T) = \frac{\Sigma_{ij}(T)}{N}, \quad i, j \in \mathbb{N}, \tag{3.5}$$

where $A_{ij}(T)$, $\Sigma_{ij}(T) \in \mathbb{R}_+$ are uniformly bounded in $i, j \in \mathbb{N}$. Furthermore, the driving noises are supposed to be independent for different particles and to enter the PMFS completely. This means that (3) and (4) hold and that $f^{N,P} = \rho^{N,P} = 0$. It is easily shown that under these specifications the rates in (5) above converge to 0 as $N \to \infty$: $r_{\rho,M}^N(T)$ and $r_f^N(T)$ are simply 0, and $r_a^N(T)$ and $r_\sigma^N(T)$ are of order $1/\sqrt{N}$ as $N \to \infty$. □

We still need to show that under assumptions (1)–(5) above, all rates $r_\iota^N(T)$, $\iota = 1, \ldots, 12$, converge to 0 as $N \to \infty$. Since $A^{N,C}(T)$ is diagonal, we have $A^{N,C}(T)^\times = 0$, and since the driving noises for different particles are independent, all covariances (or covariations) vanish outside the diagonal. Thus, we have

$$\begin{array}{lll}
r_1^N(T) \le v_X r_a^N(T), & r_2^N(T) \le v_X r_\sigma^N(T), & r_3^N(T) \le v_L r_a^N(T), \\
r_4^N(T) \le v_L r_\sigma^N(T), & r_5^N(T) \le v_b(T) r_f^N(T), & r_6^N(T) = r_{\rho,M}^N(T), \\
r_7^N(T) = 0, & r_8^N(T) = 0, & r_9^N(T) \le v_b(T) v_f(T) r_a^N(T), \\
r_{10}^N(T) \le v_b(T) v_f(T) r_\sigma^N(T), & r_{11}^N(T) \le v_{\rho,M}(T) r_a^N(T), & r_{12}^N(T) \le v_{\rho,M}(T) r_\sigma^N(T),
\end{array}$$

which all converge to 0 as $N \to \infty$ by hypothesis. The following remark continues Remark 3.3 regarding further generalizations of Theorem 3.1.

Remark 3.5. In the setting of this subsection there are actually no core relationships between different particles: every pair interaction rate tends to 0 with large N. If we even assume that

there is no dependence at all originating from the noises (i.e. $f^{N,\mathrm{P}} = \rho^{N,\mathrm{P}} = 0$ above), the propagation of chaos result can easily be extended to suitably bounded nonlinear Lipschitz interaction terms instead of the matrices a^N and σ^N (for example, the drift coefficient for X_i^N could be of the form $\frac{1}{N} \sum_{j=1}^N a_{ij}^N(X_i^N, X_j^N)$, where a_{ij}^N are uniformly bounded Lipschitz continuous functions admitting Lipschitz constants that are independent of i, j and N). Since the bound on page 175, line 13, of [36] only uses independence (but not exchangeability), and Gronwall's lemma also applies in the heterogeneous case (as our proof of Theorem 3.1 shows), the proof of [36], Theorem 1.4, can be applied with obvious changes.   □

3.2. Sparse interaction versus sparse correlation

The propagation of chaos result in the last subsection was based on two core hypotheses: asymptotically vanishing pair interaction rates and the independence of the particles' driving noises. The motivation for establishing Theorem 3.1, however, is to deal with situations where these two conditions are precisely not satisfied, that is, when the coefficients a, σ, f and ρ of (2.2) are decomposed into a core and a periphery part in a non-trivial way. In fact, in this subsection, we discuss a typical situation where the full generality of Theorem 3.1 is required. Before that, we recall that we consider networks indexed by $N \in \mathbb{N}$, and that we are interested in the cases when the rates in (3.2) vanish when N becomes large.

General assumptions
The following list of hypotheses describes the setting in this subsection.

(1) The statements (1) and (2) of Section 3.1 hold.
(2) M is an $\mathbb{R}^{\mathbb{N}}$-valued $\mathbb{F}$-Lévy process, implying that $c_{ij}(t) = \mathrm{Cov}[M_i(1), M_j(1)]t$.
(3) At stage N, the system consists of N particles, that is, we have $a_{ij}^N = \sigma_{ij}^N = f_{ij}^N = \rho_{ij}^N = X_i^N(0) = 0$ as soon as $i > N$ or $j > N$.
(4) Let $N_0 \in \mathbb{N}$ be independent of N and suppose that $N > N_0$. We assume that $\mathcal{C} := \{1, \ldots, N_0\}$ contains the *core particles*, while $\mathcal{P}^N := \{N_0 + 1, \ldots, N\}$ contains the *periphery particles*, whose number increases with N. Correspondingly, $a^{N,\mathrm{C}}$ and $\sigma^{N,\mathrm{C}}$ (resp. $a^{N,\mathrm{P}}$ and $\sigma^{N,\mathrm{P}}$) characterize the influence of the core (resp. periphery) particles in the system. In other words, we assume that $j \in \mathcal{C}$ implies that $a_{ij}^{N,\mathrm{P}}(t) = \sigma_{ij}^{N,\mathrm{P}}(t) = 0$ for all $i \in \mathbb{N}$ and $t \in \mathbb{R}_+$, while $j \in \mathcal{P}^N$ implies $a_{ij}^{N,\mathrm{C}}(t) = \sigma_{ij}^{N,\mathrm{C}}(t) = 0$ for all $i \neq j$ and $t \in \mathbb{R}_+$. We assume that the diagonals of a^N and σ^N are completely contained in $a^{N,\mathrm{C}}$ and $\sigma^{N,\mathrm{C}}$, respectively. It follows that the partitions of a^N and σ^N can be illustrated as follows (omitting all zero rows and columns, and using $*$ for all potentially non-zero elements):

$$
a^{N,\mathrm{C}}/\sigma^{N,\mathrm{C}} = \begin{array}{c} \overset{N_0}{} \qquad\qquad \overset{N}{} \\ \left(\begin{array}{ccc|cccc} * & \cdots & * & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ * & \cdots & * & 0 & \cdots & \cdots & 0 \\ \hline * & \cdots & * & * & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & 0 & \cdots & 0 & * \end{array} \right), \end{array}
$$

$$
a^{N,\mathrm{P}}/\atop \sigma^{N,\mathrm{P}} = \left(\begin{array}{ccc|cccc}
0 & \cdots & 0 & * & \cdots & \cdots & * \\
\vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & * & \cdots & \cdots & * \\
\hline
0 & \cdots & 0 & 0 & * & \cdots & * \\
\vdots & \ddots & \vdots & * & \ddots & \ddots & \vdots \\
\vdots & \ddots & \vdots & \vdots & \ddots & \ddots & * \\
0 & \cdots & 0 & * & \cdots & * & 0
\end{array}\right).
$$

(5) There is a finite number of *systematic noises*, namely $B_1, \ldots, B_{N_{00}}$ and $M_1, \ldots, M_{N_{00}}$, for some fixed $N_{00} \in \mathbb{N}$ independent of N. These noises are important to a large part of the system, and there are *idiosyncratic noises* $B_{N_{00}+i}$ and $M_{N_{00}+i}$ that only affect the specific particle $i \in \{1, \ldots, N\}$. Thus, we assume for all $i = 1, \ldots, N$ and $t \in \mathbb{R}_+$ that $\rho_{ij}^{N,\mathrm{P}}(t) = f_{ij}^{N,\mathrm{P}}(t) = 0$ for $j \in \{1, \ldots, N_{00}\} \cup \{N_{00} + i\}$ and $\rho_{ij}^{N,\mathrm{C}}(t) = f_{ij}^{N,\mathrm{C}}(t) = 0$ for the other values of j. Hence, f^N and ρ^N are of the form

$$
f^{N,\mathrm{C}}/\atop \rho^{N,\mathrm{C}} = \left(\begin{array}{ccc|cccc}
* & \cdots & * & * & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & 0 & \ddots & \ddots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \ddots & 0 \\
* & \cdots & * & 0 & \cdots & 0 & *
\end{array}\right),
$$

$$
f^{N,\mathrm{P}}/\atop \rho^{N,\mathrm{P}} = \left(\begin{array}{ccc|cccc}
0 & \cdots & 0 & 0 & * & \cdots & * \\
\vdots & \ddots & \vdots & * & \ddots & \ddots & \vdots \\
\vdots & \ddots & \vdots & \vdots & \ddots & \ddots & * \\
0 & \cdots & 0 & * & \cdots & * & 0
\end{array}\right).
$$

(6) We have for all $T \in \mathbb{R}_+$

$$
A_{ij}^{N,\mathrm{P}}(T) = \frac{\phi_{ij}^N(T)}{R_A^N}, \quad \Sigma_{ij}^{N,\mathrm{P}}(T) = \frac{\psi_{ij}^N(T)}{R_\Sigma^N}, \quad i, j = 1, \ldots, N, \tag{3.6}
$$

where the rates R_A^N and R_Σ^N satisfy

$$
\frac{R_A^N}{\sqrt{N}} \to \infty, \quad \frac{R_\Sigma^N}{\sqrt{N}} \to \infty, \quad N \to \infty, \tag{3.7}
$$

and the numbers $\phi_{ij}^N(T), \psi_{ij}^N(T) \in \mathbb{R}_+$ satisfy

$$
\phi(T) := \sup_{i,j,N \in \mathbb{N}} \phi_{ij}^N(T) < \infty, \quad \psi(T) := \sup_{i,j,N \in \mathbb{N}} \psi_{ij}^N(T) < \infty.
$$

Note that we always have $\phi_{ii}^N(T) = \psi_{ii}^N(T) = 0$.

(7) For different $i, j \in \mathbb{N}$, the noises M_i and M_j as well as B_i and B_j are uncorrelated.

(8) The rates $r_f^N(T)$ and $r_{\rho,M}^N(T)$ from Section 3.1 converge to 0 as $N \to \infty$ for all $T \in \mathbb{R}_+$.

(9) For each $N \in \mathbb{N}$, the initial values $(X_i^N(0) : i \in \mathcal{P}^N)$ are mutually uncorrelated.

Conditions (4) and (5) determine the core–periphery structure of the IPS. In practice, a fixed distinction between core and periphery particles is often not possible because a large number of

particles may be engaged in some strong and some weak linkages at the same time. As already pointed out, this does not affect the applicability of Theorem 3.1, since the concept of core and periphery refers to the linkages there. The choice of fixed core and periphery *particles* in this subsection is only a special case thereof, intended to simplify the arguments below. Next, regarding (6), one can take $R_A^N = R_\Sigma^N = N$ for concreteness, which can then be compared with Section 3.1. Furthermore, let us point out that assumption (7) is only for convenience (namely that f^N and ρ^N carry the whole correlation structure of the noises). Indeed, it is always possible (under our second-moment conditions) to replace any stochastic integral $\rho \cdot M$, where M is a Lévy process with an arbitrary correlation structure, by $\rho' \cdot M'$, where M' consists of mutually uncorrelated Lévy processes (of course, (8) would change accordingly). Finally, if $X^N(0)$ is independent of the driving noises, (9) can be enforced simply by switching to the conditional distribution given $X^N(0)$.

Under (1)–(9) it is easy to prove that the rates $r_1^N(T)$, $r_2^N(T)$, $r_5^N(T)$ and $r_6^N(T)$ converge to 0 when $N \to \infty$. For the latter two, this can be deduced in the same way as in Section 3.1 because the driving noises of different particles are uncorrelated. For the first two rates, we use that the starting random variables of periphery particles are assumed to be uncorrelated. Hence, we have by (3.7), as $N \to \infty$, that

$$r_1^N(T) = \sup_{i \in \mathbb{N}} \left(\sum_{j \in \mathcal{P}^N} (A_{ij}^{N,\mathrm{P}}(T))^2 \mathrm{Var}[X_j^N(0)] \right)^{1/2} \leq \phi(T) v_X \frac{\sqrt{N}}{R_A^N} \to 0,$$

$$r_2^N(T) = \sup_{i \in \mathbb{N}} \left(\sum_{j \in \mathcal{P}^N} (\Sigma_{ij}^{N,\mathrm{P}}(T))^2 \mathrm{Var}[X_j^N(0)] \right)^{1/2} \leq \psi(T) v_X \frac{\sqrt{N}}{R_\Sigma^N} \to 0.$$

However, the conditions above are in general *not* sufficient to imply the smallness of the other rates in (3.2). We need to add extra hypotheses.

Sparseness assumptions

For each of the remaining rates, we further examine what type of conditions are needed to make them asymptotically small. As we shall see, it is always a mixture of a sparseness condition on the interaction matrices A^N and Σ^N and a sparseness condition on the correlation matrices f^N and ρ^N.

$r_3^N(T)$ *and* $r_4^N(T)$:   We first present a counterexample to show that we have to require further conditions. Consider the simple case where $L_i = L_1$ for all $i \in \mathbb{N}$ and that $A_{ij}^{N,\mathrm{P}}(T) = 1/R_A^N$ for all $T \in \mathbb{R}_+$ and $i, j \in \{1, \ldots, N\}$ with $i \neq j$. Then

$$r_3^N(T) = \sup_{i \in \mathbb{N}} \left(\sum_{j,k \in \mathcal{P}^N \setminus \{i\}} \left(\frac{1}{R_A^N} \right)^2 \mathrm{Cov}[L_j(1), L_k(1)] \right)^{1/2} \leq v_L \frac{N}{R_A^N},$$

which need not to converge to 0 in general. A similar calculation can be done for $r_4^N(T)$. In order to make the rates $r_3^N(T)$ and $r_4^N(T)$ small, there are basically two options: we require the interaction matrices $A^{N,\mathrm{P}}$ and $\Sigma^{N,\mathrm{P}}$ to be sparse, or we require the correlation matrix of L to be sparse. Any other possibility is a suitable combination of these two.

(10a) The noises $(L_i : i \in \mathcal{P}^N)$ corresponding to periphery particles only have sparse correlation (which, in particular, includes the case of mutual independence as in Section 3.1). More

precisely, we require

$$p_L^N := \#\{(i, j) \in \mathcal{P}^N \times \mathcal{P}^N : \mathrm{Cov}[L_i(1), L_j(1)] \neq 0\}| = o\big((R_A^N)^2 \wedge (R_\Sigma^N)^2\big) \quad (3.8)$$

for large N. Then

$$r_3^N(T) = \sup_{i \in \mathbb{N}} \left(\sum_{j,k \in \mathcal{P}^N} A_{ij}^{N,\mathrm{P}}(T) A_{ik}^{N,\mathrm{P}}(T) \mathrm{Cov}[L_j(1), L_k(1)] \right)^{1/2} \leq \phi(T) v_L \frac{\sqrt{p_L^N}}{R_A^N} \to 0,$$

and, similarly, $r_4^N(T) \to 0$ as $N \to \infty$.

(10b) The matrices $A^{N,\mathrm{P}}(T)$ and $\Sigma^{N,\mathrm{P}}(T)$, which describe the influence of periphery particles on the system, are only sparsely occupied, in the sense that every particle in the system is only affected by a small number of periphery particles. In mathematical terms, this condition reads as

$$\begin{aligned} p_{A,1}^N(T) &:= \sup_{i \in \mathbb{N}} \#\{j \in \mathcal{P}^N : A_{ij}^{N,\mathrm{P}}(T) \neq 0\} = o(R_A^N), \\ p_\Sigma^N(T) &:= \sup_{i \in \mathbb{N}} \#\{j \in \mathcal{P}^N : \Sigma_{ij}^{N,\mathrm{P}}(T) \neq 0\} = o(R_\Sigma^N). \end{aligned} \quad (3.9)$$

In this case, we get

$$r_3^N(T) = \sup_{i \in \mathbb{N}} \left(\sum_{j,k \in \mathcal{P}^N} A_{ij}^{N,\mathrm{P}}(T) A_{ik}^{N,\mathrm{P}}(T) \mathrm{Cov}[L_j(1), L_k(1)] \right)^{1/2} \leq \phi(T) v_L \frac{p_{A,1}^N}{R_A^N} \to 0,$$

and similarly $r_4^N(T) \to 0$ as $N \to \infty$.

$r_7^N(T)$ *and* $r_8^N(T)$:   These two rates express the connectivity between core and periphery particles. In general, they will not become small with large N. For instance, if $A_{ij}^{N,\mathrm{C}}(T) = 1$ for all $j \in \mathcal{C}$ and $i \neq j$, and $A_{ij}^{N,\mathrm{P}}(T) = 1/R_A^N$ for all $j \in \mathcal{P}^N$ and $i \neq j$, then

$$r_7^N(T) = \sup_{i \in \mathbb{N}} \sum_{j \in \mathcal{P}^N} \sum_{k \in \mathcal{C}} A_{ij}^{N,\mathrm{P}}(T) A_{jk}^{N,\mathrm{C}}(T) = N_0 \frac{N}{R_A^N},$$

which does not necessarily converge to 0. An analogous statement holds for $r_8^N(T)$. For $r_7^N(T), r_8^N(T) \to 0$ we have to require that the lower left block of $A^{N,\mathrm{C}}$, which describes the influence of core particles on periphery particles, or the matrices $A^{N,\mathrm{P}}(T)$ and $\Sigma^{N,\mathrm{P}}(T)$, which describe the influence of periphery particles on the system, be sparse (or a combination thereof):

(11a) The influence of core on periphery particles is sparse. In other words, we suppose for the maximal number of periphery particles a single core particle interacts with through the drift:

$$p_{A,2}^N := \sup_{j \in \mathcal{C}} \#\{i \in \mathcal{P}^N : A_{ij}^{N,\mathrm{C}}(T) \neq 0\} = o(R_A^N \wedge R_\Sigma^N). \quad (3.10)$$

Then,

$$r_7^N(T) = \sup_{i \in \mathbb{N}} \sum_{j \in \mathcal{P}^N} \sum_{k \in \mathcal{C}} A_{ij}^{N,\mathrm{P}}(T) A_{jk}^{N,\mathrm{C}}(T) \leq N_0 \phi(T) v_a(T) \frac{p_{A,2}^N}{R_A^N} \to 0$$

as well as $r_8^N(T) \to 0$ as $N \to \infty$.

(11b) $A^{N,\mathrm{P}}(T)$ and $\Sigma^{N,\mathrm{P}}(T)$ are sparse in the sense of (3.9). Then $r_7^N(T), r_8^N(T) \to 0$ follow similarly.

$r_9^N(T), r_{10}^N(T), r_{11}^N(T)$ *and* $r_{12}^N(T)$:   Similar considerations as before show that these four rates do not converge to 0 in general. Instead, we again need to require some mixture of sparsely correlated driving noises and sparsely occupied matrices $A^{N,\mathrm{P}}$ and $\Sigma^{N,\mathrm{P}}$:

(12a) We assume that for all $T \in \mathbb{R}_+$

$$p_f^N(T) := \sup_{j \in \{1,\dots,N_{00}\}} \#\{i \in \mathcal{P}^N : f_{ij}^{N,\mathrm{C}} \not\equiv 0 \text{ on } [0,T]\} = o(R_A^N \wedge R_\Sigma^N), \qquad (3.11)$$

$$p_\rho^N(T) := \sup_{j \in \{1,\dots,N_{00}\}} \#\{i \in \mathcal{P}^N : \rho_{ij}^{N,\mathrm{C}} \not\equiv 0 \text{ on } [0,T]\} = o(R_A^N \wedge R_\Sigma^N). \qquad (3.12)$$

Then, recalling that the components of b and M are mutually uncorrelated,

$$
\begin{aligned}
r_9^N(T) = \sup_{i \in \mathbb{N}} \sup_{s,t \in [0,T]} \Bigg(& \sum_{j,k \in \mathcal{P}^N} \sum_{l=1}^{N_{00}} A_{ij}^{N,\mathrm{P}}(T) A_{ik}^{N,\mathrm{P}}(T) \\
& \times \left| f_{jl}^{N,\mathrm{C}}(s) f_{kl}^{N,\mathrm{C}}(t) \mathrm{Cov}[b_l(s), b_l(t)] \right| \\
& + \sum_{j \in \mathcal{P}^N} (A_{ij}^{N,\mathrm{P}}(T))^2 \left| f_{j(N_{00}+j)}^{N,\mathrm{C}}(s) f_{j(N_{00}+j)}^{N,\mathrm{C}}(t) \mathrm{Cov}[b_{N_{00}+j}(s), b_{N_{00}+j}(t)] \right| \Bigg)^{1/2} \\
\leq & \; \phi(T) v_b(T) v_f(T) \frac{\sqrt{N_{00}} p_f^N(T) + \sqrt{N}}{R_A^N} \to 0,
\end{aligned}
$$

$$
\begin{aligned}
r_{11}^N(T) = \sup_{i \in \mathbb{N}} \sup_{t \in [0,T]} \Bigg(& \sum_{j,k \in \mathcal{P}^N} \sum_{l=1}^{N_{00}} A_{ij}^{N,\mathrm{P}}(T) A_{ik}^{N,\mathrm{P}}(T) \left| \mathbb{E}[\rho_{jl}^{N,\mathrm{C}}(t) \rho_{kl}^{N,\mathrm{C}}(t)] \right| \mathrm{Var}[M_l(1)] \\
& + \sum_{j \in \mathcal{P}^N} (A_{ij}^{N,\mathrm{P}}(T))^2 \left| \mathbb{E}[(\rho_{j(N_{00}+j)}^{N,\mathrm{C}}(t))^2] \right| \mathrm{Var}[M_{N_{00}+j}(1)] \Bigg)^{1/2} \\
\leq & \; \phi(T) v_{\rho,M}(T) \frac{\sqrt{N_{00}} p_\rho^N(T) + \sqrt{N}}{R_A^N} \to 0,
\end{aligned}
$$

and similarly $r_{10}^N(T), r_{12}^N(T) \to 0$ as $N \to \infty$.

(12b) $A^{N,\mathrm{P}}(T)$ and $\Sigma^{N,\mathrm{P}}(T)$ are sparse in the sense of (3.9). Then one can deduce $r_\iota^N(T) \to 0$ for $\iota = 9, 10, 11, 12$ as before.

We conclude this subsection with two remarks.

Remark 3.6. In the sparseness conditions (3.8)–(3.12), it is not essential that the majority of entries is exactly zero. As one can see from the definition of the rates (3.2), they depend continuously on the underlying matrix entries. It suffices therefore that the matrix entries are small enough in a large proportion.   □

Remark 3.7. What can be said about Theorem 3.1 in the general case of nonlinear Lipschitz coefficients a^N and σ^N, apart from the special case discussed in Remark 3.5? In fact, a law of large numbers in the fashion of Theorem 3.1 can still be shown, but under more stringent conditions: namely we have to require condition (10b) above in addition, with $A^{N,\mathrm{P}}$ and $\Sigma^{N,\mathrm{P}}$

now containing the Lipschitz constants of the interaction terms. The reason is that (10b) suffices to make r_ι^N ($\iota \in \{3, 4, 7, \ldots, 12\}$) small. The remaining four rates are unrelated to a^N and σ^N and therefore not affected by their nonlinear structure. It is important to notice that conditions like (10a) and (12a) are no longer sufficient to make the corresponding rates small. The reason is that they are conditions of correlation type. Since correlation is a *linear* measure of dependence, it is not surprising that these conditions are not suitable for the nonlinear case. We do not go into the details at this point. □

3.3. Networks arising from preferential attachment

As demonstrated in the last subsection, the crucial criterion for the rates (3.2) in Theorem 3.1 to vanish asymptotically with growing network size is a combination of sparse interaction and sparse correlation among the particles. Condition (3.9) plays a distinguished role here: when valid, it implies that eight out of twelve rates in (3.2) are small. Moreover, it is the key factor for a nonlinear generalization of Theorem 3.1 to hold or not; see Remark 3.7. The aim of this subsection is therefore to find algorithms for the generation of the underlying networks such that the resulting interaction matrices satisfy (3.9). We will assume that $a^{N,P}(t) = a^{N,P}$ and $\sigma^{N,P}(t) = \sigma^{N,P}$ are independent of $t \in \mathbb{R}_+$, such that also $A^{N,P}(t)$ and $\Sigma^{N,P}(t)$ as well as $p_{A,1}^N(t)$ and $p_\Sigma^N(t)$ (see (3.9) for their definitions) are independent of t. Furthermore, we only concentrate on $p_{A,1}^N$ as the analysis for p_Σ^N is completely analogous.

We will base the creation of the IPS network on dynamical random graph mechanisms. Since we are mainly interested in heterogeneous graphs, we will investigate the *preferential attachment* or *scale-free* random graph [5]. There are many similar but different constructions of preferential attachment graphs; in the following, we rely on the construction of [8] for directed graphs. We remark that the random graphs to be constructed will be indexed by N, corresponding to a family of growing networks for the IPS. In particular, "time" in the random graph process must not be confused with the time t in the IPS (2.2); the correct view is rather that the IPS network has been built from the random graphs before time $t = 0$, and, of course, independently of all random variables in (2.2).

The preferential attachment algorithm works as follows: we start with $G(0) = (V, E(0))$, a given graph consisting of vertices $V = \mathbb{N}$ and edges $E(0) = \{e_1, \ldots, e_\nu\}$, where $\nu \in \mathbb{N}$ and e_i stands for a directed edge between two vertices. We allow for multiple edges and loops in our graphs. Without loss of generality, we assume that the set of vertices in $G(0)$ with at least one neighbour is given by $\{1, \ldots, n(0)\}$ with some $n(0) \in \mathbb{N}$. Furthermore, we fix $\alpha, \beta, \gamma \in \mathbb{R}_+$ with $\alpha + \beta + \gamma = 1$ and $\alpha + \gamma > 0$ and two numbers $\delta^{in}, \delta^{out} \in \mathbb{R}_+$. For $N \in \mathbb{N}$, we construct $G(N) = (V, E(N))$ from $G(N - 1)$ according to the following algorithm.

- With probability α, we create a new edge $e_{\nu+N}$ from $v = n(N - 1) + 1$ to a node w that is already connected in $G(N - 1)$. Here w is chosen randomly from $\{1, \ldots, n(N - 1)\}$ according to the probability mass function

$$\frac{d_{G(N-1)}^{in}(w) + \delta^{in}}{\nu + N - 1 + \delta^{in} n(N - 1)}, \quad w \in \{1, \ldots, n(N - 1)\},$$

  where $d_G^{in}(v)$ denotes the in-degree of vertex v in a graph G. We define $n(N) := n(N-1)+1$ and $E(N) := E(N - 1) \cup \{e_{\nu+N}\}$.
- With probability β, a new edge $e_{\nu+N}$ is formed from some vertex $v \in \{1, \ldots, n(N - 1)\}$ to some $w \in \{1, \ldots, n(N - 1)\}$ (the case $v = w$ is possible). Here v and w are chosen

independently according to the probability mass functions

$$\frac{d^{\text{out}}_{G(N-1)}(v) + \delta^{\text{out}}}{v + N - 1 + \delta^{\text{out}} n(N-1)}, \qquad \frac{d^{\text{in}}_{G(N-1)}(w) + \delta^{\text{in}}}{v + N - 1 + \delta^{\text{in}} n(N-1)}, \qquad v, w \in \{1, \ldots, n(N-1)\},$$

respectively, where $d^{\text{out}}_G(v)$ denotes the out-degree of vertex v in a graph G. Moreover, we set $n(N) := n(N-1)$ and $E(N) := E(N-1) \cup \{e_{v+N}\}$.

- With probability γ, a new edge e_{v+N} from some $v \in \{1, \ldots, n(N-1)\}$ to $w = n(N-1)+1$ is formed. Here v is chosen randomly according to the probability mass function

$$\frac{d^{\text{out}}_{G(N-1)}(v) + \delta^{\text{out}}}{v + N - 1 + \delta^{\text{out}} n(N-1)}, \qquad v \in \{1, \ldots, n(N-1)\}.$$

We set $n(N) := n(N-1) + 1$ and $E(N) := E(N-1) \cup \{e_{v+N}\}$.

Evidently, we always have $|E(N)| = v + N$ while the number $n(N)$ of non-isolated vertices in $G(N)$ is random in general.

The most important result for our purposes is the following one. We define

$$M^{\text{in}}(N) := \max\{d^{\text{in}}_{G(N)}(i) \colon i \in \mathbb{N}\}, \qquad M^{\text{out}}(N) := \max\{d^{\text{out}}_{G(N)}(i) \colon i \in \mathbb{N}\}, \qquad N \in \mathbb{N}_0,$$

as the maximal in-degree and out-degree in $G(N)$, respectively.

Lemma 3.8.   *The maximum in-degree $M^{\text{in}}(N)$ and out-degree $M^{\text{out}}(N)$ of $G(N)$ satisfy the following asymptotics:*

$$c^{\text{in}}(N) M^{\text{in}}(N) \to \mu^{\text{in}}, \qquad c^{\text{out}}(N) M^{\text{out}}(N) \to \mu^{\text{out}}, \qquad N \to \infty. \tag{3.13}$$

Here the convergence to the random variables μ^{in} and μ^{out}, respectively, holds in the almost sure as well as in the L^p-sense for all $p \in [1, \infty)$, and $(c^{\text{in}}(N))_{N \in \mathbb{N}}$ and $(c^{\text{out}}(N))_{N \in \mathbb{N}}$ are sequences of random variables which can be chosen such that for every $\epsilon \in (0, \alpha + \gamma)$ we have a.s.

$$c^{\text{in}}(N)^{-1} = \mathcal{O}\left(N^{\frac{\alpha+\beta}{1+\delta^{\text{in}}(\alpha+\gamma-\epsilon)}}\right), \qquad c^{\text{out}}(N)^{-1} = \mathcal{O}\left(N^{\frac{\beta+\gamma}{1+\delta^{\text{out}}(\alpha+\gamma-\epsilon)}}\right), \qquad N \to \infty. \tag{3.14}$$

It follows from this lemma that for every $\epsilon \in (0, \alpha + \gamma)$, we have a.s.

$$M^{\text{in}}(N) = \mathcal{O}\left(N^{\frac{\alpha+\beta}{1+\delta^{\text{in}}(\alpha+\gamma-\epsilon)}}\right), \qquad M^{\text{out}}(N) = \mathcal{O}\left(N^{\frac{\beta+\gamma}{1+\delta^{\text{out}}(\alpha+\gamma-\epsilon)}}\right), \qquad N \to \infty.$$

In particular, if $G(N)$ is used to model the underlying network of a^N (i.e. an edge in $G(N)$ from i to j is equivalent to $a^N_{ij} \neq 0$), we have

$$p^N_{A,1} \le M^{\text{out}}(N) = \mathcal{O}\left(N^{\frac{\beta+\gamma}{1+\delta^{\text{out}}(\alpha+\gamma-\epsilon)}}\right), \qquad N \to \infty.$$

In other words, the first part of condition (3.9) holds as soon as R^N_A, as specified through (3.6) and (3.7), increases in N at least with rate

$$N^{\frac{\beta+\gamma}{1+\delta^{\text{out}}(\alpha+\gamma-\epsilon)}} \tag{3.15}$$

for some small ϵ. For example, in the classical case of Example 3.4 where $R^N_A = N$, this is always true except in the case $\alpha = \delta^{\text{out}} = 0$, where all edges start from one of the initial nodes with probability one. We conclude that in all non-trivial situations of the preferential attachment model, the resulting networks are sparse enough for the law of large numbers implied by Theorem 3.1 to be in force.

4. Large deviations

In Theorem 3.1 we have established bounds on the mean squared difference between the IPS (2.2) and the PMFS (2.4). In Sections 3.1–3.3 we have given examples of dynamical networks in which these bounds converge to 0 as the network size increases. A natural question is now whether a large deviation principle holds as $N \to \infty$, which would then assure that the probability of X^N deviating strongly from $\bar{X}^N$ decreases exponentially fast in N. As in the sections before, we assume here that the network for the IPS and the PMFS is indexed by $N \in \mathbb{N}$. In the classical case of homogeneous networks, [13] is the first paper to prove a large deviation principle for the empirical measures of the processes (1.1). For heterogeneous networks, however, the empirical measure might no longer be a good quantity to investigate: the weight of a particle now depends on which particle's perspective is chosen. A sequence of differently weighted empirical measures seems to be more appropriate, but then their analysis becomes considerably more involved. Therefore, in this paper we take a more direct approach and study the large deviation behaviour of the difference $X^N - \bar{X}^N$ itself. In order to do so, we have to put stronger assumptions on the coefficients than in the previous sections. These are as follows.

(A1) $X^N(0)$, for each $N \in \mathbb{N}$, and B are deterministic.

(A2) For all $N \in \mathbb{N}$ we have $\sigma^N = 0$. All other coefficients $a^{N,C}$, $a^{N,P}$, $\rho^{N,C}$ and $\rho^{N,P}$ are constant in time. Moreover, only the first $\gamma(N)$ (resp. $\Gamma(N)$) columns of $\rho^{N,C}$ and $\rho^{N,P}$ (resp. $a^{N,C}$ and $a^{N,P}$) may contain non-zero entries, where $\gamma(N)$ is sequence increasing to infinity, and there are constants $C_1, C_2 \in \mathbb{R}_+$ such that $\Gamma(N) \leq C_1 \exp(C_2 \gamma(N))$ for all $N \in \mathbb{N}$.

(A3) All numbers in (3.1), which are indexed by N now, are bounded independently of N.

(A4) $(M_i : i \in \mathbb{N})$ is a sequence of independent mean-zero Lévy processes whose Brownian motion part has variance c_i and whose Lévy measure is ν_i. Moreover, there exists a real-valued mean-zero symmetric Lévy process M_0 with $\mathbb{E}[e^{uM_0(1)}] < \infty$ for all $u \in \mathbb{R}_+$ that dominates M_i, that is, its characteristics c_0 and ν_0 satisfy $c_i \leq c_0$ and $\nu_i(A) \leq \nu_0(A)$ for all $i \in \mathbb{N}$ and Borel sets $A \subseteq \mathbb{R}$.

(A5) Assume that $G^N(t, s) := \gamma(N) e^{a^N t} a^{N,P} e^{a^{N,C} s} \rho^{N,C}$, $s, t \in [0, T]$, converges uniformly to a limit $G(t, s) \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$:

$$\sup_{i,j=1,\dots,\gamma(N) \vee \Gamma(N)} \sup_{s,t \in [0,T]} |G^N_{ij}(t,s) - G_{ij}(t,s)| \to 0, \quad \sup_{i,j \in \mathbb{N}} \sup_{s,t \in [0,T]} |G_{ij}(t,s)| < \infty,$$
$$N \to \infty.$$

(A6) With $R^N(t) := \gamma(N) e^{a^N t} \rho^{N,P}$, $t \in [0, T]$, there exists $R(t) \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ such that

$$\sup_{i,j \in \mathbb{N}} \sup_{t \in [0,T]} |R_{ij}(t)| < \infty, \quad \sup_{i,j=1,\dots,\gamma(N) \vee \Gamma(N)} \sup_{t \in [0,T]} |R^N_{ij}(t) - R_{ij}(t)| \to 0, \quad N \to \infty.$$

(A7) The following two quantities are finite:

$$q_1 := \limsup_{N \to \infty} q_1(N) := \limsup_{N \to \infty} \sup_{i \in \mathbb{N}} \gamma(N) \sum_{j=1}^{\infty} \sum_{k \neq j} |a^{N,P}_{ij} a^{N,C}_{jk}|,$$

$$q_2 := \limsup_{N \to \infty} q_2(N) := \limsup_{N \to \infty} \sup_{i,k \in \mathbb{N}} \gamma(N) \sum_{j=1}^{\infty} |a^{N,P}_{ij} \rho^{N,C}_{jk}|.$$

(A8) Define for $m \in \mathbb{N} \cup \{0\}$

$$\Psi_m(u) := \log \mathbb{E}[e^{uM_m(1)}] = \frac{1}{2}c_m u^2 + \int_{\mathbb{R}} (e^{uz} - 1 - uz) \, v_m(dz), \quad u \in \mathbb{R}_+.$$

We assume that the following holds for every $d \in \mathbb{N}$: denoting for $m \in \mathbb{N}$, $r \in [0, T]$ and $\theta \in M_T^d$

$$H_m(\theta, r) := \int_r^T \int_s^T \sum_{i=1}^d G_{im}(t - s, s - r) \theta_i(dt) \, ds + \int_r^T \sum_{i=1}^d R_{im}(t - r) \theta_i(dt),$$

the sequence $\left(\int_0^T \Psi_m(H_m(\theta, r)) \, dr \right)_{m \in \mathbb{N}}$ is Cesàro summable, i.e. the following limit exists:

$$\lim_{N \to \infty} \frac{1}{\gamma(N)} \sum_{m=1}^{\gamma(N)} \int_0^T \Psi_m(H_m(\theta, r)) \, dr. \tag{4.1}$$

Theorem 4.1. *Let $T \in \mathbb{R}_+$. Under (A1)–(A8), the sequence $(X^N - \bar{X}^N)_{N \in \mathbb{N}}$ satisfies a large deviation principle in (D_T^∞, J_1) with a good rate function $I \colon D_T^\infty \to [0, \infty]$, that is, for every $\alpha \in \mathbb{R}_+$ the set $\{x \in D_T^\infty \colon I(x) \leq \alpha\}$ is compact in D_T^∞ (with respect to the J_1-topology), and for every $M \in \mathcal{D}_T^\infty$ we have*

$$-\inf_{x \in \text{int } M} I(x) \leq \liminf_{N \to \infty} \frac{1}{\gamma(N)} \log \mathbb{P}[X^N - \bar{X}^N \in M]$$

$$\leq \limsup_{N \to \infty} \frac{1}{\gamma(N)} \log \mathbb{P}[X^N - \bar{X}^N \in M] \leq -\inf_{x \in \text{cl } M} I(x),$$

where $\text{int } M$ and $\text{cl } M$ denote the interior and the closure of M in (D_T^∞, J_1), respectively. Moreover, the rate function I is convex, attains its minimum 0 uniquely at the origin and is infinite for $x \notin AC_T^\infty$.

Remark 4.2.

(1) We cannot drop the requirement $\sigma^N = 0$ or condition (A4) in Theorem 4.1 because otherwise, the processes X^N and $\bar{X}^N$ will typically not have exponential moments of all orders, whose existence is essential for our proof below. This kind of problem does not arise when empirical measures are considered as in [13,25] for the homogeneous case.

(2) The Cesàro summability condition (A8) accounts for the possible inhomogeneity of the coefficients and the distribution of the noises. It holds in particular for the homogeneous case. Since a convergent series is Cesàro summable with the same limit, it also holds when we have asymptotic homogeneity (in the sense that the sequence inside the sum of (4.1) converges as $m \to \infty$).

Example 4.3. We apply Theorem 4.1 to the model considered in Theorem 3.1 of [23], which in particular covers McKean's example [36]. At stage N, we have particles $1, \ldots, N$, divided into two groups Θ_1^N and Θ_2^N of $[\eta N]$ and $N - [\eta N]$ elements, respectively, where $\eta \in (0, 1)$. We assume that the attribution of particles to the two groups is stable in N, in the sense that $\Theta_1^N \uparrow \Theta_1$ and $\Theta_2^N \uparrow \Theta_2$ where Θ_1 and Θ_2 form a partition of $\mathbb{N}$ into two infinite subsets. The particles'

interaction is described by $a^{N,C} = -I$ and

$$a_{ij}^{N,P} = \begin{cases} \dfrac{1-\epsilon}{[\eta N]-1} & \text{if } i \neq j,\ i,j \in \Theta_1^N, \\[2mm] \dfrac{\epsilon}{N-[\eta N]} & \text{if } i \in \Theta_1^N,\ j \in \Theta_2^N, \\[2mm] \dfrac{1}{[\eta N]} & \text{if } i \in \Theta_2^N,\ j \in \Theta_1^N, \\[2mm] 0 & \text{if } i = j \text{ or } i,j \in \Theta_2^N, \end{cases}$$

with some $\epsilon \in (0,1)$. Here and throughout this example, we use the convention that all matrix identities at stage N are understood for the upper left $N \times N$-submatrix. All entries outside this submatrix are automatically set to zero.

Furthermore, we assume that $\rho^N = \rho^{N,C} = I$ and $\sigma^N = 0$ for all $N \in \mathbb{N}$. The initial condition and the driving noises are supposed to satisfy assumptions (A1) and (A4) of Theorem 4.1. We also suppose that the Lévy processes M_i have the same distribution among $i \in \Theta_1$, and among $i \in \Theta_2$, respectively. Then it is evident that also assumptions (A2) and (A3) hold with $\gamma(N) = \Gamma(N) = N$. Moreover, since $\rho^{N,P} = 0$, we have $R^N = 0$ and therefore (A6). Since $a^{N,C}$ and $\rho^{N,C}$ are diagonal, we deduce that $q_1 = 0$ and

$$q_2 = \limsup_{N \to \infty}\ \sup_{i,k=1,\ldots,N} N|a_{ik}^{N,P}\rho_{kk}^{N,C}| = \limsup_{N \to \infty}\ \sup_{i,k=1,\ldots,N} N|a_{ik}^{N,P}| \leq \eta^{-1} \vee (1-\eta)^{-1} < \infty,$$

which shows (A7).

Regarding (A5), we have $G^N(t,s) = N e^{a^N t} a^{N,P} e^{a^{N,C}s} \rho^{N,C} = e^{a^N t} N a^{N,P} e^{-s}$ by definition. Now classical theory for ordinary differential equations (ODE) asserts that $(G^N(t,s))_{\cdot,j}$, for $j = 1,\ldots,N$, is the solution to the ODE system

$$dx^{N,j,s}(t) = a^N x^{N,j,s}(t)\,dt, \quad x^{N,j,s}(0) = N a_{\cdot,j}^{N,P} e^{-s}. \tag{4.2}$$

By symmetry, we have $x_i^{N,j,s}(t) = y_1^{N,j,s}(t)$ for all $i \in \Theta_1^N$ and $x_i^{N,j,s}(t) = y_2^{N,j,s}(t)$ for all $i \in \Theta_2^N$, where the two-dimensional function $(y_1^{N,j,s}(t), y_2^{N,j,s}(t))$ solves the ODE

$$d\begin{pmatrix} y_1^{N,j,s} \\ y_2^{N,j,s} \end{pmatrix}(t) = -\begin{pmatrix} y_1^{N,j,s} \\ y_2^{N,j,s} \end{pmatrix}(t)\,dt + \begin{pmatrix} 1-\epsilon & \epsilon \\ 1 & 0 \end{pmatrix}\begin{pmatrix} y_1^{N,j,s} \\ y_2^{N,j,s} \end{pmatrix}(t)\,dt,$$

$$\begin{pmatrix} y_1^{N,j,s} \\ y_2^{N,j,s} \end{pmatrix}(0) = \begin{pmatrix} \dfrac{(1-\epsilon)N}{[\eta N]-1}e^{-s} \\[2mm] \dfrac{N}{[\eta N]}e^{-s} \end{pmatrix} \quad \text{if } j \in \Theta_1^N, \quad \text{and} \tag{4.3}$$

$$\begin{pmatrix} y_1^{N,j,s} \\ y_2^{N,j,s} \end{pmatrix}(0) = \begin{pmatrix} \dfrac{\epsilon N}{N-[\eta N]}e^{-s} \\ 0 \end{pmatrix} \quad \text{if } j \in \Theta_2^N.$$

The dependence on N is only via the initial condition, so clearly, as $N \to \infty$,

$$\sup_{j=1,\ldots,N}\ \sup_{s,t\in[0,T]} (|y_1^{N,j,s}(t) - y_1^{j,s}(t)| + |y_2^{N,j,s}(t) - y_2^{j,s}(t)|) \to 0, \tag{4.4}$$

where $(y_1^{j,s}, y_2^{j,s})$ satisfies the same ODE in (4.3), but with initial condition $((1-\epsilon)\eta^{-1}, \eta^{-1})e^{-s}$ if $j \in \Theta_1$, and $(\epsilon(1-\eta)^{-1}, 0)e^{-s}$ if $j \in \Theta_2$. Then (A5) holds with

$$G(t,s)_{ij} = \begin{cases} y_1^{j,s}(t) & \text{if } i \in \Theta_1, \\ y_2^{j,s}(t) & \text{if } i \in \Theta_2. \end{cases}$$

Due to symmetry, $G(t, s)_{ij}$ only depends on whether i and j belong to Θ_1 or Θ_2. In the following, we therefore write $G(t, s)_{kl}$, where $k, l \in \{1, 2\}$, if $i \in \Theta_k$ and $j \in \Theta_l$. Now suppose that $d \in \mathbb{N}$, $r \in [0, T]$ and $\theta \in M_T^d$ are given. If we write $\bar{\theta}_1 = \sum_{i \in \Theta_1^d} \theta_i$ and $\bar{\theta}_2 = \sum_{i \in \Theta_2^d} \theta_i$, then for $m \in \Theta_k$, where $k \in \{1, 2\}$,

$$H_m(\theta, r) = \bar{H}_k(\theta, r) := \int_r^T \int_s^T (G_{1k}(t - s, s - r)\bar{\theta}_1(\mathrm{d}t) + G_{2k}(t - s, s - r)\bar{\theta}_2(\mathrm{d}t))\,\mathrm{d}s.$$

Hence, the quantity in (4.1) is given by ($\bar{\Psi}_1$ and $\bar{\Psi}_2$ denote the logarithmic moment generating function of M_i for $i \in \Theta_1$ and $i \in \Theta_2$, respectively)

$$\lim_{N \to \infty} \frac{1}{N} \left(|\Theta_1^N| \int_0^T \bar{\Psi}_1(\bar{H}_1(\theta, r))\,\mathrm{d}r + |\Theta_2^N| \int_0^T \bar{\Psi}_2(\bar{H}_2(\theta, r))\,\mathrm{d}r \right)$$

$$= \eta \int_0^T \bar{\Psi}_1(\bar{H}_1(\theta, r))\,\mathrm{d}r + (1 - \eta) \int_0^T \bar{\Psi}_2(\bar{H}_2(\theta, r))\,\mathrm{d}r,$$

which is easily seen to be well defined, proving the last hypothesis (A8).

5. Proofs

We start with some preparatory results that are needed for the proof of Theorem 3.1.

Lemma 5.1. *Under the assumptions of Theorem 3.1, we have*

$$\sup_{i \in \mathbb{N}} \left\| \bar{X}_i^*(T) \right\|_{L^2} \leq V(T),$$

where $V(T)$ is given in Remark 3.2.

Proof. It is a consequence of (2.4) and the Burkholder–Davis–Gundy inequality that for all $t \in [0, T]$ and $i \in \mathbb{N}$,

$$\left\| (\bar{X}_i)^*(t) \right\|_{L^2} \leq \left\| X_i(0) \right\|_{L^2} + \int_0^t \sum_{j=1}^{\infty} A_{ij}(T) \left\| (\bar{X}_j)^*(s) \right\|_{L^2}\,\mathrm{d}s$$

$$+ 2\mathrm{Var}[L_i(1)] \left(\int_0^t \left(\sum_{j=1}^{\infty} \Sigma_{ij}(T) \left\| (\bar{X}_j)^*(s) \right\|_{L^2} \right)^2 \mathrm{d}s \right)^{1/2}$$

$$+ \int_0^t \sum_{j=1}^{\infty} \left| f_{ij}(s) \right| \left\| b_j(s) \right\|_{L^2}\,\mathrm{d}s$$

$$+ 2 \left(\sum_{j,k=1}^{\infty} \int_0^t \mathbb{E}\left[\rho_{ij}^{\mathrm{C}}(s)\rho_{ik}^{\mathrm{C}}(s)c_{jk}(s) \right] \mathrm{d}s \right)^{1/2}.$$

Therefore, if we define $w(t) := \sup_{i \in \mathbb{N}} \| (\bar{X}_i)^*(t) \|_{L^2}$, we obtain

$$w(t) \leq v_X + v_f(T)v_b(T)T + 2v_{\rho,M}(T)T^{1/2} + v_a(T) \int_0^t w(s)\,\mathrm{d}s$$

$$+ 2v_L v_\sigma(T) \left(\int_0^t (w(s))^2\,\mathrm{d}s \right)^{1/2}$$

$$\leq v_X + v_f(T)v_b(T)T + 2v_{\rho,M}(T)T^{1/2} + (v_a(T)T^{1/2}$$

$$+ 2v_L v_\sigma(T)) \left(\int_0^t (w(s))^2 \, \mathrm{d}s \right)^{1/2}.$$

Now we square the last inequality, apply the basic estimate $(a + b)^2 \leq 2(a^2 + b^2)$ and use Gronwall's inequality to deduce our claim, namely that

$$w(T) \leq \sqrt{2} e^{(v_a(T)T^{1/2} + 2v_L v_\sigma(T))^2 T} \left(v_X + v_f(T)v_b(T)T + 2v_{\rho,M}(T)T^{1/2} \right). \quad \square$$

Lemma 5.2. *Let $T \in \mathbb{R}_+$ and assume the finiteness of the numbers* (3.1). *We fix some $j \in \mathbb{N}$ throughout this lemma and define for $t \in [0, T]$,*

$$Y_j(t) := Y_j^1 + Y_j^2(t) + Y_j^3(t) + Y_j^4(t) + Y_j^5(t)$$

$$:= (X_j(0) - \mathbb{E}[X_j(0)]) + \sum_{k \neq j} \int_0^t a_{jk}^{\mathrm{C}}(s)(\bar{X}_k(s) - \mathbb{E}[\bar{X}_k(s)]) \, \mathrm{d}s$$

$$+ \sum_{k=1}^\infty \int_0^t \left(\sigma_{jk}^{\mathrm{C}}(s)\bar{X}_k(s-) + \sigma_{jk}^{\mathrm{P}}(s)\mathbb{E}[\bar{X}_k(s)] \right) \mathrm{d}L_j(s)$$

$$+ \sum_{k=1}^\infty \int_0^t f_{jk}^{\mathrm{C}}(s)(b_k(s) - \mathbb{E}[b_k(s)]) \, \mathrm{d}s + \sum_{k=1}^\infty \int_0^t \rho_{jk}^{\mathrm{C}}(s) \, \mathrm{d}M_k(s).$$

Furthermore, introduce the integrals

$$I_0^j[x](t) := x(t), \quad I_n^j[x](t) := \int_0^t a_{jj}^{\mathrm{C}}(s)I_{n-1}^j[x](s) \, \mathrm{d}s, \quad n \in \mathbb{N}, \tag{5.1}$$

where $x \colon [0, T] \to \mathbb{R}$ is a measurable function such that the integrals in (5.1) *exist for $t \in [0, T]$. Then*

$$\bar{X}_j(t) - \mathbb{E}[\bar{X}_j(t)] = \sum_{n=0}^\infty I_n^j[Y_j](t) = \sum_{\iota=1}^5 \sum_{n=0}^\infty I_n^j[Y_j^\iota](t), \quad t \in [0, T], \tag{5.2}$$

where the sums converge with respect to the maximal L^2-norm $X \mapsto \|X^*(T)\|_{L^2}$.

Proof. We deduce from (2.4) that

$$\bar{X}_j(t) - \mathbb{E}[\bar{X}_j(t)]$$

$$= (X_j(0) - \mathbb{E}[X_j(0)]) + \sum_{k \neq j} \int_0^t a_{jk}^{\mathrm{C}}(s)(\bar{X}_k(s) - \mathbb{E}[\bar{X}_k(s)]) \, \mathrm{d}s$$

$$+ \int_0^t a_{jj}^{\mathrm{C}}(s)(\bar{X}_j(s) - \mathbb{E}[\bar{X}_j(s)]) \, \mathrm{d}s + \sum_{k=1}^\infty \int_0^t \left(\sigma_{jk}^{\mathrm{C}}(s)\bar{X}_k(s-) + \sigma_{jk}^{\mathrm{P}}(s)\mathbb{E}[\bar{X}_k(s)] \right) \mathrm{d}L_j(s)$$

$$+ \sum_{k=1}^\infty \int_0^t f_{jk}^{\mathrm{C}}(s)(b_k(s) - \mathbb{E}[b_k(s)]) \, \mathrm{d}s + \sum_{k=1}^\infty \int_0^t \rho_{jk}^{\mathrm{C}}(s) \, \mathrm{d}M_k(s)$$

$$= I_1^j[\bar{X}_j - \mathbb{E}[\bar{X}_j]](t) + Y_j(t). \tag{5.3}$$

Iterating this equality n times, we obtain

$$\bar{X}_j(t) - \mathbb{E}[\bar{X}_j(t)] = \sum_{\nu=0}^{n-1} I_\nu^j[Y_j](t) + I_n^j[\bar{X}_j - \mathbb{E}[\bar{X}_j]](t), \quad t \in [0, T]. \tag{5.4}$$

Next, observe that for any càdlàg process $(X(t))_{t \in \mathbb{R}_+}$ with $\|X^*(T)\|_{L^2} < \infty$ we have

$$\left\|(I_\nu^j[X])^*(T)\right\|_{L^2} \leq \|X^*(T)\|_{L^2} \frac{(A_{jj}^{\mathrm{C}}(T))^\nu}{\nu!},$$

which is summable in ν. Thus, recalling from Lemma 5.1 that both Y_j and $\bar{X}_j - \mathbb{E}[\bar{X}_j]$ have finite maximal L^2-norm, we can let $n \to \infty$ in (5.4) and get

$$\bar{X}_j(t) - \mathbb{E}[\bar{X}_j(t)] = \sum_{\nu=0}^{\infty} I_\nu^j[Y_j](t), \quad t \in [0, T],$$

which is the first assertion. The second part of formula (5.2) holds by linearity. $\quad\square$

Proof of Theorem 3.1. The existence and uniqueness of solutions to (2.2) and (2.4) follow from the general theory of SDEs, see [33], Theorem V.7. Since the numbers (3.1) are finite, there are no difficulties in dealing with infinite-dimensional systems as in our case.

It follows from (2.2) and (2.4) that the difference between X and $\bar{X}$ satisfies the SDE

$$\begin{aligned}
\mathrm{d}(X(t) - \bar{X}(t)) &= \big(a(t)(X(t) - \bar{X}(t)) + a^{\mathrm{P}}(t)(\bar{X}(t) - \mathbb{E}[\bar{X}(t)])\big) \, \mathrm{d}t \\
&\quad + \big(\sigma(t)(X(t-) - \bar{X}(t-)) + \sigma^{\mathrm{P}}(t)(\bar{X}(t-) - \mathbb{E}[\bar{X}(t)])\big) . \mathrm{d}L(t) \\
&\quad + f^{\mathrm{P}}(t)(b(t) - \mathbb{E}[b(t)]) \, \mathrm{d}t + \rho^{\mathrm{P}}(t) \, \mathrm{d}M(t), \quad t \in \mathbb{R}_+, \\
X(0) - \bar{X}(0) &= 0.
\end{aligned}$$

Thus, denoting the left-hand side of (3.3) by $\Delta(T)$, we obtain from the Burkholder–Davis–Gundy inequality and Jensen's inequality that

$$\begin{aligned}
\Delta(T) &\leq v_a(T) \int_0^T \Delta(t) \, \mathrm{d}t + 2v_\sigma(T)v_L \left(\int_0^T (\Delta(t))^2 \, \mathrm{d}t\right)^{1/2} \\
&\quad + \left|\int_0^T \left\|a^{\mathrm{P}}(t)(\bar{X}(t) - \mathbb{E}[\bar{X}(t)])\right\|_{L^2} \mathrm{d}t\right|_\infty + \left|\left\|(\sigma^{\mathrm{P}}(\bar{X} - \mathbb{E}[\bar{X})] . L)^*(T)\right\|_{L^2}\right|_\infty \\
&\quad + T \sup_{t \in [0,T]} \left|\left\|f^{\mathrm{P}}(t)(b(t) - \mathbb{E}[b(t)])\right\|_{L^2}\right|_\infty + \left|\left\|(\rho^{\mathrm{P}} \cdot M)^*(T)\right\|_{L^2}\right|_\infty \\
&\leq (T^{1/2}v_a(T) + 2v_\sigma(T)v_L)\left(\int_0^T (\Delta(t))^2 \, \mathrm{d}t\right)^{1/2} + \sum_{\iota=1}^{4} \Delta^\iota(T),
\end{aligned} \tag{5.5}$$

where $\Delta^\iota(T)$ stands for the last four summands in the line before. So Gronwall's inequality produces the bound

$$\Delta(T) \leq K(T) \sum_{\iota=1}^{4} \Delta^\iota(T), \tag{5.6}$$

where $K(T) = \sqrt{2} \exp((T^{1/2}v_a(T) + 2v_\sigma(T)v_L)^2 T)$. We now consider each $\Delta^\iota(T)$ separately.

For $\iota = 3$ we simply have

$$\Delta^3(T) = T \sup_{t \in [0,T]} \sup_{i \in \mathbb{N}} \left(\sum_{j,k=1}^{\infty} f_{ij}^{\mathrm{P}}(t) f_{ik}^{\mathrm{P}}(t) \mathrm{Cov}[b_j(t), b_k(t)] \right)^{1/2} = T r_5(T). \tag{5.7}$$

For $\iota = 4$ another application of the Burkholder–Davis–Gundy inequality yields

$$\Delta^4(T) \leq 2 \sup_{i \in \mathbb{N}} \left(\sum_{j,k=1}^{\infty} \mathbb{E} \left[\int_0^T \rho_{ij}^{\mathrm{P}}(t) \rho_{ik}^{\mathrm{P}}(t) \, \mathrm{d}[M_j, M_k](t) \right] \right)^{1/2}$$

$$\leq 2T^{1/2} \sup_{i \in \mathbb{N}} \sup_{t \in [0,T]} \left(\sum_{j,k=1}^{\infty} \mathbb{E} \left[\rho_{ij}^{\mathrm{P}}(t) c_{jk}(t) \rho_{ik}^{\mathrm{P}}(t) \right] \right)^{1/2} = 2T^{1/2} r_6(T). \tag{5.8}$$

For $\iota = 1$, we use Lemma 5.2 including the notations introduced there as well as the fact that for all stochastic processes $(X(t))_{t \in \mathbb{R}_+}$ and $(Y(t))_{t \in \mathbb{R}_+}$ with càdlàg sample paths, we have

$$\sup_{r,s \in [0,T]} \left| \mathbb{E} \left[I_n^j[X](s) I_m^k[Y](r) \right] \right| \leq \frac{(A_{jj}^{\mathrm{C}}(T))^n}{n!} \frac{(A_{kk}^{\mathrm{C}}(T))^m}{m!} \sup_{r,s \in [0,T]} |\mathbb{E}[X(s)Y(r)]| \tag{5.9}$$

for any $j, k \in \mathbb{N}$ and $m, n \in \mathbb{N} \cup \{0\}$. In this way, we obtain

$$\Delta^1(T) = \left| \int_0^T \left\| a^{\mathrm{P}}(t)(\bar{X}(t) - \mathbb{E}[\bar{X}(t)]) \right\|_{L^2} \mathrm{d}t \right|_{\infty}$$

$$= \sup_{i \in \mathbb{N}} \int_0^T \left\| \sum_{j=1}^{\infty} a_{ij}^{\mathrm{P}}(t)(\bar{X}_j(t) - \mathbb{E}[\bar{X}_j(t)]) \right\|_{L^2} \mathrm{d}t$$

$$\leq \sum_{\iota=1}^5 \sup_{i \in \mathbb{N}} \int_0^T \left\| \sum_{j=1}^{\infty} a_{ij}^{\mathrm{P}}(t) \sum_{n=0}^{\infty} I_n^j[Y_j^\iota](t) \right\|_{L^2} \mathrm{d}t$$

$$= \sum_{\iota=1}^5 \sup_{i \in \mathbb{N}} \int_0^T \left(\sum_{j,k=1}^{\infty} \sum_{n,m=0}^{\infty} a_{ij}^{\mathrm{P}}(t) a_{ik}^{\mathrm{P}}(t) \mathbb{E} \left[I_n^j[Y_j^\iota](t) I_m^k[Y_k^\iota](t) \right] \right)^{1/2} \mathrm{d}t \tag{5.10}$$

$$\leq e^{|A^{\mathrm{C}}(T)|_{\mathrm{d}}} \sum_{\iota=1}^5 \sup_{i \in \mathbb{N}} \int_0^T \left(\sum_{j,k=1}^{\infty} A_{ij}^{\mathrm{P}}(T) A_{ik}^{\mathrm{P}}(T) \sup_{r,s \in [0,t]} \left| \mathbb{E}[Y_j^\iota(s) Y_k^\iota(r)] \right| \right)^{1/2} \mathrm{d}t$$

$$=: e^{|A^{\mathrm{C}}(T)|_{\mathrm{d}}} \sum_{\iota=1}^5 R_\iota(T).$$

Using Lemma 5.1, the five terms in (5.10) can be estimated as follows:

$$R_1(T) \leq T \sup_{i \in \mathbb{N}} \left(\sum_{j,k=1}^{\infty} A_{ij}^{\mathrm{P}}(T) A_{ik}^{\mathrm{P}}(T) |\mathrm{Cov}[X_j(0), X_k(0)]| \right)^{1/2} = T r_1(T),$$

$$R_2(T) \leq \sup_{i \in \mathbb{N}} \int_0^T \sum_{j=1}^{\infty} A_{ij}^{\mathrm{P}}(T) \sup_{s \in [0,t]} \|Y_j^2(s)\|_{L^2} \, \mathrm{d}t$$

$$\leq \sup_{i \in \mathbb{N}} \int_0^T \sum_{j=1}^{\infty} A_{ij}^{\mathrm{P}}(T) \left(\sum_{k \neq j} \int_0^t A_{jk}^{\mathrm{C}}(s) \| \bar{X}_k(s) - \mathbb{E}[\bar{X}_k(s)] \|_{L^2} \, \mathrm{d}s \right) \mathrm{d}t$$

$$\leq \frac{T^2}{2} V(T) \sup_{i \in \mathbb{N}} \sum_{j=1}^{\infty} \sum_{k \neq j} A_{ij}^{\mathrm{P}}(T) A_{jk}^{\mathrm{C}}(T) = \frac{T^2}{2} V(T) r_7(T),$$

$$R_3(T) \leq \sup_{i \in \mathbb{N}} \int_0^T \left(\sum_{j,k=1}^{\infty} A_{ij}^{\mathrm{P}}(T) A_{ik}^{\mathrm{P}}(T) \sup_{s \in [0,t]} \left| \mathbb{E}\left[\left((\sigma^{\mathrm{C}} \bar{X} + \sigma^{\mathrm{P}} \mathbb{E}[\bar{X}])_j \cdot L_j \right)(s) \right. \right. \right.$$

$$\left. \left. \left. \times \left((\sigma^{\mathrm{C}} \bar{X} + \sigma^{\mathrm{P}} \mathbb{E}[\bar{X}])_k \cdot L_k \right)(s) \right] \right| \right)^{1/2} \mathrm{d}t$$

$$\leq \sup_{i \in \mathbb{N}} \int_0^T \left(\sum_{j,k=1}^{\infty} A_{ij}^{\mathrm{P}}(T) A_{ik}^{\mathrm{P}}(T) \mathrm{Cov}[L_j(1), L_k(1)] \right.$$

$$\times \int_0^t \mathbb{E}\left[\left| \left(\sigma^{\mathrm{C}}(s) \bar{X}(s) + \sigma^{\mathrm{P}}(s) \mathbb{E}[\bar{X}(s)] \right)_j \right. \right.$$

$$\left. \left. \times \left(\sigma^{\mathrm{C}}(s) \bar{X}(s) + \sigma^{\mathrm{P}}(s) \mathbb{E}[\bar{X}(s)] \right)_k \right| \right] \mathrm{d}s \right)^{1/2} \mathrm{d}t$$

$$\leq \frac{2}{3} T^{3/2} v_\sigma(T) V(T) r_3(T),$$

$$R_4(T) \leq \sup_{i \in \mathbb{N}} \int_0^T \left(\sum_{j,k=1}^{\infty} A_{ij}^{\mathrm{P}}(T) A_{ik}^{\mathrm{P}}(T) \right.$$

$$\left. \times \int_0^t \int_0^t \left| \sum_{l,m=1}^{\infty} f_{jl}^{\mathrm{C}}(s) f_{km}^{\mathrm{C}}(r) \mathrm{Cov}[b_l(s), b_m(r)] \right| \mathrm{d}r \, \mathrm{d}s \right)^{1/2} \mathrm{d}t$$

$$\leq \frac{T^2}{2} \sup_{i \in \mathbb{N}} \left(\sum_{j,k=1}^{\infty} A_{ij}^{\mathrm{P}}(T) A_{ik}^{\mathrm{P}}(T) \sup_{s,t \in [0,T]} \left| \sum_{l,m=1}^{\infty} f_{jl}^{\mathrm{C}}(s) f_{km}^{\mathrm{C}}(t) \mathrm{Cov}[b_l(s), b_m(t)] \right| \right)^{1/2}$$

$$= \frac{T^2}{2} r_9(T),$$

$$R_5(T) \leq \sup_{i \in \mathbb{N}} \int_0^T \left(\sum_{j,k=1}^{\infty} A_{ij}^{\mathrm{P}}(T) A_{ik}^{\mathrm{P}}(T) \sup_{s \in [0,t]} \left| \sum_{l,m=1}^{\infty} \mathbb{E}\left[(\rho_{jl}^{\mathrm{C}} \cdot M_l)(s)(\rho_{km}^{\mathrm{C}} \cdot M_m)(s) \right] \right| \right)^{1/2} \mathrm{d}t$$

$$\leq \sup_{i \in \mathbb{N}} \int_0^T \left(\sum_{j,k=1}^{\infty} A_{ij}^{\mathrm{P}}(T) A_{ik}^{\mathrm{P}}(T) \int_0^t \left| \sum_{l,m=1}^{\infty} \mathbb{E}\left[\rho_{jl}^{\mathrm{C}}(s) c_{lm}(s) \rho_{km}^{\mathrm{C}}(s) \right] \right| \mathrm{d}s \right)^{1/2} \mathrm{d}t$$

$$\leq \frac{2}{3} T^{3/2} r_{11}(T).$$

The last step in the proof is the estimation of $\Delta^2(T)$. To this end, we make use of the Burkholder–Davis–Gundy inequality another time and get

$$\Delta^2(T) \leq \sup_{i \in \mathbb{N}} \left\| \left((\sigma^{\mathrm{P}}(\bar{X} - \mathbb{E}[\bar{X}]))_i \cdot L_i \right)^* (T) \right\|_{L^2}$$

$$\leq 2v_L \sup_{i \in \mathbb{N}} \left(\int_0^T \mathbb{E}\left[\left(\sigma^{\mathrm{P}}(t)(\bar{X}(t) - \mathbb{E}[\bar{X}(t)]) \right)_i^2 \right] dt \right)^{1/2}.$$

The further procedure is analogous to what we have done for $\Delta^1(T)$: instead of a^{P} we have σ^{P} here. We leave the details to the reader and only state the result, which is

$$\Delta^2(T) \leq \sum_{\iota \in \{2,4,8,10,12\}} K_\iota(T) r_\iota(T).$$

This completes the proof of Theorem 3.1. □

Our next goal is to prove Lemma 3.8 concerning the rate of growth of the maximal degree in the preferential attachment random graph as described in Section 3.3. For the undirected version as in [5], the corresponding result goes back to [30]. Indeed, the proof there basically works for our case as well, but there are some steps that require different arguments. Thus, we decided to include the proof to our lemma.

Proof of Lemma 3.8. The statement is evidently true for M^{in} when $\alpha + \beta = 0$ (resp. for M^{out} when $\beta + \gamma = 0$). In fact, for this extremal case, in every step of the random graph, a new edge is created pointing to (resp. from) a new node. This means that $M^{\mathrm{in}}(N)$ (resp. $M^{\mathrm{out}}(N)$) remains constant for all $N \in \mathbb{N}_0$, and the claim follows with $c^{\mathrm{in}} = 1$ (resp. $c^{\mathrm{out}} = 1$) identically. In the other cases, we closely follow the proof of Theorem 3.1 in [30]. In addition to the notation introduced in Section 3.3, we further define for $N \in \mathbb{N}_0$ and $\diamond \in \{\mathrm{in, out}\}$:

$$S^\diamond(N) := \nu + N + \delta^\diamond n(N),$$
$$X^\diamond(N, j) := \mathrm{d}^\diamond_{G(N)}(j) + \delta^\diamond, \quad j \in \mathbb{N},$$
$$N_j^\diamond := \inf\{N \in \mathbb{N}_0 : \mathrm{d}^\diamond_{G(N)}(j) \neq 0\}, \quad j \in \mathbb{N},$$
$$s^\diamond := \alpha \mathbb{1}_{\{\diamond = \mathrm{in}\}} + \beta + \gamma \mathbb{1}_{\{\diamond = \mathrm{out}\}},$$
$$c^\diamond(0, k) := 1, \quad c^\diamond(N + 1, k) := c^\diamond(N, k) \frac{S^\diamond(N)}{S^\diamond(N) + s^\diamond k}, \quad k \in \mathbb{R}_+,$$
$$Z^\diamond(N, j, k) := c^\diamond(N_j^\diamond + N, k) \binom{X^\diamond(N_j^\diamond + N, j) + k - 1}{k} \mathbb{1}_{\{N_j^\diamond < \infty\}}, \quad j \in \mathbb{N}, \quad k \in \mathbb{R}_+,$$
$$\mathcal{G}(N) := \sigma(e_{\nu+i} : i = 1, \ldots, N), \quad \mathcal{G}(\infty) := \sigma\left(\bigcup_{N=0}^\infty \mathcal{G}(N) \right).$$

Obviously, $\mathcal{G}(N)$ is the σ-field of all information up to step N in the preferential attachment algorithm, and $N_j^\diamond$ is a stopping time relative to the filtration $(\mathcal{G}(N))_{N \in \mathbb{N}}$ for every $j \in \mathbb{N}$. Analogously to Theorem 2.1 of [30], one can now show that for all $k \in \mathbb{R}_+$ and $j \in \mathbb{N}$ the sequence $(Z^\diamond(N, j, k))_{N \in \mathbb{N}_0}$ is a positive martingale relative to the filtration $(\mathcal{G}(N_j^\diamond + N))_{N \in \mathbb{N}_0}$. As a consequence, Doob's martingale convergence theorem implies that

$$Z^\diamond(N, j, k) \to \zeta^\diamond(j, k) \quad \text{a.s.} \tag{5.11}$$

for some random variable $\zeta^\diamond(j, k)$. The convergence in (5.11) also holds in L^p for all $p \in [1, \infty)$ because we have

$$Z^\diamond(N, j, k)^p \le C(k, p) Z^\diamond(N, j, kp) \quad \text{a.s.} \tag{5.12}$$

for some deterministic constant $C(k, p) \in \mathbb{R}_+$ that is independent of N and j. Indeed, on $\{N_j^\circ < \infty\}$ we have by definition,

$$\frac{Z^\diamond(N, j, k)^p}{Z^\diamond(N, j, kp)} = \frac{c^\diamond(N_j^\circ + N, k)^p}{c^\diamond(N_j^\circ + N, kp)} \binom{X^\diamond(N_j^\circ + N, j) + k - 1}{k}^p$$
$$\times \binom{X^\diamond(N_j^\circ + N, j) + kp - 1}{kp}^{-1},$$

where

$$\frac{c^\diamond(N, k)^p}{c^\diamond(N, kp)} = \frac{c^\diamond(N-1, k)^p}{c^\diamond(N-1, kp)} \frac{S^\diamond(N)^{p-1}(S^\diamond(N) + s^\diamond kp)}{(S^\diamond(N) + s^\diamond k)^p}$$
$$\le \frac{c^\diamond(N-1, k)^p}{c^\diamond(N-1, kp)} \le \cdots \le \frac{c^\diamond(0, k)^p}{c^\diamond(0, kp)} = 1,$$
$$\binom{x+k-1}{k}^p \binom{x+kp-1}{kp}^{-1} = \frac{\Gamma(kp+1)}{\Gamma(k+1)^p} \frac{\Gamma(k+x)^p}{\Gamma(x)^{p-1}\Gamma(kp+x)} \xrightarrow{x \to \infty} \frac{\Gamma(kp+1)}{\Gamma(k+1)^p},$$

which shows (5.12). Next, define for $N \in \mathbb{N}_0$ and $j \in \mathbb{N}$,

$$m^\diamond(N, j) := \max\{Z^\diamond(N - N_i^\circ, i, 1) : i = 1, \ldots, j, N_i^\circ \le N\}, \quad m^\diamond(N) := m^\diamond(N, n(N)),$$
$$\mu^\diamond(j) := \max\{\zeta^\diamond(i, 1) : i = 1, \ldots, j\}, \qquad \mu^\diamond := \sup\{\mu^\diamond(j) : j \in \mathbb{N}\},$$

such that in particular the relationship $m^\diamond(N) = c^\diamond(N, 1)(M^\diamond(N) + \delta^\diamond)$ holds. It is not hard to see that $(m^\diamond(N))_{N \in \mathbb{N}_0}$, as the maximum of martingale expressions, is a submartingale relative to $(\mathcal{G}(N))_{N \in \mathbb{N}_0}$. By definition, the sequence $(c^\diamond(N, k))_{N \in \mathbb{N}_0}$ decreases to 0 as $N \to \infty$; more precisely, we have

$$c^\diamond(N, k) = c^\diamond(N-1, k) \frac{S^\diamond(N-1)}{S^\diamond(N-1) + s^\diamond k}$$
$$\le c^\diamond(N-1, k) \frac{\nu + N - 1 + \delta^\diamond(n(0) + N - 1)}{\nu + N - 1 + \delta^\diamond(n(0) + N - 1) + s^\diamond k}$$
$$\le \prod_{j=0}^{N-1} \frac{(1 + \delta^\diamond)j + \delta^\diamond n(0) + \nu}{(1 + \delta^\diamond)j + \delta^\diamond n(0) + \nu + s^\diamond k}$$
$$= \frac{\Gamma\left(N + \frac{\delta^\diamond n(0) + \nu}{1 + \delta^\diamond}\right)}{\Gamma\left(N + \frac{\delta^\diamond n(0) + \nu + s^\diamond k}{1 + \delta^\diamond}\right)} \sim N^{-\frac{s^\diamond k}{1 + \delta^\diamond}}, \quad N \to \infty.$$

As a consequence, when p is large enough,

$$
\mathbb{E}[m^\diamond(N)^p] \leq \mathbb{E}\left[\sum_{i=1}^{n(N)} Z^\diamond(N - N_i^\diamond, i, 1)^p\right] \leq C(1, p)\mathbb{E}\left[\sum_{i=1}^{n(N)} Z^\diamond(N - N_i^\diamond, i, p)\right]
$$

$$
\leq C(1, p) \sum_{i=1}^{\infty} \mathbb{E}[Z^\diamond(0, i, p)] \tag{5.13}
$$

$$
\leq C(1, p)\binom{n(0) + p + \delta^\diamond - 1}{p} \sum_{i=1}^{\infty} \mathbb{E}[c^\diamond(N_i^\diamond, p)]
$$

$$
\leq C(1, p)\binom{n(0) + p + \delta^\diamond - 1}{p}\left(n(0) + \sum_{i=1}^{\infty} \mathbb{E}[c^\diamond(i, p)]\right) < \infty,
$$

independently of N. This implies that the submartingale $m^\diamond$ converges a.s. and in L^p for all $p \in [1, \infty)$. It follows from (5.13) that for $j \geq n(0)$ we have

$$
\mathbb{E}[(m^\diamond(N) - m^\diamond(N, j))^p] \leq \mathbb{E}\left[\sum_{i=j+1}^{n(N)} Z^\diamond(N - N_i^\diamond, i, 1)^p\right]
$$

$$
\leq C(1, p)\binom{n(0) + p + \delta^\diamond - 1}{p} \sum_{i=j-n(0)+1}^{\infty} \mathbb{E}[c^\diamond(i, p)]. \tag{5.14}
$$

As $N \to \infty$, the left-hand side of (5.14) converges to

$$
\mathbb{E}\left[\left(\lim_{N \to \infty} c^\diamond(N, 1)M^\diamond(N) - \mu^\diamond(j)\right)^p\right],
$$

while the right-hand side is independent of N. Now taking the limit $j \to \infty$ and again assuming that p is large, we obtain the desired result (3.13). Note at this point that $\mu^\diamond$ is indeed an a.s. finite random variable that belongs to L^p for all $p \in [1, \infty)$, which is proved using a similar argument as in (5.13).

It remains to prove (3.14). To this end, observe that by the law of large numbers, we have $(n(N) - n(0))/N \to \alpha + \gamma$ a.s. In other words, there exists for every $\epsilon \in (0, \alpha + \gamma)$, a random $\bar{N} \in \mathbb{N}$ such that for all $N \geq \bar{N}$, we have

$$
\left|\frac{n(N) - n(0)}{N} - (\alpha + \gamma)\right| \leq \epsilon,
$$

or, equivalently, $n(N) \in [n(0) + (\alpha + \gamma - \epsilon)N, n(0) + (\alpha + \gamma + \epsilon)N]$. Consequently, using the identity $\prod_{i=0}^{n-1}(a + bi) = b^n \Gamma(\frac{a}{b} + n)\Gamma(\frac{a}{b})^{-1}$, we deduce for all $k \in \mathbb{N}$ and $N \geq \bar{N}$,

$$
c^\diamond(N, k) = \prod_{i=0}^{N-1} \frac{S^\diamond(i)}{S^\diamond(i) + s^\diamond k} \geq \prod_{i=0}^{\bar{N}-1} \frac{S^\diamond(i)}{S^\diamond(i) + s^\diamond k}
$$

$$
\times \prod_{i=\bar{N}}^{N-1} \frac{\nu + i + \delta^\diamond(n(0) + (\alpha + \gamma - \epsilon)i)}{\nu + i + \delta^\diamond(n(0) + (\alpha + \gamma - \epsilon)i) + s^\diamond k}
$$

$$
= \prod_{i=0}^{\bar{N}-1} \frac{S^\diamond(i)}{S^\diamond(i) + s^\diamond k} \prod_{i=0}^{\bar{N}-1} \frac{\nu + i + \delta^\diamond(n(0) + (\alpha + \gamma - \epsilon)i) + s^\diamond k}{\nu + i + \delta^\diamond(n(0) + (\alpha + \gamma - \epsilon)i)}
$$

$$\times \prod_{i=0}^{N-1} \frac{\nu + i + \delta^\diamond (n(0) + (\alpha + \gamma - \epsilon)i)}{\nu + i + \delta^\diamond (n(0) + (\alpha + \gamma - \epsilon)i) + s^\diamond k}$$

$$= c^\diamond(\bar{N}, k) \frac{\Gamma\left(\bar{N} + \frac{\delta^\diamond n(0) + \nu + s^\diamond k}{1 + \delta^\diamond(\alpha + \gamma - \epsilon)}\right) \Gamma\left(N + \frac{\delta^\diamond n(0) + \nu}{1 + \delta^\diamond(\alpha + \gamma - \epsilon)}\right)}{\Gamma\left(\bar{N} + \frac{\delta^\diamond n(0) + \nu}{1 + \delta^\diamond(\alpha + \gamma - \epsilon)}\right) \Gamma\left(N + \frac{\delta^\diamond n(0) + \nu + s^\diamond k}{1 + \delta^\diamond(\alpha + \gamma - \epsilon)}\right)}$$

$$\sim c^\diamond(\bar{N}, k) \frac{\Gamma\left(\bar{N} + \frac{\delta^\diamond n(0) + \nu + s^\diamond k}{1 + \delta^\diamond(\alpha + \gamma - \epsilon)}\right)}{\Gamma\left(\bar{N} + \frac{\delta^\diamond n(0) + \nu}{1 + \delta^\diamond(\alpha + \gamma - \epsilon)}\right)} N^{-\frac{s^\diamond k}{1 + \delta^\diamond(\alpha + \gamma - \epsilon)}}, \qquad N \to \infty.$$

So choosing $c^\diamond(N) := c^\diamond(N, 1)$ for $N \in \mathbb{N}_0$ satisfies (3.14). $\quad\square$

Finally, we turn to the proof of the large deviation result in Section 4.

Proof of Theorem 4.1. By definition, $X^N - \bar{X}^N$ satisfies the equation

$$X^N(t) - \bar{X}^N(t) = \int_0^t a^N (X^N - \bar{X}^N)(s)\, \mathrm{d}s + \int_0^t a^{N,\mathrm{P}} (\bar{X}^N(s) - \mathbb{E}[\bar{X}^N(s)])\, \mathrm{d}s$$
$$+ \rho^{N,\mathrm{P}} M(t), \quad t \in [0, T],$$

which implies that

$$X^N(t) - \bar{X}^N(t) = \int_0^t \mathrm{e}^{a^N(t-s)} a^{N,\mathrm{P}} (\bar{X}^N(s) - \mathbb{E}[\bar{X}^N(s)])\, \mathrm{d}s$$
$$+ \int_0^t \mathrm{e}^{a^N(t-s)} \rho^{N,\mathrm{P}}\, \mathrm{d}M(s), \quad t \in [0, T]. \tag{5.15}$$

In order to establish a large deviation principle, it suffices by Theorem 4.6.1 of [15] to prove such a principle in $(D_T^d, \mathcal{D}_T^d, J_1)$ for the first d coordinates of the process $X^N - \bar{X}^N$, for all $d \in \mathbb{N}$. With the formula

$$\bar{X}_i^N(t) - \mathbb{E}[\bar{X}_i^N(t)] = \int_0^t \mathrm{e}^{a_{ii}^{N,\mathrm{C}}(t-s)} \sum_{j \neq i} a_{ij}^{N,\mathrm{C}} (\bar{X}_j^N(s) - \mathbb{E}[\bar{X}_j^N(s)])\, \mathrm{d}s$$
$$+ \sum_{j=1}^\infty \int_0^t \mathrm{e}^{a_{ii}^{N,\mathrm{C}}(t-s)} \rho_{ij}^{N,\mathrm{C}}\, \mathrm{d}M_j(s),$$

which is valid for all $i \in \mathbb{N}$ and $t \in \mathbb{R}_+$ and follows from the first identity in (5.3), the first d coordinates of $X^N - \bar{X}^N$ are given by the D_T^d-valued process

$$Y_i^N(t) := Y_i^{N,1}(t) + Y_i^{N,2}(t) + Y_i^{N,3}(t)$$

$$:= \int_0^t \int_0^s \sum_{j,k=1}^\infty \mathrm{e}^{a^N(t-s)}_{ij} a_{jk}^{N,\mathrm{P}} \mathrm{e}^{a_{kk}^{N,\mathrm{C}}(s-r)} \sum_{l \neq k} a_{kl}^{N,\mathrm{C}} (\bar{X}_l^N(r) - \mathbb{E}[\bar{X}_l^N(r)])\, \mathrm{d}r\, \mathrm{d}s$$

$$+ \int_0^t \sum_{j,k,l=1}^\infty \mathrm{e}^{a^N(t-s)}_{ij} a_{jk}^{N,\mathrm{P}} \int_0^s \mathrm{e}^{a_{kk}^{N,\mathrm{C}}(s-r)} \rho_{kl}^{N,\mathrm{C}}\, \mathrm{d}M_l(r)\, \mathrm{d}s \tag{5.16}$$

$$+ \sum_{j,k=1}^\infty \int_0^t \mathrm{e}^{a^N(t-s)}_{ij} \rho_{jk}^{N,\mathrm{P}}\, \mathrm{d}M_k(s), \quad i = 1, \dots, d, \quad t \in [0, T].$$

Actually, we will even prove the large deviation principle in $(D_T^d, \mathcal{D}_T^d, U)$, which is stronger. To this end, we introduce the notation

$$\hat{x}(t) := \sum_{k=1}^{[\gamma(N)T]-1} x\left(\frac{k}{\gamma(N)}\right) \mathbb{1}_{\left[\frac{k}{\gamma(N)}, \frac{k+1}{\gamma(N)}\right)}(t) + x\left(\frac{[\gamma(N)T]}{\gamma(N)}\right) \mathbb{1}_{\left[\frac{[\gamma(N)T]}{\gamma(N)}, T\right)}(t)$$

for $t \in [0, T]$ and $x \in D_T^d$. Then by Theorem 4.2.13 of [15] and Lemma 5.3, it suffices to show a large deviation principle for $\hat{Y}^N = \hat{Y}^{N,1} + \hat{Y}^{N,2} + \hat{Y}^{N,3}$, which is done in Lemma 5.5. The same principle will then hold for Y^N. That the rate function for $X^N - \bar{X}^N$ is convex with unique minimum 0 at 0, and can only be finite for functions in AC_T^∞, is inherited from the rate function of $\hat{Y}^N$.   □

Lemma 5.3.   *For each $d \in \mathbb{N}$ and $\iota = 1, 2, 3$, the D_T^d-valued processes $Y^{N,\iota}$ and $\hat{Y}^{N,\iota}$ are exponentially equivalent, that is, we have for all $\epsilon \in (0, 1)$,*

$$\lim_{N \to \infty} \frac{1}{\gamma(N)} \log \mathbb{P}\left[\sup_{t \in [0,T]} \sup_{i=1,\dots,d} |Y_i^{N,\iota}(t) - \hat{Y}_i^{N,\iota}(t)| > \epsilon\right] = -\infty.$$

Proof.   We start with $\iota = 1$. Writing $\hat{t} = [\gamma(N)t]/\gamma(N)$ and $\mathrm{diag}(a) := a - a^\times$, we obtain

$$\sup_{t \in [0,T]} \sup_{i=1,\dots,d} \left|Y_i^{N,1}(t) - \hat{Y}_i^{N,1}(t)\right|$$

$$\leq \sup_{t \in [0,T]} \left|(\mathrm{e}^{a^N t} - \mathrm{e}^{a^N \hat{t}}) \left(\int_0^t \int_0^s \mathrm{e}^{-a^N s} a^{N,\mathrm{P}} \mathrm{e}^{\mathrm{diag}(a^{N,\mathrm{C}})(s-r)} (a^{N,\mathrm{C}})^\times\right.\right.$$

$$\left.\left. \times (\bar{X}^N(r) - \mathbb{E}[\bar{X}^N(r)]) \, \mathrm{d}r \, \mathrm{d}s\right)\right|_\infty \tag{5.17}$$

$$+ \sup_{t \in [0,T]} \left|\mathrm{e}^{a^N \hat{t}} \left(\int_{\hat{t}}^t \int_0^s \mathrm{e}^{-a^N s} a^{N,\mathrm{P}} \mathrm{e}^{\mathrm{diag}(a^{N,\mathrm{C}})(s-r)} (a^{N,\mathrm{C}})^\times\right.\right.$$

$$\left.\left. \times (\bar{X}^N(r) - \mathbb{E}[\bar{X}^N(r)]) \, \mathrm{d}r \, \mathrm{d}s\right)\right|_\infty.$$

We can proceed with these two terms separately. Since $|\mathrm{e}^{a^N t} - \mathrm{e}^{a^N \hat{t}}|_\infty \leq v_a \mathrm{e}^{v_a T}/\gamma(N)$, we have for the first term in (5.17),

$$\mathbb{P}\left[\sup_{t \in [0,T]} \left|(\mathrm{e}^{a^N t} - \mathrm{e}^{a^N \hat{t}}) \left(\int_0^t \int_0^s \mathrm{e}^{-a^N s} a^{N,\mathrm{P}} \mathrm{e}^{\mathrm{diag}(a^{N,\mathrm{C}})(s-r)} (a^{N,\mathrm{C}})^\times\right.\right.\right.$$

$$\left.\left.\left. \times (\bar{X}^N(r) - \mathbb{E}[\bar{X}^N(r)]) \, \mathrm{d}r \, \mathrm{d}s\right)\right|_\infty > \epsilon\right]$$

$$\leq \mathbb{P}\left[\sup_{t \in [0,T]} |a^{N,\mathrm{P}}(a^{N,\mathrm{C}})^\times(\bar{X}^N(t) - \mathbb{E}[\bar{X}^N(t)])|_\infty > \frac{\epsilon \gamma(N)}{v_a(\mathrm{e}^{v_a T})^3 T^2}\right]$$

$$\leq \mathbb{P}\left[\sup_{t \in [0,T]} \sup_{i \in \mathbb{N}} \sum_{j=1}^\infty \sum_{k \neq j} |a_{ij}^{N,\mathrm{P}} a_{jk}^{N,\mathrm{C}} (\bar{X}_k^N(t) - \mathbb{E}[\bar{X}_k^N(t)])| > \frac{\epsilon \gamma(N)}{v_a(\mathrm{e}^{v_a T})^3 T^2}\right] =: p(N).$$

We note that $\xi^N := \bar{X}^N - \mathbb{E}[\bar{X}^N]$ satisfies the integral equation

$$\xi^N(t) = \int_0^t a^{N,C}\xi^N(s)\,\mathrm{d}s + \rho^{N,C}M(t), \quad t \in \mathbb{R}_+. \tag{5.18}$$

Hence, we have

$$(\xi^N)^*(t) \le \int_0^t |a^{N,C}|(\xi^N)^*(s)\,\mathrm{d}s + (\rho^{N,C}M)^*(t), \quad t \in \mathbb{R}_+,$$

or after $n \in \mathbb{N}$ iterations,

$$(\xi^N)^*(t) \le \frac{(|a^{N,C}|t)^n}{n!}(\xi^N)^*(t) + \sum_{m=0}^{n-1}\frac{(|a^{N,C}|t)^m}{m!}(\rho^{N,C}M)^*(t).$$

Since $(I - A)^{-1} = \sum_{k=0}^\infty A^k$ only has nonnegative entries as soon as A (with $|A|_\infty < 1$) does, we obtain

$$(\xi^N)^*(t) \le \left(I - \frac{(|a^{N,C}|T)^n}{n!}\right)^{-1}\sum_{m=0}^{n-1}\frac{(|a^{N,C}|t)^m}{m!}(\rho^{N,C}M)^*(t)$$

if n is large enough such that $(v_aT)^n/n! < 1$. It is not difficult to recognize that the exact value of n only affects some constants in the subsequent arguments with no impact on the final result; we therefore assume without loss of generality that $n = 1$ (i.e. $v_aT < 1$). Then

$$p(N) \le \mathbb{P}\Bigg[\sup_{i\in\mathbb{N}}\sum_{j=1}^\infty\sum_{k\ne j}\sum_{l=1}^\infty |a_{ij}^{N,P}a_{jk}^{N,C}|(I - |a^{N,C}|T)_{kl}^{-1}$$

$$\times \sup_{t\in[0,T]}\sum_{m=1}^{\gamma(N)}|\rho_{lm}^{N,C}M_m(t)| > \frac{\epsilon\gamma(N)}{v_a(e^{v_aT})^3T^2}\Bigg]$$

$$\le \mathbb{P}\Bigg[\frac{1}{\gamma(N)}\sup_{l\in\mathbb{N}}\sup_{t\in[0,T]}\sum_{m=1}^{\gamma(N)}|\rho_{lm}^{N,C}M_m(t)| > \frac{\epsilon\gamma(N)(1 - v_aT)}{q_1(N)v_a(e^{v_aT})^3T^2}\Bigg].$$

Let $\lambda(N)$ be positive numbers to be chosen later. Using Doob's maximal inequality for submartingales, and the independence of the Lévy processes M_i, we arrive at

$$p(N) \le \exp\left(-\frac{\epsilon\lambda(N)\gamma(N)(1 - v_aT)}{q_1(N)v_a(e^{v_aT})^3T^2}\right)\prod_{m=1}^{\gamma(N)}\mathbb{E}\left[\exp\left(\frac{\lambda(N)}{\gamma(N)}\sup_{l\in\mathbb{N}}|\rho_{lm}^{N,C}||M_m(T)|\right)\right]$$

$$\le 2^{\gamma(N)}\exp\left(-\frac{\epsilon\lambda(N)\gamma(N)(1 - v_aT)}{q_1(N)v_a(e^{v_aT})^3T^2}\right)\exp\left(T\gamma(N)\Psi_0\left(\frac{\lambda(N)}{\gamma(N)}\sup_{l,m\in\mathbb{N}}|\rho_{lm}^{N,C}|\right)\right).$$

Now define

$$\lambda(N) := \gamma(N)\Psi_0^{-1}(1)\Big/\left(\sup_{l,m\in\mathbb{N}}|\rho_{lm}^{N,C}|\right), \quad N \in \mathbb{N}.$$

Since Ψ_0 is an increasing, convex function with $\Psi_0(0) = 0$, its inverse Ψ_0^{-1} is increasing, concave with $\Psi^{-1}(0) = 0$, and therefore we have for large N that $\lambda(N) \ge \Psi_0^{-1}(\gamma(N))\Big/\left(\sup_{l,m\in\mathbb{N}}|\rho_{lm}^{N,C}|\right)$, which increases to infinity as $N \to \infty$. With this choice of $\lambda(N)$, it follows that

$$\lim_{N\to\infty}\frac{1}{\gamma(N)}\log p(N) = -\infty,$$

and the proof for the first term in (5.17) is complete. The second term can be treated in an analogous way: now the factor $\gamma(N)$ does not come from the difference $|e^{a^N t} - e^{a^N \hat{t}}|_\infty$, but from the domain of integration $(\hat{t}, t]$. The details are left to the reader.

For $\iota = 2$ similar methods apply. Also here we do not give the details. Instead, we sketch the proof for $\iota = 3$ where some modifications are necessary. Recalling the meaning of $\Gamma(N)$ from (A2), we have

$$
\sup_{t \in [0,T]} \sup_{i=1,\dots,d} \left| Y_i^{N,3}(t) - \hat{Y}_i^{N,3}(t) \right|
$$
$$
\leq \sup_{t \in [0,T]} \sup_{i=1,\dots,d} \left| \left((e^{a^N t} - e^{a^N \hat{t}}) \left(\int_0^t e^{-a^N s} \rho^{N,P} \, dM(s) \right) \right)_i \right|
$$
$$
+ \sup_{t \in [0,T]} \sup_{i=1,\dots,d} \left| \left(e^{a^N \hat{t}} \left(\int_{\hat{t}}^t e^{-a^N s} \rho^{N,P} \, dM(s) \right) \right)_i \right| \tag{5.19}
$$
$$
\leq |e^{a^N t} - e^{a^N \hat{t}}|_\infty \sup_{t \in [0,T]} \sup_{i=1,\dots,\Gamma(N)} \left| \left(\int_0^t e^{-a^N s} \rho^{N,P} \, dM(s) \right)_i \right|
$$
$$
+ e^{v_a T} \sup_{t \in [0,T]} \sup_{i=1,\dots,\Gamma(N)} \left| \left(\int_{\hat{t}}^t e^{-a^N s} \rho^{N,P} \, dM(s) \right)_i \right|.
$$

We can again consider these two terms separately. For the first one, we have

$$
\mathbb{P}\left[|e^{a^N t} - e^{a^N \hat{t}}|_\infty \sup_{t \in [0,T]} \sup_{i=1,\dots,\Gamma(N)} \left| \left(\int_0^t e^{-a^N s} \rho^{N,P} \, dM(s) \right)_i \right| > \epsilon \right]
$$
$$
\leq \Gamma(N) \sup_{i=1,\dots,\Gamma(N)} \mathbb{P}\left[\sup_{t \in [0,T]} \left| \left(\int_0^t e^{-a^N s} \rho^{N,P} \, dM(s) \right)_i \right| > \frac{\epsilon \gamma(N)}{v_a e^{v_a T}} \right] \tag{5.20}
$$
$$
\leq \Gamma(N) \sup_{i=1,\dots,\Gamma(N)} \exp\left(-\frac{\epsilon \lambda(N) \gamma(N)}{v_a e^{v_a T}} \right)
$$
$$
\times \prod_{k=1}^{\gamma(N)} \mathbb{E}\left[\exp\left(\frac{\lambda(N)}{\gamma(N)} \left| \int_0^T \sum_{j=1}^\infty e_{ij}^{-a^N s} \gamma(N) \rho_{jk}^{N,P} \, dM_k(s) \right| \right) \right].
$$

Observe now that by (A6), we can find a constant $C \in \mathbb{R}_+$ that is independent of i, k and N such that

$$
\sup_{i,k \in \mathbb{N}} \left| \sum_{j=1}^\infty e_{ij}^{-a^N s} \gamma(N) \rho_{jk}^{N,P} \right| \leq |e^{-a^N s} \gamma(N) \rho^{N,P}|_\infty \leq e^{v_a(T)T} \sup_{i,j \in \mathbb{N}} |R_{ij}^N(0)| \leq C.
$$

Recalling from [34], Theorem 2.7, that the stochastic integral in the last line of (5.20) has an infinitely divisible distribution with explicitly known exponential moments, we deduce that

$$
\mathbb{E}\left[\exp\left(\frac{\lambda(N)}{\gamma(N)} \left| \int_0^T \sum_{j=1}^\infty e_{ij}^{-a^N s} \gamma(N) \rho_{jk}^{N,P} \, dM_k(s) \right| \right) \right] \leq 2\mathbb{E}\left[\exp\left(\frac{\lambda(N)}{\gamma(N)} C M_0(T) \right) \right].
$$

Therefore, the remaining calculation can be completed as in the case $\iota = 1$.

For the second term in (5.19), the reasoning is the same, except that the factor $\gamma(N)$ is now due to the domain $(\hat{t}, t]$ of the stochastic integral. Observe at this point that $M_k(t) - M_k(\hat{t})$ has the same distribution as $M_k(t - \hat{t})$ and that $|t - \hat{t}| \leq 1/\gamma(N)$. Again, we do not carry out the details. $\quad\square$

Lemma 5.4. *For each $\iota = 1, 2, 3$ the processes $(\hat{Y}^{N,\iota}: N \in \mathbb{N})$ form an* exponentially tight *sequence in $(D_T^d, \mathcal{D}_T^d, U)$, that is, for every $L \in \mathbb{R}_+$, there exists a compact subset K^L of D_T^d (with respect to the uniform topology U) such that*

$$\frac{1}{\gamma(N)} \log \mathbb{P}[\hat{Y}^{N,\iota} \notin K^L] \leq -L.$$

Proof. We first consider $\iota = 1$. We will adapt the idea of Lemma 4.1 in [1] to our setting. As shown in part (I) of the proof there, it suffices to show that for every $a, \epsilon \in (0, \infty)$, there exist a compact set $H \subseteq D_T^d$, some $C \in (0, \infty)$ and $n \in \mathbb{N}$ such that for all $N \geq n$,

$$\mathbb{P}[d(\hat{Y}^{N,1}, H) > \epsilon] \leq C e^{-\gamma(N)a}, \tag{5.21}$$

where $d(f, H) := \inf\{\sup_{t \in [0,T]} \sup_{i=1,\dots,d} |f_i(t) - g_i(t)|: g \in H\}$ for $f \in D_T^d$. In order to prove (5.21), we first define for $n \in \mathbb{N}$ and $A \subseteq \mathbb{R}^d$,

$$H_n(A) := \left\{ f \in D_T^d: f = \sum_{\kappa=1}^{[\gamma(n)T]-1} x_\kappa \mathbb{1}_{\left[\frac{\kappa}{\gamma(n)}, \frac{\kappa+1}{\gamma(n)}\right)} + x_{[\gamma(n)T]} \mathbb{1}_{\left[\frac{[\gamma(n)T]}{\gamma(n)}, T\right]}, \right.$$

$$\left. x_1, \dots, x_{[\gamma(n)T]} \in A \right\}.$$

It follows from Equation (4.3) of [1] that for $N \geq n$, $A \subseteq \mathbb{R}^d$ and $f \in H_N(A)$, we have

$$d(f, H_n(A))$$

$$\leq \sup_{\kappa=0,\dots,[\gamma(n)T]-1} \sup_{\lambda \in \left[1, \frac{\gamma(N)}{\gamma(n)}+1\right)} \sup_{i=1,\dots,d} \left| f_i\left(\frac{\left[\frac{\gamma(N)\kappa}{\gamma(n)}\right] + \lambda}{\gamma(N)} \wedge T\right) - f_i\left(\frac{\left[\frac{\gamma(N)\kappa}{\gamma(n)}\right]}{\gamma(N)}\right) \right|. \tag{5.22}$$

Next, let $\beta \in (0, \infty)$ and define $K := [-1, 1]^d$ such that in particular, $H_n(\beta K)$ is a compact set. Then for every $N \geq n$, we have

$$\mathbb{P}[d(\hat{Y}^{N,1}, H_n(\beta K)) > \epsilon] \leq \mathbb{P}[\hat{Y}^{N,1} \notin H_N(\beta K)]$$
$$+ \mathbb{P}[\hat{Y}^{N,1} \in H_N(\beta K), d(\hat{Y}^{N,1}, H_n(\beta K)) > \epsilon]. \tag{5.23}$$

The first probability is bounded as follows:

$$\mathbb{P}[\hat{Y}^{N,1} \notin H_N(\beta K)]$$

$$= \mathbb{P}\left[\sup_{i=1,\dots,d} \sup_{\kappa=1,\dots,[\gamma(N)T]} \left| \int_0^{\frac{\kappa}{\gamma(N)}} \int_0^s \sum_{j,k=1}^\infty e_{ij}^{a^N(\kappa/\gamma(N)-s)} a_{jk}^{N,P} e^{a_{kk}^{N,C}(s-r)} \right. \right.$$

$$\left. \left. \times \sum_{l \neq k} a_{kl}^{N,C}(\bar{X}_l^N(r) - \mathbb{E}[\bar{X}_l^N(r)]) \, dr \, ds \right| > \beta \right]$$

$$\leq \mathbb{P}\left[\sup_{t \in [0,T]} \sup_{i=1,\dots,d} \sum_{j=1}^\infty \sum_{k \neq j} |a_{ij}^{N,P} a_{jk}^{N,C}(\bar{X}_k^N(t) - \mathbb{E}[\bar{X}_k^N(t)])| > \frac{\beta}{(e^{v_a T} T)^2} \right] =: p'(N).$$

By the same arguments as in Lemma 5.3, one obtains (again assuming $v_a T < 1$ without loss of generality)

$$p'(N) \leq 2^{\gamma(N)} \exp\left(-\frac{\beta\lambda(N)(1 - v_a T)}{q_1(N)(e^{v_a T} T)^2} \right) \exp\left(T\gamma(N) \Psi_0\left(\frac{\lambda(N)}{\gamma(N)} \sup_{l,m \in \mathbb{N}} |\rho_{lm}^{N,C}| \right) \right).$$

We choose $\lambda(N) := \gamma(N)$ this time. Then we can make $\log(p'(N))/\gamma(N)$ arbitrarily small uniformly for large N by increasing the value of β.

For the second step of the proof of (5.21), we derive from (5.22) that

$$\mathbb{P}[\hat{Y}^{N,1} \in H_N(\beta K), d(\hat{Y}^{N,1}, H_n(\beta K)) > \epsilon]$$

$$\leq \mathbb{P}\left[\sup_{\kappa=0,\ldots,[\gamma(n)T]-1} \sup_{\lambda \in \left[1, \frac{\gamma(N)}{\gamma(n)}+1\right)} \sup_{i=1,\ldots,d} \left|\hat{Y}_i^{N,1}\left(\frac{[\frac{\gamma(N)\kappa}{\gamma(n)}]+\lambda}{\gamma(N)} \wedge T\right)\right.\right.$$

$$\left.\left.-\hat{Y}_i^{N,1}\left(\frac{[\frac{\gamma(N)\kappa}{\gamma(n)}]}{\gamma(N)}\right)\right| > \epsilon\right]. \tag{5.24}$$

For the further procedure, we split the difference in the last line into two terms in the same way as in (5.17). We only treat the first corresponding term; the other term can be estimated similarly. Introducing the notation $t_{\kappa,\lambda}^{N,n}$ (resp. $t_\kappa^{N,n}$) for the time point within the first (resp. second) parentheses of (5.24), and observing that $0 \leq \widehat{t_{\kappa,\lambda}^{N,n}} - t_\kappa^{N,n} \leq 2/\gamma(N)+1/\gamma(n)$, we obtain

$$\mathbb{P}\left[\sup_{\kappa=0,\ldots,[\gamma(n)T]-1} \sup_{\lambda \in \left[1, \frac{\gamma(N)}{\gamma(n)}+1\right)} \sup_{i=1,\ldots,d} \left|\left(\left(e^{a^N \widehat{t_{\kappa,\lambda}^{N,n}}} - e^{a^N t_\kappa^{N,n}}\right)\right.\right.\right.$$

$$\left.\left.\left.\times \int_0^t \int_0^s e^{-a^N s} a^{N,\mathrm{P}} e^{\mathrm{diag}(a^{N,\mathrm{C}})(s-r)} (a^{N,\mathrm{C}})^\times (\bar{X}^N(r) - \mathbb{E}[\bar{X}^N(r)]) \, dr \, ds\right)_i\right| > \epsilon\right]$$

$$\leq \mathbb{P}\left[\sup_{t \in [0,T]} \sup_{i=1,\ldots,d} \left|\sum_{j=1}^\infty \sum_{k \neq j} a_{ij}^{N,\mathrm{P}} a_{jk}^{N,\mathrm{C}} (\bar{X}_k^N(t) - \mathbb{E}[\bar{X}_k^N(t)])\right|\right.$$

$$\left.> \frac{\epsilon}{v_a(e^{v_a T})^3 T^2 \left(\frac{2}{\gamma(N)} + \frac{1}{\gamma(n)}\right)}\right]$$

$$\leq 2^{\gamma(N)} \exp\left(-\frac{\epsilon \lambda(N)(1 - v_a T)}{q_1(N) v_a(e^{v_a T})^3 T^2 \left(\frac{2}{\gamma(N)} + \frac{1}{\gamma(n)}\right)}\right)$$

$$\times \exp\left(T\gamma(N) \Psi_0\left(\frac{\lambda(N)}{\gamma(N)} \sup_{l,m \in \mathbb{N}} |\rho_{lm}^{N,\mathrm{C}}|\right)\right),$$

where the last line follows in a similar fashion as before. With $\lambda(N) := \gamma(N)$, we can make, by taking n large enough, the logarithm of the last term divided by $\gamma(N)$ arbitrarily small for $N \geq n$. This finishes the proof for $\iota = 1$. The case $\iota = 2$ is analogous, while for $\iota = 3$ the line of argument remains the same in principle, with slight changes to account for the discretization of Lévy processes, cf. the proofs of Lemma 5.3 and Lemma 4.1 of [1]. □

Lemma 5.5. *The process* $(\hat{Y}_i^N : i = 1, \ldots, d)$ *satisfies a large deviation principle in* $(D_T^d, \mathcal{D}_T^d, U)$ *with a good convex rate function* $I_d : D_T^d \to [0, \infty]$ *such that* $I_d(x) < \infty$ *implies* $x \in AC_T^d$. *Moreover, we have* $I_d(0) = 0$ *and this minimum is unique.*

Proof. We apply the abstract Gärtner–Ellis theorem of [1], Theorems 2.1 and 2.4, to $\hat{Y}^N$ and prove the following steps.

(1) The laws of $\hat{Y}^N$, $N \in \mathbb{N}$, are exponentially tight in $(D_T^d, \mathcal{D}_T^d, U)$.

(2) For all $\theta \in M_T^d$, the limit $\Lambda(\theta) = \lim_{N \to \infty} \gamma(N)^{-1} \Lambda_N(\gamma(N)\theta)$ exists, where

$$\Lambda_N(\theta) := \log \mathbb{E}\left[\exp\left(\sum_{i=1}^d \int_0^T \hat{Y}_i^N(t)\,\theta_i(\mathrm{d}t) \right) \right].$$

(3) The mapping Λ is C_T^d-Gâteaux differentiable, in the sense that for all $\theta \in M_T^d$, there exists $x^\theta \in C_T^d$ such that for all $\eta \in M_T^d$,

$$\delta\Lambda(\theta; \eta) := \lim_{\epsilon \to 0} \frac{\Lambda(\theta + \epsilon\eta) - \Lambda(\theta)}{\epsilon} = \sum_{i=1}^d \int_0^T x_i^\theta(t)\,\eta_i(\mathrm{d}t). \tag{5.25}$$

Part of the claim is that the limit in (5.25) exists. Moreover, we have $\Lambda(0; \eta) = 0$ for all $\eta \in M_T^d$.

(4) We have $\{x \in D_T^d : \Lambda^*(x) < \infty\} \subseteq AC_T^d$, where

$$\Lambda^*(x) := \sup_{\theta \in M_T^d} \left(\sum_{i=1}^d \int_0^T x_i(t)\,\theta_i(\mathrm{d}t) - \Lambda(\theta) \right), \quad x \in D_T^d.$$

(5) For every $\alpha \in \mathbb{R}_+$, the set $\{x \in D_T^d : \Lambda^*(x) \le \alpha\}$ is compact in $(D_T^d, \mathcal{D}_T^d, U)$.

Part of the Gärtner–Ellis theorem is that the rate function I_d is given by Λ^*, the convex conjugate or Fenchel–Legendre transform of Λ. Since Λ is a convex function in θ satisfying (3), the conjugate Λ^{**} of Λ^* is again Λ, see Theorem 12 of [35]. Thus, by the first corollary to Theorem 1 in [3], we have $I_d(0) = \Lambda^*(0) = 0$ and this minimum is unique.

Let us now prove (1)–(5) above. Part (1) has been proved in Lemma 5.4. For (2) we first compute Λ. For all $\theta \in M_T^d$, we have (recall that $\hat{t} := [\gamma(N)t]/[\gamma(N)]$, and notice that we do not use (5.16), but rather (5.15) and the fact that an explicit formula for $\bar{X}^N - \mathbb{E}[\bar{X}^N]$ can be obtained by solving (5.18))

$$\begin{aligned}
\Lambda_N(\gamma(N)\theta) &= \log \mathbb{E}\Bigg[\exp\Bigg(\sum_{i=1}^d \int_0^T \Bigg(\int_0^{\hat{t}} \int_0^s \gamma(N) \sum_{j,k,l,m=1}^\infty e_{ij}^{a^N(\hat{t}-s)} a_{jk}^{N,P} \\
&\quad \times e_{kl}^{a^{N,C}(s-r)} \rho_{lm}^{N,C}\,\mathrm{d}M_m(r)\,\mathrm{d}s \\
&\quad + \gamma(N) \sum_{j,k=1}^\infty \int_0^{\hat{t}} e_{ij}^{a^N(\hat{t}-s)} \rho_{jk}^{N,P}\,\mathrm{d}M_k(s) \Bigg) \theta_i(\mathrm{d}t) \Bigg) \Bigg] \\
&= \sum_{m=1}^{\gamma(N)} \log \mathbb{E}\Bigg[\exp\Bigg(\sum_{i=1}^d \int_0^T \Bigg(\int_0^{\hat{t}} \int_0^s \sum_{j,k,l=1}^\infty \gamma(N) e_{ij}^{a^N(\hat{t}-s)} a_{jk}^{N,P} \\
&\quad \times e_{kl}^{a^{N,C}(s-r)} \rho_{lm}^{N,C}\,\mathrm{d}M_m(r)\,\mathrm{d}s \\
&\quad + \int_0^{\hat{t}} \gamma(N) \sum_{j=1}^\infty e_{ij}^{a^N(\hat{t}-s)} \rho_{jm}^{N,P}\,\mathrm{d}M_m(s) \Bigg) \theta_i(\mathrm{d}t) \Bigg) \Bigg]
\end{aligned} \tag{5.26}$$

by the independence of the processes M_m. By a stochastic Fubini argument (see Theorem IV.65

of [33]), the term within the exponential in the previous line can also be written as ($\check{s}$ denotes the smallest multiple of $\gamma(N)$ that is larger or equal to s)

$$
\int_0^{\hat{T}} \left(\int_r^{\hat{T}} \int_{\check{s}}^T \sum_{i=1}^d G_{im}^N(\hat{t} - s, s - r)\,\theta_i(\mathrm{d}t)\,\mathrm{d}s + \int_{\check{r}}^T \sum_{i=1}^d R_{im}^N(\hat{t} - r)\,\theta_i(\mathrm{d}t) \right) M_m(\mathrm{d}r),
$$

(5.27)

and has an infinitely divisible distribution such that its logarithmic Laplace exponent in (5.26) is explicitly known, see [34], Theorem 2.7. It is given by $\int_0^{\hat{T}} \Psi_m(H_m^N(\theta, r))\,\mathrm{d}r$, where $H_m^N(\theta, r)$ denotes the term within parentheses in (5.27). We claim that this term converges uniformly in m to $\int_0^T \Psi_m(H_m(\theta, r))\,\mathrm{d}r$. Indeed, by the dominating property of M_0, the claim follows as soon as we can prove that $H_m^N(\theta, r) \to H_m(\theta, r)$ as $N \to \infty$, uniformly in $m \in \{1, \ldots, \gamma(N) \vee \Gamma(N)\}$ and $r \in [0, T]$. This in turn follows from

$$
|H_m(\theta, r) - H_m^N(\theta, r)|
$$

$$
\leq \left| \int_{\hat{T}}^T \int_s^T \sum_{i=1}^d G_{im}(t - s, s - r)\,\theta_i(\mathrm{d}t)\,\mathrm{d}s \right| + \left| \int_r^{\hat{T}} \int_s^{\check{s}} \sum_{i=1}^d G_{im}(t - s, s - r)\,\theta_i(\mathrm{d}t)\,\mathrm{d}s \right|
$$

$$
+ \left| \int_r^{\hat{T}} \int_{\check{s}}^T \sum_{i=1}^d (G_{im}(t - s, s - r) - G_{im}^N(t - s, s - r))\,\theta_i(\mathrm{d}t)\,\mathrm{d}s \right|
$$

$$
+ \left| \int_r^{\hat{T}} \int_{\check{s}}^T \sum_{i=1}^d (G_{im}^N(t - s, s - r) - G_{im}^N(\hat{t} - s, s - r))\,\theta_i(\mathrm{d}t)\,\mathrm{d}s \right|
$$

$$
+ \left| \int_r^{\check{r}} \sum_{i=1}^d R_{im}(t - r)\,\theta_i(\mathrm{d}t) \right|
$$

$$
+ \left| \int_{\check{r}}^T \sum_{i=1}^d (R_{im}(t - r) - R_{im}^N(t - r))\,\theta_i(\mathrm{d}t) \right|
$$

$$
+ \left| \int_{\check{r}}^T \sum_{i=1}^d (R_{im}^N(t - r) - R_{im}^N(\hat{t} - r))\,\theta_i(\mathrm{d}t) \right|
$$

$$
\leq d \sup_{i,m \in \mathbb{N}} \sup_{s,t \in [0,T]} |G_{im}(t, s)| \left(\gamma(N)^{-1} \sup_{i=1,\ldots,d} |\theta_i|([0, T]) + \sup_{i=1,\ldots,d} \int_r^{\hat{T}} |\theta_i|([s, \check{s}))\,\mathrm{d}s \right)
$$

$$
+ d \sup_{i=1,\ldots,d} |\theta_i|([0, T]) \left(T \sup_{i,m=1,\ldots,\gamma(N)\vee\Gamma(N)} \sup_{s,t \in [0,T]} |G_{im}(t, s) - G_{im}^N(t, s)| \right.
$$

$$
+ \frac{v_a T}{\gamma(N)} \sup_{N,i,m \in \mathbb{N}} \sup_{s,t \in [0,T]} |G_{im}^N(t, s)| + \gamma(N)^{-1} \sup_{i,j \in \mathbb{N}} \sup_{t \in [0,T]} |R_{ij}(t)|
$$

$$
+ \sup_{i,j=1,\ldots,\gamma(N)\vee\Gamma(N)} \sup_{t \in [0,T]} |R_{ij}(t) - R_{ij}^N(t)| + \frac{v_a}{\gamma(N)} \sup_{N,i,j \in \mathbb{N}} \sup_{t \in [0,T]} |R_{ij}^N(t)| \right),
$$

where all terms converge to 0 by hypothesis independently of m and r (for the second summand one has to notice that the integral term equals $\int_r^{\hat{T}} \int_{\hat{t}}^t 1\,\mathrm{d}s\,|\theta_i|(\mathrm{d}t)$ and thus converges to 0 uniformly in i and r with rate $1/\gamma(N)$). Since the value of Cesàro sums remains unchanged

under uniform approximations, it follows from assumption (A8) that

$$\Lambda(\theta) = \lim_{N \to \infty} \frac{1}{\gamma(N)} \sum_{m=1}^{\gamma(N)} \int_0^T \Psi_m(H_m^N(\theta, r)) \, dr = \lim_{N \to \infty} \frac{1}{\gamma(N)} \sum_{m=1}^{\gamma(N)} \int_0^T \Psi_m(H_m(\theta, r)) \, dr.$$

Next, we prove the C_T^d-Gâteaux differentiability of Λ. First, regarding the existence of $\delta\Lambda(\theta; \eta)$ in (5.25), we note that the mappings $M_T^d \to L^\infty(\mathbb{N} \times [0, T])$, $\theta \mapsto (H_m(\theta, r): m \in \mathbb{N}, \ r \in [0, T])$, are continuous linear operators and therefore Fréchet differentiable, which is stronger than Gâteaux differentiability. Together with the fact that Ψ_m is differentiable with locally bounded derivative, and $c_m \leq c_0$ and $\nu_m \leq \nu_0$ for all $m \in \mathbb{N}$, this implies that for every θ and η,

$$\epsilon^{-1} \int_0^T \left(\Psi_m(H_m(\theta + \epsilon\eta, r)) - \Psi_m(H_m(\theta, r)) \right) \, dr$$

converges uniformly in $m \in \mathbb{N}$ as $\epsilon \to 0$. This in turn proves the Gâteaux differentiability of Λ. Moreover, it enables us to compute the derivative explicitly. Using the chain rule for Fréchet derivatives, we obtain

$$\delta\Lambda(\theta; \eta) = \lim_{N \to \infty} \frac{1}{\gamma(N)} \sum_{m=1}^{\gamma(N)} \int_0^T \lim_{\epsilon \to 0} \frac{\Psi_m(H_m(\theta + \epsilon\eta, r)) - \Psi_m(H_m(\theta, r))}{\epsilon} \, dr$$

$$= \lim_{N \to \infty} \frac{1}{\gamma(N)} \sum_{m=1}^{\gamma(N)} \int_0^T \Psi_m'(H_m(\theta, r)) H_m(\eta, r) \, dr$$

$$= \lim_{N \to \infty} \frac{1}{\gamma(N)} \sum_{m=1}^{\gamma(N)} \int_0^T \Psi_m'(H_m(\theta, r)) \left(\int_r^T \int_s^T \sum_{i=1}^d G_{im}(t - s, s - r) \, \eta_i(dt) \, ds \right.$$

$$\left. + \int_r^T \sum_{i=1}^d R_{im}(t - r) \, \eta_i(dt) \right) dr$$

$$= \lim_{N \to \infty} \frac{1}{\gamma(N)} \sum_{m=1}^{\gamma(N)} \sum_{i=1}^d \int_0^T \left(\int_0^t \int_0^s \Psi_m'(H_m(\theta, r)) G_{im}(t - s, s - r) \, dr \, ds \right.$$

$$\left. + \int_0^t R_{im}(t - r) \Psi_m'(H_m(\theta, r)) \, dr \right) \eta_i(dt)$$

$$= \sum_{i=1}^d \int_0^T \int_0^t \lim_{N \to \infty} \frac{1}{\gamma(N)} \sum_{m=1}^{\gamma(N)} \left(\int_0^s G_{im}(t - s, s - r) \Psi_m'(H_m(\theta, r)) \, dr \right.$$

$$\left. + R_{im}(t - s) \Psi_m'(H_m(\theta, s)) \right) ds \, \eta_i(dt),$$

where all interchanges of integration, summation and taking limits are justified by dominated convergence (observe that $|\Psi_m'(u)| \leq \Psi_0'(|u|)$ by hypothesis). From the last equality, we deduce the existence of $x^\theta \in C_T^d$ satisfying (5.25). Since $H_m(0, r) = 0$ and $\Psi_m'(0) = 0$, we have $\delta\Lambda(0; \eta) = 0$ identically.

Next, we demonstrate (4), namely that Λ^* only assumes finite values on the set AC_T^d, that is, $\Lambda^*(x) < \infty$ implies that for every $\epsilon \in (0, \infty)$, there exists $\delta \in (0, \infty)$ such that $\sum_{i=1}^d \sum_{j=1}^n |x_i(b_j) - x_i(a_j)| < \epsilon$ whenever $n \in \mathbb{N}$, $0 \leq a_1 < b_1 \leq \cdots \leq a_n < b_n \leq T$ and $\sum_{j=1}^n (b_j - a_j) < \delta$. In order to do so, we follow the strategy of proof in [1], Theorem 3.1.

We consider $\theta_i := \sum_{j=1}^{n} \xi_i^j (\delta_{b_j} - \delta_{a_j})$ where $\xi_i^j \in \mathbb{R}^d$ is arbitrary. Then we evidently have $\theta_i((r, T]) = \sum_{j=1}^{n} \xi_i^j \mathbb{1}_{[a_j, b_j)}(r)$. Denoting $C_T := T \sup_{s,t \in [0,T]} \sup_{i,m \in \mathbb{N}} |G_{im}(t, s)| + \sup_{t \in [0,T]} \sup_{i,m \in \mathbb{N}} |R_{im}(t)|$, we obtain

$$
\begin{aligned}
\Lambda(\theta) &\leq \sup_{m \in \mathbb{N}} \int_0^T \Psi_m(H_m(\theta, r)) \, \mathrm{d}r \leq \int_0^T \Psi_0 \left(C_T \sum_{i=1}^{d} \theta_i((r, T]) \right) \mathrm{d}r \\
&= \int_0^T \sum_{j=1}^{n} \Psi_0 \left(C_T \sum_{i=1}^{d} \xi_i^j \right) \mathbb{1}_{[a_j, b_j)}(r) \, \mathrm{d}r \leq \sup_{j=1,\ldots,n} \Psi_0 \left(C_T \sum_{i=1}^{d} |\xi_i^j| \right) \sum_{j=1}^{n} (b_j - a_j) \\
&= \sup_{j=1,\ldots,n} \Psi_0 \left(C_T \|\xi^j\|_1 \right) \sum_{j=1}^{n} (b_j - a_j),
\end{aligned}
$$

where $\|\xi^j\|_1 := \sum_{i=1}^{d} |\xi_i^j|$. As a consequence, we deduce from the definition of Λ^* that for all $\tau \in (0, \infty)$ and $\|\xi^j\|_1 \leq \tau$,

$$
\sum_{i=1}^{d} \sum_{j=1}^{n} \xi_i^j (x_i(b_j) - x_i(a_j)) \leq \Psi_0 (C_T \tau) \sum_{j=1}^{n} (b_j - a_j) + \Lambda^*(x).
$$

Taking ξ_i^j as τ times the sign of $x_i(b_j) - x_i(a_j)$, it follows that

$$
\sum_{i=1}^{d} \sum_{j=1}^{n} |x_i(b_j) - x_i(a_j)| \leq \tau^{-1} \Psi_0 (C_T \tau) \sum_{j=1}^{n} (b_j - a_j) + \tau^{-1} \Lambda^*(x). \tag{5.28}
$$

If $\Lambda^*(x) < \infty$, we can now choose τ first and then δ to make the left-hand side arbitrarily small.

It only remains to prove (5), the compactness of the level sets of Λ^*. By step (4) and the lower semicontinuity of Λ^*, its level sets are closed subsets of AC_T^d. Thus, the Arzelà–Ascoli theorem provides a compactness criterion. First, observe that for all $t \in [0, T]$, we have for $x \in AC_T^d$ with $\Lambda^*(x) \leq \alpha$ that

$$
\sum_{i=1}^{d} |x_i(t)| = \sup_{\theta \in \Theta_t} \sum_{i=1}^{d} \int_0^T x_i(t) \, \theta_i(\mathrm{d}t) \leq \alpha + \sup_{\theta \in \Theta_t} \Lambda(\theta) < \infty,
$$

where Θ_t is the finite collection of all θ for which each coordinate is either δ_t or $-\delta_t$. Second, for the proof of the uniform equicontinuity of the functions $x \in AC_T^d$ with $\Lambda^*(x) \leq \alpha$, we recall from (5.28) that

$$
\sum_{i=1}^{d} |x_i(t) - x_i(s)| \leq \tau^{-1} \Psi_0 (C_T \tau) (t - s) + \tau^{-1} \alpha,
$$

which converges to 0 independently of x when $s \uparrow t$ and $\tau \to \infty$. $\square$

Acknowledgment

References

[1] A. de Acosta, Large deviations for vector-valued Lévy processes, Stoch. Process. Appl. 51 (1) (1994) 75–115.

[2] S.M. Ahn, S.-Y. Ha, Stochastic flocking dynamics of the Cucker-Smale model with multiplicative white noises, J. Math. Phys. 51 (10) (2010) 103301.

[3] E. Asplund, R.T. Rockafellar, Gradient of convex functions, Trans. Amer. Math. Soc. 139 (1969) 443–467.

[4] J. Baladron, D. Fasoli, O. Faugeras, J. Touboul, Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons, J. Math. Neurosci. 2 (10) (2012) 50.

[5] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.

[6] S. Battiston, D. Delli Gatti, M. Gallegati, B. Greenwald, J.E. Stiglitz, Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk, J. Econom. Dynam. Control 36 (8) (2012) 1121–1141.

[7] F. Bolley, J.A. Cañizo, J.A. Carrillo, Stochastic mean-field limit: non-Lipschitz forces and swarming, Math. Model. Methods Appl. Sci. 21 (11) (2011) 2179–2210.

[8] B. Bollobás, C. Borgs, J. Chayes, O. Riordan, Directed scale-free graphs, in: Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003), ACM, New York, 2003, pp. 132–139.

[9] J.-P. Bouchaud, M. Mézard, Wealth condensation in a simple model of economy, Phys. A 282 (3–4) (2000) 536–545.

[10] R. Buckdahn, J. Li, S. Peng, Nonlinear stochastic differential games involving a major player and a large number of collectively acting minor agents, SIAM J. Control Optim. 52 (1) (2014) 451–492.

[11] A. Budhiraja, R. Wu, Some fluctuation results for weakly interacting multi-type particle systems, Stoch. Process. Appl. 126 (8) (2016) 2253–2296.

[12] R. Carmona, X. Zhu, A probabilistic approach to mean field games with major and minor players, Ann. Appl. Probab. 26 (3) (2016) 1535–1580.

[13] D.A. Dawson, J. Gärtner, Large deviations from the McKean-Vlasov limit for weakly interacting diffusions, Stochastics 20 (4) (1987) 247–308.

[14] P. Degond, J.-G. Liu, C. Ringhofer, Evolution of the distribution of wealth in an economic environment driven by local Nash equilibria, J. Stat. Phys. 154 (3) (2014) 751–780.

[15] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, second ed., Springer, New York, 2010.

[16] W. Finnoff, Law of large numbers for a general system of stochastic differential equations with global interaction, Stoch. Process. Appl. 46 (1) (1993) 153–182.

[17] W. Finnoff, Law of Large Numbers for a Heterogeneous System of Stochastic Differential Equations with Strong Local Interaction and Economic Applications, Ann. Appl. Probab. 4 (2) (1994) 494–528.

[18] J.-P. Fouque, L.-H. Sun, Systemic risk illustrated, in: J.-P. Fouque, J.A. Langsam (Eds.), Handbook on Systemic Risk, Cambridge University Press, Cambridge, 2013, pp. 444–452.

[19] N. Fournier, E. Löcherbach, On a toy model of interacting neurons, Ann. Inst. Henri Poincaré Probab. Stat. 52 (4) (2016) 1844–1876.

[20] J. Gärtner, On the McKean-Vlasov limit for interacting diffusions, Math. Nachr. 137 (1) (1988) 197–248.

[21] K. Giesecke, K. Spiliopoulos, R.B. Sowers, J. Sirignano, Large portfolio asymptotics for loss from default, Math. Finance 25 (1) (2015) 77–114.

[22] T. Ichinomiya, Bouchard-Mézard model on a random network, Phys. Rev. E 86 (3) (2012) 036111.

[23] O. Kley, C. Klüppelberg, L. Reichel, Systemic risk through contagion in a core–periphery structured banking network, in: A. Palczewski, Ł. Stettner (Eds.), Advances in Mathematics of Finance, Banach Center Publications, Warschau, 2015, pp. 133–149.

[24] C. Léonard, Some epidemic systems are long range interacting particle systems, in: J.-P. Gabriel, C. Lefèvre, P. Picard (Eds.), Stochastic Processes in Epidemic Theory, Springer, Berlin, 1990, pp. 170–183.

[25] C. Léonard, Large deviations for long range interacting particle systems with jumps, Ann. Inst. Henri Poincaré Probab. Stat. 31 (2) (1995) 289–323.

[26] H.P. McKean Jr., A class of Markov processes associated with nonlinear parabolic equations, Proc. Natl. Acad. Sci. 56 (6) (1966) 1907–1911.

[27] H.P. McKean Jr., Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas, Arch. Ration. Mech. Anal. 21 (5) (1966) 343–367.

[28] H.P. McKean Jr., Propagation of chaos for a class of non-linear parabolic equations, in: Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic University, 1967), Air Force Office of Scientific Research, Arlington, 1967, pp. 41–57.

[29] P. Del Moral, L. Kallel, J. Rowe, Modelling genetics algorithms with interacting particle systems, Rev. Mat. Teor. Apl. 8 (2) (2001) 19–77.

[30] T.F. Móri, The maximum degree of the Barabási-Albert random tree, Combin. Probab. Comput. 14 (3) (2005) 339–348.

[31] M. Nagasawa, H. Tanaka, Diffusion with interactions and collisions between coloured particles and the propagation of chaos, Probab. Theory Related Fields 74 (2) (1987) 161–198.

[32] M. Nagasawa, H. Tanaka, On the propagation of chaos for diffusion processes with drift coefficients not of average form, Tokyo J. Math. 10 (2) (1987) 403–418.

[33] P.E. Protter, Stochastic Integration and Differential Equations, second ed., Springer, Berlin, 2005.

[34] B.S. Rajput, J. Rosiński, Spectral representations of infinitely divisible processes, Probab. Theory Related Fields 82 (3) (1989) 451–487.

[35] R.T. Rockafellar, Conjugate Duality and Optimization, SIAM, Philadelphia, 1974.

[36] A.-S. Sznitman, Topics in propagation of chaos, in: P.-L. Hennequin (Ed.), École d'Été de Probabilités de Saint Flour XIX - 1989, in: Lecture Notes in Mathematics, vol. 1464, Springer, Berlin, 1991, pp. 165–251.

[37] M. Talagrand, Mean Field Models for Spin Glasses. Volume I: Basic Examples, Springer, Berlin, 2011.

[38] M. Talagrand, Mean Field Models for Spin Glasses. Volume II: Advanced Replica-Symmetry and Low Temperature, Springer, Berlin, 2011.

[39] J. Touboul, Propagation of chaos in neural fields, Ann. Appl. Probab. 24 (3) (2014) 1298–1328.

[40] J. Vaillancourt, On the existence of random McKean-Vlasov limits for triangular arrays of exchangeable diffusions, Stoch. Anal. Appl. 6 (4) (1988) 431–446.